Common gene expression signatures in Parkinson’s disease are driven by changes in cell composition

Gonzalo S. Nido1,2, Fiona Dick1,2, Lilah Toker1,2, Kjell Petersen1,3, Guido Alves4,5, Ole-Bjørn Tysnes1,2, Inge Jonassen1,3, Kristoffer Haugarvoll1,2 and Charalampos Tzoulis1,2*

Abstract

The etiology of Parkinson’s disease is largely unknown. Genome-wide transcriptomic studies in bulk brain tissue have identified several molecular signatures associated with the disease. While these studies have the potential to shed light into the pathogenesis of Parkinson’s disease, they are also limited by two major confounders: RNA post-mortem degradation and heterogeneous cell type composition of bulk tissue samples. We performed RNA sequencing following ribosomal RNA depletion in the prefrontal cortex of 49 individuals from two independent case-control cohorts. Using cell type specific markers, we estimated the cell type composition for each sample and included this in our analysis models to compensate for the variation in cell type proportions. Ribosomal RNA depletion followed by capture by random primers resulted in substantially more even transcript coverage, compared to poly(A) capture, in post-mortem tissue. Moreover, we show that cell type composition is a major confounder of differential gene expression analysis in the Parkinson’s disease brain. Accounting for cell type proportions attenuated numerous transcriptomic signatures that have been previously associated with Parkinson’s disease, including vesicle trafficking, synaptic transmission, immune and mitochondrial function. Conversely, pathways related to endoplasmic reticulum, lipid oxidation and unfolded protein response were strengthened and surface as the top differential gene expression signatures in the Parkinson’s disease prefrontal cortex. Our results indicate that differential gene expression signatures in Parkinson’s disease bulk brain tissue are significantly confounded by underlying differences in cell type composition. Modeling cell type heterogeneity is crucial in order to unveil transcriptomic signatures that represent regulatory changes in the Parkinson’s disease brain and are, therefore, more likely to be associated with underlying disease mechanisms.

Keywords: RNA sequencing, Neurodegeneration, Parkinsonism, Mitochondria, Gene expression
Introduction

Parkinson’s disease (PD) is the second most prevalent neurodegenerative disorder, affecting ~1.8% of the population above 65 years [45]. PD is a complex disorder caused by a combination of genetic and environmental factors, but the molecular mechanisms underlying its etiology remain largely unaccounted for. Genome-wide transcriptomic studies can identify expression signatures associated with PD. While not able to establish causality, these studies hold the potential to highlight important biological mechanisms, some of which may be explored as targets for therapeutic modulation.

A recent systematic review identified 33 original genome-wide transcriptomic studies in the PD brain, of which 5 were performed on laser microdissected neurons from the substantia nigra pars compacta (SNc) and the remaining in bulk tissue from various brain regions [8]. These studies show surprisingly low replicability at the level of individual genes, however, and only partial concordance for pathways. The most consistent alterations have been found in pathways related to energy metabolism/mitochondrial function and protein degradation, followed by synaptic transmission, vesicle trafficking, lysosome/autophagy and neuroinflammation [8]. While these processes commonly show differential expression signatures in PD, it remains unknown whether this is because they truly reflect the biology of PD or due to systematic bias and confounding factors. Two major sources of bias for transcriptomic studies in the human brain are the post-mortem degradation of RNA and the highly heterogeneous cell type composition of bulk tissue samples.

RNA degradation of variable extent occurs in post-mortem tissue. To further complicate the picture, it has been shown that different cell types exhibit different degrees of susceptibility to RNA degradation [32], potentially confounding differences in cellular composition with differences in RNA quality. Access to high-quality brain tissue is generally limited, and thus an optimal choice of experimental platforms becomes paramount to maximize sensitivity. While RNA microarrays are being gradually superseded by RNA-seq technology, only 3 out of the 33 studies identified by an up-to-date review [8] used RNA-seq, and all of them employed poly(A) capture, a widely used protocol (in both RNA-seq and microarray analyses) to restrict the analysis to mature mRNA [20, 30, 46]. However, this library preparation method only picks up RNA fragments with a poly-A tail, introducing substantial bias in low quality RNA samples [1, 25, 47, 56]. A well-established approach to mitigate this limitation is whole RNA-seq following active ribosomal RNA (rRNA) depletion and capture by random primers, such as the Illumina Ribo-Zero technique [31]. To our knowledge there are no genome-wide transcriptomic studies on PD brain employing active rRNA depletion methods to date.

Systematic differences in sample cell composition represent another important confounding factor. These typically originate from two sources: biological differences (e.g. secondary to neurodegeneration) and technical variation in sample dissection and preparation. Brain areas affected by neurodegeneration are characterized by neuronal loss and gliosis, resulting in a systematically increased glia-to-neurons ratio in patients. This confounder is strongest in areas with severe changes, such as the SNc, but is also present to a variable degree in less affected areas, such as the neocortex. In addition, technical sources of variation due to sampling may affect any brain region and cause an uneven distribution of gray and white matter, resulting in a variable fraction of oligodendrocytes. Thus, transcriptional signatures associated with PD in bulk brain tissue may reflect changes in cellular composition rather than disease-specific transcriptional modulation. This observation has already been put forward using neurodegenerative mouse models and re-analysis of human brain transcriptomic data [50]. Heterogeneous cell composition is, hence, a major confounder that needs to be considered and appropriately addressed in transcriptomic studies in bulk brain samples.

We report the first genome-wide transcriptomic study in the PD brain employing RNA-seq following rRNA depletion and random primer capture. We show that this approach is able to substantially mitigate the bias of post-mortem degradation, resulting in substantially better transcript coverage compared to poly(A) capture. Moreover, by estimating the relative cell type proportion in our samples, we confirm that cellular composition is a major source of variation in bulk tissue data, confounding the differential gene expression profile even in the less affected prefrontal cortex. By incorporating the estimated cell type proportions into our analysis models, we were able to unveil transcriptomic signatures which are more likely to be associated with the underlying disease mechanisms.

Material and methods

Subject cohorts

All experiments were conducted in fresh-frozen prefrontal cortex (Brodmann area 9) from a total of 49 individuals from two independent cohorts. The first cohort (n = 29) comprised individuals with idiopathic PD (n = 18) from the Park-West study (PW), a prospective population-based cohort which has been described in detail [2] and neurologically healthy controls (Ctrl, n = 11) from our brain bank for aging and neurodegeneration. Whole-exome sequencing had been performed on all patients [24] and known/predicted pathogenic mutations in genes implicated in Mendelian PD and other
monogenic neurological disorders had been excluded. None of the study participants had clinical signs of mitochondrial disease or use of medication known to influence mitochondrial function (Additional file 1). Controls had no known neurological disease and were matched for age and gender. The second cohort comprised samples from 21 individuals from the Netherlands Brain Bank (NBB) including idiopathic PD (n = 10) and demographically matched neurologically healthy controls (n = 11). Individuals with PD fulfilled the National Institute of Neurological Disorders and Stroke [26] and the UK Parkinson’s disease Society Brain Bank [54] diagnostic criteria for the disease at their final visit. Ethical permission for these studies was obtained from our regional ethics committee (REK 2017/2082, 2010/1700, 131.04).

To investigate the effect of the rRNA depletion and random primer capture protocol compared to the prevailing poly(A) method, we re-analyzed an RNA-seq dataset from a previous publication which employed a poly(A) tail selection kit on post-mortem tissue of the dataset from a previous publication which employed a prevailing poly(A) method, we re-analyzed an RNA-seq random primer capture protocol compared to the pre-ethics committee (REK 2017/2082, 2010/1700, 131.04).

**Tissue collection and neuropathology**

Brains were collected at autopsy and split sagittally along the corpus callosum. One hemisphere was fixed whole in formaldehyde and the other coronally sectioned and snap-frozen in liquid nitrogen. All samples were collected using a standard technique and fixation time of ~2 weeks. There was no significant difference in post-mortem interval (PMI) (independent t-test, PW cohort p = 0.16; NBB cohort p = 0.92), age (independent t-test, PW cohort p = 0.18; NBB cohort p = 0.074) or gender (independent t-test, PW cohort p = 0.94; NBB cohort p = 0.53) between PD subjects and controls. Subject demographics and tissue availability are provided in Additional file 1. Routine neuropathological examination including immunohistochemistry for α-synuclein, tau and beta amyloid was performed on all brains. All cases showed neuropathological changes consistent with PD including degeneration of the dopaminergic neurons of the SNC in the presence of Lewy pathology. Controls had no pathological evidence of neurodegeneration.

**RNA sequencing**

Total RNA was extracted from prefrontal cortex tissue homogenate for all samples using RNeasy plus mini kit (Qiagen) with on-column DNase treatment according to manufacturer’s protocol. Final elution was made in 65 μl of dH2O. The concentration and integrity of the total RNA was estimated by Ribogreen assay (Thermo Fisher Scientific), and Fragment Analyzer (Advanced Analytical), respectively and 500 ng of total RNA was used for downstream RNA-seq applications. First, rRNA was removed using Ribo-Zero™ Gold (Epidemiology) kit (Illumina, San Diego, CA) using manufacturer’s recommended protocol. Immediately after the rRNA removal the RNA was fragmented and primed for the first strand synthesis using the NEBNext First Strand synthesis module (New England Biolabs Inc., Ipswich, MA). Directional second strand synthesis was performed using NEBNext Ultra Directional second strand synthesis kit. Following this the samples were taken into standard library preparation protocol using NEBNext® DNA Library Prep Master Mix Set for Illumina® with slight modifications. Briefly, end-repair was done followed by poly(A) addition and custom adapter ligation. Post-ligated materials were individually barcoded with unique in-house Genomic Services Lab (GSL) primers and amplified through 12 cycles of PCR. Library quantity was assessed by Picogreen Assay (Thermo Fisher Scientific), and the library quality was estimated by utilizing a DNA High Sense chip on a Caliper Gx (Perkin Elmer). Accurate quantification of the final libraries for sequencing applications was determined using the qPCR-based KAPA Biosystems Library Quantification kit (Kapa Biosystems, Inc.). Each library was diluted to a final concentration of 12.5 nM and pooled equimolar prior to clustering. One hundred twenty-five bp Paired-End (PE) sequencing was performed on an Illumina HiSeq2500 sequencer (Illumina, Inc.). RNA quality, as measured by the RNA integrity number (RIN), varied across samples (mean = 5.3, range = 3.0–7.2 for PW; mean = 6.8, range = 3.2–9.1 for NBB), although the difference between conditions did not reach statistical significance in any of the cohorts (t-test P = 0.72 and 0.90 for PW and NBB cohorts, respectively).

**Data quality control**

FASTQ files were trimmed using Trimmomatic version 0.36 [7] to remove potential Illumina adapters and low quality bases with the following parameters: ILLUMINACLIP:truseq.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15. FASTQ files were assessed using fastQC version 0.11.5 [3] prior and following trimming. For an in-depth quality assessment, we mapped the trimmed reads using HISAT2 version 2.1.0 [34] against the hg19 human reference genome (using --rna-strandness RF option) preserving lane-specific information. To discard potential lane-specific sequencing batch effects we inspected the output of the CollectRnaSeqMetrics tool of Picard Tools version 2.6 [11]. Mapping efficiency and proportion of reads mapping to rRNA, intronic, intergenic and coding regions were obtained from the output of the CollectRnaSeqMetrics (Additional file 2: Figure S1 and S2).
For the poly(A) capture dataset [20], raw FASTQ files were obtained from the Gene Expression Omnibus (GEO:GSE68719) and analyzed exactly as described for our cohorts (with the exception of --rna-strandness in HISAT2, which was turned off to take into account that the cDNA library of this cohort was unstranded).

**RNA expression quantification and filtering**

We used Salmon version 0.9.1 [43] to quantify the abundance at the transcript level with the fragment-level GC bias correction option (--gcBias) and the appropriate option for the library type (--ISR) against the Ensembl release 75 transcriptome. Transcript-level quantification was collapsed onto gene-level quantification using the tximport R package version 1.8.0 [49] according to the gene definitions provided by the same Ensembl release. We filtered out genes in non-canonical chromosomes and scaffolds, and transcripts encoded by the mitochondrial genome. To further reduce the potential for artifacts we filtered out transcripts with unusually high expression by removing transcripts that gathered more than 1% of the reads on more than half of the samples, which resulted in the removal of 3 and 4 transcripts from the PW and NBB cohorts, respectively. Additionally, low-expressed (i.e. genes whose expression was below the median expression in at least 20% of the samples) were filtered out from downstream analyses. Samples were then marked as outliers if their median correlation in gene expression (log counts per million) was below Q1 – 1.5*IQR or above Q3 + 1.5*IQR (Tukey’s fences; Q1: first quartile, Q3: third quartile, IQR: inter-quartile range). As a result, 3 samples were marked as outliers in the PW cohort and 3 in the NBB cohort, and were not included in downstream analyses (resulting sample sizes: \( N_{PW} = 26 \), \( N_{NBB} = 18 \), Additional file 2: Figure S3).

**Estimation of marker gene profiles**

It has been previously shown that cell type-specific transcriptional signature patterns derived from bulk tissue samples (marker gene profiles, MGPs), can be used as surrogates for relative cell type abundance across samples [37]. MGPs for each cell type are calculated individually, by summarizing the concordant change in their respective marker genes via the first principal component of their expression (i.e. log-transformed counts per million (CPMs)). For the purpose of our study, we calculated MGPs for the main cortical cell types (neurons, oligodendroglia, microglia, endothelial cells, and astrocytes). Cortical cell type markers were obtained from the NeuroExpresso database [37], a comprehensive database compiled using mouse brain cell type expression datasets, and human orthologs were defined using HomoloGene [38]. To reduce the impact of outlier samples, principal component analysis was repeated 100 times on subsampled data, containing an equal number of subjects per group, and removing markers with opposite sign of the main trend. The median score for each sample was used as MGP for the downstream analyses. MGPs obtained with Neuroexpresso-based markers were highly correlated with MGPs calculated using two independent sets of markers from human brain single-cell transcriptomic studies [33, 53] (Additional file 2: Figures S4–8, Additional file 2: Table S1). To assess potential variations associated with the disease across the neuronal markers, we examined the overlap between the markers and the differentially expressed genes in four publicly available datasets of laser microdissected neurons from PD brain (SNc dopaminergic neurons [13, 22, 48] and posterior cingulate cortex pyramidal neurons [51]). We found minimal overlap (3/78 genes) between our neuronal markers and genes differentially expressed in PD dopaminergic neurons. Moreover, none of the markers were differentially expressed in PD cortical neurons [51] (Additional File 2: Figure S9). The vast majority of the cell type markers used for the calculation of MGPs changed in the same direction across our samples (Additional File 2: Figure S9), indicating that MGPs truly represent changes in global cell type-specific transcription profiles, rather than being driven by changes in specific genes.

To unravel potential complex interactions between MGPs and other experimental covariates, including disease status, we calculated the pairwise correlation between all the variables and also their association with the main axes of variation of gene expression. To assist us in choosing an optimal set of MGPs to include as covariates, we quantified the group differences in the cellular proportions between PD and controls using linear models adjusting for the known experimental covariates (i.e. RIN, PMI, sex, age, and sequencing batch). Significant association with disease status was found for oligodendrocyte MGP in the PW cohort and for microglia in the NBB cohort. Thus, these were included in the downstream analyses.

**Differential gene expression and functional enrichment analyses**

We performed differential gene expression analyses using the DESeq2 R package version 1.22.2 [35] with default parameters. Experimental covariates (sex, age, RIN, PMI, and sequencing batch) as well as oligodendrocyte and microglia MGPs were incorporated into the statistical model. Multiple hypothesis testing was performed with the default automatic filtering of DESeq2 followed by false discovery rate (FDR) calculation by the Benjamini-Hochberg procedure. Analyses were carried out independently for the two cohorts. Genes were
scored according to their significance by transforming the p-values to account for direction of change. For each gene, the up-regulated score was calculated as $S_{up} = \begin{cases} 1-p/2, & LFC < 0 \\ p/2, & LFC \geq 0 \end{cases}$, and the down-regulated score as $S_{down} = 1 - S_{up}$ where LFC corresponds to the log fold change and $p$ to the nominal p-value of the gene. Genes were then tested for enrichment using alternatively log($S_{up}$) and log($S_{down}$) scores employing the gene score resampling method implemented in the ermineR package version 1.0.1 [39], an R wrapper package for ermineJ [27] with the complete Gene Ontology (GO) database annotation [5] to obtain lists of up- and down-regulated pathways for each cohort.

In order to characterize the main biological processes affected by the cell type correction, we scored pathways based on the loss of significance caused by the addition of cellular estimates to the gene expression model. We quantified the difference in the level of significance in the up- and down-regulated enrichment results for each significant pathway as $\Delta = \log(p_0) - \log(p_{CT})$, where $p_{CT}$ and $p_0$ are the corrected enrichment p-values for the model with cell types (CT) and without (0), respectively. Only pathways that were significant in either one of the models were analyzed in this manner ($p_0 < 0.05$ or $p_{CT} < 0.05$).

The source code for the analyses is available in the GitLab repository (https://git.app.uib.no/neuromics/cell-composition-rna-pd) under the GPL public license v3.0.

**Results**

**Ribo-zero is superior to poly(A) selection in post-mortem brain**

We carried out RNA-seq using rRNA depletion and random primer capture (henceforth referred to as Ribo-Zero) in fresh-frozen prefrontal cortex (Brodmann area 9) from a total of 49 individuals from two independent cohorts: the Norwegian ParkWest study ($PW$, $n = 29$) [2] and the Netherlands Brain Bank ($NBB$, $n = 21$). Comparison of our data to a published poly(A) capture dataset of similar characteristics [20] (PA cohort) revealed important differences of mapping coverage. Mapping efficiency was slightly higher in the poly(A) dataset ($PA$: median $= 0.976$, range $= 0.971$–0.980) compared to the Ribo-Zero datasets ($PW$: median $= 0.952$, range $= 0.940$–0.962; $NBB$: median $= 0.959$, range $= 0.947$–0.965). The counts per million (CPM) of rRNA regions, as defined by Ensembl release 75, was very low in all samples ($PW$: median $= 3099$, range $= 1047$–7071; $NBB$: median $= 1583$, range $= 1129$–5024) and, as expected, significantly lower in the Ribo-Zero cohorts compared to the poly(A) dataset ($PA$: median $= 40,058$, range $= 10,701$–95,183) (Additional file 2: Figure S1).

In both datasets, the RIN was positively correlated with mapping efficiency to mRNA regions, but not to intergenic and/or intronic regions (Additional file 2: Figure S2). Despite having higher mean RIN values, the PA cohort showed a marked unevenness of transcript body coverage compared to the Ribo-Zero cohorts (Fig. 1a). The median coefficient of variation in coverage was significantly lower in the Ribo-Zero cohorts and the 5′- and 3′-ends of the transcripts showed substantially better coverage compared to the PA cohort (Fig. 1b). Moreover, in the Ribo-Zero datasets both the 3′- and 5′-end coverage loss showed a significant inverse correlation with the RIN values. In contrast, RIN showed no correlation with the 5′-bias and a positive correlation with the 3′-bias in the PA dataset (Fig. 1c). Thus, Ribo-Zero results in substantially better and more even coverage of the transcriptome in post-mortem brain tissue, providing a better alternative to poly(A) capture and minimizing the prospect of transcript quantification biases downstream.

**Cell composition is a major confounder of gene expression in bulk brain samples**

The observed gene expression profiles in bulk brain tissue can be dramatically influenced by differences in cellular composition. Such differences can be a result of variation in gray/white matter ratios introduced during tissue extraction, inter-subject variability or represent disease related alterations [14, 37, 52]. To study the contribution of various technical and biological sources of variation in our dataset we first estimated marker gene profiles (MGPs) for the major classes of cortical cell types (astrocytes, microglia, oligodendrocytes, endothelial cells and neurons) in our samples by summarizing the expression of the cell type-specific marker genes as previously described [37, 52]. Next, we examined the Pearson’s correlation between potential sources of biological variation in our data, including technical and demographic factors (RIN, PMI, sex, age, and disease status) and MGPs. MGPs for neuronal cell types were significantly anticorrelated with the other main cortical cell types in both cohorts ($p < 0.05$, Fig. 2a). In agreement with previous studies [6, 32], MGPs were also correlated with RNA quality. In both cohorts RIN was significantly correlated with neuronal (positive correlation) and astrocyte (negative correlation) MGPs. Significant negative correlation of RIN with microglia MGPs was observed in the NBB cohort (Fig. 2a). Most concerning was the detection of a significant association between the oligodendrocyte MGP and the disease status in the PW cohort (Fig. 2a). The main axis of variation in gene expression (which explained 44 and 45% of the total variance in PW and NBB, respectively) was significantly correlated with RNA quality and cellular composition in both cohorts (Fig. 2b), singling out RNA quality and
cellular composition as the main drivers of transcriptional change in bulk brain tissue.

We next looked for differences in cellular proportions between PD and controls adjusting for the known experimental covariates. In the NBB cohort, PD subjects exhibited a significant increase in the microglia MGP ($p = 0.015$, Wilcoxon test), while a significant increase in the oligodendrocyte MGP ($p = 5.5 \times 10^{-3}$, Wilcoxon test).
**Fig. 2** (See legend on next page.)

A

| Condition | PW | NBB |
|-----------|----|-----|
| Age       |    |     |
| Sex       |    |     |
| PMI       |    |     |
| RN        |    |     |
| Astrocytes| -1.73 | -1.65 |
| Endothelial | -5.62 | -5.67 |
| Microglia  | 0.43 | -0.40 |
| Oligodendrocytes | 0.48 | 0.7 |
| Neurons    | 0.68 | 0.51 |

B

| Condition | PW | NBB |
|-----------|----|-----|
| Age       | 2.7e-02 | 3.6e-06 |
| Sex       | 1.8e-02 | 2.0e-02 |
| PMI       | 1.4e-03 | 1.1e-04 |
| RN        | 1.1e-04 | 1.1e-04 |
| Astrocytes| 6.1e-06 | 9.3e-04 |
| Endothelial | 8.8e-03 | 2.3e-02 |
| Microglia  | 5.4e-02 | 4.4e-03 |
| Oligodendrocytes | 1.6e-09 | 1.2e-09 |
| Neurons    | 2.0e-07 | 3.5e-02 |

C

| Cells      | PW | NBB |
|-----------|----|-----|
| Astrocytes|    |     |
| Endothelial |    |     |
| Microglia  |    |     |
| Neurons    |    |     |
| Oligodendrocytes |    |     |

*p = 0.22, 0.14, 0.08, 0.59, 0.011, 0.056, 7.8e-10, 0.82*
was observed in PD subjects from the PW cohort. In both cohorts, these changes were accompanied by a non-significant decrease in neuronal MGPs (Fig. 2c).

MGPs of different cell types are not entirely independent from each other, since changes in one cell type can be accompanied by changes in other cell types. Thus, to ensure that neuronal, endothelial, and astrocyte MGPs do not differ between the groups, we re-estimated group differences in these MGPs while adjusting for the oligodendrocyte and microglia MGPs. This analysis showed no significant differences between the groups (Additional file 2: Figure S10). Therefore, only MGPs of oligodendrocyte and microglia were included in the statistical model of differential expression.

**Functional enrichment**
Functional enrichment analysis of the differential gene expression results without MGP adjustment indicated 476 significantly enriched (FDR < 0.05) pathways in PW (107 up-regulated and 369 down-regulated) and 992 in NBB (421 up-regulated and 571 down-regulated). MGP adjustment reduced the number of significant pathways to 89 in PW (35 up-regulated and 54 down-regulated) and 248 in NBB (115 up-regulated and 133 down-regulated). Of these, 34 pathways replicated across the two cohorts. Concordant pathways comprised protein folding, ER-related processes and lipid oxidation (Fig. 3). The complete results are provided in Additional file 5.

As expected, scoring each pathway according to the change in p-value when accounting for cellularity, revealed a marked downplay of the relevant cell type-specific functions (Table 1). In the PW cohort, which was characterized by a skewed oligodendrocyte/neurons proportion, the function with the largest attenuation (i.e. increase in p-value) was seen for up-regulation of myelination and other oligodendrocyte related functions and for down-regulation of neuronal pathways. For NBB, accounting for cell composition resulted in attenuation of immunity and neuronal pathways, consistent with the unbalanced microglial/neuronal proportions seen in that cohort (Table 1). Strikingly, pathways linked to mitochondrial respiration, including respiratory complex I, were among the down-regulated processes that lost statistical significance when controlling for cellularity. The attenuation of the mitochondrial signal was observed in both cohorts. Conversely, up-regulation of protein folding-related pathways gained significance in both cohorts (Table 2 and Fig. 3). Complete results are provided in Additional file 6.

**Discussion**
We present the first genome-wide transcriptomic study in the PD brain employing whole RNA-seq after rRNA depletion and random primer capture (Ribo-Zero). Our findings show that PD-associated differential gene expression signatures in bulk brain tissue are influenced to a great extent by the underlying differences in cell type composition of the samples. Modeling cell type heterogeneity allowed us to highlight transcriptional signatures that are likely to represent aberrant gene expression within the cells of the PD brain, rather than changes in cell composition.

Our results suggest that the Ribo-Zero approach is superior to the more commonly used poly(A) method and allows for a more accurate mapping and quantification of the transcriptome in post-mortem brain tissue. The Ribo-Zero method provides substantially higher evenness of coverage and effectively mitigates the 3′- and 5′-end coverage bias associated with poly(A) capture. Ultimately, the unevenness of coverage will influence transcript quantification, affecting the sensitivity of the
differential expression estimates. While these observations are in agreement with previous comprehensive reports [1, 25, 47, 56], we cannot rule out the contribution of experimental variables specific to each cohort, in addition to the RNA sequencing methodology. Furthermore, while the Ribo-Zero protocol shows advantages
Table 1 Loss of significance in enriched pathways

| Pathway                          | PW Up-regulated | PW Down-regulated | NBB Up-regulated | NBB Down-regulated |
|----------------------------------|----------------|------------------|------------------|--------------------|
| myelination                      | –8.85          | –9.63            | –8.67            | –9.60              |
| ensheathment of neurons          | –8.67          | –9.60            |                  |                    |
| axon ensheathment                | –8.67          | –9.58            |                  |                    |
| detection of chemical stimulus   | –7.95          | –9.48            |                  |                    |
| involved in sensory perception   | –                | –                 |                  |                    |
| of bitter taste                  | –                | –                 |                  |                    |
| oligodendrocyte differentiation  | –               | –                 | –6.58            | –9.44              |
| oligodendrocyte development      | –6.58          | –9.44            |                  |                    |
| apical junction complex          | –5.57          | –9.41            |                  |                    |
| glial cell development           | –5.54          | –9.27            |                  |                    |
| glial cell differentiation       | –5.52          | –9.19            |                  |                    |
| tight junction                   | –4.16          | –9.11            |                  |                    |
| NBB                              | -              | -                | -                | -                  |

| Pathway                          | NBB Up-regulated | NBB Down-regulated |
|----------------------------------|------------------|--------------------|
| activation of innate immune      | –8.54            | –9.09              |
| response                          |                  |                    |
| regulation of leukocyte          | –8.11            | –9.03              |
| proliferation                   |                  |                    |
| regulation of lymphocyte         | –8.01            | –8.99              |
| proliferation                   |                  |                    |
| regulation of mononuclear cell   | –7.79            | –8.92              |
| proliferation                   |                  |                    |
| innate immune response-activating| –7.43            | –8.87              |
| signal transduction             |                  |                    |
| regulation of adaptive immune    | –7.16            | –8.82              |
| response                        |                  |                    |
| response to interferon-gamma     | –7.15            | –8.77              |
| adaptive immune response based   | –7.07            | –8.70              |
| on somatic recombination of      |                  |                    |
| immunoglobulin superfamily       |                  |                    |
| domains                          |                  |                    |
| blood microparticle              | –6.75            | –8.68              |
| regulation of T cell proliferation| –6.73            | –8.43              |

Notes: Tables representing the top 10 pathways with the lowest delta for up- and down-regulated pathways for PW and NBB cohorts. The delta value represents the change in the enrichment –log10 (p-value) between the results with and without MGP adjustment (negative values of delta imply a loss of significance when accounting for cellularity). Complete results are provided in Additional file 6.

Intriguingly, accounting for cellular proportions downplayed several of the transcriptomic signatures that have been previously associated with PD. For instance, the signal from vesicle trafficking- and synaptic transmission-related processes [9, 10, 15, 21, 29, 40] was significantly attenuated in both cohorts, suggesting that the signal was primarily driven by changes in neuronal proportions between PD and controls, rather than modulation of these pathways within neurons. Moreover, we observed an attenuation in the down-regulation of mitochondrial pathways, including the respiratory chain and oxidative phosphorylation, which are among the most consistent transcriptomic signatures in PD [8, 37, 52].
The loss of transcriptional signal in these pathways is intriguing, because there is compelling evidence that decreased complex I protein levels occur in PD neurons [23]. Our results suggest that the previously reported transcriptional down-regulation of the respiratory chain is at least partly driven by altered cellular composition (due to decreased number of neurons which highly express these genes) and may therefore not be the sole mechanism by which neuronal complex I deficiency occurs in PD. Indeed, it has been suggested that complex I deficiency in PD may be mediated by proteolytic degradation by the LON-ClpP protease system, rather than transcriptional regulation [44].

Changes in the cell-composition of the affected brain regions occur in all neurodegenerative diseases, including PD, Alzheimer disease, amyotrophic lateral sclerosis (ALS) and Huntington disease. Interestingly, common and overlapping transcriptional signatures have been reported across these neurodegenerative diseases, including mitochondrial, neuronal-specific, and immunity-related pathways [4, 17]. Our findings suggest that these common transcriptional signatures of neurodegeneration may largely represent the common pattern of altered cellularity, involving neuronal loss and glial proliferation, rather than biological processes of causal nature.

Accounting for cell type composition in our samples highlighted processes related to the endoplasmic reticulum, unfolded protein response and lipid/fatty acid oxidation as the top differential gene expression signatures in the PD prefrontal cortex. Unfolded protein response is indeed one of the most consistently reported transcriptomic signatures in PD [8, 9, 20, 28, 41, 55]. Moreover, endoplasmic reticulum stress and aberrant proteostasis have been associated with the accumulation of misfolded proteins, including α-synuclein, in both in vitro studies and animal models of PD [16]. While less is known regarding the role of lipid metabolism in PD, evidence of aberrant fatty acid oxidation has been found by metabolomic studies in serum [12] and urine [36] of patients. Our results corroborate these findings and indicate that aberrant fatty acid metabolism occurs in the PD prefrontal cortex.

Based on our findings, we advocate that modeling cell type heterogeneity is crucial in order to unveil transcriptomic signatures reflecting regulatory changes in the PD brain. It is, however, important noting that modeling of cellular estimates cannot completely mitigate the cell-composition bias in bulk tissue. Moreover, cell type correction complicates the identification of transcriptional changes that are confounded with changes in cellular composition and may thus increase the false negative rate. Single-cell or cell-sorting based methods will be key to overcoming this limitation and deciphering transcriptomic signatures directly associated with underlying disease mechanisms in PD.

### Conclusions

Our findings show that differential gene expression signatures derived from bulk brain tissue of PD patients are significantly confounded by underlying differences in cell type composition. Modeling cell type heterogeneity is

| Table 2 Gain of significance in enriched pathways |
|-------------------------------------------------|
| **PW** Up-regulated                              | Down-regulated |
| Pathway                  | Delta | Pathway                  | Delta |
| protein folding         | 5.77  | DNA packaging complex    | 3.76  |
| ‘de novo’ protein folding | 5.73  | basement membrane        | 3.47  |
| unfolded protein binding | 5.54  | positive regulation of epithelial cell proliferation | 2.52 |
| chaperone-mediated protein folding | 5.32  | negative regulation of gliogenesis | 2.49 |
| ‘de novo’ postranslational protein folding | 4.68  | fatty acid beta-oxidation | 2.30 |
| heat shock protein binding | 4.34  | nucleosome               | 2.18  |
| response to unfolded protein | 4.10  | glomerulus development   | 2.16  |
| response to topologically incorrect protein | 3.53  | aorta development       | 2.06  |
| oxidoreductase activity, acting on paired... | 2.74  | endothelium development | 1.95  |
| NBB Up-regulated                              | Down-regulated |
| Pathway                  | Delta | Pathway                  | Delta |
| positive regulation of cardiac muscle tissue development | 1.99 | tertiary granule | 5.22 |
| regulation of smooth muscle cell differentiation | 1.98 | ficolin-1-rich granule membrane | 5.00 |
| negative regulation of protein serine/threonine kinase... | 1.98 | regulation of myeloid leukocyte mediated immunity | 4.55 |
| hormone-mediated signaling pathway | 1.95 | regulation of leukocyte degranulation | 4.34 |
| lung alveolus development | 1.71 | specific granule         | 4.22  |
| positive regulation of stratified muscle tissue development | 1.69 | ficolin-1-rich granule | 4.15 |
| positive regulation of muscle organ development | 1.69 | tertiary granule membrane | 3.57 |
| positive regulation of muscle tissue development | 1.62 | regulation of mast cell activation | 3.52 |
| negative regulation of MAP kinase activity | 1.48 | vacuolar lumen          | 3.05  |
| regulation of cardiac muscle cell differentiation | 1.48 | regulation of mast cell degranulation | 3.05 |

Tables representing the top 10 pathways with the highest delta for up- and down-regulated pathways for both cohorts. The delta value represents the change in the enrichment –log_{10} (p-value) between the results with and without MGP adjustment (positive values imply an increase in p-value when accounting for cellularity). Complete results are provided in Additional file 6.
crucial in order to unveil transcriptomic signatures that represent regulatory changes in the PD brain and are, therefore, more likely to be associated with underlying disease mechanisms.

Supplementary information
Supplementary information accompanies this paper at https://doi.org/10.1186/s40478-020-00932-7.

Additional file 1. Cohort demographic and experimental information.
Additional file 2: Figure S1. Read mapping efficiency; Figure S2. Read mapping statistics; Figure S3. Sample clustering; Figure S4. Neuronal MGPs and expression of neuronal markers; Figure S5. Oligodendrocyte MGPs and expression of oligodendrocyte markers; Figure S6. Microglial MGPs and expression of microglial markers; Figure S7. Astrocyte MGPs and expression of astrocyte markers; Figure S8. Endothelial MGPs and expression of endothelial markers; Figure S9. Neuroexpressed neuronal markers and overlap with single-cell neuronal DEGs; Figure S10. Cellular estimates grouped by status; Table S1. Correlations between MGPs calculated on different marker sets.

Additional file 3. Significant genes overlapping PW and NBB accounting for cell types.
Additional file 4. Complete results of the differential expression analyses.
Additional file 5. ErmineR pathway enrichment analyses before and after accounting for cell composition.
Additional file 6. Up- and down-regulated pathways ranked by the change in p-value resulting from accounting for cell composition.
Additional file 7. Read count matrix.

Abbreviations
ALS: Amyotrophic lateral sclerosis; FDR: Benjamini-Hochberg false discovery rate; GO: Gene Ontology; MGP: Marker gene profiles; NBB: Netherlands Brain Bank; PD: Parkinson’s disease; PMI: Post-mortem interval; PW: Park West; RIN: RNA integrity number; rRNA: ribosomal RNA; Snc: Substantia nigra pars compacta

Acknowledgments
Not applicable.

Authors’ contributions
GSN participated in the study conception, preprocessed the RNA-seq data, designed and performed the RNA-seq computational analyses. FD participated in the RNA-seq computational analyses. KH contributed to data interpretation and critical revision of the manuscript. UJ and KP participated in the conception and design of the study, data interpretation and critical revision of the manuscript. OBT and GA contributed part of the tissue material, supplementary data, and critical revision of the manuscript. CT conceived, designed and directed the study, contributed to data analyses and interpretation and acquired funding for the study. GSN and CT wrote the manuscript with the active input and participation of LT, KH, UJ, and KP. All authors have read and approved the manuscript.

Funding
This work is supported by grants from The Research Council of Norway (288164, ES633272) and Bergen Research Foundation (BFS2017REK05).

Availability of data and materials
The datasets supporting the conclusions of this article are included within the article and its supplementary files. The source code for the analyses is available in the GitLab repository (https://git.appuib.no/neuromics/cell-composition-rna-pd) under the GPL public license v3.0.

Ethics approval and consent to participate
Informed consent was available from all individuals. Ethical permission for these studies was obtained from our regional ethics committee (REK 2017/2082, 2010/1700, 131.04).

Competing interests
The authors declare that they have no competing interests.

Author details
1NeuroSyrdell Center of Excellence for Clinical Research in Neurological Diseases, Department of Neurology, Haukeland University Hospital, 5021 Bergen, Norway. 2Department of Clinical Medicine, University of Bergen, PB 7804, 5020 Bergen, Norway. 3Computational Biology Unit, Department of Informatics, University of Bergen, PB 7803, 5020 Bergen, Norway. 4The Norwegian Centre for Movement Disorders and Department of Neurology, Stavanger University Hospital, PB 8100, 4068 Stavanger, Norway. 5Department of Mathematics and Natural Sciences, University of Stavanger, 4062 Stavanger, Norway.

Received: 11 March 2020 Accepted: 14 April 2020
Published online: 21 April 2020

References
1. Adiconis X, Borges-Rivera D, Satija R, DeLuca DS, Busby MA, Berlin AM, Svachenko A, Thompson DA, Wysoker A, Fennell T, Gnirek A, Pochet N, Regev A, Levin JZ (2013) Comparative analysis of RNA sequencing methods for degraded or low-input samples. Nat Methods 10:623–629. https://doi.org/10.1038/nmeth.2483
2. Alves G, Muller B, Hørkofson K, HogenEsch I, Telstad W, Aarland T, Tysnes O-B, Larsen JP, for the Norwegian ParkWest study group (2009) Incidence of Parkinson’s disease in Norway: the Norwegian ParkWest study. J Neurol Neurosurg Psychiatry 80:851–857. https://doi.org/10.1136/jnnp.2008.186211
3. Andrews S (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
4. Amoros D, Zhang Y, Yang X, Narayan M (2018) Shared mechanisms among neurodegenerative diseases: from genetic factors to gene networks. J Genet 97:795–806
5. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. Nat Genet 25:25–29. https://doi.org/10.1038/75556
6. Bauer M (2007) RNA in forensic science. Forensic Sci Int Genet 1:69–74. https://doi.org/10.1016/j.fsigen.2006.11.002
7. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170
8. Borrageiro G, Haylett W, Seedar S, Kuivaniemi H, Bardien S (2018) A review of genome-wide transcriptomics studies in Parkinson’s disease. Eur J Neurosci 47:1–16. https://doi.org/10.1111/ejn.13760
9. Bossers K, Meerhoff G, Balesar R, von Dongen JW, Kruse CG, Swaab DF, Verhaagen J (2009) Analysis of gene expression in Parkinson’s disease: possible involvement of neurotrophic support and axon guidance in dopaminergic cell death. Brain Pathol Zurich Switz 19:462–468. https://doi.org/10.1002/bpa.20171
10. Botta-Orfila T, Tolosa E, Gelpi E, Sánchez-Pla A, Martí M-J, Valldeoriola F, Verhaagen J (2009) Analysis of gene expression in Parkinson’s disease and mild cognitive impairment. Mov Disord Off J Mov Disord Soc 24:927–932. https://doi.org/10.1002/mds.22695
11. Broad Institute (2018). Picard tools. Available online at: http://broadinstitute.github.io/picard/
12. Burte F, Houghtdon D, Lowes H, Pyle A, Nesbitt S, Yamall A, Yu-Wai-Man P, Burn DJ, Santibanez-Koref M, Hudson G (2017) Metabolic profiling of Parkinson’s disease and mild cognitive impairment. Mov Disord Off J Mov Disord Soc 32:927–932. https://doi.org/10.1002/mds.26992
13. Cantuti-Castelvetri I, Keller-McGandy C, Bouzou B, Aniste G, Clark TW, Frosch MP, Standaert DG (2007) Effects of gender on nigral gene expression and parkinson disease. Neurobiol Dis 26:606–614. https://doi.org/10.1016/j.nbd.2007.02.009
27. Gillis J, Mistry M, Pavlidis P (2010) Gene function analysis in complex data.
26. Gelb DJ, Oliver E, Gilman S (1999) Diagnostic criteria for Parkinson disease.
25. Gallego Romero I, Pai AA, Tung J, Gilad Y (2014) RNA-seq: impact of RNA
20. Dumitriu A, Golji J, Labadorf AT, Gao B, Beach TG, Myers RH, Longo KA,
19. Duke DC, Moran LB, Kalaitzakis ME, Deprez M, Dexter DT, Pearce RKB,
18. Dickson DW (2012) Parkinson’s disease and parkinsonism: neuropathology. Cold Spring Harb Perspect Med 2:000928–a000928. https://doi.org/10.1101/cshperspect.a00928
17. Cooper-Knock J, Kirby J, Ferriauolo L, Heath PR, Ratray M, Shaw PJ (2012)
16. Colla E (2019) Linking the endoplasmic reticulum to Parkinson disease.
15. Chandrasekaran S, Bonchev D (2013) A network view on Parkinson disease.
14. Capurro A, Bodea L-G, Schaefer P, Luthi-Carter R, Perreau VM (2014) Computational
deconvolution of genome wide expression data from Parkinson’s and Huntington’s disease brain tissues using population-specific expression analysis. Front Neurosci 8: 441. https://doi.org/10.3389/fnins.2014.00441
13. Chandrasekaran S, Bonchev D (2013) A network view on Parkinson’s disease. Comput Struct Biotechnol J 7:e201300404. https://doi.org/10.9596/csbj.2013.00404
12. Colla E (2019) Linking the endoplasmic reticulum to Parkinson’s disease and alpha-Synucleinopathy. Front Neurosci 13:560. https://doi.org/10.3389/fnins.2019.00560
11. Cooper-Knock J, Kirby J, Ferriauolo L, Heath PR, Ratray M, Shaw PJ (2012) Gene expression profiling in human neurodegenerative disease. Nat Rev Neuro 8:516–518. https://doi.org/10.1038/nrneuro.2012.156
10. Dickinson DW (2012) Parkinson’s disease and parkinsonism: neuropathology. Cold Spring Harb Perspect Med 2:000928–a000928. https://doi.org/10.1101/cshperspect.a00928
49. Soneson C, Love MI, Robinson MD (2016) Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res 4:1521. https://doi.org/10.12688/f1000research.7563.2

50. Srinivasan K, Friedman BA, Larson JL, Lauffer BE, Goldstein LD, Appling LL, Borneo J, Poon C, Ho T, Cai F, Steiner P, van der Brug MP, Modrusan Z, Kaminker JS, Hansen DV (2016) Untangling the brain’s neuroinflammatory and neurodegenerative transcriptional responses. Nat Commun 7:11295. https://doi.org/10.1038/ncomms11295

51. Stamper C, Siegel A, Liang WS, Pearson JV, Stephan DA, Shill H, Connor D, Caviness JS, Sabbath M, Beach TG, Adler CH, Dunckley T (2008) Neuronal gene expression correlates of Parkinson’s disease with dementia. Mov Disord Off J Mov Disord Soc 23:1588–1595. https://doi.org/10.1002/mds.22184

52. Toker L, Mancarci BO, Tripathy S, Pavlidis P (2018) Transcriptomic evidence for alterations in astrocytes and Parvalbumin interneurons in subjects with bipolar disorder and schizophrenia. Biol Psychiatry 84:787–796. https://doi.org/10.1016/j.biopsych.2018.07.010

53. Velmeshev D, Schirmer L, Jung D, Haeussler M, Perez Y, Mayer S, Bhaduri A, Goyal N, Rowitch DH, Kriegstein AR (2019) Single-cell genomics identifies cell type-specific molecular changes in autism. Science 364:685–689. https://doi.org/10.1126/science.aav8130

54. Ward CD, Gibb WR (1990) Research diagnostic criteria for Parkinson’s disease. Adv Neurol 53:245–249

55. Zhang Y, James M, Middleton FA, Davis RL (2005) Transcriptional analysis of multiple brain regions in Parkinson’s disease supports the involvement of specific protein processing, energy metabolism, and signaling pathways, and suggests novel disease mechanisms. Am J Med Genet B Neuropsychiatr Genet 137B:5–16. https://doi.org/10.1002/ajmg.b.30195

56. Zhao W, He X, Hoadley KA, Parker JS, Hayes D, Perou CM (2014) Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. BMC Genomics 15:419. https://doi.org/10.1186/1471-2164-15-419

57. Zheng B, Liao Z, Locascio JJ, Lesniak KA, Roderick SS, Watt ML, Ekblad AC, Zhang-James Y, Kim PD, Hauser MA, Grünblatt E, Moran LB, Mandel SA, Riederer P, Miller MA, Fedoroff HJ, Wüllner U, Papapetropoulos S, Youdim MB, Cantuti-Castelvetri I, Young AB, Vance JM, Davis RL, Hedreen JC, Adler CH, Beach TG, Graeber NB, Middleton FA, Rochet J-C, Scherzer CR, Global PD Gene Expression (GPEX) Consortium (2010) PGC-1α, a potential therapeutic target for early intervention in Parkinson’s disease. Sci Transl Med 2:52ra73. https://doi.org/10.1126/scitranslmed.3001059

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.