RNA sequencing reveals MMP2 and TGFB1 downregulation in LRRK2 G2019S Parkinson's iPSC-derived astrocytes

Heather D.E. Booth¹,¹, Frank Wessely²,¹, Natalie Connor-Robson³, Federica Rinaldi⁴, Jane Vowles⁵, Cathy Browne⁶, Samuel G. Evetts⁶,⁷, Michele T. Hu⁸,⁹, Sally A. Cowley⁸,⁹, Caleb Webber¹⁰,¹¹, Richard Wade-Martins¹,²,⁎,⁎

¹ Oxford Parkinson’s Disease Centre, University of Oxford, Oxford, UK
² UK Dementia Research Institute, Cardiff University, Maindy Road, Cardiff CF24 4HQ, UK
³ James Martin Stem Cell Facility, Sir William Dunn School of Pathology, University of Oxford, Oxford, UK
⁴ Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK

ARTICLE INFO

Keywords:
Induced pluripotent stem cells
Astrocytes
Neurodegeneration
Parkinson’s disease
LRRK2

ABSTRACT

Non-neuronal cell types such as astrocytes can contribute to Parkinson’s disease (PD) pathology. The G2019S mutation in leucine-rich repeat kinase 2 (LRRK2) is one of the most common known causes of familial PD. To characterize its effect on astrocytes, we developed a protocol to produce midbrain-patterned astrocytes from human induced pluripotent stem cells (iPSCs) derived from PD LRRK2 G2019S patients and healthy controls. RNA sequencing analysis revealed the downregulation of genes involved in the extracellular matrix in PD cases. In particular, transforming growth factor beta 1 (TGFB1), which has been shown to inhibit microglial inflammatory response in a rat model of PD, and matrix metallopeptidase 2 (MMP2), which has been shown to degrade α-synuclein aggregates, were found to be down-regulated in LRRK2 G2019S astrocytes. Our findings suggest that midbrain astrocytes carrying the LRRK2 G2019S mutation may have reduced neuroprotective capacity and may contribute to the development of PD pathology.

1. Introduction

Parkinson’s disease (PD) is the second most common neurodegenerative disorder affecting approximately 1% of people over 60 years of age (de Lau and Breteler, 2006). PD progressively manifests itself in classical motor impairments including rigidity and tremor, but patients also show non-motor symptoms like cognitive impairment, psychiatric symptoms or sleep dysfunction (Khoo et al., 2013). While the majority of PD cases are idiopathic, mutations within a small number of genes underlie monogenic forms of the disease (Hernandez et al., 2016; Klein and Westenberger, 2012). The main pathological hallmarks of PD are accumulation of α-synuclein (Spillantini et al., 1998) and the selective loss of dopaminergic neurons in the substantia nigra pars compacta (SNpc) (Damier et al., 1999; Kordower et al., 2013).

Cellular models of PD have traditionally focused on dopaminergic neurons. However, there is growing evidence that other non-neuronal cell types such as astrocytes are involved in the degeneration of dopaminergic neurons and contribute to PD pathology (Booth et al., 2017). Astrocytes represent one of the most abundant types of glial cells in the brain (von Bartheld et al., 2016), and play multiple key roles in a wide range of physiological processes ensuring healthy brain function. These processes include synaptic formation and maturation, regulation of neuronal synaptic transmission, maintenance of the blood brain barrier integrity and controlling the homeostasis of ions, neurotransmitters and energy metabolites (Khakh and Sofroniew, 2015; Oberheim et al., 2012; Sofroniew and Vinters, 2010). Several of the variant genes considered to be causative for PD, including PARK2 (Parkin) and PARK7 (DJ-1), have been found to be expressed in astrocytes and have important roles in astrocyte-specific functions such as inflammatory responses, glutamate transport and neurotrophic capacity (Booth et al., 2017). However, there is currently little knowledge of the involvement of leucine-rich repeat kinase 2 (LRRK2) in astrocyte biology. Mutations in the LRRK2 gene represent the most common known genetic cause of PD (Hernandez et al., 2005; Paisán-Ruiz et al., 2004; Zimprich et al., 2004). The most common PD-associated LRRK2 mutation, G2019S, is found in approximately 4% of familial and 1% of sporadic forms of PD, and the
risk of PD for carriers of the \textit{LRRK2} \textit{G2019S} mutation is estimated at 25–43\% \cite{Goldwurm_etal.2011, Lee_etal.2017}. Although it has been demonstrated that \textit{LRRK2} is expressed in purified human astrocytes at similar levels to neurons \cite{Zhang_etal.2016}, no studies have been conducted to investigate the function of \textit{LRRK2} in human astrocytes.

Here, we present the first study of iPSC-derived midbrain-patterned astrocytes from PD patients carrying the common \textit{LRRK2} \textit{G2019S} missense mutation. In order to understand the effect of this mutation on astrocyte function, we compared the gene expression profiles of iPSC-derived midbrain-patterned astrocytes from PD patients with those from healthy controls, while controlling for differentiation efficiency. The gene expression data reveal changes in expression of genes involved in the extracellular matrix (ECM), which are largely downregulated in PD cell lines. Transforming growth factor beta 1 (\textit{TGFBI}), which inhibits microglial inflammatory response in a rat model of PD \cite{Chen_etal.2017}, and matrix metalloproteinase 2 (\textit{MMP2}), which has recently been shown to degrade \textalpha-synuclein aggregates (Oh \etal, 2017), were found to be downregulated in PD astrocytes. These findings suggest that astrocytes in the brains of \textit{LRRK2} \textit{G2019S} carriers may have reduced neuroprotective capacity, contributing to a less inhibited inflammatory response and an increased burden of extracellular \textalpha-synuclein aggregates, which in turn could contribute to PD pathology.
A

<table>
<thead>
<tr>
<th></th>
<th>Midbrain patterning</th>
<th>Progenitor expansion</th>
<th>Differentiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHH, Pur, FGF8a</td>
<td>BGTDAC, CHIR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB, LDN</td>
<td>FGF2, Heparin, EGF</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LIF</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time (days)</th>
<th>0</th>
<th>20</th>
<th>30</th>
<th>34</th>
<th>90</th>
<th>118+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replate cells</td>
<td>Passage cells every 7 days</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B

- GFAP
- AQP4
- GLAST1
- FOXA2
- LMX1A
- PAX6
- POU5F1
- NANOG

Relative gene expression ($2^{-\Delta \Delta CT}$)

C

Day 21
- DAPI
- FOXA2
- LMX1A
- LMX1A/FOXA2

D

Day 118
- DAPI
- GLAST1
- GFAP
- DAPI/GLAST1/GFAP

(caption on next page)
Fig. 1. Differentiation of midbrain-patterned astrocytes. (A) Midbrain-patterned astrocyte differentiation protocol schematic. (B) Quantitative real-time polymerase-chain reaction (qRT-PCR) analysis of pluripotency markers (POU class 5 homeobox 1, POUSFI; homeobox protein NANOG, NANOG), midbrain and neuronal transcription factors (FOXA2, LMX1A and PAX6) and astrocyte markers (GFAP, AQP4 and GLAST1) during the differentiation; graphs show average of four lines, mean ± SEM. (C) Immunostaining at DIV 21 for the nuclear marker DAPI and the floorplate markers FOXA2 and LMX1A. (D) Immunostaining at DIV 118 for the nuclear marker DAPI, the astrocyte glutamate transporter GLAST1 and the astrocyte cytoskeletal marker GFAP. Abbreviations: LDN, LDN-193189; SB, SB-431542; SHH, sonic hedgehog C24Fl; Pur, purmorphamine; GF8b, fibroblast growth factor 8a; CHIR, CHIR99021; BDG/DAC, B: brain-derived neurotrophic factor, G: glial-derived neurotrophic factor, T: transforming growth factor beta 3, D: DAPT, A: ascorbic acid, C: N6,2′-O-dibutyryladenosine 3′,5′-cyclic monophosphate; EGF, epidermal growth factor; LIF, leukaemia inhibitory factor; GF2F, fibroblast growth factor 2.

2.6. Glutamate uptake

Sodium-dependent glutamate uptake was measured by incubating cells with 50 nM L-[3,4-3H]-glutamic acid in 135 mM NaCl, 3.1 mM KCl, 1.2 mM CaCl2, 1.2 mM MgSO4, 0.5 mM KH2PO4, and 2.0 mM glucose (all Sigma), adjusted to pH 7.4, for 10 min. Sodium-independent uptake was measured by substituting NaCl for 135 mM choline chloride (Sigma). Uptake was stopped by the addition of 500 μM cold H3-glutamic acid. Cells were lysed with 0.1 M NaOH (Sigma). 3 mL Optiphase Supermix scintillation fluid (Perkin Elmer) was added before analysis using a Tri-Carb 2800 TR liquid scintillation counter (Perkin Elmer). Data were normalized to protein content, calculated using a BCA assay.

2.7. RNA-seq library preparation

For each sample, 386 ng total RNA was used for library preparation. RNA samples were prepared using standard poly-A isolation protocol (Illumina TruSeq Stranded mRNA kit) and, following pooling, sequenced on one lane of Illumina HiSeq4000 using 75 bp paired end reads, RNA samples showed consistently high RIN (RNA integrity number) values of at least 8.8. Sequencing resulted in approximately 50–56 million reads per sample. RNA-seq data are available at GEO under accession number: GSE120306.

2.8. Confirmation of sample identity

Individual identity of RNA-seq samples was confirmed by comparing genotypes obtained from SNP-array data from undifferentiated cell lines or original fibroblast cell cultures with genotypes derived from RNA-seq data. A high percentage of identical SNPs (> 90%) called by both SNP array and transcriptomic data confirmed sample identity.

2.9. Transcriptome analysis

Salmon (0.7.2) (Patro et al., 2017) with default settings was used to transcript expression using ensembl gRch38 version 87 as the human transcriptome reference, which combined cdNA and ncRNA, and for which alternative sequences were removed before mapping. Gene annotation was obtained from the fasta headers of the ensembl reference used for mapping. Expression was summarised on the gene level using the tximport R package (1.10.1) (Soneson et al., 2015). Normalised counts were obtained from DESeq2 (1.22.2) (Love et al., 2014). Mapping rates against the human transcriptome were consistently high (90 to 93%), and the vast majority of mapped reads were assigned to protein-coding genes (93 to 95%). Differential gene expression analysis was performed with DESeq2 (1.22.2) using protein-coding genes; adjusted p-value cutoff 0.05. Two covariates were included in the design of the model: sex and differentiation efficiency as estimated by GFAP expression levels obtained from RNA-seq data. X-chromosomal genes were excluded from differential expression analysis (see below for further details). Enrichment analysis was performed with the R package ClusterProfiler (3.10.1) (Yu et al., 2012). Protein-protein interaction data were obtained from InBioMap (InWeb_InBioMap core 2016.09.12) (Li et al., 2017). The PPI network with its clusters was obtained with the R package igraph (1.2.3) (Nepusz, 2006), where proteins in clusters with fewer than 10 members were excluded from differential expression analysis (see below for further details). Enrichment analysis was performed with the R package ClusterProfiler (3.10.1) (Yu et al., 2012). Protein-protein interaction data were obtained from InBioMap (InWeb_InBioMap core 2016.09.12) (Li et al., 2017). The PPI network with its clusters was obtained with the R package igraph (1.2.3) (Nepusz, 2006), where proteins in clusters with fewer than 10 members were excluded from differential expression analysis (see below for further details). Enrichment analysis was performed with the R package ClusterProfiler (3.10.1) (Yu et al., 2012). Protein-protein interaction data were obtained from InBioMap (InWeb_InBioMap core 2016.09.12) (Li et al., 2017). The PPI network with its clusters was obtained with the R package igraph (1.2.3) (Nepusz, 2006), where proteins in clusters with fewer than 10 members.
were grouped into a cluster with the label ‘0’.
Post-hoc power analysis was performed using the R package RNASeqPower (1.22.1) (Hart et al., 2013) and applying the following parameters. Sample size and false positive rate were set to 3 and 0.05, respectively. The effect size (fold change) was tested with values of 2 and 3. The depth parameter was set to the average sum of total counts (mapped reads) across the six samples in million (46.3). The biological coefficient of variation in our dataset was estimated as 0.428, which is the square root of the estimated common dispersion computed by the estimateDisp() function of the edgeR R package (3.24.3) (Robinson et al., 2010).

2.10. Generation of heatmap with purified brain cell types

The heatmap is based on the expression levels of selected genes as used in Fig. S1G in (Sloan et al., 2017). The plot was generated with the pheatmap R package (1.0.12) (Kolde, 2015). Gene expression levels from six iPSC-derivied astrocytes and 29 purified brain cell types (Zhang et al., 2016) were vst-transformed using the vst() function from DESeq2 R package and the batch effect was removed with removeBatchEffect() from limma R package (3.38.3) (Ritchie et al., 2015). The function heatmap was called with clustering_method = “ward.D2”, scale = “row” and cluster_rows = FALSE.

2.11. Distribution of gene expression on X chromosome and autosomes

Empirical cumulative distributions of X-chromosomal and autosomal gene expression in the four female iPSC-derived astrocyte samples is based on using quantile normalised gene expression data from transcript-per-million (TPM) abundances. First, quantile normalization was performed using normalize.quantiles() from the preprocessCore R package (1.40.0) (Bolstad et al., 2003), then, resulting data were log2-transformed for plotting (stat_ecdf()). Pairwise Kolmogorov-Smirnov tests (ks.test() in R) were used to test for significant differences between samples of higher and lower XIST expression. A similar approach, except using RPKM instead of TPM, was used in Patel et al. (2017).

2.12. Randomisation for PPI network analysis

The randomisation approach takes into account both the coding sequence length and the number of PPIs of the corresponding proteins. For every input gene a random gene that has a similar coding sequence length and the number of PPIs of the corresponding proteins. For every input gene a random gene that has a similar coding sequence length and the number of PPIs of the corresponding proteins. The randomisation approach was used in (Taylor et al., 2015). A total of 1,000,000 random gene sets, of the same size as the number of differentially expressed genes, were tested for their connectivity within the PPI resource (Li et al., 2017) and the number of connections were compared to the one obtained using the list of differentially expressed genes. The PPI analysis was performed on the basis of pairs of genes (using ensemble gene IDs as annotated in the downloaded data file), which means that an interaction between two genes is reported and counted if any of the possible gene products show an interaction.

2.13. External RNA-seq datasets

Public RNA-seq data for comparison were obtained for purified brain cell types (Zhang et al., 2016) and iPSC-derived dopaminergic neurons (Sandor et al., 2017). Data were processed in the same way as RNA-seq data from this study (i.e. mapping to the identical human transcriptome using Salmon with the same settings, and summarising expression on the gene level using the tximport R package).

3. Results

3.1. Differentiation of iPSCs into midbrain-patterned astrocytes

The dopaminergic neurons that selectively degenerate in PD are located in the SNpc, in the midbrain. Although no specific markers of midbrain astrocytes have been identified, it is considered that astrocytes arise from regionally patterned glial progenitors within the brain. The SNpc arises from a region of the developing brain known as the floorplate. We therefore used a modified version of the protocol developed by Kriks et al. (Beevers et al., 2017; Kriks et al., 2011) to pattern floorplate progenitors from monolayer iPSCs (Fig. 1A). Patternning involved dual-SMAD inhibition, and activation of sonic hedgehog (SHH), fibroblast growth factor 8 (FGF8) and WNT signaling (Fig. 1A). At day in vitro (DIV 20) of the protocol, cells were passaged into medium containing epidermal growth factor (EGF) and leukemia inhibitory factor (LIF) to induce proliferation (Sanalkumar et al., 2010). Coexpression of the floorplate markers forkhead box A2 (FOXA2) and LIM homeobox transcription factor 1 alpha (LIMX1A) was confirmed by quantitative real-time polymerase-chain reaction (qRT-PCR) and immunostaining (Fig. 1B,C).

Progenitors were maintained in EGF and LIF until DIV 30, when medium was modified to contain EGF, FGF2 and Heparin (E FH medium; Fig. 1A). This combination was selected in order to maintain the proliferative state of the cells (Bonni et al., 1997; Sanalkumar et al., 2010; Tropepe et al., 1999). By DIV 33, FOXA2 and LMX1A expression were almost entirely lost, and expression of the pan-neuronal transcription marker paired box protein Pax6 (PAX6), which is expressed in tyrosine hydroxylase (TH)-negative (non-dopaminergic) cells in the developing midbrain (Ouan et al., 2013), was detected (Fig. 1B). At DIV 34, cells were passaged and maintained in EPH medium, with passaging every 7 days until DIV 90 (Fig. 1A). Pax6 expression decreased gradually until DIV 69 (Fig. 1B), potentially indicating a switch from neuronal to glial progenitors. At DIV 90, cells were passaged into unsupplemented medium for terminal differentiation and maintained for 4–5 weeks without passaging (Fig. 1A). By DIV 118 Expression of the astrocyte cytoskeletal gene glial fibrillary acid protein (GFAP), the water channel aquaporin-4 (AQP4) and glutamate aspartate transporter 1 (SLC1A3; GLAST1) were detected by qRT-PCR (Fig. 1B). Immunostaining also demonstrated the presence of GFAP and GLAST1 proteins (Fig. 1D). We found that attempts to initiate terminal differentiation earlier than DIV 90, for example at DIV 41 or DIV 69, gave rise to few, if any, GFAP-positive astrocytes (data not shown).

3.2. Differentiation and characterization of LRRK2 G2019S midbrain-patterned astrocytes

Following characterization of cells throughout the differentiation protocol, midbrain-patterned astrocytes were differentiated from iPSC lines generated from seven donors, including four patients with PD who carry the LRRK2 G2019S mutation, and three healthy control individuals (Table 1 and Supplementary Fig. 1). Batches of progenitors derived from each iPSC line were frozen at DIV 90 and used for the experiments described below. Following five weeks of terminal differentiation (DIV 125), the percentage of GFAP-positive astrocytes produced by each line was calculated (Fig. 2A,B).

In vivo, astrocytes take up glutamate in a sodium-dependent manner through their glutamate transporters (Schousboe et al., 1977). Other proteins that are implicated in PD pathogenesis have been shown to influence glutamate uptake; deficiency of DJ-1 and astrocyte-specific overexpression of α-synuclein both result in impaired glutamate uptake in astrocyte cultures (Gu et al., 2010; Kim et al., 2016). To test the functionality of the astrocytes produced using our protocol, and to investigate whether the LRRK2 G2019S mutation had an effect on glutamate uptake, we measured sodium-dependent uptake of [3H]glutamate in our cultures. Uptake was achieved by all lines; however, there was no
significant effect of the LRRK2 G2019S mutation on this process (Fig. 2C,D).

3.3. Transcriptomic analysis

To further characterize our midbrain-patterned astrocytes, and to determine whether the LRRK2 G2019S mutation leads to transcriptomic changes in these cells, we collected samples for RNA-sequencing (RNA-seq) analysis. Initial assessment of transcriptomic variability by principal components analysis (PCA) revealed PD1 as an outlying sample (Supplementary Fig. 2A). Evaluation of expression levels of typical pluripotency marker genes revealed unusually high expression of POU5F1 and NANOG in this sample. Levels were approximately 50 (POUSFI) and 30 (NANOG) times higher than the mean expression level in other samples (Supplementary Fig. 2B). These findings indicated residual pluripotency in PD1, so this sample was excluded from further analyses.

We also observed that two of the four female PD samples (PD2 and PD4) had much lower expression of X-inactive specific transcript (XIST) compared with the other female samples (Supplementary Fig. 2C). XIST is involved in the X-chromosome inactivation process in mamalian females and is expressed from the inactive X chromosome. Average expression of X-chromosomal genes showed a slight shift towards higher levels in low-XIST samples than high-XIST samples, indicating potential erosion of X chromosome inactivation (Supplementary Fig. 2D). In contrast, expression levels of autosomal genes did not show a similar deviation. This shift was not statistically significant at the 5% significance level according to pairwise two-sample Kolmogorov-Smirnov tests. However, since the disease status (controls versus PD) of the four female samples was confounded by the
observed trend, we excluded all X chromosome genes from our analysis.

3.4. iPSC-derived midbrain-patterned astrocyte gene expression profiles are similar to human adult astrocytes

We compared our iPSC-derived midbrain-patterned astrocytes with publicly available data from a set of purified human cortical cells including adult astrocytes, neurons, microglia, oligodendrocytes, endothelial cells and fetal astrocytes (Zhang et al., 2016). PCA projection demonstrated that our iPSC-derived midbrain-patterned astrocyte samples exhibit a gene expression profile that resembles human adult astrocytes and that is distinctive from other brain cell types (Fig. 3A).
We performed hierarchical clustering of our iPSC-derived midbrain-patterned astrocytes and these human samples using genes that have been shown to be differentially expressed between different cell types in the central nervous system (Sloan et al., 2017). We found that our samples clustered closely with adult astrocytes, and were distinct from the fetal astrocyte group (Fig. 3B). Secreted protein acidic and cysteine rich (SPARC), LIM homeobox 2 (LHX2) and empty spiracles homeobox 2 (EMX2) have previously been identified as differentially expressed in midbrain astrocytes compared to other brain regions (Morel et al., 2017; Xin et al., 2019). Consistent with these studies, we found that SPARC was upregulated and LHX2 and EMX2 were downregulated in our iPSC-derived midbrain-patterned astrocytes when compared with purified cortical astrocytes (Fig. 3C).

### 3.5. LRRK2 is expressed at low levels in iPSC-derived midbrain-patterned astrocytes

Data suggest that LRRK2 expression is similar in isolated human cortical astrocytes and neurons (Zhang et al., 2016). This comparison has not been made for cells in the human midbrain, or in iPSC-derived midbrain-patterned cultures. We therefore compared LRRK2 expression levels in our iPSC-derived midbrain-patterned astrocytes with data from our previously published iPSC-derived dopaminergic neurons (Sandor et al., 2017), and with purified human cortical cell types (Zhang et al., 2016). We found that among these cell types, LRRK2 gene expression levels were highest in purified microglia (Supplementary Fig. 3). In comparison, LRRK2 expression in our iPSC-derived midbrain-patterned astrocytes was present but low, and similar to iPSC-derived neurons and purified cortical astrocyte samples.

### 3.6. The extracellular matrix is perturbed in LRRK2 G2019S iPSC-derived midbrain-patterned astrocytes

Genes that are differentially expressed between iPSC-derived midbrain-patterned astrocytes from PD LRRK2 G2019S cases and controls were identified using DESeq2 (Love et al., 2014). Due to concerns about erosion of X-inactivation, genes on the X chromosome were excluded from this analysis (see above). Two covariates were included in the design matrix: sex and differentiation efficiency; both of which correlated with separation of the samples in the PCA (Supplementary Fig. 4A,B). Differentiation efficiency was controlled for using expression levels of the astrocyte marker gene GFAP, which was highly correlated with the percentage of astrocytes produced by each cell line (Supplementary Fig. 4C). A total of 279 genes were identified as differentially expressed, 80% (223) of which showed reduced expression in PD samples. Analysis of the function of differentially expressed genes by Gene Ontology enrichment analysis found that the strongest signal came from biological processes of the ECM, which are down-regulated in the PD LRRK2 G2019S samples (Fig. 4A).

### 3.7. Neuroprotective potential may be lost in LRRK2 G2019S iPSC-derived midbrain-patterned astrocytes

Given convergent extracellular functionality among the differentially expressed genes (DEGs), we wanted to test whether the proteins encoded by these DEGs physically interacted with each other more than expected by chance. The protein-protein interaction (PPI) network between the gene products of all DEGs reveals that approximately 42% of proteins encoded by the set of DEGs interact with at least one other protein within that set (Fig. 4B). No simulated gene set of the same size (controlling for node degree and gene length, empirical p-value < 1e-06) had as many interactions as the set of DEGs identified, demonstrating the unusual interconnectedness of the DEG gene set. The majority of proteins within the PPI network (98 out of 108 proteins) are encoded by genes that are down-regulated in PD LRRK2 G2019S samples. All clusters within the PPI network are enriched for ECM organization (Fig. 4C). For individual clusters, the strongest signals were detected for ECM disassembly and kinase signaling pathways (Cluster 2), collagen-related processes (Cluster 3), or cell adhesion (Cluster 4).

Of the most interconnected proteins in the PPI network, TGFβ1 and MMP2 have both been previously implicated in neuroprotective mechanisms relevant to PD. The neuroprotective role of TGFβ1 has been shown in the 1-methyl-4-phenylpyridinium (MPP+) rat model of PD, where TGFβ1 treatment led to inhibition of microglial inflammatory response (Chen et al., 2017). MMP2 has been shown to degrade extracellular α-synuclein aggregates in vitro and in vivo, leading to a reduction in cellular apoptosis as compared with controls (Oh et al., 2017). These proteins were identified as the most inter-connected within cluster 2, with 10 and 9 PPIs, respectively.

In our transcriptomic dataset, as well as being highly interconnected in the PPI, both TGFβ1 and MMP2 were expressed at significantly lower levels in midbrain-patterned astrocytes carrying the LRRK2 G2019S mutation than in healthy control samples; a finding that was confirmed by qRT-PCR (Fig. 4D). To test whether changes in expression levels of these two genes are specific to astrocytes, we reviewed the RNA-seq results of human iPSC-derived LRRK2 G2019S dopaminergic neurons (Sandor et al., 2017) and found that they were not differentially expressed. This was confirmed by qRT-PCR (Supplementary Fig. 5). Reduced expression of these genes indicates that astrocytes from patients with the LRRK2 G2019S mutation may have reduced neuroprotective capacity.

### 4. Discussion

Mutations in LRRK2 are a common cause of PD. Our study of iPSC-derived midbrain patterned astrocytes found that LRRK2 was expressed in these cells, at similar levels to purified human cortical astrocytes and iPSC-derived dopaminergic neurons, and at lower levels than purified human microglia and oligodendrocytes. The LRRK2 G2019S mutation led to downregulation of genes involved in the ECM, which is essential for cell-cell communication through extracellular signaling (Vargová and Syková, 2014). This type of communication is important for initiating astrocyte response to neuronal activity, and modification of ECM composition can have profound effects on astrocyte response to inflammatory stimuli (Johnson et al., 2015).

TGFβ1 and MMP2 are intrinsically involved in regulation of the ECM. We found that both of these genes were down-regulated in midbrain-patterned astrocytes carrying the LRRK2 G2019S mutation. Furthermore, their protein products were highly connected in our PPI network, suggesting they may have a key role in LRRK2 G2019S-mediated changes in midbrain astrocytes. MMPs modulate cell-cell and cell-ECM interactions by degradation or proteolytic activation of ECM proteins (Baker et al., 2002), and TGFβ1 has been shown to induce expression of MMP2 in a range of cell types including astrocytes (Dhar et al., 2006; Hua et al., 2016; Takahashi et al., 2014). In the presence of interleukin-1ß (IL-1ß), it has been shown that TGFβ1 induction of MMP2 proceeds through downregulation of Tissue Inhibitor of Metalloproteinase (TIMP)1 (Dhar et al., 2006). TIMP proteins (TIMP1-4) are known to inhibit the activity of MMPs (Arpino et al., 2015), and higher expression of TIMPs have been associated with lower levels of MMPs. In our model this was not the case, TIMP2 and TIMP4 were downregulated and TIMP1 and TIMP3 expression was unchanged. This finding may be explained by the fact that in addition to its role in inhibiting MMPs, TIMP2 has been shown to mediate MMP2 activation (Lafleur et al., 2003). The finding that TIMP2 is downregulated in our model may therefore reflect reduced activation, rather than reduced inhibition of MMP2.

TGFβ1 and MMP2 have both been shown to exhibit neuroprotective effects that are relevant to PD. TGFβ1 treatment has been shown to inhibit microglial inflammatory response in the MPP+ rat model of PD (Chen et al., 2017). Downregulation of TGFβ1 expression in our model may therefore point to impaired regulation of inflammatory responses.
in the brains of LRRK2 G2019S carriers. Recently, it has been shown that MMP2 degrades α-synuclein in vitro and in vivo and leads to a reduction in cellular apoptosis (Oh et al., 2017). Reduction of MMP2 expression in our LRRK2 G2019S iPSC-derived midbrain-patterned astrocytes versus controls is supported by similar findings in the substantia nigra of post-mortem brains from PD patients (Lorenzl et al., 2002).

Our study likely under-reports the degree of altered gene expression in patient-derived LRRK2 G2019S astrocytes. Power analysis shows that three samples per group provide an estimated 47% power to detect a two-fold gene expression change rising to 84% for genes with a fold change of three or more, which is the case for both MMP2 and TGFB1 (see Methods). A key issue underlying study power here is the high biological coefficient of variation across samples due in part to culture cellular heterogeneity (see Supplementary Fig. 4), which has been shown to contribute significantly to differences detected between iPSC-derived cell populations (for example, Volpato et al. (2018)). Future studies would benefit from the development of cell sorting approaches or the application of single cell sequencing to reduce this confound.

Evidence for the midbrain identity of our cultures was obtained using midbrain astrocyte markers recently identified in two different studies (Morel et al., 2017; Xin et al., 2019). Morel et al. showed up-regulation of sparcd and downregulation of lhx2 and emx2 in ventral versus dorsal mouse astrocytes, while Xin et al. identified the same trend in mouse ventral midbrain versus cortical astrocytes. We found that the human orthologs of these markers exhibited the same expression pattern in our iPSC-derived astrocytes when compared to purified cortical astrocytes.

Together, these data suggest that midbrain astrocytes of PD patients may have a reduced ability to degrade α-synuclein, thereby

---

Fig. 4. Differential gene expression and functional enrichment analysis. (A) Top 10 most enriched Gene Ontology biological processes among all differentially expressed genes between cases and controls (n = 279). (B) Protein-protein interaction network with sub-clusters highlighted by different colors. Cluster ‘0’ represents proteins with isolated interactions. Labels represent the 10 proteins with the most interactions in the network. (C) Gene Ontology biological processes enriched for expressed genes between cases and controls (n = 3) and PD samples (n = 3), p-values of a two-sided t-test are shown.

---

Differential gene expression and functional enrichment analysis. (A) Top 10 most enriched Gene Ontology biological processes among all differentially expressed genes between cases and controls (n = 279). (B) Protein-protein interaction network with sub-clusters highlighted by different colors. Cluster ‘0’ represents proteins with isolated interactions. Labels represent the 10 proteins with the most interactions in the network. (C) Gene Ontology biological processes enriched for expressed genes between cases and controls (n = 3) and PD samples (n = 3), p-values of a two-sided t-test are shown.
contributing to the build-up of toxic α-synuclein species. The neurodegeneration that results from this toxicity may then activate a cascade of inflammatory responses, which midbrain astrocytes with reduced TGFβ1 expression are unable to attenuate. Our findings thus propose a mechanism through which midbrain astrocytes influence the pathogenesis of PD.

In conclusion, our study demonstrates for the first time that LRCK2 mutations lead to transcriptomic dysregulation in midbrain astrocytes. The identified gene expression changes point to a role for LRCK2 in regulation of the astrocyte extracellular matrix, with implications for neuroprotection in PD. Further studies will be necessary to understand how LRCK2 interacts with ECM proteins, and to realize the functional implications of our findings.

Competing interests

The authors declare no competing interests.

Acknowledgments

The work was supported by the Monument Trust Discovery Award to Parkinson’s UK, a charity registered in England and Wales (2581970) and in Scotland (SC037554). HB was supported by an MRC Industrial CASE studentship. Samples and associated clinical data were supplied by the Oxford Parkinson’s Disease Centre (OPDC) Discovery Cohort, funded by the Monument Trust Discovery Award from Parkinson’s UK, with the support of the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre based at Oxford University Hospitals NHS Trust and University of Oxford, and the NIHR Comprehensive Local Research Network. The James Martin Stem Cell Facility, University of Oxford is financially supported by the Wellcome Trust WTSSSF121302 and the Oxford Martin School LCO910-004. RWM and SAC are supported by the Dementias Platform UK Stem Cell Network Capital Equipment and Partnerships Awards. The work was supported by the Innovative Medicines Initiative Joint Undertaking under grant agreement number 115439, resources of which are composed of financial contribution from the European Union’s Seventh Framework Programme (FP7/2007-2013) and EFPIA companies in kind contribution. We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics, Oxford (Funded by Wellcome Trust grant reference 090532/Z/09/Z and MRC Hub grant G0900747 91070) for the generation of Illumina genotyping and transcriptome data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jnbd.2019.05.006.

References
