Single cell transcriptomics of human PINK1 iPSC differentiation dynamics reveal a core network of Parkinson’s disease

Gabriela Novak (gabriela.novak@alumni.utoronto.ca)
University of Luxembourg

Dimitrios Kyriakis
University of Luxembourg

Kamil Grzyb
University of Luxembourg

Michela Bemini
University of Luxembourg

Steven Finkbeiner
Gladstone Institute

Alexander Skupin
University of Luxembourg

Article

Keywords: Parkinson’s disease, iPSCs, single cell RNA sequencing, network analysis, in-vitro midbrain dopaminergic neurons, mDA, PINK1, PARK6, ubiquitination, mitochondria, protein processing, RNA metabolism, secretory pathways, bioinformatics

DOI: https://doi.org/10.21203/rs.3.rs-72882/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Parkinson's disease (PD) is the second most prevalent neurodegenerative disorder, characterized by the loss of dopaminergic neurons (mDA) in the midbrain. The heterogenous pathology and complex underlying mechanisms are only partly understood and there is no treatment able to reverse PD progression. Here, we targeted the disease mechanisms by focusing on the ILE368ASN mutation within the PINK1 (PARK6) gene and systematically characterized midbrain dopaminergic neurons obtained from human induced pluripotent stem cells (iPSCs). Single-cell RNA sequencing (RNAseq) and pairwise analysis of gene expression identified genes consistently differentially expressed during the mDA neuron differentiation process. Subsequent network analysis revealed that these genes form a core network, which interacts with all known 19 protein-coding Parkinson's disease-associated genes and includes ubiquitination, mitochondrial, protein processing, RNA metabolism, and secretory pathways as important subnetworks. Our findings indicate a unified network underlying PD pathology and offers new interpretation of the phenotypic heterogeneity of PD.

Introduction

Parkinson's disease (PD) is one of the most prevalent neurological disorders, second only to Alzheimer's disease, with a prevalence of 1.8%, among persons over the age of 65 and 2.6% in the 85 to 89 age group. As the average age of the population increases, PD is expected to pose an increasing and significant burden to society. PD is characterized by the presence of motor symptoms, including bradykinesia, rigidity and tremor, but many patients also develop non-motor symptoms, such as depression or dementia. Unfortunately, there are no treatments to slow down or reverse the progression of the disease. Current treatments only temporarily ameliorate the motor symptoms, but do not slow down the progression of PD.

Most of our understanding of PD pathology is based on the identification of mutations that lead to PD, although these account for only 3-5% of PD cases, with the remaining cases being idiopathic. Despite the small fraction of cases they explain, these mutations provide an important window into the underlying molecular mechanisms of PD because they identify pathways which, when disrupted, are able to cause the disease. Many of these mutations converge on mitochondrial homeostasis, repair and mitophagy. Hence, mitochondrial dysfunction likely plays a key role in the pathophysiology of PD. An important group of these mutations lies within the PINK1 gene, a nuclear gene that codes for a kinase important for mitochondrial function. PINK1, also known as PARK6, clearly plays a role in mitophagy, but its function is much broader. The targets of this kinase are involved in many cellular functions, including neuronal maturation. The impact of the loss-of-function mutations in this important kinase has not yet been fully elucidated.

One key characteristic of PD is the death of the midbrain dopaminergic (mDA) neurons, but until recently, it was impossible to study them since 60% of these neurons have disappeared by the time of diagnosis,
and about 90% by the time the patients die\(^{11}\). As a result, research was limited to animal models\(^ {12}\), but human-like mutations in animals often do not lead to the development of comparable pathology\(^ {13}\). The development of cellular reprogramming allows nowadays for the conversion of somatic cells into induced pluripotent stem cells (iPSCs), which can subsequently be differentiated into neurons. This enables us to generate iPSCs from the skin cells of PD patients\(^ {14}\) and differentiate them into mDA neurons carrying the disease\(^ {15,16}\). Furthermore, by using cells from patients with a known PD-associated mutation, we can analyze mutation-specific pathology in the correct, permissive, genetic background. This is important because genetic background is known to significantly influence the age of onset and severity of the disease\(^ {7,17}\). Differentiating mDA neurons from iPSCs provides an almost unlimited source of neurons that allow for deep phenotyping and the elucidation of the cellular mechanisms underlying PD pathology.

Here, we generated iPSCs from the fibroblasts of a patient homozygous for the PD-associated mutation ILE368ASN in the PINK1 gene\(^ {2}\). We used an optimized differentiation protocol to specifically generate mDA neurons, as this cell type displays a unique susceptibility to cell death in PD\(^ {15,18,19}\). In contrast to mDA neurons, the effect of PD on other types of DA neurons is variable, hence their study would not be as pertinent to the elucidation of mechanisms causing PD-induced cell death\(^ {11,20}\).

The development of mDA neurons diverges from other DA neurons even before they commit to neural fate. During early neural development, neural tube stem cells generate neurons and glia, the two basic building blocks of the brain. While other DA neurons follow this direct path from neural stem cells to neurons, determined by the expression of the Pax6 transcription factor, mDA neurons develop from radial glial cells of the floor plate and are exposed to high levels of the SHH transcription factor\(^ {21}\), which prevents expression of Pax6\(^ {22}\) and sets these cells on an entirely different developmental path\(^ {18,23}\). Hence, they follow a very different signalling cascade, leading to the expression of a unique transcriptome that significantly differs from that of other DA neurons\(^ {18,20,24,25}\). Their distinct identity is reflected in their function and leads to their unique susceptibility to death in PD, which in turn leads to the classic movement symptoms of the disease\(^ {11,19,20,23,26}\).

To investigate the disease mechanisms linked to the PINK1 mutation, we performed extensive single-cell RNA expression analysis using DropSeq\(^ {27}\) at four different timepoints during mDA neuron differentiation\(^ {15,16,18}\). Differential expression analysis between four pairs, each consisting of a PINK1 and a control cell line, identified genes that were strongly and consistently dysregulated at all timepoints. Based on known protein-protein interactions, we show that these genes form a network and that its members directly interact with all 19 protein-coding PARK genes associated with PD. This suggests that other PD-associated mutations may also be acting through this common network of genes. Overall, our results indicate the existence of a common disease mechanism that potentially underlies idiopathic PD as well and may represent a unifying perspective on PD progression that will guide future intervention strategies.
Results

We performed a systematic differential expression analysis at single-cell resolution between an iPSC line carrying the PD-associated ILE368ASN mutation in the PINK1 gene and an age- and sex-matched control cell line (control 1-2 in ref 28) during their differentiation into mDA neurons (Fig. 1). After preprocessing and quality filtering, we used 4495 cells and 18,097 genes in our downstream analysis (Methods). Expression of key genes was also confirmed by qPCR and corresponding marker staining. For data integration, we performed a network analysis to identify the underlying key mechanisms of PD progression.

Single-cell RNAseq analysis reveals gene expression panel for direct classification of iPSC-status

The standard accepted procedure for determining iPSC status involves staining for iPSC markers and the use of gene expression platforms, such as Scorecard. However, we show that a standard panel of genes readily detectable by single cell analysis can be used to confirm iPSC status directly in the cells used for the single cell experiment, rather than by staining or expression analysis of an independent sample, which in some cases may not reflect the iPSC status of the experimental sample.

Fibroblasts were isolated from a 64 year-old male with PD symptom onset at 33 years of age who was homozygous for the ILE368ASN (PI368N/PI368N) mutation in the PINK1 gene (Coriell Institute, Cat. No. ND40066). The fibroblasts were confirmed to have a normal karyotype (Supplementary Fig. 1) and an episomal reprogramming method (Epi5™ Episomal iPSC Reprogramming Kit, Invitrogen USA, cat. # A15960) was used to avoid unwanted genetic modifications of the target cell. The karyotype of iPSCs was confirmed (Supplementary Fig. 2) and their iPSC status was ascertained by standard methods, including staining for the POU5F1 marker (also known as Oct4) and by using the TRA-1-80 antibody (Fig. 2a), as well as by a TaqMan iPSC Scorecard Assay, which also confirmed their trilineage potential (Fig. 2b).

Subsequent single-cell RNAseq characterization confirmed that these iPSCs expressed genes routinely used to confirm iPSC status, including SOX2, MYC, POU5F1, and NANOG (Fig. 3a, b). However, iPSCs also expressed several additional genes associated with stemness (Fig. 3a), namely TDGF-1, the expression of which was shown by Hough at al. 29 to be present only in stem cells with the highest expression of stemness markers. Additional genes expressed by the iPSCs were L1TD1, USP44, POLR3G, and TERF1 (essential for the maintenance of pluripotency in human stem cells 30–33), as well as IFITM1, DPPA4, and PRDX1 (associated with stemness 34–36). Based on these observations, the following genes are readily detectable by single-cell RNAseq and this panel should provide a reliable indication of stemness in single cell experiments: SOX2, MYC (cMyc), POU5F1 (Oct4), NANOG, LIN28A, TDGF-1, L1TD1, USP44, POLR3G, and TERF1.

In vitro differentiation of iPSC-derived mDAs recapitulates the in vivo process
To confirm that our differentiation protocol (Supplementary Table 1) is recapitulating the *in vivo* mDA differentiation path, the expression of key genes (OTX2, EN1, LMX1B, LMX1A, and FOXA2) known to drive mDA differentiation *in vivo*\(^{19,23,37}\) was confirmed using our single-cell RNAseq data or by qPCR (Fig. 3b, c, Table 1). The absence of the early, non-mDA neuron marker PAX6\(^{21,22}\) was further validated by staining (Supplementary Fig. 3 and Supplementary Table 2).

*In vivo*, the development of mDA phenotype depends on the early high expression of Sonic Hedgehog (SHH), followed by the induction of Wnt signaling and the expression mDA-specific downstream pathways\(^ {18,19,23}\). Consistent with these *in vivo* differentiation steps, among the highest-expressed genes on day 6 (D6) of the differentiation protocol were PTCH1, a receptor for SHH, and FZD7, a receptor for Wnt proteins (Fig. 3b). Following the differentiation process from iPSCs to D21 (Fig. 4a), we could see the onset of expression of the mature mDA markers TH and KCNH6 (also known as GIRK2) in the D21 cluster (Fig. 4b). By day 21, many factors that are specific to the mDA differentiation path, such as TCF12, ALCAM, PITX2, ASCL1, and DDC\(^ {20,38–41}\), were among the most highly expressed genes (Fig. 4c). The onset of TH expression indicated that the cells achieved the state of early postmitotic mDA neurons\(^ {18}\) (Fig. 4b), which was confirmed by staining (Fig. 4d) and qPCR (Fig. 4e, Supplementary Table 3). In addition, markers specific to mature mDA neurons including TH, DDC, ALDH1A1, and KCNJ6 (also known as GIRK2) also started to become expressed (Fig. 4 b-e, Table 1).

Overall, these observations confirm that our *in vitro* differentiation protocol does indeed recapitulate the *in vivo* differentiation of mDA neurons and produces mDA neurons (Pax6-, Aldh1a1+, Pitx3+, GIRK2+, Nurr1+, Lmx1a+) rather than other types of DA neurons (Pax6+, AldhA1A3+) (Supplementary Table 2).

The PINK1 ILE368ASN mutation is associated with persistently dysregulated expression of nearly 300 loci

To analyze mutation-induced changes in gene expression, we performed single-cell RNA expression analysis, at four important timepoints during the mDA neuron differentiation process, using DropSeq\(^ {27}\) (Fig. 1). After preprocessing and quality-filtering (Methods and Supplementary Fig. 4), a total of 4495 cells (2518 control and 1977 PINK1 cells) and 18,097 genes were included in our analysis). Control and PINK1 cells co-clustered together based on their differentiation stage, from iPSCs, to day 6 (D6), D15 and D21 (Fig. 4a), indicating that RNA expression was specific to differentiation stages, rather than to a cell line. The PINK1 cells at D10 showed low viability, hence D10 timepoint was not included in the pairwise analysis (Fig. 5).

The analysis of pairwise differential expression at each time point with adjusted p-values (\(p_{adj}\) < 0.01 fold changes (FC) > 0.1 resulted in 14 genes that were upregulated and 13 genes that were significantly downregulated in the PINK1 cell line compared to control (Table 2). Because iPSCs are very different from differentiating neuronal precursors, we next tested whether inclusion of iPSCs had disproportionately affected the results through the exclusion of neuron-specific genes. Repeating the analysis using D6, D15 and D21 identified 28 genes that were upregulated and 27 genes that were downregulated at all four
timepoints, including all genes previously identified (Table 2, Fig. 5, Group A). As expected, excluding iPSCs resulted in the identification of a broader range of genes because genes that are differentially expressed only in the neuronal lineage were previously excluded due to the absence of their expression in iPSCs and the requirement that DEGs have to be dysregulated at all timepoints. However, both sets are equally valuable, as genes dysregulated even in iPSCs are more likely to be genes that participate in systemic PD pathology, regardless of cell type, and may be relevant to a broader spectrum of PD pathology than the death of mDA neurons. Interestingly, most of these genes are already linked to PD, other PD mutations, or neurodegeneration (Table 2).

For an alternative definition of differentially expressed genes (DEGs), we used the maximum adjusted p-value in a pairwise combinations as adjusted p-value, and the average fold change that occurred in the pairwise comparison as fold change threshold. With this approach we retained only genes dysregulated in the same direction at all timepoints. This analysis led to 151 DEGs, including the previously identified genes of Group A, of which 65 were upregulated and 86 downregulated compared with controls ($p_{adj} < 0.01$ and FC > 0.1) (Group B, Supplementary Table 4). Taking the mean of FC of the different time points enhanced the identification of DEGs because it reduced the effect of the variability between pairs due to their different differentiation states. Repeating the same analysis for timepoints iPSCs, D6, D15 and D21, but taking into account only the absolute degree of change in iPSCs, yields 172 genes (Group C, Supplementary Table 5). Repeating the analysis using only timepoints D6, D15 and D21 identified a total of 285 DEGs (Group D) (Supplementary Table 6 and Supplementary Fig. 5). Together, when all analyses were pooled, we obtained 291 DEGs (6 genes in Group C depended on the inclusion of iPSCs and did not appear in Group D, see Supplementary Table 6).

**Data integration reveals a common PD network.**

To integrate the expression analysis and identify underlying disease mechanisms, we generated a network of interactions between the DEGs via Gephi \(^{42}\), using protein-protein interaction information obtained from the STRING and GeneMANIA databases \(^{43,44}\). The network we obtained includes 246 of the 291 DEGs, since pseudogenes and non-coding RNAs could not be integrated into a protein-protein interaction network (Supplementary Table 7), and 2122 interactions (Fig. 6, Supplementary Fig. 5c). The curated network only considers DEGs and any genes automatically added by the databases were excluded to ensure a reliable core network based solely on data. Based on known protein-protein interactions, the DEGs integrate into a close-knit core network in which several DEGs form central nodes (Fig. 6a, Supplementary Fig. 5). To evaluate the significance of the DEG-based PPI (protein-protein interaction) network produced by STIRNGdb (v10), we compared the DEG-based network with corresponding random networks generated from sets of 292 randomly chosen genes excluding DEGs. Based on 50 random networks, we show that the DEG-based network includes significantly more protein-coding genes and interactions than by chance (Fig. 6b) and that the network structure in term of degree distribution is significantly distinct as evaluated by Wilcoxon test ($p = 2.22e-16$) and indicates the mechanistic character of the network (Fig. 6c).
The network of genes dysregulated by the presence of the PINK1 mutation includes genes related to other PD-associated pathways, which is intriguing, since it was generally assumed that each PD-associated mutation leads to PD pathology via an independent, characteristic path. For example, two DEGs, GOPC and GPC3, interact with the PD-associated gene DJ-1 (PARK7). The DEG network also includes genes of the LRRK2 (PARK8) network, namely ENAH, HSPA8, MYL6, MALAT1, and SNHG5 (Table 2). SNHG5 and MALAT1 interact with LRRK2 via miR-205-5p. DLK1 and MALAT1 mediate a-synuclein accumulation. In fact, the DLK1-NURR1 interaction involved in this process may be mDA neuron-specific, highlighting the necessity to use mDA neurons for the study of PD-related pathways. Additionally, MALAT1 was shown to increase a-synuclein protein expression. In short, this suggests that interactions leading to PD pathology are more complex than one mutation leading via one path to PD, as generally thought, but it also indicates that there are likely many druggable targets that may be useful in treating PD, and that these may be universally effective for PD caused by several different mutations, and perhaps even for idiopathic PD. For example, terazosin, which is already in clinical use, was found to be associated with slower disease progression, likely by enhancing the activity of phosphoglycerate kinase 1 (PGK1), one of the top DEGs identified in our study.

For the evaluation of the relative importance of each node within the network, we applied betweenness centrality. The major nodes of these networks are formed by genes that were already shown to play an important role in PD pathology (Table 3). Next, we built a correlation network (p-value < 0.05, r > 0.1) of the 246 DEGs based on the normalized counts (Fig. 6e). By extracting the common interactions of these two networks, we obtained a network with 297 interactions (Fig. 6d, e and Supplementary Table 7), which highlights protein-protein connections that correlate with differential expression of the genes. This analysis further supports the role of the connections between these genes in mediating the resulting differential expression in the presence of the PINK1 mutation. STRING was subsequently used to highlight functional pathways represented within the DEG network (Supplementary Fig. 6 and Supplementary Table 8). Several pathways known to play a role in PD pathology are significantly represented within the network, notably ubiquitination, mitochondrial pathways, cellular response to stress, lysosomal proteins, protein metabolism (localization, modification, transport, folding and stability), RNA processing, aromatic compound metabolism, vesicle mediated transport and exocytosis, and cellular catabolic processes. Interestingly, the DEGs include genes of the KEGG-PD pathway and the CHCHD2 gene, which was recently identified as a PD-associated gene and named PARK22.

To investigate further how the identified network relates to other known PD mechanisms, PD-associated genes, also known as PARK genes (Supplementary Table 9, Fig. 7), were added to the DEG network. Next, PARK-PARK interactions were removed and only PARK-DEG interactions were retained to test how PARK genes integrate into the network. All 19 protein-coding PARK genes interact directly with at least one, but usually several DEGs (Supplementary Fig. 7). The degree of interaction of PARK genes with the DEGs of the network is illustrated by coloring (in pink) DEGs that directly interact with a PARK gene. The darker
the color, the greater the number of PARK genes the DEG interacts with. The central nodes of the network generally interact with several PARK genes, suggesting that they play a central role in linking the PARK genes to the network, but also that PARK genes may mediate PD pathology through a few central pathways of this network, and that the effects of different PARK genes converge on the same set of pathways.

Further analysis revealed that a large number of the DEGs interact with genes associated with mitochondria or ubiquitination. For this analysis, we used BioGRID \textsuperscript{44,65} to identify interactions with mitochondrial or ubiquitination proteins for the top 172 DEGs of groups A-C (Fig. 6f). These interactions were used to create a network illustrating that many of the DEGs in our study directly interact with genes involved in mitochondrial function and in ubiquitination. Thereby, only direct DEG – mitochondrial gene or DEG – ubiquitination gene interactions were included and PARK genes were added for reference (Supplementary Fig. 8).

Based on manual literature search, we determined that at least 68% of the DEGs (174 of 255 genes, not including pseudogenes and RNA genes) are already directly associated with PD, either experimentally, or as significant links in GWAS-PD, or by PD expression studies (Fig. 7, Supplementary Table 10). This is particularly true for the major nodes of the network (Supplementary Fig. 9).

**Discussion**

The aim of this study was to identify genes that were differentially expressed as a result of a mutation in the PINK1 gene, using mDA neurons differentiated from patient-derived iPSCs, a model relevant to PD. We focused on cells undergoing neural differentiation, as these are not expected to display the activation of damage control pathways induced by neurotoxicity, but are likely to identify pathways that lead to primary pathology of PD. Gene expression alterations in aging neurons as a result of the modification identified here will be the next important step to investigate the manifestation of neurodegeneration.

The single-cell expression data were analyzed in several layers. First, we identified the most strongly DEGs (group A), consistently altered in the same direction at all four timepoints including iPSCs (Fig. 6, Table 2, Supplementary Fig. 5). By choosing four different timepoints along the differentiation path and selecting only genes whose differential expression was consistent at all timepoints, we excluded pathways associated with mDA differentiation, as these were expected to change significantly during the different steps of the differentiation process. In our first analysis, we included iPSCs, in order to identify strongly DEGs due to the presence of the PINK1 mutation, regardless of cell type (Table 2, Group A “incl. iPSCs”). Our reasoning was that since iPSCs differ significantly from neural precursors and are not expected to express neuronal pathways, their inclusion would bias against the selection of neuronal pathways. A corresponding approach in which iPSCs were excluded resulted in an expanded gene list that included genes more likely dysregulated specifically in a neural cell type (Table 2) (Supplementary Fig. 5, Group B). Overall, we identified 172 DEGs (incl. iPSCs, group C, Supplementary Table 5), and a total of 285 DEGs (without iPSCs, group D, Supplementary Table 6 and Fig. 5).
Creating a protein-protein interaction network based on these groups of DEGs demonstrated that genes in Group D also formed important nodes within the interaction network and were frequently associated with PD (Supplementary, Fig. 5). The resulting network illustrates that in spite of the very different nature of PD-associated genes, there is interconnectedness among the molecular pathways through which they mediate PD pathology. It also suggests that any one mutation leads to pathology via several molecular paths. This underscores the role the genetic background plays in PD penetrance and severity, as alleles of several network genes may reduce or amplify the effect of a mutation\textsuperscript{7,17}.

The complexity of genetic interactions in PD is well illustrated by the interaction of PINK1 and Parkin. PINK1 is known to interact with Parkin directly, however, our data indicates that the presence of the PINK1 mutation results in the dysregulation of several other genes that are possibly upstream of Parkin\textsuperscript{66}, including HNRNPC\textsuperscript{67}, MTRNR2L1\textsuperscript{68}, MYL12A and SLC25A4\textsuperscript{69}, as well as LMAN1, a membrane mannose-binding lectin, which was shown to play a role in Parkin translocation\textsuperscript{70}. This suggests that the direct interaction between PINK1 and Parkin is not the only means by which PINK1 interacts with the Parkin pathway.

Analysis of the networks shows that certain DEGs are a point of convergence between pathways and form major nodes. These DEGs seem to play a central role within the network, specifically CUL3, HSPA8, EEF1A1, UQCRFS1, CNTNAR2, PSMA4, HNRNPC, and PLCB4, but also EGLN3, IPO5, IPO7, PALLD, PGD, RALGPS2, CYCS, SHH, BRCA2 and others (Supplementary Fig. 5). In fact, these genes play key roles in PD pathology (Table 3, Supplementary Fig. 9). Hence, the network derived from our analysis of a PINK1 mutation is revealing the convergence of many known, key PD-associated pathways. This convergence suggests that different mutations may feed into the same PD pathology-associated routes. These central pathways include several genes (listed in Table 3), such as CUL3, which has been linked to PD by GWAS studies and is considered a potential PD drug target\textsuperscript{71}. HSPA8 (also known as HSP73 and HSC70), which disaggregates α-synuclein amyloid fibrils and plays a role in autophagy and the catabolic pathway for α-synuclein, mediates mitophagy by regulating the stability of PINK1, and its expression was shown to be impaired in sporadic PD\textsuperscript{72–75}. It is also one of the most highly dysregulated genes in our dataset. EEF1A1 mediates activation of heat-shock transcription factor HSF1, a key player in PD\textsuperscript{76}, and prevent α-synuclein aggregation, as well as interacts with Parkin (PARK2) and HTRA2 (PARK13)\textsuperscript{65,77}. UQCRFS1 is a mitochondrial electron transport chain ubiquinol-cytochrome c reductase\textsuperscript{78}, a member of the KEGG-PD pathway (Entry K00411\textsuperscript{79,80}), and has been identified as a PD risk gene\textsuperscript{81}. CNTNAP2, which belongs to the neurexin superfamily, plays a role in triggering protein aggregates\textsuperscript{82,83}, was found to be significantly differentially expressed in the blood of PD patients with LRRK2 mutation\textsuperscript{84}, and was also associated with PD by GWAS\textsuperscript{46}. PSMA4, a proteasome subunit, is part of the KEGG-PD pathway (hsa05012, bta05012, K02728)\textsuperscript{79,80} and is a member of the ubiquitin-proteasomal pathway, which plays a key role in Parkinson's disease\textsuperscript{85}. It also interacts with Parkin (PARK2) and FBXO7 (PARK15)\textsuperscript{65}. HNRNPC interacts with both PARK2 and members of the Poly (ADP-ribose)-dependent cell death pathway implicated in PD\textsuperscript{67}. PLCB4 has been linked to PD\textsuperscript{46} and knock-out mice show motor defects consistent with ataxia\textsuperscript{86}.
These central nodes interact with several PARK genes (Supplementary Table 9), so it is possible that PD-associated mutations converge on the same main pathways, which may play a central role in PD pathology.

Currently, mutations in 23 genes or loci are linked to PD. In addition to their official names, they are named PARK1-PARK23 (Supplement Table 9). Of these, 19 are protein-coding genes. The DEGs we have identified in this study directly interact with all 19 protein-coding PARK genes (Supplementary Fig. 7). In addition, the CHCHD2 gene (PARK22) was identified as one of the DEGs. This further supports the hypothesis that perturbations due to the mutations in PARK genes may converge on the same network, which may then be responsible for the PD phenotype. This would explain why mutations in so many genes lead to a common or at least similar PD pathology.

It will be of great interest to see if cells from idiopathic patients show dysregulation of this integrated network. In fact, our analysis has identified genes which had no known connection to molecular mechanisms underlying PD pathology, though some of them are known to be involved in sporadic PD. Knowing how they integrate into the network may point to the underlying mechanism of how they cause PD pathology. For example, one of the top DEGs is LGI1. The development of antibodies to this protein leads to immunomodulated parkinsonism, yet there is no known mechanism linking it to PD pathology. In the network LGI1 has several nearest DEG neighbors it interacts with (Supplementary Fig. 10). Its most significant interaction is its co-expression with CNTNAP2, which is part of the neurexin family and is required for axon organization, and MGMT, which repairs the methylated nucleobase in DNA. From GeneMANIA alone, the strongest evidence is for interaction with GOLT1B, which plays a role in Golgi transport. Hence, LGI1-associated pathology leading to PD symptoms may be mediated through pathways which are also dysregulated by the presence of the PINK1 mutation. CNTNAP2 is another very good candidate, as it was shown to be dysregulated in PD patients carrying a mutation in the LRRK2 gene, providing additional evidence that it likely plays a role in PD pathology.

Another DEG in our dataset is CHCHD2, the PARK22 gene. Mutations in this gene are linked with autosomal dominant PD, although the precise mechanism is unknown. One hypothesis was that CHCHD2 colocalizes with the mitochondrial contact site and cristae organizing system (MICOS). However, in the DEG network, CHCHD2 has strong interactions with at least three other DEGs, SLC25A4/ANT1, GHITM, and NME4. Evidence suggests that GHITM plays a role in PINK1-mediated neurodegeneration and NME4 was shown to be downregulated in PD. SLC25A4 (also known as ANT1) plays an essential role in mitophagy and has been linked to PD pathology. Hence, the interaction of CHCHD2 with SLC25A4, GHITM, and NME4 may be relevant in mediating pathological changes in CHCHD2-associated PD.

It is of note that pathways known to play a key role in PD are profoundly and consistently dysregulated at all time points examined, far before the onset of PD pathology. For example, CHCHD2 is part of the purine metabolic pathway that produces DNA, RNA, nucleosides and nucleotides and has been shown to be
altered in PD \textsuperscript{58–61}. The DEG network illustrates that the expression of a large number of interconnected aromatic compound metabolic pathway genes is altered in cells carrying the PINK1 mutation (Supplementary Fig.6b). In total, 39 genes of the nitrogen compound metabolic process (Ncmp) and 88 genes specific to the aromatic compound metabolic process (Acmp, a subgroup of the Ncmp) are differentially expressed in our dataset (Supplementary Table 8). Many of the DEGs identified in our study are part of more than one pathway and, therefore, interconnect the various pathways known to play a role in PD, including stress and catabolic processes \textsuperscript{55,56}, aromatic compound metabolism \textsuperscript{58}, vesicle-mediated transport and exocytosis \textsuperscript{62}, RNA metabolism \textsuperscript{57}, protein transport, localization, folding, stability, and ubiquitination \textsuperscript{53} (Supplementary Fig. 6 and Table 8). This confirms observations that PD pathology involves many different pathways \textsuperscript{93}.

Additionally, when a manual literature search was performed, 68% of the DEGs were found to be associated with PD (Fig. 7, Supplementary Table 10), which is particularly true for the central nodes of the network (Table 3, Supplementary Fig. 9). This indicates that these genes may be key points of integration of the effects of PD pathology, an idea further substantiated by the convergence of the added PARK genes on these nodes (Supplementary Fig. 7). Furthermore, these nodes form a link between different functional pathways. In particular, this is true for CUL3, HSPA8 and PSMA4 (Supplementary Fig. 6 and Table 8).

We have also analyzed the correlation of expression between various gene pairs. This correlation may indicate that they are a target of the same regulatory pathway or otherwise related. In our dataset, the expression of several interaction partners shows high correlation, namely PLCB4-RALGPS-TTC3-ZNF37A, EIF3B-HSPA8 (major DEG network node, ubiquitination pathway)-PCBP1 (ubiquitination pathway). Another cluster centers on the MT-CYB gene and involves both mitochondrial and ubiquitination pathways by NME1-MT-CYB – MT-ND5 – MT-CO3 – MRPS21 interactions. Among the top pairs are also PSMD7-PSMB5, TAGLN-MYL9 and VWA5A-ZCCHC11 (Figure 6d). The interactions of these genes may, therefore, play a key role in PINK1-mediated PD pathology.

The fact that so many genes which belong to other PD mutation-related pathways were dysregulated by the presence of the PINK1 mutation suggests that pathways involved in PD pathology are far more interconnected than previously thought. It is likely that PD pathology is more a disease with a characteristic network fingerprint, than a disease caused by independent defects (mutations) (Fig. 6). This and future studies will, hopefully, provide a picture of how various mutations feed into this network and cause its dysregulation. If idiopathic PD is shown to also be based on dysregulation of this network, then we may finally be able to understand the cause of idiopathic PD, which represents 80-85\% of all PD cases \textsuperscript{2}.

**Methods**

**Generation of iPSC cell lines**
Fibroblasts (cat. No. ND40066) isolated from a 64-year-old male with PD symptom onset at 33 years of age who carried a homozygous mutation for the ILE368ASN (P.I368N/P.I368N) mutation in the PINK1 gene were obtained from the Coriell Institute (Cat. No. ND40066). Samples were collected in accordance with the US Government guidelines and are subject to an MTA issued by Coriell Institute for Medical Research NINDS Cell Repository. Fibroblasts were cultured on gelatin-coated plates (10% gelatin in PBS, coated for 10 min at room temperature) in KO DMEM +10% FBS + 1% penicillin/streptomycin (stock was 10,000 units penicillin and 10 mg streptomycin/mL) at standard culture conditions (37°C, 5% CO2).

Live adherent fibroblasts in culture media were sent to be karyotyped by Cell Line Genetics, Madison, WI, USA (Supplementary Fig. 1) and confirmed to have a normal karyotype. The reprogramming of fibroblasts into pluripotent stem cells was done according to the methods outlined by Takahashi and Yamanaka (2006) using the Epi5™ Episomal iPSC Reprogramming Kit (ThermoFisher Scientific, Waltham, MA). The iPSC clone was again analyzed using Array Comparative Genomic Hybridization (aCGH), a high-resolution karyotype analysis for the detection of unbalanced structural and numerical chromosomal alterations and confirmed to be normal and to carry the PINK1 mutation (ILE368ASN, P.I368N/P.I368N)(Supplemental Fig. 2). The resulting iPSC cell lines were maintained on Geltrex matrix (Gibco) in mTeSR™1 media (StemCell Technologies) under standard incubator conditions of 5% CO2 and humidity. The protocol was approved by the Committee on Human Research at the University of California San Francisco.

**Analysis of iPSC status and trilineage potential by TaqMan iPSC Scorecard Assay.**

We also performed a TaqMan iPSC Scorecard Assay, which confirmed their iPSC status and their trilineage potential (Fig. 2b). The protocol was followed as described by the manufacturer of the TaqMan hPSC Scorecard Assay (ThermoFisher Scientific).

**Immunocytochemistry**

Adherent colonies were fixed in 4% PFA for 10 min, washed and permeabilized with 0.1% Triton X-100 in 1X PBS for 15 min, then washed and incubated in a blocking solution of 2% BSA in 1X PBS for 1 hour. They were then incubated with a primary antibody for POU5F1 (also known as Oct3/4, R&D systems MAB1759) and Tra-1-60 (MAB4360, Merck Millipore) at 1/500 dilution in blocking solution, overnight at 4°C (Fig. 2a). The next day they were washed 3 times with PBS and a secondary antibody (AlexaFluor 488, ThermoFisher) was applied at a 1/1000 dilution in blocking solution and incubated for 1 hour at room temperature. The cells were then washed 3 times with PBS and imaged. Differentiated cells were stained for microtubule-associated protein 2 (MAP2, MAB3418, Merck Millipore), tyrosine hydroxylase (TH, Pel-Freez Biologicals P40101-150), Pax6 (901301, Imtec diagnostics) at 1/500 dilution, and GDAP1L1 (1:250, TA503153, Origene).

**Differentiation of iPSCs into mDA neurons**
The protocol used to differentiate iPSCs into mDA neurons was modified from \textsuperscript{16,94} (Table 1). The iPSCs were grown to 95% confluence, dissociated using accutase, and 1.5 wells were combined into one well at day -1. They were allowed to recover in the presence of ROCK inhibitor for about 8 hours and then in mTeSR without ROCK inhibitor for about 16 hours. After this, day 0 media were applied (Table 1). Cells were fed fresh media daily, 36 ml per 6-well plate or as needed, judging consumption from media colour and replacing media whenever it turned yellow, using the appropriate media and factor mix for that day.

**Real-time quantitative PCR of mDA and non-mDA markers.**

Total RNA was extracted from a cell pellet of a 12-well plate well using the RNeasy Plus Universal Kit Mini (50), Catalog no. 73404), as per manufacturer instructions. The RNA concentration was determined through absorption at 260 nm using the Nanodrop instrument (Fisher Scientific). The Superscript IIITM First-Strand Synthesis System for RT-PCR (Invitrogen) was used to prepare cDNA, using oligo(dT)20 and 2 ug of total RNA as per manufacturer instructions. The cDNA was stored at -20°C.

Primers were designed using Primer Blast\textsuperscript{95} and synthesized by Eurogentec Belgium). The primers used are listed in Supplementary Table 3. Standard templates of 90–150 bp in length were generated by PCR, purified using Invitrogen Pure Link PCR Micro Kit (K310050) and their concentration determined using NanoDrop Spectrophotometer. The cDNA levels within samples were determined using quantitative real-time PCR (QRT-PCR) on a Roche Lightcycler 480 using the Maxima® SYBR Green/ROX qPCR Master Mix (2×) (cat. #K0223). This was followed by a dissociation curve to confirm that only one PCR product was present. The values were standardized per expression of the housekeeping gene GAPDH.

**Statistics**

In real-time qPCR graphs, each timepoint consists of three independently differentiated samples of the ND40066 (PINK1) cell line (n=3) and the control (n=3). Each of the samples was amplified in triplicate. Each sample value was an average of the experimental triplicate. Standard error was calculated using GraphPad, based on n=3 independently differentiated samples for each cell line at each timepoint.

**Single-cell RNA seq**

On the day of collection, cells were dissociated using accutase. The single-cell suspension was spun down and cells were washed with (PBS, 2%BSA) twice, then passed through a 40mm filter to remove larger cell clumps. The sample was then counted and viability was determined using (Vi-CELL XR, Cell Counter, Beckman Coulter). Cells were required to have at least a 90% viability. Most samples showed 95% viability. They were then diluted in PBS with 2%BSA to a final concentration of 190,000 cells/ml. 3ml were then used for single-cell analysis. Subsequently, cells were processed by the DropSeq approach as described previously \textsuperscript{27,96,97} and sequenced.

**NGS preparation for Drop-seq libraries**
The 3’ end-enriched cDNA libraries were prepared by tagmentation reaction of 600 pg cDNA library using the standard Nextera XT tagmentation kit (Illumina). Reactions were performed according to the manufacturer's instructions. The PCR amplification cycling programme used was 95°C 30 s, and twelve cycles of 95°C (10 s), 55°C (30 s) and 72°C (30 s), followed by a final extension step of 72°C (5 min). Libraries were purified twice to reduce primers and short DNA fragments with 0.6× and 1× Agencourt AMPure XP beads (Beckman Coulter), respectively, in accordance with the manufacturer's protocol. Finally, purified libraries were eluted in 10 μl Molecular Grade Water. Quality and quantity of the tagmented cDNA library were evaluated using Bioanalyzer High Sensitivity DNA Chip. The average size of the tagmented libraries prior to sequencing was between 400 and 700 bps.

Purified Drop-seq cDNA libraries were sequenced using Illumina NextSeq 500 with the recommended sequencing protocol except for 6pM of custom primer (GCCTGTCCGGAAGCAGTGGTATCAACGCAGCAGTAC) applied for priming of read 1. Paired-end sequencing of 20 bases (covering the 1–12 bases of random cell barcode and the remaining 13–20 bases of random unique molecular identifier (UMI)) was performed for read 1, and of 50 bases of the genes for read 2.

Bioinformatics processing and data analysis

The FASTQ files were assembled from the raw BCL files using Illumina's bcl2fastq converter and run through the FASTQC codes (Babraham bioinformatics; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to check for consistency in library qualities. The monitored quality assessment parameters monitored were (i) per base sequence quality (especially for the read 2 of the gene), (ii) per base N content, (iii) per base sequence content, and (iv) over-represented sequences. The FASTQ files were then merged and converted into binaries using PICARD’s FastqToSam algorithm. The sequencing reads were converted into a digital gene expression matrix using the Drop-seq bioinformatics pipeline.

Single-cell RNA-seq data analysis

The identification of low quality cells was done separately for each data set. In order to select only the highest quality data, we sorted the cells by their cumulative gene expression. Only cells with the highest cumulative expression were considered for the analysis.

In addition to this filtering we defined cells as low-quality based on three criteria for each cell. The number of expressed genes must be more than 200 and 2 median-absolute- deviations (MADs) above the median; the total number of counts has to be 2 MADs above or below the median, and the percentage of counts to mitochondrial genes has to be 1.5 MADs above the median. Cells failing at least one criteria were considered as low quality cells and filtered out from further analysis. Similar to the cell filtering, we filtered out low quality genes, identified by being expressed in less than 10 cells in the data.

The integration of the filtered matrices of the different datasets was performed using scTransform on a Seurat object based on the treatment. The final gene expression matrix, which was used for the
downstream analysis, consisted of 4495 cells and 39,194 genes with a median total number of mRNA counts of 7750 and a median number of expressed genes of 3521. Principal component analysis (PCA) was computed using the 5000 most variable genes of the integrated data. The clustering of data was performed using Louvain clustering. The resolution of the clustering was selected based on the best silhouette score of the different resolutions. A short list of manually curated markers was used to validate the different stages of the differentiation process.

We then performed differential expression analysis between the two treatments (control and PINK1) at each time point. The differential expression analysis was done using MAST on the normalized counts using the total number of transcripts in each cell as a covariate and the Bonferroni correction to correct for multiple hypothesis testing (Padj). In addition, we tried to find conserved markers among the different time points using MAST again and the total number of transcripts in each cell as a latent variable. Genes with fold changes of the same sign in the fold change were then identified across the different time points and the average fold change was calculated. The genes with average fold change > 0.1 and maximum adjusted p-value < 0.01 were selected as differentially expressed.

**Network analysis**

We extracted protein-protein interaction information between the DEGs from STRING and from GeneMANIA. We excluded indirect association, such as text mining, co-occurrence and neighborhood from STRING, and coexpression, colocalization, shared protein domains and predicted interactions from GeneMANIA, retaining only genetic interactions, pathways and physical interactions (2122 interactions in total). We deletes any any genes or interactions that were added by these databases, to focus only on DEGs and interactions among them. The network diameter was calculated and betweenness centrality was used to illustrate the relative importance of each node within the network. As a control, we selected the same number of genes at random, using the list of genes detected by our RNA-seq analysis, excluding DEGs. This control set did not produce a network and led to a mostly disconnected array of genes (Supplementary Fig. 11). Networks were also generated using the STRING and GeneMANIA inputs independently (Supplementary Fig. 5 a, b). We constructed a correlation network based on the correlation of expression of DEGs (p-value < 0.05, correlation > 0.1) and identified edges that are common to the two networks. This network consisted of 860 interactions (Fig. 6). We then extracted shared interactions of these two networks, which amounted to 297 interactions (Fig. 6).

In order to validate the PPI network produced by STIRNGdb (v10), we created 50 PPI (protein protein interaction) networks using 292 random genes (same as the number of DEGs). We then compared the number of detected proteins, the number of interactions between the genes and the distribution of the node degrees. We performed the Wilcoxon test to access if the two-degree distributions are different from one another in a statistically significant manner, which showed a statistically significant difference (p=2.22e-16) (Fig.5b,c.).

**Data availability.** Single cell RNAseq data will be uploaded to GEO upon acceptance of the manuscript.
**Ethics.** Patient-derived cell lines were handled according to the ethics guidelines set out by the National Ethics Board of Luxembourg, (Comité National d’Ethique dans la Recherche; CNER).

**Declarations**

**Acknowledgements**

We thank Dr. Rudi Balling for stimulating discussions. We would like to acknowledge the excellent support by the LCSB Bio-Imaging Facility through Aymeric d’Hérouël and the outstanding service by the LCSB-Sequencing Platform through Rashi Halder.

D.K. and M.B. were financially supported by the PRIDE program of the Luxembourg National Research Fund through PRIDE17/12244779/PARK-QC and PRIDE/10907093/CRITICS, respectively. This work was also made possible in part by support to S.F. from the Michael J Fox Foundation through Head Start Program and Parkin Consortium grants and from the National Institutes on Aging (RF1 AG058476 and P01 AG54407).

**Author Contributions** (Articles only)

GN and AS designed the project. GN generated the iPSC cell lines and performed iPSC characterization, as well as mDA differentiation, sample preparation for Single Cell RNA Seq, qPCR experiments and ICC. KG performed single cell analysis RNAseq experiments. DK designed the bioinformatic analysis pipeline, performed the bioinformatic analysis and contributed article figures. MB artistically rendered Figures 1 and 2, assisted with cell culture and supplied PAX6 staining figure. GN designed and performed the network analysis, generated Tables and Supplementary Figures. GN and AS supervised the work and wrote the manuscript. The iPSC cell lines were generated under the supervision of S.F., who also helped edit the manuscript.

**Competing Interests statement.**

Authors declare no competing interests.

**References**

1. Mayeux, R. Epidemiology of neurodegeneration. *Annu. Rev. Neurosci.* **26**, 81–104 (2003).
2. Klein, C. & Westenberger, A. Genetics of Parkinson’s disease. *Cold Spring Harb. Perspect. Med.* (2012) doi:10.1101/cshperspect.a008888.
3. Abbas, M. M., Xu, Z. & Tan, L. C. S. Epidemiology of Parkinson’s Disease—East Versus West. *Movement Disorders Clinical Practice* (2018) doi:10.1002/mdc3.12568.
4. Chaudhuri, K. R., Healy, D. G. & Schapira, A. H. V. Non-motor symptoms of Parkinson’s disease: Diagnosis and management. *Lancet Neurology* (2006) doi:10.1016/S1474-4422(06)70373-8.
5. Verschuur, C. V. M. *et al.* Randomized delayed-start trial of levodopa in Parkinson's disease. *N. Engl. J. Med.* (2019) doi:10.1056/NEJMoA1809983.

6. Larsen, S. B., Hanss, Z. & Krüger, R. The genetic architecture of mitochondrial dysfunction in Parkinson's disease. *Cell and Tissue Research* (2018) doi:10.1007/s00441-017-2768-8.

7. Schneider, S. A. & Klein, C. *PINK1 Type of Young-Onset Parkinson Disease. GeneReviews®* (1993).

8. Dagda, R. K. *et al.* Beyond the mitochondrion: Cytosolic PINK1 remodels dendrites through Protein Kinase A. *J. Neurochem.* (2014) doi:10.1111/jnc.12494.

9. Furlong, R. M. *et al.* The Parkinson's disease gene PINK1 activates Akt via PINK1 kinase-dependent regulation of the phospholipid PI(3,4,5)P3. *J. Cell Sci.* **132**, (2019).

10. Pryde, K. R., Smith, H. L., Chau, K. Y. & Schapira, A. H. V. PINK1 disables the anti-fission machinery to segregate damaged mitochondria for mitophagy. *J. Cell Biol.* (2016) doi:10.1083/jcb.201509003.

11. Giguère, N., Nanni, S. B. & Trudeau, L. E. On cell loss and selective vulnerability of neuronal populations in Parkinson's disease. *Frontiers in Neurology* (2018) doi:10.3389/fneur.2018.00455.

12. Le, W., Sayana, P. & Jankovic, J. Animal Models of Parkinson's Disease: A Gateway to Therapeutics? *Neurotherapeutics* (2014) doi:10.1007/s13311-013-0234-1.

13. Konnova, E. A. & Swanberg, M. *Parkinson's Disease: Pathogenesis and Clinical Aspects.* (Codon Publications, 2018). doi:10.15586/codonpublications.parkinsonsdisease.2018.ch5.

14. Tanabe, K., Takahashi, K. & Yamanaka, S. Induction of pluripotency by defined factors. *Proceedings of the Japan Academy Series B: Physical and Biological Sciences* (2014) doi:10.2183/pjab.90.83.

15. Kriks, S. *et al.* Floor plate-derived dopamine neurons from hESCs efficiently engraft in animal models of PD. *Nature* (2012) doi:10.1038/nature10648.Floor.

16. Kriks, S. *et al.* Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease. *Nature* (2011) doi:10.1038/nature10648.

17. Zanon, A., Pramstaller, P. P., Hicks, A. A. & Pichler, I. Environmental and genetic variables influencing mitochondrial health and Parkinson's disease penetrance. *Parkinson's Disease* (2018) doi:10.1155/2018/8684906.

18. Arenas, E., Denham, M. & Villaescusa, J. C. How to make a midbrain dopaminergic neuron. *Dev.* (2015) doi:10.1242/dev.097394.

19. Hegarty, S. V., Sullivan, A. M. & O'Keeffe, G. W. Midbrain dopaminergic neurons: A review of the molecular circuitry that regulates their development. *Developmental Biology* (2013) doi:10.1016/j.ydbio.2013.04.014.

20. Anderegg, A., Poulin, J. F. & Awatramani, R. Molecular heterogeneity of midbrain dopaminergic neurons - Moving toward single cell resolution. *FEBS Letters* (2015) doi:10.1016/j.feslets.2015.10.022.

21. Wurst, W., Bally-Cuif, L. & Bally-Cuif, L. Neural plate patterning: Upstream and downstream of the isthmic organizer. *Nat. Rev. Neurosci.* (2001) doi:10.1038/35053516.
22. Corbin, J. G., Rutlin, M., Gaiano, N. & Fishell, G. Combinatorial function of the homeodomain proteins Nkx2.1 and Gsh2 in ventral telencephalic patterning. *Development* (2003) doi:10.1242/dev.00717.

23. Blaess, S. & Ang, S. L. Genetic control of midbrain dopaminergic neuron development. *Wiley Interdisciplinary Reviews: Developmental Biology* (2015) doi:10.1002/wdev.169.

24. Przybilla, J., Rohlf, T., Loeffler, M. & Galle, J. Understanding epigenetic changes in aging stem cells - a computational model approach. *Aging Cell* (2014) doi:10.1111/acel.12177.

25. Bonilla, S. *et al.* Identification of midbrain floor plate radial glia-like cells as dopaminergic progenitors. *Glia* (2008) doi:10.1002/glia.20654.

26. Roeper, J. Dissecting the diversity of midbrain dopamine neurons. *Trends in Neurosciences* (2013) doi:10.1016/j.tins.2013.03.003.

27. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* (2015) doi:10.1016/j.cell.2015.05.002.

28. Schöndorf, D. C. *et al.* IPSC-derived neurons from GBA1-associated Parkinson's disease patients show autophagic defects and impaired calcium homeostasis. *Nat. Commun.* (2014) doi:10.1038/ncomms5028.

29. Hough, S. R., Laslett, A. L., Grimmond, S. B., Kolle, G. & Pera, M. F. A continuum of cell states spans pluripotency and lineage commitment in human embryonic stem cells. *PLoS One* (2009) doi:10.1371/journal.pone.0007708.

30. Emani, M. R. *et al.* The L1TD1 protein interactome reveals the importance of post-transcriptional regulation in human pluripotency. *Stem Cell Reports* (2015) doi:10.1016/j.stemcr.2015.01.014.

31. Lund, R. J. *et al.* RNA Polymerase III Subunit POLR3G Regulates Specific Subsets of PolyA+ and SmallRNA Transcriptomes and Splicing in Human Pluripotent Stem Cells. *Stem Cell Reports* (2017) doi:10.1016/j.stemcr.2017.04.016.

32. Liu, Q. *et al.* The miR-590/Acvr2a/Terf1 Axis Regulates Telomere Elongation and Pluripotency of Mouse iPSCs. *Stem Cell Reports* (2018) doi:10.1016/j.stemcr.2018.05.008.

33. Suresh, B., Lee, J., Kim, H. & Ramakrishna, S. Regulation of pluripotency and differentiation by deubiquitinating enzymes. *Cell Death and Differentiation* (2016) doi:10.1038/cdd.2016.53.

34. Fu, Y. *et al.* IFITM1 suppresses expression of human endogenous retroviruses in human embryonic stem cells. *FEBS Open Bio* (2017) doi:10.1002/2211-5463.12246.

35. Madan, B. *et al.* The Pluripotency-Associated Gene Dppa4 Is Dispensable for Embryonic Stem Cell Identity and Germ Cell Development but Essential for Embryogenesis. *Mol. Cell. Biol.* (2009) doi:10.1128/mcb.01970-08.

36. Kwon, S. C. *et al.* The RNA-binding protein repertoire of embryonic stem cells. *Nat. Struct. Mol. Biol.* (2013) doi:10.1038/nsmb.2638.

37. Ang, S. L. Transcriptional control of midbrain dopaminergic neuron development. *Development* (2006) doi:10.1242/dev.02501.
38. Bye, C. R., Rytova, V., Alsanie, W. F., Parish, C. L. & Thompson, L. H. Axonal Growth of Midbrain Dopamine Neurons is Modulated by the Cell Adhesion Molecule ALCAM Through Trans-Heterophilic Interactions with L1cam, Chl1, and Semaphorins. *J. Neurosci.* (2019) doi:10.1523/JNEUROSCI.0278-19.2019.

39. Hoekstra, E. J. *et al.* Lmx1a Encodes a Rostral Set of Mesodiencephalic Dopaminergic Neurons Marked by the Wnt/B-Catenin Signaling Activator R-spondin 2. *PLoS One* (2013) doi:10.1371/journal.pone.0074049.

40. Mesman, S. & Smidt, M. P. Tcf12 is involved in early cell-fate determination and subset specification of midbrain dopamine neurons. *Front. Mol. Neurosci.* (2017) doi:10.3389/fnmol.2017.00353.

41. Raina, A., Mahajani, S., Bähr, M. & Kügler, S. Neuronal Trans-differentiation by Transcription Factors Ascl1 and Nurr1: Induction of a Dopaminergic Neurotransmitter Phenotype in Cortical GABAergic Neurons. *Mol. Neurobiol.* (2020) doi:10.1007/s12035-019-01701-x.

42. Bastian, M., Heymann, S. & Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Third Int. AAAI Conf. Weblogs Soc. Media* (2009) doi:10.1136/qshc.2004.010033.

43. Szklarczyk, D. *et al.* STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* (2019) doi:10.1093/nar/gky1131.

44. Warde-Farley, D. *et al.* The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* (2010) doi:10.1093/nar/gkq537.

45. Qiu, B. *et al.* DJ-1 promotes development of DEN-induced hepatocellular carcinoma and proliferation of liver cancer cells. *Oncotarget* (2017) doi:10.18632/oncotarget.14293.

46. Rouillard, A. D. *et al.* The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database (Oxford).* (2016) doi:10.1093/database/baw100.

47. Puschmann, A. New Genes Causing Hereditary Parkinson's Disease or Parkinsonism. *Current Neurology and Neuroscience Reports* (2017) doi:10.1007/s11910-017-0780-8.

48. Oliveira, L. M. A. *et al.* Elevated α-synuclein caused by SNCA gene triplication impairs neuronal differentiation and maturation in Parkinson's patient-derived induced pluripotent stem cells. *Cell Death Dis.* (2015) doi:10.1038/cddis.2015.318.

49. Cruz-Monteagudo, M. *et al.* Efficient and biologically relevant consensus strategy for Parkinson's disease gene prioritization. *BMC Med. Genomics* (2016) doi:10.1186/s12920-016-0173-x.

50. Volakakis, N. *et al.* Nurr1 and retinoid X receptor ligands stimulate ret signaling in dopamine neurons and can alleviate α-synuclein disrupted gene expression. *J. Neurosci.* (2015) doi:10.1523/JNEUROSCI.1155-15.2015.

51. Zhang, Q. S. *et al.* Beta-asarone protects against MPTP-induced Parkinson's disease via regulating long non-coding RNA MALAT1 and inhibiting α-synuclein protein expression. *Biomed. Pharmacother.* (2016) doi:10.1016/j.biopharma.2016.06.017.
52. Cai, R. et al. Enhancing glycolysis attenuates Parkinson's disease progression in models and clinical databases. *J. Clin. Invest.* (2019) doi:10.1172/JCI129987.

53. Walden, H. & Muqit, M. M. K. Ubiquitin and Parkinson's disease through the looking glass of genetics. *Biochemical Journal* (2017) doi:10.1042/BCJ20160498.

54. Matheoud, D. et al. Parkinson's Disease-Related Proteins PINK1 and Parkin Repress Mitochondrial Antigen Presentation. *Cell* (2016) doi:10.1016/j.cell.2016.05.039.

55. Puspita, L., Chung, S. Y. & Shim, J. W. Oxidative stress and cellular pathologies in Parkinson's disease. *Molecular Brain* (2017) doi:10.1186/s13041-017-0340-9.

56. Lin, K. J. et al. The overcrowded crossroads: Mitochondria, alpha-synuclein, and the endo-lysosomal system interaction in Parkinson's disease. *International Journal of Molecular Sciences* (2019) doi:10.3390/ijms20215312.

57. Martin, I. Decoding Parkinson's Disease Pathogenesis: The Role of Deregulated mRNA Translation. *Journal of Parkinson's Disease* (2016) doi:10.3233/JPD-150738.

58. Garcia-Esparcia, P., Hernández-Ortega, K., Ansoleaga, B., Carmona, M. & Ferrer, I. Purine metabolism gene deregulation in Parkinson's disease. *Neuropathol. Appl. Neurobiol.* (2015) doi:10.1111/nan.12221.

59. Krupke, D. M., Begley, D. A., Sundberg, J. P., Bult, C. J. & Eppig, J. T. The mouse tumor biology database. *Nature Reviews Cancer* (2008) doi:10.1038/nrc2390.

60. Bult, C. J. et al. Mouse Genome Database (MGD) 2019. *Nucleic Acids Res.* (2019) doi:10.1093/nar/gky1056.

61. Finger, J. H. et al. The mouse Gene Expression Database (GXD): 2011 update. *Nucleic Acids Res.* (2011) doi:10.1093/nar/gky1132.

62. Ebanks, K., Lewis, P. A. & Bandopadhyay, R. Vesicular Dysfunction and the Pathogenesis of Parkinson’s Disease: Clues From Genetic Studies. *Front. Neurosci.* 13, (2020).

63. Funayama, M. et al. CHCHD2 mutations in autosomal dominant late-onset Parkinson’s disease: A genome-wide linkage and sequencing study. *Lancet Neurol.* (2015) doi:10.1016/S1474-4422(14)70266-2.

64. Lee, R. G. et al. Early-onset Parkinson disease caused by a mutation in CHCHD2 and mitochondrial dysfunction. *Neurol. Genet.* (2018) doi:10.1212/nxg.0000000000000276.

65. Stark, C. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* (2006) doi:10.1093/nar/gkj109.

66. Zanon, A. et al. Profiling of Parkin-binding partners using tandem affinity purification. *PLoS One* (2013) doi:10.1371/journal.pone.0078648.

67. Lee, Y. et al. Poly (ADP-ribose) in the pathogenesis of Parkinson's disease. *BMB Reports* (2014) doi:10.5483/BMBRep.2014.47.8.119.

68. Guo, B. et al. Humanin peptide suppresses apoptosis by interfering with Bax activation. *Nature* (2003) doi:10.1038/nature01627.
69. Hoshino, A. et al. The ADP/ATP translocase drives mitophagy independent of nucleotide exchange. *Nature* (2019) doi:10.1038/s41586-019-1667-4.

70. Hasson, S. A. et al. High-content genome-wide RNAi screens identify regulators of parkin upstream of mitophagy. *Nature* (2013) doi:10.1038/nature12748.

71. Canning, P. & Bullock, A. N. New strategies to inhibit KEAP1 and the Cul3-based E3 ubiquitin ligases. *Biochem. Soc. Trans.* (2014) doi:10.1042/BST20130215.

72. Gao, X. et al. Human Hsp70 Disaggregase Reverses Parkinson's-Linked α-Synuclein Amyloid Fibrils. *Mol. Cell* (2015) doi:10.1016/j.molcel.2015.07.012.

73. Sala, G. et al. Rotenone down-regulates HSPA8/hsc70 chaperone protein in vitro: A new possible toxic mechanism contributing to Parkinson’s disease. *Neurotoxicology* (2016) doi:10.1016/j.neuro.2016.04.018.

74. Zheng, Q. et al. Hsp70 participates in PINK1-mediated mitophagy by regulating the stability of PINK1. *Neurosci. Lett.* (2018) doi:10.1016/j.neulet.2017.10.051.

75. Mandel, S. et al. Gene expression profiling of sporadic Parkinson's disease substantia nigra pars compacta reveals impairment of ubiquitin-proteasome subunits, SKP1A, aldehyde dehydrogenase, and chaperone HSC-70. in *Annals of the New York Academy of Sciences* (2005). doi:10.1196/annals.1344.031.

76. Liangliang, X. et al. Dominant-positive HSF1 decreases alpha-synuclein level and alpha-synuclein-induced toxicity. *Mol. Biol. Rep.* (2010) doi:10.1007/s11033-009-9623-2.

77. Ekimova, I. V. et al. New HSF1 inducer as a therapeutic agent in a rodent model of Parkinson's disease. *Exp. Neurol.* (2018) doi:10.1016/j.expneurol.2018.04.012.

78. Hernandez, S. M., Tikhonova, E. B. & Karamyshev, A. L. Protein-Protein Interactions in Alpha-Synuclein Biogenesis: New Potential Targets in Parkinson’s Disease. *Front. Aging Neurosci.* 12, (2020).

79. Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* (2000) doi:10.1093/nar/28.1.27.

80. Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. & Tanabe, M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* (2019) doi:10.1093/nar/gky962.

81. Feng, Y. & Wang, X. Systematic analysis of microarray datasets to identify Parkinson’s disease-associated pathways and genes. *Mol. Med. Rep.* (2017) doi:10.3892/mmr.2017.6124.

82. La Cognata, V., Morello, G., D’Agata, V. & Cavallaro, S. Copy number variability in Parkinson’s disease: assembling the puzzle through a systems biology approach. *Human Genetics* (2017) doi:10.1007/s00439-016-1749-4.

83. Varea, O. et al. Synaptic abnormalities and cytoplasmic glutamate receptor aggregates in contactin associated protein-like 2/Caspr2 knockout neurons. *Proc. Natl. Acad. Sci. U. S. A.* (2015) doi:10.1073/pnas.1423205112.
84. Infante, J. *et al.* Identification of candidate genes for Parkinson's disease through blood transcriptome analysis in LRRK2-G2019S carriers, idiopathic cases, and controls. *Neurobiol. Aging* (2015) doi:10.1016/j.neurobiolaging.2014.10.039.

85. Chung, K. K., Dawson, V. L. & Dawson, T. M. The role of the ubiquitin-proteasomal pathway in Parkinson's disease and other neurodegenerative disorders. *Trends in neurosciences* (2001) doi:10.1016/s0166-2236(00)01998-6.

86. Kim, D. *et al.* Phospholipase C isozymes selectively couple to specific neurotransmitter receptors. *Nature* (1997) doi:10.1038/38508.

87. Inamdar, N., Arulmozhi, D., Tandon, A. & Bodhankar, S. Parkinsons Disease: Genetics and Beyond. *Curr. Neuropharmacol.* 5, 99–113 (2007).

88. Kurtis, M. M., Toledano, R., García-Morales, I. & Gil-Nagel, A. Immunomodulated parkinsonism as a presenting symptom of LGI1 antibody encephalitis. *Parkinsonism and Related Disorders* (2015) doi:10.1016/j.parkreldis.2015.08.014.

89. Lin, A., Wang, R. T., Ahn, S., Park, C. C. & Smith, D. J. A genome-wide map of human genetic interactions inferred from radiation hybrid genotypes. *Genome Res.* (2010) doi:10.1101/gr.104216.109.

90. Zhou, W. *et al.* PD-linked CHCHD2 mutations impair CHCHD10 and MICOS complex leading to mitochondria dysfunction. *Hum. Mol. Genet.* (2019) doi:10.1093/hmg/ddy413.

91. Sen, N. E. *et al.* Search for SCA2 blood RNA biomarkers highlights Ataxin-2 as strong modifier of the mitochondrial factor PINK1 levels. *Neurobiol. Dis.* (2016) doi:10.1016/j.nbd.2016.09.002.

92. Martin, L. J., Semenkow, S., Hanaford, A. & Wong, M. The mitochondrial permeability transition pore regulates Parkinson's disease development in mutant α-synuclein transgenic mice. *Neurobiol. Aging* (2014) doi:10.1016/j.neurobiolaging.2013.11.008.

93. Agarwal, D. *et al.* A single-cell atlas of the human substantia nigra reveals cell-specific pathways associated with neurological disorders. *Nat. Commun.* 11, 4183 (2020).

94. Tomishima, M. *StemBook.* (Harvard Stem Cell Institute, 2012). doi:10.3824/stembook.1.70.1.

95. Ye, J. *et al.* Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* (2012) doi:10.1186/1471-2105-13-134.

96. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* (2014) doi:10.1038/nbt.2859.

97. Sousa, C. *et al.* Single-cell transcriptomics reveals distinct inflammation-induced microgli signatures. *EMBO Rep.* (2018) doi:10.15252/embr.201846171.

98. James, N. A. & Matteson, D. S. ecp: An R package for nonparametric multiple change point analysis of multivariate data. *J. Stat. Softw.* (2015) doi:10.18637/jss.v062.i07.

99. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* (2018) doi:10.1038/nbt.4096.
100. Finak, G. et al. MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* (2015) doi:10.1186/s13059-015-0844-5.

101. Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* (1987) doi:10.1016/0377-0427(87)90125-7.

102. Gaare, J. J. et al. Rare genetic variation in mitochondrial pathways influences the risk for Parkinson's disease. *Mov. Disord.* (2018) doi:10.1002/mds.64.

103. Carpanini, S. M., Torvell, M. & Morgan, B. P. Therapeutic inhibition of the complement system in diseases of the central nervous system. *Frontiers in Immunology* (2019) doi:10.3389/fimmu.2019.00362.

104. Kuang, X. L. et al. Reductions of the components of the calreticulin/calnexin quality-control system by proteasome inhibitors and their relevance in a rodent model of Parkinson's disease. *J. Neurosci. Res.* (2014) doi:10.1002/jnr.23413.

105. Klebe, S. et al. The Val158Met COMT polymorphism is a modifier of the age at onset in Parkinson's disease with a sexual dimorphism. *J. Neurol. Neurosurg. Psychiatry* (2013) doi:10.1136/jnnp-2012-304475.

106. Hoeffer, C. A. et al. Altered mTOR signaling and enhanced CYFIP2 expression levels in subjects with fragile X syndrome. *Genes, Brain Behav.* (2012) doi:10.1111/j.1601-183X.2012.00768.x.

107. Tan, C., Liu, X. & Chen, J. Microarray Analysis of the Molecular Mechanism Involved in Parkinson’s Disease. *Parkinsons. Dis.* (2018) doi:10.1155/2018/1590465.

108. Jacobs, F. M. J. et al. Identification of Dlk1, Ptpru and Klhl1 as novel Nurtr1 target genes in mesodiencephalic dopamine neurons. *Development* (2009) doi:10.1242/dev.037556.

109. Lindenbach, D., Conti, M. M., Ostock, C. Y., Dupre, K. B. & Bishop, C. Alterations in primary motor cortex neurotransmission and gene expression in hemi-parkinsonian rats with drug-induced dyskinesia. *Neuroscience* (2015) doi:10.1016/j.neuroscience.2015.09.018.

110. Lavoy, S., Chittoor-Vinod, V. G., Chow, C. Y. & Martin, I. Genetic modifiers of neurodegeneration in a drosophila model of parkinson's disease. *Genetics* (2018) doi:10.1534/genetics.118.301119.

111. Martin-Urdiroy, M., Deeks, M. J., Horton, C. G., Dawe, H. R. & Jourdain, I. The exocyst complex in health and disease. *Frontiers in Cell and Developmental Biology* (2016) doi:10.3389/fcell.2016.00024.

112. Chen, Q., Huang, X. & Li, R. Lncrna MALAT1/miR-205-5p axis regulates MPP+-induced cell apoptosis in MN9d cells by directly targeting LRRK2. *Am. J. Transl. Res.* (2018).

113. Loeffler, D. A., Smith, L. M., Coffey, M. P., Aasly, J. O. & LeWitt, P. A. CSF Nrf2 and HSPA8 in Parkinson's disease patients with and without LRRK2 gene mutations. *J. Neural Transm.* (2016) doi:10.1007/s00702-015-1479-0.

114. Sun, Y. et al. MLF1 is a proapoptotic antagonist of HOP complex-mediated survival. *Biochim. Biophys. Acta - Mol. Cell Res.* (2017) doi:10.1016/j.bbamcr.2017.01.016.
115. Emmanuele, V. et al. A novel mutation in the mitochondrial DNA cytochrome b gene (MTCYB) in a patient with mitochondrial encephalomyopathy, lactic acidosis, and strokelike episodes syndrome. J. Child Neurol. (2013) doi:10.1177/0883073812445787.

116. Odgerel, Z. et al. Whole genome sequencing and rare variant analysis in essential tremor families. PLoS One (2019) doi:10.1371/journal.pone.0220512.

117. Mamik, M. K. & Power, C. Inflammasomes in neurological diseases: Emerging pathogenic and therapeutic concepts. Brain (2017) doi:10.1093/brain/awx133.

118. Itoh, Y. & Voskuhl, R. R. Cell specificity dictates similarities in gene expression in multiple sclerosis, Parkinson’s disease, and Alzheimer’s disease. PLoS One (2017) doi:10.1371/journal.pone.0181349.

119. Sakaue, S. et al. Early-onset parkinsonism in a pedigree with phosphoglycerate kinase deficiency and a heterozygous carrier: do PGK-1 mutations contribute to vulnerability to parkinsonism? npj Park. Dis. (2017) doi:10.1038/s41531-017-0014-4.

120. Morales-Briceño, H. et al. Parkinsonism in PGK1 deficiency implicates the glycolytic pathway in nigrostriatal dysfunction. Park. Relat. Disord. (2019) doi:10.1016/j.parkreldis.2019.04.004.

121. He, B., Bai, Y., Kang, W., Zhang, X. & Jiang, X. LncRNA SNHG5 regulates imatinib resistance in chronic myeloid leukemia via acting as a CeRNA against MiR-205-5p. Am. J. Cancer Res. (2017).

122. Lippolis, R. et al. Altered protein expression pattern in skin fibroblasts from parkin-mutant early-onset Parkinson’s disease patients. Biochim. Biophys. Acta - Mol. Basis Dis. (2015) doi:10.1016/j.bbadis.2015.06.015.

123. Epping, M. T. et al. TSPYL5 suppresses p53 levels and function by physical interaction with USP7. Nat. Cell Biol. (2011) doi:10.1038/ncb2142.

124. Hashimoto, M., Takeda, A., Hsu, L. J., Takenouchi, T. & Masliah, E. Role of cytochrome c as a stimulator of α-synuclein aggregation in Lewy body disease. J. Biol. Chem. (1999) doi:10.1074/jbc.274.41.28849.

125. Rocha, S. et al. Biological implications of differential expression of mitochondrial-shaping proteins in Parkinson’s disease. Antioxidants (2018) doi:10.3390/antiox7010001.

126. Dutta, D. et al. Low Levels of Prohibitin in Substantia Nigra Makes Dopaminergic Neurons Vulnerable in Parkinson’s Disease. Mol. Neurobiol. (2017) doi:10.1007/s12035-016-0328-y.

127. Lee, J. H. et al. Injury-stimulated Sonic hedgehog expression in microglia contributes to neuroinflammatory response in the MPTP model of Parkinson’s disease. Biochem. Biophys. Res. Commun. (2017) doi:10.1016/j.bbrc.2016.11.144.

128. Cheng, X. et al. The BRCC3 regulated by Cdk5 promotes the activation of neuronal NLRP3 inflammasome in Parkinson’s disease models. Biochem. Biophys. Res. Commun. (2020) doi:10.1016/j.bbrc.2019.11.141.

129. Bandres-Ciga, S. et al. The endocytic membrane trafficking pathway plays a major role in the risk of Parkinson’s disease. Mov. Disord. (2019) doi:10.1002/mds.27614.
130. Zhen, Y. & Li, W. Impairment of autophagosome-lysosome fusion in the buff mutant mice with the VPS33AD251E mutation. *Autophagy* (2015) doi:10.1080/15548627.2015.1072669.

131. Fu, J., Menzies, K., Freeman, R. S. & Taubman, M. B. EGLN3 prolyl hydroxylase regulates skeletal muscle differentiation and myogenin protein stability. *J. Biol. Chem.* (2007) doi:10.1074/jbc.M608748200.

132. Tiklová, K. *et al.* Single-cell RNA sequencing reveals midbrain dopamine neuron diversity emerging during mouse brain development. *Nat. Commun.* (2019) doi:10.1038/s41467-019-08453-1.

### Tables

**Table 1.** Key pathways expressed during *in-vivo* mDA differentiation stages. These pathways were used to confirm that the correct mDA differentiation path was followed in vitro. Based on refs 19,23.

| Stage 1 | Stage 2 | Stage 3 | Stage 4 |
|---------|---------|---------|---------|
| Regional specification | Floor Plate VM DA NPs | Differentiating VM DA neurons | Differentiated mDAs |
| OTX2 | LMX1A | NGN2 (Neurog) | SLC18A2 (VMAT2) |
| WNT1 | FOXA2 | NR4A2 (Nurr1) | SLC6A3 (DAT) |
| EN 1/2 | MSX1 | PITX3 | ALDH1A1 |
| LMX1B | | TH | KCNJ6 (GIRK2) |
| | | | SNCA (a-synuclein) |

**Table 2.** The top genes dysregulated consistently in PINK1 vs. control cells across differentiation stages. Pairwise differential expression analysis of each time point (iPSCs, D6, D15, D21), resulted in 14 genes that were upregulated and 13 genes that were downregulated in the PINK1 cell line, compared to control (*p_val* adj $< 0.01$ & $\text{abs(avg\_logFC)} > 0.1$); these genes are marked with “x” in column “Incl. iPSCs”. 29 additional genes were identified in an analysis that included D6, D15, and D21, but not iPSCs. The “Excluded” column explains why a gene was not included in the protein-protein interaction network.
| GENE                  | Incl. iPSCs | Excluded | PD association | Ref. | GENE                  | incl. iPSCs | Excluded | PD association | Ref. |
|----------------------|-------------|----------|----------------|------|----------------------|-------------|----------|----------------|------|
| AC009245.3           | X           | Pseudogene |                |      | ACTN1                | X           |          | PD             | 65   |
| ADGRG7               | X           |          | Rare var., mito | 102  | C6ORF48              | X           |          |                |      |
| BBS2                 | X           |          |                |      | CCDC144NL.AS1        | X           | RNA Gene |                |      |
| CALM2P2              | X           | Pseudogene |                |      | CD59                 | X           |          | PD             | 103  |
| CALR                 | PD          |          |                | 104  | COMT                 | PD          |          |                | 105  |
| CECR1                |             |          |                |      | CRYZ                 | X           |          | GWAS, PD       | 46   |
| CMTM8                |             |          | GWAS, PD       | 46   | CYFIP1               | X           |          | Via mTOR       | 106  |
| EFCAB2               | X           |          | Microarray     | 107  | DLK1                 | PD          |          |                | 108  |
| FOS                  |             |          | Rat, L-DOPA    | 109  | ENAH                 | GWAS, binds LRRK2 |          |                | 110  |
| GOPC                 |             |          | PARK7 inter.   | 46   | EXOC5                | PD Map      |          |                | 111  |
| HNRNPC               | X           |          | Binds Parkin   | 66   | GPC3                 | Reduced in DJ-1 (PARK7) mutant. | 45   |
| MALAT1               | PD          |          |                | 112  | HSPA8                | X           |          | PD, LRRK2      | 113  |
| MINOS1P3             | X           | Pseudogene |                |      | LGI1                 | PD          |          |                | 88   |
| MLF1                 |             |          | Via HtrA2      | 114  | LMAN1                | Parkin translocation | 70   |
| MORF4L1P1            | X           | Pseudogene |                |      | MYL12A               | Binds Parkin | 66   |
| MYC                  | Mito.       |          |                | 115  | MYL6                 | X           |          | Binds LRRK2    | 65   |
| MTRNR2L1             | X           |          | binds Parkin   | 66   | NIPA2                | Tremor      | 116  |
| NAP1L5               |             |          |                |      | OSBPL8               | via ZNF746  | 65   |
| NLRP2                | X           |          | Inflammasome   | 117  | PALLD                | PD          |          |                | 118  |
| PTGR1                |             |          |                |      | PGK1                 | X           |          | PD             | 119,120 |
| RP11.692N5.2         | X           | Pseudogene |                |      | RANBP1               |             |          |                |      |
| RP4.765C7.2          | X           | Pseudogene |                |      | SLC25A4              | binds Parkin | 66   |
| RSRP1                |             |          |                |      | SNHG5                | X           | RNA Gene | miR-205, LRRK2 | 112,121 |

Page 26/34
Table 3. Central nodes of the DEG network are associated with PD (Supplementary Fig. 9). Central nodes were determined using the Gephi visualization platform. They represent points of convergence of the network (Supplementary Fig. 5). Since these nodes have already been linked to PD pathways, many more DEGs might also contribute to PD pathology through these pathways. These nodes not only provide a point of convergence for DEGs identified in our study, but they also interact with several PARK genes, suggesting that PARK proteins may also converge on the pathways identified here (Supplementary Fig. 7).

| Gene     | PD     | RNA Gene | X     | 122  | RNA Gene |
|----------|--------|----------|-------|------|----------|
| S100A6   |        |          |       |      |          |
| TCEAL5   |        |          | TYW3  | X    |          |
| TSPYL5   |        | Ubiquitination | ZNF37A | X    |          |
| ZNF280D  |        | Nurr1 target | ZNF880 | X    |          |
| ZNF528   |        |          |       |      |          |
| ZNF528.AS1 | RNA Gene |          |       |      |          |
| **Node gene** | **Role in Parkinson’s disease** |
|---------------|--------------------------------|
| **HSPA8** (also known as HSP73, HSC70) | Disaggregation of α-synuclein amyloid fibrils 72  |
| | Autophagy, part of catabolic pathway for α-synuclein 73  |
| | Mediates mitophagy by regulating the stability of PINK1 74  |
| | Impaired gene expression in sporadic PD 75  |
| **EEF1A1** | Mediates activation of heat-shock transcription factor HSF1, prevents α-synuclein aggregation 77  |
| | Interacts with Parkin (PARK2) 65  |
| **HNRNPC** | Interacts with Parkin (PARK2) 65  |
| | Part of the poly ADP-ribose (PAR) cell death pathway accountable for selective dopaminergic neuronal loss 67  |
| **PSMA4** | Part of the Parkinson’s disease KEGG pathway 79,80  |
| | Interacts with Parkin (PARK2) and FBX07 (PARK15) 65  |
| **CYCS** | Role in aggregation of alpha-synuclein 124  |
| | CTD Gene-Disease Associations - Parkinson Disease Gene Set 46  |
| **ACTN1** | Interacts with DJ-1 (PARK7) 65  |
| | It is a binding partner of mitochondrial-shaping proteins 125  |
| **PGK1** | PGK-1 mutation causes vulnerability to parkinsonism 119  |
| | Activation of PGK1 partially restored motor function & slowed disease progression 52  |
| **PHB** | Regulates dopaminergic cell death in substantia nigra 126  |
| **SHH** | Play a role in neuroinflammatory response in the MPTP Model of Parkinson’s Disease 127  |
| **BRCA2** | Deubiquitinase, plays a role in neuronal inflammation 128  |
| **VPS39** | It is part of the endocytic membrane trafficking pathway involved in PD and its methylation rates are associated with Parkinson’s disease risk 129  |
| | Plays complex functions in endocytic and autophagic pathways 130  |
| **UQCRFS1** | KEGG pathway, Parkinson disease 79,80  |
| **CNTNAP2** | Differentially expressed in the presence of LRRK2 G2019S mutation, associated with PD 84  |
| | GWASdb SNP-Disease Associations, Parkinson’s disease Gene Set 46  |
| | Plays a role in formation of protein aggregates & PD 82,83  |
| **CUL3** | Ubiquitin ligase, potential drug target for Parkinson’s disease 71  |
| **PLCB4** | GWAS - Parkinson’s disease 46  |
| | Motor defect consistent with ataxia in Plcb4-null mice 86  |
| **EGLN3** | GEO Signatures of Differentially Expressed Genes for Diseases - Parkinson’s Disease Substantia Nigra 46  |
| | Prolyl hydroxylase, targets substrates for ubiquitination 131  |
| **RALGPS2** | Targets include Nurr1, which is associated with Parkinson disease 46  |
Figure 1

Experimental design. Fibroblasts were used to generate human induced pluripotent stem cells (iPSCs), which were then used to generate mDA neurons. Differentiation was initiated at three different times to obtain cells at different stages of differentiation that could all be collected and analyzed by single-cell RNA-seq on the same day, thus avoiding batch effects. “P+1” indicates that the iPSCs were passaged before new differentiation was initiated. Since D10 was not used in pairwise analysis, we indicated “P+2” between D15 and D6 differentiation initiation.
Figure 2

Generation and classification of iPS cell lines. a. Immunocytochemistry. Staining for the iPSC markers Oct3/4 and TRA-180. DAPI was used to stain cell nuclei as a reference. b. Results of Scorecard analysis of iPSCs and embryonic bodies (EBs). iPSCs are expected to show high expression of self-renewal genes (Self-renew +) and low expression of mesoderm, ectoderm and endoderm markers (Ecto -, Meso-, Endo -). EBs include cells at an early stage of spontaneous differentiation. Scorecard analysis of EBs determines the iPSC line's potential to differentiate into the three germ layers: ectoderm, mesoderm, and endoderm. EBs are expected to express few or no self-renewal genes (Self-renew -) and to show expression of some mesoderm, ectoderm and endoderm markers: Ecto +/-, Meso +/-, Endo +/-.

![Image](image1)

Figure 3

RNA-seq analysis of iPSC status and early differentiation. a. Heatmap showing a gradual transition of gene expression between differentiation stages, at timepoints D0 (iPSCs), D6, D10, D15 and D21. As expected, D6 and D10 are quite similar, as there is only a small change in the factors the cells are exposed to (Supp. Table 1). b. Expression of key iPSC markers and transition to the expression of genes belonging to pathways characteristic of early differentiation. c. Expression of OTX2 was confirmed by absolute quantitative real-time PCR (qPCR) (bars represent SE, n=3 independently differentiated samples, qPCR of each sample was done in triplicate). EN1, an important marker of mDA differentiation, was measured by qPCR (n=3, same qPCR procedure) in bulk samples, as its expression is too low to be detected by single cell RNA seq. Primers are listed in Supp. Table 3.

![Image](image2)
Figure 4

Clustering of cells and expression of mDA markers. a. Gene expression profiles led to cells clustering according to their developmental stage. b. Mature mDA neuronal markers KCNJ6 and TH begin to be expressed in the D21 cluster, indicating the appearance of early postmitotic neurons. c. Expression of additional genes that are characteristic of the mDA differentiation path. d. Immunohistochemistry staining of D25 culture with the neuronal marker MAP2 (red) and the DA neuron marker TH (green). e. Expression of TH, determined by qPCR. TH becomes expressed by early postmitotic neurons (D21) indicating that D21 neurons in culture are roughly equivalent to E12 neurons in mice. Expression of ALDH1A1, a gene specifically expressed by mDA neurons. ALDH1A1 becomes expressed earlier in development than TH132.
Figure 5

Differentially expressed genes (DEGs) in a cell line homozygous for a mutation in the PINK1 gene compared to a control cell line, at three timepoints during the differentiation of mDA neurons (D6, D15 and D21). a. Heatmap of the top DEGs. Each column corresponds to a timepoint for either control (left 3) or PINK1 cells (right 3); each row shows the expression of one gene in individual cells. b. Top DEGs. The minimum fold change was increased to highlight the top differentially expressed genes. The top Venn diagram illustrates upregulated genes, the bottom one downregulated genes. c. Volcano plots of differential expression at the three developmental timepoints.
Figure 6

Ubiquitination and mitochondrial pathways represented within the DEG network, and expression correlation between genes. a. Mitochondrial and ubiquitination pathways represented in the DEG network, as identified by STRING 43. Genes that are part of the ubiquitination pathway are in green and genes associated with mitochondria are in blue; yellow indicates genes belonging to both pathways (also Supplementary Fig.8). b. Upper Plot: Density plot of the number of genes detected as known proteins in STRINGdb (v10) of 50 random networks. The red line represents the number of differentially expressed genes that were detected as known proteins in STRINGdb. Lower Plot: Density plot of the number of protein-protein interactions of 50 random networks. The red line represents the number of interactions found in the PPI network of the differentially expressed genes. c. Distribution of the degree of all genes (nodes) in 51 networks (50 random networks, 1 network using DEGs). The dashed lines represent the means of the two distributions. The boxplot shows the difference between these two distributions (Wilcoxon test, p=2e-16). d. Identification of gene partners based on correlation of expression. These gene partners are dysregulated to the same degree and in the same direction by the presence of the PINK1 mutation. e. Overlay of common edges of the graphs in a. and b., identifying genes that are both known to interact (based on information in STRING and GeneMania databases) and whose expression correlates in our RNA-seq data. f. The 171 top DEGs (groups A-C) and their interaction with ubiquitination genes or mitochondrial genes (based on BioGrid 65). Only direct protein-protein interactions (based on
experimental evidence) were included in the analysis (betweenness Centrality is illustrated in Supplementary Figure 8).

Figure 7

Association with PD and expression correlation of the DEGs. Over 68% of the DEGs were already known to be associated with PD (Supplementary Table 10). However, it is likely that our literature search has omitted several genes. PD-associated DEGs are in blue and expression correlations (Supplementary Table 7) in red. (Betweenness Centrality is illustrated in Supplementary Figure 9).

Supplementary Files

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

- SUPPLEMENTNovaketalNCOMMS2035714492020.pdf