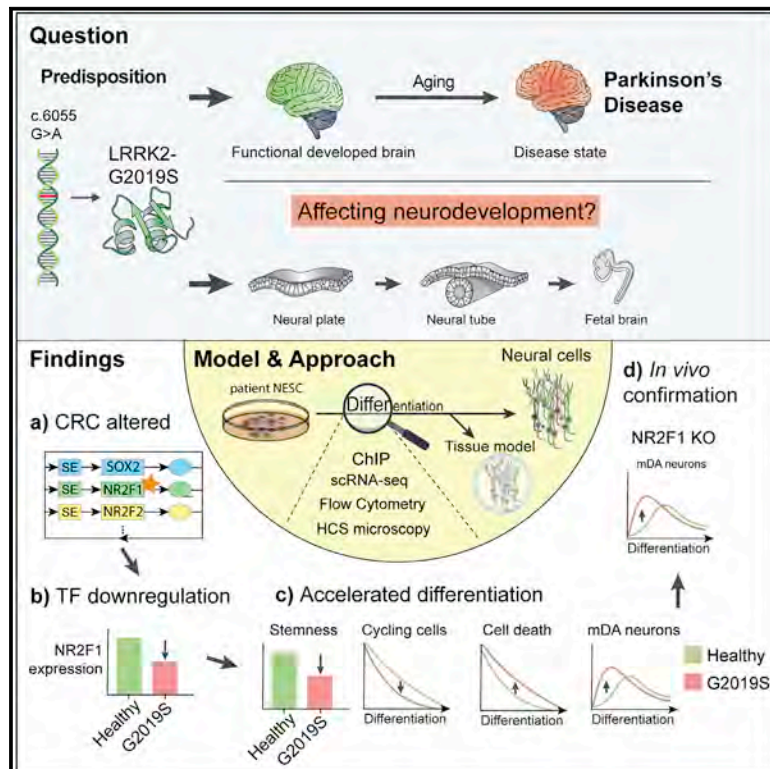**Article**

The Parkinson's-disease-associated mutation LRRK2-G2019S alters dopaminergic differentiation dynamics via NR2F1

Graphical abstract

Authors

Jonas Walter, Silvia Bolognin,
Suresh K. Poovathingal, ...,
Lasse Sinkkonen, Alexander Skupin,
Jens C. Schwamborn

Correspondence

jens.schwamborn@uni.lu

In brief

Walter et al. show that differentiating iPSC-derived neurons carrying the LRRK2-G2019 mutation are characterized by faster initial neuronal differentiation and cell-cycle exit, as well as increased cell death, compared to controls. Downregulation of the transcription factor *NR2F1* appears key to mediate LRRK2-G2019S-dependent phenotypes.

Highlights

- LRRK2-G2019S accelerates cell-cycle exit and dopaminergic differentiation

- LRRK2-G2019S reduces the viability of differentiating dopaminergic neurons

- LRRK2-G2019S downregulates the core regulatory circuit transcription factor *NR2F1*

- Dopaminergic differentiation is accelerated in *NR2F1* mutant mouse embryos

Article

The Parkinson's-disease-associated mutation LRRK2-G2019S alters dopaminergic differentiation dynamics via NR2F1

Jonas Walter,[1,12] Silvia Bolognin,[1,12] Suresh K. Poovathingal,[1,10,13] Stefano Magni,[1,13] Deborah Gérard,[2] Paul M.A. Antony,[1] Sarah L. Nickels,[1,2] Luis Salamanca,[1] Emanuel Berger,[1] Lisa M. Smits,[1] Kamil Grzyb,[1] Rita Perfeito,[3] Fredrik Hoel,[4] Xiaobing Qing,[1] Jochen Ohnmacht,[1,2] Michele Bertacchi,[5] Javier Jarazo,[1] Tomasz Ignac,[1] Anna S. Monzel,[1] Laura Gonzalez-Cano,[1,11] Rejko Krüger,[1,6,7] Thomas Sauter,[2] Michèle Studer,[5] Luis Pereira de Almeida,[3,8] Karl J. Tronstad,[4] Lasse Sinkkonen,[2] Alexander Skupin,[1,9] and Jens C. Schwamborn[1,14,*]

[1]Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 4362 Belvaux, Luxembourg
[2]Department of Life Science and Medicine (DLSM), University of Luxembourg, 4362 Belvaux, Luxembourg
[3]CNC-Center for Neuroscience and Cell Biology, University of Coimbra, Rua Larga, Coimbra 3004-504, Portugal
[4]Department of Biomedicine, University of Bergen, Jonas Lies Vei 91, 5009 Bergen, Norway
[5]Université Côte d'Azur, CNRS, Inserm, 06108 Nice, France
[6]Transversal Translational Medicine, Luxembourg Institute of Health, Strassen, Luxembourg
[7]Centre Hospitalier de Luxembourg, Luxembourg City, Luxembourg
[8]Faculty of Pharmacy, University of Coimbra, Coimbra 3000-548, Portugal
[9]Center for Research of Biological Systems, University of California, San Diego, La Jolla, CA 92093, USA
[10]Present address: Single Cell Analytics & Microfluidics Core, VIB-KU Leuven, Herestraat 49, 3000 Leuven, Belgium
[11]Present address: Department of Cell and Developmental Biology, Max Planck Institute for Molecular Biomedicine, Münster, Germany
[12]These authors contributed equally
[13]These authors contributed equally
[14]Lead contact
*Correspondence: jens.schwamborn@uni.lu
https://doi.org/10.1016/j.celrep.2021.109864

SUMMARY

Increasing evidence suggests that neurodevelopmental alterations might contribute to increase the susceptibility to develop neurodegenerative diseases. We investigate the occurrence of developmental abnormalities in dopaminergic neurons in a model of Parkinson's disease (PD). We monitor the differentiation of human patient-specific neuroepithelial stem cells (NESCs) into dopaminergic neurons. Using high-throughput image analyses and single-cell RNA sequencing, we observe that the PD-associated LRRK2-G2019S mutation alters the initial phase of neuronal differentiation by accelerating cell-cycle exit with a concomitant increase in cell death. We identify the NESC-specific core regulatory circuit and a molecular mechanism underlying the observed phenotypes. The expression of *NR2F1*, a key transcription factor involved in neurogenesis, decreases in LRRK2-G2019S NESCs, neurons, and midbrain organoids compared to controls. We also observe accelerated dopaminergic differentiation *in vivo* in *NR2F1*-deficient mouse embryos. This suggests a pathogenic mechanism involving the LRRK2-G2019S mutation, where the dynamics of dopaminergic differentiation are modified via NR2F1.

INTRODUCTION

Parkinson's disease (PD) is the second most prevalent neurodegenerative disorder after Alzheimer's disease. It is histopathologically characterized by the progressive loss of midbrain dopaminergic neurons (mDAs) in the substantia nigra (Przedborski, 2017; Schapira et al., 2017). However, the pathology's wide spectrum of clinical manifestations is only partially explained by the progressive loss of mDAs. Several lines of evidence have substantiated the hypothesis that PD might have a neurodevelopment component. This is, for example, suggested by an analysis of PD autoptic brain samples (Huisman et al., 2004). The number of mDAs has been found to be decreased in the sub-

stantia nigra but strongly increased in the olfactory bulb (Huisman et al., 2004). Neurogenesis has been reported to occur in human olfactory bulb until the first 18 months after birth (Sanai et al., 2011; Wang et al., 2011). Thus, it is reasonable to conclude that the increased number of mDAs was specified during development. Further evidence for a neurodevelopmental component of PD comes from leucine-rich repeat serine/threonine-protein kinase 2 (LRRK2). Mutations in the gene encoding for LRRK2 have been associated with both familial and sporadic PD (Funayama et al., 2002; Lesage et al., 2006; Ozelius et al., 2006). LRRK2 is highly expressed during development (Zechel et al., 2010), and it binds to proteins of the dishevelled (DLV) family (Sancho et al., 2009), key mediators of the Wnt signaling

pathway. The Wnt pathway is essential in embryonic development (Croce and McClay, 2008), controlling several key processes such as cell fate specification (Logan and Nusse, 2004), including that of mDAs (Arenas, 2014). LRRK2 mutations in the Roc domain have been shown to disrupt its interactions with DLV1–3 (Sancho et al., 2009). Consistent with the neurodevelopmental hypothesis, LRRK2 has been shown to regulate synaptogenesis (Parisiadou et al., 2014; Winner et al., 2011).

The best-characterized PD-associated LRRK2 mutation is the autosomal-dominant genomic mutation c.6055G>A, which results in a LRRK2-G2019S substitution. It also constitutes the most prevalent genetic risk factor for PD (Funayama et al., 2002; Paisán-Ruíz et al., 2004; Zimprich et al., 2004). Functionally, the LRRK2-G2019S mutation causes increased LRRK2 kinase activity (West et al., 2007), which leads to the activation of several pathways (Cookson, 2010).

The penetrance of the LRRK2-G2019S mutation is incomplete and variable (Healy et al., 2008), and it is modulated by age, individual genetic background, and environmental factors (Chen et al., 2017; Hill-Burns et al., 2016; Lorenz et al., 2017). It has already been shown that genetic mutations can interfere with development, resulting in defects of variable severity and diverse clinical manifestations (Marín, 2016; Poduri et al., 2013). Increasing evidence suggests that neural stem cell (NSC) deregulation and the resulting compromised neurodevelopment play a role in PD (Le Grand et al., 2015; Marxreiter et al., 2013; Schwamborn, 2018). The existence of a neurodevelopmental predisposition to PD could potentially explain also some of the extra-nigral symptoms associated with PD (Schapira et al., 2017).

The use of PD-patient-derived human induced pluripotent stem cells (hiPSCs) represents an enabling tool to model developmental alterations in an *in vitro* human-based system (Gammill and Bronner-Fraser, 2003; Reinhardt et al., 2013b). hiPSC-derived cells resemble embryonic identity at the gene expression level and recapitulate early developmental events (Mariani et al., 2012). hiPSC-derived neural cells carrying the LRRK2-G2019S mutation reproduced several *in vivo* PD phenotypes, including reduced neurite ramification (Reinhardt et al., 2013b), synaptic alterations (Borgs et al., 2016; Reinhardt et al., 2013b; Sánchez-Danés et al., 2012), mitochondrial abnormalities (Cooper et al., 2012; Sanders et al., 2014), reduced stress sensitivity (Reinhardt et al., 2013b), autophagosomal dysfunctions (Sánchez-Danés et al., 2012), and stem cell alterations (Liu et al., 2012; Sanders et al., 2014). Previous studies mainly characterized the mature differentiated mDAs without considering potential alterations in the transition dynamics from NSCs to neurons. In contrast, here, we addressed this gap and extensively study the impact of the LRRK2-G2019S mutation on the dynamics of cell fate specification throughout neuronal differentiation of patient-specific neuroepithelial stem cells (NESCs).

RESULTS

In vitro differentiation of human NESCs follows an *in vivo* developmental process

To evaluate whether LRRK2-G2019S mutation could influence the differentiation dynamics of dopaminergic neurons, we cultured NESCs from an early-onset patient (PD2/G2019S) and

the corresponding isogenic control line, where the mutation was corrected (PD2/WT) (Figure S1A). We induced dopaminergic differentiation and collected the cells after 4, 10, 14, and 42 days. To analyze neuronal differentiation at a high-level view, we performed droplet-based single-cell RNA sequencing (scRNA-seq) at the mentioned time points.

To assess the quality and identity of our *in vitro* preparation, we performed a prototype analysis using a machine-learning tool and the information contained in a single-cell atlas of the human developing ventral midbrain (La Manno et al., 2016). The algorithm compares the transcriptomes measured from our *in vitro* model to the *in vivo* reference human fetal cell types. The approach allows evaluating the similarity of each cell to many reference cell types. The classifier assigns each cell a probability to belong to each of the prototypes. An analysis of the top 18 expressed candidate genes relevant to midbrain progenitors, neuroblast medial (NbM) stem cells, and mDAs was conducted. The analysis revealed that the *in vitro* mDA differentiation is incomplete. However, the *in vitro* cells resemble good-quality NbM and progenitors (Figure S1A). Hence, the used cell culture model represents a suitable tool to address early developmental processes.

LRRK2-G2019S causes accelerated neuronal differentiation

After proving the *in vivo* relevance of our *in vitro* model, we evaluated whether the LRRK2-G2019S mutation altered cell fate acquisition. We used 13 iPSC-derived NESC lines generated from three patients carrying the LRRK2-G2019S mutation and four age- and gender-matched healthy individuals. Four isogenic NESC lines (two where the mutation was introduced and two where the mutation was corrected) were also used (Figures 1A and S1A). We used two different clones for patient 1 (P1 and P1.1). The derivation of NESCs from iPSCs and their mDA differentiation and characterization have been previously described (Figure S1A) (Reinhardt et al., 2013a).

Throughout the study, we grouped the cells in two ways; we indicated the disease condition of the donor as healthy (H) or PD and the absence (WT) or presence (G2019S) of the LRRK2-G2019S mutation. H/WT indicates cells from a healthy individual, while H/G2019S are healthy cells after the introduction of the LRRK2-G2019S mutation (isogenic control). The PD/G2019S lines are cells from PD patients carrying the LRRK2-G2019S mutation, and they are annotated as PD/WT after the gene correction. We analyzed all NESC lines according to the following groupings: (1) H⁺PD/G2019S versus H⁺PD/WT), (2) PD/G2019S versus H/WT, (3) PD/G2019S versus PD/WT, and (4) (H/WT versus H/G2019S).

To study the potential contribution of neurodevelopmental alterations to PD, we analyzed the temporal dynamics of neuronal and mDA differentiation via automated image analysis. We focused the analysis on days 4, 10, and 14 of differentiation (Figure 1B). We analyzed the appearing DA neuronal marker tyrosine hydroxylase (TH⁺) and neuron-specific class III beta tubulin (TUJ1⁺) cells in combination with the LIM homeobox transcription factor 1-alpha (LMX1A) as a marker for mDA (Andersson et al., 2006). LMX1A and TH staining confirmed the midbrain identity of the cells during differentiation and showed the higher
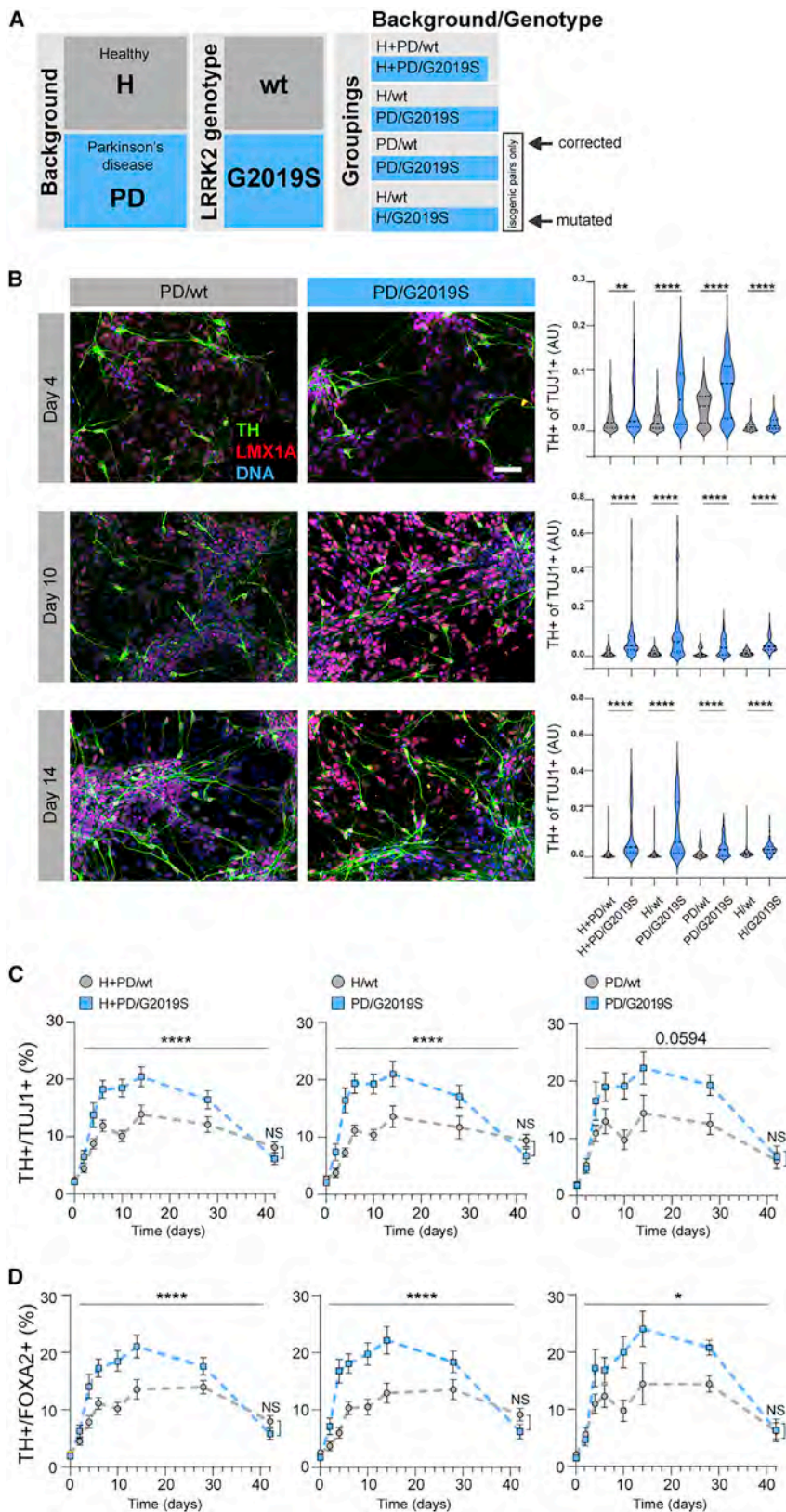
(A) Cell line nomenclature and applied groupings.

(B) Representative confocal images of cells stained for TH, LMX1A, and DNA at days 4, 10, and 14 (scale bar, 50 μm). Quantification of TH$^+$/TUJ1$^+$ pixels at days 4, 10, and 14. Statistical analysis performed with Mann-Whitney test (experiments were performed three times, with one technical replicate).

(C) Cytometry analysis of TH$^+$/TUJ1$^+$ neurons quantified over time.

(D) Cytometry analysis of TH$^+$/FOXA2$^+$ neurons quantified over time, two-way ANOVA followed by Bonferroni post hoc test (experiments were performed at least four times, with one technical replicate).

All data are presented as mean ± SEM. *p < 0.05; ****p < 0.0001.

number of neurons and mDA in LRRK2-G2019S compared to control cultures (Figures 1B and S2A–S2C). LRRK2-G2019S correction in PD-patient-derived cells rescued the dynamics of mDA specification to healthy control levels. Furthermore, we observed a significant increase in the mDA fraction in healthy genetic background lines when the LRRK2-G2019S mutation was introduced (Figure S2A). We also analyzed the ratio of TUJ1$^+$/Hoechst (Figure S2B) and TH$^-$-TUJ1$^+$/Hoechst (Figure S2C) and observed a similar tendency for neurons and non-mDA.

To gain further insights into the exact dynamics of this phenotype, we analyzed the cultures with a flow-cytometry-based approach (fluorescence-activated cell sorting [FACS]). We immunostained the cells using antibodies against TUJ1 and TH. TH$^+$/TUJ1$^+$ neurons were measured prior induction of mDA differentiation (day 0) and at days 2, 4, 6, 8, 10, 14, 28, and 42 of differentiation (Figures 1C, S3A, and S3B). LRRK2-G2019S cultures at days 4, 10, and 14 differentiated into mDAs more rapidly than control cultures. However, at the endpoint of differentiation (day 42), no significant difference in the number of mDA was detectable when comparing LRRK2-G2019S and control cultures. Interestingly, correction of LRRK2-G2019S mutation in patient-derived cells rescued the impaired mDA differentiation dynamics to the levels observed in healthy controls (Figure 1C). We also quantified TH in combination with the mDA marker forkhead box protein A2 (FOXA2) (Sasaki and Hogan, 1994). FOXA2 was present in the majority of the cells (e.g., ~95%) at day 14 (Figures 1D, S3C, and S3D). Quantification of TH$^+$/FOXA2$^+$ neurons confirmed the initial increase of mDA in LRRK2-G2019S cultures compared to controls. Thus, we conclude that the LRRK2-G2019S mutation increases the rate of neuron and mDA formation during differentiation.

To analyze cell identities during neuronal specification, we reverted to the scRNA-seq and analyzed the global transcriptome changes due to LRRK2-G2019S during neuronal differentiation. We generated targeted gene lists specifically for progenitor cells, mDAs, astrocytes, microglia, oligodendrocytes, endothelial cells, and cell cycle (Table S1) based on the available literature (La Manno et al., 2016; Reinhardt et al., 2013a; Whitfield et al., 2002). We profiled 2,000 quality-controlled cells at the four differentiation time points (250 cells per condition on days 0, 10, 14, and 42). After pre-processing of the data (Figure S4A), we calculated cumulative gene expression scores for the targeted cells or cellular processes (NESCs and mDA-based defined gene lists). The analysis of cumulative gene expression distributions and differentially regulated gene expression (Figure S4B; Table S2) showed significant differences between the genotypes for most time points and gene lists. We pooled the data obtained from both genotypes and performed a principal-component analysis. We used the 20 first principal components in t-distributed stochastic neighbor embedding (t-SNE) for further dimensionality reduction (van der Maaten and Hinton, 2008). The resulting t-SNE plots revealed separated cell clusters, particularly on days 10 and 14 (Figure 2A). To characterize the clustering structures and the relation of the cells with clusters of similar gene expression, we fitted a Gaussian mixture model to the t-SNE plots. This resulted in the identification of defined clusters per each time point, as shown by the gray shadowed re-

gions. We mapped individual cells on this clustered structure using color-coded cluster assignments. We observed an increasing specification and heterogeneity of the cells during differentiation. The rather diffuse arrangement of cells over the clusters in the NESCs became more constrained at later time points, forming more distinct clusters. Interestingly, this heterogeneity converged to a more homogeneous gene expression pattern at differentiation day 42. Collectively, the results of the cluster-based scRNA-seq analyses verified the accelerated mDA specification in LRRK2-G2019S-expressing neurons. Furthermore, the pattern of initially diverging and eventually converging gene profiles was clearly visualized by the distinct clusters (Figure 2A). Cellular identity mapping revealed highly specific neuronal differentiation and highlighted the tendency of LRRK2-G2019S NESC cultures toward differentiation. The panel shows dimensionality reduction of the scRNA-seq, clustering, and identification of cell types based on lists of genes. The panel shows dimensionality reduction of the scRNA-seq, clustering, and identification of cell types based on lists of genes for both PD2/WT and PD2/G2019S. Cells identified as neurons are colored in yellow, while cells identified more specifically as mDA are in green. There is one table per time point. The first line of the table reports which percentage of cells belong to which cell type. We can see that neurons represent 5.1% (day 0), 21.6% (day 10), 34% (day 14), and 28.4% (day 42), while mDAs represent 1.8% (day 0), 8.6% (day 10), 11% (day 14), and 5% (day 42) of all cell types. Importantly, the fraction of neurons/mDAs is growing over time, except for the last time point. The second line of the table reports which percentage of cells within every cell type belong to the PD2/G2019S sample. For neurons, these percentages are 88.2% (day 0), 76.8% (day 10), 75.8% (day 14), and 70.8% (day 42); for mDAs, these percentages are 77.7% (day 0), 74.4% (day 10), 84% (day 14), and 22.8% (day 42). Thus, both neurons and mDA are considerably more numerous in the PD2/G2019S cultures than in PD2/WT at every time point (except for mDAs at day 42).

Coherently, LRRK2-G2019S-expressing NESCs appeared to be more prone to differentiate even before neuronal induction, as indicated by the significant increased number of mDA-specific transcripts (Figure 2B). Furthermore, the cumulative expression analysis statistically verified the early upregulation of mDA-related transcripts at days 10 and 14 of differentiation.

To quantify and compare the developmental state between the genotypes, we used the branching and pseudotime analysis provided by *Monocle* (Trapnell et al., 2014; Qiu et al., 2017a, 2017b). The branching analysis identifies three branching points, which reflect the differentiation process. It becomes evident that cells carrying the LRRK2-G2019S mutation originate from the same cell state and that both cell lines differentiate toward similar cell states (Figures 2C and S4F). Moreover, LRRK2-G2019S preceded LRRK2-WT cultures, suggesting a faster differentiation. To quantify this observation, we compared the pseudotimes as a measure of the developmental state of individual cells from both cultures, LRRK2-G2019S and LRRK2-WT, at each day. Figure 2D shows the distribution of pseudotimes of LRRK2-WT and LRRK2-G2019S cells in gray and blue, respectively. In the NESCs, the two distributions coincided
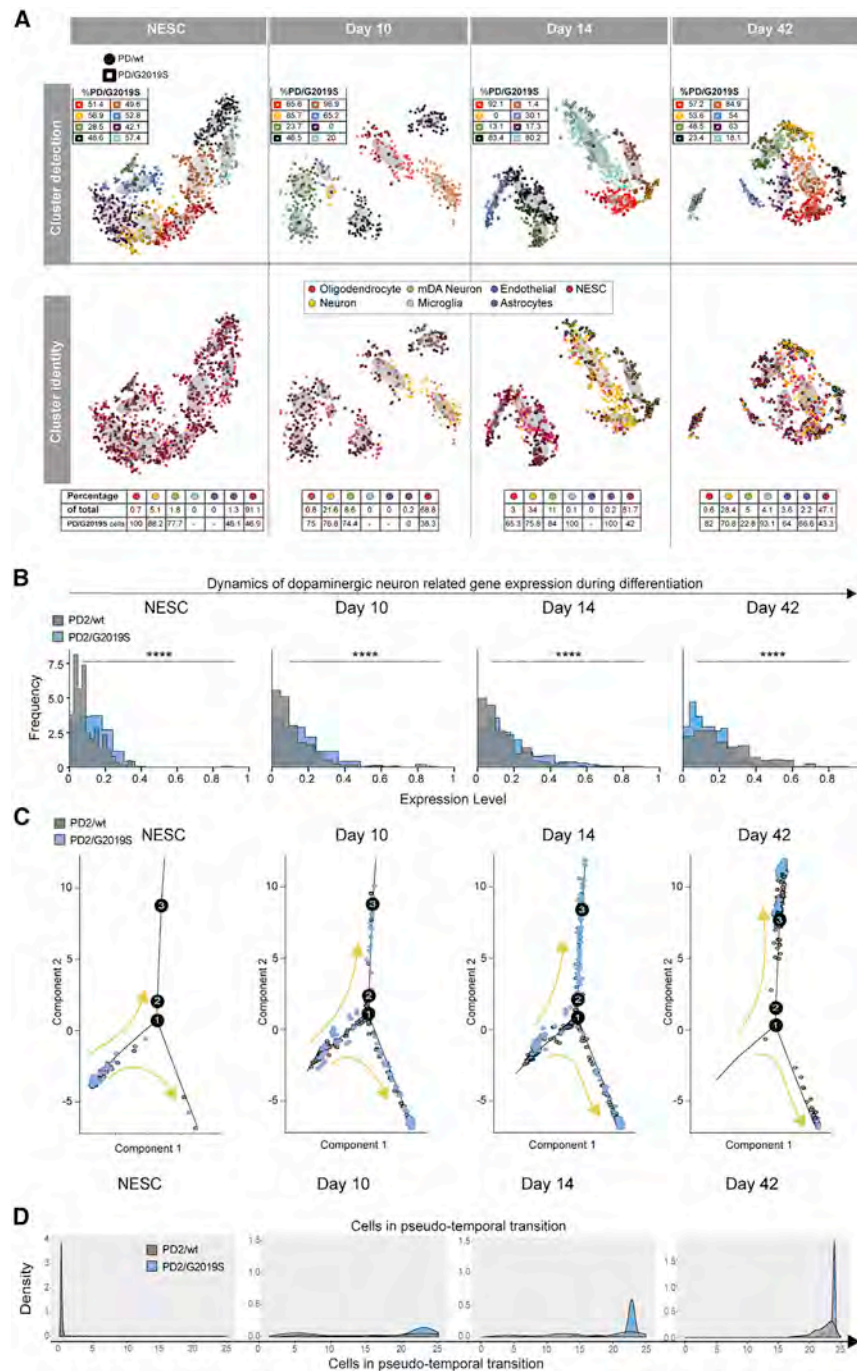
Figure 2. Single-cell RNA sequencing (scRNA-seq) and pseudo-temporal ordering confirms LRRK2-G2019S-dependent acceleration of neuronal differentiation

(A) PCA-t-SNE combination plots showing cluster detection and color-coded cell-type detection based on gene profiles. The gray underlying ellipsoids represent a Gaussian mixture model (x = center). The tables show the fraction of distinct cell types in the total population (upper row) and indicate the percentage of the G2019S genotype in each particular cell type (lower row). ● = control (PD2/WT) and ■ ■ ■ = patient (PD2/G2019S).

(B) Histograms showing cumulative gene expression distributions, for mDA-specific genes, during differentiation. 250 cells per population per day were analyzed. Statistical analysis was performed via z-test followed by Bonferroni post hoc test, ****p < 0.0001. Differentially expressed genes and percentages of particular lists are presented in Figure S4B and in Table S2.

(C) Branching analysis of cell transitions using the Monocle 2.0 R package based on the complete pool of data.

(D) Distribution of corresponding cellular pseudo-times for PD2/WT and PD2/G2019S. Experiments were performed one time with one replicate per time point.

specification. Additionally, these results demonstrate that prior to differentiation, the gene expression pattern in LRRK2-G2019S NESCs already differs from that of the isogenic control cells.

LRRK2-G2019S induces early cell-cycle exit, loss of stemness, and decreased viability in differentiating mDA

To better understand the dynamics of LRRK2-G2019S-associated early enrichment of mDAs during differentiation, we performed a cell-cycle exit analysis. Using automated image analysis, we quantified the fraction of cycling cells using Ki-67 staining (Scholzen and Gerdes, 2000). We observed that the aberrant enrichment of neurons in LRRK2-G2019S cultures was accompanied by an earlier exit of LRRK2-G2019S NESCs from the cell cycle. At days 4, 10, and 14 of differentiation, there were significantly fewer Ki-67[+] cells in LRRK2-G2019S than in wild-type cultures (Figures 3A and S5A).

We then used the scRNA-seq data to investigate differences in cell-cycle-related gene expression (Figures 3B, 3C, and S4D). The resulting expression matrices visualized the loss of cycling cells during differentiation and highlighted the more rapid loss of cell-cycle-related gene expression in LRRK2-G2019S compared to LRRK2-WT cultures (Figure 3B).

demonstrating that cells started from a very similar state. At day 10, the mutant cells exhibited already a significant fraction of cells close to the final differentiation state, while the majority of control cells still exhibited a stem cell state. The faster differentiation of LRRK2-G2019S compared LRRK2-WT cells was maintained at day 14, but eventually, the distribution of both cell lines converged at day 42.

Overall, these data show that the LRRK2-G2019S mutation causes an acceleration of neuronal differentiation and mDA
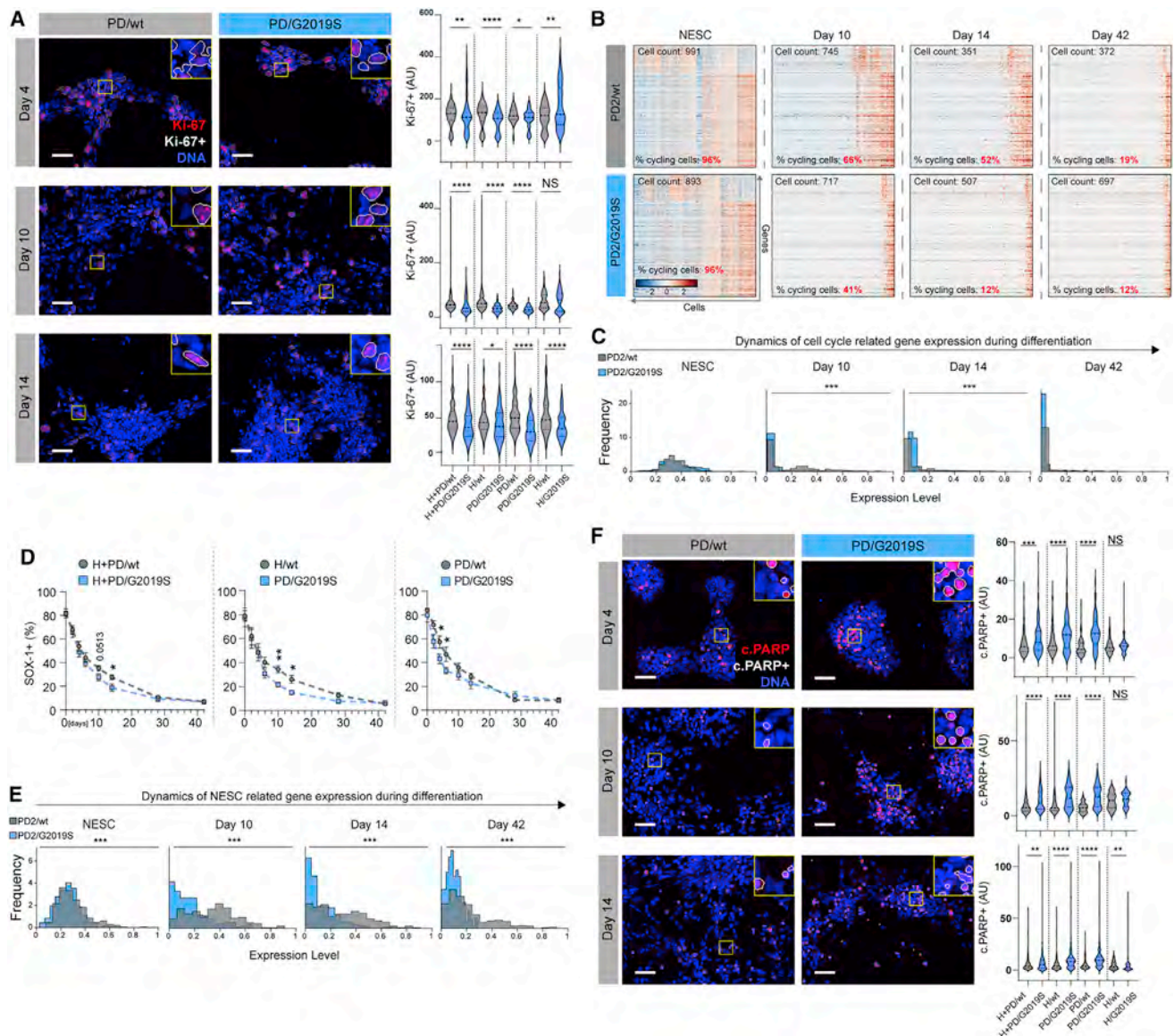
Figure 3. LRRK2-G2019S induces early cell-cycle exit, loss of stemness, and cell death

(A) Representative confocal images of Ki-67 staining and automated quantification of Ki67+ cells at days 4, 10, and 14 of differentiation (scale bar 50 μm). Each data point represents one analyzed micrograph (experiments were performed 3 times, 1 technical replicate). Data were analyzed via Mann-Whitney test.

(B) Expression matrices quantifying cell-cycle exit and related gene expression per cell. The expression levels are color-coded as indicated in the bar; red indicates upregulation, and blue indicates downregulation.

(C) Histograms visualizing loss of cumulative expression of cell-cycle-specific genes during differentiation. Population histograms were analyzed via z-test followed by Bonferroni post hoc test.

(D) Dynamics of SOX-1+ loss during differentiation quantified by FACS (experiments were performed at least four times, with one technical replicate). Data were analyzed via unpaired t test.

(E) Histograms showing cumulative gene expression distributions for NESC-specific genes. Population histograms were analyzed via z-test followed by Bonferroni post hoc test. Differentially expressed genes and percentages of particular lists are presented in Figure S4B and Table S2.

(F) Representative confocal images of cleaved PARP (c.PARP) staining and c.PARP detection at days 4, 10, and 14 of differentiation (scale bar, 50 μm). Quantification of c.PARP+ pixels at days 4, 10, and 14. Each data point represents one micrograph (experiments were performed three times, with one technical replicate).

For all panels, the data are presented as the mean ± SEM. *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001. The scRNA-seq experiment was performed one time with one replicate per time point.

We then quantified the differences in the expression of cell-cycle-related genes in a cumulative expression approach (Figure 3C).

Since the exit from the cell cycle is associated with loss of stemness, we measured the fraction of cells positive for the NSC marker SRY-box 1 (SOX-1) (Bylund et al., 2003; Pevny

et al., 1998). In NESC maintenance cultures (day 0), more than ~85% of the cells were positive for SOX-1 (Figures 3D, S5B, and S5C). In all comparisons, the starting NESC cultures were similar with respect to SOX-1 expression. As expected, the fraction of SOX-1$^+$ cells decreased markedly during neuronal differentiation. The reduction in the fraction of SOX-1$^+$ cells was more prominent in LRRK-G2019S than in LRRK2-WT cells (Figure 3D). We further confirmed the accelerated loss of stemness in LRRK2-G2019S compared to LRRK2-WT cells by considering the cell distributions over gene expression for stemness genes (Figure 3E).

We then investigated whether the increased apoptosis was the reason for the progressive time-dependent decrease in mDA number in LRRK2-G2019S cultures. We quantified the presence of cleaved poly(ADP-ribose) polymerase (c.PARP)$^+$ cells (Duriez and Shah, 1997) at days 4, 10, and 14 of neuronal differentiation. The automated image analysis revealed a significant increase of apoptotic cells during differentiation in LRRK2-G2019S cultures (Figures 3F and S5D). The isogenic controls confirmed the clear dependency of the c.PARP phenotype to the LRRK2-G2019S mutation. The increased apoptosis at day 42 was also confirmed by scRNA-seq (Figures S4B–S4I). In Figure S4I, we compute the distributions of gene expressions for pro-apoptotic, anti-apoptotic, and caspases across cells and compute the corresponding fold change (FC) between the expression in the PD2/G2019S and PD2/WT cells. FC > 0 indicates overexpression in PD2/G2019S (compared to PD2/WT), FC < 0 indicates under-expression in PD2/G2019S (compared to PD2/WT), FC = 1 indicates that PD2/G2019S cells express (on average) twice as much as PD2/WT cells, and FC = 0 indicates that PD2/G2019S and PD2/WTcells express the same levels. We observed that at day 10, PD2/G2019S cells significantly overexpressed anti-apoptotic genes compared to PD2/WT. At day 42, PD2/G2019S cells overexpressed both pro-apoptotic and caspase genes, indicating that at day 42, increased cell death is particularly occurring in PD2/G2019S cells.

Overall, LRRK2-G2019S NESCs differentiated into mDAs significantly more rapidly than healthy cells. This accelerated differentiation was highlighted by a faster loss of proliferation capacity and stemness, and it is accompanied by increased apoptosis.

The core regulatory circuit of NESCs is deregulated in PD-patient-derived cells

As NESCs carrying the LRRK2-G2019S mutation showed an increased propensity to differentiate into mDAs, we next sought to investigate the molecular cause of this alteration. Different cell types have been shown to maintain their cell identity through core regulatory circuits (CRCs) of transcription factors. These factors are often controlled by super-enhancers and bind to each other's regulatory regions to establish feedback mechanisms maintaining cell-type-specific gene expression profiles (Boyer et al., 2005; Hnisz et al., 2013). To derive the NESC-specific CRC, we performed chromatin immunoprecipitation sequencing (ChIP-seq) analysis of histone H3 lysine 27 acetylation (H3K27ac) in NESC lines H1 and H4 from healthy individuals and identified between 225 and 256 active super-enhancers in

NESCs, depending on the cell line (Tables S3 and S4). Next, the transcription factors that are expressed and associated with an active super-enhancer in NESCs were identified, and the obtained data were used as an input for CRC mapper (Saint-André et al., 2016). With CRC mapper, the transcription factors that contain their DNA-binding sequence motif in their own super-enhancer region, thus forming an autoregulatory loop, are first identified. Next, the super-enhancer regions of these autoregulated transcription factors are scanned for the presence of the DNA-binding motifs of other similarly controlled transcription factors. Finally, the identified feedback loops are used to form the cell-line-specific CRCs (Figure 4). The analysis identified a CRC of six transcription factors in each cell line, with five transcription factors (NR2F1, NR2F2, SOX2, POU3F2, and POU3F3) contributing to the CRC reproducibly in each cell line, thus forming the consensus CRC of NESCs.

To test whether the CRC genes were altered in LRRK2-G2019S NESCs and whether this could underlie the observed differences in dynamics of neuronal fate acquisition, we used the scRNA-seq data to compare the expression levels of the CRC genes across the cell populations.

We considered the gene expression across the cell populations (of 250 cells each) for the five transcription factors of the core CRC. We compared the expression for each of these genes in PD2/WT and PD2/G2019S cells at days 0, 10, 14, and 42. This was performed by computing the gene expression FCs of PD2/G2019S with respect to PD2/WT cells (Figure 5A). NR2F1 was significantly downregulated in PD2/G2019S compared to PD2/WT cells. NR2F2 was also significantly downregulated at days 10, 14, and 42. Furthermore, we computed the gene-gene correlation among the expression levels of these five transcription factors, retained only those correlations that were statistically significant (p value < 0.05), and averaged them over the number of gene-gene combinations. As shown in Figure 5B, the value of the average gene-gene correlation across the CRC genes in the NESCs was similar between PD2/WT and PD2/G2019S cells. During differentiation, its value increased in the PD2/WT cells, indicating that these genes are co-expressed in PD2/WT. In contrast, it decreased to zero in PD2/G2019S cultures at days 14 and 42. This indicates no statistically significant correlation between the expressions of these genes in PD2/G2019S cells upon differentiation. This further supports the idea that the CRC is dysregulated by PD2/G2019S mutation upon differentiation. Moreover, we observed in both PD2/WT and PD2/G2019S groups that at day 10 of differentiation the average correlation is higher than the values at day 0 and at day 14 in the corresponding group. This behavior is consistent with the fact that the average gene-gene correlation is expected to be higher when cells are undergoing a transition between cell states than when they are in a well-defined cell state. The majority of PD2/G2019S cells showed a clearly reduced expression of NR2F1, while only a few of the PD2/WT cells show any reduction in its expression (Figure 5C). qRT-PCR confirmed the decreased expression of NR2F1 in both NESC and dopaminergic cultures after 8 days of differentiation (Figures 5D and 5E). NR2F1 expression decreased in H lines when LRRK2-G2019S mutation was introduced and increased in PD lines when it was corrected. Overall, these data show that LRRK2-G2019S mutation altered NESC CRC compared to
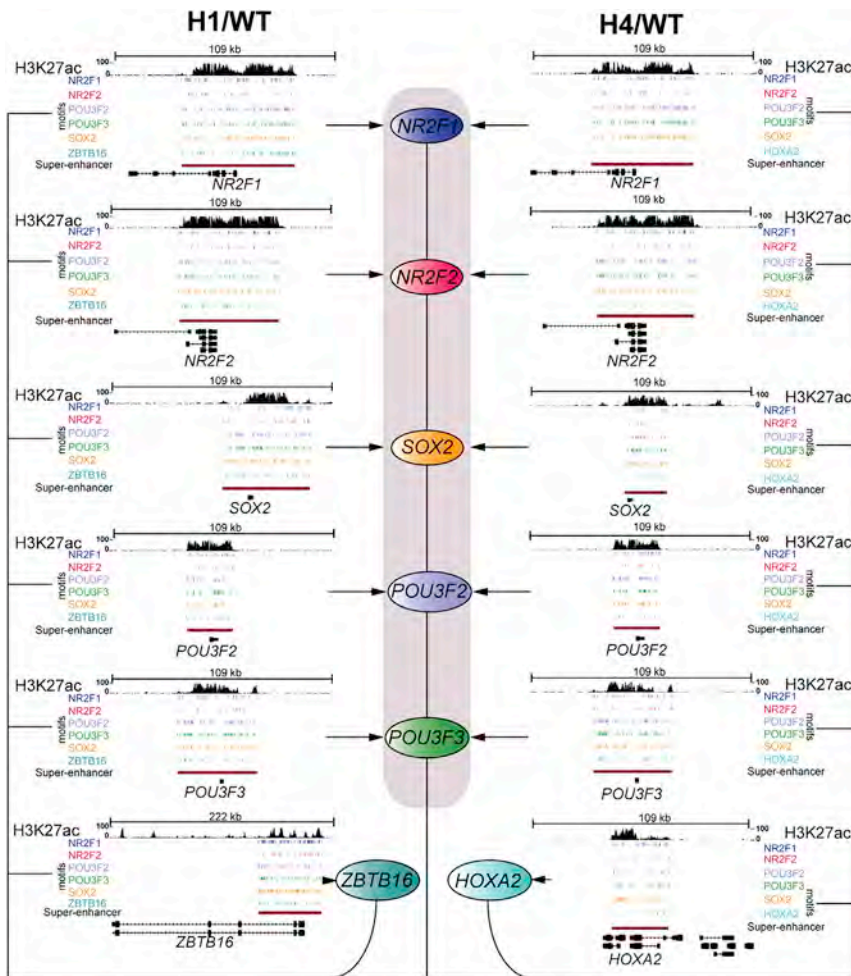
Figure 4. Identification of the core regulatory circuit (CRC) of the NESC state

Representation of the CRC derived for two healthy donor NESC lines (one technical replicate) using CRC mapper (Saint-André et al., 2016). The five shared TFs forming the consensus circuit are highlighted via underlying gray color. For each locus encoding for the indicated TFs, the detected H3K27ac ChIP-seq signal is shown. The called corresponding super-enhancer region is indicated with a burgundy bar, and the locations of the different DNA-binding motifs for the core transcription factors are indicated in the respective colors. For both cell lines, one additional CRC TF (ZBTB16 and HOXA2) was mapped as indicated.

LRRK2-WT cultures. It also highlights the peculiar role of LRRK2-G2019S in downregulating *NR2F1* expression levels.

To evaluate *NR2F1* downregulation in a more complex *in vitro* system, we used human midbrain organoids (hMOs). We have previously shown that midbrain organoids are composed of dopaminergic neurons producing and secreting dopamine (Monzel et al., 2017; Smits et al., 2019). hMOs derived from PD patients carrying the LRRK2-G2019S mutation presented a decrease in the number and dendritic tree complexity of mDAs compared to control organoids ((Smits et al., 2019)). We looked into the scRNA-seq in (Smits et al., 2020) describing the transcriptomic profile of hMOs generated from the healthy line H3/WT in which the mutation was introduced (H3/G2019S). We observed that *NR2F1* mRNA expression is downregulated not only in mDAs but also in glia, stem cell progenitors, and non-dopaminergic neurons in H3/G2019S organoids compared to H3/WT (Figure 5F). The decrease in *NR2F1* expression was observed after both 35 and 70 days of differentiation.

Next, we wanted to understand how the reduced expression of *NR2F1* might be mediated. First, we analyzed existing proximity ligation-assisted ChIP-seq (PLAC-seq) data from primary human cortical neurons to identify the regulatory regions associated with NR2F1 gene via chromatin looping interactions *in vivo* (Figure 6A) (Nott et al., 2019). The identified regulatory regions showed high consistency with the active enhancer regions identified in our NESCs, with most interactions clustering at the *NR2F1* super-enhancer region and some chromatin loops connecting from several hundred kilobases away (Figure 6). Then we performed further ChIP-seq analysis in additional cell lines to map the differentially active enhancer regions between LRRK2-WT and LRRK2-G2019S cells. In detail, we performed ChIP-seq for H3K27ac in two of the patient-derived cell lines (PD1 and PD2) and compared them either to an age-matched healthy NESC line (H1 versus PD1/G2019S) or tthe isogenic control line used in the scRNA-seq analysis (PD2/G2019S versus P2/WT) using THOR, a tool for differential peak calling (Allhoff et al., 2016). This analysis confirmed the presence of a large super-enhancer flanking the *NR2F1* locus also in the additional cell lines, but it did not reveal any consistent changes in it between the two comparisons (Figures 6B and 6C). However, the analysis did reveal the presence of an additional super-enhancer almost 300 kb downstream of *NR2F1* that was also interacting with *NR2F1* transcription start site according to the PLAC-seq data. Interestingly, the differential peak calling revealed a consistent reduction in the H3K27ac signal along a 5-kb region of the distal super-enhancer in the patient-derived stem cells (Figure 6B) in a LRRK2-G2019S-dependent manner (Figure 6C). This suggests that the downregulation of NR2F1 in PD patient cells could be due to LRRK2-dependent alterations in the activity of a distal downstream super-enhancer.

Dopaminergic differentiation is accelerated *in vivo* in Nr2f1-deficient mouse embryos

NR2F1 (Nr2f1 in mouse) is a key factor for the development of several neural structures (Armentano et al., 2007; Bertacchi et al., 2019; Parisot et al., 2017; Tomassy et al., 2010), and its mutations or deletions have been implicated in human pathological
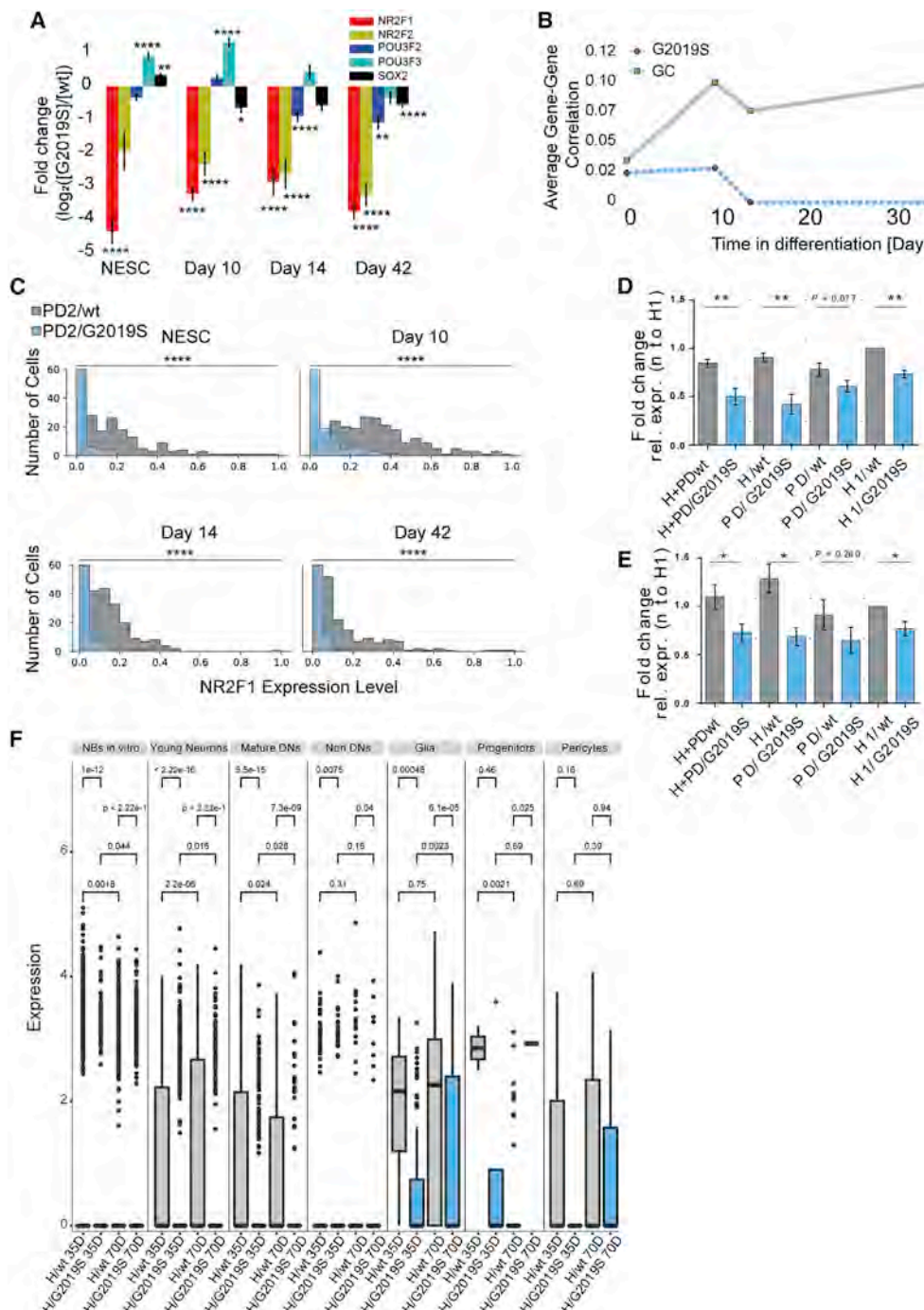
Figure 5. *NR2F1* mRNA levels are downregulated in LRRK2-G2019S cultures compared to controls

(A) Gene expression fold change for the five genes of the CRC of NESCs. For each gene and time point, this was defined as the logarithm in base two of the ratio between the average expression (across cells) of that gene in LRRK2-G2019S and in LRRK2-WT cultures. Negative numbers indicate that LRRK2-G2019S expresses that gene less than LRRK2-WT, and positive numbers indicate the opposite. p values were generated by applying a z-test followed by Bonferroni post hoc test.

(B) Average gene-gene correlations among the five genes of the consensus core CRC. Correlations were filtered by their p value (required to be < 0.05, and Bonferroni correction was applied), thus assuming no correlation exists for couples of genes where the correlation was not statistically significant.

(C and D) Distributions of cells populations across gene expression of *NR2F1* in the scRNA-seq data. The expression of *NR2F1*, normalized within each day, was expressed on the horizontal axes. p values were generated by applying a z-test followed by Bonferroni post hoc test. The scRNA-seq experiment was performed one time with one replicate per time point. NR2F1 mRNA expression levels were measured by qRT-PCR in NESCs (C) and after 8 days of mDA differentiation (D).

conditions (Bertacchi et al., 2018; Bosch et al., 2014; Chen et al., 2016). To test Nr2f1 possible implication in the development of mDA, we first investigated its co-expression pattern with midbrain-specific marker FoxA2 in mouse embryos at embryonic day 12.5 (E12.5) (Figure S6A and S6B), when both neural progenitors (NPs) and neurons start accumulating at the ventral side of the mesencephalon (Ang, 2006; Gale and Li, 2008). Nr2f1 is expressed in virtually all FoxA2[+] cells, at higher level in differentiating neurons compared to NPs (Figures S6A and S6A′). The expression of Nr2f1 can be modulated by using heterozygous (*HET*) or homozygous mouse model (Armentano et al., 2006), in which one or both alleles are removed; consequently, Nr2f1 protein level can be roughly half in *HET* mice or completely abolished in *KO* ones (Figures S6A″–S6B). At E12.5, TH[+] differentiated dopaminergic neurons found at the ventral-most midbrain region widely express Nr2f1 (Figures S6C–S6E), with more than 64.4% ± 2.7% showing high protein levels (Figure S6D). In *HET* embryos, the loss of one *NR2F1* allele is sufficient to lower Nr2f1 protein level, with the majority of the TH[+] population (65.2% ± 3.4%) showing low levels of Nr2f1 (Figure S6E). Taken together, our data show that Nr2f1 expression follows the differentiative process of dopaminergic neurons and can be efficiently modulated with a genetic approach.

As NR2F1 is one of the CRC genes downregulated in LRRK2-G2019S cultures, we exploited the mouse model to test the possible effect of Nr2f1 loss on mDA differentiation *in vivo*. At early time points (E10.5), *Nr2f1 HET* and *KO* embryos already show decreased FoxA2[+] progenitors and a parallel increased production of the first FoxA2[+]TUJ1[+] neurons exiting the cell cycle (Figures 7A–7D), which we reconfirmed 2 days later (Figures 7E–7H). To further investigate early DA differentiation, we analyzed the appearing DA marker TH in the different genotypes and found an increased DA neuronal number also at E13.5 (Figures 7I–7L). As for *in vitro* human cell differentiation, a double staining with cleaved-caspase3 revealed that TH[+] neurons undergo apoptosis when Nr2f1 is downregulated (Figure 7M). Consistently, at E15.5, the number of TH[+] neurons in mutants was back to a level that is comparable to *WT* brains (Figure 7N), suggesting that aberrantly differentiated dopaminergic cells are removed by apoptotic process. All in all, the downregulation of Nr2f1 triggers aberrant early mDA differentiation, followed by increased apoptosis, implying that Nr2f1 is key for dopaminergic differentiation. This highlights the *in vivo* relevance of our findings.

DISCUSSION

Cells in the early neurodevelopmental phase are particularly vulnerable to insults (Rice and Barone, 2000). Changes in the brain microenvironment during the sensitive developmental period might have deleterious effects that become apparent over time (Hübener and Bonhoeffer, 2014). Even subtle develop-

mental alterations might increase the susceptibility toward the development of certain diseases, especially those associated with aging. The contribution of such alterations toward the development of several brain-related diseases has been previously demonstrated (Marín, 2016; Poduri et al., 2013). In this context, we suggest that the PD-associated LRRK2-G2019S mutation alters mDA developmental dynamics via NR2F1.

To model the potential neurodevelopmental contribution of the PD-associated LRRK2-G2019S mutation, we used human iPSC-based neuronal cultures, which are particularly suitable for modeling early human embryonic neurodevelopment (Mariani et al., 2012). The rejuvenation process during reprogramming, which normally represents a major obstacle when studying the age-related aspects of neurodegenerative diseases (Studer et al., 2015), is in this case beneficial. We analyzed the dynamics of the early transition phase of neuronal differentiation from NESCs to mDAs. We observed an accelerated neuronal and mDA fate specification in cells carrying LRRK2-G2019S mutation compared to control cultures. Previous studies have reported that the presence of the LRRK2-G2019S mutation does not significantly affect the number of mDAs at a time point that is comparable to our endpoint (Reinhardt et al., 2013b; Sánchez-Danés et al., 2012). Also, in our study, the amount of mDAs at the last time point assessed is comparable between LRRK2-G2019S and LRRK2-WT cultures. However, the dynamics by which this comparable proportion is reached are significantly different. We previously studied the LRRK2 mutation R1441G and found that murine NSCs carrying the mutation showed increased cell death and decreased neuronal differentiation, as shown by doublecortin staining after 4 days of differentiation (Bahnassawy et al., 2013). The R1441G and G2019S mutations are located in the Roc and kinase domains of the LRRK2 protein, respectively. It is likely that the dynamics of the dopaminergic differentiation might be differentially affected in models carrying the two mutations considering the location of the two-point mutations. Furthermore, differences in the utilized cell culture models as well as species differences might play a role.

An alteration in the temporal specification of mDAs might have subtle but important consequences in the substantia nigra. The timing of brain development influences the signals (growth factors, chemokines, and neurotransmitters) that a neuron receives and impacts the establishment of proper neuronal connections (Jiang and Nardelli, 2016). Neurodevelopment is a competitive and selective process (Yamaguchi and Miura, 2015), and the presence of an altered number of mDAs during specific phases of neurodevelopment might impact correct neuronal integration (Bergami and Berninger, 2012). Such subtle changes in the specification dynamics probably are not sufficient to cause severe abnormalities at that specific time point but might increase the susceptibility toward the onset of PD after a second hit (Sulzer, 2007). This second hit can consist of additional

(E) Data were normalized to the NR2F1 expression level of H1. Statistical significance was assessed with unpaired t test or Mann-Whitney test according to normality test. The comparison H1/WT versus H1/G2019S was assessed with a one-sample t test. *p < 0.05, **p < 0.01. qRT-PCR experiments were performed three times in triplicate.
(F) Gene expression data of *NR2F1* levels from scRNA-seq data of human midbrain organoids (hMOs) generated from the line H3/WT with its isogenic control H3/G2019S. hMOs were cultured for 35 and 70 days; 30 organoids were used for each genotype and time point.
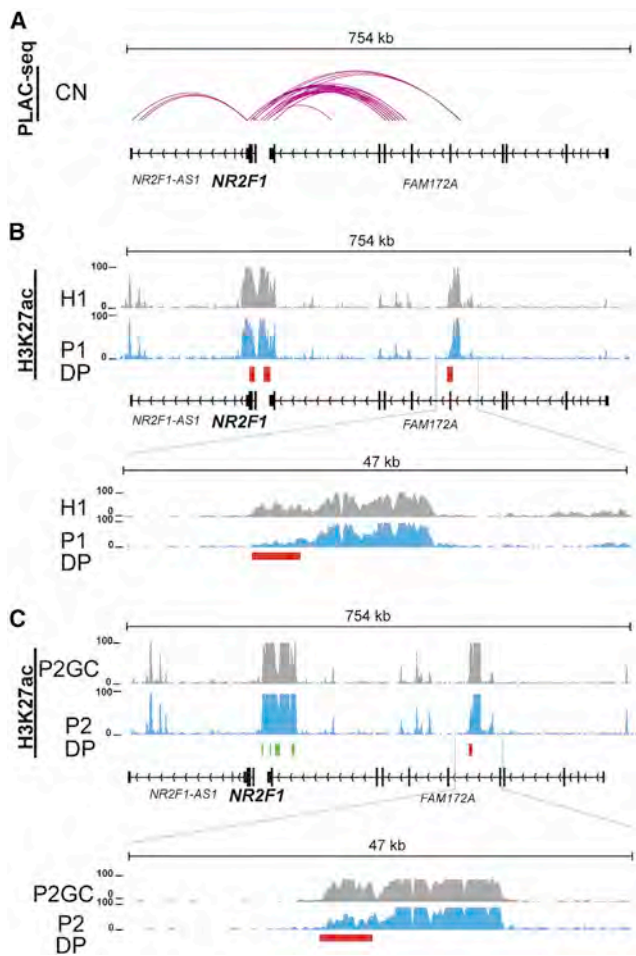
CellPress
OPEN ACCESS

Figure 6. *NR2F1* downregulation is accompanied by reduced enhancer activity at a distal super-enhancer

(A) Long-distance chromatin interactions at *NR2F1* locus were derived from published PLAC-seq data of primary human cortical neurons (CNs) (Nott et al., 2019). The purple lines represent detected chromatin looping interactions between H3K4me3-enriched chromatin and distal regulatory regions.

(B and C) Overview depicting the enrichment of H3K27ac at the *NR2F1* locus in the indicated cell lines. The ChIP-seq tracks from LRRK2-WT cells (H1 and PD2/WT) are shown in gray and tracks from LRRK2-G2019S cells (PD1/G2019S and PD2/G2019S) in blue (1 technical replicate per line). Differential peaks (DPs) with >1,000 raw sequencing read difference called by THOR are indicated by red bars and green bars for regions with significantly (p < 0.05) decreased or increased signal in LRRK2-G2019S cells, respectively.

mutations or genetic variations, infection, trauma, or environmental factors.

In agreement with the literature (Liu et al., 2012; Reinhardt et al., 2013b), we detected a loss of viability in LRRK2-G2019S cultures and confirmed this phenotype throughout several stages of mDA differentiation. The loss of viability and the altered cellular fate transition of LRRK2-G2019S mDAs could result from multiple factors. Increased cell death was in fact observed also in LRRK2-G2019S NESCs compared to controls (Walter et al., 2019; Nickels et al., 2020). According to the scRNA-seq analysis, the presence of the LRRK2-G2019S mutation induced the acquisition of a pro-differentiation state in NESCs. This is highlighted by the significant upregulation of neuronal and mDA transcripts and the downregulation of NESC transcripts. Overall, LRRK2-G2019S cultures showed increased and accelerated mDA generation accompanied by increased cell death.
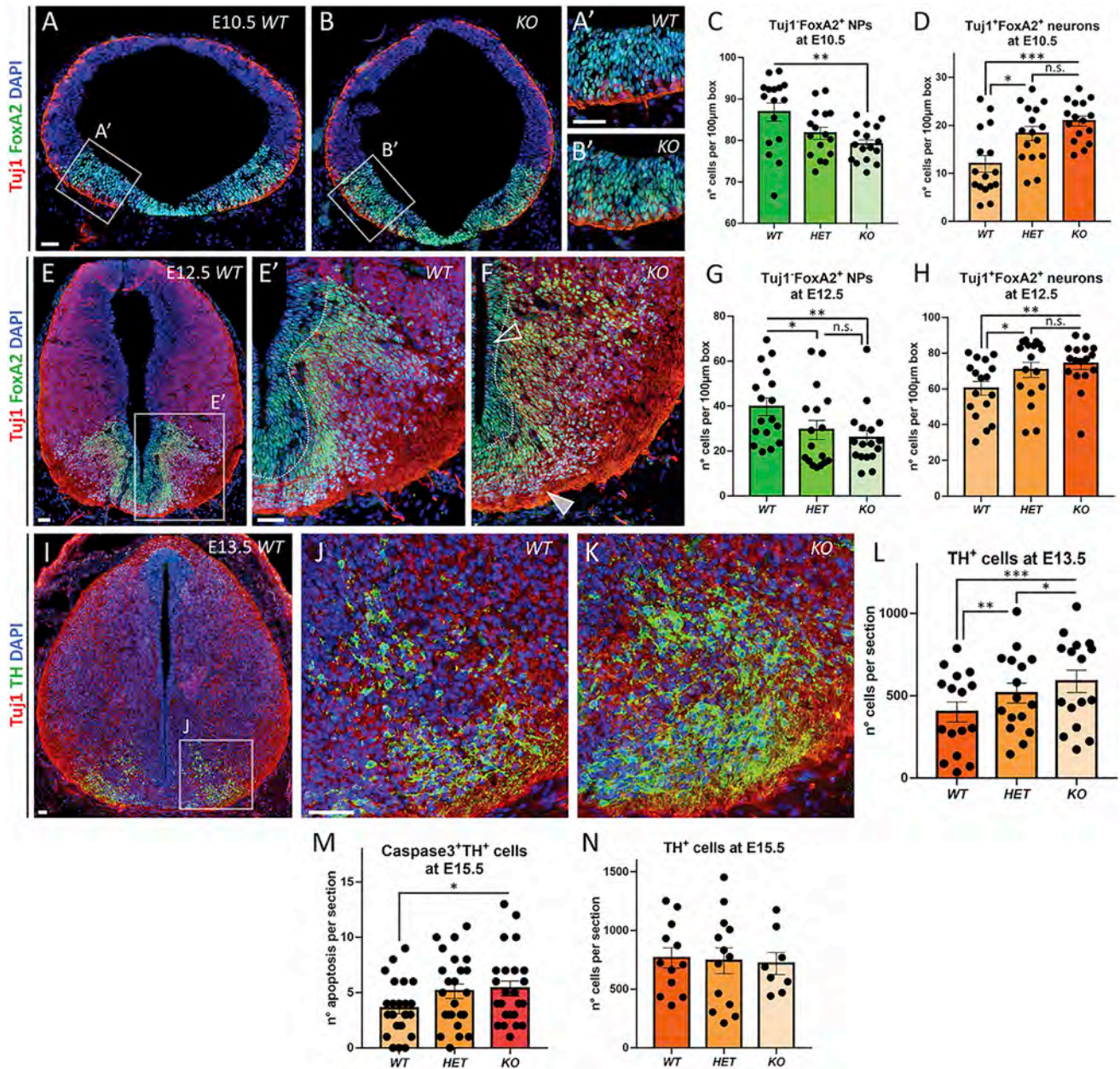
Strikingly, the transcription factor NR2F1 appears to be involved in this accelerated differentiation. NR2Fs, also called chicken ovalbumin upstream promoter transcriptional factors (COUP-TFs), are a family of orphan receptors whose putative molecular ligands are still elusive (Bertacchi et al., 2018). NR2F1 and NR2F2 are the major homologs of the family identified in vertebrates and thought to have similar functions.

The transcription factor Nr2f1 has been shown to play a key role in development, in particular during neocortical area patterning (Greig et al., 2013; Terrigno et al., 2018). Its expression levels seem to regulate temporal competence in NPs (Bertacchi et al., 2018). Nr2f1, along with Nr2f2, regulates the timing of the switch of NSCs from neurogenesis to gliogenesis (Naka et al., 2008). In addition, it regulates neuronal differentiation by coordinating several processes within the neuroepithelium, including cell-cycle exit, neurogenesis, and fate determination (Faedo et al., 2008). These processes seem to be mediated via receptor tyrosine kinase signaling and β-catenin-mediated Wnt signaling (Faedo et al., 2008).

By knocking out Nr2f1, we were able to phenocopy the accelerated mDA differentiation phenotype *in vivo* in the mouse mesencephalon. Interestingly, Nr2f1 downregulation in mouse to half dosage (*HET* mutants) only shows a trend toward accelerated DA neuron differentiation. Only following complete loss of Nr2f1 expression *in vivo* (i.e., full *KO* mutants) were a significant increase of early differentiation and increased apoptosis found. This discrepancy could be due to the fact that Nr2f1 is only one out of five CRC factors for DA development. The loss of Nr2f1 alone in mouse development could be compensated by other factors of the same network, notably Nr2f2, as redundancy between the two homologs has been demonstrated in other brain areas (Flore et al., 2017; Tang et al., 2010; Zhou et al., 2015). *Nr2f1-2* double knockdown was shown to cause sustained neurogenesis in stem cells at the expense of gliogenesis both *in vitro* and *in vivo* in developing mouse brain (Naka et al., 2008). The *in vivo* relevance of NR2F1 in the context of dopaminergic neurons comes also from a study investigating NR2F1 expression in the olfactory bulb of mice (Bovetti et al., 2013). The study proved that odor deprivation resulted in decreased expression of TH and Nr2f1. Considering that hyposmia is an early clinical manifestation of PD, it is notable that Nr2f1 is involved in the regulation of sensory-dependent plasticity in the adult olfactory bulb.

LIMITATIONS OF THE STUDY

This study contains convincing pieces of evidence describing abnormalities in the differentiation trajectory of LRRK2-G2019S dopaminergic neurons mediated by NR2F1. The LRRK2-G2019S-dependent decrease in NR2F1 levels is associated with altered distal super-enhancer activity, but further studies are necessary to clarify precisely the molecular signaling connecting the two proteins. NR2F1 contains several putative phosphorylation sites, some of which have been proven to be

Figure 7. Nr2f1 loss triggers early differentiation and apoptosis of mesencephalic dopaminergic cells

(A–D) FoxA2 (green, dopaminergic progenitors) and TUJ1 (red, neurons) immunostaining at E10.5 in the mesencephalon of *wild-type* (A and A′) versus *Nr2f1* mutant (B and B′) brains. Partial (*HET*) or complete (*KO*) NR2F1 loss triggers early dopaminergic neuron differentiation; as a consequence, NPs are reduced (C; *WT/HET*: *p = 0.017; *WT/KO*: **p = 0.0012), while TUJ1$^+$ neurons increase in mutants (D; *WT/HET*: *p = 0.039; *WT/KO*: **p = 0.0041).

(E–H) FoxA2 (green) and TUJ1 (red) immunostaining at E12.5 in the mesencephalon of *wild-type* (E and E′) versus mutant (F) brains. Dotted lines in (E′ and F) highlight the border between NP and differentiated neuron domains. Average number of NPs and neurons are shown in (G; *WT/HET*: p = 0.12 [not significant (n.s.)]; *WT/KO*: *p = 0.016) and (H; *WT/HET*: p = 0.15 [n.s.]; *WT/KO*: *p = 0.028), respectively. Note the accumulation of increased number of differentiated cells (white arrowhead in F) in the ventral-most mesencephalic region of mutant brains and the reduction of NPs along the ventricular surface (empty arrowhead in F).

(I–L) TH (green, dopaminergic neurons) and TUJ1 (red) immunostaining at E13.5 in control (I and J′) and mutant (K) brains. The average number of TH$^+$ neurons is shown in (L; *WT/HET*: *p = 0.032; *WT/KO*: *p = 0.017).

(M) Average number of apoptotic TH$^+$ neurons, quantified in the ventral mesencephalon after double TH/cleaved-caspase-3 immunostaining. *WT/HET*: p = 0.058 (n.s.); *WT/KO*: *p = 0.026.

(N) Graph shows the average number of TH$^+$ cells at E15.5 in the mesencephalon of different genotypes as indicated. *WT/HET*: p = 0.27 (n.s.); *WT/KO*: p = 0.12 (n.s.).

(legend continued on next page)

phosphorylated *in vivo* (Gay et al., 2002). From here, it is tempting to speculate that NR2F1 can be targeted by cellular kinase, such as LRRK2.

A limitation of this study is that the *in vivo* relevance of the observed LRRK2-G2019S-dependent neurodevelopmental defects has yet to be mechanistically proven in the context of PD. The key point that remains unsolved is whether the accelerated differentiation is the major driver of the increased cell death observed in LRRK2-G2019S differentiating neurons or whether other factors are contributing. The latter scenario seems more likely, as cell death has been observed also in LRRK2-G2019S NESCs (Walter et al., 2018; Nickels et al., 2019), accompanied by altered mitochondrial morphology and functionality (Walter et al., 2018). A major experimental challenge to fully address the role of LRRK2-G2019S in dopaminergic neurogenesis and its impact on PD pathology remains how to recapitulate *in vitro* complex phenomena ranging from brain development to age-associated neurodegeneration, which *in vivo* take place over decades. The use of aging inducers might certainly help to recreate on dish the desired time-dependent cellular changes within a reasonable experimental setup. An additional strategy would be the possibility of gene editing postmitotic neurons to, for example, correct the LRRK2-G2019 mutation only in fully differentiated neurons. This would allow us to evaluate whether these neurons are still susceptible to cell death as a consequence of the altered neurogenesis even though the mutation was corrected. These types of studies currently still present technical challenges, but they are the necessary step forward to unravel key unresolved questions in the field of neurodegeneration.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Material availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Cell Culture
  - Neuronal differentiation
  - Chromatin Immunoprecipitation (ChIP)
  - ChIP-seq
  - Identification of the core regulatory circuit (CRC)
  - Identification of differential H3K27ac peaks
  - Single-cell RNaseq using DropSeq
  - Analysis of differentially regulated genes
  - Cell cycle analysis
  - Pseudo-temporal analysis
  - Cluster detection, cell identity mapping, and quantification
  - Data pre-processing scRNA-seq in organoids
  - Cell type identification scRNA-seq in organoids
  - Characterization of the core CRC genes expression by sc-RNaseq
  - *In vivo* prototype mapping
  - Immunocytochemistry
  - Image analysis
  - Flow Cytometry Analyses
  - Quantitative real time PCR
  - Immunohistochemistry
- QUANTIFICATION AND STATISTICAL ANALYSIS

In (C), (D), (G), and (H), the number of positive cells was quantified in 100-μm-wide boxes randomly placed across the ventral-most region of mesencephalon, while in (L)–(N), the total number of cells per cryostat section was counted. Nuclei (blue) were stained with DAPI. In graphs, data are represented as mean ± SEM (Student's t test: *p < 0.05; **p < 0.01. Scale bars, 50 μm. For cell percentage/number quantification, measurements were performed on at least nine sections from two or three different animals per genotype.

REFERENCES

Allhoff, M., Seré, K., F Pires, J., Zenke, M., and G Costa, I. (2016). Differential peak calling of ChIP-seq signals with replicates with THOR. Nucleic Acids Res. *44*, e153.

Andersson, E., Tryggvason, U., Deng, Q., Friling, S., Alekseenko, Z., Robert, B., Perlmann, T., and Ericson, J. (2006). Identification of intrinsic determinants of midbrain dopamine neurons. Cell *124*, 393–405.

Ang, S.L. (2006). Transcriptional control of midbrain dopaminergic neuron development. Development *133*, 3499–3506.

Arenas, E. (2014). Wnt signaling in midbrain dopaminergic neuron development and regenerative medicine for Parkinson's disease. J. Mol. Cell Biol. *6*, 42–53.

Armentano, M., Chou, S.J., Tomassy, G.S., Leingartner, A., O'Leary, D.D., and Studer, M. (2007). COUP-TFI regulates the balance of cortical patterning between frontal/motor and sensory areas. Nat. Neurosci. *10*, 1277–1286.

Armentano, M., Filosa, A., Andolfi, G., and Studer, M. (2006). COUP-TFI is required for the formation of commissural projections in the forebrain by regulating axonal growth. Development *133*, 4151–4162.

Bahnassawy, L., Nicklas, S., Palm, T., Menzl, I., Birzele, F., Gillardon, F., and Schwamborn, J.C. (2013). The parkinson's disease-associated LRRK2 mutation R1441G inhibits neuronal differentiation of neural stem cells. Stem Cells Dev *22*, 2487–2496.

Becht, E., McInnes, L., Healy, J., Dutertre, C.A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. Nat. Biotechnol. https://doi.org/10.1038/nbt.4314.

Bergami, M., and Berninger, B. (2012). A fight for survival: the challenges faced by a newborn neuron integrating in the adult hippocampus. Dev. Neurobiol. *72*, 1016–1031.

Bertacchi, M., Parisot, J., and Studer, M. (2018). The pleiotropic transcriptional regulator COUP-TFI plays multiple roles in neural development and disease. Brain Res.

Bertacchi, M., Parisot, J., and Studer, M. (2019). The pleiotropic transcriptional regulator COUP-TFI plays multiple roles in neural development and disease. Brain Res *1705*, 75–94.

Bishop, C.M. (2006). Pattern Recognition and Machine Learning (Springer-Verlag New York).

Borgs, L., Peyre, E., Alix, P., Hanon, K., Grobarczyk, B., Godin, J.D., Purnelle, A., Krusy, N., Maquet, P., Lefebvre, P., et al. (2016). Dopaminergic neurons differentiating from LRRK2 G2019S induced pluripotent stem cells show early neuritic branching defects. Sci. Rep. *6*, 33377.

Bosch, D.G., Boonstra, F.N., Gonzaga-Jauregui, C., Xu, M., de Ligt, J., Jhangiani, S., Wiszniewski, W., Muzny, D.M., Yntema, H.G., Pfundt, R., et al. (2014). NR2F1 mutations cause optic atrophy with intellectual disability. Am. J. Hum. Genet. *94*, 303–309.

Bovetti, S., Bonzano, S., Garzotto, D., Giannelli, S.G., Iannielli, A., Armentano, M., Studer, M., and De Marchis, S. (2013). COUP-TFI controls activity-dependent tyrosine hydroxylase expression in adult dopaminergic olfactory bulb interneurons. Development *140*, 4850–4859.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947–956.

Bylund, M., Andersson, E., Novitch, B.G., and Muhr, J. (2003). Vertebrate neurogenesis is counteracted by Sox1-3 activity. Nat. Neurosci. *6*, 1162–1168.

Chen, C.A., Bosch, D.G., Cho, M.T., Rosenfeld, J.A., Shinawi, M., Lewis, R.A., Mann, J., Jayakar, P., Payne, K., Walsh, L., et al. (2016). The expanding clinical phenotype of Bosch-Boonstra-Schaaf optic atrophy syndrome: 20 new cases and possible genotype-phenotype correlations. Genet. Med. *18*, 1143–1150.

Chen, H., Kwong, J.C., Copes, R., Tu, K., Villeneuve, P.J., van Donkelaar, A., Hystad, P., Martin, R.V., Murray, B.J., Jessiman, B., et al. (2017). Living near major roads and the incidence of dementia, Parkinson's disease, and multiple sclerosis: a population-based cohort study. Lancet *389*, 718–726.

Cookson, M.R. (2010). The role of leucine-rich repeat kinase 2 (LRRK2) in Parkinson's disease. Nat. Rev. Neurosci. *11*, 791–797.

Cooper, O., Seo, H., Andrabi, S., Guardia-Laguarta, C., Graziotto, J., Sundberg, M., McLean, J.R., Carrillo-Reid, L., Xie, Z., Osborn, T., et al. (2012). Pharmacological rescue of mitochondrial deficits in iPSC-derived neural cells from patients with familial Parkinson's disease. Sci. Transl. Med. *4*, 141ra90.

Croce, J.C., and McClay, D.R. (2008). Evolution of the Wnt pathways. Methods Mol. Biol. *469*, 3–18.

Duriez, P.J., and Shah, G.M. (1997). Cleavage of poly(ADP-ribose) polymerase: a sensitive parameter to study cell death. Biochem. Cell Biol. *75*, 337–349.

Faedo, A., Tomassy, G.S., Ruan, Y., Teichmann, H., Krauss, S., Pleasure, S.J., Tsai, S.Y., Tsai, M.J., Studer, M., and Rubenstein, J.L. (2008). COUP-TFI coordinates cortical patterning, neurogenesis, and laminar fate and modulates MAPK/ERK, AKT, and beta-catenin signaling. Cereb. Cortex *18*, 2117–2131.

Flore, G., Di Ruberto, G., Parisot, J., Sannino, S., Russo, F., Illingworth, E.A., Studer, M., and De Leonibus, E. (2017). Gradient COUP-TFI Expression Is Required for Functional Organization of the Hippocampal Septo-Temporal Longitudinal Axis. Cereb. Cortex *27*, 1629–1643.

Funayama, M., Hasegawa, K., Kowa, H., Saito, M., Tsuji, S., and Obata, F. (2002). A new locus for Parkinson's disease (PARK8) maps to chromosome 12p11.2-q13.1. Ann. Neurol. *51*, 296–301.

Gale, E., and Li, M. (2008). Midbrain dopaminergic neuron fate specification: Of mice and embryonic stem cells. Mol. Brain *1*, 8.

Gammill, L.S., and Bronner-Fraser, M. (2003). Neural crest specification: migrating into genomics. Nat. Rev. Neurosci. *4*, 795–805.

Gay, F., Barath, P., Desbois-Le Peron, C., Metivier, R., Le Guevel, R., Birse, D., and Salbert, G. (2002). Multiple phosphorylation events control chicken ovalbumin upstream promoter transcription factor I orphan nuclear receptor activity. Mol. Endocrinol. *16*, 1332–1351.

Gérard, D., Schmidt, F., Ginolhac, A., Schmitz, M., Halder, R., Ebert, P., Schulz, M.H., Sauter, T., and Sinkkonen, L. (2019). Temporal enhancer profiling of parallel lineages identifies AHR and GLIS1 as regulators of mesenchymal multipotency. Nucleic Acids Res *47*, 1141–1163.

Greig, L.C., Woodworth, M.B., Galazo, M.J., Padmanabhan, H., and Macklis, J.D. (2013). Molecular logic of neocortical projection neuron specification, development and diversity. Nat. Rev. Neurosci. *14*, 755–769.

Haghverdi, L., Büttner, M., Wolf, F.A., Buettner, F., and Theis, F.J. (2016). Diffusion pseudotime robustly reconstructs lineage branching. Nat. Methods *13*, 845–848.

Healy, D.G., Falchi, M., O'Sullivan, S.S., Bonifati, V., Durr, A., Bressman, S., Brice, A., Aasly, J., Zabetian, C.P., Goldwurm, S., et al.; International LRRK2 Consortium (2008). Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. Lancet Neurol. *7*, 583–590.

Hill-Burns, E.M., Ross, O.A., Wissemann, W.T., Soto-Ortolaza, A.I., Zareparsi, S., Siuda, J., Lynch, T., Wszolek, Z.K., Silburn, P.A., Mellick, G.D., et al. (2016). Identification of genetic modifiers of age-at-onset for familial Parkinson's disease. Hum. Mol. Genet. *25*, 3849–3862.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Hübener, M., and Bonhoeffer, T. (2014). Neuronal plasticity: beyond the critical period. Cell *159*, 727–737.

Huisman, E., Uylings, H.B., and Hoogland, P.V. (2004). A 100% increase of dopaminergic cells in the olfactory bulb may explain hyposmia in Parkinson's disease. Mov. Disord. *19*, 687–692.

Jiang, X., and Nardelli, J. (2016). Cellular and molecular introduction to brain development. Neurobiol. Dis. *92 (Pt A)*, 3–17.

La Manno, G., Gyllborg, D., Codeluppi, S., Nishimura, K., Salto, C., Zeisel, A., Borm, L.E., Stott, S.R.W., Toledo, E.M., Villaescusa, J.C., et al. (2016). Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. Cell *167*, 566–580.e19.

Le Grand, J.N., Gonzalez-Cano, L., Pavlou, M.A., and Schwamborn, J.C. (2015). Neural stem cells in Parkinson's disease: a role for neurogenesis defects in onset and progression. Cell. Mol. Life Sci. *72*, 773–797.

Lesage, S., Dürr, A., Tazir, M., Lohmann, E., Leutenegger, A.L., Janin, S., Pollak, P., and Brice, A.; French Parkinson's Disease Genetics Study Group (2006). LRRK2 G2019S as a cause of Parkinson's disease in North African Arabs. N. Engl. J. Med. *354*, 422–423.

Liu, G.H., Qu, J., Suzuki, K., Nivet, E., Li, M., Montserrat, N., Yi, F., Xu, X., Ruiz, S., Zhang, W., et al. (2012). Progressive degeneration of human neural stem cells caused by pathogenic LRRK2. Nature *491*, 603–607.

Logan, C.Y., and Nusse, R. (2004). The Wnt signaling pathway in development and disease. Annu. Rev. Cell Dev. Biol. *20*, 781–810.

Lorenz, C., Lesimple, P., Bukowiecki, R., Zink, A., Inak, G., Mlody, B., Singh, M., Semtner, M., Mah, N., Auré, K., et al. (2017). Human iPSC-Derived Neural Progenitors Are an Effective Drug Discovery Model for Neurological mtDNA Disorders. Cell Stem Cell *20*, 659–674.e9.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214.

Mariani, J., Simonini, M.V., Palejev, D., Tomasini, L., Coppola, G., Szekely, A.M., Horvath, T.L., and Vaccarino, F.M. (2012). Modeling human cortical development in vitro using induced pluripotent stem cells. Proc. Natl. Acad. Sci. USA *109*, 12770–12775.

Marín, O. (2016). Developmental timing and critical windows for the treatment of psychiatric disorders. Nat. Med. *22*, 1229–1238.

Marxreiter, F., Regensburger, M., and Winkler, J. (2013). Adult neurogenesis in Parkinson's disease. Cell. Mol. Life Sci. *70*, 459–473.

Monzel, A.S., Smits, L.M., Hemmer, K., Hachi, S., Moreno, E.L., van Wuellen, T., Jarazo, J., Walter, J., Bruggemann, I., Boussaad, I., et al. (2017). Derivation of Human Midbrain-Specific Organoids from Neuroepithelial Stem Cells. Stem Cell Rep *8*, 1144–1154.

Naka, H., Nakamura, S., Shimazaki, T., and Okano, H. (2008). Requirement for COUP-TFI and II in the temporal specification of neural stem cells in CNS development. Nat. Neurosci. *11*, 1014–1023.

Nickels, S.L., Modamio, J., Mendes-Pinheiro, B., Monzel, A.S., Betsou, F., and Schwamborn, J.C. (2020). Reproducible generation of human midbrain organoids for in vitro modeling of Parkinson's disease. Stem Cell Res *46*, 101870.

Nott, A., Holtman, I.R., Coufal, N.G., Schlachetzki, J.C.M., Yu, M., Hu, R., Han, C.Z., Pena, M., Xiao, J., Wu, Y., et al. (2019). Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. Science *366*, 1134–1139.

Ozelius, L.J., Senthil, G., Saunders-Pullman, R., Ohmann, E., Deligtisch, A., Tagliati, M., Hunt, A.L., Klein, C., Henick, B., Hailpern, S.M., et al. (2006). LRRK2 G2019S as a cause of Parkinson's disease in Ashkenazi Jews. N. Engl. J. Med. *354*, 424–425.

Paisán-Ruíz, C., Jain, S., Evans, E.W., Gilks, W.P., Simón, J., van der Brug, M., López de Munain, A., Aparicio, S., Gil, A.M., Khan, N., et al. (2004). Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. Neuron *44*, 595–600.

Parisiadou, L., Yu, J., Sgobio, C., Xie, C., Liu, G., Sun, L., Gu, X.L., Lin, X., Crowley, N.A., Lovinger, D.M., and Cai, H. (2014). LRRK2 regulates synaptogenesis and dopamine receptor activation through modulation of PKA activity. Nat. Neurosci. *17*, 367–376.

Parisot, J., Flore, G., Bertacchi, M., and Studer, M. (2017). COUP-TFI mitotically regulates production and migration of dentate granule cells and modulates hippocampal Cxcr4 expression. Development *144*, 2045–2058.

Pevny, L.H., Sockanathan, S., Placzek, M., and Lovell-Badge, R. (1998). A role for SOX1 in neural determination. Development *125*, 1967–1978.

Poduri, A., Evrony, G.D., Cai, X., and Walsh, C.A. (2013). Somatic mutation, genomic variation, and neurological disease. Science *341*, 1237758.

Przedborski, S. (2017). The two-century journey of Parkinson disease research. Nat. Rev. Neurosci. *18*, 251–259.

Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y.A., and Trapnell, C. (2017a). Single-cell mRNA quantification and differential analysis with Census. Nat. Methods *14*, 309–315.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H., and Trapnell, C. (2017b). Reversed graph embedding resolves complex single-cell developmental trajectories. bioRxiv. https://doi.org/10.1038/nmeth.4402.

Reinhardt, P., Glatza, M., Hemmer, K., Tsytsyura, Y., Thiel, C.S., Höing, S., Moritz, S., Parga, J.A., Wagner, L., Bruder, J.M., et al. (2013a). Derivation and expansion using only small molecules of human neural progenitors for neurodegenerative disease modeling. PLoS ONE *8*, e59252.

Reinhardt, P., Schmid, B., Burbulla, L.F., Schöndorf, D.C., Wagner, L., Glatza, M., Höing, S., Hargus, G., Heck, S.A., Dhingra, A., et al. (2013b). Genetic correction of a LRRK2 mutation in human iPSCs links parkinsonian neurodegeneration to ERK-dependent changes in gene expression. Cell Stem Cell *12*, 354–367.

Rice, D., and Barone, S., Jr. (2000). Critical periods of vulnerability for the developing nervous system: evidence from humans and animal models. Environ. Health Perspect. *108 (Suppl 3)*, 511–533.

Saint-André, V., Federation, A.J., Lin, C.Y., Abraham, B.J., Reddy, J., Lee, T.I., Bradner, J.E., and Young, R.A. (2016). Models of human core transcriptional regulatory circuitries. Genome Res. *26*, 385–396.

Sanai, N., Nguyen, T., Ihrie, R.A., Mirzadeh, Z., Tsai, H.H., Wong, M., Gupta, N., Berger, M.S., Huang, E., Garcia-Verdugo, J.M., et al. (2011). Corridors of migrating neurons in the human brain and their decline during infancy. Nature *478*, 382–386.

Sánchez-Danés, A., Richaud-Patin, Y., Carballo-Carbajal, I., Jiménez-Delgado, S., Caig, C., Mora, S., Di Guglielmo, C., Ezquerra, M., Patel, B., Giralt, A., et al. (2012). Disease-specific phenotypes in dopamine neurons from human iPS-based models of genetic and sporadic Parkinson's disease. EMBO Mol. Med. *4*, 380–395.

Sancho, R.M., Law, B.M., and Harvey, K. (2009). Mutations in the LRRK2 Roc-COR tandem domain link Parkinson's disease to Wnt signalling pathways. Hum. Mol. Genet. *18*, 3955–3968.

Sanders, L.H., Laganière, J., Cooper, O., Mak, S.K., Vu, B.J., Huang, Y.A., Paschon, D.E., Vangipuram, M., Sundararajan, R., Urnov, F.D., et al. (2014). LRRK2 mutations cause mitochondrial DNA damage in iPSC-derived neural cells from Parkinson's disease patients: reversal by gene correction. Neurobiol. Dis. *62*, 381–386.

Sasaki, H., and Hogan, B.L. (1994). HNF-3 beta as a regulator of floor plate development. Cell *76*, 103–115.

Schapira, A.H.V., Chaudhuri, K.R., and Jenner, P. (2017). Non-motor features of Parkinson disease. Nat. Rev. Neurosci. *18*, 435–450.

Scholzen, T., and Gerdes, J. (2000). The Ki-67 protein: from the known and the unknown. J. Cell. Physiol. *182*, 311–322.

Schwamborn, J.C. (2018). Is Parkinson's Disease a Neurodevelopmental Disorder and Will Brain Organoids Help Us to Understand It? Stem Cells Dev. *27*, 968–975.

Smits, L.M., Magni, S., Kinugawa, K., Grzyb, K., Luginbuhl, J., Sabate-Soler, S., Bolognin, S., Shin, J.W., Mori, E., Skupin, A., et al. (2020). Single-cell transcriptomics reveals multiple neuronal cell types in human midbrain-specific organoids. Cell Tissue Res *382*, 463–476.

Smits, L.M., Reinhardt, L., Reinhardt, P., Glatza, M., Monzel, A.S., Stanslowsky, N., Rosato-Siri, M.D., Zanon, A., Antony, P.M., Bellmann, J., et al.

(2019). Modeling Parkinson's disease in midbrain-like organoids. NPJ Parkinsons Dis 5, 5.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell 177, 1888-1902 e1821.

Studer, L., Vera, E., and Cornacchia, D. (2015). Programming and Reprogramming Cellular Age in the Era of Induced Pluripotency. Cell Stem Cell 16, 591–600.

Sulzer, D. (2007). Multiple hit hypotheses for dopamine neuron loss in Parkinson's disease. Trends Neurosci. 30, 244–250.

Tang, K., Xie, X., Park, J.I., Jamrich, M., Tsai, S., and Tsai, M.J. (2010). COUP-TFs regulate eye development by controlling factors essential for optic vesicle morphogenesis. Development 137, 725–734.

Terrigno, M., Bertacchi, M., Pandolfini, L., Baumgart, M., Calvello, M., Cellerino, A., Studer, M., and Cremisi, F. (2018). The microRNA miR-21 Is a Mediator of FGF8 Action on Cortical COUP-TFI Translation. Stem Cell Rep. 16, 591–600.

Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., 2nd, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science 352, 189–196.

Tomassy, G.S., De Leonibus, E., Jabaudon, D., Lodato, S., Alfano, C., Mele, A., Macklis, J.D., and Studer, M. (2010). Area-specific temporal control of corticospinal motor neuron differentiation by COUP-TFI. Proc. Natl. Acad. Sci. USA. 107, 3576–3581.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. 32, 381–386.

van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605.

Walter, J., Bolognin, S., Antony, P.M.A., Nickels, S.L., Poovathingal, S.K., Salamanca, L., Magni, S., Perfeito, R., Hoel, F., Qing, X., et al. (2019). Neural Stem Cells of Parkinson's Disease Patients Exhibit Aberrant Mitochondrial Morphology and Functionality. Stem Cell Reports 12, 878–889.

Wang, C., Liu, F., Liu, Y.Y., Zhao, C.H., You, Y., Wang, L., Zhang, J., Wei, B., Ma, T., Zhang, Q., et al. (2011). Identification and characterization of neuroblasts in the subventricular zone and rostral migratory stream of the adult human brain. Cell Res. 21, 1534–1550.

Warren Andersen, S., Trentham-Dietz, A., Gangnon, R.E., Hampton, J.M., Figueroa, J.D., Skinner, H.G., Engelman, C.D., Klein, B.E., Titus, L.J., and Newcomb, P.A. (2013). The associations between a polygenic score, reproductive and menstrual risk factors and breast cancer risk. Breast Cancer Res. Treat. 140, 427–434.

West, A.B., Moore, D.J., Choi, C., Andrabi, S.A., Li, X., Dikeman, D., Biskup, S., Zhang, Z., Lim, K.L., Dawson, V.L., and Dawson, T.M. (2007). Parkinson's disease-associated mutations in LRRK2 link enhanced GTP-binding and kinase activities to neuronal toxicity. Hum. Mol. Genet. 16, 223–232.

Whitfield, M.L., Sherlock, G., Saldanha, A.J., Murray, J.I., Ball, C.A., Alexander, K.E., Matese, J.C., Perou, C.M., Hurt, M.M., Brown, P.O., and Botstein, D. (2002). Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol. Biol. Cell 13, 1977–2000.

Winner, B., Melrose, H.L., Zhao, C., Hinkle, K.M., Yue, M., Kent, C., Braithwaite, A.T., Ogholikhan, S., Aigner, R., Winkler, J., et al. (2011). Adult neurogenesis and neurite outgrowth are impaired in LRRK2 G2019S mice. Neurobiol. Dis. 41, 706–716.

Yamaguchi, Y., and Miura, M. (2015). Programmed cell death in neurodevelopment. Dev. Cell 32, 478–490.

Zechel, S., Meinhardt, A., Unsicker, K., and von Bohlen Und Halbach, O. (2010). Expression of leucine-rich-repeat-kinase 2 (LRRK2) during embryonic development. Int. J. Dev. Neurosci. 28, 391–399.

Zhou, X., Liu, F., Tian, M., Xu, Z., Liang, Q., Wang, C., Li, J., Liu, Z., Tang, K., He, M., et al. (2015). Transcription factors COUP-TFI and COUP-TFII are required for the production of granule cells in the mouse olfactory bulb. Development 142, 1593–1605.

Zimprich, A., Biskup, S., Leitner, P., Lichtner, P., Farrer, M., Lincoln, S., Kachergus, J., Hulihan, M., Uitti, R.J., Calne, D.B., et al. (2004). Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron 44, 601–607.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
(1) Anti-beta III Tubulin (TUJ1)	BioLegend	Cat#801201; RRID:AB_2313773
(2) Ki67 anti-human	BD Biosciences	Cat# 550609, RRID:AB_393778
(3) LAMP2 (H4B4)	DSHB	Cat# H4B4, RRID:AB_2134755
(4) Tyrosine Hydroxylase TH (H-196)	Santa Cruz	Cat# sc-14007, RRID:AB_671397
(5) Anti-beta III Tubulin (TUJ1)	Abcam	Cat# ab107216, RRID:AB_10899689
(6) TOM20 (FL-145)	Santa Cruz	Cat# sc-11415, RRID:AB_2207533
(7) human SOX1	R&D systems	Cat# AF3369, RRID:AB_2239879
(8) HNF-3β (FoxA2 (RY7))	Santa Cruz	Cat# sc-101060, RRID:AB_1124660
(9) LMX1A	Abcam	Cat# ab139726, RRID:AB_2827684
(10) Anti-Cleaved PARP (Asp214)	BD	Cat# 552596, RRID:AB_394437
(11) LC3b	Cell Signaling	Cat# 2775, RRID:AB_915950
(12) SQSTM1	Cell Signaling	Cat# 5114, RRID:AB_10624872
(13) Beclin-1	BD Transduction Laboratories	Cat# 612113, RRID:AB_399484
(14) Anti- NR2F1	Abcam	Cat# ab181137, RRID:AB_2890250
(15) Anti- cleaved Caspase-3	Cell Signaling	Cat# 9661, RRID:AB_2341188
(16) Anti- Tyrosine Hydroxylase TH	Millipore	Cat# AB152, RRID:AB_390204
(17) Anti- FoxA2	Santa Cruz	Cat# sc-101060, RRID:AB_1124660
(18) Anti-rabbit β-III (Tubulin Tuj1)	Covance	Cat# MRB-435P-100, RRID:AB_663339
(19) Anti-mouse β-III (Tubulin Tuj1)	Sigma	Cat# T8660, RRID:AB_477590
(20) Anti-rabbit H3K27ac	Abcam	Cat# ab4729, RRID:AB_2118291
(21) Donkey anti-Mouse IgG 647	Thermo Fisher Scientific	Cat# A-31571, RRID:AB_162542
(22) Donkey anti-goat IgG 568	Thermo Fisher Scientific	Cat# A-11057, RRID:AB_142581
(23) Donkey anti-rabbit IgG 488	Thermo Fisher Scientific	Cat# A-21206, RRID:AB_2535792
(24) Goat IgG Isotype CTRL	R&D systems	Cat# AB-108-C, RRID:AB_354267
(25) Mouse IgG2a Isotype CTRL	Sigma	Cat# M5409, RRID:AB_1163691
(26) Rabbit IgG Isotype CTRL	Santa Cruz	Cat# sc-3888, RRID:AB_737196
(27) Goat anti-Chicken IgG 647	Thermo Fisher Scientific	21449
(28) Goat anti-Mouse IgG 568	Thermo Fisher Scientific	Cat# A-11031, RRID:AB_144696
(29) Goat anti-Rabbit IgG 488	Thermo Fisher Scientific	11034
(30) Goat anti-Mouse IgG1 647	Thermo Fisher Scientific	Cat# A-21240, RRID:AB_141658
Chemicals, peptides, and recombinant proteins		
hBDNF	PeproTech EC Ltd.	450-02
N-2 supplement	Thermo Fisher Scientific	17502048
E8 iPSC medium	Thermo Fisher Scientific	A1517001
DMEM/F12	Thermo Fisher Scientific	12634010
Neurobasal Medium	Thermo Fisher Scientific	21103049
Y-27632	Merck Millipore	688000-100mg
B-27 Supplement	Thermo Fisher Scientific	17504044
L-Ascorbic acid	Sigma-Aldrich	A4544-100 g
dbcAMP	Sigma-Aldrich	D0627
CHIR99021	Axon Medchem BV	AXON1386
SB 431542	Abcam	ab120163
Hoechst33342	Thermo Fisher Scientific	H21492

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
hGDNF	PeproTech EC Ltd.	450-44B
DMEM/F-12, w/o phenol red	Thermo Fisher Scientific	21041-025
Purmorphamin	Enzo Life science	ALX-420-045-M005
Dorsomorphin dihydrochlorid	Tocris	3093
B-27 Supplement Minus Vitamin A (50X), Liquid	Thermo Fisher Scientific	12587-010
Penicillin-Streptomycin, liquid	Thermo Fisher Scientific	15140122
Accutase	Thermo Fisher Scientific	A11105-01
GelTrex	Thermo Fisher Scientific	A1413302
TGF-β3	PeproTech EC Ltd.	100-36E
Paraformaldehyde	Sigma-Aldrich	P6148
Antimycin A	Abcam	ab141904
D-(+)- Glucose	Sigma-Aldrich	G8270-100 g
Dulbecco's Modified Eagle's Medium	Sigma-Aldrich	D5030-10X1L
CyQuant	Thermo Fisher Scientific	C7026
NUNCLON cell culture ware	Thermo Fisher Scientific	—–
MatriGel	Corning, Inc.	354277
Tween20	Sigma-Aldrich	P7949
TritonX	Carl Roth	3051.3
Bovine Serum Albumin	Sigma-Aldrich	A4503
Fetal Bovine Serum (FBS)	Thermo Fisher Scientific	10270106
Fluoromount-G	Southern Biotech	SOUT0100-01
EBSS	Thermo Fisher Scientific	24010043
HBSS (Hanks Balanced Salt Solution w/ Ca and Mg)	Thermo Fisher Scientific	14025-092
L-Glutamine	Thermo Fisher Scientific	25030-024
Sodium Pyruvate 100 mM	Thermo Fisher Scientific	11360-039
Deposited data		
Single cell RNA sequencing data	This paper	GEO: GSE128040
Figures and Source codes	This paper	https://doi.org/10.17881/y9k6-xa72
Experimental models: Cell lines		
iPSC line H1	Reinhardt et al., 2013b	N/A
iPSC line H3	Dr. Bill Skarnes (The Wellcome Trust Sanger Institute, Cambridge, UK)	N/A
iPSC line H2	Reinhardt et al., 2013b	N/A
iPSC line H4	Coriell	GM23338
iPSC line P1	Reinhardt et al., 2013b	N/A
iPSC line P1.1	Reinhardt et al., 2013b	N/A
iPSC line P2	Reinhardt et al., 2013b	N/A
iPSC line P4	Coriell	ND35367
Software and algorithms		
SEURAT R package	https://satijalab.org/seurat/	-
Monocle R package	http://cole-trapnell-lab.github.io/monocle-release/	2.4.0
GraphPad Prism	GraphPad Software, Inc.	7.03
MATLAB	The MathWorks, Inc.	R2017a (9.2.0.556244)
Illustrator	Adobe Systems, Inc.	19.2.1
Photoshop	Adobe Systems, Inc.	2015.1.2 20160113.r.355
ImageView	PerkinElmer, Inc.	1.1.0.0

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
FASTQC	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/	-
FlowJo	FlowJo, LLC	10.1r7
Destiny R package	https://bioconductor.org/packages/release/bioc/html/destiny.html	N/A
t-SNE	van der Maaten and Hinton, 2008	https://cran.r-project.org/web/packages/Rtsne/index.html
WAVE	Seahorse Bioscience, Inc.	2.3.0.20
Chemidoc Imaging System	BioRad	-
THOR	http://www.regulatory-genomics.org/thor-2/basic-intrstruction/	0.9.9

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Jens Schwamborn (jens.schwamborn@uni.lu). Certain materials are shared with research organizations for research and educational purposes only under an MTA to be discussed in good faith with the recipient.

Material availability
This study did not generate new unique reagents.

Data and code availability
scRNA-seq data have been deposited at GEO (GSE128040) and are publicly available. All the data of this manuscript is openly available https://doi.org/10.17881/y9k6-xa72.

The codes used in the paper are available on GitHub through the page https://doi.org/10.17881/y9k6-xa72.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human iPSCs were used in this study. Their use was approved by the local ethical committee (Ethic Review Panel of the University of Luxembourg, PDiPS project and CNER No 201305/04).

All mouse experiments were conducted in accordance with relevant national and international guidelines and regulations (European Union rules; 2010/63/UE), and have been approved by the local ethical committee in France (CIEPAL NCE/2019-548). *Nr2f1* heterozygous (*HET*) and homozygous (*KO*) mice were generated and genotyped as previously described (Armentano et al., 2006). Littermates of *HET* and *KO* mice with normal *Nr2f1* alleles were used as control mice (herein called *WT*). Midday of the day of the vaginal plug was considered as embryonic day 0.5 (E0.5). Control and mutant mice were bred in a 129S2/SvPas background. Both male and female embryos were used in this study; age is specified for each embryo used in specific experiments. Standard housing conditions were approved by local ethical committee in France (CIEPAL NCE/2019-548); briefly, adult mice were kept on a 12 hours light-dark cycle and housed three per cage with the recommended environmental enrichment (wooden cubes, cotton pad, igloo) with food and water *ad libidum*.

METHOD DETAILS

Cell Culture
The work with human iPSCs was approved by the local ethical committee (Ethic Review Panel of the University of Luxembourg, PDiPS project and CNER No 201305/04). The gender of each line is reported in Figure S1A. Cells were maintained under standard cell culture condition. iPSCs were maintained in E8 condition, on GelTrex matrix while splits were performed using Accutase. After splitting cells were incubated overnight (o/n) with 5 μM Y-27632 (Merck Millipore). NESCs were derived from iPSC as described and depicted in Figure S1C (Reinhardt et al., 2013a). Cells were maintained on MatriGel matrix NUNCLON cell culture ware or Cell carrier-96 plates for imaging. N2B27 maintenance media formulation: Neurobasal, DMEM-F12 (1:1), 1x P/S, 1x L-Glutamine, B27 (1:100), N2 (1:200) (ThermoFisher) freshly supplemented with 3 μM CHIR(−99021) (Axon Medchem), 0.75 μM Purmorphamine (PMA) (Enzo Life Science) and 150 μM ascorbic acid (AA) (Sigma) (key resources table). Medium was changed every other day. Cells were maintained with initial seeds at $5*10^4$ cells/cm^2 (counted in a Countess II, AMQAX1000 ThermoFisher). Weekly splits using 6 min Accutase digestion.

Neuronal differentiation

Cells were seeded at $5*10^4$ cells/cm^2 density for long-term (42 days) differentiations and $5.5*10^4$ cells/cm^2 for short-term (up to 14 days). Either on NUNCLON cell culture ware or Cell Carrier-96 plates from Perkin Elmer for HCS analyses, MatriGel matrix coated. Seeding densities were kept constant within a specific experiment. mDA differentiation was initiated two days after initial cell seeding by addition of differentiation media consisting of N2B27 freshly supplemented with: 10 ng/ml hBDNF (Peprotech), 10 ng/ml hGDNF (Peprotech), 500 μM dbcAMP (Peprotech), 200 μM ascorbic acid (Sigma), 1 ng/ml TGF-β3 (Peprotech), and 1 μM PMA (Enzo Life Science). PMA was withdrawn from the media after six days.

Chromatin Immunoprecipitation (ChIP)

The chromatin immunoprecipitation of hNESCs was performed as previously described (Gérard et al., 2019). For immunoprecipitation 30 μg of chromatin was used per reaction with 5 μg of an antibody against H3K27ac (Abcam, ab4729) and incubated overnight.

ChIP-seq

The sequencing of the ChIP samples was performed at the Genomics Core Facility in EMBL Heidelberg, Germany or at the sequencing platform of the Luxembourg Centre for Systems Biomedicine (LCSB) at the University of Luxembourg with single-end and unstranded reads sequenced in an Illumina HiSeq 2000 machine with read length of 50 bp or in an Illumina NextSeq 500 machine with read length of 75 bp, respectively. In total approximately 30 million raw reads per sample were obtained. The read quality control and mapping were performed as previously described (Gérard et al., 2019) using human reference genome GRCh38.

The enhancer and super-enhancer regions were called with Hypergeometric Optimization of Motif EnRichment (HOMER) version 4.7.2 (Hnisz et al., 2013) using the respective input samples as controls. For super-enhancer calling a stitching distance of 10 kb was used.

Identification of the core regulatory circuit (CRC)

ChIP-seq data of H3K27ac enrichment from NESCs and the HOMER called super-enhancer coordinates were used as an input for ROSE (Warren Andersen et al., 2013) to associate super-enhancers to actively expressed genes and to derive a list of super-enhancer controlled transcription factors in NESCs using the default parameters. The obtained matrix was in turn used as an input for CRC mapper (Saint-André et al., 2016) to derive a core regulatory circuit per NESC line using the default parameters.

Identification of differential H3K27ac peaks

THOR (Allhoff et al., 2016) was used to identify differential H3K27ac regions between conditions. H3K27ac signal normalization was performed using the TMM approach implemented in THOR and differential H3K27ac regions were considered significant at a p value of 0.05. Only differential H3K27ac peaks that show an absolute difference in read count superior or equal to 1000 were kept for downstream analysis.

Single-cell RNaseq using DropSeq

Microfluidics devices were fabricated using a previously published design (Macosko et al., 2015; Walter et al., 2019). All the procedures for obtaining the single-cell suspension as well the processing of the Dropseq data were performed as previously described (Walter et al., 2019).

Analysis of differentially regulated genes

The output of the DropSeq experimental setup (Figure S4A) is the expression matrix where rows represent genes, and each column is an individual cell. The elements of this matrix represent the measured intensity of gene expression, i.e., the measured number of mRNA molecules leading to the gene expression matrix.

To extract the relevant information on the developmental status of cells from the large dataset, we defined five lists of specific genes (Table S1). In particular, these lists contain genes covering stemness (NESC), dopaminergic neurons (mDA neurons), cell cycle, and related to pro-apoptotic and caspases genes.

For each of these lists, we compared the expression of the corresponding genes between genotypes within sampling points. Since the gene expression levels were measured at the single cell level, we analyzed the distribution of gene expression across cells. For this purpose, we consider in Figures 2A, 3A, and S4C histograms showing the distributions for cells of each group of the cumulative gene expression, for each of the lists considered above (rows) and for each day (columns), where cumulative gene expression corresponds to the sum of all mRNAs of listed genes leading to a single cumulative score for each cell. Since total numbers of mRNAs are not comparable between days the values in the horizontal axis are normalized, separately for each day and each list of genes, to the maximum of the cumulative gene expression (across the genotypes GC and G2019S) for that particular day and list. Thus, 1 corresponds to the maximal cumulative gene expression within one day and for one list of genes, while 0 corresponds to no expression for all of the genes on that list on that day.

We applied a z-test to assess a statistically significant difference between the means of the cumulative gene expression of the genotypes. The test was corrected with Bonferroni for multiple testing. Since between Figures 2B, 3C, 3E, and S4C the test was applied

to five lists of genes and across four different days, which means to correct for a total of 20 repetitions of the test. Standard p values thresholds are indicated on the panels as p values * < 0.05,** < 0.01, *** < 0.001, **** < 0.0001.

Furthermore, for each day we determined how many genes were differentially expressed between control and mutant, by applying to all genes (independently for each gene) the following three statistical tests: one-way ANOVA test, a one-way ANOVA test on ranks (Kruskal-Wallis test), and a Mutual Information based test. The minimum p value obtained by each gene across these three tests was retained, and any gene was considered differentially expressed with statistical significance when the corresponding p value was below the threshold of 0.01. In order to account for multiple testing, Bonferroni correction was included, which represents a conservative choice. The test was applied for all the approximately 20000 genes for each time point. This means that the individual p value for each gene is required to be < 0.01/20000 in order to claim statistical significance with an overall p value < 0.01. Figure S4B shows the number of genes that were differentially expressed between genotypes with a statistical significance corresponding to a p value < 0.01, for each day. In the tables also the percentages were reported, indicating how many of the genes of each individual list (of the five lists outlined above) are differentially expressed at each day. The colors indicate the list of genes to which the percentages refer to, and the color coding is the same as in Figure S4C.

Cell cycle analysis

To characterize the cellular maturity, we subsequently investigate the cell cycle in more detail. Figure 3B shows a subset of the expression matrices corresponding to cell cycle genes of Figure 3C for each day (columns), and each group (upper and lower row respectively), where the color scale indicates the intensity of expression. We further investigate cell cycle by means of dimensionality reduction, in particular, applying tSNE, in Figure S4D.

Normalization to the mean gene expression value was further performed to obtain the relative expression levels. This was done by subtracting the average expression value ($\log_2$(TPM+1)) of each gene from all the cells of the DGE matrix. Previous cell cycle analysis have shown two prominent gene expression programs (G1/S and G2/M) to contain several gene expressions of different cell cycle phases to overlap among the two (Macosko et al., 2015; Tirosh et al., 2016; Whitfield et al., 2002). Furthermore, these gene expression patterns have also been observed in the G1/S and G2/M cell cycle phases, in the bulk sample analysis of the differentially synchronized HeLa cells (Whitfield et al., 2002). Based on this data, a core set of 100 genes (G1/S) and 133 genes (G2/M) was considered for the cell cycle analysis (Table S1).

Due to the sparsity of the scRNaseq data, the expression for each cell cycle phase was refined by first evaluating the correlation between each of the genes in the scRNaseq data with the average gene expression values of all the genes involved in the respective cell cycle program (G1/S & G2/M), and by then including all the genes with high correlation value (R^2 > 0.3; p value < 0.05). Hierarchical clustering of the data demonstrates that some cells are cycling with high relative expression of most of the genes included in either of the cell cycle program or both of the programs, while other cells show basal expression for most of these genes (Figure 3D).

To determine the cell cycle score for individual cells and to set a threshold for classifying the cells as cycling cells, we proceed as follows. First, to determine the cell cycle score for individual cells (Figure S4D), the average value of the relative expression of all the genes involved in both cell cycle programs were evaluated. Second, the average basal expression score for the cell cycle program was calculated for all the genes involved in both cell cycle programs (G1/S and G2/M) for the 10% of the cell population showing the least expression magnitude. Finally, the classification of a cell as cycling or non-cycling was performed using the t test statistics on the expression value of different genes in both cycling programs, considering the averaged basal cell cycle score (cycling cells; FDR < 0.05).

Pseudo-temporal analysis

To cross-validate our analysis and to quantify the developmental differences between the PD2/WT and PD2/G2019S cell lines, we also employed the single cell R analysis package Monocle (Qiu et al., 2017). For a cross time point analysis, data of all 3055 cells from both cell lines and all time points were combined into one expression matrix shown in Figure S4A. We then performed the standard quality control steps of Monocle for gene filtering, considering a minimal expression threshold of 10% and only genes expressed in at least 10 cells, leading to a reduction of the number of considered genes from 20,766 to 16,992. To remove outliers and doublets, only cells with at least 2123 and less than 23350 transcripts were considered leading to 2954 cells. Using the same gene lists led to a very similar population composition compared to our customized analysis (see below).

For the cluster analysis across days, we reduced batch effects by excluding residuals caused by the number of expressed genes. The resulting clustering shown in Figure S4E demonstrates that cells at day 0 do not differ between the two cell lines but diversify during the differentiation process.

To quantify and compare the developmental state between the genotypes, we used the branching and pseudotime analysis provided by Monocle. For this purpose, we restricted our analysis to the most variable genes identified by the dispersion characteristics calculated by Monocle. Subsequently, we again reduced the dimensionality for stable calculation of cell trajectories based on reverse graph embedding (Qiu et al., 2017). Based on the resulting trajectories, Monocle identifies branching points and calculates pseudotimes by ordering cells on the resulting branches based on shortest paths similar to the diffusion map approach (Haghverdi et al., 2016).

Cluster detection, cell identity mapping, and quantification

For a deeper analysis of cell states and the differentiation process, we developed a custom analysis pipeline using MATLAB (version R2017a; Mathworks) scripts. To preprocess the gene expression matrices as described. We restricted the number of cells (columns)

to a subset of the most maximally expressed cells, according to various factors such as the minimum level of expression, the cumulative level of expression as described. This led to the number of cells used for each group and day as indicated (Figures S4G and S4H). Then, we performed a cluster analysis of this high dimensional data to investigate cell type-specific gene expression across different days, and for the different cell lines. We first reduced the dimensionality of the data by principle component analysis (PCA) (Bishop, 2006). This technique allowed obtaining the orthogonally transformed space that better captures the variance of the data. For the sake of data analysis and visualization, we kept only the first 20 principal components (PCs), reflecting more than 75% of the variance for all days.

Next, the resulting matrices for each day containing cells of both genotypes were used as inputs to the t-SNE (t-distributed stochastic neighbor embedding) algorithm (van der Maaten and Hinton, 2008). While PCA is transforming the data by capturing the highest mode of variation, allowing to preserve as much variance as desired and still reduce the dimensionality, t-SNE rather looks at finding a non-linear transformation that preserves the neighboring structure of nearby points in the original data. Therefore, its purpose its merely enabling the visualization of the data, originally in a high dimensional space, by projecting them to 2 or 3 dimensions. Ideally, the projected data may possess some non-random structure, such as separated clusters of points organized together, that will allow unveiling different properties and relations between the samples, in our case, the different cells. Using the results of PCA as input and mapping the data to three dimensions we selected the best viewpoint in terms of clearness of the existing clusters. When running t-SNE, we used the Barnes-hut algorithm for the approximate computation of the joint distributions (van der Maaten and Hinton, 2008), and Euclidean distance as a measure of similarity between points.

From the scatterplots obtained by t-SNE we already got some insights into the structure of the data. But to more deeply investigate the data, we designed a fully automated and unbiased pipeline for performing the different analyses presented in Figure 2A. In the next step, a cluster analysis on the resulting 3D data was performed by fitting a Gaussian mixture model using expectation maximization (Bishop, 2006). Each cluster is represented by one resulting ellipsoid, whose centroid and shape depends on the mean and covariance matrix of the underlying 3D Gaussian distribution fitted to the neighboring points. We fitted eight components in all conditions, as shown in Figure 2A.

In Figure 2A we color-coded the cells by their cell type identity. The seven potential identities considered were: oligodendrocyte, neuron, mDA neuron, microglia, endothelial, astrocytes, and NESC. Each of these phenotypic identities was defined by a list of genes known to be highly expressed in the corresponding cell types (Table S1). For each cell, we obtain then seven scores (one per identity) as the mean value of the expression level of all genes contained in each particular list. This way, we compared unbiasedly all these scores and just selected the one with the highest value as the mostly expressed identity, which finally enabled the generation of all scatterplots in Figure 2A.

Similarly, we have a different list of genes that defines different phenotypes and/or states during cell development (Table S1). Using these lists, we first obtained for each cell a mean expression value. Then, using these scores, we compared all the cells of one cluster to the cells in the remaining clusters by means of the Mann-Whitney U test in order to obtain a p value from the comparison. This test was chosen due to the fact that the normality assumption was not fulfilled. This was done for all possible comparisons, eight in our case, the number of fitted components. Then, we discarded all the comparisons where the mean expression level of all cells in the chosen cluster is smaller than the mean expression level of the remaining ones. Finally, among the comparisons left, we chose the one with the lowest p value. This way, we led to one cluster that we define as the "cluster with the highest expression," as opposed to the remaining cells. Additionally, the size of the marker depends on the expression level.

Additionally, we provide the percentages of either PD2/G2019S or PD2/WT cells in the highlighted cluster with respect to the total number of cells of that type, indicated with square and rounded markers respectively. The p value provides the statistical significance which of the genotypes is more highly concentrated within the highlighted cluster. This p value was obtained by running a permutation test on the data, i.e., we randomly shuffle the elements of the label vector that contains the correspondence between each cell and the group it belongs to. Then, using this shuffled vector, we counted again the number of cells of each genotype contained in the particular cluster. We repeated this permutation 20,000 times, finally leading to the distributions of a number of cells for each group in the cluster. Hence, given these distributions and the original number of cells of each genotype in the cluster, we computed a p value that accounts for an abnormal number of cells of one specific type in the highlighted cluster.

Data pre-processing scRNA-seq in organoids

From midbrain organoids datasets, cells having unique feature counts over 2,500 were removed as probable doublets or multiplets. Similarly, low-quality cells or empty droplets, were further filtered out with unique feature counts below 100 (for day35 data) - 200 (for day70 data), and mitochondrial transcripts above 30% (Figure S1). MT genes were removed from midbrain organoid datasets after QC by deletion of all row names with pattern 'MT'. After QC, WT35 midbrain organoids included 2864 cells, WT70 - 2005 cells, MUT35 - 2946 cells and MUT70 - 2660 cells. To better transmit the biological information between *in vivo* and *in vitro* ventral midbrain datasets, midbrain organoid data (WT35, WT70, MUT35 and MUT70) were integrated with embryonic midbrain data (La Manno et al., 2016) using Seurat integration analysis workflow (Stuart et al., 2019). Integration was performed based on the top 20 dimensions. RNA assay data of integrated object were log normalized and scaled to 10'000 transcripts per cell.

Cell type identification scRNA-seq in organoids

After integration of embryonic midbrain and midbrain organoid datasets, integrated object was scaled and principal component analysis (PCA) was applied. Cell clustering was performed based on the top 20 principal components using Louvain algorithm modularity optimization with a resolution of 0.5. UMAP was used for cell cluster visualization (Becht et al., 2018). Nine distinct cell clusters were identified in UMAP plot. Clusters 0 and 7 were present only in midbrain organoids, and located in a close proximity to each other in UMAP plot, indicating their high similarity and *in vitro* specificity. Because of this overclustering both clusters were pulled, resulting in eight distinct cellular identities. For cell type identification, binarized gene list across cell types from (La Manno et al., 2016) was used. For more details on how this list is generated please refer to (La Manno et al., 2016).

Characterization of the core CRC genes expression by sc-RNaseq

To investigate via scRNaseq the difference in expression of the five genes of the core CRC, we considered the gene expression across the cell populations (of 250 cells each) for the five transcription factors. We compared the expression for each of these genes within the PD2/WT cells with the one within the PD2/G2019S cells, for day 0, 10, 14 and 42. This was done by computing the gene expression fold-change of the G2019S compared to the WT (Figure 5A). For each gene and day, this is obtained as the logarithm in base two of the ratio between the average expression (across cells) of that gene in the G2019S, and that in WT, so negative numbers indicate that PD2/G2019S expresses that gene less than PD2/WT, positive numbers the converse. The number of cells considered is 250 per population and per day, and the p value comes from applying a z-test, and accounting for repeated hypothesis testing by applying Bonferroni correction for the total number of time the test was repeated for all the genes of the consensus. Error bars are obtained considering the error on each of the two averages (PD2/WT, PD2/G2019S) by the standard error of the mean (SEM), and combining the two using the variance formula for calculating error propagation. Therein, the statistical significance was assessed with a z-test, including Bonferroni correction for multiple hypothesis testing.

The expression levels of the transcription factor *NR2F1* are shown in detail in terms of cells distributions across gene expression in Figure 5C. These histograms exhibit the normalized *NR2F1* expression on the horizontal axes. They show the details of the distributions of cells corresponding to the red bars in Figure 5A. Statistical significance is again assessed by means of a z-test corrected by Bonferroni correction accounting for the repetition of this test on the full-list of consensus and non-consensus genes of the core CRC tested.

Furthermore, the average gene-gene correlation between the five genes of the consensus core CRC was computed and we retained only those correlations which were statistically significant (p value < 0.05). These were averaged over the number of gene-gene combinations. Gene-gene correlation is computed for each day and condition among each couple of genes, then averaged over genes. Correlations are filtered by their p value, thus assuming no correlation for couples of genes where the correlation is not statistically significant, i.e., whenever the corresponding p value is < 0.05 after applying Bonferroni correction for multiple hypothesis testing across all genes, days and conditions.

In vivo prototype mapping

To compare our *in vitro* cell preparation to *in vivo* cells embryonic human development we used, as a reference, a single-cell atlas of the embryonic human ventral midbrain (described in La Manno et al., 2016). To perform the comparison we used "prototype analysis" a machine learning approach that is described in more detail in that paper. In brief, the analysis strategy is to compare every single-cell transcriptome to a set of predefined cell types using a machine learning classifier. First a set of cell type identities, also "prototypes," are each defined by taking in consideration a group of cells from the human embryonic atlas. Then, a regularized logistic regression classifier is trained to discriminate cells from different prototypes. The model formulation is probabilistic and can output a probability for each possible assignment. The assigned prototype is the one of maximum likelihood.

With respect to the original paper (La Manno et al., 2016), we made the following minor changes. We performed a normalization of the NESC dataset. This was a normalization by total molecules across the cells of the sample and normalized the whole dataset to the same scale of the training dataset (the atlas). In particular, the normalization is this is: $\tilde{x}_{ij} = \sum \frac{\mu_j x_{ij}}{\sum_i x_{ij}}$ where $\mu_{y_i} = mean\left(\sum_i y_{ij}\right)$. x_{ij} and y_{ij} indicate the value of the ith gene for the jth cell for the test and train set respectively. This size normalization was meant to buffer two effects: cell-to-cell variation in RNA detection that in the DropSeq can be stark, and the difference in average sequencing depth of the test with respect to the reference dataset since they were generated with different technologies (STRT versus DropSeq).

Prototypes that were used to train the classifier are so defined: Embryonic stem cells (eES) prototype was learned by the model using eSCa, eSCb and eSCc as a reference. Radial glia-like cells type 1 (Rgl1) were trained on hRgl1; radial glia-like cells type 2 (Rgl2) trained on hRgl2, hRgl2b, hRgl2c; radial glia-like cells type 3 (Rgl3) trained on hRgl3, Neural progenitors (NProg) on hNProg, neuroblast mediolateral type 1 on hNbML1, serotonergic neurons (Sert) trained on hSert; dopaminergic neurons (DA) trained on hDA0, hDA1 and hDA2 populations, neuroblast medial (NbM) on hNbM; Early progenitor cells (Prog) on hProgFPM, hProgFPL, hProgM and hProgBP. The expression of the top 18 genes for the twenty best cells identified as DA, NbM and Prog are compared to their corresponding embryo standards in Figure S1.

Immunocytochemistry

Cells were fixed using 4% paraformaldehyde (PFA) in 1x phosphosaline buffer (1xPBS), pH 7.4, for 15min at room temperature (RT). Permeabilization was done using 0.3% Triton X-100 in 1xPBS for 5 min at RT. Unspecific antibody (AB) binding was avoided by blocking buffer incubation (5% FCS, 0.1%Tween20 in 1xTBST) for 1h at RT. Primary AB binding was performed o/n at 4°C using indicated primary AB in blocking buffer. Following incubation and 3x 1xPBS washing steps, secondary AB binding was performed for 1h at RT, using all secondary AB at 1:1000 dilution, including 1:1000 Hoechst33342. Following incubation, cells were washed 3x using 1xPBS. For HCS analysis cells were imaged while covered with 1xPBS. Image acquisition was done within 14 days after ICC in automated unbiased manner in an OPERA HCS microscope (PerkinElmer). Primary ABs used with this setup: 2, 4, 5, 9, 10. Images were acquired on a spinning disc confocal microscope (Opera QEHS, Perkin Elmer). Four fluorescence channels were measured sequentially. First, Hoechst was excited with a 405 nm laser and emitted fluorescence was detected behind a 450/50 bandpass filter. Second, Alexa 488 was excited with a 488 nm laser and detected behind a 520/35 bandpass filter. Third, Alexa568 was excited with a 561 nm laser and detected behind a 600/40 bandpass filter. Finally, Alexa647 was excited with a 640 nm laser and detected behind a 690/70 bandpass filter. For each channel, a 20x water immersion objective with N.A. 0.7 was used in combination with camera binning 2.

Image analysis

Nuclei were detected using pre-processing of the Hoechst channel with a difference of Gaussians and thresholding. For the difference of Gaussians, the foreground image was convolved with a Gaussian of size 10 and standard deviation 2 and the background image was convolved with a Gaussian of size 60 and standard deviation 20 (NucDoG). The nuclei mask was defined via thresholding of NucDoG (> 20), and by removing connected components with less than 200 pixels (NucMask). Only images with a total nuclei area of at least 15000 pixels were considered for further analysis. To quantify apoptosis, the raw nuclei channel was low pass filtered with an average filter of side length 5 and thresholded (> 1000, NucMaskHigh).

For segmentation of TH positive structures, local and global thresholding approaches were combined. For global thresholding, the TH channel was low pass filtered with a Gaussian filter of size 10 and standard deviation 1, and thresholded (> 300, TH_Mask_Global). For local thresholding, a difference of Gaussians was applied. The foreground image was convolved with a Gaussian of size 11 and standard deviation 1 while the background image was convolved with a Gaussian of size 21 and standard deviation 7 (TH_DoG). The resulting image was thresholded (> 200, TH_Mask_Local). The final TH mask was defined via Boolean OR operation between TH_Mask_Global and TH_Mask_Lokal, and by excluding nuclei and connected components with less than 100 pixels (TH_Mask).

For the segmentation of neurons based on Tuj1 fluorescence, local and global analyses were combined. For global thresholding, the Tuj1 channel was low pass filtered with a Gaussian of size 10 and standard deviation 3 (Tuj1LP). The resulting image was thresholded (> 150, Tuj1_GlobalMask). For the local approach, a background image convolved with a Gaussian of size 20 and standard deviation 6 was subtracted from a foreground image convolved with a Gaussian of size 10 and standard deviation 3 (Tuj1DoG). This difference of Gaussians was thresholded (> 3, Tuj1_LocalMask). To define the neuronal mask (NeuroMask), Tuj1_GlobalMask and Tuj1_LocalMask were combined via boolean OR operation and connected components with less than 200 pixels were removed.

PARP fluorescence was segmented via global thresholding. The PARP channel was pre-processed via low pass filtering with a Gaussian of size 10 and standard deviation 1 (PARP_LP). After thresholding (> 50), connected components with less than 50 pixels were removed (PARP_Mask).

Ki67 was also segmented using global thresholding. The Ki67 channel was low pass filtered with a Gaussian filter of size 10 and standard deviation 1 (Ki67_LP). This low pass filtered Ki67 image was thresholded (> 50), and in the resulting mask, connected components with less than 50 pixels were removed (Ki67_Mask).

Flow Cytometry Analyses

Cell preparation for stainings was done under sterile conditions. At indicated time points and stages, cells were washed using warm 1xPBS and dissociated using Accutase digestion at 37°C, incubation time dependent on densities, minimal 10 min and maximal 30min. Cells were pelleted for 5 min at 300 r*cf.* The single cell suspension was prepared by dislodging the pellet in cold 1xPBS and the cell clumps and debris were excluded from the suspension using 40 μm Nylon cell strainer (BD). Cells were pelleted at 400 r*cf.* for 5 min at 4°C. The resulting pellets were dissociated in 100 μL 10% FBS-PBS. While vortexing gently 500 μL 4%PFA-1xPBS were added drop-wise to the suspension, cells in fixative were incubated for 15 min at RT. Fixed cells were pelleted at 800 r*cf.* And washed twice in 10% FCS-PBS.

Samples were subjected to flow cytometric imaging protocol. Samples were split for different antibody staining combinations, and a similar number of cells for each sample was stained (ABs used: 1, 4, 7, 8, 20-25, 15, 16, 17, 18, 19 key resources table). Cell permeabilization was performed for 20 min at 4°C using permeabilization buffer 1 mL 0.05% Saponin-1% BSA-1xPBS. Next, cells were subjected to titrated primary antibodies and isotype controls in equal concentration in permeabilization buffer for 1h at 4°C. After two 10%FBS-1xPBS wash steps cells were resuspended in secondary antibody mix and incubated for 30 min at 4°C. Following two 10% FBS-1xPBS washing steps, cells were resuspended in 1xPBS and analyzed in a Fortessa flow cytometry analyzer (BD Biosciences), Becton Dickinson Biosciences. Resulting flow cytometric data was further processed using FlowJo, LLC software. First, we gated-out doublets using conservative double gating via SSH-A/SSH-H and SSH-A/SSC-W). Samples were further analyzed while considering appropriate negative, compensation, and isotype controls. Gated quantitative numbers were subjected to different groupings

and further statistically analyzed as indicated. Cells were dissociated using Accutase incubation for 8 min, at 37°C. Accutase activity was reduced using DMEM/F12 w/o phenol-red. After 5 min 400 r*cf.* centrifugation cells were washed once in 1 mL ice cold DMEM/F12 w/o phenol-red. In the last step, cells were dissociated in 300 μl DMEM/F12 w/o phenol-red and stained with final concentration propidium iodide 0.33 μg/ml for 2 min at 4°C.

Quantitative real time PCR

NESCs and differentiated neurons (day eight of differentiation) were collected by scraping in cold PBS and lysed in RLT buffer (QIAGEN) containing 5% β-mercaptoethanol. For each cell line used, three independent samples from three different passages were taken (n = 3 per line). RNA isolation was performed using the RNeasy Mini kit (QIAGEN) including an on-column DNA digestion (QIAGEN) according to the manufacturer's instructions. RNA quantities were measured using the NanoDrop 2000c Spectrophotometer (Thermo Scientific). 500ng RNA per sample were reversely transcribed to cDNA using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). cDNA samples were diluted 1:10 before 5 ul per sample were subjected to quantitative real-time PCR using the TaqMan® Gene Expression Assays (Applied Biosystems) for NR2F1 (assay ID Hs00818842_m1) and RPL37 (assay ID Hs03044965_g1) as reference gene. FAM signal of gene-specific probes was measured on the AriaMx Real-Time PCR System (Agilent Technologies). The ΔΔct method was applied to transform individual cycle threshold (ct) values into the relative expression of NR2F1 (fold change) referenced to RPL37 and normalized to the H1/WT group.

Immunohistochemistry

Mouse embryonic brains/whole heads were dissected and fixed in 4% paraformaldehyde (PFA) at 4°C for 3 h in agitation, then washed in PBS 1X and dehydrated in 25% sucrose overnight at 4°C. Primary antibodies used: NR2F1 (Abcam ab181137, 1:1000, rabbit; R&D H8132, 1:1000, mouse); cleaved Caspase-3 (Cell signaling #9661, 1:2000, rabbit; Tuj1 (β-III Tubulin, Covance MRB-435P, 1:1000, rabbit; or Sigma T8660, 1:1000, mouse); FoxA2 (Santa Cruz sc-101060, 1:1000, mouse); TH (Millipore AB152, 1:1000, rabbit). All antibodies required antigen retrieval (10 minutes at 95°C in pH = 6 Citric acid solution). Alexa Fluor 488, 555, 594 and 647 anti-mouse or anti-rabbit IgG conjugates (Thermo Fisher scientific, all 1:500) were used as secondary antibodies. Images were acquired at an Apotome Zeiss, using the AxioVision software. For cell percentage/number quantification after immunofluorescence (IF), measurements were performed on at least 9 sections coming from 2 to 3 different animals. Fixed embryos with damaged tissues were excluded from any further analysis/processing. Microscope images were processed with Photoshop or ImageJ software, by randomly overlapping fixed-width (100μm) rectangular boxes on the area of interest, then quantifying positive cells or pixel intensity inside the boxes. When calculating percentages over the total cell number, the latter was quantified by counting DAPI⁺ nuclei, unless otherwise specified.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis was assay and purpose dependent as indicated in the figure legends. We applied: unpaired or paired (isogenic controls) Student's t test when sample size was small and data normally distributed. For large datasets like HCS imaging data that was not normally distributed Mann-Whitney-Test without further cleaning of the data was applied. Two-way ANOVA with Bonferroni correction was applied when indicated. Tests applied indicated in the particular legends. All data are presented as mean ± SEM. Significance levels were set at*/† $p < 0.05$, **/†† $p < 0.01$, ***/††† $p < 0.001$, ****/†††† $p < 0.0001$. For statistical analysis, GraphPad Prism was used. Statistical analysis of single-cell RNaseq data was handled differently and as indicated. In particular, the statistical significance of the differences in cumulative gene expressions (Figures 2B, 3C, and S4C) were assessed using a z-test, with Bonferroni correction to account for multiple testing. One-way ANOVA together with one-way ANOVA on ranks and Mutual Information where used to identify genes which are differentially expressed with statistical significance, again considering Bonferroni correction to account for multiple testing. The details of these different statistical analyses of single-cell RNaseq data are explained in the section titled "analysis of differentially regulated genes" of this supplementary material. Further statistical aspects of the analysis of the single-cell RNaseq data are explained in detail in the sections "Cell cycle analysis," "Pseudo-temporal analysis," "Cluster detection, cell identity mapping, and quantification" and "Characterization of the core CRC genes expression by sc-RNaseq."

For the scRNaseq data in organoids, the statistical analysis was performed with RStudio R version 3.6.2 using ggplot2 package. For all comparison non-parametric Kruskal Wallis test was performed. Statistical significance between comparisons are represented with asterisks: $p < 0.05$ *, $p < 0.01$**, $p < 0.001$***, $p < 0.00001$****.

For the immunohistochemistry analysis in animals, data were statistically analyzed and graphically represented using Microsoft Office Excel software and GraphPad Prism (version 7.00). Quantitative data are shown as the mean ± standard error (SEM). Data were compared by 2-way ANOVA (analysis of variance; for comparison of three or more groups) and statistical significance was set as follows: * = $p \leq 0.05$; ** = $p \leq 0.01$; *** = $p \leq 0.001$.