UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

A Functional Genomics Dissection of Parkinson's Disease Risk Genes in Human Dopaminergic Neurons Using Isogenic Stem Cell Models and CRISPRi Perturb-seq

Permalink

https://escholarship.org/uc/item/24z5p8kp

ISBN

9798189280047

Author

Dunnack, Jesse

Publication Date

2026-05-01

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

A Functional Genomics Dissection of Parkinson's Disease Risk Genes in Human Dopaminergic Neurons Using Isogenic Stem Cell Models and CRISPRi Perturb-seq

By

Jesse Dunnack

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Dirk Hockemeyer, Co-Chair

Professor Helen Bateup, Co-Chair

Professor David Schaffer

Professor Liana Lareau

Professor Daniela Kaufer

Spring 2026

Abstract

A Functional Genomics Dissection of Parkinson's Disease Risk Genes in Human Dopaminergic Neurons Using Isogenic Stem Cell Models and CRISPRi Perturb-seq

by

Jesse Dunnack

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Dirk Hockemeyer, Co-Chair

Professor Helen Bateup, Co-Chair

Parkinson's disease (PD) is the second most common neurodegenerative disorder, largely defined by the progressive degeneration of midbrain dopaminergic neurons. While about 10% of PD cases can be attributed to heritable, highly-penetrant mutations in genes such as *SNCA*, *LRRK2*, *PRKN*, and *PINK1*, a vast majority of cases are idiopathic and have comparatively elusive genetic risk contribution. Genome-wide association studies (GWAS) have identified more than 90 risk loci in the human genome which nominate over 300 genes for their potential role in the etiology of PD. Despite this expanding catalog, the affected cell types and molecular mechanisms by which these putative risk genes may contribute to disease remain unclear. A central question is whether the cellular mechanisms nominated by monogenic PD genes generalize to the broader set of GWAS-nominated genes.

To address this question, our group established iSCORE-PD, a collection of 65 human pluripotent stem cell lines (hPSCs), engineered in an isogenic background to harbor 13 pathogenic mutations across 11 monogenic-PD genes. We additionally developed an inducible CRISPRi perturbation system in the same hPSC background, facilitating a large-scale Perturb-seq screen of both familial and GWAS-nominated risk genes. Across two experiments in midbrain dopaminergic neurons and other PD-relevant cell types, we profiled over 85,000 single cells and identified perturbation- and cell-type-specific signatures which implicate mitochondrial quality control, proteostasis, and autophagy pathways across multiple familial and GWAS nominated PD risk genes. Together, these data work to resolve the genetic heterogeneity of PD into a tractable set of cell-state-dependent transcriptional programs shared among familial and GWAS-PD risk genes.

A Dedication:

First and foremost, **To my parents:**

Since I was a kid, you always told me I could do anything I put my mind to.

Though I'd like to believe it's true, it understates the role your boundless love and support have played at each and every step along the way.

<u>There are no words to properly express my gratitude.</u>

My appreciation extends to every person that's supported me, inspired me, or helped preserve my sanity during this endeavor:

This is for Kaitlyn, Brooklyn, (little) Weston, and the even-littler-one on the way.

It's for Alex, (big) Weston and for Luke.

For Shayan, Antonio, Grayson, Kim and Jenny.

And of course:

It's for Lady, Scout, Kitty, and every other animal

that put a smile on my face

when I needed it

the most.

<3

Table of Contents

Chapter 1 — Introduction 1

1.1 Parkinson's disease: clinical features and neuropathology 1

1.1.1 Clinical presentation and epidemiology 1

1.1.2 Neuropathological hallmarks of PD 1

1.2 Selective vulnerability of DA neurons in the substantia nigra 2

1.3 The genetic architecture of Parkinson's disease 4

1.3.1 Monogenic Parkinson's disease 4

1.3.2 Idiopathic Parkinson's Disease and GWAS 7

1.4 Molecular mechanisms of Parkinson's disease pathogenesis 9

1.5 hPSC–derived neurons as a model system for Parkinson's disease 13

1.6 CRISPRi and Perturb-seq as tools for functional genomics 15

1.7 Dissertation framework and thesis organization 16

References 18

Chapter 2: iSCORE-PD — An Isogenic Stem Cell Collection to Research Parkinson's Disease 32

2.1 Chapter Preamble 33

2.3 Results 36

2.3.1 - Characterization of the WIBR3 hESC line 36

2.3.2 - WIBR3 cells differentiate into PD-relevant cell types 39

2.3.3 - Genetic engineering of PD-associated mutations 44

2.3.4 – Quality control of genetically modified hPSCs 46

2.3.5 – Composition of the iSCORE-PD cell line collection 48

2.3.6 – WGS reveals minimal editing-induced genetic variability 50

2.3.7 – Off-target effects are rare in CRISPR/Cas9-edited and absent in prime-edited clones 50

2.4 – Discussion 53

Limitations of this study 55

2.5 Chapter summary and transition to Chapter 3 56

2.6 – Materials and Methods 58
2.6.1 – hPSC culture 58
2.6.2 – Array genotyping 58
2.6.3 – Long-read sequencing and variant analysis 58
2.6.4 – Molecular cloning for genome editing 59
2.6.5 – Genome editing of hESCs 59
2.6.6 – Editing pipelines 60
2.6.7 – NGS-based identification and validation of targeted clonal lines 60
2.6.8 – Zygosity confirmation and Southern blotting 60
2.6.9 – Cortical spheroid differentiation and p53 irradiation assay 61
2.6.10 – Dopaminergic neuron differentiation 62
2.6.11 – scRNA-seq library preparation and analysis of DA neurons 62
2.6.12 – Microglia differentiation 63
2.6.13 – Immunocytochemistry and qRT-PCR 63
2.6.14 – Data and code availability 64
2.6.15 – Short-read whole-genome sequencing pipeline 64
2.6.16 – Analysis pipeline to identify clone-specific SNVs/indels 64
2.6.17 – Off-target analysis 65
2.A.1 – Control cell lines 66
2.A.2 – *SNCA* (PARK1) 66
2.A.3 – *PRKN* (PARK2) 66
2.A.4 – *PINK1* (PARK6) 67
2.A.5 – DJ-1 (*PARK7*) 67
2.A.6 – *LRRK2* (PARK8) 67
2.A.7 – *ATP13A2* (PARK9) 68
2.A.8 – *FBXO7* (PARK15) 68
2.A.9 – *DNAJC6* (PARK19) 69
2.A.10 – *SYNJ1* (PARK20) 69
2.A.11 – *VPS13C* (PARK23) 69

2.A.12 – *GBA1* 70
References 71
Chapter 3: CRISPRi identifies cell-type-specific signatures of PD risk genes 84
3.1 Introduction 84
3.2 Establishing a Perturb-seq system in hPSC-derived dopaminergic neurons. 85
3.2.1 Generation and validation of CRISPRi hPSC lines 85
3.2.2 Design, production + delivery of familial-PD and GWAS-PD gRNA libraries 88
3.2.3 DA neuron differentiation, perturbation induction, and scRNA-Seq 89
3.3 Characterization of the familial-PD CRISPRi dataset 91
3.3.1 Dataset generation and quality-control metrics 91
3.3.2 Cluster identity and reference-dataset comparison 93
3.3.3 Distribution of perturbations and evaluation of on-target knockdown 94
3.3.4 Differential expression and pathway enrichment 95
3.4 Mixscale reveals cell, perturbation-specific signatures of familial-PD genes 98
3.4.1 Analysis workflow and interpretive framework 98
3.4.2 *VPS35* 101
3.4.3 *PARK7* / DJ-1 106
3.4.4 *FBXO7* 111
3.4.5 *VPS13C* 119
3.4.6 Abridged summary of additional familial-PD perturbation signatures 121
GIGYF2 121
SYNJ1 123
POLG 123
PRKN (PARK2) 124
3.5.4 Perturbation signatures unique to the late-progenitor cluster 124
NR4A2 perturbation in late progenitors activates senescence programs 124
***EIF4G1* perturbation in late progenitors downregulates KEGG PD pathway** 125
3.5 CRISPRi Perturb-seq of GWAS-nominated PD risk genes 126
3.5.1 Characterization of the GWAS Perturb-seq dataset 126

3.5.2 Perturbation quality control and analysis overview 127
3.5.3 Overview of notable GWAS-PD perturbation signatures 128
SYT4 128
ATP6V0A1 129
3.6 Convergent themes across Perturb-seq experiments 131
3.6.1 V-ATPases / lysosomal acidification 134
3.6.2 Autophagy and protein-aggregate clearance 134
3.6.3 Mitochondrial dysfunction 135
3.6.4 Translation and ribosome biogenesis 136
3.7 Study limitations 137
Direction-specific transcriptional responses across familial-PD perturbations 139
3.8 Implications and future directions 141
Cross-platform convergence 141
Nuances of allele-specific effects not captured in CRISPRi knockdown 142
Technical factors and experiment improvements 142
Extending Perturb-seq to other PD-relevant cell-types 142
3.9 Materials and Methods 143
Parental hESC line and culture conditions 143
Generation of Cre-inducible CRISPRi hESC lines 144
Clone validation in hESCs 144
gRNA library design 145
Oligo synthesis and library cloning 146
Lentiviral packaging 146
Viral titer and transduction 146
Dopaminergic neuron differentiation 147
Cre-mediated induction of CRISPRi 148
Immunocytochemistry 148
FACS and sample preparation 148
Single-cell RNA-Seq library preparation and sequencing 149

scRNA-Seq initialization 149
Differential Expression analysis with Mixscale 151
Gene Set Enrichment Analysis (GSEA) 152
Western Blot 153
Software and data availability 153
References 155

List of Figures

Figure 2.1A/B 37
Figure 2.1C 38
Figure 2.1D/E 39
Figure 2.2A/B 40
Figure 2.2C/D/E 41
Figure S2A/B/C. 42
Figure S3A/B/C 43
Figure 2.2F-N 44
Figure 2.3A/B 45
Figure 2.3C 46
Figure 2.3D 47
Figure 2.5 52
Figure 3.1 86
Figure 3.2 86
Figure 3.3 87
Figure 3.4 87
Figure 3.5 88
Figure 3.6 89
Figure 3.7 90
Figure 3.8 90
Figure 3.9 91
Figure 3.10. 92
Figure 3.11 93
Figure 3.12. 94
Figure 3.13. 95
Figure 3.14 96

Figure 3.15. 97
Figure 3.16. 99
Figure 3.17. 100
Figure 3.18 102
Figure 3.19 103
Figure 3.20 104
Figure 3.21 105
Figure 3.22 106
Figure 3.23. 107
Figure 3.24 108
Figure 3.25. 109
Figure 3.26. 110
Figure 3.27. 112
Figure 3.28 113
Figure 3.29. 114
Figure 3.30. 115
Figure 3.31. 117
Figure 3.32 118
Figure 3.33. 120
Figure 3.34 121
Figure 3.35 122
Figure 3.36. 127
Figure 3.37 128
Figure 3.38. 130
Figure 3.39. 132
Figure 3.40 133

List of Tables

Table 2.1 49

Chapter 1 — Introduction

1.1 Parkinson's disease: clinical features and neuropathology

1.1.1 Clinical presentation and epidemiology

Parkinson's disease (PD) is a progressive neurodegenerative disorder defined clinically by a core of motor deficits (bradykinesia, resting tremor, rigidity, postural instability) together with a widening recognition of non-motor features including autonomic dysfunction, REM-sleep behavior disorder, and late-stage cognitive decline[1]. PD is the second most common neurodegenerative disease and the fastest-growing neurological disorder by prevalence, affecting an estimated 6 to 10 million people worldwide with incidence rising sharply after age 60[2,3]. Current therapies are largely limited to symptom management: pharmacological supplementation of striatal dopamine with L-DOPA remains the standard of care a half-century since its introduction, and no treatment to date has been shown to slow or halt the underlying neurodegenerative process[4]. The development of disease-modifying interventions for PD requires both a refined understanding of the molecular events which render specific neurons vulnerable and, ideally, reliable biomarkers that allow for intervention before these critical populations are lost.

1.1.2 Neuropathological hallmarks of PD

The defining cellular pathology of PD is the selective, progressive loss of dopaminergic (DA) neurons within the substantia nigra pars compacta (SNc) of the midbrain. These cells compose the nigrostriatal tract, where cell bodies situated in the SNc extend long-range axonal projections with extensive arbors which project onto the dorsal striatum. In the healthy brain, nigrostriatal DA neurons provide tonic release of dopamine onto GABAergic medium spiny neurons in the striatum, supporting its central role in the selection and initiation of voluntary movement[1,5]. Aside from DA neuron loss, a second pathological hallmark of PD is the intraneuronal accumulation of Lewy bodies, inclusions of insoluble protein aggregates composed predominantly, though not exclusively, of aggregated α-synuclein[6]. Lewy body pathology follows a stereotyped spatiotemporal pattern described by Braak and colleagues, ascending from dorsal motor and olfactory nuclei through midbrain and limbic structures into the neocortex[7]. This anatomical progression generally parallels the clinical course in many patients, though progression is heterogeneous. Symptoms may begin with prodromal autonomic dysfunction, followed by advancing motor deficits as dopamine release onto the striatum becomes compromised, and as Lewy body pathology spreads to the cortex in late-stage disease, cognitive decline[1,7]. This asymmetry extends to the cellular pathology itself: at the time of motor diagnosis, approximately 60-70% of SNc DA neuron cell bodies are lost[8–11], while striatal DA axon terminal density is typically depleted by 50 to 80% in the putamen[8,12], supporting the hypothesis that distal axons degenerate substantially before the cell bodies they belong to[12].

The exact relationship between α-synuclein aggregation and neuronal death is an area of active debate: whether Lewy bodies represent a direct driver of cellular injury, a protective sequestration response, or simply a consequence of upstream cellular dysfunction remains to be settled[13,14].

1.2 Selective vulnerability of DA neurons in the substantia nigra

Several cell-autonomous features of nigrostriatal DA neurons are thought to contribute to their particular vulnerability in PD. First is their distinct physiology: these cells demonstrate autonomous tonic firing (~2 to 10 Hz) in the absence of excitatory input, with characteristically broad action potentials and periodic activation of voltage-gated calcium channels, which results in sustained oscillations of cytoplasmic Ca^{2+} concentrations[15]. Although intracellular Ca^{2+} levels are generally buffered by calcium-sequestering proteins like calbindin (CALB1), the persistent firing by DA neurons in the SNc demands a pool of free Ca^{2+} ions to maintain activity. When compared to DA neurons located in the ventral tegmental area (VTA) which exhibit similar tonic firing activity, cells in SNc show less CALB1 expression and consequently, comparatively less Ca^{2+} buffering capacity. Mitochondrial uptake of cytoplasmic Ca^{2+} promotes oxidative phosphorylation (OXPHOS), which despite supplementing cellular levels of ATP, serves to increase mitochondrial oxidative burden[15]. In addition, nigrostriatal DA neurons have remarkable axonal arborization, with substantial bioenergetic demand, and exhibit elevated basal levels of mitochondrial respiration and an increased oxidative stress burden as compared to cells in VTA[16,17]. Taken together, these features are thought to account for why cells in SNc are disproportionately affected in PD. They fail, however, to explain a gradient of vulnerability among DA neurons in SNc, which generally share the features described above.

Deeper investigations of DA neuron loss in PD have made clear that vulnerability is graded rather than uniform within SNc. Several groups working with post-mortem human SNc from PD patients and matched controls have worked to resolve this gradient at single-cell resolution. In the most comprehensive profiling effort to date, Kamath and colleagues sequenced 22,048 DA neuron nuclei, identifying ten transcriptionally defined subtypes along an established *SOX6*$^{+}$*/CALB1*$^{+}$ axis of variation which reflects the specification of SNc and VTA, respectively[18–22]. Of these, the authors identified a particular subpopulation in ventral SNc (*SOX6*$^{+}$*/AGTR1*$^{+}$) which was significantly depleted in PD, and found to be enriched in the expression of GWAS-nominated PD risk genes[18]. Independent work corroborates the vulnerability signature: Martirosyan and colleagues similarly report depletion of an AGTR1+ DA cluster in PD patients[23], and Smajić and colleagues identified a disease-associated CADPS2^{HIGH}, TH^{LOW} cluster[24]. Dorsal CALB1+ subtypes also decline in PD, though more slowly; the apparent proportional increase of CALB1+ cells in some datasets reflects faster SOX6_AGTR1 collapse, not absolute sparing[18]. This asymmetric loss compounds a general limitation of post-mortem sampling: the most vulnerable cells are already depleted by the time tissue is available, leaving the earliest-dying populations underrepresented in every PD dataset. Selective vulnerability in PD is therefore best

understood as a continuous gradient: all SNc DA neurons are at risk, but a ventrolateral subpopulation degenerates earliest and most extensively.

While the transcriptional basis of regional vulnerability in DA neurons has yielded crucial insights, it is also clear that non-cell autonomous factors also play a role. Non-neuronal cell types in the midbrain, along with peripheral immune and enteric populations, contribute to the pathological environment in which DA neurons degenerate, and human genetic evidence increasingly implicates these populations as active contributors to disease risk. Cell-type enrichment analyses of PD GWAS summary statistics have repeatedly identified non-neuronal signal alongside or exceeding the DA neuron signal: Bryois and colleagues found enrichment for genetic associations in cholinergic and monoaminergic neurons but also in enteric neurons and oligodendrocytes when PD heritability was partitioned across the mouse nervous system[25]; Reynolds and colleagues showed that PD heritability is enriched in a curated lysosomal gene set most highly expressed in astrocytes, microglia, and oligodendrocyte subtypes rather than in a single brain-specific cell type[26]; and Andersen and colleagues, applying stratified LD score regression to cell-type-specific open chromatin, localized the strongest PD heritability signal to microglia, nominating *P2RY12* as a candidate microglial effector at a genome-wide-significant locus[27]. Independent experimental work identifies specific non-cell-autonomous mechanisms. Additional work has identified α-synuclein-specific T cells in the blood of PD patients, with one study showing approximately 40% of PD patients demonstrated an immune response to α-synuclein epitopes, consistent with an adaptive immune contribution to DA neuron loss[28]. Injection of α-synuclein preformed fibrils into the gastric and duodenal muscularis of wild-type mice induces caudorostral spread of phosphorylated α-synuclein pathology from the dorsal motor nucleus of the vagus to the SNc, with accompanying DA neuron loss and motor deficits. Importantly, truncal vagotomy prevents this spread[29].

One central contributor to the specific vulnerability of DA neurons comes from dopamine itself. Cytosolic dopamine auto-oxidizes to reactive quinones and superoxide and, via monoamine oxidase, is metabolized to the highly reactive aldehyde DOPAL, species that covalently modify cysteines on proteins including α-synuclein, parkin, DJ-1, and UCH-L1[30,31]. Across the human lifespan the SNc accumulates neuromelanin, a catecholamine-derived polymer that sequesters cytosolic catechols and binds iron at concentrations far above surrounding tissue; upon cell death neuromelanin is released into the extracellular space where it can activate microglia[32]. Iron accumulation in the SNc has attracted particular attention as a potential link between dopamine metabolism, oxidative stress, and ferroptotic cell death, a connection that reappears in the Perturb-seq results presented in Chapter 3 of this dissertation.

Though cell-autonomous and non-cell-autonomous factors have been studied extensively in animal models, the extent to which rodent and primate systems recapitulate human DA neuron vulnerability has important limits. Rodent midbrain accumulates far less neuromelanin than human midbrain, a species difference widely considered to contribute

to the weaker neurodegenerative phenotypes observed when human PD genes are modeled in mice[32]. More broadly, many PD-associated mutations that cause clear disease in humans produce subtle or no overt neurodegenerative phenotypes in rodent models, a limitation widely acknowledged to reflect fundamental species differences in DA neuron biology, gene expression, electrophysiology, and connectivity[33,34]. Although primate models may more accurately reflect certain electrophysiological, neuroanatomical and behavioral hallmarks of PD as compared to rodent models, they fail to faithfully recapitulate Lewy body pathology, and this goes without mentioning the ethical and financial considerations regarding primate studies[33]. These constraints motivate the experimental approach taken in this dissertation: perturbing PD-implicated genes in isogenic human DA neurons *in vitro*, where early transcriptional responses can be read out in cells that have not yet degenerated.

1.3 The genetic architecture of Parkinson's disease

Cases of PD can be classified into one of two categories based on the genetic drivers of the disease, classified either as monogenic or idiopathic PD[35,36]. Present understandings of PD genetics largely come from the former, where studies of families with heritable forms of the disease have identified a number of highly penetrant mutations in approximately 20 genes that are sufficient to cause the disease. While monogenic, or "familial," instances make up about 10% of all diagnoses, the majority of PD cases are idiopathic and can be attributed to a complex interplay of aging, environmental stressors, and subtle but compounding risk contributed by genetic variants with comparatively low effect size.[37]. The following sections explore each category further to illustrate how genetic studies have both established and developed present understandings of PD etiology.

1.3.1 Monogenic Parkinson's disease

The first gene identified for its causal role behind PD was *SNCA*. In an initial study of Italian and Greek families with early-onset cases showing an autosomal dominant inheritance pattern, Polymeropoulos *et al*. first identified the A53T missense mutation in *SNCA*[38]. In the same year, Spillantini and colleagues demonstrated that α-synuclein (the protein encoded by *SNCA*) is a principal structural component of Lewy bodies, linking the newly identified genetic cause to the long-recognized neuropathological hallmark of the disease[6]. Additional dominant missense mutations (A30P, E46K) and genomic *SNCA* rearrangements were identified shortly thereafter. Singleton and colleagues identified a triplication of the *SNCA* locus causing severe phenotype, and Chartier-Harlin *et al*. reported heterozygous duplications associated with a milder, later-onset form of PD[39,40]. Critically, these two studies established that dosage of wild-type *SNCA* is sufficient to cause PD, even in the absence of a coding mutation. Together, these observations anchored the current pathogenic framework in which the abundance and aggregation state of α-synuclein protein, rather than any single mutation, is the unifying driver of this dominant form of PD.

Mutations in *LRRK2* are the most common genetic cause of autosomal-dominant PD and were first identified by two independent groups in 2004[41,42]. The gene encodes a large

multidomain protein containing both GTPase (Roc-COR) and serine/threonine kinase domains along with multiple protein-interaction repeats. The G2019S substitution in the kinase domain accounts for the majority of *LRRK2*-associated PD and is most frequently identified among Ashkenazi Jewish and North African Berber populations, where it accounts for 10-40% of familial PD depending on the affected cohort[1,43]. Functional studies established that pathogenic *LRRK2* variants like G2019S increase kinase activity, and a phosphoproteomic screen by Steger and colleagues identified the RAB GTPases Rab8A, Rab10, and Rab12 as physiological LRRK2 substrates whose phosphorylation is elevated in cells expressing pathogenic alleles[44]. *LRRK2* occupies a distinctive position in PD genetics: the same gene harbors both highly penetrant coding variants that cause familial disease and common non-coding variants associated with risk for idiopathic PD through GWAS, blurring the conventional separation between monogenic and polygenic architectures.

Loss-of-function mutations in *PRKN*, encoding the E3 ubiquitin ligase parkin, were identified by Kitada and colleagues as the cause of autosomal-recessive juvenile parkinsonism[45], and loss-of-function mutations in *PINK1* were subsequently reported by Valente and colleagues[46]. Biochemical work established that these two proteins act in a common pathway in which PINK1 kinase accumulates on damaged mitochondria and recruits parkin to ubiquitinate outer-membrane proteins, targeting the organelle for autophagic clearance in a process termed mitophagy[47]. The PINK1-parkin axis has become one of the most thoroughly characterized mechanistic links from a PD gene to a specific cellular process, and the iSCORE-PD collection of genetically engineered human embryonic stem cells our group has established (introduced in Chapter 2) includes engineered *PRKN* and *PINK1* loss-of-function lines to interrogate it in isogenic human cells[48].

PARK7, encoding the protein DJ-1, was identified as a third recessive PD gene by Bonifati and colleagues on the basis of consanguineous Dutch and Italian pedigrees carrying a homozygous exon 1-5 deletion and the L166P missense variant[49]. DJ-1 is a small, highly conserved protein whose canonical function rests on a single cysteine residue at position 106 (Cys106), which is selectively susceptible to oxidation to cysteine-sulfinic acid. This oxidation drives conformational rearrangement and is required for several DJ-1 activities, including mitochondrial relocalization under oxidative stress; DJ-1 is therefore best understood as a cysteine redox sensor rather than a conventional enzyme[50]. A proposal that DJ-1 acts as a protein deglycase has been disputed by multiple independent groups[51–54], and current consensus treats DJ-1 as having at most weak glyoxalase activity. DJ-1 loss produces mitochondrial phenotypes in multiple systems, but no mechanism analogous to the PINK1-parkin kinase-ligase axis has been established[55]. A separate finding by Yamaguchi and colleagues proposed that DJ-1 acts as a transcriptional co-activator of SREBP-2 at the sterol regulatory element of the *LDLR* promoter, with consequences for cholesterol homeostasis[56]. This proposal has not been independently replicated in neurons and is noted here because it represents the sole prior hypothesis linking DJ-1 to cholesterol regulation, a connection that becomes relevant to the Perturb-seq results presented in Chapter 3.

Two independent groups used exome sequencing in 2011 to identify mutations in *VPS35* as a cause of autosomal-dominant PD[57,58]. VPS35 is a core subunit of the retromer complex, a heterotrimer that mediates the sorting and recycling of transmembrane cargo from endosomes to the trans-Golgi network and the plasma membrane. The D620N missense substitution is the most common and best-characterized pathogenic variant. Functional studies have shown that D620N impairs the association of retromer with the WASH complex, disrupting endosomal actin dynamics and autophagosome formation[59], and causes mitochondrial fragmentation through altered recycling of the fission factor DLP1[60,61]. VPS35 thus sits at the intersection of endosomal trafficking and mitochondrial dynamics, two pathways repeatedly implicated in PD.

Loss-of-function mutations in *FBXO7* cause an autosomal-recessive parkinsonian-pyramidal syndrome (PARK15), first identified through a homozygous R378G missense mutation in an Iranian family with pyramidal features preceding parkinsonism[62], and subsequently confirmed with additional mutations (R498X, compound heterozygous T22M/splice-site) in Italian and Dutch families[63]. FBXO7 encodes an F-box subunit of an SCF-type E3 ubiquitin ligase. It has established roles in two proteostasis-related processes: proteasome assembly, where FBXO7 facilitates the formation of the 26S complex and its loss reduces proteasomal chymotryptic activity by approximately 20% in mouse brain[64], and mitophagy, where FBXO7 interacts directly with both parkin and PINK1 to promote the clearance of damaged mitochondria[65].

Mutations in *VPS13C* were identified by Lesage and colleagues as the cause of autosomal-recessive early-onset PD with rapid progression and early cognitive decline[66]. VPS13C encodes a large bridge-like lipid transfer protein that localizes to ER-lysosome membrane contact sites, where it facilitates non-vesicular lipid exchange required for lysosomal membrane integrity[67]. Loss of VPS13C increases PINK1/parkin-dependent mitophagy, placing this gene at the intersection of the endolysosomal and mitochondrial quality-control axes[66]. Mutations in *ATP13A2* cause Kufor-Rakeb syndrome, a rare juvenile-onset parkinsonism with dementia[68]. ATP13A2 encodes a lysosomal P5-type ATPase that mediates polyamine and cation transport; its loss leads to lysosomal dysfunction, impaired autophagosome clearance, and iron accumulation. Together with VPS35, these genes anchor the endolysosomal axis of monogenic PD biology.

Beyond the mitochondrial and endolysosomal axes, mutations in two genes encoding synaptic vesicle recycling machinery have been linked to early-onset recessive parkinsonism. *DNAJC6* (PARK19) encodes auxilin, a clathrin-uncoating co-chaperone whose mutations were first linked to juvenile parkinsonism with seizures in 2012[69], and subsequent work in hESC-derived neurons demonstrated that auxilin loss causes clathrin accumulation, impaired vesicular transport, and dopaminergic neurodegeneration[70]. *SYNJ1* (PARK20) encodes synaptojanin-1, a phosphoinositide phosphatase enriched at presynaptic terminals; homozygous mutations in its Sac1-like domain were identified in 2013 as a cause of early-onset PD with generalized seizures[71]. These genes establish a

synaptic vesicle recycling axis of monogenic PD that is distinct from, though intersecting with, the endolysosomal dysfunction pathway.

Several additional loci have been designated as PARK genes on the basis of initial linkage or sequencing studies but have not withstood independent replication. *UCHL1* (PARK5) was proposed as a dominant PD gene based on a single I93M variant in one German family with two affected individuals, but a large case-control study found no association of *UCHL1* variation with PD risk[72,73]. *GIGYF2* (PARK11), *EIF4G1* (PARK18), *HTRA2* (PARK13), and *DNAJC13* (PARK21) have each been reported in isolated families or small cohorts but failed replication in larger or multi-ethnic studies. This dissertation does not treat these loci as established monogenic PD genes, though several (*UCHL1*, *GIGYF2*, *EIF4G1*, *DNAJC13*) are represented in the CRISPRi Perturb-seq library and their perturbation signatures are reported where informative in Chapter 3.

1.3.2 Idiopathic Parkinson's Disease and GWAS

While monogenic mutations provide high-confidence entry points into disease biology, the overwhelming majority of PD cases (>90% by most estimates) are clinically classified as idiopathic (also termed sporadic) and are presumed to arise from the combined effects of common genetic variation, aging, and environmental exposure[37]. Common-variant genome-wide association studies (GWAS) have progressively enlarged the catalog of loci associated with idiopathic PD risk. Early studies identified 28 loci[74], a 2017 meta-analysis increased this to 41[35], and the 2019 meta-analysis of Nalls and colleagues, pooling data from approximately 1.5 million participants including 37,700 cases, identified 90 independent risk signals across 78 genomic regions[36]. The Nalls 2019 study is the GWAS from which the gRNA library used for the Perturb-seq screen in this dissertation was designed, and its mechanics merit close treatment. Subsequent work has continued to expand the catalog: a multi-ancestry meta-analysis incorporating East Asian, Latin American, and African cohorts identified 78 loci including 12 novel ones[75], and the Global Parkinson's Genetics Program (GP2) GWAS has identified 134 risk loci, 59 of which had not been previously reported[76].

GBA1, encoding the lysosomal enzyme glucocerebrosidase, sits across the boundary between the monogenic and common-variant architectures of PD. Biallelic loss-of-function mutations cause Gaucher disease, an autosomal-recessive lysosomal storage disorder; in the heterozygous state, the same mutations confer the single largest common genetic risk for PD, with odds ratios ranging from ~1.4 to >10 across variants[77]. Approximately 5-15% of PD patients across ancestries carry a *GBA1* risk variant, making *GBA1* the most prevalent genetic contributor to PD at the population level. *GBA1*, together with *LRRK2* noted in the preceding section, exemplifies a pattern central to the motivation for this dissertation: individual genes can harbor both highly penetrant monogenic mutations and common lower-penetrance variants, and a complete account of PD genetics cannot treat familial and idiopathic disease as entirely separate categories.

The first step in interpreting a PD GWAS is to recognize what the association test actually identifies. A GWAS tests single-nucleotide polymorphisms (SNPs) for statistical association with disease status; the output is a set of sentinel SNPs, not a set of genes. In the Nalls 2019 meta-analysis, approximately 90% of the 90 risk sentinel SNPs fall in non-coding regions of the genome and are presumed to act through effects on gene regulation rather than on protein structure[36]. Translating a regulatory SNP into a candidate effector gene is a second, inferential step that depends on the method used. Nalls and colleagues used two complementary approaches: the first defined a candidate-gene universe by proximity, collecting genes within ±1 Mb in linkage disequilibrium ($r^2 > 0.5$) with each index SNP, yielding 305 candidate genes under the GWAS peaks; the second applied summary-based Mendelian randomization (SMR) to integrate PD GWAS summary statistics with expression and methylation QTL reference datasets, nominating 151 of the 305 candidates at 70 of the 90 signals as showing evidence for a causal mediating effect of a regulatory variant on gene expression[36]. The 2019 analysis did not employ Bayesian fine-mapping or colocalization testing at the per-locus level; subsequent work has retrospectively added fine-mapping, narrowing credible-variant sets at a subset of signals[78]. The 305 and 151 figures therefore represent complementary candidate pools rather than definitive causal gene lists; in the CRISPRi Perturb-seq screen described in Chapter 3, the full 305-gene universe was perturbed so as not to pre-commit the analysis to the SMR-nominated subset.

Translating a GWAS candidate gene into a biological mechanism requires functional evidence beyond statistical association. This gap between catalog and cell biology, often termed the GWAS-to-function problem, is not specific to PD; it is the defining challenge of post-GWAS functional genomics across complex diseases[79]. For PD it is particularly acute because the cellular context in which PD risk variants are presumed to act (midbrain DA neurons and their supporting non-neuronal cell types) has been difficult to access experimentally at the scale that functional interrogation requires. Closing this gap in a disease-relevant cell type is a primary motivation for the work described in this dissertation.

Three interpretive cautions follow from how GWAS results are generated and how they will be treated in this work. First, GWAS-nominated candidate genes are not equivalent to familial PD genes. Familial PD mutations typically confer large individual risk effects and segregate in pedigrees with inheritance patterns consistent with a single-gene causal mechanism, warranting the term causal when describing the link between the gene and disease. Common GWAS-nominated variants confer small individual risk effects, are identified in aggregate across tens of thousands of cases, and act in the genetic background of countless other common variants. Language reserved for high-penetrance Mendelian genes (most directly, the term causal) is not applied to GWAS-nominated genes in this dissertation; these are described instead as candidate risk genes, GWAS-nominated, or PD-implicated.

Second, the cell type through which a GWAS-nominated gene contributes to PD risk is not necessarily the DA neuron. GWAS-based gene-set and cell-type enrichment analyses have repeatedly implicated pathways including autophagy-lysosomal function, vesicular

trafficking, mitochondrial biology, and immune and inflammatory signaling[26,36], and PD-associated signal at many loci co-localizes with regulatory elements active in microglia and monocytes rather than in neurons[27]. A DA-neuron perturbation experiment such as the one described in Chapter 3 is therefore a necessary but not sufficient functional test: it captures the cell-autonomous contribution of a candidate gene within the DA lineage, while leaving open the contribution of the same gene in astrocytes, microglia, or peripheral immune cells to PD risk. This cell-autonomy caveat is revisited in §3.7 (Study limitations).

Third, and relevant to the generalizability of any functional result, the Nalls 2019 reference used to define the gRNA library in this dissertation is approximately 99% European-ancestry. Kim and colleagues recently demonstrated that multi-ancestry meta-analysis identifies novel risk loci that are absent from or underpowered in European-only analysis, and Rizig and colleagues reported African-specific *GBA1* haplotypes associated with PD that had not been detectable in prior GWAS[75,80]. The CRISPRi Perturb-seq library used in this dissertation reflects the Nalls 2019 candidate pool and therefore inherits its ancestry composition, a limitation revisited in §3.7.

Despite the inferential distance between a GWAS sentinel SNP and a validated biological mechanism, a question that naturally arises is whether the familial and common-variant architectures of PD converge at the level of biological pathways. Two kinds of evidence are typically cited in support of convergence. The first is overlap at the locus level: *GBA1* and *LRRK2*, discussed above, each harbor both highly penetrant coding variants and common non-coding risk variants, placing the same gene in both arms of the architecture. The second comes from pathway enrichment of GWAS summary statistics, where gene-based tests such as MAGMA aggregate SNP-level p-values and test for enrichment against curated pathway databases, repeatedly identifying terms related to endolysosomal function, autophagy, and mitochondrial biology that overlap with the pathways defined by monogenic PD genes[36,37]. These enrichment analyses, however, are constrained by pre-annotated databases and average across cell types; a gene assigned to one canonical pathway may contribute to disease through an unannotated function, and a pathway enrichment detected in bulk tissue may not hold in a specific cell population. A direct experimental test of whether familial and GWAS-nominated PD genes converge on shared transcriptional programs when perturbed in the same cell type has not previously been performed; this test is the central aim of the Perturb-seq experiments described in Chapter 3 and is formalized as the organizing hypothesis in §1.7.

1.4 Molecular mechanisms of Parkinson's disease pathogenesis

The genes that have been directly associated with monogenic PD appear to converge on a limited number of cellular pathways whose dysfunction, either individually or in combination, is thought to drive dopaminergic neurodegeneration. Understanding these pathways and the extent to which they interact is essential context for the Perturb-seq experiments presented in Chapter 3, where unbiased transcriptomic readout of gene

perturbation in DA neurons may recover signatures that map onto, or extend, the pathway biology summarized here.

The best-characterized molecular axis in monogenic PD centers on mitochondrial quality control. As described in §1.3.1, PINK1 and parkin operate in a common mitophagy pathway in which PINK1 kinase accumulates on depolarized mitochondria and recruits parkin to ubiquitinate outer-membrane proteins, targeting the damaged organelle for autophagic clearance[47]. DJ-1 contributes to mitochondrial homeostasis through its Cys106-dependent redox-sensing function; its loss increases mitochondrial reactive oxygen species and reduces membrane potential, though the mechanism by which DJ-1 protects mitochondria remains less precisely defined than the PINK1-parkin axis[49,55]. FBXO7 has been proposed to interact with both PINK1 and parkin to promote mitophagy[65], though this role was not confirmed in iPSC-derived neurons and HeLa cells by Kraus and colleagues[81], leaving the mitophagy function of FBXO7 contested. Pathogenic *LRRK2* variants also impinge on mitochondrial biology: G2019S kinase hyperactivity promotes DRP1-mediated mitochondrial fragmentation and increases mitochondrial superoxide production, compounding the oxidative burden that DA neurons already carry from dopamine metabolism and neuromelanin accumulation (§1.2)[32]. These mitochondrial defects are thought to create a feed-forward loop in which impaired electron-transport-chain function generates reactive oxygen species that further damage mitochondrial DNA, lipids, and the respiratory complexes themselves.

A second major axis implicates the endolysosomal system. VPS35, a core subunit of the retromer complex, mediates endosomal cargo sorting and recycling; the D620N mutation impairs the association of retromer with the WASH complex, disrupting endosomal actin dynamics and autophagosome formation[59], and causes mitochondrial fragmentation through altered recycling of the fission factor DLP1[60,61]. *ATP13A2* encodes a lysosomal P5-type ATPase that maintains cation homeostasis; its loss causes lysosomal dysfunction, impaired autophagosome clearance, and iron accumulation[68]. VPS13C functions as a bridge-like lipid transfer protein at ER-lysosome membrane contact sites, where it mediates non-vesicular lipid exchange required for lysosomal membrane integrity[67]; loss of *VPS13C* secondarily increases PINK1/parkin-dependent mitophagy, linking endolysosomal failure to the mitochondrial axis[66]. *GBA1* encodes glucocerebrosidase, whose reduced activity in heterozygous carriers impairs glycosphingolipid hydrolysis and promotes α-synuclein accumulation, establishing a direct biochemical connection between lysosomal dysfunction and protein aggregation[77]. LRRK2 is also a major player in this pathway: it localizes to early and late endosomes and to lysosomal membranes, and pathogenic variants hyperphosphorylate RAB GTPases (Rab8A, Rab10, Rab12), disrupting their membrane-cytosol cycling and impairing endolysosomal cargo sorting[44]. *LRRK2* knockout in mice results in enlarged lamellar bodies with lipofuscin accumulation, consistent with a role in lysosomal homeostasis. The interaction between VPS35 and LRRK2 is particularly noteworthy: the VPS35 D620N mutation has been shown to enhance LRRK2 kinase activity at endolysosomes, suggesting that these two PARK genes operate not merely in the same pathway but in direct functional interaction[82]. The

convergence of multiple monogenic PD genes on endolysosomal function has led to the characterization of PD as, in part, a lysosomal disorder.

Impaired protein homeostasis represents a third axis that intersects with both mitochondrial and endolysosomal dysfunction. α-Synuclein aggregation is fundamentally a proteostasis problem: the ubiquitin-proteasome system (UPS) and autophagy are the two principal cellular mechanisms for clearing misfolded and aggregated proteins, and both are compromised in PD[6,38]. Parkin, in addition to its mitophagy role, ubiquitinates cytosolic substrates for proteasomal degradation; its loss reduces the clearance capacity of the UPS[45]. FBXO7 facilitates 26S proteasome assembly through non-degradative K63-ubiquitination of the proteasomal subunit PSMA2, and its loss reduces proteasomal chymotryptic activity by approximately 20% in mouse brain[64]. LRRK2 kinase hyperactivity disrupts autophagosome transport along axonal microtubules and impairs autophagic flux, adding a trafficking dimension to the proteostasis defect[83]. Beyond the classical UPS and autophagy framework, genome-scale CRISPRi Perturb-seq in K562 cells has revealed that perturbation of diverse essential genes frequently converges on a shared integrated stress response, characterized by eIF2α phosphorylation and ATF4 activation[84,85]. Whether PD gene perturbations in DA neurons elicit analogous translational stress programs, and what form those programs take in a post-mitotic neuronal context, is a question addressed directly by the Perturb-seq experiments in Chapter 3.

A less canonical but increasingly recognized axis involves lipid metabolism, particularly the cholesterol biosynthesis and membrane-remodeling pathways. *GBA1* loss alters sphingolipid and cholesterol metabolism through impaired lysosomal hydrolysis of glucosylceramide, and the relationship between lysosomal cholesterol sensing and mTORC1 signaling has been defined biochemically by Castellano and colleagues[77,86]. VPS13C, through its ER-lysosome lipid transfer function, is expected to influence membrane lipid composition when lost[67]. DJ-1 has been proposed as a transcriptional co-activator of SREBP-2 at the *LDLR* promoter, with consequences for cholesterol homeostasis, though this finding has not been independently replicated in neurons[56]. *PLA2G6* (PARK14) encodes a calcium-independent phospholipase A2 that hydrolyzes the sn-2 acyl chain from membrane phospholipids; its loss disrupts membrane phospholipid remodeling and is associated with lipid profile alterations and iron accumulation in the basal ganglia[87]. Cholesterol biosynthesis is not a canonical PD pathway in the current literature, and its appearance in functional genomic data, if observed, would represent a relatively novel connection between PD genetics and lipid homeostasis.

Dysfunction at the presynaptic terminal constitutes a distinct axis that is consistent with the dying-back hypothesis introduced in §1.1.2, in which DA axon terminals degenerate before cell bodies are lost. α-Synuclein normally associates with synaptic vesicle membranes and participates in SNARE complex assembly; its aggregation impairs vesicle clustering, docking, and exocytosis[6,38]. LRRK2 phosphorylates endophilin A at S75, modulating its function between synaptic vesicle endocytosis and autophagy, and also phosphorylates Rab3A/B/C/D, which regulate synaptic vesicle docking and fusion at

presynaptic terminals[44]. This positions LRRK2 as a kinase hub connecting synaptic vesicle dynamics to both endolysosomal and autophagy pathways. *DNAJC6* (PARK19) encodes auxilin, a clathrin-uncoating co-chaperone that recruits Hsc70 to clathrin-coated vesicles after fission; loss of auxilin causes clathrin accumulation, impaired vesicular transport, and DA neurodegeneration in hESC-derived neurons[69,70]. *SYNJ1* (PARK20) encodes synaptojanin-1, a phosphoinositide phosphatase enriched at nerve terminals whose Sac1-like domain activity is required for clathrin-coat disassembly and synaptic vesicle recycling[71]. These genes demonstrate that disruption of the synaptic vesicle cycle is sufficient to cause parkinsonism, and the synaptic axis intersects with the endolysosomal pathway at the point where clathrin-mediated endocytosis feeds internalized membrane into the endosomal sorting system.

Finally, neuroinflammation has emerged as both a consequence and a potential amplifier of PD pathology. LRRK2 is the strongest genetic link between PD and innate immunity: it is highly expressed in monocytes, macrophages, and microglia, and G2019S increases inflammatory cytokine release from activated microglia. The association of *LRRK2* with Crohn's disease and leprosy through independent GWAS supports a broader role for LRRK2 in innate immune regulation that extends beyond the nervous system[37]. α-Synuclein-specific T cells have been identified in the blood of PD patients and recognize epitopes presented by MHC class I on dopaminergic neurons, consistent with an adaptive immune contribution to DA neuron targeting (§1.2)[28]. Cell-type enrichment analyses of PD GWAS summary statistics have repeatedly identified signal in microglia and peripheral immune populations alongside neurons (§1.3.2), and post-mortem studies consistently report activated microglia and elevated pro-inflammatory cytokines in the PD substantia nigra. Whether immune dysregulation is a primary driver of DA neuron loss or a secondary response to upstream cellular dysfunction remains unresolved, but the genetic evidence, particularly from *LRRK2*, argues that at least some component of neuroinflammation is causally upstream rather than purely reactive.

These pathways are not independent. Mitochondrial dysfunction generates reactive oxygen species that accelerate α-synuclein misfolding; impaired lysosomes fail to clear both damaged mitochondria and aggregated proteins; altered lipid composition disrupts the membranes required for vesicle trafficking and lysosomal function; and synaptic vesicle recycling defects overload the endosomal sorting system. *LRRK2* illustrates this crosstalk most clearly: a single kinase whose hyperactivation produces measurable consequences in endolysosomal, mitochondrial, synaptic, autophagy, and immune compartments. The extensive interconnection among these pathways means that perturbation of a gene assigned to one axis may produce transcriptional signatures in another, and that the downstream transcriptional response to a given perturbation may not map neatly onto the gene's primary annotated function. It is additionally worth noting that transcriptional responses observed may be compensatory or adaptive response to perturbation, and may not necessarily reflect the biological effect of perturbation at the protein level. This complexity motivates the unbiased, transcriptome-wide approach of Perturb-seq as a

complement to single-pathway biochemical studies, and it is the central rationale for the experiments described in Chapter 3.

1.5 hPSC–derived neurons as a model system for Parkinson's disease

The genetic and anatomical observations outlined above establish what is known about PD from human tissue and population genetics; they also define what those approaches cannot provide. Each of the main experimental systems historically used to study PD makes trade-offs that limit its utility for dissecting gene-specific disease mechanisms. Understanding these trade-offs motivates the experimental approach used in this dissertation and the two complementary resources it builds on: the iSCORE-PD collection (Chapter 2) and the CRISPRi Perturb-seq screen (Chapter 3).

Human patient-derived iPSCs offer the ability to study disease-gene function in human cells, at physiologically relevant expression levels, in cell types directly relevant to disease pathology[88]. However, the full promise of patient-iPSC-based disease modeling has been constrained by genetic background variation. When comparing a PD-patient line to an unrelated control line, observed transcriptional differences reflect both the PD-associated mutation and thousands of unrelated polymorphisms distributed across the genome. Large-scale iPSC consortia have documented that inter-individual variability in differentiation efficiency, cell-type composition, and baseline transcriptional state typically exceeds the effect sizes of most disease mutations, making it difficult to isolate disease-specific signatures from background noise[89]. This is particularly problematic for PD, where *in vitro* perturbation effect sizes are expected to be modest. It is worth considering this alongside the fact that hPSC-derived models more generally reflect embryonic or fetal cell states rather than mature adult ones, which additionally impedes a faithful recapitulation of a decades-long disease course.

The solution adopted by several large-scale initiatives, and central to the work of this dissertation, is the use of isogenic cell line collections: disease-associated mutations engineered into a single common parental hPSC background, such that all phenotypic comparisons are made against cells that differ only at the locus of interest[88]. This design eliminates background genetic variation as a confound, improving the signal-to-noise ratio for detecting disease-associated transcriptional signatures. It also enables pooled perturbation screens that are impossible in patient-iPSC comparisons: because all cells share the same genetic background, gRNA-mediated perturbation can be applied to a pooled population, and the transcriptional consequence of each perturbation read out in single cells from the same culture. The two projects presented in this dissertation exploit this isogenic framework in complementary ways. Chapter 2 describes the iSCORE-PD collection, which characterizes the transcriptional consequences of endogenous PD mutations across 11 familial PD genes engineered into the WIBR3 background; Chapter 3 presents the Perturb-seq screen, which perturbs 41 familial and 310 GWAS-nominated PD risk genes via CRISPRi in the same parental line.

Both human embryonic stem cells (hESCs) and human induced pluripotent stem cells (hiPSCs) can be expanded indefinitely, genetically modified, and differentiated to DA neurons. For large-scale isogenic mutation-panel studies, where consistency across many engineered daughter lines is paramount, hESCs offer practical advantages over hiPSCs: they avoid reprogramming-associated somatic mutations and epigenetic artifacts, and tend to maintain more stable karyotypes over extended culture[88,90].

Two large-scale isogenic collections have been generated for PD research. The iPSC Neurodegenerative Disease Initiative (iNDI), anchored at the NIH Center for Alzheimer's and Related Dementias, is engineering a mutation panel spanning both Alzheimer's and Parkinson's disease in the male KOLF2.1J iPSC line, with hundreds of daughter lines planned including knock-ins, knockouts, and reporter lines[91]. The iSCORE-PD collection, which is the central resource of Chapter 2 and was conducted as part of the ASAP Collaborative Research Network, takes a complementary approach focused specifically on PD: 65 clonal cell lines carrying 13 mutations in 11 familial PD genes (*SNCA, LRRK2, PRKN, PINK1, PARK7, ATP13A2, FBXO7, DNAJC6, SYNJ1, VPS13C, and GBA1*), all engineered in the female hESC line WIBR3[48].

WIBR3 is a female hESC line (NIH registry NIHhESC-10-0079) that has been used extensively in human stem cell biology and PD research[92–94]. Its selection as the iSCORE-PD parental line reflects a confluence of practical advantages: hESC derivation, karyotypic stability across extended passage, efficient CRISPR editing, and robust differentiation to midbrain DA neurons. As a female line, WIBR3 also provides sex complementarity to the male KOLF2.1J used by iNDI, which is relevant given documented sex differences in PD epidemiology and biology[95]. While female lines permit the investigation of sex differences in disease, these lines can undergo erosion of X-chromosome inactivation after prolonged culture, leading to experimental confounds due to resulting changes the heterochromatin landscape and in the dosage of X-chromosome genes [96].

The efficient generation of midbrain-lineage DA neurons from hPSCs requires recapitulating the signaling events that pattern the ventral midbrain floor plate during embryonic development. The landmark protocol of Kriks and colleagues established that dual-SMAD inhibition combined with WNT activation (via the GSK3β inhibitor CHIR99021) and Sonic hedgehog (SHH) signaling induces a midbrain floor plate progenitor fate that subsequently gives rise to DA neurons expressing TH, FOXA2, and LMX1A[97]. Subsequent refinements, including the biphasic WNT signaling modulation protocol of Kim and colleagues, in which a controlled early CHIR pulse is followed by a second, later WNT activation phase, have improved both the yield of DA neurons and the proportion expressing substantia-nigra-specific markers (ALDH1A1, GIRK2/KCNJ6) indicative of SNc-like identity rather than VTA-like identity[98]. The experiments described in this dissertation employ the Kim et al. biphasic WNT protocol adapted for WIBR3 hESCs, with DA neurons harvested at day 45 of differentiation after ten days of CRISPRi induction (detailed in Chapter 3). While these protocols generate neurons that express canonical midbrain DA markers and recapitulate key electrophysiological properties, important limitations remain: *in vitro* DA neurons are transcriptionally and functionally immature relative to adult human neurons, and monoculture systems lack the non-neuronal cell types (astrocytes, microglia,

oligodendrocytes) whose contributions to PD pathology were discussed in §1.2. These limitations constrain the biological conclusions that can be drawn from any *in vitro* perturbation experiment and are revisited in §3.7.

1.6 CRISPRi and Perturb-seq as tools for functional genomics

CRISPR interference (CRISPRi) and CRISPR activation (CRISPRa), introduced by Gilbert, Qi, and colleagues, repurpose catalytically inactive Cas9 (dCas9) as a programmable transcriptional modulator that directs an effector domain to a specified genomic locus without inducing DNA double-strand breaks[99,100]. For CRISPRi, dCas9 is fused to the Krüppel-associated box (KRAB) repression domain, which recruits the KAP1/TRIM28 corepressor complex and establishes repressive H3K9me3 chromatin at the targeted gene promoter, producing robust but incomplete transcriptional silencing typically ranging from 50% to 95% knockdown depending on guide selection and target[101]. For CRISPRa, dCas9 is fused to activation domains such as VP64-p65-Rta (VPR), which recruit the endogenous transcriptional machinery to upregulate expression of target genes[102]. For studying PD genetics specifically, CRISPRi offers several advantages over gene knockout as a perturbation modality. Many familial PD genes act through haploinsufficient or graded loss-of-function mechanisms: heterozygous mutations in *GBA1* confer substantial disease risk without complete loss of enzyme activity; dominant mutations in LRRK2 increase kinase activity without ablating the gene; and many autosomal recessive PD alleles produce partial rather than total loss of function[37,41,77]. Partial, tunable transcriptional knockdown via CRISPRi more faithfully models these graded loss-of-function phenotypes than does frameshift-mediated knockout. CRISPRi also avoids the confounds of double-strand-break-induced DNA damage response activation, off-target indels, and frameshift-generated truncation products that can complicate interpretation of CRISPR-Cas9 knockout screens[103], and it spares essential genes whose complete loss is lethal but whose partial knockdown is experimentally tractable.

An additional design consideration specific to hPSC-derived neuron systems is the timing of perturbation relative to differentiation. Constitutive expression of dCas9-KRAB from the pluripotent stem cell stage would subject cells to CRISPRi-mediated gene knockdown throughout the extended (35–45 day) differentiation protocol, perturbing developmental trajectories and potentially preventing the establishment of a stable DA neuron identity. The Cre-inducible CRISPRi system used in this dissertation addresses this by placing a loxP-flanked STOP cassette upstream of the dCas9-KRAB expression cassette integrated into the AAVS1 safe-harbor locus, such that dCas9-KRAB is not expressed until Cre recombinase excises the cassette. In practice, gRNA-transduced hPSCs are differentiated to day 35 DA neurons with the CRISPRi machinery silenced; Cre is then delivered to induce CRISPRi, and cells are harvested ten days later at day 45 for single-cell profiling. This design restricts perturbation to a defined ten-day window in post-mitotic DA neurons, avoiding developmental confounds and ensuring that transcriptional responses reflect acute gene knockdown rather than accumulated developmental mispatterning. The technical details of the Cre-inducible CRISPRi platform are presented in Chapter 3.

Perturb-seq, introduced independently by Dixit and colleagues and Adamson and colleagues in 2016, pairs pooled CRISPR-based genetic screens with single-cell RNA sequencing to measure the transcriptomic consequences of individual perturbations at single-cell resolution[104,105]. In a typical experiment, a population of cells is transduced with a lentiviral gRNA library at low multiplicity of infection, ensuring most cells carry a single perturbation. After a defined perturbation window, cells are captured on a droplet-based single-cell platform and both the transcriptome and gRNA identity of each cell are sequenced in parallel. Computational methods then associate perturbation identity with transcriptomic phenotype, yielding differentially expressed gene sets and enriched pathway signatures for each targeted gene. Perturb-seq has since been scaled to genome-wide screens profiling millions of cells[84], applied to the discovery of genetic interaction manifolds[106], and adapted to a growing range of cell types including hPSC-derived neurons and microglia[107,108]. The decision to use Perturb-seq rather than a series of bulk RNA-seq experiments reflects two specific requirements of this work. First, pooling all perturbations in a single culture ensures every gRNA is represented in the same differentiation batch, sample preparation, and sequencing run, dramatically reducing batch-effect confounding at a scale (41 familial + 310 GWAS genes) that would be prohibitively expensive in a one-per-well format. Second, DA neuron cultures are genuinely heterogeneous: even optimized differentiation protocols produce a mixture of progenitors, neuroblasts, and post-mitotic DA neurons of varying maturity. Single-cell readout, combined with computational cell-type assignment, enables the same perturbation to be read out independently in each cell population within the same culture, a capability that is central to this dissertation.

CRISPRi screens in hPSC-derived neurons have established precedent for the approach taken here. Tian, Kampmann, and colleagues applied genome-wide CRISPRi survival screens to hPSC-derived neurons to identify pathways controlling neuronal vulnerability, defining roles for lysosomal function and iron handling in neuronal ferroptosis susceptibility[107]. Dräger and colleagues extended this framework to hPSC-derived microglia[108], and more recent work has applied CRISPRi Perturb-seq to hPSC-derived cortical neurons to profile the transcriptional consequences of candidate Alzheimer's disease risk genes[109,110]. The work described in this dissertation builds on this foundation by applying CRISPRi Perturb-seq, to our knowledge for the first time, to a systematically selected panel of familial and GWAS-nominated PD risk genes in isogenic human midbrain DA neurons.

1.7 Dissertation framework and thesis organization

This dissertation is organized around a central hypothesis: that the genetic heterogeneity of Parkinson's disease, dozens of familial genes and hundreds of GWAS-nominated risk genes spanning diverse cellular pathways, obscures a smaller number of shared transcriptional programs which can be revealed by systematic perturbation in a controlled genetic background. The preceding sections have established the biological and genetic context that motivates this claim: the selective vulnerability of DA neurons in the SNc (§1.2), the monogenic and common-variant genetic architecture of PD (§1.3), the

convergence of multiple PD genes on a limited set of cellular pathways (§1.4), and the advantages of isogenic hPSC-derived DA neurons as a system for functional interrogation of PD risk genes at scale (§1.5–1.6). Two hypotheses follow from this framework. First, mechanistically distinct PD risk genes, both familial and GWAS-nominated, converge on a small number of shared transcriptional axes in human DA neurons; despite acting through ostensibly unrelated cellular machineries, their perturbation produces signatures that can be organized into functional groups defined by convergent downstream programs. Second, the transcriptional consequence of perturbing a given PD risk gene is cell-type- and cell-state-specific: the same perturbation can produce qualitatively different responses depending on the developmental or functional state of the cell, and single-cell readout is required to resolve this specificity.

These hypotheses are tested through two CRISPRi Perturb-seq experiments conducted in isogenic WIBR3-derived midbrain DA neurons. The first experiment targets 41 familial PD genes with a focused gRNA library; the second targets 310 GWAS-nominated PD candidate risk genes derived from the Nalls 2019 meta-analysis. Together, these experiments profile the transcriptional consequences of perturbation across more than 85,000 single cells, enabling both per-perturbation and cross-perturbation analyses of pathway convergence with cell-type resolution. Although the readout of these experiments is strictly transcriptional and cannot detect functional changes at the protein level (such as post-translational modifications, degradation or localization), our goals are to identify the biological pathways with evident transcriptional disruption so that we may refine and help prioritize certain perturbations for follow-up studies.

The dissertation is organized as follows. Chapter 2 presents the iSCORE-PD isogenic hPSC collection, adapted from a co-first-authored publication[48]. This chapter introduces the collection of 65 clonal lines carrying 13 mutations in 11 familial PD genes engineered into the WIBR3 background, the quality-control pipeline, and the genomic characterization of parental and edited lines. It establishes the isogenic framework and parental-line context against which the Perturb-seq experiments of Chapter 3 are interpreted. Chapter 3 presents the CRISPRi Perturb-seq screen. It begins with the development and validation of the Cre-inducible CRISPRi platform, describes the design and delivery of both gRNA libraries, and reports the primary results organized around the two hypotheses: convergent transcriptional axes across familial and GWAS perturbations, and cell-state-specific perturbation signatures across DA-lineage cell populations. The chapter terminates with a discussion of study limitations and implications for future work, including integration with next-generation CRISPRi effectors, extension to non-neuronal cell types, and orthogonal validation against the iSCORE-PD mutant lines.

References

1. Kalia LV, Lang AE. Parkinson's disease. Lancet. 2015 Aug;386(9996):896–912. doi:10.1016/S0140-6736(14)61393-3 PubMed PMID: 25904081.

2. GBD 2016 Parkinson's Disease Collaborators. Global, regional, and national burden of Parkinson's disease, 1990-2016: A systematic analysis for the Global Burden of Disease Study 2016. The Lancet Neurology. 2018 Nov;17(11):939–53. doi:10.1016/S1474-4422(18)30295-3 PubMed PMID: 30287051; PubMed Central PMCID: PMC6191528.

3. GBD 2021 Nervous System Disorders Collaborators. Global, regional, and national burden of disorders affecting the nervous system, 1990-2021: A systematic analysis for the Global Burden of Disease Study 2021. The Lancet Neurology. 2024 Apr;23(4):344–81. doi:10.1016/S1474-4422(24)00038-3 PubMed PMID: 38493795; PubMed Central PMCID: PMC10949203.

4. Lang AE, Espay AJ. Disease Modification in Parkinson's Disease: Current Approaches, Challenges, and Future Considerations. Movement Disorders: Official Journal of the Movement Disorder Society. 2018 May;33(5):660–77. doi:10.1002/mds.27360 PubMed PMID: 29644751.

5. Matsuda W, Furuta T, Nakamura KC, Hioki H, Fujiyama F, Arai R, et al. Single nigrostriatal dopaminergic neurons form widely spread and highly dense axonal arborizations in the neostriatum. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience. 2009 Jan;29(2):444–53. doi:10.1523/JNEUROSCI.4029-08.2009 PubMed PMID: 19144844; PubMed Central PMCID: PMC6664950.

6. Spillantini MG, Schmidt ML, Lee VM, Trojanowski JQ, Jakes R, Goedert M. Alpha-synuclein in Lewy bodies. Nature. 1997 Aug;388(6645):839–40. doi:10.1038/42166 PubMed PMID: 9278044.

7. Braak H, Del Tredici K, Rüb U, de Vos RAI, Jansen Steur ENH, Braak E. Staging of brain pathology related to sporadic Parkinson's disease. Neurobiology of Aging. 2003;24(2):197–211. doi:10.1016/s0197-4580(02)00065-9 PubMed PMID: 12498954.

8. Fearnley JM, Lees AJ. Ageing and Parkinson's disease: Substantia nigra regional selectivity. Brain: A Journal of Neurology. 1991 Oct;114 (Pt 5):2283–301. doi:10.1093/brain/114.5.2283 PubMed PMID: 1933245.

9. Ma SY, Röyttä M, Rinne JO, Collan Y, Rinne UK. Correlation between neuromorphometry in the substantia nigra and clinical features in Parkinson's disease using disector counts. Journal of the Neurological Sciences. 1997 Oct;151(1):83–7. doi:10.1016/s0022-510x(97)00100-7 PubMed PMID: 9335015.

10. Greffard S, Verny M, Bonnet AM, Beinis JY, Gallinari C, Meaume S, et al. Motor score of the Unified Parkinson Disease Rating Scale as a good predictor of Lewy body-associated neuronal loss in the substantia nigra. Archives of Neurology. 2006 Apr;63(4):584–8. doi:10.1001/archneur.63.4.584 PubMed PMID: 16606773.

11. Cheng HC, Ulane CM, Burke RE. Clinical progression in Parkinson disease and the neurobiology of axons. Annals of Neurology. 2010 Jun;67(6):715–25. doi:10.1002/ana.21995 PubMed PMID: 20517933; PubMed Central PMCID: PMC2918373.

12. Kordower JH, Olanow CW, Dodiya HB, Chu Y, Beach TG, Adler CH, et al. Disease duration and the integrity of the nigrostriatal system in Parkinson's disease. Brain: A Journal of Neurology. 2013 Aug;136(Pt 8):2419–31. doi:10.1093/brain/awt192 PubMed PMID: 23884810; PubMed Central PMCID: PMC3722357.

13. Bendor JT, Logan TP, Edwards RH. The function of α-synuclein. Neuron. 2013 Sep;79(6):1044–66. doi:10.1016/j.neuron.2013.09.004 PubMed PMID: 24050397; PubMed Central PMCID: PMC3866954.

14. Mahul-Mellier AL, Burtscher J, Maharjan N, Weerens L, Croisier M, Kuttler F, et al. The process of Lewy body formation, rather than simply α-synuclein fibrillization, is one of the major drivers of neurodegeneration. Proceedings of the National Academy of Sciences of the United States of America. 2020 Mar;117(9):4971–82. doi:10.1073/pnas.1913904117 PubMed PMID: 32075919; PubMed Central PMCID: PMC7060668.

15. Surmeier DJ, Obeso JA, Halliday GM. Selective neuronal vulnerability in Parkinson disease. Nature Reviews Neuroscience. 2017 Jan;18(2):101–13. doi:10.1038/nrn.2016.178 PubMed PMID: 28104909; PubMed Central PMCID: PMC5564322.

16. Bolam JP, Pissadaki EK. Living on the edge with too many mouths to feed: Why dopamine neurons die. Movement Disorders: Official Journal of the Movement Disorder Society. 2012 Oct;27(12):1478–83. doi:10.1002/mds.25135 PubMed PMID: 23008164; PubMed Central PMCID: PMC3504389.

17. Pacelli C, Giguère N, Bourque MJ, Lévesque M, Slack RS, Trudeau LÉ. Elevated Mitochondrial Bioenergetics and Axonal Arborization Size Are Key Contributors to

the Vulnerability of Dopamine Neurons. Current biology: CB. 2015 Sep;25(18):2349–60. doi:10.1016/j.cub.2015.07.050 PubMed PMID: 26320949.

18. Kamath T, Abdulraouf A, Burris SJ, Langlieb J, Gazestani V, Nadaf NM, et al. Single-cell genomic profiling of human dopamine neurons identifies a population that selectively degenerates in Parkinson's disease. Nature Neuroscience. 2022 May;25(5):588–95. doi:10.1038/s41593-022-01061-1

19. Panman L, Papathanou M, Laguna A, Oosterveen T, Volakakis N, Acampora D, et al. Sox6 and Otx2 control the specification of substantia nigra and ventral tegmental area dopamine neurons. Cell Reports. 2014 Aug;8(4):1018–25. doi:10.1016/j.celrep.2014.07.016 PubMed PMID: 25127144.

20. Damier P, Hirsch EC, Agid Y, Graybiel AM. The substantia nigra of the human brain. I. Nigrosomes and the nigral matrix, a compartmental organization based on calbindin D(28K) immunohistochemistry. Brain: A Journal of Neurology. 1999 Aug;122 (Pt 8):1421–36. doi:10.1093/brain/122.8.1421 PubMed PMID: 10430829.

21. Damier P, Hirsch EC, Agid Y, Graybiel AM. The substantia nigra of the human brain. II. Patterns of loss of dopamine-containing neurons in Parkinson's disease. Brain: A Journal of Neurology. 1999 Aug;122 (Pt 8):1437–48. doi:10.1093/brain/122.8.1437 PubMed PMID: 10430830.

22. Brichta L, Greengard P. Molecular determinants of selective dopaminergic vulnerability in Parkinson's disease: An update. Frontiers in Neuroanatomy. 2014;8:152. doi:10.3389/fnana.2014.00152 PubMed PMID: 25565977; PubMed Central PMCID: PMC4266033.

23. Martirosyan A, Ansari R, Pestana F, Hebestreit K, Gasparyan H, Aleksanyan R, et al. Unravelling cell type-specific responses to Parkinson's Disease at single cell resolution. Molecular Neurodegeneration. 2024 Jan;19(1):7. doi:10.1186/s13024-023-00699-0 PubMed PMID: 38245794; PubMed Central PMCID: PMC10799528.

24. Smajić S, Prada-Medina CA, Landoulsi Z, Ghelfi J, Delcambre S, Dietrich C, et al. Single-cell sequencing of human midbrain reveals glial activation and a Parkinson-specific neuronal state. Brain: A Journal of Neurology. 2022 Apr;145(3):964–78. doi:10.1093/brain/awab446 PubMed PMID: 34919646; PubMed Central PMCID: PMC9050543.

25. Bryois J, Skene NG, Hansen TF, Kogelman LJA, Watson HJ, Liu Z, et al. Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease. Nature Genetics. 2020 May;52(5):482–93.

doi:10.1038/s41588-020-0610-9 PubMed PMID: 32341526; PubMed Central PMCID: PMC7930801.

26. Reynolds RH, Botía J, Nalls MA, International Parkinson's Disease Genomics Consortium (IPDGC), Noyce AJ, Nicolas A, et al. Moving beyond neurons: The role of cell type-specific gene regulation in Parkinson's disease heritability. npj Parkinson's Disease. 2019 Apr;5(1):6. doi:10.1038/s41531-019-0076-6

27. Andersen MS, Bandres-Ciga S, Reynolds RH, Hardy J, Ryten M, Krohn L, et al. Heritability Enrichment Implicates Microglia in Parkinson's Disease Pathogenesis. Annals of Neurology. 2021 May;89(5):942–51. doi:10.1002/ana.26032 PubMed PMID: 33502028; PubMed Central PMCID: PMC9017316.

28. Sulzer D, Alcalay RN, Garretti F, Cote L, Kanter E, Agin-Liebes J, et al. T cells from patients with Parkinson's disease recognize α-synuclein peptides. Nature. 2017 Jun;546(7660):656–61. doi:10.1038/nature22815 PubMed PMID: 28636593; PubMed Central PMCID: PMC5626019.

29. Kim S, Kwon SH, Kam TI, Panicker N, Karuppagounder SS, Lee S, et al. Transneuronal Propagation of Pathologic α-Synuclein from the Gut to the Brain Models Parkinson's Disease. Neuron. 2019 Aug;103(4):627–641.e7. doi:10.1016/j.neuron.2019.05.035 PubMed PMID: 31255487; PubMed Central PMCID: PMC6706297.

30. Sulzer D. Multiple hit hypotheses for dopamine neuron loss in Parkinson's disease. Trends in Neurosciences. 2007 May;30(5):244–50. doi:10.1016/j.tins.2007.03.009 PubMed PMID: 17418429.

31. Mor DE, Tsika E, Mazzulli JR, Gould NS, Kim H, Daniels MJ, et al. Dopamine induces soluble α-synuclein oligomers and nigrostriatal degeneration. Nature Neuroscience. 2017 Nov;20(11):1560–8. doi:10.1038/nn.4641 PubMed PMID: 28920936; PubMed Central PMCID: PMC5893155.

32. Zucca FA, Segura-Aguilar J, Ferrari E, Muñoz P, Paris I, Sulzer D, et al. Interactions of iron, dopamine and neuromelanin pathways in brain aging and Parkinson's disease. Progress in Neurobiology. 2017 Aug;155:96–119. doi:10.1016/j.pneurobio.2015.09.012 PubMed PMID: 26455458; PubMed Central PMCID: PMC4826627.

33. Blesa J, Przedborski S. Parkinson's disease: Animal models and dopaminergic cell vulnerability. Frontiers in Neuroanatomy. 2014;8:155. doi:10.3389/fnana.2014.00155 PubMed PMID: 25565980; PubMed Central PMCID: PMC4266040.

34. Dawson VL, Dawson TM. Animal Models of Parkinson's Disease. Cold Spring Harbor Perspectives in Medicine. 2025 Dec;15(12):a041644. doi:10.1101/cshperspect.a041644 PubMed PMID: 40983494; PubMed Central PMCID: PMC12667399.

35. Chang D, Nalls MA, Hallgrímsdóttir IB, Hunkapiller J, van der Brug M, Cai F, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. Nature Genetics. 2017 Oct;49(10):1511–6. doi:10.1038/ng.3955 PubMed PMID: 28892059; PubMed Central PMCID: PMC5812477.

36. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: A meta-analysis of genome-wide association studies. The Lancet Neurology. 2019 Dec;18(12):1091–102. doi:10.1016/S1474-4422(19)30320-5

37. Blauwendraat C, Nalls MA, Singleton AB. The genetic architecture of Parkinson's disease. The Lancet Neurology. 2020 Feb;19(2):170–8. doi:10.1016/S1474-4422(19)30287-X PubMed PMID: 31521533; PubMed Central PMCID: PMC8972299.

38. Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, et al. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science. 1997 Jun;276(5321):2045–7. doi:10.1126/science.276.5321.2045 PubMed PMID: 9197268.

39. Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, et al. Alpha-Synuclein locus triplication causes Parkinson's disease. Science. 2003 Oct;302(5646):841. doi:10.1126/science.1090278 PubMed PMID: 14593171.

40. Chartier-Harlin MC, Kachergus J, Roumier C, Mouroux V, Douay X, Lincoln S, et al. Alpha-synuclein locus duplication as a cause of familial Parkinson's disease. Lancet. 2004--1;364(9440):1167–9. doi:10.1016/S0140-6736(04)17103-1 PubMed PMID: 15451224.

41. Zimprich A, Biskup S, Leitner P, Lichtner P, Farrer M, Lincoln S, et al. Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron. 2004 Nov;44(4):601–7. doi:10.1016/j.neuron.2004.11.005 PubMed PMID: 15541309.

42. Paisán-Ruíz C, Jain S, Evans EW, Gilks WP, Simón J, van der Brug M, et al. Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. Neuron. 2004 Nov;44(4):595–600. doi:10.1016/j.neuron.2004.10.023 PubMed PMID: 15541308.

43. Healy DG, Falchi M, O'Sullivan SS, Bonifati V, Durr A, Bressman S, et al. Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: A case-control study. The Lancet Neurology. 2008 Jul;7(7):583–90. doi:10.1016/S1474-4422(08)70117-0 PubMed PMID: 18539534; PubMed Central PMCID: PMC2832754.

44. Steger M, Tonelli F, Ito G, Davies P, Trost M, Vetter M, et al. Phosphoproteomics reveals that Parkinson's disease kinase LRRK2 regulates a subset of Rab GTPases. eLife. 2016 Jan;5:e12813. doi:10.7554/eLife.12813 PubMed PMID: 26824392; PubMed Central PMCID: PMC4769169.

45. Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, Minoshima S, et al. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature. 1998 Apr;392(6676):605–8. doi:10.1038/33416 PubMed PMID: 9560156.

46. Valente EM, Abou-Sleiman PM, Caputo V, Muqit MMK, Harvey K, Gispert S, et al. Hereditary early-onset Parkinson's disease caused by mutations in PINK1. Science. 2004 May;304(5674):1158–60. doi:10.1126/science.1096284 PubMed PMID: 15087508.

47. Pickrell AM, Youle RJ. The roles of PINK1, parkin, and mitochondrial fidelity in Parkinson's disease. Neuron. 2015 Jan;85(2):257–73. doi:10.1016/j.neuron.2014.12.007 PubMed PMID: 25611507; PubMed Central PMCID: PMC4764997.

48. Busquets O, Li H, Syed KM, Jerez PA, Dunnack J, Bu RL, et al. iSCORE-PD: An isogenic stem cell collection to research Parkinson's Disease. bioRxiv: The Preprint Server for Biology. 2025 Mar;2024.02.12.579917. doi:10.1101/2024.02.12.579917 PubMed PMID: 38405931; PubMed Central PMCID: PMC10888955.

49. Bonifati V, Rizzu P, van Baren MJ, Schaap O, Breedveld GJ, Krieger E, et al. Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science. 2003 Jan;299(5604):256–9. doi:10.1126/science.1077209 PubMed PMID: 12446870.

50. Canet-Avilés RM, Wilson MA, Miller DW, Ahmad R, McLendon C, Bandyopadhyay S, et al. The Parkinson's disease protein DJ-1 is neuroprotective due to cysteine-sulfinic acid-driven mitochondrial localization. Proceedings of the National Academy of Sciences of the United States of America. 2004 Jun;101(24):9103–8. doi:10.1073/pnas.0402959101 PubMed PMID: 15181200; PubMed Central PMCID: PMC428480.

51. Pfaff DH, Fleming T, Nawroth P, Teleman AA. Evidence Against a Role for the Parkinsonism-associated Protein DJ-1 in Methylglyoxal Detoxification. The Journal

of Biological Chemistry. 2017 Jan;292(2):685–90. doi:10.1074/jbc.M116.743823 PubMed PMID: 27903648; PubMed Central PMCID: PMC5241742.

52. Andreeva A, Bekkhozhin Z, Omertassova N, Baizhumanov T, Yeltay G, Akhmetali M, et al. The apparent deglycase activity of DJ-1 results from the conversion of free methylglyoxal present in fast equilibrium with hemithioacetals and hemiaminals. The Journal of Biological Chemistry. 2019 Dec;294(49):18863–72. doi:10.1074/jbc.RA119.011237 PubMed PMID: 31653696; PubMed Central PMCID: PMC6901308.

53. Mazza MC, Shuck SC, Lin J, Moxley MA, Termini J, Cookson MR, et al. DJ -1 is not a deglycase and makes a modest contribution to cellular defense against methylglyoxal damage in neurons. Journal of Neurochemistry. 2022 Aug;162(3):245–61. doi:10.1111/jnc.15656

54. Coukos JS, Lee CW, Pillai KS, Shah H, Moellering RE. PARK7 Catalyzes Stereospecific Detoxification of Methylglyoxal Consistent with Glyoxalase and Not Deglycase Function. Biochemistry. 2023 Nov;62(21):3126–33. doi:10.1021/acs.biochem.3c00325 PubMed PMID: 37884446; PubMed Central PMCID: PMC10634309.

55. Krebiehl G, Ruckerbauer S, Burbulla LF, Kieper N, Maurer B, Waak J, et al. Reduced basal autophagy and impaired mitochondrial dynamics due to loss of Parkinson's disease-associated protein DJ-1. PloS One. 2010 Feb;5(2):e9367. doi:10.1371/journal.pone.0009367 PubMed PMID: 20186336; PubMed Central PMCID: PMC2826413.

56. Yamaguchi S, Yamane T, Takahashi-Niki K, Kato I, Niki T, Goldberg MS, et al. Transcriptional activation of low-density lipoprotein receptor gene by DJ-1 and effect of DJ-1 on cholesterol homeostasis. PloS One. 2012;7(5):e38144. doi:10.1371/journal.pone.0038144 PubMed PMID: 22666465; PubMed Central PMCID: PMC3364227.

57. Vilariño-Güell C, Wider C, Ross OA, Dachsel JC, Kachergus JM, Lincoln SJ, et al. VPS35 mutations in Parkinson disease. American Journal of Human Genetics. 2011 Jul;89(1):162–7. doi:10.1016/j.ajhg.2011.06.001 PubMed PMID: 21763482; PubMed Central PMCID: PMC3135796.

58. Zimprich A, Benet-Pagès A, Struhal W, Graf E, Eck SH, Offman MN, et al. A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. American Journal of Human Genetics. 2011 Jul;89(1):168–75. doi:10.1016/j.ajhg.2011.06.008 PubMed PMID: 21763483; PubMed Central PMCID: PMC3135812.

59. Zavodszky E, Seaman MNJ, Moreau K, Jimenez-Sanchez M, Breusegem SY, Harbour ME, et al. Mutation in VPS35 associated with Parkinson's disease impairs WASH complex association and inhibits autophagy. Nature Communications. 2014 May;5:3828. doi:10.1038/ncomms4828 PubMed PMID: 24819384; PubMed Central PMCID: PMC4024763.

60. Tang FL, Liu W, Hu JX, Erion JR, Ye J, Mei L, et al. VPS35 Deficiency or Mutation Causes Dopaminergic Neuronal Loss by Impairing Mitochondrial Fusion and Function. Cell Reports. 2015 Sep;12(10):1631–43. doi:10.1016/j.celrep.2015.08.001 PubMed PMID: 26321632; PubMed Central PMCID: PMC4565770.

61. Cutillo G, Simon DK, Eleuteri S. VPS35 and the mitochondria: Connecting the dots in Parkinson's disease pathophysiology. Neurobiology of Disease. 2020 Nov;145:105056. doi:10.1016/j.nbd.2020.105056 PubMed PMID: 32853677.

62. Shojaee S, Sina F, Banihosseini SS, Kazemi MH, Kalhor R, Shahidi GA, et al. Genome-wide linkage analysis of a Parkinsonian-pyramidal syndrome pedigree by 500 K SNP arrays. American Journal of Human Genetics. 2008 Jun;82(6):1375–84. doi:10.1016/j.ajhg.2008.05.005 PubMed PMID: 18513678; PubMed Central PMCID: PMC2427312.

63. Di Fonzo A, Dekker MCJ, Montagna P, Baruzzi A, Yonova EH, Correia Guedes L, et al. FBXO7 mutations cause autosomal recessive, early-onset parkinsonian-pyramidal syndrome. Neurology. 2009 Jan;72(3):240–5. doi:10.1212/01.wnl.0000338144.10967.2b PubMed PMID: 19038853.

64. Vingill S, Brockelt D, Lancelin C, Tatenhorst L, Dontcheva G, Preisinger C, et al. Loss of FBXO7 (PARK15) results in reduced proteasome activity and models a parkinsonism-like phenotype in mice. The EMBO journal. 2016 Sep;35(18):2008–25. doi:10.15252/embj.201593585 PubMed PMID: 27497298; PubMed Central PMCID: PMC5282834.

65. Burchell VS, Nelson DE, Sanchez-Martinez A, Delgado-Camprubi M, Ivatt RM, Pogson JH, et al. The Parkinson's disease-linked proteins Fbxo7 and Parkin interact to mediate mitophagy. Nature Neuroscience. 2013 Sep;16(9):1257–65. doi:10.1038/nn.3489 PubMed PMID: 23933751; PubMed Central PMCID: PMC3827746.

66. Lesage S, Drouet V, Majounie E, Deramecourt V, Jacoupy M, Nicolas A, et al. Loss of VPS13C Function in Autosomal-Recessive Parkinsonism Causes Mitochondrial Dysfunction and Increases PINK1/Parkin-Dependent Mitophagy. American Journal of Human Genetics. 2016 Mar;98(3):500–13. doi:10.1016/j.ajhg.2016.01.014 PubMed PMID: 26942284; PubMed Central PMCID: PMC4800038.

67. Wang X, Xu P, Bentley-DeSousa A, Hancock-Cerutti W, Cai S, Johnson BT, et al. The bridge-like lipid transport protein VPS13C/PARK23 mediates ER-lysosome contacts following lysosome damage. Nature Cell Biology. 2025 May;27(5):776–89. doi:10.1038/s41556-025-01653-6 PubMed PMID: 40211074; PubMed Central PMCID: PMC12081312.

68. Ramirez A, Heimbach A, Gründemann J, Stiller B, Hampshire D, Cid LP, et al. Hereditary parkinsonism with dementia is caused by mutations in ATP13A2, encoding a lysosomal type 5 P-type ATPase. Nature Genetics. 2006 Oct;38(10):1184–91. doi:10.1038/ng1884 PubMed PMID: 16964263.

69. Edvardson S, Cinnamon Y, Ta-Shma A, Shaag A, Yim YI, Zenvirt S, et al. A deleterious mutation in DNAJC6 encoding the neuronal-specific clathrin-uncoating co-chaperone auxilin, is associated with juvenile parkinsonism. PloS One. 2012;7(5):e36458. doi:10.1371/journal.pone.0036458 PubMed PMID: 22563501; PubMed Central PMCID: PMC3341348.

70. Wulansari N, Darsono WHW, Woo HJ, Chang MY, Kim J, Bae EJ, et al. Neurodevelopmental defects and neurodegenerative phenotypes in human brain organoids carrying Parkinson's disease-linked DNAJC6 mutations. Science Advances. 2021 Feb;7(8):eabb1540. doi:10.1126/sciadv.abb1540 PubMed PMID: 33597231; PubMed Central PMCID: PMC7888924.

71. Krebs CE, Karkheiran S, Powell JC, Cao M, Makarov V, Darvish H, et al. The Sac1 domain of SYNJ1 identified mutated in a family with early-onset progressive Parkinsonism with generalized seizures. Human Mutation. 2013 Sep;34(9):1200–7. doi:10.1002/humu.22372 PubMed PMID: 23804563; PubMed Central PMCID: PMC3790461.

72. Leroy E, Boyer R, Auburger G, Leube B, Ulm G, Mezey E, et al. The ubiquitin pathway in Parkinson's disease. Nature. 1998 Oct;395(6701):451–2. doi:10.1038/26652 PubMed PMID: 9774100.

73. Healy DG, Abou-Sleiman PM, Casas JP, Ahmadi KR, Lynch T, Gandhi S, et al. UCHL-1 is not a Parkinson's disease susceptibility gene. Annals of Neurology. 2006 Apr;59(4):627–33. doi:10.1002/ana.20757 PubMed PMID: 16450370.

74. Nalls MA, Pankratz N, Lill CM, Do CB, Hernandez DG, Saad M, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nature Genetics. 2014 Sep;46(9):989–93. doi:10.1038/ng.3043 PubMed PMID: 25064009; PubMed Central PMCID: PMC4146673.

75. Kim JJ, Vitale D, Otani DV, Lian MM, Heilbron K, 23andMe Research Team, et al. Multi-ancestry genome-wide association meta-analysis of Parkinson's disease.

Nature Genetics. 2024 Jan;56(1):27–36. doi:10.1038/s41588-023-01584-8 PubMed PMID: 38155330; PubMed Central PMCID: PMC10786718.

76. The Global Parkinson's Genetics Program (GP2), Leonard HL. Novel Parkinson's Disease Genetic Risk Factors Within and Across European Populations. bioRxiv; 2025. doi:10.1101/2025.03.14.24319455

77. Sidransky E, Nalls MA, Aasly JO, Aharon-Peretz J, Annesi G, Barbosa ER, et al. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. The New England Journal of Medicine. 2009 Oct;361(17):1651–61. doi:10.1056/NEJMoa0901281 PubMed PMID: 19846850; PubMed Central PMCID: PMC2856322.

78. Schilder BM, Raj T. Fine-mapping of Parkinson's disease susceptibility loci identifies putative causal variants. Human Molecular Genetics. 2022 Mar;31(6):888–900. doi:10.1093/hmg/ddab294 PubMed PMID: 34617105; PubMed Central PMCID: PMC8947317.

79. Gallagher MD, Chen-Plotkin AS. The Post-GWAS Era: From Association to Function. American Journal of Human Genetics. 2018 May;102(5):717–30. doi:10.1016/j.ajhg.2018.04.002 PubMed PMID: 29727686; PubMed Central PMCID: PMC5986732.

80. Rizig M, Bandres-Ciga S, Makarious MB, Ojo OO, Crea PW, Abiodun OV, et al. Identification of genetic risk loci and causal insights associated with Parkinson's disease in African and African admixed populations: A genome-wide association study. The Lancet Neurology. 2023 Nov;22(11):1015–25. doi:10.1016/S1474-4422(23)00283-1 PubMed PMID: 37633302; PubMed Central PMCID: PMC10593199.

81. Kraus F, Goodall EA, Smith IR, Jiang Y, Paoli JC, Adolf F, et al. PARK15/FBXO7 is dispensable for PINK1/Parkin mitophagy in iNeurons and HeLa cell systems. EMBO reports. 2023 Aug;24(8):e56399. doi:10.15252/embr.202256399 PubMed PMID: 37334901; PubMed Central PMCID: PMC10398645.

82. McCarron KR, Elcocks H, Mortiboys H, Urbé S, Clague MJ. The Parkinson's disease related mutant VPS35 (D620N) amplifies the LRRK2 response to endolysosomal stress. The Biochemical Journal. 2024 Feb;481(4):265–78. doi:10.1042/BCJ20230492 PubMed PMID: 38299383; PubMed Central PMCID: PMC10903469.

83. Boecker CA, Goldsmith J, Dou D, Cajka GG, Holzbaur ELF. Increased LRRK2 kinase activity alters neuronal autophagy by disrupting the axonal transport of autophagosomes. Current biology: CB. 2021 May;31(10):2140–2154.e6.

doi:10.1016/j.cub.2021.02.061 PubMed PMID: 33765413; PubMed Central PMCID: PMC8154747.

84. Replogle JM, Saunders RA, Pogson AN, Hussmann JA, Lenail A, Guna A, et al. Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. Cell. 2022 Jul;185(14):2559–2575.e28. doi:10.1016/j.cell.2022.05.013 PubMed PMID: 35688146; PubMed Central PMCID: PMC9380471.

85. Costa-Mattioli M, Walter P. The integrated stress response: From mechanism to disease. Science. 2020 Apr;368(6489):eaat5314. doi:10.1126/science.aat5314 PubMed PMID: 32327570; PubMed Central PMCID: PMC8997189.

86. Castellano BM, Thelen AM, Moldavski O, Feltes M, van der Welle REN, Mydock-McGrane L, et al. Lysosomal cholesterol activates mTORC1 via an SLC38A9-Niemann-Pick C1 signaling complex. Science. 2017 Mar;355(6331):1306–11. doi:10.1126/science.aag1417 PubMed PMID: 28336668; PubMed Central PMCID: PMC5823611.

87. Paisan-Ruiz C, Bhatia KP, Li A, Hernandez D, Davis M, Wood NW, et al. Characterization of PLA2G6 as a locus for dystonia-parkinsonism. Annals of Neurology. 2009 Jan;65(1):19–23. doi:10.1002/ana.21415 PubMed PMID: 18570303; PubMed Central PMCID: PMC9016626.

88. Soldner F, Jaenisch R. Stem Cells, Genome Editing, and the Path to Translational Medicine. Cell. 2018 Oct;175(3):615–32. doi:10.1016/j.cell.2018.09.010 PubMed PMID: 30340033; PubMed Central PMCID: PMC6461399.

89. Jerber J, Seaton DD, Cuomo ASE, Kumasaka N, Haldane J, Steer J, et al. Population-scale single-cell RNA-seq profiling across dopaminergic neuron differentiation. Nature Genetics. 2021 Mar;53(3):304–12. doi:10.1038/s41588-021-00801-6 PubMed PMID: 33664506; PubMed Central PMCID: PMC7610897.

90. Liang G, Zhang Y. Genetic and epigenetic variations in iPSCs: Potential causes and implications for application. Cell Stem Cell. 2013 Aug;13(2):149–59. doi:10.1016/j.stem.2013.07.001 PubMed PMID: 23910082; PubMed Central PMCID: PMC3760008.

91. Pantazis CB, Yang A, Lara E, McDonough JA, Blauwendraat C, Peng L, et al. A reference human induced pluripotent stem cell line for large-scale collaborative studies. Cell Stem Cell. 2022 Dec;29(12):1685–1702.e22. doi:10.1016/j.stem.2022.11.004 PubMed PMID: 36459969; PubMed Central PMCID: PMC9782786.

92. Lengner CJ, Gimelbrant AA, Erwin JA, Cheng AW, Guenther MG, Welstead GG, et al. Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations. Cell. 2010 May;141(5):872–83. doi:10.1016/j.cell.2010.04.010 PubMed PMID: 20471072.

93. Soldner F, Laganière J, Cheng AW, Hockemeyer D, Gao Q, Alagappan R, et al. Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell. 2011 Jul;146(2):318–31. doi:10.1016/j.cell.2011.06.019 PubMed PMID: 21757228; PubMed Central PMCID: PMC3155290.

94. Soldner F, Stelzer Y, Shivalila CS, Abraham BJ, Latourelle JC, Barrasa MI, et al. Parkinson-associated risk variant in distal enhancer of α-synuclein modulates target gene expression. Nature. 2016 May;533(7601):95–9. doi:10.1038/nature17939 PubMed PMID: 27096366; PubMed Central PMCID: PMC5042324.

95. Nass RD, Hampel KG, Elger CE, Surges R. Blood Pressure in Seizures and Epilepsy. Frontiers in Neurology. 2019;10:501. doi:10.3389/fneur.2019.00501 PubMed PMID: 31139142; PubMed Central PMCID: PMC6527757.

96. Vallot C, Ouimette JF, Makhlouf M, Féraud O, Pontis J, Côme J, et al. Erosion of X Chromosome Inactivation in Human Pluripotent Cells Initiates with XACT Coating and Depends on a Specific Heterochromatin Landscape. Cell Stem Cell. 2015 May;16(5):533–46. doi:10.1016/j.stem.2015.03.016

97. Kriks S, Shim JW, Piao J, Ganat YM, Wakeman DR, Xie Z, et al. Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease. Nature. 2011 Nov;480(7378):547–51. doi:10.1038/nature10648 PubMed PMID: 22056989; PubMed Central PMCID: PMC3245796.

98. Kim TW, Piao J, Koo SY, Kriks S, Chung SY, Betel D, et al. Biphasic Activation of WNT Signaling Facilitates the Derivation of Midbrain Dopamine Neurons from hESCs for Translational Use. Cell Stem Cell. 2021 Feb;28(2):343–355.e5. doi:10.1016/j.stem.2021.01.005 PubMed PMID: 33545081; PubMed Central PMCID: PMC8006469.

99. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. 2013 Jul;154(2):442–51. doi:10.1016/j.cell.2013.06.044 PubMed PMID: 23849981; PubMed Central PMCID: PMC3770145.

100. Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, et al. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell.

2014 Oct;159(3):647–61. doi:10.1016/j.cell.2014.09.029 PubMed PMID: 25307932; PubMed Central PMCID: PMC4253859.

101. Alerasool N, Segal D, Lee H, Taipale M. An efficient KRAB domain for CRISPRi applications in human cells. Nature Methods. 2020 Nov;17(11):1093–6. doi:10.1038/s41592-020-0966-x PubMed PMID: 33020655.

102. Chavez A, Scheiman J, Vora S, Pruitt BW, Tuttle M, P R Iyer E, et al. Highly efficient Cas9-mediated transcriptional programming. Nature Methods. 2015 Apr;12(4):326–8. doi:10.1038/nmeth.3312 PubMed PMID: 25730490; PubMed Central PMCID: PMC4393883.

103. Haapaniemi E, Botla S, Persson J, Schmierer B, Taipale J. CRISPR-Cas9 genome editing induces a P53-mediated DNA damage response. Nature Medicine. 2018 Jul;24(7):927–30. doi:10.1038/s41591-018-0049-z PubMed PMID: 29892067.

104. Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell. 2016 Dec;167(7):1853–1866.e17. doi:10.1016/j.cell.2016.11.038 PubMed PMID: 27984732; PubMed Central PMCID: PMC5181115.

105. Adamson B, Norman TM, Jost M, Cho MY, Nuñez JK, Chen Y, et al. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell. 2016 Dec;167(7):1867–1882.e21. doi:10.1016/j.cell.2016.11.048 PubMed PMID: 27984733; PubMed Central PMCID: PMC5315571.

106. Norman TM, Horlbeck MA, Replogle JM, Ge AY, Xu A, Jost M, et al. Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. Science. 2019 Aug;365(6455):786–93. doi:10.1126/science.aax4438

107. Tian R, Abarientos A, Hong J, Hashemi SH, Yan R, Dräger N, et al. Genome-wide CRISPRi/a screens in human neurons link lysosomal failure to ferroptosis. Nature Neuroscience. 2021 Jul;24(7):1020–34. doi:10.1038/s41593-021-00862-0

108. Dräger NM, Sattler SM, Huang CTL, Teter OM, Leng K, Hashemi SH, et al. A CRISPRi/a platform in human iPSC-derived microglia uncovers regulators of disease states. Nature Neuroscience. 2022 Sep;25(9):1149–62. doi:10.1038/s41593-022-01131-4

109. Wessels HH, Méndez-Mancilla A, Hao Y, Papalexi E, Mauck WM, Lu L, et al. Efficient combinatorial targeting of RNA transcripts in single cells with Cas13 RNA Perturb-

seq. Nature Methods. 2023 Jan;20(1):86–94. doi:10.1038/s41592-022-01705-x PubMed PMID: 36550277; PubMed Central PMCID: PMC10030154.

110. Xu Z, Sziraki A, Lee J, Zhou W, Cao J. Dissecting key regulators of transcriptome kinetics through scalable single-cell RNA profiling of pooled CRISPR screens. Nature Biotechnology. 2024 Aug;42(8):1218–23. doi:10.1038/s41587-023-01948-9

Chapter 2: iSCORE-PD — An Isogenic Stem Cell Collection to Research Parkinson's Disease

This chapter is adapted from co-first-authored research article our group shared on bioRxiv, and was recently accepted at *Nature Communications:*

Busquets, O.*; Li, H.*; Syed, K.M.*; Alvarez Jerez, P. *; Dunnack, J. *; Lo Bu, R.; Verma, Y.; Pangilinan, G.R.; Martin, A.; Straub, J.; Du, Y.; Simon, V.M.; Poser, S.; Bush, Z.; Diaz, J.; Sahagun, A.; Gao, J.; Hernandez, D.G.; Levine, K.S.; Pochet, N.; Booth, E.O.; Bateup, H.S.; Rio, D.C.; Hockemeyer, D.; Blauwendraat, C.; Soldner, F. iSCORE-PD: an isogenic stem cell collection to research Parkinson's Disease. *Nature Communications* (2026).

*These authors contributed equally.

The manuscript is adapted in accordance with UC Berkeley's Graduate Division policy on the inclusion of published and co-authored material in doctoral dissertations.

Abstract

Genome-edited human pluripotent stem cells (hPSCs) provide a powerful platform to study complex diseases such as Parkinson's disease (PD). Here, we describe iSCORE-PD, an isogenic collection of 65 genome-edited hPSC lines carrying disease-causing or high-risk variants in 11 PD-linked genes (*SNCA, PRKN, PINK1, DJ1/PARK7, LRRK2, ATP13A2, FBXO7, DNAJC6, SYNJ1, VPS13C,* and *GBA1*). All lines are derived from a well-characterized female hESC line and subjected to extensive quality control. Whole-genome sequencing reveals that genetic variation between lines, largely confined to non-coding regions, is minimal relative to inter-individual differences in patient-derived hiPSCs, with most variation arising from random mutations acquired during cell culture rather than genome-editing-induced off-target effects. Including multiple independently derived clones per mutation controls for this random genetic drift. Our systematic approach ensures high quality of this publicly available iSCORE-PD resource, highlights the advantages of prime editing over conventional CRISPR/Cas9 methods, and establishes best practices for generating disease-modeling hPSC collections.

2.1 Chapter Preamble

My Contributions

The iSCORE-PD cell line collection represents a panel of isogenic hESC lines which harbor 13 mutations across 11 familial-PD genes. My primary contribution to iSCORE-PD manuscript centered on the single-cell RNA sequencing (scRNA-Seq) of dopaminergic (DA) neurons differentiated from three subclones of the WIBR3 parental line. I assisted in assaying these samples and their scRNA-Seq library prep, and I led the computational analysis, cluster annotation, and cross-reference comparison against the FOUNDIN-PD reference dataset. This work is outlined in Figure 2 and Supplemental Figures 3 and 4.

In addition, I contributed to the genome-editing arm of the collection by assisting with the design of prime-editing and CRISPR/Cas9 reagents for the VPS13C W395C and A444P mutations, helping to isolate and quality control multiple subclones for each via immunocytochemistry and alkaline phosphatase staining to assess markers of pluripotency.

All other work – gene-editing for other PD mutations, whole-genome sequencing and structural variant analysis, microglia differentiation experiments, and the overall coordination of the iSCORE-PD collection – represent the contributions of the other co-first authors and collaborators. All authors reviewed and approved the final manuscript, and all co-first and lead authors have consented to inclusion of the full work in this dissertation.

Context within this dissertation

iSCORE-PD provides an isogenic resource which complements the CRISPRi Perturb-seq experiments presented in Chapter 3. Both the iSCORE-PD cell lines and the CRISPRi clone used in Chapter 3 were derived from the same WIBR3 hESC parental background (see section 3.7 for a caveat about isogenicity). The dopaminergic-neuron atlas established here — cell-type annotations, cluster identities, and correlation to the FOUNDIN-PD reference — directly informs the cell-type annotation framework applied to the Perturb-seq dataset in Chapter 3, where the mature DA neuron clusters serve as an analytical focus for Mixscale-based differential expression testing.

Finally, the iSCORE-PD collection and the CRISPRi Perturb-seq dataset constitute complementary experimental approaches to the same underlying biological question. The work in this chapter surveys the transcriptional consequences of endogenous, disease-causing mutations across eleven familial PD genes. Chapter 3 applies acute, dosage-controlled CRISPRi knockdown to a partially overlapping and substantially expanded gene set (41 Familial PD genes plus 310 GWAS-nominated candidates) in the same genetic background. Systematic comparison of mutation signatures and CRISPRi perturbation signatures is the next step toward elucidating cell-type-specific effects resulting from the disruption of PD risk genes.

2.2 Introduction

As outlined in Chapter 1, Parkinson's disease is pathologically defined by a relatively narrow set of cellular phenotypes, namely the selective and progressive degeneration of dopaminergic (DA) neurons in the substantia nigra pars compacta (SNc) and the accumulation of Lewy body protein aggregates[1]. The proposed mechanisms behind these phenotypes, however, are remarkably broad, spanning mitochondrial quality control, endolysosomal trafficking, synaptic vesicle recycling, the ubiquitin-proteasome system, lipid metabolism, protein translation, and lysosomal degradation machinery[1,2]. Although advancements in genome editing, human cell culture systems and multi-omics tools have presented new opportunities to elucidate the genetic underpinnings of cellular dysfunction in PD, limited genetic control in certain experimental systems has precluded the ability to bridge this genotype-phenotype gap.

Even in cases of monogenic PD, individuals carrying the same pathogenic variant frequently show heterogeneous clinical and pathological presentation, including variable age at onset, incomplete penetrance, and diverse patterns of Lewy-body pathology — even its complete absence in some carriers[1]. This phenotypic heterogeneity highlights the additional contributions to PD risk made by other genetic and environmental factors, with effects that may compound or modify those of the primary mutation. This poses a fundamental challenge for the study of PD mechanisms: in an experimental system where the genetic backgrounds of the samples are themselves confounded, distinguishing the effects of a highly penetrant pathogenic variant from those of common, low-effect-size modifiers is difficult, and phenotypic differences attributed to the disease-causing mutation may instead reflect background heterogeneity.

Genome editing of human pluripotent stem cells (hPSCs) has emerged as an approach which addresses this confound. Using CRISPR/Cas9 or prime editing, disease-causing mutations can be introduced into (or corrected in) a single parental hPSC line, generating a pair of cell lines which differ only at the engineered locus[3,4]. Isogenic pairs of this kind have been used to characterize individual PD-associated mutations in patient-specific genetic backgrounds[5,6]. However, when each isogenic pair is generated on a distinct, patient-derived parental line, systematic comparison of mutations across different genetic backgrounds remains difficult and partially confounded by line-to-line heterogeneity. What the field has lacked is the systematic deployment of a single, well-characterized cell line into which a panel of PD-associated mutations is engineered, and with rigorous quality control.

An important recent example of this approach is the iPSC Neurodegenerative Disease Initiative (iNDI) led by the NIH Center for Alzheimer's and Related Dementias, which employed the KOLF2.1J hiPSC line as a common parental background for a panel of neurodegenerative disease mutations[7]. KOLF2.1J is derived from a male donor and has been extensively characterized as a benchmark reference line for the field. Alongside the scientific case for expanding the genetic and sex diversity of available reference lines, the

broader literature on hiPSCs raises a distinct set of concerns: somatic mutations present in the donor before reprogramming, reprogramming-induced genetic and epigenetic alterations, and aberrant genomic imprinting have all been documented in iPSC-derived lines[3,8,9]. Human embryonic stem cell (hESC) lines, derived directly from preimplantation embryos rather than from reprogrammed somatic cells, avoid this subset of concerns. For the work described here, we selected the female hESC line WIBR3[10] as the parental line for a complementary isogenic PD collection. WIBR3 has been characterized at the karyotypic and functional levels, maintains stable ploidy over prolonged in vitro culture, has been used in multiple prior PD modeling studies[5,6], and through previous work in our laboratory has been established as a substrate for efficient CRISPR/Cas9, TALEN, and prime-editing approaches[4,11]. Starting from early-passage WIBR3 cells and three independently derived single-cell subclones (WIBR3-S1, S2, S3), we assembled the iSCORE-PD (Isogenic Stem Cell Collection to Research Parkinson's Disease) resource as part of the Aligning Science Across Parkinson's (ASAP) Collaborative Research Network[12].

The goals of the work described in this chapter were fourfold: first, to comprehensively characterize the WIBR3 parental line using long-read whole-genome sequencing, high-density array genotyping, polygenic risk assessment, and functional assays of pluripotency and of the p53-dependent DNA damage response, in order to establish the baseline genomic and cellular state on which all downstream editing would be built. Second, we sought to validate that WIBR3 and its subclones differentiate efficiently into PD-relevant cell types and to benchmark the resulting populations against an established reference dataset (FOUNDIN-PD)[13]. Third, we looked to engineer pathogenic variants in eleven high-confidence monogenic PD genes using a combination of CRISPR/Cas9, TALEN, and prime editing approaches: *SNCA* (A53T, A30P), *PRKN* (Exon 3 deletion), *PINK1* (Q129X), *PARK7*/DJ-1 (Exon 1–5 deletion), *LRRK2* (G2019S), *ATP13A2* (frameshift), *FBXO7* (R498X/frameshift), *DNAJC6* (c.801-2 A>G/frameshift), *SYNJ1* (R258Q/frameshift), *VPS13C* (A444P and W395C, each with accompanying frameshift alleles), and *GBA1* (IVS2+1 splice-site mutation). Fourth, we focused on applying a uniform quality-control framework to each clonal cell line in the collection, including array comparative genomic hybridization (aCGH), high-density Illumina Infinium Global Diversity Array / Neuro Booster Array genotyping, zygosity analysis to detect loss of heterozygosity at edited loci, and short-read whole-genome sequencing to quantify the genetic variability that arises both from the editing process as well as from routine clonal expansion.

The 65 clonal cell lines which passed quality control constitute the iSCORE-PD cell-line resource, shared alongside three wild-type WIBR3 subclones and a set of "edited-wild-type" control lines which emerged from editing pipelines without incurring edits at the targeted loci. All lines have been deposited with the WiCell Research Institute and are available to the scientific community through standard material transfer agreements. The work in this chapter sets the stage for Chapter 3, where we perform a CRISPRi Perturb-seq screen of 351 PD risk genes in dopaminergic neuron cultures from the same WIBR3 background.

The remainder of this chapter contains an abridged adaptation of the iSCORE-PD manuscript. The Results sections describing the parental WIBR3 characterization, dopaminergic and microglia differentiation, the genome-editing pipelines, and the quality-control framework are presented in full. The whole-genome sequencing analysis of genetic variability across the collection and the off-target characterization (presented as Figures 4 and 5 of the published manuscript) are summarized at reduced depth in §2.3.6 and §2.3.7. Per-gene editing strategies, genotyping, and zygosity analyses for each of the eleven targeted PD genes are reproduced in Appendix (§2.A) rather than in the main results below.

2.3 Results

2.3.1 - Characterization of the WIBR3 hESC line

As outlined in §2.2, there are several limitations patient-derived iPSCs, including genetic alterations present in donor cells before reprogramming, reprogramming-induced genetic alterations, incomplete epigenetic reprogramming, and aberrant genomic imprinting[3,8]. In consideration of these and in addressing the need to expand the genetic and sex diversity of available reference lines, we selected the female hESC line WIBR3 (NIH approval number NIHhESC-10-0079) as the parental line for the iSCORE-PD collection. WIBR3 has been previously demonstrated to maintain a stable karyotype over prolonged in vitro culture and has been widely used to model human diseases including PD[4,5,10,14]. We acquired early-passage WIBR3 cells (P14) and initially generated three independent single-cell-derived subclones (WIBR3-S1, WIBR3-S2, WIBR3-S3). The parental line and its subclones showed regular growth and morphology when cultured on mouse fibroblast feeders (MEFs) and under feeder-free conditions in mTeSR Plus media (Figure 2.1A). We validated the pluripotency of all cell lines through detection of pluripotency markers using immunocytochemistry and qRT-PCR (Figure 2.1B,C). Furthermore, we analyzed the genomic integrity of all cell lines using standard array comparative genomic hybridization (aCGH) and a modified high-density Illumina Infinium Global Diversity Array (GDA) Neuro Booster Array (NBA). This analysis confirmed a normal karyotype and the absence of larger structural alterations (greater than approximately 500 kb) for both the parental WIBR3 line and its derived subclones.

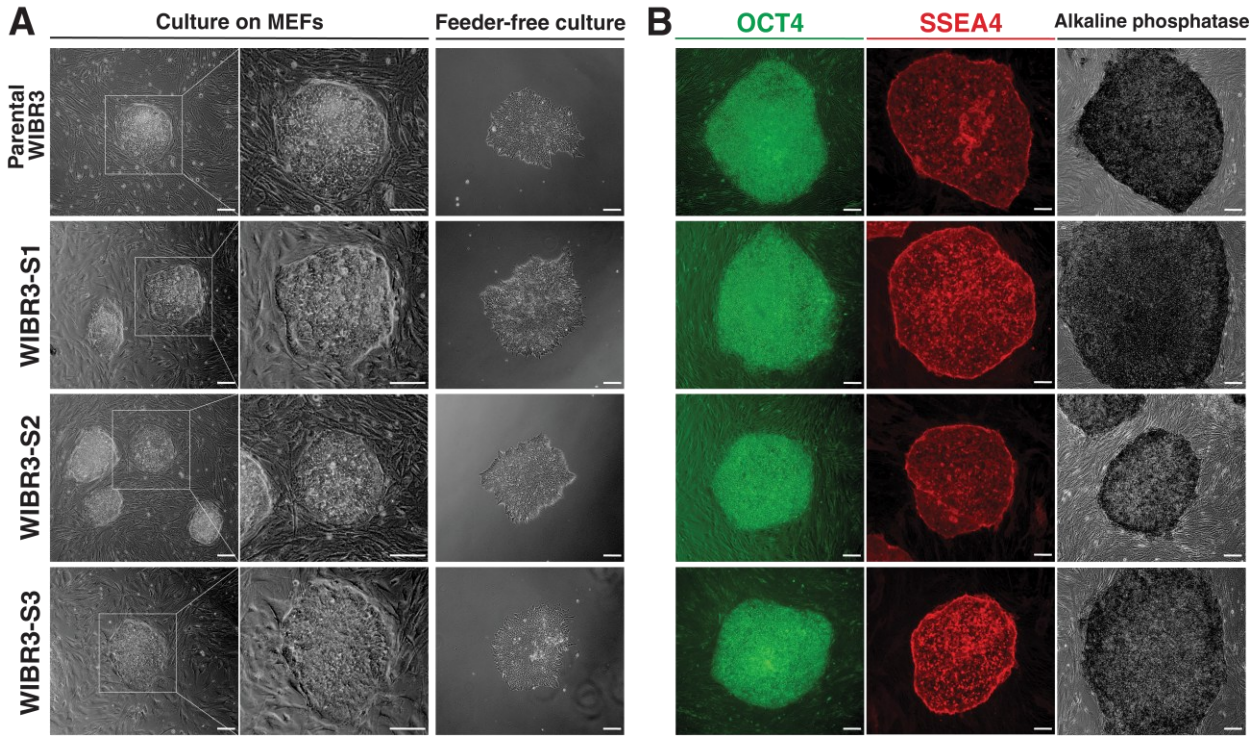


Figure 2.1A and 2.1B – (A) *Phase contrast images of parental WIBR3 hESCs and subclones WIBR3-S1, WIBR3-S2, and WIBR3-S3, grown on mouse embryonic fibroblast (MEF) and feeder-free conditions. Scale bar 100 µm.* ***(B)*** *Immunocytochemistry for pluripotency markers OCT4 (green) and SSEA4 (red) and staining for alkaline phosphatase (black) of WIBR3 parental hESCs and subclones cultured on MEFs. Scale bar = 100 µm.*

To determine the presence of insertions and deletions (indels) at higher resolution and to identify potential pathogenic single nucleotide variants (SNVs) in the parental WIBR3 hESC line, we performed long-read whole genome sequencing (WGS) combining Pacific Biosciences (PacBio, average coverage 28.39X, median read length 18 kb) and Oxford Nanopore Technologies (Nanopore, average coverage 43X, median read length 84 kb). The initial analysis for structural variants, using the Truvari algorithm to integrate the PacBio and Nanopore datasets, identified a total of 20,561 high-confidence structural variants in the WIBR3 parental line compared to the reference human genome (GRCh38/hg38). Among these, 109 were localized to coding exons, impacting 102 genes. The number and distribution of these structural variants is comparable to those observed in the general human population[15,16]. As our goal was to model PD and related neurodegenerative diseases, we determined that none of these structural variants affects genes with known pathogenic mutations in PD, Alzheimer's disease (AD) and AD-related dementias (ADRD), or risk genes identified in GWAS associated with these diseases[2,17,18]. Additionally, we annotated the integrated structural variant calls using SVAnna to determine if any variant was of high priority for the phenotype terms HP:0002180 (neurodegeneration) and HP:0000707 (abnormality of the nervous system). SVAnna did not identify any structural variant as high priority for either of these terms.

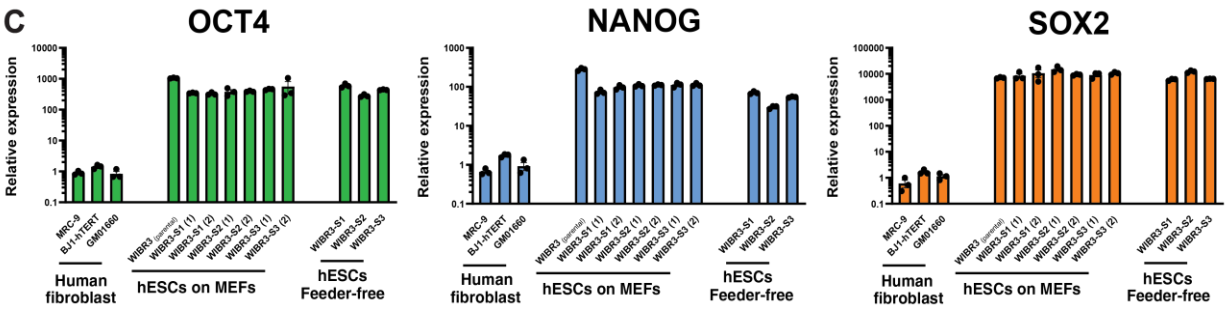


Figure 2.1C - *qRT-PCR analysis for the relative expression of pluripotency markers OCT4, NANOG, and SOX2 in human primary fibroblasts (MRC-9, BJ1-hTERT, GM01660), WIBR3 parental hESCs, and subclones cultured on MEFs and in feeder-free conditions. Relative expression levels normalized to expression in primary fibroblasts. (N=3; Mean ± SEM).*

Next, we identified the number and distribution of coding, missense, frameshift and predicted loss-of-function (LOF) SNVs in the parental WIBR3 cell line compared to the reference human genome. This analysis identified 6,613 missense SNVs in 3,933 coding genes and 120 potential loss-of-function mutations (including 15 startloss, 54 stopgain, 13 stoploss, 18 frameshift deletion, and 20 frameshift insertion). Overall, the number and distribution of these variants is comparable to other sequenced hPSC lines[7] and to the human population in the gnomAD database[19,20].

To determine the presence of potentially pathogenic variants in the parental WIBR3 cell line, we annotated all SNVs using ClinVar. Collectively, these analyses revealed 47 variants listed as pathogenic or with conflicting interpretations of pathogenicity; however, none of these are in a neurological phenotype of interest or are convincingly pathogenic. As we aim to provide our cell collection to study PD and related neurodegenerative diseases, we analyzed the WGS to calculate the polygenic risk score (PRS) based on the cumulative number of GWAS risk variants for PD (Figure 2.1D)[2]. The analysis indicated that the PRS of the parental WIBR3 line falls within the range observed in the normal population. Next, we specifically focused on identifying high-risk variants in known neurodegenerative disease-associated genes including *APOE*, *TMEM106B* and *MAPT*. This analysis determined that WIBR3 is heterozygous for the *APOE* ε4 allele (a risk factor for Alzheimer's disease), heterozygous for rs3173615 (TMEM106B p.T185S)[21] (reported as a modifier of frontotemporal dementia), and homozygous for the *MAPT* H1 allele, a gene of interest in several diseases[22,23].

It is widely recognized that hPSCs can accumulate genetic alterations over time that provide a growth advantage in cell culture. Notably, mutations in the p53 tumor suppressor pathway have been frequently observed in various hPSC lines[24,25]. To evaluate the function of the p53 pathway in the parental WIBR3 cell line, we analyzed the p53-dependent DNA damage response following irradiation (Figure 2.1E). The induction of DNA damage response genes *CDKN1A* and *RPS27L* confirmed a robust p53 response in undifferentiated cells and in *in vitro*-derived cortical spheroids (Figure 2.1E).

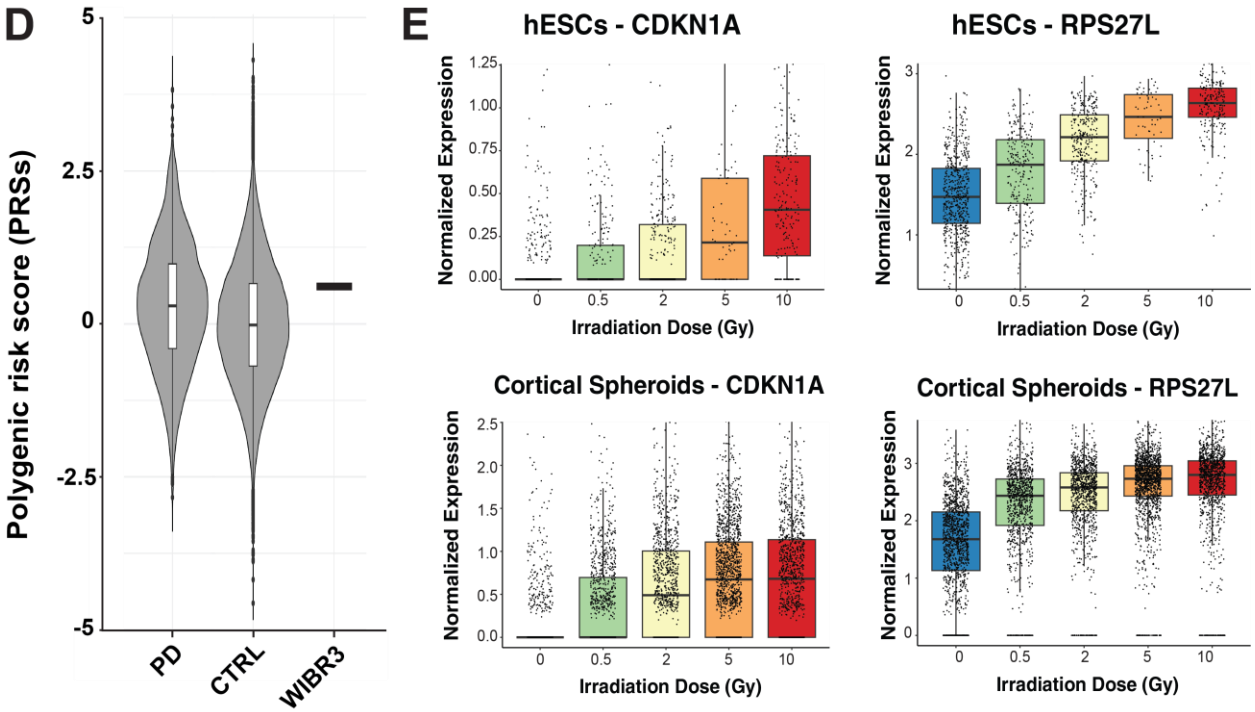


Figure 2.1D and 2.1E - ***(D)*** *Polygenic risk scores (PRSs) for PD comparing WIBR3 hESCs to population-centered Z-score distribution for PD PRSs in individuals with PD and the normal population from the UK Biobank.* ***(E)*** *Assessment of p53 pathway activity following irradiation (0.5, 2, 5, and 10 Gy) of WIBR3 parental hESCs (1,464 cells) and WIBR3-derived cortical spheroids (5,920 cells) by scRNA-seq analysis for the expression of DNA damage response genes CDKN1A and RPS27L (box plot showing interquartile range with a line at the median).*

2.3.2 - WIBR3 cells differentiate into PD-relevant cell types

Given that PD is characterized by the chronic progressive loss of dopaminergic (DA) neurons in the substantia nigra, the ability to effectively generate these cell types is crucial for *in vitro* modeling of PD. Therefore, we implemented a previously established protocol[26,27] to differentiate the three independently derived WIBR3 subclones (WIBR3-S1, WIBR3-S2, WIBR3-S3) into midbrain-specified dopaminergic neurons. Briefly, WIBR3 hESCs were subjected to dual SMAD inhibition for neural induction along with sonic hedgehog (SHH) agonist exposure for midbrain-specific patterning for 11 days. Concentrations of the GSK-3 inhibitor (CHIR99021) were increased from days 4 to 11. From day 11, committed midbrain neural progenitors were differentiated into dopamine neurons until day 35 (Figure 2.2A). For each subclone, we determined the efficacy of neural induction into neural precursor cells and dopaminergic neurons by analyzing the expression of midbrain-specific genes. At early time points (day 11 and day 25), the in vitro differentiated cultures expressed the expected midbrain floor plate markers at levels comparable to those in concurrently differentiated KOLF2.1J hiPSCs (Supplemental Figure 2.2B). At day 35 of differentiation, greater than 80% of cells expressed FOXA2, a marker for early midbrain floorplate neuronal precursors, and approximately 20% expressed the dopaminergic neuron marker tyrosine hydroxylase (TH), indicating the generation of midbrain-specified dopaminergic neurons (Figure 2.2B).

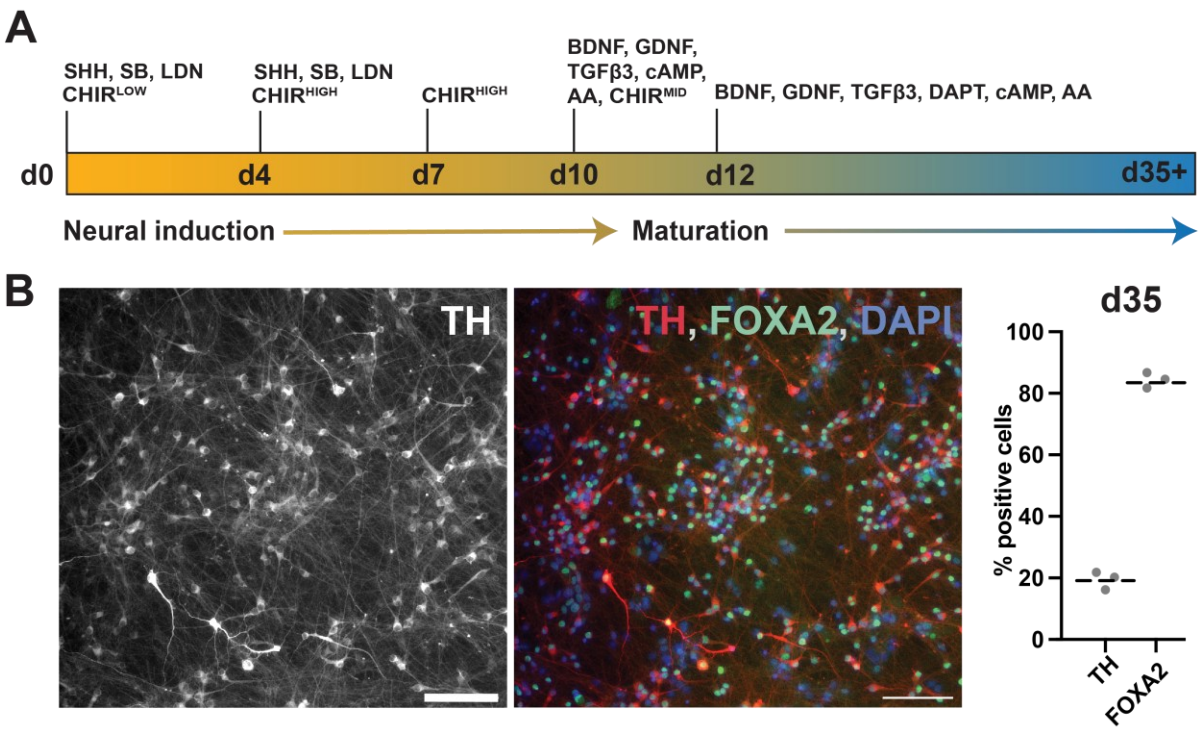


Figure 2.2A and 2.2B – (A) *Schematic depicting the protocol for in vitro differentiation of dopaminergic neurons from WIBR3 hESCs.* ***(B)*** *Immunocytochemistry and quantification of TH- and FOXA2-expressing cells in WIBR3 parental hESC-derived dopaminergic neurons at day 35. Scale bar 100 µm (N=3).*

To further characterize the DA neuron cultures, we performed single-cell RNA sequencing (scRNA-seq) at day 35–37 post-differentiation from hESCs. We profiled a total of 10,097 cells across the three independently assayed subclones, yielding an aggregate dataset. Following standard approaches for dimensionality reduction and unsupervised clustering, seven distinct clusters were present in the integrated dataset, each composed of cells belonging to all subclones (Figure 2.2C, Figure 2.6A). Out of the seven clusters, clusters 0, 1, 3, 4, and 6 expressed canonical dopaminergic neuron markers (KCNJ6, TH, NR4A2), whereas clusters 2 and 5 expressed midbrain progenitor markers (CALB1, SOX6 and CORIN) (Figure 2.2D, Figure S2B). To further compare the differentiation propensity of WIBR3 cells against other cell lines with different genetic backgrounds, we compared our dataset with the recently published Foundational Data Initiative for Parkinson Disease (FOUNDIN-PD) data (Figure 2.2E, Figure S3A-C)[13]. The FOUNDIN-PD dataset includes single-cell RNA-seq data from midbrain DA neuron cultures at day 65, derived from 80 distinct iPSC lines using a comparable in vitro differentiation protocol[13]. This comparison revealed that the clusters representing DA neuron populations (clusters 0, 1, 3, 4, and 6) in our dataset showed the highest Spearman correlation scores with iDA1, iDA2, iDA3, and iDA4 neuron clusters identified in the FOUNDIN-PD data, and that clusters 2 and 5 from our dataset have higher similarity to progenitor populations. Together, this analysis indicates a high similarity in the expression profiles of our WIBR3-derived cell types to the midbrain-derived cell types found in the FOUNDIN-PD dataset, currently considered a benchmark for in vitro-derived midbrain-specific DA neurons. This cross-dataset correspondence is

relevant as it allows the integration of data generated from the iSCORE-PD collection with FOUNDIN-PD datasets.

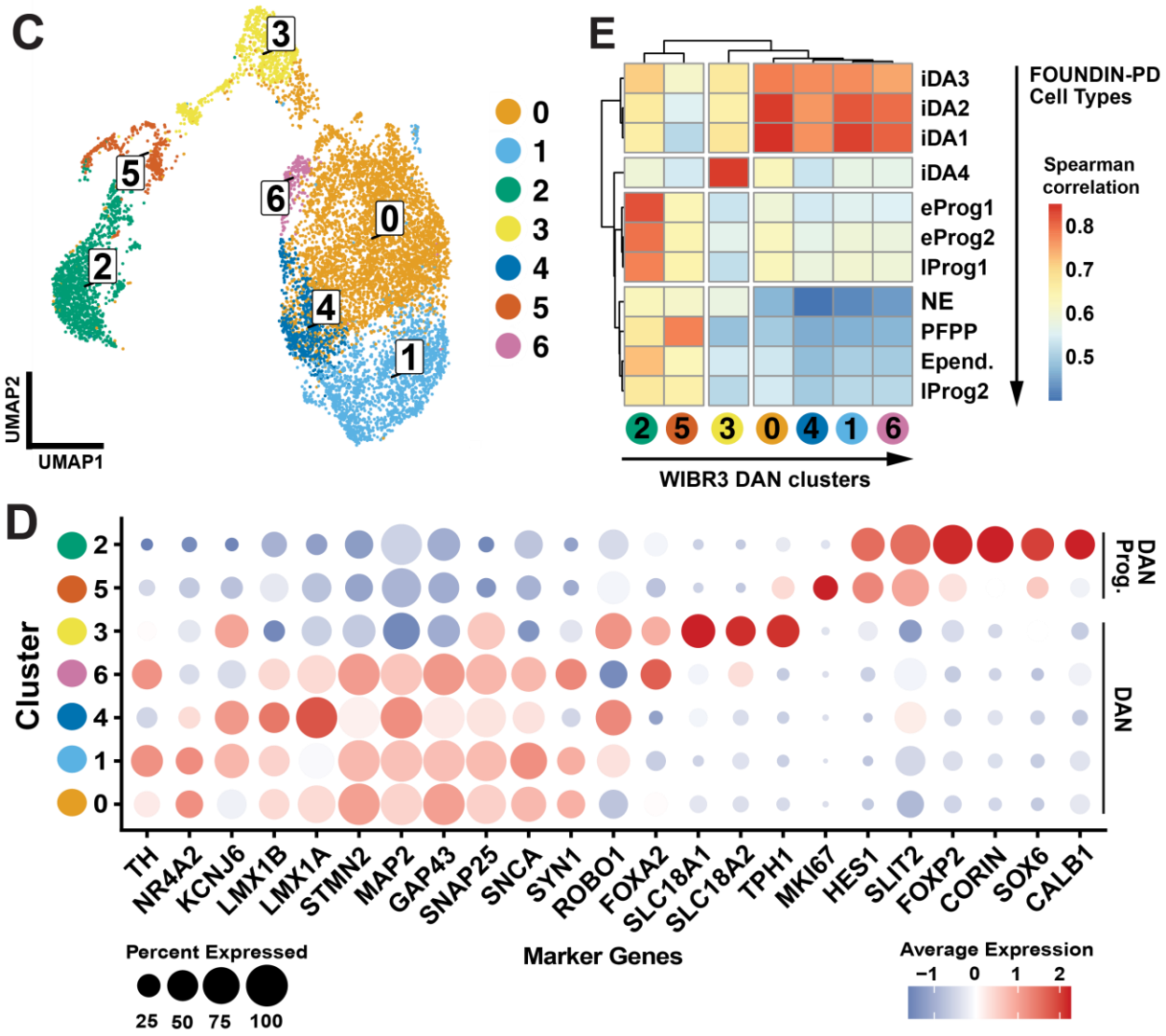


Figure 2.2C, 2.2D, and 2.2E - (C) *UMAP plot of scRNA-seq analysis at day 35–37 of dopaminergic neuron differentiation from WIBR3-S1, WIBR3-S2, and WIBR3-S3 hESCs, showing 10,097 cells separated into 7 coarse clusters.* **(D)** *Dot plot showing expression of key midbrain progenitor and dopamine neuron marker genes across cluster identities, indicating that clusters identified in (C) represent midbrain progenitors and dopaminergic neurons at different developmental stages. Dot size indicates the proportion of cells in a cluster expressing a given gene; color intensity indicates average expression.* **(E)** *Heatmap depicting Spearman correlation coefficients between pseudo-bulk expression profiles of the 7 WIBR3 clusters in (C) and pseudo-bulk expression profiles of 11 FOUNDIN-PD cell types.*

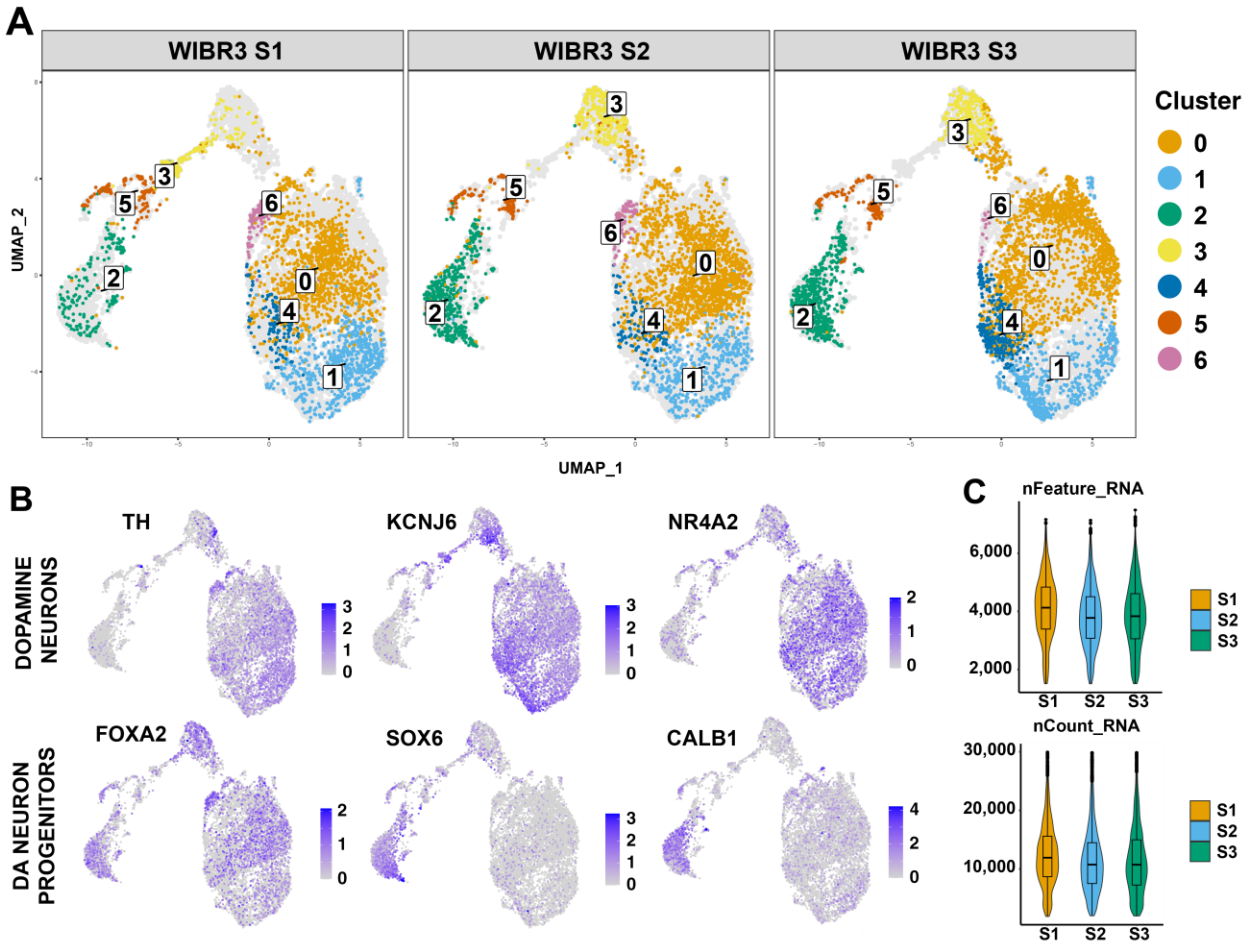


Figure S2A, S2B and S2C - (A) *UMAP representation of WIBR3 dopamine neurons, split to show distribution of cells from three subclones.* **(B)** *UMAP feature plots showing expression patterns of midbrain progenitor and mature dopamine neuron markers across cell clusters.* **(C)** *Quality control plots of nCount RNA reads (UMIs) and feature RNA reads (genes) detected among all cells for each subclone.*

The reproducibility of this differentiation across the three independent WIBR3 subclones is supported by the per-subclone composition of the integrated dataset (Figure S2A), in which all seven clusters are populated by cells from each of WIBR3-S1, WIBR3-S2, and WIBR3-S3. Canonical midbrain DA-neuron and floor-plate-progenitor markers show the expected cluster-specific expression patterns: TH, KCNJ6, and NR4A2 mark the mature DA-neuron clusters, while FOXA2, SOX6, and CALB1 distinguish the floor-plate-progenitor populations (Figure S2B). Per-cell quality metrics (nFeature_RNA and nCount_RNA) are comparable across subclones, indicating consistent capture and library complexity (Figure S2C). To formalize the FOUNDIN-PD cross-comparison at single-cell resolution, we applied SingleR-based label transfer using FOUNDIN-PD as the reference. The SingleR score heatmap (Figure S3A) confirms that the iSCORE-PD DA-neuron clusters score most strongly against the FOUNDIN-PD iDA1–iDA4 populations, while the progenitor clusters score against the FOUNDIN-PD progenitor populations. Projection of the SingleR-assigned FOUNDIN-PD labels onto the iSCORE-PD UMAP coordinates (Figure S3B) recovers the same

cluster identities established by manual annotation, and the proportional distribution of SingleR-assigned cell types is consistent across subclones (Figure S3C). Together, these analyses support the conclusion that WIBR3-derived DA neurons recapitulate the canonical midbrain dopaminergic differentiation trajectory and are quantitatively comparable to the FOUNDIN-PD reference at the cell-type-assignment level.

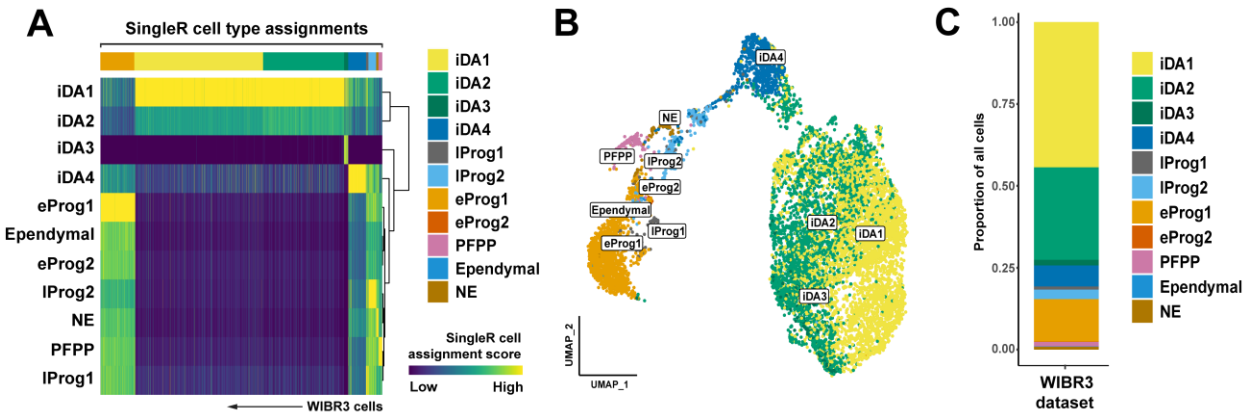


Figure S3A, S3B and S3C– (A) *Heatmap depicting the cell type classification of 10,097 WIBR3 cells (rows) into a list of cognate cell-type labels as defined in the FOUNDIN-PD reference dataset. Plotted values represent normalized cell-type assignment scores as calculated by the SingleR package in R (Wilcoxon rank-sum test). Any cells with no assignment after SingleR classification ("NA") are omitted from downstream analyses.* **(B)** *UMAP plot of all 10,097 WIBR3 cells profiled, with labels representing a cell's assignment to a corresponding cell-type as defined in the FOUNDIN-PD reference dataset. Labels for each cell are identified using the SingleR classifier in R (Wilcoxon rank sum test).* **(C)** *Stacked barplot depicting the proportion of FOUNDIN-PD cell-type labels applied to the 10,097 cell WIBR3 dataset.*

Recent evidence suggests that the impact of PD-associated mutations extends beyond neurons, with effects on cell types such as microglia playing a critical role in the pathogenesis of PD[28,29]. In particular, chronic microglial activation has been proposed as a key pathophysiological feature of many neurodegenerative disorders including PD[30]. Following a previously described protocol[31], we differentiated the subclones of WIBR3 into microglia-like cells (iMGs). In this protocol, hPSCs are initially induced to myeloid intermediates and subsequently differentiated to microglia through the addition of cytokines normally secreted from neurons and astrocytes, including IL-34, M-CSF, and TGF-β1 (Figure 2.2F). All WIBR3 subclones robustly generated microglial precursors (iMPs) at high purity, evidenced by the presence of 92.5% CD11b/CD45 and 73.2% CD14/CD16 double-positive cells (Figure 2.2G–J). Moreover, terminal differentiation yielded cells expressing key markers for mature microglia such as IBA1, CX3CR1, and P2RY12, with FACS analysis demonstrating 97% CX3CR1/P2RY12 double-positive cells and confirmed through immunostaining (Figure 2.2K–N). Collectively, these findings underscore the suitability of WIBR3 hESCs as a model to study the contribution of different cell types to PD pathology.

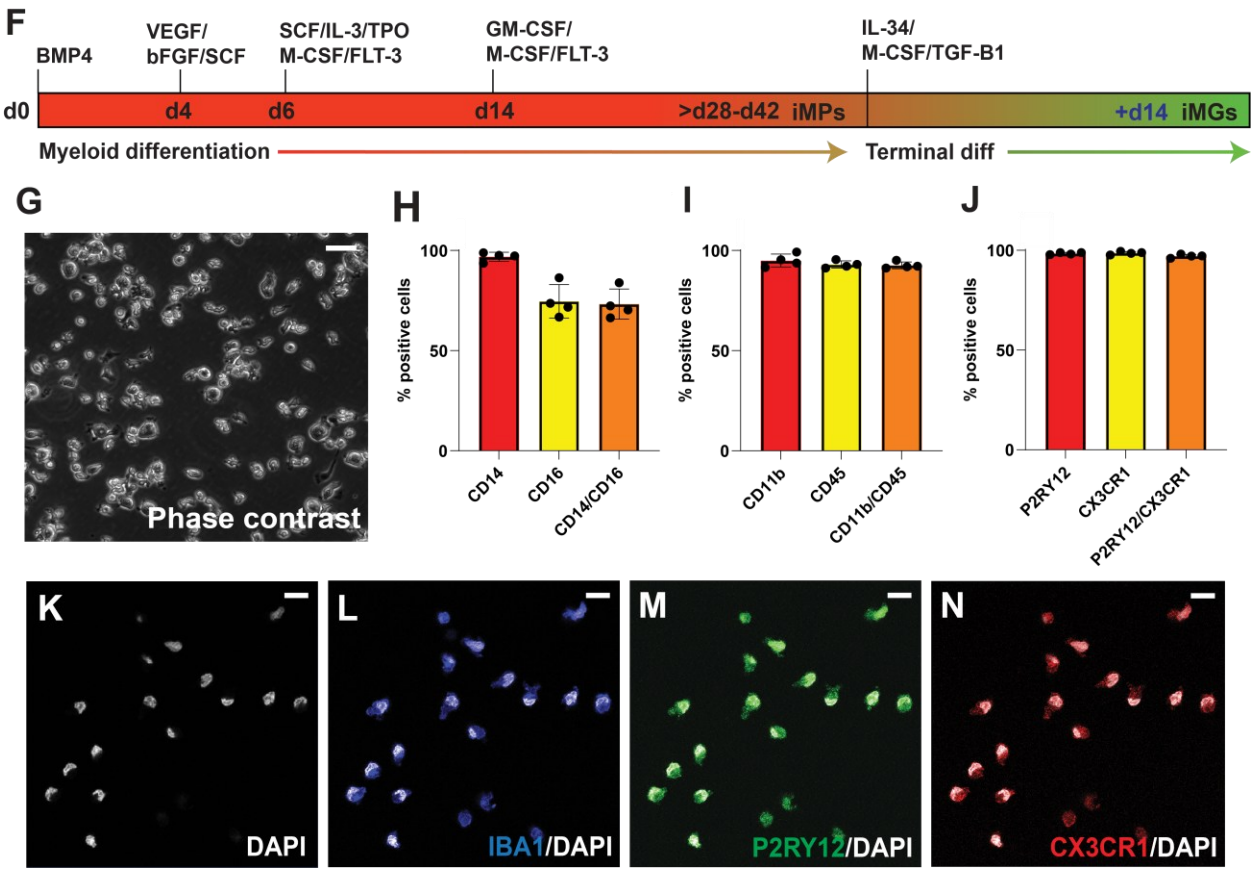


Figure 2.2F through 2.2N - ***(F)*** *Schematic depicting in vitro microglia differentiation protocol from WIBR3 hESCs.* ***(G–J)*** *Representative flow cytometry (FACS) analysis (G,I) and quantification (H,J) of CD14/CD16 and CD11b/CD45 expression in hESC-derived iMPs from subclones WIBR3-S1, S2, S3.* ***(K–N)*** *Representative phase contrast (K, scale bar 50 µm) and immunostaining (L–N, scale bar 10 µm) images of in vitro differentiated microglia derived from subclone WIBR3-S1 for microglia-specific markers IBA1 (blue), P2RY12 (green), and CX3CR1 (red) (terminal differentiation day 14).*

2.3.3 - Genetic engineering of PD-associated mutations

The goal of this work was to establish a comprehensive collection of cell lines carrying mutations associated with PD within a well-defined genetic background. This requires precise, robust gene-editing approaches to insert the desired genetic alterations in hPSCs, followed by rigorous quality control. We previously demonstrated that WIBR3 cells can be efficiently genetically modified using CRISPR/Cas9, TALEN, or prime editing-based genome engineering approaches[4,11,14]. CRISPR/Cas9-based editing is effective for introducing targeted genomic deletions and biallelic alterations, while prime editing is highly efficient in introducing heterozygous modifications required for modeling dominantly inherited disease-associated alleles[4]. To create cell lines carrying PD-associated genetic alterations in WIBR3 hESCs, we employed two different editing pipelines. Pipeline A (Figure 2.3A) uses FACS-enrichment post-nucleofection to purify effectively transfected cells, followed by clonal expansion and genotyping to allow the isolation of clonal, correctly targeted lines (estimated time 30–35 days). Pipeline B (Figure 2.3B) uses nucleofection, limited dilution, and next-generation sequencing (NGS)-based genotyping to identify

desirable edits in a 96-well plate system. This integrated workflow allows the efficient isolation and purification of correctly edited clonal cell lines, even at low frequency, within a shorter timeframe compared to previous approaches (21–35 days).

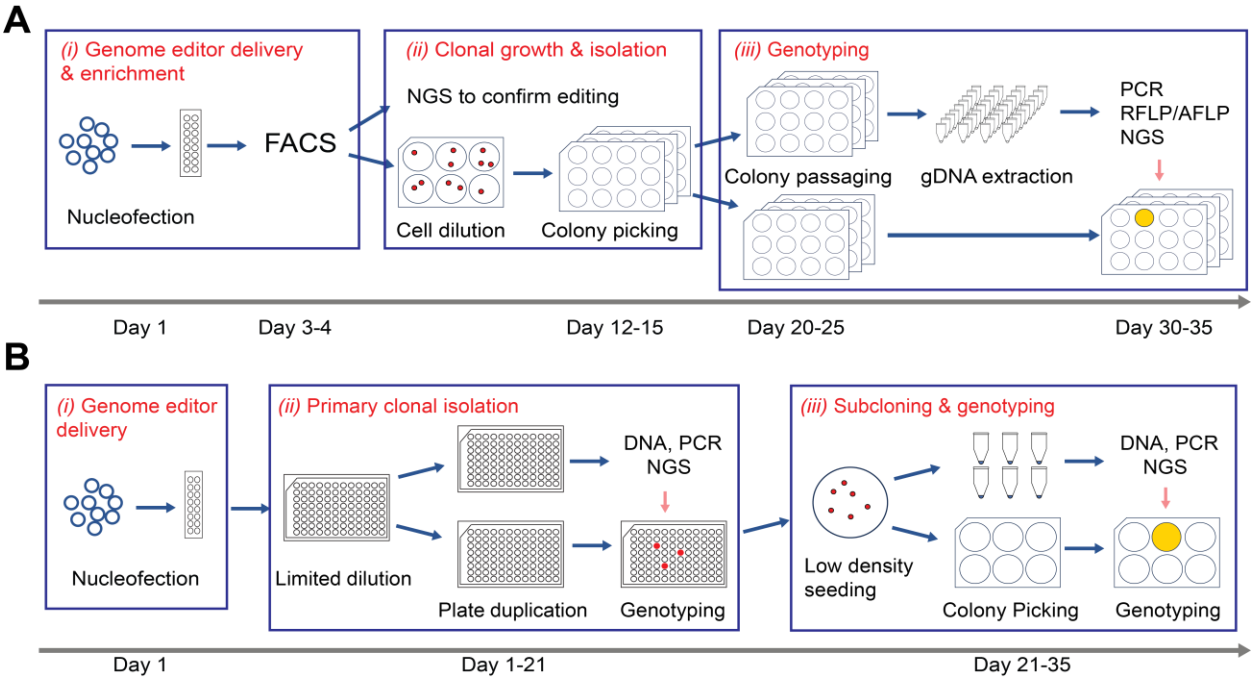


Figure 2.3A and 2.3B – (A) *Schematic illustrating genome editing Pipeline A, involving (i) FACS-based enrichment of nucleofected cells containing gene editing reagents including a fluorescent reporter, (ii) isolation of clonally expanded cell lines, and (iii) NGS-based genotyping to identify correctly edited cell lines.* **(B)** *Schematic illustrating genome editing Pipeline B, a high-throughput cell isolation system comprising (i) nucleofection of gene editing reagents, (ii) plating of cells in limited dilution (~10 cells/well) to isolate wells containing correctly targeted cells by NGS, and (iii) subcloning, expansion, and NGS-based genotyping.*

Given the identification of several disease-causing alleles within most PD-linked genes, the selection and prioritization of specific alleles for gene editing within each gene was based on the confidence of pathogenicity for each mutation, the allele prevalence, and the feasibility of the editing strategy. As described in detail below for each gene, we employed five general editing strategies to recreate the genomic alterations identified as causal or high-risk factors for PD (Figure 2.3C): (i) CRISPR/Cas9-facilitated homology-directed repair (HDR) using ssODNs as templates; (ii) competing HDR templates carrying synonymous mutations in the gRNA-target site to favor heterozygous over homozygous mutations; (iii) TALEN-facilitated HDR; (iv) prime editing for insertion of point mutations; and (v) dual-guide RNA CRISPR/Cas9 deletions to create large genomic deletions.

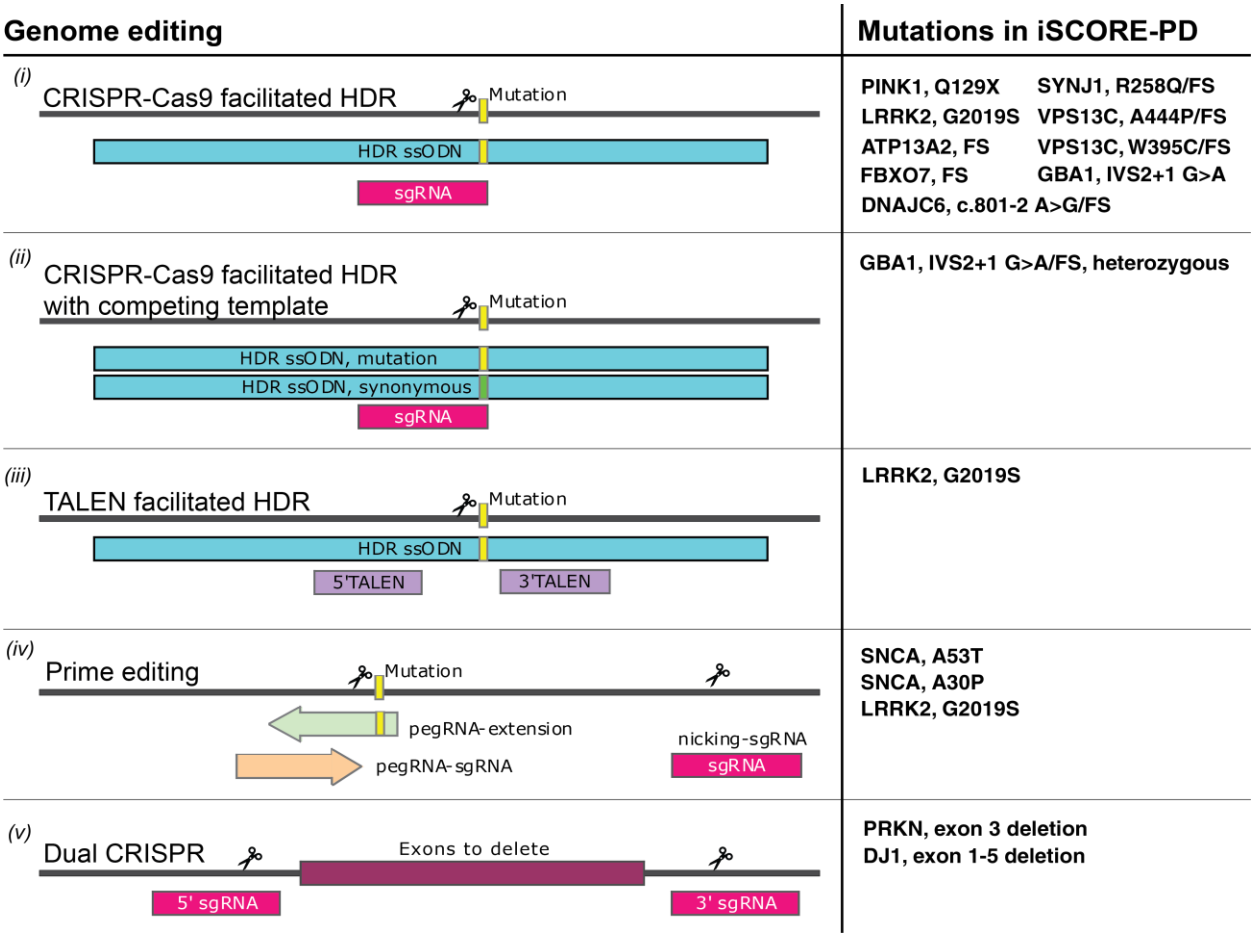


Figure 2.3C - *Table summarizing the gene-editing strategies used to generate the iSCORE-PD collection:* **(i)** *CRISPR/Cas9-facilitated HDR using ssODN repair templates;* **(ii)** *CRISPR/Cas9 HDR with competing synonymous-mutation ssODN templates to favor heterozygous over homozygous mutations;* **(iii)** *TALEN-facilitated HDR;* **(iv)** *prime editing for PD-associated point mutations;* **(v)** *dual CRISPR approach using 3' and 5' sgRNAs flanking the desired deletion.*

2.3.4 – Quality control of genetically modified hPSCs

One significant challenge associated with any genome-editing approach is the risk of introducing unintended on- and off-target genetic modifications in the edited cell lines. Additionally, it is well-established that clonal expansion and in vitro culture of hPSCs can lead to the acquisition of genetic alterations that provide growth advantages[24,25,32–36]. Consequently, there is a consensus in the field that all gene-edited hPSC-derived disease models should undergo a rigorous quality-control process to validate the pluripotent properties of the edited cell lines and ensure the absence of major gene editing- or culture-induced genetic alterations. As part of this collection, all genome-edited cell lines underwent a comprehensive quality-control process (Figure 2.3D). Following CRISPR/Cas9 or prime-editing-based genome editing and subsequent clonal expansion, individual correctly targeted clones were initially identified using either Sanger sequencing or NGS. Subsequently, all correctly targeted clonal lines were expanded, cryopreserved at a low passage number post-genome editing, and tested for the expression of pluripotency

markers OCT4, SSEA4, and alkaline phosphatase by immunocytochemistry. To confirm a normal karyotype and assess overall genomic integrity, genome-edited clonal hESC lines underwent standard aCGH genotyping and were analyzed using the modified high-density Illumina Infinium Global Diversity Array / Neuro Booster Array. This analysis aimed to exclude cell lines with large genome-editing-induced structural alterations or complete chromosomal loss compared to the genome of the parental WIBR3 cell line.

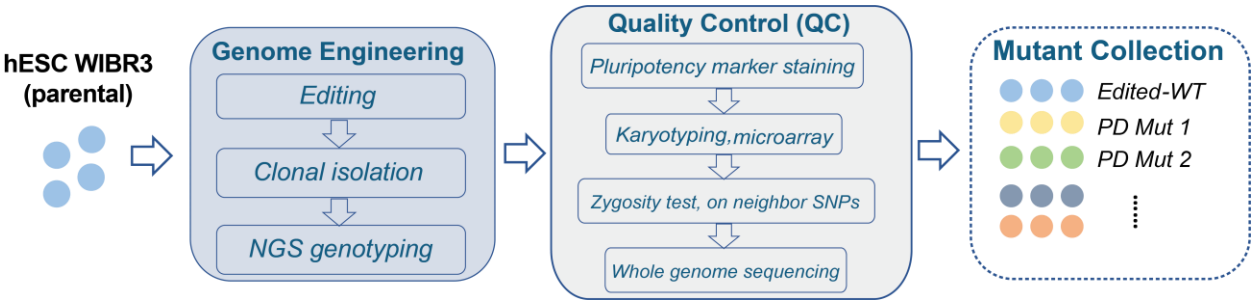


Figure 2.3D - (D) *Overview of genome engineering and quality-control steps in the full iSCORE-PD collection.*

A frequently overlooked challenge associated with genotyping approaches based on PCR amplification of the target locus is the common failure to detect loss of heterozygosity (LOH)[37]. LOH can result from large deletions or the loss of entire chromosome fragments distal to the targeting site, induced by the genome-editing process. To rule out LOH in cell lines that appeared to be homozygously edited based on the detection of a single allele by NGS, we introduced an additional quality-control step using either a Southern blot or SNP-PCR-based analysis to validate the presence of two alleles at the targeting site. Using this analysis, we identified LOH in 2 out of 18 (11.11%) tested clonal cell lines that were initially classified as correctly edited with two identical alleles at the target site. These data underscore that LOH is a significant complication arising from genome editing, emphasizing the importance of incorporating LOH testing as a critical component of the quality-control process for genome-edited hPSC lines.

A summarized list of the genes, mutations, and number of cell lines in the iSCORE-PD collection can be found in Table 2.1. For detailed information on each cell line in the iSCORE-PD collection, see Supplemental Table 2.3 and the per-gene Appendix sections (§2.A.1–§2.A.12), which describe the pipeline used, the gene and mutation targeted, the genome-engineering technique, the genetic description, the expected protein products, and the quality-control results. Overall, 19.75% (16 out of 81) of isolated clonal cell lines with correct NGS-confirmed genotype were excluded from our collection. The reasons include chromosomal and structural alterations (16.05%, 13 out of 81), lack of expression of a pluripotency marker (1.47%, 1 out of 68), and LOH (11.11%, 2 out of 18 LOH-tested lines). Importantly, the frequency of chromosomal and large structural abnormalities was higher in clones generated by double-strand-break-based genome editing (CRISPR and TALEN: 20.34%, 12 out of 59) compared to prime editing (5.56%, 1 out of 18). Similarly, deletions resulting in LOH at the targeted locus were observed only in double-strand-break-based genome-edited cell lines and were absent in prime-edited cell lines. Given that any genetic

alteration induced by cell culture or genome editing can impact the biological properties of hPSCs and disease phenotypes, these data emphasize the necessity for comprehensive quality control in genome engineering. Considering that clonally derived cell lines are susceptible to cell culture-induced genetic drift and acquire additional genetic and epigenetic alterations over time, we advocate for the use of multiple independently gene-edited clonal lines for each genotype to account for this variability. In addition, we recommend performing routine quality-control analysis to detect long-term-culture-induced genetic aberrations. This approach ensures a robust assessment of disease-relevant phenotypes in vitro, acknowledging the potential variability that may arise during prolonged cell culture and genome-editing processes.

2.3.5 – Composition of the iSCORE-PD cell line collection

For our initial iSCORE-PD collection, we prioritized engineering cell lines carrying highly penetrant mutations in monogenic familial PD genes[1]. The specific modifications for each gene were selected based on information in the MDSGene database[38] (https://www.mdsgene.org) and the currently available literature. The collection comprises 65 clonal cell lines carrying pathogenic variants in 11 monogenic PD-linked genes generated using the editing pipelines and quality-control framework described in §2.3.3 and §2.3.4. Specific allelic modifications include *SNCA* A53T and A30P, *PRKN* Ex3del, *PINK1* Q129X, DJ1/*PARK7* Ex1–5del, *LRRK2* G2019S, *ATP13A2* frameshift, *FBXO7* R498X/frameshift, *DNAJC6* c.801-2 A>G/frameshift, *SYNJ1* R258Q/frameshift, *VPS13C* A444P and W395C (each with accompanying frameshift alleles), and *GBA1* IVS2+1 splice-site mutation. Per-gene editing strategies, genotyping, and zygosity-validation results for each line are reproduced in the chapter Appendix (§2.A.1 through §2.A.12).

Wild-type controls are provided alongside the mutant collection. These include the parental WIBR3 line, three independent single-cell-derived subclones (WIBR3-S1, S2, S3), and eight edited-wild-type (EWT) clones recovered from the genome-editing pipeline without on-target modification. Of the EWT lines, six were derived from prime-editing experiments (EWT_S1-3 from Pipeline B, EWT_S6-8 from Pipeline A) and two from CRISPR/Cas9 experiments (EWT_S4-5 from Pipeline B). EWT lines provide the closest experimental controls for any given mutant comparison, as they capture the editing-experiment-specific founder background acquired during clonal expansion (further analyzed in §2.3.6). All cell lines in the collection have been deposited at the WiCell Research Institute (https://www.wicell.org/) and are summarized in Table 2.1.

	MUTATION	INHERITANCE	FREQUENCY[2]	PROPOSED DISEASE MECHANISM[2]	EDITING	HETERO-ZYGOUS CLONES	HOMO-ZYGOUS CLONES
(PARK1) SNCA	A53T	Autosomal dominant	Very rare	Gain of function or overexpression	Prime editing	**3**	**-**
	A30P				Prime editing	**3**	**1**
(PARK2) PRKN	EX3DEL	Autosomal recessive	Rare	Loss of function	CRISPR/Cas9 (dual guide)	**-**	**3**
(PARK6) PINK1	Q129X	Autosomal recessive	Rare	Loss of function	CRISPR/Cas9 (HDR)	**-**	**3**
(PARK7) DJ1	EX1-5DEL	Autosomal recessive	Very rare	Loss of function	CRISPR/Cas9 (dual guide)	**5**	**7**
(PARK8) LRRK2	G2019S	Autosomal dominant	Common	Gain of function	Prime editing, TALEN, CRISPR/Cas9 (HDR)	**4**	**1**
(PARK9) ATP13A2	FS	Autosomal recessive	Very rare	Loss of function	CRISPR/Cas9 (HDR)	**-**	**4**
(PARK15) FBXO7	R498X/FS	Autosomal recessive	Very rare	Loss of function	CRISPR/Cas9 (HDR)	**-**	**2**
(PARK19) DNAJC6	c.801-2 A>G/FS	Autosomal recessive	Very rare	Loss of function	CRISPR/Cas9 (HDR)	**-**	**2**
(PARK20) SYNJ1	R258Q/FS	Autosomal recessive	Very rare	Loss of function	CRISPR/Cas9 (HDR)	**-**	**2**
(PARK23) VPS13C	W395C/FS	Autosomal recessive	Rare	Loss of function	CRISPR/Cas9 (HDR)	**-**	**4**
	A444P/FS	Autosomal recessive			CRISPR/Cas9 (HDR)	**1**	**2**
GBA1	IVS2+1/FS	Autosomal dominant[1]	Common	Likely loss of function	CRISPR/Cas9 (HDR)	**4**	**2**
Wild type	-	-	-	-	-	Parental + 3 independent clones	
Edited wild type (EWT)	-	-	-	-	Prime editing, CRISPR/Cas9 (HDR)	Prime editing (6) + CRISPR/Cas9 (2)	

Table 2.1. iSCORE-PD cell line collection (overview). *[1] Mixed inheritance; [2] Adapted from Blauwendraat et. al., 2022.*

2.3.6 – WGS reveals minimal editing-induced genetic variability

An important question for any genome-edited isogenic resource is whether engineered cell lines truly differ from their parental line only at the targeted locus, or whether genome-editing reagents and routine cell culture introduce additional variation that could confound downstream phenotyping. To address this, whole-genome sequencing (WGS) was performed on 61 of the 65 cell lines in the iSCORE-PD collection at approximately 30x coverage, and clone-specific single-nucleotide variants and indels (SNVs/indels) absent from the parental WIBR3 reference were identified using a custom analysis pipeline, detailed in §2.5.15–§2.5.16[12].

Across the 61 sequenced lines, the average per-clone variant burden was 216.1 ± 55.5 (mean ± SD) SNVs/indels above the WIBR3 baseline. Of these, only 1.3 ± 1.3 were localized to protein-coding exons (excluding synonymous variants), and the pipeline recovered every engineered mutation in the collection except one. The number of clone-specific variants per line scaled with passage number (R^2 = 0.382), corresponding to an estimated rate of 17.3 mutations per passage acquired during standard hPSC culture (95% CI: 11.52 to 23.08), consistent with previously reported rates for hPSCs[33]. Critically, this rate was indistinguishable between gene-edited cell lines and the unedited control lines (WIBR3-S1-3 and EWT_S1-8), indicating that the genome-editing process itself contributes minimally to total variant burden.

When SNVs/indels are shared across multiple clones from the same editing experiment, the variants could in principle reflect either editing-induced common off-target effects or preexisting variants in the founder cell population that became fixed during clonal expansion. The data favor the latter explanation: shared variants are observed not only between correctly-targeted clones but also between targeted clones and the untargeted EWT controls from the same editing experiment, which by definition were never exposed to a successful editing event. Per-editing-group founder-variant breakdown is provided in §2.A. The practical implication is that the inclusion of multiple independently-targeted clones per genotype, alongside matched EWT controls from the same editing experiment, provides the strongest control for residual genetic variability when the collection is used in downstream phenotyping.

2.3.7 – Off-target effects are rare in CRISPR/Cas9-edited and absent in prime-edited clones

Off-target editing is a residual concern for any genome-engineering approach. The WGS dataset described in §2.3.6 enables a systematic test of off-target frequency in the iSCORE-PD collection. For each gRNA and pegRNA used in the collection, candidate off-target sites were identified using Cas-OFFinder[39] under a deliberately permissive threshold (up to five mismatches), and SNVs/indels within a 100-bp window of any predicted site were enumerated across all sequenced clones (analysis detailed in §2.5.17). Despite the low stringency of the prediction threshold, only five candidate variants were observed across 54 assessed cell lines, of which four were classified as genuine off-target events based on their

location 2–5 bases upstream of the NGG protospacer-adjacent motif consistent with a Cas9-mediated cleavage signature. Stratified by editing modality, the four genuine off-targets all occurred in CRISPR/Cas9-edited lines (4 of 41 lines, ~10%); zero off-target events were detected across the 18 prime-edited lines. All off-target events were heterozygous, intronic, and non-coding.

A phylogenetic analysis of the gene-edited cell lines, computed from each clone's unique SNVs/indels, places the eight EWT control lines distributed across the genetic space spanned by the targeted lines (Figure 2.5). This distribution validates the use of EWT clones from a given editing experiment as the closest available genetic controls for that experiment's targeted lines.

Together, the variability and off-target findings reinforce the structural-aberration and loss-of-heterozygosity observations in §2.3.4: across all three failure modes (large structural alterations, LOH at the targeted locus, and predicted-site off-target events), prime editing exhibits substantially lower rates than CRISPR/Cas9 in this collection, consistent with the mechanistic difference between a double-strand-break and single-strand-nick repair intermediate[40]. These findings support prime editing as the preferred modality for generating isogenic disease-modeling collections of hPSCs, as employed for several of the iSCORE-PD lines and adopted for the Cre-inducible CRISPRi clone described in Chapter 3.

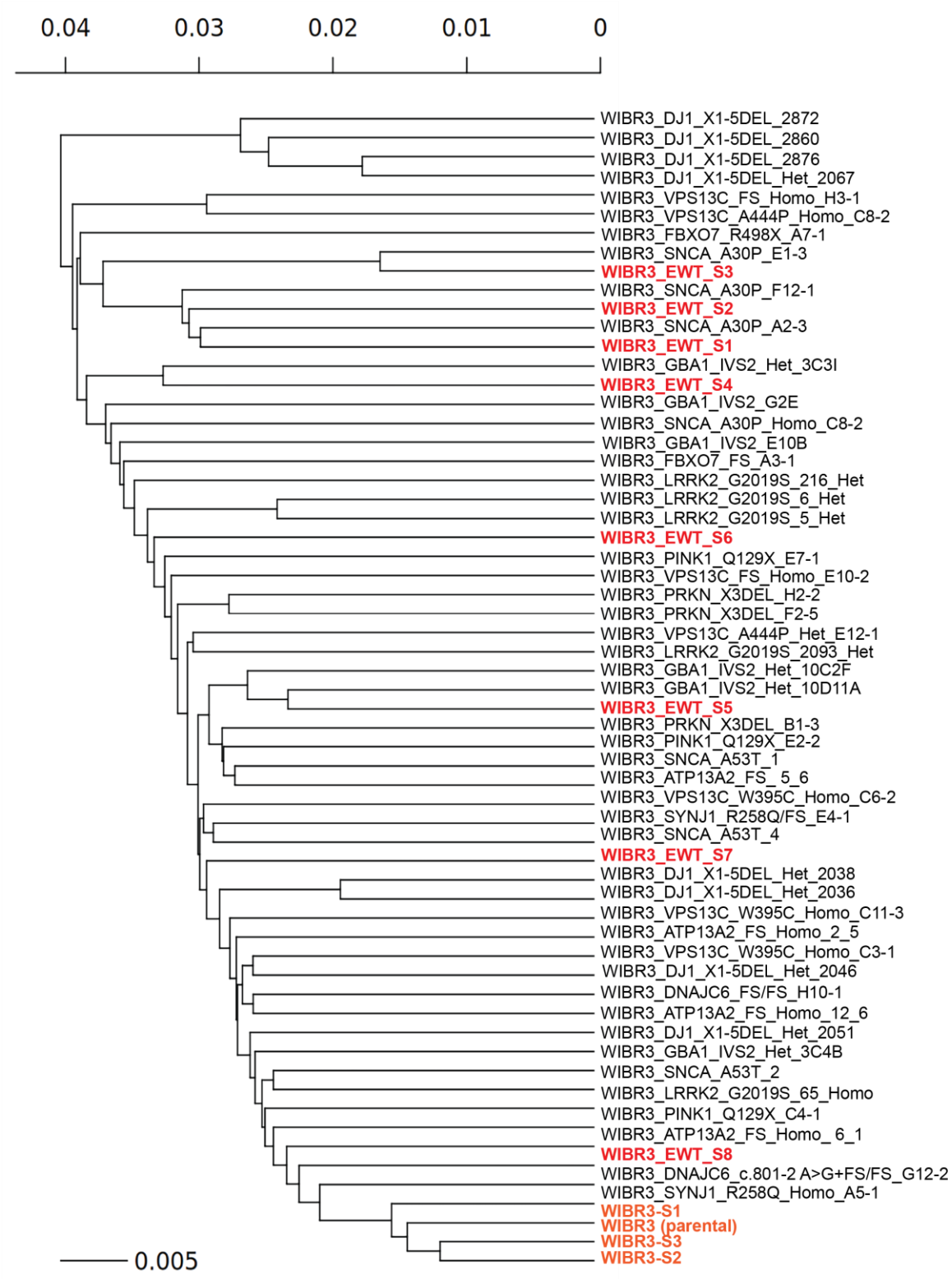


Figure 2.5 - *Phylogenetic relationship of all analyzed cell lines in iSCORE-PD. Cell lines displayed in orange represent WIBR3 (parental) + WIBR3-S1,2,3 subclones and light red represents untargeted EWT_S1-8 cell lines.*

2.4 – Discussion

Population genetics and advances in sequencing technologies have substantially contributed to our understanding of the genetic architecture of complex diseases and have identified numerous genetic variants linked to the development and progression of diseases such as PD. While each of these variants could provide valuable insights into the pathophysiology of the underlying disease, unraveling their function within a genetically heterogeneous population remains challenging due to additional effects from undefined genetic-modifier loci. To overcome this limitation, we present an experimentally tractable human cellular model system that facilitates molecular and cellular studies in a genetically controlled platform. For this, we have generated a collection of isogenic hESC lines that carry monogenic or high-risk PD-associated mutations. Similar to the development of inbred animal models, which have proven instrumental in establishing robust genotype-phenotype correlations and enabled the comparison of such phenotypes across research groups, our isogenic cell-line collection offers the opportunity to directly compare the phenotypic effect of PD-associated mutations in a genetically controlled system across genes and laboratories.

The establishment of an isogenic cell-line collection involves two important components, both crucial for the effective implementation of this approach: (i) thorough characterization of the parental hPSC line, and (ii) the establishment of a genome-editing platform enabling the efficient engineering of genetic alterations similar to those found in patients. Regarding the hPSC line, we have conducted a comprehensive analysis of the parental WIBR3 hESC line and demonstrated that the genomic integrity of this cell line can be sustained over extended periods in culture. We also show that WIBR3 cells are amenable to multiple rounds of clonal expansion and genome engineering. In addition, high-density genotyping and long-read WGS show that the WIBR3 cell line does not carry major structural or genetic alterations impacting genes with known relevance to PD. We further demonstrate that WIBR3 cells can efficiently generate PD-relevant cell types in vitro using established differentiation protocols. WIBR3 cells thus serve as a highly characterized female reference hESC line that complements current hPSC lines for modeling neurodegenerative diseases.

To generate this collection, we established CRISPR/Cas9, TALEN, and prime-editing pipelines in hPSCs, enabling the highly efficient and multiplexed introduction of a broad range of genetic alterations known to be associated with a given disease, ranging from heterozygous and homozygous single nucleotide variants to large structural genomic deletions. During the process of establishing this collection, we made several key observations. Notably, we recognized that both CRISPR/Cas9 and prime-editing approaches necessitate a comprehensive quality-control (QC) process beyond the validation of the intended genomic modification. Consequently, all the described hESC lines underwent a rigorous quality-control procedure, which included the validation of pluripotency using immunohistochemistry, the exclusion of karyotypic and structural aberrations using standard aCGH arrays and high-density genotyping arrays, and zygosity analysis at the targeted genomic locus. The need for this quality control is evident from the

high number of genome-edited cell lines that were excluded due to structural aberrations and loss of heterozygosity (Supplemental Table 4).

Whole-genome sequencing of 61 of the 65 cell lines in the collection further revealed that genetic variation between any two clones is small relative to the inter-individual variation that confounds patient-derived iPSC comparisons, and most of this variation is non-coding. Critically, the editing process itself contributes minimally to total variant burden: the rate at which clones acquire variants per passage is indistinguishable between gene-edited lines and the unedited control lines. Most shared variants between clones from the same editing experiment are observed not only between targeted clones but also between targeted clones and the untargeted EWT controls, indicating that they were already present in the parental population at the time of nucleofection rather than introduced by the editing reagents. The practical consequence is that multi-step cloning strategies — in which an already-edited line is retargeted to introduce a second modification — propagate the founder background of the first cloning step into all downstream lines, and require careful control selection to interpret phenotypes from such lineages.

This finding has implications for one common validation strategy: reverting the engineered mutation back to wild-type in a second editing step, with the corrected line used as a confirmatory control. Mutation reversion is highly useful for validating mutation-specific phenotypes when the phenotype of interest is robust and well-established. However, when subtle or novel phenotypes are being characterized, the corrected line carries its own additional ~17 mutations per passage acquired during the second clonal expansion, and a small number of these will fall in coding regions. Reverting a single clone is therefore not sufficient as a stand-alone validation strategy for novel phenotypes; multiple independently-targeted clones — with matched EWT controls from the same editing experiment — provide stronger phenotypic discrimination.

A key finding across the derivation of all cell lines was that the frequency of karyotypic and structural aberrations, as well as LOH of the edited locus, was substantially higher in CRISPR/Cas9 than in prime-edited cell lines. In fact, LOH was seen exclusively in CRISPR/Cas9-edited clonal hESCs. While this is the first time that this has been formally reported across a large cohort of gene edits, the finding might not be unexpected. In contrast to CRISPR/Cas9-based genome editing, prime editing does not require a double-strand break (DSB) at the target site to insert genetic modifications[40]. Instead, prime editing directly repairs a prime-editor-induced single-strand DNA break[40]. These results confirm our previous observations that prime editing has substantial advantages over CRISPR/Cas9-based approaches for introducing point mutations and small structural modifications in hPSCs[4]. Furthermore, we strongly recommend including a zygosity analysis, specifically to exclude LOH at the target locus, as a critical step in the quality-control pipeline of genome-engineered hPSCs.

A further consideration specific to female hPSC lines is the phenomenon of X-chromosome erosion, in which the inactive X chromosome progressively loses repressive

epigenetic marks during prolonged in vitro culture, leading to aberrant biallelic expression of X-linked genes. WIBR3 is a female line that retains a normal XCI pattern at the early passage from which the iSCORE-PD collection was derived, but as with any female hPSC line, the integrity of X-inactivation should be monitored over time. Continued monitoring of XCI status is particularly important when comparing phenotypes across editing groups assayed at different culture passages, as drift in X-linked gene dosage could plausibly contribute to apparent phenotypic differences. Among the practical strategies that mitigate this risk are the use of low-passage stocks for downstream phenotyping, the inclusion of multiple independent clones per genotype, and routine validation of XCI markers as part of the QC pipeline.

While the cell lines described here currently focus on coding risk variants with large effect size linked to monogenic PD[1], we envision that our laboratory and other researchers can expand this collection to eventually incorporate GWAS-identified risk variants with lower effect size. Such an expansion could provide functional insights into how these primarily non-coding sequence variants affect similar cellular and molecular pathways as those implicated in monogenic PD. To facilitate such efforts, all generated cell lines are made available with the support of the Aligning Science Across Parkinson's (ASAP) initiative and the Michael J. Fox Foundation through the WiCell Research Institute. In parallel work, the iSCORE-PD collection has already been used to characterize alternative pre-mRNA splicing patterns across the eleven monogenic PD mutations in midbrain dopaminergic neurons, identifying mutation-associated splicing changes that overlap with patterns observed in postmortem PD brain tissue[41]. The collection has also enabled work mapping the relationship between LRRK2 G2019S, TDP-43 dysfunction, and aberrant splicing in PD-relevant cell types[42]. Subsequent biological analysis of this collection and its future expanded forms, involving research groups with diverse expertise, can contribute to a unified understanding of how genetic risk variants functionally contribute to the pathogenesis of PD.

Limitations of this study

Each genome is inherently unique, carrying a distinct combination of sequence variants and genetic alterations that can influence the development and pathology of complex diseases such as PD. Consequently, no single cellular model can fully recapitulate all the molecular and cellular features of such disorders. Given this limitation, it will become necessary to expand the described approach to include additional cell lines with diverse genetic backgrounds to fully dissect the pathobiology of PD. Moreover, it is widely recognized that genetic engineering and in vitro culture of cell lines lead to the acquisition of genetic and epigenetic modifications, potentially affecting the phenotypical analysis of in vitro-derived cell types, irrespective of the disease genotype. While using multiple independently edited clonal cell lines alongside continuous quality-control measures can mitigate many such random alterations, we cannot entirely exclude remaining systematic genetic and epigenetic off-target modifications associated with specific gene-editing approaches.

2.5 Chapter summary and transition to Chapter 3

The work presented in this chapter delivered four related outputs. First, a comprehensively-characterized female hESC parental line, WIBR3, validated at the level of long-read whole-genome sequence, structural-variant content, polygenic risk distribution, and p53-dependent DNA damage response, and confirmed to differentiate efficiently into midbrain dopaminergic neurons and microglia-like cells. Second, a collection of 65 clonal, quality-controlled, isogenic hESC lines carrying pathogenic mutations in eleven monogenic PD genes — a collection now banked and distributed through the WiCell Research Institute and openly available to the PD research community. Third, a systematic comparison of genome-editing approaches at scale, demonstrating that prime editing introduces substantially fewer on-target structural aberrations and no detectable loss of heterozygosity relative to CRISPR/Cas9-based approaches, with direct implications for how future isogenic collections should be generated. Fourth, a dopaminergic-neuron single-cell atlas correlating WIBR3-derived neurons to the FOUNDIN-PD cross-laboratory reference dataset, establishing cluster-level transcriptional annotations that are portable across experiments.

Two of these outputs — the dopaminergic-neuron differentiation platform and the single-cell atlas — are the direct experimental substrate for the work presented in Chapter 3. Chapter 3 describes the derivation, validation, and genomic characterization of a Cre-inducible CRISPRi clone (C12) in the same WIBR3 parental background, along with the 10x Genomics Chromium single-cell RNA-seq and CRISPR-capture workflow used to generate the Perturb-seq dataset. The quality-control framework developed for iSCORE-PD carries over almost entirely: the same aCGH genotyping, high-density array analysis, and zygosity-based validation are applied to the CRISPRi parental clone, and — consistent with the observation reported here that routine clonal expansion is a meaningful source of genomic drift — a specific set of clone-specific copy-number variants present in the C12 CRISPRi lineage is identified and discussed in relation to downstream interpretation.

Chapter 3 then reports the Perturb-seq screen itself: a pooled CRISPRi perturbation of forty-one high-confidence familial PD genes (many overlapping with the iSCORE-PD mutation panel described here) and a second, larger pool of over three hundred GWAS-nominated candidate risk genes, assayed by single-cell RNA sequencing in matured dopaminergic-neuron cultures. The cell-type annotations established in the iSCORE-PD atlas — midbrain floor-plate progenitors, early dopaminergic neurons, and mature dopaminergic neurons — are used directly to define the cell populations analyzed by MixScale differential expression testing.

A note on the relationship between the two resources is warranted. iSCORE-PD presents endogenous, physiologically-relevant disease-causing mutations that recapitulate the genetics of familial PD directly. The CRISPRi Perturb-seq screen, by contrast, applies acute, dosage-controlled transcriptional knockdown to a much larger gene panel — trading physiological fidelity for throughput and for the statistical power that comes from interrogating hundreds of genes in parallel in a single experiment. The two

approaches are complementary rather than redundant. Neither alone answers the question of whether the heterogeneous genetic architecture of PD converges on a smaller number of shared cellular programs; together, with appropriate cross-comparison, they begin to. Systematic integration of iSCORE-PD mutation signatures with the Perturb-seq knockdown signatures reported in Chapter 3 is a natural forward direction for the laboratory but lies beyond the scope of this dissertation.

With the isogenic platform established in this chapter, the dissertation now turns to the methodological and scientific core of the work that follows: the generation, validation, and genomic characterization of the Cre-inducible CRISPRi cell lines, and the single-cell experimental and computational pipelines used to acquire and process the Perturb-seq dataset that forms the basis of Chapter 3.

2.6 – Materials and Methods

The methods presented here are abridged from the published iSCORE-PD manuscript. Protocols specific to the work reported in this chapter are retained below with brief technical detail. These include long-read whole-genome sequencing of the WIBR3 parental line, structural-variant and small-variant calling pipelines, array comparative genomic hybridization, SNP-based zygosity/LOH detection, Southern blot validation, CRISPR/Cas9, TALEN and prime-editing pipelines, the two editing/genotyping workflows (Pipeline A and Pipeline B), cortical-spheroid differentiation, MULTI-seq irradiation experiments for p53-pathway assessment, and microglia differentiation. Protocols that overlap substantially with methods used in Chapters 3 include hPSC culture on mouse embryonic fibroblast (MEF) feeders and under feeder-free conditions, midbrain-specified dopaminergic-neuron differentiation using the biphasic WNT signaling protocol of Kim and colleagues[26,27], and the 10x Genomics Chromium scRNA-seq library preparation. These sections are described in full in Chapter 3 and referenced briefly here.

2.6.1 – hPSC culture

WIBR3 hESCs and their derived subclones were maintained on irradiated or mitomycin-C-inactivated MEF monolayers and under feeder-free conditions on Matrigel-coated plates in mTeSR Plus medium (Stem Cell Technologies), as further described in Chapter 3. All cultures were maintained under low-oxygen conditions (95% CO2, 5% O2). The DA-neuron differentiation and scRNA-seq profiling reported in this chapter were performed on feeder-free-adapted cultures starting from mTeSR. Detailed step-by-step protocols for all hPSC culture conditions used in this chapter are deposited at protocols.io (https://doi.org/10.17504/protocols.io.b4pbqvin;
https://doi.org/10.17504/protocols.io.b4msqu6e;
https://doi.org/10.17504/protocols.io.b4mcqu2w).

2.6.2 – Array genotyping

Genomic DNA was isolated from hESC cell pellets using the DNeasy Blood and Tissue Kit (QIAGEN #69504). High-density genotyping was performed using the Neuro Booster Array (NBA) following best-practices guidelines for the Infinium Global Diversity Array (Bandres-Ciga et al., 2023, medRxiv 2023.11.06.23298176). Genotyping data was processed using GenomeStudio; subsequent genotype calls, B-allele frequency, and LogR ratio values were used for genomic-integrity assessment. For gene-edited cell lines present on the NBA, genotype calls were compared to confirm the engineered edit. Genome-wide genotyping calls were compared with PacBio HiFi WGS variants to assess large genomic events across the two data types using PLINK (v1.9)[43]. B-allele frequency and LogR ratio values were plotted using the GWASTools package in R (v3.6.1)[44].

2.6.3 – Long-read sequencing and variant analysis

Ultra-high-molecular-weight DNA was extracted from 5×10^6 parental WIBR3 hESCs using the Nanobind CBB Kit and UHMW DNA Aux Kit (Pacific Biosciences) following the

manufacturer's protocol. Oxford Nanopore libraries were prepared using the SQK-ULK001 Kit and Nanobind Ultra Long Library Preparation Kit, split across three PromethION R9.4.1 flow cells, with library reloading every 24 hours for a total of nine flow cells. PacBio HiFi libraries were prepared from 2×10^6 cells using the Nanobind CBB Kit and SRE size-selection, sheared to 18-20 kb on a Megaruptor 3 (Diagenode), and library-prepared using SMRTbell Prep Kit 3.0 for sequencing on a Revio flow cell with a 24-hour movie time.

Nanopore data were basecalled using Guppy v6.1.2 in super-accuracy mode with methylation tag preservation (dna_r9.4.1_450bps_modbases_5mc_cg_sup_prom.cfg), mapped to hg38 using Minimap2 v2.24[45], and merged across flow cells using SAMtools v1.17[46]. Small variants were called and phased using PEPPER-Margin-DeepVariant v0.8[47], and structural variants called using Sniffles2 v2.2[48]. Haplotype-specific methylation calls were generated using Modbamtools v0.4.8 (Razaghi et al., 2022, bioRxiv 2022.07.07.499188) and modbam2bed. PacBio HiFi data were mapped with Minimap2 (PacBio flags), with small variants called by Clair3 v1.0.4 and structural variants by Sniffles2. Small variants were filtered for DP>15 and GQ>20 using BCFtools[46] and annotated with ANNOVAR v2022-06-08[49]. PD and AD polygenic risk scores (excluding UK Biobank summary statistics) were calculated using PLINK v2.0 and compared to UK Biobank diagnosed participants[50]. Structural variant calls labelled 'PASS' for both ONT and PacBio were merged using Truvari v4.4.0, annotated with ANNOVAR, and further annotated for phenotype relevance (HP:0002180 neurodegeneration; HP:0000707 abnormality of the nervous system) using SVAnna v1.0.4. Variant-type distributions were generated using SURVIVOR v1.0.7.

2.6.4 – Molecular cloning for genome editing

Molecular cloning followed standard protocols (Green & Sambrook, 2012, Molecular Cloning: A Laboratory Manual, 4th ed., CSH Press) and our previously published procedures for prime-editing and CRISPR/Cas9 reagent construction[4]. Prime-editing pegRNA plasmids were cloned by ligating annealed oligonucleotide pairs into the BsaI-digested pU6-pegRNA-GG-acceptor plasmid (Addgene #132777, gift from David Liu). Prime-editing nicking guide (ngRNA) plasmids were cloned into BsmBI-digested pBPK1520 (Addgene #65777, gift from Keith Joung)[51]. CRISPR/Cas9 gRNA plasmids were cloned into BbsI-digested pX330-GFP or pX330-mCherry plasmids as described previously. TALEN pairs for LRRK2 G2019S were used as previously reported[14]. Oligonucleotide sequences used for cloning and subsequent genotyping are provided in the published manuscript Supplemental Table 5[12].

2.6.5 – Genome editing of hESCs

Genome editing was performed using plasmid- or ribonucleoprotein (RNP)-based CRISPR/Cas9, TALEN, or prime-editing approaches as described previously[4]. hESCs cultured on MEFs were pre-treated with 10 μM ROCK inhibitor Y27632 for 2-24 hours before nucleofection. Cells were dissociated with collagenase IV followed by Accutase, and 5×10^5 to 1×10^6 cells were nucleofected in 20 μL P3 Primary Cell 4D-Nucleofector solution (Lonza) using program CA-137. Editing reagent combinations were selected according to the strategy required for each genomic alteration: CRISPR/Cas9 facilitated HDR (plasmid or

RNP); dual CRISPR/Cas9 for genomic deletions; CRISPR/Cas9 facilitated HDR with competing templates carrying synonymous mutations (used for heterozygous GBA1 IVS2+1 editing); TALEN-facilitated HDR (used for LRRK2 G2019S); and RNA- or plasmid-based prime editing (used for SNCA A53T, A30P, and LRRK2 G2019S). Specific reagent quantities per nucleofection are detailed in the published manuscript[12] and in the protocols.io deposition (dx.doi.org/10.17504/protocols.io.e6nvwkkewvmk/v2).

2.6.6 – Editing pipelines

Pipeline A used FACS enrichment post-nucleofection to purify effectively transfected cells, followed by clonal expansion and Sanger/NGS-based genotyping (full timeline: 30-35 days). Nucleofected hESCs were plated on MEFs in hESC medium containing 10 µM ROCK inhibitor at high density; 48-72 hours later, Accutase-dissociated single cells were FACS-sorted for the expression of a fluorescent marker protein and plated at clonal density (250-350 cells/cm^2). Individual colonies were picked 7-14 days after electroporation, and correctly targeted clones identified by Sanger or NGS sequencing (protocols.io: dx.doi.org/10.17504/protocols.io.b4piqvke). Pipeline B used low-cell-number nucleofection, limited dilution, and NGS-based genotyping in a 96-well plate format for high-throughput isolation of correctly edited clones within 21-35 days. Nucleofected cells were directly seeded onto MEFs in 96-well plates at 1,000 cells/plate. At day 14, plates were duplicated for (1) maintenance and (2) DNA extraction for NGS genotyping (protocols.io: dx.doi.org/10.17504/protocols.io.b4mmqu46;
dx.doi.org/10.17504/protocols.io.b4n3qvgn).

2.6.7 – NGS-based identification and validation of targeted clonal lines

Cell lysates from Pipeline B plate duplication were processed for DNA extraction in 100 µL 1X lysis buffer (50 mM KCl, 2 mM MgCl2, 0.45% NP-40, 0.45% Tween-20, 250 µg/mL proteinase K, 10 mM Tris-HCl pH 8) with overnight 50°C incubation followed by 95°C for 10 minutes. A ~300 bp genomic region covering the designed mutation was amplified using Titan DNA polymerase with primers containing NGS barcode attachment sites. Amplicons were indexed and sequenced on an iSeq (UC Berkeley Innovative Genomics Institute NGS core, 150PE). CRISPResso2[52] was used to analyze NGS data in prime-editing mode and identify wells containing the designed mutation according to defined criteria (heterozygous candidates: mutant allele frequency 20-70%, indel frequency <5%; homozygous candidates: mutant allele frequency >70%, indel frequency <5%). Wells containing the desired edit were subcloned and genotyped clonally to confirm clonality.

2.6.8 – Zygosity confirmation and Southern blotting

To exclude loss of heterozygosity at the targeted locus in homozygously-edited cell lines, the closest informative SNP to each editing site was identified from the whole-genome sequencing data of the parental WIBR3 hESC line. A genomic DNA region flanking both the SNP and the editing site was amplified by PCR and sequenced by Sanger sequencing or NGS. Clones showing LOH were excluded from the collection. For cell lines with larger genomic deletions or for which SNP-PCR was inconclusive, Southern blot analysis was

performed using standard protocols (Cold Spring Harbor Protocols 2021, doi:10.1101/pdb.top100396). Southern blot probes were generated by PCR amplification of 150-600 bp regions 3' and 5' of the targeted genomic region and radiolabelled using the Prime-it Random Primer Labeling Kit (Agilent). Restriction-digested genomic DNA was separated on 0.8% agarose gels, transferred to nylon membranes (Amersham), and hybridised with ^{32}P-labelled probes. Probe sequences and restriction-enzyme information are provided in the published manuscript[12] (protocols.io: https://doi.org/10.17504/protocols.io.bp2l6xe6dlqe/v1).

2.6.9 – Cortical spheroid differentiation and p53 irradiation assay

For assessment of p53 pathway function, WIBR3 hESCs were differentiated into early cortical spheroids following an adaptation of published protocols[53,54]. Briefly, hESC colonies were dissociated and plated at 1.8×10^6 cells per well of pre-coated 6-well Aggrewell 800 plates in hESC medium with 10 µM ROCK inhibitor. The next day, aggregates were transferred to ultralow-adherence plates with hESC medium supplemented with 5 µM Dorsomorphin and 10 µM SB431542 (daily media changes). From day 6, cultures were switched to neural-precursor expansion medium (Neurobasal + B27 without vitamin A, Penicillin-Streptomycin, GlutaMAX, HEPES, 20 ng/mL FGF2, 20 ng/mL EGF) and maintained until day 25 (protocols.io: https://doi.org/10.17504/protocols.io.5jyl8po57g2w/v1, https://doi.org/10.17504/protocols.io.5jyl8po57g2w/v1.

hESCs were transferred from MEF to feeder-free Matrigel substrate for 2 weeks prior to irradiation; cortical spheroids were irradiated at day 25 of differentiation. Both cell types were irradiated at 0, 0.5, 2, 5, and 10 Gy using a discrete cesium source (protocols.io: https://doi.org/10.17504/protocols.io.bp2l6xwbzlqe/v1). At 24 hours post-irradiation, each irradiation condition was labelled with lipid-modified MULTI-seq barcoded oligonucleotides following McGinnis and colleagues[55]: a 1:1 molar ratio of lipid-modified Anchor:Barcode oligo was incubated with cells in PBS on ice for 5 minutes, followed by addition of equimolar lipid-modified co-anchor and another 5-minute incubation on ice, followed by two washes in ice-cold PBS + 1% BSA. Labelled cells were pooled for single-cell sequencing using the Chromium Single Cell 3' Reagent Kit v3 with Feature Barcoding Technology (10x Genomics) (protocols.io:https://doi.org/10.17504/protocols.io.kxygx3xzkg8j/v1, https://doi.org/10.17504/protocols.io.kxygx3xzkg8j/v1.

For cortical spheroids, single-cell suspensions were FACS-sorted to remove debris and aggregates, and 10x Genomics scRNA-seq was performed using the Chromium Single Cell 3' Reagent Kit v3 targeting 2,000 cells per irradiation condition with one 10x lane per condition. MULTI-seq barcode deconvolution was performed using the published MULTI-seq R package (https://github.com/chris-mcginnis-ucsf/MULTI-seq/). Downstream analysis used Seurat v4[56] with default normalization and integration parameters. Droplets with more than 15% mitochondrial reads were excluded. Plots were generated with ggplot2 in R. Note on protocol distinctions relevant to Chapter 4: the MULTI-seq lipid-modified oligo labelling used here (day-25 cortical spheroids + day-35/37 DA neurons, single-experiment p53 assay)

differs from the cell-multiplexing strategy used in the Perturb-seq experiments of Chapter 4, which combine 10x Genomics CellPlex Cell Multiplexing Oligos (CMOs) for sample multiplexing with Feature Barcoding capture of guide-RNA sequences from a CROP-seq backbone. The MULTI-seq protocol described in this chapter was used only for the iSCORE-PD p53 irradiation assay and was not applied to any Perturb-seq experiment.

2.6.10 – Dopaminergic neuron differentiation

Feeder-free-adapted WIBR3 hESCs were differentiated into midbrain-specified dopaminergic neurons using the biphasic WNT signaling protocol of Kim and colleagues[26,27], with slight modifications. Briefly, hESC colonies were dissociated into single cells and seeded onto Matrigel-coated plates at 400,000-600,000 cells per well of a 6-well plate in mTeSR supplemented with 10 μM ROCK inhibitor. Differentiation was induced sequentially through four defined media conditions (Media A-D) containing combinations of SHH C25II (100-200 ng/mL), CHIR99021 (graduated from 0.7 μM to 7.5 μM to 3 μM), LDN (250 nM), SB431542 (10 μM), BDNF, GDNF, ascorbic acid, dibutyryl-cAMP, TGF-β3, and DAPT over 25 days, followed by replating and maturation in Neurobasal + B27 + L-Glutamine + Penicillin-Streptomycin + BDNF + GDNF + ascorbic acid + dibutyryl-cAMP + TGF-β3 + DAPT through day 35. Full medium compositions and replating schedules are provided at protocols.io https://doi.org/10.17504/protocols.io.3byl4q8yovo5/v1

Note on protocol distinctions relevant to Chapter 3: the iSCORE-PD scRNA-seq experiments of this chapter were performed on day 35-37 mature DA neurons derived from WIBR3-S1, S2, S3 subclones carrying no CRISPRi machinery, with each subclone harvested on a different day. The CRISPRi Perturb-seq experiments of Chapter 3 were performed on day 45 mature DA neurons derived from Cre-inducible CRISPRi clone C12 of WIBR3, carrying a pooled gRNA library, with Cre induction initiated at day 35 (a 10-day knockdown window prior to assay). The media composition and small-molecule patterning regimen from day 0 to day 35 are identical between the two chapters; the only protocol differences are the presence of the CRISPRi module and later timepoint assayed in Chapter 3. Full details of the Chapter 3 protocol are described in the methods section for Chapter 3.

2.6.11 – scRNA-seq library preparation and analysis of DA neurons

DA neurons were harvested with Accutase on day 35, 36, or 37 of culture and labelled with 10x Genomics CellPlex Cell Multiplexing Oligos (CMOs) following manufacturer recommendations (10x Genomics protocol CG000391 Rev B). Samples were pooled and loaded for 10x Chromium 3' v3.1 library preparation (CG000388 Rev C) targeting 30,000 cells per lane. Sequencing was performed on an Illumina NovaSeq 6000.

FASTQ files were processed with 10x Genomics CellRanger (v8.0.1) Count matrices were filtered to remove low-quality cells (cells with fewer than 1,500 genes, greater than 30,000 RNA counts, or greater than 10% mitochondrial reads were excluded). Filtered datasets were individually processed with Seurat v4[56] using SCTransform for normalization and variance stabilization, and integrated to generate the aggregate 10,097-cell dataset

analyzed in Section 2.4.2 (Figures 2C-E, Supplemental Figures 3-4). Cluster annotation, dot-plot construction, and cross-dataset comparison against the FOUNDIN-PD reference[13] followed standard Seurat workflows and the SingleR classifier (Wilcoxon rank-sum test) for cell-type assignment. Scripts for all scRNA-seq analyses presented in this chapter are publicly available at https://github.com/ASAP-Team-Rio/iSCORE-PD_DAN_scRNASeq.

2.6.12 – Microglia differentiation

In-vitro-differentiated microglia-like cells (iMGs) were generated using an adapted version of the Douvaras protocol[31]. Undifferentiated feeder-free WIBR3 colonies in mTeSR were seeded at low density onto reduced-growth-factor-Matrigel-coated T75 flasks (30 colonies/flask) by manual passaging. Myeloid differentiation was achieved through four sequential media steps over ~28 days (Step 1: mTeSR + 80 ng/mL BMP4, 4 days; Step 2: StemPro-34 SFM + 80 ng/mL VEGF + 25 ng/mL FGF + 100 ng/mL SCF, 3 days; Step 3: + 50 ng/mL SCF + 50 ng/mL IL-3 + 5 ng/mL TPO + 50 ng/mL M-CSF + 50 ng/mL Flt3, 9 days; Step 4: + 50 ng/mL M-CSF + 50 ng/mL Flt3 + 25 ng/mL GM-CSF, 14 days). Microglia progenitors were isolated and plated on Primaria plates (Corning) for ≥2 weeks of maturation in Neurobasal + N2 Neuroplex + GEM21 Neuroplex + 0.2% AlbuMAX I + 50 mM NaCl + sodium pyruvate + glutaMAX + Penicillin-Streptomycin + 50 ng/mL TGF-β1 + 100 ng/mL IL-34 + 12.5 ng/mL M-CSF (protocols.io: https://doi.org/10.17504/protocols.io.4r3l22zbjl1y/v1). Marker expression (CD16, CD45, CX3CR1, P2RY12, CD11b, CD14, IBA1, PU.1) was validated by immunocytochemistry and FACS analysis (FlowJo v10.8.0; protocols.io:
https://doi.org/10.17504/protocols.io.yxmvm3146l3p/v1;
https://doi.org/10.17504/protocols.io.81wgbxokqlpk/v1

2.6.13 – Immunocytochemistry and qRT-PCR

Immunocytochemistry of pluripotency markers (OCT4, SSEA4, DSHB), DA neuron markers (TH, Pelfreeze; FOXA2, R&D Systems), and microglia markers (IBA1, Abcam; P2RY12, Sigma; CX3CR1, Biolegend; PU.1, Cell Signaling Technology) followed standard PFA fixation and optional 0.03% Triton permeabilisation. Alkaline phosphatase staining used the Vector Black Substrate Kit (Vector Laboratories). Imaging was performed on a Zeiss ZEN 3.8 fluorescence microscope or epifluorescence/confocal microscopes with 10× objectives. For qRT-PCR, total RNA was isolated using the RNeasy kit (QIAGEN), 1-2 µg used for cDNA synthesis (High-capacity reverse transcriptase kit, ThermoFisher), and amplification performed on a QuantStudio 6 Flex using PowerUp SYBR green master mix. Expression was normalised to GAPDH and compared against human fibroblast controls (MRC-9, BJ1-hTERT, GM01660). Primer sequences are listed in the published manuscript Supplemental Table 5[12]. Plots were generated in GraphPad Prism v10.1.2. Full protocols are found at the links below:
https://doi.org/10.17504/protocols.io.yxmvm3146l3p/v1;
https://doi.org/10.17504/protocols.io.b4yyqxxw;
https://doi.org/10.17504/protocols.io.4r3l22r9pl1y/v1.

2.6.14 – Data and code availability

Genotyping, long-read sequencing (Oxford Nanopore and PacBio HiFi), NGS and Sanger sequencing data for targeted genotyping and zygosity analysis, and single-cell RNA-seq data from dopaminergic neuron differentiation and p53 pathway analysis are available at https://www.amp-pd.org/ via GP2 data-sharing agreements. Additional materials including aCGH reports, ICC images, Southern blot raw films, qRT-PCR results, and FACS analysis are deposited at Zenodo (10.5281/zenodo.10511722). All cell lines in the iSCORE-PD collection are banked and distributed by the WiCell Research Institute (https://www.wicell.org/) under standard material transfer agreements. Analysis scripts for the scRNA-seq DA neuron characterization are available at https://github.com/ASAP-Team-Rio/iSCORE-PD_DAN_scRNASeq.

2.6.15 – Short-read whole-genome sequencing pipeline

DNA samples were sequenced with Illumina short-read WGS at Psomagen (Rockville, MD), with a mean coverage of 30x. Data was processed using standard GP2 WGS pipelines. In brief, 150-bp paired-end reads were aligned to the human reference genome (GRCh38 build) using BWA-mem (https://github.com/lh3/bwa) following the functional equivalence pipeline[57]. Sample processing and variant calling were performed using DeepVariant v1.6.1[58]. Joint-genotyping was performed using GLnexus v1.4.3 with the preset DeepVariant WGS configuration[59]. A detailed description of the WGS pipeline can be found at https://github.com/GP2code/releases/tree/main/BETA-APR2022/wgs_var_calling and at https://github.com/GP2code/GP2-WorkingGroups/tree/main/MN-DAWG-Monogenic-Data-Analysis/Terra_wdl/variant_calling/deepvariant.

2.6.16 – Analysis pipeline to identify clone-specific SNVs/indels

To identify SNVs/indels for each group of cell lines modified by a given set of CRISPR/Cas9 or prime-editing reagents, the variants identified with the WGS pipeline were first filtered to only consider calls with a GLnexus quality score greater than or equal to 30. Then, to identify SNVs/indels specific to only one edited group, the SNVs were further filtered to remove any calls found in any of the other cell lines in the collection. The remaining variants account for each editing group's specific SNVs/indels and were used to determine unique and shared variants for each clonal cell line within each edited group. These unique and shared SNVs/indels were used to further characterize their contribution to coding or non-coding regions, as well as their effect on the coding sequences, by leveraging tools within the BioConductor DeepVariant package[60] against the TxDb.Hsapiens.UCSC.hg38.knownGene_3.18.0 human transcript annotation package (based on the UCSC hg38 genome and knownGene table). A detailed description of the pipeline can be found at https://github.com/hockemeyer-ucb/pd-sv-analysis. The group-unique SNVs/indels were used to generate a phylogeny tree using the BioConductor package fastreeR and visualized using the R package ape[61,62].

2.6.17 – Off-target analysis

To test whether any group-specific SNVs could be linked to potential off-target effects triggered by the CRISPR/Cas9 or prime-editing reagents, all putative off-target sites were identified using Cas-OFFinder[39] (https://github.com/snugel/cas-offinder) and used to identify nearby SNVs/indels that could have resulted from the editing strategies. A detailed description of the pipeline can be found at https://github.com/hockemeyer-ucb/pd-sv-analysis.

2.A - Appendix: Per-gene editing strategies and quality control

This appendix reproduces the per-gene editing details described in Supplementary Note 1 of the published iSCORE-PD manuscript[12]. Each subsection below describes the rationale for the chosen mutation, the editing strategy employed, and the quality-control results validating the resulting cell lines. Cell lines are summarized in Table 2.1. All references to supplemental figures point the reader to the published manuscript, since these supplemental figures are not included within this dissertation.

2.A.1 – Control cell lines

Together with the parental WIBR3 cell line, we provide a set of subclones with this collection (WIBR3-S1, S2, S3; Figure 2.1, Supplemental Figure 2.1). In addition, we included WIBR3 cell lines that were isolated as part of the standard genome-editing pipeline but did not exhibit any genetic modifications at the targeted locus. We consider these as "edited wild-type" cells (EWT), which are the best experimental control to account for any non-specific changes caused by the gene-editing process (EWT1-3: prime-editing controls, Pipeline B; EWT4-5: CRISPR/Cas9 controls, Pipeline B; EWT6-8: prime-editing controls, Pipeline A) (Table 2.1).

2.A.2 – *SNCA* (PARK1)

The *SNCA* gene encodes the α-synuclein protein. The discovery that mutations and copy-number increases of the *SNCA* gene linked to familial PD, along with the identification of non-coding variants in the *SNCA* locus as a risk factor for sporadic PD, indicate a central pathogenic role for this protein[63]. Moreover, fibrillar α-synuclein is the major component of Lewy bodies, and changes in the dosage, aggregation, and clearance properties of α-synuclein are thought to be central pathogenic drivers of PD[1,2,63,64]. We established prime-editing reagents to efficiently introduce the A30P or A53T mutation in the SNCA gene, both linked to autosomal dominant forms of PD[4,64,65]. Using this approach (Supplemental Figure 2.A.7A,D), we generated cell lines for the iSCORE-PD collection that now includes 3 cell lines carrying the A53T mutation (heterozygous) (Supplemental Figure 2.A.7B,C) and 4 cell lines carrying the A30P mutation (3 heterozygous and 1 homozygous) (Supplemental Figure 2.A.7E,F). All cell lines passed the above-described quality-control steps. In addition, we performed SNP-PCR followed by Sanger sequencing-based zygosity analysis to exclude LOH in the homozygous *SNCA* A30P line (Supplemental Figure 2.A.6G).

2.A.3 – *PRKN* (PARK2)

Mutations in the *PRKN* gene, which encodes the E3 ubiquitin-protein ligase Parkin, are the most frequent cause of autosomal recessive PD[66]. The first genetic alterations in *PRKN* linked to PD were distinct large-scale deletions[67]. Since then, a wide range of deletions and point mutations have been identified[68]. Molecular alterations in Parkin impact a wide range of molecular and cellular functions including responses to oxidative stress, mitochondrial membrane potential, mitochondrial motility, contact sites with the endoplasmic reticulum, and the regulation of inflammatory responses[69,70]. Homozygous

and compound heterozygous deletions of exon 3 (Ex3del) are repeatedly found in families with PD, and Ex3del is now a confirmed pathogenic variant of *PRKN*[68,71]. We used a dual CRISPR/Cas9 approach (Figure 2.5C(v)) targeting each side of exon 3 to generate 3 cell lines that are homozygous for the Ex3del (Supplemental Figure 2.A.8A-C). Subsequent Southern blot analysis to exclude LOH revealed that one cell line (WIBR3_PRKN_X3DEL_B1-3) carries a larger Ex3del (~3.8 kb, Supplemental Figure 2.A.8D,E), which we fully characterized by NGS sequencing (Supplemental Figure 2.A.8F). We included this cell line in the iSCORE-PD collection since the flanking exons 2 and 4 are not affected by the deletion and thus no significant consequences are expected from the extended deletion. No heterozygous clones have been included in this collection, since these genotypes are frequent in control populations and not associated with a higher risk of PD[72].

2.A.4 – *PINK1* (PARK6)

Mutations in *PINK1* were first associated with PD in 2004 following earlier studies linking the PARK6 genomic region with increased risk for PD[73]. Subsequently, a large number of point mutations, frameshift mutations, and deletions have been identified, predominantly affecting the activity of the kinase domain[68]. This indicates that PINK1 loss-of-function is the cause for early-onset autosomal recessive PD. Serine/threonine-protein kinase PINK1 plays a crucial role in mitochondrial quality control by regulating mitochondrial homeostasis and clearance[74]. Based on the segregation of the Q129X and a related Q129fsX157 mutation in PINK1 observed in two large families with PD[75], we generated 3 cell lines that are homozygous for the Q129X mutation in PINK1 (Supplemental Figure 2.A.9A-C). To exclude LOH, we performed SNP-PCR followed by Sanger sequencing-based zygosity analysis (Supplemental Supplemental Figure 2.A.9D). Heterozygous *PINK1* mutant cell lines were not included in the iSCORE-PD collection, as the heterozygous genotype is not considered a risk factor for PD.

2.A.5 – DJ-1 (*PARK7)*

Since its initial association with PD[76], multiple point mutations and genomic rearrangements in *PARK7* (DJ1), encoding the Parkinson disease protein 7, have been identified as a rare cause for autosomal recessive PD[68]. While the exact function of this enzyme remains largely unclear, DJ-1 is implicated in regulating transcription, cell growth, and oxidative stress response pathways linked to redox sensing, cell survival and apoptosis[77,78]. Based on the identification of homozygous deletions of either exon 5[79] or exons 1 to 5[80] in families with PD, we used a dual CRISPR/Cas9 approach to closely recapitulate the exon 1-5 deletion (Supplemental Figure 2.A.10A-C). Southern blot analysis was performed to exclude LOH in the homozygous exon 1\-5 deleted cell lines (Supplemental Figure 2.A.10D,E).

2.A.6 – *LRRK2* (PARK8)

Mutations in the *LRRK2* gene, encoding leucine-rich repeat serine/threonine-protein kinase 2, were first linked to PD in 2004[81]. Over the years, more than 100 different variants of the gene have been described[82], establishing *LRRK2* coding variants as the most frequently

mutated gene linked to dominant and sporadic forms of PD. Among these mutations, G2019S is the most common substitution identified across populations[83], found in approximately 4% of dominantly inherited familial PD cases, in both heterozygous and homozygous forms, and around 1% of sporadic PD cases[84]. This variant increases the kinase activity of LRRK2, affecting a wide range of cellular and molecular processes including vesicular trafficking, cytoskeleton dynamics, autophagy and lysosomal degradation, neurotransmission, mitochondrial function, and immune and microglial responses[85,86]. We previously established CRISPR/Cas9, TALEN, and prime-editing reagents to efficiently introduce the G2019S mutation in the *LRRK2* gene[4]. Using this approach (Supplemental Supplemental Figure 2.A.11A), we generated additional cell lines so that the iSCORE-PD collection now includes 5 clones carrying G2019S (4 heterozygous and 1 homozygous) (Supplemental Figure 2.A.11B,C). We performed SNP-PCR followed by Sanger sequencing-based zygosity analysis to exclude LOH in the homozygous *LRRK2* G2019S line (Supplemental Figure 2.A.11D).

2.A.7 – *ATP13A2* (PARK9)

Mutations in the *ATP13A2* gene, encoding the polyamine-transporting ATPase 13A2 (ATP13A2), were identified as a cause for PD in the PARK9 locus in 2006, with this genomic region previously associated with Kufor-Rakeb disease[87,88]. Various mutations in the gene result in a truncated ATP13A2 protein, causing mis-localization from the lysosome to the endoplasmic reticulum, where it accumulates before being eventually targeted for proteasomal degradation in the cytoplasm[87,88]. Loss of ATP13A2 is linked to mitochondrial and lysosomal dysfunction, as well as to the accumulation of α-synuclein due to the dysregulation of proteasomal and autophagy-mediated protein degradation[89,90]. Based on the identification of a homozygous frameshift mutation (T367RfsX29) in a family with recessive PD[91], we used CRISPR/Cas9-based genome editing to engineer the frameshift T367Rfs mutation in the *ATP13A2* gene (Supplemental Figure 2.A.12A). Given the recessive inheritance pattern of the gene, our collection contains 5 clones with the T367Rfs frameshift mutation, predicted to result in a truncated loss-of-function protein similar to that observed in PD patients (Supplemental Figure 2.A.12B,C). We performed SNP-PCR followed by NGS sequencing-based zygosity analysis to exclude LOH in the homozygous edited cell lines (Supplemental Figure 12D).

2.A.8 – *FBXO7* (PARK15)

Mutations in the *FBXO7* gene, encoding F-box only protein 7 (FBXO7), were first linked to autosomal recessive PD in 2008[92]. The most notable homozygous *FBXO7* variant, found in multiple family pedigrees with PD, is the truncating R498X premature-stop mutation located in the proline-rich region of the *FBXO7* protein[93]. This mutation disrupts the interaction of *FBXO7* with PINK1 and Parkin[94,95] and results in abnormal localization and reduced stability of the truncated FBXO7 protein. This causes disruption of mitophagy and leads to mitochondrial aggregation[94,96,97]. We used CRISPR/Cas9-based genome editing to engineer a frameshift mutation that leads to a premature stop and is predicted to result in truncated *FBXO7* similar to the protein in patients carrying the R498X mutation

(Supplemental Figure 2.A.13A-C). We performed Southern blot and SNP-PCR followed by NGS sequencing-based zygosity analysis to exclude LOH in the homozygous edited cell line (Supplemental Figure 2.A.13D-F).

2.A.9 – *DNAJC6* (PARK19)

The *DNAJC6* gene encodes auxilin, which belongs to the DNAJ/HSP40 family of proteins that regulate molecular chaperone activity. *DNAJC6* was initially linked to very early-onset autosomal recessive PD in 2012[98] through the identification of a mutation that affects splicing and the expression of auxilin. Consistent with a role as co-chaperone to recruit *HSC70* to clathrin-coated vesicles, hESC models show that *DNAJC6* alterations result in loss of auxilin protein, leading to the accumulation of clathrin, reduced vesicular transport, and the degeneration of midbrain dopaminergic neurons[99]. Furthermore, these alterations are associated with α-synuclein aggregation, mitochondrial and lysosomal dysfunction, and lipid defects[99,100]. We used a CRISPR/Cas9-based approach to recapitulate the effect of the c.801-2 A>G splice-acceptor-site mutation by either precise insertion of this mutation or a frameshift modification (Supplemental Figure 2.A.14). Similar frameshift modifications were previously shown to recapitulate the loss of DNAJC6 expression in hESC-derived neuronal cells[99], similar to what is observed in DNAJC6 variant carriers[98].

2.A.10 – *SYNJ1* (PARK20)

SYNJ1 encodes synaptojanin-1, and was first linked to early-onset autosomal recessive PD in 2013[101,102]. Synaptojanin-1 is predominantly expressed in neurons and concentrated in presynaptic terminals. Homozygous *SYNJ1* mutations are linked to alterations in lipid metabolism and vesicle trafficking[101,102], as well as to defects in autophagosome maturation[103]. We used a CRISPR/Cas9 approach to insert the R258Q substitution, which was identified in a homozygous state in several independent families with PD[101,102,104] (Supplemental Figure 2.A.15A). The iSCORE-PD collection includes clones that are homozygous for the R258Q mutation in *SYNJ1* or compound heterozygous for R258Q and a frameshift allele at the same location. Both genotypes are expected to recapitulate the modification in SYNJ1 associated with disease (Supplemental Figure 2.A.15B,C). We performed Southern blot-based zygosity analysis to exclude LOH in the homozygous edited cell line (Supp. Figure 2.A.15D-F).

2.A.11 – *VPS13C* (PARK23)

Mutations in *VPS13C* were initially identified as a cause of autosomal recessive early-onset PD in 2016[105]. The initial functional analysis revealed that disruption of intermembrane lipid-transfer protein *VPS13C* causes decreased mitochondrial membrane potential, mitochondrial fragmentation, increased respiration rates, exacerbated *PINK1*/Parkin-dependent mitophagy, and transcriptional upregulation of *PRKN* in response to mitochondrial damage[105]. As *VPS13C* is critical for the transport of lipids between the ER and endosome/lysosome, as well as for lipid-droplet formation[106,107], loss of *VPS13C* causes the accumulation of lysosomes and altered lipid profiles[108]. We used CRISPR/Cas9-based editing to introduce the W395C and A444P variants into the *VPS13C* gene (Supplemental

Figure 2.A.16A,E), both of which are found as homozygous or compound heterozygous mutations in PD patients[109,110]. While *VPS13C* is thought to cause PD through a loss-of-function mechanism, we included additional frameshift alleles in the iSCORE-PD collection to allow further investigation of the loss-of-function mechanism (Supplemental Figure 2.A.16B,C,F,G). We performed SNP-PCR followed by Sanger sequencing-based zygosity analysis to exclude LOH in the homozygous *VPS13C* A444P lines.

2.A.12 – *GBA1*

The *GBA1* gene codes for the enzyme glucocerebrosidase, which is essential for maintaining glycosphingolipid homeostasis. While homozygous or compound heterozygous pathogenic variants in *GBA1*, associated with reduced glucocerebrosidase activity, cause autosomal recessive Gaucher disease, heterozygous carriers of pathogenic *GBA1* variants have a high risk of developing PD[111,112]. *GBA1* mutations are currently considered the strongest risk factor for PD and Lewy body dementia, with odds ratios between 1.4 and greater than 10[1]. Given that *GBA1* mutations are present in approximately 3-20% of sporadic PD patients across different populations, *GBA1* represents the most prevalent genetic risk factor for PD[1]. Over 300 mutations in *GBA1* with variable risks for developing PD have been reported[112]. Among them, the splice-site IVS2+1 mutation, causing mis-splicing and loss of GBA1 expression, represents one of the most pathogenic alleles for PD[111,112]. To insert IVS2+1 in the *GBA1* gene, we devised a CRISPR/Cas9-based targeting strategy that allows specific targeting of the *GBA1* gene and not the nearby highly homologous *GBAP1* pseudogene (Supplemental Figure 2.A.17A). To identify correctly targeted clones, we used a genotyping strategy that can conclusively distinguish between *GBA1* and the *GBAP1* pseudogene based on small sequence variation. Using this approach, we generated 2 heterozygous and 2 homozygous edited cell lines carrying the IVS2+1 mutation in *GBA1*. In addition, to allow comparison between the IVS2+1 mutation and a loss-of-function allele, we also included cell lines with a frameshift mutation at the same genomic location in the iSCORE-PD collection.

References

1. Blauwendraat C, Nalls MA, Singleton AB. The genetic architecture of Parkinson's disease. The Lancet Neurology. 2020 Feb;19(2):170–8. doi:10.1016/S1474-4422(19)30287-X PubMed PMID: 31521533; PubMed Central PMCID: PMC8972299.

2. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: A meta-analysis of genome-wide association studies. The Lancet Neurology. 2019 Dec;18(12):1091–102. doi:10.1016/S1474-4422(19)30320-5

3. Soldner F, Jaenisch R. Stem Cells, Genome Editing, and the Path to Translational Medicine. Cell. 2018 Oct;175(3):615–32. doi:10.1016/j.cell.2018.09.010 PubMed PMID: 30340033; PubMed Central PMCID: PMC6461399.

4. Li H, Busquets O, Verma Y, Syed KM, Kutnowski N, Pangilinan GR, et al. Highly efficient generation of isogenic pluripotent stem cell models using prime editing. eLife. 2022 Sep;11:e79208. doi:10.7554/eLife.79208

5. Soldner F, Laganière J, Cheng AW, Hockemeyer D, Gao Q, Alagappan R, et al. Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell. 2011 Jul;146(2):318–31. doi:10.1016/j.cell.2011.06.019 PubMed PMID: 21757228; PubMed Central PMCID: PMC3155290.

6. Ryan SD, Dolatabadi N, Chan SF, Zhang X, Akhtar MW, Parker J, et al. Isogenic Human iPSC Parkinson's Model Shows Nitrosative Stress-Induced Dysfunction in MEF2-PGC1α Transcription. Cell. 2013 Dec;155(6):1351–64. doi:10.1016/j.cell.2013.11.009

7. Pantazis CB, Yang A, Lara E, McDonough JA, Blauwendraat C, Peng L, et al. A reference human induced pluripotent stem cell line for large-scale collaborative studies. Cell Stem Cell. 2022 Dec;29(12):1685–1702.e22. doi:10.1016/j.stem.2022.11.004 PubMed PMID: 36459969; PubMed Central PMCID: PMC9782786.

8. Liang G, Zhang Y. Genetic and epigenetic variations in iPSCs: Potential causes and implications for application. Cell Stem Cell. 2013 Aug;13(2):149–59. doi:10.1016/j.stem.2013.07.001 PubMed PMID: 23910082; PubMed Central PMCID: PMC3760008.

9. Bar S, Schachter M, Eldar-Geva T, Benvenisty N. Large-Scale Analysis of Loss of Imprinting in Human Pluripotent Stem Cells. Cell Reports. 2017 May;19(5):957–68. doi:10.1016/j.celrep.2017.04.020

10. Lengner CJ, Gimelbrant AA, Erwin JA, Cheng AW, Guenther MG, Welstead GG, et al. Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations. Cell. 2010 May;141(5):872–83. doi:10.1016/j.cell.2010.04.010 PubMed PMID: 20471072.

11. Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP, et al. Genetic engineering of human pluripotent cells using TALE nucleases. Nature Biotechnology. 2011 Jul;29(8):731–4. doi:10.1038/nbt.1927 PubMed PMID: 21738127; PubMed Central PMCID: PMC3152587.

12. Busquets O, Li H, Syed KM, Jerez PA, Dunnack J, Bu RL, et al. iSCORE-PD: An isogenic stem cell collection to research Parkinson's Disease. bioRxiv: The Preprint Server for Biology. 2025 Mar;2024.02.12.579917. doi:10.1101/2024.02.12.579917 PubMed PMID: 38405931; PubMed Central PMCID: PMC10888955.

13. Bressan E, Reed X, Bansal V, Hutchins E, Cobb MM, Webb MG, et al. The Foundational Data Initiative for Parkinson Disease: Enabling efficient translation from genetic maps to mechanism. Cell Genomics. 2023 Mar;3(3):100261. doi:10.1016/j.xgen.2023.100261 PubMed PMID: 36950378; PubMed Central PMCID: PMC10025424.

14. Soldner F, Stelzer Y, Shivalila CS, Abraham BJ, Latourelle JC, Barrasa MI, et al. Parkinson-associated risk variant in distal enhancer of α-synuclein modulates target gene expression. Nature. 2016 May;533(7601):95–9. doi:10.1038/nature17939 PubMed PMID: 27096366; PubMed Central PMCID: PMC5042324.

15. Kolmogorov M, Billingsley KJ, Mastoras M, Meredith M, Monlong J, Lorig-Roach R, et al. Scalable Nanopore sequencing of human genomes provides a comprehensive view of haplotype-resolved variation and methylation. Nature Methods. 2023 Oct;20(10):1483–92. doi:10.1038/s41592-023-01993-x PubMed PMID: 37710018; PubMed Central PMCID: PMC11222905.

16. Liao WW, Asri M, Ebler J, Doerr D, Haukness M, Hickey G, et al. A draft human pangenome reference. Nature. 2023 May;617(7960):312–24. doi:10.1038/s41586-023-05896-x PubMed PMID: 37165242; PubMed Central PMCID: PMC10172123.

17. Bellenguez C, Küçükali F, Jansen IE, Kleineidam L, Moreno-Grau S, Amin N, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. Nature Genetics. 2022 Apr;54(4):412–36. doi:10.1038/s41588-022-01024-z PubMed PMID: 35379992; PubMed Central PMCID: PMC9005347.

18. Scheltens P, De Strooper B, Kivipelto M, Holstege H, Chételat G, Teunissen CE, et al. Alzheimer's disease. Lancet. 2021 Apr;397(10284):1577–90.

doi:10.1016/S0140-6736(20)32205-4 PubMed PMID: 33667416; PubMed Central PMCID: PMC8354300.

19. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016 Aug;536(7616):285–91. doi:10.1038/nature19057 PubMed PMID: 27535533; PubMed Central PMCID: PMC5018207.

20. Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, et al. A genomic mutational constraint map using variation in 76,156 human genomes. Nature. 2024 Jan;625(7993):92–100. doi:10.1038/s41586-023-06045-0 PubMed PMID: 38057664; PubMed Central PMCID: PMC11629659.

21. Van Deerlin VM, Sleiman PMA, Martinez-Lage M, Chen-Plotkin A, Wang LS, Graff-Radford NR, et al. Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. Nature Genetics. 2010 Mar;42(3):234–9. doi:10.1038/ng.536 PubMed PMID: 20154673; PubMed Central PMCID: PMC2828525.

22. Simón-Sánchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, Berg D, et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nature Genetics. 2009 Dec;41(12):1308–12. doi:10.1038/ng.487 PubMed PMID: 19915575; PubMed Central PMCID: PMC2787725.

23. Strang KH, Golde TE, Giasson BI. MAPT mutations, tauopathy, and mechanisms of neurodegeneration. Laboratory Investigation; a Journal of Technical Methods and Pathology. 2019 Jul;99(7):912–28. doi:10.1038/s41374-019-0197-x PubMed PMID: 30742061; PubMed Central PMCID: PMC7289372.

24. Ihry RJ, Worringer KA, Salick MR, Frias E, Ho D, Theriault K, et al. P53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. Nature Medicine. 2018 Jul;24(7):939–46. doi:10.1038/s41591-018-0050-6 PubMed PMID: 29892062.

25. Merkle FT, Ghosh S, Kamitaki N, Mitchell J, Avior Y, Mello C, et al. Human pluripotent stem cells recurrently acquire and expand dominant negative P53 mutations. Nature. 2017 May;545(7653):229–33. doi:10.1038/nature22312 PubMed PMID: 28445466; PubMed Central PMCID: PMC5427175.

26. Kim TW, Piao J, Koo SY, Kriks S, Chung SY, Betel D, et al. Biphasic Activation of WNT Signaling Facilitates the Derivation of Midbrain Dopamine Neurons from hESCs for Translational Use. Cell Stem Cell. 2021 Feb;28(2):343–355.e5. doi:10.1016/j.stem.2021.01.005 PubMed PMID: 33545081; PubMed Central PMCID: PMC8006469.

27. Piao J, Zabierowski S, Dubose BN, Hill EJ, Navare M, Claros N, et al. Preclinical Efficacy and Safety of a Human Embryonic Stem Cell-Derived Midbrain Dopamine Progenitor Product, MSK-DA01. Cell Stem Cell. 2021 Feb;28(2):217–229.e7. doi:10.1016/j.stem.2021.01.004 PubMed PMID: 33545080; PubMed Central PMCID: PMC7903922.

28. Andersen MS, Bandres-Ciga S, Reynolds RH, Hardy J, Ryten M, Krohn L, et al. Heritability Enrichment Implicates Microglia in Parkinson's Disease Pathogenesis. Annals of Neurology. 2021 May;89(5):942–51. doi:10.1002/ana.26032 PubMed PMID: 33502028; PubMed Central PMCID: PMC9017316.

29. Langston RG, Beilina A, Reed X, Kaganovich A, Singleton AB, Blauwendraat C, et al. Association of a common genetic variant with Parkinson's disease is mediated by microglia. Science Translational Medicine. 2022 Jul;14(655):eabp8869. doi:10.1126/scitranslmed.abp8869 PubMed PMID: 35895835; PubMed Central PMCID: PMC9809150.

30. Heneka MT, McManus RM, Latz E. Inflammasome signalling in brain function and neurodegenerative disease. Nature Reviews Neuroscience. 2018 Oct;19(10):610–21. doi:10.1038/s41583-018-0055-7 PubMed PMID: 30206330.

31. Douvaras P, Sun B, Wang M, Kruglikov I, Lallos G, Zimmer M, et al. Directed Differentiation of Human Pluripotent Stem Cells to Microglia. Stem Cell Reports. 2017 Jun;8(6):1516–24. doi:10.1016/j.stemcr.2017.04.023 PubMed PMID: 28528700; PubMed Central PMCID: PMC5470097.

32. Haapaniemi E, Botla S, Persson J, Schmierer B, Taipale J. CRISPR-Cas9 genome editing induces a P53-mediated DNA damage response. Nature Medicine. 2018 Jul;24(7):927–30. doi:10.1038/s41591-018-0049-z PubMed PMID: 29892067.

33. Kuijk E, Jager M, van der Roest B, Locati MD, Van Hoeck A, Korzelius J, et al. The mutational impact of culturing human pluripotent and adult stem cells. Nature Communications. 2020 May;11(1):2493. doi:10.1038/s41467-020-16323-4 PubMed PMID: 32427826; PubMed Central PMCID: PMC7237696.

34. Thompson O, von Meyenn F, Hewitt Z, Alexander J, Wood A, Weightman R, et al. Low rates of mutation in clinical grade human pluripotent stem cells under different culture conditions. Nature Communications. 2020 Mar;11(1):1528. doi:10.1038/s41467-020-15271-3 PubMed PMID: 32251294; PubMed Central PMCID: PMC7089967.

35. Andrews PW, Ben-David U, Benvenisty N, Coffey P, Eggan K, Knowles BB, et al. A ssessing the Safety of Human Pluripotent Stem Cells and Their Derivatives for Clinical Applications. Stem Cell Reports. 2017 Jul;9(1):1–4.

doi:10.1016/j.stemcr.2017.05.029 PubMed PMID: 28700896; PubMed Central PMCID: PMC7206225.

36. Närvä E, Autio R, Rahkonen N, Kong L, Harrison N, Kitsberg D, et al. High-resolution DNA analysis of human embryonic stem cell lines reveals culture-induced copy number changes and loss of heterozygosity. Nature Biotechnology. 2010 Apr;28(4):371–7. doi:10.1038/nbt.1615 PubMed PMID: 20351689.

37. Weisheit I, Kroeger JA, Malik R, Klimmt J, Crusius D, Dannert A, et al. Detection of Deleterious On-Target Effects after HDR-Mediated CRISPR Editing. Cell Reports. 2020 May;31(8):107689. doi:10.1016/j.celrep.2020.107689 PubMed PMID: 32460021.

38. Lill CM, Mashychev A, Hartmann C, Lohmann K, Marras C, Lang AE, et al. Launching the movement disorders society genetic mutation database (MDSGene). Movement Disorders: Official Journal of the Movement Disorder Society. 2016 May;31(5):607–9. doi:10.1002/mds.26651 PubMed PMID: 27156390.

39. Bae S, Park J, Kim JS. Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics. 2014 May;30(10):1473–5. doi:10.1093/bioinformatics/btu048

40. Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature. 2019 Dec;576(7785):149–57. doi:10.1038/s41586-019-1711-4 PubMed PMID: 31634902; PubMed Central PMCID: PMC6907074.

41. Lee YJ, Syed K, Busquets O, Li H, Dunnack J, Bateup HS, et al. Alternative pre-mRNA Splicing and Gene Expression Patterns in Midbrain Lineage Cells Carrying Familial Parkinson's Disease Mutations. bioRxiv; 2024. doi:10.1101/2024.02.28.582420

42. Brenton JW, Follett J, Nirujogi R, Toomey CE, Lopez-Garcia P, Evans JR, et al. TDP-43 loss of function drives aberrant splicing in Parkinson's disease. bioRxiv; 2025. doi:10.1101/2025.09.04.673943

43. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. GigaScience. 2015;4:7. doi:10.1186/s13742-015-0047-8 PubMed PMID: 25722852; PubMed Central PMCID: PMC4342193.

44. Gogarten SM, Bhangale T, Conomos MP, Laurie CA, McHugh CP, Painter I, et al. GWASTools: An R/Bioconductor package for quality control and analysis of genome-wide association studies. Bioinformatics. 2012 Dec;28(24):3329–31. doi:10.1093/bioinformatics/bts610 PubMed PMID: 23052040; PubMed Central PMCID: PMC3519456.

45. Li H. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics. 2018 Sep;34(18):3094–100. doi:10.1093/bioinformatics/bty191 PubMed PMID: 29750242; PubMed Central PMCID: PMC6137996.

46. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021 Feb;10(2):giab008. doi:10.1093/gigascience/giab008 PubMed PMID: 33590861; PubMed Central PMCID: PMC7931819.

47. Shafin K, Pesout T, Chang PC, Nattestad M, Kolesnikov A, Goel S, et al. Haplotype-aware variant calling with PEPPER-Margin-DeepVariant enables high accuracy in nanopore long-reads. Nature Methods. 2021 Nov;18(11):1322–32. doi:10.1038/s41592-021-01299-w PubMed PMID: 34725481; PubMed Central PMCID: PMC8571015.

48. Smolka M, Paulin LF, Grochowski CM, Horner DW, Mahmoud M, Behera S, et al. Detection of mosaic and population-level structural variants with Sniffles2. Nature Biotechnology. 2024 Oct;42(10):1571–80. doi:10.1038/s41587-023-02024-y PubMed PMID: 38168980; PubMed Central PMCID: PMC11217151.

49. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research. 2010 Sep;38(16):e164. doi:10.1093/nar/gkq603 PubMed PMID: 20601685; PubMed Central PMCID: PMC2938201.

50. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018 Oct;562(7726):203–9. doi:10.1038/s41586-018-0579-z PubMed PMID: 30305743; PubMed Central PMCID: PMC6786975.

51. Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature. 2015 Jul;523(7561):481–5. doi:10.1038/nature14592 PubMed PMID: 26098369; PubMed Central PMCID: PMC4540238.

52. Clement K, Rees H, Canver MC, Gehrke JM, Farouni R, Hsu JY, et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nature Biotechnology. 2019 Mar;37(3):224–6. doi:10.1038/s41587-019-0032-3 PubMed PMID: 30809026; PubMed Central PMCID: PMC6533916.

53. Paşca AM, Sloan SA, Clarke LE, Tian Y, Makinson CD, Huber N, et al. Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture. Nature Methods. 2015 Jul;12(7):671–8. doi:10.1038/nmeth.3415 PubMed PMID: 26005811; PubMed Central PMCID: PMC4489980.

54. Yoon SJ, Elahi LS, Pasca AM, Marton RM, Gordon A, Revah O, et al. Reliability of human cortical organoid generation. Nature Methods. 2019 Jan;16(1):75–8. doi:10.1038/s41592-018-0255-0 PubMed PMID: 30573846; PubMed Central PMCID: PMC6677388.

55. McGinnis CS, Patterson DM, Winkler J, Conrad DN, Hein MY, Srivastava V, et al. MULTI-seq: Sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. Nature Methods. 2019 Jul;16(7):619–26. doi:10.1038/s41592-019-0433-8 PubMed PMID: 31209384; PubMed Central PMCID: PMC6837808.

56. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021 Jun;184(13):3573–3587.e29. doi:10.1016/j.cell.2021.04.048 PubMed PMID: 34062119; PubMed Central PMCID: PMC8238499.

57. Regier AA, Farjoun Y, Larson DE, Krasheninina O, Kang HM, Howrigan DP, et al. Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. Nature Communications. 2018 Oct;9(1):4038. doi:10.1038/s41467-018-06159-4

58. Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T, Ku A, et al. A universal SNP and small-indel variant caller using deep neural networks. Nature Biotechnology. 2018 Oct;36(10):983–7. doi:10.1038/nbt.4235

59. Yun T, Li H, Chang PC, Lin MF, Carroll A, McLean CY. Accurate, scalable cohort variant calls using DeepVariant and GLnexus, editorRobinson P. Bioinformatics. 2021 Apr;36(24):5582–9. doi:10.1093/bioinformatics/btaa1081

60. Obenchain V, Lawrence M, Carey V, Gogarten S, Shannon P, Morgan M. VariantAnnotation : A Bioconductor package for exploration and annotation of genetic variants. Bioinformatics. 2014 Jul;30(14):2076–8. doi:10.1093/bioinformatics/btu168

61. Gkanogiannis A, Gazut S, Salanoubat M, Kanj S, Brüls T. A scalable assembly-free variable selection algorithm for biomarker discovery from metagenomes. BMC Bioinformatics. 2016 Aug;17(1):311. doi:10.1186/s12859-016-1186-3

62. Paradis E, Schliep K. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R, editorSchwartz R. Bioinformatics. 2019 Feb;35(3):526–8. doi:10.1093/bioinformatics/bty633

63. Devine MJ, Gwinn K, Singleton A, Hardy J. Parkinson's disease and α-synuclein expression. Movement Disorders: Official Journal of the Movement Disorder Society. 2011 Oct;26(12):2160–8. doi:10.1002/mds.23948 PubMed PMID: 21887711; PubMed Central PMCID: PMC4669565.

64. Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, et al. Mutation I n the alpha-synuclein gene identified in families with Parkinson's disease. Science. 1997 Jun;276(5321):2045–7. doi:10.1126/science.276.5321.2045 PubMed PMID: 9197268.

65. Krüger R, Kuhn W, Müller T, Woitalla D, Graeber M, Kösel S, et al. Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. Nature Genetics. 1998 Feb;18(2):106–8. doi:10.1038/ng0298-106 PubMed PMID: 9462735.

66. Klein C, Lohmann-Hedrich K, Rogaeva E, Schlossmacher MG, Lang AE. Deciphering the role of heterozygous mutations in genes associated with parkinsonism. The Lancet Neurology. 2007 Jul;6(7):652–62. doi:10.1016/S1474-4422(07)70174-6 PubMed PMID: 17582365.

67. Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, Minoshima S, et al. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature. 1998 Apr;392(6676):605–8. doi:10.1038/33416 PubMed PMID: 9560156.

68. Kasten M, Hartmann C, Hampf J, Schaake S, Westenberger A, Vollstedt EJ, et al. Genotype-Phenotype Relations for the Parkinson's Disease Genes Parkin, PINK1, DJ1: MDSGene Systematic Review. Movement Disorders: Official Journal of the Movement Disorder Society. 2018 May;33(5):730–41. doi:10.1002/mds.27352 PubMed PMID: 29644727.

69. Kamienieva I, Duszyński J, Szczepanowska J. Multitasking guardian of mitochondrial quality: Parkin function and Parkinson's disease. Translational Neurodegeneration. 2021 Jan;10(1):5. doi:10.1186/s40035-020-00229-8 PubMed PMID: 33468256; PubMed Central PMCID: PMC7816312.

70. Pickrell AM, Youle RJ. The roles of PINK1, parkin, and mitochondrial fidelity in Parkinson's disease. Neuron. 2015 Jan;85(2):257–73. doi:10.1016/j.neuron.2014.12.007 PubMed PMID: 25611507; PubMed Central PMCID: PMC4764997.

71. Lücking CB, Abbas N, Dürr A, Bonifati V, Bonnet AM, de Broucker T, et al. Homozygous deletions in parkin gene in European and North African families with autosomal recessive juvenile parkinsonism. The European Consortium on Genetic Susceptibility in Parkinson's Disease and the French Parkinson's Disease Genetics Study Group. Lancet. 1998 Oct;352(9137):1355–6. doi:10.1016/s0140-6736(05)60746-5 PubMed PMID: 9802278.

72. Zhu W, Huang X, Yoon E, Bandres-Ciga S, Blauwendraat C, Billingsley KJ, et al. Heterozygous PRKN mutations are common but do not increase the risk of Parkinson's disease. Brain: A Journal of Neurology. 2022 Jun;145(6):2077–91.

doi:10.1093/brain/awab456 PubMed PMID: 35640906; PubMed Central PMCID: PMC9423714.

73. Valente EM, Abou-Sleiman PM, Caputo V, Muqit MMK, Harvey K, Gispert S, et al. Hereditary early-onset Parkinson's disease caused by mutations in PINK1. Science. 2004 May;304(5674):1158–60. doi:10.1126/science.1096284 PubMed PMID: 15087508.

74. Gonçalves FB, Morais VA. PINK1: A Bridge between Mitochondria and Parkinson's Disease. Life. 2021 Apr;11(5):371. doi:10.3390/life11050371 PubMed PMID: 33919398; PubMed Central PMCID: PMC8143285.

75. Ishihara-Paul L, Hulihan MM, Kachergus J, Upmanyu R, Warren L, Amouri R, et al. PINK1 mutations and parkinsonism. Neurology. 2008 Sep;71(12):896–902. doi:10.1212/01.wnl.0000323812.40708.1f PubMed PMID: 18685134; PubMed Central PMCID: PMC2676945.

76. van Duijn CM, Dekker MC, Bonifati V, Galjaard RJ, Houwing-Duistermaat JJ, Snijders PJ, et al. Park7, a novel locus for autosomal recessive early-onset parkinsonism, on chromosome 1p36. American Journal of Human Genetics. 2001 Sep;69(3):629–34. doi:10.1086/322996 PubMed PMID: 11462174; PubMed Central PMCID: PMC1235491.

77. Macedo MG, Anar B, Bronner IF, Cannella M, Squitieri F, Bonifati V, et al. The DJ-1L166P mutant protein associated with early onset Parkinson's disease is unstable and forms higher-order protein complexes. Human Molecular Genetics. 2003 Nov;12(21):2807–16. doi:10.1093/hmg/ddg304 PubMed PMID: 12952867.

78. Niki T, Endo J, Takahashi-Niki K, Yasuda T, Okamoto A, Saito Y, et al. DJ-1-binding compound B enhances Nrf2 activity through the PI3-kinase-Akt pathway by DJ-1-dependent inactivation of PTEN. Brain Research. 2020 Feb;1729:146641. doi:10.1016/j.brainres.2019.146641 PubMed PMID: 31891690.

79. Darvish H, Movafagh A, Omrani MD, Firouzabadi SG, Azargashb E, Jamshidi J, et al. Detection of copy number changes in genes associated with Parkinson's disease in Iranian patients. Neuroscience Letters. 2013 Sep;551:75–8. doi:10.1016/j.neulet.2013.07.013 PubMed PMID: 23880019.

80. Bonifati V, Rizzu P, van Baren MJ, Schaap O, Breedveld GJ, Krieger E, et al. Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science. 2003 Jan;299(5604):256–9. doi:10.1126/science.1077209 PubMed PMID: 12446870.

81. Zimprich A, Biskup S, Leitner P, Lichtner P, Farrer M, Lincoln S, et al. Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology.

Neuron. 2004 Nov;44(4):601–7. doi:10.1016/j.neuron.2004.11.005 PubMed PMID: 15541309.

82. Kalogeropulou AF, Purlyte E, Tonelli F, Lange SM, Wightman M, Prescott AR, et al. Impact of 100 LRRK2 variants linked to Parkinson's disease on kinase activity and microtubule binding. The Biochemical Journal. 2022 Sep;479(17):1759–83. doi:10.1042/BCJ20220161 PubMed PMID: 35950872; PubMed Central PMCID: PMC9472821.

83. Gilks WP, Abou-Sleiman PM, Gandhi S, Jain S, Singleton A, Lees AJ, et al. A common LRRK2 mutation in idiopathic Parkinson's disease. Lancet. 2005--4;365(9457):415–6. doi:10.1016/S0140-6736(05)17830-1 PubMed PMID: 15680457.

84. Healy DG, Falchi M, O'Sullivan SS, Bonifati V, Durr A, Bressman S, et al. Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: A case-control study. The Lancet Neurology. 2008 Jul;7(7):583–90. doi:10.1016/S1474-4422(08)70117-0 PubMed PMID: 18539534; PubMed Central PMCID: PMC2832754.

85. Bonet-Ponce L, Cookson MR. LRRK2 recruitment, activity, and function in organelles. The FEBS journal. 2022 Nov;289(22):6871–90. doi:10.1111/febs.16099 PubMed PMID: 34196120; PubMed Central PMCID: PMC8744135.

86. Tolosa E, Vila M, Klein C, Rascol O. LRRK2 in Parkinson disease: Challenges of clinical trials. Nature Reviews Neurology. 2020 Feb;16(2):97–107. doi:10.1038/s41582-019-0301-2 PubMed PMID: 31980808.

87. Ramirez A, Heimbach A, Gründemann J, Stiller B, Hampshire D, Cid LP, et al. Hereditary parkinsonism with dementia is caused by mutations in ATP13A2, encoding a lysosomal type 5 P-type ATPase. Nature Genetics. 2006 Oct;38(10):1184–91. doi:10.1038/ng1884 PubMed PMID: 16964263.

88. Williams DR, Hadeed A, al-Din ASN, Wreikat AL, Lees AJ. Kufor Rakeb disease: Autosomal recessive, levodopa-responsive parkinsonism with pyramidal degeneration, supranuclear gaze palsy, and dementia. Movement Disorders: Official Journal of the Movement Disorder Society. 2005 Oct;20(10):1264–71. doi:10.1002/mds.20511 PubMed PMID: 15986421.

89. Fujii T, Nagamori S, Wiriyasermkul P, Zheng S, Yago A, Shimizu T, et al. Parkinson's disease-associated ATP13A2/PARK9 functions as a lysosomal H+,K+-ATPase. Nature Communications. 2023 Apr;14(1):2174. doi:10.1038/s41467-023-37815-z PubMed PMID: 37080960; PubMed Central PMCID: PMC10119128.

90. Murphy KE, Cottle L, Gysbers AM, Cooper AA, Halliday GM. ATP13A2 (PARK9) protein levels are reduced in brain tissue of cases with Lewy bodies. Acta Neuropathologica

Communications. 2013 May;1(1):11. doi:10.1186/2051-5960-1-11 PubMed PMID: 24252509; PubMed Central PMCID: PMC4046687.

91. Paisán-Ruiz C, Guevara R, Federoff M, Hanagasi H, Sina F, Elahi E, et al. Early-onset L-dopa-responsive parkinsonism with pyramidal signs due to ATP13A2, PLA2G6, FBXO7 and spatacsin mutations. Movement Disorders: Official Journal of the Movement Disorder Society. 2010 Sep;25(12):1791–800. doi:10.1002/mds.23221 PubMed PMID: 20669327; PubMed Central PMCID: PMC6005705.

92. Shojaee S, Sina F, Banihosseini SS, Kazemi MH, Kalhor R, Shahidi GA, et al. Genome-wide linkage analysis of a Parkinsonian-pyramidal syndrome pedigree by 500 K SNP arrays. American Journal of Human Genetics. 2008 Jun;82(6):1375–84. doi:10.1016/j.ajhg.2008.05.005 PubMed PMID: 18513678; PubMed Central PMCID: PMC2427312.

93. Conedera S, Apaydin H, Li Y, Yoshino H, Ikeda A, Matsushima T, et al. FBXO7 mutations in Parkinson's disease and multiple system atrophy. Neurobiology of Aging. 2016 Apr;40:192.e1–5. doi:10.1016/j.neurobiolaging.2016.01.003 PubMed PMID: 26882974.

94. Zhao T, De Graaff E, Breedveld GJ, Loda A, Severijnen LA, Wouters CH, et al. Loss of nuclear activity of the FBXO7 protein in patients with parkinsonian-pyramidal syndrome (PARK15). PloS One. 2011 Feb;6(2):e16983. doi:10.1371/journal.pone.0016983 PubMed PMID: 21347293; PubMed Central PMCID: PMC3037939.

95. Nelson DE, Randle SJ, Laman H. Beyond ubiquitination: The atypical functions of Fbxo7 and other F-box proteins. Open Biology. 2013 Oct;3(10):130131. doi:10.1098/rsob.130131 PubMed PMID: 24107298; PubMed Central PMCID: PMC3814724.

96. Zhou ZD, Xie SP, Sathiyamoorthy S, Saw WT, Sing TY, Ng SH, et al. F-box protein 7 mutations promote protein aggregation in mitochondria and inhibit mitophagy. Human Molecular Genetics. 2015 Nov;24(22):6314–30. doi:10.1093/hmg/ddv340 PubMed PMID: 26310625.

97. Burchell VS, Nelson DE, Sanchez-Martinez A, Delgado-Camprubi M, Ivatt RM, Pogson JH, et al. The Parkinson's disease-linked proteins Fbxo7 and Parkin interact to mediate mitophagy. Nature Neuroscience. 2013 Sep;16(9):1257–65. doi:10.1038/nn.3489 PubMed PMID: 23933751; PubMed Central PMCID: PMC3827746.

98. Edvardson S, Cinnamon Y, Ta-Shma A, Shaag A, Yim YI, Zenvirt S, et al. A deleterious mutation in DNAJC6 encoding the neuronal-specific clathrin-uncoating co-chaperone auxilin, is associated with juvenile parkinsonism. PloS One.

2012;7(5):e36458. doi:10.1371/journal.pone.0036458 PubMed PMID: 22563501; PubMed Central PMCID: PMC3341348.

99. Wulansari N, Darsono WHW, Woo HJ, Chang MY, Kim J, Bae EJ, et al. Neurodevelopmental defects and neurodegenerative phenotypes in human brain organoids carrying Parkinson's disease-linked DNAJC6 mutations. Science Advances. 2021 Feb;7(8):eabb1540. doi:10.1126/sciadv.abb1540 PubMed PMID: 33597231; PubMed Central PMCID: PMC7888924.

100. Jacquemyn J, Kuenen S, Swerts J, Pavie B, Vijayan V, Kilic A, et al. Parkinsonism mutations in DNAJC6 cause lipid defects and neurodegeneration that are rescued by Synj1. NPJ Parkinson's disease. 2023 Feb;9(1):19. doi:10.1038/s41531-023-00459-3 PubMed PMID: 36739293; PubMed Central PMCID: PMC9899244.

101. Quadri M, Fang M, Picillo M, Olgiati S, Breedveld GJ, Graafland J, et al. Mutation in the SYNJ1 gene associated with autosomal recessive, early-onset Parkinsonism. Human Mutation. 2013 Sep;34(9):1208–15. doi:10.1002/humu.22373 PubMed PMID: 23804577.

102. Krebs CE, Karkheiran S, Powell JC, Cao M, Makarov V, Darvish H, et al. The Sac1 domain of SYNJ1 identified mutated in a family with early-onset progressive Parkinsonism with generalized seizures. Human Mutation. 2013 Sep;34(9):1200–7. doi:10.1002/humu.22372 PubMed PMID: 23804563; PubMed Central PMCID: PMC3790461.

103. Vanhauwaert R, Kuenen S, Masius R, Bademosi A, Manetsberger J, Schoovaerts N, et al. The SAC1 domain in synaptojanin is required for autophagosome maturation at presynaptic terminals. The EMBO journal. 2017 May;36(10):1392–411. doi:10.15252/embj.201695773 PubMed PMID: 28331029; PubMed Central PMCID: PMC5430236.

104. Olgiati S, De Rosa A, Quadri M, Criscuolo C, Breedveld GJ, Picillo M, et al. PARK20 caused by SYNJ1 homozygous Arg258Gln mutation in a new Italian family. Neurogenetics. 2014 Aug;15(3):183–8. doi:10.1007/s10048-014-0406-0 PubMed PMID: 24816432.

105. Lesage S, Drouet V, Majounie E, Deramecourt V, Jacoupy M, Nicolas A, et al. Loss of VPS13C Function in Autosomal-Recessive Parkinsonism Causes Mitochondrial Dysfunction and Increases PINK1/Parkin-Dependent Mitophagy. American Journal of Human Genetics. 2016 Mar;98(3):500–13. doi:10.1016/j.ajhg.2016.01.014 PubMed PMID: 26942284; PubMed Central PMCID: PMC4800038.

106. Kumar N, Leonzino M, Hancock-Cerutti W, Horenkamp FA, Li P, Lees JA, et al. VPS13A and VPS13C are lipid transport proteins differentially localized at ER contact sites. The Journal of Cell Biology. 2018 Oct;217(10):3625–39.

doi:10.1083/jcb.201807019 PubMed PMID: 30093493; PubMed Central PMCID: PMC6168267.

107. Cai S, Wu Y, Guillén-Samander A, Hancock-Cerutti W, Liu J, De Camilli P. In situ architecture of the lipid transport protein VPS13C at ER-lysosome membrane contacts. Proceedings of the National Academy of Sciences of the United States of America. 2022 Jul;119(29):e2203769119. doi:10.1073/pnas.2203769119 PubMed PMID: 35858323; PubMed Central PMCID: PMC9303930.

108. Hancock-Cerutti W, Wu Z, Xu P, Yadavalli N, Leonzino M, Tharkeshwar AK, et al. ER-lysosome lipid transfer protein VPS13C/PARK23 prevents aberrant mtDNA-dependent STING signaling. The Journal of Cell Biology. 2022 Jul;221(7):e202106046. doi:10.1083/jcb.202106046 PubMed PMID: 35657605; PubMed Central PMCID: PMC9170524.

109. Smolders S, Philtjens S, Crosiers D, Sieben A, Hens E, Heeman B, et al. Contribution of rare homozygous and compound heterozygous VPS13C missense mutations to dementia with Lewy bodies and Parkinson's disease. Acta Neuropathologica Communications. 2021 Feb;9(1):25. doi:10.1186/s40478-021-01121-w PubMed PMID: 33579389; PubMed Central PMCID: PMC7881566.

110. Li W, Fu Y, Halliday GM, Sue CM. PARK Genes Link Mitochondrial Dysfunction and Alpha-Synuclein Pathology in Sporadic Parkinson's Disease. Frontiers in Cell and Developmental Biology. 2021;9:612476. doi:10.3389/fcell.2021.612476 PubMed PMID: 34295884; PubMed Central PMCID: PMC8291125.

111. Aharon-Peretz J, Rosenbaum H, Gershoni-Baruch R. Mutations in the glucocerebrosidase gene and Parkinson's disease in Ashkenazi Jews. The New England Journal of Medicine. 2004 Nov;351(19):1972–7. doi:10.1056/NEJMoa033277 PubMed PMID: 15525722.

112. Gan-Or Z, Amshalom I, Kilarski LL, Bar-Shira A, Gana-Weisz M, Mirelman A, et al. Differential effects of severe vs mild GBA mutations on Parkinson disease. Neurology. 2015 Mar;84(9):880–7. doi:10.1212/WNL.0000000000001315 PubMed PMID: 25653295; PubMed Central PMCID: PMC4351661.

Chapter 3: CRISPRi Perturb-seq identifies cell-type-specific signatures of PD risk genes

3.1 Introduction

Although studies of families with heritable Parkinson's disease (PD) have identified highly penetrant mutations in nearly two dozen genes which are sufficient to drive the disease[1], the genetic risk factors behind idiopathic cases that make up the majority of PD diagnoses are less well-defined. Recent genome wide association studies (GWAS) have nominated ~90 candidate risk loci which nominate over 300 genes for their potential role in driving the disease[2], but whether their effects converge with familial PD genes upon common biological pathways remains unresolved. The functional validation of GWAS-nominated genes has been largely precluded by the need for tools that allow investigation at scale in human cell systems with genetic control and in a cell-type specific manner. The work outlined in this chapter aims to bridge this gap by developing CRISPRi-edited human pluripotent stem cells (hPSCs) and performing a Perturb-Seq screen of PD risk genes in differentiated, disease-relevant cell types.

Among the modalities considered for gene perturbation, CRISPRi-mediated repression was selected for reasons outlined in §1.6: transcriptional repression more faithfully models the haploinsufficient and graded loss-of-function mechanisms common to PD risk alleles than does Cas9 knockout, and it avoids the p53-mediated DNA damage response that Cas9 cutting elicits in hPSCs[3,4]. This choice carries an interpretive caveat. For gain-of-function mechanisms in PD such as *LRRK2* G2019S kinase hyperactivation or dosage-driven pathology for *SNCA*, repression has a functionally inverse effect to that of the mutant allele. Despite this, signatures for such examples can still be informative. Even where CRISPRi knockdown does not model the direction of effect for a disease allele, perturbation still informs the cellular processes through which disease alleles act, even if the direction of effect is reversed. For GWAS-nominated genes, the causal gene and direction of regulatory effect are often unresolved; CRISPRi tests the consequence of reduced expression of the candidate gene, leaving alleles that act through increased expression, altered splicing, or post-transcriptional effects outside the scope of this screen.

Two hypotheses organize the analyses presented in this chapter. First, mechanistically distinct PD risk genes from both the familial and GWAS-nominated arms converge on a small number of shared transcriptional programs in human dopaminergic (DA) neurons. Second, the transcriptional consequence of perturbing a given PD risk gene is cell-type- and cell-state-specific, with the same perturbation capable of producing qualitatively different responses in mature neurons versus progenitor populations.

To investigate these hypotheses, we performed CRISPRi Perturb-seq in hPSC-derived midbrain cell types. Two experiments were conducted: the first (C12-FPD-24) targeted 41

genes, comprising familial PD genes alongside additional genes of interest; the second (C12-GWAS-24) targeted 310 candidate risk genes nominated by the Nalls 2019 GWAS meta-analysis[2]. In both experiments, gRNA-transduced hPSCs were differentiated to day 35 DA neurons before Cre-mediated activation of CRISPRi, restricting perturbation to a defined ten-day window. Cells were harvested at day 45 and underwent fluorescence-activated cell sorting (FACS) to enrich for perturbed cells, before being assayed via single-cell RNA-sequencing (scRNA-Seq) with the 10x Genomics platform. In the resulting datasets, experiments yielded 23,389 singlets and 62,245 singlets for the FPD and GWAS experiments respectively. Differential expression (DE) was computed using Mixscale, a tool developed for CRISPRi perturbations which leverages heterogeneity in gRNA performance to apply weights in DE testing and prioritize gRNAs which have a larger effect on the transcriptome and deprioritize poor performers[5].

In this chapter, Section 3.2 describes the establishment and validation of the CRISPRi Perturb-seq system. Section 3.3 characterizes the familial-PD dataset, including cluster identity, perturbation assignment, and the analytical framework used thereafter. Section 3.4 presents perturbation signatures for four familial-PD genes of particular interest (*PARK7*, *VPS35, FBXO7*, *VPS13C*), followed by abridged results for additional noteworthy perturbations. Section 3.5 reports on the GWAS experiment, focusing on signatures from two perturbations (*SYT4*, and *ATP6V0A1*). Section 3.6 synthesizes convergent biological themes across both experiments and Sections 3.7 and 3.8 address study limitations and implications for future work.

3.2 Establishing a Perturb-seq system in hPSC-derived dopaminergic neurons.

3.2.1 Generation and validation of CRISPRi hPSC lines

To establish a system for the perturbation of genes associated with Parkinson's disease (PD) in disease-relevant cell types, we generated human pluripotent stem cells (hPSCs) with targeted integration of CRISPRi machinery into the AAVS1 safe-harbor locus. Temporal control over perturbation induction is achieved by incorporating a loxP-flanked STOP cassette within the transgene, ensuring that the conditional construct (dCas9-KRAB-P2A-mCherry) is only expressed once cells are exposed to Cre recombinase (Figure 3.1).

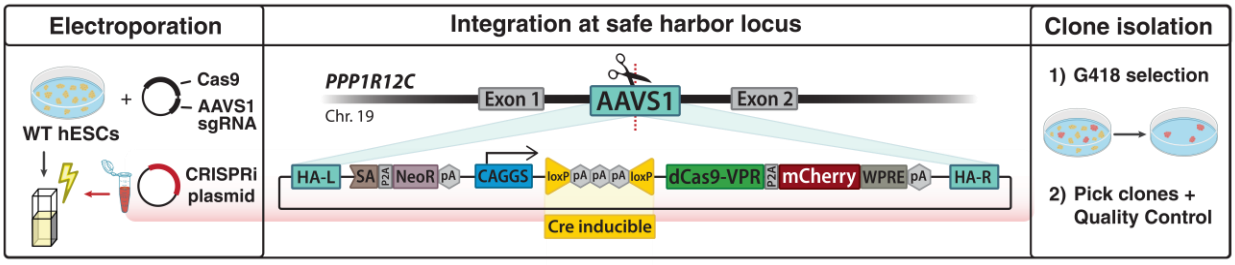


Figure 3.1 - *Schematic of CRISPRi WIBR3 hPSC generation. Wild-type WIBR3 hPSCs were electroporated with two plasmids; one encoding Cas9 and a gRNA targeting the AAVS1 safe harbor locus, the other an HDR donor plasmid carrying a dCas9-KRAB-P2A-mCherry expression cassette flanked by AAVS1 homology arms. Cre-mediated excision of the floxed-STOP cassette is required for expression of dCas9-KRAB-P2A-mCherry. Clones with successful integration are neomycin resistant, as conferred by a splice-acceptor-NeoR cassette.*

Following neomycin selection to enrich for cells with successful AAVS1 integration, targeted clones were subcloned and subjected to additional quality control. Proper AAVS1 integration was determined by junction PCRs spanning both homology arms (Figure 3.2). One CRISPRi clone, "C12" was selected and subcloned once more to ensure monoclonality. This cell line retained pluripotency, as evidenced by stemness markers OCT4 and SSEA4, alkaline phosphatase staining, and characteristic stem cell morphology (Figure 3.3).

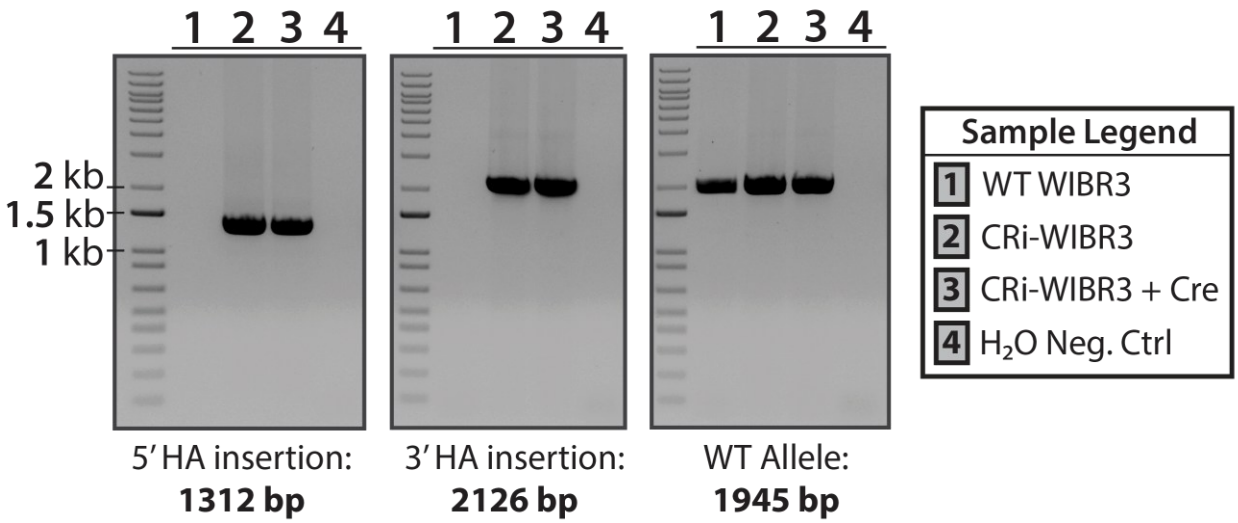


Figure 3.2 – PCR genotyping of transgene insertion of the inducible CRISPRi expression cassette into the AAVS1 safe harbor locus. **LEFT**: Inside-outside PCR spanning the 5' AAVS1 homology arm. **MIDDLE**: Inside-outside PCR spanning the 3' AAVS1 homology arm. **RIGHT**: Outside-outside PCR spanning the WT AAVS1 locus

Next, we confirmed that Cre delivery induced expression of the dCas9-KRAB-P2A-mCherry cassette, as determined by Western blot for Cas9 and by mCherry fluorescence (Figure 3.4). Finally, to functionally validate inducible CRISPRi knockdown, we transduced CRISPRi hPSCs with a lentivirus conferring transgenic expression of a gRNA targeting the *B2M* gene and subsequently delivered Cre recombinase to a subset of cells. Ten days later, flow cytometry revealed that in cells having received Cre and B2M gRNA, but not in those with gRNA alone, there was a marked decrease in B2M protein on the surface of assayed

cells (Figure 3.5). In full, our CRISPRi hPSCs were determined to have Cre-inducible expression of CRISPRi machinery which can successfully effectuate transcriptional repression of targeted genes.

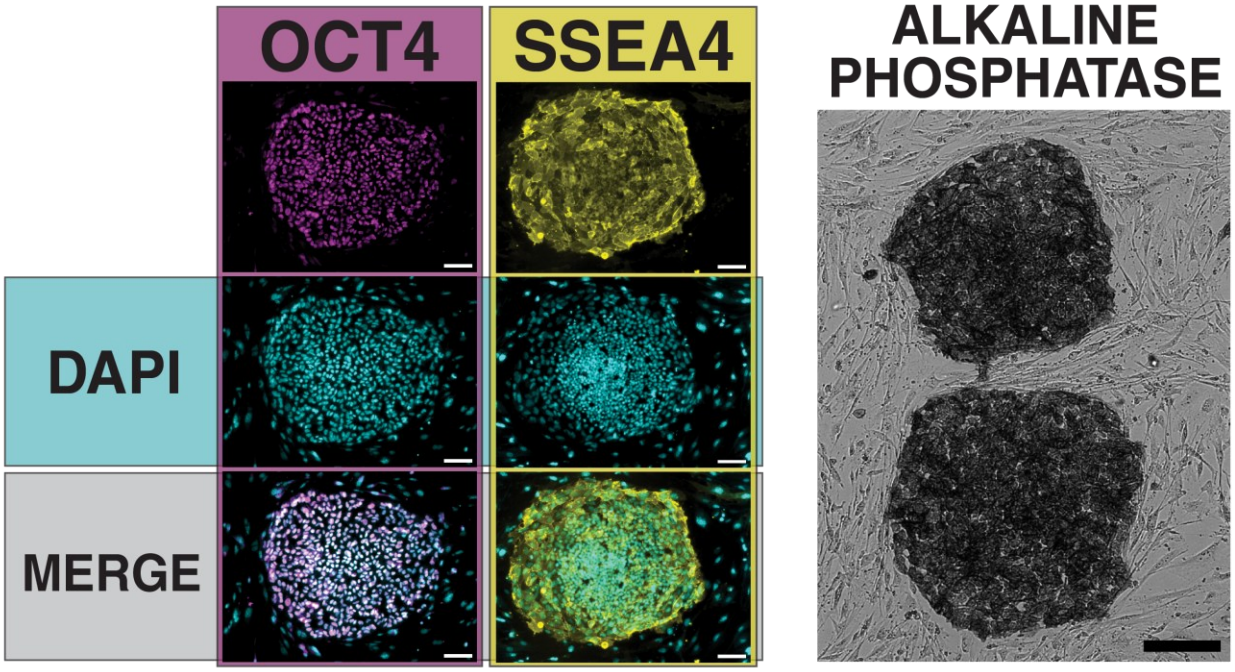


Figure 3.3 – LEFT: Fluorescence immunocytochemistry of CRISPRi WIBR3 hESCs cultured on mouse embryonic fibroblasts indicating expression of pluripotency markers OCT4 and SSEA4. Scale bar = 100 μM. **RIGHT**: Chromogenic immunocytochemistry of CRISPRi WIBR3 hESCs cultured on mouse embryonic fibroblasts, showing expression of the pluripotency marker alkaline phosphatase. Scale bar = 250 μM.

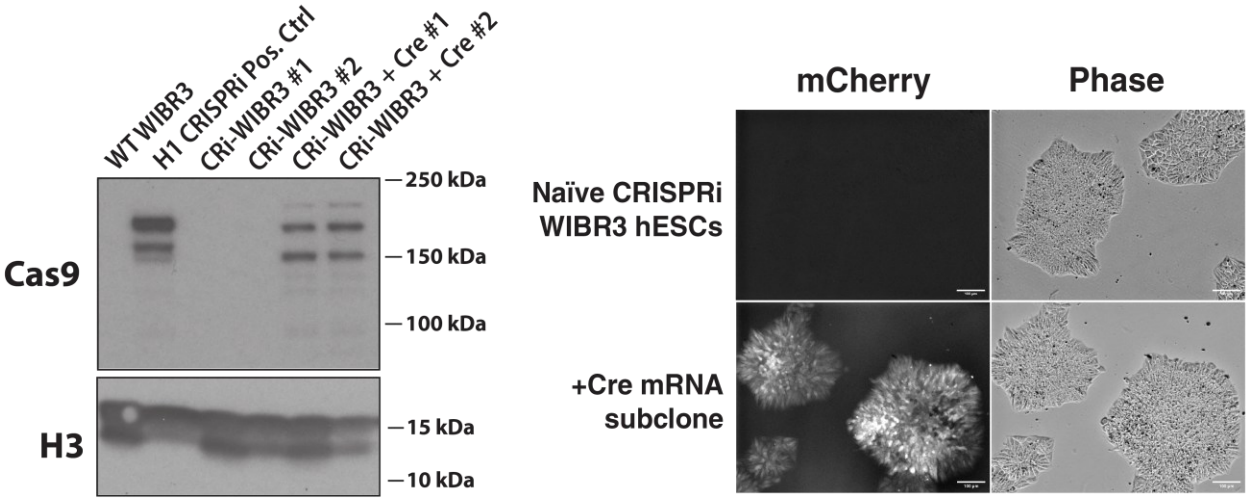


Figure 3.4 – LEFT: Western blot indicating Cre-conditional expression of CRISPRi machinery. Naïve CRISPRi WIBR3 hESCs do not express Cas9 protein (2 replicates, lane 3 + 4), but following exposure to Cre mRNA and isolation of an mCherry-positive subclone, are positive for Cas9 protein, indicating successful induction of the dCas9-KRAB-P2A-mCherry cassette. **RIGHT**: Fluorescence micrograph of CRISPRi WIBR3 hESCs adapted to feeder-free culture. Naïve cells do not express mCherry, but following exposure to Cre mRNA and isolation of a fluorescent subclone, show robust expression of mCherry. Scale bar = 100 μM.

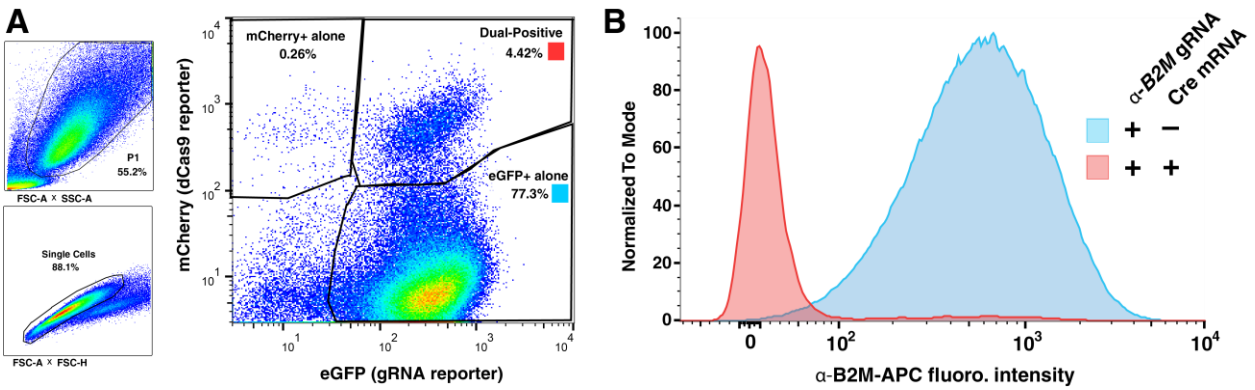


Figure 3.5– *Flow cytometry data showing Cre-inducible knockdown of the cell surface protein B2M in hESCs.* ***(A)*** *– Gating strategy used to define populations of interest based on eGFP intensity (reporter for the lentivirus-delivered B2M gRNA) and mCherry (reporter for Cre-inducible expression of dCas9-KRAB CRISPRi machinery). Approximately 4.7% of hESCs sparsely transfected with Cre mRNA appear to have successfully induced expression of dCas9-KRAB and the mCherry reporter.* ***(B)*** *Histogram of fluorescence intensity from the fluorophore-conjugated B2M antibody, shown for two populations of cells in the same sample shown in (A). Dual-positive cells which received a B2M gRNA and which express CRISPRi machinery show a marked decrease in B2M-antibody fluorescence compared to cells with B2M gRNA alone (no Cre), indicating a depletion of B2M protein on the surface of perturbed hESCs.*

3.2.2 Design, production, and delivery of familial-PD and GWAS-PD gRNA libraries

We designed two gRNA libraries for use with our CRISPRi hPSCs. The familial-PD library (FPD) targeted 41 genes, consisting of familial PD genes, genes with putative associations with familial PD, as well as midbrain transcription factors and other genes of interest, each with four independent gRNAs per transcription start site for each perturbation target, plus five non-targeting control gRNAs[6]. Non-targeting controls by design have no predicted genomic target and serve as the reference population for all downstream differential expression testing. The GWAS library targeted 310 candidate risk genes nominated by the Nalls 2019 meta-analysis[2], using the same four-guides-per-TSS design plus negative controls. Both libraries were cloned into a lentiviral vector designed for use with 10x Genomics reagents[7], with minor modifications to prevent Cre-loopout of the gRNA expression cassette itself (see methods). In addition to conferring mU6-mediated gRNA expression, this vector also expresses PURO-T2A-eGFP, allowing for antibiotic selection and expression of a fluorescent reporter (Figure 3.6). Following validation of the cloned plasmid libraries via next-generation sequencing (NGS), CRISPRi hPSCs were transduced at low multiplicity of infection (MOI = 0.3) to ensure that most cells received a single gRNA before selection with puromycin to enrich for successfully transduced cells.

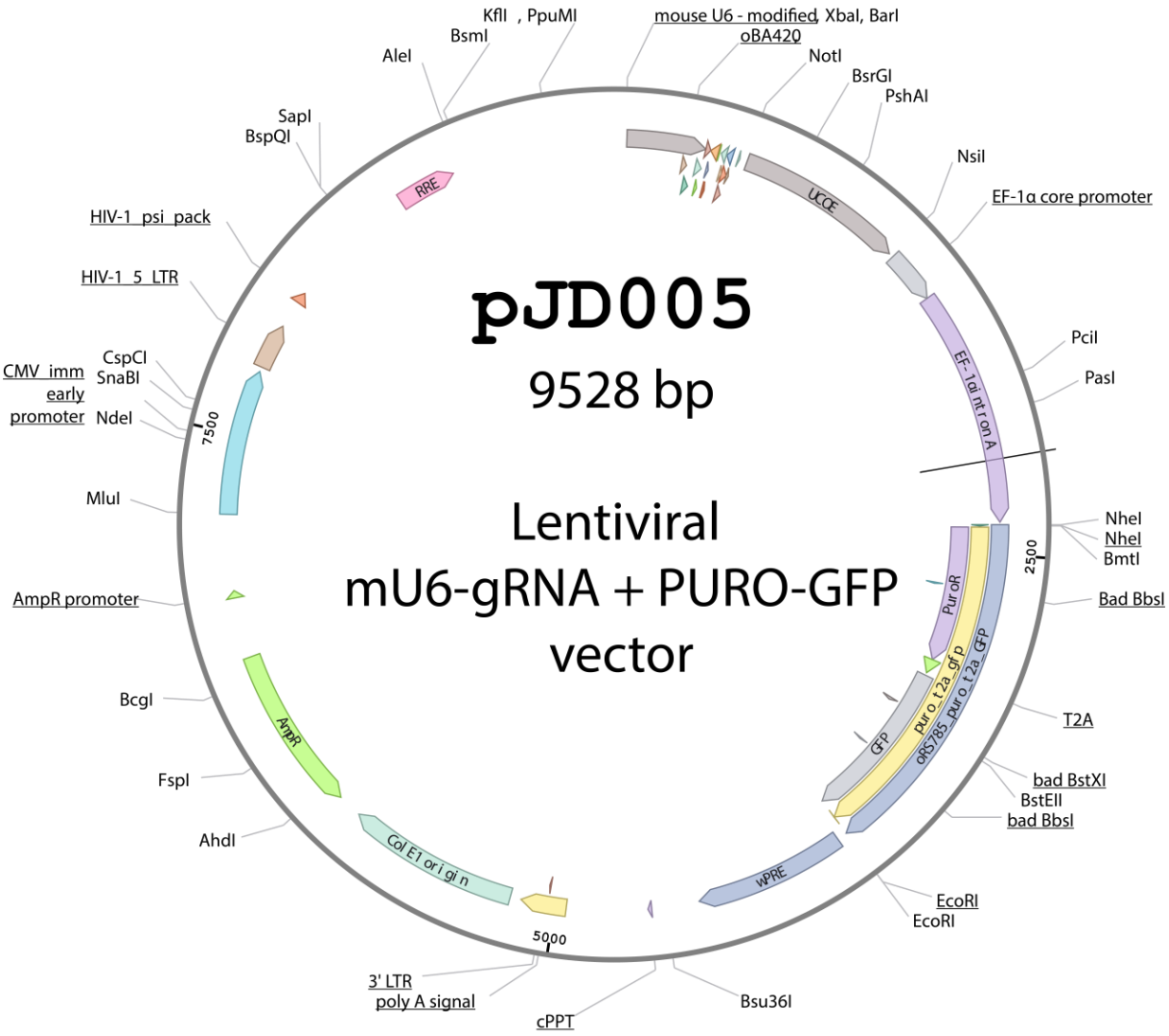


Figure 3.6 – *Plasmid map of the lentiviral gRNA vector used in our Perturb-seq experiments. Adapted from pJR101[99] to delete internal LoxP sites, this vector delivers eGFP and PURO resistance alongside an mU6 driven gRNA cassette. Libraries are cloned into this vector using the restriction enzymes BstXI and BlpI.*

3.2.3 DA neuron differentiation, perturbation induction, and scRNA-Seq

We differentiated gRNA-transduced CRISPRi hPSCs into midbrain DA neurons using an established protocol[8] (Figure 3.7) On day 35 of differentiation, cultures showed clear expression of midbrain floor-plate marker FOXA2 and canonical DA neuron marker TH (Figure 3.8). Cultures were transduced with Cre lentivirus at day 35 to drive expression of the dCas9-KRAB-P2A-mCherry cassette, which was confirmed by the presence of sparse mCherry expression 3 days after delivery of Cre. At day 45, following 10 days of CRISPRi perturbation, cells were dissociated and FACS-sorted to isolate mCherry+ (CRISPRi-induced) and GFP+ (gRNA-expressing) double-positive cells. Sorted cells were assayed using the 10x Genomics Chromium platform (v3.1 with Feature Barcoding technology for CRISPR Screening) to generate single-cell gene expression and gRNA libraries. Sequencing was performed on Illumina NovaSeq X, following 10x recommendations.

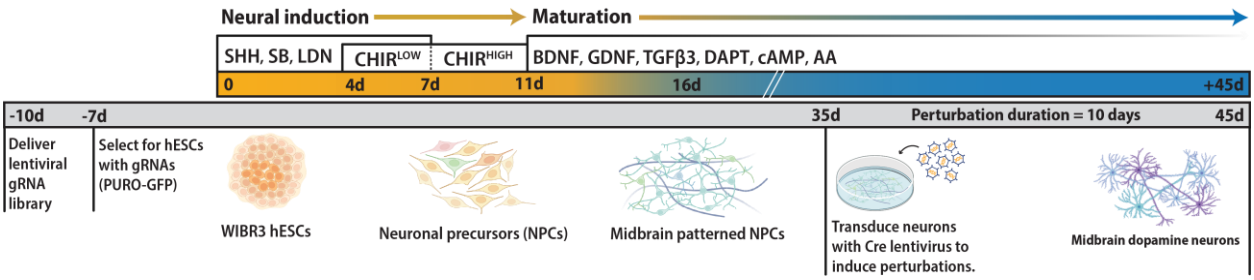


Figure 3.7 – *A schematic outlining the differentiation protocol (upper timeline) and CRISPRi perturbation strategy (lower timeline) used in the Perturb-seq experiments from this chapter. See methods for exact details on media composition, lentiviral gRNA library delivery, or single-cell RNA-seq assay.*

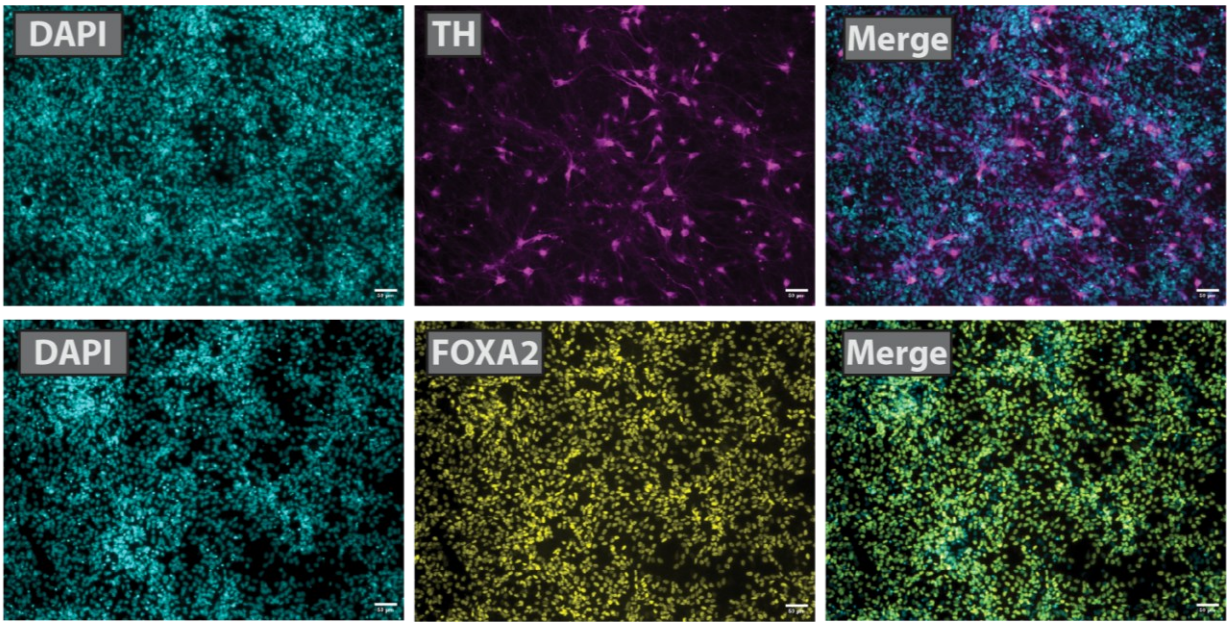


Figure 3.8 – *Immunocytochemistry of midbrain-specified dopaminergic neuron cultures at day 35 of in vitro differentiation. Top row: staining for the dopaminergic neuron marker tyrosine hydroxylase (TH) with nuclei labeled with DAPI. Bottom row: a separate well of the same differentiation showing nuclear expression of the midbrain floorplate marker FOXA2, which colocalizes with DAPI. Scale bar for all images: 50 μM.*

3.3 Characterization of the familial-PD CRISPRi dataset

3.3.1 Dataset generation and quality-control metrics

Reads were mapped with CellRanger v8.0.1 (10x Genomics) and count matrices were initialized in R using Seurat v5[9]. Across six 10x Chromium lanes, CellRanger identified 63,151 cells. Because CellRanger's feature-barcode algorithm showed limited sensitivity for gRNA detection, we reprocessed the gRNA count matrices with scMAGeCK[10] to assign perturbation identities. Of the 63,151 cells assayed, 26,757 (42.4%) were assigned a single gRNA, 8,028 (12.7%) received 2 or more gRNAs, and the remaining 28,366 cells (44.9%) could not be confidently assigned a gRNA perturbation (Figure 3.9). After doublet removal with DoubletFinder[11] and trimming of cells in the top and bottom 1.5% of nCount_RNA and nFeature_RNA, the final dataset consisted of 23,389 perturbed cells.

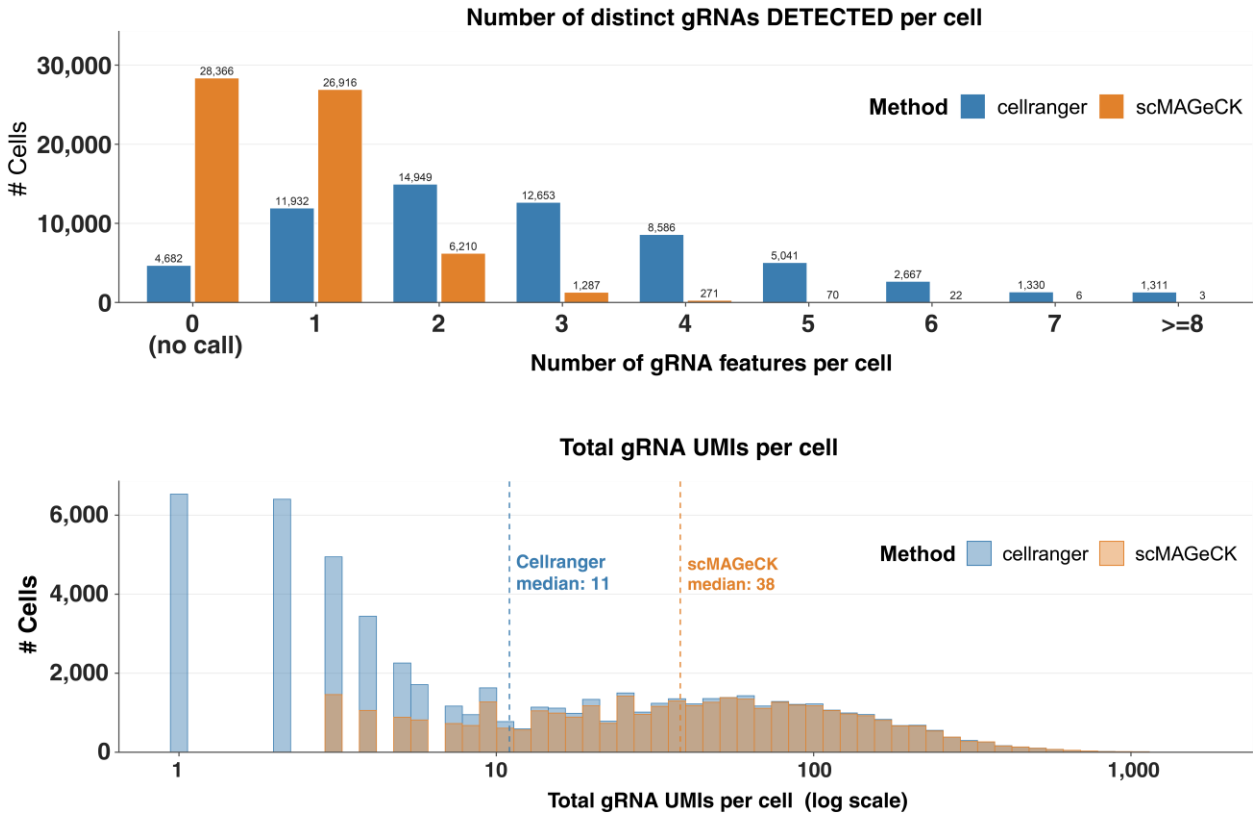


Figure 3.9 – *Quality control plots that helped inform the choice to use the barcode calling algorithm from scMAGeCK rather than Cellranger for assigning perturbation identities based on the number of unique gRNA molecules sequenced within each cell. TOP – the number of unique gRNA species confidently detected above background noise using either the cellranger (blue) or the scMAGeCK (orange) assignment algorithm. BOTTOM- distribution of the number of unique gRNA molecules captured and sequenced within each cell.*

Resulting cells were normalized, integrated, and clustered following standard Seurat workflows (see methods), and clustering at resolution 0.15 resolved six clusters (Figure 3.10B). Both components of the CRISPRi system were detected across all clusters by

reverse transcription and sequencing of polyadenylated mRNA from either transgene (Figure 3.10C). Read counts from the PURO-eGFP reporter delivered by the lentiviral gRNA vector and the dCas9-KRAB-P2A-mCherry cassette were detected in 17% and 45% of cells, respectively. Although only 9.6% of all cells had non-zero expression for both transgenes, these rates are consistent with the sparsity inherent in scRNA-seq data. Our use of negative controls to define FACS gates and the universal presence of direct-capture gRNA counts provided additional confidence that all cells our dataset had in fact experienced CRISPRi perturbation.

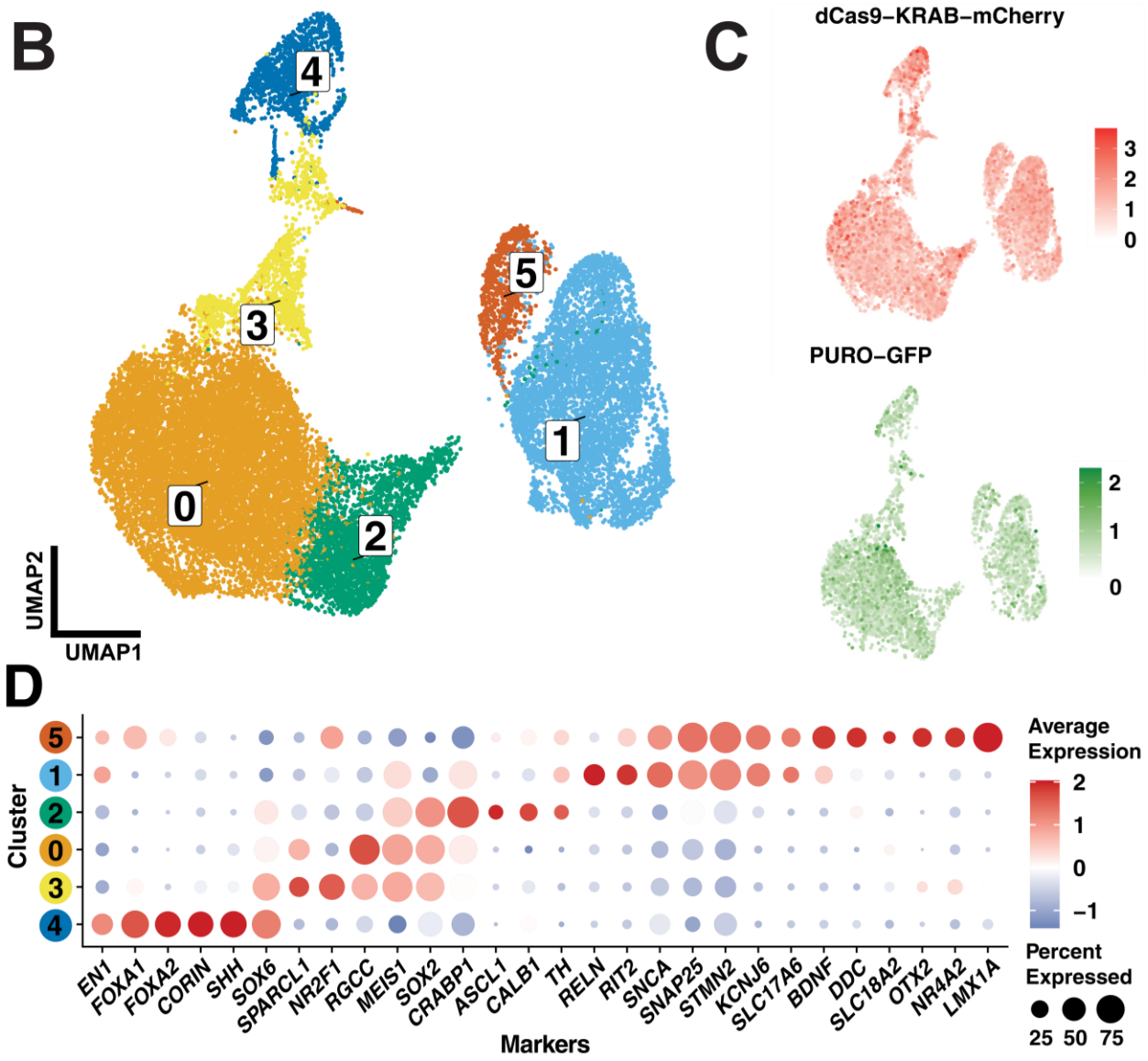


Figure 3.10 – ***(B)*** *UMAP reduction depicting all 23,389 singlets from the familial-PD Perturb-seq experiments. Clustering identifies 6 cell-type clusters, representing midbrain specified neurons (clusters 1 and 5), among other progenitor cell-types in other clusters.* ***(C)*** *Feature plots showing expression of the 2 transgenes in the experiment required for Perturb-seq activity; the dCas9-KRAB-P2A-mCherry cassette (TOP) and the PURO-eGFP reporter delivered as part of our lentiviral gRNA library vector.* ***(D)*** *Dotplot of marker gene expression across all clusters of the dataset. Left = midbrain progenitor markers, right = midbrain DA neuron markers.*

3.3.2 Cluster identity and reference-dataset comparison

Analysis of the marker genes defining each of the 6 clusters identified two neuronal populations and four progenitor clusters (Figure 3.10D). The mDA cluster (Cluster 5, 976 cells) expressed canonical midbrain DA markers (*TH, NR4A2, LMX1A, KCNJ6, DDC, SLC18A2, OTX2*), comprising our primary cell type of interest. A second, mixed-neuron cluster (Cluster 1, 5,724 cells) expressed several pan-neuronal markers (*SNAP25, STMN2, BDNF*) alongside a heterogeneous mixture of GABAergic (*GAD1*), glutamatergic (*SLC17A6*), and other neuronal markers (*RELN, RIT2*), indicating a mixture of neuron identities. Together, the two neuronal clusters comprised ~28.6% of all cells. The remaining clusters made up a progenitor gradient: an early midbrain progenitor cluster (Cluster 4 - *EN1, FOXA1/2, SOX6*), intermediate midbrain progenitors (Cluster 3 – *SOX6, NR2F1, RGCC*), a late midbrain progenitor cluster (Cluster 0 - *MEIS1, CRABP1, SOX2*), and finally a midbrain neuroblast cluster (Cluster 2 - *ASCL1, CALB1, TH*).

To evaluate our manually curated cell-type labels, we benchmarked our data against multiple reference datasets by pseudobulk Spearman correlation (Figure 3.11). The FOUNDIN-PD atlas[12] is the largest published collection of iPSC-derived midbrain DA neurons across multiple genetic backgrounds, many with mutations in PD risk genes. Our mDA and mixed-neuron clusters showed strong correlation with the iDA1 to iDA3 populations from the FOUNDIN-PD dataset (max Spearman rho= 0.80). Agarwal et al. profiled DA neurons from post-mortem human substantia nigra[13], both neuronal clusters correlated with their DA neuron clusters at max rho = 0.56. Kamath et al. resolved adult substantia nigra DA neurons into ten distinct subtypes including one *SOX6+/AGTR1+* subtype most depleted in PD[14]; our neuronal clusters showed moderate correlation with both *SOX6+* ventral and *CALB1+* dorsal DA populations. Neither of our clusters correlated strongly with the vulnerable *SOX6+/AGTR1+* subtype. Birtele et al.[15] and La Manno et al.[16] profiled human fetal midbrain; benchmarking against these atlases confirmed the midbrain floor-plate developmental origin and DA-lineage commitment of our neuronal clusters.

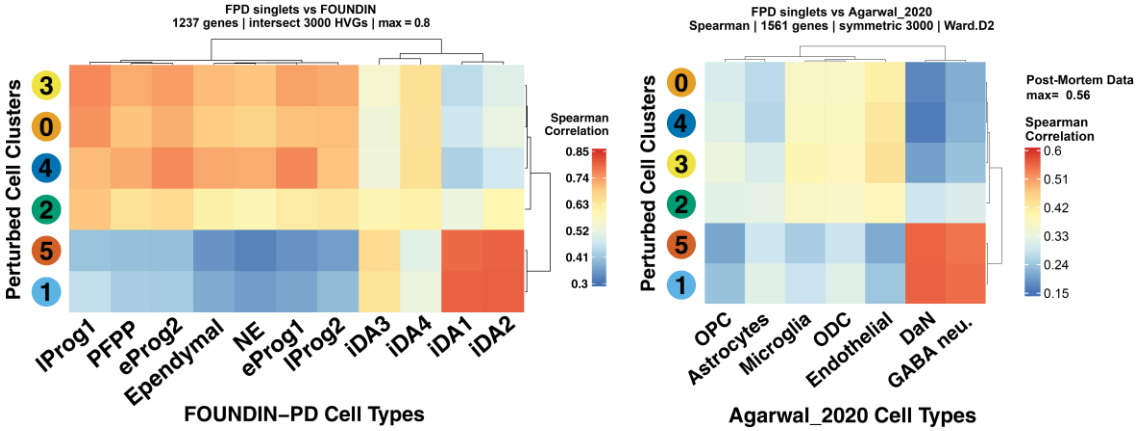


Figure 3.11 – *Heatmaps showing the spearman correlations between the pseudobulk expression profiles of the top 3000 highly-variable genes (HVGs) among all cells in each of our familial-PD Perturb-seq clusters (0 through 5), as compared to the pseudobulk expression profiles of the top 3000 HVGs for all cells within each of the reference dataset's assigned clusters/cell-types.* ***LEFT*** *– compared against all cell-types identified in the FOUNDIN-PD reference dataset[12], or* ***RIGHT*** *– compared to human post-mortem midbrain snRNA-Seq[13].*

3.3.3 Distribution of perturbations and evaluation of on-target knockdown

Per-perturbation cell counts varied across the dataset but did not show obvious cluster bias (Figure 3.12). In the pooled-neuron analysis combining both the mixed-neuron and mDA clusters (6,700 cells, 214 NTCs), the median number of cells per perturbation was 146, ranging from 38 cells for *NEUROG1*, up to 296 cells for *LRRK2*.

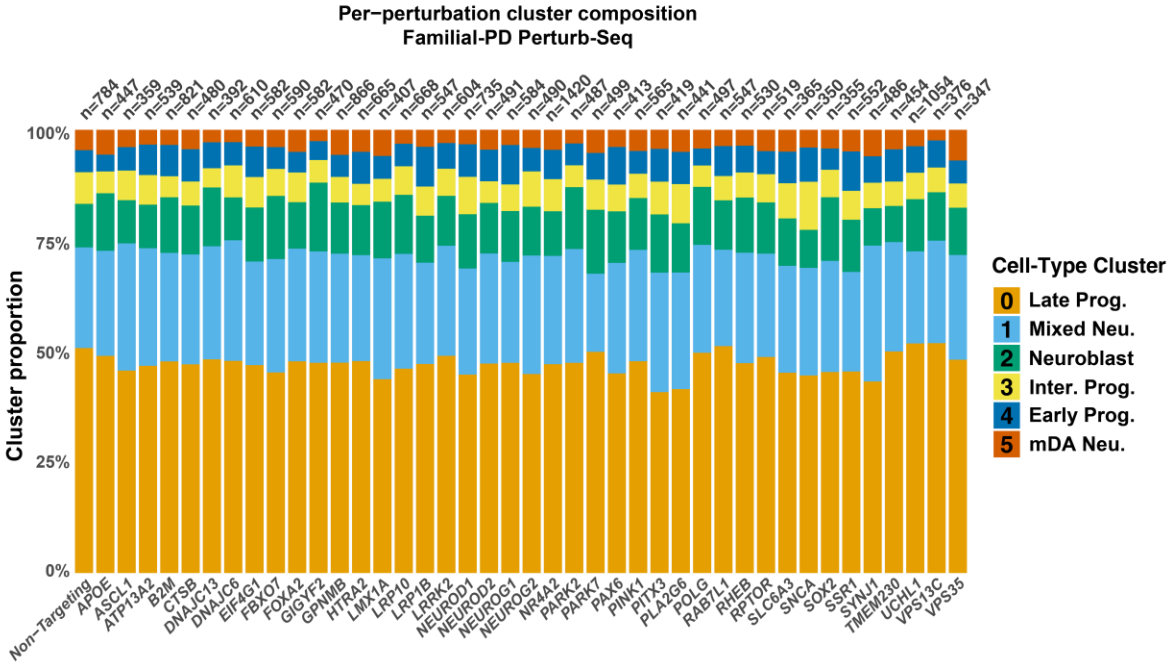


Figure 3.12 – Stacked bar chart showing the proportion of perturbed cells identified in each of the cell-type clusters of the Familial PD Perturb-seq experiment. Each bar represents the set of all cells having received gRNAs that targeting a particular gene for CRISPRi perturbation.

As an initial step in evaluating the functional efficacy of our CRISPRi system, we first identified the on-target effects of our perturbations. Working with the combined neuron clusters, we first grouped all cells based on the specific gRNA they received. Among these groups, we examined the average expression of the genes targeted in our familial-PD gRNA library, and quantified changes in expression as compared to the average of all 214 non-targeting control cells. These comparisons (Figure 3.13) confirmed that individual gRNAs produce on-target knockdown with heterogeneous effect; certain perturbations like *VPS35* had multiple gRNA species which strongly reduced expression of *VPS35* past 50% as compared to control. For other perturbations, like *GIGYF2*, only one of the 4 gRNAs appeared to reduce *GIGYF2* to a similar magnitude.

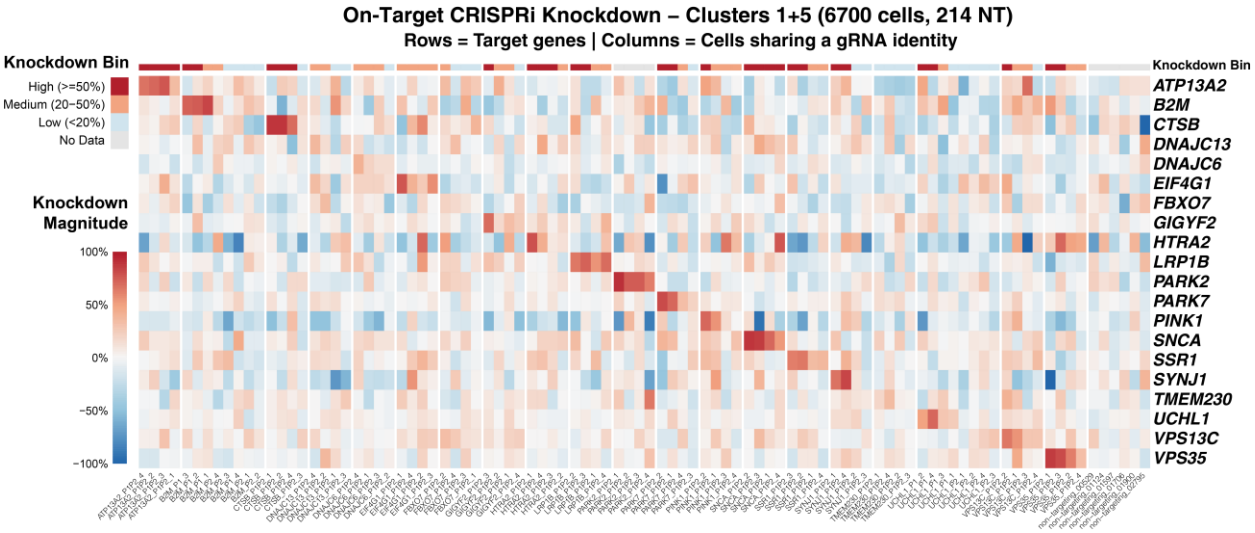


Figure 3.13 – *Heatmap showing the change in expression for a set of genes targeted in the familial-PD experiment (rows), as seen within all cells sharing a unique gRNA identity (columns) compared to non-targeting control cells, shown at right. On-target knockdown is evident on the diagonal, and this plot also demonstrates the heterogeneity in gRNA performances; some cells with a particular gRNA show strong knockdown for the targeted gene, while cells with a different gRNA targeting the same gene for CRISPRi may have minimal effect.*

3.3.4 Differential expression and pathway enrichment

Per-perturbation differential expression was computed using Mixscale[5], a CRISPRi-tailored framework that quantifies the magnitude of transcriptional perturbation on a per-cell basis and leverages these weights in differential expression testing. For each perturbed cell, Mixscale first builds a "perturbation signature" by subtracting the mean expression vector of the cell's N nearest non-targeting-control neighbors, removing technical and cell-state variation. Our analysis of combined neuron clusters 1 + 5 employed N=25 nearest neighbors, while all other cluster-specific analyses used N=20. From here, the framework then computes a "perturbation vector" to capture the average direction of transcriptional change across the full transcriptome, and projects each cell's perturbation signature onto that vector to yield a continuous Mixscale score, indicating how strongly a given cell's global transcriptome resembles the average perturbation response. Cells with stronger projection scores carry greater weight in the downstream weighted multivariate regression (wmvReg), which returns nominal weighted p-values for each tested gene. A Mixscale score ridgeplot stratified by perturbation and cluster (Figure 3.14) illustrates the range of per-cell scores across a subset of perturbations in the Familial-PD library. We applied Benjamini-Hochberg correction to the weighted p-values within each perturbation to control the false discovery rate (FDR), and we define differentially expressed genes (DEGs) throughout §3.4 and §3.5 by BH-adjusted $p < 0.05$ and $|\log2FC| > 0.25$.

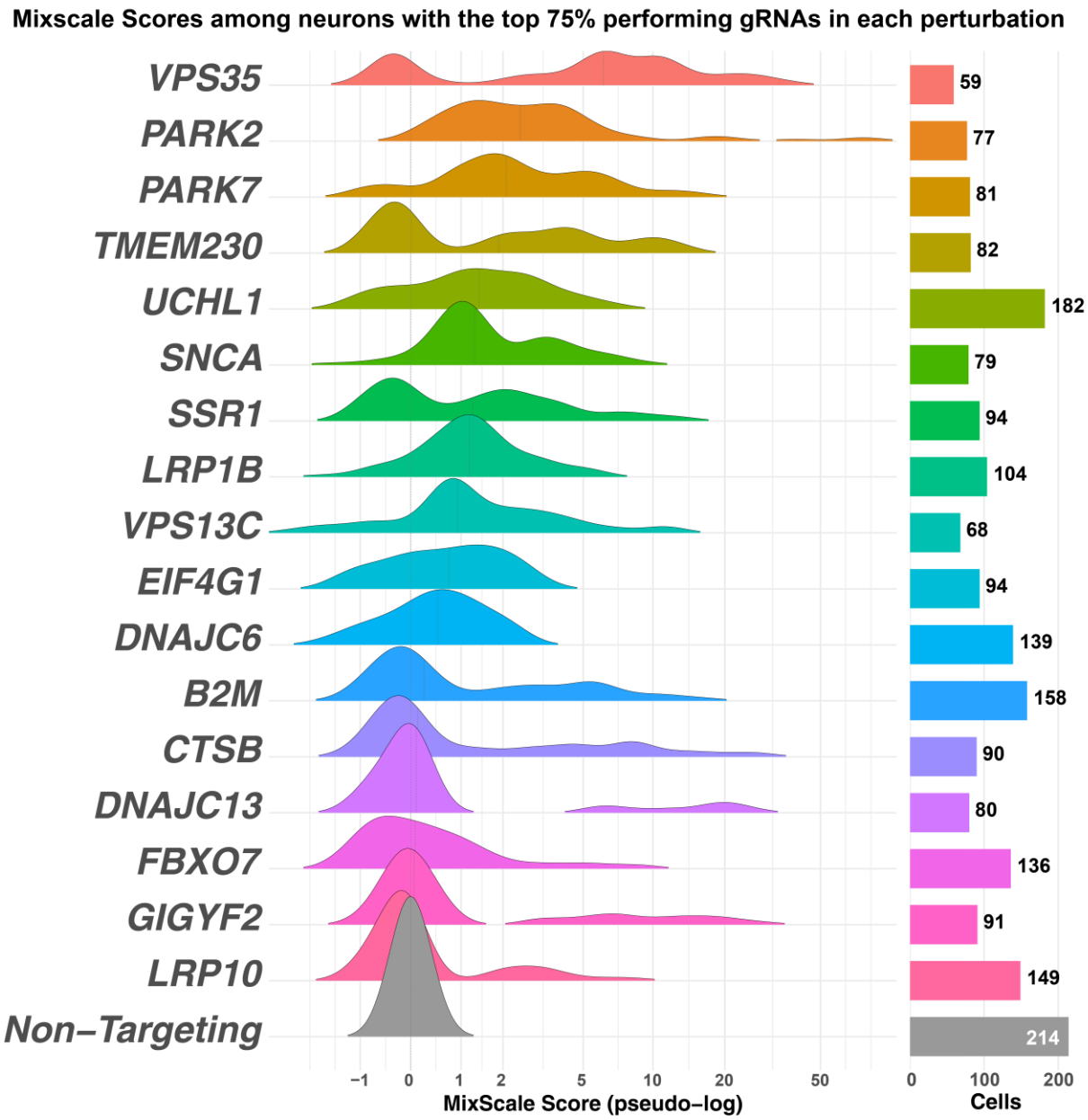


Figure 3.14 – *Ridgeplot depicting the distribution of Mixscale scores across a subset of perturbations in the neurons from the familial-PD experiment. Data shown are filtered to omit cells containing each perturbation's single "worst" gRNA, as determined by comparing median Mixscale score among cells from other gRNAs. Perturbations near top tend to have a greater Mixscale score among analyzed cells, indicating a larger magnitude in transcriptional perturbation as compared to non-targeting control cells shown in grey. Cell counts indicating the number of cells in each group post-filter are shown at right.*

Pre-ranked gene set enrichment analysis (GSEA) was computed against eight databases (GO Biological Process, GO Cellular Component, GO Molecular Function, KEGG pathways, KEGG modules, Reactome, WikiPathways, MSigDB Hallmark) using fgsea[17] and clusterProfiler[18], with genes ranked by sign(log2FC) × −log10(BH-adjusted p). For each perturbation, GSEA was run twice: once with the target gene retained in the ranked list (withSelf) and once with it excluded (noSelf). Rationale for this analysis is outlined in §3.4.1.

Applying this framework cross-cluster, knockdown efficiency was cluster-dependent (Figure 3.15): 13 FPD perturbations reached significant self-knockdown (BH-adjusted p < 0.05, log2FC < −0.25) in at least one cluster, including *PARK7, VPS35, VPS13C, SNCA, GIGYF2, DNAJC6,* and *B2M* in the pooled-neuron group, and *ATP13A2, EIF4G1, LRP10,* and *LRP1B* with their strongest signal in the late-progenitor cluster. The late-progenitor cluster exhibited the broadest coverage of significant self-knockdown events, reflecting its larger per-perturbation cell counts.

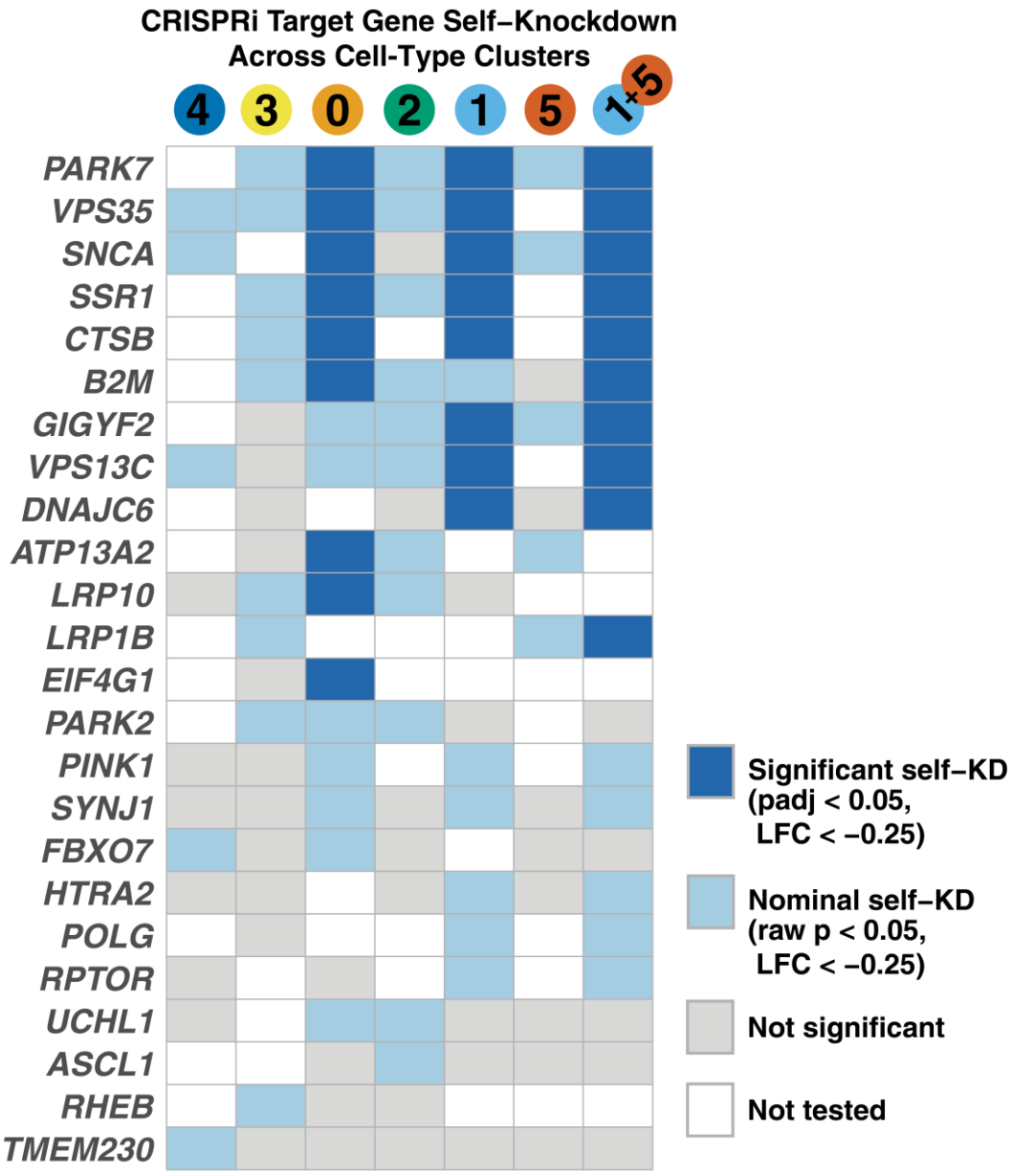


Figure 3.15 – *Heatmap indicating all perturbations where Mixscale identified >1 case of significant self-knockdown in any of the clusters analyzed in the familial-PD Perturb-seq experiment. Light blue indicates significant self-knockdown at LFC < 0.25 and nominal p-value < 0.05. Dark blue indicates cases of which remain significant after Benjamini-Hochberg FDR correction. Not tested indicates cases where a gene's sparsity or heterogeneity prevented Mixscale from performing DE tests for that gene.*

The absence of statistically significant self-knockdown for a given perturbation does not necessarily mean that CRISPRi failed to reduce target-gene expression. For genes with low basal expression in neurons (such as *FBXO7*, whose self-knockdown log2FC = −0.36 did not reach significance despite 174 perturbed cells), heterogeneous per-cell knockdown efficiency disperses the test statistic and reduces power. Such perturbations remain analyzable through their downstream transcriptional response, with interpretive caveats described in §3.4.1.

3.4 Mixscale reveals cell-type and perturbation-specific signatures of familial-PD genes

3.4.1 Analysis workflow and interpretive framework

In the familial PD Perturb-seq experiment, two factors ultimately informed the approach and interpretation of our results. First is the presence of two distinct neuron clusters in the data. As outlined in §3.3.2, we defined two clear neuron populations; first a well-defined mDA cluster (976 cells; *LMX1A*+/*NR4A2*+) and another with a mixture of varied neuron cell types (5,724 cells; a heterogeneous GABAergic/glutamatergic/*RELN*+/*RIT2*+). Given the transcriptional diversity that has been appreciated in SNc as well as adjacent midbrain regions[14,16], we initially intended to analyze perturbation signatures in each respective cluster. However, technical constraints attributed primarily to neuron differentiation efficiency and the cost/scale limitations of droplet-based scRNA-Seq experiments ultimately restricted the number of perturbed cells assayed in either neuron subtype. To illustrate, our mixed-neuron cluster had a median of 125 cells/perturbation, while our mDA cluster had just 21. In short, these factors greatly limited statistical power for individual neuron cluster analyses.

To address this constraint while still permitting an investigation of pan-neuronal effects of gene perturbations, we decided to focus our initial analyses on a combined, 6,700-cell “pooled-neuron” group, with a combined median of 146 cells-per-perturbation and a total of 214 non-targeting control (NTC) cells. Pooling yielded roughly four- to six-fold more cells per perturbation than either sub-cluster alone. This combination provided additional statistical power, thus allowing for the detection of significant differentially expressed genes and/or GSEA results for many of the familial-PD perturbations which otherwise failed to produce significant results when the neuron clusters were analyzed independently. Overviews of the number and directionality of differentially expressed genes for each perturbation analyzed in the combined neuron analysis is depicted in Figure 3.16, and a similar depiction of GSEA results is shown in Figure 3.17.

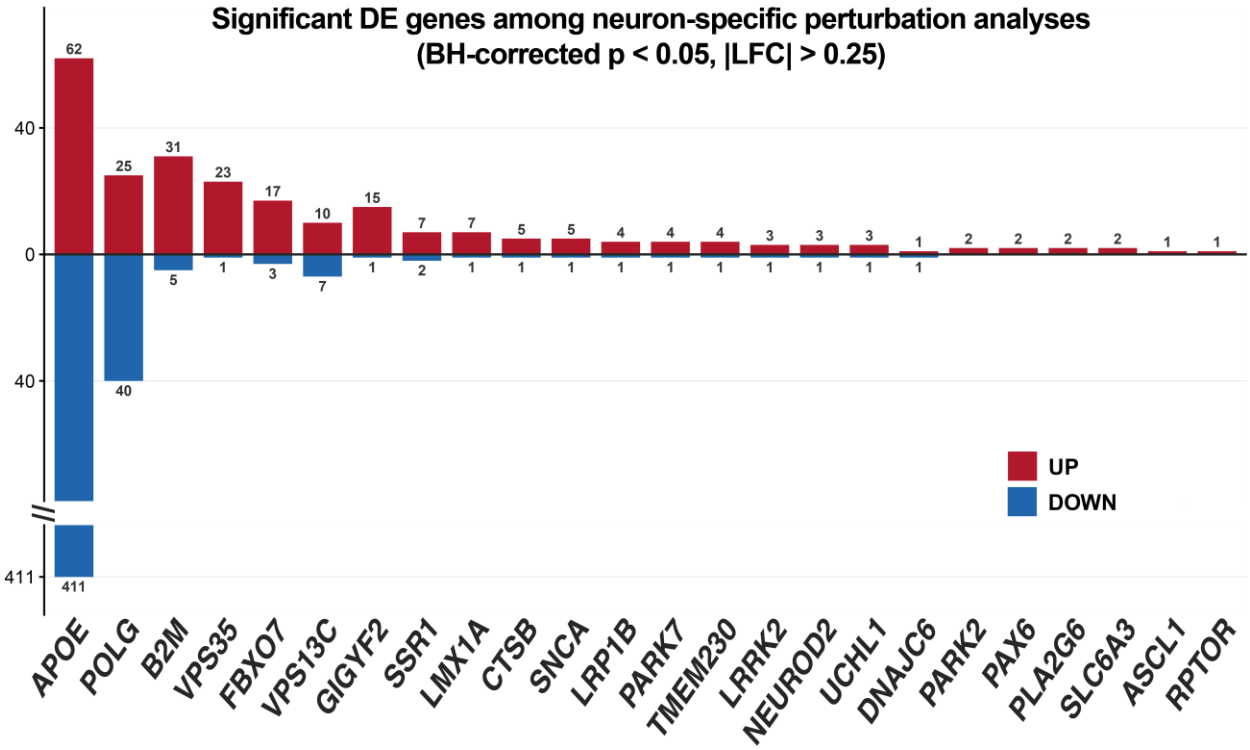


Figure 3.16 – Barplot depicting the number of significant DE genes (BH-corrected p < 0.05, |LFC| > 0.25) for each perturbation tested in the combined neuron analysis.

In addition to the pooled-neuron analyses mentioned above, we additionally investigated perturbation effects observed in additional non-neuronal clusters, most notably the late midbrain progenitors. This cluster was comparatively more populated, with a median of 249 cells for each perturbation. The late-progenitor analysis revealed cell-type-specific transcriptional responses for some perturbations and resolved on-target knockdown for others (notably *ATP13A2, EIF4G1, LRP10*) where neuron-scope cell counts were insufficient.

Per-perturbation differential expression used Mixscale weighted multivariate regression (wmvReg)[5] as described in §3.3.4, with Benjamini-Hochberg (BH) correction applied to DE results (DEGs) for each perturbation. DEG lists described in the following sections is defined with cutoffs of BH_pAdj < 0.05 and log2FC > |0.25|. Gene set enrichment analysis(GSEA) was performed against eight databases (GO BP/CC/MF, KEGG pathways, KEGG modules, Reactome, WikiPathways, MSigDB Hallmark) using fgsea[17] and clusterProfiler[18], with genes ranked by sign(log2FC) × −log10(BH-adjusted p). Where this chapter reports significant pathway counts for a perturbation, the count is the number of GSEA terms reaching BH-adjusted p < 0.05 within that perturbation, summed across the eight databases. Magnitude and direction of GSEA terms are indicated by the Normalized Enrichment Score, or NES. Leading edge genes, when mentioned, refer to the genes which contributed most to a given term's enrichment score.

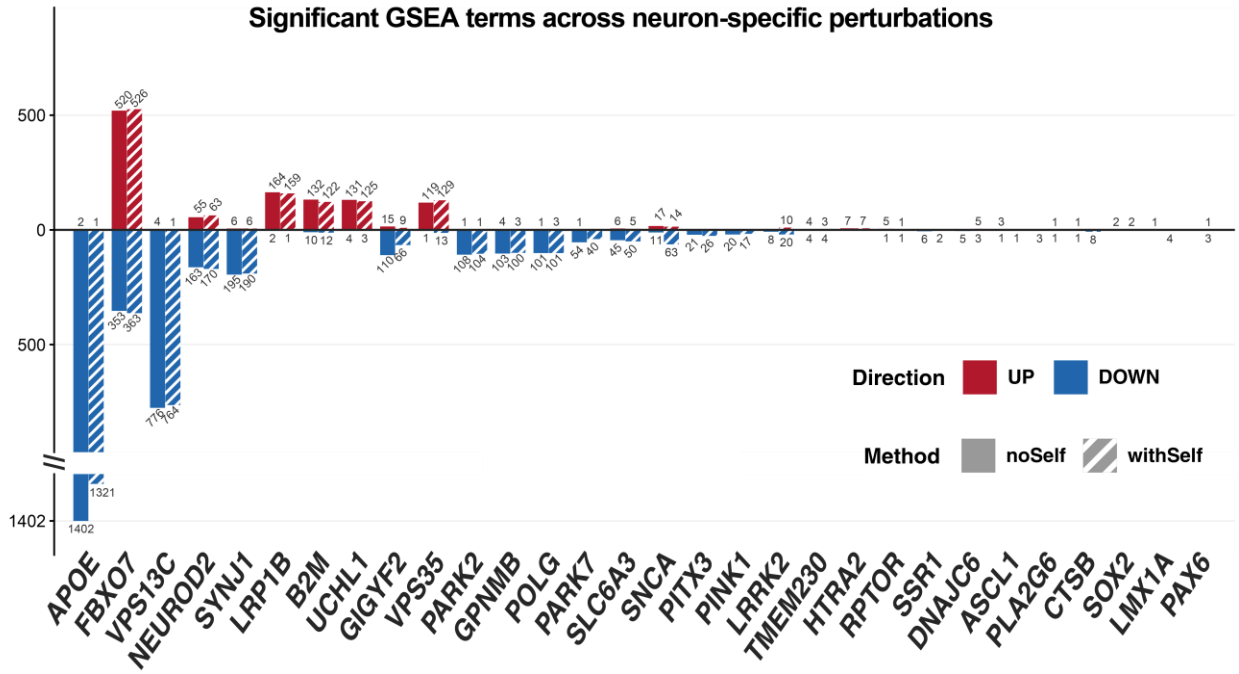


Figure 3.17 – Barplot depicting the number of significant GSEA terms (padj < 0.05) for each perturbation tested in the combined neuron analysis. Results from withSelf and noSelf results are depicted with hashed and solid bars, respectively.

A set of several hundred terms for a single perturbation is still difficult to interpret as biology in its own right. To facilitate biological interpretation, we manually defined twenty curated biological categories (e.g. mitochondrial function, lysosome/autophagy, protein homeostasis, synapse, or uncategorized) and assigned each significant term to the most applicable category by keyword matching on the term description, with manual review of edge cases.

One concern when running GSEA on a strongly-perturbed gene is that knockdown of the gene itself could exaggerate the enrichment signals, and that significant pathways may simply reflect annotations of the perturbed gene rather than the broader transcriptional response. To address this, GSEA analyses are run twice for each perturbation: once including the targeted gene kept in the ranked list (deemed “withSelf” analysis), and once more with the perturbed gene excluded (“noSelf” analysis). The noSelf analysis is meant to identify significant pathways that are driven by broader transcriptional changes rather than the on-target perturbation itself.

To categorize the changes in significant GSEA terms nominated by a perturbation’s withSelf/noSelf analysis, pathways may be referred in the following chapter sections as robust, self-driven, or emergent. A pathway is robust when it is found to be statistically significant in both noSelf and withSelf results, self-driven if the term is significant only in the withSelf approach, and emergent if only significant for noSelf. As a conservative approach to interpret the effect of perturbation on GSEA results, biological claims rest on the noSelf analysis; the withSelf analysis is reported only where the contrast between the two modes

is itself informative (§3.4.2, PARK7, is the clearest example in this chapter) or where the target gene is mechanistically central to the pathway in question.

3.4.2 *VPS35*

Background

VPS35 encodes the cargo-recognition subunit of the heterotrimeric mammalian retromer complex with *VPS26* and *VPS29*, which serves to facilitate recycling of transmembrane proteins in the endosome to the trans-Golgi network and plasma membrane[19,20]. The heterozygous D620N missense mutation in *VPS35* has been shown to cause autosomal-dominant Parkinson's disease[21,22]. This mutation causes partial loss-of-function and disrupts retromer-dependent retrieval of multiple cargoes, including the cation-independent mannose-6-phosphate receptor (CIMPR)[23], AMPA receptor trafficking[24], and the Wnt receptor Wntless[25]. Additionally, *VPS35* loss in dopamine neurons has been associated with mitochondrial fragmentation through *MUL1-MFN2* and *DLP1* dysregulation[26,27].

DE genes and GSEA overview

VPS35 perturbation in the pan-neuron analysis identified very significant self-knockdown within the 106 cells tested (log2FC = −1.223, BH-adjusted $p = 6.2 \times 10^{-10}$). Results included 24 DEGs, with *VPS35* itself being the sole downregulated gene. Among the upregulated DEGs were *SENP5,* known to control mitochondrial fission dynamics in neurons through deSUMOylation of *DRP1*[28,29]*, CRTC1*, a transcriptional coactivator of *CREB* which mediates PAK4-driven protection of DA neurons in rat models of PD[30], *NFE2L3* (*NRF3*) which regulates proteasome assembly and sterol biosynthesis[31–33], and the histone methyltransferases *SETD4* and *KMT2C* (Figure 3.18). A network plot of significant noSelf GSEA terms is shown in Figure 3.21.

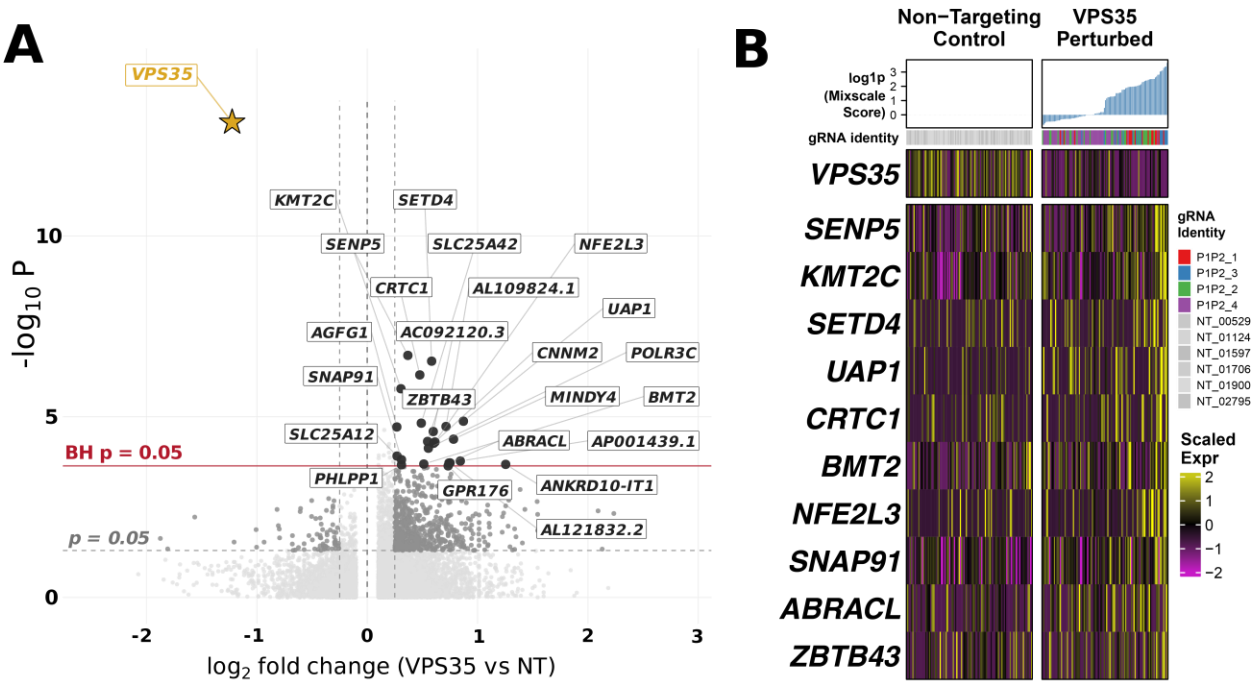


Figure 3.18 – *Overview of differentially expressed genes identified by Mixscale within VPS35-perturbed neurons from the familial-PD Perturb-seq experiment.* ***(A)*** *– Volcano plot indicating log2FC vs. p-value. Grey dotted horizontal line indicates a nominal p-value of 0.05, the red horizontal line denotes the Benjamini-Hochberg corrected pAdj of 0.05. Self knockdown of VPS35 is indicated with a gold star.* ***(B)*** *– Heatmap showing the expression of VPS35 itself and a selection of 10 significantly upregulated DEGs in VPS35 perturbed neurons. Columns at left represent non-targeting control cells, while columns on the right indicate all VPS35-perturbed cells, ranked from left-to-right by increasing Mixscale score.*

Upregulation of mitochondrial genes

A coherent set of terms significantly upregulated in *VPS35*-perturbed neurons centered on oxidative phosphorylation and mitochondrial biogenesis. Representative terms like **oxidative phosphorylation** (KEGG, NES = +1.60, padj = 8.8×10^{-3}) were driven by leading-edge genes which spanned the four respiratory complexes and ATP synthase, including *NDUFA13, NDUFS8, COX5A, ATP5F1A,* and *UQCRC2*. Related terms like **transcriptional activation of mitochondrial biogenesis** (Reactome, NES = +1.73, padj = 3.1×10^{-3}), included leading edge genes like *NCOA6, MED1, ATF2, SIRT5,* and *PPARGC1A* which encodes PGC-1α, a master regulator of mitochondrial biogenesis (Figure 3.19). The pairing of oxidative phosphorylation and mitochondrial biogenesis signatures suggest a coordinated upregulation of the mitochondrial gene-expression program in neurons with *VPS35* perturbation.

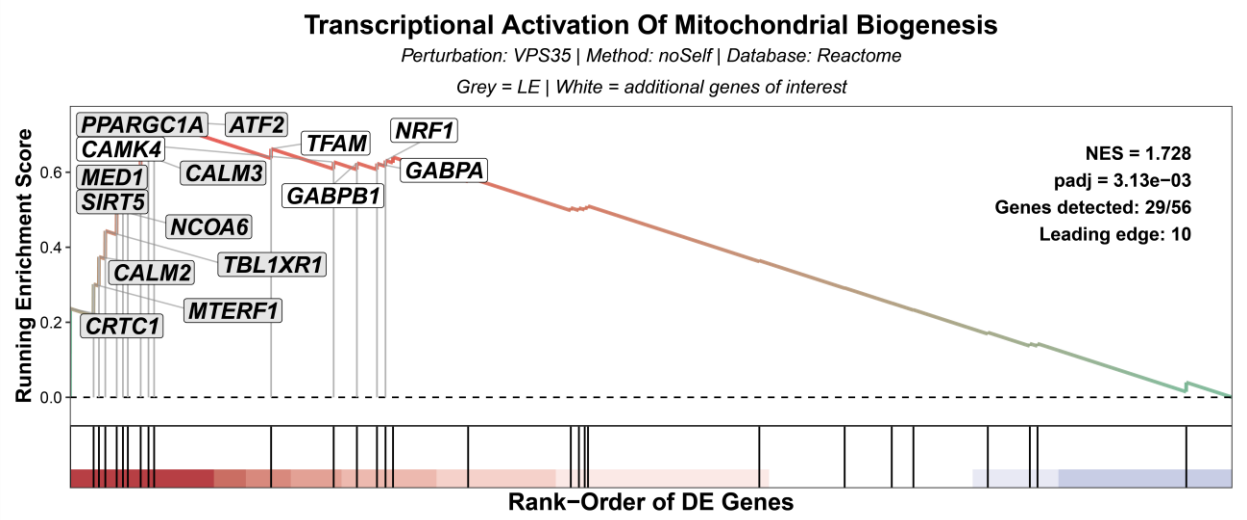


Figure 3.19 – *GSEA plot for the term "Transcriptional Activation of Mitochondrial Biogenesis" which is significantly upregulated in VPS35-perturbed neurons. Leading edge genes are labeled in grey, and additional genes associated with this term are shown white.*

Upregulation of endolysosome terms

Another group of noSelf terms implicating *VPS35*-related biology capture endolysosomal-trafficking biology apart from the retromer complex. The term **membrane trafficking** (Reactome, NES = +1.31, padj = 0.015) is the broadest representative, with 87 leading-edge genes including the SNX-BAR retromer component *SNX2*, GARP retrograde-trafficking complex subunits *VPS53* and *VPS54*, the class-I endosomal-trafficking regulator *VPS45*, and the ESCRT-III disassembly ATPase *VPS4A* (Figure 3.20). Additional signatures involving Rho-GTPase signaling were captured by terms like **RHO GTPase effectors** (Reactome, NES = +1.43, padj = 0.022), with leading-edge genes including *CDC42, IQGAP1, PAK1, PAK2,* and *LIMK1,* actin-cytoskeleton regulators of endosomal tubulation and cargo sorting.

Cathepsin signature within a downregulated collagen related term

The sole downregulated term in *VPS35*'s noSelf results is **collagen catabolic process** (GO_BP, NES = −2.29, padj = 0.016), with 6 leading-edge genes: cathepsins *CTSD,* and *CTSL,* as well as matrix metalloproteinases *MMP10, MMP15, MMP24,* in addition to *CST3*. Cathepsin delivery to lysosomes depends on retromer-mediated trafficking of the cation-independent mannose-6-phosphate receptor, providing a mechanistic link to retromer biology[23]. Downregulation of *CTSD* in particular is concordant with lysosomal-failure in PD biology. Although reports have implicated metalloproteinases in relation to neurodegenerative disease, direct relevance of matrix metalloproteinase downregulation in *VPS35*-perturbed cells, however, remains unclear.

Terms unique to withSelf results

The number of downregulated terms expands to 13 when considering *VPS35*'s withSelf results. Although these terms are all rendered non-significant when *VPS35* is excluded, they nevertheless nominate a canonical phenotype of retromer-loss. The **regulation of presynapse organization** term (GO_BP, NES = −2.55, padj = 0.011) is the only one with a substantial non-VPS35 leading edge (*NLGN1, NLGN3, SLITRK1, PTPRD,* and *APP*), implicating a number of transmembrane synaptic-adhesion proteins consistent with retromer's documented role in the recycling of synaptic machinery[24]. Reactome's **WNT ligand biogenesis and trafficking** (NES = −2.17, padj = 1.9×10^{-3}) nominates retromer's involvement in recycling Wntless from endosomes to the trans-Golgi for Wnt secretion[25].

Upregulation of ribosomes and translation

The dominant signatures among both noSelf and withSelf GSEA results for *VPS35* were nominated by over 40 terms which clearly indicated a strong upregulation of ribosomal and translation-related genes. A representative example is the WikiPathways term **cytoplasmic ribosomal proteins** (NES = +2.04, padj = 3.9×10^{-8}), which has a leading edge almost entirely composed of cytosolic ribosome genes (*RPL* / RPS**). Because translation-and-ribosome upregulation is a recurrent axis across multiple perturbations discussed in this chapter, the cross-perturbation interpretation of this signal is deferred to §3.6 (convergent themes).

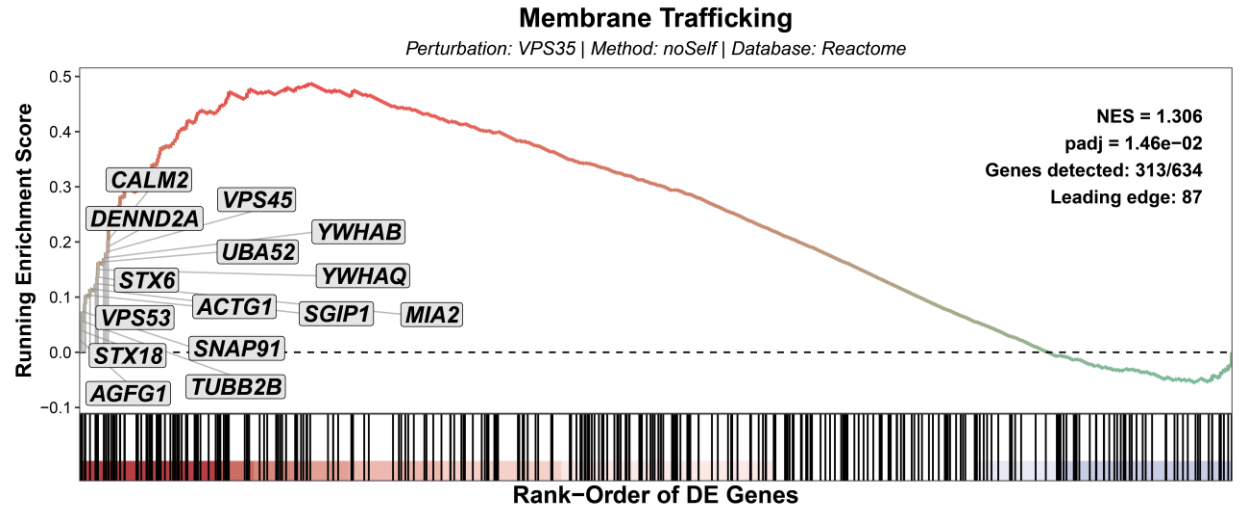


Figure 3.20 – *GSEA plot for the term "Membrane Trafficking" which is significantly upregulated in VPS35-perturbed neurons. Leading edge genes are labeled in grey.*

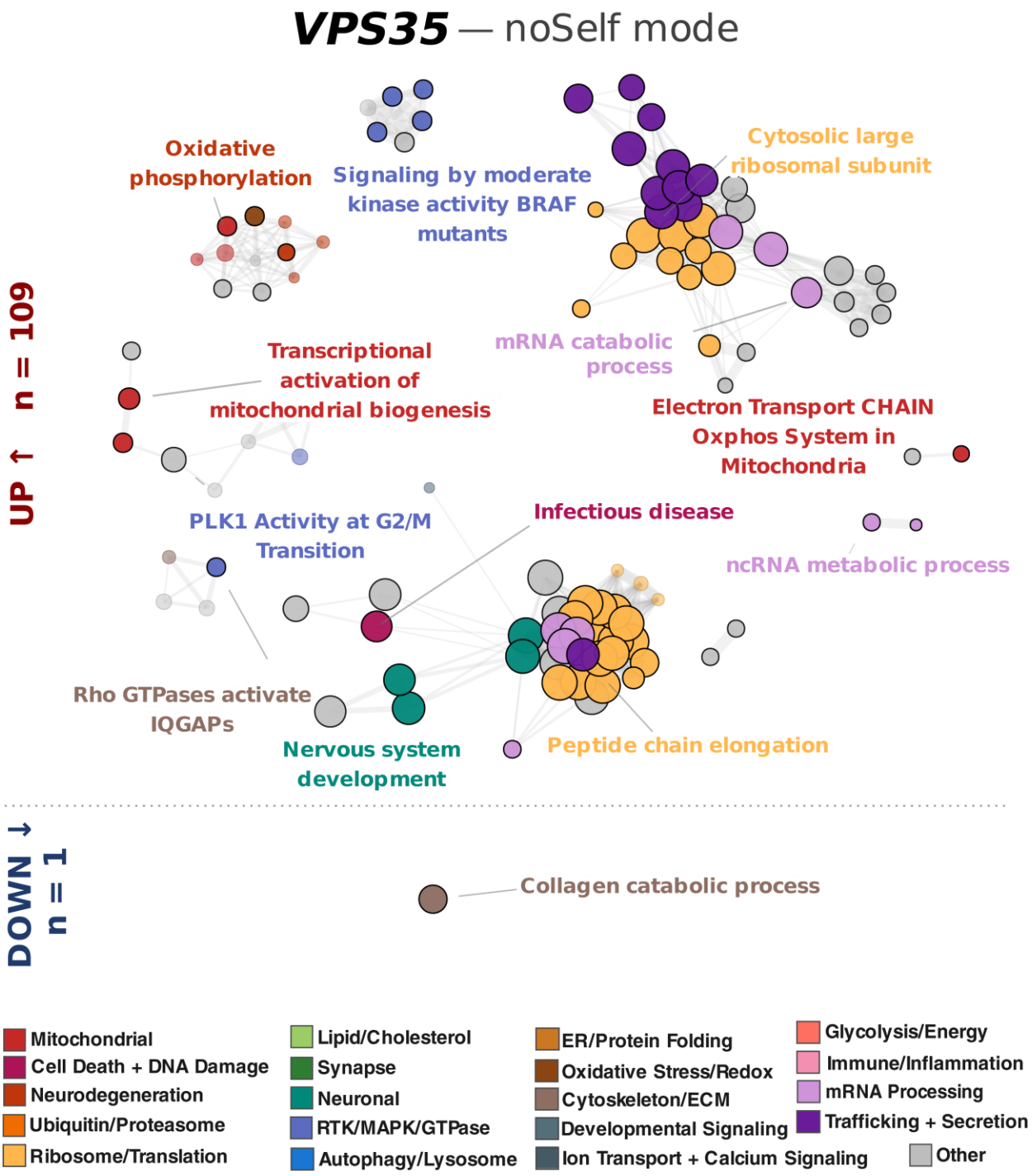


Figure 3.21 - Network plot of significant GSEA terms nominated by *VPS35* perturbation in neurons using noSelf method. Nodes represent significant GSEA results, connections between them represent the degree of overlap behind related gene sets. Terms are manually assigned to 20 biological categories, depicted as color. Circles with a black border are GSEA tests results which remain significant after multiple testing correction. Dot size indicates the adjusted p-value of the GSEA result: larger dots are more significant.

3.4.3 *PARK7* / DJ-1

Background

PARK7 encodes DJ-1, a small redox-sensitive protein whose loss-of-function mutations cause early-onset autosomal recessive Parkinson's disease[34]. DJ-1 functions as an oxidative-stress sensor that regulates antioxidant gene expression and stabilizes the antioxidant master regulator NRF2[35,36]. In DA neurons within SNc, loss of DJ-1 elevates the mitochondrial oxidative stress generated by L-type calcium-channel-driven autonomous pacemaking[37] and leads to a reduction in striatal dopamine release[38].

DE genes and GSEA overview

In the analysis of our combined neuron clusters, comparison of 114 perturbed cells against 214 non-targeting controls indicated significant self-knockdown for *PARK7* (log2FC = −1.187; BH-adjusted p = 3.46×10^{-9}). Four additional DEGs were found to be upregulated in *PARK7* cells: *HBEGF*, *KATNAL2*, *AGBL3*, and *THUMPD3* (Figure 3.22) Gene set enrichment analysis identified 40 significant terms (0 up, 40 down) for the withSelf approach where *PARK7* was retained in the ranked DEG list, and 55 significant terms for the noSelf approach where it was omitted (1 up, 54 down). A network plot of significant GSEA terms is shown in Figure 3.26.

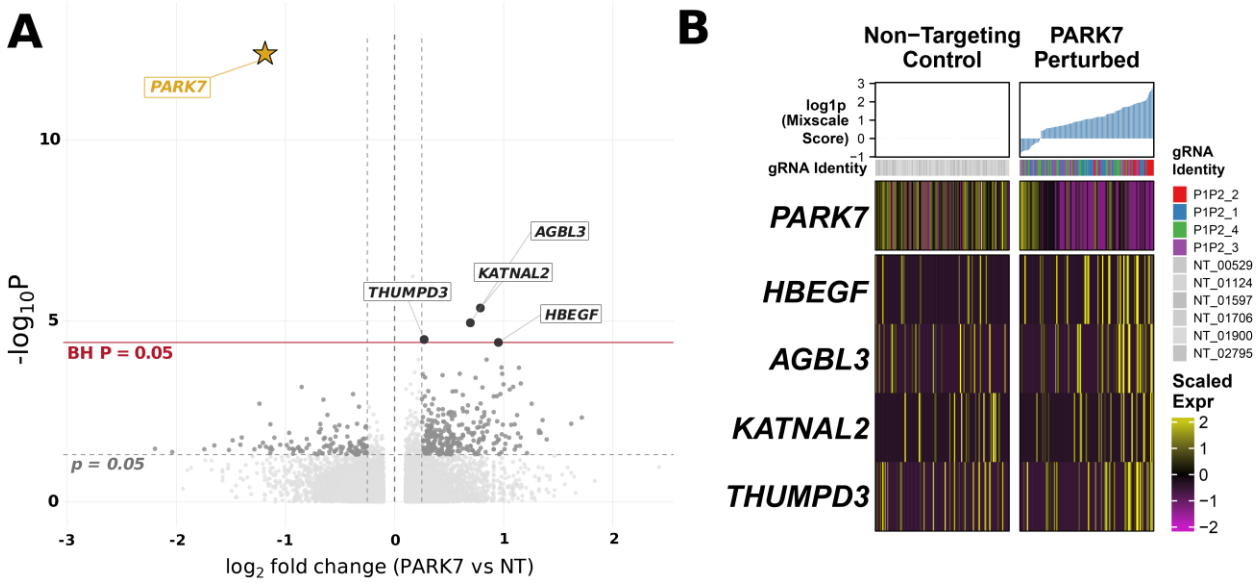


Figure 3.22 – *Overview of differentially expressed genes identified by Mixscale within PARK7-perturbed neurons from the familial-PD Perturb-seq experiment.* ***(A)*** *– Volcano plot indicating log2FC vs. p-value. Grey dotted horizontal line indicates a nominal p-value of 0.05, the red horizontal line denotes the Benjamini-Hochberg corrected pAdj of 0.05. Self knockdown of PARK7 is indicated with a gold star.* ***(B)*** *– Heatmap showing the expression of PARK7 itself alongside all four other DEGs that are significantly upregulated in PARK7-perturbed neurons. Columns at left represent non-targeting control cells, while columns on the right indicate all PARK7-perturbed cells, ranked from left-to-right by increasing Mixscale score.*

Upregulation of P53 targets

There is a single positively enriched term among all of *PARK7*'s results, appearing only in the noSelf approach. This term, **p53 pathway** (Hallmark, NES = +1.80, padj = 0.025), has a 16-gene leading edge containing p53 regulators like tumor-suppressor *RB1* and the stress-response transcription factor *FOXO3*, as well as transcriptional targets of p53 including antiproliferative factor *BTG1*, p53-stabilizing ribosomal protein *RPS27L,* and redox/metabolism regulator *TXNIP* (Figure 3.23). Another notable gene in this leading edge is the growth factor *HBEGF,* which was also identified as a significantly DEG upregulated in neurons with *PARK7*-perturbation. *HBEGF* is a direct transcriptional target of p53 and a documented survival factor for midbrain DA neurons[39,40], and exogenous HB-EGF has been shown to protect nigrostriatal DA neurons in a 6-OHDA rat model of PD[41]. Its induction in our *PARK7*-perturbed neurons is consistent with engagement of a p53-driven survival program, potentially driven in response to impaired redox-buffering capacity. The p53 pathway enrichment also calls back to work from Kamath and colleagues, which identified a transcriptionally defined population of DA neurons which are particularly vulnerable in PD: transcriptional targets of *TP53*/p53 were significantly enriched in the vulnerable *SOX6+/AGTR1+* subpopulation, but not in other DA neuron subtypes[14].

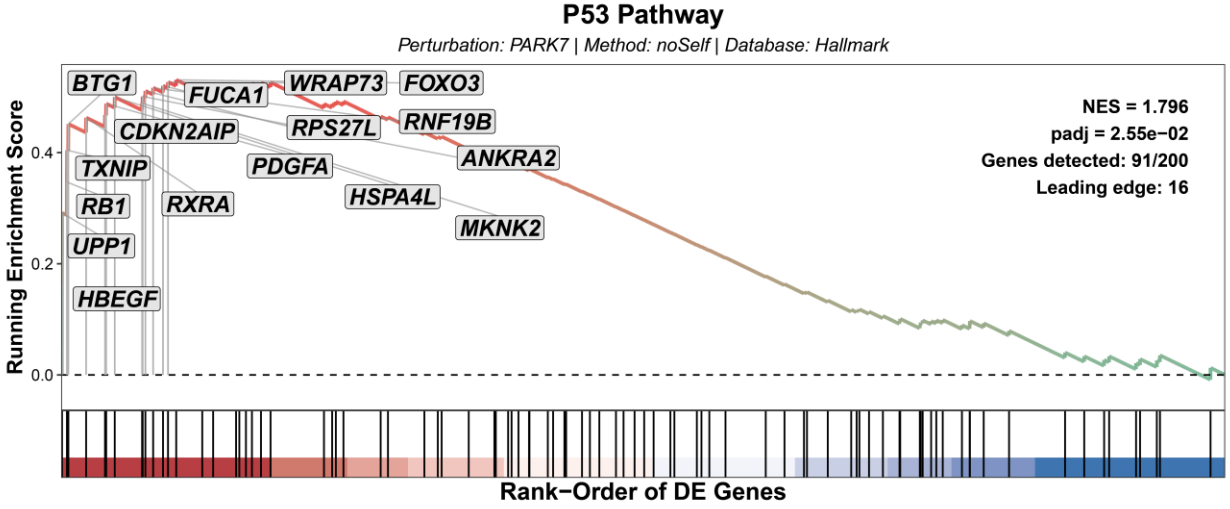


Figure 3.23 – *GSEA plot for the term "P53 Pathway" which is found to be significantly upregulated in PARK7-perturbed neurons. Leading edge genes that make the largest contribution to the signal are labeled in grey.*

Downregulation of cholesterol homeostasis

In evaluating both the noSelf and withSelf enrichment results for our *PARK7*-knockdown neurons, we observed a consistent downregulation of terms involving cholesterol and sterol-biosynthesis. A representative example is **cholesterol homeostasis** (Hallmark, NES = -1.92, padj = 0.015), with 22 genes in its leading edge implicating decrease in genes from the mevalonate (*HMGCR*, *HMGCS1*, *MVD*) and sterol-biosynthesis (*FDFT1*, *SQLE*, *CYP51A1*, *DHCR7*, *EBP*, and *TM7SF2*) branches of the cholesterol pathway. Two additional leading edge genes, *LDLR* and *SREBF2*, have been implicated in *PARK7*'s

reported role in the regulation of cholesterol homeostasis. Expression of *LDLR* is driven through binding of *SREBF2's* gene product, SREBP2, to the sterol regulatory element (SRE). Co-immunoprecipitation of DJ-1 and SREBP2 has been shown in human SH-SY5Y neuroblastoma cells, and chromatin immunoprecipitation experiments in mouse fibroblasts have identified DJ-1's association at the same SRE of *LDLR* as well[42]. DJ-1 knockout mice showed reduced *LDLR* expression in the liver, and elevated serum LDL[42]. Although prior evidence in human cells was confined to biochemical assays in SH-SY5Y cells, our results suggest a possible link between *PARK7* and cholesterol homeostasis in hPSC-derived midbrain neurons. While cholesterol-related terms were also downregulated for *GIGYF2* and *B2M*-perturbed neurons, neither has a reported link to cholesterol, and *PARK7* terms were the only ones with *LDLR* or *SREBF2* in the leading edge. The downregulation of cholesterol-pathway genes is consistent with epidemiologic evidence that lower serum LDL cholesterol has been associated with elevated PD risk[43,44].

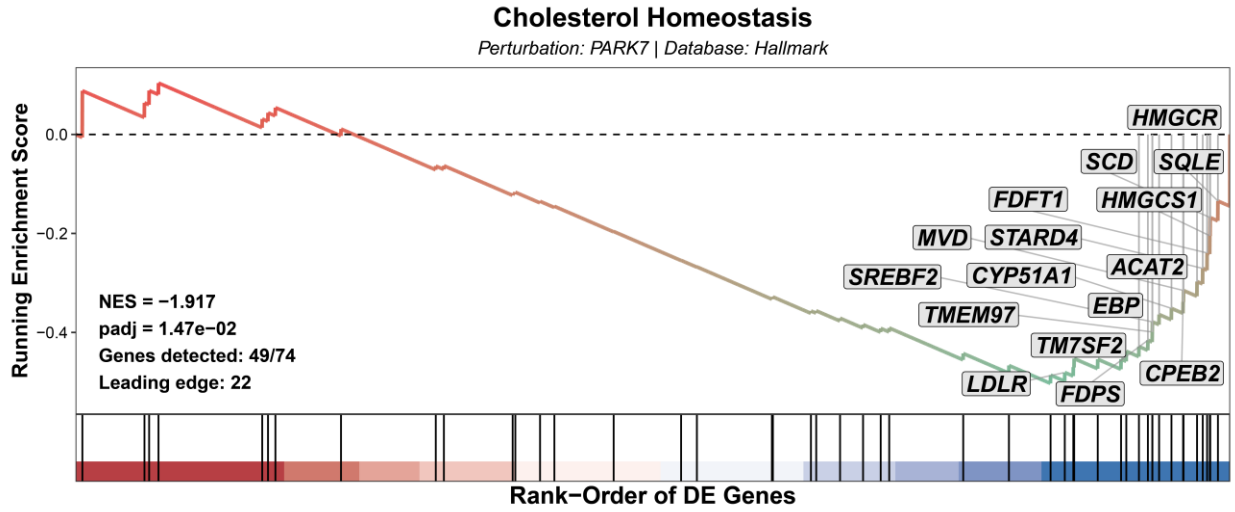


Figure 3.24 – *GSEA plot for the term "Cholesterol Homeostasis" which is significantly downregulated in PARK7-perturbed neurons. Leading edge genes that make the largest contribution to the signal are labeled in grey.*

mTORC1 downregulation

The strongest noSelf signal by significance is the downregulated **mTORC1 signaling** term (Hallmark, NES = −2.21, padj = 1.43 × 10^{-6}), which has a 59-gene leading edge dominated by many of the same cholesterol-biosynthesis enzymes (*HMGCR*, *HMGCS1*, *SQLE*, *CYP51A1*, *DHCR7*, *IDI1*, *EBP*, and *TM7SF2*) alongside *LDLR* and the SREBP regulator *INSIG1*. Thirteen of the 22 leading-edge genes from **cholesterol homeostasis** are present in the mTORC1 leading edge as well. The remaining genes driving this signature span glycolysis (*GAPDH*, *TPI1*, *ENO1*, *PFKL*, and *PGK1*), amino-acid transport and biosynthesis (*SLC7A5*, *SLC7A11*, *MTHFD2*, and *PSAT1*), and ER protein-folding components (*CANX*, *CALR*, *SSR1*, and *RPN1*), reflecting the anabolic programs regulated by mTORC1.

Downregulation of synapses and secretion

A coherent downward-shift in pathways related to secretion is also identified in both modes. In noSelf results, **negative regulation of secretion** (GO_BP, NES = −2.14, padj = 7.4

× 10^{-3}) is accompanied by several related terms and contains a 23-gene leading edge that includes *SNCA, SYT4, CHGA, MAOB,* and *HMGCR*. The same family of terms appears in the withSelf results as well, with the appearance of **protein secretion** (GO_BP, NES = −2.37, padj = 0.018), which adds leading edge genes *SCG2, GAPDH, S100A13, RAB3B,* and *LRP1*. *PARK7* knockdown in neurons additionally drove a downregulation of synaptic and calcium channel related terms. The term **negative regulation of synaptic transmission** (GO_BP, NES = −1.95, padj = 0.042) carries a small but noteworthy leading edge with several known PD risk genes: *SNCA, MAPT, SYT4, SHANK2,* and *HTR2A*. A final related term **regulation of cation channel activity** (GO_BP, NES = −1.90, padj = 0.025) nominates the calcium-channel signature again, with leading-edge genes including *CACNA2D1,* and *CACNB2* as well as *GRIA2, SHANK3,* and *NEFL*.

Downregulation of mitophagy

A single withSelf term, **positive regulation of autophagy of mitochondrion** (GO_BP, NES = −2.60, padj = 0.018), had a noteworthy leading edge driven by two genes, *BNIP3* and *PARK7* itself (Figure 3.25). The enrichment is clearly driven strongly by *PARK7*'s extreme rank order in the gene set, but the co-occurrence of *BNIP3*, a canonical mitophagy receptor, is nevertheless consistent with the established DJ-1↔autophagy↔mitochondrial-clearance axis, where loss of DJ-1 has been shown to impair lysosomal and mitophagic clearance in neurons[45]. Additional withSelf terms driven by *PARK7* included a cluster of redox terms related to DJ-1 (**glutathione metabolism**, **peroxide-acting oxidoreductase activity**, **hydrogen-peroxide and superoxide responses**) as well as dopamine-metabolism / dopaminergic-synaptic-transmission terms. While the presence of these terms clearly point to well-established functions of *PARK7*/DJ-1, their disappearance in the noSelf approach suggests that the signal is predominately driven by the strong self-knockdown of *PARK7* that place it at the very end of the full rank-ordered gene list.

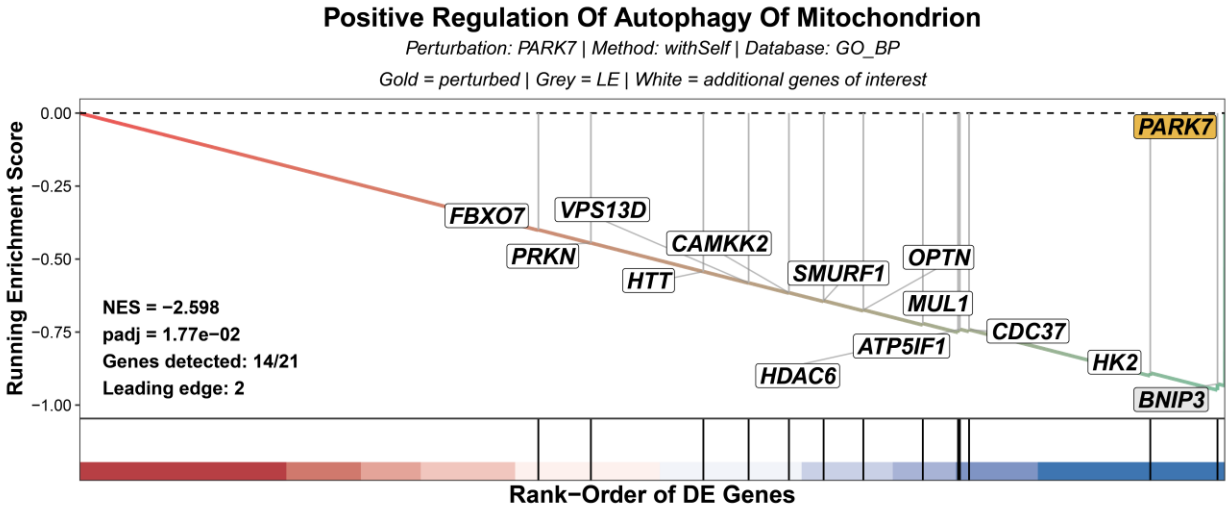


Figure 3.25 – *GSEA plot for the term "Positive Regulation of Autophagy of Mitochondrion" which is significantly downregulated in PARK7-perturbed neurons when PARK7 itself is retained in the rank-ordered input. PARK7 is labeled in gold, leading edge genes are labeled in grey, and additional genes associated with this term in white.*

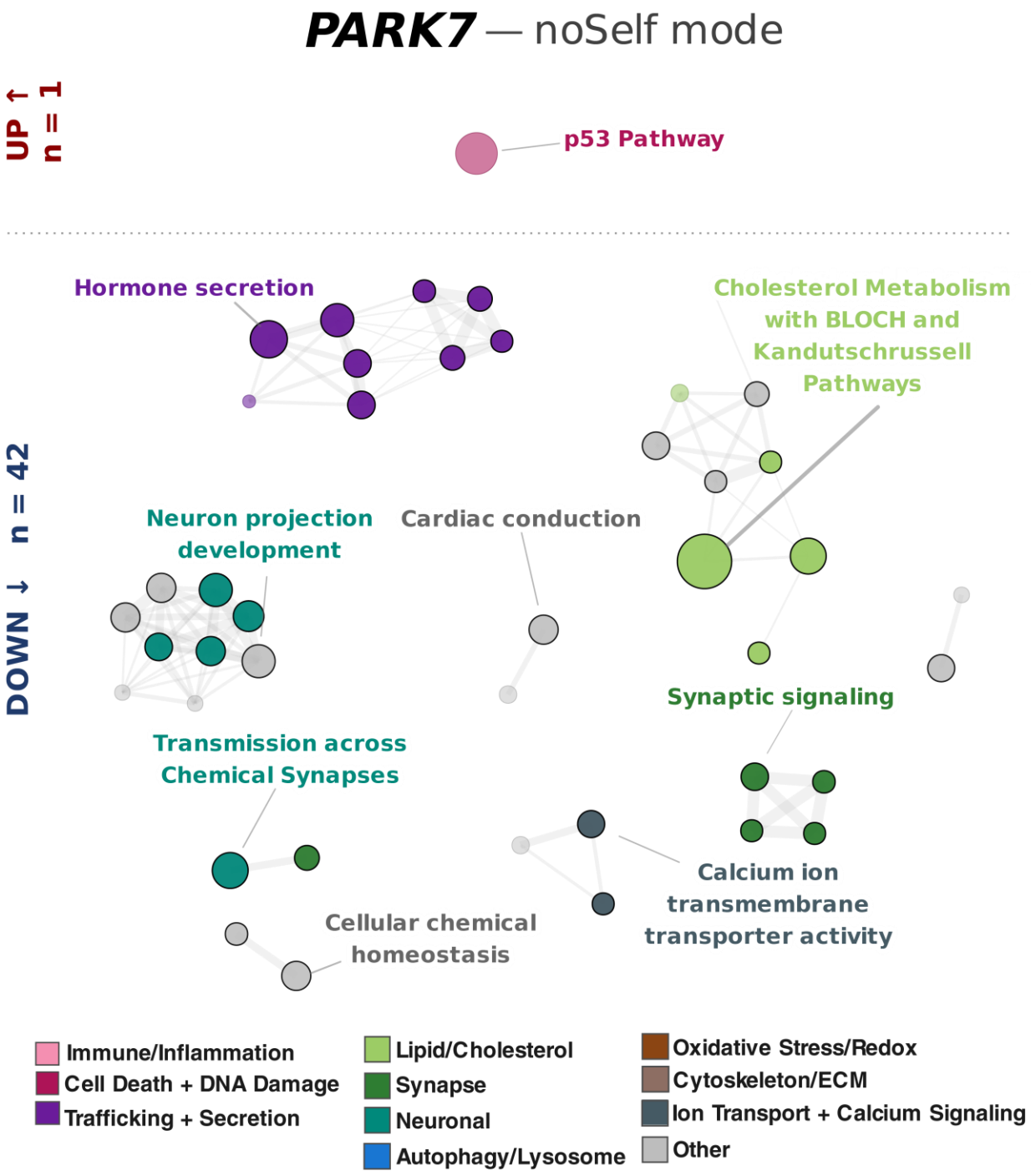


Figure 3.26 - Network plot of significant GSEA terms nominated by *PARK7* perturbation in neurons using noSelf method. Nodes represent significant GSEA results, connections between them represent the degree of overlap behind related gene sets. Terms are manually assigned to 20 biological categories, depicted as color. Circles with a black border are GSEA tests results which remain significant after multiple testing correction. Dot size indicates the adjusted p-value of the GSEA result: larger dots are more significant.

3.4.4 *FBXO7*

Background

FBXO7 encodes F-box-only-7 protein, a subunit of the SCF (Skp1-Cul1-F-box) E3 ubiquitin ligase complex responsible for substrate recognition, and loss-of-function mutations in this gene have been associated with autosomal-recessive early-onset forms of PD[46–48]. *FBXO7* is involved in proteasome assembly and has been shown to ubiquitinate proteasomal core-particle subunit *PSMA2*[49]. In addition, FBXO7 has been shown to negatively regulate NF-κB signaling in non-neuronal cells via ubiquitination of TNF-receptor signaling components TRAF2 and cIAP1 and the NF-κB transcriptional cofactor UXT-V2[50,51]. Global knockout of *FBXO7* in mice causes proteasomal impairment, motor deficits and failure to thrive[49], and its conditional deletion in midbrain DA neurons causes their progressive loss through an RPL23-MDM2-p53 stress pathway[52]. Previous reports suggested that *FBXO7* acts as a positive regulator of PINK1-Parkin-mediated mitophagy via direct interaction with both proteins[53], but more recent work in *FBXO7*-null HeLa cells and iNeurons found no defect in PINK1/Parkin pathway activation, mitophagic flux, or mitochondrial clearance, arguing against a general role for *FBXO7* in PINK1/Parkin-dependent mitophagy[54].

DE genes and GSEA overview

In the 174 *FBXO7*-perturbed cells analyzed across both neuron clusters, self-knockdown was modest but failed to pass significance thresholds (log2FC = -0.36, BH-adjusted p = 0.62). Despite this, *FBXO7* knockdown produced one of the largest sets of significant GSEA terms of all perturbations, with 873 noSelf terms, 520 up and 353 down. Terms nominated by the noSelf and withSelf modes were nearly identical, showing 96% overlap and changes predominantly in neuron-development terms where *FBXO7* itself was in the leading edge. Whereas many perturbations' signatures showed bias in the balance of term directionality, significant terms identified for *FBXO7* were notably bidirectional. Network plots of UP/DOWN GSEA terms for *FBXO7* perturbation are shown in Figures 3.31 and 3.32, respectively.

Twenty DEGs were identified in *FBXO7*-knockdown neurons, 17 up, and 3 down (Figure 3.27). Among the up genes were mitochondrial Complex I core subunits *NDUFA2* and *NDUFAB1*, as well as Complex I assembly factor *NDUFAF3*. Additionally, we saw that mitochondrial outer-membrane import receptor *TOMM22* and the mito-ribosomal subunit *MRPS24* had significantly increased expression. The greatest positive fold change among protein-coding genes belonged to *IKBKG* (+1.05) which encodes NEMO, a subunit of the IκB kinase complex responsible for the activation of NF-κB signaling[55]. Among the downregulated DEGs were *NEFL*, a component of the neuronal cytoskeleton and *PDE10A*, a cAMP/cGMP phosphodiesterase expressed most prominently in striatal medium spiny neurons and in midbrain dopamine neurons to a lesser extent[56]. Protein levels of *PDE10A* in the striatum have been reported to be decreased in PD patients as a function of motor severity[57].

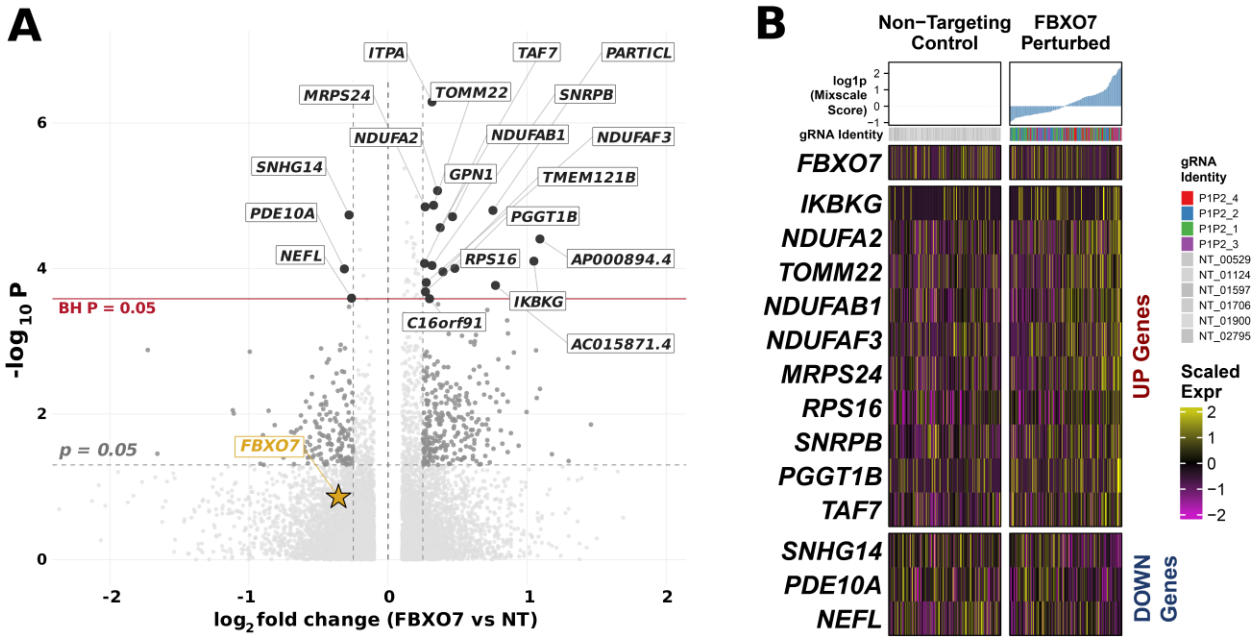


Figure 3.27 – *Overview of differentially expressed genes identified by Mixscale within FBXO7-perturbed neurons from the familial-PD Perturb-seq experiment.* ***(A)*** *– Volcano plot indicating log2FC vs. p-value. Grey dotted horizontal line indicates a nominal p-value of 0.05, the red horizontal line denotes the Benjamini-Hochberg corrected pAdj of 0.05. Self knockdown of FBXO7 is indicated with a gold star despite failing to pass significance thresholds.* ***(B)*** *– Heatmap showing the expression of FBXO7 itself alongside a subset of 10 upregulated DEGs and all three downregulated DEGs detected in FBXO7-perturbed neurons. Columns at left represent non-targeting control cells, while columns on the right indicate all FBXO7-perturbed cells, ranked from left-to-right by increasing Mixscale score.*

Proteasome upregulation

The family of upregulated terms with the clearest links to *FBXO7* biology surround ubiquitination and the proteasome. The most representative term, **SCF-dependent proteasomal ubiquitin-dependent protein catabolic process** (GO_BP, NES = +1.84, padj = 9.7×10^{-3}), has a leading edge which includes the SCF-scaffold component *RBX1,* three additional F-box proteins (*FBXO9, FBXL18,* and *FBXL16),* seven 20S core-particle subunits, three 19S regulatory-particle subunits, the proteasome assembly chaperone *DMAC2*, and proteasome activator *PSME3*. While loss of *FBXO7* has been shown to decrease proteasome activity in neurons[49], the transcription of proteasome genes in our *FBXO7*-knockdown cells is markedly upregulated. Considering *FBXO7*'s role in mediating proteasome assembly, it seems likely that the increase in proteasomal gene expression at the RNA level is a compensatory mechanism in efforts to restore deficient proteasomal function.

Upregulation of NF-κB pathway

The upregulation of proteasomal genes coincides with significant enrichment of NF-κB-pathway terms, consistent with *FBXO7*'s documented role as a negative regulator of NF-κB[50]. Two related terms, **Interleukin-1 signaling** (Reactome, NES = +2.04, padj = 9.8×10^{-5}) and **IL-1-mediated signaling pathway** (GO_BP, NES = +2.12, padj = 6.3×10^{-5}) both share a leading edge of 17 genes, including twelve PSM-family components, the differentially

expressed gene *IKBKG*/NEMO and four additional NF-κB activation components: *MAPK3* (ERK1), *CHUK* (IKKα), *SQSTM1*, and *RBX1* (Figure 3.28). A final related term, **cellular response to tumor necrosis factor** (GO_BP, NES = +1.70, padj = 1.4×10^{-2}), additionally nominates *TP53*, the linear ubiquitin chain assembly complex (LUBAC) component *RBCK1*/HOIL-1L, the NF-κB scaffold *TANK*, and Tumor Necrosis Factor Receptor superfamily member *TNFRSF12A*. In non-neuronal systems (U2OS, BJET fibroblasts, HEK293 cell lines), *FBXO7* has been shown to ubiquitinate cIAP1 and TRAF2, decreasing RIP1 ubiquitination and stabilizing IκBα, leading to a suppression of NF-κB signaling[50]. Whether the same regulation operates in neurons has not been directly tested, but the upregulation of NF-κB-pathway transcriptional signatures in our *FBXO7*-knockdown neurons is consistent with this paradigm despite limitations in the statistical significance of its self-knockdown in our experiments.

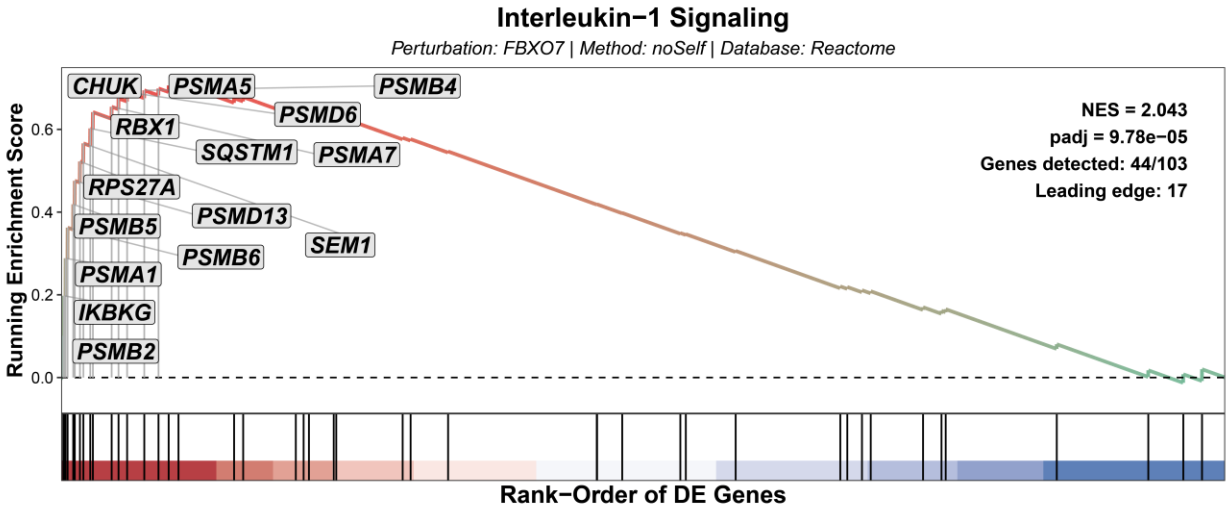


Figure 3.28 – *GSEA plot for the term "Interleukin-1 Signaling" which is significantly upregulated in FBXO7-perturbed neurons.*

Upregulation of mitochondrial and proteostasis terms

Another group of significant GSEA terms involves the convergence of mitochondrial and proteostasis genes in PD-relevant disease pathways. The term **Parkinson's disease** (KEGG, NES = +1.96, padj = 5.3×10^{-5}) is positively enriched, with a 42-gene leading edge spanning Complex I subunits (*NDUFA2/A4/A9*, *NDUFB2/B3/B5/B6/B9/B10*, *NDUFS2/S3/S5/S6*, *NDUFV2*, and *NDUFAB1*), proteasome subunits (*PSMA1/A5/A7*, *PSMB2/B4/B5/B6, PSMD6/D12/D13,* and *SEM1*), ATP synthase (*ATP5F1C* and *ATP5MC2/MC3*), Complex IV (*COX7B*), Complex III (*UQCR10, UQCRFS1*), and additional noteworthy signaling components *PINK1, TP53, ATF4, EIF2S1, UBE2L3, UBE2L6*, and *UBE2J2*, and the intracellular calcium channel *RYR3* (Figure 3.29). The convergence of mitochondrial-complex genes, proteasome subunits, and mitophagy regulator *PINK1* within a single curated PD-pathway leading edge is consistent with a disruption of mitochondrial quality control and proteostasis. The *TP53/ATF4/EIF2S1* stress-response components in this leading edge have direct mechanistic grounding: conditional deletion of *Fbxo7* in

midbrain dopamine neurons activates an RPL23-MDM2-TP53 stress pathway with a pro-apoptotic transcriptional signature, and this axis is required for the dopaminergic degeneration phenotype in the conditional-knockout mouse[52]. The TP53 callout also parallels the results seen for *PARK7* knockdown, where the p53 pathway emerged as the single positively enriched term — both perturbations engage TP53-dependent stress programs through distinct upstream mechanisms.

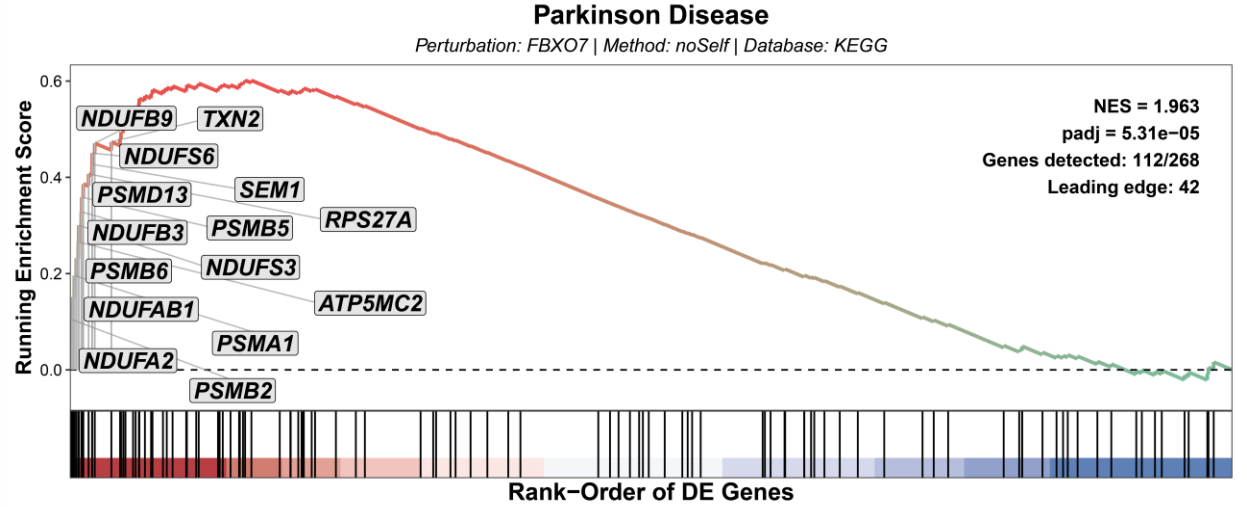


Figure 3.29 – *GSEA plot for the KEGG term "Parkinson Disease" which is significantly upregulated in FBXO7-perturbed neurons*.

Terms upregulated in FBXO7-perturbed neurons also implicate oxidative phosphorylation and electron transport chain (ETC) biogenesis. The representative term **oxidative phosphorylation** (KEGG, NES = +2.19, padj = 5.1×10^{-7}) anchors this cluster of terms, with leading-edge genes spanning Complex I subunits (*NDUFA2, NDUFAB1, NDUFA4, NDUFA9, NDUFB2/3/5/6/9/10, NDUFS2/3/5/6,* and *NDUFV2*), Complex III components (*UQCRFS1* and *UQCR10*), Complex IV (*COX7B*), and ATP synthase (*ATP5MC2/3, ATP5ME, ATP5F1C,* and *ATP5MG*). Terms involving an increase in ETC biogenesis are exemplified by **Complex I biogenesis** (Reactome, NES = +2.16, padj = 1.3×10^{-5}), whose 18 leading edge genes contain the aforementioned DEGs *NDUFA2, NDUFAB1,* and *NDUFAF3* alongside fifteen additional Complex I subunits and assembly factors (*TIMMDC1, NDUFS3, NDUFB3, NDUFS6, NDUFB9, NDUFS5, NDUFS2, NDUFV2, NDUFAF4, NDUFB2, NDUFB5, TMEM126B, NDUFA9, NDUFB10,* and *NDUFB6*).

The term **PINK1-PRKN mediated mitophagy** (NES = +1.71, padj = 2.0×10^{-2}) contains *TOMM22, TOMM6, RPS27A,* and *SQSTM1* in its leading edge, the first of which was also identified as a significantly upregulated DE gene for *FBXO7*-perturbed neurons (Figure 3.30). In addition, the broader term **macroautophagy** from GO_BP includes *PINK1, SQSTM1, ATG5, ATG2B, GABARAP, NEDD4, TP53,* and *SIRT1*. Although *FBXO7* has been previously nominated as a positive regulator of PINK1-Parkin mitophagy, its role in this pathway has been contested. More recent work involving CRISPR-knockout of *FBXO7* in HeLa cells and iNeurons showed no defect in mitophagic flux[54]. The compensatory-upregulation pattern

observed here could be consistent with either model: the signal we see could reflect transcriptional buffering of a directly perturbed mitophagy axis, or transcriptional response to broader proteostasis stress that secondarily engages mitochondrial-quality-control machinery via NRF1-mediated SQSTM1/GABARAPL1 induction. The transcriptomic results presented here are insufficient to distinguish between these mechanisms.

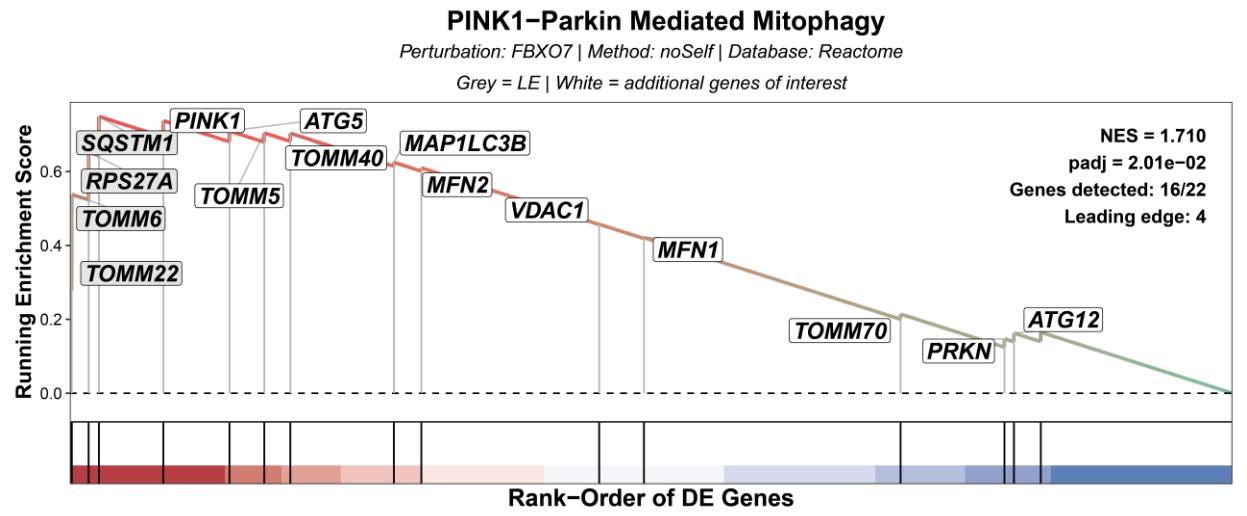


Figure 3.30 - *GSEA plot for the Reactome term "PINK1-Parkin Mediated Mitophagy" which is significantly upregulated in FBXO7-perturbed neurons*.

Upregulation of mRNA splicing terms

A coherent signature of spliceosomal-machinery upregulation is also seen following FBXO7 pertubation as well. The term **Processing of Capped Intron-Containing Pre-mRNA** (NES = +2.10, padj = 2.8×10^{-7}) is the broadest representative, with a leading edge spanning the SNRP small-nuclear-ribonucleoprotein family (*SNRPB, SNRPE, SNRPF, SNRPG,* and *SNRPA*), the SF3B and SF3A spliceosome subcomplexes (*SF3B5/6,* and *SF3A2*), the exon-junction-complex component *MAGOH*, and the deubiquitinase *USP39* among others. This signature is also supported by the presence of *SNRPB* in the list of differentially expressed genes following *FBXO7* knockdown.

Downregulation of synapse-related terms

The dominant downregulated cluster in the FBXO7 signature is a coordinated impairment of synaptic-and-neuronal-architecture programs that spans assembly, projection morphology, neurotransmitter handling, and ion-channel function. Within that broader cluster, the neurotransmitter-handling and ion-channel terms name canonical dopaminergic-neuron markers in their leading edges, making the functional-output portion of the signature specifically dopaminergic rather than generically neuronal. The strongest term by NES is the term **synapse assembly** (GO_BP, NES = −2.42, padj = 3.8×10^{-8}), with a 48-gene leading edge containing the neuronal-adhesion molecules *NRXN1, NRXN2, NLGN1, NLGN2, NRCAM, CNTN5, CDH2, NTRK2, NTRK3*, the receptor protein-tyrosine phosphatases *PTPRD* and *PTPRS*, the ephrin receptor *EPHB2*, the SLITRK family (*SLITRK4*,

SLITRK5), the IL-1 receptor accessory proteins *IL1RAPL1*/*IL1RAPL2*, and notably the PD-risk-genes *SNCA* and *APP*. Notably, the significantly downregulated DE gene *NEFL* (neurofilament light chain, log2FC = −0.26) appears in 77 terms' leading edges.

The functional-output arm of this cluster names the canonical DA-neuron machinery directly. The term **synaptic vesicle cycle** (GO_BP, NES = −2.13, padj = 1.3×10^{-5}) has a leading edge containing *TH* (tyrosine hydroxylase), *DDC* (DOPA decarboxylase), *SNCA*, *CACNA1D*, the canonical syntaxin family (STX1B), the synaptotagmin family (*SYT4*, *SYT5*, *SYT6*, *SYT9*), the SNARE machinery (*SNAP29, SNAP47, VAMP4, DOC2B*), and *SV2A*. **Neurotransmitter secretion** (GO_BP, NES = −2.02, padj = 3.4×10^{-4}) recovers the same dopaminergic-functional-output signature, with *PRKN* joining *SNCA*, *CACNA1D*, *STX1B*, and the synaptotagmin and SNARE machinery in the leading edge. The term **neurotransmitter transport** (GO_BP, NES = −1.91, padj = 3.2×10^{-4}) extends the signature with the SLC6 neurotransmitter transporters (*SLC6A17*, *SLC6A6*) alongside *PRKN*, *DDC*, *TH*, *SNCA*, and *CACNA1D*. The convergent appearance of the two dopamine biosynthesis enzymes *TH* and *DDC* together with *PRKN*, *SNCA*, and *CACNA1D* (the L-type calcium channel that underlies SNc DA-neuron pacemaking and creates the calcium-driven oxidative-stress vulnerability of the SNc DA population[37])— strongly suggests that the synaptic downregulation observed could be functionally dopaminergic. This is reinforced by the term **calcium ion transport** (GO_BP, NES = −1.85, padj = 4.9×10^{-4}), whose leading edge includes *CACNA1D, CACNA1A, CACNA1E, RYR2, ATP2B1, SLC8A2,* and *SNCA,* capturing the Ca^{2+}-handling machinery of dopaminergic synaptic biology. It must be

This signature has direct experimental support in *FBXO7*-knockout mice. Conditional deletion of *Fbxo7* in cortical glutamatergic neurons has been shown to alter the protein composition of the corticostriatal presynaptic terminals, identified by quantitative mass spectrometry of striatal synaptic-membrane preparations with significant changes in proteins related to endo/exocytosis pathways as well as to glutamatergic, GABAergic, and dopaminergic signaling[50]. At the mechanistic level, FBXO7 binds and stabilizes the proteasome-binding partner PI31, which couples proteasomes to dynein light-chain proteins (DYNLL1/DYNLL2) for microtubule-dependent axonal transport. Inactivation of PI31 disrupts proteasome motility in axons and causes defects in synaptic protein homeostasis, structure, and function in Drosophila motor neurons[59]. The pan-neuron downregulation of synaptic genes and dopaminergic programs observed here is consistent with two complementary FBXO7-dependent mechanisms; direct SCF-FBXO7-mediated regulation of synaptic protein homeostasis[50] and indirect impairment of axonal proteasome delivery via the FBXO7-PI31 axis[59]. Both mechanisms converge on synaptic proteostasis, with the additional layer that NRF1-driven aggrephagy upregulation observed here could represent a compensatory response to the change in proteostasis gene expression.

Upregulation of ribosomal genes

The dominant noSelf signature for *FBXO7* based purely on significance is a coordinated upregulation of cytosolic translation. The most specific representative, **cytoplasmic ribosomal proteins** (WikiPathways, NES = +2.33, padj = 8.9×10^{-8}), has a

leading edge composed almost entirely of cytosolic ribosomal-protein genes (*RPL*/RPS**). Because translation-and-ribosome upregulation is a recurrent axis across multiple §3.4 perturbations, the interpretation of this signal is deferred to §3.6 (convergent themes).



Figure 3.31 – Network plot of significantly upregulated GSEA terms after *FBXO7* perturbation in neurons using noSelf method. Nodes represent significant GSEA results, connections between them represent the degree of overlap behind related gene sets. Terms are manually assigned to 20 biological categories, depicted as color. Circles with a black border are GSEA tests results which remain significant after multiple testing correction. Dot size indicates the adjusted p-value of the GSEA result: larger dots are more significant.

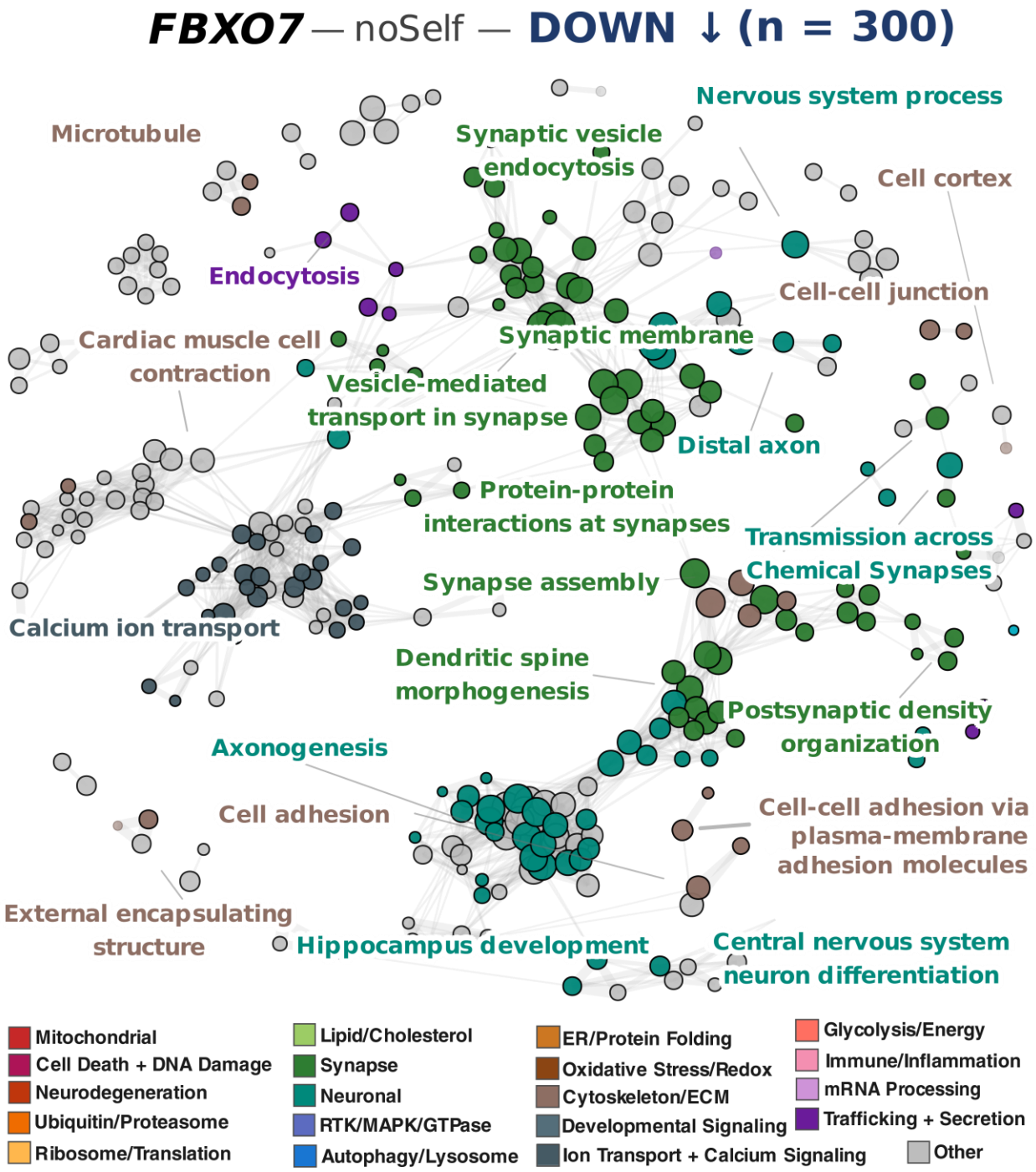


Figure 3.32 – Network plot of significantly downregulated GSEA terms after *FBXO7 p*erturbation in neurons using noSelf method. Nodes represent significant GSEA results, connections between them represent the degree of overlap behind related gene sets. Terms are manually assigned to 20 biological categories, depicted as color. Circles with a black border are GSEA tests results which remain significant after multiple testing correction. Dot size indicates the adjusted p-value of the GSEA result: larger dots are more significant.

3.4.5 *VPS13C*

Background

VPS13C encodes a bridge-like lipid transfer protein which facilitates the transfer of lipids at membrane contact sites between lysosomes and the endoplasmic reticulum. Following the detection of lysosomal damage, *VPS13C* is rapidly recruited to the surface of lysosomes to facilitate membrane repair[60–62]. Biallelic *VPS13C* loss-of-function alleles cause autosomal-recessive early-onset PD with rapid clinical progression[60,63].

DE genes and GSEA overview

In the pooled neuron analysis, 96 *VPS13C*-perturbed cells were compared against 214 non-targeting controls. Mixscale identified significant self-knockdown (log2FC = -1.17, BH-adjusted p = 2.60 x 10^{-3}). A total of 17 DEGs were identified (10 up, 7 down), with notable downregulated genes including several ribosomal genes (*RPL17, RPS4X, RPL7A, RPS3*) and the glycolytic enzyme TPI1. The ferritin heavy-chain transcript *FTH1* was also down-regulated (log2FC = -0.27, BH-adjusted p = 0.05) (Figure 3.33).

Pre-ranked enrichment returned 361 significant noSelf terms (358 down, 3 up), making *VPS13C* the perturbation with the third-largest term set in this experiment. Two themes dominate. First, coordinated downregulation of cytosolic translation and SRP-dependent ER targeting: **cytosolic ribosome** (NES = -2.87), **cytoplasmic ribosomal proteins** (NES = -2.92), **rRNA processing** (NES = -2.78), **SRP-dependent cotranslational protein targeting** (NES = -2.98), **protein localization to ER** (NES = -2.99). This translation-DOWN signature loads on the same shared translational-machinery axis discussed in §3.7 but in the opposite direction from VPS35 (§3.4.2), which drives ribosome-UP; the two perturbations target related endolysosomal machinery yet produce opposite translational-axis responses at the GSEA level.

Downregulation of V-type ATPases

The central observation for *VPS13C*, however, is a second theme: *VPS13C*-perturbed pooled neurons transcriptionally repress expression of V-ATPase and lysosomal-acidification programs. The term **Proton transmembrane transport** (NES = -2.25, p = 4.40 x 10^{-7}) has a 36-gene leading edge containing 12 V-ATPase subunits (*ATP6V0B, ATP6V0C, ATP6V0E1, ATP6V0E2, ATP6V1D, ATP6V1E2, ATP6V1F, ATP6V1G1, ATP6V1G2, ATP6V1H, ATP6AP1,* and *ATP6AP1L*). The closely related terms **Vacuolar V-type ATPase complex** (NES = -1.79, BH-adjusted p = 0.01) and **phagosome acidification** (NES = -1.87, BH-adjusted p = 0.014) reinforce the same observation, with largely overlapping leading-edge genes. The *VPS13C*-mediated down-regulation of *FTH1* and *FTL* appears within the leading edge of Reactome's iron uptake and transport term, but roughly half of that term's leading-edge genes (10 of 21) are themselves V-ATPase subunits, because endosomal acidification is mechanistically required for transferrin-mediated Fe^{3+} release. The iron signal is therefore best understood as a ferritinophagy readout nested within the V-ATPase/lysosomal-acidification observation, not as an independent ferroptotic phenotype. This V-ATPase signature serves as an anchor for the cross-experiment convergence discussed in §3.6. In

the independent GWAS experiment, *VPS13C*-perturbed neurons reproduce a concordant lysosomal-acidification signature; the Jaccard index for V-ATPase-subunit leading-edge content for the term **proton transmembrane transporter activity** is 0.45, with 5 shared subunits across both experiments (*ATP6V0B, ATP6V0C, ATP6V0E2, ATP6V1D, ATP6V1F*).

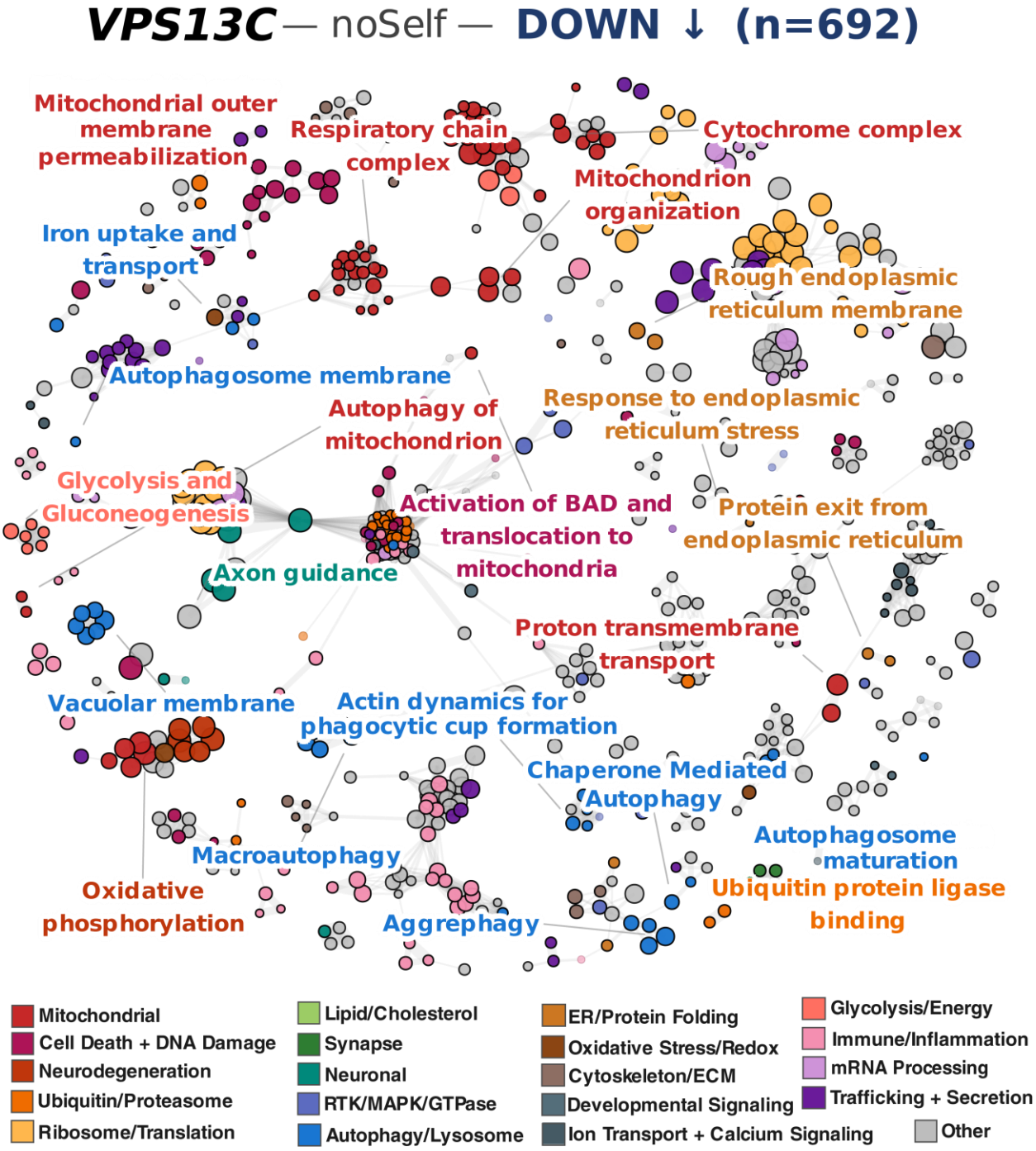


Figure 3.33 – Network plot of significantly downregulated GSEA terms nominated by *VPS13C* perturbation in neurons. Nodes represent significant GSEA results, and connections between them represent the degree of overlap behind related gene sets. Terms are manually assigned to 20 biological categories, depicted as color. Circles with a black border are GSEA tests results that remain significant after multiple testing correction. Dot size indicates the adjusted p-value of the GSEA result: larger dots are more significant.

3.4.6 Abridged summary of additional familial-PD perturbation signatures

GIGYF2

GIGYF2 encodes a translational repressor partnering with *4EHP* in ribosome-associated mRNA surveillance[64,65], and although initially linked to PD[66], its status as a familial PD gene is contested[67,68]. In our analysis 130 *GIGYF2*-perturbed cells produced significant self-knockdown (log2FC = -0.738, pAdj = 1.54×10^{-3}) with 15 upregulated genes (Figure 3.34). DEGs were dominated by ribosome and translation-related genes (*TSR1, NOP16, DRG1,* and *GTF3C5*), with corresponding GSEA terms like **ribonucleoprotein complex biogenesis** (NES = +1.77, padj = 1.7×10^{-3}) and **peptide biosynthetic process** (NES = +1.54, padj = 2.1×10^{-2}). Other notable upregulated DEGs include ferritin subunit *FTL*, contained in the term **cellular iron-ion homeostasis** (NES = +1.86, padj = 3.7×10^{-2}), and regulator of lysosome biogenesis, *TMEM251*, associated with the term **secondary lysosome** (NES = +1.84, padj = 3.0×10^{-2}) alongside *SQSTM1* and *SLC48A1*.

Another noteworthy group of downregulated signatures in *GIGYF2*-perturbed neurons comprised eight significant cholesterol and sterol biosynthesis terms. The most specific representative is **cholesterol metabolism with Bloch and Kandutsch-Russell pathways** (WikiPathways, NES = -2.29, padj = 4.6×10^{-3}). This term and its seven related terms share a 12-gene leading edge (*HMGCR, MVD, DHCR7, DHCR24, SQLE, EBP, MSMO1, TM7SF2, CYP51A1, FDPS, SCAP,* and *GPAM*) spanning the post-mevalonate sterol pathway. Ten of these twelve genes overlap the *PARK7*-perturbation's cholesterol-biosynthesis leading edge as described in §3.4.3, while *DHCR24* and *SCAP* are specific to the *GIGYF2*.

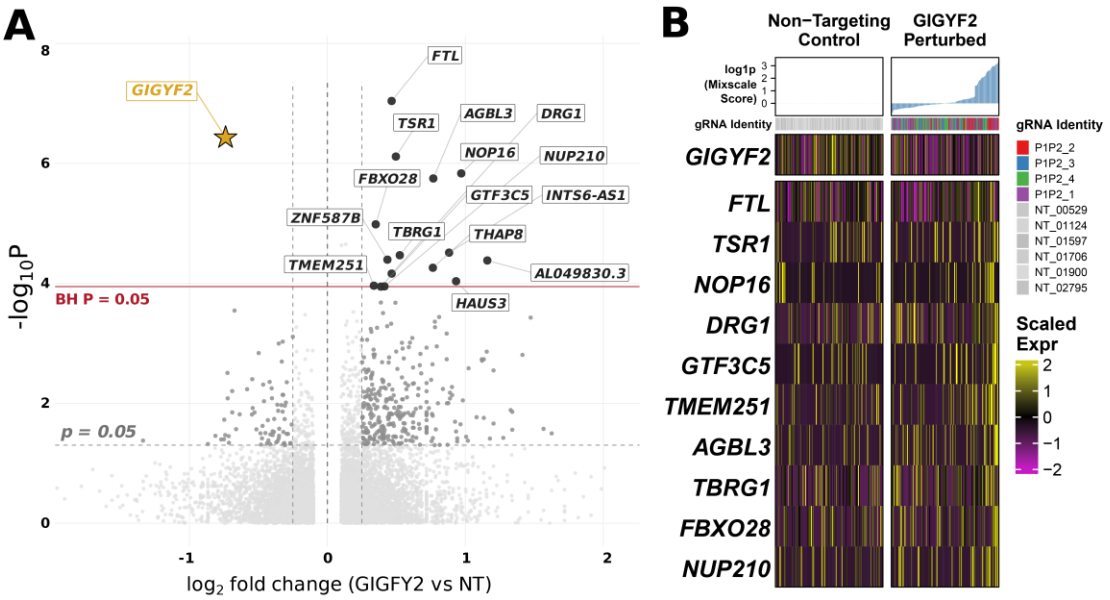


Figure 3.34 – *Overview of differentially expressed genes identified by Mixscale within GIGYF2-perturbed neurons from the familial-PD Perturb-seq experiment.* ***(A)*** *– Volcano plot indicating log2FC vs. p-value. Grey dotted horizontal line indicates a nominal p-value of 0.05, the red horizontal line denotes the Benjamini-Hochberg corrected pAdj of 0.05. Self knockdown of GIGYF2 is indicated with a gold star.* ***(B)*** *– Heatmap showing the expression of GIGYF2 itself alongside a subset of 10 upregulated DEGs detected in GIGYF2-perturbed neurons. Columns at left represent non-targeting control cells, while columns on the right indicate all GIGYF2-perturbed cells, ranked from left-to-right by increasing Mixscale score.*

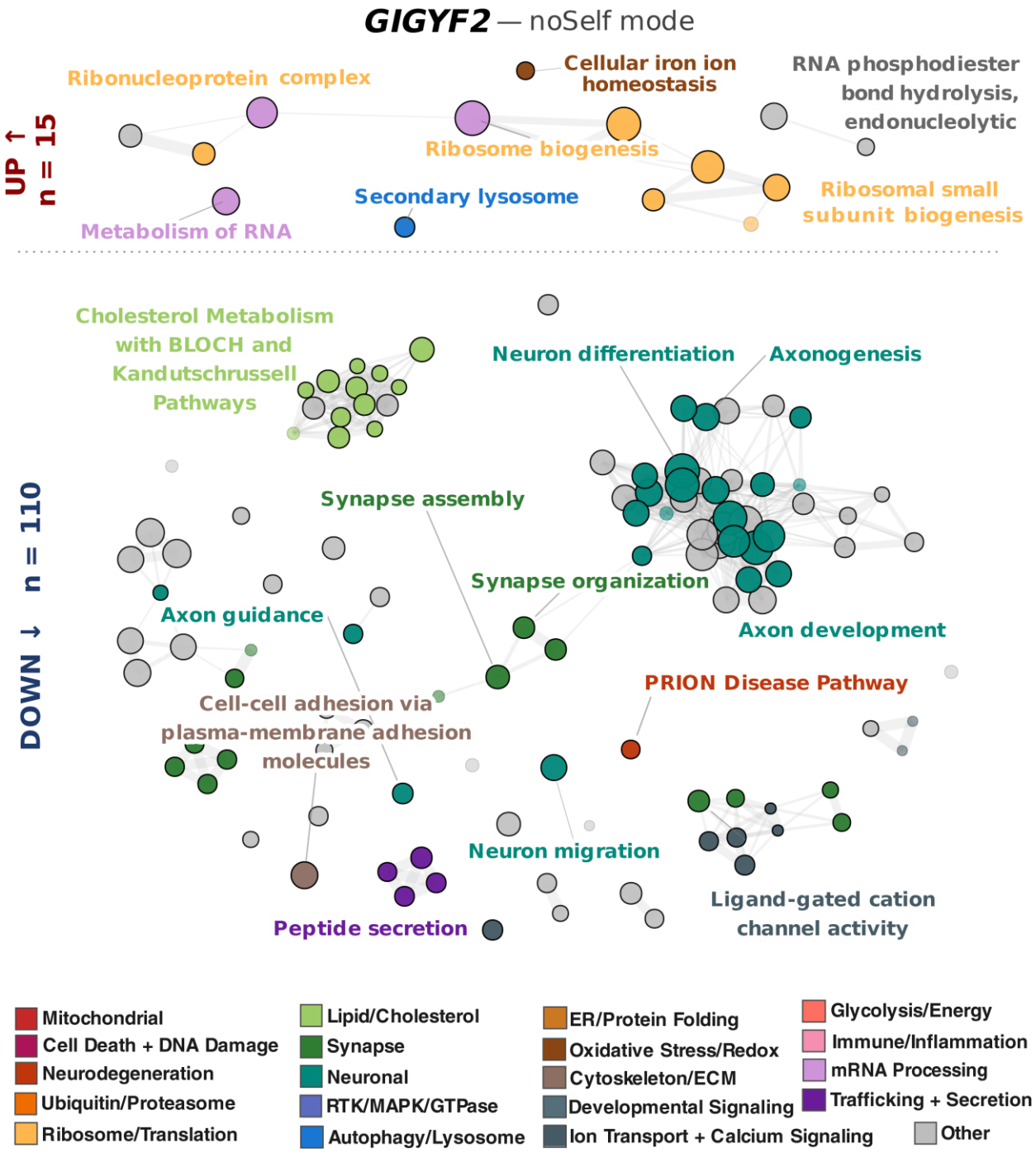


Figure 3.35 - Network plot of significant GSEA terms nominated by *GIGFY2* perturbation in neurons. Nodes represent significant GSEA results, and connections between them represent the degree of overlap behind related gene sets. Terms are manually assigned to 20 biological categories, depicted as color. Circles with a black border are GSEA tests results which remain significant after multiple testing correction. Dot size indicates the adjusted p-value of the GSEA result: larger dots are more significant.

SYNJ1

SYNJ1 encodes synaptojanin 1, a polyphosphoinositide phosphatase central to synaptic-vesicle endocytosis; biallelic loss-of-function mutations cause autosomal-recessive early-onset PD[69,70]. In the pooled neuron analysis, 178 *SYNJ1*-perturbed cells returned a modest but non-significant self-knockdown (log2FC = -0.83, BH-corrected p = 0.31) and zero significant DEGs, but 89 significant GSEA terms (4 up, 85 down). The largest group of terms captures secretory-granule and exocytosis biology, anchored by the term **secretory granule** (GO_CC, NES = -2.64, padj = 3.2×10^{-7}) with a leading edge including SNAREs (*VAMP2, SNAP29*), the synaptotagmin *SYT4*, Rab GTPases (*RAB26, RAB27B*), V-ATPase subunits (*ATP6V0C, ATP6V1A*), and the chaperones *DNAJC5* and *DNAJC13*; sister terms **exocytosis** (GO_BP, NES = -2.42, padj = 6.8×10^{-6}) and **regulated exocytosis** (GO_BP, NES = -2.46, padj = 1.0×10^{-5}) recover the same axis, consistent with *SYNJ1*'s role in synaptic-vesicle recycling.

Another group of downregulated terms capture cytosolic translation and cotranslational ER targeting, exemplified by **cytoplasmic ribosomal proteins** (WikiPathways, NES = -2.77, padj = 1.3×10^{-3}) with a leading edge consisting strictly of ribosomal genes (*RPL*/RPS**). A similar signature recurs in **protein targeting to ER** (GO_BP, NES = -3.12, padj = 6.8×10^{-6}), whose leading edge adds Sec61 translocon components (*SEC61A1, SEC61B, SEC61G*) and SRP factors (*SRP9, SRPRA*) to the same core of ribosomal genes. Finally, another set of terms nominate autophagy and lysosome biology, with terms like **chaperone mediated autophagy** (Reactome, NES = -2.35, padj = 4.7×10^{-3}) driven by leading edge genes like *HSP90AA1, HSP90AB1, HDAC6,* and *PARK7* (DJ-1).

POLG

POLG encodes the catalytic subunit of mitochondrial DNA polymerase γ; loss-of-function mutations cause a spectrum of mitochondrial disorders including progressive external ophthalmoplegia and Alpers-Huttenlocher syndrome, with parkinsonism reported as a late-onset feature, hence its inclusion in this gRNA library[71,72]. In the pooled neuron analysis, 142 *POLG*-perturbed cells produced a substantial but non-significant self-knockdown (log2FC = -2.87, BH-corrected p = 0.077), 65 DEGs (25 up, 40 down), and 41 significant noSelf GSEA terms (1 up, 40 down).

The strongest downregulated signal is a cytosolic translation and cotranslational ER-targeting program, anchored by **cytosolic ribosome** (GO_CC, NES = -2.19, padj = 3.1×10^{-7}) with a leading edge of large- and small-subunit ribosomal proteins (*RPL*/RPS** genes) plus *RACK1, EIF2A, ABCE1,* and *EEF1A1*. Similar terms like **SRP-dependent cotranslational protein targeting to membrane** (GO_BP, NES = -2.13, padj = 4.2×10^{-5}), additionally contain *SRP14*, *SRP19*, and *SSR3*. These signatures recapitulate the downregulation of translation seen for *SYNJ1* and *PARK2* (and opposite the upregulation in *PARK7*, *VPS35*, and *FBXO7*); cross-perturbation interpretation is developed in §3.6.

Interestingly, the single coherent upregulated signal is a myogenic / smooth-muscle program. **Smooth muscle contraction** (GO_BP, NES = +1.97, padj = 3.9×10^{-2}) is driven by

a leading edge *of NMU, MYOCD, MYLK, ACTA2, PDE4D, PRKG1,* and *SULF1*, and the strict upregulated DEGs anchor the same biology with large effect sizes (NMU log2FC = +2.65, MYOCD = +2.71, NNAT = +3.70, MYLK = +1.77, ACTA2 = +1.01). This pattern is unusual in a perturbation of a mitochondrial replication factor, and we report it as an unexpected transcriptional response rather than a mechanistically interpreted axis.

PRKN (PARK2)

PRKN encodes the E3 ubiquitin ligase Parkin, the downstream effector of the PINK1-Parkin mitophagy pathway (§1.3.1)[73,74]. In the pooled neuron analysis, determination of self-knockdown within *PARK2*-perturbed cells failed to return a log2FC or p-value for *PARK2* itself. Despite this, Mixscale still identified two significant upregulated DEGs including the mitoribosomal protein *MRPS9* and the lncRNA *ZNF32-AS3*. Nonetheless, enrichment analyses returned a substantial signature: 52 significant noSelf terms, 1 up, 51 down.

The dominant downregulated signal is a cytosolic translation and cotranslational ER-targeting program. **Cytosolic ribosome** (GO_CC, NES = -2.57, padj = 7.0×10^{-5}) is the most specific representative, with a leading edge of large- and small-subunit ribosomal proteins (*RPL7, RPL10, RPL12, RPL38, RPS2, RPS6, RPS9, RPS11, RPS15A*) plus *RACK1* and ubiquitin gene *UBA52*. The same axis recurs in cytoplasmic ribosomal proteins (WikiPathways, NES = -2.30, padj = 2.0×10^{-3}), **translational initiation** (GO_BP, NES = -2.34, padj = 7.1×10^{-4}), and **protein targeting to** ER (GO_BP, NES = -2.54, padj = 4.2×10^{-4}). *PARK2* thereby joins *SYNJ1* in showing a translation-axis downregulation, opposite the upregulation seen in *PARK7*, *VPS35*, and *FBXO7*; cross-perturbation interpretation is developed in §3.6.

A second, more PD-direct cluster nominates monoaminergic vesicular biology. **Monoamine transport** (GO_BP, NES = -2.21, padj = 2.7×10^{-2}) is driven by a leading edge including the synucleins *SNCA* and *SNCG*, the vesicular monoamine transporter *SLC18A1* (VMAT1), monoamine oxidase B (*MAOB*), and the synaptotagmins *SYT4*, *SYT5*, and *SYT6*. The related term **neuron death** (GO_BP, NES = -1.76, padj = 1.5×10^{-2}) shares the *SNCA*/*SNCG* anchor and adds *AIMP2* (an established Parkin substrate)[75], the synaptic chaperone *DNAJC5*, and the ER chaperones *HSPA5* (BiP) and *HSPD1* (HSP60).

3.5.4 Perturbation signatures unique to the late-progenitor cluster

NR4A2 perturbation in late progenitors activates senescence and cell-death programs

NR4A2 (NURR1) is a transcription factor and master regulator of midbrain DA fate required for the differentiation and maintenance of midbrain DA neurons. In the combined neuron analysis of 411 *NR4A2*-perturbed cells, Mixscale failed to produce a signature due to incomplete convergence during its expectation-maximization scoring step. In cases like this for sparse or transcriptionally heterogeneous perturbations, cells are not assigned a Mixscale perturbation score and fail to undergo differential expression testing. Thus, we could not recover a differential expression signature for *NR4A2*-perturbed neurons.

In the late midbrain progenitor cluster, however, the analysis of 669 perturbed cells against 398 non-targeting controls yielded 7 significantly upregulated DEGs (*FOS, KLF10, JUN, ATF3, SLC2A3, EGR1,* and *FAM49A*) as well as 267 significant noSelf enrichment terms, all upregulated.

The 51 upregulated pathways are dominated by senescence, inflammatory secretion, and pro-apoptotic regulation. The term **senescence-associated secretory phenotype** (Reactome R-HSA-2559582) reaches NES = +1.97 (padj = 0.005); **positive regulation of cell death** (GO:0010942) is found at NES = +1.84 (padj = 0.031); **TNFa Signaling Via NFKB** reaches NES = +2.11; **Hypoxia** reaches NES = +2.09; **P53 Pathway** reaches NES = +2.05; and **TGFbeta Signaling Pathway** reaches NES = +2.12 (the strongest single magnitude in the set). The combined pattern is consistent with developmental senescence engaged by cells that cannot complete the mDA fate-acquisition program: loss of NR4A2 in cells still committing to the DA lineage produces simultaneous activation of pluripotency programs (loss of fate restriction), senescence (cell-cycle exit without mature differentiation), pro-inflammatory secretion (SASP), and pro-apoptotic priming. This phenotype is not visible in the pooled-neuron analysis because the pooled neurons have already passed through the developmental window in which *NR4A2* acts. The *NR4A2* signature observed is presented as an internal control of sorts, demonstrating a cell-state specific signature that is uniquely detected in late midbrain progenitors but not in our *NR4A2*-perturbed neurons.

EIF4G1 perturbation in late progenitors downregulates the KEGG Parkinson disease pathway

EIF4G1 (PARK18) is a translation-initiation scaffolding factor whose perturbation in the late-progenitor cluster produces a distinct and PD-relevant signature. *EIF4G1* is the only perturbation in any cluster of either familial-PD or GWAS-PD experiment which significantly downregulates the KEGG **Parkinson disease** pathway (hsa05012) at padj_global < 0.05, both with and without target gene removal: noSelf NES = −2.12 (padj_global = 0.005, padj_db = 0.001); withSelf NES = −2.08 (padj_global = 0.015, padj = 0.003). The withSelf-noSelf concordance indicates that the signal is not driven by *EIF4G1* itself appearing in the leading edge. The leading edge for the KEGG PD pathway in *EIF4G1*-perturbed late progenitors is composed of mitochondrial respiratory chain genes (*NDUFA12, NDUFA8, NDUFB3, NDUFA6, NDUFA7, NDUFC2, UQCR11, COX7B, ATP5MC1, ATP5PB, ATP5MC3, ATP5PF, SLC25A5, VDAC3*), proteasome subunits (*PSMC3, PSMD13, PSMD8*), the antioxidant gene *SOD1*, two tubulin isoforms (*TUBA1A*, *TUBA1B*), the calmodulin gene *CALM2*, the G protein gene GNAI2, and the ubiquitin-conjugating enzyme *UBE2L3*. The leading-edge composition is consistent with a translation-dependent program in which loss of *EIF4G1* disproportionately affects ribosome-association and translation efficiency of the KEGG PD pathway transcript set, with potential downstream consequences for the expression of mitochondrial respiratory chain components. *EIF4G1* was nominated as a familial-PD gene (PARK18) by Chartier-Harlin and colleagues based on a missense variant segregating with autosomal-dominant late-onset PD[76], though subsequent meta-analyses have called the

gene's penetrance and pathogenicity into question. The specific signature reported here does not address the question of whether *EIF4G1* variation contributes to PD risk in patients, but does demonstrate that *EIF4G1* knockdown in hPSC-derived late progenitors produces transcriptional convergence on the canonical KEGG PD pathway through a translation-dependent mechanism, consistent with EIF4G1's role in cap-dependent translation initiation. The finding is reported as a single-perturbation observation in the cell-state-conditional analysis, parallel to the *FBXO7* synapse-DOWN observation in pooled neurons (§3.4.4), and motivates further targeted analysis of *EIF4G1* in additional cellular contexts.

3.5 CRISPRi Perturb-seq of GWAS-nominated PD risk genes

3.5.1 Characterization of the GWAS Perturb-seq dataset

The companion screen of GWAS nominated PD risk genes perturbed 310 PD-nominated candidate risk genes in the same WIBR3-derived DA neuron system as the familial-PD experiment (hereafter “FPD”). Raw 10x output comprised 103,281 cells; after singlet filtering and QC, 62,245 single cells entered clustering at Leiden resolution 0.10. Six clusters resolved at this resolution. Mixscale analyses in this chapter were performed on cluster 2 (13,163 cells, 437 non-targeting controls), which contained midbrain specified DA-neurons. The remaining clusters (22,619, 20,626, 3,811, 1,397, and 629 cells) comprised earlier midbrain progenitor states and minor off-target lineage fates totaling 5,837 cells (9.4% of the dataset) in the three smallest clusters; these were excluded from the perturbation analyses (Figure 3.36).

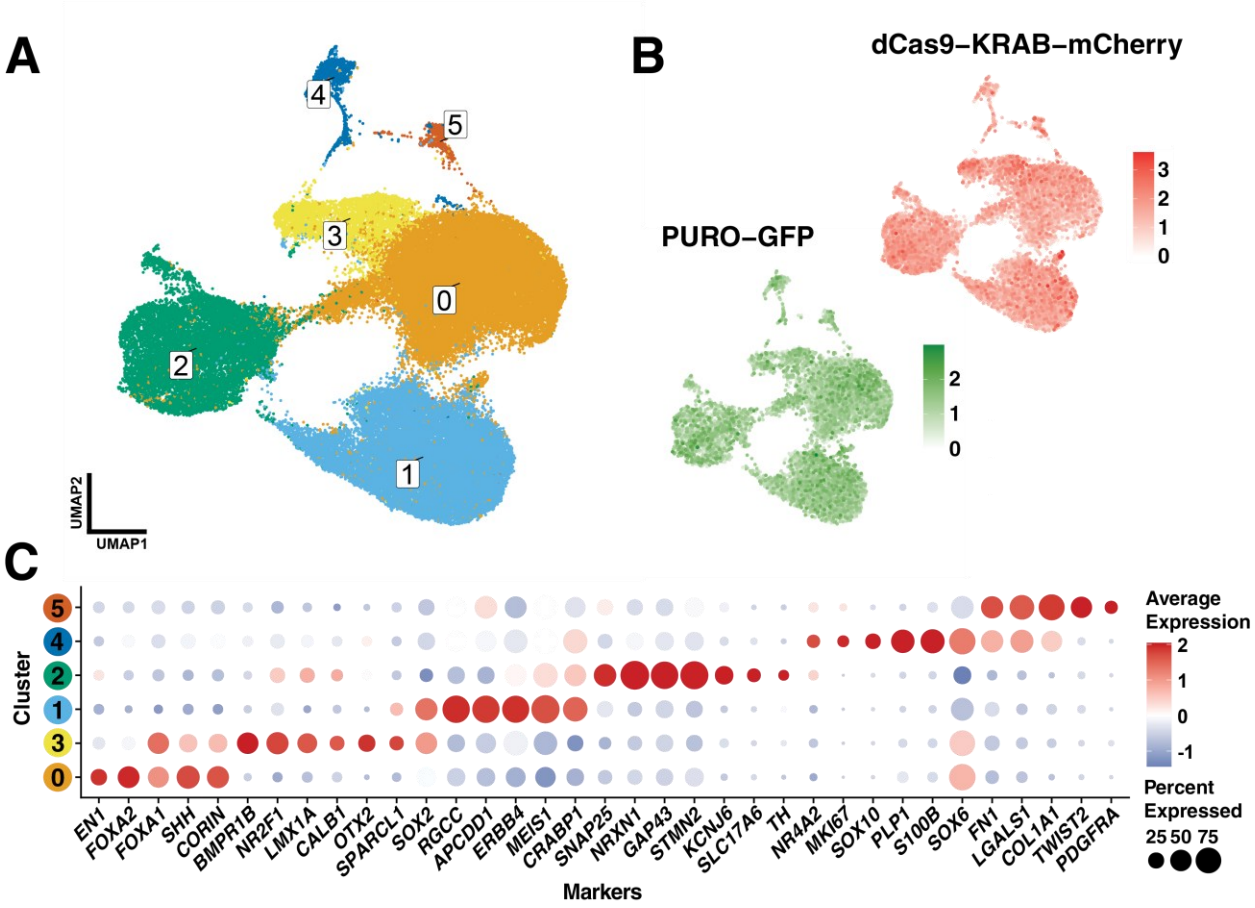


Figure 3.36 – *Overview of the GWAS-PD Perturb-seq experiment, covering 62,245 cells with single-gRNA for 310 perturbations.* ***(A)*** UMAP reduction of all singlets, showing 6 distinct cell type clusters. ***(B)*** Feature plots showing expression of the dCas-KRAB-mCherry transgene or the PURO-eGFP reporter delivered alongside gRNAs. ***(C)*** *Dotplot of cell-type markers expressed among the 6 clusters identified in our GWAS-PD experiment. This experiment identified similar progenitor and mixed-neuron identity clusters, in addition to two cell-types absent in the famililal-PD dataset; cluster 4 appears to be oligodendrocyte precursor cell (OPC) identity, while cluster 5 appears to be stromal.*

3.5.2 Perturbation quality control and analysis overview

Per-perturbation differential expression and pre-ranked GSEA followed the same Mixscale plus eight-database pipeline used for the familial-PD experiment ("FPD", §3.4.1), with noSelf as the default mode. Power constraints are more severe than in FPD: across all 310 perturbations targeting the DA-lineage cluster, the median per-perturbation cell count is 37 (IQR 28-52, range 5-123), a consequence of distributing guide representation across 310 targets rather than 41. Mixscale analysis requires a sufficient number of cells to properly model the effects of perturbation compared to non-targeting controls; perturbations falling below this threshold cannot be analyzed and are excluded from the perturbation analyses below. We focus on two in-depth cases below where self-knockdown of the perturbed gene itself is found to be signficant, at BH-corrected p_weight < 0.05 and LFC < 0.25 (Figure 3.37).

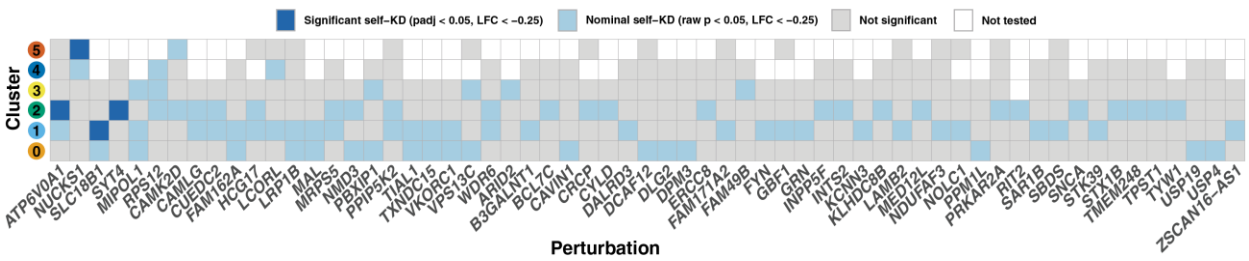


Figure 3.37 – *Heatmap indicating perturbations with significant self-knockdown in the GWAS-PD Perturb-seq experiment, with results shown for each perturbation's cluster-specific analysis. Perturbations shown represent those within the full GWAS-PD library which had self-knockdown at nominal p-value cutoff in at least one cluster (raw p < 0.05, LFC < −0.25). Dark blue indicates cases where self-knockdown was still significant after Benjamini-Hochberg correction of the Mixscale-derived p-value (pAdj < 0.05, LFC < −0.25).*

3.5.3 Overview of notable GWAS-PD perturbation signatures

The Mixscale and GSEA analyses for perturbations in the neuron cluster of the GWAS-PD experiment yielded a vast set of results. Due to the size of the library and limitations in statistical power, this section focuses on just two perturbations which had adequate cell numbers and indicated significant self-knockdown.

SYT4

SYT4 encodes synaptotagmin-4, which is required for activity-dependent somatodendritic dopamine release in ventral midbrain DA neurons[77,78]. Turnover of SYT4 protein is regulated via parkin-mediated polyubiquitination, and *Prkn*-knockout mice have been shown to accumulate striatal SYT4, and exhibit a reduction in evoked striatal dopamine release[79]. In the DA-lineage neuron cluster, analysis of 54 *SYT4*-perturbed cells against 437 non-targeting controls indicated significant self-knockdown (log2FC = -1.74, BH-adjusted p = 1.23×10^{-2}) with 18 significantly upregulated genes and expression *SYT4* itself being the only gene with a significant decrease. Among the 18 upregulated DEGs, is the autophagy initiator *ATG7*, regulator of synaptic-vesicle endocytosis *FCHSD2* and the Golgi-tethering subunit *COG6*. The mitochondrial ribosomal component *MRPL48* and mediator of assembly *MTG2* are significantly upregulated as well. GSEA analysis for *SYT4* identified 225 significant noSelf terms (5 up, 220 down).

The down-direction signature among GSEA terms is dominated by cytoskeletal genes and ubiquitin-proteasome signatures. **Structural constituent of cytoskeleton** (NES = −3.00 padj = 6.7×10^{-8}), has a leading edge that includes microtubule subunits (*TUBA1A, TUBB, TUBB2A, TUBB2B, TUBB3, TUBA1B*), neurofilaments (*INA, NEFM, NEFL*), and actin / Arp2/3 components (*ACTB, ACTG1, ARPC5*). The term **Parkin–ubiquitin–proteasomal pathway** (WikiPathways, NES = −2.69, padj = 3.4×10^{-4}), includes *PRKN* itself in the leading edge alongside *HSPA8*, the proteasome subunits *PSMD8* and *PSMC5*, and the E1 ubiquitin-activating enzyme *UBA1*. The KEGG terms **Parkinson's disease** (NES = −2.61, padj = 2.0×10^{-7}) and **substantia nigra development** (NES = −2.83, p = 1.7×10^{-5}) are also significantly downregulated in *SYT4*-perturbed neurons.

ATP6V0A1

ATP6V0A1 encodes the a1 isoform of the V0 sector of the vacuolar H^+-ATPase, the proton pump that acidifies lysosomes, late endosomes, and synaptic vesicles. In the DA-lineage neuron cluster, 52 *ATP6V0A1*-perturbed cells were analyzed against 437 non-targeting controls. Mixscale identified significant self-knockdown (log2FC = −1.90, BH-adjusted p = 0.005), with 20 significantly upregulated DEGs and GSEA yielding 153 significant terms with the noSelf approach (130 down, 23 up) (Figure 3.38).

Translation and ribosome signatures are suppressed in *ATP6V0A1*-perturbed neurons. Representative terms like **eukaryotic translation initiation** (GO_BP, NES = −2.27 padj = 2.4×10^{-5}) are primarily driven by *RPL*/RPS** genes. Expression of the V-ATPase complex itself is repressed, with the term **proton-transporting two-sector ATPase complex** (NES = −1.99, padj = 7.5×10^{-3}) showed decreased expression, as well as **mitochondrial proton-transporting ATP-synthase complex** (NES = −1.94). Oxidative phosphorylation and electron-transport chain related terms are coordinately down-regulated (Hallmark, **Oxidative Phosphorylation**, NES = −2.24, p = 2.0×10^{-7}) as well.

The 20 upregulated DEGs for *ATP6V0A1* suggest a potential compensatory synaptic-membrane and cell-adhesion program, consistent with the positive enrichment seen for **synaptic-membrane** (NES = +1.86, p = 5.5×10^{-4}) and **cell-cell adhesion** (NES = +1.65 to +1.87) terms in the noSelf landscape. The strongest up-regulated DEGs are the synaptic glypican *GPC4*, synaptic adhesion molecules (*NCAM2*, *PCDH11X*, *PCDHB9*) and the postsynaptic muscarinic receptor *CHRM2*. Two additional upregulated DEGs, *ATP2B1* and *PLK2* implicate mediators of calcium homeostasis and synaptic plasticity, respectively. Notably, *PLK2* has been shown to phosphorylate α-synuclein at serine residue S129, a post-translational modification strongly associated with α-synuclein aggregation, with about 90% of α-synuclein immunoreactivity in Lewy bodies reflecting the S129 form[80,81].

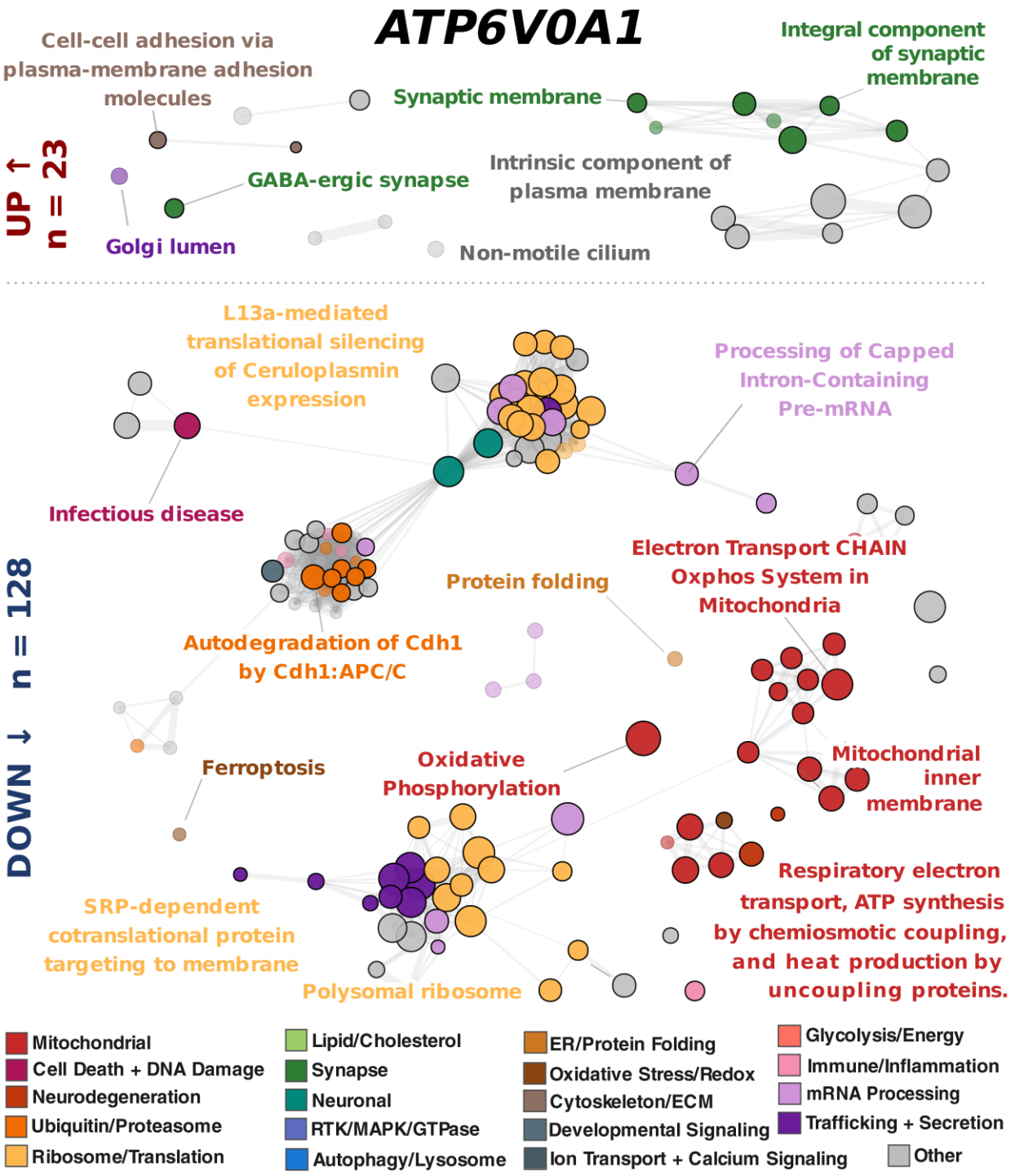


Figure 3.38 - Network plot of significant GSEA terms nominated by *ATP6V0A1* perturbation in neurons of the GWAS-PD experiment (cluster 2). Nodes represent significant GSEA results, connections between them represent the degree of overlap behind related gene sets. Terms are manually assigned to 20 biological categories, depicted as color. Circles with a black border are GSEA tests results that remain significant after multiple testing correction.

3.6 Convergent themes across Familial-PD and GWAS-PD Perturb-seq experiments

Ten perturbations appear with results in both the Familial-PD and GWAS-PD experiments: *APOE, CTSB, GPNMB, LRP1B, LRRK2, PITX3, SLC6A3, SNCA, SSR1,* and *VPS13C*. The familial PD experiment's signatures (§3.3–§3.4) define a transcriptional reference map of sorts, nominating the biological pathways that are disrupted by particular familial-PD perturbations in our hPSC-derived midbrain neurons. In §3.6.1–§3.6.4, we ask which GWAS-PD perturbations engage the same biological axes in that experiment's corresponding set of neuron-specific analyses, with focus on the ten perturbations present in both experiments as well as recurrent enrichment signatures seen across the broader set of all familial-PD and GWAS-PD perturbations.

We begin with *VPS13C*, one of the chapter's central familial-PD perturbations discussed in §3.4.5. In both experiments, *VPS13C* knockdown produced strong, concordant downregulation of the GO_BP term **V-ATPase / proton transmembrane transport** (NES = -2.25, padj = 4.4×10^{-7} in familial-PD neurons; NES = -2.10, padj = 1.1×10^{-3} in GWAS-PD neurons) as well as a decrease in signatures of mitochondrial respiration and cytosolic translation. The recurrence of these axes as defined in the familial-PD analysis within an entirely independent experiment served as proof-of-concept that motivated the broader comparisons developed in this section.

The cross-dataset evaluation however, is bounded by three key considerations: subtleties in cell-type composition, sequencing depth (lower in the GWAS-PD experiment), and per-perturbation cell counts. These implications are discussed in depth within §3.7.

The KEGG term **Parkinson's Disease** (Figure 3.39 and 3.40) is the clearest and most compelling signature across both experiments. Five of the perturbations described in this chapter engage this term at padj < 0.05. Perturbed neurons from *FBXO7* and *VPS35* from the familial-PD experiment show upregulation of this signature, while *SYT4* and *ATP6V0A1* in GWAS-PD show the signature is downregulated, as does *VPS13C* perturbation in both experiments. The *FBXO7* and *VPS35* upregulation is consistent with disruption of clearance pathways producing compensatory transcriptional changes rather than reduced gene expression; these directionality differences across familial-PD perturbations are discussed in §3.7. The following sections of text examine several noteworthy biological axes detected across neurons from both experiments: V-ATPase suppression and lysosomal acidification (§3.6.1); autophagy and protein-aggregate clearance (§3.6.2); mitochondrial dysfunction (§3.6.3); and translation and ribosome biogenesis (§3.6.4). Finally, §3.6.5 closes with limitations regarding the interpretation of these cross-experiment signals.

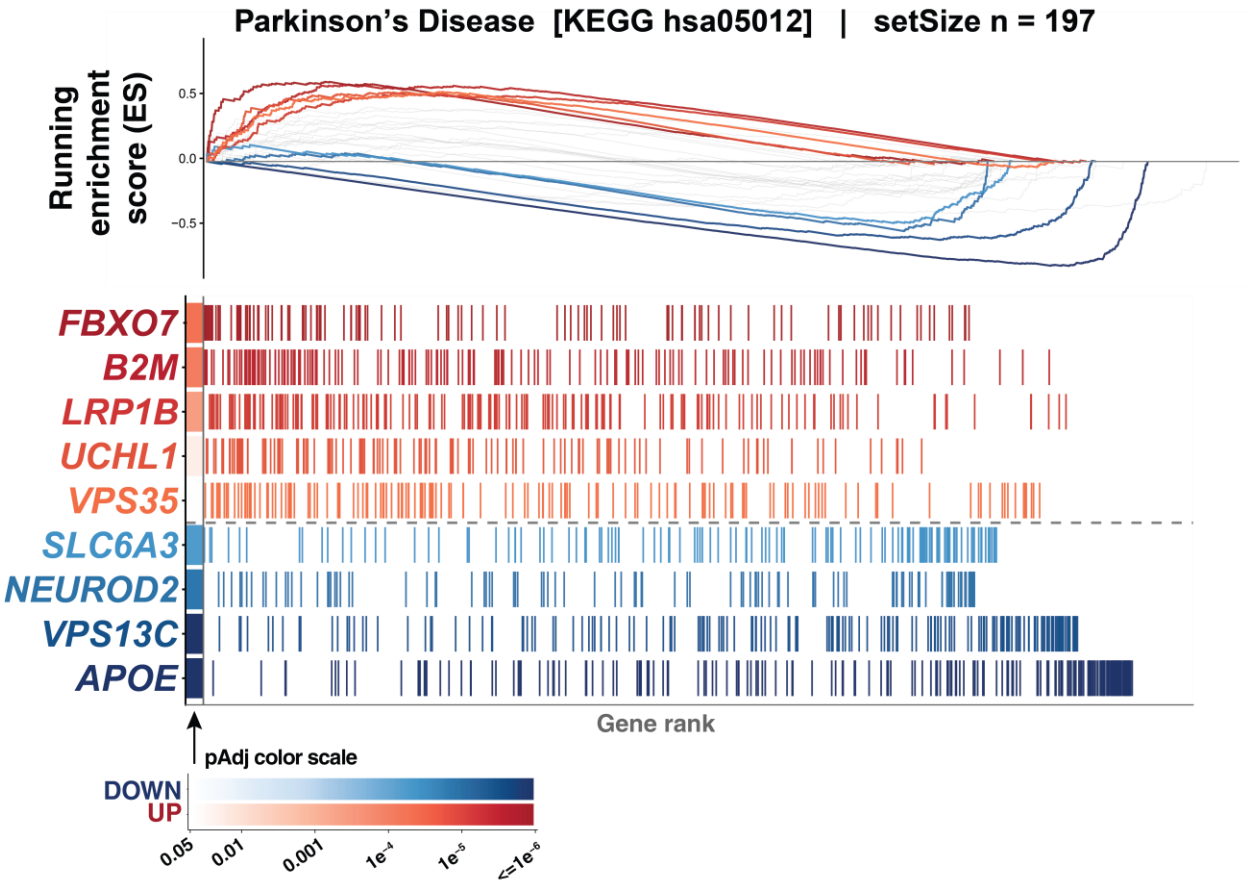


Figure 3.39 – *GSEA results for the KEGG term* "Parkinson's Disease", *representing all significant hits among perturbed* ***neurons of the familial-PD experiment****. Only perturbations with significant enrichment (red) or depletion (blue) of this gene set's expression are shown. "Barcode plots" on the bottom indicate where in each rank-ordered gene list we encounter genes belonging to this term's gene set.*

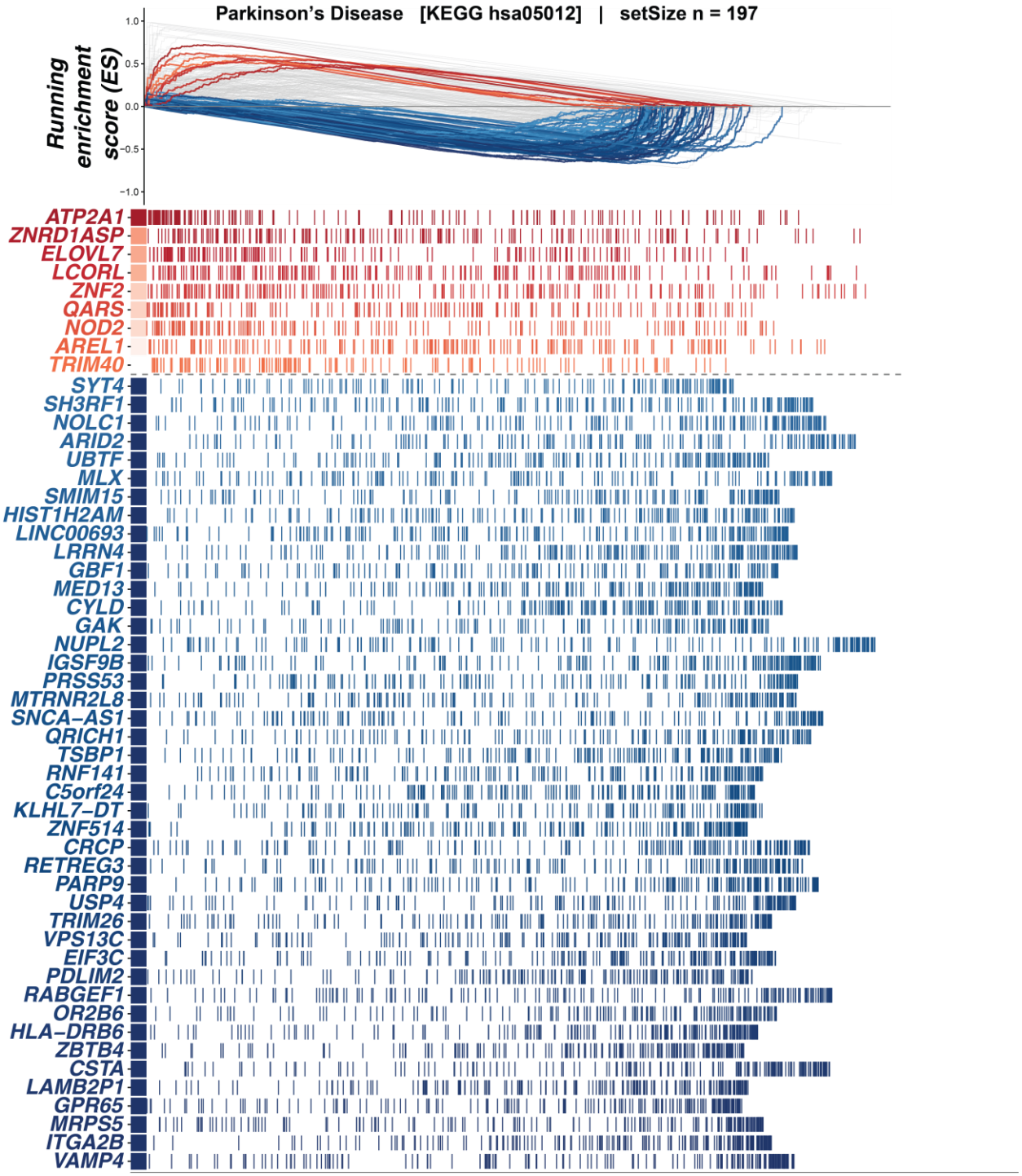


Figure 3.40 – *GSEA results for the KEGG term* "Parkinson's Disease", *representing all significant hits among* ***perturbed neurons of the GWAS-PD*** *experiment. Total of 9 UP and 78 DOWN results significant among all perturbation tested. Only a subset of the top ~40 perturbations with most significant suppression of this term's signature are shown, to save space. "Barcode plots" on the bottom indicate where in each rank-ordered gene list we encounter genes belonging to this term's gene set.*

3.6.1 V-ATPases / lysosomal acidification

VPS13C perturbation in both experiments produced strong, concordant downregulation of the **V-ATPase / proton transmembrane transport** gene set (GO_BP, NES = -2.25, padj = 4.4×10^{-7} in familial-PD neurons; NES = -2.10, padj = 1.1×10^{-3} in GWAS-PD neurons). The leading edges of these enrichment results contained the same 13 V-ATPase subunits in both experiments (V0 sector: *ATP6V0A1*, *ATP6V0A2*, *ATP6V0B*, *ATP6V0C,* and *ATP6V0D1*; as well as the V1 sector: *ATP6V1A*, *ATP6V1B2*, *ATP6V1C1*, *ATP6V1D*, *ATP6V1E1*, *ATP6V1F*, *ATP6V1G1*, and *ATP6V1H*).

VPS13C encodes a lipid-transfer protein which facilitates lipid transfer from the endoplasmic reticulum to lysosome and is rapidly recruited to lysosomes following sensing of lysosomal damage[60–62,82]. Loss-of-function studies of *VPS13C* provide support for this observation: lentiviral shRNA knockdown of *VPS13C* in iPSC-derived DA neurons impairs lysosomal hydrolytic activity and acidification, alongside disruption of phospho-Rab10-mediated lysosomal stress response[82]. The V-ATPase suppression we observe after CRISPRi perturbation is consistent with these observations of lysosomal dysfunction, though we cannot evaluate functional effects of these responses with transcriptional results alone.

Direct perturbation of a V-ATPase subunit gene predictably elicits this same signature. *ATP6V0A1,* which encodes the A1 isoform of the V0 sector of the V-type ATPase was one of the few perturbations to show significant self-knockdown in the GWAS-PD experiment (log2FC = -1.90, BH-adjusted p = 5.0×10^{-3}). The transcriptional signature of *ATP6V0A1* perturbation consequently indicated a coordinated suppression of V-ATPase complex expression, as nominated by the term **proton-transporting two-sector ATPase complex** (GO_CC, NES = -1.99, padj = 7.5×10^{-3}).

Two additional familial-PD perturbations nominate this V-ATPase signature. *SYNJ1* knockdown similarly downregulates the **proton-transporting two-sector ATPase complex** term (GO_CC, NES = -2.07, padj = 3.0×10^{-2}), broadening the support for V-ATPase as a point of convergence within this experiment. *FBXO7* knockdown, however, produces a paradoxical upregulation of V-type ATPase-specific terms (V-type ATPase complex, GO:0033176, GO_CC, NES = +1.86, padj = 2.0×10^{-3}; vacuolar V-type ATPase complex, GO:0016471, NES = +1.84, padj = 3.6×10^{-3}); this directional shift is consistent with the broader pattern of directionality differences across familial-PD perturbations discussed in §3.7, where gene-set-level upregulation following certain perturbations could reflect compensatory transcriptional responses to proteasome disruption rather than an increase in V-ATPase activity itself.

3.6.2 Autophagy and protein-aggregate clearance

Aggrephagy, which involves the selective autophagic clearance of protein aggregates[83,84] is identified as one of the clearest autophagy-related signatures in both our familial-PD and GWAS-PD experiments. In the GWAS-PD experiment, *SYT4* and *GPNMB* knockdown both indicated downregulation of the Reactome term **aggrephagy** (*SYT4* NES = -2.53, padj =

6.1×10^{-5} and *GPNMB* NES = -2.4, padj = 3.7×10^{-5}), with 47 other GWAS-PD perturbations nominating the term in the same direction. Aggrephagy is a suppressed signature in the familial-PD experiment as well, nominated by both *SYNJ1* and *VPS13C* perturbations (NES = -2.23 padj = 1.5×10^{-2}; NES = -1.89, padj = 1.3×10^{-3}, respectively). Additionally, *VPS13C* and *SYNJ1* knockdown in the familial-PD experiment each indicate suppression of multiple autophagy sub-pathway gene sets including **Aggrephagy**, **Macroautophagy**, **Selective autophagy**, and **Chaperone-Mediated Autophagy**.

Gene-level findings add nuance to these signatures. The presence of *PRKN* in the leading edge for *SYT4*'s aggrephagy signature points to the *SYT4* parkin-substrate biology highlighted in §3.5.3. Alongside it are *HSP90AA1*, the ubiquitin-conjugating enzyme *UBE2V1*, and ubiquitin itself (*UBC*), together with the microtubule and dynein machinery responsible for trafficking ubiquitinated aggregates to autophagosomes (*TUBA1A*, *TUBA1B*, *TUBB2A*, *TUBB2B*, *TUBB3*, *TUBB4A*, *TUBB4B*, *DYNC1I2*, *DYNC1LI2*, and *DYNLL1*). The strongest significant DE genes in *SYT4* and *GPNMB* perturbations, however, oppose the directionality of the aggrephagy signature. For example, *ATG7* (autophagy initiation; log2FC = +0.86, BH-adjusted BH-corrected p = 2.4×10^{-4}) is significantly upregulated under *SYT4* knockdown, and *PRKAA1* (AMPK alpha-1; log2FC = +1.11, BH-corrected p = 0.001) is upregulated under *GPNMB* knockdown. The aggrephagy gene-set signal appears to reflect a coordinated small-magnitude downregulation of many genes involved in the aggrephagy, while specific upstream autophagy regulators are upregulated at substantially larger magnitudes. Together these results suggest an activation of autophagy gene expression even though downstream clearance machinery is coordinately suppressed.

The familial-PD experiment reproduces this axis to a similar extent. *VPS13C* knockdown suppresses four particular autophagy-related terms (**Aggrephagy** NES = -1.89; **Macroautophagy** NES = -1.88; Selective **autophagy** NES = -1.78; and **Chaperone-Mediated Autophagy** NES = -1.89; all padj < 0.05), and *SYNJ1* knockdown suppresses three at similar magnitudes (**Aggrephagy** NES = -2.23; Selective **autophagy** NES = -1.93; **Chaperone-Mediated Autophagy** NES = -2.35). Although our initial thinking was that the knockdown of genes involved in lysosome homeostasis (*VPS13C*), endosomal recycling (*SYNJ1*) might rather elicit upregulation of aggrephagy genes as a compensatory mechanism, we are limited by what our transcriptional results alone can explain. Without follow-up experiments to replicate these findings and explore possible explanations, we cannot definitively attribute the downregulation to a specific functional outcome.

3.6.3 Mitochondrial dysfunction

Results from both our familial-PD and GWAS-PD screens recover an established signature of Mitochondrial Complex I deficiency[85]. *VPS13C* knockdown produced strong downregulation of Complex I genes in both experiments, exemplified by the term **mitochondrial electron transport, NADH to ubiquinone** (GO_BP, NES = -2.06, padj = 3.8×10^{-4} in familial-PD neurons; NES = -2.04, padj = 1.9×10^{-2} in GWAS-PD neurons). Additionally, a broader **oxidative-phosphorylation** gene set was similarly suppressed

(Hallmark, NES = -2.05, padj = 2.9×10^{-7} for familial-PD experiment; NES = -2.45, padj = 2.5×10^{-9} in GWAS-PD). The leading edges of these results contained canonical Complex I subunits (*NDUFA*, *NDUFB*, *NDUFS*, *NDUFV* family genes) alongside *PARK7* and *PINK1* in both experiments as well. *VPS13C*-perturbation was also found to suppress the PINK1-PRKN mediated mitophagy gene set (R-HSA-5205685, Reactome) in the GWAS-PD experiment (NES = -1.81, padj = 4.4×10^{-2}).

Additional within-experiment perturbations contribute to this axis. In the GWAS-PD experiment, *ATP6V0A1* knockdown in neurons produced strong **OXPHOS** suppression (Hallmark, NES = -2.24, padj = 2.0×10^{-7}; KEGG hsa00190 NES = -2.11, padj = 6.2×10^{-4}), with *LRRK2* knockdown producing a more moderate OXPHOS suppression (NES = -1.73, padj = 4.2×10^{-2}; KEGG hsa00190 NES = -1.94, padj = 1.4×10^{-2}). In the familial-PD experiment, *SLC6A3* and *PITX3* knockdown produced suppresion of OXPHOS genes as well (*SLC6A3* NES = -2.13, padj = 2.8×10^{-4}; *PITX3* NES = -1.84, padj = 5.7×10^{-3}).

A difference in directionality was apparent within the familial-PD experiment. *FBXO7* knockdown produced an *upregulation* of the same **OXPHOS** and **PINK1-PRKN mitophagy** gene sets that *VPS13C* knockdown downregulates (Hallmark, NES = +1.91, padj = 9.0×10^{-5}; and NES = +1.71, padj = 2.0×10^{-2}, respectively). Knockout of *Fbxo7* in mice and cellular models previously identified an impairment in the recruitment of Parkin to depolarized mitochondria, with reduced mitophagic clearance of damaged mitochondria[53]. The transcriptional *upregulation* of **PINK1-PRKN Mitophagy** observed under *FBXO7* CRISPRi knockdown likely reflects compensatory transcriptional responses to deficient mitochondrial-quality-control function.

3.6.4 Translation and ribosome biogenesis

Both our familial-PD and GWAS-PD screens recover broad transcriptional disruption of translation and ribosome-biogenesis machinery, with 17 of 30 familial-PD perturbations and 140 of 282 GWAS-PD perturbations reaching padj < 0.05 on at least one translation- or ribosome-related gene set. *VPS13C* knockdown produced strong concordant downregulation of translation in both experiments, exemplified by Eukaryotic Translation Elongation (Reactome, NES = -2.92, padj = 2.9×10^{-9} in familial-PD neurons; NES = -3.31, padj = 2.9×10^{-9} in GWAS-PD neurons). The leading edges of the results in both experiments were dominated by the canonical *RPL* and *RPS* ribosomal-protein gene families, alongside translation elongation factors (*EEF1A1, EEF1B2, EEF2*).

The pan-perturbation pattern is overwhelmingly skewed toward downregulation in the GWAS-PD experiment. Of the 128 perturbations reaching significance for **Eukaryotic Translation Elongation** (Reactome, padj < 0.05), 117 of them (91%) indicated downregulation. The strongest GWAS-PD downregulators include *PRSS53* (NES = -4.06), *EIF3C* (NES = -4.00), *SH2B1* (NES = -3.97), *KLHL7-DT* (NES = -3.73), and *HLA-DQA2* (NES = -3.68). Among the perturbations described earlier in this chapter section, *LRRK2* knockdown

reached NES = -2.52 (padj = 1.6×10^{-6}), *ATP6V0A1* reached NES = -2.28 (padj = 2.4×10^{-5}), with *SNCA* and *CTSB* showing a more moderate effect with NES = -1.84 and -1.94 respectively.

This difference in directionality for ribosome and translation related signatures was more apparent and evenly split within the familial-PD experiment. *VPS13C*, *SYNJ1*, *PARK2*, *SLC6A3*, *APOE*, and *POLG* knockdowns suppressed translation (NES from -2.16 to -3.26), while *FBXO7*, *LRP1B*, *UCHL1*, *VPS35*, *SNCA*, and *B2M* knockdowns identified increased expression for this signature (NES from +1.86 to +2.38; all padj < 0.05). This split is reproducible across the broader set of translation-related GSEA terms and constitutes the most well-defined difference in directionality across familial-PD perturbations.

We next looked to identify if these consistently nominated signals implicating ribosomes and translation were associated with other signatures or stress response pathways like the ATF4-mediated Integrated Stress Response (ISR). The Reactome term **Response of EIF2AK4 (GCN2) to amino acid deficiency** is significantly enriched across the same 12 familial-PD perturbations described in this chapter as the term **Eukaryotic Translation Elongation**, seen with the same directionality split: *VPS13C*, *SYNJ1*, *PARK2*, *SLC6A3*, *APOE*, and *POLG* show downregulation (NES from -2.08 to -2.85), while *FBXO7*, *LRP1B*, *UCHL1*, *VPS35*, *B2M*, and *SNCA* show upregulation (NES from +1.81 to +2.38). Individual ATF4 transcriptional targets (*DDIT3*, *ASNS*, *TRIB3*, *PPP1R15A*, *SESN2*, *ATF3*) trend in directions concordant with this split: in the *LRP1B*, *B2M*, and *FBXO7* perturbed neurons, ATF4 targets show coordinate small-magnitude upregulation (mean log2FC +0.32 to +0.49 across measured targets); in the downregulating perturbations *VPS13C*, *APOE*, and *POLG*, ATF4 targets show coordinate small-magnitude downregulation (mean log2FC -0.31 to -0.49). None of these individual ATF4-target log2FC shifts are associated with statistically significant BH-corrected p-values in any familial-PD chapter perturbation, and the modest signal magnitudes preclude strong claims about whether ATF4-mediated integrated stress response is differentially active across the directionality split. The pattern is consistent with the broader direction-specific transcriptional responses across both libraries used in our Perturb-seq experiments. In the next section of this chapter, we discuss considerations regarding the statistical-power and effect sizes that nominate these signals.

3.7 Study limitations

A few points of consideration regarding the observations presented within this study:

First, technical and cost constraints hampered the scale of our experiments. To an extent, our results cover a breadth of 351 perturbations at cost of depth (perturbed cells assayed for each targeted genes). Consequently, per-perturbation cell counts in both experiments are modest. The pooled familial-PD neurons carry a median of approximately 146 cells per perturbation; -cluster representations are lower, and the GWAS-PD DA-lineage neuron cluster has a median of approximately 37 cells per perturbation across all 310 perturbations targeting the cluster (rising to approximately 66 among perturbations passing the Mixscale minimum-cell-count filter). Claims anchored to perturbations clearing the

chapter-default self-knockdown threshold (BH-adjusted $p < 0.05$) are the most directly interpretable; weaker-knockdown perturbations are reported only where the signature is supported by the withSelf/noSelf audit, mechanistic concordance with known biology, or cross-experiment replication.

The C12 subclone of the WIBR3 hESC line used these Perturb-seq experiments carries a 20q11.21 amplification of approximately 1.23 Mb spanning *BCL2L1* and a 12q24.21-q24.22 deletion of approximately 1.09 Mb encompassing *MED13L*, *HRK*, and *NOS1* (cell-line CNV characterization by Cell Line Genetics aCGH 100, sample C12.4 p5; report CLG-47397, 16 Aug 2022). The 20q11.21 gain is a well-documented adaptation of human pluripotent stem cells to prolonged culture, with *BCL2L1* established as the driver gene through its anti-apoptotic activity[86,87]. The 12q24 deletion is hemizygous loss of a region containing *HRK* (Harakiri, a pro-apoptotic BH3-only Bcl-2 family member); the directional pairing of *BCL2L1* gain and *HRK* loss both shift the apoptotic threshold away from cell death. Neither CNV is expected to drive the convergent axes reported here, but both plausibly shift baseline apoptotic sensitivity, and the conclusions of this chapter should be interpreted as specific to the WIBR3 background until replicated in an additional line.

All perturbations were induced in 2D monolayer cultures enriched for mDA-lineage cells but lacking glia and circuit-level connectivity. An appreciable component of PD-relevant biology plausibly depends on astrocyte, microglia, and oligodendrocyte contributions, thus the cell-autonomous responses measured here are a necessary but not sufficient test of each gene's PD-relevant function.

CRISPRi produces a partial mRNA-level knockdown. For PD genes whose pathogenic mechanism is a gain-of-function missense change (*SNCA* A53T, *LRRK2* G2019S), the knockdown model captures the loss-of-function axis only. For loss-of-function genes with severity-graded alleles, including *GBA1* (N370S vs L444P), CRISPRi modeling is direction-appropriate but cannot capture allelic-series quantitative differences in residual enzyme activity. Complementary iSCORE-PD mutant DA neuron experiments (Chapter 2, §2.4.5) address the gain-of-function axis and allelic-series biology for a subset of these genes.

Self-knockdown efficiency varies by approximately two-fold in median magnitude between the familial-PD and GWAS-PD libraries (familial-PD median $|\log2FC| = 0.84$ across 41 perturbations; GWAS-PD median = 0.36 across 310 perturbations) and across roughly an order-of-magnitude span within each library. For well-expressed targets, mRNA depletion is consistent across the four gRNAs targeting that gene, and the perturbation signature is directly interpretable. For less-expressed targets the self-measurement is weaker, but the signature may still be informative where enrichment is concordant across withSelf and noSelf analyses, where leading-edge genes are traceable to the target's known biology, or where the finding is supported by cross-experiment replication. *FBXO7* (§3.4.4) illustrates this: its self-measurement fails at every correction tier, yet the bidirectional signature with ribosome biogenesis upregulated alongside synapse-related gene sets downregulated is coherent with *FBXO7*'s proteasome-assembly biology[49].

Direction-specific transcriptional responses across familial-PD perturbations

Beyond the per-perturbation signatures developed in §3.4, the familial-PD pooled-neuron data exhibit a structural feature that bears on the interpretation of cross-pathway convergence. Across the 2,244 unique GSEA terms reaching significance in at least one familial-PD perturbation at padj < 0.05, the chapter-anchor perturbations partition into two coherent direction subsets: an upregulating subset comprising *FBXO7*, *LRP1B*, *B2M*, *VPS35*, and *UCHL1*, and a downregulating subset comprising *APOE*, *VPS13C*, *SYNJ1*, and *POLG*. The downregulating subset is particularly direction-pure at the per-perturbation level: *APOE* shows 2 significant upregulating terms across 1,404 total significant terms; *VPS13C* shows 4 of 780; *SYNJ1* shows 6 of 201; *POLG* shows 1 of 102. The upregulating subset is more variable: *LRP1B* (164 upregulating / 2 downregulating of 166 significant terms), *VPS35* (119 / 1 of 120), and *UCHL1* (131 / 4 of 135) are nearly direction-pure, while *FBXO7* alone shows substantial bidirectional signal (520 / 353 of 873 significant terms), consistent with its dual-direction signature developed in §3.4.4 (ribosome biogenesis upregulated alongside synapse downregulated).

The direction asymmetry is descriptive rather than mechanistic. Whether it reflects shared upstream regulatory axes that bind same-direction perturbations together, distinct cellular states accessed by different perturbations, or a correlation of perturbation magnitude with directional consistency is not resolved by the present data. The composition is also not biology-guided: *B2M* is treated elsewhere in this chapter as a positive-control perturbation rather than a familial-PD allele, and the asymmetry persists when *B2M* is excluded. The pattern is reported here primarily because it provides a within-experiment organizing structure for §3.6: cross-pathway convergences anchored on downregulating-subset perturbations (V-ATPase axis, autophagy axis, mitochondrial axis, translation axis) are observed precisely where the per-perturbation direction-purity is highest, motivating the methodological caveats developed below.

APOE perturbation illustrates a systematic failure mode for baseline-silent targets in CRISPRi screens. *APOE* is transcribed in fewer than 4% of hPSC-DA cells at day 45, with the Mixscale self-log2FC returning NA. CRISPRi cannot down-regulate a transcript that is not expressed. Despite this, the *APOE*-targeted population produces 1,404 significant noSelf GSEA terms at padj < 0.05 (2 upregulating, 1,402 downregulating), driven largely by a single-sgRNA outlier effect (*APOE_P1P2_4* retains 38 cells versus 126-153 for the other three sgRNAs), a reduction in median library complexity in *APOE*-targeted neurons (median nCount_RNA = 12,910 versus 14,047 for non-targeting controls in the neuron-cluster scope), and a heavy loading on the dataset's shared transcriptional axis. *APOE* PC1 = -65.7 places the perturbation an order of magnitude beyond any other perturbation in the familial-PD library (next-most-negative *VPS13C* PC1 = -6.2). For these reasons *APOE* is excluded from any §3.6 claim of signature convergence for lack of additional validation experiments.

The entire interpretive layer of this chapter rests on mRNA measurements, while several of the canonical stress, signaling, and translational-control programs in PD biology operate primarily at the protein and post-translational levels. The integrated stress response

is the clearest example: eIF2α phosphorylation, ATF4 translational de-repression, and downstream target-gene induction are post-translational events largely invisible to a transcriptome readout. When we examined the 41-perturbation familial-PD dataset for canonical ATF4 transcriptional output, no perturbation produced a BH-corrected significant change in *ATF4* mRNA or its canonical target genes (*DDIT3*, *ASNS*, *ATF3*, *TRIB3*, *PPP1R15A*, *SESN2*). At the sub-threshold log2FC level, however, ATF4-target gene expression shifts trend in directions concordant with the directionality split documented in §3.6.4 (mean log2FC across measured targets ranges from +0.32 to +0.49 in the upregulating chapter perturbations *LRP1B*, *B2M*, and *FBXO7*, and from -0.31 to -0.49 in the downregulating chapter perturbations *VPS13C*, *APOE*, and *POLG*). These sub-threshold magnitudes are not interpretable as definitive evidence of differential integrated stress response activity, because they fall near the noise floor of the per-perturbation expression estimates and cannot distinguish primary transcriptional regulation by ATF4 from secondary effects of broader translation-machinery shutdown. Where this chapter cites pathways under headings such as mTORC1 signaling or unfolded-protein response (§3.6), the claims are observation-level claims about transcriptional output, not mechanism-level claims about the protein pathways themselves.

To examine cross-perturbation transcriptional structure in the familial-PD pooled neuron analysis, we aggregated single-cell expression counts across all cells carrying each perturbation's gRNA into per-perturbation pseudobulk expression profiles, then computed principal-component decomposition on these profiles. The first principal component captures approximately 53% of cross-perturbation variance and is loaded most strongly on translation- and ribosomal-machinery gene expression, mirroring the broad signature documented at the per-perturbation NES level in §3.6.4. Coherent modular co-regulation of translation- and ribosomal-protein gene programs across perturbations has been documented in genome-scale Perturb-seq screens in K562 and RPE1 cells[7], and genome-wide CRISPRi/a screens in human iPSC-derived neurons recover similar functional clustering of perturbation responses[88]. The 41-perturbation familial-PD pooled neuron dataset is smaller-scale than these published Perturb-seq screens, and the apparent dominance of this single axis may be partially amplified by per-perturbation cell-count noise.

Two consequences for the §3.6 framing follow. First, perturbations with extreme position on this axis contribute disproportionately to any multi-pathway convergence claim: *APOE* sits at PC1 = -65.7, an order of magnitude beyond any other perturbation in the familial-PD library, and is excluded from §3.6 for this reason. Second, *VPS13C* is the next-most-extreme perturbation along the same axis (PC1 = -6.2, the largest residual loading in the pooled neuron analysis after *APOE*), and *VPS13C* anchors three of the four cross-experiment biological axes developed in §3.6 (V-ATPase / lysosomal acidification, §3.6.1; mitochondrial dysfunction, §3.6.3; translation, §3.6.4). Together, these observations argue that the four cross-experiment axes are mechanistically connected rather than arising independently, since a single perturbation reproduces three of them; they also nominate VPS13C as the strongest single anchor for the convergent biology described in this chapter.

Despite *VPS13C*'s extreme position on this axis, two within-experiment observations argue that the §3.6 cross-experiment biological axes reflect perturbation-specific biology rather than translation/ribosome-axis artifacts. *VPS13C* knockdown produces concordant downregulation of the V-ATPase, mitochondrial, and translation axes in both familial-PD and GWAS-PD experiments, reproducing signatures across two independently designed experiments. Additionally, each §3.6 axis is anchored by multiple perturbations within each experiment: *ATP6V0A1* (a direct V-ATPase subunit perturbation) and *SYNJ1* contribute to V-ATPase suppression alongside *VPS13C* in the familial-PD experiment; *ATP6V0A1*, *LRRK2*, *SLC6A3*, and *PITX3* contribute to mitochondrial suppression across both experiments. These patterns motivated the within-experiment-anchored convergence approach developed in §3.6.

3.8 Implications and future directions

Two complementary hPSC resources for isogenic control in PD biology

This dissertation contributes two complementary resources for studying the biology of PD in a human cellular system. First is the iSCORE-PD isogenic mutant cell-line collection developed in Chapter 2[89], providing a panel of pluripotent stem cell lines bearing PD-associated alleles in an isogenic background. Second is the CRISPRi Perturb-seq screen reported in this chapter, profiling loss-of-function transcriptional consequences across 351 PD-relevant genes (41 familial-PD plus 310 GWAS-PD risk-locus-nominated) at moderate per-gene depth in DA-lineage neurons. A third resource has very recently emerged from our group: a single-cell atlas of cells harboring thirteen familail-PD mutations in the iSCORE-PD collection differentiated into DA neurons[90] in the same WIBR3 hESC background. Together these three resources address human PD biology along orthogonal dimensions: many genes at moderate depth (CRISPRi screen), an allelic-series isogenic catalog (iSCORE-PD), and detailed single-cell characterization of fourteen specific mutations (the companion atlas).

Cross-platform convergence

Two of the four convergent transcriptional programs identified by the CRISPRi screen in §3.6 (V-ATPase/lysosomal acidification, autophagy/protein-aggregate clearance, mitochondrial dysfunction, and translation/ribosome biogenesis) also appear as convergent signatures across the thirteen mutations characterized in the iSCORE-PD single-cell DA-neuron atlas. Mitochondrial dysfunction notably emerges as a convergent signature in both datasets. An endolysosomal-degradation signature in the iSCORE-PD dataset combines what the CRISPRi screen nominates as two separate signatures described first in §3.6.1 (V-ATPase/lysosomal acidification) and later in §3.6.2 (autophagy/aggregate clearance). Importantly, these two experiments leverage mechanistically distinct perturbation mechanisms: the CRISPRi screen applies acute transcriptional knockdown, whereas the iSCORE-PD dataset represents the transcriptional state of cells which chronically carry endogenous disease alleles. The fact that the two converge on the same cellular programs supports the indication that these programs reflect genuine PD biology rather than perturbation-modality artifacts.

Divergent signatures between CRISPRi and the iSCORE-PD DA-neuron dataset

Two signatures diverge between the two datasets in informative ways. First, the translation and ribosome biogenesis program identified by the CRISPRi screen (§3.6.4) does not appear as a convergent signature in the iSCORE-PD dataset, potentially reflecting the difference between acute and chronic perturbation. Acute CRISPRi-mediated knockdown of a substantially expressed gene rapidly reduces its protein output, which may presumably trigger an integrated-stress-response-like translation shutdown as a generic acute-perturbation signature. By contrast, cells with engineered mutations which carry persitent disease alleles in the iSCORE-PD dataset have instead reached a steady state, in which an acute translation-shutdown response would have presumably resolved. The iSCORE-PD dataset identifies an iron-handling and ferroptosis signature as a third convergent program that the CRISPRi screen does not (per §3.6).

Nuances of allele-specific effects not captured in CRISPRi knockdown

The CRISPRi screen and the companion iSCORE-PD atlas address distinct aspects of PD biology. For example, gain-of-function disease alleles (*SNCA* A53T aggregation; *LRRK2* G2019S kinase hyperactivation) cannot be modeled by CRISPRi knockdown of the wild-type gene, which captures only the loss-of-function direction. Severity-graded loss-of-function alleles in genes such as *GBA1* cannot be reproduced by uniform CRISPRi knockdown, because the disease-relevant biology of *GBA1* N370S versus L444P differs in residual enzyme activity rather than in transcript abundance. Single-gene allelic-series comparisons (*VPS13C* A444P versus W395C in the iSCORE-PD dataset) similarly require allele-specific perturbations rather than gene-level knockdown.

Technical factors and experiment improvements

Three methodological refinements would substantially increase the biological yield of a future iteration of this screen. First, the gRNA library could be refined to omit guides whose self-knockdown signal is undetectable across the four-guide set; in the present screen, perturbations with undetectable self-log2FC contribute cellular sequencing capacity but no interpretable signal at the perturbation level. Second, migration from 10x Chromium to a split-pool combinatorial-barcoding scRNA-seq platform[104] (e.g., Parse Biosciences SPLiT-seq) would lift the per-pool perturbation ceiling that constrained the present screen to multi-lane batched processing of the 310-perturbation GWAS-PD library, supporting 1,000+ perturbation libraries reported in genome-scale Perturb-seq studies. In this approach, FACS-sorted perturbation-positive cells can frozen in batches and combinatorially barcoded across batches at the time of the scRNA-seq experiment[6]. Third, sensitization conditions applied prior to scRNA-seq harvest could convert baseline-silent transcriptional responses into phenotypes: mitochondrial stressors such as CCCP, rotenone, or MPTP would test whether perturbations in mitochondrial-quality-control genes (*PARK7*, *PINK1*, *PRKN*, *FBXO7*, *VPS13C*) produce transcriptional changes that converge on a stronger signature when cells face a mitochondrial-membrane-potential challenge.

Extending Perturb-seq to other PD-relevant cell-types

Extending the CRISPRi platform to non-neuronal cell types would enable the study of cell-autonomous gene-perturbation effects in cell types beyond DA-lineage neurons, complementing the present DA-neuron data with PD-relevant cell types whose biology cannot be captured by neuronal monoculture. Microglia carry significant common-variant heritability for PD risk[92], and several PD risk genes (notably GBA1) drive lysosomal and neuroinflammatory phenotypes in microglia[105] not testable in DA-neuron monocultures. Methodological infrastructure exists: human pluripotent stem cells can be differentiated into microglia[92], and a CRISPRi platform in iPSC-derived microglia has been developed and applied to neurodegeneration biology[93]. A glial-CRISPRi screen using the same C12 hESC background, with astrocyte, microglia, and oligodendrocyte derivatives in parallel, would provide a directly comparable Perturb-seq dataset for the glia-relevant subset of PD risk genes.

3.9 Materials and Methods

Parental hESC line and culture conditions

The CRISPRi cell line used in this study ("C12") was derived from the WIBR3 human embryonic stem cell (hESC) line (NIH Human Embryonic Stem Cell Registry approval number NIHhESC-10-0079; RRID:CVCL_9767), a female line that has been extensively characterized for genomic integrity and differentiation capacity[94]. WIBR3 was selected as the parental line to maintain isogenicity with the companion iSCORE-PD collection of familial PD mutations generated in the same genetic background[94,95]. High-density genotyping confirmed that our WIBR3 stock carries a heterozygous APOE ε4/ε3 genotype[94]. Array comparative genomic hybridization (aCGH; Agilent 60K Standard, hg38, reference Agilent Euro Female; report ID CLG-47397 dated 8/16/22, performed by Cell Line Genetics) identified two copy number variants in the later-passage WIBR3 stock used to derive the CRISPRi clone employed in this study: a ~1.23 Mb gain at 20q11.21 (chr20:31,300,674–32,533,660; 36 probes; $\log_2$ ratio +0.520) spanning *BCL2L1* and a ~1.09 Mb loss at 12q24.21–q24.22 (chr12:116,194,528–117,288,105; 23 probes; $\log_2$ ratio −0.646) spanning *MED13L* and *MAP1LC3B2* among others. Mycoplasma testing was performed monthly using an in-house PCR assay; all lines tested negative prior to experiments.

WIBR3 hESCs were initially maintained on irradiated or mitomycin-C-inactivated mouse embryonic fibroblast (MEF) monolayers in hESC medium consisting of DMEM/F12 (Thermo Fisher Scientific) supplemented with 15% fetal bovine serum (HyClone), 5% KnockOut Serum Replacement (Thermo Fisher Scientific), 1 mM L-glutamine (Invitrogen), 1% non-essential amino acids (Thermo Fisher Scientific), 0.1 mM β-mercaptoethanol (Sigma), 4 ng/mL basic fibroblast growth factor (Thermo Fisher Scientific or PeproTech), and 1× Penicillin-Streptomycin (Thermo Fisher Scientific)[94]. Cultures were maintained at 37 °C, 5% CO_2, and 5% O_2 (low-oxygen conditions) and passaged as aggregates every 5–7 days using collagenase IV (Thermo Fisher Scientific).

For all downstream experiments (lentiviral transduction, clone validation by flow cytometry, and dopaminergic neuron differentiation), hESCs were adapted to feeder-free culture conditions. Feeder-free WIBR3 hESCs were maintained on iMatrix-511 (Takara #T303), and fed with mTeSR Plus medium (STEMCELL Technologies #100-0276) at 37 °C, 5% CO_2. Cultures were passaged as single cells using Accutase (Innovative Cell Technologies #AT104). To promote cell survival following single-cell dissociation, ROCK inhibitor Y-27632 was added at 10 μM to the culture medium for 24 hours after each passage, and was also included at every subsequent step where cells were rendered as single-cell suspensions, including electroporation and FACS sorting.

Generation of Cre-inducible CRISPRi hESC lines

To enable conditional CRISPRi-based transcriptional perturbation in differentiated cell types, we engineered hESC lines carrying a Cre-inducible CRISPRi allele targeted to the AAVS1 safe harbor locus. WIBR3 hESCs were dissociated to single cells with Accutase and electroporated using the Bio-Rad Gene Pulser Xcell system in 4-mm cuvettes with an exponential decay program (250 V, 500 μF, infinite resistance) with two plasmids: (1) pX330 (Addgene #42230,[96]), encoding Cas9 and an sgRNA targeting the AAVS1 locus, and (2) a custom homology-directed repair (HDR) donor template containing a CAGGS promoter, a loxP-flanked transcriptional STOP cassette upstream of dCas9-KRAB-mCherry, and a neomycin resistance cassette coupled to a splice acceptor element at the AAVS1 locus, such that G418 resistance is conferred only by correct in-frame integration at the target site[97]. Following electroporation, cells were plated onto DR4 MEFs (feeders with NeoR), in hESC media supplemented with 10 μM Y-27632 and subjected to G418 selection (100 μg/mL) to enrich for correctly targeted clones. Approximately 40 clonal colonies were manually picked, expanded, and screened for correct integration. Several clones were validated for CRISPRi activity, but a single clone was used for this study.

Clone validation in hESCs

Genomic integration at the AAVS1 locus was verified by three PCR reactions: (i) a 5' HA insertion PCR (1,312 bp expected product) using primers **JD165** (5'-GAACTCTGCCCTCTAACGCT-3') and **JD166** (5'-CCCGCTTCAGTGACAACGTC-3'); (ii) a 3' HA insertion PCR (2,126 bp product) using primers **JD072** (5'-CACTACGACGCTGAGGTCAA-3') and **JD168** (5'-TACCCCGAAGAGTGAGTTTGC-3'); and (iii) a wild-type AAVS1 outside/outside PCR (1,945 bp expected product) using JD165 + JD168, which produces the wild-type band only in non-targeted alleles. Cre-inducibility was confirmed by transient delivery of Cre mRNA (Miltenyi #130-101-113) via StemFect reagent, and mCherry fluorescence was assessed by epifluorescence microscopy 48–72 hours after Cre delivery. Inducible expression of dCas9-KRAB protein was confirmed by Western blot using an anti-Cas9 antibody (Abcam, rabbit monoclonal clone EPR18991, #ab189390, 1:5000), comparing uninduced and Cre-induced C12 hESC lysates with Histone H3 (Cell Signaling Technology, mouse monoclonal clone 96C10, #3638, 1:2000) as loading control.

Functional CRISPRi activity in hESCs was assessed by flow cytometry on a BD LSR Fortessa (UC Berkeley MCB Flow Cytometry Facility). C12 hESCs were transduced with a lentiviral vector encoding an sgRNA targeting *B2M* (β2-microglobulin) — designated **B2M_P1_3** with protospacer sequence 5'-GGCCACGGAGCGAGACATCT-3' — a ubiquitously expressed cell-surface protein amenable to antibody-based detection. A non-targeting control sgRNA **ntCTRL_02784** (protospacer 5'-GGCGCCGACCGGAGCCCCTA-3') served as the negative control. Following lentiviral delivery of either gRNA, cells underwent puromycin selection to enrich for transduced cells. Resulting populations were transfected with Cre mRNA ((Miltenyi #130-101-113 via StemFect reagent), and harvested for flow cytometry after 10 days post-Cre delivery. Cre-induced and uninduced populations were treated with Human BD Fc Block (BD Pharmingen) per manufacturer recommendations followed by APC-conjugated mouse anti-human β_2-microglobulin (BioLegend clone 2M2, #316306, used at manufacturer-recommended 5 µL per 10^6 cells in 100 µL). Flow cytometry confirmed robust B2M protein depletion in Cre-induced CRISPRi cells relative to NTC and uninduced controls.

gRNA library design

Two pooled gRNA libraries were designed to target (1) familial PD (FPD) genes and related controls and (2) GWAS-nominated PD risk genes. Guide RNA candidates were selected from the Horlbeck et al. CRISPRi-v2 library resource[98] based on predicted on-target efficiency scores. For each target gene, guides were ranked according to the number of annotated transcription start sites (TSSs): genes with a single TSS received the top 4 ranked guides, while genes with two annotated TSSs received up to 8 guides (4 per TSS) to ensure coverage of alternative promoter usage.

The **FPD library** comprised 186 unique sgRNA species targeting 41 genes plus **6 non-targeting control (NTC) sgRNAs**. The core of the library consisted of approximately 20 genes with established associations to familial or monogenic PD, including *SNCA*, *PARK7*, *PINK1*, *PRKN* (PARK2), *LRRK2*, *VPS13C*, *VPS35*, *FBXO7*, *ATP13A2*, *DNAJC6*, *SYNJ1*, *EIF4G1*, *HTRA2*, *PLA2G6*, *POLG*, *GIGYF2*, and *TMEM230*. In addition, the library included transcription factors critical for midbrain DA neuron specification and identity (*FOXA2*, *LMX1A*, *NR4A2*, *PITX3*, *SOX2*, *PAX6*, *ASCL1*, *NEUROD1*, *NEUROD2*, *NEUROG1*, *NEUROG2*), DA neuron marker genes (*SLC6A3*/DAT), mTOR pathway components (*RHEB*, *RPTOR*), and several additional PD risk genes nominated by collaborators within the ASAP Collaborative Research Network (*SSR1*, *CTSB*, *GPNMB*, *LRP1B*, *LRP10*).
The **GWAS library** comprised 1,362 unique sgRNA species targeting approximately 307 genes nominated by the Nalls et al. 2019 PD GWAS meta-analysis[2], plus **46 Non-Targeting Control (NTC) sgRNAs**. Both libraries used NTC sequences with no predicted genomic target. The complete gene lists and guide sequences are provided within the author's Github repo for this manuscript https://github.com/jessedunnack/berkeley_dissertation.

Oligo synthesis and library cloning

Pooled oligonucleotide libraries encoding the sgRNA protospacer sequences with flanking adapter sequences were synthesized by Twist Bioscience. Upon receipt, the oligo pools were PCR-amplified using NEB Q5 Hot Start High-Fidelity DNA Polymerase (NEB #M0493) and cloned into the lentiviral sgRNA expression vector **pJD005**, yielding the cloned plasmid libraries designated G5O11 (FPD) and G5O13 (GWAS). pJD005 drives sgRNA expression from a mouse U6 promoter and co-expresses eGFP and a puromycin resistance cassette (eGFP-T2A-PuroR) from an EF1α promoter. **pJD005 was derived from the lentiviral gRNA vector pJR101**[99] from the Weissman laboratory by deletion of internal LoxP sites to permit downstream use with Cre-inducible systems. Restriction-enzyme cloning used BstXI (NEB #R0113) and BlpI (NEB #R0585). Library electroporation into Lucigen Endura electrocompetent *E. coli* was performed to maintain library complexity. Library representation and complexity were verified by next-generation sequencing of the plasmid pools.

Lentiviral packaging

Lentiviral libraries were produced by transient transfection of HEK293T cells (in-house lab stock, routinely propagated in DMEM + 10% FBS + Penicillin-Streptomycin) using a second-generation packaging system. HEK293T cells were transfected with the sgRNA library transfer vector (pJD005-G5O11 or pJD005-G5O13) together with the packaging plasmid psPAX2 (Addgene #12260) and the VSV-G envelope plasmid pMD2.G (Addgene #12259) using Polysciences PEI MAX (linear 25 kDa transfection-grade, Cat #24765) at 1 µg/µL working stock with a 3:1 PEI:DNA mass ratio in serum-free Opti-MEM medium (Thermo Fisher Scientific). Following transfection, the medium was replaced with serum-containing HEK293T growth medium. Viral supernatant was collected at 24 and 48 hours post-transfection, pooled, snap-frozen in liquid nitrogen, and stored at −80 °C until use.

For sgRNA library virus, no concentration step was performed; unconcentrated supernatant was sufficient to achieve the target multiplicity of infection (MOI) of 0.3. For Cre recombinase lentivirus, produced from the pLV-CMV-Puro-Cre vector (Addgene #17408) using the same HEK293T/PEI packaging protocol, viral supernatant was concentrated using Amicon Ultra-15 centrifugal filter units with a 100-kDa molecular weight cutoff (MilliporeSigma #UFC9100). Cre lentivirus titer on neurons was optimized in a separate pilot DA neuron differentiation using mCherry fluorescence as the readout for Cre-mediated excision of the loxP-STOP-loxP cassette.

Viral titer and transduction

Functional titers of the sgRNA library lentivirus were determined by limited dilution transduction assays. Serial volumes of unconcentrated viral supernatant were applied to target cells in the presence of 2 µg/mL polybrene (Sigma #H9268) to enhance transduction efficiency. Transduction efficiency was assessed 30–48 hours post-infection by flow cytometry, quantifying the percentage of eGFP-positive cells at each dilution to calculate functional titer and determine the volume required for a multiplicity of infection (MOI) of 0.3.

For Perturb-seq experiments, WIBR3 CRISPRi hESCs (clone C12) were transduced with sgRNA library lentivirus at MOI 0.3 in the presence of 2 µg/mL polybrene to favor single-sgRNA integration. Transduction was performed at the hESC stage, prior to differentiation. Following transduction, cells were subjected to puromycin selection at 0.6 µg/mL (added at 300 µL of a 1 mg/mL stock per 500 mL of medium) for 5–7 days to eliminate non-transduced cells. Following selection and one additional passage, cells were seeded into dopaminergic neuron differentiation (Section 3.3). ROCK inhibitor (Y-27632, 10 µM) was included whenever cells were dissociated to single cells during this process.

Dopaminergic neuron differentiation

Dopaminergic (DA) midbrain neurons were generated from C12 hESCs using a protocol adapted from Kim et al.[8] and applied in our companion iSCORE-PD atlas work[95]. Feeder-free hESCs already transduced with sgRNA library lentivirus and puromycin-selected (Section 3.2.4) were dissociated to single cells with Accutase (Innovative Cell Technologies #AT104) in the presence of Y-27632 (Chemdea #CD0141, 10 µM) and seeded at day 0 onto Matrigel (Corning #354230, growth-factor-reduced)–coated 6-well plates at 4–6 × 10^5 cells per well in mTeSR Plus (STEMCELL #100-0276) supplemented with 10 µM Y-27632.

Floor plate induction was carried out in differentiation medium consisting of Neurobasal (Gibco), N-2 supplement (Gibco), B-27 supplement (Gibco), 2 mM GlutaMAX (Gibco), Penicillin-Streptomycin (Gibco), and non-essential amino acids (NEAA, Gibco). From day 0 to day 3, cells were exposed to 500 ng/mL SHH C25II (R&D #1845-SH-100), 250 nM LDN193189 (Stemgent #04-0074-10), 10 µM SB431542 (Tocris #1614/10), 0.7 µM CHIR99021 (R&D #4423/10), and 10 µM Y-27632 (day 0 only; withdrawn day 1). On day 4, CHIR99021 was boosted to 7.5 µM (biphasic WNT activation per[8]). On day 7, LDN, SB, and SHH were withdrawn. On day 10, the medium was changed to neuron maturation medium [Neurobasal/B-27/GlutaMAX/Pen-Strep/NEAA] supplemented with 20 ng/mL BDNF (PeproTech #450-02-100), 20 ng/mL GDNF (PeproTech #450-10-100), 0.2 µM ascorbic acid (Sigma #A4034), 1 ng/mL TGFβ3 (R&D #8420-B3-025), 0.5 mM dibutyryl-cAMP (Selleck #S7858), and 3 µM CHIR99021. On day 11, cells were dissociated with Accutase and replated at high density on Cultrex Mouse Laminin I (R&D #3400-010-02) + poly-L-ornithine (Sigma #P3655)–coated dishes in maturation medium. DAPT (Tocris #2634, 10 µM) was added on day 12. On day 16, cells were dissociated and re-plated on PLO/laminin under the same conditions; on day 25, cells were dissociated and re-plated at low density in maturation medium. Detailed medium compositions, replating schedules, and densities are deposited at protocols.io (https://doi.org/10.17504/protocols.io.3byl4q8yovo5/v1), the same protocol used in the companion iSCORE-PD atlas work[95].

At approximately day 35 of differentiation, a time when cultures had adopted a post-mitotic neuronal morphology and expressed canonical floor-plate (FOXA2) and DA neuron (TH) markers, Cre recombinase lentivirus was delivered to activate CRISPRi. Cells were

maintained for an additional 10 days in maturation medium. At day 45, neurons were harvested for single-cell RNA sequencing. Cultures were dissociated to single cells using Accutase in the presence of Y-27632 and processed immediately for FACS and 10x Chromium library preparation (Sections 3.5–3.6).

Cre-mediated induction of CRISPRi

CRISPRi activity in C12 cultures was initiated by delivery of concentrated Cre recombinase lentivirus (pLV-CMV-Puro-Cre, Addgene #17408) at day 35 of differentiation. The optimal volume of concentrated Cre lentivirus was determined empirically by titration on a pilot DA neuron differentiation, using mCherry fluorescence (indicating successful Cre-mediated excision of the loxP-STOP-loxP cassette and activation of the dCas9-KRAB-mCherry transgene) as the readout. Upon Cre delivery, the loxP-flanked transcriptional STOP cassette is excised, permitting expression of dCas9-KRAB-mCherry from the integrated AAVS1 locus. In cells that had previously been transduced with sgRNA library lentivirus, dCas9-KRAB is then recruited to the sgRNA's target gene promoter to repress transcription. The 10-day window between Cre delivery (day 35) and harvest (day 45) allowed sufficient time for dCas9-KRAB expression and for downstream transcriptional consequences to manifest in the single-cell transcriptome.

Immunocytochemistry

To characterize the dopaminergic neuron identity of differentiated cultures, immunocytochemistry was performed on parallel cultures not used for Perturb-seq. Cells were fixed in 4% paraformaldehyde (PFA) in PBS for 10 minutes at room temperature, washed in PBS, permeabilized with 0.1% Triton X-100 in PBS for 10 minutes at room temperature, and blocked with 2% bovine serum albumin (BSA) in PBS for 1 hour at room temperature. Primary antibodies were applied overnight at 4 °C: mouse anti-FOXA2 (Abcam #ab60721, 1:250) and rabbit anti-TH (Pel-Freez Biologicals #P40101-150, 1:1000). Following washes, cells were incubated with Donkey anti-mouse Alexa Fluor 488 (Invitrogen #A-21202) and Donkey anti-rabbit Alexa Fluor 594 (Invitrogen #A-21207), both at 1:500, for 1 hour at room temperature. Nuclei were counterstained with Hoechst 33342 (Thermo Fisher). Cells were imaged directly in PBS in the original culture plate (no slide mounting) on a Nikon Ti-E inverted widefield epifluorescence microscope. FOXA2 expression confirmed floor-plate-derived midbrain identity, while TH positivity confirmed dopaminergic neuron differentiation.

FACS and sample preparation

Following Accutase dissociation at day 45, single-cell suspensions were sorted on a BD FACSAria Fusion cell sorter (UC Berkeley MCB Flow Cytometry Core Facility). The gating strategy proceeded as follows: forward scatter and side scatter were used to exclude debris; doublet exclusion (FSC-H vs. FSC-W) restricted analysis to singlets; the singlet population was gated for dual fluorescence (eGFP-positive for sgRNA cassette expression + mCherry-positive for Cre-mediated activation of the dCas9-KRAB-mCherry transgene). This $eGFP^+/mCherry^+$ double-positive population — approximately 5% of the total live cell

population across experiments — was collected for 10x Chromium library preparation. No viability dye was used; live cells were enriched by FSC/SSC gating in combination with dual fluorescent-marker gating, and post-sort viability was confirmed by trypan blue exclusion on a Countess II automated cell counter (Thermo Fisher Scientific) prior to 10x loading.

Sort collection tubes (polypropylene) were pre-coated with PBS + 0.5% BSA, and cells were collected into the same buffer to minimize cell loss to tube walls and maintain post-sort viability. Y-27632 was not included during the FACS sorting step. Following sorting, cells were pelleted by centrifugation at 300 × *g* for 5 minutes, resuspended in a defined volume, and assessed for viability and concentration on the Countess II.

For the C12_FPD-24 experiment, the post-sort eGFP$^+$/mCherry$^+$ population showed 87% viability at 827 cells/µL. Sorted cells were loaded at 16,000 cells per lane onto the 10x Chromium controller, targeting a recovery of approximately 10,000 cells per lane.

Single-cell RNA-Seq library preparation and sequencing

Single-cell gene expression and sgRNA identity libraries were generated using the 10x Genomics Chromium Next GEM Single Cell 3' Reagent Kit v3.1 with CRISPR Screening capture (10x Genomics Protocol CG000205 Rev D). Each 10x lane produced two parallel libraries: a gene expression (GEX) library and a CRISPR sgRNA capture library (feature barcode). The sgRNA capture library enables direct sequencing of the sgRNA protospacer present in each cell, linking perturbation identity to transcriptomic profile within the same single-cell partition[99].

For the C12_FPD-24 experiment, **6 lanes** were processed, all containing eGFP$^+$/mCherry$^+$ sorted, actively perturbed cells, yielding 12 sequencing libraries (6 GEX + 6 sgRNA capture). For the C12_GWAS-24 experiment, 12 lanes were processed (24 libraries). Libraries were sequenced on the Illumina NovaSeq X Plus platform (10B flow cell) at the UC Berkeley QB3 Genomics Core (Vincent J. Coates Genomics Sequencing Laboratory) with read configuration 28 (R1) × 91 (R2) × 10 (i7), per 10x specifications for the 3' v3.1 CRISPR Screening kit. Feature reference CSV files specifying sgRNA sequences and target gene assignments were designated G5O11 (FPD library) and G5O13 (GWAS library).

scRNA-Seq initialization

Raw sequencing reads were processed using CellRanger v8.0.1 (10x Genomics) to perform read alignment, barcode demultiplexing, and UMI counting. Gene expression reads were aligned to a custom human reference transcriptome derived from the 10x GRCh38-2020-A release with the addition of *eGFP* and *mCherry* transgene sequences (to enable read assignment for the sgRNA cassette and dCas9-KRAB-mCherry transcripts, respectively). sgRNA capture reads were aligned to custom feature reference files specifying the protospacer sequences and associated target gene identities for each library (deemed G5O11 for the FPD library; G5O13 for the GWAS library). Cell Ranger simultaneously

produces a gene expression count matrix and a sgRNA feature barcode count matrix for each lane, enabling downstream assignment of perturbation identity to individual cells.

Perturbation identity for each cell was assigned using the scMAGeCK R package[10] (selected over the default Cell Ranger sgRNA assignment for improved sensitivity to single-sgRNA cells). The Cell Ranger–derived sgRNA UMI count matrix was processed using a custom wrapper that converts the matrix to a triplet-format barcode table, which was then passed to the scMAGeCK pre_processRDS() function. A UMI threshold of >2 (i.e., ≥3 UMIs per sgRNA per cell) was applied to define positive sgRNA assignment. Cells with exactly one sgRNA above this threshold were classified as singlets; cells with two or more were multiplets; cells with none were unassigned.

The scMAGeCK output appends three metadata columns to each cell: scMAGeCK_gRNA_assignment (specific guide RNA identity), scMAGeCK_gene_assignment (target gene), and scMAGeCK_multiplet_id (singlet/multiplet/unassigned). Only singlet cells were retained for all downstream analyses to ensure unambiguous assignment of perturbation identity. Cells assigned to Non-Targeting Control (NTC) sgRNAs served as the reference population for differential expression testing.

Quality control filtering was applied independently to each sequencing lane to account for lane-to-lane variation in sequencing depth and cell quality. For each of the 6 FPD lanes, cells were removed if they fell below the 1.5th percentile or above the 98.5th percentile of either nCount_RNA (total UMI counts) or nFeature_RNA (number of detected genes); the union of both metrics was applied. An additional 194 extreme outlier cells were manually removed upon inspection of quality-metric distributions. This filtering reduced the total FPD cell count from 31,100 to 29,801 (1,299 cells removed: 1,105 by percentile thresholds + 194 manual).

Computational doublet detection was performed using DoubletFinder[11] with parameter sweep optimization (paramSweep_v3) and an expected doublet rate calculated as (nCells / 1000) × 0.8 / 100, per 10x Genomics recommendations. Identified doublets were flagged in the cell metadata and combined with scMAGeCK-based singlet classification (Section 3.9) to produce the final singlet-filtered dataset.

Cells were retained only if classified as singlet by scMAGeCK, not flagged as doublet by DoubletFinder, and passing per-lane QC thresholds. This intersection filter reduced the FPD dataset from 29,801 QC-filtered cells to **23,389 singlets** (78.5% retention). The GWAS dataset was similarly reduced from 103,281 to **62,245 singlets**. All downstream analyses — integration, clustering, MixScale differential expression, and GSEA — were performed exclusively on these singlet-filtered populations.

The 23,389 FPD singlet cells were integrated using the Seurat v5[9] and Harmony[100] frameworks. Data were normalized using NormalizeData (log-normalization with scale factor 10,000), and the top 3,000 variable features were identified with

FindVariableFeatures. Expression values were scaled using ScaleData with regression on nCount_RNA. Principal component analysis was performed using RunPCA with 50 components. Batch correction across the 6 sequencing lanes used Harmony integration (IntegrateLayers with method=HarmonyIntegration, group.by="sampleID"). A shared nearest-neighbor graph was constructed from the Harmony-corrected embedding using FindNeighbors (reduction="harmony", dims=1:50) and cells were clustered at resolution 0.15 using FindClusters, yielding 6 clusters (clusters 0–5). The GWAS dataset was integrated identically and canonically clustered at resolution 0.10, yielding 6 clusters.

At resolution 0.15 (FPD), clusters were annotated based on canonical midbrain DA neuron marker genes (*TH*, *DDC*, *SLC18A2*, *KCNJ6*, *LMX1A*) alongside early- and late-progenitor markers (*SOX2*, *PAX6*, *FOXA2*, *FOXA1*, *MEIS1*, *CRABP1*, *RGCC*). Clusters 1 and 5 were neuronal: cluster 5 (Neu-1/mDA) expressed the strongest midbrain DA markers; cluster 1 (Neu-2) was a larger, more heterogeneous neuronal population including GABAergic and glutamatergic subtypes in addition to DA-like cells. Remaining clusters were annotated as: cluster 0 = late progenitors (late-Prog; *MEIS1*, *CRABP1*, *RGCC* high; *FOXA1*/*FOXA2*/*SOX2* reduced but present); cluster 2 = neuroblasts (nBlast); cluster 3 = mid progenitors (mid-Prog); cluster 4 = early progenitors (early-Prog). Cluster-averaged expression profiles were correlated with published external references including FOUNDIN-PD iPSC-derived DA neurons[12] and postmortem human substantia nigra single-cell atlases[14,16] using Spearman correlation. After subsetting to clusters 1 and 5 and filtering to scMAGeCK-classified singlets, the final FPD neuron dataset for perturbation analysis comprised 6,700 cells (cluster 1: 5,724; cluster 5: 976) with 214 NTC cells.

Differential Expression analysis with Mixscale

Perturbation effects were quantified using the MixScale framework[5], an extension of Mixscape[101]. MixScale uses a Gaussian mixture model to distinguish genuinely perturbed cells from “escapees” and assigns each cell a continuous perturbation score (0–1).
After subsetting to the 6,700 FPD neuron cells, the data were re-preprocessed within the subset: layers were split by sample ID, normalized, the top 3,000 variable features were identified, data were scaled, PCA was performed with 50 components, and Harmony integration was applied. Cell-level perturbation signatures were computed using CalcPerturbSig with parameters: assay="RNA", slot="data", gd.class="scMAGeCK_gene_assignment", nt.cell.class="Non-Targeting", reduction="harmony", ndims=50, and num.neighbors=25 (K=25) for the combined-neuron analysis. For each perturbed cell, this function identifies the K=25 nearest NTC cells in Harmony embedding space and computes the perturbation signature as the difference between the perturbed cell’s expression profile and the mean expression of its local NTC neighbors. The resulting signatures are stored in a new assay (PRTB) within the Seurat object.

Because cluster 1 contributes approximately 85% of the cells in the K=25 combined-neuron analysis (5,724 of 6,700 cells), this analysis is dominated by cluster 1 biology. To

recover cluster-specific effects, separate per-cluster MixScale runs were performed at K=20 for each individual cluster (cl0_K20 through cl5_K20). Per-cluster analyses are particularly important for cluster 0 (late progenitors), where perturbations such as *EIF4G1* and *ATP13A2* produce strong signatures that are absent or attenuated in the combined-neuron analysis. The K=25 combined-neuron analysis retains its utility for self-knockdown detection and genome-wide DE gene counting where larger cell numbers provide greater statistical power.

Differential expression between each perturbation and the NTC population was performed using Run_wmvRegDE, which implements a weighted mean-variance regression test based on the glmGamPoi negative-binomial framework[102]. Parameters: assay="RNA", slot="counts", labels="scMAGeCK_gene_assignment", nt.class.name="Non-Targeting", logfc.threshold=0.1, full.results=TRUE. Each perturbed cell was weighted by its mixscale_score. The model includes log(nCount_RNA) as an estimated covariate (with size_factors=FALSE) to adjust for library size variation. DE testing was performed independently for each of the 38 testable perturbations at K=25, with NTC cells as the shared reference. A random seed of 12345 was used for reproducibility.

The MixScale DE output provides nominal p-values for each gene within each perturbation. To control the false discovery rate, Benjamini-Hochberg (BH) correction was applied independently to each perturbation's DE results using p.adjust(p_weight, method="BH"), generating a p_weight_BH column. Genes were considered differentially expressed if they met two criteria: p_weight_BH < 0.05 and $|\log_2 FC| > 0.25$. Only the BH-corrected p-values were used for downstream analyses.

Gene Set Enrichment Analysis (GSEA)

For each perturbation, all tested genes were ranked using a composite score that incorporates both the direction and statistical significance of differential expression: rank_score = sign(log2FC) × (−log10(p_weight_BH)). This ranking preserves the sign of the fold change while weighting by significance. The ranked list was sorted in decreasing order, with gene symbols as identifiers.

GSEA was performed in two modes per perturbation. In withSelf mode, the perturbed gene's own expression values are included in the ranked list. In noSelf mode, the perturbed gene is removed from the ranked list prior to enrichment analysis. The rationale for noSelf is that the perturbed gene's strong downregulation by CRISPRi can artificially inflate enrichment scores for gene sets containing that gene; noSelf provides an unbiased assessment of downstream pathway effects, while withSelf serves as a diagnostic for target-gene-driven inflation. The noSelf mode is the default source for manuscript and dissertation claims unless otherwise specified.

Enrichment analysis was performed against 8 gene set databases using two complementary tools: clusterProfiler[18] and fgsea[103]. The clusterProfiler package was used for Gene Ontology Biological Process (gseGO, ont="BP"), Gene Ontology Molecular

Function (gseGO, ont="MF"), Gene Ontology Cellular Component (gseGO, ont="CC"), KEGG pathways (gseKEGG), Reactome pathways (gsePathway from ReactomePA, requiring Entrez ID conversion via bitr), and WikiPathways (gseWP). The fgsea package was used for the MSigDB Hallmark (H) gene set collection and the MSigDB curated Reactome (C2:CP:REACTOME) collection, both accessed via the msigdbr package. Common parameters across all databases: minimum gene set size of 5, maximum gene set size of 500, Benjamini-Hochberg p-value adjustment (padj), p-value cutoff of 1.0 (all results retained for post-hoc filtering). The clusterProfiler eps parameter was set to 1e-10 (default), establishing a p-value floor for the most statistically significant terms. Significance was assessed at padj < 0.05 (the fgsea/clusterProfiler-native column).

GSEA across multiple databases and GO hierarchies frequently produces highly redundant enrichment results. To address this for figure visualization, enrichment terms were categorized into biological categories (e.g., Mitochondrial, Autophagy/Lysosome, Dopamine, Synapse) using keyword-based pattern matching. Note**:** all citation counts of significant pathway terms in the results sections refer to the full set of padj < 0.05 hits.

Western Blot

Western blot analysis was used to confirm dCas9-KRAB protein expression in Cre-induced versus uninduced C12 hESCs. Cells were lysed in home-made RIPA buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with Halt Protease and Phosphatase Inhibitor Cocktail (Thermo #78440). Total protein was quantified using the Pierce BCA Protein Assay Kit (Thermo #23227). 30–50 µg of total protein per lane was separated by SDS-PAGE on Bio-Rad 4–15% Mini-PROTEAN TGX gels (#4561084) and transferred to nitrocellulose membranes using the Bio-Rad Trans-Blot Turbo semi-dry transfer system (#1704155) with the Trans-Blot Turbo Mid Mol Wt transfer kit. Membranes were blocked in 5% (w/v) non-fat milk in TBST (Tris-buffered saline + 0.1% Tween-20) for 1 hour at room temperature. Primary antibodies were incubated overnight at 4 °C: rabbit anti-Cas9 (Abcam clone EPR18991, #ab189390, 1:5000) and mouse anti-Histone H3 (Cell Signaling clone 96C10, #3638, 1:2000) as a loading control. Following washes, membranes were incubated with HRP-conjugated secondary antibodies (Bio-Rad goat anti-rabbit IgG-HRP #1706515 and goat anti-mouse IgG-HRP #1706516, both 1:5000) for 1 hour at room temperature. Protein bands were visualized using Bio-Rad Clarity Western ECL substrate (#1705061) and imaged on a Thermo Fisher iBright FL1500 imaging system.

Software and data availability

All computational analyses were performed in R (v4.4.3) using the Seurat v5 single-cell analysis framework. Read alignment and UMI counting used Cell Ranger v8.0.1. Computationally intensive steps (Harmony integration, MixScale DE, GSEA enrichment) were executed on the UC Berkeley Savio high-performance computing cluster, while visualization and post-processing were performed locally. Key R packages used in the analysis pipeline: Seurat, Harmony, glmGamPoi, scMAGeCK, MixScale, DoubletFinder,

clusterProfiler, fgsea, ReactomePA, msigdbr, org.Hs.eg.db, GOSemSim, and ComplexHeatmap.

Raw sequencing data (FASTQ files) and processed Seurat objects (RDS files) for all Perturb-seq experiments will be deposited with the Aligning Science Across Parkinson's (ASAP) Collaborative Research Network (CRN) Cloud. Gene expression and sgRNA capture libraries are structured as paired FASTQ files per lane. Processed data include Cell Ranger output matrices, integrated Seurat objects with perturbation metadata, MixScale DE results, and complete GSEA enrichment results. GEO accession number will be assigned at submission.

All analysis scripts, including custom MixScale pipeline code, enrichment-analysis workflows, and visualization scripts, will be deposited on GitHub at https://github.com/jessedunnack/berkeley_dissertation

References

1. Blauwendraat C, Nalls MA, Singleton AB. The genetic architecture of Parkinson's disease. The Lancet Neurology. 2020 Feb;19(2):170–8. doi:10.1016/S1474-4422(19)30287-X PubMed PMID: 31521533; PubMed Central PMCID: PMC8972299.

2. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: A meta-analysis of genome-wide association studies. The Lancet Neurology. 2019 Dec;18(12):1091–102. doi:10.1016/S1474-4422(19)30320-5

3. Haapaniemi E, Botla S, Persson J, Schmierer B, Taipale J. CRISPR-Cas9 genome editing induces a P53-mediated DNA damage response. Nature Medicine. 2018 Jul;24(7):927–30. doi:10.1038/s41591-018-0049-z PubMed PMID: 29892067.

4. Ihry RJ, Worringer KA, Salick MR, Frias E, Ho D, Theriault K, et al. P53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. Nature Medicine. 2018 Jul;24(7):939–46. doi:10.1038/s41591-018-0050-6 PubMed PMID: 29892062.

5. Jiang L, Dalgarno C, Papalexi E, Mascio I, Wessels HH, Yun H, et al. Systematic reconstruction of molecular pathway signatures using scalable single-cell perturbation screens. Nature Cell Biology. 2025 Mar;27(3):505–17. doi:10.1038/s41556-025-01622-z

6. Replogle JM, Bonnar JL, Pogson AN, Liem CR, Maier NK, Ding Y, et al. Maximizing CRISPRi efficacy and accessibility with dual-sgRNA libraries and optimal effectors. eLife. 2022 Dec;11:e81856. doi:10.7554/eLife.81856

7. Replogle JM, Saunders RA, Pogson AN, Hussmann JA, Lenail A, Guna A, et al. Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. Cell. 2022 Jul;185(14):2559–2575.e28. doi:10.1016/j.cell.2022.05.013 PubMed PMID: 35688146; PubMed Central PMCID: PMC9380471.

8. Replogle JM, Norman TM, Xu A, Hussmann JA, Chen J, Cogan JZ, et al. Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. Nature Biotechnology. 2020 Aug;38(8):954–61. doi:10.1038/s41587-020-0470-y

9. Kim TW, Piao J, Koo SY, Kriks S, Chung SY, Betel D, et al. Biphasic Activation of WNT Signaling Facilitates the Derivation of Midbrain Dopamine Neurons from hESCs for Translational Use. Cell Stem Cell. 2021 Feb;28(2):343–355.e5. doi:10.1016/j.stem.2021.01.005 PubMed PMID: 33545081; PubMed Central PMCID: PMC8006469.

10. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nature Biotechnology. 2024 Feb;42(2):293–304. doi:10.1038/s41587-023-01767-y PubMed PMID: 37231261; PubMed Central PMCID: PMC10928517.

11. Yang L, Zhu Y, Yu H, Cheng X, Chen S, Chu Y, et al. scMAGeCK links genotypes with multiple phenotypes in single-cell CRISPR screens. Genome Biology. 2020 Dec;21(1):19. doi:10.1186/s13059-020-1928-4

12. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Systems. 2019 Apr;8(4):329–337.e4. doi:10.1016/j.cels.2019.03.003 PubMed PMID: 30954475; PubMed Central PMCID: PMC6853612.

13. Bressan E, Reed X, Bansal V, Hutchins E, Cobb MM, Webb MG, et al. The Foundational Data Initiative for Parkinson Disease: Enabling efficient translation from genetic maps to mechanism. Cell Genomics. 2023 Mar;3(3):100261. doi:10.1016/j.xgen.2023.100261 PubMed PMID: 36950378; PubMed Central PMCID: PMC10025424.

14. Agarwal D, Sandor C, Volpato V, Caffrey TM, Monzón-Sandoval J, Bowden R, et al. A single-cell atlas of the human substantia nigra reveals cell-specific pathways associated with neurological disorders. Nature Communications. 2020 Aug;11(1):4183. doi:10.1038/s41467-020-17876-0 PubMed PMID: 32826893; PubMed Central PMCID: PMC7442652.

15. Kamath T, Abdulraouf A, Burris SJ, Langlieb J, Gazestani V, Nadaf NM, et al. Single-cell genomic profiling of human dopamine neurons identifies a population that selectively degenerates in Parkinson's disease. Nature Neuroscience. 2022 May;25(5):588–95. doi:10.1038/s41593-022-01061-1

16. Birtele M, Storm P, Sharma Y, Kajtez J, Wahlestedt JN, Sozzi E, et al. Single-cell transcriptional and functional analysis of dopaminergic neurons in organoid-like cultures derived from human fetal midbrain. Development. 2022 Dec;149(23):dev200504. doi:10.1242/dev.200504 PubMed PMID: 36305490; PubMed Central PMCID: PMC10114107.

17. La Manno G, Gyllborg D, Codeluppi S, Nishimura K, Salto C, Zeisel A, et al. Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. Cell. 2016 Oct;167(2):566–580.e19. doi:10.1016/j.cell.2016.09.027 PubMed PMID: 27716510; PubMed Central PMCID: PMC5055122.

18. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. bioRxiv; 2016. doi:10.1101/060012

19. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Cambridge (Mass)). 2021 Aug;2(3):100141. doi:10.1016/j.xinn.2021.100141 PubMed PMID: 34557778; PubMed Central PMCID: PMC8454663.

20. Seaman MNJ. Cargo-selective endosomal sorting for retrieval to the Golgi requires retromer. The Journal of Cell Biology. 2004 Apr;165(1):111–22. doi:10.1083/jcb.200312034 PubMed PMID: 15078902; PubMed Central PMCID: PMC2172078.

21. Seaman MNJ. The retromer complex - endosomal protein recycling and beyond. Journal of Cell Science. 2012 Oct;125(Pt 20):4693–702. doi:10.1242/jcs.103440 PubMed PMID: 23148298; PubMed Central PMCID: PMC3517092.

22. Vilariño-Güell C, Wider C, Ross OA, Dachsel JC, Kachergus JM, Lincoln SJ, et al. VPS35 mutations in Parkinson disease. American Journal of Human Genetics. 2011 Jul;89(1):162–7. doi:10.1016/j.ajhg.2011.06.001 PubMed PMID: 21763482; PubMed Central PMCID: PMC3135796.

23. Zimprich A, Benet-Pagès A, Struhal W, Graf E, Eck SH, Offman MN, et al. A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. American Journal of Human Genetics. 2011 Jul;89(1):168–75. doi:10.1016/j.ajhg.2011.06.008 PubMed PMID: 21763483; PubMed Central PMCID: PMC3135812.

24. Seaman MNJ. Retromer and the cation-independent mannose 6-phosphate receptor-Time for a trial separation? Traffic. 2018 Feb;19(2):150–2. doi:10.1111/tra.12542 PubMed PMID: 29135085; PubMed Central PMCID: PMC5814863.

25. Munsie LN, Milnerwood AJ, Seibler P, Beccano-Kelly DA, Tatarnikov I, Khinda J, et al. Retromer-dependent neurotransmitter receptor trafficking to synapses is altered by the Parkinson's disease VPS35 mutation p.D620N. Human Molecular Genetics. 2015 Mar;24(6):1691–703. doi:10.1093/hmg/ddu582 PubMed PMID: 25416282.

26. Belenkaya TY, Wu Y, Tang X, Zhou B, Cheng L, Sharma YV, et al. The retromer complex influences Wnt secretion by recycling wntless from endosomes to the trans-Golgi network. Developmental Cell. 2008 Jan;14(1):120–31. doi:10.1016/j.devcel.2007.12.003 PubMed PMID: 18160348.

27. Tang FL, Liu W, Hu JX, Erion JR, Ye J, Mei L, et al. VPS35 Deficiency or Mutation Causes Dopaminergic Neuronal Loss by Impairing Mitochondrial Fusion and Function. Cell Reports. 2015 Sep;12(10):1631–43. doi:10.1016/j.celrep.2015.08.001 PubMed PMID: 26321632; PubMed Central PMCID: PMC4565770.

28. Wang W, Wang X, Fujioka H, Hoppel C, Whone AL, Caldwell MA, et al. Parkinson's disease-associated mutant VPS35 causes mitochondrial dysfunction by recycling DLP1 complexes. Nature Medicine. 2016 Jan;22(1):54–63. doi:10.1038/nm.3983 PubMed PMID: 26618722; PubMed Central PMCID: PMC4826611.

29. Zunino R, Schauss A, Rippstein P, Andrade-Navarro M, McBride HM. The SUMO protease SENP5 is required to maintain mitochondrial morphology and function. Journal of Cell Science. 2007 Apr;120(Pt 7):1178–88. doi:10.1242/jcs.03418 PubMed PMID: 17341580.

30. Yamada S, Sato A, Ishihara N, Akiyama H, Sakakibara SI. Drp1 SUMO/deSUMOylation by Senp5 isoforms influences ER tubulation and mitochondrial dynamics to regulate brain development. iScience. 2021 Dec;24(12):103484. doi:10.1016/j.isci.2021.103484 PubMed PMID: 34988397; PubMed Central PMCID: PMC8710555.

31. Won SY, Park MH, You ST, Choi SW, Kim HK, McLean C, et al. Nigral dopaminergic PAK4 prevents neurodegeneration in rat models of Parkinson's disease. Science Translational Medicine. 2016 Nov;8(367):367ra170. doi:10.1126/scitranslmed.aaf1629 PubMed PMID: 27903866.

32. Waku T, Nakamura N, Koji M, Watanabe H, Katoh H, Tatsumi C, et al. NRF3-POMP-20S Proteasome Assembly Axis Promotes Cancer Development via Ubiquitin-Independent Proteolysis of P53 and Retinoblastoma Protein. Molecular and Cellular Biology. 2020 Apr;40(10):e00597–19. doi:10.1128/MCB.00597-19 PubMed PMID: 32123008; PubMed Central PMCID: PMC7189095.

33. Waku T, Hagiwara T, Tamura N, Atsumi Y, Urano Y, Suzuki M, et al. NRF3 upregulates gene expression in SREBP2-dependent mevalonate pathway with cholesterol uptake and lipogenesis inhibition. iScience. 2021 Oct;24(10):103180. doi:10.1016/j.isci.2021.103180 PubMed PMID: 34667945; PubMed Central PMCID: PMC8506969.

34. Waku T, Kobayashi A. Pathophysiological Potentials of NRF3-Regulated Transcriptional Axes in Protein and Lipid Homeostasis. International Journal of Molecular Sciences. 2021 Nov;22(23):12686. doi:10.3390/ijms222312686 PubMed PMID: 34884489; PubMed Central PMCID: PMC8657584.

35. Bonifati V, Rizzu P, van Baren MJ, Schaap O, Breedveld GJ, Krieger E, et al. Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science. 2003 Jan;299(5604):256–9. doi:10.1126/science.1077209 PubMed PMID: 12446870.

36. Clements CM, McNally RS, Conti BJ, Mak TW, Ting JP-Y. DJ-1, a cancer- and Parkinson's disease-associated protein, stabilizes the antioxidant transcriptional

master regulator Nrf2. Proceedings of the National Academy of Sciences of the United States of America. 2006 Oct;103(41):15091–6. doi:10.1073/pnas.0607260103 PubMed PMID: 17015834; PubMed Central PMCID: PMC1586179.

37. Im JY, Lee KW, Woo JM, Junn E, Mouradian MM. DJ-1 induces thioredoxin 1 expression through the Nrf2 pathway. Human Molecular Genetics. 2012 Jul;21(13):3013–24. doi:10.1093/hmg/dds131 PubMed PMID: 22492997; PubMed Central PMCID: PMC3373246.

38. Guzman JN, Sanchez-Padilla J, Wokosin D, Kondapalli J, Ilijic E, Schumacker PT, et al. Oxidant stress evoked by pacemaking in dopaminergic neurons is attenuated by DJ-1. Nature. 2010 Dec;468(7324):696–700. doi:10.1038/nature09536 PubMed PMID: 21068725; PubMed Central PMCID: PMC4465557.

39. Goldberg MS, Pisani A, Haburcak M, Vortherms TA, Kitada T, Costa C, et al. Nigrostriatal dopaminergic deficits and hypokinesia caused by inactivation of the familial Parkinsonism-linked gene DJ-1. Neuron. 2005 Feb;45(4):489–96. doi:10.1016/j.neuron.2005.01.041 PubMed PMID: 15721235.

40. Fang L, Li G, Liu G, Lee SW, Aaronson SA. P53 induction of heparin-binding EGF-like growth factor counteracts P53 growth suppression through activation of MAPK and PI3K/Akt signaling cascades. The EMBO journal. 2001 Apr;20(8):1931–9. doi:10.1093/emboj/20.8.1931 PubMed PMID: 11296226; PubMed Central PMCID: PMC125417.

41. Farkas LM, Krieglstein K. Heparin-binding epidermal growth factor-like growth factor (HB-EGF) regulates survival of midbrain dopaminergic neurons. Journal of Neural Transmission. 2002 Mar;109(3):267–77. doi:10.1007/s007020200022 PubMed PMID: 11956950.

42. Hanke M, Farkas LM, Jakob M, Ries R, Pohl J, Sullivan AM. Heparin-binding epidermal growth factor-like growth factor: A component in chromaffin granules which promotes the survival of nigrostriatal dopaminergic neurones in vitro and in vivo. Neuroscience. 2004;124(4):757–66. doi:10.1016/j.neuroscience.2003.12.033 PubMed PMID: 15026116.

43. Yamaguchi S, Yamane T, Takahashi-Niki K, Kato I, Niki T, Goldberg MS, et al. Transcriptional activation of low-density lipoprotein receptor gene by DJ-1 and effect of DJ-1 on cholesterol homeostasis. PloS One. 2012;7(5):e38144. doi:10.1371/journal.pone.0038144 PubMed PMID: 22666465; PubMed Central PMCID: PMC3364227.

44. Fu X, Wang Y, He X, Li H, Liu H, Zhang X. A systematic review and meta-analysis of serum cholesterol and triglyceride levels in patients with Parkinson's disease. Lipids

in Health and Disease. 2020 May;19(1):97. doi:10.1186/s12944-020-01284-w PubMed PMID: 32430016; PubMed Central PMCID: PMC7236933.

45. Benn M, Nordestgaard BG, Frikke-Schmidt R, Tybjærg-Hansen A. Low LDL cholesterol, PCSK9 and HMGCR genetic variation, and risk of Alzheimer's disease and Parkinson's disease: Mendelian randomisation study. BMJ. 2017 Apr;357:j1648. doi:10.1136/bmj.j1648 PubMed PMID: 28438747; PubMed Central PMCID: PMC5421439.

46. Krebiehl G, Ruckerbauer S, Burbulla LF, Kieper N, Maurer B, Waak J, et al. Reduced basal autophagy and impaired mitochondrial dynamics due to loss of Parkinson's disease-associated protein DJ-1. PloS One. 2010 Feb;5(2):e9367. doi:10.1371/journal.pone.0009367 PubMed PMID: 20186336; PubMed Central PMCID: PMC2826413.

47. Shojaee S, Sina F, Banihosseini SS, Kazemi MH, Kalhor R, Shahidi GA, et al. Genome-wide linkage analysis of a Parkinsonian-pyramidal syndrome pedigree by 500 K SNP arrays. American Journal of Human Genetics. 2008 Jun;82(6):1375–84. doi:10.1016/j.ajhg.2008.05.005 PubMed PMID: 18513678; PubMed Central PMCID: PMC2427312.

48. Di Fonzo A, Dekker MCJ, Montagna P, Baruzzi A, Yonova EH, Correia Guedes L, et al. FBXO7 mutations cause autosomal recessive, early-onset parkinsonian-pyramidal syndrome. Neurology. 2009 Jan;72(3):240–5. doi:10.1212/01.wnl.0000338144.10967.2b PubMed PMID: 19038853.

49. Paisán-Ruiz C, Guevara R, Federoff M, Hanagasi H, Sina F, Elahi E, et al. Early-onset L-dopa-responsive parkinsonism with pyramidal signs due to ATP13A2, PLA2G6, FBXO7 and spatacsin mutations. Movement Disorders: Official Journal of the Movement Disorder Society. 2010 Sep;25(12):1791–800. doi:10.1002/mds.23221 PubMed PMID: 20669327; PubMed Central PMCID: PMC6005705.

50. Vingill S, Brockelt D, Lancelin C, Tatenhorst L, Dontcheva G, Preisinger C, et al. Loss of FBXO7 (PARK15) results in reduced proteasome activity and models a parkinsonism-like phenotype in mice. The EMBO journal. 2016 Sep;35(18):2008–25. doi:10.15252/embj.201593585 PubMed PMID: 27497298; PubMed Central PMCID: PMC5282834.

51. Kuiken HJ, Egan DA, Laman H, Bernards R, Beijersbergen RL, Dirac AM. Identification of F-box only protein 7 as a negative regulator of NF-kappaB signalling. Journal of Cellular and Molecular Medicine. 2012 Sep;16(9):2140–9. doi:10.1111/j.1582-4934.2012.01524.x PubMed PMID: 22212761; PubMed Central PMCID: PMC3822984.

52. Spagnol V, Oliveira CAB, Randle SJ, Passos PMS, Correia CRSTB, Simaroli NB, et al. The E3 ubiquitin ligase SCF(Fbxo7) mediates proteasomal degradation of UXT isoform 2 (UXT-V2) to inhibit the NF-κB signaling pathway. Biochimica Et Biophysica Acta General Subjects. 2021 Jan;1865(1):129754. doi:10.1016/j.bbagen.2020.129754 PubMed PMID: 33010352; PubMed Central PMCID: PMC8063000.

53. Stott SR, Randle SJ, Al Rawi S, Rowicka PA, Harris R, Mason B, et al. Loss of FBXO7 results in a Parkinson's-like dopaminergic degeneration via an RPL23-MDM2-TP53 pathway. The Journal of Pathology. 2019 Oct;249(2):241–54. doi:10.1002/path.5312 PubMed PMID: 31144295; PubMed Central PMCID: PMC6790581.

54. Burchell VS, Nelson DE, Sanchez-Martinez A, Delgado-Camprubi M, Ivatt RM, Pogson JH, et al. The Parkinson's disease-linked proteins Fbxo7 and Parkin interact to mediate mitophagy. Nature Neuroscience. 2013 Sep;16(9):1257–65. doi:10.1038/nn.3489 PubMed PMID: 23933751; PubMed Central PMCID: PMC3827746.

55. Kraus F, Goodall EA, Smith IR, Jiang Y, Paoli JC, Adolf F, et al. PARK15/FBXO7 is dispensable for PINK1/Parkin mitophagy in iNeurons and HeLa cell systems. EMBO reports. 2023 Aug;24(8):e56399. doi:10.15252/embr.202256399 PubMed PMID: 37334901; PubMed Central PMCID: PMC10398645.

56. Rothwarf DM, Zandi E, Natoli G, Karin M. IKK-gamma is an essential regulatory subunit of the IkappaB kinase complex. Nature. 1998 Sep;395(6699):297–300. doi:10.1038/26261 PubMed PMID: 9751060.

57. Fujishige K, Kotera J, Michibata H, Yuasa K, Takebayashi S, Okumura K, et al. Cloning and characterization of a novel human phosphodiesterase that hydrolyzes both cAMP and cGMP (PDE10A). The Journal of Biological Chemistry. 1999 Jun;274(26):18438–45. doi:10.1074/jbc.274.26.18438 PubMed PMID: 10373451.

58. Niccolini F, Foltynie T, Reis Marques T, Muhlert N, Tziortzi AC, Searle GE, et al. Loss of phosphodiesterase 10A expression is associated with progression and severity in Parkinson's disease. Brain: A Journal of Neurology. 2015 Oct;138(Pt 10):3003–15. doi:10.1093/brain/awv219 PubMed PMID: 26210536.

59. Wang J, Joseph S, Vingill S, Dere E, Tatenhorst L, Ronnenberg A, et al. Loss of the parkinsonism-associated protein FBXO7 in glutamatergic forebrain neurons in mice leads to abnormal motor behavior and synaptic defects. Journal of Neurochemistry. 2023 Oct;167(2):296–317. doi:10.1111/jnc.15962 PubMed PMID: 37753846.

60. Liu K, Jones S, Minis A, Rodriguez J, Molina H, Steller H. PI31 Is an Adaptor Protein for Proteasome Transport in Axons and Required for Synaptic Development. Developmental Cell. 2019 Aug;50(4):509–524.e10.

doi:10.1016/j.devcel.2019.06.009 PubMed PMID: 31327739; PubMed Central PMCID: PMC6702053.

61. Lesage S, Drouet V, Majounie E, Deramecourt V, Jacoupy M, Nicolas A, et al. Loss of VPS13C Function in Autosomal-Recessive Parkinsonism Causes Mitochondrial Dysfunction and Increases PINK1/Parkin-Dependent Mitophagy. American Journal of Human Genetics. 2016 Mar;98(3):500–13. doi:10.1016/j.ajhg.2016.01.014 PubMed PMID: 26942284; PubMed Central PMCID: PMC4800038.

62. Hancock-Cerutti W, Wu Z, Xu P, Yadavalli N, Leonzino M, Tharkeshwar AK, et al. ER-lysosome lipid transfer protein VPS13C/PARK23 prevents aberrant mtDNA-dependent STING signaling. The Journal of Cell Biology. 2022 Jul;221(7):e202106046. doi:10.1083/jcb.202106046 PubMed PMID: 35657605; PubMed Central PMCID: PMC9170524.

63. Wang X, Xu P, Bentley-DeSousa A, Hancock-Cerutti W, Cai S, Johnson BT, et al. The bridge-like lipid transport protein VPS13C/PARK23 mediates ER-lysosome contacts following lysosome damage. Nature Cell Biology. 2025 May;27(5):776–89. doi:10.1038/s41556-025-01653-6 PubMed PMID: 40211074; PubMed Central PMCID: PMC12081312.

64. Schormair B, Kemlink D, Mollenhauer B, Fiala O, Machetanz G, Roth J, et al. Diagnostic exome sequencing in early-onset Parkinson's disease confirms VPS13C as a rare cause of autosomal-recessive Parkinson's disease. Clinical Genetics. 2018 Mar;93(3):603–12. doi:10.1111/cge.13124 PubMed PMID: 28862745.

65. Morita M, Ler LW, Fabian MR, Siddiqui N, Mullin M, Henderson VC, et al. A novel 4EHP-GIGYF2 translational repressor complex is essential for mammalian development. Molecular and Cellular Biology. 2012 Sep;32(17):3585–93. doi:10.1128/MCB.00455-12 PubMed PMID: 22751931; PubMed Central PMCID: PMC3422012.

66. Peter D, Weber R, Sandmeir F, Wohlbold L, Helms S, Bawankar P, et al. GIGYF1/2 proteins use auxiliary sequences to selectively bind to 4EHP and repress target mRNA expression. Genes & Development. 2017 Jun;31(11):1147–61. doi:10.1101/gad.299420.117 PubMed PMID: 28698298; PubMed Central PMCID: PMC5538437.

67. Lautier C, Goldwurm S, Dürr A, Giovannone B, Tsiaras WG, Pezzoli G, et al. Mutations in the GIGYF2 (TNRC15) gene at the PARK11 locus in familial Parkinson disease. American Journal of Human Genetics. 2008 Apr;82(4):822–33. doi:10.1016/j.ajhg.2008.01.015 PubMed PMID: 18358451; PubMed Central PMCID: PMC2427211.

68. Lesage S, Condroyer C, Lohman E, Troiano A, Tison F, Viallet F, et al. Follow-up study of the GIGYF2 gene in French families with Parkinson's disease. Neurobiology of Aging. 2010 Jun;31(6):1069-1071; discussion 1072-1074. doi:10.1016/j.neurobiolaging.2009.06.008 PubMed PMID: 20004041.

69. Zimprich A, Schulte C, Reinthaler E, Haubenberger D, Balzar J, Lichtner P, et al. PARK11 gene (GIGYF2) variants Asn56Ser and Asn457Thr are not pathogenic for Parkinson's disease. Parkinsonism & Related Disorders. 2009 Aug;15(7):532–4. doi:10.1016/j.parkreldis.2009.01.005 PubMed PMID: 19250854.

70. Krebs CE, Karkheiran S, Powell JC, Cao M, Makarov V, Darvish H, et al. The Sac1 domain of SYNJ1 identified mutated in a family with early-onset progressive Parkinsonism with generalized seizures. Human Mutation. 2013 Sep;34(9):1200–7. doi:10.1002/humu.22372 PubMed PMID: 23804563; PubMed Central PMCID: PMC3790461.

71. Quadri M, Fang M, Picillo M, Olgiati S, Breedveld GJ, Graafland J, et al. Mutation in the SYNJ1 gene associated with autosomal recessive, early-onset Parkinsonism. Human Mutation. 2013 Sep;34(9):1208–15. doi:10.1002/humu.22373 PubMed PMID: 23804577.

72. Luoma P, Melberg A, Rinne JO, Kaukonen JA, Nupponen NN, Chalmers RM, et al. Parkinsonism, premature menopause, and mitochondrial DNA polymerase gamma mutations: Clinical and molecular genetic study. Lancet. 2004--Tenth century;364(9437):875–82. doi:10.1016/S0140-6736(04)16983-3 PubMed PMID: 15351195.

73. Davidzon G, Greene P, Mancuso M, Klos KJ, Ahlskog JE, Hirano M, et al. Early-onset familial parkinsonism due to POLG mutations. Annals of Neurology. 2006 May;59(5):859–62. doi:10.1002/ana.20831 PubMed PMID: 16634032.

74. Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, Minoshima S, et al. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature. 1998 Apr;392(6676):605–8. doi:10.1038/33416 PubMed PMID: 9560156.

75. Narendra D, Tanaka A, Suen DF, Youle RJ. Parkin is recruited selectively to impaired mitochondria and promotes their autophagy. The Journal of Cell Biology. 2008 Dec;183(5):795–803. doi:10.1083/jcb.200809125 PubMed PMID: 19029340; PubMed Central PMCID: PMC2592826.

76. Ko HS, von Coelln R, Sriram SR, Kim SW, Chung KKK, Pletnikova O, et al. Accumulation of the authentic parkin substrate aminoacyl-tRNA synthetase cofactor, P38/JTV-1, leads to catecholaminergic cell death. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience. 2005

Aug;25(35):7968–78. doi:10.1523/JNEUROSCI.2172-05.2005 PubMed PMID: 16135753; PubMed Central PMCID: PMC6725452.

77. Chartier-Harlin MC, Dachsel JC, Vilariño-Güell C, Lincoln SJ, Leprêtre F, Hulihan MM, et al. Translation initiator EIF4G1 mutations in familial Parkinson disease. American Journal of Human Genetics. 2011 Sep;89(3):398–406. doi:10.1016/j.ajhg.2011.08.009 PubMed PMID: 21907011; PubMed Central PMCID: PMC3169825.

78. Mendez JA, Bourque MJ, Fasano C, Kortleven C, Trudeau LE. Somatodendritic dopamine release requires synaptotagmin 4 and 7 and the participation of voltage-gated calcium channels. The Journal of Biological Chemistry. 2011 Jul;286(27):23928–37. doi:10.1074/jbc.M111.218032 PubMed PMID: 21576241; PubMed Central PMCID: PMC3129174.

79. Delignat-Lavaud B, Ducrot C, Kouwenhoven W, Feller N, Trudeau LÉ. Implication of synaptotagmins 4 and 7 in activity-dependent somatodendritic dopamine release in the ventral midbrain. Open Biology. 2022 Mar;12(3):210339. doi:10.1098/rsob.210339 PubMed PMID: 35232250; PubMed Central PMCID: PMC8889187.

80. Kabayama H, Tokushige N, Takeuchi M, Kabayama M, Fukuda M, Mikoshiba K. Parkin promotes proteasomal degradation of synaptotagmin IV by accelerating polyubiquitination. Molecular and Cellular Neurosciences. 2017 Apr;80:89–99. doi:10.1016/j.mcn.2017.02.006 PubMed PMID: 28254618.

81. Inglis KJ, Chereau D, Brigham EF, Chiou SS, Schöbel S, Frigon NL, et al. Polo-like kinase 2 (PLK2) phosphorylates alpha-synuclein at serine 129 in central nervous system. The Journal of Biological Chemistry. 2009 Jan;284(5):2598–602. doi:10.1074/jbc.C800206200 PubMed PMID: 19004816; PubMed Central PMCID: PMC2631975.

82. Ramalingam N, Haass C, Dettmer U. Physiological roles of α-synuclein serine-129 phosphorylation - not an oxymoron. Trends in Neurosciences. 2024 Jul;47(7):480–90. doi:10.1016/j.tins.2024.05.005 PubMed PMID: 38862330; PubMed Central PMCID: PMC11999472.

83. Schrder LF, Peng W, Gao G, Wong YC, Schwake M, Krainc D. VPS13C regulates phospho-Rab10-mediated lysosomal function in human dopaminergic neurons. The Journal of Cell Biology. 2024 May;223(5):e202304042. doi:10.1083/jcb.202304042 PubMed PMID: 38358348; PubMed Central PMCID: PMC10868123.

84. Lamark T, Johansen T. Aggrephagy: Selective disposal of protein aggregates by macroautophagy. International Journal of Cell Biology. 2012;2012:736905.

doi:10.1155/2012/736905 PubMed PMID: 22518139; PubMed Central PMCID: PMC3320095.

85. Tikhonova MA. A new avenue for treating Parkinson's disease targeted at aggrephagy modulation and neuroinflammation: Insights from in vitro and animal studies. EBioMedicine. 2020 Jan;51:102575. doi:10.1016/j.ebiom.2019.11.036 PubMed PMID: 31901571; PubMed Central PMCID: PMC6940682.

86. Schapira AH, Cooper JM, Dexter D, Clark JB, Jenner P, Marsden CD. Mitochondrial complex I deficiency in Parkinson's disease. Journal of Neurochemistry. 1990 Mar;54(3):823–7. doi:10.1111/j.1471-4159.1990.tb02325.x PubMed PMID: 2154550.

87. International Stem Cell Initiative, Amps K, Andrews PW, Anyfantis G, Armstrong L, Avery S, et al. Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage. Nature Biotechnology. 2011 Nov;29(12):1132–44. doi:10.1038/nbt.2051 PubMed PMID: 22119741; PubMed Central PMCID: PMC3454460.

88. Avery S, Hirst AJ, Baker D, Lim CY, Alagaratnam S, Skotheim RI, et al. BCL-XL mediates the strong selective advantage of a 20q11.21 amplification commonly found in human embryonic stem cell cultures. Stem Cell Reports. 2013;1(5):379–86. doi:10.1016/j.stemcr.2013.10.005 PubMed PMID: 24286026; PubMed Central PMCID: PMC3841249.

89. Tian R, Abarientos A, Hong J, Hashemi SH, Yan R, Dräger N, et al. Genome-wide CRISPRi/a screens in human neurons link lysosomal failure to ferroptosis. Nature Neuroscience. 2021 Jul;24(7):1020–34. doi:10.1038/s41593-021-00862-0

90. Busquets O, Li H, Syed KM, Jerez PA, Dunnack J, Bu RL, et al. iSCORE-PD: An isogenic stem cell collection to research Parkinson's Disease. bioRxiv. 2025 Mar;2024.02.12.579917. doi:10.1101/2024.02.12.579917 PubMed PMID: 38405931; PubMed Central PMCID: PMC10888955.

91. Syed KM, Dunnack J, Paatz S, Ajjarapu K, Sahagun A, Rio D, et al. An isogenic single-cell atlas of familial Parkinson's disease mutations reveals convergent changes in dopamine neurons. bioRxiv; 2026. doi:10.64898/2026.05.04.722458

92. Andersen MS, Bandres-Ciga S, Reynolds RH, Hardy J, Ryten M, Krohn L, et al. Heritability Enrichment Implicates Microglia in Parkinson's Disease Pathogenesis. Annals of Neurology. 2021 May;89(5):942–51. doi:10.1002/ana.26032 PubMed PMID: 33502028; PubMed Central PMCID: PMC9017316.

93. Douvaras P, Sun B, Wang M, Kruglikov I, Lallos G, Zimmer M, et al. Directed Differentiation of Human Pluripotent Stem Cells to Microglia. Stem Cell Reports.

2017 Jun;8(6):1516–24. doi:10.1016/j.stemcr.2017.04.023 PubMed PMID: 28528700; PubMed Central PMCID: PMC5470097.

94. Dräger NM, Sattler SM, Huang CTL, Teter OM, Leng K, Hashemi SH, et al. A CRISPRi/a platform in human iPSC-derived microglia uncovers regulators of disease states. Nature Neuroscience. 2022 Sep;25(9):1149–62. doi:10.1038/s41593-022-01131-4

95. Busquets O, Li H, Syed KM, Jerez PA, Dunnack J, Bu RL, et al. iSCORE-PD: An isogenic stem cell collection to research Parkinson's Disease. bioRxiv. 2025 Mar;2024.02.12.579917. doi:10.1101/2024.02.12.579917 PubMed PMID: 38405931; PubMed Central PMCID: PMC10888955.

96. Syed KM, Dunnack J, Paatz S, Ajjarapu K, Sahagun A, Rio D, et al. An isogenic single-cell atlas of familial Parkinson's disease mutations reveals convergent changes in dopamine neurons. bioRxiv; 2026. doi:10.64898/2026.05.04.722458

97. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013 Feb;339(6121):819–23. doi:10.1126/science.1231143 PubMed PMID: 23287718; PubMed Central PMCID: PMC3795411.

98. DeKelver RC, Choi VM, Moehle EA, Paschon DE, Hockemeyer D, Meijsing SH, et al. Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. Genome Research. 2010 Aug;20(8):1133–42. doi:10.1101/gr.106773.110 PubMed PMID: 20508142; PubMed Central PMCID: PMC2909576.

99. Horlbeck MA, Gilbert LA, Villalta JE, Adamson B, Pak RA, Chen Y, et al. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. eLife. 2016 Sep;5:e19760. doi:10.7554/eLife.19760 PubMed PMID: 27661255; PubMed Central PMCID: PMC5094855.

100. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nature Methods. 2019 Dec;16(12):1289–96. doi:10.1038/s41592-019-0619-0 PubMed PMID: 31740819; PubMed Central PMCID: PMC6884693.

101. Papalexi E, Mimitou EP, Butler AW, Foster S, Bracken B, Mauck WM, et al. Characterizing the molecular regulation of inhibitory immune checkpoints with multimodal single-cell screens. Nature Genetics. 2021 Mar;53(3):322–31. doi:10.1038/s41588-021-00778-2

102. Ahlmann-Eltze C, Huber W. glmGamPoi: Fitting Gamma-Poisson generalized linear models on single cell count data. Bioinformatics. 2021 Apr;36(24):5701–2.

doi:10.1093/bioinformatics/btaa1009 PubMed PMID: 33295604; PubMed Central PMCID: PMC8023675.

103. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. bioRxiv; 2016. doi:10.1101/060012

104. Rosenberg AB, Roco CM, Muscat RA, Kuchina A, Sample P, Yao Z, et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science. 2018 Apr 13;360(6385):176–182. doi:10.1126/science.aam8999 PubMed PMID: 29545511

105. Aflaki E, Stubblefield BK, Maniwang E, Lopez G, Moaven N, Goldin E, et al. Macrophage models of Gaucher disease for evaluating disease pathogenesis and candidate drugs. Science Translational Medicine. 2014 Oct 22;6(258):258ra144. doi:10.1126/scitranslmed.3009314 PubMed PMID: 25318935