Machine learning discriminates parkinsonian movement disorders in zebrafish

Gideon Llywelyn Hughes

PhD

University of York

Biology

September 2019

Abstract

Parkinson's disease is caused by a progressive loss of dopamine neurons in the substantia nigra. A loss of motor control occurs in patients, with symptoms including bradykinesia, resting tremor and muscle rigidity. There is currently no cure for Parkinson's disease and the commonly used treatment, L-DOPA, can cause side effects including dyskinesia. To identify potential treatments, drug screening using a suitable animal model is required before reaching clinical trials. Zebrafish are a vertebrate model organism well suited for high throughput drug screens, and genome editing can be used to create heritable mutations in causative genes to model human disease.

This thesis presents five genetic models of Parkinson's disease in the zebrafish created by CRISPR/Cas9 targeting (*pink1*, *parkin*, *dj-1*, *fbxo7*, and *gba*). Molecular analyses show a loss of dopamine neurons in the brain of the DJ-1 deficient zebrafish. This thesis also presents findings from a transcriptomic analysis of the *dj-1* mutant brain revealing dysregulated genes consistent with known parkinsonian defects.

An important focus of this work is the development of a novel computational method to analyse the movement phenotype of a zebrafish. The method developed uses high-speed recordings of zebrafish swimming, processed with a new fish tracking software. By measuring spatial coordinates and angles along the spine, swimming movement was converted into data suitable for computational analysis. The movement data was subjected to unbiased analysis employing a white-box supervised machine learning method (an *Evolutionary Algorithm*) which successfully discriminated the *dj-1* mutant from wild type.

This thesis concludes that the DJ-1 deficient zebrafish is a representative animal model of Parkinson's disease, and that machine learning can be used to classify the model based on movement data alone. It is proposed that the techniques developed here, have the potential for drug screens on the *dj-1* mutant using evolved classifiers to assess treatment effectiveness.

List of Contents

Abstract……2
List of Contents……..……..3
List of Tables 8
List of Figures 9
Acknowledgments 11
Declaration 12
1 Introduction 13
1.1 Parkinson's disease 14
1.1.1 Prevalence 14
1.1.2 Pathology 15
1.2 Cellular basis of PD 16
1.2.1 Impaired autophagy 16
1.2.2 Mitochondrial dysfunction and oxidative stress 17
1.2.3 Dopamine metabolism and oxidation 18
1.2.4 Selective loss of the nigrostriatal dopamine neurons 19
1.3 Neurotoxins associated with Parkinson's disease 20
1.4 The genetic basis of PD 21
1.4.1 Dominant forms of inherited PD 21
1.4.2 Recessive forms of inherited PD 23
1.4.3 Risk factors for PD 27
1.5 Animal models of Parkinson's disease 27
1.5.1 Drosophila models 28
1.5.2 Rodent models 31
1.5.3 Zebrafish models 36
1.5.4 Objective analysis of the movement phenotype 41
1.6 This thesis 42
2 Materials and Methods 43
2.1 Zebrafish 44

2.1.1 Husbandry 44
2.2 CRISPR/Cas9 targeting in zebrafish embryos 44
2.2.1 Guide RNA design 44
2.2.2 Generation of guide RNA 45
2.2.3 Microinjecting gRNA and Cas9 protein 46
2.3 Nucleic acid extraction 47
2.3.1 Genomic DNA extraction for sequencing 47
2.3.2 RNA extraction from zebrafish brains 47
2.4 Polymerase chain reaction (PCR) 48
2.4.1 cDNA synthesis 48
2.4.2 PCR/RT-PCR 48
2.4.3 DNA gel electrophoresis 50
2.4.4 Quantitative PCR 50
2.5 Sequencing mutant lines 52
2.5.1 PCR product clean-up 52
2.5.2 Ligation of PCR product into plasmid 53
2.5.3 Transformation of competent E.coli 53
2.5.4 Liquid cultures 54
2.5.5 Isolating plasmid DNA from bacteria 54
2.5.6 Sequencing in plasmid 54
2.5.7 Heteroduplex analysis 55
2.5.8 Genotyping by restriction enzyme digestion 57
2.6 Whole mount in situ hybridisation (ISH) 58
2.6.1 Embryo fixation 58
2.6.2 Synthesis of ISH probes 58
2.6.3 Whole mount ISH protocol 59
2.6.4 Imaging zebrafish embryos 60
2.7 RNA-Seq analysis 60
2.7.1 Preparation and sequencing of RNA samples 60
2.7.2 Computational analysis 61
2.7.3 Gene set enrichment analysis 61
2.8 Western Blot analysis 63

2.9 Immunofluorescence 64
2.9.1 Vibratome sectioning of zebrafish brains 64
2.9.2 Immunofluorescence for TH positive cells 65
3 Generating zebrafish models of PD 66
3.1 Introduction 67
3.1.1 Genome editing to model PD in the zebrafish 67
3.1.2 PD-associated genes targeted for knockout 69
3.1.3 Aims of this chapter 70
3.2 Results 72
3.2.1 Orthologs of the PD-associated genes in zebrafish 72
3.2.2 Synthesising in situ probes for the target genes 72
3.2.3 Expression profiles of PD-associated target genes 74
3.2.4 Synthesising guide RNA for CRISPR/Cas9 targeting 78
3.2.5 Validating CRISPR/Cas9 mutation of target genes in the F0 . 78
3.2.6 Identifying frameshift mutations in the F1 81
3.2.7 In situ hybridisation analysis of larval models 94
3.2.8 Analysing PD-models by Western blot 96
3.2.9 Immunofluorescent labelling of dopamine neurons 96
3.3 Discussion 101
3.3.1 Expression of the PD-associated genes targets 101
3.3.2 Effective targeting of the PD-associated genes in zebrafish. 101
3.3.3 Genetic mechanisms to recover target gene function 103
3.3.4 ISH analysis to measure DA neuron loss 105
3.3.5 DJ-1 deficient zebrafish effective at modelling PD 106
3.4 Conclusions 106
4 Transcriptomic analysis of the dj-1$^{-/-}$ mutant line 108
4.1 Introduction 109
4.1.1 DJ-1 in cancer and Parkinson's disease 109
4.1.2 Signalling pathways downstream of DJ-1 110
4.1.3 Transcriptional regulation by DJ-1 113
4.1.4 Altered gene expression in previous DJ-1 models 115
4.1.5 Aims of this chapter 116

4.2 Results ... 116
4.2.1 RNA-Seq analysis of differential gene expression ... 116
4.2.2 Validation of DJ-1 downstream targets ... 118
4.2.3 Identifying enriched gene sets from the RNA-Seq data ... 125
4.2.4 Validation of enriched gene sets ... 126
4.3 Discussion ... 131
4.3.1 Reduced expression of gpx3 ... 131
4.3.2 Genes implicated in Parkinson's disease ... 132
4.3.3 Enrichment of gene sets implicated in PD ... 136
4.4 Conclusion ... 138
5 Analysis of the movement phenotype ... 140
5.1 Introduction ... 141
5.1.1 Classifying by machine learning ... 141
5.1.2 Artificial neural networks: a black box classifier ... 141
5.1.3 Evolutionary algorithms: a white box classifier ... 143
5.1.4 Evolutionary algorithms for diagnosis of PD ... 144
5.1.5 Aims of this chapter ... 144
5.2 Methods ... 146
5.2.1 Recording tank design ... 146
5.2.2 Zebrafish recording protocol ... 146
5.2.3 Processing zebrafish recordings ... 147
5.2.4 Extracting the features of movement ... 149
5.2.5 Comparing features of movement ... 153
5.2.6 Statistical classification using features of movement ... 153
5.2.7 Evolving classifiers using the features of movement ... 154
5.2.8 Cartesian genetic programming ... 157
5.2.9 Fitness function: area under the curve ... 160
5.2.10 Organising data sets ... 161
5.2.11 Implementing an evolutionary algorithm ... 162
5.2.12 Evolving classifiers using the raw movement data ... 164
5.2.13 Principal component analysis ... 164
5.2.14 Evolving an ICRCGP sliding window classifier ... 164
5.3 Results ... 167

5.3.1 Extracted features of movement 167
5.3.2 Classifiers evolved using extracted features of movement ... 170
5.3.3 Sliding window classifiers evolved using the raw data 174
5.4 Discussion 176
5.4.1 Locomotor defects in the DJ-1 and PINK1 deficient lines 176
5.4.2 Evolved classifiers recognise dj-1$^{-/-}$ zebrafish movement 177
5.4.3 IRCCPG classifiers effectively recognise the PD models 178
5.4.4 Classifiers need a movement phenotype to discriminate 178
6 Discussion 180
6.1 Summary of findings 181
6.2 Translational models of disease in the zebrafish 182
6.3 Altered metabolism in the DJ-1 deficient zebrafish brain 185
6.4 Enrichment of cell cycle-related gene sets in the dj-1$^{-/-}$ brain 186
6.5 Future work 188
6.6 Perspectives 188
Appendix 190
References 194

List of Tables

Table 1.1 Genetic models of PD in the Drosophila 30
Table 1.2 Neurotoxin-based rodent models of Parkinson's disease 33
Table 1.3 Genetic mouse models of Parkinson's disease..... 34
Table 1.4 Genetic models of Parkinson's disease in the rat 35
Table 1.5 Neurotoxin-based models of PD in the zebrafish 39
Table 1.6 Gene knockdown/knockout models of PD in the zebrafish 40
Table 2.1 Target and primer sequences for generating gRNA template 45
Table 2.2 Primer sequences for PCR / RT-PCR 49
Table 2.3 Primer sequences for qPCR 51
Table 2.4 Restriction enzymes for genotyping mutants 57
Table 2.5 Restriction enzymes used for ISH probe templates 59
Table 2.6 Antibodies used for protein detection in Western blot 64
Table 3.1 F0 generation rate of mutation……81
Table 3.2 F1 generation mutation transmission……82
Table 4.1 Significantly downregulated transcripts in the dj-1-/- brain 123
Table 4.2 Significantly upregulated transcripts in the dj-1-/- brain..... 124
Table 4.3 Hallmark gene sets enriched in the dj-1-/- zebrafish brain 129
Table 5.1 Features of movement based on previous works..... 151
Table 5.2 Scores of classifiers evolved using extracted features..... 173
Table 5.3 Scores of the sliding window classifiers evolved 174
Appendix Table 1……..190
Appendix Table 2……..191
Appendix Table 3……..193

List of Figures

Figure 2.1 Heteroduplex analysis of an F1 generation..... 58
Figure 3.1 Diagram of CRISPR/Cas9 gene targeting..... 69
Figure 3.2 Generating in situ probe templates 73
Figure 3.3 Spatial expression of PD-associated genes in zebrafish 76
Figure 3.4 Spatial/temporal expression of the target genes……77
Figure 3.5 Location of CRISPR target sequences……80
Figure 3.6 Mutation of parkin and the effect on DA neuron markers……..84
Figure 3.7 Mutation of pink1 and the effect on DA neuron markers……..86
Figure 3.8 Mutation of dj-1 and the effect on DA neuron markers..... 88
Figure 3.9 Mutation of fbxo7 and effect on DA neuron markers..... 90
Figure 3.10 Mutation of gba and the effect on DA neuron markers..... 93
Figure 3.11 ISH analysis of dat expression in the PD models 95
Figure 3.12 TH expression in parkin-/- and dj-1-/- zebrafish brains 97
Figure 3.13 Counting the dopamine neurons in the posterior tuberculum….98
Figure 3.14 Labelling of TH+ cells in wild type and dj-$1^{-/-}$ brains…….99
Figure 3.15 Labelling of TH+ cells in $pink1^{-/-}$ and $parkin^{-/-}$ brains 100
Figure 4.1 Diagrams of transcriptional regulation by DJ-1 114
Figure 4.2 Integrity of RNA samples used in the RNA-Seq analysis..... 120
Figure 4.3 Differential gene expression in the *dj-1-/-* zebrafish brain..... 121
Figure 4.4 qRT-PCR validation of the RNA-Seq results..... 122
Figure 4.5 Enriched gene sets implicated in Parkinson's disease 130
Figure 4.6 Binding sites upstream of human GPX3 132
Figure 5.1 Artificial neural networks 142
Figure 5.2 Recording tank setup and extracted movement data..... 148
Figure 5.3 Calculating tailbeat frequency and tail bend amplitude..... 152
Figure 5.4 Evolution of a classifier 156
Figure 5.5 Example of a CGP network..... 158
Figure 5.6 A CGP network evolved using the features of movement..... 159
Figure 5.7 A Receiver Operating Characteristic (ROC) curve..... 161
Figure 5.8 Process of evolving a classifier..... 163
Figure 5.9 Evolving a sliding window classifier 166
Figure 5.10 pink1-/- features of movement ..... .168
Figure 5.11 dj-1-/- extracted features of movement……169

Figure 5.12 CGP network of the highest scoring dj-1-/- classifier……173
Figure 5.13 Patterns in the time series data used for discrimination……175
Appendix Figure 1 192

Acknowledgments

I'd like to thank my supervisor Betsy Pownall for taking on Harv's confused looking undergraduate student and somehow getting me to where I am today at the end of my PhD. I'd also like to thank my co-supervisor Stephen Smith for all he's taught me about machine learning and how positive he's been throughout my PhD, helping me to reach this point. I couldn't have asked for more supportive supervisors and I'm thankful for your love of beer which helped dream up this project. Thank you also to Harv Isaacs who I did my original undergraduate project with and has always been helpful with advice in lab group meetings.

I'd like to also thank the members of the Fish (and Frog) lab, past and present. To Lewis for all of his support and for helping me to keep calm under the pressure, you've kept me going through it all. To Caitlin for being like a big sister in the lab and a role model during my PhD, always up for a chat and willing to help. To Ali for his general merriment, appreciation of good coffee and discussions on transcriptomic data. To Laura and Alex, the newest additions to the lab, for their positive energy and senses of humour, keeping me laughing this last year. All of you have been a pleasure to work alongside.

Thanks to Sam for all of the advice and encouragement he's given throughout my PhD, Laura for being such a supportive friend that I can always count on and Ashley for being a great listener when I am on one of my many rants.

I'd like to thank my family for all of the support and encouragement over the past four years especially through the last few months with desks exploding and deadlines to meet. Thank you for providing for me, whether it be a place to stay, breakfast waffles or the countless cups of coffee. You've all been my rock throughout this.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Signed……  ……

1 Introduction

1.1 Parkinson's disease

Parkinson's disease (PD) is the second most common neurodegenerative disorder after Alzheimer's affecting approximately 1 in every 500 people in the UK (Parkinson's UK, 2017). PD is characterised by a movement disorder with four main symptoms: resting tremor, bradykinesia (slowing of movement), rigidity and loss of postural stability (Jankovic, 2008). Symptoms usually occur above the age of 50 with only around 4% of cases having an earlier onset (Van Den Eeden, 2003). The loss of motor control seen with PD is due to a progressive loss of dopamine (DA) neurons in a region of the brain called the substantia nigra (Greffard et al., 2006). The substantia nigra DA neurons project on to the striatum in the basal ganglia which is directly involved in action selection and initiation (Balleine, Delgado and Hikosaka, 2007). Currently there is no known cure for PD and the most commonly used treatment, L-DOPA, acts to replace the DA lost and reduce the symptoms of PD (Poewe et al., 2010). However, with continued use L-DOPA can cause side effects such as dyskinesia, a separate movement disorder characterised by involuntary movements (Jenner, 2008). The majority of PD cases are sporadic with around 10% being inherited (Thomas and Beal, 2007). Familial PD shares similarities with idiopathic PD including loss of the substantia nigra DA neurons and responsiveness to L-DOPA treatment (Cookson, 2012; Adams et al., 2005). However, the onset of symptoms is generally earlier with heritable PD occurring before the age of 50 (Puschmann, 2013).

1.1.1 Prevalence

The number of PD cases has risen dramatically over the last few decades with 2.5 million around the globe in 1990 rising to 6.1 million in 2016 (Dorsey et al., 2018). In the UK alone, the number of PD cases occurring in those over the age of 20 rose from an estimated 137,000 people to 145,519 between the years of 2017 and 2018 (Parkinson's UK, 2017, 2018). The increase in PD cases will partly be due to the growing population that is living longer (Partridge, Deelen and Slagboom, 2018). PD is an age-related disease and this is apparent in the age-specific prevalence rates. In 2017, an estimated 1.8 people per 100,000

aged 20-29 in the UK had PD and this figure rose with age to 1,696 in the 80-84 age bracket (Parkinson's UK, 2017). Similarly, a previous study in the Netherlands saw a rise in prevalence with age up to around 4% in the 85 - 94 year age bracket (De Rijk et al., 1995). However, increased life expectancy is not the only contributing factor to the increased prevalence of PD. The number of PD cases has risen for all age groups with age-standardised prevalence rates increasing by around 20% globally between 1990 and 2016 (Dorsey et al., 2018). PD has been reported as the only neurological disorder with increasing age-standardised prevalence (Feigin et al., 2017). Furthermore, environmental factors can also increase the chances of an individual developing sporadic PD with exposure to pesticides, herbicides, solvents and rural living all reported as significant risk factors (Pezzoli and Cereda, 2013).

1.1.2 Pathology

The pathology of PD is characterised by a progressive loss of dopamine (DA) neurons projecting from the substantia nigra in the midbrain to the striatum in the forebrain (Poewe et al., 2017). Symptoms are thought to occur around 5 years after the start of neurodegeneration with an estimated 50% of substantia nigra DA neurons lost (Fearnley and Lees, 1991). However, after adjusting for age-related loss the percentage of substantia nigra DA neurons lost at onset becomes 30% and a further 7% are estimated to degenerate year on year (Greffard et al., 2006). Whereas a reported 7% of DA neurons in the substantia nigra are lost per decade in healthy individuals (Ma et al., 1999). Post-mortem, a 55-80% loss of DA neurons in the substantia nigra has been observed in PD patients compared to controls (Ma et al., 1997; Pakkenberg et al., 1991; Damier et al., 1999). A second pathological feature of PD is the formation of intraneural protein aggregates known as Lewy bodies, consisting mainly of the protein alpha-synuclein (Spillantini et al., 1998). However, this is not essential for the loss of DA neurons and movement phenotype characterising PD as seen in a number of inherited cases (Johansen et al., 2018; Poulopoulos, Levy and Alcalay, 2012).

1.2 Cellular basis of PD

Several converging pathogenic mechanisms have been suggested to cause the death of substantia nigra DA neurons in PD including impaired autophagy, mitochondrial dysfunction and oxidative stress (Hauser and Hastings, 2013; Lynch-Day et al., 2012). Sporadic cases of PD are thought to be caused by a combination of genetic risk factors and environmental factors that contribute to the underlying causes of PD (Warner et al., 2003).

1.2.1 Impaired autophagy

Reduced autophagy has been suggested as a reason for the accumulation of alpha-synuclein and damaged mitochondria seen in PD (Lynch-Day et al., 2012). The alpha-synuclein protein is degraded by two different autophagy pathways, chaperone-mediated autophagy (CMA) and macroautophagy (Vogiatzi et al., 2008). In CMA proteins are delivered to the lysosome by chaperones such as hsc70, where the LAMP-2A receptor on the lysosome surface recognises a specific peptide sequence in the protein and internalises it for degradation (Majeski and Fred Dice, 2004). A loss of CMA markers, such as LAMP-2A and hsc70, has been observed in the substantia nigra of PD patient's post-mortem, implicating dysfunctional CMA in the pathology (Alvarez-Erviti et al., 2010). Studies linking disrupted autophagy to alpha-synuclein accumulation support this theory. Inhibition of CMA has been shown to increase alpha-synuclein levels in neuronal cultures (Xilouri, Vogiatzi and Stefanis, 2008). Similarly, an accumulation of alpha-synuclein was seen in the presynaptic terminals of DA neurons in a mouse model of PD with disrupted autophagy (Friedman et al., 2012). Increased levels of alpha-synuclein further exacerbate the disruption of autophagy, seen when overexpressing alpha-synuclein in human cells (Winslow et al., 2010). Mutations in alpha-synuclein are also known to cause autosomal dominant PD and the mutant protein binds to LAMP-2A inhibiting CMA (Xilouri et al., 2009; Alvarez-Erviti et al., 2010). In fact, a number of PD-associated genes are involved in autophagy. The most commonly mutated gene to cause heritable PD, *LRRK2*, has been associated with the regulation of autophagy and its mutation can inhibit CMA (Orenstein et al.,

2013; Manzoni, 2017). Three genes linked to autosomal recessive PD, *PINK1*, *PARKIN* and *FBXO7* are all involved in the same mitophagy pathway (Burchell et al., 2013), and the biggest genetic risk factor for developing PD is mutation of the *GBA* gene, which encodes a lysosomal enzyme (Gegg et al., 2012). Loss of *GBA* function inhibits autophagy and promotes alpha-synuclein accumulation (Xu et al., 2011).

1.2.2 Mitochondrial dysfunction and oxidative stress

Mitochondria play major roles in calcium homeostasis and the production of ATP for the cell through oxidative phosphorylation (Giorgi, Marchi and Pinton, 2018). In oxidative phosphorylation an electron travels along the electron transport chain (ETC) made up of complexes which pump protons across the inner mitochondrial membrane (Sazanov, 2015). The proton gradient created drives the phosphorylation of ADP by ATP synthase. However, complex I of the ETC is known to leak electrons which can reduce oxygen and produce superoxide anions, a damaging reactive oxygen species (Kushnareva, Murphy and Andreyev, 2002). Mitochondrial dysfunction, especially defects effecting complex I, can increase the production of ROS (Pitkänen and Robinson, 1996). Oxidative stress is highly associated with the neurodegeneration seen in PD (Blesa et al., 2015), and post-mortem analysis has previously revealed dysfunctional complex I activity in the substantia nigra of PD patients (Schapira et al., 1990). The majority of neurotoxins that cause a loss of DA neurons in the substantia nigra also act through inhibition of complex I in the ETC (Betarbet et al., 2000; Giordano et al., 2012). Furthermore, oxidative stress can damage the mitochondrial DNA (mtDNA) increasing mitochondrial dysfunction (Yakes and Van Houten, 1997). A higher number of mtDNA deletions are seen in the substantia nigra DA neurons of PD patients, around 50% compared to the 40% seen in aged controls (Bender et al., 2006). Due to the high energy demands of neurons carrying an action potential the length of their axons and the calcium involved in neuronal electrical activity mitochondrial dysfunction can further contribute to neurodegeneration by reducing ATP production and impairing calcium homeostasis (Bolam and Pissadaki, 2012; Schapira, 2013). A number of the PD-associated genes are also involved in the turnover of damaged mitochondria further implicating mitochondrial dysfunction as an underlying

molecular cause of PD (Narendra et al., 2010; Burchell et al., 2013). Additionally, the oxidative stress produced by damaged mitochondria can activate DJ-1, a PD-associated gene and major regulator of antioxidant genes (Clements et al., 2006). Overall, mitochondrial dysfunction can contribute to the neurodegeneration in PD by increasing oxidative stress, impairing calcium homeostasis and reducing ATP production.

1.2.3 Dopamine metabolism and oxidation

The metabolism and oxidation of DA can contribute to the loss of DA neurons seen in PD. DA is metabolised by the enzyme monoamine oxidase to produce the reactive oxygen species, hydrogen peroxide (Spina and Cohen, 1989). The oxidation of DA to DA-quinone, either by autoxidation or the activity of a peroxidase enzyme, can further damage the DA neurons by forming harmful adducts with proteins in the cell (Muñoz et al., 2012). These adducts are generally seen in the DA expressing areas of the brain suggesting the autooxidation of DA is a regular occurrence (Fornstedt et al., 1990). DA oxidation to form DA-quinones occurs in the cytosol, which has a pH of approximately 7, whereas DA is stabilised by the acidic environment inside synaptic vesicles (around pH 5) (Umek et al., 2018). This prevents DA-quinones forming within synaptic vesicles and being transported across the membrane into the synaptic junction and neighbouring neurons/astrocytes. DA-quinones are also unstable at cytosolic pH and need to form adducts with nearby proteins in the cytosol or they will rapidly undergo cyclization to form aminochrome (Bisaglia, Mammi and Bubacco, 2007). DA-quinones are likely trafficked with the proteins they form adducts with or DA-quinones can be transported into autophagic lysosomes where they are converted to neuromelanin permanently (Sulzer et al., 2000). DA-quinone can contribute to PD pathogenesis by forming adducts with alpha-synuclein and promoting the more harmful protofibril structure of alpha-synuclein over a fibrillar arrangement (Conway et al., 2001). Furthermore, alpha-synuclein modified by DA-quinone in the substantia nigra has been found to inhibit chaperone-mediated autophagy contributing to its accumulation (Martinez-Vicente et al., 2008). DA-quinone also forms an adduct with Parkin, a PD-associated protein. The Parkin adduct has decreased solubility and a loss of function, interestingly, a similar decrease in Parkin

solubility has been observed in the brains of PD patients (LaVoie et al., 2005). The combined effects of DA metabolism and oxidation can therefore contribute to the neurodegeneration seen in PD increasing oxidative stress and producing harmful protein adducts.

1.2.4 Selective loss of the nigrostriatal dopamine neurons

A selective loss of DA neurons in the substantia nigra is observed in PD patients (Surmeier, Obeso and Halliday, 2017). Neurons in general are more susceptible to a loss of ATP and calcium buffering than other cells due to their high energy demand and requirement of calcium for synaptic transmission (Haddad and Nakamura, 2015). DA neurons in the substantia nigra have the added vulnerability of extremely long axons and a much greater number of synaptic interactions (Bolam and Pissadaki, 2012). In the rat brain, a single DA neuron of the substantia nigra was estimated to interact with 75,000 neurons in the striatum (Matsuda et al., 2009). The nigrostriatal DA neurons are also unmyelinated so require a greater amount of energy to maintain a membrane potential and for the propagation of action potentials (Pissadaki and Bolam, 2013). Myelination involves glial cells wrapping around neuronal axons, with frequent gaps in myelination called the nodes of Ranvier (Seidl, 2014). Saltatory conduction, with sodium ion influx, occurs at the nodes of Ranvier rather than along the entire axonal membrane, allowing an action potential to rapidly travel the length of an axon. Maintaining calcium homeostasis further adds to the vulnerability of nigrostriatal DA neurons. A continuous influx of calcium ions through L-type calcium channels maintains an autonomous pacemaking activity in the DA neurons of the substantia nigra (Puopolo, Raviola and Bean, 2007). To maintain calcium homeostasis the Ca^{2+} ions have to be constantly removed from the neuron requiring ATP, and the associated increase in mitochondrial activity produces more oxidative stress (Surmeier and Schumacker, 2013). ATP synthase uses the proton gradient created by the electron transport chain to generate ATP and an increase in mitochondrial membrane potential ($\Delta\Psi$) increases the rate of ATP synthesis by ATP synthase (Kaim and Dimroth, 1999). Mitochondria extracted from the hearts and brains of adult rats have been shown to increase the production of hydrogen peroxide, a ROS, with an increase in mitochondrial membrane potential ($\Delta\Psi m$) (Starkov and Fiskum,

2003; Korshunov, Skulachev and Starkov, 1997). Antagonising the L-type calcium channels has been shown to reduce the oxidative stress produced by mitochondria (Guzman et al., 2010). Overall, DA neurons of the substantia nigra are exceptionally energy expensive cells and are therefore susceptible to disruption of metabolism.

1.3 Neurotoxins associated with Parkinson's disease

The first neurotoxin associated with PD development was MPTP, discovered in 1983, when drug users injected heroin that happened to be contaminated with the toxin (Langston et al., 1983). Since the discovery of MPTP several other neurotoxins that cause a PD pathology have been discovered including the pesticide, rotenone, the herbicide, paraquat, and 6-hydroxydopamine (6-OHDA), a hydroxylated analogue of DA. The MPTP metabolite, MPP+, rotenone and 6-OHDA can all act through inhibition of complex I in the electron transport chain to increase oxidative stress, and 6-OHDA can also be oxidised to produce hydrogen peroxide (Betarbet et al., 2000; Nicklas et al., 1987; Simola, Morelli and Carta, 2007). Inhibition of complex I activity has also been shown to cause the redistribution of phosphorylated tau protein from axons to the cell body, reduce the oxidation of NADH to NAD+, reduce ATP production in the cell and decrease cell survival (Höllerhage et al., 2009). Tau and phosphorylated tau have previously been shown to enhance alpha-synuclein aggregation in mouse primary cortical neurons in areas including the cell body (Badiola et al., 2011). Therefore, the redistribution of p-Tau could increase alpha-synuclein aggregation in the cell body. The reduced oxidation of NADH to NAD+ is expected as complex I is known to transfer an electron from NADH to ubiquinone producing NAD+, and this is required to create the electrochemical gradient for ATP synthesis (Brandt et al., 2003). The loss of cell survival is likely due to a combination of increased oxidative stress and reduced ATP production. Paraquat causes neurodegeneration by activating the Jun N-terminal kinase (JNK) cell death pathway and caspase-3, an enzyme in apoptosis (Ramachandiran et al., 2007). Overall, the majority of neurotoxins associated with PD development act through mitochondrial dysfunction and increased oxidative stress.

1.4 The genetic basis of PD

Only about 10% of PD cases are inherited, however identifying and understanding the genetic basis of heritable PD has led to a much better understanding of the molecular basis of the disease (Wood-Kaczmar, Gandhi and Wood, 2006). PD was originally found to have a genetic component in 1996 after linkage analysis was performed on genome data from members of an Italian family with inherited PD, revealing a shared genetic marker (Polymeropoulos et al, 1996). This first genetic marker was shortly discovered to be a mutation in the *SNCA* gene (Polymeropoulos et al., 1997). Since then a number of other genes have been associated with PD and as of 2018, 19 PD causing genes had been confirmed (Deng, Wang and Jankovic, 2018). Mutation of a single one of these genes causes a monogenic case of PD, which account for approximately 5-10% of all PD cases (Lill CM, 2016). About 30% of inherited cases are monogenic and between 3 - 5% of sporadic cases (Klein and Westenberger, 2012). In addition to the genes that cause monogenic PD another gene of interest is *GBA,* mutations in *GBA* being the most common genetic risk factor for developing PD (Sidransky and Lopez, 2012). Animal models of PD, created by transgenic expression or knockdown/knockout of PD-associated genes have helped improve understanding of the pathogenic mechanisms in PD (Table 1.1, Table 1.3 and Table 1.4). A number of these genes and their role in PD pathogenesis are described below.

1.4.1 Dominant forms of inherited PD

SNCA

In 1997, a mutation in *SCNA* (A53T) was originally identified as a cause of autosomal dominant PD in four families from Greece and Italy (Polymeropoulos et al., 1997). A year later a second mutation in the *SNCA* gene (A30P) was also discovered to cause autosomal dominant PD (Krüger et al., 1998). Around the same time alpha-synuclein, encoded by the *SNCA* gene, was identified as a major component of the intraneural Lewy bodies associated with PD pathology (Spillantini et al., 1997; Mezey et al., 1998). Multiple functions have been suggested for alpha-synuclein including a chaperone activity and roles in

membrane curvature, monoamine homeostasis and neurotransmission. Manda et al. (2014) reported that alpha-synuclein has a chaperone activity preventing the heat-induced aggregation of certain proteins and that this activity is dependent on its C-terminal. Monomeric alpha-synuclein has been shown to induce membrane curvature by inserting into a membrane and this is thought to occur during the endocytosis and exocytosis of synaptic vesicles (Westphal and Chandra, 2013). Alpha-synuclein can also affect dopamine homeostasis by forming a complex with the dopamine transporter and promoting its clustering at the synapse increasing neuronal dopamine uptake (Lee et al., 2001). Burré et al. (2010) found that alpha-synuclein can interact with phospholipids in synaptic vesicles and the protein synaptobrevin-2 to promote SNARE-complex assembly at the presynaptic nerve terminal thereby assisting neurotransmission. The A53T and A30P mutant forms of alpha-synuclein were discovered to inhibit chaperone-mediated autophagy preventing their degradation and contributing to alpha-synuclein accumulation in the neurons of PD patients (Cuervo et al., 2004). Furthermore, the A53T and A30P mutations in alpha-synuclein promote its aggregation forming oligomers at a faster rate compared to the wild type protein and the A53T mutant rapidly forms fibrils of alpha-synuclein (Conway et al., 2000). In addition to the mutations in *SNCA,* duplications and triplications of the gene have been identified as causes of autosomal dominant PD in a number of cases (Singleton et al., 2003; Chartier-Harlin et al., 2004). They observed a distinction in the disease phenotype, with *SNCA* triplications causing an earlier onset and faster disease progression, suggesting the severity was dose-dependent. Overall, the mutations in *SNCA* and duplication/triplications of the gene cause an increase in alpha-synuclein, which aggregates to form Lewy bodies (Mezey et al., 1998).

LRRK2

In 2004, mutations in *LRRK2* were identified as a cause of late onset autosomal dominant PD (Zimprich et al., 2004). Mutations in *LRRK2* are now known to be the most common cause of inherited PD and the clinical phenotype closely matches idiopathic cases (Marras et al., 2011; Healy et al., 2008). *LRRK2* encodes a large multidomain protein including a kinase domain, a GTPase domain (Roc) and various protein-protein interaction domains allowing LRRK2 to act as a scaffold for establishing protein complexes (Miklossy et al., 2006;

Esteves and Cardoso, 2016). The G2019S mutation and other common PD-causing mutations in *LRRK2*, have been shown to increase its kinase activity causing a gain in function (West et al., 2005; Gloeckner et al., 2006). LRRK2 is thought to have roles in autophagy, mitochondrial function, vesicular trafficking and cytoskeletal dynamics (Biskup et al., 2006; Kett et al., 2012; Manzoni et al., 2013; Ludtmann et al., 2019). The LRRK2 protein localises to lysosomes and endosomes correlating with its role in autophagy and the G2019S mutant has an inhibitory effect on chaperone-mediated autophagy preventing the degradation of lysosomal substrates (Biskup et al., 2006; Orenstein et al., 2013). LRRK2 also plays a role in mitochondrial function with overexpression in human cells causing fragmentation of the mitochondria, a heightened effect was also seen with the G2019S mutation (Wang et al., 2012). LRRK2 is further associated with cytoskeletal dynamics phosphorylating microtubules and actin-binding proteins to promote microtubule polymerisation and cytoskeletal remodelling, respectively (Gillardon, 2009; Parisiadou et al., 2009). The kinase activity of LRRK2 plays a role in vesicle trafficking, with overexpression of the LRRK2(G2019S) mutant altering vesicle distribution in neurons and increasing DA release (Migheli et al., 2013). In summation, mutations in *LRRK2* could contribute to PD development through a number of dysfunctional pathways including autophagy, mitochondrial function, vesicular trafficking and cytoskeletal dynamics.

1.4.2 Recessive forms of inherited PD

PARKIN

PARKIN was first identified in 1998 as the gene responsible for autosomal recessive juvenile PD (Kitada et al., 1998). Juvenile PD being cases with onset before the age of 20 (Quinn, Critchley and Marsden, 1987). Mutations in *PARKIN* are the most common cause of autosomal recessive PD and the second most common for monogenic PD cases (Dawson et al., 2010). They are observed in around 49% of families with early onset PD affecting a family member (Lücking et al., 2000). PD cases caused by mutation of *PARKIN* generally lack a Lewy body pathology (Johansen et al., 2018; Morrison, 2003). *PARKIN* encodes a ubiquitously expressed protein with heightened levels observed in the brain and substantia nigra (Kitada et al., 1998). The Parkin

protein is an E3 ubiquitin ligase consisting of an N-terminal ubiquitin like domain (Ubl), and RING0, RING1, in between ring (IBR) and RING2 domains at the C-terminal (Riley et al., 2013). The RING1, IBR and RING2 domains are required for the ubiquitin ligase activity of Parkin (Narendra et al., 2010). Parkin is recruited to damaged mitochondria where it ubiquitinates substrates on the mitochondrial surface to signal for mitochondrial turnover (Narendra et al., 2008; Chan et al., 2011). However, the Parkin protein is generally found in an inactive state, caused by RING0 blocking the RING1 and RING2 domains. A conformational change and phosphorylation of the protein are required to activate its ubiquitin ligase activity (Koyano and Matsuda, 2015). The majority of PD-associated mutations in *PARKIN* either reduce translocation to the mitochondria or its ubiquitin ligase activity (Shimura et al., 2000; Matsuda et al., 2010). Overall, a loss of Parkin increases mitochondrial dysfunction and this was observed in the Parkin deficient mouse (Palacino et al., 2004).

PINK1

PINK1 was originally identified as a downstream target of the tumour suppressor PTEN (Unoki and Nakamura, 2001). In 2004, mutations in the kinase domain of PINK1 were identified as a cause of autosomal recessive PD in three families (Valente et al., 2004). Mutations in *PINK1* are now reported to be the second most common cause of autosomal recessive PD (Kawajiri et al., 2011). PINK1 is ubiquitously expressed and acts upstream of Parkin in the same mitophagy pathway (Clark et al., 2006; Unoki and Nakamura, 2001). The PINK1 protein consists of an N-terminal mitochondrial targeting sequence, a transmembrane domain and a kinase domain (Zhou et al., 2008). Following localisation of PINK1 to the mitochondria, the protein is normally cleaved signalling for its degradation, however, with mitochondrial dysfunction and depolarisation of the membrane potential cleavage is inhibited causing PINK1 to accumulate on the outer mitochondrial membrane (Narendra et al., 2010). PINK1 autophosphorylates itself on the mitochondrial membrane recruiting Parkin to the damaged mitochondria (Okatsu et al., 2012). PINK1 also phosphorylates ubiquitin, which subsequently binds to Parkin, causing a conformational change in the protein (Gladkova et al., 2018). This rearrangement of the Parkin protein untethers the RING2 domain, allowing PINK1 to phosphorylate the Ubl domain and activate the ubiquitin ligase activity

of Parkin. PINK1 further phosphorylates a protein on the mitochondrial surface, Mfn2, which Parkin subsequently ubiquitinates signalling for mitophagy (Chen and Dorn, 2013). Similar to Parkin a loss of PINK1 has been shown to cause mitochondrial dysfunction in the mouse (Gispert et al., 2009).

DJ-1

DJ-1 was initially identified as an oncogene working with Ras to promote the malignant transformation of cells (Nagakubo et al., 1997). In 2003, mutations in *DJ-1* were discovered to cause early onset autosomal recessive PD in two families (Bonifati et al., 2003). *DJ-1* mutations are now estimated to occur in around 1% of early onset cases of PD (Abou-Sleiman et al., 2003). The *DJ-1* gene encodes a redox sensitive multifunctional protein, with a neuroprotective role and ubiquitous expression pattern (Nagakubo et al., 1997; Dolgacheva et al., 2019). DJ-1 expression is upregulated in response to oxidative stress and a conserved cysteine residue, C106, is oxidised to activate the neuroprotective function of DJ-1 (Canet-Avilés et al., 2004; Lev et al., 2009). Oxidative stress is thought to be a major pathological mechanism contributing to PD and accordingly the expression and oxidation of DJ-1 is increased in PD patients (Waragai et al., 2006, 2007; Choi et al., 2006; Henchcliffe and Beal, 2008). DJ-1 exerts a neuroprotective effect by combatting oxidative stress (Ariga et al., 2013). This was seen with the overexpression of DJ-1 in a neurotoxin-based model of PD in the rat, reducing oxidative stress and neurodegeneration in the substantia nigra (Inden et al., 2006). Whereas a loss of DJ-1 has been shown to increase the vulnerability of neuronal cells to oxidative stress (Mullett and Hinkle, 2009). Once oxidised, DJ-1 localises to the mitochondria and subsequently the nucleus (Junn et al., 2009). These are two locations where DJ-1 exerts a neuroprotective effect. Specifically, DJ-1 localises to complex I in the mitochondria under oxidative stress conditions to maintain complex I activity (Hayashi et al., 2009). Loss of DJ-1 in the mouse has been shown to cause mitochondrial dysfunction and a progressive loss of ATP production in muscle (Hao, Giasson and Bonini, 2010). Whereas mitochondrial targeting of the DJ-1 protein, in neuronal cell cultures under oxidative stress, has been shown to enhance its neuroprotective effect (Junn et al., 2009). DJ-1 further combats oxidative stress via the transcriptional regulation of antioxidant genes, (Clements et al., 2006; Gu et al., 2009). The translocation of oxidised DJ-1 to

the nucleus is thought to play an important role in its regulation of transcription (Kim et al., 2012). In addition to its role in the oxidative stress response, DJ-1 acts as a molecular chaperone preventing the toxic aggregation of proteins including alpha-synuclein, the main component of Lewy bodies in PD (Shendelman et al., 2004). Overall, DJ-1 exerts a neuroprotective effect by activating an oxidative stress response and inhibiting toxic protein aggregation.

FBXO7

Mutations in *FBXO7* were identified to cause Parkinsonian-pyramidal syndrome in 2008 (Shojaee et al., 2008). Parkinsonian-pyramidal syndrome involves additional neurodegeneration in the pallidum, residing in the basal ganglia, and the corticospinal tract which can cause muscle weakness (Paisán-Ruiz et al., 2010). The *FBXO7* gene encodes an F-box protein which interacts with Parkin and plays a role in the PINK1/Parkin mitophagy pathway (Burchell et al., 2013). F-box proteins recognise specific phosphorylated substrates and present them to an E3 ubiquitin ligase, which in turn marks them for proteasomal degradation through ubiquitination (Skowyra et al., 1997). The domain structure of Fbxo7 consists of an N-terminal ubiquitin-like (Ub1) domain, an F-box motif and a C-terminal proline-rich region for substrate binding (Laman et al., 2005). PINK1 is thought to recruit Fbxo7 to the mitochondria, which in turn recruits Parkin through its Ubl domain (Burchell et al., 2013). In humans there are two Fbxo7 isoforms, isoform 1 with an N-terminal Ubl domain and a shorter second isoform lacking Ubl (Zhou et al, 2016). PD causing mutations in *FBXO7* include a mutation in the Ubl domain (T22M), causing a specific loss of isoform 1 expression, and a mutation in the proline-rich region (R498X) causing a truncation (Zhao et al., 2011). This implicates the loss of Fbxo7 interaction with Parkin, and recognition of ubiquitination substrates, in PD development. Similar to Parkin and PINK1, a loss of Fbxo7 is expected to increase mitochondrial dysfunction and this was seen in Fbxo7-deficient cells, reducing ATP production and increasing oxidative stress (Delgado-Camprubi et al., 2017).

1.4.3 Risk factors for PD

GBA

GBA encodes the lysosomal enzyme glucocerebrosidase (GCase), that degrades glucocerebroside (GlcCer), and homozygous mutation of *GBA* causes the lysosomal storage disorder Gauchers disease (Hruska et al., 2008). In the early 2000s, it became evident that Gauchers disease was associated with an increased risk of developing PD with a Parkinsonian phenotype reported in multiple cases (Neudorfer et al., 1996; Tayebi et al., 2001; Várkonyi et al., 2003). A high occurrence of *GBA* mutations, both homozygous and heterozygous, was later observed in PD patients post-mortem, with a reported frequency of 14% (Lwin et al., 2004). A loss of GCase activity has been seen in the substantia nigra of sporadic PD patients, in addition to those carrying *GBA* mutations, further implicating dysfunctional GCase in the PD pathology (Gegg et al., 2012). The loss of GCase activity causes substrates to accumulate in the lysosome and impair autophagy (Bae et al., 2015). An accumulation of alpha-synuclein and dysfunctional mitochondria, associated with the development of PD, has been observed with GCase inhibition in neuronal cell cultures (Cleeter et al., 2013). The GlcCer that avoids degradation also promotes the formation of Lewy bodies by stabilising aggregates of alpha-synuclein (Taguchi et al., 2017). In turn, the aggregated alpha-synuclein can impair the trafficking of any wild type GCase through the cell, exacerbating autophagy impairment in a positive feedback loop (Mazzulli et al., 2011). Overall, mutation of *GBA* contributes to the development of PD by impairing autophagy.

1.5 Animal models of Parkinson's disease

In vivo animal models of PD have the advantage of the actual brain environment surrounding the DA neurons compared to *in vitro* cellular models. This includes the different neuron subtypes and neuronal circuitry, the downstream networks and the supporting glial cells. Using an animal model also allows us to notice the first signs of dysfunction by symptoms, follow the disease progression and identify possible targets for treatment.

1.5.1 Drosophila models

The *Drosophila* is a widely used model organism for studying human disease, including neurodegenerative disorders (Marsh and Thompson, 2006). As an invertebrate the *Drosophila* body plan and organ system differs more from humans than vertebrate models, however, they share a lot of the same molecular pathways underlying cellular functions (Ugur, Chen and Bellen, 2016). The *Drosophila* genome has been sequenced and found to have orthologs for 77% of human disease causing genes (Reiter et al., 2001). One of the major advantages of using the *Drosophila* to model disease is their genetic tractability. Genome editing tools such as the GAL4-UAS system can be used to ectopically express human disease genes in specific cell types of the *Drosophila* (Brand and Perrimon, 1993). Creating knockout mutations in the *Drosophila* genome is also well established using transposable elements (Ryder and Russell, 2003). Various genetic models of PD have been generated in the *Drosophila* by the transgenic expression and knockdown/knockout of genes orthologous to those causative for PD in humans (Table 1.1). The transgenic expression of mutant human *LRRK2* (G2019S) in the DA neurons has been widely used to model PD in the *Drosophila* (Cording et al., 2017; Liu et al., 2008; Ng et al., 2009; Venderova et al., 2009). A phenotype characteristic of PD has been reported in this model, with a loss of the DA neurons, reduced DA levels in the brain, locomotor defects and L-dopa responsiveness. The movement phenotype was further characterised by Cording et al. (2017) using high-speed video capture of the proboscis extension response (PER). The *Drosophila* extends its proboscis as a coordinated motor response to feed when it detects a food source (Gordon and Scott, 2009). A movement phenotype with aspects of bradykinesia (slowness of movement), akinesia (failure to initiate movement), hypokinesia (reduced amplitude of movement), and tremor was reported in the PER of the *LRRK2*(G2019S) transgenic model (Cording et al., 2017). These are all characteristics of the movement disorder seen in PD patients indicating that a loss of DA neurons in the *Drosophila* can produce a movement phenotype similar to PD (Mazzoni, Shabbott and Cortés, 2012). Transgenic expression of human *SNCA* in the *Drosophila* was similarly able to cause a loss of DA neurons, with alpha-synuclein intraneural inclusions, and locomotor defects (Feany and Bender, 2000). These genetic models of

autosomal dominant PD in the *Drosophila* have therefore been effective at reproducing aspects of the PD pathology and phenotype. Whereas the knockdown and knockout models of autosomal recessive PD in the *Drosophila* have had varying levels of success (Table 1.1). Loss of *Parkin* caused a specific loss of DA neurons in the PPL1 cluster and locomotor defects, however, muscle degeneration was also observed which could impact the movement phenotype (Greene et al., 2003; Whitworth et al., 2005). Similarly, muscle structure was affected in each of the *Pink1* knockdown and knockout models along with abnormal wing structure (Park et al., 2006; Clark et al., 2006; Yang et al., 2006). All three of the *Pink1* models were reported to have locomotor defects, but only two of the three were observed with a loss of DA neurons. This suggests that the effects on muscle and wing structure contribute to the movement phenotype, rather than solely a loss of DA neurons. *DJ-1* has two orthologs in the *Drosophila* genome, and the knockout of both orthologs has been shown to have no effect on the number of DA neurons (Meulener et al., 2006). Contrastingly, knockdown of a single *Dj-1* ortholog was reported to cause DA neurodegeneration (Yang et al., 2005). Overall, the transgenic expression of genes associated with autosomal dominant PD was more effective for modelling PD in the *Drosophila*. The *Drosophila* has potential to be used as a PD model in drug screens as part of the drug discovery pipeline. However, an effective vertebrate model of PD, for screening potential therapies, would be required before moving on to later clinical trials.

Gene	Tg expression / KD / KO	Features	Reference
LRRK2	Tg expression of mutant hLRRK2 (G2019S) in DA neurons	Loss of DA neurons, reduced DA, impaired climbing ability, L-dopa responsive	(Cording et al., 2017; Liu et al., 2008; Venderova et al., 2009; Ng et al., 2009)
SNCA	Tg expression of wild type / mutant alpha-synuclein in neurons	Loss of DA neurons, alpha-synuclein inclusions, impaired climbing ability	(Feany and Bender, 2000)
Parkin	KO	Specific loss of DA neurons in the PPL1 cluster, muscle degeneration, impaired flight and climbing	(Greene et al., 2003; Whitworth et al., 2005)
Pink1	KO	Loss of DA neurons, muscle degeneration, abnormal wings, impaired flight and climbing ability	(Park et al., 2006)
Pink1	KO	No loss of DA neurons, abnormal muscle structure, abnormal wings, impaired flight	(Clark et al., 2006)
Pink1	KD	Loss of DA neurons, reduced DA, muscle degeneration, abnormal wings, impaired flight and climbing ability	(Yang et al., 2006)
Dj-1	KO (both *Dj-1* A and B)	No loss of DA neurons	(Meulener et al., 2005)
Dj-1	KD (A)	Loss of DA neurons, reduced DA	(Yang et al., 2005)

Table 1.1 Genetic models of PD in the *Drosophila*

Examples of genetic models of PD in the *Drosophila.* Tg = transgenic, KD = knockdown (transient loss of expression), KO = knockout (permanent loss of expression). *PINK1* and *PARKIN* have single orthologs in the *Drosophila*. There are two *DJ-1* orthologs in the *Drosophila* (*Dj-1a* and *Dj-1b*) (flybase.org).

1.5.2 Rodent models

The mouse and rat have long been used to model human disease, as both are mammals they share a similar physiology to humans and highly similar genomes with orthologs for nearly all human disease genes (Huang et al., 2004). They are relatively simple to maintain in a laboratory and the protocols for creating transgenic and knockout lines, administering drugs and assessing locomotor ability are well established (Konnova and Swanberg, 2018). Various neurotoxin-based models of PD have been generated in the mouse and rat by treatment with MPTP, rotenone, paraquat and 6-OHDA either orally or by injection. The majority of these neurotoxin-based models have been shown to cause neurodegeneration in the substantia nigra and locomotor defects (Table 1.2). One such model in the rat was reported to have a movement phenotype characterised by hypokinesia and rigidity (Betarbet et al., 2000). These characteristics are commonly seen in PD patients showing how a loss of the nigrostriatal DA neurons in the rat can produce a similar movement phenotype to PD (Mazzoni, Shabbott and Cortés, 2012). However, neurotoxins such as MPTP cause a rapid loss of the DA neurons in rodents (Jackson-Lewis and Przedborski, 2007), contrasting with the progressive loss seen in PD patients (Greffard et al., 2006). Furthermore, loss of the DA neurons in neurotoxin-based models may act through different pathogenic mechanisms to PD in humans making them less reliable for drug screening purposes (Chesselet, 2008).

Genetic models of PD have also been generated in the mouse and rat using transgenic expression, knockdown and knockout methods (Table 1.3 and Table 1.4). The transgenic expression of genes associated with autosomal dominant PD in humans, *LRRK2* (G2019S) and *SNCA*, was shown to cause a loss of nigrostriatal DA neurons and locomotor defects in the mouse, reproducing aspects of PD in humans (Ramonet et al., 2011; Janezic et al., 2013). However, there have been studies that report aggregates of alpha-synuclein forming in the motor neurons of transgenic mice overexpressing alpha-synuclein, which could impact on the movement phenotype (van der Putten et al., 2000). In contrast with transgenic models of PD in the mouse, models of autosomal recessive PD in the mouse have generally failed to produce a PD phenotype (Table 1.3). The loss of *Pink1*, *Parkin*, *Dj-1* or *Fbxo7* in the mouse has been

shown to have no effect on the number of nigrostriatal DA neurons or locomotor ability (Gispert et al., 2009; Kim et al., 2005; Perez and Palmiter, 2005; Vingill et al., 2016). A further triple knockout of *Pink1, Parkin* and *Dj-1* in the mouse failed to produce a phenotype with aspects of PD (Kitada et al., 2009). They reported no effects on the number of DA neurons or their morphology in the substantia nigra but neglected to study the levels of mitophagy and oxidative stress in the triple knockout mice. Whereas, models of autosomal recessive PD in the rat have generally been better at recapitulating the PD pathology and phenotype (Table 1.4). Dave et al. (2014) reported a loss of DA neurons in the substantia nigra, and locomotor defects, with knockouts of *Dj-1* and *Pink1* in the rat. However, no significant effects were observed on the number of DA neurons, or locomotor ability of the *Parkin* knockout rat. Overexpression of *LRRK2*(G2019S) and *SNCA* in the rat brain has been able to cause a loss of the nigrostriatal DA neurons in some cases (Dusonchet et al., 2011; Lo Bianco et al., 2002). Overall, more of the genetic models of PD in the rat have been successful at producing a PD pathology and phenotype than the genetic models in the mouse. Genetic models of PD in the rat could be beneficial for testing potential PD therapies. However, there are limitations to using a rat model in drug screens. Rats are less suited to high-throughput drug screens than other model organisms such as the *Drosophila* or zebrafish, which are much smaller, have lower maintenance costs and produce much larger broods. Ideally, an animal model of PD better suited for high-throughput drug screens would be used in the drug discovery pipeline before a mammalian model such as the rat.

Species	Neurotoxin	Administration	Features	Reference
Mouse	6-OHDA	Injection of SN	Loss of SN DA neurons, locomotor defects	(Bagga, Dunnett and Fricker, 2015)
Mouse	MPTP	Intraperitoneal injection	Loss of SN DA neurons, reduced DA, locomotor defects	(Shimoji et al., 2005; Paul et al., 2017)
Mouse	Paraquat	Oral administration	Loss of SN DA neurons, reduced DA, locomotor defects	(Ren, Zhao and Sun, 2009)
Mouse	Rotenone	Oral administration	Loss of SN DA neurons	(Chiu et al., 2015)
Rat	6-OHDA	Injection of SN	Loss of SN DA neurons, reduced DA, locomotor defects	(Ferro et al., 2005)
Rat	MPTP	Injection of SN	Loss of SN DA neurons, reduced DA, increased locomotor activity	(Ferro et al., 2005)
Rat	Paraquat	Intraperitoneal injection	Loss of SN DA neurons, locomotor defects	(Muthukumaran et al., 2014)
Rat	Rotenone	Systemic administration	Loss of the SN DA neurons, locomotor defects	(Betarbet et al., 2000)

Table 1.2 Neurotoxin-based rodent models of Parkinson's disease

Examples of neurotoxin-based mouse and rat models of PD. SN = substantia nigra.

Gene	Tg expression / KD / KO	Features	Reference
LRRK2	Tg expression of mutant hLRRK2 (G2019S) in the neurons	Loss of SN DA neurons, reduced DA, no locomotor defects	(Ramonet et al., 2011)
SNCA	Tg overexpression of wild type human a-synuclein	Loss of SN DA neurons, reduced DA, locomotor defects	(Janezic et al., 2013)
Parkin	KO	No loss of SN DA neurons, no locomotor defects	(Perez and Palmiter, 2005)
Pink1	KO	No loss of SN DA neurons, locomotor defects	(Gispert et al., 2009)
Dj-1	KO	No loss of SN DA neurons, no locomotor defects	(Kim et al., 2005)
Parkin/Pink1/Dj-1	Triple KO	No loss of SN DA neurons, no loss of DA	(Kitada et al., 2009)
Fbxo7	KO	No loss of SN DA neurons, no loss of DA, locomotor defects	(Vingill et al., 2016)

Table 1.3 Genetic mouse models of Parkinson's disease

Examples of genetic mouse models of PD. Tg = transgenic, KD = knockdown (transient loss of expression), KO = knockout (permanent loss of expression), SN = substantia nigra.

Gene	Tg expression / Vector / KD / KO	Features	Reference
LRRK2	Tg inducible expression of mutant hLRRK2 (G2019S)	No loss of SN DA neurons, no loss of DA, impaired DA reuptake, enhanced locomotor activity	(Zhou et al., 2011)
LRRK2	Vector driving neuronal expression of mutant hLRRK2 (G2019S) injected into the striatum	Loss of SN DA neurons	(Dusonchet et al., 2011)
SNCA	Vector driving alpha-synuclein expression injected into SN	Loss of SN DA neurons	(Bianco et al., 2002)
Parkin	KO	No loss of SN DA neurons, no locomotor defects	(Dave et al., 2014)
Pink	KO	Loss of SN DA neurons, locomotor defects	(Dave et al., 2014)
Dj-1	KO	Loss of SN DA neurons, locomotor defects	(Dave et al., 2014)

Table 1.4 Genetic models of Parkinson's disease in the rat

Examples of genetic rat models of PD. Tg = transgenic, KD = knockdown (transient loss of expression), KO = knockout (permanent loss of expression), SN = substantia nigra.

1.5.3 Zebrafish models

In recent years the zebrafish has grown in popularity as a model organism for studying human disease (Bradford et al., 2017). Firstly, zebrafish are a vertebrate which means they are more closely related to humans than other model organisms such as *C.elegans* and *Drosophila*. Technological advances have also made it easier to both generate and analyse disease models in zebrafish and compared to other vertebrate models zebrafish are better suited for high-throughput drug screening (Lieschke and Currie, 2007). In particular, zebrafish are useful for modelling neurodegenerative diseases because (as vertebrates) they share a similar neuroanatomy to humans and employ the same neurotransmitter pathways, including the dopaminergic system (Rink and Wullimann, 2001; Sager et al., 2010). The zebrafish is especially well suited for modelling PD, with certain DA neurons found in the posterior forebrain of the zebrafish homologous to those in the substantia nigra of humans (Rink and Wullimann, 2001; Blandini and Armentero, 2012). These neurons project to the zebrafish striatum, which shares similarities with the mammalian striatum. Furthermore, orthologs of PD-associated genes have been identified in the zebrafish genome, which are useful for creating genetic models of PD in the zebrafish (Anichtchik et al., 2008; Bretaud et al., 2007; Flinn et al., 2009; Keatinge et al., 2015; Zhao et al., 2012). Transient models of PD in the zebrafish have also been successful at recapitulating aspects of the pathology (Table 1.5 and Table 1.6).

Neurotoxin-based models

Neurotoxins have previously been used to create transient models of PD in the zebrafish. Treatment with MPTP or 6-OHDA has been shown to impair DA neurotransmission and cause locomotor defects in both larval and adult zebrafish (Table 1.5). A marked decrease in locomotor activity and altered complexity in swimming pattern, with larger turn angles, was observed in adults injected with either MPTP or 6-OHDA (Anichtchik et al., 2004). Decreased mobility and freezing bouts were also seen in MPTP injected adults, thought to represent the bradykinesia and dyskinesia seen in PD patients (Babu et al., 2016). They were also responsive to a PD treatment, deprenyl, which rescued the movement phenotype. Altered DA neurotransmission and locomotor defects

were also seen in adult zebrafish injected with rotenone (Table 1.5). These models support the premise that DA deficiency in the zebrafish brain can produce a similar phenotype to that seen in PD patients, with the loss of motor control and bradykinesia present (Jankovic, 2008). However, the loss of DA and motor control in the neurotoxin-based models is rapid (Anichtchik et al., 2004), contrasting with the progressive loss seen in PD patients (Greffard et al., 2006). Neurotoxins also have a transient effect on the DA neurons in zebrafish (Anichtchik et al., 2004; Sallinen et al., 2009), and they may act through different pathogenic mechanisms to PD in humans (Chesselet, 2008). Therefore, neurotoxin-based models of PD in the zebrafish are not the most suitable for drug screening purposes.

Gene knockdown/knockout models

Post-transcriptional knockdown of PD-associated genes is another widely used method of modelling PD in the zebrafish. Morpholino oligonucleotides have been used to inhibit the translation of Lrrk2, Parkin, Pink1, Dj-1 and Fbxo7 in the zebrafish (Table 1.6). A loss of DA neurons was reported with *lrrk2* knockdown, however, the neurodegeneration was not specific and the embryos had developmental defects (Prabhudesai et al., 2016). There were also two zebrafish models with a loss of the WD40 domain in LRRK2 by splice-blocking morpholino. Sheng et al. (2010) reported a loss of DA neurons and responsiveness to L-DOPA treatment, however, a year later Ren et al. (2011) reproduced the model and saw no loss of the DA neurons. This shows how the results of a morpholino knockdown can vary. A similar effect was seen with the morpholino knockdown of *pink1.* Anichtchik et al. (2008) reported developmental defects and a loss of the DA neurons, whereas Xi et al. (2010) saw only altered patterning of the DA neurons. A loss of DA neurons was seen in the *pink1* knockout embryo but no developmental defects were reported (Flinn et al., 2013), suggesting a possible off-target effect in the *pink1* knockdown by Anichtchik et al. (2008). MOs can bind to similar RNA sequences blocking splicing or translation, with increasing likelihood at higher doses, and developmental defects are a possible off-target effect (Bill et al., 2009; Kok et al., 2015). No loss of DA neurons was seen with knockdown of *dj-1* expression (Baulac et al., 2009; Bretaud et al., 2007). A loss was reported in zebrafish embryos with knockdowns of *parkin* and *fbxo7* (Table 1.6). However, *parkin*

morphants had no motor impairment (Flinn et al., 2009), and whilst *fbxo7* morphants did have locomotor deficiencies they also suffered from developmental defects (Zhao et al., 2012). Out of all of these models, there was no loss of the DA neurons and motor impairment reported, without the zebrafish embryo also having developmental defects. Therefore, morpholino knockdown was not effective for generating models of PD in the zebrafish. A further issue is the short-lived effect of MO knockdown when using this approach to model progressive disorders such as PD (Nasevicius and Ekker, 2000; Greffard et al., 2006). This was observed with knockdown of *dj-1*, where the MO efficiency at removing exons was lost at 5dpf and the Dj-1 protein was only absent until 3dpf (Bretaud et al., 2007).

Larval/Adult	Neurotoxin	Administration	Assessment	Features	Reference
Adult	6-OHDA	Intramuscular injection	Anti-TH immunostain / WB, distance travelled, velocity, turn angle	No loss of DA neurons, reduced DA, locomotor defects	(Anichtchik et al., 2004)
Adult	6-OHDA	Injection into ventral diencephalon	Anti-TH immunostain, distance travelled, swimming speed	Loss of DA neurons, locomotor defects	(Vijayanathan et al., 2017)
Adult	MPTP	Intramuscular injection	Anti-TH immunostain / WB, distance travelled, velocity, turn angle	No loss of DA neurons, reduced DA, locomotor defects	(Anichtchik et al., 2004)
Adult	MPTP	Intraperitonial injection	Anti-TH immunostain / WB, distance travelled, swimming pattern, freezing episodes, freezing duration	Loss of DA neurons, locomotor defects	(Babu et al., 2016)
Adult	Rotenone	4 week exposure to rotenone dissolved in water	Anti-TH WB, distance travelled and duration swimming at fast speed	Loss of TH expression, reduced DA, locomotor defects	(Wang et al., 2017)
Larval	6-OHDA	Exposure from 2-4/5dpf dissolved in water	Anti-TH ISH / immunostain, distance travelled, swimming pattern, time spent at lower depths	Loss of DA neurons, locomotor defects	(Feng et al., 2014; Parng et al., 2007)
Larval	MPTP	Exposure from 1-5dpf dissolved in water	Anti-TH ISH, neuronal fluorescent reporter, swimming speed	Loss of DA neurons, locomotor defects	(Bretaud, Lee and Guo, 2004; Wen et al., 2008)
Larval	Paraquat	Exposure from 1-5dpf dissolved in water	Anti-TH ISH, swimming speed	No loss of DA neurons, no locomotor defects	(Bretaud, Lee and Guo, 2004)
Larval	Rotenone	Exposure from 1-5dpf dissolved in water	Anti-TH ISH, swimming speed	No loss of DA neurons, no locomotor defects	(Bretaud, Lee and Guo, 2004)

Table 1.5 Neurotoxin-based models of PD in the zebrafish

Examples of neurotoxin-based models of Parkinson's disease in the zebrafish. TH = tyrosine hydroxylase (expressed in catecholaminergic neurons), ISH = *in situ* hybridisation (RNA expression), WB = western blot (protein expression).

Gene	Targeting	Analysis	Features	Reference
lrrk2	Morpholino knockdown	Anti-TH immunostain, neuronal fluorescent reporter gene	Developmental defects, general neurodegeneration, loss of DA neurons	(Prabhudesai et al., 2016)
lrrk2	Loss of WD40 domain by splice-blocking morpholino	Anti-TH ISH / WB, anti-*dat* ISH, neuronal fluorescent reporter gene, distance travelled	Midbrain neurodegeneration, loss of DA neurons, locomotor defects, L-DOPA responsive	(Sheng et al., 2010)
lrrk2	Loss of WD40 domain by splice-blocking morpholino	Anti-*th* ISH, anti-*dat* ISH	No loss of DA neurons	(Ren et al., 2011)
parkin	Morpholino knockdown	Anti-TH ISH / immunostain, distance travelled	Loss of DA neurons, no locomotor defects	(Flinn et al., 2009)
pink1	Morpholino knockdown	Anti-TH immunostain	Developmental defects, loss of DA neurons	(Anichtchik et al., 2008)
pink1	Morpholino knockdown	Anti-TH ISH / immunostain, touch response, distance travelled	Altered DA neuron patterning, locomotor defects	(Xi et al., 2010)
pink1	TILLING knockout	Anti-TH ISH / immunostain, anti-*dat* ISH	Loss of DA neurons	(Flinn et al., 2013)
dj-1	Morpholino knockdown	Anti-TH ISH / immunostain	No loss of DA neurons	(Bretaud et al., 2007; Baulac et al., 2009)
fbxo7	Morpholino knockdown	Anti-TH ISH, swimming speed	Developmental defects, loss of DA neurons, locomotor defects	(Zhao et al., 2012)

Table 1.6 Gene knockdown/knockout models of PD in the zebrafish

Examples of gene knockdown and knockout models of Parkinson's disease in the larval zebrafish. TH = tyrosine hydroxylase (expressed in catecholaminergic neurons), DAT = dopamine transporter (expressed in dopamine neurons), ISH = *in situ* hybridisation (RNA expression), WB = western blot (protein expression).

1.5.4 Objective analysis of the movement phenotype

PD is characterised by a specific movement phenotype of bradykinesia and resting tremor caused by a loss of DA neurons in the substantia nigra region of the brain (Lees et al., 2009). Therefore, an effective zebrafish model of PD is expected to have a movement phenotype distinct from that of a wild type zebrafish. The exact features of movement that would characterise a zebrafish model of PD are difficult to predict and identifying the changes in swimming movement is challenging. Analyses to discern these subtle features by human observation are too subjective; the chosen features to study and how they are measured would differ from person to person. Four individuals, experienced in zebrafish movement, independently categorised larval zebrafish movements and differed in their category choices by up to 32% (Mirat et al., 2013). This demonstrates the subjectivity of people when studying zebrafish movement. Researchers have analysed video recordings of larval zebrafish, frame by frame, and measured the angle along the tail when a tail bend has been completed (Budick and O'Malley, 2000). Here individual investigators could select different frames for completed tail bends and when measuring the angles they could end up with dissimilar values. Again, humans being subjective would affect the results. Machines, however, are objective and therefore more reliable for identifying changes in movement to provide more unbiased, reproducible results (Lones et al., 2014). This would involve digitising the swimming movements of fish and extracting features that can objectively describe the movement characteristics.

Statistical approaches, such as Principal Component Analysis (PCA), are commonly used to process extracted movement data and identify patterns in data sets with a large number of variables (Molenaar, Wang and Newell, 2013). In PCA, variables are condensed down into principal components (PCs) by linear combination of any that show correlation (Lever, Krzywinski and Altman, 2017). The complexity of the data is reduced whilst any trends are retained in the PCs making patterns easier to identify. In a study to describe the movement characteristics of PD in humans, it was found that traditional statistical approaches were not the most effective for identifying patterns in the complex nonlinear data obtained (Smith et al., 2007). For this reason a more effective

machine learning approach to recognise a pattern in PD patient movement was adopted. In this thesis, a similar objective method of machine learning was used to identify any movement phenotype distinguishing a zebrafish PD model from the wild type with a measured level of accuracy. Machine learning was chosen for the task because of its ability to automatically analyse large data sets and 'learn' to recognise differences between the classes of data (Alpaydin, 2016). The data was subsequently used to model and define the movement phenotype in zebrafish PD models.

1.6 This thesis

There are two main objectives of this thesis. The first is to develop a genetic model of PD in the zebrafish with molecular and cellular signatures of the human disease (Chapters 3 and 4). The second is to assess the movement phenotype using selected features of movement as well as unbiased analysis of movement data (Chapter 5). Evolutionary algorithms, a method of machine learning proven effective at diagnosing PD in humans (Lones et al., 2014), were used to objectively characterise the movement phenotype.

The aims of this thesis:

1. To create genetic models of PD in the zebrafish by CRISPR/Cas9 targeting of PD-associated gene orthologs
2. To characterise the expression of each target gene in the zebrafish
3. To describe the mutation in each target gene and the translated protein
4. To determine if there has been nonsense-mediated decay of the mutant transcript in each PD model
5. To assess whether the DA neurons have been affected in the PD models
6. To analyse the movement phenotype of PD models by calculating select features of movement
7. To evolve classifiers that can discriminate zebrafish modelling PD using the features of movement and raw movement data

2 Materials and Methods

2.1 Zebrafish

2.1.1 Husbandry

All zebrafish used in this study originate from the London wild type (LWT) strain. Zebrafish were kept at 28°C on a 14 hour light / 10 hour dark cycle, with regular feeding three times a day. Any embryos collected were incubated at 28°C in E3 media (5mM NaCl, 0.17mM KCl, 0.33mM $CaCl_2$, 0.33mM $MgSO_4$) with 0.1% methylene blue for up to five days post fertilisation (dpf). To anaesthetise adult zebrafish 50ml of 0.4% (w/v) MS-222 dissolved in dH_2O was added to 200ml of tank water before adding the fish.

The experimental procedures carried out in this thesis are in accordance with the UK Home Office Animals (Scientific Procedures) Act 1986 under project licence PPL POF245295 held by Dr Betsy Pownall and Personal Licence PIL I3351E3A2 held by Gideon Hughes.

2.2 CRISPR/Cas9 targeting in zebrafish embryos

2.2.1 Guide RNA design

The online CRISPR design tool ChopChop (https://chopchop.rc.fas.harvard.edu/) was used to identify a target sequence, directly upstream of a 3bp PAM site (5' NGG 3') and near the start of the coding region, for each target gene. Each 20bp target sequence was incorporated into a forward primer for guide RNA (gRNA) production. Additionally, a promoter sequence for increasing transcription efficiency with T7 RNA polymerase was added to the 5' end of the forward primer along with an additional G nucleotide necessary for the T7 polymerase to work (Nakayama et al., 2014).

Target gene	Target sequence	gRNA synthesis forward primer
parkin	GGAGCAGAGCACGGTTCATG	GCAGCTAATACGACTCACTATAG GGAGCAGAGCACGGTTCATG GTTTTAGAGCTAGAAATAGCAAG
pink1(1)	GGTGAAGCAGAAAGTCGAAG	GCAGCTAATACGACTCACTATAG GGTGAAGCAGAAAGTCGAAG GTTTTAGAGCTAGAAATAGCAAG
pink1(2)	CGCAGCTGTTTATGAGGCTG	GCAGCTAATACGACTCACTATAG CGCAGCTGTTTATGAGGCTG GTTTTAGAGCTAGAAATAGCAAG
dj-1	GCCGGTTCAGTGCAGCCGTG	GCAGCTAATACGACTCACTATAG GCCGGTTCAGTGCAGCCGTG GTTTTAGAGCTAGAAATAGCAAG
fbxo7	GCTGTCTCTGAATGGTAAAG	GCAGCTAATACGACTCACTATAG GCTGTCTCTGAATGGTAAAG GTTTTAGAGCTAGAAATAGCAAG
gba	GTCTCTGTCTTCCGGAGCTC	GCAGCTAATACGACTCACTATAG GTCTCTGTCTTCCGGAGCTC GTTTTAGAGCTAGAAATAGCAAG

Table 2.1 Target and primer sequences for generating gRNA template

The reverse primer used for gRNA synthesis was shared between all of the gRNAs: 5'-AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATT TTAACTTGCTATTTCTAGCTCTAAAAC-3' (Nakayama et al., 2014)

2.2.2 Generation of guide RNA

A DNA template of each gRNA was generated by PCR of two overlapping primers, the site specific forward primer (Table 2.1) and the non-variable reverse primer containing a scaffold sequence. Primers were ordered and diluted to a 10μM working stock before making up PCR reaction mixtures containing 5μl forward primer, 6.5μl reverse primer, 1μl dNTPs (10μM), 10μl 5x HF buffer, 1μl Phusion polymerase and 27μl ddH_2O. Reactions were heated to 98°C for 30 seconds before 35 cycles of denaturation, annealing and amplification at 98°C for 10 seconds, 60°C for 30 seconds and 72°C for 15 seconds, before a final extension at 72°C for 10 minutes. 5μl from each PCR

was ran on a 2% agarose gel containing TAE buffer to validate the product. The gRNAs were then transcribed from 4µl of DNA template using the Ambion MEGAshortscript T7 Transcription Kit following manufacturers instructions. 2µl of each gRNA was ran on a 2% agarose gel containing TBE buffer to validate. 1µl of Turbo DNase was then added to every gRNA before incubating the mixtures at 37°C for 15 minutes to degrade any leftover template. Subsequently, 400µl of phenol chloroform, 340µl of dH_2O and 40µl of 5M ammonium acetate was added to each gRNA before vortexing for one minute and centrifuging at 13,000 rpm for five minutes. The aqueous phase was isolated from each mixture and added to an equal volume of chloroform before vortexing for one minute and centrifuging at 13,000 rpm for a further five. Again the aqueous phase was extracted and added to an equal volume of 100% ethanol before vortexing to mix and storing at -20°C overnight to precipitate. The following day solutions were centrifuged at 13,000 rpm for 30 minutes to form a gRNA pellet and the supernatant was removed. Pellets were washed by vortexing in 500µl of 70% ice cold ethanol and then centrifuged at 13,000 rpm for 15 minutes. The supernatant was removed from each gRNA pellet before desiccating for 15 minutes to remove any leftover ethanol. gRNAs were resuspended in 20µl of ddH_2O by heating to 80°C for one minute and vortexing for a further minute. 1µl of each gRNA was ran on a 2% agarose gel containing TBE buffer to check that purification was successful and then quantified using a NanoDrop ND1000 spectrophotometer.

2.2.3 Microinjecting gRNA and Cas9 protein

Zebrafish embryos were injected with 2nl of a 300pg gRNA/1ng Cas9/300mM KCl mixture into the yolk at the one cell stage. This was consistent for all target genes apart from *pink1* where two gRNAs were injected, both at 250pg.

2.3 Nucleic acid extraction

2.3.1 Genomic DNA extraction for sequencing

100µl of 50mM NaOH was added to each zebrafish embryo or fin clipping before heating at 95°C for 15 minutes. 25µl of 1M Tris pH 8.0 was then added to each tube, pipetting up and down to break up the tissue and neutralise the solution.

2.3.2 RNA extraction from zebrafish brains

Adult zebrafish were euthanized and brains were extracted in PBS. Brains were immediately pulverised in 1ml of Tri Reagent (Sigma) and left at room temperature for 5 minutes. 200µl of chloroform was added to each tube before shaking vigorously for 15 seconds. The mixtures were then left at room temperature for 10 minutes before centrifuging at 13,000 rpm for 10 minutes at 4°C. 400µl was taken from the aqueous layer in each tube and transferred into a clean 1.5ml centrifuge tube. 400µl of isopropanol was then added to each tube before vortexing to mix and leaving at room temperature for 5 minutes. The mixtures were centrifuged at 13,000 rpm for 10 minutes at 4°C to pellet the nucleic acids. The supernatant was removed from each tube and 1ml of ice-cold 70% ethanol was added to the pellet before vortexing. Solutions were centrifuged at 13,000 rpm for 10 minutes at 4°C to repellet the nucleic acids again. Pellets were dessicated for 15 minutes to remove any leftover ethanol and resuspended in 45µl of dH_2O before vortexing, heating at 80°C for 1 minute and vortexing again to mix. 5µl of each sample was ran on a 1.5% agarose gel containing TBE buffer to check for the presence of 18S and 28S ribosomal RNA bands.

2.4.3 RNA clean-up

RNA samples were cleaned up using the Zymo RNA Clean & Concentrator kit following the manufacturers instructions. This included the optional DNase I treatment before RNA purification and samples were eluted in 15µl of dH_2O for

the final step. 1µl of each RNA sample was ran on a 1.5% agarose gel containing TBE buffer to check for the presence of 18S and 28S ribosomal RNA bands. Samples were quantified using a NanoDrop ND1000 spectrophotometer and stored at -80°C until use.

2.4 Polymerase chain reaction (PCR)

2.4.1 cDNA synthesis

cDNA was synthesised from 1µg of each RNA sample using Superscript IV (Thermo-Fisher) reverse transcriptase following the manufacturers instructions. Random hexamers were used rather than oligo $(dT)_{20}$ primers or gene-specific reverse primers in the reaction mixture.

2.4.2 PCR/RT-PCR

PCRs were carried out using Promega PCR Master Mix with either 1µl of extracted gDNA or 2µl of cDNA as template in the following 20µl reaction mixture:

- 10µl of 2X PCR Master Mix
- 1µl of 10mM Forward Primer
- 1µl of 10mM Reverse Primer
- 1µl gDNA / 2µl cDNA
- dH_2O up to a total volume of 20ul

DNA was initially denatured for 2 minutes at 95°C before 30 PCR cycles with the following conditions:

- 95°C for 30 seconds
- 5°C below Tm° for 30 seconds
- 72°C for 60 seconds/kb product

Followed by a final extension at 72°C for 10 minutes.

Primers are listed:

Gene	Forward primer	Reverse primer	Task
parkin	ATGCCAAATCTCTGAGGTGTATT	CTCGGGTATCGTTCCAACTAAAG	Sequencing in plasmid / Restriction digest
pink1	GGCTGTATTTAGAAAGAAGAAGTTTCAG	GCAGCACAGTACAATTGTCAACTATAAA	Sequencing in plasmid / PCR genotyping
dj-1	GGCTCTGGCCATCATTACTAT	GTAAAGTCAGACCTGTTTGTGTG	Sequencing in plasmid / Restriction digest
fbxo7	ATGCTTTTTATTGATTATTG	GAGGGGGTTTACATTTCT	Sequencing in plasmid / Restriction digest
gba	GTGGCTGAAGAATCAGGCTTA	GTCTTCTGGAGTGTCAGCATAC	Sequencing in plasmid / Restriction digest
parkin	AAAGCCCACAGCAGTTTCTA	CCACAGCCGAGTCCATTATT	RT-PCR / ISH probe
pink1	GGCTACAGGCTGGAAGATTATG	CCAGGAGGATGTTGTCTGATTT	RT-PCR / ISH probe
dj-1	CGGAGGAGATGGAGACTGTAA	CTCCACTATAGTCAAGGCAAACTC	RT-PCR / ISH probe
fbxo7	CGGCTGTCTCTGAATGGTAAA	CTCCAAACTCGTCTCCATCTTC	RT-PCR / ISH probe
gba	CTCGCCGGAATAATCACCACAGC	GTCAGCCCCCAGAACGAAAGATTA	RT-PCR / ISH probe
dat	CCGTCCGGACAGCAGCGTGAGA	GTGGCGTTGGGGTCGGAGCAGT	RT-PCR / ISH probe
th	TGTTGGGAGGCGGCAGAGTTTGAT	CTCCCGCCATGTTCCGATTTCTTC	RT-PCR / ISH probe
syn2a	CCCGCGCCAGGAGAGGAA	AGCGTACTGGAGGCCGATGATTA	RT-PCR / ISH probe
smyhc	CAGGCTCCCAGAATGAGATTGAAG	AGCTTCATTCCTGCTGCGAG	RT-PCR / ISH probe
ef1a	CTGGAGGCCAGCTCAAACAT	ATCAAGAAGAGTAGTACCGCTAGCATTAC	RT-PCR

Table 2.2 Primer sequences for PCR / RT-PCR

Amplicons were used for genotyping purposes either by sequencing in plasmid, genotyping by PCR product length or genotyping by restriction digest assay. RT-PCRs and ISH (in situ hybridisation) probes were used to assess spatiotemporal expression of a gene.

2.4.3 DNA gel electrophoresis

1-1.5% agarose gels containing either TAE (50X: 242g/L Tris, 57.1ml/L Glacial Acetic acid, 100ml/L 0.5M EDTA pH8.0) or TBE buffer (10X: 108g/L Tris, 55g/L Boric acid, 40ml/L 0.5M EDTA pH8.0) and 0.05µg/ml ethidium bromide were made. DNA / RNA samples were loaded on the agarose gels with NEB 6x gel loading dye alongside 3µl of NEB 2-log DNA ladder and ran at 150-180V.

2.4.4 Quantitative PCR

Primers were designed using NCBI Primer-BLAST with PCR products of between 50 and 150bp. At least one of the primers in a pair had to span an exon-exon junction to limit amplification to cDNA rather than gDNA. Primer melting temperatures ranged from 58 to 62°C with a maximum Tm° difference of 2°C. Primer size varied between 18 and 24bp and GC content ranged between 35 and 65%. The maximum number of nucleotide repeats in a row was 3 and the maximum number of G or C nucleotides allowed in the last 5 bases at the 3’ end was 3.

Gene	Forward primer	Reverse primer
parkin	CGGTGCAAAGACTGCAAACA	GTAGAACTCCGCCAGTCT
pink1	CCACAGACTGATGTGCAGGA	CTGATGACGTTCAGCTGGTG
dj-1	CGCACACAAACAGGGTCCATA	AAGCAAACCTCCAGGCAGAA
fbxo7	TCACTCTGGAGACACCTGCT	TCCAGTCTGTGTCTCTGT
gba	CGAGTAGGTGTGTCCTTTAGTC	CCTCCTCTTCAGACCTGTTTAG
dat	CGTCACCAACGGTGGAATCTA	TGCCGATGGCCTCAATTAGTA
th	GACGGAAGATGATCGGAGACA	CCGCCATGTTCCGATTTCT
pitx3	GACAACAGTGACACAGAGAAGT	TGTCGGGATAACGGTTTCTC
syn2a	CAAGGTTGTGAGGTCATTTAAGC	GTACTGGAGGCCGATGATTAAG
ef1a	TTGAGAAGAAAATCGGTGGTGCTG	GGAACGGTGTGATTGAGGGAAATTC
pkma	CCGCACACAGCACATGATTG	ACTGCATCTGAAGGGTCACG
gpx3	ACCATAAACGGGACGCAGTT	AGTGCATTCAGTTCCACGTACT
gatm	TCCAGAGGAGTCAGGTTACGA	ATGTTGAAGGTGGCGTCGAT
vamp3	GAGCTGGATGATCGTGCTGA	ATGGCCCACATCTTGACGTT
brk1	GCAGGTCTCGTTTAGCCACT	TTGTGACTCTGGCTTCGATGT
abcf2a	CGGAACAGGCAAATCCATGC	GCTCTGGAATTGGCACCTCT
kcnk3	AGGTCGTGTTGCTCCTGAAG	CATGGCCATAACCTATGGTTGTTA
ccna2	AAAGCAGCTAACAACAGGACAGT	GGTTTACACGCAATTATCTGTGG
ccnb1	AAGGTTCAACATGGCGTTAAG	CCGGAGGTCTCCATAGGAA
ccnb2	GGTCCACTACAACCCATCTGA	CATAGGTGGAGTAGTGTTGCTGA
ccnd1	TTCCTTGCCAAACTGCCTAT	GCTTGCGA TGAAGTTGACG
ccne	TCTTCAACCCAAAATGAGAGC	CCCAAATAAAATGTTTCTCTGTGTAA

Table 2.3 Primer sequences for qPCR
The following primer pairs were obtained from previously published works: *pink1* (Flinn et al., 2013), *dat* and *th* (Chen et al., 2012), *pitx3* (Sanchez-Simon et al., 2010), *ef1a* (Fan et al., 2010).

A qPCR standard curve assay was performed with serial dilutions of cDNA, for each pair of qPCR primers, to ensure amplification efficiencies fell between 90 and 110%.

cDNA was synthesised from 1µg of each RNA sample in a 20µl reaction mixture. qPCRs were then carried out using Fast SYBR Green 2X Master Mix (Applied Biosystems) with 3µl of cDNA as template using the following 20µl reaction mixture:

- 10µl of 2X SYBR Green Master Mix
- 0.5µl of 10mM Forward Primer
- 0.5µl of 10mM Reverse Primer
- 3µl cDNA
- 17µl dH_2O

Reactions were prepared in 96 well 0.1ml qPCR plates and the qPCR was ran in a QuantStudio 3 RealTime PCR System (ThermoFisher Scientific) using the Fast Run mode and Comparative CT (ΔΔCT) settings.

Each qPCR analysis was carried out with three biological replicates and the mean cycle threshold (Ct) value was normalised to the Ct of reference gene ef1α generating a delta-Ct (dCt) value for genes of interest. Delta-delta-Ct (ddCt) values were then calculated by subtracting the dCt values of wild type siblings from mutants and the relative fold change was equal to 2^(-ddCt). Student t-tests were carried out comparing the dCt values of mutants and wild type siblings for each gene. GraphPad Prism5 was used to generate graphs showing the relative fold change with error bars representing standard error of the mean (SEM).

2.5 Sequencing mutant lines

2.5.1 PCR product clean-up

PCR products were cleaned up using the Monarch DNA Gel Extraction Kit (NEB). dH_2O was added to each PCR product to reach a total volume of 100µl

before adding 400µl of the Gel Dissolving Buffer. Samples were loaded into the provided spin columns inside collection tubes. Tubes were centrifuged at 13,000rpm for 1 minute before discarding the flow-through. 200µl of DNA Wash Buffer was then added to each spin column before centrifuging for a further minute and discarding the flow-through. The DNA Wash Buffer step was repeated before transferring the spin column to a clean 1.5ml centrifuge tube and adding 10µl of dH_2O. Tubes were left for 1 minute at room temperature before centrifuging for 1 minute to elute the DNA sample.

2.5.2 Ligation of PCR product into plasmid

PCR products were ligated into the pGEM T-Easy Vector system (Promega) using the following 10µl reaction mixture:

- 5µl of 2X Rapid Ligation Buffer
- 3µl of PCR product
- 1µl of pGEM T-Easy Vector
- 1µl of T4 DNA Ligase

The ligation mixture was gently vortexed to mix and stored overnight at 4°C.

2.5.3 Transformation of competent E.coli

100µl of *dam*$^-$/*dcm*$^-$ competent *E.coli* was transformed with 10µl of plasmid. The plasmid was added to the *E.coli* before heatshocking at 42°C for 30 seconds, storing on ice for 5 minutes, then adding 900µl of LB broth (1L: 5g yeast extract, 10g tryptone, 10g NaCl) and incubating at 37°C for 1 hour. Several 10-fold dilutions were made and 100µl was plated from each dilution on 100µg/ml ampicillin LB agar plates (1.5% agarose in LB broth) using asceptic technique. The plates were incubated at 37°C overnight.

2.5.4 Liquid cultures

Single colonies were selected from plates using sterile pipette tips and transferred to 3ml of LB broth with ampicillin (1µg/ml). The tips in LB broth were incubated at 37°C overnight with agitation to grow liquid cultures.

2.5.5 Isolating plasmid DNA from bacteria

Liquid cultures were centrifuged at 10,000 rpm for 10 minutes at 4°C to pellet the bacteria before isolating plasmids using the QiaPrep Spin Miniprep kit (Qiagen) following the manufacturers instructions and eluting in 50µl of dH_2O. The concentration of each plasmid was determined using a Nanodrop ND1000 spectrophotometer.

2.5.6 Sequencing in plasmid

Plasmids were digested with EcoRI (Promega) to check for insert of the PCR products. Restriction digests were set up using the following 20µl reaction mixture:

- 2µl of 10X Buffer H (Promega)
- 1µl of EcoRI (Promega)
- 5µl of plasmid
- 12µl of dH2O

Digestion products were ran on a 1% agarose gel containing TAE. Plasmids with a confirmed insert were diluted to 100ng/µl. 5µl of the diluted plasmid was then combined with 5µl of SP6 sequencing primer (5µM) and sent for Sanger sequencing.

SP6 sequencing primer: AGGTGACACTATAGAATACTCAAGC

2.5.7 Heteroduplex analysis

Heteroduplex analysis was used to identify heterozygous mutants in the F1 generation of each mutant line. The target gene was PCR amplified from extracted gDNA using the same primers as those used for sequencing in plasmid (Table 2.2). 15µl of each PCR product was loaded onto a 15% polyacrylamide gel containing TBE (30ml: 15ml 30% acrylamide mix, 3ml 10X TBE, 11.66ml dH_2O, 300µl 10% APS, 30µl TEMED) and ran at 150V for 2.5 hours in 1X TBE. Gels were then soaked in 1X TBE containing ethidium bromide 0.5µg/ml for 10 minutes before imaging.

The denaturation and annealing of homozygous DNA during PCR produces a homoduplex, whilst PCR annealing of heterozygous DNA creates two different homoduplexes and two different heteroduplexes (Figure 2.1A). A homoduplex is double stranded DNA (dsDNA) where the strands are identical, a heteroduplex is dsDNA where the strands differ. The difference in sequence causes a looping out of the unpaired region in a heteroduplex (Nemudryi et al., 2014). Using this technique to identify heterozygous mutants relies on the fact that heteroduplexes migrate slower through the gel than homoduplexes due to the looping out. Multiple slower travelling bands of heteroduplex DNA are seen when the individual is a heterozygous mutant (Figure 2.1B).

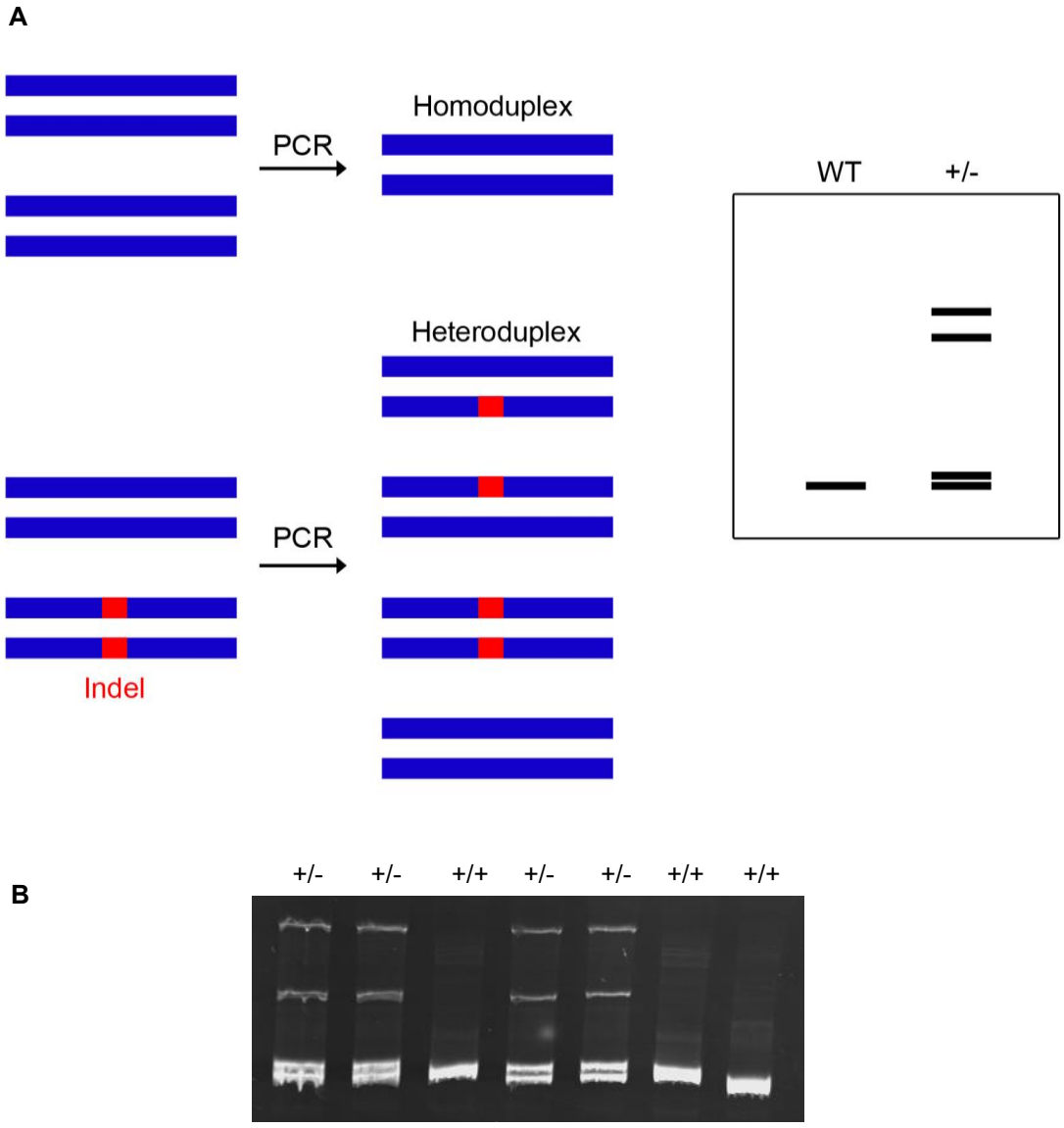


Figure 2.1 Heteroduplex analysis of an F1 generation

(A) Schematic of the heteroduplex analysis. PCR annealing from homozygous DNA produces a single homoduplex, PCR annealing from heterozygous DNA produces two homoduplexes and two heteroduplexes (left). PCR products ran on a polyacrylamide gel (right). A single band is seen for wild type, multiple bands are seen for heterozygous mutants. (B) Example of PCR products amplified from the CRISPR target of an F1 generation ran on a 15% acrylamide gel containing TBE. Heterozygous mutants (lanes indicated by +/-) are identified by the presence of four bands on the gel, wild type zebrafish (+/+) are identified by a single band.

2.5.8 Genotyping by restriction enzyme digestion

In order to genotype the progeny of a mutant incross, the target gene was amplified from gDNA extracted from fin clippings before using a restriction digest assay. Restriction sites present in the target genes, but with a differing number of sites between the mutant and wild type alleles, were identified.

Restriction enzymes used are listed:

Target gene	Restriction enzyme	Wild type allele digestion products	Mutant allele digestion products
parkin	BsaJI	464bp	360bp, 109bp
dj-1	BbvI	214bp, 55bp	286bp
fbxo7	MnlI	244bp, 155bp, 48bp, 42bp	244bp, 199bp, 42bp
gba	SacI	519bp, 210bp	724bp

Table 2.4 Restriction enzymes for genotyping mutants

The following 20µl reaction mixture was set up for each restriction digest:

- 10µl of PCR product
- 7µl of dH_2O
- 2µl of 10x NEB Buffer
- 1µl of restriction enzyme

Each reaction was incubated at 37°C for 3 hours to ensure digestion of the PCR products. 10µl from each digest was then ran on a 1.2% agarose gel containing TBE buffer to check the length of digestion products.

2.6 Whole mount *in situ* hybridisation (ISH)

2.6.1 Embryo fixation

Embryos were demembraned and fixed in MEMFA (0.1M MOPS pH7.4, 2mM EGTA, 1mM MgSO4, 3.7% formaldehyde) for 1 hour before washing twice for 5 minutes in PBS. Embryos were then stored in 100% methanol at -20°C until use.

2.6.2 Synthesis of ISH probes

In order to characterise the expression profile of each target gene and certain markers in zebrafish RT-PCR and *in situ* hybridisation (ISH) were used. Primers were designed to amplify a sequence of approximately 400-600bp in the exonic region of each gene (Table 2.2). These primers were used for RT-PCR analysis on cDNA generated from different zebrafish stages/tissues (Figure 3.4). Products amplified from wild type cDNA were ligated into pGEM T Easy vector (Promega) for ISH probe synthesis (Figure 3.2). The plasmids were sequenced to determine the orientation of each insert and linearised with either NcoI or SalI. The enzyme had to digest at the 5' end of each insert. RNA probes were then transcribed with either T7 or SP6 polymerase in the antisense direction, incorporating DIG-UTP.

The enzyme used for template linearization and RNA polymerase used for transcription are listed for each RNA probe:

Gene	Restriction enzyme	RNA polymerase
parkin	SalI	T7
pink1	NcoI	SP6
dj-1	NcoI	SP6
fbxo7	NcoI	SP6
gba	NcoI	SP6
syn2a	NcoI	SP6
smyhc	NcoI	SP6

Table 2.5 Restriction enzymes used for ISH probe templates

2.6.3 Whole mount ISH protocol

Embryos were rehydrated through a methanol/PBST gradient of washes (75% methanol / 25% PBST, 50% methanol / 50% PBST, 100% PBST). This was followed by Proteinase K treatment to permeabilise the embryos. 10µg/ml of Proteinase K in PBST was added to the embryos before gentle agitation for 10 minutes multiplied by the number of days post fertilisation. Embryos were washed twice in 5ml of 0.1M Triethanolamine pH 7.8, 5 minutes for the first wash, 5 minutes in the second wash with an added 12.5µl of acetic anhydride (Sigma) before adding an additional 12.5µl for a further 5 minutes. Embryos were then washed twice for 5 minutes in PBST before refixing in 10% formaldehyde / 90% PBST for 20 minutes and washing a further five times in PBST. Next embryos were transferred into hybridisation buffer (50% formamide, 5X SSC, 100µg/ml heparin, 1X Denhart's, 0.1% Tween-20, 0.1% CHAPS,

10mM EDTA) containing 1mg/ml of boiled total yeast RNA and rocked at 60°C for 2 hours. DIG probes were heated at 80°C for 3 minutes before adding to hybridization buffer (2-4µl per ml) and replacing the hybridization buffer covering the embryos. Embryos were incubated at 60°C overnight with the DIG probe in hybridization buffer while rocking. The following day the hybridization buffer with probe was removed and embryos were washed twice for 10 minutes in hybridization buffer at 60°C. Embryos were then washed three times for 20 minutes in 2X SSC + 0.1% Tween before a further three washes for 30 minutes in 0.2X SSC + 0.1% Tween all at 60°C. Next embryos were washed twice for 15 minutes in maleic acid buffer (MAB: 100mM maleic acid, 150mM NaCl, 0.1% Tween-20 pH 7.8) before incubating in MAB + 20% heat treated lamb serum (heated to 60°C for 30 minutes) + 2% Boehringer Mannheim Blocking Reagent (BMB) for 2 hours, all at room temperature. This was replaced with more of the same blocking solution with a 1/2000 dilution of affinity purified anti-DIG fragments coupled to alkaline phosphatase (Roche) and rolled overnight at 4°C. The following day embryos were washed three times for 5 minutes, followed by three hour long washes in MAB. Embryos were then washed twice in alkaline phosphatase buffer (100mM Tris pH 9.5, 50mM $MgCl_2$, 100mM NaCl, 0.1% Tween-20) first for 3 minutes, then for 10 minutes. The alkaline phosphatase buffer was then replaced with 1ml of BM Purple staining solution (Roche) and left until colour developed in the embryos. After staining embryos were washed twice for 15 minutes in PBST and stored in MEMFA.

2.6.4 Imaging zebrafish embryos

A SPOT SP401-230 camera (Diagnostic Instruments Inc.) attached to a Leica MZFLIII microscope and a computer running the SPOT advanced software was used to photograph embryos.

2.7 RNA-Seq analysis

2.7.1 Preparation and sequencing of RNA samples

(Undertaken by Sally James, Technology Facility, University of York)

The integrity of each RNA sample was measured using an Agilent 2100 Bioanalyzer (Figure 4.2). An NEBNext Poly(A) mRNA Magnetic Isolation Module was then used to isolate the mRNAs from 0.5µg of each total RNA sample. cDNA libraries were generated using the NEBNext Ultra RNA Library Prep Kit for Illumina following the manufacturer's instructions. Unique adaptor sequences were added to the fragments in each cDNA library before pooling them together. The pooled cDNA libraries were then sequenced using 2 x 150bp paired end reads on 1 lane of an Illumina HiSeq 3000 system at the University of Leeds Next Generation Sequencing Facility.

2.7.2 Computational analysis

(Undertaken by Katherine Newling, Technology Facility, University of York) Approximately 300 million reads were produced by Illumina sequencing equating to ~50 million reads per sample. FastQC analysis was used to check the quality of reads in the FASTQ files generated and Cutadapt 1.16 was used to trim the adaptor sequences from each read. The FASTQ files were then aligned to genome assembly GRCz11, the transcriptome was annotated using RefSeq (NCBI) and transcript abundance was quantified using Salmon 0.10.2. Transcripts Per Million (TPM) values were calculated to normalize the reads for gene length and sequencing depth. Differential expression analysis was then carried out on the aligned transcripts using Sleuth 0.30.0. The likelihood ratio test was used to calculate statistical significance and the Wald test was used to calculate the beta values of transcripts, analogous to fold change. 32 transcripts had significantly altered expression ($q < 0.05$). A heatmap representing the changes in expression of a curated set of these transcripts was created by uploading the TPM values to https://software.broadinstitute.org/morpheus/.

2.7.3 Gene set enrichment analysis

A gene set enrichment analysis (GSEA) was performed using the GSEA 4.0.0 (Broad Institute) software on data from the RNA-Seq analysis. GSEA determines whether there are changes in the expression of large gene sets that have been grouped by biological function. In GSEA the genes from an expression analysis are ordered by correlation between expression and

genotype before determining whether the genes in each gene set are overrepresented at the top or bottom of the ordered list (Subramanian et al., 2005). The main advantages of using GSEA to analyse transcriptomic data are that it uses all of the genes in a data set, rather than just those with a significant difference, and gene-gene dependency is maintained during the analysis (Tamayo et al., 2016). GSEA also has more sensitivity when it comes to recognising coordinated regulation of genes grouped by function compared to the study of single differentially expressed genes. The results of GSEA are also easier to interpret and more reproducible than just studying the significantly different genes from a transcriptomic analysis (Subramanian et al., 2005). One disadvantage of GSEA is that many gene sets can overlap because of genes playing roles in multiple biological processes (Simillion et al., 2017). Using the HALLMARK gene sets, which was created using a computational method to reduce any potential overlap between the gene sets, helps overcome this problem (Liberzon et al., 2015). Another possible issue with using GSEA is the effect sample source bias can have on the results, this is most apparent when gene sets specific to the tissue type being analysed are tested for significance against genes from the entire genome (Simillion et al., 2017). Following the instructions from the Broad Institute website (software.broadinstitute.org/gsea) an expression dataset file (.gct) was created containing the TPM values for each transcript in a sample, and a phenotype labels file (.cls) was created to label the phenotype of each sample. The Zebrafish.chip file was chosen to translate zebrafish transcript names in the expression dataset file to HUGO (Human Genome Organisation) gene symbols. The HALLMARK gene sets file was chosen from the Molecular Signature Database (Broad Institute) for analysis. GSEA was then performed to detect enriched HALLMARK gene sets in the *dj-1*$^{-/-}$ samples. Gene set permutation was used to assess statistical significance of the enrichment scores as the number of samples in a phenotype was less than 7 (software.broadinstitute.org/gsea). Gene sets with a normalised enrichment score (NES) ≥ 1.5, FDR q-value ≤ 0.05 and nominal p-value ≤ 0.05 were considered significant.

2.8 Western Blot analysis

Zebrafish brains were dissected out and homogenised in 100µl of PhosphoSafe Extraction Reagent (Novagen) using a plastic homogenizer before centrifuging at 13,000rpm for 15 minutes and transferring 70µl of supernatant over to a new centrifuge tube. 70µl of 2X sample buffer was added to each sample before heating to 95°C for 2 minutes to denature the proteins. 25µl of each sample was loaded onto a 12% acrylamide/SDS-PAGE resolving gel (20ml 12% resolving gel: 8ml 30% Acrylamide, 5ml 1.5M Tris-HCl pH8.8, 6.6ml H_2O, 200µl 10% SDS, 200µl 10% APS, 50µl TEMED) with a 4% acrylamide/SDS-PAGE stacking gel (10ml 4% stacking gel: 6.8ml H_2O, 1.7ml 30% Acrylamide, 1.25ml 1M Tris-HCl pH6.8, 100µl 10% SDS, 100µl 10% APS, 25µl TEMED).

Samples were ran at 180V for 2 hours alongside 10µl PageRuler Prestained Protein Ladder (26616, ThermoFisher) in Tris-Glycine gel running buffer (3g/L Tris, 14.4g/L Glycine, 10ml/L 10% SDS). Gels were then soaked in Tris-Glycine transfer buffer (3g/l Tris, 14.4g/l Glycine, 10ml/l 10% SDS, 100ml/l methanol) for 10 minutes. During this PVDF membranes were soaked in 100% methanol for 15 seconds before washing in dH_2O for 2 minutes and finally soaking in the Tris-Glycine transfer buffer for 5 minutes.
The protein bands on the acrylamide gels were transferred onto PVDF membranes at 90V for 3 hours at room temperature. Membranes were then blocked with 5% milk (Marvel) in PBST (1X PBS, 0.1% Tween 20) for 1 hour. Membranes were incubated at 4°C overnight on a roller with the primary antibodies diluted in the same blocking solution (Table 2.6). The following day membranes were washed six times in PBST for 15 minutes. Membranes where then blocked again for 30 minutes before incubating for 1 hour at RT in secondary antibodies diluted in blocking solution. Membranes were washed a further six times in PBST for 15 minutes. BM Chemiluminescence (Roche) substrate was used to develop the membranes before exposing them to Hyperfilm ECL (Amersham) until expression was visualised.

Membranes were stripped following the Bio-Rad Blot Stripping and Reprobing protocol (biorad.com) based on Legocki and Verma (1981). Membranes were incubated in acidic glycine stripping buffer (0.1M glycine, 20mM magnesium

acetate, 50mM KCl, pH2.2) for 10 minutes at room temperature with agitation. The incubation was repeated a second time before washing the membranes in PBS three times for 5 minutes. The membranes were blocked again for 1 hour in 5% milk (Marvel)/PBST before incubating with primary antibody dilutions overnight and continuing the Western Blot protocol.

Target	Primary antibody dilution	Secondary antibody dilution
DJ-1	1:50,000 (PA5-72638, Invitrogen)	1:2000 anti-rabbit HRP (#7074, Cell Signaling Technology)
GAPDH	1:50,000 (G8795, Sigma-Aldrich)	1:4000 anti-mouse HRP (62-650, Invitrogen)
TH	1:50,000 (22941, Immunostar)	1:4000 anti-mouse HRP (62-650, Invitrogen)
P-AKT	1:5000 (9271S, Cell Signaling Technology)	1:2000 anti-rabbit HRP (#7074, Cell Signaling Technology)
ATP5F1B	1:5000 (HPA001528, Atlas Antibodies)	1:2000 anti-rabbit HRP (#7074, Cell Signaling Technology)

Table 2.6 Antibodies used for protein detection in Western blot

2.9 Immunofluorescence

2.9.1 Vibratome sectioning of zebrafish brains

Adult zebrafish were euthanised and whole brains were extracted by dissection following the jove protocol (Gupta and Mullins, 2010). Brains were then fixed overnight at 4°C in a 4% PFA / 1X PBS (w/v) solution. Following fixation brains were washed twice in 1X PBS for ten minutes. A 3% agarose / 1X PBS solution (w/v) was then prepared and cooled in a 60°C water bath before use. The agarose solution was used to embed each brain in a separate mold. Once the agarose had set each agarose block was removed from its mold and trimmed down using a razor blade to leave a small excess of agarose around each brain. The brains in agarose were then superglued to specimen disks of the

vibratome, orientated so the most caudal region of the brain was facing the specimen disk. The specimen disk was inserted into the buffer tray of the vibratome which was then filled with 1X PBS solution. Each brain was sectioned transversely at a thickness of 100μM. Floating sections were moved from the buffer tray to 6-well plates using a Pasteur pipette.

2.9.2 Immunofluorescence for TH positive cells

The PBS in each well containing brain sections was removed and replaced with a 10 mM sodium citrate buffer (pH 8.5). Sections were then incubated in the sodium citrate buffer for 2 hours at 80°C. Following incubation sections were washed twice in 1X PBS with 1% TritonX-100 for 15 minutes. Sections were then blocked in 2% BSA / 1X PBS with 1% TritonX-100 (w/v) for 30 minutes. After blocking, sections were incubated overnight at 4°C with a 1:500 dilution of mouse anti-TH (#22941, Immunostar) in 2% BSA / 1X PBS with 1% TritonX-100 (w/v). The following day sections were washed four times in 1X PBS with 1% TritonX-100 for 30 minutes. Sections were then incubated overnight at 4°C with a 1:500 dilution of goat anti-mouse IgG1 conjugated to Alexa Fluor 488 (A21121, Invitrogen) in 2% BSA / 1X PBS with 1% TritonX-100 (w/v). The following day sections were washed four times in 1X PBS with 1% TritonX-100 for 30 minutes. Staining for nuclei was carried out by incubating the sections in 1μg/ml Hoecshst 33342 in PBS with 1% TritonX-100 for 10 minutes. Sections were washed four times in PBS with 1% TritonX-100 for 30 minutes. Sections were transferred to microscope slides with a 15mm cavity before mounting with Hydromount (National Diagnostics) and a coverslip. Slides were then analysed using a Zeiss LSM 710 confocal microscope.

3 Generating zebrafish models of PD

3.1 Introduction

3.1.1 Genome editing to model PD in the zebrafish

The overall aims of modelling human disease in animals are to (i) gain a better understanding of the molecular basis of the condition by creating an experimental model, and (ii) in the case of zebrafish in particular, to contribute to the drug design pathway to develop new therapies. In this thesis, the aim was to evolve a classifier to distinguish the PD model as having a movement disorder when compared to healthy control zebrafish. A future study could then attempt to rescue the movement phenotype, with the possibility of a chemical screen to identify new candidates for treating the PD pathology. In order to do this, the model generated needs to be heritable. Here, genome editing was used to create both permanent and heritable models of PD in zebrafish.

Genome editing methods in zebrafish

There are three well-known methods of genome editing used in zebrafish: zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENS) and the relatively new CRISPR/Cas9 (Doyon et al., 2008; Cade et al., 2012; Hwang et al., 2013). ZFNs and TALENs are both DNA-binding proteins, containing a Fok I nuclease domain, used to create a targeted double-strand break in the DNA (Chandrasegaran et al., 1996; Joung and Sander, 2013). The DNA-binding domain of a ZFN recognises a unique 9-18bp sequence whilst a TALEN recognises a sequence of 32-35bp giving increased specificity (Shimizu et al., 2011; Gaj et al., 2013). Additionally, each ZFN or TALEN only cuts through a single strand of DNA so a pair of these reagents is required to recognise and cleave adjacent sense and antisense sequences (Bitinaite et al., 1998; Smith, 2000). Working as a pair increases the specificity of ZFNs or TALENs reducing the chances of off target effects. However, both methods require protein engineering to create the unique DNA-binding domains, which is complex and costly, whereas CRISPR/Cas9 targeting is simple and much more cost-effective (Gupta and Musunuru, 2014). For each gene targeted using CRISPR/Cas9 a single RNA molecule containing a unique sequence is required while the protein cleaving the DNA, Cas9, is always the

same and generates a double-strand break (Chang et al., 2013). The guide RNA (gRNA) contains a 19-23bp target sequence identical to a region in the target gene directly upstream of a 3bp protospacer adjacent motif (PAM site). The gRNA guides the Cas9 protein to the target sequence which is cleaved upstream of the PAM site (Figure 3.1). Compared to ZFNs and TALENs, generating gRNA is simple, efficient and economical (Chang et al., 2013). However, the drawback of CRISPR/Cas9 is the specificity. The target sequence is only ~20bp and only one gRNA with Cas9 is needed to produce a double-strand break rather than a pair, increasing the chances of off target effects (Gupta and Musunuru, 2014). Nevertheless, CRISPR/Cas9 has proven to be an effective method to specifically edit genomes in many species (Gratz et al., 2013; Nakayama et al., 2013; Wang et al., 2013; Woong Y Hwang et al., 2013) and was the method used here to create five mutant lines of zebrafish for modelling PD. Several generations of out-crossings are required to generate each mutant line; this means that the probability of any off-target mutations being carried along with the selected mutation in the homozygous mutants is low. Therefore, the CRISPR/Cas9 advantages of simplicity and low cost outweigh any drawbacks compared to TALENs and ZFNs, especially when producing multiple mutant lines.

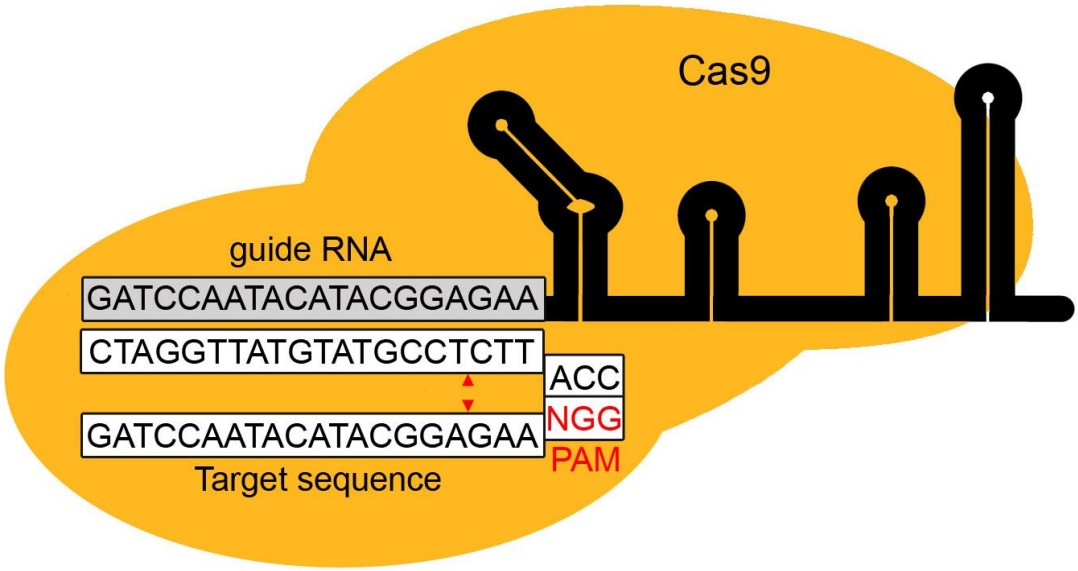


Figure 3.1 Diagram of CRISPR/Cas9 gene targeting

The Cas9/guide RNA complex base pairs with the complementary target sequence in genomic DNA. Cleavage (indicated by red arrows) occurs upstream of the PAM site (NGG).

3.1.2 PD-associated genes targeted for knockout

The genes targeted for mutation were orthologous to human genes considered either causative for autosomal recessive PD or a major PD risk factor. Target genes included *parkin, pink1, dj-1, fbxo7* and *gba. parkin* was chosen because mutations in human *PARKIN* are the most common cause of autosomal recessive PD, accounting for ~50% of cases (Lücking et al., 2000). *PARKIN* mutations were also seen in 77% of sporadic PD cases with onset before the age of 20 (Lücking et al., 2000). *PINK1* is the second most commonly mutated gene to cause autosomal recessive PD accounting for 1-7% of cases (Lohmann and Klein, 2008). Mutations in *DJ-1* are the third most common cause of autosomal recessive PD found in approximately 1-2% of cases (Klein and Westenberger, 2012). An additional reason for selecting DJ-1 was its involvement in the oxidative stress response (Kahle, Waak and Gasser, 2009). Oxidative stress is highly associated with the death of DA neurons in PD patients (Blesa et al., 2015). DJ-1 is a major transcriptional regulator of antioxidant genes and not all of its downstream targets have been fully

investigated (Clements et al., 2006; Wang et al., 2011). Therefore, a DJ-1 mutant line was of high interest to study the molecular mechanisms contributing to PD. Out of the genes targeted *FBXO7* is the rarest cause of autosomal recessive PD, with mutations seen in 18 patients from 5 families (Zhou et al., 2015). FBXO7 co-operates with PINK1 and PARKIN in the breakdown of damaged mitochondria (Burchell et al., 2013). Therefore, it would be of interest to see whether zebrafish with mutations in *PINK1*, *PARKIN* or *FBXO7* are phenotypically similar. The final gene targeted in zebrafish was *gba*. *GBA* mutations are the largest risk factor for developing PD and cases associated with heterozygous mutation of GBA are clinically very similar to idiopathic PD (Migdalska-Richards and Schapira, 2016). However, homozygous mutation of *GBA* causes the autosomal recessive lysosomal storage disorder Gaucher's disease in humans (Hruska et al., 2008). Homozygous and heterozygous *gba* mutant zebrafish were studied as both have an increased risk of developing PD in humans (Lwin et al., 2004).

3.1.3 Aims of this chapter

Previous models of PD in zebrafish have used transient methods including neurotoxins to ablate DA neurons and knockdown methods using antisense MOs (Table 1.5 and Table 1.6). Neurotoxins have generally caused a decrease in dopamine and locomotion in the zebrafish but with a fast recovery and the rapid loss goes against the progressive loss seen in PD (Sallinen et al., 2009). These models are also poor for studying the molecular mechanisms that underlie PD pathogenesis. Knockdown of PD-associated genes in zebrafish is better for studying the pathways involved but again the models are transient (Bretaud et al., 2007). In order to generate permanent heritable models of PD in zebrafish genome editing was used. CRISPR/Cas9 was chosen as the method to generate mutant lines.

The aims of this chapter are:

- To identify and characterize the temporal/spatial expression of a set of PD-associated genes in zebrafish (*parkin, pink1, dj-1, fbxo7 and gba).*

- To disrupt each PD-associated gene in zebrafish using CRISPR/Cas9 and sequence the mutation generated.
- To evaluate the effect of each mutation on the translated protein.
- To determine whether there is nonsense-mediated decay of the target gene transcripts.
- To analyse the effect each mutation has on the DA neurons in the zebrafish brain.

3.2 Results

3.2.1 Orthologs of the PD-associated genes in zebrafish

A BLAST search was used to identify the orthologs in zebrafish for each of the PD-associated genes being targeted, and to compare the amino acid sequences of the encoded proteins. A single *PARKIN* orthologue, containing 12 exons, was identified in zebrafish on chromosome 13 (ENSDARG00000021555). The 458 amino acid protein encoded by *parkin* in zebrafish had 62% identity with human PARKIN. *PINK1* has a single orthologue, containing 8 exons, in zebrafish on chromosome 23 (ENSDARG00000001929). Zebrafish Pink1 was 574 amino acids long and shared 62% identity with human PINK1. A single *DJ-1* orthologue, containing 6 exons, was identified in zebrafish on chromosome 11 (ENSDARG00000116835). The 189 amino acid protein encoded by *dj-1* in zebrafish had 83% identity with human DJ-1. *FBXO7* has a single orthologue, containing 10 exons, in zebrafish on chromosome 4 (ENSDARG00000099833). Zebrafish Fbxo7 was 486 amino acids long and shared 38% identity with human FBXO7. A single *GBA* orthologue, containing 10 exons, was identified in zebrafish on chromosome 16 (ENSDARG00000076058). The 518 amino acid protein encoded by *gba* in zebrafish had 57% identity with human GBA.

3.2.2 Synthesising *in situ* probes for the target genes

To characterise the expression profiles of each of the PD-associated genes being targeted in zebrafish *in situ* hybridisation (ISH) and RT-PCR assays were carried out (see Methods in Chapter 2 for full detail). Primers were designed in the coding sequence of each gene, with each primer residing in a separate exon (Figure 3.2A). The primer pair for each gene was then used for cloning a region of the coding sequence for *in situ* probe synthesis and for RT-PCR analysis. PCR products amplified from the cDNA of wild type zebrafish (Figure 3.2B) were T-cloned into the pGEM T-Easy vector system (Figure 3.2C) before sequencing to determine the orientation of each insert. DIG-labelled antisense RNA probes were then synthesized by transcription.

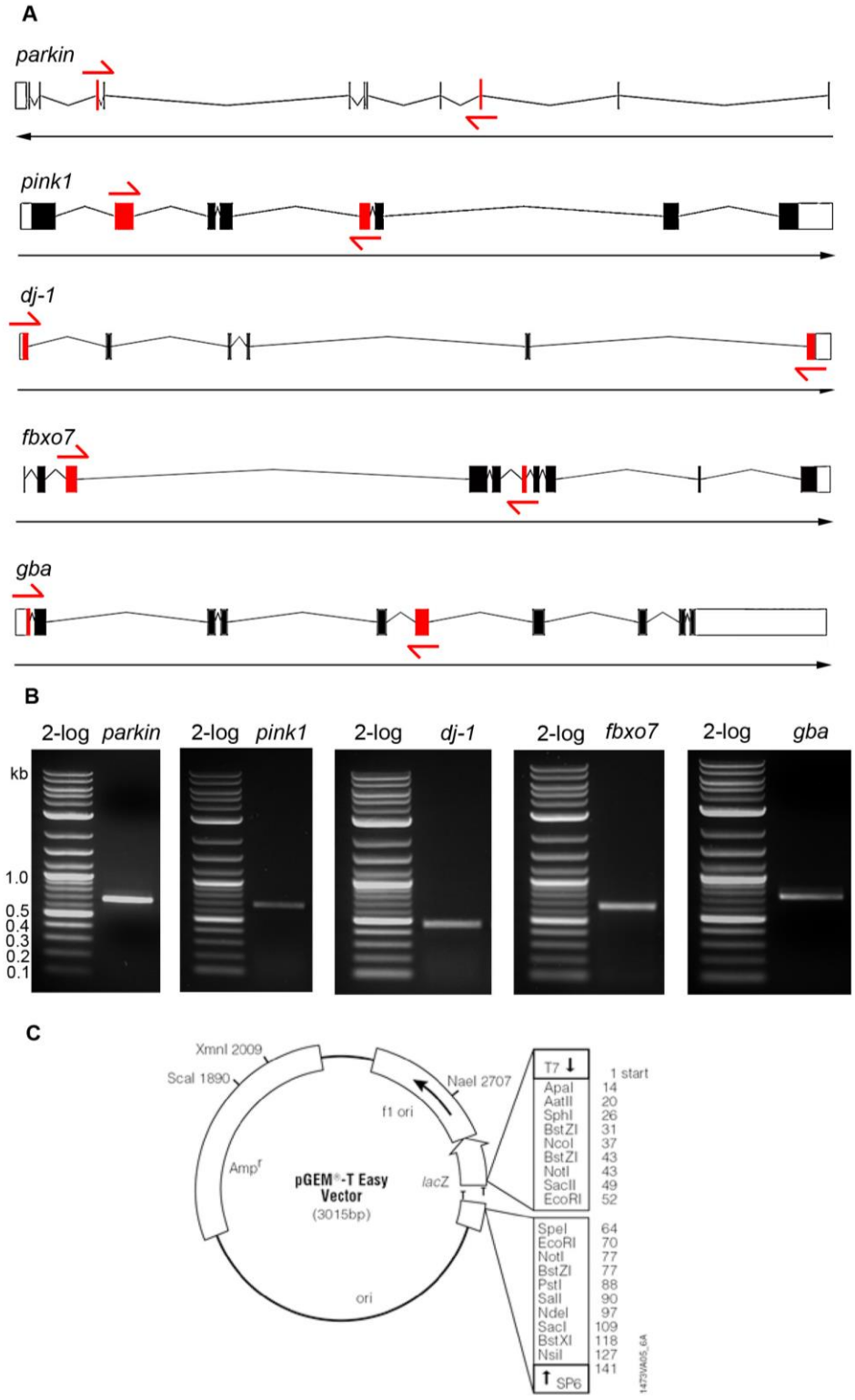


Figure 3.2 Generating *in situ* probe templates

(A) The location of primers used to amplify a region of the coding sequence for *parkin*, *pink1, dj-1, fbxo7* and *gba* (adapted from ensembl.org) (B) PCR products amplified from the coding sequence of each PD-associated gene. (C) The plasmid map for the pGEM T-Easy vector system (Promega) PCR products amplified from each PD-associated gene were T-cloned into the plasmid before linearising and transcribing each ISH probe.

3.2.3 Expression profiles of PD-associated target genes

To investigate the spatial expression of the target genes in zebrafish *in situ* hybridisation (ISH) was carried out on wild type embryos at 4 dpf (see Methods in Chapter 2 for more details). Digoxigenin-labeled RNA probes hybridised to the transcripts of interest and a colorimetric reaction produced staining in the areas of expression. In addition to the analyses of *parkin, pink1, dj-1, fbxo7* and *gba*, the expression of markers for DA neurons (*dat*), catecholaminergic neurons (*th*), general neurons (*syn2a*) and muscle (*smyhc*) were also studied. *dat* expression was highly specific to the olfactory bulb (OB), pretectum (P) and diencephalic dopaminergic clusters (DC) in the brain (Figure 3.3A). *parkin, pink1, dj-1* and *gba* were all found to be widely expressed in the zebrafish embryo including a ring of expression around the retina (R) (Figure 3.3C-E,G). *fbxo7* had low levels of ubiquitous expression (Figure 3.3F). *parkin, dj-1* and *gba* all showed heightened expression in the intestines (I) (Figure 3.3C,E,G). Additionally, *dj-1* had greater expression in the brain and liver (Figure 3.3E), whereas *gba* was expressed higher in the brain as well as the caudal haematopoietic region (CHR) (Figure 3.3G), a region of haematopoeisis (Murayama et al., 2006). *syn2a* expression was highly specific to the brain, including a ring around the retina (Figure 3.3H), and *smyhc* was highly expressed specifically in muscle tissue (Figure 3.3I). The results of these analyses agree with the results of previous expression analyses in the zebrafish. *dat* and *th* expression matching the expression patterns shown in Figure 3.3A and 3.3B has been seen in multiple reports (Decker et al., 2014; Xia et al., 2017; Chen et al., 2012; Flinn et al., 2013, 2009; Zhao et al., 2012; Bretaud et al., 2007). Thisse and Thisse (2004) saw a basal level of expression throughout the zebrafish embryo for *parkin*, *pink1*, *dj-1 and fbxo7* with heightened expression in the brain and around the eye for *dj-1.* Bai et al. (2006) saw heightened expression of *dj-1* in the intestines of zebrafish embryos. Keatinge et al. (2015) reported heightened expression of *gba* in the brain and liver of the zebrafish embryo. A loss of *gba* expression has also been reported to increase the number of haematopoietic precursor cells in the zebrafish embryo, which are found in the CHR (Zancan et al., 2015). This suggests that *gba* is expressed in the CHR to have an effect on haematopoietic precursors. High expression of *syn2a* in the brain of the zebrafish embryo was reported by

Garbarino et al. (2014) and Thisse and Thisse (2004) saw *smyhc* expression in the tail of the zebrafish and in the extraocular muscles. Therefore, the staining seen for each of the genes in Figure 3.3 appeared successful. *In situ* hybridisation for each of the genes was only carried out once on 4dpf embryos, however, staining on 3dpf and 5dpf embryos had similar patterns of expression.

To study any spatial restriction of gene expression in adult zebrafish RT-PCR was used. RNA samples extracted from the eye, brain, liver, muscle and tail of wild type adults were used. Similar to the results of the ISH all of the target genes were expressed in all tissues examined, whilst *dat* and *syn2a* expression was restricted to the brain with faint expression in the eye (Figure 3.4A). *smyhc* was expressed at high levels in the muscle and tail tissue with a lower level of expression in the eye, likely from the extraocular muscles (Kasprick et al., 2011). To study the temporal expression during zebrafish development RT-PCR was carried out using RNA extracted from zebrafish 1 to 6 dpf. Each target gene assessed was expressed from 1 dpf whilst *dat* expression was first detected at 2dpf (Figure 3.4B).

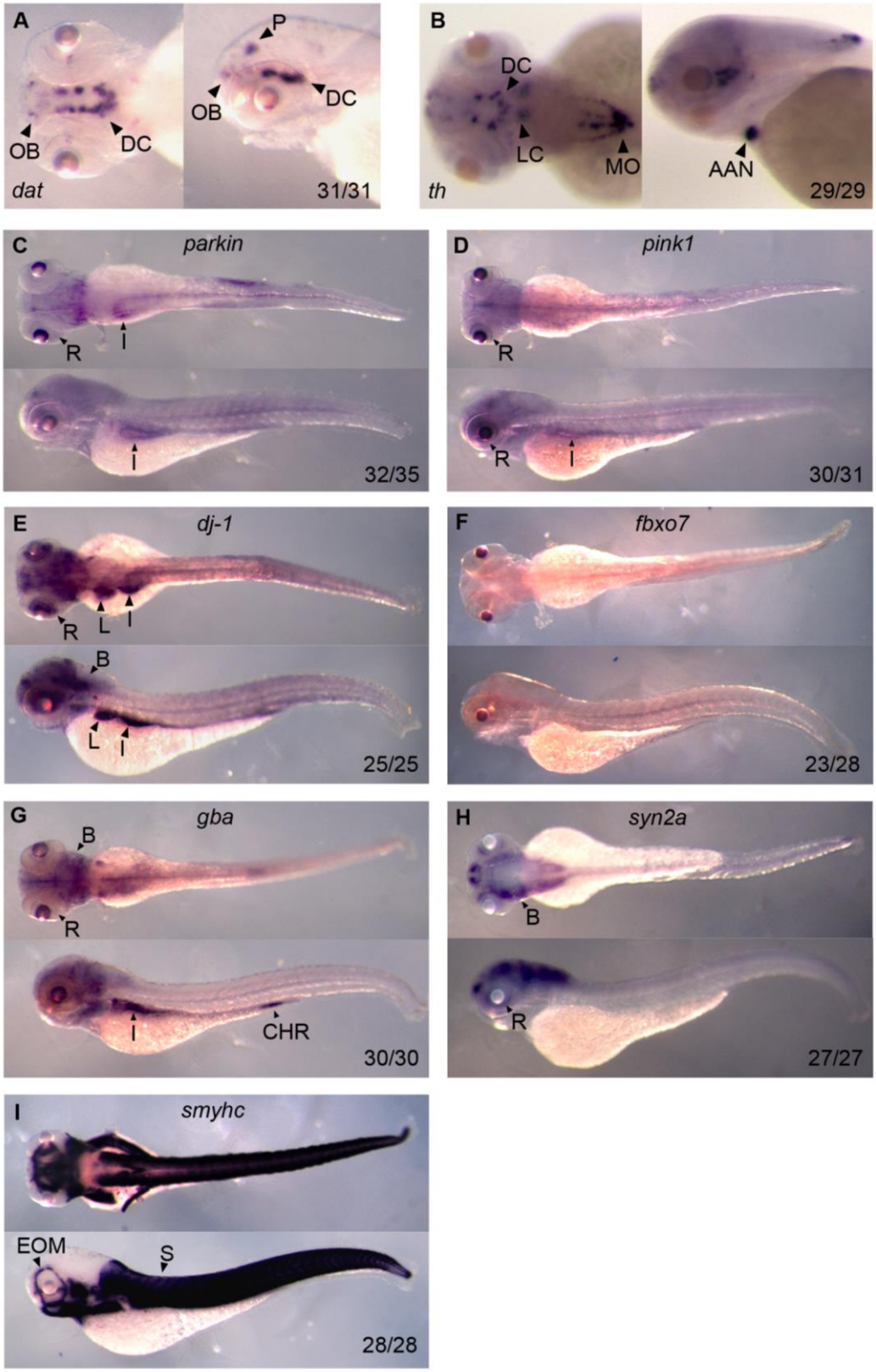


Figure 3.3 Spatial expression of PD-associated genes in zebrafish

Whole mount *in situ* hybridisations on zebrafish at 4dpf for *dat* (A), *th* (B) *parkin* (C), *pink1* (D), *dj-1* (E), *fbxo7* (F), *gba* (G), *syn2a* (H) and *smyhc* (I). Upper image is the dorsal view and lower image is the lateral view for all figures. Olfactory bulb (OB), dopaminergic diencephalic clusters (DC), pretectum (P), retina (R), intestines (I), liver (L), brain (B) and caudal haematopoietic region (CHR) are indicated. The fraction of embryos with the staining shown is indicated in the bottom right corner of each image.

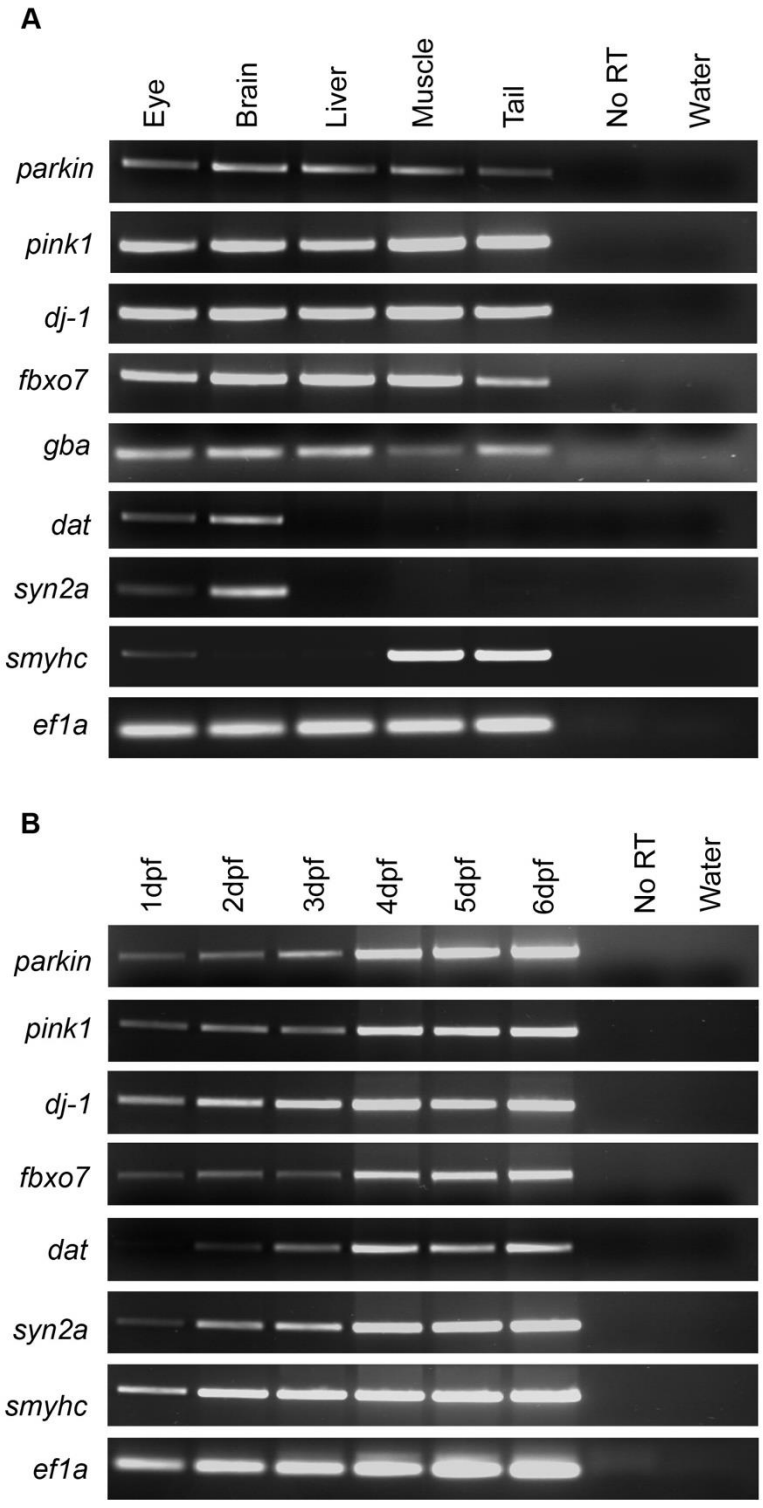


Figure 3.4 Spatial/temporal expression of the target genes

(A) RT-PCR analysis comparing expression across tissues of PD-associated target genes, DA neuron marker (*dat*), neuronal marker (*syn2a*) and muscle marker (*smyhc*). (B) RT-PCR analysis studying temporal expression of *parkin, pink1, dj-1, fbxo7, dat*, *syn2a* and *smyhc. ef1a* was used as a loading control for both RT-PCR analyses.

3.2.4 Synthesising guide RNA for CRISPR/Cas9 targeting

A 20bp target sequence was identified in each target gene and incorporated into the gene specific forward primer for amplification of the gRNA template (see Methods in Chapter 2 for full detail). Target sequences were selected in exons close to the 5' end of the gene to increase the probability of an early stop codon near the 5' end and truncation of the protein with loss of the important functional domains. Sequences with a GC content between 40 and 70% were chosen based on previous CRISPR/Cas9 studies that found gRNAs with these GC levels to be more effective (Wang et al., 2014; Tsai et al., 2015). To reduce the chance of off-target effects only sequences lacking off-target sites (with 1-2 nucleotide mismatches) were chosen. Furthermore, the position of the target sequence in the gene was important for PCR amplification and sequencing of the mutations generated. Space was allowed upstream and downstream of the CRISPR target in the gene to amplify products of 400-800bp in length. PCR primers started at least 80bp from the target sequence to avoid the mutations generated occurring in the low quality base call region when sequencing in plasmid (Yao et al., 2006). These PCR primer requirements meant that target sequences in the second or third exon ended up being selected rather than sequences in the first exon. In an attempt to remove a large section of the gene in *pink1* a second target sequence was identified approximately 100bp downstream from the first and used to generate a gRNA template. The location of each target sequence in the relevant PD-associated gene is indicated in (Figure 3.5) and the primers used for generating the gRNA templates are listed in (Table 2.1). gRNAs were transcribed from each template using T7 RNA polymerase and cleaned up by phenol-chloroform extraction and ethanol precipitation. gRNAs were then injected alongside Cas9 protein into one cell stage embryos.

3.2.5 Validating CRISPR/Cas9 mutation of target genes in the F0

To confirm that CRISPR/Cas9 was effective at generating mutations in the target genes of the founder generation (F0), DNA sequencing was carried out on the genomic DNA of multiple F0 individuals. However, CRISPR/Cas9 is known to produce mosaic mutants where members of the F0 can bear multiple

populations of cells with different mutant alleles of the target gene in each population (Mehravar et al., 2019). Therefore, 5 embryos from each F0 generation were lysed to extract their gDNA, PCR amplified products of the target region were ligated into the pGEM T-easy plasmid and used to transform competent *E.coli* (see Methods in Chapter 2 for full detail). The primers used for each PCR amplification are listed in (Table 2.2) and the location in each target gene is displayed in (Figure 3.5). PCR primers were selected using the online tool Primer-Blast (www.ncbi.nlm.nih.gov/tools/primer-blast/) to produce amplicons of 400-800bp containing the CRISPR target with each primer starting at least 80bp from the target sequence. The maximum range in melting temperature for a primer pair was 6°C (for *fbxo7*) and the remaining primer pairs had a maximum range of 2°C ensuring simultaneous binding and efficient PCR amplification. The GC content for the majority of primers ranged from 35-50% (the *fbxo7* forward primer had a 20% GC content), a higher GC content increasing the strength of annealing (Kamel, 2003). Multiple colonies from each ligation were picked to investigate the mosaic of sequences present in each of the 5 founder fish analysed. Each colony could feasibly represent a different mutant allele from the same gRNA/CRISPR injected individual. Plasmids were isolated from each liquid culture and sequenced.

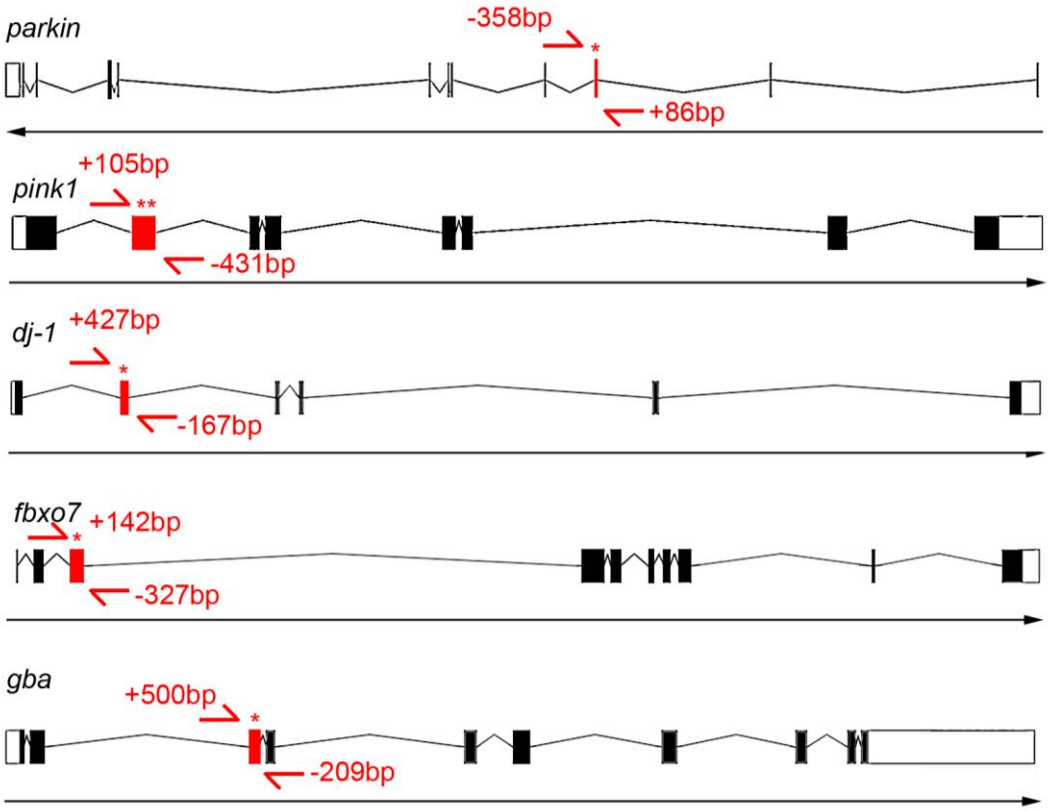


Figure 3.5 Location of CRISPR target sequences and PCR primers

The location of CRISPR target sequences in the exons of *parkin*, *pink1, dj-1, fbxo7* and *gba* (indicated by a *). Primers for the amplification of each CRISPR target region for sequencing purposes are indicated by red arrows (adapted from ensembl.org). The start of each PCR primer in relation to the target sequence is given in bp with + indicating upstream in the gene and – indicating downstream in the gene.

Target Gene		F0 embryo				
		1	2	3	4	5
parkin		2/3 (2)	1/1	3/3 (3)	2/3 (2)	2/2 (2)
pink1	gRNA 1	1/1	4/4 (3)	4/4 (3)	0/3	0/2
	gRNA 2	1/1	3/4 (2)	4/4 (2)	0/3	0/2
dj-1		2/2 (2)	1/1	1/1	2/2 (2)	1/3
fbxo7		4/4 (3)	3/3 (3)	4/4 (3)	3/3 (3)	1/1
gba		0/1	0/1	2/2 (2)	2/2 (2)	1/1

Table 3.1 F0 generation rate of mutation

The rate of mutation in the F0 generation for each target gene and gRNA. 5 F0 embryos were lysed to extract their genomic DNA, the target region was PCR amplified and ligated into plasmid before transforming competent bacteria with the ligations. Four colonies were selected and sequencing on their four corresponding plasmids was attempted. The rate of mutation is displayed as the number of mutant alleles identified out of successfully sequenced plasmids. The number of mutant allele variants is displayed in brackets when at least two sequenced plasmids carried a mutation. This acts as a measure of mosaicism in the F0 generation.

3.2.6 Identifying frameshift mutations in the F1

Individuals from the F0 generation were outcrossed with wild type zebrafish to produce a heterozygous F1 generation for each mutant line. For this to be successful the F0 individual in the outcross had to be carrying a mutant allele of the target gene in its germ cells (Hwang et al., 2013b). Because fertilisation involves one sperm with a single haploid genome, the problem with mosaicism stops at this point and if germ line transmission of a mutation is successful, then each F1 individual will be heterozygous for that allele. To identify heterozygous mutations in the F1 that caused a frameshift and an early stop codon, both copies of the target gene had to be sequenced in F1 individuals. Each F1 generation was raised to 3 mpf and fin clipped. Clippings were lysed to extract the genomic DNA for PCR of the target genes. A heteroduplex analysis was then carried out on the PCR products to identify any carriers of a mutated target gene (see Methods in Chapter 2). The amplified target genes of any confirmed heterozygous mutants were ligated into plasmid and used to transform competent *E.coli*. Multiple colonies were picked from each transformation; 50%

of the colonies are expected to carry the mutation. Plasmids were retrieved by miniprep and sequenced.

Target Gene	Heterozygous mutants in the F1
parkin	7/9
pink1	9/14
dj-1	8/13
fbxo7	3/6
gba	5/12

Table 3.2 F1 generation mutation transmission

The number of fish in the F1 generation identified with a heterozygous mutation in the target gene out of those successfully sequenced.

Parkin

In the *parkin* F1 generation seven heterozygous mutants were identified by heteroduplex analysis and sequenced, six carried a 5bp frameshift in *parkin* and one fish had a 1bp insert. The 5bp frameshift in exon 2 at position 224 of NM_00101763 was caused by a 4bp deletion followed by a 9bp insertion (Figure 3.6A). This resulted in a premature stop codon 85bp downstream at position 309. The premature stop codon would cause truncation of the Parkin protein to 96aa and loss of the RING1, IBR and RING2 domains (Figure 3.6B) (based on structure by Fett et al., 2010). The mutant parkin sequence was found to gain a BsaJI restriction site allowing genotyping of *parkin* mutants by restriction digest assay. PCR amplification of *parkin* using the primers in (Table 2.2) creates a ~500bp product. Digestion of the mutant amplicon with BsaJI generates two products of ~400bp and ~100bp, whereas the wild type amplicon contains no BsaJI restriction site. Gel electrophoresis of the *parkin* digestion products was used to genotype (Figure 3.6C). qRT-PCR analysis was carried out on RNA extracted from the brains of 3 *parkin*$^{-/-}$ mutants and 3 wild type siblings at 16 wpf to study the expression of *parkin* and markers for DA neurons (*dat*), catecholaminergic neurons (*th)*, DA neuron development (*pitx3)* and a general neuronal marker (*syn2a*) (Nunes et al., 2003; Rink and Wullimann, 2002; Holzschuh et al., 2001; Garbarino et al., 2014). No significant difference

was seen in *parkin*, *dat* or *th* levels. Whereas, a marked reduction of *pitx3* (~50%) and increase of *syn2a* (~40%) was seen in the *parkin*$^{-/-}$ mutants (Figure 3.6D). Therefore, no nonsense-mediated decay of the *parkin* transcript has occurred and according to *dat* and *th* expression there is no overall change in the number of DA neurons. The loss of *pitx3* suggests issues with the development and maintenance of DA neurons (Nunes et al., 2003; Li, Dani and Le, 2009) and the increase in *syn2a* expression could indicate increased synapse formation (Ferreira et al., 1995; Han et al., 1991). In conclusion, the mutation of *parkin* has had an effect on the maintenance and synaptic function of neurons, including the DA neurons, but there has been no loss of the DA neurons.

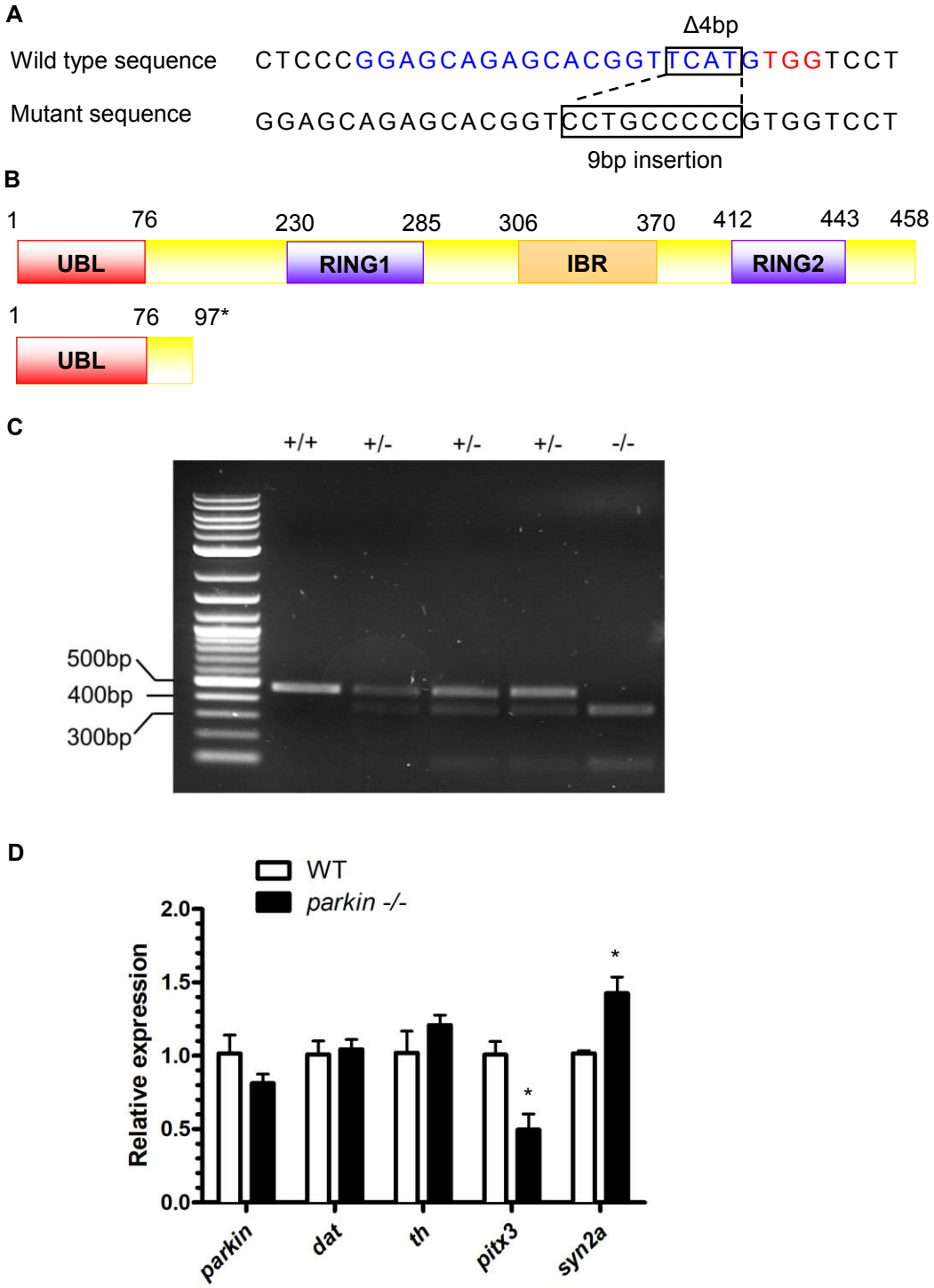


Figure 3.6 Mutation of *parkin* and the effect on DA neuron markers

(A) The *parkin* target sequence (blue) and PAM site (red) in the wild type and *parkin* mutant. A 4bp deletion followed by a 9bp insertion occurred in the *parkin* mutant. (B) Wild type Parkin (above) and truncated Parkin protein (below) predicted in the *parkin* mutant (C) Gel electrophoresis of *parkin* PCR products amplified from gDNA and digested with BsaJI. Genotypes of zebrafish are labelled above each lane. (D) qRT-PCR analysis comparing *parkin, dat, th*, *pitx3* and *syn2a* expression in *parkin-/-* (n=3) and wild type sibling (n=3) brains at 16 wpf. Error bars represent SEM, * = P<0.05.

PINK1

In the *pink1* F1 generation nine heterozygous mutants were identified by heteroduplex analysis and sequenced. Seven carried a 6 bp deletion in *pink1* and two individuals carried a 101 bp deletion between the two target sequences. The 101bp deletion in exon 2 at position 591 of NM_001008628 caused a shift in the reading frame resulting in a premature stop codon 71bp downstream (Figure 3.7A). The premature stop codon would cause truncation of the translated Pink1 protein from 575 aa to 184 aa and loss of the majority of the kinase domain (Figure 3.7B) (based on Flinn et al., 2013). Genotyping of *pink1* mutants by PCR was simple due to the 101bp deletion in the mutant allele. Amplification of the wild type allele using the primers in (Table 2.2) produced a ~700bp product, whereas PCR of the mutant allele generated a ~600bp product. Gel electrophoresis of the *pink1* PCR products was used to genotype *pink1* mutants (Figure 3.7C) qRT-PCR analysis was carried out on RNA extracted from the brains of 4 *pink1*$^{-/-}$ mutants and 3 wild type siblings at 16 wpf to study the expression of *pink1, dat, th, pitx3* and *syn2a.* A 90% loss of *pink1* expression was seen in the *pink1*$^{-/-}$ mutant brain (Figure 3.7D). There was no significant change in expression of *dat, th*, *pitx3* or *syn2a.*

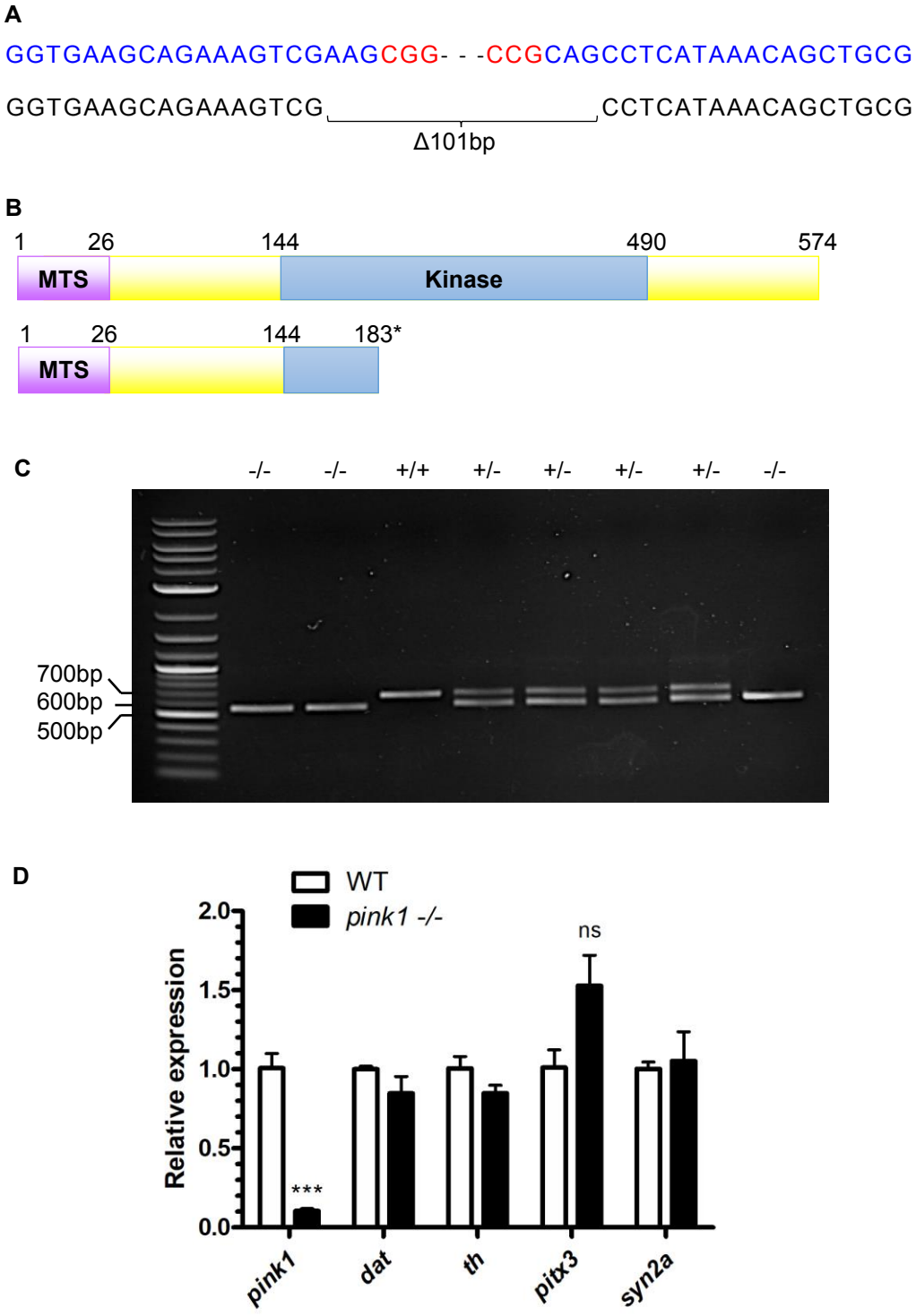


Figure 3.7 Mutation of *pink1* and the effect on DA neuron markers

(A) The *pink1* target sequence (blue) and PAM site (red) in the wild type (above) and *pink1* mutant (below). A 101bp deletion between the two target sequences caused a frameshift in the *pink1* mutant. (B) Wild type Pink1 (above) and truncated Pink1 protein (below) predicted in the *pink1* mutant (C) Gel electrophoresis of *pink1* PCR products amplified from gDNA. Genotypes of zebrafish are labelled above each lane. (D) qRT-PCR analysis comparing *pink1, dat, th*, *pitx3* and *syn2a* expression in *pink1-/-* (n=4) and wild type sibling (n=3) brains at 16 wpf. Error bars represent SEM, *** = P<0.001.

DJ-1

In the *dj-1* F1 generation eight heterozygous mutants were identified and sequenced, six carried a 3 bp deletion in *dj-1* and two carried a 17 bp frameshift. The 17 bp frameshift was the result of a 2bp deletion followed by a 19bp insertion in exon 2 at position 200 in NM_001005938 (Figure 3.8A). The shift in reading frame produced a premature stop codon at position 227. Dj-1 protein translated from the mutant transcript would be 56 aa in length missing the crucial C106 residue (Figure 3.8B) (based on Bai et al., 2006). The mutant *dj-1* sequence was found to lose a BbvI restriction site allowing genotyping of *dj-1* mutants by restriction digest assay. PCR amplification of *dj-1* using the primers in (Table 2.2) created a ~300bp product. Digestion of the wild type *dj-1* amplicon with BbvI created a ~200bp product and another fragment less than 100bp in length. The mutant *dj-1* amplicon lacks the BbvI restriction site so no digestion with BbvI occurs. Gel electrophoresis of the *dj-1* digestion products was used to genotype (Figure 3.8C). qRT-PCR analysis was carried out on RNA extracted from the brains of 5 *dj-1*$^{-/-}$ mutants and 3 wild type siblings at 16 wpf to study the expression of *dj-1, dat*, *th*, *pitx3* and *syn2a.* A 95% loss of *dj-1* was seen in the mutants alongside a ~40% reduction in *dat, th* and *pitx3* expression (Figure 3.8D). No significant loss of *syn2a* expression was observed. Western blot analysis was carried out to compare DJ-1 protein expression in the brains of 3 *dj-1*$^{-/-}$ mutants and 4 wild type siblings at 16 wpf using an antibody (PA5-72638) that recognizes amino acids 40-69 in the DJ-1 protein. The expression was analysed in the homozygous mutants because DJ-1 is associated with autosomal recessive PD. There was no DJ-1 protein detected in the *dj-1*$^{-/-}$ brain using the PA5-72638 antibody (Figure 3.8E).

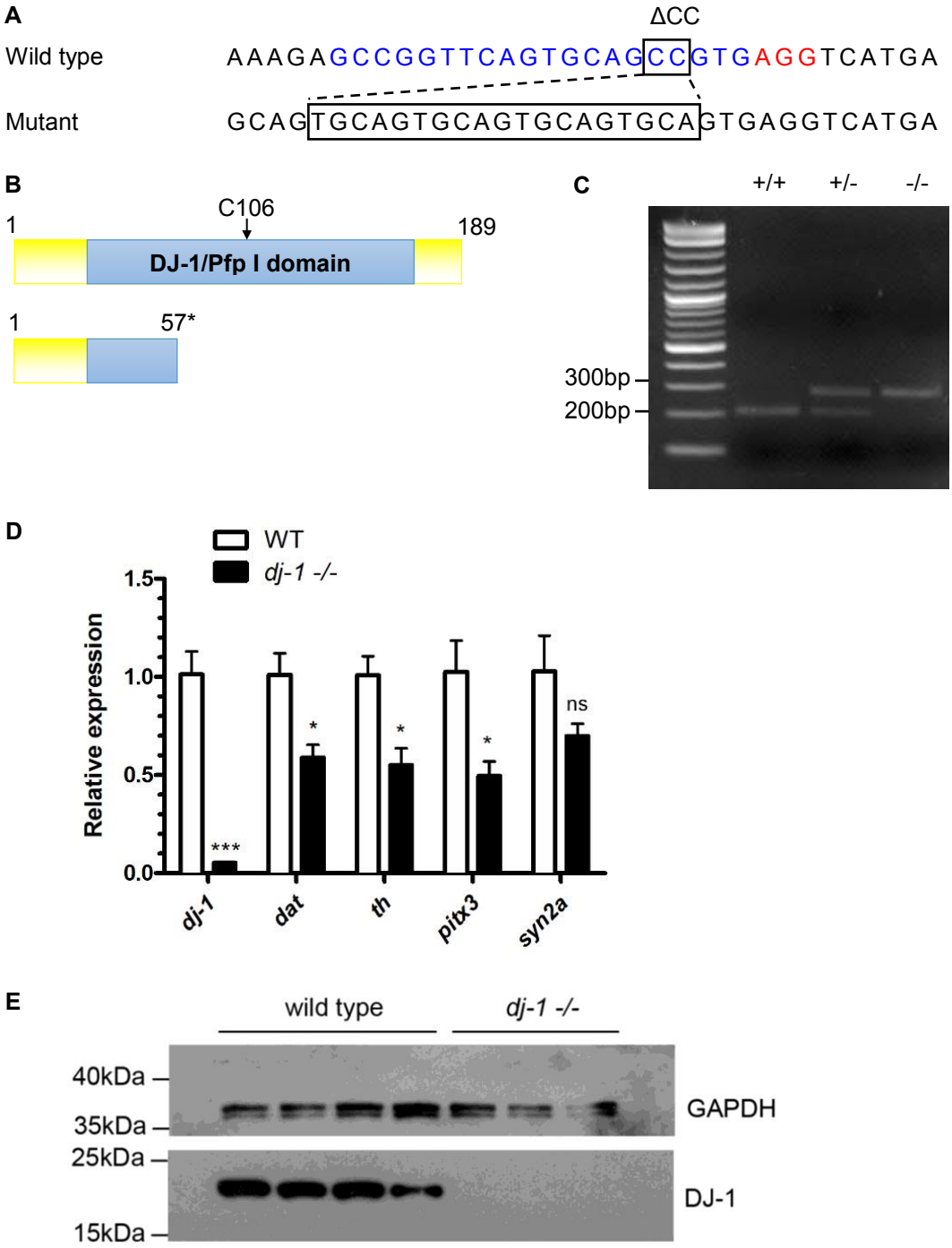


Figure 3.8 Mutation of *dj-1* and the effect on DA neuron markers

(A) The *dj-1* target sequence (blue) and PAM site (red) in the wild type and *dj-1* mutant. A 2bp deletion (ΔCC) followed by a 19bp insertion occurred in the *dj-1* mutant. (B) Wild type Dj-1 (above) and truncated Dj-1 protein (below) predicted in the *dj-1* mutant. (C) Gel electrophoresis of *dj-1* PCR products amplified from gDNA and digested with BbvI. Genotypes of zebrafish are labelled above each lane. (D) qRT-PCR analysis comparing *dj-1, dat, th*, *pitx3* and *syn2a* expression in *dj-1-/-* (n=5) and wild type sibling (n=3) brains at 16 wpf. Error bars represent SEM, * = P<0.05, *** = P<0.001. (E) Western blot analysis of DJ-1 protein expression in the brains of 3 *dj-1-/-* zebrafish and 4 wild type siblings at 16 wpf. GAPDH was used as a loading control.

FBXO7

In the *fbxo7* F1 generation three heterozygous mutants were identified and sequenced, all three were found to carry a 4 bp deletion in *fbxo7*. The 4bp deletion in exon 2 at position 190 of NM_001025499 caused a shift in the reading frame resulting in a premature stop codon 57bp downstream in the same exon (Figure 3.9A). The premature stop codon would cause truncation of the translated protein from 486 aa to 71 aa and loss of both the F-box domain and proline rich region (Figure 3.9B) (based on Zhao et al., 2012 and predicted F-box domain from prosite.expasy.org). qRT-PCR analysis was carried out on RNA extracted from the brains of 3 *fbxo7*$^{-/-}$ mutants and 3 wild type siblings at 16 wpf to study the expression of *fbxo7, dat*, *th*, *pitx3* and *syn2a.* No significant difference was seen in the expression of *fbxo7* or any neuronal markers (Figure 3.9C).

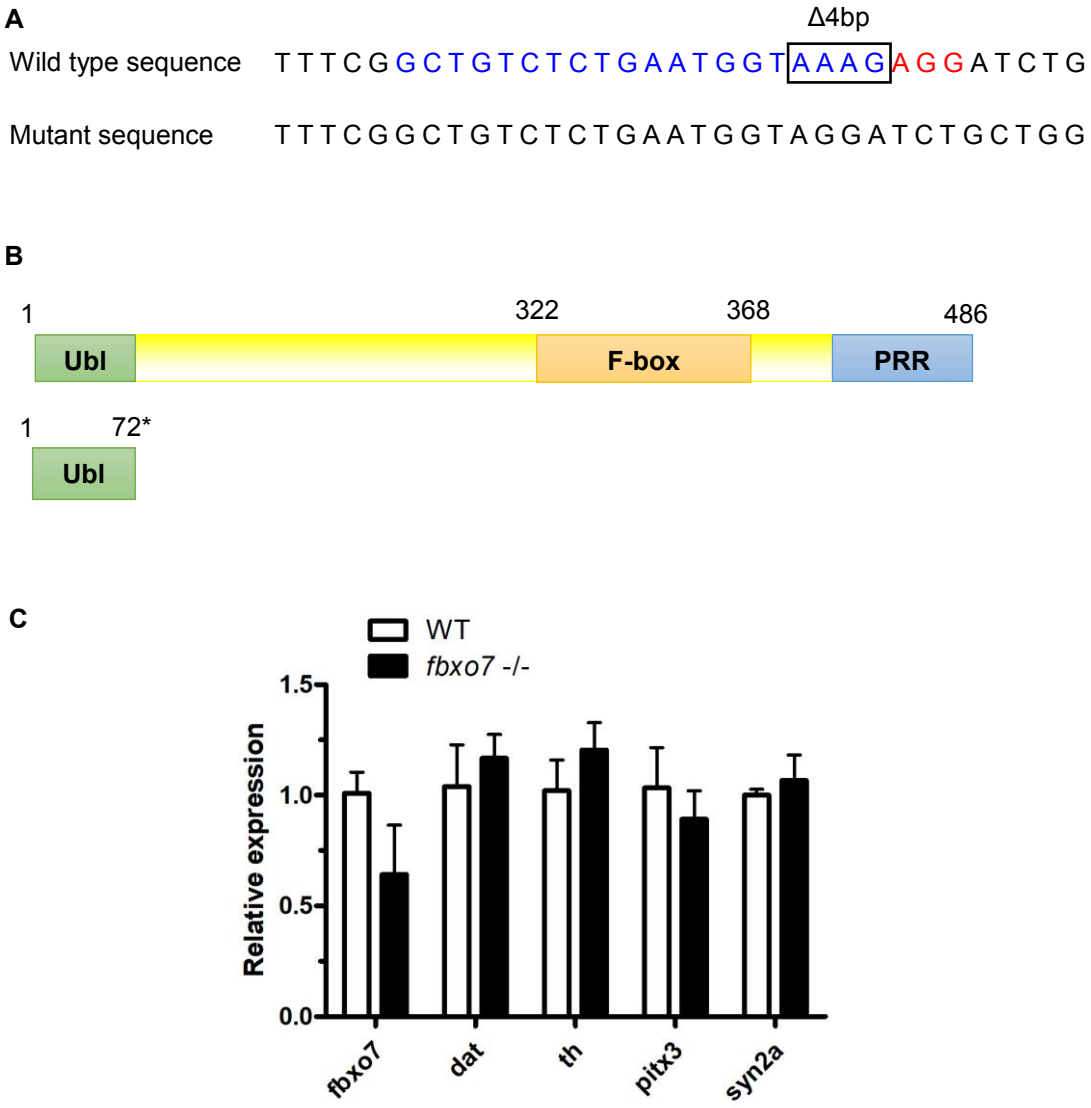


Figure 3.9 Mutation of *fbxo7* and effect on DA neuron markers

(A) The *fbxo7* target sequence (blue) and PAM site (red) in the wild type and *fbxo7* mutant. A 4bp deletion occurred in the *fbxo7* mutant. (B) Wild type Fbxo7 (above) and truncated Fbxo7 protein (below) predicted in the *fbxo7* mutant. (C) qRT-PCR analysis comparing *fbxo7, dat, th*, *pitx3* and *syn2a* expression in *fbxo7-/-* (n=3) and wild type sibling (n=3) brains at 16 wpf. Error bars represent SEM.

GBA

In the *gba* F1 generation five heterozygous mutants were identified and sequenced, four carried a 5 bp deletion in *gba* and a single individual carried an 8 bp deletion. The 5 bp deletion in exon 3 at position 615 of XM_682379 caused a shift in the reading frame resulting in a premature stop codon 76 bp downstream in exon 4 (Figure 3.10A). The premature stop codon would cause truncation of the translated Gba protein from 518 aa to 70 aa and loss of the majority of the predicted glycosidase domain (Figure 3.10B) (based on predicted glycosidase domain from uniprot.org). The mutant *gba* sequence was found to lose a SacI restriction site allowing genotyping of *gba* mutants by restriction digest assay. PCR amplification of *gba* using the primers in (Table 2.2) creates a ~700bp product. Digestion of the wild type *gba* amplicon with SacI creates two fragments of ~500bp and ~200bp. Whereas the mutant *gba* amplicon lacks the SacI restriction site so no digestion with SacI occurs. Gel electrophoresis of the *gba* digestion products was used to genotype (Figure 3.10C). qRT-PCR analysis was carried out on RNA extracted from the brains of 3 *gba* homozygous mutants, 3 heterozygous carriers and 3 wild type siblings at 16 wpf to study the expression of *gba, dat*, *th, pitx3* and *syn2a* (Figure 3.10D). Homozygous and heterozygous mutants were studied as both have an increased risk of developing PD in humans (Lwin et al., 2004). A significant loss of all the genes analysed apart from *pitx3* was seen in the homozygous mutant. No significant loss was seen for any of the genes analysed in the heterozygous *gba* mutant.

Further molecular analysis of the DA neurons was carried out on the *parkin, pink1* and *dj-1* mutant lines, detailed in the rest of this chapter. The *parkin* line was chosen because of its altered expression of *pitx3*, involved in the development and maintenance of DA neurons, and *syn2a*, a general neuronal marker. The *pink1* line was studied further because of the loss of *pink1* expression indicating nonsense-mediated decay of the transcript and the previously reported loss of DA neurons in *pink1* deficient embryos (Flinn et al., 2013). The *dj-1* line was studied because of the loss of *dj-1* expression and the loss of multiple DA neuron markers. Analysis of the homozygous *gba* mutant was discontinued because of the severe movement phenotype observed

matching the description of a previous model of Gauchers disease in the zebrafish (Keatinge et al., 2015)

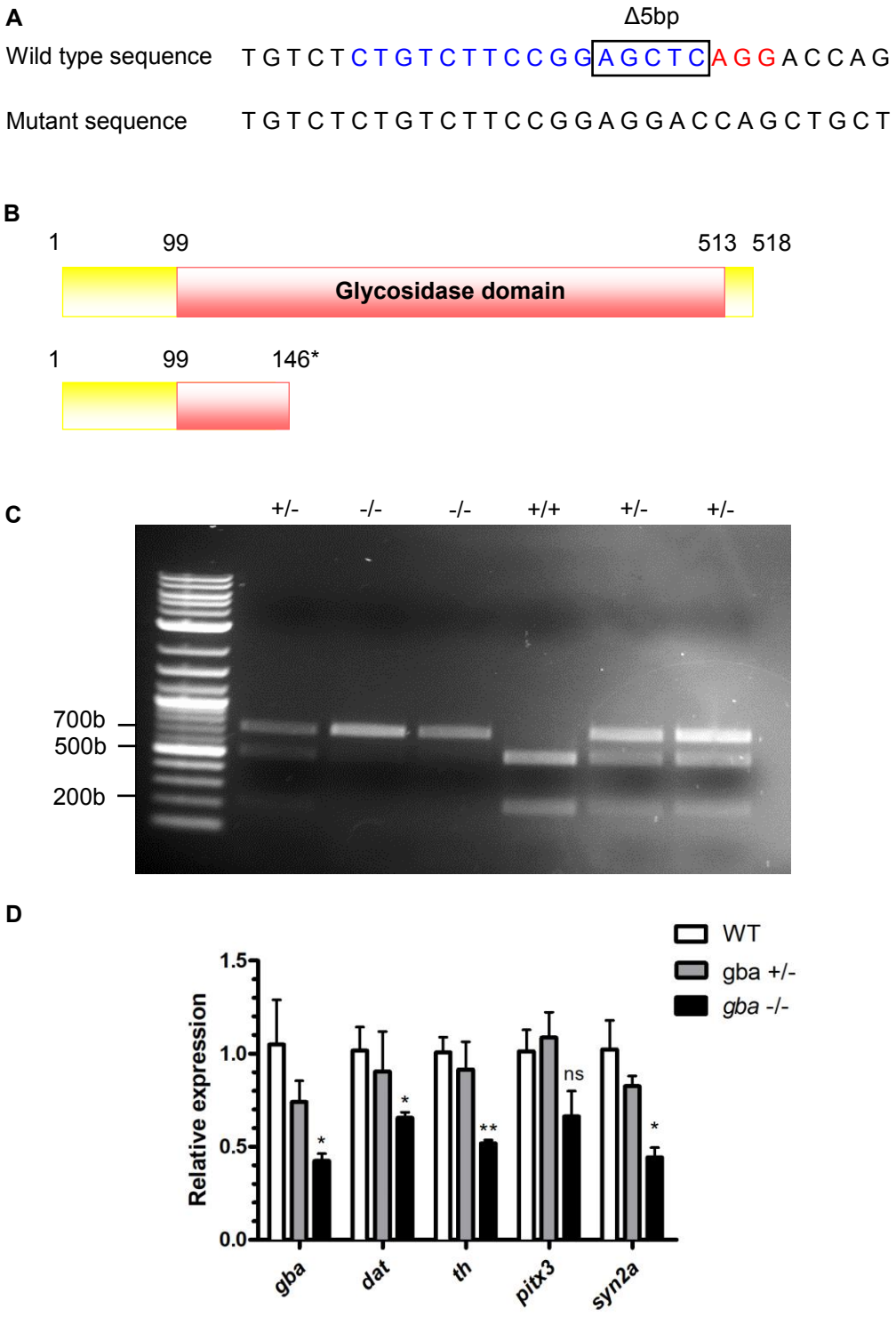


Figure 3.10 Mutation of *gba* and the effect on DA neuron markers

(A) The *gba* target sequence (blue) and PAM site (red) in the wild type and *gba* mutant. A 5bp deletion occurred in the *gba* mutant. (B) Wild type Gba (above) and truncated Gba protein (below) predicted in the *gba* mutant. (C) Gel electrophoresis of *gba* PCR products, amplified from gDNA and digested with SacI. Genotypes of zebrafish are labelled above each lane. (D) qRT-PCR analysis comparing *gba, dat, th*, *pitx3* and *syn2a* expression in *gba -/-* (n=3), gba +/- (n=3) and wild type sibling (n=3) brains at 16 wpf. Error bars represent SEM, * = P<0.05, ** = P<0.01.

3.2.7 *In situ* hybridisation analysis of larval models

In previous work modelling PD in the zebrafish, embryos were studied rather than adults and *in situ* hybridisation was used to analyse the expression of dopaminergic (*dat*) or catecholaminergic (*th*) neuron markers (Bretaud et al., 2007; Flinn et al., 2009; Zhao et al., 2012; Flinn et al., 2013; Priyadarshini et al., 2013). In order to compare the genetic mutants to previous knockdown studies, ISH was used to investigate whether mutation of *parkin*, *pink1* or *dj-1* had an effect on the DA neurons in a zebrafish embryo. For each mutant line, a homozygous and heterozygous mutant were crossed to produce a progeny where half of the individuals were homozygous mutant. This was in an attempt to reduce variability by using sibling controls and the exact same experimental treatment of mutant and control embryos. As all the genes targeted (except *GBA*) are known to cause a phenotype when homozygous recessive, an effect on the DA neurons was expected in 50% of embryos from these out-crosses. The embryos were raised to 3dpf, without genotyping, before fixing and staining for *dat* expression. In addition, to test whether homozygous mutation of *parkin, pink1* or *dj-1* increased the susceptibility of these embryos to oxidative stress half of the individuals were treated with hydrogen peroxide. Based on previous studies, at 2 dpf embryos were treated with 5mM of H_2O_2 for 24 hours before fixing and staining for *dat* (Bretaud et al., 2007, Priyadarshini et al., 2013). No obvious loss of DA neurons projecting to the striatum in diencephalic clusters 2 or 4 was seen in any of the experimental groups (Tay et al., 2011; Rink and Wullimann, 2002). Individuals were therefore separated into two classes, high or low expression of *dat,* based on staining intensity (Figure 3.11A). The percentage of individuals with low *dat* expression was lower than the 50% of predicted homozygous mutants (Figure 3.11B).

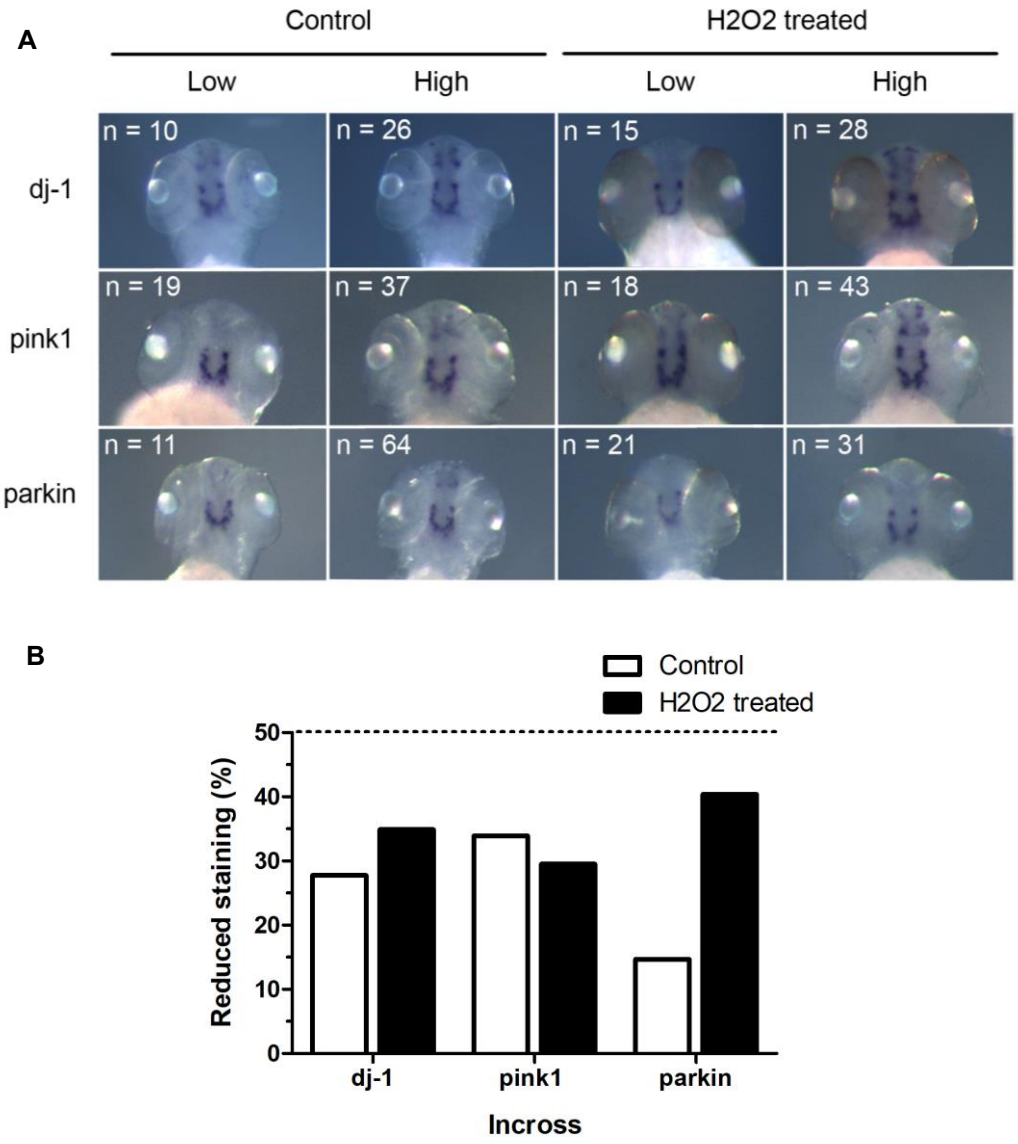


Figure 3.11 ISH analysis of *dat* expression in the PD models

(A) ISH analysis of *dat* expression in the brains of homozygous mutants and sibling controls with control and H_2O_2 treatments. A homozygous mutant was crossed with a heterozygous carrier for each mutant line (*dj-1, pink1, parkin*) to produce a progeny with a 50:50 ratio of mutant to carrier. Half of the individuals from each progeny were treated with 5mM H2O2 for 24 hours to investigate whether increased oxidative stress augmented a loss of DA neurons in the mutants. ISH was used to stain for *dat* in all embryos at 3dpf. Individuals were classified as low or high expression depending on the intensity of staining. Examples of low and high staining for each mutation and treatment combination are shown. (B) A bar chart displays the percentage of low *dat* expressing embryos identified in each scenario for the single analysis.

3.2.8 Analysing PD-models by Western blot

Western blot analysis was carried out to study TH expression in the brains of 3 *parkin*$^{-/-}$ mutants and 3 wild type sibling brains at 16 wpf (Figure 3.12A) Stripping and reprobing with anti-TH antibody was carried out on the western blot membrane containing protein from 3 *dj-1*$^{-/-}$ and 4 wild type sibling brains at 16 wpf (Figure 3.12B). No significant difference in TH expression was seen in either the *parkin*$^{-/-}$ or *dj-1*$^{-/-}$ mutants compared to wild type.

3.2.9 Immunofluorescent labelling of dopamine neurons

In order to specifically analyse the DA neurons that project to the striatum in each PD model, immunofluorescence was used on adult brain sections. Three brains were taken from wild type, *parkin*$^{-/-}$, *pink1*$^{-/-}$ and *dj-1*$^{-/-}$ zebrafish at 12 mpf before sectioning through the posterior tuberculum (pT) in the diencephalon. Figure 3.13A shows a diagram of an adult zebrafish brain with the level of the sections analysed indicated by a line and a further diagram of the sections themselves. The pT is highlighted in red and this is the area of the brain Rink and Wullimann (2001) found to contain the DA neurons that project to the striatum via retrograde tracing. They described the location and shape of these DA neurons as periventricular (adjacent to the ventricle in the pT) with large-pear shaped cells. Figure 3.13B shows a diagram of the pear-shaped cells adjacent to the ventricle based on their work. Cell bodies in the posterior tuberculum positive for TH and having a pear-shaped appearance and a periventricular position were identified as pT DA neurons and counted in each of the sections. Figure 3.13C indicates the mean number of pT DA neurons identified for each genotype. In the 3 wild type brains Figure 3.14A-D a range of 4-5 of these cells were identified; this number is consistent with the retrograde tracing study by Rink and Wullimann (2001). The mean number of pT DA neurons seen in a *dj-1*$^{-/-}$ section was 2 (Figure 3.14E-H), significantly lower than the 4-5 counted in wild type. No significant difference was observed in the number of pT DA neurons counted in *pink1*$^{-/-}$ or *parkin*$^{-/-}$ sections. The number of pT DA neurons ranged from 1-5 for *pink1*$^{-/-}$ (Figure 3.15A-D) and 2-5 for *parkin*$^{-/-}$ (Figure 3.15E-H).

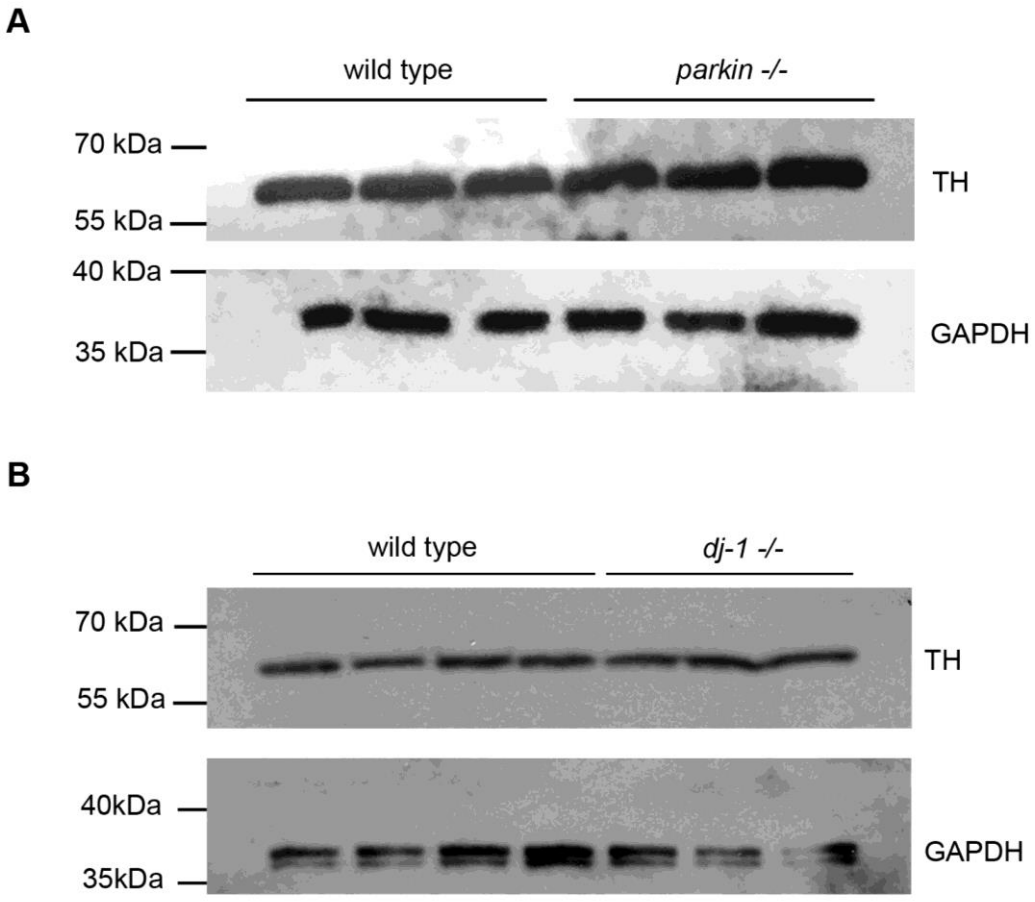


Figure 3.12 TH expression in *parkin-/-* and *dj-1-/-* zebrafish brains

Western blots comparing TH levels in the brains of (A) *parkin-/-* and (B) *dj-1-/-* zebrafish to their wild type siblings at 16 wpf. GAPDH was used as a loading control.

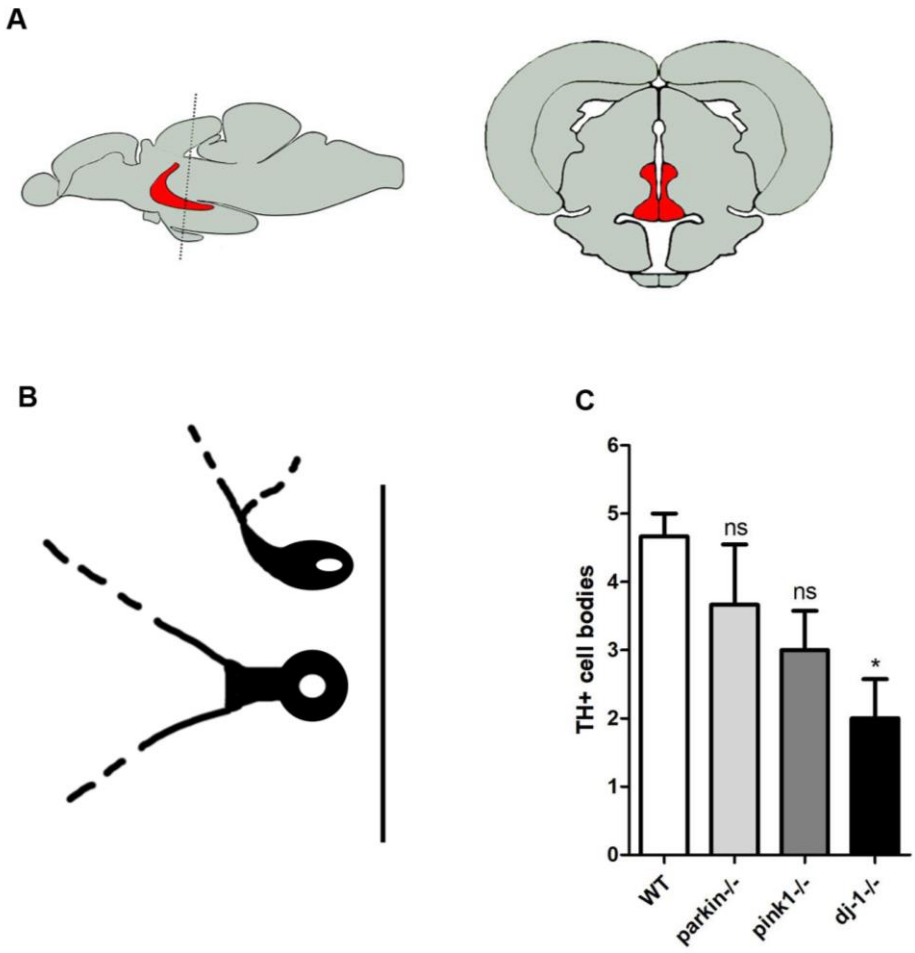


Figure 3.13 Counting the dopamine neurons of the posterior tuberculum

(A) The left diagram is a lateral view of the adult zebrafish brain. The right diagram is a cross section through the adult zebrafish brain. Highlighted is the posterior tuberculum (red). (B) A diagram of the pear-shaped dopamine neuron cell bodies found next to the ventricle of the posterior tuberculum that project to the striatum, based on the work of Rink and Wullimann (2001). (C) A graph plotting the mean number of TH positive cell bodies seen in sections of the posterior tuberculum for wild type, *parkin-/-*, *pink1-/-* and *dj-1-/-* zebrafish (n=3).

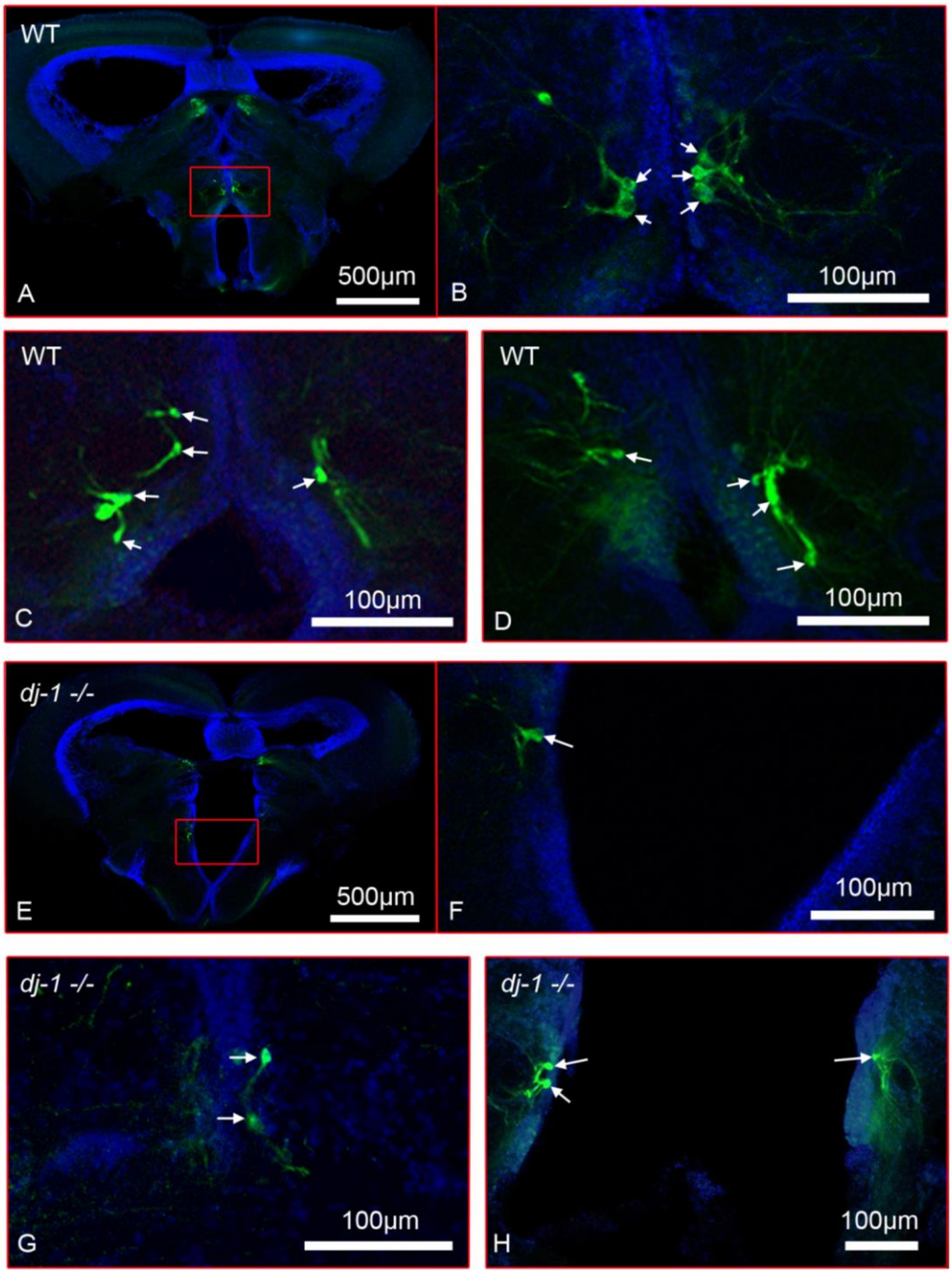


Figure 3.14 Labelling of TH+ cells in wild type and *dj-1*$^{-/-}$ brains

Immunofluorescently-labelled TH positive cells (green) in the posterior tuberculum (pT) of wild type and *dj-1*$^{-/-}$ zebrafish at 12 mpf. Hoechst staining of nuclei is in blue. Arrows indicate the cell bodies of the paraventricular dopamine neurons. (A) TH positive cells in a section through the pT (red box) of a wild type brain. (B) A close up of the dopamine neurons in the pT from A. (C, D) The dopamine neurons from the pT of two further wild type fish. (E) TH positive cells in a section through the pT (red box) of a *dj-1*$^{-/-}$ brain. (F) A close up of the dopamine neurons in the pT from E. (G, H) The dopamine neurons from the pT of two further *dj-1*$^{-/-}$ fish.

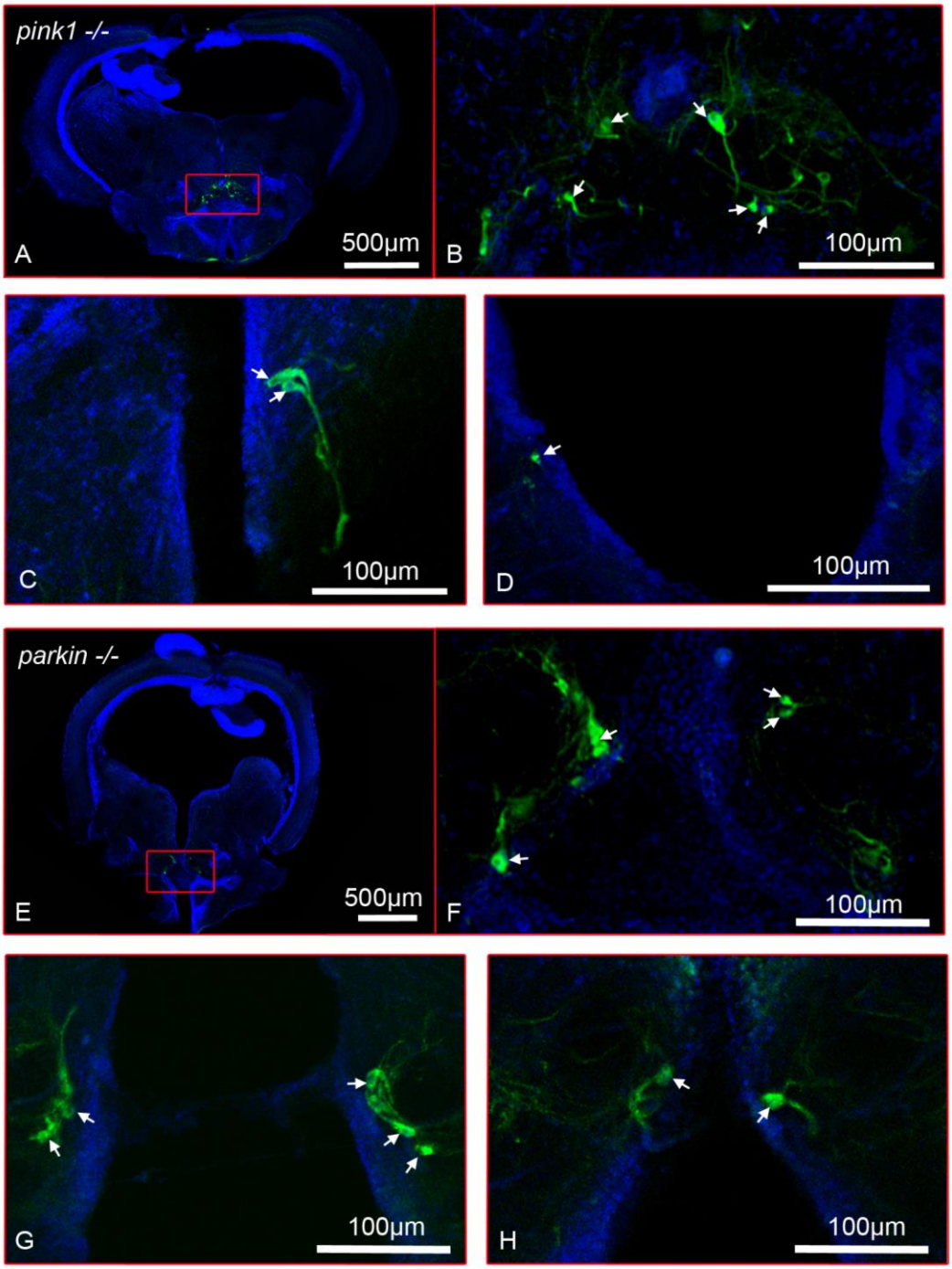


Figure 3.15 Labelling of TH+ cells in *pink1*$^{-/-}$ and *parkin*$^{-/-}$ brains

Immunofluorescently-labelled TH positive cells (green) in the posterior tuberculum (pT) of *pink1*$^{-/-}$ and *parkin*$^{-/-}$ zebrafish at 12 mpf. Hoechst staining of nuclei is in blue. Arrows indicate the cell bodies of the paraventricular dopamine neurons. (A) TH positive cells in a section through the pT (red box) of a *pink1*$^{-/-}$ brain. (B) A close up of the dopamine neurons in the pT from A. (C, D) The dopamine neurons from the pT of two further *pink1*$^{-/-}$- fish. (E) TH positive cells in a section through the pT (red box) of a *parkin*$^{-/-}$ brain. (F) A close up of the dopamine neurons in the pT from E. (G, H) The dopamine neurons from the pT of two further *parkin*$^{-/-}$ fish.

3.3 Discussion

3.3.1 Expression of the PD-associated genes targets

As previously reported, parkinsonian genes are known to be ubiquitously expressed (Kitada et al., 1998; Ilyin et al., 2000; Bandopadhyay et al., 2004; Gandhi et al., 2006). The findings of this chapter demonstrate that the expression of each target gene analysed was seen in the eye, brain, muscle, liver and tail tissue of adult zebrafish and at all stages of development analysed. The DA neuron marker *dat* and general neuronal marker *syn2a*, both revealed strong expression in the brain, as anticipated (Figure 3.4A). When studying temporal expression, all five of the target genes and both the neuronal and muscle marker were seen from 1dpf, whereas *dat* was only expressed from 2dpf (Figure 3.4B). This indicates that the PD-associated genes are active throughout development of the dopamine neurons and a loss of function might be expected to affect the earliest stages of the dopaminergic system.

3.3.2 Effective targeting of the PD-associated genes in zebrafish

CRISPR/Cas9 was shown to be effective at targeting the PD-associated genes in zebrafish, producing frameshift mutations that resulted in premature stop codons downstream. The predicted truncation of each protein caused a loss of the important functional domains. PARKIN loses its RING1, IBR and RING2 domains (Figure 3.6B), the RING2 domain being essential for parkin-mediated ubiquitination (Hampe et al., 2006). Pink1 loses the majority of its kinase domain (Figure 3.7B), which is required to activate Parkin by phosphorylation (Kane et al., 2014). Additionally, a *pink1* mutant line of zebrafish has previously been reported on with a loss of the C-terminal 60 residues of the kinase domain (Flinn et al., 2013). They observed a 65% reduction in *pink1* transcript and loss of kinase activity when modelling the mutation in an insect orthologue. The *pink1* mutation in this study translates to a Pink1 protein missing a much larger part of its C-terminus (392 aa) and almost the whole of the kinase domain; no kinase activity would be expected from this truncated protein. A 90% loss of the *pink1* transcript was also observed (Figure 3.7C), a greater reduction than in the

previous mutant and indicative of nonsense mediated decay (Flinn et al., 2013). The mutant developed for DJ-1 loses the cysteine residue at position 106 (Figure 3.8B). C106 is highly conserved in DJ-1 suggesting that it is important for function and mutations changing the cysteine residue are selected against (Bandyopadhyay and Cookson, 2004). The partial oxidation of C106 is also known to activate DJ-1 and studies have shown that the residue is essential for the neuroprotective function of DJ-1 (Zhou et al., 2006; Canet-Avilés et al., 2004). Therefore, loss of the C106 residue in the *dj-1*$^{-/-}$ zebrafish generated is expected to cause a loss of function. The truncation of FBXO7 causes a loss of the predicted F-box and PRR domains but retains at least part of the Ubl domain (Figure 3.9B). FBXO7 in humans interacts with SCF-type (Skp1–Cul1–F-box) ubiquitin ligase complexes through the F-box domain and recruits substrates for ubiquitination through the PRR (Bai et al., 1996; Hsu et al., 2004). Whereas PARKIN is known to interact with FBXO7 through the Ubl domain (Burchell et al., 2013). Previous PD-associated mutations include one in the F-box domain (R378G) and a premature stop codon in the PRR (R498X) (Shojaee et al., 2008; Fonzo et al., 2009). The mutation in the F-box domain has been shown to impair ubiquitination of substrates on the mitochondrial surface and the premature stop codon in the PRR lessens the translocation of Parkin to mitochondria (Burchell et al., 2013). Therefore, the mutated Fbxo7 may still interact with Parkin through the Ubl domain but the translocation of Parkin and ubiquitination of substrates on mitochondria are expected to be reduced. However, as *FBXO7* mutations cause a rarer form of PD, and there was no decrease in DA neuron markers or nonsense-mediated decay of the *fbxo7* transcript, the analysis of this line was not continued any further. The mutation in *gba* would cause truncation to 145 aa (Figure 3.10B). Recently, another *gba* mutant line of zebrafish was generated with a premature stop codon in exon 7 truncating the protein to 379 aa (Keatinge et al., 2015). A similar loss of *gba* expression was seen and GCase activity was reduced by >50% in the homozygous mutant. A comparable loss of activity is expected in the *gba* mutant generated in this study. There was reduced expression of the DA neuron markers in the *gba*$^{-/-}$ mutant, however, the previously generated *gba*$^{-/-}$ mutant, which had the same severe movement phenotype, was found to have enlarged macrophages recapitulating the pathology of Gaucher's disease (Hruska et al., 2008). Analysis of the heterozygous *gba* mutant revealed no

significant difference in expression of dopaminergic or catecholaminergic neuron markers. As this line has a severe phenotype in the homozygous condition that goes beyond PD pathology work on this line was also stopped at this point. The rest of this thesis focuses on developing zebrafish models of PD with the *parkin, pink1* and *dj-1* mutant lines described in this chapter.

3.3.3 Genetic mechanisms to recover target gene function

A homozygous null mutation seems to have been generated in each of the PD-associated genes targeted. However, cells harbouring mutations can occasionally use genetic mechanisms to recover target gene function. One such mechanism is genetic compensation, where genes are upregulated to rescue the phenotype of a mutant. This is seen with nonsense-mediated decay (NMD) of a mutant transcript but not with gene knockdown or the expression of a dominant negative protein (Rossi et al., 2015). El-Brolosy and Stainier (2017) have proposed two mechanisms by which genetic compensation may occur. Firstly, chromatin remodelling in response to DNA damage could be increasing the expression of genes with a similar function. Chromatin remodelling is widely accepted as a mechanism to increase the transcription of genes by allowing transcription factors access to their promoter regions (Cairns, 2009). El-Brolosy et al. (2019) have shown that the knockdown of chromatin remodelling factors, including WDR5, reduces genetic compensation in mutants. They further identified WDR5 at the transcriptional start site (TSS) of upregulated compensatory genes, strengthening the hypothesis that chromatin remodelling is involved. Ma et al. (2019) have also reported increased levels of WDR5 and H3K4me3, an epigenetic modification to enhance gene expression, at the TSS of compensatory genes upregulated in *capn3a* mutant embryos. Knockdown of WDR5 again reduced genetic compensation in the *capn3a* mutant. Epigenetic modification and chromatin remodelling are therefore strongly associated with the upregulation of compensatory genes. However, DNA damage alone is unlikely to be the trigger activating genetic compensation. Mutations in *hbegfa* and *vcla* where the transcripts had reduced degradation lacked the same upregulation of compensatory genes as mutations causing NMD (El-Brolosy et al., 2019). If genetic compensation was solely in response to DNA damage then the level of transcript degradation should not determine the response. The

second mechanism proposed by El-Brolosy and Stainier (2017) involves the transcript fragments from NMD guiding epigenetic machinery to genes with similar sequences to promote chromatin remodelling and their upregulation. This follows logic as the upregulated genes are normally paralogues or family members of the degraded transcripts, sharing much of the same sequence (El-Brolosy et al., 2019). The correlation between level of transcript degradation and upregulation of compensatory genes further supports this theory. Overall, the proposed mechanism of genetic compensation, involving transcript fragments guiding epigenetic machinery to similar genes, is a strong hypothesis. The *pink1*, *dj-1* and *gba* lines, which had nonsense-mediated decay of the target gene, would be susceptible to genetic compensation. However, both the *dj-1*$^{-/-}$ and *gba*$^{-/-}$ mutants had a loss of DA neuron markers and TH expressing cells were diminished in the *dj-1*$^{-/-}$ posterior tuberculum. Additionally, the *gba*$^{-/-}$ mutant had a severe movement phenotype, identical to a *gba*$^{-/-}$ mutant from a previous study (Keatinge et al., 2015). Enlarged macrophages were seen in the brain of the previously generated *gba*$^{-/-}$ mutant, a pathological hallmark of Gauchers disease which is caused by homozygous mutation of *GBA* in humans (Hruska et al., 2008). Furthermore, genetic compensation is usually seen to upregulate the expression of paralogues in the zebrafish (El-Brolosy and Stainier, 2017; El-Brolosy et al., 2019; Rossi et al., 2015). Zebrafish have no paralogues of *dj-1* or *gba* reducing the likelihood of genetic compensation with nonsense-mediated decay of their transcripts (Keatinge et al., 2015; Bai et al., 2006). Genetic compensation to rescue the DA neurodegeneration phenotype in the *dj-1*$^{-/-}$ and *gba*$^{-/-}$ mutants has therefore not occurred.

An alternative mechanism that can recover target gene function is exon skipping. Exon skipping is the splicing of pre-mRNA to remove exons harbouring a mutation from the mRNA (Dietz et al., 1993). The argument for exon skipping is a strong one. A splicing donor site and a splicing acceptor site, near the 5’ end and 3’ end of an intron, respectively, are required for splicing to occur. There is further regulation of splicing by the splicing signals, which include exonic splicing enhancers (ESEs) and exonic splicing silencers (ESSs) that recruit factors to either promote or inhibit splicing, respectively (Dvinge, 2018). Mutations in splice donor sites, splice acceptor sites, and ESEs can

cause exon skipping and these events have been widely observed in cancer and multiple genetic disorders (Jung et al., 2015; Anna and Monika, 2018). The IVS10–6T>G mutation in the ATM gene occurs in a splice acceptor site and causes the skipping of exon 11 in the mRNA, a shift in the reading frame, truncation of the protein and an increased risk of breast cancer (Broeks et al., 2003, 2000). The genetic disorder, neurofibromatosis 1, can be caused by the mutation of a predicted ESE in exon 7, leading to the skipping of exon 7 in the mRNA (Colapietro et al., 2003). Furthermore, exon skipping can occur with CRISPR/Cas9-mediated mutation of genes. Mou et al. (2017) reported skipping of exon 3 in the *Ctnnb1* gene with CRISPR/Cas9 targeting of exon 3 in a human cell line. Furthermore, they used 11 different sgRNAs all targeting regions in exon 3 of *Ctnnb1* and the 8 most effective sgRNAs caused skipping of exon 3. This suggests that exon skipping can be a regular occurrence with targeted mutagenesis of genes. Similarly, CRISPR/Cas9 targeting of exon 3 in the *FLOT1* gene of human cells has been reported to cause skipping of multiple exons in the mRNA (Kapahnke, Banning and Tikkanen, 2016). Of particular interest, CRISPR/Cas9 targeting of genes in the zebrafish can lead to exon skipping in the mRNA. This was seen by Prykhozhij et al. (2017) when CRISPR/Cas9 targeting produced a mutation in an ESE site of the *pycr1a* gene. Overall, exon skipping can occur with the mutation of particular splice-associated sites in a gene. Exon skipping could rescue the function of a gene if the exon removed from a transcript harboured a mutation and the exon was not essential for the proteins function. However, it can only work to recover the function of the target gene if there is no nonsense-mediated decay of the transcript after splicing. A lack of nonsense-mediated decay was observed in the *fbxo7*$^{-/-}$ and *parkin*$^{-/-}$ mutants so there was potential for exon skipping to have an effect here. A previous study in the mouse saw reduced mitophagy with knockdown of *Parkin*, whereas no reduction was seen with knockout of the *Parkin* gene suggesting recovery of the gene function (Williams et al., 2015). A similar effect could be occurring in the *parkin*$^{-/-}$ mutant zebrafish in this study.

3.3.4 ISH analysis to measure DA neuron loss

Based on previous research ISH was used to measure DA neuron loss in embryos modelling PD. When interpreting the results of the experiment, no

obvious loss of DA neurons projecting to the striatum was observed. Furthermore, no obvious correlation was seen between the number of homozygous mutants predicted in a progeny and the number with lower *dat* expression (Figure 3.11B). A loss of DA neurons is unlikely at the larval stage with the age-related nature of PD (Fearnley and Lees, 1991). Overall, using ISH on embryos to analyse the DA neuron loss was deemed ineffective. ISH is a colorimetric assay and staining intensity can vary between experiments depending on experimental variables. It can be used to see a complete loss or gain of expression in a tissue, but is not able to absolutely quantify the levels of mRNA present.

3.3.5 DJ-1 deficient zebrafish effective at modelling PD

Western blot analysis showed no significant changes of TH expression in the *dj-1*$^{-/-}$ or *parkin*$^{-/-}$ brain at 16 wpf (Figure 3.12). However, this experiment was on whole brain samples and TH is a marker of all of the catecholaminergic neurons not just DA neurons (Rink and Wullimann, 2002) In order to analyse DA neurons and specifically those that are known to project to the striatum immunofluorescence was carried out on sections from the posterior tuberculum (Rink and Wullimann, 2001). A significant loss of TH positive cells was only seen in the posterior tuberculum of *dj-1*$^{-/-}$ zebrafish (Figure 3.14). Another recent study showed that *dj-1*$^{-/-}$ knockout zebrafish have decreased TH expression at 16 mpf (Edson et al., 2019). This supports the result of this study with DA neuron loss in the *dj-1*$^{-/-}$ zebrafish at adulthood. Therefore, the *dj-1*$^{-/-}$ mutant line was deemed the most effective at modelling PD.

3.4 Conclusions

This chapter has shown the effective targeting of PD-associated genes in zebrafish to create homozygous genetic nulls. Premature stop codons were observed in each target gene and the resultant proteins were shown to lack important functional domains, nonsense-mediated decay of the mutant transcript was also seen for *pink1, dj-1* and *gba*. Western blot analysis confirmed a complete loss of expression with the mutation of *dj-1*. To determine

how effectively each mutant line modelled PD, experiments were carried out to study the effect had on the DA neurons. An initial qRT-PCR analysis revealed a loss of DA neuron markers in the *dj-1*$^{-/-}$ and *gba*$^{-/-}$ mutant brains. However, the *gba*$^{-/-}$ zebrafish is suspected to model Gauchers disease which could confound future PD research using the mutant. ISH analysis of the DA neurons in larvae was decided against due to the subjectivity of the method and the age-related nature of PD. Western blot analysis of *parkin*$^{-/-}$ and *dj-1*$^{-/-}$ zebrafish at 16 wpf revealed no overall effect on the catecholaminergic neurons in the brain. Further analysis of the *parkin, pink1* and *dj-1* mutant lines revealed a specific loss of TH positive neurons in the posterior tuberculum of the *dj-1*$^{-/-}$ brain. The *dj-1*$^{-/-}$ zebrafish was therefore determined to be the most effective model of PD.

4 Transcriptomic analysis of the *dj-1*$^{-/-}$ mutant line

4.1 Introduction

4.1.1 DJ-1 in cancer and Parkinson's disease

DJ-1 was initially discovered as an oncogene cooperating with Ras to promote the malignant transformation of cells (Nagakubo et al., 1997). Since then overexpression of the DJ-1 protein has been observed in both cancer cell lines and multiple types of tumour (MacKeigan et al., 2003; Hod, 2004; Davidson et al., 2008; Tian et al., 2008). In contrast, mutations causing a DJ-1 loss of function have been found to cause autosomal recessive PD (Bonifati et al., 2003). Initially, how DJ-1 could be linked to the development of both cancer and PD, two very separate conditions, was unclear. Over time, however, research into the function of DJ-1 has revealed the role it plays in both.

DJ-1 is part of a highly conserved DJ-1/ThiJ/Pfp superfamily present in humans to archaea with diverse functions as kinases, proteases and chaperones (Wilson et al., 2003). The protein itself is redox sensitive, activated by the oxidation of a cysteine residue (C106) (Canet-Avilés et al., 2004). ROS induce oxidation of the cysteine thiol (SH) at C106 to produce a sulfenic acid (SOH) that goes on to form disulphide bonds with cysteine residues in other DJ-1 monomers or target proteins. These disulphide dimers are reduced by thioredoxin, which itself becomes oxidised (Fernandez-Caggiano et al., 2016). Oxidised thioredoxin is reduced by thioredoxin reductase which gains a hydrogen from NADH (Meyer et al., 2009). Through this mechanism DJ-1 can quench ROS. Further oxidation of the cysteine thiol at C106 can produce sulfinic acid (SO_2H). The two oxygen molecules can then form hydrogen bonds with the residues E18, G75 and A107 in the DJ-1 protein and there is a change in the isoelectric point (pI) of the protein from ~6.6 to the more acidic ~5.8 (Canet-Avilés et al., 2004). The oxidation of C106, hydrogen bonds formed and change in net charge of the protein could affect its conformation, protein-protein interactions and activity. C106 oxidation has been found to increase DJ-1 binding affinity for p53, reducing p53 DNA binding activity and the expression of its downstream targets (Kato et al., 2013). Oxidation of C106 has also been seen to increase the interaction of DJ-1 with ASK1, suppressing the pro-

apoptotic signalling of ASK1 and conferring a cytoprotective effect (Waak et al., 2009). The bond formed between C106 and E18 is important for stabilising C106-SO_2H and for its recruitment to mitochondria under oxidative stress conditions to prevent mitochondrial fragmentation (Blackinton et al., 2009). C106 has also been proposed to have a catalytic activity, which is blocked by the C-terminal of DJ-1 and freed by conformational changes after oxidation (Honbou et al., 2003). Finally, the C106 residue of DJ-1 can be irreversibly hyperoxidised to form a sulfonic acid group (SO_3H) which destabilises DJ-1 and targets it for degradation (Song et al., 2016). Once activated DJ-1 promotes the transcription of antioxidant genes and regulates signalling pathways involved in cell survival, proliferation and apoptosis conferring a cytoprotective effect (Clements et al., 2006; Junn et al., 2005; Kim et al., 2005a). These exact same pathways tend to be dysfunctional in cancer with uncontrolled proliferation of cells and a loss of pro-apoptotic signals (Hanahan and Weinberg, 2011). Whereas mutations in DJ-1 are thought to cause neurodegeneration in PD by loss of its cytoprotective function against oxidative stress (Ariga et al., 2013). Oxidative stress is highly associated with the neurodegeneration seen in PD (Blesa et al., 2015) and correspondingly increases in DJ-1 expression and its oxidation have been observed in the brains of PD patients post mortem (Choi et al., 2006; Saito et al., 2014). Oxidative stress is also heightened in cancer cells and the increased expression of DJ-1 is thought to confer a protective effect in them (Hileman et al., 2004)

4.1.2 Signalling pathways downstream of DJ-1

DJ-1 has the ability to regulate a number of signalling pathways involved in the development of both PD and cancer. ERK1/2 and Akt signalling are two of the pathways activated by DJ-1 which promote cell survival, proliferation and progression through the cell cycle (Kim et al., 2005; Yang et al., 2005; Gu et al., 2009; Vasseur et al., 2009). Dysregulation of ERK1/2 and Akt signalling are known to contribute to the uncontrolled growth of cancer cells. Dysregulation of the Akt signalling pathway is thought to be oncogenic based on reports of increased AKT activity (Ringel et al., 2001; Sun et al., 2001) or a loss of PTEN, an inhibitor of Akt signalling, in multiple tumours and cancer cell lines (Dahia et al., 1997; Sakai et al., 1998; Zhou et al., 2000). Increased ERK1/2 activity is

associated with cancer due to the mutation of upstream activators BRAF (Davies et al., 2002) and Ras (Prior, Lewis and Mattos, 2012), observed in multiple tumour types and cancer cell lines, which increase ERK activity (Wan et al., 2004; Xu et al., 2017). Similarly, both ERK and Akt have altered activity in PD. Increased levels of phosphorylated ERK (p-ERK) have been seen in the brains of PD patients, with mitochondrial aggregation in the substantia nigra DA neurons (Zhu et al., 2006). In PD patients with functioning DJ-1, DJ-1 activated by oxidative stress may be phosphorylating ERK as a neuroprotective response. Alternative mechanisms could also be phosphorylating ERK in PD patients. The increased levels of p-ERK could be due to ROS oxidising important cysteine residues in the phosphatases that dephosphorylate p-ERK, inhibiting their activity (Levinthal and DeFranco, 2005).The inverse effect is seen with phosphorylated Akt (p-Akt), detected in the substantia nigra DA neurons of healthy individuals and reduced in PD patients (Timmons et al., 2009). Patients with mutations in *DJ-1* would have reduced AKT phosphorylation under oxidative stress conditions (Aleyasin et al., 2010). Similarly, mutations in *PARKIN* or *PINK1* may reduce AKT activity. *PARKIN* mutations in human cells and in mice have both been shown to reduce AKT signalling (Fallon et al., 2006). PINK1 enhances AKT activity by phosphorylating mTORC2 which is upstream of AKT so a loss of PINK1 could reduce AKT phosphorylation (Murata et al., 2011). AKT signalling is also reduced in a 6-OHDA cell model of PD, in an ROS independent manner, the loss of p-AKT in PD patient DA neurons could therefore occur in a similar matter independent of oxidative stress (Rodriguez-Blanco et al., 2008). ERK1/2 signalling, activated by DJ-1 overexpression in mammalian cells, has been implicated in protecting the cell from oxidative stress (Gu et al., 2009; Kulich et al., 2007). The accumulation of p-ERK in the mitochondria is also associated with increased mitochondrial ROS production (Kulich et al., 2007). The translocation of p-ERK into the mitochondria is thought to occur through VDAC1 (voltage-dependent anion channel 1), the major transport channel across mitochondrial membranes, which p-ERK has been shown to interact with in human cells (Galli et al., 2009). ERK1/2 signalling downstream of DJ-1 can promote the expression of superoxide dismutase (SOD1) an enzyme converting damaging superoxide anions to hydrogen peroxide which can be further converted into oxygen and water (Wang et al., 2011). However, there are a number of regulatory elements

in the human *SOD1* promoter including binding sites for NF-κB, NRF2 and AP-1 allowing these factors to regulate the transcription of SOD1 (Rojo et al., 2004; Park and Rho, 2002; Afonso et al., 2006). A further site binding the proteins SP1 and EGR-1 has been shown to be essential for the basal expression of SOD1 (Minc et al., 1999). One of the transcription factors, AP-1, induced by the pro-inflammatory cytokine TNF-α downregulates SOD1 by preventing SP1 from binding to its promoter (Afonso et al., 2006). NF-κB and NRF2 positively regulate the transcription of SOD1 (Park and Rho, 2002; Rojo et al., 2004). Therefore, an interplay of these transcription factors will determine whether ERK1/2 signalling can upregulate SOD1. The activity of SOD1 can further be regulated by post-translational modifications, including phosphorylation of SOD1 by mTORC1, which has an inhibitory effect (Tsang et al., 2018). ERK1/2 signalling also activates Nurr1, a transcription factor involved in the development and maintenance of DA neurons (Lu et al., 2012; Kadkhodaei et al., 2009). Similarly, Akt signalling has a neuroprotective function, its stimulation has been shown to reduce neurodegeneration in rodent models of PD (Quesada, Lee and Micevych, 2008). DJ-1 is necessary for the activation of AKT signalling as a neuroprotective response to oxidative stress (Aleyasin et al., 2010). In addition to activating pathways for cell survival and proliferation, DJ-1 inhibits ASK1 (apoptosis signal-regulating kinase 1) and MEKK1, both members of the pro-apoptotic JNK signalling pathway. A study in mammalian cells has shown that DJ-1 interacts with Daxx, an upstream activator of ASK1, preventing its translocation into the nucleus and reducing ASK1 activation (Junn et al., 2005). Whereas, DJ-1 has been shown to directly interact with MEKK1 in human cells, inhibiting its activity and preventing its translocation into the nucleus (Mo et al., 2008). Interestingly, the neurotoxin paraquat is known to activate the JNK signalling pathway to cause a selective loss of SN DA neurons (Ramachandiran et al., 2007) Suppression of apoptosis is highly associated with cancer development (Evan and Vousden, 2001), whereas an increase in pro-apoptotic proteins has been seen in the substantia nigra of PD patients post mortem (Tatton, 2000).

4.1.3 Transcriptional regulation by DJ-1

Although DJ-1 itself is not a transcription factor (Yamaguchi et al., 2012), its ability to interact with other proteins means that it plays a role in gene regulation, for instance by stabilising transcription factors, inhibiting repressors and binding co-activators of gene expression. Nrf2, a well-known regulator of the oxidative stress response is stabilised by DJ-1 (Nguyen, Nioi and Pickett, 2009). The cytoplasmic protein Keap1 binds to Nrf2 preventing it from translocating into the nucleus and promotes its ubiquitination, signalling for Nrf2 degradation (Kobayashi et al., 2004; Itoh et al., 1999). DJ-1 has been shown to inhibit the interaction of Nrf2 with Keap1 and reduce its ubiquitination (Figure 4.1A), leading to increased expression of its downstream targets including NQO1 (Clements et al., 2006). NQO1 acts to reduce quinones, created by DA oxidation, which can go on to generate reactive oxygen species (Dinkova-Kostova and Talalay, 2010). Nrf2 signalling downstream of DJ-1 also promotes thioredoxin 1 expression, which acts to reduce the oxidation of other proteins (Im et al., 2012). DJ-1 further regulates gene expression through interactions with transcriptional co-activator p54nrb (Figure 4.1B) and the repressor protein PSF (pro-apoptotic protein-associated splicing factor) (Figure 4.1C) (Xu et al., 2005). p54nrb is thought to activate an antioxidant response in cooperation with DJ-1 (Xu et al., 2005), whereas the binding of DJ-1 to PSF inhibits its repression of tyrosine hydroxylase an enzyme involved in DA synthesis (Zhong et al., 2006). DJ-1 also activates the NF-κB pathway by binding to an inhibitor of NF-κB called Cezanne (Figure 4.1D) (McNally et al., 2011). This promotes the expression of uncoupling proteins, such as UCP4 found in the mitochondria, which act to slightly depolarise the mitochondria and reduce the production of reactive oxygen species during ATP production (Guzman et al., 2010; Xu et al., 2018). DJ-1 is also known to activate HIF-1 signalling by binding to the Von Hippel Lindau protein, a negative regulator of the HIF-1α subunit, preventing HIF-1α ubiquitination and subsequent degradation (Figure 4.1E) (Parsanejad et al., 2014). HIF-1 is a transcription factor that activates gene expression in cells under hypoxic conditions and altered regulation of HIF-1 activity has been implicated in both cancer and PD (Pezzuto and Carico, 2018; Zhang et al., 2011).

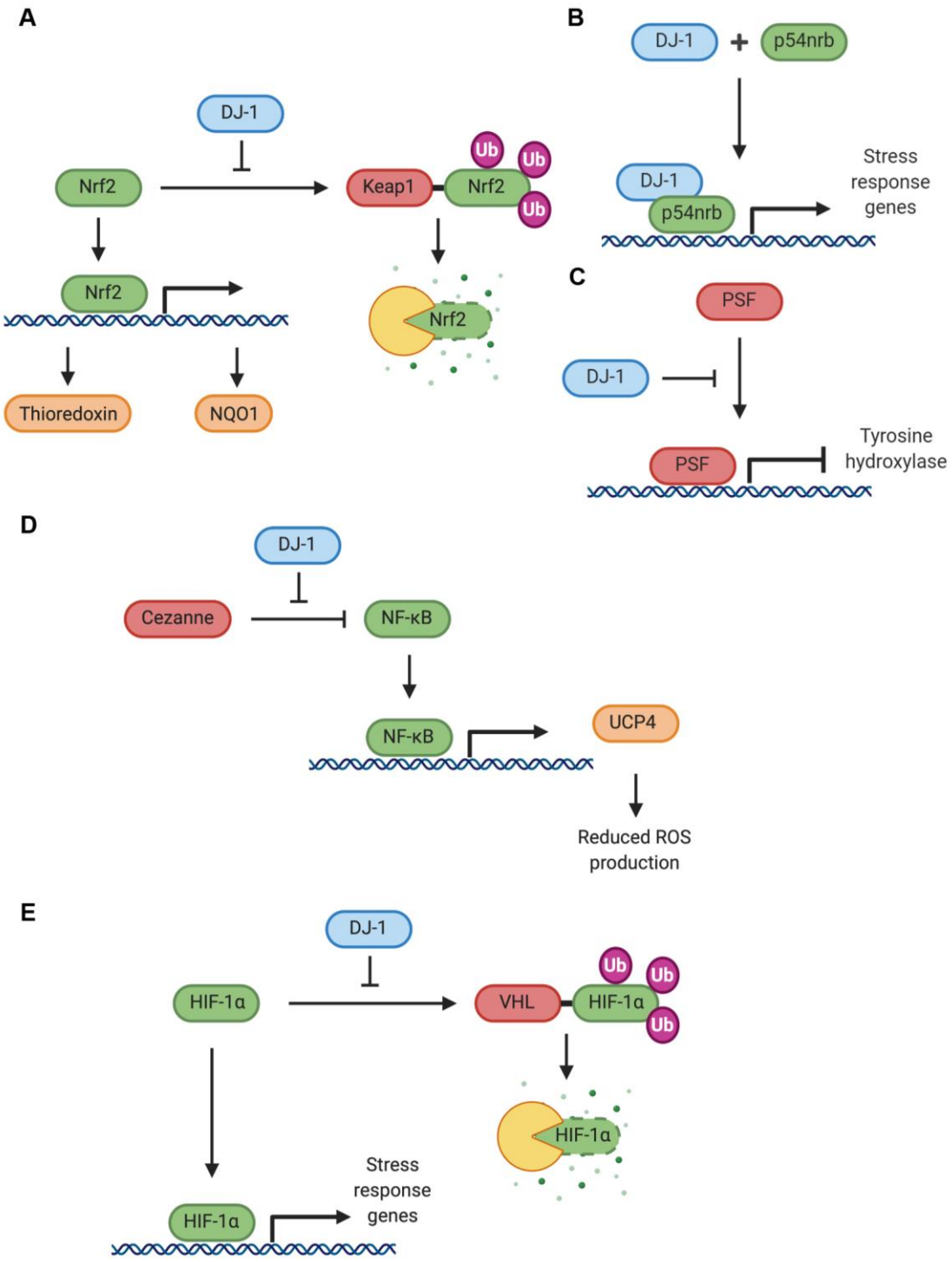


Figure 4.1 Diagrams of transcriptional regulation by DJ-1

(A) DJ-1 inhibits the interaction of Keap1 with Nrf2 which would ubiquitinate Nrf2 targeting it for proteasomal degradation. Nrf2 is then free to activate the transcription of oxidative stress response genes including thioredoxin and NQO1. (B) DJ-1 transcriptionally co-activates stress response genes with p54nrb. (C) DJ-1 inhibits the activity of PSF which is a transcriptional repressor of tyrosine hydroxylase. (D) DJ-1 inhibits Cezanne which is an inhibitor of NF-κB. NF-κB promotes the expression of uncoupling protein UCP4 which depolarizes mitochondria and reduces the production of ROS. (E) DJ-1 prevents the Von Hippel Lindau (VHL) protein from binding and ubiquitinating HIF-1α targeting it for proteasomal degradation. HIF-1α promotes the transcription of stress response genes.

4.1.4 Altered gene expression in previous DJ-1 models

Recent works investigating the downstream targets of DJ-1 include an overexpression study in zebrafish and *Dj-1* knockout models in the rat and zebrafish (Hauser et al., 2017; Frøyset et al., 2018; Edson et al., 2019). A transgenic line of zebrafish, overexpressing DJ-1 in the astrocytes, was generated by Frøyset et al (2018) and treated with the PD-associated neurotoxin MPTP. This drug is used to detect genes upregulated in response to the oxidative stress and energy failure caused by MPTP inhibiting mitochondrial respiration. MPTP is converted to the metabolite MPP+ which is taken up by DA neurons and inhibits complex I in the electron transport chain (Nicklas, Vyas and Heikkila, 1985), increasing the production of ROS (Sriram et al., 1997) and reducing ATP production (Chan et al., 1991). However, MPTP can also alter the activity of pro-apoptotic and anti-apoptotic genes. MPTP treatment has been found to increase the phosphorylation and activation of JNK which signals for apoptosis (Saporito, Thomas and Scott, 2002). Furthermore, MPTP has been shown to upregulate Bcl-2, an anti-apoptotic protein, in cultured DA neurons (Itano and Nomura, 1995) and Bax, a pro-apoptotic protein, in the mouse SN (Itano and Nomura, 1995; Hassouna et al., 1996). Increased cytochrome c levels and caspase-3 activity have also been seen in MPTP treated neuronal cultures (Du et al., 1997), and mice treated with MPTP have increased caspase 3 activity in the SN (Yamada et al., 2010). Cytochrome c is released from the mitochondria and leads to the activation of caspase 3, which promotes programmed cell death (PCD) (Seervi et al., 2011). Therefore, in addition to increased ROS and reduced ATP, the use of MPTP will alter the expression of genes involved in apoptosis and PCD. Proteomic analysis of the isolated astrocytes revealed increased expression of genes involved in the oxidative stress response, mitochondrial function, inflammation and glutathione metabolism (Frøyset et al., 2018). A number of the antioxidant proteins were downstream targets of Nrf2 suggesting its activation by DJ-1. Transcriptomic analysis of the *Dj-1* knockout rat brain revealed altered expression of many genes downstream of Akt signalling and reduced phosphorylation of Akt was confirmed by Western blot (Hauser et al., 2017). Additional gene sets involved in metabolism, MAPK signalling, cancer pathways, calcium signalling and the immune response were found enriched in the *Dj-1*$^{-/-}$ rat brain. During the

timeperiod of my PhD project, a DJ-1 deficient line of zebrafish was generated by another lab in a separate study and a proteomic analysis was carried out comparing protein expression in the brains of wild type and *dj-1*$^{-/-}$ zebrafish (Edson et al., 2019). They reported that the largest group of upregulated proteins were those in the mitochondrial transport chain and the largest group of downregulated proteins were those involved in the stress response and vesicular trafficking. In summation, targets of DJ-1, linked to PD pathology, have been identified in previous models including genes involved in the antioxidant response, mitochondrial function, vesicular trafficking and Akt signalling. Similar changes in gene expression are expected in the transcriptomic analysis of the *dj-1*$^{-/-}$ zebrafish from this study.

4.1.5 Aims of this chapter

The aim of this chapter is to identify changes in gene expression and cell signalling related to PD pathology with a loss of DJ-1 in zebrafish. RNA-Seq analysis was used to identify differentially expressed transcripts in the *dj-1*$^{-/-}$ brain and gene set enrichment analysis (GSEA) (Subramanian et al., 2005) was carried out on the RNA-Seq data to recognise enriched gene sets. The altered expression and enriched gene sets are discussed in relation to PD pathology.

The aims of this chapter are:

- To identify downstream targets of DJ-1 in the zebrafish by RNA-Seq analysis of the *dj-1*$^{-/-}$ brain
- To validate the results of the RNA-Seq by qRT-PCR analysis
- To recognise gene sets enriched in the *dj-1*$^{-/-}$ brain by GSEA

4.2 Results

4.2.1 RNA-Seq analysis of differential gene expression

To investigate any global changes in gene expression associated with a loss of DJ-1, RNA-Seq was carried out on RNA extracted from the brains of 3 *dj-1*$^{-/-}$

mutants and 3 wild type siblings at 16 weeks post fertilisation. The quality of extracted RNA was determined using the Agilent Bioanalyzer 2000 which returned an RNA integrity number (RIN) for each sample. The RIN scale assesses RNA quality from 1 to 10, with a score of 1 indicating degradaded RNA and 10 indicating the sample is fully intact, a commonly used RIN threshold suitable for a transcriptomic analysis is 8 (Imbeaud et al., 2005) and each sample used in this experiment was rated above 9 (Figure 4.2). cDNA libraries were prepared and next-generation sequencing resulted in approximately 50 million reads per sample. The complete data set is available at GEO accession number (135271). Reads were aligned to zebrafish genome assembly GRCz11 and transcript abundance was quantified and expressed in TPM (transcripts per million). The data was subjected to statistical analysis using the likelihood ratio test to identify differentially expressed transcripts in the *dj-1*$^{-/-}$ brain. An overview of the changes in gene expression with a loss of DJ-1 are shown as a volcano plot (MATLAB) with an x axis of log2(fold change) and a y axis of –log10(P-value) (Figure 4.3A). Transcripts in orange that appear above the dotted line are differentially expressed with a P-value of <0.05 and meet fold change thresholds of <0.5 for downregulation or >2 for upregulation. A list of gene targets was created using a significance threshold of q<0.05, adjusting for false discovery rate (FDR), and fold change thresholds of 0.8 and 1.2 to class a transcript as downregulated or upregulated, respectively. The FDR is a measure of the number of false positives or type I errors (Benjamini and Hochberg, 1995). Adjusting for FDR is important when analysing RNA-Seq data sets to take into account the high number of genes and multiple hypotheses being testing (Li et al., 2012). When testing a single hypothesis, using a p-value of 0.05 as a significance threshold, 5% of tests would produce false positives. Whereas, the number of false positives increases with multiple hypothesis testing. The true negative rate is multiplied by itself for the number of hypotheses being tested to produce a new true negative rate (Bland and Altman, 1995). E.g. if 10 hypotheses are being tested, using a p-value of 0.05, the true negative rate becomes 0.95^{10} or 0.60. This means the false positive rate would actually be 0.40 when using a p-value of 0.05. Adjusting for FDR maintains the false positive rate at 0.05 when multiple hypothesis testing and using a q-value of 0.05.32 differentially expressed transcripts (DETs) were identified from the RNA-Seq data (presented in Table 4.1 and Table 4.2). To

present a visual overview of the DETs, a heatmap representing the changes in expression of a curated set of 26 transcripts was created (Figure 4.3B).

4.2.2 Validation of DJ-1 downstream targets

Validation of a set of these targets by qRT-PCR analysis confirmed their altered gene expression (Figure 4.4). RNA samples were extracted from the brains of 3 *dj-1*$^{-/-}$ zebrafish and 3 wild type siblings and reverse transcribed to generate cDNA. An n of 3 was initially chosen to see if the genes that had significantly altered expression in the RNA-Seq data, which used an n of 3, would also have significantly altered expression when analyzing the same number of samples by qRT-PCR. Three biological replicates is also generally accepted for carrying out a qRT-PCR to allow statistical analysis of the data (Udvardi, Czechowski and Scheible, 2008). A further 2 RNA samples from both *dj-1*$^{-/-}$ and wild type sibling brains were analysed by qRT-PCR to identify a significant difference in the expression of *kcnk3b.* A number of DETs identified in the RNA-Seq were selected and primers were designed to amplify their coding regions with 100 ± 10% efficiency. The expression of *abcf2a, brk1, gatm, gpx3, kcnk3, pkma* and *vamp3* was then quantified in each RNA sample. These genes were chosen based on their association with DJ-1, disrupted pathways in PD, DA neurotransmission and neurodegeneration. They also had a fold change of < -0.5 or > 0.5 in the RNA-Seq data increasing the chances of observing their altered expression by qRT-PCR analysis. *abcf2a* is a member of the ATP-binding cassette family of proteins and its human ortholog is a downstream target of DJ-1 (Bao et al., 2017; Clements et al., 2006). Furthermore, *ABCE1*, another member of the ATP-binding cassette family is downregulated in PD patients (Wu et al., 2018), and a separate DJ-1 deficient zebrafish line (Edson et al., 2019). *brk1* is important for inducing actin polymerisation (Derivery et al., 2008) and the actin cytoskeleton is essential for endocytic recycling of dopamine transporter (DAT) in the DA neurons (Gabriel et al., 2013). Dopamine transporter is involved in the reuptake of DA from the synaptic cleft, therefore, dysfunctional recycling of DAT could affect the neurotransmission of DA (Pramod et al., 2013). *Gatm* encodes glycine amidinotransferase, a mitochondrial enzyme involved in generating creatine, which is an energy store (Sandell et al., 2003). A loss of *gatm* may indicate mitochondrial dysfunction

and reduce the energy available to DA neurons which DA neurons are particularly vulnerable to (Bolam and Pissadaki, 2012). *gpx3* encodes glutathione peroxidase 3, an enzyme involved in the oxidative stress response (Chung et al., 2009). DA neurodegeneration is highly associated with oxidative stress and DJ-1 regulates the expression of antioxidant genes (Clements et al., 2006). *Kcnk3b* encodes a potassium (K+) channel, and altered activity of K+ channels can affect the pacemaking activity of SN DA neurons and increase their vulnerability (Du et al., 2016; Wolfart et al., 2001). *pkma* encodes pyruvate kinase, the rate limiting enzyme in glycolysis (Israelsen and Vander Heiden, 2015), and its loss could indicate metabolic dysfunction which SN DA neurons are particularly susceptible to (Anandhan et al., 2017). Finally, vamp3 is a part of the SNARE protein complex required for vesicular transport from endosomes to the trans-Golgi network (Mallard et al., 2001). SN DA neurons have increased vulnerability to dysfunctional intracellular trafficking as vesicles have to travel extreme lengths down the axon and to an increased number of synapses (Bolam and Pissadaki, 2012; Giguère et al., 2019). A further gene identified with altered expression and a role in neuronal development, *sema3fb*, was excluded from the qRT-PCR analysis (Kolk et al., 2009). The simultaneous increase and decrease of different *sema3fb* transcripts would make it difficult to draw conclusions on overall *sema3fb* activity. Furthermore, *sema3fb* is associated with neuronal development rather than the neurodegeneration seen in PD. All seven of the transcripts analysed had significantly altered expression, consistent with the change in expression detected by RNA-Seq.

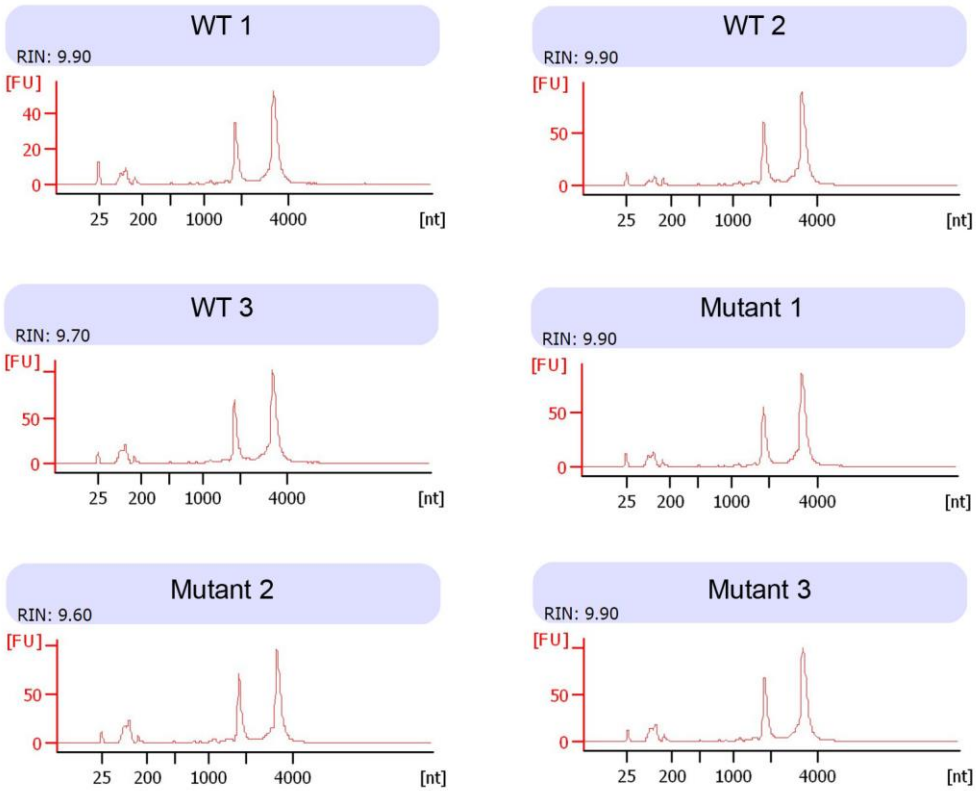


Figure 4.2 Integrity of RNA samples used in the RNA-Seq analysis

Bioanalyzer results for RNA samples extracted from the brains of *dj-1*$^{-/-}$ zebrafish (mutant) and wild type siblings (WT). The RNA integrity number (RIN) is a measure of each RNA samples quality from 1 to 10. RIN of 1 = degraded, RIN of 10 = fully intact.

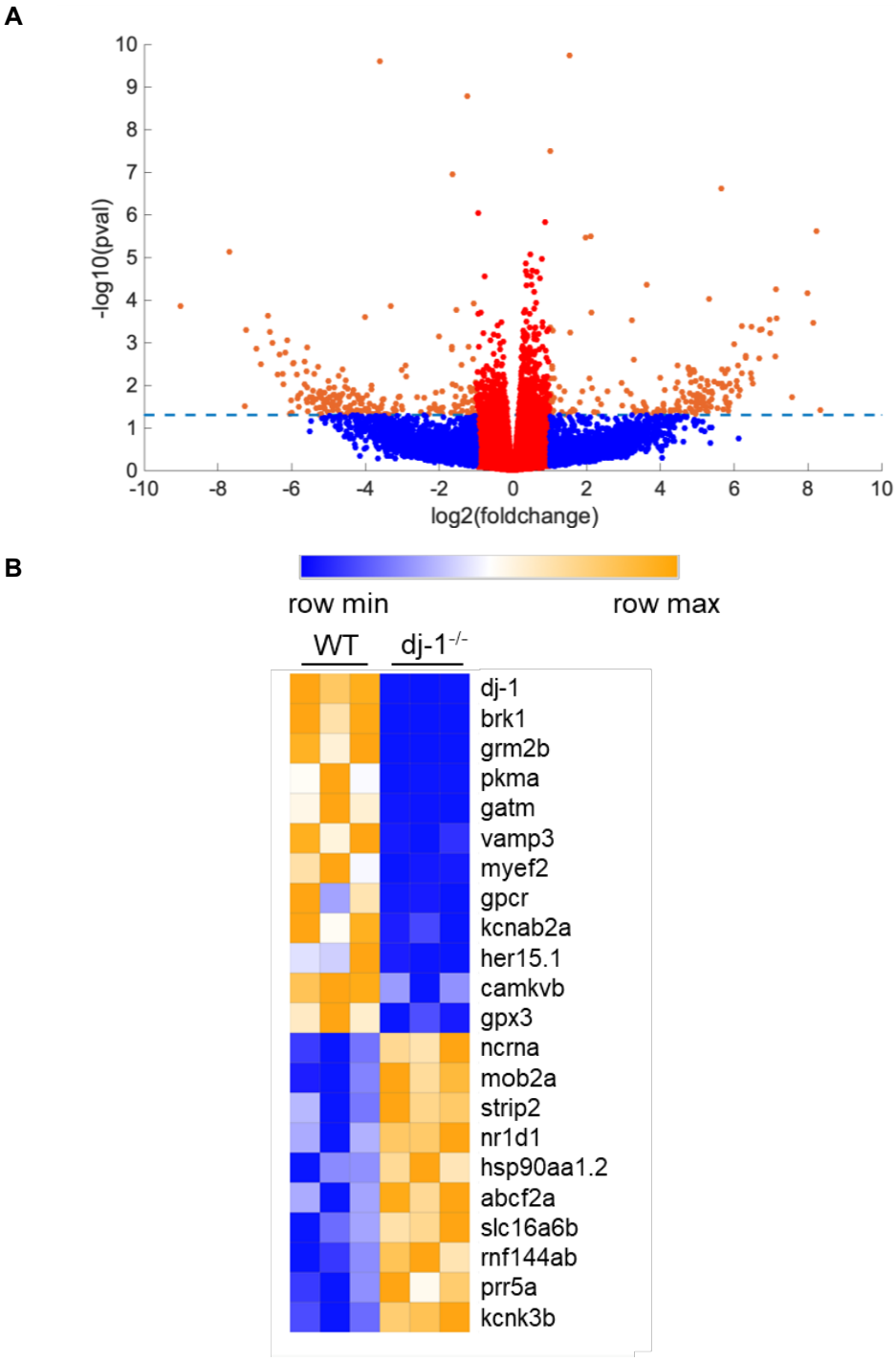


Figure 4.3 Differential gene expression in the *dj-1-/-* zebrafish brain
(A) A volcano plot of the global changes in gene expression observed in the *dj-1-/-* brain by RNA-Seq analysis. The fold change (log2) of each transcript was plot against the P-value (-log10). Transcripts in red do not meet the fold change thresholds of <0.5 for downregulation or >2 for upregulation. Transcripts in orange that appear above the dotted line are differentially expressed with a P-value of <0.05. Transcripts in blue have a P-value of >=0.05. (B) A heatmap showing select genes differentially expressed with a Q-value of <0.05. Relative expression is represented on a colour scale from blue (low) to orange (high).

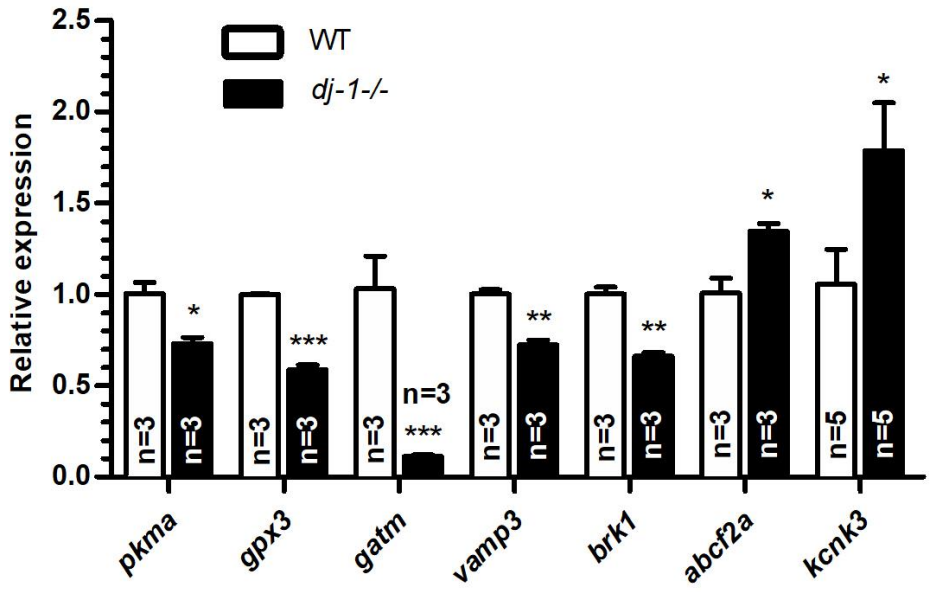


Figure 4.4 qRT-PCR validation of the RNA-Seq results

RT-qPCR analysis validating the results seen in the RNA-Seq. An n of at least 3 biological replicates was used for each assay with 3 technical repeats for each. Student t-tests were used to compare the ΔCT values for the *dj-1-/-* and wild type siblings. Error bars represent SEM, * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$.

Gene Name	NCBI Transcript ID	Fold Change (log2)	q-value	Avg. WT tpm	Avg. *dj-1-/-* tpm
grm2b	NM_001287547.1	-11.41	0.00	60.20	0.02
sema3fb	XM_009306194.3	-10.45	0.00	9.93	0.00
si:ch211-102c2.8 (ncRNA)	XM_021476476.1	-7.68	0.01	1.42	0.00
brk1	XM_005172903.4	-6.49	0.00	117.14	1.28
park7 (dj-1)	NM_001005938.1	-4.17	0.00	160.17	8.87
pkma	NM_199333.1	-3.62	0.00	496.18	38.60
si:ch211-213o11.11 (predicted GPCR)	XM_017358408.2	-3.62	0.00	8.62	0.66
vamp3	NM_001002073.1	-2.79	0.00	12.86	1.87
gatm	NM_199531.1	-2.56	0.00	33.80	5.57
her15.1	XM_003199526.5	-1.63	0.00	13.18	4.05
cldn19	XM_017358226.2	-1.36	0.00	38.59	15.09
kcnab2a	XM_009306291.3	-1.23	0.00	19.41	8.20
myef2	NM_001037423.1	-1.20	0.00	37.69	16.17
camkvb	NM_200450.1	-0.93	0.00	98.64	52.36
gpx3	NM_001137555.1	-0.76	0.04	33.62	19.65

Table 4.1 Significantly downregulated transcripts in the *dj-1-/-* brain

A list of significantly downregulated transcripts identified in the *dj-1-/-* zebrafish brain by RNA-Seq analysis. A significance threshold of q<0.05 was used.

Gene Name	NCBI Transcript ID	Fold Change (log2)	q-value	Avg. WT tpm	Avg. *dj-1-/-* tpm
Nfic	XM_009303916.3	8.22	0.01	0.00	0.70
sema3fb	XM_021479878.1	5.65	0.00	0.18	8.99
mslna	XM_021474714.1	2.14	0.00	0.44	2.02
LOC110437990 (ncRNA)	XR_002455805.1	2.12	0.01	7.87	32.20
si:ch211-213o11.11 (predicted GPCR)	XM_021479881.1	1.97	0.01	1.19	4.42
cldn19	NM_001017736.1	1.54	0.00	5.61	15.91
zgc:194629 (ncRNA)	NR_120380.1	1.01	0.00	55.40	110.82
mob2a	XM_005170681.4	0.88	0.00	9.92	18.14
adamts1	XM_021475923.1	0.79	0.02	2.68	4.60
kcnk3b	XM_694909.8	0.74	0.04	7.40	12.31
slc16a6b	XM_685174.9	0.65	0.03	9.49	14.80
abcf2a	NM_201315.2	0.52	0.03	54.02	77.35
prr5a	XM_005163058.4	0.49	0.04	19.10	26.79
strip2	XM_021475407.1	0.48	0.02	19.89	27.58
rnf144ab	XM_005160676.4	0.38	0.04	50.12	65.35
nr1d1	NM_205729.2	0.35	0.02	84.96	108.17
hsp90aa1.2	NM_001045073.1	0.34	0.03	127.68	161.99

Table 4.2 Significantly upregulated transcripts in the dj-1-/- brain

A list of significantly upregulated transcripts identified in the dj-1-/- zebrafish brain by RNA-Seq analysis. A significance threshold of q<0.05 was used.

4.2.3 Identifying enriched gene sets from the RNA-Seq data

Identifying individual target genes and manually curating a 'gene list' may not give a complete picture of the changes in biology that occur in the absence of DJ-1. Therefore, a statistical approach called Gene Set Enrichment Analysis was carried out on the transcriptomic data to uncover small changes in large sets of genes, grouped by biological function (Subramanian et al., 2005). An advantage of using GSEA over gene ontology analysis is that all of the genes in a gene set are used to test for enrichment in the expression data, rather than just those with statistically significant changes in expression (Yoon, Kim and Nam, 2016). The genes from an expression analysis are ordered by correlation between expression and genotype before determining whether each gene set being studied is overrepresented at the top or bottom of that list (Subramanian et al., 2005). When studying the individual genes that are differentially expressed in a transcriptomic analysis there can also be large biological and technical variability (McIntyre et al., 2011). GSEA helps to overcome this issue by determining activated pathways/functions from the overall changes in expression. The results become more reproducible and comparable between experiments (Subramanian et al., 2005). Here, the RNA-Seq data was interrogated by GSEA to identify Hallmark gene sets enriched in the *dj-1*$^{-/-}$ zebrafish brain. The hallmark gene sets, from the Molecular Signature Database (Broad Institute), represent well defined biological processes (Liberzon et al., 2015). Gene sets with normalised enrichment scores (NES) of ≥ 1.5, FDR q-values of ≤ 0.05 and nominal p-values of ≤ 0.05 were considered significant (Table 4.3). Of particular interest, gene sets involved in progression of the cell cycle (HALLMARK_G2M_CHECKPOINT, HALLMARK_E2F_TARGETS), Akt signalling (HALLMARK_PI3K_AKT_MTOR_SIGNALLING) and oxidative phosphorylation (HALLMARK_OXIDATIVE_PHOSPHORYLATION) were enriched (Figure 4.5A). The cell cycle related gene sets were interesting because re-entry of neurons into the cell cycle has been associated with neurodegeneration (Herrup et al., 2004; Herrup and Yang, 2007; Wang et al., 2009). Cell cycle markers have also been identified in both the SN DA neurons of PD patients post-mortem (Höglinger et al., 2007), and the SN DA neurons of a PD model in the rat (El-Khodor et al., 2003). PI3K AKT MTOR signalling was an interesting gene set because dysfunctional AKT signalling is associated with PD (Timmons

et al., 2009) and altered AKT signalling has been seen in DJ-1 deficient mice (Hauser et al., 2017). The oxidative phosphorylation gene set was of interest because it occurs in the mitochondria and mitochondrial dysfunction is highly associated with DA neurodegeneration (Winklhofer and Haass, 2010), and activation of the DJ-1 oxidative stress response (Aleyasin et al., 2010).

4.2.4 Validation of enriched gene sets

qRT-PCR analysis was carried out to investigate the enrichment of cell cycle related gene sets. Based on an experiment by Nguyen et al. (2017) the expression of cyclins *ccna2*, *ccnb1*, *ccnb2*, *ccnd1* and *ccne* were compared in 3 *dj-1*$^{-/-}$ and 3 wild type sibling brains at 16 wpf. Cyclins are involved in regulating the cell cycle and have altered expression at different stages. D and E-type cyclins (e.g. *ccnd1* and *ccne*) are induced by mitogenic growth factors in the G1 phase of the cell cycle (Mann et al., 1997). They go on to form complexes with cyclin-dependent kinases (CDK) 4 and 6, and with CDK2, respectively (Bates et al., 1994; Koff et al., 1992). There is sequential phosphorylation of the retinoblastoma protein (Rb) by the CDK4/6-Cyclin D and CDK2-Cyclin E complexes preventing Rb from binding to the E2F transcription factors (Lundberg and Weinberg, 1998). The uninhibited E2Fs are then free to promote transcription of downstream targets involved in cell cycle progression including A and B-type cyclins (e.g *ccna2*, *ccnb1* and *ccnb2*), which accumulate throughout S phase and into the G2 phase (Zhu, Giangrande and Nevins, 2004), whilst cyclin E is degraded (Clurman et al., 1996). Cyclin A forms a complex with CDK2 and phosphorylates and inhibits substrates including CDC6 and MCM4, which are involved in genome duplication, to prevent DNA re-replication at the end of the S phase (Petersen et al., 1999; Ishimi et al., 2000). Cyclin A further interacts with CDK1 to phosphorylate and activate downstream targets that go on to trigger mitosis and one of these targets, Plx1, goes on to activate the CDK1/Cyclin B complex (Vigneron et al., 2018). Active CDK1/Cyclin B complex is an essential component of the M-phase promoting factor (MPF) required for cells to enter mitosis (Hara et al., 2012). Therefore, the cyclins act as markers for different stages in the cell cycle. No significant difference was seen for any of the cyclins; however, as only a specific set of cells in the posterior tuberculum are thought to be affected any differences in cell cycling

were unlikely to be detected in the whole brain. In the Nguyen et al. (2017) paper, cells were sorted by fluorescence before analysing the expression of different cyclins and detecting increased G2/M checkpoint markers.

Two gene sets of particular interest that were found significantly enriched in the *dj-1*$^{-/-}$ zebrafish brain were the oxidative phosphorylation and PI3K AKT MTOR signalling gene sets. Oxidative phosphorylation occurs in the mitochondria and enrichment of associated genes is interesting as mitochondrial dysfunction is highly associated with PD pathology and activation of the DJ-1 oxidative stress response (Winklhofer and Haass, 2010; Aleyasin et al., 2010). Furthermore, mitochondrial dysfunction and increased expression of mitochondrial complex I components has been seen in a previous DJ-1 deficient line of zebrafish (Edson et al., 2019). PI3K AKT MTOR signalling is an interesting gene set because dysfunctional AKT signalling has been associated with PD and a DJ-1 loss of function in mice (Timmons et al., 2009; Hauser et al., 2017). Western blot analysis was carried out as a simple assay to determine whether the enrichment of oxidative phosphorylation and PI3K AKT MTOR signalling indicated an increased number of mitochondria and altered AKT signalling. The levels of phosphorylated Akt (P-AKT) and a mitochondrial marker (ATP5F1B) were compared between whole brain lysates of 3 *dj-1*$^{-/-}$ and 3 wild type siblings at 16 wpf. P-AKT was studied to determine if there was altered Akt signalling in the whole brain. The mitochondrial marker was analysed to determine whether enrichment of the oxidative phosphorylation gene set was due to an increase in mitochondria. ImageJ analysis of the protein bands revealed no significant difference in P-AKT or the mitochondrial marker with a loss of DJ-1 (Figure 4.5C). Again, with no way of isolating the DA neurons of the posterior tuberculum, the level of sensitivity offered by Western blotting would be unlikely to detect any change in these functions in whole brains. While the methods used in this chapter to detect changes in expression have utilised the whole brain, rather than just the DA neurons of the posterior tuberculum, the data extracted is still relevant. DJ-1 is expressed ubiquitously in both humans (Bonifati et al., 2003) and zebrafish (Bretaud et al., 2007), its expression is not specific to the dopamine neurons. However, a loss of DJ-1 causes a specific degeneration of the dopamine neurons in humans (Bonifati et al., 2003). DJ-1 is also a major regulator of transcription in the oxidative stress response

(Clements et al., 2006). A loss of DJ-1 would affect the transcription of its downstream targets in the whole brain. A transcriptomic analysis on RNA extracted from the whole brain would therefore be able to detect these changes in all neuronal cell types. However, these changes in expression can contribute to the degeneration of dopamine neurons, which have increased vulnerability to oxidative stress (Kim et al., 2005b). The genes that are specifically altered in the DA neurons may go undetected but those differentially expressed throughout the brain and contributing to DA neurodegeneration would be identified. Altered AKT signaling and re-entry into the cell cycle may be effects that specifically occur in the degenerating dopamine neurons. This would be similar to reports of altered AKT signaling and cell cycle markers in occurring specifically in the SN DA neurons of PD patients (Timmons et al., 2009; Jordan-Sciutto et al., 2003).

Name	Gene set size	NES	p-val	FDR-adjusted q-val
G2M Checkpoint	125	2.11	0.00	0.00
Oxidative Phosphorylation	139	2.04	0.00	0.00
E2F Targets	131	1.87	0.00	0.00
Epithelial Mesenchymal Transition	98	1.82	0.00	0.01
Cholesterol Homeostasis	44	1.82	0.00	0.00
PI3K AKT MTOR Signaling	63	1.69	0.00	0.02
MYC Targets V1	135	1.62	0.00	0.03
MTORC1 Signaling	137	1.60	0.00	0.03
Mitotic Spindle	121	1.57	0.00	0.04
TGF beta Signaling	25	1.53	0.02	0.05
Androgen Response	64	1.53	0.01	0.04

Table 4.3 Hallmark gene sets enriched in the *dj-1-/-* zebrafish brain

A list of Hallmark gene sets found significantly enriched in the *dj-1-/-* zebrafish brain. The thresholds for significance were NES >1.5, p-value<0.05 and FDR-adjusted q-value<0.05. NES = normalized enrichment score.

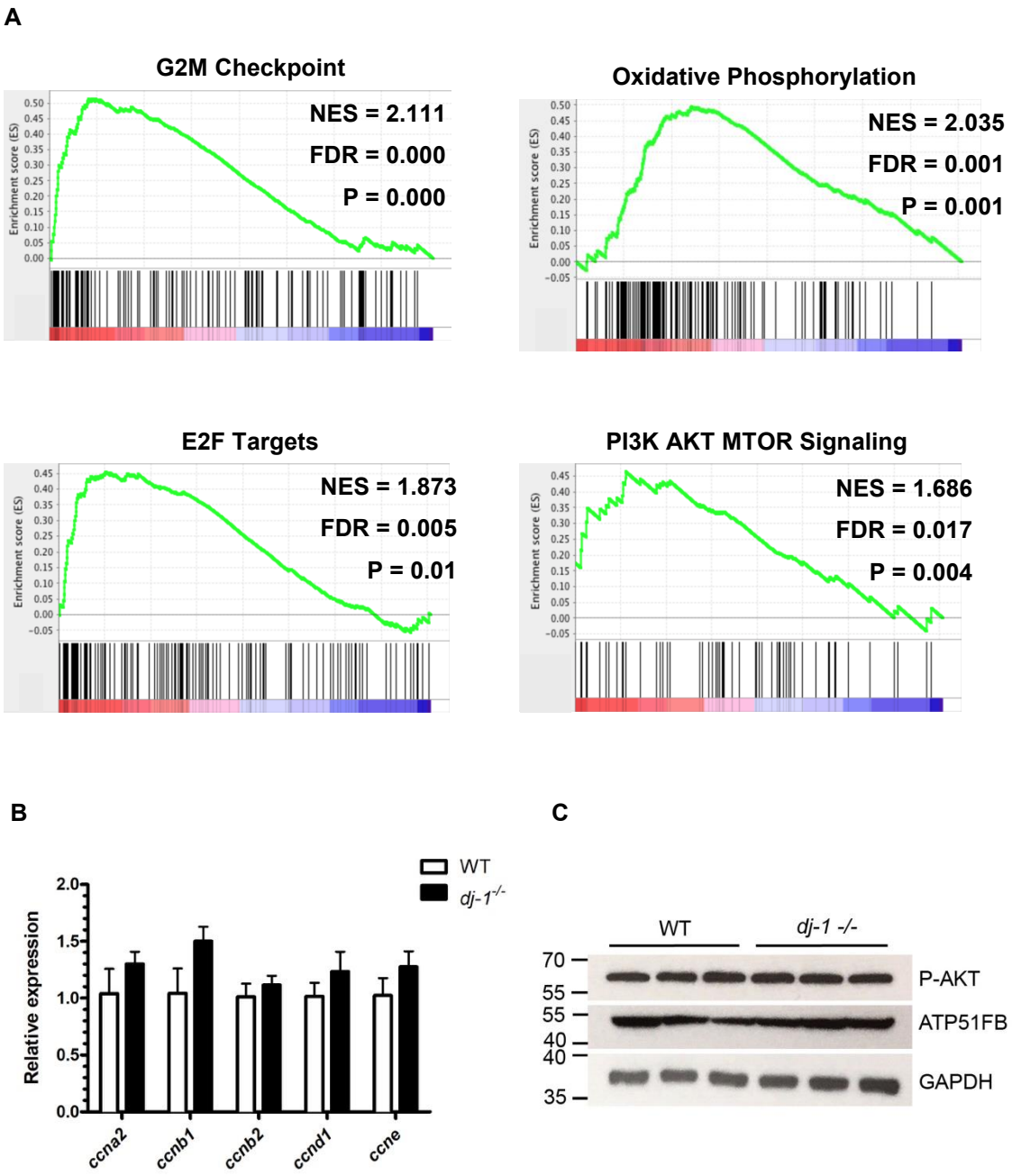


Figure 4.5 Enriched gene sets implicated in Parkinson's disease

(A) GSEA enrichment plots for Hallmark gene sets: G2M checkpoint, oxidative phosphorylation, E2F transcription factors and PI3K/AKT/mTOR signalling. NES = normalised enrichment score. FDR = false discovery rate adjusted P-value. P = nominal P-value. (B) qRT-PCR analysis comparing the expression of cyclins *ccna2, ccnb1, ccnb2*, *ccnd1* and *ccne* in the brains of *dj-1-/-* zebrafish (n=3) and wild type siblings (n=3). Student t-tests were used to compare the ΔCT values of *dj-1-/-* and wild type samples. Error bars represent SEM. (C) Western blot analysis comparing P-AKT and ATP5F1B levels in the *dj-1*$^{-/-}$ and wild type brain. GAPDH was used as a loading control.

4.3 Discussion

A number of pathogenic mechanisms contribute to the loss of substantia nigra DA neurons in PD including oxidative stress and mitochondrial dysfunction (Henchcliffe and Beal, 2008). DJ-1 is a major regulator of transcription and signal transduction that plays a role in the oxidative stress response and mitochondrial maintenance to protect the cell (Clements et al., 2006; Hao, Giasson and Bonini, 2010). However, there are still downstream targets of DJ-1 that mediate its neuroprotective function yet to be discovered. In this study a mutant line of zebrafish lacking DJ-1 was generated and a transcriptomic analysis was carried out on the brains at adulthood to identify downstream targets of DJ-1 that may play a role in the neurodegeneration of PD.

4.3.1 Reduced expression of *gpx3*

Among the genes with significantly altered expression was glutathione peroxidase 3 (*gpx3*), an enzyme that reduces hydrogen peroxide, protecting the cell from oxidative stress (Chung et al., 2009). In addition to its role in the antioxidant response, *gpx3* was of interest because of the regulation of its human ortholog. The promoter of human *GPX3* is known to contain an antioxidant response element (ARE), a peroxisome proliferator response element (PPRE) and a HIF-1 (hypoxia-inducible factor-1) binding site (Figure 4.6) (Bierl et al., 2004; Chung et al., 2009). Nrf2 is known to activate the transcription of oxidative stress response genes through binding to the ARE linking *GPX3* expression to DJ-1 because of its role in stabilising and activating Nrf2 (Clements et al., 2006; Venugopal and Jaiswal, 1996). The PPRE in *GPX3* is another binding site that could make it a downstream target of DJ-1. A peroxisome proliferator-activated receptor (PPAR) dimerised with a retinoic acid receptor (RXR) can bind to a PPRE site activating downstream transcription (Tyagi et al., 2011). PPARγ and RXRα have been shown to promote the expression of *GPX3* in human and mouse cells suggesting conservation across species (Chung et al., 2009). DJ-1 is thought to activate PPARγ expression with overexpression of DJ-1 in human cells increasing PPARγ expression (Han et al., 2018), and conversely DJ-1 knockdown reducing the levels of PPARγ seen

in human cells (Kim et al., 2014). A further binding site in the human *GPX3* promoter is HIF-1 (Bierl et al., 2004). The HIF-1 transcription factor binds to the HIF-1 site, under hypoxic conditions, to promote the transcription of genes that can reduce oxidative stress (Semenza, 2011). However, the HIF-1α subunit of HIF-1 is normally ubiquitinated by the Von Hippel Lindau (VHL) protein, signalling for its degradation (Groulx and Lee, 2002). DJ-1 has been shown to interact with the VHL protein in human cells, preventing the ubiquitination of HIF-1α and accordingly DJ-1 knockout causes a loss of HIF-1α (Parsanejad et al., 2014). Depending on which binding sites are conserved upstream of *gpx3* in the zebrafish DJ-1 could be activating its expression through Nrf2, PPARγ or HIF-1.Overall, *gpx3* is likely a downstream target of DJ-1 in the oxidative stress reponse of zebrafish.

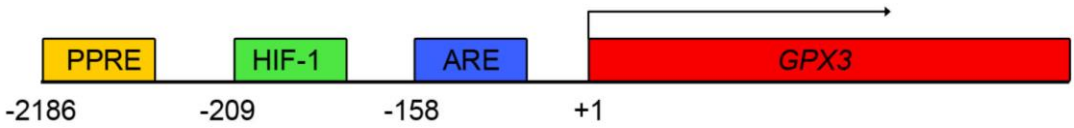


Figure 4.6 Binding sites upstream of human *GPX3*

Diagram of the PPRE, HIF-1 and ARE binding sites in the promoter of human *GPX3* (position of binding sites informed by Bierl et al. (2004) and Chung et al. (2009)).

4.3.2 Genes implicated in Parkinson's disease

Some of the genes significantly affected by a loss of DJ-1 are related to cell functions known to be disrupted in PD. One of the downregulated genes in the *dj-1*$^{-/-}$ brain encodes, *grm2b*, the ortholog of *GRM2* in humans (zfin.org) which encodes glutamate metabotropic receptor 2 (mGluR2) (Joo et al., 2001). This is of particular interest because altered glutamatergic neurotransmission is thought to play a role in the neurodegeneration of PD (Gasparini, Di Paolo and Gomez-Mancilla, 2013). Furthermore, a loss of glutamatergic synapses has been seen in the striatum of a rodent model of PD (Day et al., 2006). mGluR2 is known to reduce glutamate release in areas of the brain including the striatum (Carrillo-Mora, Silva-Adaya and Villaseñor-Aguayo, 2013). Therefore, reduced expression of mGluR2 could contribute to the altered glutamatergic neurotransmission seen in PD.

The differentially expressed genes also include those related to neuronal development, structure and function. For instance, a decrease in expression was observed for *myef2,* a transcriptional repressor of the myelin basic protein which is involved in the myelination of neurons (Haas et al., 1995; Brösamle and Halpern, 2002). Altered expression was also seen for *brk1*, *strip2* and *camkvb,* genes involved in cytoskeletal organisation which is important in neuronal development and dendritic morphology. (Bai et al., 2011; Derivery et al., 2008; Konietzny, Bär and Mikhaylova, 2017). Knockdown of *STRIP2* in human cells causes a loss of filamentous actin bundles and elongation of the cell (Bai et al., 2011). Whereas knockdown of the *camkvb* ortholog in mouse neurons has been shown to reduce filamentous actin and cause a loss of dendritic spines (Liang et al., 2016). Brk1 is part of the WAVE protein complex that activates the Arp2/3 complex to induce actin polymerisation for neuronal development (Derivery et al., 2008; Korobova and Svitkina, 2008). Complete loss of Brk1 expression in a cell causes a loss of cytoskeletal structures including lamellipodia and ruffles (Derivery et al., 2008). Recently, the importance of cytoskeletal dysregulation in PD was indicated in a study of the PD-associated gene, LRRK2 (Bardai et al., 2018). The actin cytoskeleton also plays a role in the endocytic recycling of dopamine transporter (DAT) subsequent to neurotransmission across the synapse (Gabriel et al., 2013). DAT is involved in the reuptake of DA from the synaptic cleft and its endocytic recycling regulates DA neurotransmission (Sorkina et al., 2009). Another differentially expressed gene, *vamp3*, is part of a SNARE complex, found on the vesicle membrane and required for vesicular transport from early/recycling endosomes to the trans Golgi network (Mallard et al., 2001). The substantia nigra DA neurons in humans have increased vulnerability to dysfunctional intracellular trafficking, as vesicles transport cargo great lengths and to a high number of synapses (Bolam and Pissadaki, 2012).

Mitochondrial dysfunction is accepted as central to PD pathophysiology (Hu and Wang, 2016) and in the transcriptomic analysis, glycine amidinotransferase (*gatm*), a mitochondrial enzyme involved in creatine synthesis (Sandell et al., 2003), was found significantly downregulated in the *dj-1*$^{-/-}$ mutant brain. As well as being an energy store, creatine can also play a role in neurotransmission binding to both the GABA-A and NMDA receptors and has an anti-oxidative

function scavenging for reactive oxygen species (Koga et al., 2005; Royes et al., 2008; Lawler et al., 2002). A further gene involved in supplying the cell with energy was found with altered expression in the *dj-1*$^{-/-}$ zebrafish brain. The ortholog of pyruvate kinase M1/2 (*pkma*), a key regulator of aerobic glycolysis, was downregulated reflecting dysfunctional metabolism in the *dj-1*$^{-/-}$ brains consistent with a parkinsonian disorder (Anandhan et al., 2017; Stone et al., 2018; Israelsen and Vander Heiden, 2015).

The differential expression in the *dj-1*$^{-/-}$ zebrafish brain was further compared with gene expression in the SN DA neurons, altered expression in the SN of PD patients and differential gene expression in the brain of a separate *dj-1*$^{-/-}$ zebrafish line. Previous transcriptomic analyses have found that genes associated with metabolism, mitochondria and vesicle trafficking have higher expression in the SN DA neurons than other DA neurons (Chung et al., 2005; Greene, Dingledine and Greenamyre, 2005). These results are indicative of the high-energy requirements of the SN DA neurons (Pissadaki and Bolam, 2013), and their reliance on effective vesicle trafficking due to extensive axonal arborisation (Giguère et al., 2019). Specific genes identified with higher expression in the SN DA neurons include the glycolytic enzyme pyruvate kinase, mitochondrial enzyme *Gatm,* reducing enzyme glutathione peroxidase 3 (*Gpx3*), a metabotropic glutamate receptor, mGluR1, and the vesicle trafficking protein *Vamp2* (Hook et al., 2018; Grimm et al., 2004; Greene, Dingledine and Greenamyre, 2005). Interestingly, pyruvate kinase, *gatm, gpx3,* the metabotropic glutamate receptor, mGluR2, and the vesicle trafficking protein *vamp3* all had reduced expression in the brain of the *dj-1*$^{-/-}$ zebrafish described in this thesis (Table 4.1). mGluR1 and mGluR2 are both metabotropic glutamate receptors and increased activation of the SN DA neurons by excitatory glutamatergic input has been associated with the neurodegeneration in PD (Marino et al., 2001). Furthermore, mGluR2 is a glutamate receptor that has an inhibitory effect on activation of the SN DA neurons (Bradley et al., 2000), therefore its loss could increase the chances of SN DA over excitation. VAMP2 and VAMP3 are both SNARE proteins involved in the docking and exocytosis of synaptic vesicles (Hong, 2005). The reduced expression of genes associated with the SN DA neurons could reflect a loss of SN DA neurons in the *dj-1*$^{-/-}$ zebrafish brain. Transcriptomic analyses comparing gene expression in the SN

of PD patients and healthy controls postmortem identified decreased expression of genes associated with ubiquitination, vesicle trafficking, the mitochondria and cytoskeletal maintenance (Hauser et al., 2005; Miller et al., 2006). Specific genes with reduced expression in the SN of PD patients included the SNARE proteins, *VAMP1* and Syntaxin-1A. Contrasting with the SN of PD patients, an E3 ubiquitin ligase, *rnf144ab* , was upregulated in the *dj-1*$^{-/-}$ mutant brain, which could increase ubiquitination activity (Table 4.2) (Ho et al., 2014). However, there was decreased expression of *vamp3*, which is a SNARE protein, the mitochondrial enzyme *gatm,* and altered expression of genes associated with cytoskeletal maintenance (*brk1*, *strip2* and *camkvb*) (Table 4.1 and 4.2). The differential expression in the *dj-1*$^{-/-}$ zebrafish brain therefore shares some similarities with gene expression in the SN of PD patients. Differential gene expression has also been studied in the brain of a separate *dj-1*$^{-/-}$ mutant line by mass spectrometry (Edson et al., 2019). Proteins in the electron transport chain (ETC) were seen upregulated whilst those involved in inflammation, the stress response and vesicle trafficking were downregulated. No ETC genes were differentially expressed in the *dj-1*$^{-/-}$ line described in this thesis, however, the mitochondrial enzyme *gatm,* involved in creatine synthesis (Sandell et al., 2003), was downregulated. Similarly, creatine kinase, a mitochondrial enzyme which phosphorylates creatine (Lahiri et al., 2002), had reduced expression in the *dj-1*$^{-/-}$ line produced by Edson et al. (2019). Inflammation associated genes had no significantly altered expression in the *dj-1*$^{-/-}$ line generated for this project, but there was reduced expression of a stress response gene, *gpx3,* and a vesicle trafficking gene, *vamp3.* Interestingly, *vamp3* is a SNARE protein and one of the genes downregulated in the *dj-1*$^{-/-}$ line reported by Edson et al. (2019), *picalma* has a SNARE binding activity (Miller et al., 2011). Overall, when comparing the differential expression in the two *dj-1*$^{-/-}$ mutant lines, they shared reduced expression of mitochondrial enzymes involved in creatine synthesis/phosphorylation, genes involved in the stress response and genes associated with vesicle trafficking. The two *dj-1*$^{-/-}$ lines differed when it came to the expression of ETC genes and inflammation associated genes.

4.3.3 Enrichment of gene sets implicated in PD

RNA-Seq is useful for identifying significant changes in gene expression, however, it can be difficult to interpret the biological significance of each differentially expressed transcript. Furthermore, DJ-1 regulates signalling pathways and the RNA-Seq data alone was not able to convey whether specific pathways had been affected by the loss of DJ-1 (Oh and Mouradian, 2017). In order to take a broader look at the biological functions sensitive to a loss of DJ-1, Gene Set Enrichment Analysis (GSEA) was used to analyse the RNA-Seq data (Subramanian et al., 2005). GSEA identified enrichment of gene sets consistent with disrupted metabolism, Akt signalling and cell cycle regulation (Figure 4.5).

Metabolic regulation in neurons is crucial given the high and dynamic energy demands of the vertebrate brain (Magistretti and Allaman, 2015). These demands in normal neurons can result in a transient uncoupling of glycolysis and oxidative phosphorylation (Yellen, 2018), similar to the Warburg effect seen in cancer cells (Koppenol, Bounds and Dang, 2011). The increase in glycolysis over oxidative phosphorylation enhances carbon and NADPH production, which are both required to create the amino acids, nucleotides and other components needed for proliferation in cancer cells (Heiden, Cantley and Thompson, 2009). Cancer cells can therefore prioritise proliferation over ATP production through the Warburg effect. Similarly, neurons are thought to use the Warburg effect to supply biosynthetic components for synaptic growth, and supporting this increased levels of aerobic glycolysis have been reported in areas of the brain with heightened expression of genes involved in synapse formation (Goyal et al., 2014). The Warburg effect is also less efficient at producing ATP than glycolysis coupled to oxidative phosphorylation, however, it requires less steps to produce ATP. Therefore, the increase in aerobic glycolysis with neuronal excitation is thought to rapidly supply neurons with the energy they need once stimulated before switching back to the more efficient oxidative phosphorylation (Manlio Díaz-García et al., 2017). Here, in the *dj-1*$^{-/-}$ zebrafish brain the gene set for oxidative phosphorylation was enriched (Figure 4.5) whilst the *pkma* gene encoding a major enzyme in glycolysis was downregulated (Figure 4.4). This suggests a dysregulation of the mechanism to transiently uncouple

glycolysis and oxidative phosphorylation and disruption of the responsive metabolism required by neurons in zebrafish brains lacking DJ-1. A western blot analysis was carried out to determine whether the enrichment of oxidative phosphorylation was due to an altered number of mitochondria in the *dj-1*$^{-/-}$ brain. However, no significant difference was seen in the expression of mitochondrial marker ATP5F1B between *dj-1*$^{-/-}$ and wild type sibling brains at 16 wpf (Figure 4.5C). Interestingly, proteomic analysis of a separate *dj-1*$^{-/-}$ zebrafish mutant revealed increased expression of complex I components, which are involved in oxidative phosphorylation, but reduced complex I activity (Edson et al., 2019). Overall, the disrupted metabolism suggested by the transcriptomic data may create a neurodegenerative state in the *dj-1*$^{-/-}$ brain.

The enrichment of gene sets associated with PI3K/AKT/mTOR and mTORC1 signalling (Table 4.3) is similar to previous reports of altered AKT and mTOR signalling in PD (Lan et al., 2017; Timmons et al., 2009). Increased levels of mTOR have also been seen in mouse models of PD (Wills et al., 2012) and dysfunctional AKT signalling was observed in the *Dj-1* knockout mouse (Hauser et al., 2017). DJ-1 is also known to negatively regulate the tumour suppressor PTEN, which is part of the Akt signal transduction pathway (Kim et al., 2005). A western blot analysis was carried out to determine whether a loss of DJ-1 increased the phosphorylation of Akt in the zebrafish brain (Figure 4.5C). However, no change in P-Akt was observed when comparing *dj-1*$^{-/-}$ zebrafish and wild type siblings at 16 wpf.

Three cell cycle-related gene sets were enriched in the *dj-1*$^{-/-}$ zebrafish brain: G2M checkpoint, E2F Targets and Mitotic Spindle. The G2M checkpoint prevents a cell entering mitosis with damaged DNA (Stark and Taylor, 2004), E2F are a family of transcription factors involved in the G1/S phase of the cell cycle (Bertoli, Skotheim and De Bruin, 2013), and mitotic spindle is required to separate sister chromosomes during mitosis (Walczak and Heald, 2008). These enriched gene sets suggest an increase in the number of cells actively cycling in the *dj-1*$^{-/-}$ brain. This may seem surprising, since DJ-1 was first identified by its interaction with c-Myc and indeed can transform 3T3 cells when co-expressed with c-Myc or ras (Nagakubo et al., 1997). While this suggested DJ-1 can promote the cell cycle, the GSEA indicates enrichment of cell cycle genes

in brains lacking DJ-1. Interestingly, entering the cell cycle has been linked to cell death in neurons (Herrup et al., 2004; Herrup and Yang, 2007; Wang et al., 2009) and moreover, Höglinger et al. (2007) found that the pRb/E2F pathway was activated and caused neuronal cell death in a chemical model of PD. Similarly, in a chemoablation model of PD in rats, dying substantia nigra DA neurons showed markers for the S phase and G2 phase of the cell cycle (El-Khodor et al., 2003). There has been some speculation about the underlying cause of neurons re-entering the cell cycle during neurodegeneration. Kruman et al. (2004) proposes that re-entry of postmitotic neurons into the cell cycle is an essential part of their DNA-damage response, which ends with apoptosis after reaching a cell-cycle checkpoint. Oxidative stress, which DJ-1 acts as a sensor for, is known to cause DNA damage (Barzilai and Yamamoto, 2004). The mitotic gene sets enriched in mutant samples could therefore reflect a neurodegenerative state in the *dj-1*$^{-/-}$ brain. Alternatively the enrichment of cell cycle-related gene sets may be due to increased proliferation of progenitor cells to replace the neurons lost or it could be a sign of cancer development in the brain (Evan and Vousden, 2001; Kroehne et al., 2011). These possible reasons for the enrichment of cell-cycle related gene sets in the *dj-1*$^{-/-}$ brain will be discussed in further detail in Chapter 6. A qRT-PCR analysis was carried out to see whether there was altered expression of any cell cycle markers in the *dj-1*$^{-/-}$ brain, however, no significant difference was seen in the whole brain (Figure 4.5C). This could be due to cell heterogeneity in the whole brain lysate masking an increase seen in a certain population.

4.4 Conclusion

Using transcriptomic analysis 32 transcripts were found differentially expressed in the DJ-1 deficient zebrafish brain including genes involved in the oxidative stress response and metabolism. These were particularly interesting as DJ-1 promotes antioxidant gene expression and the DA neurons have extremely high energy demands a(Bolam and Pissadaki, 2012; Ariga et al., 2013). Loss of DJ-1 in the zebrafish brain was further shown to enrich gene sets involved in oxidative phosphorylation, the cell cycle, and Akt signalling; reflecting some known parkinsonian defects and indicative of others . Overall, the transcriptomic

analysis revealed effects on gene expression consistent with PD and, together with a loss of DA neurons in the posterior tuberculum, data indicates that creating a DJ-1 null zebrafish has provided an animal model with molecular and physiological hallmarks of Parkinson’s disease.

5 Analysis of the movement phenotype

5.1 Introduction

5.1.1 Classifying by machine learning

Machine learning is a growing strand of artificial intelligence, that is, software that can modify its own behaviour to better suit the requirements of a given task (Alpaydin, 2016). With machine learning, instead of writing a program to solve a particular problem, a solution is generated automatically over a number of iterations (Tarca et al., 2007). One criterion for selecting the machine learning method was that it would be able to learn to classify the mutant as distinct from wild type fish based on features of movement. In order to do this the algorithm had to know what data came from which class of fish. This is called 'supervised learning', a type of machine learning where a supervisor gives the desired output for every input when trying to generate an optimised classifier (Ayodele, 2010). In this instance, the input is all the features of movement extracted from the recording of a fish and the desired output would be the class of fish, either mutant or wild type. There are multiple different types of supervised machine learning, all with their own representation of models, the criteria they use for optimising the model and the way in which parameters are adjusted in the optimisation process (Alpaydin, 2016). Machine learning can be further separated into black box or white box models depending on the transparency of the classifier generated in regards to its inner workings (Olden and Jackson, 2002; Dreiseitl and Ohno-Machado, 2002).

5.1.2 Artificial neural networks: a black box classifier

Artificial neural networks (ANN) are a widely used form of supervised machine learning inspired by the human brain (Jain and Mao, 1996). The following description of evolving solutions using an ANN is informed by Alpaydin (2016), Ding et al. (2013) and Yao (1999). Each node in an ANN represents an artificial neuron and the neurons form connections with each other similar to synapses in the brain, which is the model representation (Figure 5.1). The neurons form connections with different weightings determining the effect they have on subsequent neurons and they can be excitatory or inhibitory, having a positive

or negative weighting. A neuron becomes active when the sum of all weightings acting on it reaches a threshold and it produces an output. The weightings are adjusted by a learning algorithm to reduce the number of differences occurring between outputs and the answers provided by the supervisor, termed the error function. Therefore, weightings are the adjustable parameters and reducing the error function is the criteria for optimising. However ANNs, like many forms of machine learning are 'black box' systems which means they do not share much insight on the internal functions of the classifier generated (Olden and Jackson, 2002). This makes ANNs and any other black box machine learning methods unviable for the purpose of identifying and describing a movement phenotype.

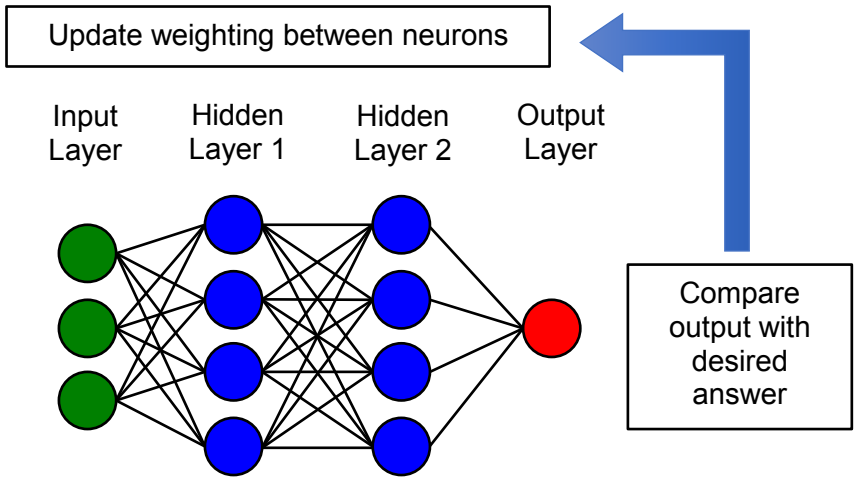


Figure 5.1 Artificial neural networks
Example of an ANN representation. The hidden layers contain the processing neurons, which can be excitatory or inhibitory and need to have an input above their specific thresholds to be activated. The output is compared with the desired answer to determine the error function and the weightings between neurons are updated to reduce the difference.

5.1.3 Evolutionary algorithms: a white box classifier

Evolutionary algorithms (EA) are another form of supervised machine learning inspired by Darwinian evolution (Smith et al., 2015). The following description of evolving solutions using an EA is informed by Eiben and Smith (2015), Smith et al. (2015, 2007) and Vikhar (2016). Due to the occasionally confusing use of biological terms to describe EAs, italics will be used to indicate when a word has an alternative meaning. An EA is a *population*-based algorithm designed to optimise the candidate solution to a defined problem. In an EA a *population* of candidate solutions or *individuals* is evaluated using a *fitness* function, an objective mathematical measure that determines which solution is the most effective, and that *individual* is used to spawn a new *population* for the next generation. The EA goes through multiple generations and the candidate solution becomes increasingly optimised with each round of evolution. This culminates in an effective solution being produced and expressed as a discrete mathematical equation that describes the solution. One advantage of EAs is their simplicity compared to other types of machine learning: they are initialised, there is iterative variation and selection is based on a fitness function (Sivanandam and Deepa, 2008). They also perform a global search in the space of candidate solutions to a problem, rather than a local search, making them more likely to find an optimal solution but taking more steps to get there (Maimon and Rokach, 2008). EAs can be applied to a broad range of problems, anything where a function needs to be optimised and can be represented by operators that are varied with each generation (Sivanandam and Deepa, 2008). They can also be dynamic, adapting to changes in circumstance whereas other methods might need starting over (Branke, 2012). Furthermore, EAs are useful for modelling complex, poorly understood processes such as neurological disorders where an EA can recognise minute differences that human observation and logic might overlook (Lones et al., 2014). Importantly, the internal functions applied to the input data in a classifier evolved through an EA can be seen (Sette and Boullart, 2001). This is essential for discovering the movement phenotype that characterises a zebrafish model of PD.

5.1.4 Evolutionary algorithms for diagnosis of PD

EAs have already been proven successful at classifying PD patients based on movement (Lones et al., 2014). The clinical diagnosis of PD by movement disorder experts has only been accurate around 80% of the time (Hughes et al., 1992; Rizzo et al., 2016). There is a need for a more accurate, less subjective method of diagnosing PD. In an attempt to diagnose PD in an automated manner 49 PD patients and 41 age-matched controls wore electromagnetic tracking sensors on the finger and thumb and performed a finger tapping task (Lones et al., 2014). Simple comparison of the gross features of movement between PD patients and controls showed that the mean amplitude was the most effective feature for diagnosing PD with an accuracy of 78%. However, this was no more effective than a clinician's diagnosis, so classifiers were then optimised using EAs and the same data was used to train them. The best classifiers produced had diagnostic accuracies of around 90%. EAs have therefore been successful at optimising classifiers for diagnosing PD, with accuracies higher than that of clinician diagnosis. The discrete equations generated by the best classifiers also managed to describe the movement characteristics of PD in the finger tapping exercise. Classifiers have been evolved here to identify mutant zebrafish based on data that depicts the swimming movement. The data was processed to make it suitable for computational analysis similar to that used to diagnose PD in humans.

5.1.5 Aims of this chapter

Previous work modelling PD in zebrafish and analysing the movement phenotype has focused on swimming at the larval stage (Flinn et al., 2009; Xi et al., 2010; Zhao et al., 2012). The features of movement studied and the methods used to measure them have been subjective. Subjective assessments of zebrafish movement have already been shown to differ widely between individuals (Mirat et al., 2013), whereas objective methods increase reproducibility and can be used to identify novel features of movement (Lones et al., 2014; Smith et al., 2007). Additionally, studying the locomotion of adult zebrafish modelling PD would be more suitable due to the progressive neurodegenerative nature of the disorder (Greffard et al., 2006). In order to

better describe the movement phenotype characterising PD in zebrafish, objective machine learning methods were used to classify adults models. Evolutionary algorithms were chosen as the type of machine learning to optimise the classifiers.

The aims of this chapter are:

- To compare the features of movement in zebrafish modelling PD with age-matched controls.
- To evolve classifiers that recognise zebrafish modelling PD based on their extracted features of movement.
- To evolve sliding window classifiers that identify zebrafish modelling PD from the raw movement data.

5.2 Methods

5.2.1 Recording tank design

An Aquatics Habitat breeding tank was modified to record zebrafish swimming whilst keeping the fish in frame and eliminating any reflection or shadow. To achieve this a white plastic insert of 145mm x 90mm was designed to fit the bottom of the tank and a hollow square frustum with openings at the top and bottom was created for the fish to swim in whilst allowing a camera to record from above (Figure 5.2A). The insert permitted background subtraction whilst the frustum kept the fish in frame and reduced reflections and shadows. A raised platform was added to the lid to hold the camera in place and it was adjusted to the correct height to keep fish in frame at all times.

5.2.2 Zebrafish recording protocol

Individual fish were filmed by a GoPro Hero 3 camera positioned above the recording tank (Figure 5.2B) for 5 minutes after moving the fish into the tank and allowing a 10-minute acclimation time. All recordings were at a rate of 100 frames per second with a resolution of 1280 x 720 pixels. The zebrafish recorded were known mutants and age-matched wild type controls. This allowed a greater number of mutant and wild type controls to be recorded. If the progeny of a single heterozygous mutant incross were used, only 25% would be homozygous mutant and 25% would be wild type according to mendelian inheritance (Harel et al., 2015). Heterozygous mutants were avoided in case of any intermediate movement phenotype differing from the wild type. Importantly, using mutants and wild type fish from multiple progenies reduced the chances of evolving a classifier that recognised a pattern occurring in a single progeny. Similarly, Lones et al. (2014) used PD patients and unrelated age-matched controls to evolve effective classifiers of PD and they avoided smaller sample sizes to prevent the classifiers responding to a pattern over-represented in the sample. Furthermore, the classifiers evolved would have increased generality, from training with mutants and wild type fish from multiple progenies. The high

generality should improve classification accuracy when the classifiers are used to discriminate future mutants.

5.2.3 Processing zebrafish recordings

A 30-second period from each 5-minute recording was processed using fish tracking software developed by a previous student, Matthew Bedder. The fish was traced in each frame, identified by background subtraction of a recording of the same tank when empty. An additional 30 seconds of recording were processed for the *dj-1*$^{-/-}$ mutants and age-matched controls to increase the amount of movement data available for evolving classifiers. The software divided the trace into 6 sections of equal length creating 5 equidistant vertices where the sections met (Figure 5.2C). The x and y coordinates at both ends of the trace and all five vertices were extracted from each frame of recording. Additionally, the angles at each vertex, created between two sections, were extracted from each frame (Figure 5.2D). Angles along the trace were assigned a positive or negative value depending on whether the angle being analysed was bending in the same direction as the previous angle. This meant that each time the tail returned from a bend on one side of the body to a neutral position where the spine was straight the signage of the angles would reset. If the tail were to then bend on the other side of the body the angles would fail to switch to the opposite signage. Therefore, it was decided that the amplitude and frequency of the tail bend would be investigated rather than the direction of tail bend, which would be less reliable. Furthermore, the size of the recording tank meant that the fish would have to be turning regularly at the corners and sides in order to keep swimming. This would introduce more tail bends on one side of the body in a row than if the fish were recorded in an open space, which could confound results. The raw data was recorded and stored in a text file in comma separated values format (CSV).

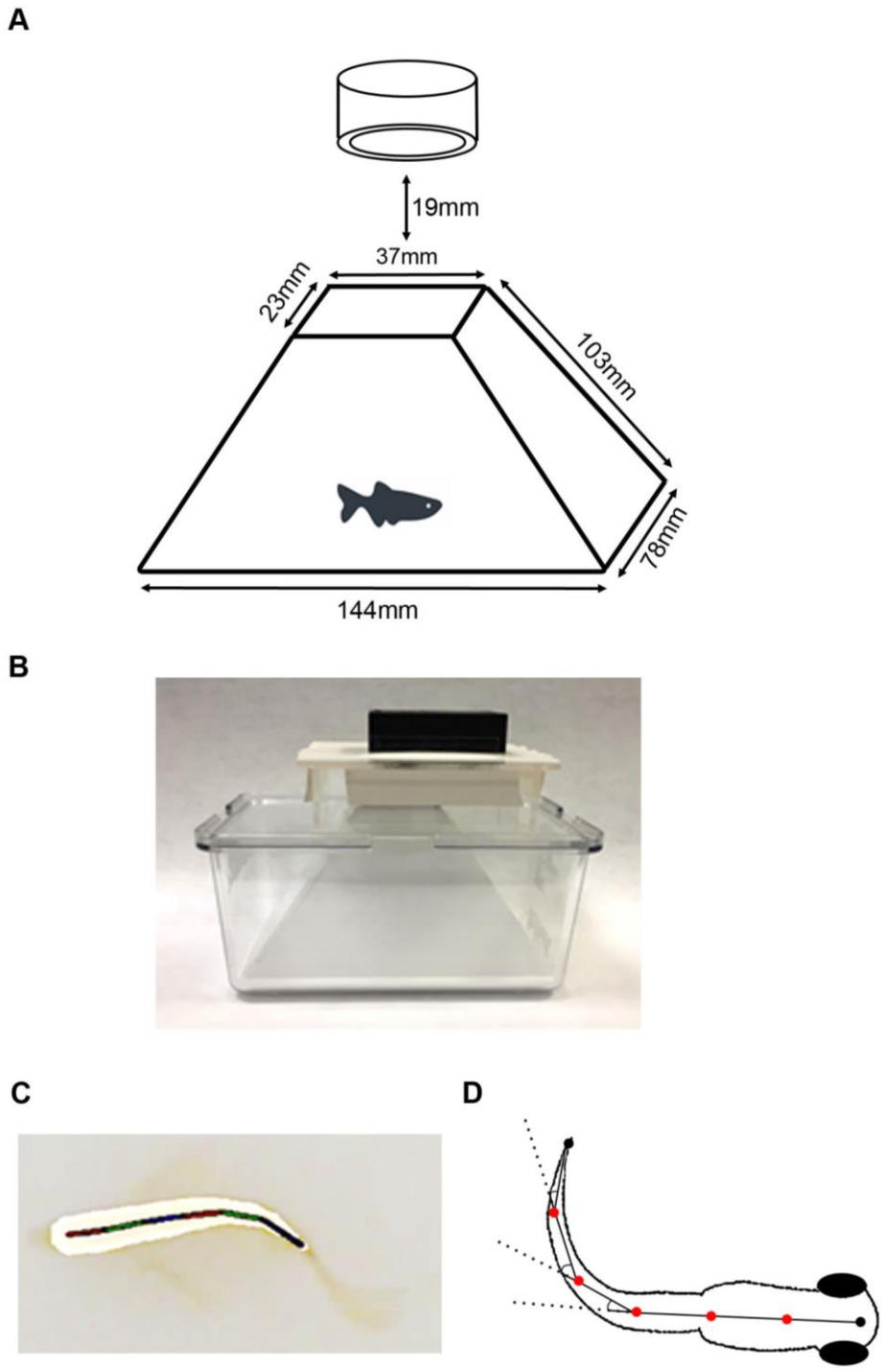


Figure 5.2 Recording tank setup and extracted movement data

(A) A diagram of the recording setup, made to fit in a breeding tank, with measurements. (B) A photograph of the recording setup in an Aquatics Habitat breeding tank. (C) A close up of a zebrafish in a recording after processing with the FishTracker software. The trace has been divided into 6 sections of equal length in the repeating colour pattern of red, green and blue. (D) A diagram of the angles measured along the zebrafish trace, at the vertices, when there is a bend in the tail.

5.2.4 Extracting the features of movement

Raw movement data from the CSV files was input into MATLAB as a matrix and functions were applied to extract features of movement. The features of movement being extracted were inspired by previous research investigating movement phenotypes in zebrafish (Table 5.1). The Pythagoras theorem ($a^2 + b^2 = c^2$) was applied to the absolute coordinates of the centre of a fish 100 frames apart in order to work out the displacement over a second. Visual assessment of the video clips showed that zebrafish very rarely covered a distance much larger than the displacement calculated for each second. Calculating the displacement for each second was therefore deemed suitable for generating an accurate total distance travelled. This method will also be more efficient to process than calculating the displacement over smaller time periods. The values calculated were then added together to get the total distance travelled in a recording. Based on previous work by Keatinge et al. (2015) the zebrafish were categorised as swimming at high, medium or low speed for each second of a recording. However, when trying to label the swimming speeds of fish preliminary testing found that zebrafish recorded in the breeding tank rarely reached speeds of >7cm/s. Consequently, the range of values for high, medium and low speed categorisation were reduced to match the speeds reached in the breeding tank. Low, medium and high swimming speeds were defined by fish travelling $0.5 \leq x < 2$cm, $2 \leq x < 4$ cm and $x \geq 4$cm over a one second period, respectively. In a previous study by Lambert, et al. (2012) a larval zebrafish was classed as swimming if it reached speeds of >3mm/s and this was used as a criteria when measuring the duration of swimming episodes. Similarly, speeds of >5mm/s were used as a criteria in this study for measuring the duration of swimming episodes, scaled up slightly for adult zebrafish. The percentage of time spent moving, inspired by Godoy et al. (2015), was calculated using the same criteria. The velocity in this study was the total distance travelled divided by the number of seconds spent swimming, similar to the swim speed used by Ingebretson and Masino (2013). Budicke et al. (2000) studied tailbeat frequency in larval zebrafish by going through recordings frame by frame and subjectively identifying the frames where the tail was at maximum bend. Tailbeat frequency was given as the number of tailbeat cycles occurring per second and a tailbeat cycle consisted of a tail bend on both

sides of the larvae. To identify maximum tail bends in this study, the sum of angles along a zebrafish trace were plotted over time with peaks on the graph indicating the timepoints of maximum tail bend (Figure 5.3A). A running average over 5 frames was used to remove noise from the data and a minimum distance between peaks and peak prominence of 5 frames/degrees was used to prevent multiple peaks being identified with the occurrence of a single tail bend. In order to include any unilateral tail bends that may be useful in discriminating a movement phenotype the tailbeat frequency calculated in this study was the number of maximum tail bends occurring per second. Tail bend amplitude was another feature of movement inspired by Budick et al. (2000). In the previous work tail bend amplitude was calculated by subtracting the head-tail angle at maximum tail bend from 180°. The head-tail angle being the angle between tangents drawn at the most rostral and most caudal parts of the curved larval body (Figure 5.3B). In this study, the sum of angles along the zebrafish trace (Figure 5.3C) was given as the tail bend amplitude, which should be a more accurate measure of the curvature of the zebrafish tail. Furthermore, the software identifying angles along the zebrafish trace will be more objective than a human measuring an angle between tangents drawn on top of a video frame (Budick and O'Malley, 2000). Overall, 12 features of movement were generated for evolving classifiers: distance travelled, velocity, time spent moving, mean swimming episode duration, mean tailbeat frequency, mean tail bend amplitude and the tailbeat frequencies and tail bend amplitudes at low, medium and high speed.

Feature	Based on	Stage	Principles in previous work	Principles used here
Distance travelled	Keatinge, et al., 2015	Adult	Total displacement of zebrafish over recording (taken side on)	Total displacement of zebrafish over recording (taken from above)
Velocity	Ingebretson and Masino, 2013	Larval	Centre of mass used to determine velocity in swimming episodes	Distance travelled divided by number of seconds spent swimming
Time spent moving	Godoy et al., 2015	Larval	Percentage of time spent moving over a 10-minute period	Percentage of time spent swimming at ≥ 5mm/s
Low/medium/high speed swimming	Keatinge, et al., 2015	Adult	Low speed: <5cm/s Medium speed: 5<x<7cm/s High speed: >7cm/s	Low speed: 0.5cm<x<2cm/s Medium speed: 2<x<4cm/s High speed: >4cm/s
Mean swimming episode duration	Lambert et al., 2012	Larval	Speed of ≥ 3mm/s for swimming episode	Speed of ≥ 5mm/s for swimming episode
Tail bend amplitude	Budick et al., 2000	Larval	Angle between rostral and caudal tangents subtracted from 180°	Sum of angles along zebrafish trace at maximum tail bend (low/medium/high speed).
Tailbeat frequency	Budick et al., 2000	Larval	Frequency of tail beat cycles in a second	Frequency of maximum tail bends in a second (low/medium/high speed).

Table 5.1 Features of movement based on previous works

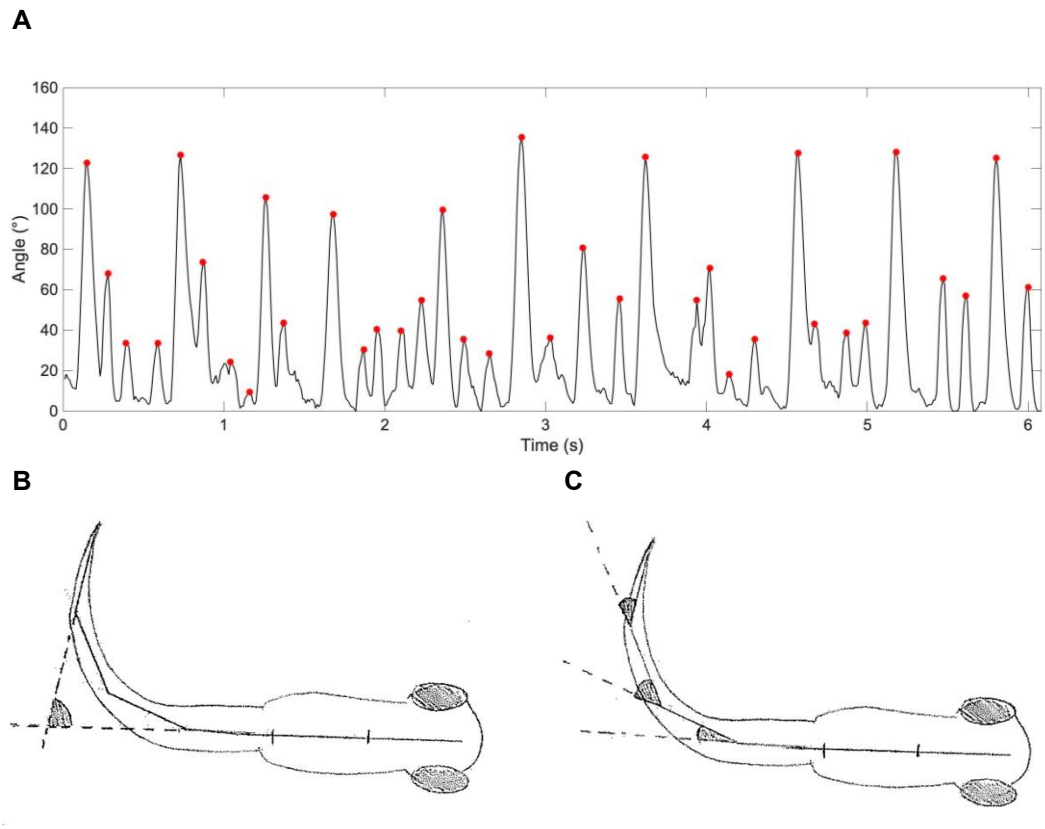


Figure 5.3 Calculating tailbeat frequency and tail bend amplitude

(A) A graph plotting the sum of the angles along a zebrafish trace over time. The peaks (circled red) are the timepoints when a maximum tail bend has occurred. (B) A diagram of the head-tail angle measured between tangents drawn at the most rostral and most caudal parts of the zebrafish body (adapted from Budicke et al. (2000)). (C) A diagram to show the angles captured along the tail of the zebrafish when bent.

5.2.5 Comparing features of movement

To assess whether the extracted features of movement were significantly different between zebrafish modelling PD and age-matched controls the correct statistical test had to be used. The features of movement calculated were examined using the Shapiro-Wilk test to see whether they followed a normal distribution. If the data followed a normal distribution the interquartile range (IQR) was used to identify suspected outliers. Any values at least 1.5 x IQR below the lower quartile or greater than the upper quartile were identified as outliers and removed prior to statistical analysis. Features were compared between classes using either a Student's t-test if the data was normally distributed or a Mann-Whitney U test if the data was nonparametric.

5.2.6 Statistical classification using features of movement

Binary logistic regression would seem an appropriate statistical classifier for recognising mutants using the features of movement. The class of fish, mutant or wild type, being the binary dependent variable and the features of movement being the independent variables (Kleinbaum and Klein, 2010). The output of a binary logistic regression model is the probability, between 0 and 1, that a set of independent variables belong to a certain class. However, there are a set of assumptions that need to be met for binary logistic regression to be suitable. Amongst these was the assumption that there should be no collinearity between the independent variables (Stoltzfus, 2011). The Belsey collinearity test was carried out on MATLAB to determine whether there was any collinearity between the features of movement. Distance travelled, mean tailbeat frequency, tailbeat frequency at medium speed and mean swimming episode duration were found to be highly correlated. These features of movements would therefore not meet the assumptions for logistic regression and their removal would leave eight remaining features of movement. A further requirement is a sample size large enough to produce an effective logistic regression model (Nemes et al., 2009). A widely used rule for estimating the minimum sample size for logistic regression is the 10 events per variable (EPV) rule where the number of independent variables is divided by the probability of the least likely class outcome, and then multiplied by 10 (Peduzzi et al., 1996).

Using the remaining 8 independent variables and a 0.5 probability of a fish being mutant, the minimum sample size would be 160. 0.5 was used as the probability because an equal number of *dj-1*$^{-/-}$ mutants to age-matched controls would be used. However, more recently a 20EPV rule has been suggested to calculate the minimum sample size for logistic regression (Austin and Steyerberg, 2017). Using this rule the minimum sample size here would be 320. In contrast, a classifier diagnosing PD has previously been evolved, using an EA, with a sample size of 22 (Smith et al., 2007). Use of the correlated independent variables would reduce the statistical power of the test (Sainani, 2010), and even with removal of these variables the available sample size would still lead to a low statistical power (Nemes et al., 2009). The desired power of 95%, which would be the probability of classifying a mutant correctly, would be unachievable using the number of available recordings to generate a binary logistic regression classifier (Hsieh, Bloch and Larsen, 1998). Overall, the correlation between independent variables, and the sample size, would make binary logistic regression unreliable for classification purposes because of a low statistical power.

5.2.7 Evolving classifiers using the features of movement

Informed by the work of Smith et al. (2015) and Lones et al. (2014) a classifier was evolved through multiple generations, using an EA (Figure 5.4). Firstly, the input data was class labelled and divided into three data sets: training, validation and test. Then an initial *population* of classifiers or *individuals* was randomly generated. The training dataset was applied to the population and the *fitness* of each *individual*, i.e. how effective they were at classifying, was evaluated using a *fitness function*. The *fitness* function used is discussed later in this chapter. The *fittest individual* was selected for and reproduced with modifications (*mutations*) producing a progeny with slight differences in the functions applied to the input data. The *individuals* of this new generation were subjected to the same process: selection of the *fittest* and reproduction with *mutation*. The process was repeated for a set number of generations (3000) or until the EA began to overfit to the training data. Overfitting is when the *individuals* evolved start to perform excessively well at classifying the training data but have reduced *fitness* when classifying the data as a whole (the

validation and test data sets) (Chicco, 2017). An overfit classifier would have a higher training score than previous generations but would have lower validation and test scores. The validation data was used to test the *fitness* of the classifier evolved at each generation as a measure of generality making sure that the EA was not overfitting to the training data. The classifier produced should correctly identify the class of any new inputs even if they vary slightly from the training data. The classifier would be overfit if it had a high *fitness* when evaluated on the training data but a lower *fitness* than other classifiers on the data as a whole (Maimon and Rokach, 2008). If the EA was found to start overfitting then training was stopped. The test set, not used in the training or validation stages, was used to get an unbiased measure of the final classifiers discriminative power. Implementation of the EA used to evolve classifiers will be described in more depth later in this chapter.

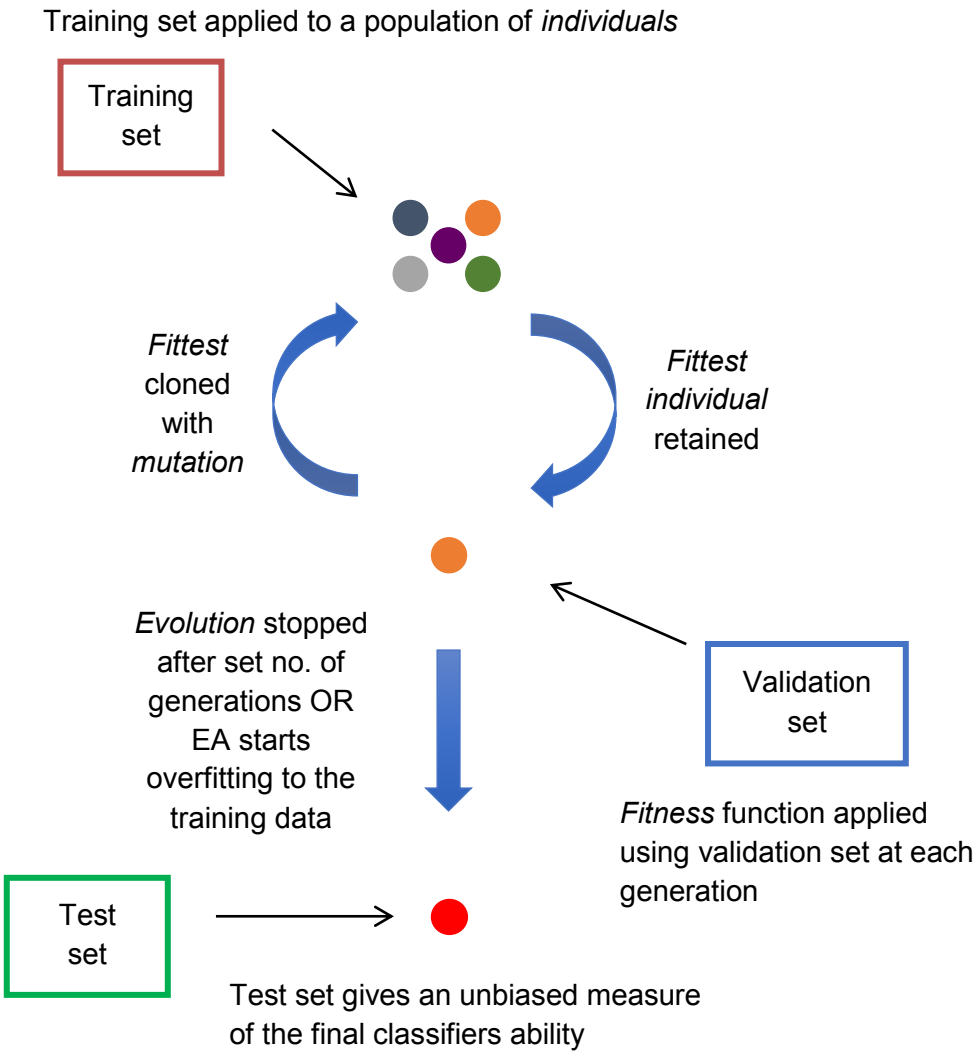


Figure 5.4 Evolution of a classifier

A diagram to show the evolution of a classifier with a standard generational EA. The initial *population* of classifiers (*individuals*) attempt to label the training data. The *fittest* (or most effective) *individual* is retained whilst the others are discarded. The validation set is used to make sure the classifier is not overfitting to the training data. This cycle continues until a set number of generations has occurred or the EA starts overfitting to the training data. The test set of data is then used to test the discriminatory ability of the final classifier.

5.2.8 Cartesian genetic programming

The type of EA used was a Genetic Programming (GP) algorithm which follows the routine of generating an initial population of individuals, selecting based on *fitness*, *mutating* to produce the next *population* and repeating the cycle until the stopping criteria is met (Maimon and Rokach, 2008). The ability of GP algorithms to apply a set of functions to input data allows them to change the representation of the original data and extract knowledge from it. Cartesian Genetic Programming (CGP) was the form of GP used where programs are represented as directed graphs with two-dimensional grids of computational nodes (Miller, 2011).

The following description of a CGP network is informed by Miller et al. (2008) and Smith et al. (2007). CGP networks consist of data input nodes, processing nodes and a singular output node, encoded as a list of integers (termed *genes)* forming a *chromosome*. The *genes* identify the input nodes that each processing node gets data from, which function (selected from a predefined set) is applied to that data and where the output data is retrieved from. The *genes* are arranged in groups of three with each triplet relating to one of the processing nodes and a final *gene* at the end of the chromosome refers to the node containing the output data. The first two *genes* in a triplet indicate the nodes containing data to be processed and the third *gene* determines the function applied. When evolving a CGP network to create a mathematical equation the arithmetic operators plus, minus, multiply and divide are commonly used as functions applied to the input data. These different functions can manipulate the input data in alternative ways to produce the output value for class labelling. For instance, multiply and divide differ from plus and minus as they can alter the sign of the input data and scale values up or down to a different degree. Including all four of these arithmetic operators allows greater manipulation of the input data, increases the number of possible classifier variations and widens the search for the most effective classifier. The chromosome that would configure the CGP network in Figure 5.5 is ‘011 112 020 233 5’. The first triplet in the chromosome relates to node 2, the initial processing node, which receives data from nodes 0 and 1 specified by those first two *genes* ‘01’ and the third *gene* ‘1’ discloses the function to be applied to

the data. The final *gene* in the *chromosome* ‘5’ reveals that the output data is retrieved from node 5. The diagram in Figure 5.5 is a simplified example of a CGP network to increase understanding and not a representation of the actual network that would be evolved using the features of movement. The actual CGP network generated to classify mutant fish would contain 12 data input nodes correlating to the 12 features of movement extracted from recordings, there would be 10 rows and 3 columns of processing nodes where selected functions are applied and a single output deciding the class label (Figure 5.6).

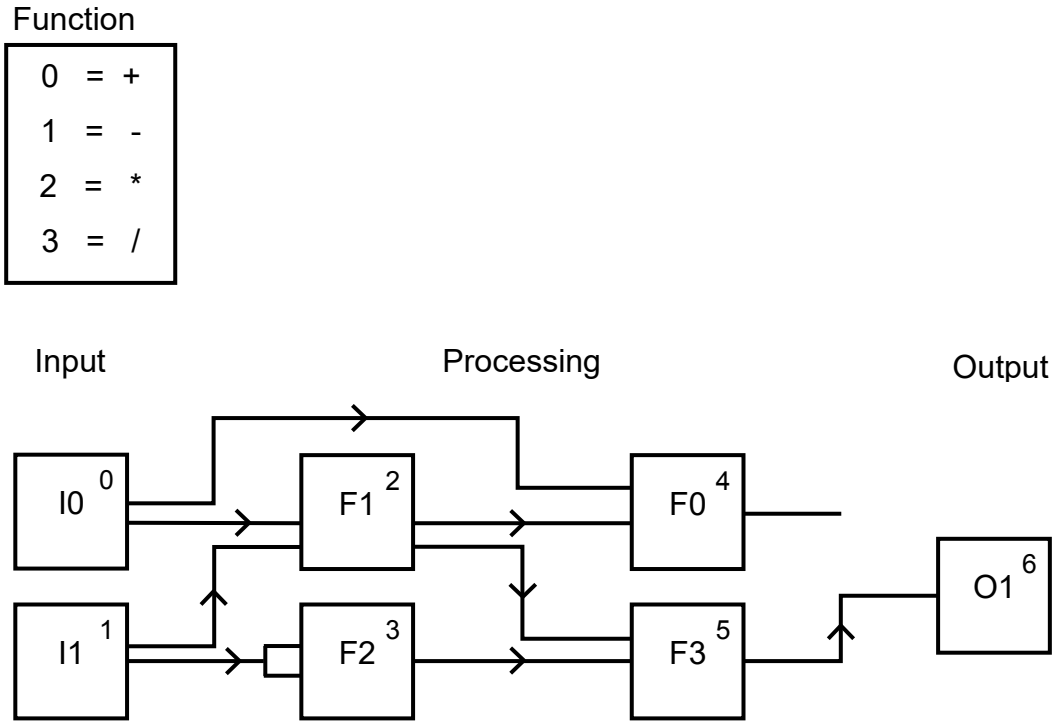


Figure 5.5 Example of a CGP network

An example of an evolved CGP network. The node number is indicated in the top right corner of each node and the node function is found in the centre. The table of functions in the top left corner determines which integers select which functions when configuring the network.

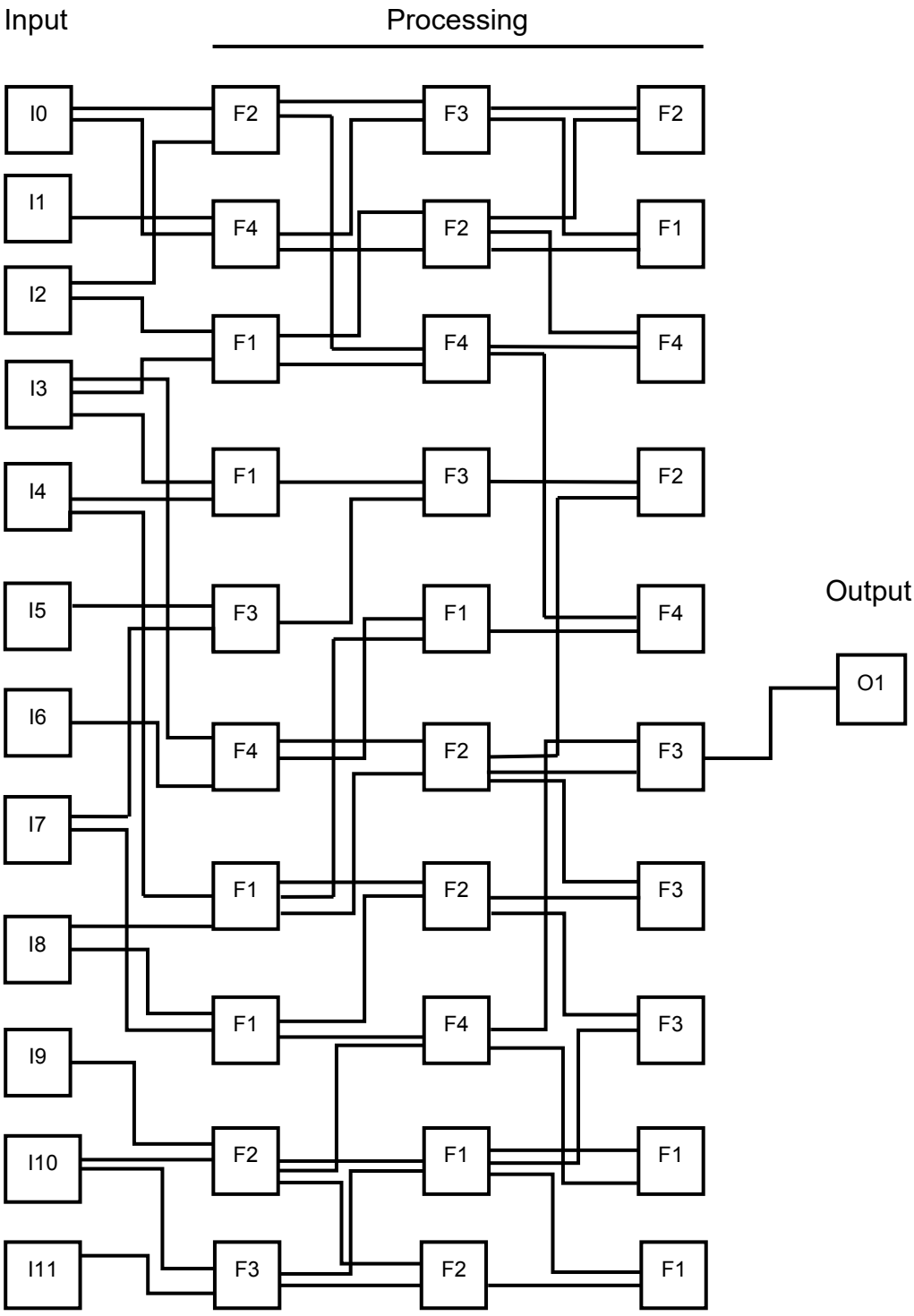


Figure 5.6 A CGP network evolved using the features of movement

An example of the evolutionary network that would be created to classify mutant lines of zebrafish with 12 inputs, 3 columns and 10 rows of processing nodes containing functions and a single output node.

5.2.9 Fitness function: area under the curve

The continuous value, between 0 and 1, output by each classifier is mapped to a class label using a threshold. Altering the threshold varies the sensitivity and specificity of a classifier; sensitivity being the true positive rate and specificity being the true negative rate (Lones et al., 2014). The sensitivity plotted against 1 - specificity produces a Receiver Operating Characteristic (ROC) curve (Bradley, 1997). An example ROC curve is shown in Figure 5.7. Each point along the solid line represents the sensitivity and specificity a classifier achieves for a certain class label threshold (from 0 to 1). The dotted line represents a classifier with no discriminative ability where for any point on the line sensitivity is equal to 1 – specificity. The area under the curve (AUC) is calculated between the solid line and the x axis, this gives a value between 0 and 1, representing the fitness of the classifier. If the curve can be represented by an equation integration can be used to calculate the area. The curve in Figure 5.7 was created using the statistics software GraphPad Prism (GraphPad Software, Inc) and the AUC can be generated in the software package. The AUC was automatically generated for each classifier evolved by the evolutionary algorithm and used to compare fitness. The AUC for the classifier with no discriminative ability (in Figure 5.7) would be equal to 0.5. A curve that reaches the upper left corner will have a high true positive rate to false positive rate ratio for most thresholds and will produce a higher AUC value (Zweig and Campbell, 1993). The area under the curve (AUC) calculated gives the probability that the classifier will rank a positive value higher than a negative value, in other terms the likelihood it correctly classifies the input data. An AUC value of 0.5 would mean the classification is down to chance so a value greater than 0.5 is desired (Fawcett, 2006). The AUC was used as the *fitness* function to select for the most effective classifier in a *population*.

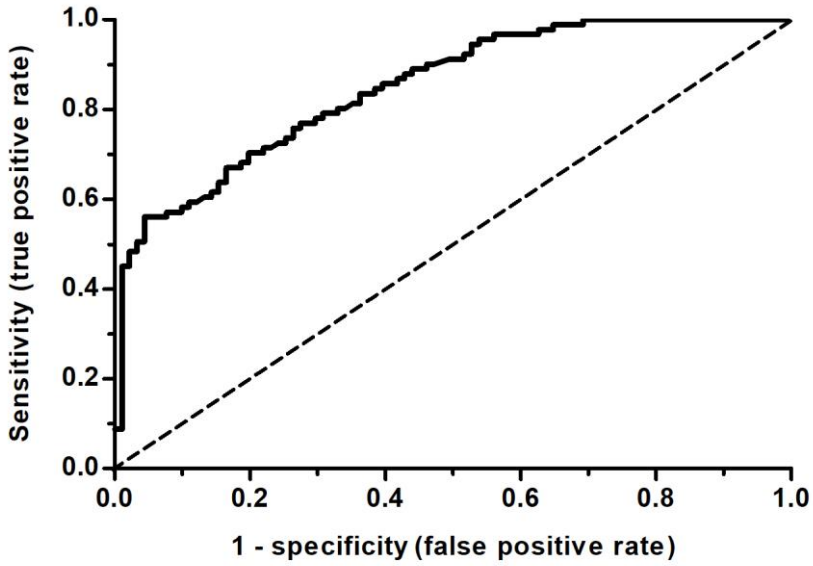


Figure 5.7 A Receiver Operating Characteristic (ROC) curve

An example of a Receiver Operating Characteristic curve used to calculate the Area Under the Curve value for a classifier. The solid line is plotted using the sensitivity and specificity achieved by a classifier with each class label threshold. The dotted line represents the sensitivity and specificity achieved by a classifier with no discriminatory ability and an AUC value of 0.5.

5.2.10 Organising data sets

The data used to train each EA was organised into three data sets using MATLAB. Extracted data was first marked to identify the class of fish (mutant or wild type) and divided in a 3:1:1 ratio between training, validation and test data sets with the data for each fish being randomly allocated between the three. Each time a classifier was evolved, the test data, specific to the fold of data sets used in training and validation, was unseen by the evolutionary algorithm during the training and validation stages. The test data was used to calculate an unbiased discriminatory score for the final classifier evolved using that fold of data sets only. The training data was normalised with regard to the mean for each extracted feature, centred to 0, and the standard deviation (SD) was scaled to ±1. The validation and test data sets were also normalised using the mean and SD obtained from the training set. The same data was randomly organised into the three data sets a further 19 times to generate 20 folds of the data. The 20-folds of data were used to get an average discriminatory score of the classifiers evolved when the same data was organised differently into the three data sets.

5.2.11 Implementing an evolutionary algorithm

A standard generational evolutionary algorithm was used to evolve the CGP classifiers. The *population* size used was 5, with a generation size of 3000, and the highest scoring *individual* in each *population*, determined using the *fitness* function, was used to spawn the next generation. The first occurrence of a classifier with the highest validation score over the 3000 generations was selected as the final classifier for that fold of datasets. This ensured that the EA wasn't overfitting the classifier to the training data and it retained generality. The test scores of all final classifiers, evolved using the 20 folds of dataset, were averaged to get the mean test score. The mean test score acted as an unbiased measure of the average classifiers discriminatory ability not dependent on the organisation of data into the three data sets. The final classifiers evolved were saved as .dot files which were loaded into Graphviz (AT&T Labs) software to visualise the CGP network evolved. The entire process from recording zebrafish swimming to evolving classifiers is visualised in (Figure 5.8).

Figure 5.8 Process of evolving a classifier

A flow chart detailing the process from start to finish of recording zebrafish swimming and evolving classifiers to recognise the movement of mutant lines. Boxes in orange were executed by writing a script in MATLAB.

5.2.12 Evolving classifiers using the raw movement data

The following work was carried out in collaboration with Dr Michael Lones from Heriot-Watt University.

5.2.13 Principal component analysis

Principal Component Analysis (PCA) was carried out on the bend angles from the raw movement data to reduce their dimensionality and transform the multivariate time series into a single time series whilst retaining any patterns in the data. Linear combinations of the angles in each frame produced new variables called principal components (PCs) that together explained 100% of the variance in the data (Ringnér, 2008). The first PC (PC1) captured only broad movements whilst the second PC (PC2) retained the more subtle movements useful for characterising a movement phenotype. PC2 was also found to be the most effective for evolving a classifier to recognise the movements of zebrafish modelling PD. To identify the PC most effective for evolving a classifier the raw angle data from each processed video clip was converted into each of the PCs. 20 folds of dataset (containing training, validation and test sets) were then organised for each of the PCs and used to evolve sliding window classifiers. The mean validation score of the 20 classifiers evolved using a particular PC was used to identify the PC most effective for evolving a classifier. The classifiers evolved using PC2 were found to achieve the greatest mean validation score. The mean weighting for the first and last bend angles in PC2 across all the fish used to evolve the DJ-1 classifier was 0.63 and 0.65, respectively. The mean weightings for angles 2, 3 and 4 were 0.26, 0.09 and 0.21, respectively. Therefore, the first and last bend angles had the greatest weightings in PC2, with relatively small weightings for the other angles, suggesting that it primarily measured the degree of head and tail deflection during swimming.

5.2.14 Evolving an ICRCGP sliding window classifier

Sliding window classifiers, encoded using Implicit Context Representation CGP (ICRCGP), were evolved to classify the PD models using the PC2 time series

as input data. IRCGP generates mathematical equations to describe local patterns in the movement data that can be used for discrimination (Lones et al., 2014). IRCGP differs from standard CGP because it has positional independence, the location of nodes within the network having no effect on the output (Cai, Smith and Tyrrell, 2005). Similar to the organisation of data for evolving a standard CGP classifier, the PC2 time series data was organised into training, validation and test data sets in a 3:1:1 ratio. Again, the training set was used to determine the most effective classifier during training, the validation set was used to stop training when a model began to overfit to the training data and the test set was used to get an unbiased measure of the final classifiers discriminatory ability. The PC2 time series data was reorganised into the three datasets 20 times to evolve multiple models and generate average training and test scores across the 20 folds of data. The population size used was 500, with a generation size of 50 and grid size of 5 x 5. The mutation rate was 6% for functions, 3% for dimensions and the mutation:crossover ratio was 50:50.

The window in a sliding window classifier contains the data from a specific range in a time series, which becomes the input data for the evolved ICRCGP algorithm (Figure 5.9). Here, 20 PC2 values equating to 20 frames in a recording were used as the input data. The algorithm applies predefined functions to the input data and produces a continuous value between 0 and 1 (Lones et al., 2014). Following this the window slides along one position in the time series to get the next overlapping range of data points. The algorithm is applied to the new input data and produces a further output value, the process is repeated until the sliding window has reached the end of the time series. The mean of all the output values is the classifier output used to label the data as belonging to one of two classes depending on a threshold value.

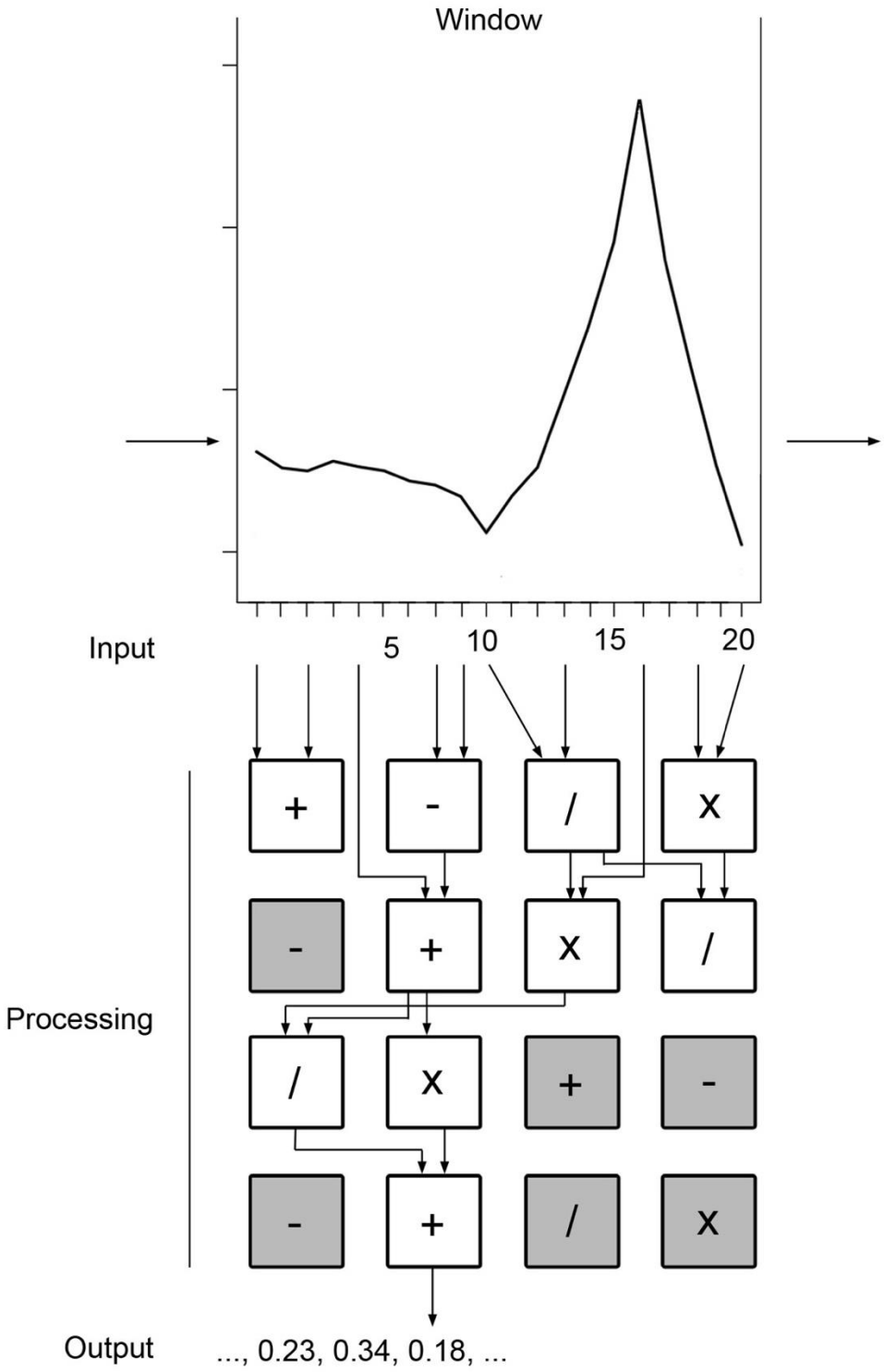


Figure 5.9 Evolving a sliding window classifier

An example diagram of an Implicit context Representation CGP network (adapted from Lones et al. (2014)). The input values are principle component values taken from a 20 frame window. Functions are applied to the input data in processing nodes and an output between 0 and 1 is generated for each 20 frame window before sliding one frame further along the time series. The ouput values are averaged across the time series and used to label the zebrafish as either mutant or wild type.

5.3 Results

The movement phenotype of the *pink1*$^{-/-}$ mutant was analysed as it was the first mutant line to reach homozygosity with the ease of genotyping *pink1*$^{-/-}$ fish. Furthermore, a previous study has reported a loss of DA neurons in the *pink1*$^{-/-}$ zebrafish (Flinn et al., 2013). The movement phenotype of the *dj-1*$^{-/-}$ mutant was analysed as it was found to be the most effective model of PD, with cellular and molecular signatures of the disease (Chapters 3 and 4).

5.3.1 Extracted features of movement

pink1$^{-/-}$ *features of movement*

MATLAB was used to extract features of movement from the raw movement data for 39 recordings of *pink1*$^{-/-}$ zebrafish at 14 wpf and 44 recordings of age-matched wild type controls. The distance travelled by *pink1*$^{-/-}$ zebrafish was approximately 10% less than wild type (Figure 5.10A). Whereas no significant difference was seen in the velocity, time spent moving or mean swimming episode duration of *pink1*$^{-/-}$ fish (Figure 5.10B,C,D). Tail beat frequency and tail bend amplitude were both reduced at medium and high swimming speeds (Figure 5.10E,F).

dj-1$^{-/-}$ *features of movement*

18 *dj-1*$^{-/-}$ zebrafish at 12 wpf and 32 age-matched wild type controls were recorded swimming for 5 minutes. Two 30 second clips were processed from each 5-minute recording and MATLAB was used to extract the features of movement from the raw movement data. The distance travelled, velocity, time spent moving and mean swimming episode duration were all significantly reduced with *dj-1*$^{-/-}$ mutants compared to wild type (Figure 5.11A-D). Tailbeat frequency was decreased at low swimming speeds in the *dj-1*$^{-/-}$ fish but remained unaltered at medium and high speeds (Figure 5.11E). The tail bend amplitude was reduced at low speeds but increased at high speeds when comparing the *dj-1*$^{-/-}$ fish to wild type (Figure 5.11F).

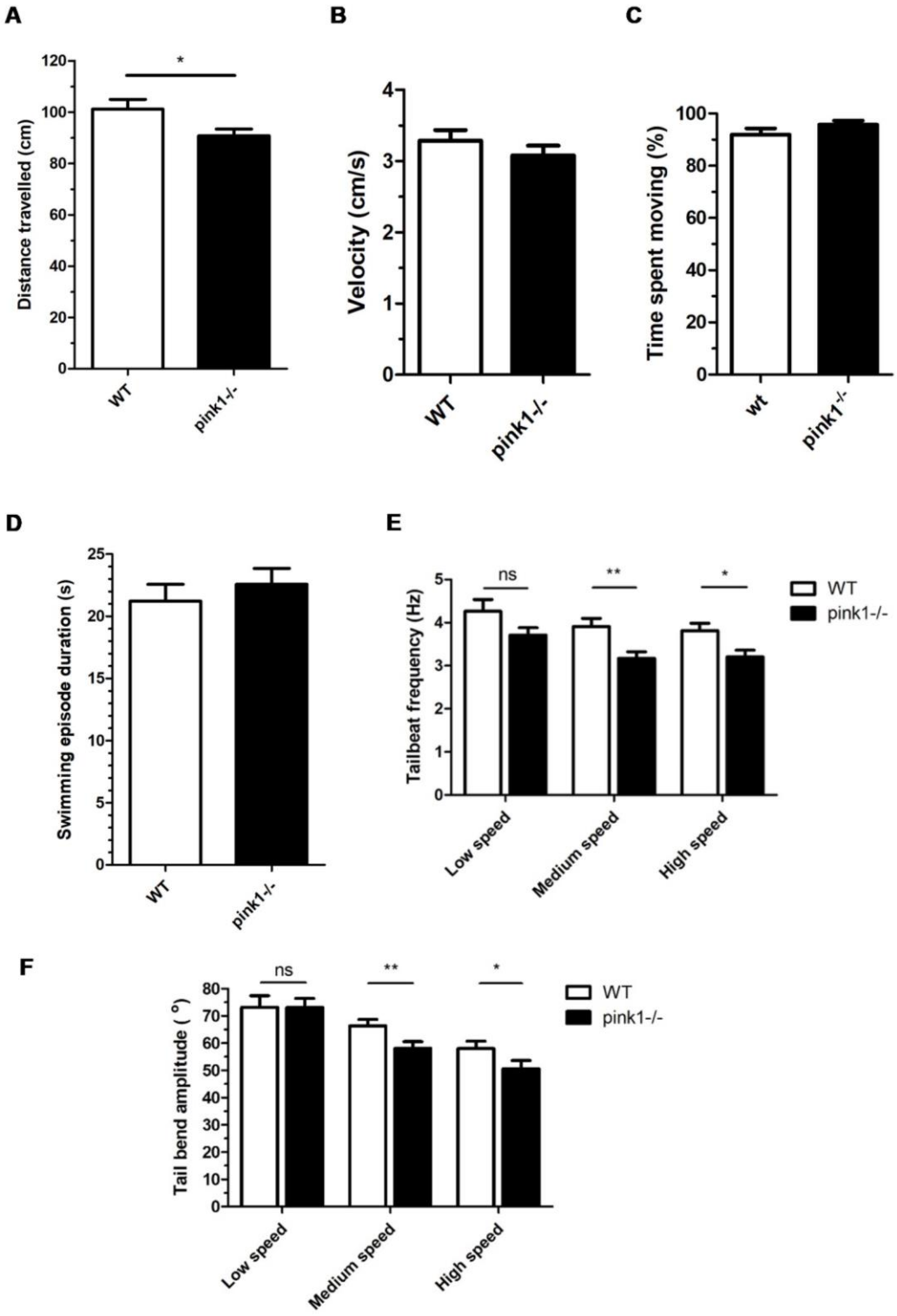


Figure 5.10 *pink1-/-* features of movement

The features of movement compared between *pink1-/-* and wild type (WT) at 14 wpf including (A) distance travelled, (B) velocity, (C) percentage of time spent moving, (D) mean duration of a swimming episode, (E) tailbeat frequency at low, medium and high swimming speeds, (F) tail bend amplitude at low, medium and high swimming speeds. * = $p < 0.05$, ** = $p < 0.01$, ns = not significant.

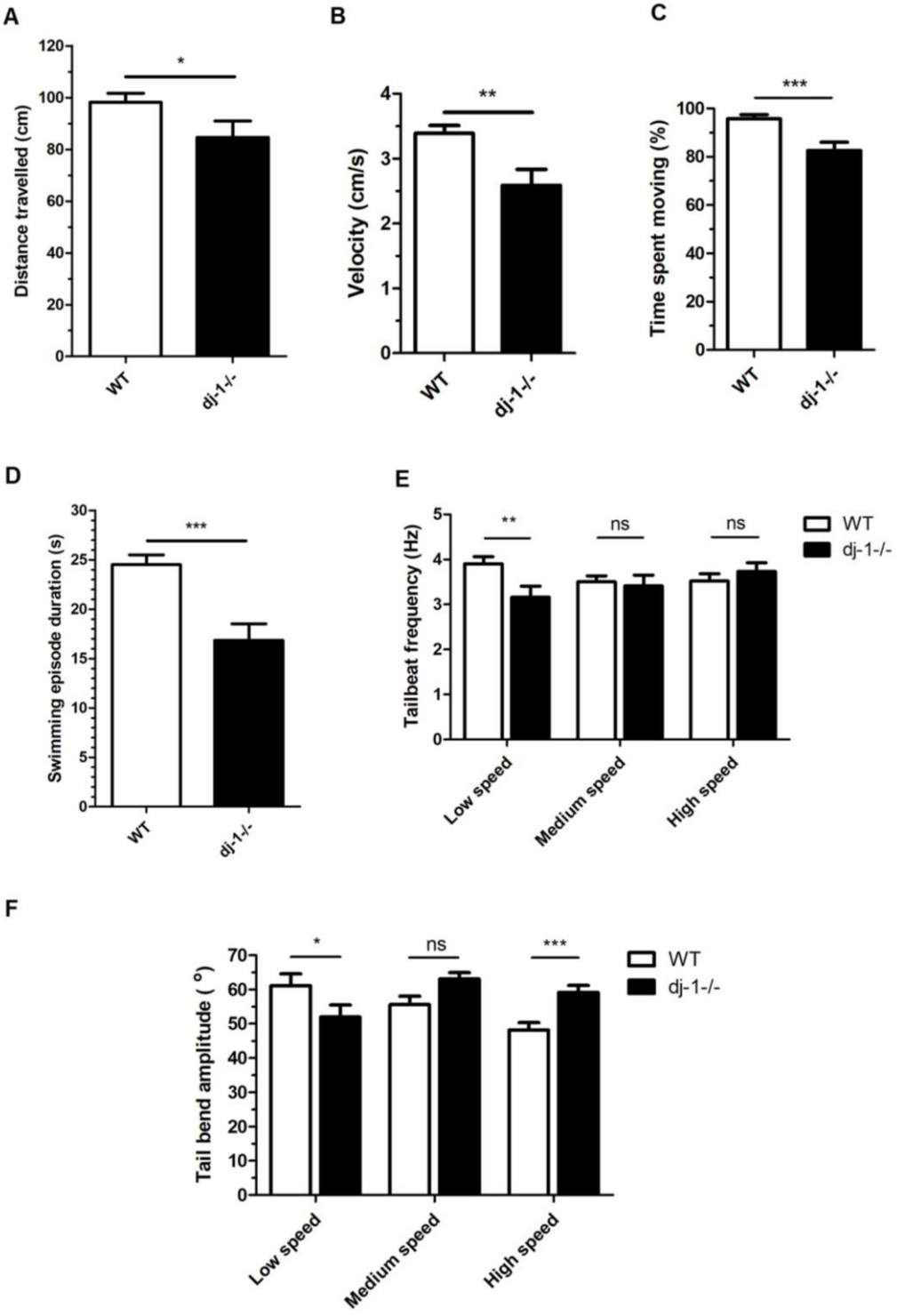


Figure 5.11 *dj-1-/-* extracted features of movement

The features of movement compared between *dj-1-/-* and wild type (WT) at 12 wpf including (A) distance travelled, (B) velocity, (C) percentage of time spent moving, (D) mean duration of a swimming episode, (E) tailbeat frequency at low, medium and high swimming speeds, (F) tail bend amplitude at low, medium and high swimming speeds. * = $p < 0.05$, ** = $p < 0.01$, ns = not significant.

5.3.2 Classifiers evolved using extracted features of movement

The average training and test scores over the 20 folds of dataset are shown in Table 5.2 for classifiers evolved to recognise zebrafish modelling PD. The results of the highest scoring classifiers, evolved using one of the 20 folds, were also recorded. An equal number of recordings (30-second clips) needed to be present in each class and each had to have a value present for the features calculated (e.g. low/medium/high speed features). Using 37 recordings of *pink1*$^{-/-}$ zebrafish at 14wpf and 37 recordings of age-matched controls the evolved classifiers averaged a test score of 51.6% across the 20 folds of dataset. The classifiers evolved using 30 recordings of *dj-1*$^{-/-}$ zebrafish at 12 wpf and 30 recordings of age-matched controls achieved an average test score of 70%. The three features of movement that occurred most often in the equations evolved to classify *dj-1*$^{-/-}$ were velocity, tail beat frequency and low speed tail beat frequency, occurring in at least 7 classifiers each. The highest scoring classifier evolved using one of the dataset folds, illustrated in Figure 5.12, achieved a test score of 92.9% demonstrating highly effective classification of *dj-1*$^{-/-}$ mutants. The number of columns used for processing nodes has been reduced from 3 to 2 in Figure 5.12. One of the three columns has been omitted as the nodes output was not used in the calculation of the output value, this is a common occurrence in CGP programs and these nodes are referred to as 'non-coding' (Miller, 2011). Although the mean swimming episode duration was significantly lower in the *dj-1*$^{-/-}$ fish compared to wild type, tailbeat frequency, which lacked a significant difference, was used in the classifier shown in Figure 5.13. This is because the classifier evolved is not a statistical test, looking for a significant difference between all of the values in one class compared to another. Instead, the classifier uses the features of movement from a single fish at a time, to calculate an output value either above or below a threshold to classify as either mutant or wild type. Therefore, the relationship between the features of movement for each fish, and how they can be manipulated to produce the correct output value is what determines the features used in the CGP network. A reason why tailbeat frequency was included in the classifier over swimming episode duration may be that for wild type and *dj-1*$^{-/-}$ fish a strong ($R = 0.7285$) and significant ($p<0.05$) correlation was seen between swimming episode duration and time spent moving using Pearson's correlation

coefficient. There will therefore be some redundancy between these two features of movement making the classifier less likely to use both in its equation. No such correlation was seen between tailbeat frequency and time spent moving. Additionally, when either class had a swimming episode duration lasting the length of the recording, they had a time spent moving value of 100%. This occurred regularly for both wild type (24/30) and *dj-1*$^{-/-}$ fish (10/30). Therefore, addition of these two features would regularly result in the same output value from processing node 1 for both classes of fish, which would not aid classification. Even if you were to change the function applied there would still be an output value from processing node 1 shared by many *dj-1*$^{-/-}$ and wild type fish. The use of tail beat frequency helps manipulate the data to increase the likelihood of a lower final output for wild type and a greater final output for *dj-1*$^{-/-}$ mutants. The output from processing node 1 had a different signage to velocity for wild type fish (13/30) more often than *dj-1*$^{-/-}$ mutants (1/29) (One *dj-1*$^{-/-}$ mutant had a velocity of zero after normalisation) (Appendix Tables 1 and 2). This meant that multiplication of the output by velocity had a higher chance of producing a negative final output value for wild type than *dj-1*$^{-/-}$. A negative final output value for a wild type fish would result in the correct classification. Furthermore, over half (9/17) of the output values from processing node 1 that shared the same signage with velocity for wild type had a value less than 1 (Appendix Table 1). Multiplication of the velocity by a value less than 1 would reduce the size of the final output value and increase the chances of a wild type classification (below the 0.5 threshold). Whereas, over half (20/28) of the output values from processing node 1 that shared the same signage with velocity for *dj-1*$^{-/-}$ fish had a magnitude >±1 (Appendix Table 2). Multiplication of the velocity by a value greater than 1 would increase the size of the final output value and increase the chances of a *dj-1*$^{-/-}$ mutant classification (above the 0.5 threshold). Furthermore, out of those fish that shared a signage between the output from processing node 1 and velocity, *dj-1*$^{-/-}$ fish had a higher proportion of velocity values with a magnitude (17/28) compared to wild type (5/17) (Appendix Tables 1 and 2). Multiplying the output value from processing node 1 by a velocity with the same signage and a magnitude > ±1 would increase the size of the final output and the chance of a *dj-1*$^{-/-}$ classification (above the 0.5 threshold). This further explains why the + and * operators were used in this classifier. Addition between time spent moving and tail beat frequency more frequently resulted in

a value > ±1 and a signage the same as velocity for *dj-1*$^{-/-}$ fish. The opposite can be said for wild type fish. Multiplication of the output value from processing node 1 by velocity more frequently produced a positive value and increased its magnitude for *dj-1*$^{-/-}$ fish compared to the wild type. Together these functions increased the likelihood of a final output value less than 0.5 for wild type and greater than 0.5 for *dj-1*$^{-/-}$ mutant fish. Overall, swimming episode duration was highly correlated with time spent moving adding redundancy, whereas tailbeat frequency was more useful for manipulating the data and improving classification accuracy.

To examine whether related classifiers had similar accuracies, the fish in the 20 test data sets were labelled using two classifiers nearly identical to the classifier in Figure 5.13. One of the related classifiers had the multiply function replaced with an addition function and the other classifier had the addition replaced by a multiply. The classifier with two addition functions achieved a mean test score of 37.1% whilst the classifier with two multiply functions achieved a mean test score of 55.4%. A test score of approximately 50% suggests that the classifier has a similar accuracy to labelling the *dj-1*$^{-/-}$ fish by chance (because the test data has a 50:50 ratio of mutants to wild type). A score of less than 50% indicates a classifier with low labelling accuracy. Therefore, related classifiers with a single difference in the functions applied can vary in classification accuracy and here the related classifiers were shown to perform less effectively than the classifier in Figure 5.13. The functions applied to the input data contribute to the accuracy of the evolved classifier. The classifier in Figure 5.12 was also evolved separately using two other folds of dataset where the data was organised differently into the training, validation and test data. The test scores of these matching classifiers were both 64.3%, indicating that the test score can vary with the same classifier depending on how the data was organised into the three data sets. However, all three identical classifiers still achieved accuracies above 50% when classifying the test data, showing that they are more effective at labelling the *dj-1*$^{-/-}$ fish than if classification were down to chance.

Mutant line	Age	No. of recordings	Avg. of 20 folds		Highest scores	
			Train	Test	Train	Test
pink1 -/-	14 wpf	37 vs 37	64.6%	51.6%	76.2%	75%
dj-1 -/-	12 wpf	30 vs 30	75.9%	70.0%	73.7%	92.9%

Table 5.2 Scores of classifiers evolved using extracted features

The mean training and test scores for classifiers evolved to recognise *pink1*$^{-/-}$ and *dj-1*$^{-/-}$ zebrafish over 20 folds of dataset and the results of the highest scoring classifiers, evolved using one of the 20 folds of dataset.

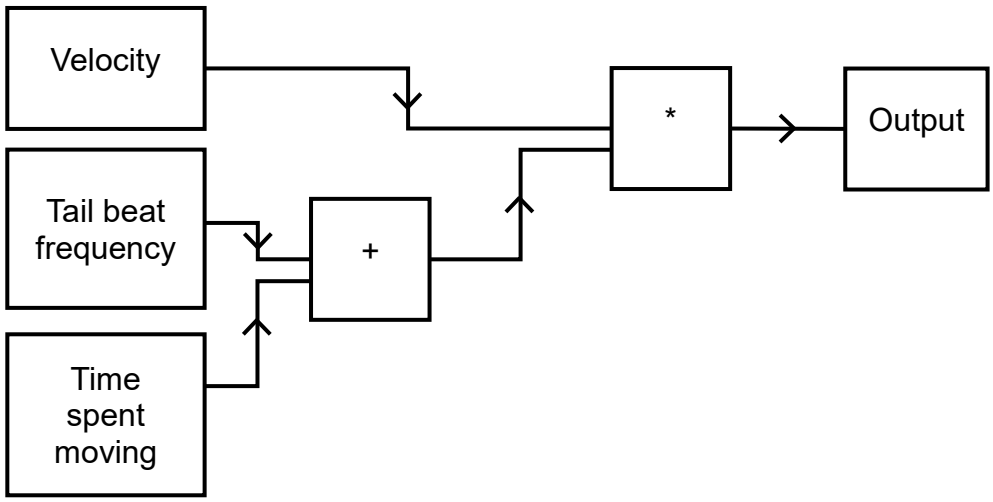


Figure 5.12 CGP network of the highest scoring *dj-1-/-* classifier

The CGP network of the highest scoring *dj-1-/-* classifier evolved using one of the 20 folds of dataset. The classifier achieved training and test scores of 73.7% and 92.9%, respectively.

5.3.3 Sliding window classifiers evolved using the raw data

Sliding window classifiers were evolved to recognise the *pink1*$^{-/-}$ and *dj-1*$^{-/-}$ mutant zebrafish using the PC2 time series data. The average training and test scores over the 20 folds of dataset are shown in Table 5.. The results of the highest scoring classifiers evolved using one of the 20 folds of dataset are also recorded. Using 39 *pink1*$^{-/-}$ zebrafish at 14wpf and 44 age-matched controls the evolved classifiers averaged a test score of 80% across the 20 folds of dataset. Classifiers evolved using 38 recordings of *dj-1*$^{-/-}$ zebrafish at 12 wpf and 64 recordings of age-matched controls achieved an average test score of 84%. The mean plots of the 20 windows most useful for discriminating *pink1*$^{-/-}$ and *dj-1*$^{-/-}$ movement can be seen in (Figure 5.13).

Mutant line	Age	No. of recordings	Avg. of 20 folds		Highest scores	
			Train	Test	Train	Test
pink1-/-	14 wpf	39 vs 44	85%	80%	86%	81%
dj-1-/-	12 wpf	38 vs 64	92%	84%	94%	89%

Table 5.3 Scores of the sliding window classifiers evolved

A table showing the average training and test scores for sliding window classifiers evolved to recognise *pink1*-/- and *dj-1-/-* zebrafish and the results of the highest scoring classifiers evolved.

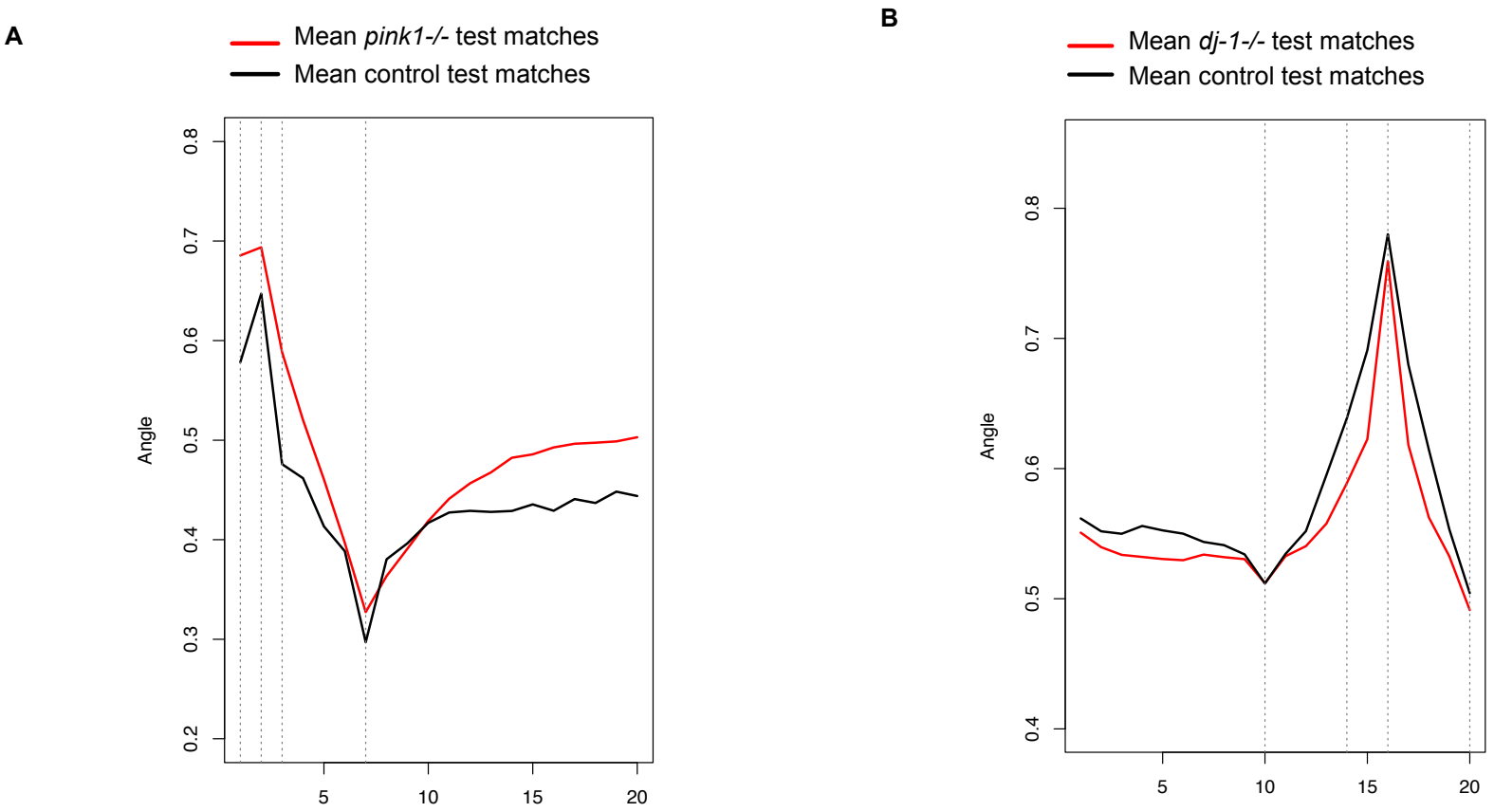


Figure 5.13 Patterns in the time series data used for discrimination

The mean plots of the 20 windows that were most useful for discriminating PD models from age-matched controls (e.g. the windows that created the highest output values). (A) The PC2 times series data used for discriminating *pink1-/-* mutants (red) and age-matched controls (black) in the test data sets. (B) The PC2 time series data used for discriminating *dj-1-/-* mutants and age-matched controls (black) in the test data sets.

5.4 Discussion

5.4.1 Locomotor defects in the DJ-1 and PINK1 deficient lines

Marked locomotor defects were observed in the DJ-1 deficient zebrafish at 12 weeks post fertilisation. There was an overall loss of movement seen with the distance travelled, velocity, time spent moving and duration of a swimming episode, not dissimilar to the bradykinesia (slowing of movement) observed in PD patients (Jankovic, 2008). Furthermore, the shorter duration of swimming episodes indicated more frequent periods of inactivity. This could be akin to the freezing episodes observed in PD where a patient's movement is halted for a few seconds (Chee et al., 2009). The tail beat frequency remained largely unaffected by DJ-1 deficiency apart from a slight loss at lower speeds (Figure 5.11E). However, when comparing the mean tail bend amplitudes at different speeds there were significant differences at low and high swimming speeds (Figure 5.11F). A trend was seen in the wild type zebrafish for the tail bend amplitude to decrease with an increase in swimming speed. Similar to previous reports this suggests the wild type zebrafish is relying on momentum to coast once higher speeds have been reached, reducing tail bend amplitude and energy expenditure (McHenry and Lauder, 2005). A greater tail bend amplitude was observed in the DJ-1 deficient zebrafish at high speeds compared to the wild type (Figure 5.11F). As the *dj-1*$^{-/-}$ mutant fish stop more frequently (Figure 5.11D), the resulting inertia may impede them from gaining the momentum required for coasting. Alternatively, the *dj-1*$^{-/-}$ zebrafish could lack the fine motor control required to moderate their swimming at speed. This is similar to the hypothesis that PD patients fail to appropriately scale the size of their movements to complete a given task (Mazzoni, Shabbott and Cortés, 2012). Overall, the distinctly altered movement of the *dj-1*$^{-/-}$ zebrafish correlates well with the loss of DA neuron markers seen at 16 wpf (Figure 3.8D). Comparatively, a loss of movement of a much smaller magnitude was observed in the PINK1 deficient zebrafish at 14 weeks post fertilisation. The distance travelled was decreased along with reductions in tailbeat frequency and tail bend amplitude at medium and high swimming speeds (Figure 5.10E,F). A loss of DA neurons markers has previously been reported in the PINK1 deficient

zebrafish (Flinn et al., 2013), however, no significant loss was seen in this *pink1*$^{-/-}$ mutant at 16 wpf (Figure 3.7D) or late adulthood (Figure 3.15).

5.4.2 Evolved classifiers recognise *dj-1*$^{-/-}$ zebrafish movement

The classifiers evolved using the features of movement were able to recognise the *dj-1*$^{-/-}$ zebrafish achieving an average test score of 70% across the 20 folds of data (Table 5.2). In contrast, the classifiers evolved to identify the *pink1*$^{-/-}$ zebrafish achieved an average test score of 51.6% indicating no real discriminatory ability. An equal number of mutants to wild type was included in each data set, therefore a classifier recognising approximately 50% of mutants in the test set is no more effective than classifying every single zebrafish as a mutant. The increased discriminatory ability of the *dj-1*$^{-/-}$ classifier makes sense as more features of movement were significantly different for the *dj-1*$^{-/-}$ fish. Figure 5.12 is a diagram of the highest scoring CGP network evolved to classify the *dj-1*$^{-/-}$ zebrafish using one of the 20 folds of dataset. The velocity, tailbeat frequency and time spent moving were used to label a zebrafish as either *dj-1*$^{-/-}$ mutant or wild type in that classifier. Similarly, velocity, tail beat frequency and low speed tail beat frequency occurred most frequently in the 20 *dj-1*$^{-/-}$ classifiers evolved. The velocity and tailbeat frequency must therefore be useful for discriminating DJ-1 deficient zebrafish. However, the classifier in Figure 5.12 was evolved separately using two other folds of dataset and achieved varying test scores. Further work on evolving classifiers using the features of movement could include recording more *dj-1*$^{-/-}$ and wild type fish to produce a further test set unseen by any of the final classifiers. The further test set could be used to get an unbiased discriminatory score for each classifier and would help in comparing how effectively they label the same data. Separating the current recordings (30 wild type vs 30 *dj-1*$^{-/-}$) into the three data sets and a further test set would reduce the numbers in training, validation and test data to a level where training would be insufficient and validation and test scores would be less precise.

5.4.3 IRCCPG classifiers effectively recognise the PD models

The IRCGP classifiers, evolved using the raw movement data, improved upon the results of classifiers evolved using features of movement. The classifiers evolved to recognise either the PINK1 or DJ-1 deficient zebrafish achieved average test scores of 80 and 84%, respectively. The increase in discriminatory ability is likely due to the more objective method of evolving a classifier from the raw movement data rather than select features of movement. The algorithm evolving a classifier using the features of movement is unbiased, however, humans originally decided upon the features of movement to study therefore making them subjective. The IRCGP classifiers were able to identify patterns in the PC2 value, over the time series, that were most useful for discriminating the mutants (Figure 5.13). For both classifiers the PC2 value was a linear combination of the 5 angles along the zebrafish trace with the most weighting given to angles 1 and 5. Therefore, the rapid change in angle seen entering and leaving the peak/trough in each plot is mainly representing the change in angles at the extremities. In other terms, deflection of the head and tail during swimming occurs at faster rate in the PD models than the wild type. Overall, a new feature of movement, characteristic of a zebrafish modelling PD, has been identified. Furthermore, the aim of evolving classifiers to discriminate between zebrafish modelling PD and age-matched controls based on movement data has been achieved. There is potential to use the classifiers evolved in a drug screen for future PD therapies, seeing whether the classifier will recognise the movement of a zebrafish modelling PD after treatment.

5.4.4 Classifiers need a movement phenotype to discriminate

As part of a separate study an attempt was made to see whether standard CGP classifiers could be evolved to recognise heterozygous mutant zebrafish with no observable movement phenotype. Homozygous mutation of the dystrophin gene in humans causes Duchenne muscular dystrophy (DMD) (Hoffman and Kunkel, 1989). The *sapje* zebrafish has a homozygous mutation in the dystrophin gene ortholog (*dmd*), causing detachment of muscle fibres at the somite boundary, and making it a representative animal model of DMD (Bassett et al., 2003). To determine whether the $dmd^{ta222a/+}$ heterozygous mutant, which

displays no overt phenotype, had any subtle differences in swimming movement they were recorded swimming from above in the same recording tank setup described in this chapter (Figure 5.2). 25 dmd$^{ta222a/+}$ heterozygous mutants and 25 age-matched wild type controls were recorded swimming at 12 wpf. The features of movement were extracted from each recording and used to evolve a standard CGP classifier, following the same method described in this chapter. There was no significant difference in any of the features of movement calculated (Appendix Figure 1) and the classifiers evolved were not able to discriminate the dmd$^{ta222a/+}$ zebrafish with an average test score of 43% (Appendix Table 3). The dmd$^{ta222a/+}$ zebrafish acted as a negative control showing that classifiers evolved using the features of movement fail to average a test score above ~50% when the two classes of zebrafish have no difference in movement phenotype.

6 Discussion

6.1 Summary of findings

The aims of this PhD thesis were to develop a genetic model of PD in the zebrafish that presents the molecular and cellular signatures of the disease, and then use an evolutionary algorithm to evolve a classifier that can recognise any movement phenotype. The data presented in chapter 3 shows the generation of five genetic models of PD in zebrafish and a confirmed loss of the DA neurons in one of them. Chapter 4 presents data from a transcriptomic analysis characterising the changes in gene expression in the *dj-1*$^{-/-}$ mutant line and outlines how this data is consistent with PD pathology. Finally, chapter 5 uses evolutionary algorithms to recognise the altered movement phenotype in zebrafish models of PD.

CRISPR/Cas9 targeting of *dj-1* in the zebrafish caused nonsense-mediated decay of the transcript and a subsequent loss of the DJ-1 protein. DJ-1 deficiency has been shown to cause a loss of DA neuron markers at the transcript level in the adult zebrafish brain. A further loss of TH positive neurons was seen in the posterior tuberculum, homologous to the substantia nigra in humans (Rink and Wullimann, 2001). Transcriptomic analysis of the *dj-1*$^{-/-}$ zebrafish has revealed dysregulation of genes involved in oxidative phosphorylation, the cell cycle, and Akt signalling; reflecting some known parkinsonian defects and indicative of others. Furthermore, I have been able to show that features of movement significantly differ in the *dj-1*$^{-/-}$ zebrafish at adulthood. An overall loss of movement was observed, with reduced velocity and more frequent periods of inactivity, suggesting a phenotype similar to the bradykinesia and freezing episodes seen in PD patients (Mazzoni, Shabbott and Cortés, 2012). I was then able to show that a classifier could be evolved to recognise the *dj-1*$^{-/-}$ zebrafish using the features of movement. As part of a collaboration, a further classifier was evolved using the raw movement data identifying a new characteristic of the *dj-1*$^{-/-}$ movement phenotype.

6.2 Translational models of disease in the zebrafish

In this thesis, an animal model of PD was generated in the zebrafish. The zebrafish was chosen because of its suitability for disease modelling and its track record in identifying potential therapies (Berger et al., 2010; Patton et al., 2005; Langenau et al., 2003). As a vertebrate, zebrafish share more in common with humans than widely used invertebrate organisms such as *C. elegans* or *Drosophila.* They have a similar body plan to humans and develop many of the same major organs by 5dpf (Santoriello and Zon, 2012). The zebrafish genome presents orthologs of 71% of human genes and 82% of known disease-causing genes (Howe et al., 2013). The externally developing zebrafish embryo allows one to carry out gene overexpression and knockdown studies by simple injection of mRNA or morpholino oligonucleotides, respectively (Rosen, Sweeney and Mably, 2009). Likewise, genome editing is also simple and effective in the zebrafish and can be used to produce genetic models of disease and transgenic lines expressing fluorescent reporter genes (Liu et al., 2017). The transparency of the zebrafish embryo allows the expression of fluorescent reporter genes to be examined in the live embryo (Long et al., 1997). Transgenic lines are therefore very useful for analysing any effects of gene manipulation. Furthermore, molecular methods such as *in situ* hybridisation and immunohistochemistry can be carried out on whole embryos to study gene expression. These methods are not as easily undertaken in mammalian models of disease which lack transparent externally developing embryos. Zebrafish are also well suited for high-throughput drug screens: producing hundreds of embryos at a time with rapid external development; they can be stocked at high density, with low maintenance costs, and treated with potential therapies dissolved in their water (Lieschke and Currie, 2007).

There are many examples of the power of zebrafish for modelling human disease; for instance, there are good zebrafish models of melanoma, leukemia and muscular dystrophy (Patton et al., 2005; Langenau et al., 2003; Berger et al., 2010). In 2005, a zebrafish model of melanoma was produced, with knockout of the p53 tumour suppressor and transgenic expression of the mutant BRAF(V600E) gene, which is one of the most common mutations in melanoma (Ascierto et al., 2012; Patton et al., 2005). A screen of 2,000

chemicals together with gene expression analysis, using *in situ* hybridisation and fluorescent reporters (to study melanocytes and their progenitors), allowed the identification of leflunomide as an effective suppressor of melanoma development (White et al., 2011). Leflunomide is a drug inhibiting transcriptional elongation and effective at reducing the elongation of c-Myc targets during melanoma development (White et al., 2011). Leflunomide has since been found to effectively treat human melanoma cell lines including those without the BRAF(V600E) mutation (Hanson et al., 2018). Leflunomide is currently involved in clinical trails for the treatment of breast cancer and myeloma (www.cancer.gov). Transgenic expression of the *MYC* oncogene in lymphoid cells has been used to generate a zebrafish model of T-cell leukaemia (Langenau et al., 2003). Two large scale drug screens have been carried out on the zebrafish model and fluorescent reporter genes expressed in the T-cells were used as a diagnostic (Ridges et al., 2012; Gutierrez et al., 2014). Lenaldekar and perphenazine, both regulators of Akt signalling, were identified as effective treatments in the zebrafish model and human leukaemia cell lines.

Zebrafish have also been used to model Duchenne muscular dystrophy (DMD) which is caused by mutation of the dystrophin gene (Hoffman and Kunkel, 1989). A dystrophin (*dmd*) zebrafish mutant produced in a forward genetic screen was identified by Bassett et al. (2003). They characterised the detachment of muscle fibres at the somite boundary with gene expression analysis, using a fluorescent reporter, *in situ* hybridisation and immunohistochemistry. A loss of the muscle specific isoform of *dmd* was revealed explaining why the phenotype was muscle specific. Birefringence assay and time lapse photography have since been used to show that muscle activity promotes muscle fibre detachment in the DMD zebrafish model (Li and Arner, 2015; Berger et al., 2010). The birefringence assay takes advantage of the transparency of the zebrafish embryo and uses light to study muscle integrity, with its structure rotating the beams of light (Berger, Sztal and Currie, 2012). Two large chemical screens have been carried out on the DMD zebrafish model using the birefringence assay as a diagnostic to see whether the muscle phenotype was rescued (Kawahara et al., 2011; Waugh et al., 2014). Aminophylline, an activator of PKA signalling, and fluoxetine, an inhibitor of serotonin reuptake, were identified as effective treatments from the screens.

Morpholino knockdown of a serotonin transporter in the zebrafish model further implicated serotonin reuptake in the pathology of DMD (Waugh et al., 2014). Using the expression of a fluorescent reporter mitochondrial fragmentation has also been identified in the pathology of the DMD zebrafish model (Giacomotto et al., 2013). A further study found that mitochondria in the DMD model opened their permeability transition pores (PTPs) and produced bursts of ROS with increased frequency (Zhang et al., 2015). Opening of the mitochondrial PTPs is involved in the cell death pathway and this is regulated by the protein cyclophilin D (Giorgio et al., 2010). Correspondingly, cyclosporin A, an inhibitor of cyclophilin D, was found to partially rescue muscular degeneration in the DMD zebrafish model (Zhang et al., 2015). More recently, Alisporivir, another inhibitor of cylophilin D, has been shown to partially rescue the phenotype in the DMD zebrafish model and in muscle biopsies from DMD patients (Schiavone et al., 2017). A zebrafish model of DMD has also been used to examine exon skipping as a potential therapy (Berger et al., 2011). A common mutation causing DMD is a deletion in exon 50 causing a premature stop codon in exon 51 (Bladen et al., 2015). The idea being that if exon 51 were skipped in translation the dystrophin protein generated may be functional and the muscle phenotype could be rescued. In a zebrafish model of DMD, with a premature stop codon in exon 32, exon-skipping increased dystrophin expression and an increase up to 30-40% of control levels was found to improve muscle function (Berger et al., 2011). Comparably, exon-skipping as a potential therapy for DMD managed to reach phase III clinical trails and was shown to improve the walking distance of less impaired DMD patients (Goemans et al., 2018).

In each of these examples the use of zebrafish for modelling human disease contributed to the identification of potential therapies. Using zebrafish made it possible to rapidly screen thousands of drugs and study the phenotypic effects with fluorescent transgenic lines and molecular methods such as *in situ* hybridisation. The therapies identified were also effective in human cells showing the promise of zebrafish models in clinical translation and, indeed, many have gone to clinical trials. Furthermore, the models helped improve understanding of the underlying molecular basis of disease. Transcriptional elongation of c-Myc targets was found to be essential in melanoma (White et al., 2011) and the regulation of Akt signalling was important in leukemia

(Langenau et al., 2003). In the DMD zebrafish model PKA signalling, serotonin reuptake and regulation of the mitochondrial PTP were all identified to play a role in the pathology (Waugh et al., 2014; Zhang et al., 2015; Kawahara et al., 2011). In much the same way, transcriptomic analysis of the PD model described in this thesis has indicated dysregulation of genes/pathways involved in metabolism and cell cycle contributing to our understanding of the molecular mechanisms that underpin PD.

6.3 Altered metabolism in the DJ-1 deficient zebrafish brain

Altered metabolism was seen in the DJ-1 deficient brain with reduced expression of a major enzyme in glycolysis, pyruvate kinase (Sun et al., 2011), and enrichment of the oxidative phosphorylation gene set (Chapter 4). In glycolysis, one molecule of glucose is converted into 2 ATP molecules, whereas, in oxidative phosphorylation up to 38 ATP molecules can be produced from the same amount of glucose (Fadaka et al., 2017). Glycolysis is therefore a much less efficient method of producing energy. However, the rate of glycolysis is much more rapid than oxidative phosphorylation allowing for a faster supply of energy when needed (Epstein, Gatenby and Brown, 2017). This is known to occur in muscle during periods of intense exercise when energy is in high demand (Baker, McCormick and Robergs, 2010). Neurons have also been shown to increase glycolysis over oxidative phosphorylation in response to stimulation (Díaz-García et al., 2017). This makes sense as neurons require a lot of energy to carry an action potential the length of their axons (Attwell and Laughlin, 2001). The DA neurons projecting from the posterior tuberculum to the striatum in the zebrafish brain would be even more energy reliant due to the distance covered (Rink and Wullimann, 2004). Speculatively, the ability to swap from oxidative phosphorylation to glycolysis could be an important mechanism to allow for the high energy requirements of DA neurons. The altered gene expression seen in the DJ-1 deficient zebrafish brain could indicate a loss of this ability contributing to the loss of DA neurons observed. Furthermore, an enzyme involved in creatine synthesis, glycine amidinotransferase (*gatm*) (Humm et al., 1997), was downregulated in the *dj-1*$^{-/-}$ zebrafish brain (Figure 4.3). Creatine can be phosphorylated to act as an

energy store in both the brain and muscle (Tarnopolsky and Beal, 2001). Therefore, loss of creatine synthesis could contribute to the issue of reduced energy available on demand for the DA neurons.

6.4 Enrichment of cell cycle-related gene sets in the *dj-1*$^{-/-}$ brain

An enrichment of cell cycle-related gene sets was identified in the DJ-1 deficient zebrafish brain alongside gene sets for epithelial mesenchymal transition, Akt signalling, Myc targets and TGF beta signalling (Chapter 5) which are all pathways associated with cancer development (Dang, 2012; Massagué, 2008; Thiery, 2002; Altomare and Testa, 2005). In 1997, DJ-1 was first identified in a screen for proteins interacting with c-Myc in HeLa cells and was shown to cooperate with Ras to transform NIH3T3 cells (Nagakubo et al., 1997). It was therefore surprising that loss of DJ-1 in the zebrafish brain was associated with the enrichment of these cancer related gene sets. My interpretation is that the enrichment of these gene sets is not really indicative of a cancerous state in the *dj-1*$^{-/-}$ brain. The evidence for this is that a major enzyme in glycolysis, pyruvate kinase (Israelsen and Vander Heiden, 2015), was downregulated whereas the oxidative phosphorylation gene set was enriched (Chapter 5). Cancer cells are known to use a distinct form of metabolism referred to as the Warburg effect. This is where cancer cells use glycolysis in lieu of oxidative phosphorylation for generating ATP despite the presence of oxygen (Heiden, Cantley and Thompson, 2009). The hallmarks of cancerous growth seen in the GSEA could instead be the result of forced re-entry of postmitotic neurons into the cell cycle in the *dj-1*$^{-/-}$ brain. This has been suggested as an apoptotic mechanism in neurodegeneration with increased expression of cell cycle markers observed in the neurons of both Alzheimer's and PD patients (Höglinger et al., 2007; Yang, Mufson and Herrup, 2003). Re-entry of neurons into the cell cycle is a specific response to DNA damage and increased production of genotoxic reactive oxygen species causes DNA damage in the substantia nigra DA neurons of PD patients (Zhang et al., 1999; Kruman et al., 2004). A loss of DJ-1 is likely to increase oxidative stress due to its role in activating the expression of antioxidant genes (Im et al., 2012). The decreased expression of antioxidant gene *gpx3* (Chung et al., 2009) seen in the *dj-1*$^{-/-}$ brain supports this (Figure

4.3). The cell cycle proteins implicated in this apoptotic mechanism include the cyclin dependent kinases (Cdks), the retinoblastoma protein (Rb) and the E2F transcription factors. Cdks 4 and 6 have been shown to increase in neuronal rat cultures following DNA damage and before the start of apoptosis (Park et al., 1998). The Cdks promote the cell cycle by phosphorylating their substrates including Rb (Suryadinata, Sadowski and Sarcevic, 2010). Active Rb sequesters the E2F transcription factors, preventing the cell from entering the S phase of the cell cycle. Its sequential phosphorylation by Cdk4/6 and Cdk2 inactivates the protein causing the release of the E2F transcription factors (Ezhevsky et al., 2001; Harbour et al., 1999). Interestingly, an increase in E2F1 has been observed in the substantia nigra DA neurons of PD patients post-mortem and inhibiting its transcription has been shown to have a neuroprotective effect in a neurotoxin-based mouse model of PD (Höglinger et al., 2007). E2F1 overexpression has further been shown to cause apoptosis in cultured mouse neurons (Hou et al., 2001). These reports suggest that E2F1 plays an active role in the death of neurons following cell cycle re-entry. An enrichment of the E2F targets gene set was also seen in the *dj-1*$^{-/-}$ brain (Figure 4.5) indicating increased activity of the E2F transcription factors. The enrichment of cell cycle-related gene sets could also indicate increased proliferation of progenitor cells to replace the neurons lost in the *dj-1*$^{-/-}$ zebrafish brain. The adult zebrafish brain has multiple zones containing actively cycling progenitors (Grandel et al., 2006). In response to injury, progenitor cells of the zebrafish brain have been shown to increase proliferation and migrate to replace the neurons lost (Kroehne et al., 2011), and more specifically this has been shown for a loss of catecholaminergic neurons (Caldwell et al., 2019). The enrichment of TGF beta signalling gene sets could also be involved in the regenerative process. TGF beta signalling has been observed during the proliferation of progenitor cells and regeneration of the retina in the adult zebrafish (Lenkowski et al., 2013). Furthermore, TGF beta signalling has been shown to promote the dendritic growth of DA neurons (Luo et al., 2016). A further gene upregulated in the *dj-1*$^{-/-}$*, abcf2a*, has previously been implicated in the regeneration of zebrafish hair cells (Jiang et al., 2014). Overall, the enrichment of cell cycle related gene sets may be due to forced re-entry of neurons in to the cell cycle as part of an apoptotic mechanism, or it could be that progenitor cells have increased proliferation to regenerate the neurons lost.

6.5 Future work

There are several avenues that could be explored with continuation of the research project. Firstly, to build upon the classifiers evolved here, more recordings of the wild type fish and PD models could be generated. This would allow the creation of a further test set, unseen by any of the 20 classifiers evolved using the 20 folds of dataset. The further test set would be used to get an unbiased discriminatory score from the classifiers for comparing how accurately they each label the same data. A clear next step in the project would be testing new PD therapies on the zebrafish models of PD and seeing whether the classifiers then label them as wild type. This would indicate that the drug rescues the PD movement phenotype in zebrafish and could lead to further research studying the effects on the DA neurons. It would also be interesting to study the molecular effects of DJ-1 deficiency on the DA neurons further. Cell sorting by flow cytometry could be used to extract the DA neurons and markers of the cell cycle could be examined to see whether there is forced re-entry of the DA neurons into the cell cycle. Additionally, the levels of phosphorylated AKT could be studied to see if there is altered AKT signalling specific to the DA neurons. The activity of the different mitochondrial complexes in DA neurons would be interesting to study, especially the activity of complex I which was reported as significantly reduced in a separate DJ-1 deficient line of zebrafish (Edson et al., 2019). Overall, future work would involve further enhancement of the evolved classifiers, using the classifiers for screening potential PD therapies, and molecular analysis of the DA neurons in *dj-1*$^{-/-}$ fish based on the enriched gene sets.

6.6 Perspectives

To date the majority of zebrafish models of PD have been transient models in the larvae. These have been generated using neurotoxins such as MPTP which cause a rapid loss of the DA neurons (Bretaud, Lee and Guo, 2004; Feng et al., 2014) or by morpholino knockdown of PD-associated genes (Bretaud et al., 2007; Flinn et al., 2009; Zhao et al., 2012). Testing potential therapies on zebrafish models of PD has been rare and the effects on movement phenotype

were assessed by measuring the distance travelled and velocity of larvae (Cronin and Grealy, 2017; Sheng et al., 2010). PD in humans is both age-related in nature and characterised by a movement phenotype of bradykinesia (slowing of movement), resting tremor and muscle rigidity (Jankovic, 2008). A model of PD in the adult zebrafish, with a progressive loss of the DA neurons, is therefore more suitable and evolving classifiers with the movement data provides a more comprehensive assessment of the movement phenotype. In this thesis, a genetic model of PD in the zebrafish, with a loss of the DA neurons, was raised to adulthood. In addition, high resolution video capture and machine learning allowed the discrete movement phenotype of a PD model in zebrafish to be assessed. There is also the future benefit of using the classifiers evolved for testing potential therapies and seeing whether they rescue the movement phenotype. This is an example of how machine learning will become increasingly important in the field of medicine. The use of machine learning is already growing in healthcare, with highly effective classifiers generated for the diagnosis of diseases such as PD, cancer and glaucoma (Kim, Cho and Oh, 2017; Lones et al., 2014; Sato et al., 2019). Machine learning has also been implemented in drug discovery, with biological and chemical datasets being utilised to identify disease-causing targets, virtually screen potential therapies and see whether they interact with the targets, and to predict how effective a therapy will be on a particular cell line (Vamathevan et al., 2019; Korkmaz, Zararsiz and Goksuluk, 2015; Menden et al., 2013). The novel use of machine learning described in this thesis has the potential to further impact on the drug discovery pipeline, with more comprehensive drug screening on animal models of movement disorders.

Appendix

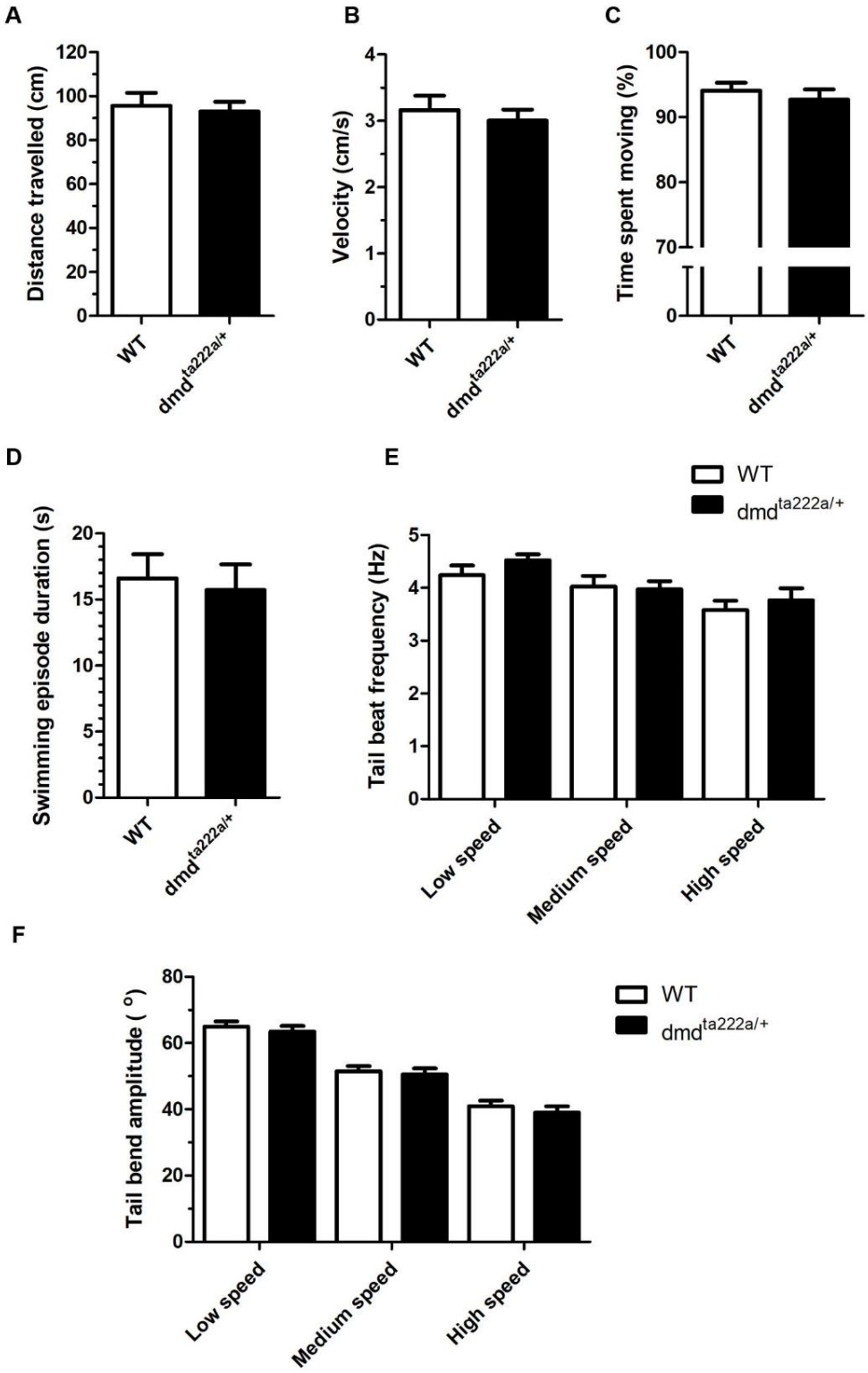
Velocity (V)	Mean tailbeat frequency (TBF)	Time spent moving (TSM)	Swimming episode duration (SED)	Class label	TBF + TSM	(TBF+TSM)*V	SED + TSM	(SED+TSM)*V
-0.33	0.62	0.02	-1.28	0	0.64	-0.21	-1.26	0.42
0.03	0.22	0.58	0.84	0	0.80	0.02	1.42	0.04
0.58	-0.62	0.58	0.84	0	-0.04	-0.02	1.42	0.82
0.50	0.59	0.58	0.84	0	1.17	0.59	1.42	0.71
1.74	-0.43	0.58	0.84	0	0.15	0.26	1.42	2.47
-0.59	0.33	0.30	-0.75	0	0.63	-0.37	-0.45	0.27
0.11	0.29	0.58	0.84	0	0.87	0.10	1.42	0.16
0.35	-0.83	0.58	0.84	0	-0.25	-0.09	1.42	0.50
1.04	-0.30	0.58	0.84	0	0.28	0.29	1.42	1.48
-0.34	0.08	0.58	0.84	0	0.66	-0.22	1.42	-0.48
0.50	-0.90	0.58	0.84	0	-0.32	-0.16	1.42	0.71
0.02	-0.09	0.58	0.84	0	0.49	0.01	1.42	0.03
1.19	-1.13	0.58	0.84	0	-0.55	-0.65	1.42	1.69
0.48	0.57	0.30	-0.75	0	0.87	0.42	-0.45	-0.22
0.47	-0.90	0.58	0.84	0	-0.32	-0.15	1.42	0.67
-0.45	3.46	0.58	0.84	0	4.04	-1.82	1.42	-0.64
0.86	-0.68	0.58	0.84	0	-0.10	-0.09	1.42	1.22
1.46	1.13	0.58	0.84	0	1.71	2.50	1.42	2.07
0.10	-0.62	0.58	0.84	0	-0.04	0.00	1.42	0.14
0.16	0.67	0.58	0.84	0	1.25	0.20	1.42	0.23
0.23	2.02	0.58	0.84	0	2.60	0.60	1.42	0.33
0.67	1.35	0.58	0.84	0	1.93	1.29	1.42	0.95
0.84	1.11	0.58	0.84	0	1.69	1.42	1.42	1.19
1.20	0.21	0.58	0.84	0	0.79	0.95	1.42	1.70
0.67	-0.07	0.58	0.84	0	0.51	0.34	1.42	0.95
0.98	1.24	0.58	0.84	0	1.82	1.78	1.42	1.39
0.68	0.48	0.30	-0.75	0	0.78	0.53	-0.45	-0.31
1.37	-0.77	0.58	0.84	0	-0.19	-0.26	1.42	1.95
-1.12	1.58	0.30	-0.75	0	1.88	-2.11	-0.45	0.50
1.48	1.40	0.30	-0.75	0	1.7	2.52	-0.45	-0.67

Appendix Table 1
A table showing the normalised velocity, tailbeat frequency, time spent moving and swimming episode duration for the wild type fish used to evolve a *dj-1*$^{-/-}$ classifier. The class label and possible output values from nodes in the highest scoring classifier are also shown.

Velocity (V)	Mean tailbeat frequency (TBF)	Time spent moving (TSM)	Swimming episode duration (SED)	Class label	TBF + TSM	(TBF+TSM)*V	SED + TSM	(SED+TSM)*V
-0.67	-0.87	0.02	0.63	1	-0.85	0.57	0.65	-0.44
-1.41	1.52	-3.04	-1.39	1	-1.52	2.14	-4.43	6.25
0.60	0.22	0.58	0.84	1	0.80	0.48	1.42	0.85
0.89	1.13	0.58	0.84	1	1.71	1.52	1.42	1.26
1.35	1.16	0.30	-0.75	1	1.46	1.97	-0.45	-0.61
-0.07	0.12	-0.25	-0.86	1	-0.13	0.01	-1.11	0.08
1.45	1.13	0.58	0.84	1	1.71	2.48	1.42	2.06
-1.15	-0.87	-0.53	-1.36	1	-1.40	1.61	-1.89	2.17
-0.41	-0.85	-0.25	-0.86	1	-1.10	0.45	-1.11	0.46
-1.13	-0.94	0.30	-0.75	1	-0.64	0.72	-0.45	0.51
-2.23	-0.38	-3.87	-1.78	1	-4.25	9.48	-5.65	12.60
-0.55	0.15	-0.81	-1.39	1	-0.66	0.36	-2.20	1.21
-1.52	-1.18	-0.25	-1.32	1	-1.43	2.17	-1.57	2.39
-1.94	-0.79	-1.09	-1.43	1	-1.88	3.65	-2.52	4.89
1.01	1.24	-0.53	0.42	1	0.71	0.72	-0.11	-0.11
-1.00	0.03	-1.09	-1.02	1	-1.06	1.06	-2.11	2.11
0.00	-0.94	0.58	0.84	1	-0.36	0.00	1.42	0.00
-0.87	-1.31	-0.53	-1.36	1	-1.84	1.60	-1.89	1.64
1.32	2.38	0.58	0.84	1	2.96	3.91	1.42	1.87
0.26	0.62	0.58	0.84	1	1.20	0.31	1.42	0.37
0.50	-0.68	0.58	0.84	1	-0.10	-0.05	1.42	0.71
-2.32	-0.86	-2.76	-1.79	1	-3.62	8.40	-4.55	10.56
-0.67	0.36	-0.53	-0.91	1	-0.17	0.11	-1.44	0.96
2.02	2.15	0.58	0.84	1	2.73	5.51	1.42	2.87
-1.95	0.29	-3.31	-1.44	1	-3.02	5.89	-4.75	9.26
2.20	2.97	0.58	0.84	1	3.55	7.81	1.42	3.12
-0.11	-0.33	-0.25	-0.86	1	-0.58	0.06	-1.11	0.12
-2.39	-0.70	-3.04	-1.82	1	-3.74	8.94	-4.86	11.62
2.50	2.49	0.58	0.84	1	3.07	7.68	1.42	3.55
-1.46	-0.40	-1.65	-1.50	1	-2.05	2.99	-3.15	4.60

Appendix Table 2

A table showing the normalised velocity, tailbeat frequency, time spent moving and swimming episode duration for the *dj-1*$^{-/-}$ fish used to evolve a *dj-1*$^{-/-}$ classifier. The class label and possible output values from nodes in the highest scoring classifier are also shown.



Appendix Figure 1

The features of movement compared between *dmd*^{ta222a}/+ and wild type (WT) at 12 wpf including (A) distance travelled, (B) velocity, (C) percentage of time spent moving, (D) mean duration of a swimming episode, (E) tailbeat frequency at low, medium and high swimming speeds and (F) tail bend amplitude at low, medium and high swimming speeds.

Mutant line	Age	No. of recordings	Avg. of 20 folds		Highest scores	
			Train	Test	Train	Test
dmd$^{ta222a/+}$	12 wpf	25 vs 25	66.0%	43.5%	73.3%	80%

Appendix Table 3

A table showing the average training and test scores for classifiers evolved to recognise *dmd*$^{ta222a/+}$ zebrafish using the features of movement and the results of the highest scoring classifier evolved using one of the dataset folds.

References

Abou-Sleiman, P. M., Healy, D. G., Quinn, N., Lees, A. J. and Wood, N. W. (2003). The role of pathogenic DJ-1 mutations in Parkinson's disease. *Ann. Neurol.* **54**, 283–286.

Adams, J. R., Van Netten, H., Schulzer, M., Mak, E., Mckenzie, J., Strongosky, A., Sossi, V., Ruth, T. J., Lee, C. S., Farrer, M., et al. (2005). PET in LRRK2 mutations: Comparison to sporadic Parkinson's disease and evidence for presymptomatic compensation. *Brain* **128**, 2777–2785.

Afonso, V., Santos, G., Collin, P., Khatib, A. M., Mitrovic, D. R., Lomri, N., Leitman, D. C. and Lomri, A. (2006). Tumor necrosis factor-α down-regulates human Cu/Zn superoxide dismutase 1 promoter via JNK/AP-1 signaling pathway. *Free Radic. Biol. Med.* **41**, 709–721.

Aleyasin, H., Rousseaux, M. W. C., Marcogliese, P. C., Hewitt, S. J., Irrcher, I., Joselin, A. P., Parsanejad, M., Kim, R. H., Rizzu, P., Callaghan, S. M., et al. (2010). DJ-1 protects the nigrostriatal axis from the neurotoxin MPTP by modulation of the AKT pathway. *Proc Natl Acad Sci USA* **107**, 3186–3191.

Alpaydin, E. (2016). *Machine learning: the new AI*. Cambridge, Massachusetts: MIT Press.

Altomare, D. A. and Testa, J. R. (2005). Perturbations of the AKT signaling pathway in human cancer. *Oncogene* **24**, 7455–7464.

Alvarez-Erviti, L., Rodriguez-Oroz, M. C., Cooper, J. M., Caballero, C., Ferrer, I., Obeso, J. A. and Schapira, A. H. V. (2010). Chaperone-mediated autophagy markers in Parkinson disease brains. *Arch. Neurol.* **67**, 1464–1472.

Anandhan, A., Jacome, M. S., Lei, S., Hernandez-Franco, P., Pappa, A., Panayiotidis, M. I., Powers, R. and Franco, R. (2017). Metabolic Dysfunction in Parkinson's Disease: Bioenergetics, Redox Homeostasis and Central Carbon Metabolism. *Brain Res. Bull.* **133**, 12–30.

Anichtchik, O., Diekmann, H., Fleming, A., Roach, A., Goldsmith, P. and Rubinsztein, D. C. (2008). Loss of PINK1 function affects development and results in neurodegeneration in zebrafish. *J. Neurosci.* **28**, 8199–

8207.

Anichtchik, O. V, Kaslin, J., Peitsaro, N., Scheinin, M. and Panula, P. (2004). Neurochemical and behavioural changes in zebrafish Danio rerio after systemic administration of 6-hydroxydopamine. *J. Neurochem.* **88**, 443–453.

Anna, A. and Monika, G. (2018). Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J. Appl. Genet.* **59**, 253–268.

Ariga, H., Takahashi-Niki, K., Kato, I., Maita, H., Niki, T. and Iguchi-Ariga, S. M. M. (2013). Neuroprotective Function of DJ-1 in Parkinson's Disease. *Oxid. Med. Cell. Longev.* **2013**, 1–9.

Ascierto, P. A., Kirkwood, J. M., Grob, J. J., Simeone, E., Grimaldi, A. M., Maio, M., Palmieri, G., Testori, A., Marincola, F. M. and Mozzillo, N. (2012). The role of BRAF V600 mutation in melanoma. *J. Transl. Med.* **10**, 1–9.

Attwell, D. and Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab.* **21**, 1133–1145.

Austin, P. C. and Steyerberg, E. W. (2017). Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. *Stat. Methods Med. Res.* **26**, 796–808.

Ayodele, T. O. (2010). Types of Machine Learning Algorithms. In: *New Advances in Machine Learning*. London: IntechOpen. 19–49.

Babu, N. S., Lakshmi, C., Murthy, N., Kakara, S. and Sharma, R. (2016). 1-Methyl- 4- phenyl- 1,2,3,6- tetrahydropyridine induced Parkinson's disease in zebrafish. *Proteomics* **16**, 1407–1420.

Badiola, N., de Oliveira, R. M., Herrera, F., Guardia-Laguarta, C., Gonçalves, S. A., Pera, M., Suárez-Calvet, M., Clarimon, J., Outeiro, T. F. and Lleó, A. (2011). Tau Enhances α-Synuclein Aggregation and Toxicity in Cellular Models of Synucleinopathy. *PLoS One* **6**, e26609.

Bae, E. J., Yang, N. Y., Lee, C., Lee, H. J., Kim, S., Sardi, S. P. and Lee, S. J. (2015). Loss of glucocerebrosidase 1 activity causes lysosomal dysfunction and α-synuclein aggregation. *Exp. Mol. Med.* **47**, e153.

Bagga, V., Dunnett, S. B. and Fricker, R. A. (2015). The 6-OHDA mouse model of Parkinson's disease - Terminal striatal lesions provide a

superior measure of neuronal loss and replacement than median forebrain bundle lesions. *Behav. Brain Res.* **288**, 107–117.

Bai, C., Partha, S., Hofmann, K., Lei, M., Goebl, M., Harper, J. W. and Elledge, S. J. (1996). SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. *Cell* **86**, 263–274.

Bai, Q., Mullett, S. J., Garver, J. A., Hinkle, D. A. and Burton, E. A. (2006). Zebrafish DJ-1 is evolutionarily conserved and expressed in dopaminergic neurons. *Brain Res.* **1113**, 33–44.

Bai, S. W., Herrera-Abreu, M. T., Rohn, J. L., Racine, V., Tajadura, V., Suryavanshi, N., Bechtel, S., Wiemann, S., Baum, B. and Ridley, A. J. (2011). Identification and characterization of a set of conserved and new regulators of cytoskeletal organization, cell morphology and migration. *BMC Biol.* **9**, 1–18.

Baker, J. S., McCormick, M. C. and Robergs, R. A. (2010). Interaction among skeletal muscle metabolic energy systems during intense exercise. *J. Nutr. Metab.* **2010**, 1–13.

Balleine, B. W., Delgado, M. R. and Hikosaka, O. (2007). The Role of the Dorsal Striatum in Reward and Decision-Making. *J. Neurosci.* **27**, 8161–8165.

Bandyopadhyay, S. and Cookson, M. R. (2004). Evolutionary and functional relationships within the DJ I superfamily. *BMC Evol. Biol.* **4**, 1–9.

Bao, L., Wu, J., Dodson, M., Ning, Y., Zhang, Z., Yao, M., Zhang, D. D., Xu, C., Hospital, G., Hospital, G., et al. (2017). ABCF2, an Nrf2 target gene, contributes to cisplatin resistance in ovarian cancer cells. *Mol Carcinog* **56**, 1543–1553.

Bardai, F. H., Ordonez, D. G., Bailey, R. M., Hamm, M., Lewis, J. and Feany, M. B. (2018). Lrrk promotes tau neurotoxicity through dysregulation of actin and mitochondrial dynamics. *PLoS Biol.* **16**, e2006265.

Barzilai, A. and Yamamoto, K.-I. (2004). DNA damage responses to oxidative stress. *DNA Repair* **3**, 1109–1115.

Bassett, D. I., Bryson-Richardson, R. J., Daggett, D. F., Gautier, P., Keenan, D. G. and Currie, P. D. (2003). Dystrophin is required for the formation of stable muscle attachments in the zebrafish embryo. *Development* **130**, 5851–5860.

Bates, S., Bonetta, L., MacAllan, D., Parry, D., Holder, A., Dickson, C. and Peters, G. (1994). CDK6 (PLSTIRE) and CDK4 (PSK-J3) are a distinct subset of the cyclin-dependent kinases that associate with cyclin D1. *Oncogene* **9**, 71–79.

Baulac, S., Lu, H., Strahle, J., Yang, T., Goldberg, M. S., Shen, J., Schlossmacher, M. G., Lemere, C. A., Lu, Q. and Xia, W. (2009). Increased DJ-1 expression under oxidative stress and in Alzheimer's disease brains. *Mol. Neurodegener.* **4**, 1–14.

Bender, A., Krishnan, K. J., Morris, C. M., Taylor, G. A., Reeve, A. K., Perry, R. H., Jaros, E., Hersheson, J. S., Betts, J., Klopstock, T., et al. (2006). High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease. *Nat. Genet.* **38**, 515–517.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc B* **57**, 289–300.

Berger, J., Berger, S., Hall, T. E., Lieschke, G. J. and Currie, P. D. (2010). Dystrophin-deficient zebrafish feature aspects of the Duchenne muscular dystrophy pathology. *Neuromuscul. Disord.* **20**, 826–832.

Berger, J., Berger, S., Jacoby, A. S., Wilton, S. D. and Currie, P. D. (2011). Evaluation of exon-skipping strategies for Duchenne muscular dystrophy utilizing dystrophin-deficient zebrafish. *J. Cell. Mol. Med.* **15**, 2643–2651.

Berger, J., Sztal, T. and Currie, P. D. (2012). Quantification of birefringence readily measures the level of muscle damage in zebrafish. *Biochem. Biophys. Res. Commun.* **423**, 785–788.

Bertoli, C., Skotheim, J. M. and De Bruin, R. A. M. (2013). Control of cell cycle transcription during G1 and S phases. *Nat. Rev. Mol. Cell Biol.* **14**, 518–528.

Betarbet, R., Sherer, T. B., MacKenzie, G., Garcia-Osuna, M., Panov, A. V. and Greenamyre, J. T. (2000). Chronic systemic pesticide exposure reproduces features of Parkinson's disease. *Nat. Neurosci.* **3**, 1301–1306.

Lo Bianco, C., Ridet, J. L., Schneider, B. L., Déglon, N. and Aebischer, P. (2002). α-synucleinopathy and selective dopaminergic neuron loss in a rat lentiviral-based model of Parkinson's disease. *Proc Natl Acad Sci USA* **99**, 10813–10818.

Bierl, C., Voetsch, B., Jin, R. C., Handy, D. E. and Loscalzo, J. (2004). Determinants of Human Plasma Glutathione Peroxidase (GPx-3) Expression. *J Biol Chem* **279**, 26839–26845.

Bill, B. R., Petzold, A. M., Clark, K. J., Schimmenti, L. A. and Ekker, S. C. (2009). A primer for morpholino use in zebrafish. *Zebrafish* **6**, 69–77.

Bisaglia, M., Mammi, S. and Bubacco, L. (2007). Kinetic and structural analysis of the early oxidation products of dopamine: Analysis of the interactions with α-synuclein. *J. Biol. Chem.* **282**, 15597–15605.

Biskup, S., Moore, D. J., Celsi, F., Higashi, S., West, A. B., Andrabi, S. A., Kurkinen, K., Yu, S. W., Savitt, J. M., Waldvogel, H. J., et al. (2006). Localization of LRRK2 to membranous and vesicular structures in mammalian brain. *Ann. Neurol.* **60**, 557–569.

Bitinaite, J., Wah, D. A., Aggarwal, A. K. and Schildkraut, I. (1998). FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci USA* **95**, 10570–10575.

Blackinton, J., Lakshminarasimhan, M., Thomas, K. J., Ahmad, R., Greggio, E., Raza, A. S., Cookson, M. R. and Wilson, M. A. (2009). Formation of a stabilized cysteine sulfinic acid is critical for the mitochondrial function of the parkinsonism protein DJ-1. *J. Biol. Chem.* **284**, 6476–6485.

Bladen, C. L., Salgado, D., Monges, S., Foncuberta, M. E., Kekou, K., Kosma, K., Dawkins, H., Lamont, L., Roy, A. J., Chamova, T., et al. (2015). The TREAT-NMD DMD global database: Analysis of more than 7,000 duchenne muscular dystrophy mutations. *Hum. Mutat.* **36**, 395–402.

Bland, J. M. and Altman, D. G. (1995). Multiple significance tests: The Bonferroni method. *BMJ* **310**, 170.

Blandini, F. and Armentero, M.-T. (2012). Animal models of Parkinson's disease. *FEBS J.* **279**, 1156–1166.

Blesa, J., Trigo-Damas, I., Quiroga-Varela, A. and Jackson-Lewis, V. R. (2015). Oxidative stress and Parkinson's disease. *Front. Neuroanat.* **9**, 1–9.

Bolam, J. P. and Pissadaki, E. K. (2012). Living on the edge with too many mouths to feed: Why dopamine neurons die. *Mov. Disord.* **27**, 1478–1483.

Bonifati, V., Rizzu, P., van Baren, M. J., Schaap, O., Breedveld, G. J., Krieger, E., Dekker, M. C. J., Squitieri, F., Ibanez, P., Joosse, M., et al. (2003). Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. *Science* **299**, 256–259.

Bradford, Y. M., Toro, S., Ramachandran, S., Ruzicka, L., Howe, D. G., Eagle, A., Kalita, P., Martin, R., Moxon, S. A. T., Schaper, K., et al. (2017). Zebrafish models of human disease: Gaining insight into human disease at ZFIN. *ILAR J.* **58**, 4–16.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* **30**, 1145–1159.

Bradley, S. R., Marino, M. J., Wittmann, M., Rouse, S. T., Awad, H., Levey, A. I. and Conn, P. J. (2000). Activation of group II metabotropic glutamate receptors inhibits synaptic excitation of the substantia nigra pars reticulata. *J. Neurosci.* **20**, 3085–3094.

Brand, A. H. and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415.

Brandt, U., Kerscher, S., Dröse, S., Zwicker, K. and Zickermann, V. (2003). Proton pumping by NADH:ubiquinone oxidoreductase. A redox driven conformational change mechanism? *FEBS Lett.* **545**, 9–17.

Branke, J. (2012). *Evolutionary optimization in dynamic environments*. Boston, Massachusetts: Springer.

Bretaud, S., Allen, C., Ingham, P. W. and Bandmann, O. (2007). p53-dependent neuronal cell death in a DJ-1-deficient zebrafish model of Parkinson's disease. *J. Neurochem.* **100**, 1626–1635.

Bretaud, S., Lee, S. and Guo, S. (2004). Sensitivity of zebrafish to environmental toxins implicated in Parkinson's disease. *Neurotoxicol. Teratol.* **26**, 857–864.

Broeks, A., Urbanus, J. H. M., Floore, A. N., Dahler, E. C., Klijn, J. G. M., Rutgers, E. J. T., Devilee, P., Russell, N. S., Van Leeuwen, F. E. and Van 't Veer, L. J. (2000). ATM-heterozygous germline mutations contribute to breast cancer-susceptibility. *Am. J. Hum. Genet.* **66**, 494–500.

Broeks, A., Urbanus, J. H. M., Knijff, P. de, Devilee, P., Nicke, M., Klöpper,

K., Dörk, T., Floore, A. N. and van't Veer, L. J. (2003). IVS10–6T>G, an ancient ATM germline mutation linked with breast cancer. *Hum. Mutat.* **21**, 521–528.

Brösamle, C. and Halpern, M. E. (2002). Characterization of myelination in the developing zebrafish. *Glia* **39**, 47–57.

Budick, S. A. and O'Malley, D. M. (2000). Locomotor Repertoire of the Larval Zebrafish: Swimming, Turning and Prey Capture. *J. Exp. Biol.* **203**, 2565–2579.

Burchell, V. S., Nelson, D. E., Sanchez-Martinez, A., Delgado-Camprubi, M., Ivatt, R. M., Pogson, J. H., Randle, S. J., Wray, S., Lewis, P. a, Houlden, H., et al. (2013). The Parkinson's disease-linked proteins Fbxo7 and Parkin interact to mediate mitophagy. *Nat. Neurosci.* **16**, 1257–1265.

Burré, J., Sharma, M., Tsetsenis, T., Buchman, V., Etherton, M. R. and Südhof, T. C. (2010). α-Synuclein promotes SNARE-complex assembly in vivo and in vitro. *Science* **329**, 1663–1667.

Cai, X., Smith, S. L. and Tyrrell, A. M. (2005). Benefits of Employing an Implicit Context Representation on Hardware Geometry of CGP. In: *Evolvable Systems: From Biology to Hardware*. Berlin: Springer. 143–154.

Cairns, B. R. (2009). The logic of chromatin architecture and remodelling at promoters. *Nature* **461**, 193–198.

Caldwell, L. J., Davies, N. O., Cavone, L., Mysiak, K. S., Semenova, S. A., Panula, P., Armstrong, J. D., Becker, C. G., Becker, T., Caldwell, A. L. J., et al. (2019). Regeneration of dopaminergic neurons in adult zebrafish depends on immune system activation and differs for distinct populations. *J. Neurosci.* **39**, 4694–4713.

Canet-Avilés, R. M., Wilson, M. A., Miller, D. W., Ahmad, R., McLendon, C., Bandyopadhyay, S., Baptista, M. J., Ringe, D., Petsko, G. A. and Cookson, M. R. (2004). The Parkinson's disease protein DJ-1 is neuroprotective due to cysteine-sulfinic acid-driven mitochondrial localization. *Proc Natl Acad Sci USA* **101**, 9103–9108.

Carrillo-Mora, P., Silva-Adaya, D. and Villaseñor-Aguayo, K. (2013). Glutamate in Parkinson's disease: Role of antiglutamatergic drugs. *Basal Ganglia* **3**, 147–157.

Chan, N. C., Salazar, A. M., Pham, A. H., Sweredoski, M. J., Kolawa, N. J., Graham, R. L. J., Hess, S. and Chan, D. C. (2011). Broad activation of the ubiquitin-proteasome system by Parkin is critical for mitophagy. *Hum. Mol. Genet.* **20**, 1726–1737.

Chan, P., DeLanney, L. E., Irwin, I., Langston, J. W. and Monte, D. (1991). Rapid ATP Loss Caused by 1-Methyl-4-Phenyl-1,2,3,6-Tetrahydropyridine in Mouse Brain. *J. Neurochem.* **57**, 348–351.

Chang, N., Sun, C., Gao, L., Zhu, D., Xu, X., Zhu, X., Xiong, J.-W. and Xi, J. J. (2013). Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. *Cell Res.* **23**, 465–472.

Chartier-Harlin, M. C., Kachergus, J., Roumier, C., Mouroux, V., Douay, X., Lincoln, S., Levecque, C., Larvor, L., Andrieux, J., Hulihan, M., et al. (2004). α-synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet* **364**, 1167–1169.

Chee, R., Murphy, A., Danoudis, M., Georgiou-Karistianis, N. and Iansek, R. (2009). Gait freezing in Parkinson's disease and the stride length sequence effect interaction. *Brain* **132**, 2151–2160.

Chen, Y. C., Sundvik, M., Rozov, S., Priyadarshini, M. and Panula, P. (2012). MANF regulates dopaminergic neuron development in larval zebrafish. *Dev. Biol.* **370**, 237–249.

Chen, Y. and Dorn, G. W. (2013). PINK1-Phosphorylated Mitofusin 2 Is a Parkin Receptor for Culling Damaged Mitochondria. *Science* **340**, 471–475.

Chesselet, M. F. (2008). In vivo alpha-synuclein overexpression in rodents: A useful model of Parkinson's disease? *Exp. Neurol.* **209**, 22–27.

Chicco, D. (2017). Ten quick tips for machine learning in computational biology. *BioData Min.* **10**, 1–17.

Chiu, C. C., Yeh, T. H., Lai, S. C., Wu-Chou, Y. H., Chen, C. H., Mochly-Rosen, D., Huang, Y. C., Chen, Y. J., Chen, C. L., Chang, Y. M., et al. (2015). Neuroprotective effects of aldehyde dehydrogenase 2 activation in rotenone-induced cellular and animal models of parkinsonism. *Exp. Neurol.* **263**, 244–253.

Choi, J., Sullards, M. C., Olzmann, J. A., Rees, H. D., Weintraub, S. T., Bostwick, D. E., Gearing, M., Levey, A. I., Chin, L. S. and Li, L. (2006). Oxidative damage of DJ-1 is linked to sporadic Parkinson and

Alzheimer diseases. *J. Biol. Chem.* **281**, 10816–10824.

Chung, C. Y., Seo, H., Sonntag, K. C., Brooks, A., Lin, L. and Isacson, O. (2005). Cell type-specific gene expression of midbrain dopaminergic neurons reveals molecules involved in their vulnerability and protection. *Hum. Mol. Genet.* **14**, 1709–1725.

Chung, S. S., Kim, M., Youn, B.-S., Lee, N. S., Park, J. W., Lee, I. K., Lee, Y. S., Kim, J. B., Cho, Y. M., Lee, H. K., et al. (2009). Glutathione Peroxidase 3 Mediates the Antioxidant Effect of Peroxisome Proliferator-Activated Receptor in Human Skeletal Muscle Cells. *Mol. Cell. Biol.* **29**, 20–30.

Clark, I. E., Dodson, M. W., Jiang, C., Cao, J. H., Huh, J. R., Seol, J. H., Yoo, S. J., Hay, B. A. and Guo, M. (2006). Drosophila pink1 is required for mitochondrial function and interacts genetically with parkin. *Nature* **441**, 1162–1166.

Clements, C. M., McNally, R. S., Conti, B. J., Mak, T. W. and Ting, J. P. (2006). DJ-1, a cancer- and Parkinson's disease-associated protein, stabilizes the antioxidant transcriptional master regulator Nrf2. *Proc Natl Acad Sci USA* **103**, 15091–15096.

Clurman, B. E., Sheaff, R. J., Thress, K., Groudine, M. and Roberts I, J. M. (1996). Turnover of cyclin E by the ubiquitin-proteasome pathway is regulated by cdk2 binding andcyclin phosphorylation. *Genes Dev.* **10**, 1979–1990.

Colapietro, P., Gervasini, C., Natacci, F., Rossi, L., Riva, P. and Larizza, L. (2003). NF1 exon 7 skipping and sequence alterations in exonic splice enhancers (ESEs) in a neurofibromatosis 1 patient. *Hum. Genet.* **113**, 551–554.

Conway, K. A., Lee, S. J., Rochet, J. C., Ding, T. T., Williamson, R. E. and Lansbury, P. T. (2000). Acceleration of oligomerization, not fibrillization, is a shared property of both α-synuclein mutations linked to early-onset Parkinson's disease: Implications for pathogenesis and therapy. *Proc Natl Acad Sci USA* **97**, 571–576.

Conway, K. A., Rochet, J. C., Bieganski, R. M. and Lansbury, J. (2001). Kinetic stabilization of the α-synuclein protofibril by a dopamine-α-synuclein adduct. *Science* **294**, 1346–1349.

Cookson, M. R. (2012). Parkinsonism due to mutations in PINK1, Parkin, and

DJ-1 and oxidative stress and mitochondrial pathways. *Cold Spring Harb. Perspect. Med.* **2**, a009415.

Cording, A. C., Shiaelis, N., Petridi, S., Middleton, C. A., Wilson, L. G. and Elliott, C. J. H. (2017). Targeted kinase inhibition relieves slowness and tremor in a Drosophila model of LRRK2 Parkinson's disease. *npj Park. Dis.* **34**, 1–10.

Cronin, A. and Grealy, M. (2017). Neuroprotective and Neuro-restorative Effects of Minocycline and Rasagiline in a Zebrafish 6-Hydroxydopamine Model of Parkinson's Disease. *Neuroscience* **367**, 34–46.

Cuervo, A. M., Stafanis, L., Fredenburg, R., Lansbury, P. T. and Sulzer, D. (2004). Impaired degradation of mutant α-synuclein by chaperone-mediated autophagy. *Science* **305**, 1292–1295.

Dahia, P. L. M., Marsh, D. J., Zheng, Z., Zedenius, J., Komminoth, P., Frisk, T., Wallin, G., Parsons, R., Longy, M., Larsson, C., et al. (1997). Somatic deletions and mutations in the Cowden disease gene, PTEN, in sporadic thyroid tumors. *Cancer Res.* **57**, 4710–4713.

Damier, P., Hirsch, E. C., Agid, Y. and Graybiel, A. M. (1999). The substantia nigra of the human brain: II. Patterns of loss of dopamine-containing neurons in Parkinson's disease. *Brain* **122**, 1437–1448.

Dang, C. V. (2012). MYC on the path to cancer. *Cell* **149**, 22–35.

Dave, K. D., De Silva, S., Sheth, N. P., Ramboz, S., Beck, M. J., Quang, C., Switzer, R. C., Ahmad, S. O., Sunkin, S. M., Walker, D., et al. (2014). Phenotypic characterization of recessive gene knockout rat models of Parkinson's disease. *Neurobiol. Dis.* **70**, 190–203.

Davidson, B., Hadar, R., Schlossberg, A., Sternlicht, T., Slipicevic, A., Skrede, M., Risberg, B., Flørenes, V. A., Kopolovic, J. and Reich, R. (2008). Expression and clinical role of DJ-1, a negative regulator of PTEN, in ovarian carcinoma. *Hum. Pathol.* **39**, 87–95.

Davies, H., Bignell, G. R., Cox, C., Stephens, P., Edkins, S., Clegg, S., Teague, J., Woffendin, H., Garnett, M. J., Bottomley, W., et al. (2002). Mutations of the BRAF gene in human cancer. *Nature* **417**, 949–954.

Day, M., Wang, Z., Ding, J., An, X., Ingham, C. A., Shering, A. F., Wokosin, D., Ilijic, E., Sun, Z., Sampson, A. R., et al. (2006). Selective elimination of glutamatergic synapses on striatopallidal neurons in Parkinson disease models. *Nat. Neurosci.* **9**, 251–259.

Decker, A. R., McNeill, M. S., Lambert, A. M., Overton, J. D., Chen, Y. C., Lorca, R. A., Johnson, N. A., Brockerhoff, S. E., Mohapatra, D. P., MacArthur, H., et al. (2014). Abnormal differentiation of dopaminergic neurons in zebrafish trpm7 mutant larvae impairs development of the motor pattern. *Dev. Biol.* **386**, 428–439.

Delgado-Camprubi, M., Esteras, N., Soutar, M. P., Plun-Favreau, H. and Abramov, A. Y. (2017). Deficiency of Parkinson's disease-related gene Fbxo7 is associated with impaired mitochondrial metabolism by PARP activation. *Cell Death Differ.* **24**, 120–131.

Deng, H., Wang, P. and Jankovic, J. (2018). The genetics of Parkinson disease. *Ageing Res. Rev.* **42**, 72–85.

Derivery, E., Fink, J., Martin, D., Houdusse, A., Piel, M. and Theresia, E. (2008). Free Brick1 Is a Trimeric Precursor in the Assembly of a Functional Wave Free Brick1 Is a Trimeric Precursor in the Assembly of a Functional Wave Complex. *PLoS One* **3**, e2462.

Díaz-García, C. M., Mongeon, R., Lahmann, C., Koveal, D., Zucker, H. and Yellen, G. (2017). Neuronal Stimulation Triggers Neuronal Glycolysis and Not Lactate Uptake. *Cell Metab.* **26**, 361–374.

Dietz, H. C., Valle, D., Francomano, C. A., Kendzior, R. J., Pyeritz, R. E. and Cutting, G. R. (1993). The Skipping of Constitutive Exons in Vivo Induced by Nonsense Mutations. *Science* **259**, 680–683.

Ding, S., Li, H., Su, C., Yu, J. and Jin, F. (2013). Evolutionary artificial neural networks: a review. *Artif. Intell. Rev.* **39**, 251–260.

Dinkova-Kostova, A. T. and Talalay, P. (2010). NAD(P)H:quinone acceptor oxidoreductase 1 (NQO1), a multifunctional antioxidant enzyme and exceptionally versatile cytoprotector. *Arch. Biochem. Biophys.* **501**, 116–123.

Dolgacheva, L. P., Berezhnov, A. V., Fedotova, E. I., Zinchenko, V. P. and Abramov, A. Y. (2019). Role of DJ-1 in the mechanism of pathogenesis of Parkinson's disease. *J. Bioenerg. Biomembr.* **51**, 175–188.

Dorsey, E. R., Elbaz, A., Nichols, E., Abd-Allah, F., Abdelalim, A., Adsuar, J. C., Ansha, M. G., Brayne, C., Choi, J. Y. J., Collado-Mateo, D., et al. (2018). Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* **17**, 939–953.

Doyon, Y., McCammon, J. M., Miller, J. C. and Faraji, F. (2008). Heritable Targeted Gene Disruption in Zebrafish Using Designed Zinc Finger Nucleases. *Nat. Biotechnol.* **26**, 702–708.

Dreiseitl, S. and Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: A methodology review. *J. Biomed. Inform.* **35**, 352–359.

Du, X., Xu, H., Shi, L., Jiang, Z., Song, N., Jiang, H. and Xie, J. (2016). Activation of ATP-sensitive potassium channels enhances DMT1-mediated iron uptake in SK-N-SH cells in vitro. *Sci. Rep.* **6**, 1–10.

Du, Y., Dodel, R. C., Bales, K. R., Jemmerson, R., Hamilton-Byrd, E. and Paul, S. M. (1997). Involvement of a Caspase-3-Like Cysteine Protease in 1-Methyl-4-Phenylpyridinium-Mediated Apoptosis of Cultured Cerebellar Granule Neurons. *J. Neurochem.* **69**, 1382–1388.

Dusonchet, J., Kochubey, O., Stafa, K., Young, S. M., Zufferey, R., Moore, D. J., Schneider, B. L. and Aebischer, P. (2011). A rat model of progressive nigral neurodegeneration induced by the Parkinson's disease-associated G2019S mutation in LRRK2. *J. Neurosci.* **31**, 907–912.

Dvinge, H. (2018). Regulation of alternative mRNA splicing: old players and new perspectives. *FEBS Lett.* **592**, 2987–3006.

Edson, A. J., Hushagen, H. A., Frøyset, A. K., Elda, I., Khan, E. A., Di Stefano, A. and Fladmark, K. E. (2019). Dysregulation in the Brain Protein Profile of Zebrafish Lacking the Parkinson's Disease-Related Protein DJ-1. *Mol. Neurobiol.* **56**, 1–17.

Van Den Eeden, S. K. (2003). Incidence of Parkinson's Disease: Variation by Age, Gender, and Race/Ethnicity. *Am. J. Epidemiol.* **157**, 1015–1022.

Eiben, A. E. and Smith, J. (2015). From evolutionary computation to the evolution of things. *Nature* **521**, 476–482.

El-Brolosy, M. A., Kontarakis, Z., Rossi, A. and Kuenne, C. (2019). Genetic compensation triggered by mutant mRNA degradation. *Nature* **568**, 193–197.

El-Brolosy, M. A. and Stainier, D. Y. R. (2017). Genetic compensation: A phenomenon in search of mechanisms. *PLoS Genet.* **13**, 1–17.

El-Khodor, B. F., Oo, T. F., Kholodilov, N. and Burke, R. E. (2003). Ectopic Expression of Cell Cycle Markers in Models of Induced Programmed Cell

Death in Dopamine Neurons of the Rat Substantia Nigra Pars Compacta. *Exp. Neurol.* **179**, 17–27.

Epstein, T., Gatenby, R. A. and Brown, J. S. (2017). The Warburg effect as an adaptation of cancer cells to rapid fluctuations in energy demand. *PLoS One* **12**, e0185085.

Esteves, A. R. and Cardoso, S. M. (2017). LRRK2 at the Crossroad Between Autophagy and Microtubule Trafficking: Insights into Parkinsons Disease. *Neurosci.* **23**, 16–26.

Evan, G. I. and Vousden, K. H. (2001). Proliferation, cell cycle and apoptosis in cancer. *Nature* **411**, 342–348.

Ezhevsky, S. A., Ho, A., Becker-Hapak, M., Davis, P. K. and Dowdy, S. F. (2001). Differential Regulation of Retinoblastoma Tumor Suppressor Protein by G1 Cyclin-Dependent Kinase Complexes In Vivo. *Mol. Cell. Biol.* **21**, 4773–4784.

Fadaka, A., Ajiboye, B., Ojo, O., Adewale, O., Olayide, I. and Emuowhochere, R. (2017). Biology of glucose metabolization in cancer cells. *J. Oncol. Sci.* **3**, 45–51.

Fallon, L., Bélanger, C. M. L., Corera, A. T., Kontogiannea, M., Regan-Klapisz, E., Moreau, F., Voortman, J., Haber, M., Rouleau, G., Thorarinsdottir, T., et al. (2006). A regulated interaction with the UIM protein Eps15 implicates parkin in EGF receptor trafficking and PI(3)K-Akt signalling. *Nat. Cell Biol.* **8**, 834–842.

Fan, C. Y., Cowden, J., Simmons, S. O., Padilla, S. and Ramabhadran, R. (2010). Gene expression changes in developing zebrafish as potential markers for rapid developmental neurotoxicity screening. *Neurotoxicol. Teratol.* **32**, 91–98.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognit. Lett.* **27**, 861–874.

Feany, M. B. and Bender, W. W. (2000). A Drosophila model of Parkinson's disease. *Nature* **404**, 394–398.

Fearnley, J. M. and Lees, A. J. (1991). Ageing and parkinson's disease: Substantia nigra regional selectivity. *Brain* **114**, 2283–2301.

Feigin, V. L., Krishnamurthi, R. V., Theadom, A. M., Abajobir, A. A., Mishra, S. R., Ahmed, M. B., Abate, K. H., Mengistie, M. A., Wakayo, T., Abd-Allah, F., et al. (2017). Global, regional, and national burden of

neurological disorders during 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Neurol.* **16**, 877–897.

Feng, C. W., Wen, Z. H., Huang, S. Y., Hung, H. C., Chen, C. H., Yang, S. N., Chen, N. F., Wang, H. M., Hsiao, C. Der and Chen, W. F. (2014). Effects of 6-hydroxydopamine exposure on motor activity and biochemical expression in zebrafish (Danio Rerio) larvae. *Zebrafish* **11**, 227–239.

Fernandez-Caggiano, M., Schröder, E., Cho, H. J., Burgoyne, J., Barallobre-Barreiro, J., Mayr, M. and Eaton, P. (2016). Oxidant-induced interprotein disulfide formation in cardiac protein DJ-1 occurs via an interaction with peroxiredoxin 2. *J. Biol. Chem.* **291**, 10399–10410.

Ferreira, A., Han, H. Q., Greengard, P. and Kosik, K. S. (1995). Suppression of synapsin II inhibits the formation and maintenance of synapses in hippocampal culture. *Proc Natl Acad Sci USA* **92**, 9225–9229.

Ferro, M. M., Bellissimo, M. I., Anselmo-Franci, J. A., Angellucci, M. E. M., Canteras, N. S. and Da Cunha, C. (2005). Comparison of bilaterally 6-OHDA- and MPTP-lesioned rats as models of the early phase of Parkinson's disease: Histological, neurochemical, motor and memory alterations. *J. Neurosci. Methods* **148**, 78–87.

Fett, M. E., Pilsl, A., Paquet, D., van Bebber, F., Haass, C., Tatzelt, J., Schmid, B. and Winklhofer, K. F. (2010). Parkin is protective against proteotoxic stress in a transgenic zebrafish model. *PLoS One* **5**, e11783.

Flinn, L., Mortiboys, H., Volkmann, K., Kster, R. W., Ingham, P. W. and Bandmann, O. (2009). Complex i deficiency and dopaminergic neuronal cell loss in parkin-deficient zebrafish (Danio rerio). *Brain* **132**, 1613–1623.

Flinn, L. J., Keatinge, M., Bretaud, S., Mortiboys, H., Matsui, H., De Felice, E., Woodroof, H. I., Brown, L., McTighe, A., Soellner, R., et al. (2013). *TigarB* causes mitochondrial dysfunction and neuronal loss in PINK1 deficiency. *Ann. Neurol.* **74**, 837–847.

Fonzo, A. Di, Dekker, M. C. J., Montagna, P., Baruzzi, A., Yonova, E. H., Guedes, L. C., Szczerbinska, A., Zhao, T., Dubbel-Hulsman, L. O. M., Wouters, C. H., et al. (2009). FBXO7 mutations cause autosomal recessive, early-onset parkinsonian-pyramidal syndrome. *Neurology* **72**, 240–245.

Fornstedt, B., Bergh, I., Rosengren, E. and Carlsson, A. (1990). An Improved HPLC-Electrochemical Detection Method for Measuring Brain Levels of 5-S-Cysteinyldopamine, 5-S-Cysteinyl-3,4-Dihydroxyphenylalanine, and 5-S-Cysteinyl-3,4-Dihydroxyphenylacetic Acid. *J. Neurochem.* **54**, 578–586.

Friedman, L. G., Lachenmayer, M. L., Wang, J., He, L., Poulose, S. M., Komatsu, M., Holstein, G. R. and Yue, Z. (2012). Disrupted autophagy leads to dopaminergic axon and dendrite degeneration and promotes presynaptic accumulation of α-synuclein and LRRK2 in the brain. *J. Neurosci.* **32**, 7585–7593.

Frøyset, A. K., Edson, A. J., Gharbi, N., Khan, E. A., Dondorp, D., Bai, Q., Tiraboschi, E., Suster, M. L., Connolly, J. B., Burton, E. A., et al. (2018). Astroglial DJ-1 over-expression up-regulates proteins involved in redox regulation and is neuroprotective in vivo. *Redox Biol.* **16**, 237–247.

Gabriel, L. R., Wu, S., Kearney, P., Bellvé, K. D., Standley, C., Fogarty, K. E. and Melikian, H. E. (2013). Dopamine transporter endocytic trafficking in striatal dopaminergic neurons: Differential dependence on dynamin and the actin cytoskeleton. *J. Neurosci.* **33**, 17836–17846.

Gaj, T., Gersbach, C. A. and Barbas III, C. F. (2013). ZFN, TALEN and Crispr-Cas-based methods for genome engineering. *Trends Biotechnol* **31**, 397–405.

Galli, S., Jahn, O., Hitt, R., Hesse, D., Opitz, L., Plessmann, U., Urlaub, H., Poderoso, J. J., Jares-Erijman, E. A. and Jovin, T. M. (2009). A New Paradigm for MAPK: Structural Interactions of hERK1 with Mitochondria in HeLa Cells. *PLoS One* **4**, e7541.

Gandhi, S., Muqit, M. M. K., Stanyer, L., Healy, D. G., Abou-Sleiman, P. M., Hargreaves, I., Heales, S., Ganguly, M., Parsons, L., Lees, A. J., et al. (2006). PINK1 protein in normal human brain and Parkinson's disease. *Brain* **129**, 1720–1731.

Garbarino, G., Costa, S., Pestarino, M. and Candiani, S. (2014). Differential expression of synapsin genes during early zebrafish development. *Neuroscience* **280**, 351–367.

Gasparini, F., Di Paolo, T. and Gomez-Mancilla, B. (2013). Metabotropic glutamate receptors for parkinson's disease therapy. *Parkinsons. Dis.* 196028.

Gegg, M. E., Burke, D., Heales, S. J. R., Cooper, J. M., Hardy, J., Wood, N. W. and Schapira, A. H. V. (2012). Glucocerebrosidase deficiency in substantia nigra of parkinson disease brains. *Ann. Neurol.* **72**, 455–463.

Giacomotto, J., Brouilly, N., Walter, L., Mariol, M.-C., Berger, J., Ségalat, L., Becker, T. S., Currie, P. D. and Gieseler, K. (2013). Chemical genetics unveils a key role of mitochondrial dynamics, cytochrome c release and IP3R activity in muscular dystrophy. *Hum. Mol. Genet.* **22**, 4562–4578.

Giguère, N., Delignat-Lavaud, B., Herborg, F., Voisin, A., Li, Y., Jacquemet, V., Anand-Srivastava, M., Gether, U., Giros, B. and Trudeau, L.-É. (2019). Increased vulnerability of nigral dopamine neurons after expansion of their axonal arborization size through D2 dopamine receptor conditional knockout. Petrou, S. (Ed). *PLOS Genet.* **15**, e1008352.

Gillardon, F. (2009). Leucine-rich repeat kinase 2 phosphorylates brain tubulin-beta isoforms and modulates microtubule stability - A point of convergence in Parkinsonian neurodegeneration? *J. Neurochem.* **110**, 1514–1522.

Giordano, S., Lee, J., Darley-Usmar, V. M. and Zhang, J. (2012). Distinct Effects of Rotenone, 1-methyl-4-phenylpyridinium and 6-hydroxydopamine on Cellular Bioenergetics and Cell Death. *PLoS One* **7**, e44610.

Giorgi, C., Marchi, S. and Pinton, P. (2018). The machineries, regulation and cellular functions of mitochondrial calcium. *Nat. Rev. Mol. Cell Biol.* **19**, 713–730.

Giorgio, V., Soriano, M. E., Basso, E., Bisetto, E., Lippe, G., Forte, M. A. and Bernardi, P. (2010). Cyclophilin D in mitochondrial pathophysiology. *Biochim. Biophys. Acta - Bioenerg.* **1797**, 1113–1118.

Gispert, S., Ricciardi, F., Kurz, A., Azizov, M., Hoepken, H.-H., Becker, D., Voos, W., Leuner, K., Müller, W. E., Kudin, A. P., et al. (2009). Parkinson phenotype in aged PINK1-deficient mice is accompanied by progressive mitochondrial dysfunction in absence of neurodegeneration. *PLoS One* **4**, e5777.

Gladkova, C., Maslen, S. L., Skehel, J. M. and Komander, D. (2018). Mechanism of parkin activation by PINK1. *Nature* **559**, 410–414.

Gloeckner, C. J., Kinkl, N., Schumacher, A., Braun, R. J., O'Neill, E.,

Meitinger, T., Kolch, W., Prokisch, H. and Ueffing, M. (2006). The Parkinson disease causing LRRK2 mutation I2020T is associated with increased kinase activity. *Hum. Mol. Genet.* **15**, 223–232.

Godoy, R., Noble, S., Yoon, K., Anisman, H. and Ekker, M. (2015). Chemogenetic ablation of dopaminergic neurons leads to transient locomotor impairments in zebrafish larvae. *J. Neurochem.* **135**, 249–260.

Goemans, N., Mercuri, E., Belousova, E., Komaki, H., Dubrovsky, A., McDonald, C. M., Kraus, J. E., Lourbakos, A., Lin, Z., Campion, G., et al. (2018). A randomized placebo-controlled phase 3 trial of an antisense oligonucleotide, drisapersen, in duchenne muscular dystrophy. *Neuromuscul. Disord.* **28**, 4–15.

Gordon, M. D. and Scott, K. (2009). Motor Control in a Drosophila Taste Circuit. *Neuron* **61**, 373–384.

Goyal, M. S., Hawrylycz, M., Miller, J. A., Snyder, A. Z. and Raichle, M. E. (2014). Aerobic glycolysis in the human brain is associated with development and neotenous gene expression. *Cell Metab.* **19**, 49–57.

Grandel, H., Kaslin, J., Ganz, J., Wenzel, I. and Brand, M. (2006). Neural stem cells and neurogenesis in the adult zebrafish brain: Origin, proliferation dynamics, migration and cell fate. *Dev. Biol.* **295**, 263–277.

Gratz, S. J., Cummings, A. M., Nguyen, J. N., Hamm, D. C., Donohue, L. K., Harrison, M. M., Wildonger, J. and O'connor-Giles, K. M. (2013). Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. *Genetics* **194**, 1029–1035.

Greene, J. C., Whitworth, A. J., Kuo, I., Andrews, L. A., Feany, M. B. and Pallanck, L. J. (2003). Mitochondrial pathology and apoptotic muscle degeneration in Drosophila parkin mutants. *Proc Natl Acad Sci USA* **100**, 4078–4083.

Greene, J. G., Dingledine, R. and Greenamyre, J. T. (2005). Gene expression profiling of rat midbrain dopamine neurons: Implications for selective vulnerability in parkinsonism. *Neurobiol. Dis.* **18**, 19–31.

Greffard, S., Verny, M., Bonnet, A.-M. and Beinis, J.-Y. (2006). Motor score of the unified Parkinson disease rating scale as a good predictor of lewy body-associated neuronal loss in the substantia nigra. *Arch. Neurol.* **63**, 584–588.

Grimm, J., Mueller, A., Hefti, F. and Rosenthal, A. (2004). Molecular basis for

catecholaminergic neuron diversity. *Proc Natl Acad Sci USA* **101**, 13891–13896.

Groulx, I. and Lee, S. (2002). Oxygen-dependent ubiquitination and degradation of hypoxia-inducible factor requires nuclear-cytoplasmic trafficking of the von Hippel-Lindau tumor suppressor protein. *Mol. Cell. Biol.* **22**, 5319–5336.

Gu, L., Cui, T., Fan, C., Zhao, H., Zhao, C., Lu, L. and Yang, H. (2009). Involvement of ERK1/2 signaling pathway in DJ-1-induced neuroprotection against oxidative stress. *Biochem. Biophys. Res. Commun.* **383**, 469–474.

Gupta, R. M. and Musunuru, K. (2014). Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *J. Clin. Invest.* **124**, 4154–4161.

Gupta, T. and Mullins, M. C. (2010). Dissection of organs from the adult zebrafish. *J. Vis. Exp.* **37**, e1717.

Gutierrez, A., Pan, L., Groen, R. W. J., Baleydier, F., Kentsis, A., Marineau, J., Grebliunaite, R., Kozakewich, E., Reed, C., Pflumio, F., et al. (2014). Phenothiazines induce PP2A-mediated apoptosis in T cell acute lymphoblastic leukemia. *J. Clin. Invest.* **124**, 644–655.

Guzman, J. N., Sanchez-padilla, J., Wokosin, D., Kondapalli, J., Ilijic, E., Schumacker, P. T. and Surmeier, D. J. (2010). Oxidant stress evoked by pacemaking in dopaminergic neurons is attenuated by DJ-1. *Nature* **468**, 696–700.

Haas, S., Steplewski, A., Siracusa, L. D., Amini, S. and Khalili, K. (1995). Identification of a sequence-specific single-stranded DNA binding protein that suppresses transcription of the mouse myelin basic protein gene. *J. Biol. Chem.* **270**, 12503–12510.

Haddad, D. and Nakamura, K. (2015). Understanding the susceptibility of dopamine neurons to mitochondrial stressors in Parkinson's disease. *FEBS Lett.* **589**, 3702–3713.

Hampe, C., Ardila-osorio, H., Fournier, M., Brice, A., Corti, O., Curie, M. and Vi, P. (2006). Biochemical analysis of Parkinson's disease-causing variants of Parkin , an E3 ubiquitin – protein ligase with monoubiquitylation capacity. *Hum. Mol. Genet.* **15**, 2059–2075.

Han, H. Q., Nichols, R. A., Rubin, M. R., Bähler, M. and Greengard, P. (1991). Induction of formation of presynaptic terminals in neuroblastoma

cells by synapsin IIb. *Nature* **349**, 697–700.

Hanson, K., Robinson, S. R., Al-Yousuf, K., Hendry, A. E., Sexton, D. W., Sherwood, V. and Wheeler, G. N. (2018). The anti-rheumatic drug, leflunomide, synergizes with MEK inhibition to suppress melanoma growth. *Oncotarget* **9**, 3815–3829.

Hao, L. Y., Giasson, B. I. and Bonini, N. M. (2010). DJ-1 is critical for mitochondrial function and rescues PINK1 loss of function. *Proc Natl Acad Sci USA* **107**, 9747–9752.

Hara, M., Abe, Y., Tanaka, T., Yamamoto, T., Okumura, E. and Kishimoto, T. (2012). Greatwall kinase and cyclin B-Cdk1 are both critical constituents of M-phase-promoting factor. *Nat. Commun.* **3**, 1–9.

Harbour, J. W., Luo, R. X., Dei Santi, A., Postigo, A. A. and Dean, D. C. (1999). Cdk phosphorylation triggers sequential intramolecular interactions that progressively block Rb functions as cells move through G1. *Cell* **98**, 859–869.

Harel, T., Pehlivan, D., Caskey, C. T. and Lupski, J. R. (2015). Mendelian, Non-Mendelian, Multigenic Inheritance, and Epigenetics. In: *Rosenberg's Molecular and Genetic Basis of Neurological and Psychiatric Disease: Fifth Edition*. Cambridge, Massachusetts: Academic Press. 3–27.

Hassouna, I., Wickert, H., Zimmermann, M. and Gillardon, F. (1996). Increase in bax expression in substantia nigra following 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) treatment of mice. *Neurosci. Lett.* **204**, 85–88.

Hauser, D. N., Mamais, A., Conti, M. M., Primiani, C. T., Kumaran, R., Dillman, A. A., Langston, R. G., Beilina, A., Garcia, J. H., Diaz-Ruiz, A., et al. (2017). Hexokinases link DJ-1 to the PINK1/parkin pathway. *Mol. Neurodegener.* **12**, 1–17.

Hauser, D. N. and Hastings, T. G. (2013). Mitochondrial dysfunction and oxidative stress in Parkinson's disease and monogenic parkinsonism. *Neurobiol. Dis.* **51**, 35–42.

Hauser, M. A., Li, Y. J., Xu, H., Noureddine, M. A., Shao, Y. S., Gullans, S. R., Scherzer, C. R., Jensen, R. V., McLaurin, A. C., Gibson, J. R., et al. (2005). Expression profiling of substantia nigra in Parkinson disease, progressive supranuclear palsy, and frontotemporal dementia with parkinsonism. *Arch. Neurol.* **62**, 917–921.

Hayashi, T., Ishimori, C., Takahashi-Niki, K., Taira, T., Kim, Y. chul, Maita, H., Maita, C., Ariga, H. and Iguchi-Ariga, S. M. M. (2009). DJ-1 binds to mitochondrial complex I and maintains its activity. *Biochem. Biophys. Res. Commun.* **390**, 667–672.

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

Heiden, M. G. Vander, Cantley, L. C. and Thompson, C. B. (2009). Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. *Science* **324**, 1029–1033.

Henchcliffe, C. and Beal, F. M. (2008). Mitochondrial biology and oxidative stress in Parkinson disease pathogenesis. *Nat. Clin. Pract. Neurol.* **4**, 600–609.

Herrup, K., Neve, R., Ackerman, S. L. and Copani, A. (2004). Divide and die: Cell cycle events as triggers of nerve cell death. *J. Neurosci.* **24**, 9232–9239.

Herrup, K. and Yang, Y. (2007). Cell cycle regulation in the postmitotic neuron: Oxymoron or new biology? *Nat. Rev. Neurosci.* **8**, 368–378.

Hileman, E. O., Liu, J., Albitar, M., Keating, M. J. and Huang, P. (2004). Intrinsic oxidative stress in cancer cells: A biochemical basis for therapeutic selectivity. *Cancer Chemother. Pharmacol.* **53**, 209–219.

Ho, S. R., Mahanic, C. S., Lee, Y. J. and Lin, W. C. (2014). RNF144A, an E3 ubiquitin ligase for DNA-PKcs, promotes apoptosis during DNA damage. *Proc Natl Acad Sci USA* **111**, E2646–E2655.

Hod, Y. (2004). Differential control of apoptosis by DJ-1 in prostate benign and cancer cells. *J. Cell. Biochem.* **92**, 1221–1233.

Hoffman, E. P. and Kunkel, L. M. (1989). Dystrophin abnormalities in Duchenne/Becker muscular dystrophy. *Neuron* **2**, 1019–1029.

Höglinger, G. U., Breunig, J. J., Depboylu, C., Rouaux, C., Michel, P. P., Alvarez-fischer, D., Boutillier, A., Degregori, J., Oertel, W. H., Rakic, P., et al. (2007). The pRb / E2F cell-cycle pathway mediates cell death in Parkinson's disease. *Proc Natl Acad Sci USA* **104**, 3585–3590.

Höllerhage, M., Matusch, A., Champy, P., Lombès, A., Ruberg, M., Oertel,

W. H. and Höglinger, G. U. (2009). Natural lipophilic inhibitors of mitochondrial complex I are candidate toxins for sporadic neurodegenerative tau pathologies. *Exp. Neurol.* **220**, 133–142.

Holzschuh, J., Ryu, S., Aberger, F. and Driever, W. (2001). Dopamine transporter expression distinguishes dopaminergic neurons from other catecholaminergic neurons in the developing zebrafish embryo. *Mech. Dev.* **101**, 237–243.

Honbou, K., Suzuki, N. N., Horiuchi, M., Niki, T., Taira, T., Ariga, H. and Inagaki, F. (2003). The crystal structure of DJ-1, a protein related to male fertility and Parkinson's disease. *J. Biol. Chem.* **278**, 31380–31384.

Hong, W. (2005). SNAREs and traffic. *Biochim. Biophys. Acta* **1744**, 493–517.

Hook, P. W., McClymont, S. A., Cannon, G. H., Law, W. D., Morton, A. J., Goff, L. A. and McCallion, A. S. (2018). Single-Cell RNA-Seq of Mouse Dopaminergic Neurons Informs Candidate Gene Selection for Sporadic Parkinson Disease. *Am. J. Hum. Genet.* **102**, 427–446.

Hou, S. T., Cowan, E., Walker, T., Ohan, N., Dove, M., Rasqinha, I. and MacManus, J. P. (2001). The transcription factor E2F1 promotes dopamine-evoked neuronal apoptosis by a mechanism independent of transcriptional activation. *J. Neurochem.* **78**, 287–297.

Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., Collins, J. E., Humphray, S., McLaren, K., Matthews, L., et al. (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503.

Hruska, K. S., LaMarca, M. E., Scott, C. R. and Sidransky, E. (2008). Gaucher disease: Mutation and polymorphism spectrum in the glucocerebrosidase gene (GBA). *Hum. Mutat.* **29**, 567–583.

Hsieh, F. Y., Bloch, D. A. and Larsen, M. D. (1998). A simple method of sample size calculation for linear and logistic regression. *Stat. Med.* **17**, 1623–1634.

Hsu, J. M., Lee, Y. C. G., Yu, C. T. M. and Huang, C. Y. F. (2004). Fbx7 functions in the SCF complex regulating Cdk1-cyclin B-phosphorylated hepatoma up-regulated protein (HURP) proteolysis by a proline-rich region. *J. Biol. Chem.* **279**, 32592–32602.

Hu, Q. and Wang, G. (2016). Mitochondrial dysfunction in Parkinson's disease. *Transl. Neurodegener.* **5**, 1–8.

Huang, H., Winter, E. E., Wang, H., Weinstock, K. G., Xing, H., Goodstadt, L., Stenson, P. D., Cooper, D. N., Smith, D., Albà, M. M., et al. (2004). Evolutionary conservation and selection of human disease gene orthologs in the rat and mouse genomes. *Genome Biol.* **5**, R47.

Hughes, A. J., Daniel, S. E., Kilford, L. and Lees, A. J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: A clinico-pathological study of 100 cases. *J. Neurol. Neurosurg. Psychiatry* **55**, 181–184.

Humm, A., Fritsche, E., Steinbacher, S. and Huber, R. (1997). Crystal structure and mechanism of human L-arginine:glycine amidinotransferase: A mitochondrial enzyme involved in creatine biosynthesis. *EMBO J.* **16**, 3373–3385.

Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., Peterson, R. T., Yeh, J.-R. J. and Joung, J. K. (2013a). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 227–229.

Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Kaini, P., Sander, J. D., Joung, J. K., Peterson, R. T. and Yeh, J. R. J. (2013b). Heritable and Precise Zebrafish Genome Editing Using a CRISPR-Cas System. *PLoS One* **8**, 1–9.

Ilyin, G. P., Rialland, M., Pigeon, C. and Guguen-Guillouzo, C. (2000). cDNA cloning and expression analysis of new members of the mammalian F-box protein family. *Genomics* **67**, 40–47.

Im, J., Lee, K., Woo, J., Junn, E. and Mouradian, M. M. (2012). DJ-1 induces thioredoxin 1 expression through the Nrf2 pathway. *Hum. Mol. Genet.* **21**, 3013–3024.

Imbeaud, S., Graudens, E., Boulanger, V., Barlet, X., Zaborski, P., Eveno, E., Mueller, O., Schroeder, A. and Auffray, C. (2005). Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. *Nucleic Acids Res.* **33**, 1–12.

Inden, M., Taira, T., Kitamura, Y., Yanagida, T., Tsuchiya, D., Takata, K., Yanagisawa, D., Nishimura, K., Taniguchi, T., Kiso, Y., et al. (2006). PARK7 DJ-1 protects against degeneration of nigral dopaminergic neurons in Parkinson's disease rat model. *Neurobiol. Dis.* **24**, 144–158.

Ingebretson, J. J. and Masino, M. A. (2013). Quantification of locomotor

activity in larval Zebrafish: Considerations for the design of high-throughput behavioral studies. *Front. Neural Circuits* **7**, 1–9.

Ishimi, Y., Komamura-Kohno, Y., You, Z., Omori, A. and Kitagawa, M. (2000). Inhibition of Mcm4,6,7 helicase activity by phosphorylation with cyclin A/Cdk2. *J. Biol. Chem.* **275**, 16235–16241.

Israelsen, W. J. and Vander Heiden, M. G. (2015). Pyruvate kinase: Function, regulation and role in cancer. *Semin. Cell Dev. Biol.* **43**, 43–51.

Itano, Y. and Nomura, Y. (1995). 1-Methyl-4-phenyl-pyridinium ion (MPP+) causes DNA fragmentation and increases the Bcl-2 expression in human neuroblastoma, SH-SY5Y cells, through different mechanisms. *Brain Res.* **704**, 240–245.

Itoh, K., Wakabayashi, N., Katoh, Y., Ishii, T., Igarashi, K., Engel, J. D. and Yamamoto, M. (1999). Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain. *Genes Dev.* **13**, 76–86.

Jackson-Lewis, V. and Przedborski, S. (2007). Protocol for the MPTP mouse model of Parkinson's disease. *Nat. Protoc.* **2**, 141–151.

Jain, A. K. and Mao, J. (1996). Artificial Neural Networks: A Tutorial. *Computer (Long. Beach. Calif).* **29**, 31–44.

Janezic, S., Threlfell, S., Dodson, P. D., Dowie, M. J., Taylor, T. N., Potgieter, D., Parkkinen, L., Senior, S. L., Anwar, S., Ryan, B., et al. (2013). Deficits in dopaminergic transmission precede neuron loss and dysfunction in a new Parkinson model. *Proc Natl Acad Sci USA* **110**, E4016–E4025.

Jankovic, J. (2008). Parkinson's disease : clinical features and diagnosis. *J. Neurol. Neurosurg. Physiatry* **79**, 368–376.

Jenner, P. (2008). Molecular mechanisms of L-DOPA-induced dyskinesia. *Nat. Rev. Neurosci.* **9**, 665–677.

Jiang, L., Romero-carvajal, A., Haug, J. S., Seidel, C. W. and Piotrowski, T. (2014). Gene-expression analysis of hair cell regeneration in the zebrafish lateral line. *Proc Natl Acad Sci USA* **111**, E1383–E1392.

Johansen, K. K., Torp, S. H., Farrer, M. J., Gustavsson, E. K. and Aasly, J. O. (2018). A Case of Parkinson's Disease with No Lewy Body Pathology due to a Homozygous Exon Deletion in Parkin. *Case Rep. Neurol. Med.* **2018**, 6838965.

Joo, A., Shibata, H., Ninomiya, H., Kawasaki, H., Tashiro, N. and Fukumaki, Y. (2001). Structure and polymorphisms of the human metabotropic glutamate receptor type 2 gene (GRM2): Analysis of association with schizophrenia. *Mol. Psychiatry* **6**, 186–192.

Jordan-Sciutto, K. L., Dorsey, R., Chalovich, E. M., Hammond, R. R. and Achim, C. L. (2003). Expression patterns of retinoblastoma protein in Parkinson disease. *J. Neuropathol. Exp. Neurol.* **62**, 68–74.

Joung, J. K. and Sander, J. D. (2013). TALENs: a widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.* **14**, 49–55.

Jung, H., Lee, D., Lee, J., Park, D., Kim, Y. J., Park, W. Y., Hong, D., Park, P. J. and Lee, E. (2015). Intron retention is a widespread mechanism of tumor-suppressor inactivation. *Nat. Genet.* **47**, 1242–1248.

Junn, E., Taniguchi, H., Jeong, B. S., Zhao, X., Ichijo, H. and Mouradian, M. M. (2005). Interaction of DJ-1 with Daxx inhibits apoptosis signal-regulating kinase 1 activity and cell death. *Proc Natl Acad Sci USA* **102**, 9691–9696.

Junn, E., Jang, W. H., Zhao, X., Jeong, B. S. and Mouradian, M. M. (2009). Mitochondrial localization of DJ-1 leads to enhanced neuroprotection. *J. Neurosci. Res.* **87**, 123–129.

Kadkhodaei, B., Ito, T., Joodmardi, E., Mattsson, B., Rouillard, C., Carta, M., Muramatsu, S. I., Sumi-Ichinose, C., Nomura, T., Metzger, D., et al. (2009). Nurr1 is required for maintenance of maturing and adult midbrain dopamine neurons. *J. Neurosci.* **29**, 15923–15932.

Kahle, P. J., Waak, J. and Gasser, T. (2009). DJ-1 and prevention of oxidative stress in Parkinson's disease and other age-related disorders. *Free Radic. Biol. Med.* **47**, 1354–1361.

Kaim, G. and Dimroth, P. (1999). ATP synthesis by F-type ATP synthase is obligatorily dependent on the transmembrane voltage. *EMBO J.* **18**, 4118–4127.

Kamel, A. A.-E. (2003). Bioinformatic tools and guideline for PCR primer design. *African J. Biotechnol.* **2**, 91–95.

Kane, L. A., Lazarou, M., Fogel, A. I., Li, Y., Yamano, K., Sarraf, S. A., Banerjee, S. and Youle, R. J. (2014). *PINK1 phosphorylates ubiquitin to activate Parkin E3 ubiquitin ligase activity*. **205**, 143–153.

Kapahnke, M., Banning, A. and Tikkanen, R. (2016). Random Splicing of Several Exons Caused by a Single Base Change in the Target Exon of CRISPR/Cas9 Mediated Gene Knockout. *Cells* **5**, 1–12.

Kasprick, D. S., Kish, P. E., Junttila, T. L., Ward, L. A., Bohnsack, B. L. and Kahana, A. (2011). Microanatomy of adult zebrafish extraocular muscles. *PLoS One* **6**, e27095.

Kato, I., Maita, H., Takahashi-Niki, K., Saito, Y., Noguchi, N., Iguchi-Ariga, S. M. M. and Ariga, H. (2013). Oxidized DJ-1 Inhibits p53 by Sequestering p53 from Promoters in a DNA-Binding Affinity-Dependent Manner. *Mol. Cell. Biol.* **33**, 340–359.

Kawahara, G., Karpf, J. A., Myers, J. A., Alexander, M. S., Guyone, J. R. and Kunkel, L. M. (2011). Drug screening in a zebrafish model of duchenne muscular dystrophy. *Proc Natl Acad Sci USA* **108**, 5331–5336.

Kawajiri, S., Saiki, S., Sato, S. and Hattori, N. (2011). Genetic mutations and functions of PINK1. *Trends Pharmacol. Sci.* **32**, 573–580.

Keatinge, M., Bui, H., Menke, A., Chen, Y.-C., Sokol, A. M., Bai, Q., Ellett, F., Da Costa, M., Burke, D., Gegg, M., et al. (2015). Glucocerebrosidase 1 deficient *Danio rerio* mirror key pathological aspects of human Gaucher disease and provide evidence of early microglial activation preceding alpha-synuclein-independent neuronal cell death. *Hum. Mol. Genet.* **24**, 6640–6652.

Kett, L. R., Boassa, D., Ho, C. C. Y., Rideout, H. J., Hu, J., Terada, M., Ellisman, M. and Dauer, W. T. (2012). LRRK2 Parkinson disease mutations enhance its microtubule association. *Hum. Mol. Genet.* **21**, 890–899.

Kim, R. H., Peters, M., Jang, Y., Shi, W., Pintilie, M., Fletcher, G. C., Deluca, C., Liepa, J., Zhou, L., Snow, B., et al. (2005a). DJ-1, a novel regulator of the tumor suppressor PTEN. *Cancer Cell* **7**, 263–273.

Kim, R. H., Smith, P. D., Aleyasin, H., Hayley, S., Mount, M. P., Pownall, S., Wakeham, A., You-Ten, A. J., Kalia, S. K., Horne, P., et al. (2005b). Hypersensitivity of DJ-1-deficient mice to 1-methyl-4-phenyl-1,2,3,6-tetrahydropyrindine (MPTP) and oxidative stress. *Proc Natl Acad Sci USA* **102**, 5215–5220.

Kim, S. J., Park, Y. J., Hwang, I. Y., Youdim, M. B. H., Park, K. S. and Oh, Y. J. (2012). Nuclear translocation of DJ-1 during oxidative stress-induced

neuronal cell death. *Free Radic. Biol. Med.* **53**, 936–950.

Kim, S. J., Cho, K. J. and Oh, S. (2017). Development of machine learning models for diagnosis of glaucoma. *PLoS One* **12**, e0177726.

Kim, Y.-G., Cha, J. and Chandrasegaran, S. (1996). Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci USA* **93**, 1156–1160.

Kitada, T., Asakawa, S., Hattori, N., Matsumine, H., Yamamura, Y., Minoshima, S., Yokochi, M., Mizuno, Y. and Shimizu, N. (1998). Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature* **392**, 605–608.

Kitada, T., Tong, Y., Gautier, C. A. and Shen, J. (2009). Absence of nigral degeneration in aged parkin/DJ-1/PINK1 triple knockout mice. *J. Neurochem.* **111**, 696–702.

Klein, C. and Westenberger, A. (2012). Genetics of Parkinson's disease. *Cold Spring Harb. Perspect. Med.* **2**, a008888.

Kleinbaum, D. G. and Klein, M. (2010). Introduction to Logistic Regression. In: *Logistic Regression*. New York: Springer. 1–39.

Kobayashi, A., Kang, M.-I., Okawa, H., Ohtsuji, M., Zenke, Y., Chiba, T., Igarashi, K. and Yamamoto, M. (2004). Oxidative Stress Sensor Keap1 Functions as an Adaptor for Cul3-Based E3 Ligase To Regulate Proteasomal Degradation of Nrf2. *Mol. Cell. Biol.* **24**, 7130–7139.

Koff, A., Giordano, A., Desai, D., Yamashita, K., Harper, J. W., Elledge, S., Nishimoto, T., Morgan, D. O., Franza, B. R. and Roberts, J. M. (1992). Formation and activation of a cyclin E-cdk2 complex during the G 1 phase of the human cell cycle. *Science* **257**, 1689–1694.

Koga, Y., Takahashi, H., Oikawa, D., Tachibana, T., Denbow, D. M. and Furuse, M. (2005). Brain creatine functions to attenuate acute stress responses through GABAnergic system in chicks. *Neuroscience* **132**, 65–71.

Kok, F. O., Shin, M., Ni, C. W., Gupta, A., Grosse, A. S., vanImpel, A., Kirchmaier, B. C., Peterson-Maduro, J., Kourkoulis, G., Male, I., et al. (2015). Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Dev. Cell* **32**, 97–108.

Kolk, S. M., Gunput, R.-A. F., Tran, T. S., van den Heuvel, D. M. A., Prasad,

A. A., Hellemons, A. J. C. G. M., Adolfs, Y., Ginty, D. D., Kolodkin, A. L., Burbach, J. P. H., et al. (2009). Semaphorin 3F Is a Bifunctional Guidance Cue for Dopaminergic Axons and Controls Their Fasciculation, Channeling, Rostral Growth, and Intracortical Targeting. *J. Neurosci.* **29**, 12542–12557.

Konietzny, A., Bär, J. and Mikhaylova, M. (2017). Dendritic actin cytoskeleton: Structure, functions, and regulations. *Front. Cell. Neurosci.* **11**, 147.

Konnova, E. A. and Swanberg, M. (2018). Animal Models of Parkinson's Disease. In: *Parkinson's Disease: Pathogenesis and Clinical Aspects*. Brisbane, Australia: Codon Publications. 83–106.

Koppenol, W. H., Bounds, P. L. and Dang, C. V. (2011). Otto Warburg's contributions to current concepts of cancer metabolism. *Nat. Rev. Cancer* **11**, 325–337.

Korkmaz, S., Zararsiz, G. and Goksuluk, D. (2015). MLViS: A web tool for machine learning-based virtual screening in early-phase of drug discovery and development. *PLoS One* **10**, e0124600.

Korobova, F. and Svitkina, T. (2008). Arp2/3 complex is important for filopodia formation, growth cone motility, and neuritogenesis in neuronal cells. *Mol. Biol. Cell* **19**, 1561–1574.

Korshunov, S. S., Skulachev, V. P. and Starkov, A. A. (1997). High protonic potential actuates a mechanism of production of reactive oxygen species in mitochondria. *FEBS Lett.* **416**, 15–18.

Koyano, F. and Matsuda, N. (2015). Molecular mechanisms underlying PINK1 and Parkin catalyzed ubiquitylation of substrates on damaged mitochondria. *Biochim. Biophys. Acta - Mol. Cell Res.* **1853**, 2791–2796.

Kroehne, V., Freudenreich, D., Hans, S., Kaslin, J. and Brand, M. (2011). Regeneration of the adult zebrafish brain from neurogenic radial glia-type progenitors. *Development* **138**, 4831–4841.

Krüger, R., Kuhn, W., Müller, T., Woitalla, D., Graeber, M., Kösel, S., Przuntek, H., Epplen, J. T., Schöls, L. and Riess, O. (1998). Ala30Pro mutation in the gene encoding α-synuclein in Parkinson's disease. *Nat. Genet.* **18**, 106–108.

Kruman, I. I., Wersto, R. P., Cardozo-Pelaez, F., Smilenov, L., Chan, S. L., Chrest, F. J., Emokpae, R., Gorospe, M. and Mattson, M. P. (2004).

Cell Cycle Activation Linked to Neuronal Cell Death Initiated by DNA Damage. *Neuron* **41**, 549–561.

Kulich, S. M., Horbinski, C., Patel, M. and Chu, C. T. (2007). 6-Hydroxydopamine induces mitochondrial ERK activation. *Free Radic. Biol. Med.* **43**, 372–383.

Kushnareva, Y., Murphy, A. N. and Andreyev, A. (2002). Complex I-mediated reactive oxygen species generation: Modulation by cytochrome c and NAD(P)+ oxidation-reduction state. *Biochem. J.* **368**, 545–553.

Lahiri, S. D., Wang, P.-F., Babbitt, P. C., McLeish, M. J., Kenyon, G. L. and Allen, K. N. (2002). The 2.1 Å Structure of Torpedo californica Creatine Kinase Complexed with the ADP-Mg2+ −NO 3- −Creatine Transition-State Analogue Complex. *Biochemistry* **41**, 13861–13867.

Laman, H., Funes, J. M., Ye, H., Henderson, S., Galinanes-Garcia, L., Hara, E., Knowles, P., McDonald, N. and Boshoff, C. (2005). Transforming activity of Fbxo7 is mediated specifically through regulation of cyclin D/cdk6. *EMBO J.* **24**, 3104–3116.

Lambert, A. M., Bonkowsky, J. L. and Masino, M. A. (2012). The Conserved Dopaminergic Diencephalospinal Tract Mediates Vertebrate Locomotor Development in Zebrafish Larvae. *J. Neurosci.* **32**, 13488–13500.

Lan, A. P., Chen, J., Zhao, Y., Chai, Z. and Hu, Y. (2017). mTOR Signaling in Parkinson's Disease. *Neuromolecular Med.* **19**, 1–10.

Langenau, D. M., Traver, D., Ferrando, A. A., Kutok, J. L., Aster, J. C., Kanki, J. P., Lin, S., Prochownik, E., Trede, N. S., Zon, L. I., et al. (2003). Myc-induced T cell leukemia in transgenic zebrafish. *Science* **299**, 887–890.

Langston, J., Ballard, P., Tetrud, J. and Irwin, I. (1983). Chronic Parkinsonism in humans due to a product of meperidine-analog synthesis. *Science* **219**, 979–980.

LaVoie, M. J., Ostaszewski, B. L., Weihofen, A., Schlossmacher, M. G. and Selkoe, D. J. (2005). Dopamine covalently modifies and functionally inactivates parkin. *Nat. Med.* **11**, 1214–1221.

Lawler, J. M., Barnes, W. S., Wu, G., Song, W. and Demaree, S. (2002). Direct antioxidant properties of creatine. *Biochem. Biophys. Res. Commun.* **290**, 47–52.

Lee, F. J. S., Liu, F., Pristupa, Z. B. and Niznik, H. B. (2001). Direct binding

and functional coupling of α-synuclein to the dopamine transporters accelerate dopamine-induced apoptosis. *FASEB J.* **15**, 916–926.

Lees, A. J., Hardy, J., Revesz, T. and Lila, R. (2009). Parkinson ' s disease. *Lancet* **373**, 2055–2066.

Legocki, R. P. and Verma, D. P. S. (1981). Multiple immunoreplica technique: Screening for specific proteins with a series of different antibodies using one polyacrylamide gel. *Anal. Biochem.* **111**, 385–392.

Lenkowski, J. R., Qin, Z., Sifuentes, C. J., Thummel, R., Soto, C. M., Moens, C. B. and Raymond, P. A. (2013). Retinal regeneration in adult zebrafish requires regulation of TGFβ signaling. *Glia* **61**, 1687–1697.

Lev, N., Ickowicz, D., Barhum, Y., Lev, S., Melamed, E. and Offen, D. (2009). DJ-1 protects against dopamine toxicity. *J. Neural Transm.* **116**, 151–160.

Lever, J., Krzywinski, M. and Altman, N. (2017). Points of Significance: Principal component analysis. *Nat. Methods* **14**, 641–642.

Levinthal, D. J. and DeFranco, D. B. (2005). Reversible oxidation of ERK-directed protein phosphatases drives oxidative toxicity in neurons. *J. Biol. Chem.* **280**, 5875–5883.

Li, J., Witten, D. M., Johnstone, I. M. and Tibshirani, R. (2012). Normalization, testing, and false discovery rate estimation for RNA-sequencing data. *Biostatistics* **13**, 523–538.

Li, J., Dani, J. A. and Le, W. (2009). The role of transcription factor Pitx3 in dopamine neuron development and Parkinson's disease. *Curr. Top. Med. Chem.* **9**, 855–859.

Li, M. and Arner, A. (2015). Immobilization of dystrophin and laminin α2-chain deficient zebrafish larvae in vivo prevents the development of muscular dystrophy. *PLoS One* **10**, e0139483.

Liang, Z., Zhan, Y., Shen, Y., Wong, C. C. L., Yates, J. R., Plattner, F., Lai, K. O. and Ip, N. Y. (2016). The pseudokinase CaMKv is required for the activity-dependent maintenance of dendritic spines. *Nat. Commun.* **7**, 1–13.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P. and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* **1**, 417–425.

Lieschke, G. J. and Currie, P. D. (2007). Animal models of human disease:

zebrafish swim into view. *Nat. Rev. Genet.* **8**, 353–367.

Lill CM. (2016). Genetics of Parkinson's disease, Molecular and Cellular Probes. *Handb. Clin. Neurol.* **147**, 211–227.

Liu, J., Zhou, Y., Qi, X., Chen, J., Chen, W., Qiu, G., Wu, Z. and Wu, N. (2017). CRISPR/Cas9 in zebrafish: an efficient combination for human genetic diseases modeling. *Hum. Genet.* **136**, 1–12.

Liu, Z., Wang, X., Yu, Y., Li, X., Wang, T., Jiang, H., Ren, Q., Jiao, Y., Sawa, A., Moran, T., et al. (2008). A Drosophila model for LRRK2-linked parkinsonism. *Proc Natl Acad Sci USA* **105**, 2693–2698.

Lohmann, K. and Klein, C. (2008). Genetics of Parkinson disease. *Contin. Lifelong Learn. Neurol.* **14**, 90–113.

Lones, M. ., Smith, S. L., Alty, J. E., Lacy, S. E., Possin, K. L., Jamieson, D. R. S. and Tyrrell, A. (2014). Evolving Classifiers to Recognize the Movement Characteristics of Parkinson's Disease Patients. *IEEE Trans. Evol. Comput.* **18**, 559–576.

Long, Q., Meng, A., Wang, H., Jessen, J. R., Farrell, M. J. and Lin, S. (1997). GATA-1 expression pattern can be recapitulated in living transgenic zebrafish using GFP reporter gene. *Development* **124**, 4105–4111.

Lu, L., Sun, X., Liu, Y., Zhao, H., Zhao, S. and Yang, H. (2012). DJ-1 upregulates tyrosine hydroxylase gene expression by activating its transcriptional factor Nurr1 via the ERK1/2 pathway. *Int. J. Biochem. Cell Biol.* **44**, 65–71.

Lücking, C. B., Dürr, A., Bonifati, V., Vaughan, J., De Michele, G., Gasser, T., Harhangi, B. S., Meco, G., Denèfle, P., Wood, N. W., et al. (2000). Association between early-onset Parkinson's disease and mutations in the parkin gene. *N. Engl. J. Med.* **342**, 1560–1567.

Ludtmann, M. H. R., Kostic, M., Horne, A., Gandhi, S., Sekler, I. and Abramov, A. Y. (2019). LRRK2 deficiency induced mitochondrial Ca 2+ efflux inhibition can be rescued by Na + /Ca 2+ /Li + exchanger upregulation. *Cell Death Dis.* **10**, 265.

Lundberg, A. S. and Weinberg, R. A. (1998). Functional inactivation of the retinoblastoma protein requires sequential modification by at least two distinct cyclin-cdk complexes. *Mol. Cell. Biol.* **18**, 753–761.

Luo, S. X., Timbang, L., Kim, J. I., Shang, Y., Sandoval, K., Tang, A. A., Whistler, J. L., Ding, J. B. and Huang, E. J. (2016). TGF-β Signaling in

Dopaminergic Neurons Regulates Dendritic Growth, Excitatory-Inhibitory Synaptic Balance, and Reversal Learning. *Cell Rep.* **17**, 3233–3245.

Lwin, A., Orvisky, E., Goker-Alpan, O., LaMarca, M. E. and Sidransky, E. (2004). Glucocerebrosidase mutations in subjects with parkinsonism. *Mol. Genet. Metab.* **81**, 70–73.

Lynch-Day, M. A., Mao, K., Wang, K., Zhao, M. and Klionsky, D. J. (2012). The Role of Autophagy in Parkinson's Disease. *Cold Spring Harb. Perspect. Med.* **2**, a009357.

Ma, S. Y., Röyttä, M., Rinne, J. O., Collan, Y. and Rinne, U. K. (1997). Correlation between neuromorphometry in the substantia nigra and clinical features in Parkinson's disease using disector counts. *J. Neurol. Sci.* **151**, 83–87.

Ma, S. Y., Ciliax, B. J., Stebbins, G., Jaffar, S., Joyce, J. N., Cochran, E. J., Kordower, J. H., Mash, D. C., Levey, A. I. and Mufson, E. J. (1999). Dopamine transporter-immunoreactive neurons decrease with age in the human substantia nigra. *J. Comp. Neurol.* **409**, 25–37.

Ma, Z., Zhu, P., Shi, H., Guo, L., Zhang, Q., Chen, Y., Chen, S., Zhang, Z., Peng, J. and Chen, J. (2019). PTC-bearing mRNA elicits a genetic compensation response via Upf3a and COMPASS components. *Nature* **568**, 259–263.

MacKeigan, J. P., Clements, C. M., Lich, J. D., Pope, R. M., Hod, Y. and Ting, J. P.-Y. (2003). Proteomic Profiling Drug-Induced Apoptosis in Non-Small Cell Lung Carcinoma: Identification of RS/DJ-1 and RhoGDIalpha. *Cancer Res.* **63**, 6928–6934.

Magistretti, P. J. and Allaman, I. (2015). A Cellular Perspective on Brain Energy Metabolism and Functional Imaging. *Neuron* **86**, 883–901.

Maimon, O. and Rokach, L. (2008). *Soft computing for knowledge discovery and data mining*. New York: Springer.

Majeski, A. E. and Fred Dice, J. (2004). Mechanisms of chaperone-mediated autophagy. *Int. J. Biochem. Cell Biol.* **36**, 2435–2444.

Mallard, F., Tang, B. L., Galli, T., Tenza, D., Saint-pol, A., Yue, X., Antony, C., Hong, W., Goud, B. and Johannes, L. (2001). Early / recycling endosomes-to-TGN transport involves two SNARE complexes and a Rab6 isoform. *J. Cell Biol.* **156**, 653–664.

Manda, K. M., Yedlapudi, D., Korukonda, S., Bojja, S. and Kalivendi, S. V.

(2014). The Chaperone-Like Activity of α-Synuclein Attenuates Aggregation of Its Alternatively Spliced Isoform, 112-Synuclein In Vitro: Plausible Cross-Talk between Isoforms in Protein Aggregation. *PLoS One* **9**, e98657.

Manlio Díaz-García, C., Mongeon, R., Lahmann, C., Koveal, D., Zucker, H. and Yellen, G. (2017). Neuronal Stimulation Triggers Neuronal Glycolysis and Not Lactate Uptake. *Cell Metab.* **26**, 361–374.

Mann, D. J., Higgins, T., Jones, N. C. and Rozengurt, E. (1997). Differential control of cyclins D1 and D3 and the cdk inhibitor p27(Kip1) by diverse signalling pathways in Swiss 3T3 cells. *Oncogene* **14**, 1759–1766.

Manzoni, C., Mamais, A., Dihanich, S., Abeti, R., Soutar, M. P. M., Plun-Favreau, H., Giunti, P., Tooze, S. A., Bandopadhyay, R. and Lewis, P. A. (2013). Inhibition of LRRK2 kinase activity stimulates macroautophagy. *Biochim. Biophys. Acta* **1833**, 2900–2910.

Manzoni, C. (2017). The LRRK2-macroautophagy axis and its relevance to Parkinson's disease. *Biochem. Soc. Trans.* **45**, 155–162.

Marino, M. J., Wittmann, M., Bradley, S. R., Hubert, G. W., Smith, Y. and Conn, P. J. (2001). Activation of group I metabotropic glutamate receptors produces a direct excitation and disinhibition of GABAergic projection neurons in the substantia nigra pars reticulata. *J. Neurosci.* **21**, 7001–7012.

Marras, C., Schüle, B., Schuele, B., Munhoz, R. P., Rogaeva, E., Langston, J. W., Kasten, M., Meaney, C., Klein, C., Wadia, P. M., et al. (2011). Phenotype in parkinsonian and nonparkinsonian LRRK2 G2019S mutation carriers. *Neurology* **77**, 325–333.

Marsh, J. L. and Thompson, L. M. (2006). Drosophila in the Study of Neurodegenerative Disease. *Neuron* **52**, 169–178.

Martinez-Vicente, M., Talloczy, Z., Kaushik, S., Massey, A. C., Mazzulli, J., Mosharov, E. V, Hodara, R., Fredenburg, R., Wu, D.-C., Follenzi, A., et al. (2008). Dopamine-modified alpha-synuclein blocks chaperone-mediated autophagy. *J. Clin. Invest.* **118**, 777–788.

Massagué, J. (2008). TGFβ in Cancer. *Cell* **134**, 215–230.

Matsuda, N., Sato, S., Shiba, K., Okatsu, K., Saisho, K., Gautier, C. A., Sou, Y. S., Saiki, S., Kawajiri, S., Sato, F., et al. (2010). PINK1 stabilized by mitochondrial depolarization recruits Parkin to damaged mitochondria

and activates latent Parkin for mitophagy. *J. Cell Biol.* **189**, 211–221.

Matsuda, W., Furuta, T., Nakamura, K. C., Hioki, H., Fujiyama, F., Arai, R. and Kaneko, T. (2009). Single nigrostriatal dopaminergic neurons form widely spread and highly dense axonal arborizations in the neostriatum. *J. Neurosci.* **29**, 444–453.

Mazzoni, P., Shabbott, B. and Cortés, J. C. (2012). Motor control abnormalities in Parkinson's disease. *Cold Spring Harb. Perspect. Med.* **2**, a009282.

Mazzulli, J. R., Xu, Y.-H., Sun, Y., Knight, A. L., McLean, P. J., Caldwell, G. A., Sidransky, E., Grabowski, G. A. and Krainc, D. (2011). Gaucher Disease Glucocerebrosidase and α-Synuclein Form a Bidirectional Pathogenic Loop in Synucleinopathies. *Cell* **146**, 37–52.

McHenry, M. J. and Lauder, G. V. (2005). The mechanical scaling of coasting in zebrafish (Danio rerio). *J. Exp. Biol.* **208**, 2289–2301.

McIntyre, L. M., Morse, A. M., Amin, V., Lopiano, K. K., Young, L. J., Oberg, A. L. and Nuzhdin, S. V. (2011). RNA-seq: Technical variability and sampling. *BMC Genomics* **12**, 1–13.

McNally, R. S., Davis, B. K., Clements, C. M., Accavitti-Loper, M. A., Mak, T. W. and Ting, J. P. Y. (2011). DJ-1 enhances cell survival through the binding of Cezanne, a negative regulator of NF-κB. *J. Biol. Chem.* **286**, 4098–4106.

Mehravar, M., Shirazi, A., Nazari, M. and Banan, M. (2019). Mosaicism in CRISPR/Cas9-mediated genome editing. *Dev. Biol.* **445**, 156–162.

Menden, M. P., Iorio, F., Garnett, M., McDermott, U., Benes, C. H., Ballester, P. J. and Saez-Rodriguez, J. (2013). Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLoS One* **8**, e61318.

Meulener, M., Whitworth, A. J., Armstrong-Gold, C. E., Rizzu, P., Heutink, P., Wes, P. D., Pallanck, L. J. and Bonini, N. M. (2005). Drosophila DJ-1 mutants are selectively sensitive to environmental toxins associated with Parkinson's disease. *Curr. Biol.* **15**, 1572–1577.

Meulener, M. C., Xu, K., Thompson, L., Ischiropoulos, H. and Bonini, N. M. (2006). Mutational Analysis of DJ-1 in Drosophila Implicates Functional Inactivation by Oxidative. *Proc Natl Acad Sci USA* **103**, 12517–12522.

Meyer, Y., Buchanan, B. B., Vignols, F. and Reichheld, J.-P. (2009).

Thioredoxins and Glutaredoxins: Unifying Elements in Redox Biology. *Annu. Rev. Genet.* **43**, 335–367.

Mezey, E., Dehejia, A. M., Harta, G., Suchy, S. F., Nussbaum, R. L., Brownstein, M. J. and Polymeropoulos, M. H. (1998). Alpha synuclein is present in Lewy bodies in sporadic Parkinson's disease. *Mol. Psychiatry* **3**, 493–499.

Migdalska-Richards, A. and Schapira, A. H. V. (2016). The relationship between glucocerebrosidase mutations and Parkinson disease. *J. Neurochem.* **139**, 77–90.

Migheli, R., Del Giudice, M. G., Spissu, Y., Sanna, G., Xiong, Y., Dawson, T. M., Dawson, V. L., Galioto, M., Rocchitta, G., Biosa, A., et al. (2013). LRRK2 Affects Vesicle Trafficking, Neurotransmitter Extracellular Level and Membrane Receptor Localization. *PLoS One* **8**, e77198.

Miklossy, J., Arai, T., Guo, J. P., Klegeris, a, Yu, S., McGeer, E. G. and McGeer, P. L. (2006). LRRK2 expression in normal and pathologic human brain and in human cell lines. *J. Neuropathol. Exp. Neurol.* **65**, 953–963.

Miller, J. F. (2011). *Cartesian Genetic Programming (Natural Computing Series)*. Berlin: Springer.

Miller, R. M., Kiser, G. L., Kaysser-Kranich, T. M., Lockner, R. J., Palaniappan, C. and Federoff, H. J. (2006). Robust dysregulation of gene expression in substantia nigra and striatum in Parkinson's disease. *Neurobiol. Dis.* **21**, 305–313.

Miller, S. E., Sahlender, D. A., Graham, S. C., Höning, S., Robinson, M. S., Peden, A. A. and Owen, D. J. (2011). The molecular basis for the endocytosis of small R-SNAREs by the clathrin adaptor CALM. *Cell* **147**, 1118–1131.

Minc, E., De Coppet, P., Masson, P., Thiery, L., Phanie Dutertre, S., Amor-Gué, M. and Jaulin, C. (1999). The Human Copper-Zinc Superoxide Dismutase Gene (SOD1) Proximal Promoter Is Regulated by Sp1, Egr-1, and WT1 via Non-canonical Binding Sites*. *J. Biol. Chem.* **274**, 503–509.

Mirat, O., Sternberg, J. R., Severi, K. E. and Wyart, C. (2013). ZebraZoom: an automated program for high-throughput behavioral analysis and categorization. *Front. Neural Circuits* **7**, 1–12.

Mo, J. S., Kim, M. Y., Ann, E. J., Hong, J. A. and Park, H. S. (2008). DJ-1

modulates UV-induced oxidative stress signaling through the suppression of MEKK1 and cell death. *Cell Death Differ.* **15**, 1030–1041.

Molenaar, P. C. M., Wang, Z. and Newell, K. M. (2013). Compressing movement information via principal components analysis (PCA): contrasting outcomes from the time and frequency domains. *Hum. Mov. Sci.* **32**, 1495–1511.

Morrison, K. E. (2003). Parkin mutations and early onset parkinsonism. *Brain* **126**, 1250–1251.

Mou, H., Smith, J. L., Peng, L., Yin, H., Moore, J., Zhang, X. O., Song, C. Q., Sheel, A., Wu, Q., Ozata, D. M., et al. (2017). CRISPR/Cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion. *Genome Biol.* **18**, 1–8.

Mullett, S. J. and Hinkle, D. A. (2009). DJ-1 knock-down in astrocytes impairs astrocyte-mediated neuroprotection against rotenone. *Neurobiol. Dis.* **33**, 28–36.

Muñoz, P., Huenchuguala, S., Paris, I. and Segura-Aguilar, J. (2012). Dopamine Oxidation and Autophagy. *Parkinsons. Dis.* **2012**, 920953.

Murata, H., Sakaguchi, M., Jin, Y., Sakaguchi, Y., Futami, J. I., Yamada, H., Kataoka, K. and Huh, N. H. (2011). A new cytosolic pathway from a Parkinson disease-associated kinase, BRPK/PINK1: Activation of AKT via MTORC2. *J. Biol. Chem.* **286**, 7182–7189.

Murayama, E., Kissa, K., Zapata, A., Mordelet, E., Briolat, V., Lin, H.-F., Handin, R. I. and Herbomel, P. (2006). Tracing hematopoietic precursor migration to successive hematopoietic organs during zebrafish development. *Immunity* **25**, 963–975.

Muthukumaran, K., Leahy, S., Harrison, K., Sikorska, M., Sandhu, J. K., Cohen, J., Keshan, C., Lopatin, D., Miller, H., Borowy-Borowski, H., et al. (2014). Orally delivered water soluble Coenzyme Q10 (Ubisol-Q10) blocks on-going neurodegeneration in rats exposed to paraquat: Potential for therapeutic application in Parkinson's disease. *BMC Neurosci.* **15**, 21.

Nagakubo, D., Taira, T., Kitaura, H., Ikeda, M., Tamai, K., Iguchi-Ariga, S. M. and Ariga, H. (1997). DJ-1, a novel oncogene which transforms mouse NIH3T3 cells in cooperation with ras. *Biochem. Biophys. Res. Commun.* **231**, 509–513.

Nakayama, T., Fish, M. B., Fisher, M., Oomen-Hajagos, J., Thomsen, G. H. and Grainger, R. M. (2013). Simple and efficient CRISPR/Cas9-mediated targeted mutagenesis in Xenopus tropicalis. *Genesis* **51**, 835–843.

Nakayama, T., Blitz, I. L., Fish, M. B., Odeleye, A. O., Manohar, S., Cho, K. W. Y. and Grainger, R. M. (2014). Cas9-based genome editing in Xenopus tropicalis. *Methods Enzymol.* **546**, 355–375.

Narendra, D., Tanaka, A., Suen, D. F. and Youle, R. J. (2008). Parkin is recruited selectively to impaired mitochondria and promotes their autophagy. *J. Cell Biol.* **183**, 795–803.

Narendra, D. P., Jin, S. M., Tanaka, A., Suen, D. F., Gautier, C. A., Shen, J., Cookson, M. R. and Youle, R. J. (2010). PINK1 is selectively stabilized on impaired mitochondria to activate Parkin. *PLoS Biol.* **8**, e1000298.

Nasevicius, A. and Ekker, S. C. (2000). Effective targeted gene 'knockdown' in zebrafish. *Nat. Genet.* **26**, 216–220.

Nemes, S., Jonasson, J. M., Genell, A. and Steineck, G. (2009). Bias in odds ratios by logistic regression modelling and sample size. *BMC Med. Res. Methodol.* **9**, 56.

Nemudryi, A. A., Valetdinova, K. R., Medvedev, S. P. and Zakian, S. M. (2014). TALEN and CRISPR/Cas Genome Editing Systems: Tools of Discovery. *Acta Naturae* **6**, 19–40.

Neudorfer, O., Giladi, N., Elstein, D., Abrahamov, A., Turezkite, T., Aghai, E., Reches, A., Bembi, B. and Zimran, A. (1996). Occurrence of Parkinson's syndrome in type I Gaucher disease. *QJM* **89**, 691–694.

Ng, C. H., Mok, S. Z. S., Koh, C., Ouyang, X., Fivaz, M. L., Tan, E. K., Dawson, V. L., Dawson, T. M., Yu, F. and Lim, K. L. (2009). Parkin protects against LRRK2 G2019S mutant-induced dopaminergic neurodegeneration in Drosophila. *J. Neurosci.* **29**, 11257–11262.

Nguyen, P. D., Gurevich, D. B., Sonntag, C., Hersey, L., Alaei, S., Nim, H. T., Siegel, A., Hall, T. E., Rossello, F. J., Boyd, S. E., et al. (2017). Muscle Stem Cells Undergo Extensive Clonal Drift during Tissue Growth via Meox1-Mediated Induction of G2 Cell-Cycle Arrest. *Cell Stem Cell* **21**, 107–119.

Nguyen, T., Nioi, P. and Pickett, C. B. (2009). The Nrf2-antioxidant response element signaling pathway and its activation by oxidative stress. *J. Biol.*

Chem. **284**, 13291–13295.

Nicklas, W. J., Youngster, S. K., Kindt, M. V. and Heikkila, R. E. (1987). MPTP, MPP+ and mitochondrial function. *Life Sci.* **40**, 721–729.

Nicklas, W. J., Vyas, I. and Heikkila, R. E. (1985). Inhibition of NADH-linked oxidation in brain mitochondria by 1-methyl-4-phenyl-pyridine, a metabolite of the neurotoxin, 1-methyl-4-phenyl-1,2,5,6-tetrahydropyridine. *Life Sci.* **36**, 2503–2508.

Nunes, I., Tovmasian, L. T., Silva, R. M., Burke, R. E. and Goff, S. P. (2003). Pitx3 is required for development of substantia nigra dopaminergic neurons. *Proc Natl Acad Sci USA* **100**, 4245–4250.

Oh, S. E. and Mouradian, M. M. (2017). Regulation of Signal Transduction by DJ-1. *Adv Exp Med Biol.* **1037**, 97–131.

Ohtsuka, T. (1999). Hes1 and Hes5 as Notch effectors in mammalian neuronal differentiation. *EMBO J.* **18**, 2196–2207.

Okatsu, K., Oka, T., Iguchi, M., Imamura, K., Kosako, H., Tani, N., Kimura, M., Go, E., Koyano, F., Funayama, M., et al. (2012). PINK1 autophosphorylation upon membrane potential dissipation is essential for Parkin recruitment to damaged mitochondria. *Nat. Commun.* **3**, 1–10.

Olden, J. D. and Jackson, D. A. (2002). Illuminating the 'black box': A randomization approach for understanding variable contributions in artificial neural networks. *Ecol. Modell.* **154**, 135–150.

Orenstein, S. J., Kuo, S.-H., Tasset, I., Arias, E., Koga, H., Fernandez-Carasa, I., Cortes, E., Honig, L. S. and Dauer, W. (2013). Interplay of LRRK2 with chaperone-mediated autophagy. *Nat. Neurosci.* **16**, 394–406.

Paisán-Ruiz, C., Guevara, R., Federoff, M., Hanagasi, H., Sina, F., Elahi, E., Schneider, S. A., Schwingenschuh, P., Bajaj, N., Emre, M., et al. (2010). Early-onset L-dopa-responsive Parkinsonism with pyramidal signs due to ATP13A2, PLA2G6, FBXO7 and Spatacsin mutations. *Mov. Disord.* **25**, 1791–1800.

Pakkenberg, B., Moller, A., Gundersen, H. J. G., Dam, A. M. and Pakkenberg, H. (1991). The absolute number of nerve cells in substantia nigra in normal subjects and in patients with Parkinson's disease estimated with an unbiased stereological method. *J. Neurol. Neurosurg. Psychiatry* **54**, 30–33.

Palacino, J. J., Sagi, D., Goldberg, M. S., Krauss, S., Motz, C., Wacker, M., Klose, J. and Shen, J. (2004). Mitochondrial Dysfunction and Oxidative Damage in parkin-deficient Mice. *J. Biol. Chem.* **279**, 18614–18622.

Parisiadou, L., Xie, C., Cho, H. J., Lin, X., Gu, X.-L., Long, C.-X., Lobbestael, E., Baekelandt, V., Taymans, J.-M., Sun, L., et al. (2009). Phosphorylation of ezrin/radixin/moesin proteins by LRRK2 promotes the rearrangement of actin cytoskeleton in neuronal morphogenesis. *J. Neurosci.* **29**, 13971–13980.

Park, D. S., Morris, E. J., Padmanabhan, J., Shelanski, M. L., Geller, H. M. and Greene, L. A. (1998). Cyclin-dependent kinases participate in death of neurons evoked by DNA-damaging agents. *J. Cell Biol.* **143**, 457–467.

Park, E. Y. and Rho, H. M. (2002). The transcriptional activation of the human copper/zinc superoxide dismutase gene by 2,3,7,8-tetrachlorodibenzo-p-dioxin through two different regulator sites, the antioxidant responsive element and xenobiotic responsive element. *Mol. Cell. Biochem.* **240**, 47–55.

Park, J., Lee, S. B., Lee, S., Kim, Y., Song, S., Kim, S., Bae, E., Kim, J., Shong, M., Kim, J. M., et al. (2006). Mitochondrial dysfunction in Drosophila PINK1 mutants is complemented by parkin. *Nature* **441**, 1157–1161.

Parkinson's UK. (2017). *The prevalence and incidence of Parkinson's in the UK*. London: Parkinson's UK.

Parkinson's UK. (2018). *The incidence and prevalence of Parkinsons's in the UK*. London: Parkinson's UK.

Parng, C., Roy, N. M., Ton, C., Lin, Y. and McGrath, P. (2007). Neurotoxicity assessment using zebrafish. *J. Pharmacol. Toxicol. Methods* **55**, 103–112.

Parsanejad, M., Zhang, Y., Qu, D., Irrcher, I., Rousseaux, M. W. C., Aleyasin, H., Kamkar, F., Callaghan, S., Slack, R. S., Mak, T. W., et al. (2014). Regulation of the VHL/HIF-1 pathway by DJ-1. *J. Neurosci.* **34**, 8043–8050.

Partridge, L., Deelen, J. and Slagboom, P. E. (2018). Facing up to the global challenges of ageing. *Nature* **561**, 45–56.

Patton, E. E., Widlund, H. R., Kutok, J. L., Kopani, K. R., Amatruda, J. F., Murphey, R. D., Berghmans, S., Mayhall, E. A., Traver, D., Fletcher,

C. D. M., et al. (2005). BRAF mutations are sufficient to promote nevi formation and cooperate with p53 in the genesis of melanoma. *Curr. Biol.* **15**, 249–254.

Paul, R., Choudhury, A., Kumar, S., Giri, A., Sandhir, R. and Borah, A. (2017). Cholesterol contributes to dopamineneuronal loss in MPTP mouse model of Parkinson's disease: Involvement of mitochondrial dysfunctions and oxidative stress. *PLoS One* **12**, e0171285.

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R. and Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *J. Clin. Epidemiol.* **49**, 1373–1379.

Perez, F. A. and Palmiter, R. D. (2005). Parkin-deficient mice are not a robust model of parkinsonism. *Proc Natl Acad Sci USA* **102**, 2174–2179.

Petersen, B. O., Lukas, J., Sørensen, C. S., Bartek, J. and Helin, K. (1999). Phosphorylation of mammalian CDC6 by cyclin A/CDK2 regulates its subcellular localization. *EMBO J.* **18**, 396–410.

Pezzoli, G. and Cereda, E. (2013). Exposure to pesticides or solvents and risk of Parkinson disease. *Neurology* **80**, 2035–2041.

Pezzuto, A. and Carico, E. (2018). Role of HIF-1 in Cancer Progression: Novel Insights. A Review. *Curr. Mol. Med.* **18**, 343–351.

Pissadaki, E. K. and Bolam, J. P. (2013). The energy cost of action potential propagation in dopamine neurons: Clues to susceptibility in Parkinson's disease. *Front. Comput. Neurosci.* **7**, 1–17.

Pitkänen, S. and Robinson, B. H. (1996). Mitochondrial complex I deficiency leads to increased production of superoxide radicals and induction of superoxide dismutase. *J. Clin. Invest.* **98**, 345–351.

Poewe, W., Antonini, A., Zijlmans, J. C., Burkhard, P. R. and Vingerhoets, F. (2010). Levodopa in the treatment of Parkinson's disease: an old drug still going strong. *Clin. Interv. Aging* **5**, 229–238.

Poewe, W., Seppi, K., Tanner, C. M., Halliday, G. M., Brundin, P., Volkmann, J., Schrag, A.-E. and Lang, A. E. (2017). Parkinson disease. *Nat. Rev. Dis. Prim.* **3**, 17013.

Polymeropoulos, M. H., Lavedan, C., Leroy, E., Ide, S. E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R., et al. (1997). Mutation in the α-synuclein gene identified in families with Parkinson's disease. *Science* **276**, 2045–2047.

Poulopoulos, M., Levy, O. A. and Alcalay, R. N. (2012). The neuropathology of genetic Parkinson's disease. *Mov. Disord.* **27**, 831–842.

Prabhudesai, S., Bensabeur, F. Z., Abdullah, R., Basak, I., Baez, S., Alves, G., Holtzman, N. G., Larsen, J. P. and Møller, S. G. (2016). LRRK2 knockdown in zebrafish causes developmental defects, neuronal loss, and synuclein aggregation. *J. Neurosci. Res.* **94**, 717–735.

Pramod, A. B., Foster, J., Carvelli, L. and Henry, L. K. (2013). SLC6 transporters: Structure, function, regulation, disease association and therapeutics. *Mol. Aspects Med.* **34**, 197–219.

Prior, I. A., Lewis, P. D. and Mattos, C. (2012). A comprehensive survey of ras mutations in cancer. *Cancer Res.* **72**, 2457–2467.

Prykhozhij, S. V., Steele, S. L., Razaghi, B. and Berman, J. N. (2017). A rapid and effective method for screening, sequencing and reporter verification of engineered frameshift mutations in zebrafish. *Dis. Model. Mech.* **10**, 811–822.

Puopolo, M., Raviola, E. and Bean, B. P. (2007). Roles of subthreshold calcium current and sodium current in spontaneous firing of mouse midbrain dopamine neurons. *J. Neurosci.* **27**, 645–656.

Puschmann, A. (2013). Monogenic Parkinson's disease and parkinsonism: Clinical phenotypes and frequencies of known mutations. *Park. Relat. Disord.* **19**, 407–415.

van der Putten, H., Wiederhold, K.-H., Probst, A., Barbieri, S., Mistl, C., Danner, S., Kauffmann, S., Hofele, K., Spooren, W. P. J. M., Ruegg, M. A., et al. (2000). Neuropathology in Mice Expressing Human α-Synuclein. *J. Neurosci.* **20**, 6021–6029.

Quesada, A., Lee, B. Y. and Micevych, P. E. (2008). PI3 kinase/Akt activation mediates estrogen and IGF-1 nigral DA neuronal neuroprotection against a unilateral rat model of Parkinson's disease. *Dev. Neurobiol.* **68**, 632–644.

Quinn, N., Critchley, P. and Marsden, C. D. (1987). Young onset Parkinson's disease. *Mov. Disord.* **2**, 73–91.

Ramachandiran, S., Hansen, J. M., Jones, D. P., Richardson, J. R. and Miller, G. W. (2007). Divergent mechanisms of paraquat, MPP + , and rotenone toxicity: Oxidation of thioredoxin and caspase-3 activation. *Toxicol. Sci.* **95**, 163–171.

Ramonet, D., Daher, J. P. L., Lin, B. M., Stafa, K., Kim, J., Banerjee, R., Westerlund, M., Pletnikova, O., Glauser, L., Yang, L., et al. (2011). Dopaminergic Neuronal loss, Reduced Neurite Complexity and Autophagic Abnormalities in Transgenic Mice Expressing G2019S Mutant LRRK2. *PLoS One* **6**, e18568.

Reiter, L. T., Potocki, L., Chien, S., Gribskov, M. and Bier, E. (2001). A systematic analysis of human disease-associated gene sequences in Drosophila melanogaster. *Genome Res.* **11**, 1114–1125.

Ren, G., Xin, S., Li, S., Zhong, H. and Lin, S. (2011). Disruption of LRRK2 Does Not Cause Specific Loss of Dopaminergic Neurons in Zebrafish. *PLoS One* **6**, e20630.

Ren, J. P., Zhao, Y. W. and Sun, X. J. (2009). Toxic influence of chronic oral administration of paraquat on nigrostriatal dopaminergic neurons in C57BL/6 mice. *Chin. Med. J. (Engl).* **122**, 2366–2371.

Ridges, S., Heaton, W. L., Joshi, D., Choi, H., Eiring, A., Batchelor, L., Choudhry, P., Manos, E. J., Sofla, H., Sanati, A., et al. (2012). Zebrafish screen identifies novel compound with selective toxicity against leukemia. *Blood* **119**, 5621–5631.

De Rijk, M. C., Breteler, M. M. B., Graveland, G. A., Ott, A., Grobbee, D. E., Van Der Meche, F. G. A. and Hofman, A. (1995). Prevalence of parkinson's disease in the elderly: The rotterdam study. *Neurology* **45**, 2143–2146.

Riley, B. E., Lougheed, J. C., Callaway, K., Velasquez, M., Brecht, E., Nguyen, L., Shaler, T., Walker, D., Yang, Y., Regnstrom, K., et al. (2013). Structure and function of Parkin E3 ubiquitin ligase reveals aspects of RING and HECT ligases. *Nat. Commun.* **4**, 1–9.

Ringel, M. D., Hayre, N., Saito, J., Saunier, B., Schuppert, F., Burch, H., Bernet, V., Burman, K. D., Kohn, L. D. and Saji, M. (2001). Overexpression and Overactivation of Akt in Thyroid Carcinoma 1. *Cancer Res.* **61**, 6105–6111.

Ringnér, M. (2008). What is principal component analysis? *Nat. Biotechnol.* **26**, 303–304.

Rink, E. and Wullimann, M. F. (2001). The teleostean (zebrafish) dopaminergic system ascending to the subpallium (striatum) is located in the basal diencephalon (posterior tuberculum). *Brain Res.* **889**, 316–330.

Rink, E. and Wullimann, M. F. (2002). Development of the catecholaminergic system in the early zebrafish brain: An immunohistochemical study. *Dev. Brain Res.* **137**, 89–100.

Rink, E. and Wullimann, M. F. (2004). Connections of the ventral telencephalon (subpallium) in the zebrafish (Danio rerio). *Brain Res.* **1011**, 206–220.

Rizzo, G., Copetti, M., Arcuti, S., Martino, D., Fontana, A. and Logroscino, G. (2016). Accuracy of clinical diagnosis of Parkinson disease. *Neurology* **86**, 566–576.

Rodriguez-Blanco, J., Martín, V., Herrera, F., García-Santos, G., Antolín, I. and Rodriguez, C. (2008). Intracellular signaling pathways involved in post-mitotic dopaminergic PC12 cell death induced by 6-hydroxydopamine. *J. Neurochem.* **107**, 127–140.

Rojo, A. I., Salinas, M., Martín, D., Perona, R. and Cuadrado, A. (2004). Regulation of Cu/Zn-superoxide dismutase expression via the phosphatidylinositol 3 kinase/Akt pathway and nuclear factor-κB. *J. Neurosci.* **24**, 7324–7334.

Rosen, J. N., Sweeney, M. F. and Mably, J. D. (2009). Microinjection of zebrafish embryos to analyze gene function. *J. Vis. Exp.* **25**, e1115.

Rossi, A., Kontarakis, Z., Gerri, C., Nolte, H., Hölper, S., Krüger, M. and Stainier, D. Y. R. (2015). Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* **524**, 230–233.

Royes, L. F. F., Fighera, M. R., Furian, A. F., Oliveira, M. S., Fiorenza, N. G., Ferreira, J., da Silva, A. C., Priel, M. R., Ueda, É. S., Calixto, J. B., et al. (2008). Neuromodulatory effect of creatine on extracellular action potentials in rat hippocampus: Role of NMDA receptors. *Neurochem. Int.* **53**, 33–37.

Ryder, E. and Russell, S. (2003). Transposable elements as tools for genomics and genetics in Drosophila. *Briefings Funct. Genomics Proteomics* **2**, 57–71.

Sager, J. J., Bai, Q. and Burton, E. A. (2010). Transgenic zebrafish models of neurodegenerative diseases. *Brain Struct. Funct.* **214**, 285–302.

Sainani, K. (2010). The Importance of Accounting for Correlated Observations. *PM R* **2**, 858–861.

Saito, Y., Hatsuta, H., Hayashi, K., Mita, Y., Noguchi, N., Miyasaka, T.,

Ihara, Y., Murayama, S., Kusano-Arai, O., Iwanari, H., et al. (2014). Immunostaining of oxidized DJ-1 in human and mouse brains. *J. Neuropathol. Exp. Neurol.* **73**, 714–728.

Sakai, A., Thieblemont, C., Wellmann, A., Jaffe, E. S. and Raffeld, M. (1998). PTEN Gene Alterations in Lymphoid Neoplasms. *Blood* **92**, 3410–3415.

Sallinen, V., Torkko, V., Sundvik, M., Reenilä, I., Khrustalyov, D., Kaslin, J. and Panula, P. (2009). MPTP and MPP+ target specific aminergic cell populations in larval zebrafish. *J. Neurochem.* **108**, 719–731.

Sanchez-Simon, F. M., Zhang, X. X., Loh, H. H., Law, P.-Y. and Rodriguez, R. E. (2010). Morphine regulates dopaminergic neuron differentiation via miR-133b. *Mol. Pharmacol.* **78**, 935–942.

Sandell, L. L., Guan, X., Ingram, R. and Tilghman, S. M. (2003). Gatm, a creatine synthesis enzyme, is imprinted in mouse placenta. *Proc Natl Acad Sci USA* **100**, 4622–4627.

Santoriello, C. and Zon, L. I. (2012). Hooked! modeling human disease in zebrafish. *J. Clin. Invest.* **122**, 2337–2343.

Saporito, M. S., Thomas, B. A. and Scott, R. W. (2002). MPTP Activates c-Jun NH2-Terminal Kinase (JNK) and Its Upstream Regulatory Kinase MKK4 in Nigrostriatal Neurons In Vivo. *J. Neurochem.* **75**, 1200–1208.

Sato, M., Morimoto, K., Kajihara, S., Tateishi, R., Shiina, S., Koike, K. and Yatomi, Y. (2019). Machine-learning Approach for the Development of a Novel Predictive Model for the Diagnosis of Hepatocellular Carcinoma. *Sci. Rep.* **9**, 1–7.

Sazanov, L. A. (2015). A giant molecular proton pump: Structure and mechanism of respiratory complex I. *Nat. Rev. Mol. Cell Biol.* **16**, 375–388.

Schapira, A. H. V., Cooper, J. M., Dexter, D., Clark, J. B., Jenner, P. and Marsden, C. D. (1990). Mitochondrial Complex I Deficiency in Parkinson's Disease. *J. Neurochem.* **54**, 823–827.

Schapira, A. H. V. (2013). Calcium dysregulation in Parkinson's disease. *Brain* **136**, 2015–2016.

Schiavone, M., Zulian, A., Menazza, S., Petronilli, V., Argenton, F., Merlini, L., Sabatelli, P. and Bernardi, P. (2017). Alisporivir rescues defective mitochondrial respiration in Duchenne muscular dystrophy. *Pharmacol.*

Res. **125**, 122–131.

Seervi, M., Joseph, J., Sobhan, P. K., Bhavya, B. C. and Santhoshkumar, T. R. (2011). Essential requirement of cytochrome c release for caspase activation by procaspase-activating compound defined by cellular models. *Cell Death Dis.* **2**, e207.

Seidl, A. H. (2014). Regulation of conduction time along axons. *Neuroscience* **276**, 126–134.

Semenza, G. L. (2011). Hypoxia-inducible factor 1: Regulator of mitochondrial metabolism and mediator of ischemic preconditioning. *Biochim. Biophys. Acta - Mol. Cell Res.* **1813**, 1263–1268.

Sette, S. and Boullart, L. (2001). Genetic programming: Principles and applications. *Eng. Appl. Artif. Intell.* **14**, 727–736.

Shendelman, S., Jonason, A., Martinat, C., Leete, T. and Abeliovich, A. (2004). DJ-1 is a redox-dependent molecular chaperone that inhibits alpha-synuclein aggregate formation. *PLoS Biol.* **2**, e362.

Sheng, D., Qu, D., Kwok, K. H. H., Ng, S. S., Lim, A. Y. M., Aw, S. S., Lee, C. W. H., Sung, W. K., Tan, E. K., Lufkin, T., et al. (2010). Deletion of the WD40 Domain of LRRK2 in Zebrafish Causes Parkinsonism-Like Loss of Neurons and Locomotive Defect. *PLoS Genet.* **6**, e1000914.

Shimizu, Y., Sollu, C., Meckler, J. F. and Adriaenssens, A. (2011). Adding Fingers To An Engineered Zinc Finger Nuclease Can Reduce Activity. *Biochemistry* **50**, 5033–5041.

Shimoji, M., Zhang, L., Mandir, A. S., Dawson, V. L. and Dawson, T. M. (2005). Absence of inclusion body formation in the MPTP mouse model of Parkinson's disease. *Mol. Brain Res.* **134**, 103–108.

Shimura, H., Hattori, N., Kubo, S. I., Mizuno, Y., Asakawa, S., Minoshima, S., Shimizu, N., Iwai, K., Chiba, T., Tanaka, K., et al. (2000). Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase. *Nat. Genet.* **25**, 302–305.

Shojaee, S., Sina, F., Banihosseini, S. S., Kazemi, M. H., Kalhor, R., Shahidi, G. A., Fakhrai-Rad, H., Ronaghi, M. and Elahi, E. (2008). Genome-wide Linkage Analysis of a Parkinsonian-Pyramidal Syndrome Pedigree by 500 K SNP Arrays. *Am. J. Hum. Genet.* **82**, 1375–1384.

Sidransky, E. and Lopez, G. (2012). The link between the GBA gene and parkinsonism. *Lancet Neurol.* **11**, 986–998.

Simillion, C., Liechti, R., Lischer, H. E. L., Ioannidis, V. and Bruggmann, R. (2017). Avoiding the pitfalls of gene set enrichment analysis with SetRank. *BMC Bioinformatics* **18**, 151.

Simola, N., Morelli, M. and Carta, A. R. (2007). The 6-hydroxydopamine model of Parkinson's disease. *Neurotox. Res.* **11**, 151–167.

Singleton, A. B., Farrer, M., Johnson, J., Singleton, A., Hague, S., Kachergus, J., Hulihan, M., Peuralinna, T., Dutra, A., Nussbaum, R., et al. (2003). α-Synuclein Locus Triplication Causes Parkinson's Disease. *Science* **302**, 841.

Sivanandam, S. N. and Deepa, S. N. (2008). Genetic Algorithms. In: *Introduction to Genetic Algorithms*. Berlin: Springer. 15–37.

Skowyra, D., Craig, K. L., Tyers, M., Elledge, S. J. and Harper, J. W. (1997). F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. *Cell* **91**, 209–219.

Smith, J. (2000). Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res.* **28**, 3361–3369.

Smith, S. L., Gaughan, P., Halliday, D. M., Ju, Q., Aly, N. M. and Playfer, J. R. (2007). Diagnosis of Parkinson's disease using evolutionary algorithms. *Genet. Program. Evolvable Mach.* **8**, 433–447.

Smith, S. L., Lones, M. A., Ivanoiu, D., Bedder, M., Cording, A., Maguire, R. J., Pownall, M. E., Lyle, C., Alty, J. E., Elliott, C. J. H., et al. (2015). Computational approaches for understanding the diagnosis and treatment of Parkinson's disease. *IET Syst. Biol.* **9**, 226–233.

Song, I. K., Lee, J. J., Cho, J. H., Jeong, J., Shin, D. H. and Lee, K. J. (2016). Degradation of redox-sensitive proteins including peroxiredoxins and DJ-1 is promoted by oxidation-induced conformational changes and ubiquitination. *Sci. Rep.* **6**, 1–15.

Sorkina, T., Richards, T. L., Rao, A., Zahniser, N. R. and Sorkin, L. (2009). Negative regulation of dopamine transporter endocytosis by membrane-proximal N-terminal residues. *J. Neurosci.* **29**, 1361–1374.

Spillantini, M. G., Schmidt, M. L., Lee, V. M. Y., Trojanowski, J. Q., Jakes, R. and Goedert, M. (1997). α-synuclein in Lewy bodies. *Nature* **388**, 839–840.

Spillantini, M. G., Crowther, R. A., Jakes, R., Hasegawa, M. and Goedert, M.

(1998). alpha-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with lewy bodies. *Proc Natl Acad Sci USA* **95**, 6469–6473.

Spina, M. B. and Cohen, G. (1989). Dopamine turnover and glutathione oxidation: Implications for Parkinson disease. *Proc Natl Acad Sci USA* **86**, 1398–1400.

Sriram, K., Pai, K. S., Boyd, M. R. and Ravindranath, V. (1997). Evidence for generation of oxidative stress in brain by MPTP: in vitro and in vivo studies in mice. *Brain Res.* **749**, 44–52.

Stark, G. R. and Taylor, W. R. (2004). Analyzing the G2/M checkpoint. *Methods Mol. Biol.* **280**, 51–82.

Starkov, A. A. and Fiskum, G. (2003). Regulation of brain mitochondrial H2O2 production by membrane potential and NAD(P)H redox state. *J. Neurochem.* **86**, 1101–1107.

Stoltzfus, J. C. (2011). Logistic regression: A brief primer. *Acad. Emerg. Med.* **18**, 1099–1104.

Stone, O. A., El-Brolosy, M., Wilhelm, K., Liu, X., Romão, A. M., Grillo, E., Lai, J. K. H., Günther, S., Jeratsch, S., Kuenne, C., et al. (2018). Loss of pyruvate kinase M2 limits growth and triggers innate immune signaling in endothelial cells. *Nat. Commun.* **9**, 1–12.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**, 15545–15550.

Sulzer, D., Bogulavsky, J., Larsen, K. E., Behr, G., Karatekin, E., Kleinman, M. H., Turro, N., Krantz, D., Edwards, R. H., Greene, L. A., et al. (2000). Neuromelanin biosynthesis is driven by excess cytosolic catecholamines not accumulated by synaptic vesicles. *Proc Natl Acad Sci USA* **97**, 11869–11874.

Sun, M., Wang, G., Paciga, J. E., Feldman, R. I., Yuan, Z. Q., Ma, X. L., Shelley, S. A., Jove, R., Tsichlis, P. N., Nicosia, S. V., et al. (2001). AKT1/PKBα kinase is frequently elevated in human cancers and its constitutive activation is required for oncogenic transformation in NIH3T3 cells. *Am. J. Pathol.* **159**, 431–437.

Sun, Q., Chen, X., Ma, J., Peng, H., Wang, F., Zha, X., Wang, Y., Jing, Y., Yang, H., Chen, R., et al. (2011). Mammalian target of rapamycin up-regulation of pyruvate kinase isoenzyme type M2 is critical for aerobic glycolysis and tumor growth. *Proc Natl Acad Sci USA* **108**, 4129–4134.

Surmeier, D. J., Obeso, J. A. and Halliday, G. M. (2017). Selective neuronal vulnerability in Parkinson disease. *Nat. Rev. Neurosci.* **18**, 101–113.

Surmeier, D. J. and Schumacker, P. T. (2013). Calcium, bioenergetics, and neuronal vulnerability in Parkinson's disease. *J. Biol. Chem.* **288**, 10736–10741.

Suryadinata, R., Sadowski, M. and Sarcevic, B. (2010). Control of cell cycle progression by phosphorylation of cyclin-dependent kinase (CDK) substrates. *Biosci. Rep.* **30**, 243–255.

Taguchi, Y. V., Liu, J., Ruan, J., Pacheco, J., Zhang, X., Abbasi, J., Keutzer, J., Mistry, P. K. and Chandra, S. S. (2017). Glucosylsphingosine promotes α-synuclein pathology in mutant GBA-associated parkinson's disease. *J. Neurosci.* **37**, 9617–9631.

Tamayo, P., Steinhardt, G., Liberzon, A. and Mesirov, J. P. (2016). The limitations of simple gene set enrichment analysis assuming gene independence. *Stat. Methods Med. Res.* **25**, 472–487.

Tarca, A. L., Carey, V. J., Chen, X. wen, Romero, R. and Drăghici, S. (2007). Machine learning and its applications to biology. *PLoS Comput. Biol.* **3**, e116.

Tarnopolsky, M. A. and Beal, M. F. (2001). Potential for creatine and other therapies targeting cellular energy dysfunction in neurological disorders. *Ann. Neurol.* **49**, 561–574.

Tatton, N. A. (2000). Increased caspase 3 and Bax immunoreactivity accompany nuclear GAPDH translocation and neuronal apoptosis in Parkinson's disease. *Exp. Neurol.* **166**, 29–43.

Tay, T. L., Ronneberger, O., Ryu, S., Nitschke, R. and Driever, W. (2011). Comprehensive catecholaminergic projectome analysis reveals single-neuron integration of zebrafish ascending and descending dopaminergic systems. *Nat. Commun.* **2**, 1–12.

Tayebi, N., Callahan, M., Madike, V., Stubblefield, B. K., Orvisky, E., Krasnewich, D., Fillano, J. J. and Sidransky, E. (2001). Gaucher disease and parkinsonism: A phenotypic and genotypic characterization.

Mol. Genet. Metab. **73**, 313–321.

Thiery, J. P. (2002). Epithelial–mesenchymal transitions in tumour progression. *Nat. Rev. Cancer* **2**, 442–454.

Thisse, B. and Thisse, C. (2004). *Fast Release Clones: A High Throughput Expression Analysis.*

Thomas, B. and Beal, M. F. (2007). Parkinson's disease. *Hum. Mol. Genet.* **16**, R183–R194.

Tian, M., Cui, Y. Z., Song, G. H., Zong, M. J., Zhou, X. Y., Chen, Y. and Han, J. X. (2008). Proteomic analysis identifies MMP-9, DJ-1 and A1BG as overexpressed proteins in pancreatic juice from pancreatic ductal adenocarcinoma patients. *BMC Cancer* **8**, 1–11.

Timmons, S., Coakley, M. F., Moloney, A. M. and O'Neill, C. (2009). Akt signal transduction dysfunction in Parkinson's disease. *Neurosci. Lett.* **467**, 30–35.

Tsai, S. Q., Zheng, Z., Nguyen, N. T., Liebers, M., Topkar, V. V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A. J., Le, L. P., et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–198.

Tsang, C. K., Chen, M., Cheng, X., Qi, Y., Chen, Y., Das, I., Li, X., Vallat, B., Fu, L. W., Qian, C. N., et al. (2018). SOD1 Phosphorylation by mTORC1 Couples Nutrient Sensing and Redox Regulation. *Mol. Cell* **70**, 502–515.

Tyagi, S., Gupta, P., Saini, A., Kaushal, C. and Sharma, S. (2011). The peroxisome proliferator-activated receptor: A family of nuclear receptors role in various diseases. *J. Adv. Pharm. Technol. Res.* **2**, 236–240.

Udvardi, M. K., Czechowski, T. and Scheible, W. R. (2008). Eleven golden rules of quantitative RT-PCR. *Plant Cell* **20**, 1736–1737.

Ugur, B., Chen, K. and Bellen, H. J. (2016). Drosophila tools and assays for the study of human diseases. *Dis. Model. Mech.* **9**, 235–244.

Umek, N., Geršak, B., Vintar, N., Šoštarič, M. and Mavri, J. (2018). Dopamine autoxidation is controlled by acidic pH. *Front. Mol. Neurosci.* **11**, 1–8.

Unoki, M. and Nakamura, Y. (2001). Growth-suppressive effects of BPOZ and EGR2, two genes involved in the PTEN signaling pathway. *Oncogene* **20**, 4457–4465.

Valente, E. M., Abou-Sleiman, P. M., Caputo, V., Muqit, M. M. K., Harvey, K.,

Gispert, S., Ali, Z., Del Turco, D., Bentivoglio, A. R., Healy, D. G., et al. (2004). Hereditary early-onset Parkinson's disease caused by mutations in PINK1. *Science* **304**, 1158–1160.

Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., et al. (2019). Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* **18**, 463–477.

Várkonyi, J., Rosenbaum, H., Baumann, N., MacKenzie, J. J., Simon, Z., Aharon-Peretz, J., Walker, J. M., Tayebi, N. and Sidransky, E. (2003). Gaucher disease associated with parkinsonism: four further case reports. *Am. J. Med. Genet. A.* **116A**, 348–351.

Vasseur, S., Afzal, S., Tardivel-Lacombe, J., Park, D. S., Iovanna, J. L. and Mak, T. W. (2009). DJ-1/PARK7 is an important mediator of hypoxia-induced cellular responses. *Proc Natl Acad Sci USA* **106**, 1111–1116.

Venderova, K., Kabbach, G., Abdel-Messih, E., Zhang, Y., Parks, R. J., Imai, Y., Gehrke, S., Ngsee, J., Lavoie, M. J., Slack, R. S., et al. (2009). Leucine-rich repeat kinase 2 interacts with Parkin, DJ-1 and PINK-1 in a Drosophila melanogaster model of Parkinson's disease. *Hum. Mol. Genet.* **18**, 4390–4404.

Venugopal, R. and Jaiswal, A. K. (1996). Nrf1 and Nrf2 positively and c-Fos and Fra1 negatively regulate the human antioxidant response element-mediated expression of NAD(P)H:quinone oxidoreductase1 gene. *Proc Natl Acad Sci USA* **93**, 14960–14965.

Vigneron, S., Sundermann, L., Labbé, J. C., Pintard, L., Radulescu, O., Castro, A. and Lorca, T. (2018). Cyclin A-cdk1-Dependent Phosphorylation of Bora Is the Triggering Factor Promoting Mitotic Entry. *Dev. Cell* **45**, 637–650.

Vijayanathan, Y., Lim, F. T., Lim, S. M. and Long, C. M. (2017). 6-OHDA-Lesioned Adult Zebrafish as a Useful Parkinson ' s Disease Model for Dopaminergic Neuroregeneration. *Neurotox. Res.* **32**, 496–508.

Vikhar, P. A. (2016). Evolutionary Algorithms : A Critical Review and its Future Prospects. In: *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)*. 2016. Jalgaon, India: IEEE. 261–265.

Vingill, S., Brockelt, D., Lancelin, C., Tatenhorst, L., Dontcheva, G.,

Preisinger, C., Schwedhelm-Domeyer, N., Joseph, S., Mitkovski, M., Goebbels, S., et al. (2016). Loss of FBXO7 (PARK 15) results in reduced proteasome activity and models a parkinsonism-like phenotype in mice. *EMBO J.* **35**, 2008–2025.

Vogiatzi, T., Xilouri, M., Vekrellis, K. and Stefanis, L. (2008). Wild type α-synuclein is degraded by chaperone-mediated autophagy and macroautophagy in neuronal cells. *J. Biol. Chem.* **283**, 23542–23556.

Waak, J., Weber, S. S., Görner, K., Schall, C., Ichijo, H., Stehle, T. and Kahle, P. J. (2009). Oxidizable residues mediating protein stability and cytoprotective interaction of DJ-1 with apoptosis signal-regulating kinase 1. *J. Biol. Chem.* **284**, 14245–14257.

Walczak, C. E. and Heald, R. (2008). Mechanisms of mitotic spindle assembly and function. *Int. Rev. Cytol.* **265**, 111–158.

Wan, P. T. C., Garnett, M. J., Roe, S. M., Lee, S., Niculescu-Duvaz, D., Good, V. M., Project, C. G., Jones, C. M., Marshall, C. J., Springer, C. J., et al. (2004). Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. *Cell* **116**, 855–867.

Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F. and Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/cas-mediated genome engineering. *Cell* **153**, 910–918.

Wang, T., Wei, J. J., Sabatini, D. M. and Lander, E. S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84.

Wang, W., Bu, B., Xie, M., Zhang, M., Yu, Z. and Tao, D. (2009). Neural cell cycle dysregulation and central nervous system diseases. *Prog. Neurobiol.* **89**, 1–17.

Wang, X., Yan, M. H., Fujioka, H., Liu, J., Wilson-delfosse, A., Chen, S. G., Perry, G., Casadesus, G. and Zhu, X. (2012). LRRK2 regulates mitochondrial dynamics and function through direct interaction with DLP1. *Hum. Mol. Genet.* **21**, 1931–1944.

Wang, Y., Liu, W., Yang, J., Wang, F., Sima, Y., Zhong, Z. min, Wang, H., Hu, L. F. and Liu, C. F. (2017). Parkinson's disease-like motor and non-motor symptoms in rotenone-treated zebrafish. *Neurotoxicology* **58**, 103–109.

Wang, Z., Liu, J., Chen, S., Wang, Y., Cao, L., Zhang, Y., Kang, W., Li, H., Gui, Y., Chen, S., et al. (2011). DJ-1 modulates the expression of Cu/Zn-superoxide dismutase-1 through the Erk1/2-Elk1 pathway in neuroprotection. *Ann. Neurol.* **70**, 591–600.

Waragai, M., Wei, J., Fujita, M., Nakai, M., Ho, G. J., Masliah, E., Akatsu, H., Yamada, T. and Hashimoto, M. (2006). Increased level of DJ-1 in the cerebrospinal fluids of sporadic Parkinson's disease. *Biochem. Biophys. Res. Commun.* **345**, 967–972.

Waragai, M., Nakai, M., Wei, J., Fujita, M., Mizuno, H., Ho, G., Masliah, E., Akatsu, H., Yokochi, F. and Hashimoto, M. (2007). Plasma levels of DJ-1 as a possible marker for progression of sporadic Parkinson's disease. *Neurosci. Lett.* **425**, 18–22.

Warner, T. T., Schapira, A. H. V., Tatton, Rascol, Kordower, Olanow, Beal, Marek, Stocchi and Isacson. (2003). Genetic and environmental factors in the cause of Parkinson's disease. *Ann. Neurol.* **53**, S16–S25.

Waugh, T. A., Horstick, E., Hur, J., Jackson, S. W., Davidson, A. E., Li, X. and Dowling, J. J. (2014). Fluoxetine prevents dystrophic changes in a zebrafish model of duchenne muscular dystrophy. *Hum. Mol. Genet.* **23**, 4651–4662.

Wen, L., Wei, W., Gu, W., Huang, P., Ren, X., Zhang, Z., Zhu, Z., Lin, S. and Zhang, B. (2008). Visualization of monoaminergic neurons and neurotoxicity of MPTP in live transgenic zebrafish. *Dev. Biol.* **314**, 84–92.

West, A. B., Moore, D. J., Biskup, S., Bugayenko, A., Smith, W. W., Ross, C. A., Dawson, V. L. and Dawson, T. M. (2005). Parkinson's disease-associated mutations in leucine-rich repeat kinase 2 augment kinase activity. *Proc Natl Acad Sci USA* **102**, 16842–16847.

Westphal, C. H. and Chandra, S. S. (2013). Monomeric synucleins generate membrane curvature. *J. Biol. Chem.* **288**, 1829–1840.

White, R. M., Cech, J., Ratanasirintrawoot, S., Lin, C. Y., Rahl, P. B., Burke, C. J., Langdon, E., Tomlinson, M. L., Mosher, J., Kaufman, C., et al. (2011). DHODH modulates transcriptional elongation in the neural crest and melanoma. *Nature* **471**, 518–522.

Whitworth, A. J., Theodore, D. A., Greene, J. C., Beneš, H., Wes, P. D. and Pallanck, L. J. (2005). Increased glutathione S-transferase activity rescues dopaminergic neuron loss in a Drosophila model of Parkinson's

disease. *Proc Natl Acad Sci USA* **102**, 8024–8029.

Williams, J. A., Ni, H. M., Haynes, A., Manley, S., Li, Y., Jaeschke, H. and Ding, W. X. (2015). Chronic deletion and acute knockdown of Parkin have differential responses to acetaminophen-induced mitophagy and liver injury in mice. *J. Biol. Chem.* **290**, 10934–10946.

Wills, J., Credle, J., Oaks, A. W., Duka, V., Lee, J. H., Jones, J. and Sidhu, A. (2012). Paraquat, but not maneb, induces synucleinopathy and tauopathy in striata of mice through inhibition of proteasomal and autophagic pathways. *PLoS One* **7**, e30745.

Wilson, M. A., Collins, J. L., Hod, Y., Ringe, D. and Petsko, G. A. (2003). The 1.1-Å resolution crystal structure of DJ-1, the protein mutated in autosomal recessive early onset Parkinson's disease. *Proc Natl Acad Sci USA* **100**, 9256–9261.

Winklhofer, K. F. and Haass, C. (2010). Mitochondrial dysfunction in Parkinson's disease. *BBA - Mol. Basis Dis.* **1802**, 29–44.

Winslow, A. R., Chen, C. W., Corrochano, S., Acevedo-Arozena, A., Gordon, D. E., Peden, A. A., Lichtenberg, M., Menzies, F. M., Ravikumar, B., Imarisio, S., et al. (2010). α-Synuclein impairs macroautophagy: Implications for Parkinson's disease. *J. Cell Biol.* **190**, 1023–1037.

Wolfart, J., Neuhoff, H., Franz, O. and Roeper, J. (2001). Differential expression of the small-conductance, calcium-activated potassium channel SK3 is critical for pacemaker control in dopaminergic midbrain neurons. *J. Neurosci.* **21**, 3443–3456.

Wood-Kaczmar, A., Gandhi, S. and Wood, N. W. (2006). Understanding the molecular causes of Parkinson's disease. *Trends Mol. Med.* **12**, 521–528.

Wu, Z., Wang, Y., Lim, J. and Liu, B. (2018). Ubiquitination of ABCE1 by NOT4 in Response to Mitochondrial Damage Links Co-translational Quality Control to PINK1- Directed Mitophagy. *Cell Metab.* **28**, 130–144.

Xi, Y., Ryan, J., Noble, S., Yu, M., Yilbas, A. E. and Ekker, M. (2010). LRRK2 knockdown in zebrafish causes developmental defects, neuronal loss, and synuclein aggregation. *Eur. J. Neurosci.* **31**, 623–633.

Xia, Z., Wei, J., Li, Y., Wang, J., Li, W., Wang, K., Hong, X., Zhao, L., Chen, C., Min, J., et al. (2017). Zebrafish slc30a10 deficiency revealed a novel

compensatory mechanism of Atp2c1 in maintaining manganese homeostasis. Gitlin, J. (Ed). *PLOS Genet.* **13**, e1006892.

Xilouri, M., Vogiatzi, T., Vekrellis, K., Park, D. and Stefanis, L. (2009). Abberant α-synuclein confers toxicity to neurons in part through inhibition of chaperone-mediated autophagy. *PLoS One* **4**, e5515.

Xilouri, M., Vogiatzi, T. and Stefanis, L. (2008). Alpha-synuclein degradation by autophagic pathways: A potential key to Parkinson's Disease pathogenesis. *Autophagy* **4**, 917–919.

Xu, J., Zhong, N., Wang, H., Elias, J. E., Kim, C. Y., Woldman, I., Pifl, C., Gygi, S. P., Geula, C. and Yankner, B. A. (2005). The Parkinson's disease-associated DJ-1 protein is a transcriptional co-activator that protects against neuronal apoptosis. *Hum. Mol. Genet.* **14**, 1231–1241.

Xu, J., Pfarr, N., Endris, V., Mai, E. K., Md Hanafiah, N. H., Lehners, N., Penzel, R., Weichert, W., Ho, A. D., Schirmacher, P., et al. (2017). Molecular signaling in multiple myeloma: Association of RAS/RAF mutations and MEK/ERK pathway activation. *Oncogenesis* **6**, e337.

Xu, S., Yang, X., Qian, Y. and Xiao, Q. (2018). Parkinson's disease-related DJ-1 modulates the expression of uncoupling protein 4 against oxidative stress. *J. Neurochem.* **145**, 312–322.

Xu, Y. H., Sun, Y., Ran, H., Quinn, B., Witte, D. and Grabowski, G. A. (2011). Accumulation and distribution of α-synuclein and ubiquitin in the CNS of Gaucher disease mouse models. *Mol. Genet. Metab.* **102**, 436–447.

Yakes, F. M. and Van Houten, B. (1997). Mitochondrial DNA damage is more extensive and persists longer than nuclear DNA damage in human cells following oxidative stress. *Proc Natl Acad Sci USA* **94**, 514–519.

Yamada, M., Kida, K., Amutuhaire, W., Ichinose, F. and Kaneki, M. (2010). Gene disruption of caspase-3 prevents MPTP-induced Parkinson's disease in mice. *Biochem. Biophys. Res. Commun.* **402**, 312–318.

Yamaguchi, S., Yamane, T., Takahashi-Niki, K., Kato, I., Niki, T., Goldberg, M. S., Shen, J., Ishimoto, K., Doi, T., Iguchi-Ariga, S. M. M., et al. (2012). Transcriptional activation of low-density lipoprotein receptor gene by DJ-1 and effect of DJ-1 on cholesterol homeostasis. *PLoS One* **7**, e38144.

Yang, Y., Gehrke, S., Haque, M. E., Imai, Y., Kosek, J., Yang, L., Beal, M. F., Nishimura, I., Wakamatsu, K., Ito, S., et al. (2005). Inactivation of

Drosophila DJ-1 leads to impairments of oxidative stress response and phosphatidylinositol 3-kinase/Akt signaling. *Proc Natl Acad Sci USA* **102**, 13670–13675.

Yang, Y., Gehrke, S., Imai, Y., Huang, Z., Ouyang, Y., Wang, J.-W., Yang, L., Beal, M. F., Vogel, H. and Lu, B. (2006). Mitochondrial pathology and muscle and dopaminergic neuron degeneration caused by inactivation of Drosophila Pink1 is rescued by Parkin. *Proc Natl Acad Sci USA* **103**, 10793–10798.

Yang, Y., Mufson, E. J. and Herrup, K. (2003). Neuronal Cell Death Is Preceded by Cell Cycle Events at All Stages of Alzheimer's Disease. *J. Neurosci.* **23**, 2557–2563.

Yao, F., Zhang, R., Zhu, Z., Xia, K. and Liu, C. (2006). MutScreener: Primer design tool for PCR-direct sequencing. *Nucleic Acids Res.* **34**, W660–W664.

Yao, X. (1999). Evolving artificial neural networks. *Proc. IEEE* **87**, 1423–1447.

Yellen, G. (2018). Fueling thought: Management of glycolysis and oxidative phosphorylation in neuronal metabolism. *J. Cell Biol.* **217**, 2235–2246.

Yoon, S., Kim, S. Y. and Nam, D. (2016). Improving gene-set enrichment analysis of RNA-Seq data with small replicates. *PLoS One* **11**, e0165919.

Zancan, I., Bellesso, S., Costa, R., Salvalaio, M., Stroppiano, M., Hammond, C., Argenton, F., Filocamo, M. and Moro, E. (2015). Glucocerebrosidase deficiency in zebrafish affects primary bone ossification through increased oxidative stress and reduced Wnt/β-catenin signaling. *Hum. Mol. Genet.* **24**, 1280–1294.

Zhang, J., Perry, G., Smith, M. A., Robertson, D., Olson, S. J., Graham, D. G. and Montine, T. J. (1999). Parkinson's disease is associated with oxidative damage to cytoplasmic DNA and RNA in substantia nigra neurons. *Am. J. Pathol.* **154**, 1423–1429.

Zhang, M., Sun, T., Jian, C., Lei, L., Han, P., Lv, Q., Yang, R., Zhou, X., Xu, J., Hu, Y., et al. (2015). Remodeling of mitochondrial flashes in muscular development and dystrophy in zebrafish. *PLoS One* **10**, e0132567.

Zhang, Z., Yan, J., Chang, Y., Yan, S. S. and Shi, H. (2011). Hypoxia Inducible Factor-1 as a Target for Neurodegenerative Diseases. *Curr. Med. Chem.* **18**, 4335–4343.

Zhao, T., Zondervan-van der Linde, H., Severijnen, L.-A., Oostra, B. A., Willemsen, R. and Bonifati, V. (2012). Dopaminergic Neuronal Loss and Dopamine-Dependent Locomotor Defects in Fbxo7-Deficient Zebrafish. *PLoS One* **7**, e48911.

Zhong, N., Kim, C. Y., Rizzu, P., Geula, C., Porter, D. R., Pothos, E. N., Squitieri, F., Heutink, P. and Xu, J. (2006). DJ-1 transcriptionally up-regulates the human tyrosine hydroxylase by inhibiting the sumoylation of pyrimidine tract-binding protein-associated splicing factor. *J. Biol. Chem.* **281**, 20940–20948.

Zhou, C., Huang, Y., Shao, Y., May, J., Prou, D., Perier, C., Dauer, W., Schon, E. a and Przedborski, S. (2008). The kinase domain of mitochondrial PINK1 faces the cytoplasm. *Proc Natl Acad Sci USA* **105**, 12022–12027.

Zhou, H., Huang, C., Tong, J., Hong, W. C., Liu, Y.-J. and Xia, X.-G. (2011). Temporal expression of mutant LRRK2 in adult rats impairs dopamine reuptake. *Int. J. Biol. Sci.* **7**, 753–761.

Zhou, W., Zhu, M., Wilson, M. A., Petsko, G. A. and Fink, A. L. (2006). The oxidation state of DJ-1 regulates its chaperone activity toward alpha-synuclein. *J. Mol. Biol.* **356**, 1036–1048.

Zhou, X. P., Gimm, O., Hampel, H., Niemann, T., Walker, M. J. and Eng, C. (2000). Epigenetic PTEN silencing in malignant melanomas without PTEN mutation. *Am. J. Pathol.* **157**, 1123–1128.

Zhou, Z. D., Xie, S. P., Sathiyamoorthy, S., Ting, W., Sing, T. Y., Ng, S. H., Pek, H., Chua, H., Mei, A., Tang, Y., et al. (2015). F-box protein 7 mutations promote protein aggregation in mitochondria and inhibit mitophagy. *Hum. Mol. Genet.* **24**, 6314–6330.

Zhu, J.-H., Guo, F., Shelburne, J., Watkins, S. and Chu, C. T. (2006). Localization of Phosphorylated ERK/MAP Kinases to Mitochondria and Autophagosomes in Lewy Body Diseases. *Brain Pathol.* **13**, 473–481.

Zhu, W., Giangrande, P. H. and Nevins, J. R. (2004). E2Fs link the control of G1/S and G2/M transcription. *EMBO J.* **23**, 4615–4626.

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

Zweig, M. H. and Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin. Chem.* **39**, 561–577.