

## RESEARCH ARTICLE

# Machine learning discriminates a movement disorder in a zebrafish model of Parkinson's disease

Gideon L. Hughes<sup>1</sup>, Michael A. Lones<sup>2</sup>, Matthew Bedder<sup>1,5</sup>, Peter D. Currie<sup>3</sup>, Stephen L. Smith<sup>4,5</sup> and Mary Elizabeth Pownall<sup>1,4,\*</sup>

## ABSTRACT

Animal models of human disease provide an *in vivo* system that can reveal molecular mechanisms by which mutations cause pathology, and, moreover, have the potential to provide a valuable tool for drug development. Here, we have developed a zebrafish model of Parkinson's disease (PD) together with a novel method to screen for movement disorders in adult fish, pioneering a more efficient drug-testing route. Mutation of the *PARK7* gene (which encodes DJ-1) is known to cause monogenic autosomal recessive PD in humans, and, using CRISPR/Cas9 gene editing, we generated a Dj-1 loss-of-function zebrafish with molecular hallmarks of PD. To establish whether there is a human-relevant parkinsonian phenotype in our model, we adapted proven tools used to diagnose PD in clinics and developed a novel and unbiased computational method to classify movement disorders in adult zebrafish. Using high-resolution video capture and machine learning, we extracted novel features of movement from continuous data streams and used an evolutionary algorithm to classify parkinsonian fish. This method will be widely applicable for assessing zebrafish models of human motor diseases and provide a valuable asset for the therapeutics pipeline. In addition, interrogation of RNA-seq data indicate metabolic reprogramming of brains in the absence of Dj-1, adding to growing evidence that disruption of bioenergetics is a key feature of neurodegeneration.

This article has an associated First Person interview with the first author of the paper.

**KEY WORDS:** DJ-1, *PARK7*, Artificial intelligence, Gene targeting, Video capture, Parkinson's disease

## INTRODUCTION

Parkinson's disease (PD) is common, affecting about 1% of the population over 60 (Tysnes and Storstein, 2017), and it has no cure. The incidence of PD increases with age; therefore, this statistic is likely to become worse owing to an ageing population. Most cases of PD are not inherited, arising sporadically; however, several

monogenic forms of PD have been identified (reviewed in Hernandez et al., 2016). Studying the genes disrupted in patients with inherited forms of PD has been instrumental for progress in understanding the molecular basis of the disease (Clark et al., 2006; Clements et al., 2006; Shimura et al., 2000).

A valuable approach to understanding mechanisms by which gene mutations can result in human pathogenesis relies on *in vivo* experiments using animal models. Invertebrate models such as *Caenorhabditis elegans* (Ved et al., 2005) and *Drosophila* (Lu and Vogel, 2009) provide important platforms for studying neurodegenerative diseases in the context of a complete nervous system; however, the genetically tractable vertebrate *Danio rerio* shares more features of neuroanatomy with humans. Retrograde tracing has been used to identify dopaminergic (DA) neurons in zebrafish that project from the ventral diencephalon to the ventral telencephalon (Rink and Wullmann, 2001). These neurons are sensitive to the neurotoxin 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) (known to induce PD in humans) and considered analogous to the ascending midbrain DA neurons of the mammalian nigrostriatal pathway. In addition to anatomical similarities, vertebrate genomes share a high level of conservation such that orthologues of PD genes have been identified and manipulated in zebrafish (Anichtchik et al., 2008; Bretau et al., 2007; Flinn et al., 2009; Keatinge et al., 2015; Zhao et al., 2012). However, most of the zebrafish models of PD have used transient post-transcriptional gene inhibition protocols focusing on the analysis of larval DA neurons (Bretau et al., 2007; Flinn et al., 2009; Zhao et al., 2012) and swimming behaviour (Flinn et al., 2009; Zhao et al., 2012). In this study, we have developed an adult zebrafish model of PD that presents movement disorders typical of human patients.

Loss-of-function mutations in *PARK7* cause a rare form of early-onset PD (Bonifati et al., 2003). *PARK7* codes for the protein DJ-1, which has multiple roles in protecting cells from toxic protein aggregation and oxidative stress, and loss of these functions contributes to the presentation of parkinsonian pathology. DJ-1 protein is a member of a deeply conserved DJ-1/ThiJ/PfpI superfamily that includes related proteins from human to bacteria and archae (Bandyopadhyay and Cookson, 2004). Although these homologues share a core structure, their diverse functions include acting as kinases, proteases and chaperones. In humans, DJ-1 has been found to prevent the aggregation of  $\alpha$ -synuclein (Zondler et al., 2014), regulate transcription of oxidative stress response genes (Xu et al., 2005) and maintain mitochondrial function (Thomas et al., 2011; Winklhofer and Haass, 2010). High levels of oxidative stress in DA neurons are characteristic of PD patients, so it is interesting that DJ-1 itself is activated by the oxidation of a specific cysteine residue (C106) that is essential for its translocation to the mitochondria and neuroprotective function (Canet-Avilés et al., 2004; Wilson, 2011). Structural biology (reviewed in Cookson, 2003) and *Drosophila* genetics (Oswald et al., 2018) have been particularly informative

<sup>1</sup>Department of Biology, University of York, York YO10 5DD, UK. <sup>2</sup>School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK. <sup>3</sup>Australian Regenerative Medicine Institute, Monash University, Victoria 3800, Australia. <sup>4</sup>York Biomedical Research Institute, University of York, York YO10 5DD, UK. <sup>5</sup>Department of Electronic Engineering, University of York, York YO10 5DD, UK.

\*Author for correspondence (betsy.pownall@york.ac.uk)

 M.E.P., 0000-0003-2329-5844

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

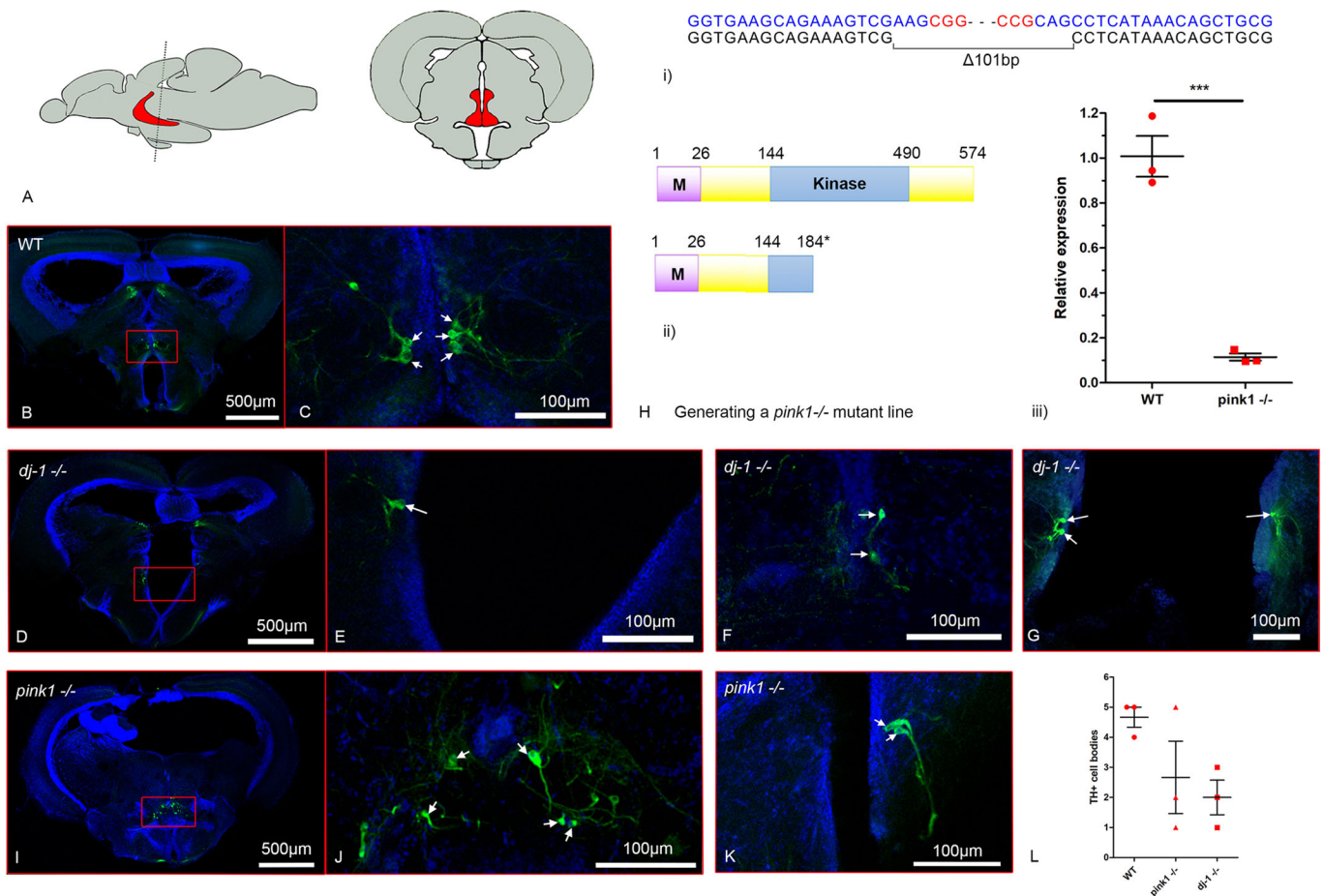
Handling Editor: E. Elizabeth Patton  
Received 18 May 2020; Accepted 17 August 2020



decay of the transcript (Fig. 1D). Genes known to be expressed in (DA) neurons – *dopamine transporter* (*dat*; also known as *slc6a3*), *tyrosine hydroxylase* (*th*) and *paired-like homeodomain 3* (*pitx3*) – were also found to be downregulated in *dj-1*<sup>-/-</sup> mutant brains compared to their wild-type siblings. *dat* and *th* expression were assessed as both are markers of DA neurons; *dat* is specific to DA neurons, whereas *th* is expressed in all catecholaminergic neurons (Holzschuh et al., 2001). *pitx3* is a transcription factor involved in the development and maintenance of DA neurons (Filippi et al., 2007). We conclude that our targeting of *dj-1* has resulted in a line of a genetic null zebrafish.

A *PTEN-induced kinase 1* null (*pink1*<sup>-/-</sup>) mutant line was generated by CRISPR/Cas9 targeting of the *PINK1* orthologue in zebrafish with two guide RNAs (gRNAs) producing a 101 bp deletion in exon 2 of *pink1* (at position 591 of NM\_001008628)

(Fig. 2H). The predicted Pink1 protein translated from the mutant transcript would be 183 amino acids in length, losing the majority of the kinase domain, and qRT-PCR analysis of RNA extracted from *pink1*<sup>-/-</sup> (*n*=4) and wild-type sibling (*n*=3) adult brains revealed a ~90% loss of *pink1* expression, indicative of nonsense-mediated decay of the mRNA. An immunohistochemical analysis was used to detect DA neurons (Fig. 2); three brains were dissected from wild-type, *dj-1*<sup>-/-</sup> and *pink1*<sup>-/-</sup> zebrafish at 12 months post-fertilisation (mpf), and 100 µm sections were cut with a vibratome through the posterior tuberculum (pT) in the diencephalon. In zebrafish, the pT includes a subset of DA neurons that project to the subpallium; these cells are considered homologous to the anterior-most ascending mesodiencephalic DA neurons in mammals (groups A8-A10), which are affected in PD (Rink and Wullimann, 2002, 2001). To investigate whether these cells are impacted in our zebrafish models



**Fig. 2. Reduction of dopaminergic neurons in the posterior tuberculum of PD zebrafish.** Immunohistochemical detection of Tyrosine hydroxylase (Th) in *dj-1*<sup>-/-</sup> and *pink1*<sup>-/-</sup> brains. A section through the posterior tuberculum (pT) was identified based on the shape of the brain according to Wullimann et al. (1996). The large pear-shaped Th-positive cells located next to the ventricle in the pT have previously been identified as part of the zebrafish dopaminergic system projecting to the striatum (Rink and Wullimann, 2001). Therefore, the pear-shaped Th-positive cells identified in the periventricular pT location were counted. (A) Lateral view of the adult zebrafish brain (left) and cross-section through the adult zebrafish brain (right). Highlighted in red is the pT. (B-G, I-K) Immunofluorescently labelled Th-positive cells (green) in the pT of wild-type, *dj-1*<sup>-/-</sup> and *pink1*<sup>-/-</sup> zebrafish (*n*=3 biological replicates) at 12 months post-fertilisation (mpf). Hoechst staining of nuclei is in blue. Arrows indicate the cell bodies of the paraventricular DA neurons. (B) Th-positive cells in a section through the pT (red box) of a wild-type brain. (C) A close-up of the Th-positive cells in the pT from A. (D) Th-positive cells in a section through the pT (red box) of a *dj-1*<sup>-/-</sup> brain. (E) A close-up of the Th-positive cells in the pT from D. (F, G) Close-ups of Th-positive cells in the pT of two further *dj-1*<sup>-/-</sup> brains. (H) (i) CRISPR/Cas9 target sequences (blue), PAM sites (red) and the 101 bp deletion generated in exon 2 of *pink1* (at position 591 of NM\_001008628). (ii) Wild-type Pink1 (above) and the truncated Pink1 protein (below) predicted in the *pink1* mutant. (iii) qRT-PCR analysis comparing *pink1* expression in brains extracted from *pink1*<sup>-/-</sup> zebrafish (*n*=4, biological replicates) and their wild-type siblings (*n*=3) at 16 wpf. Student's *t*-tests (two-tailed, unpaired) were used to compare the dCt values for *pink1*<sup>-/-</sup> and wild-type samples. Data are mean ± s.e.m., \*\*\*P<0.001. (I) Th-positive cells in a section through the pT (red box) of a *pink1*<sup>-/-</sup> brain. (J) A close-up of the Th-positive in the pT from I. (K) A close-up of the Th-positive cells in the pT of a further *pink1*<sup>-/-</sup> brain. (L) Counts of Th-positive cell bodies seen in the pT (single 100 µm section) for wild-type, *dj-1*<sup>-/-</sup> and *pink1*<sup>-/-</sup> zebrafish (*n*=3 biological replicates) at 12 mpf.



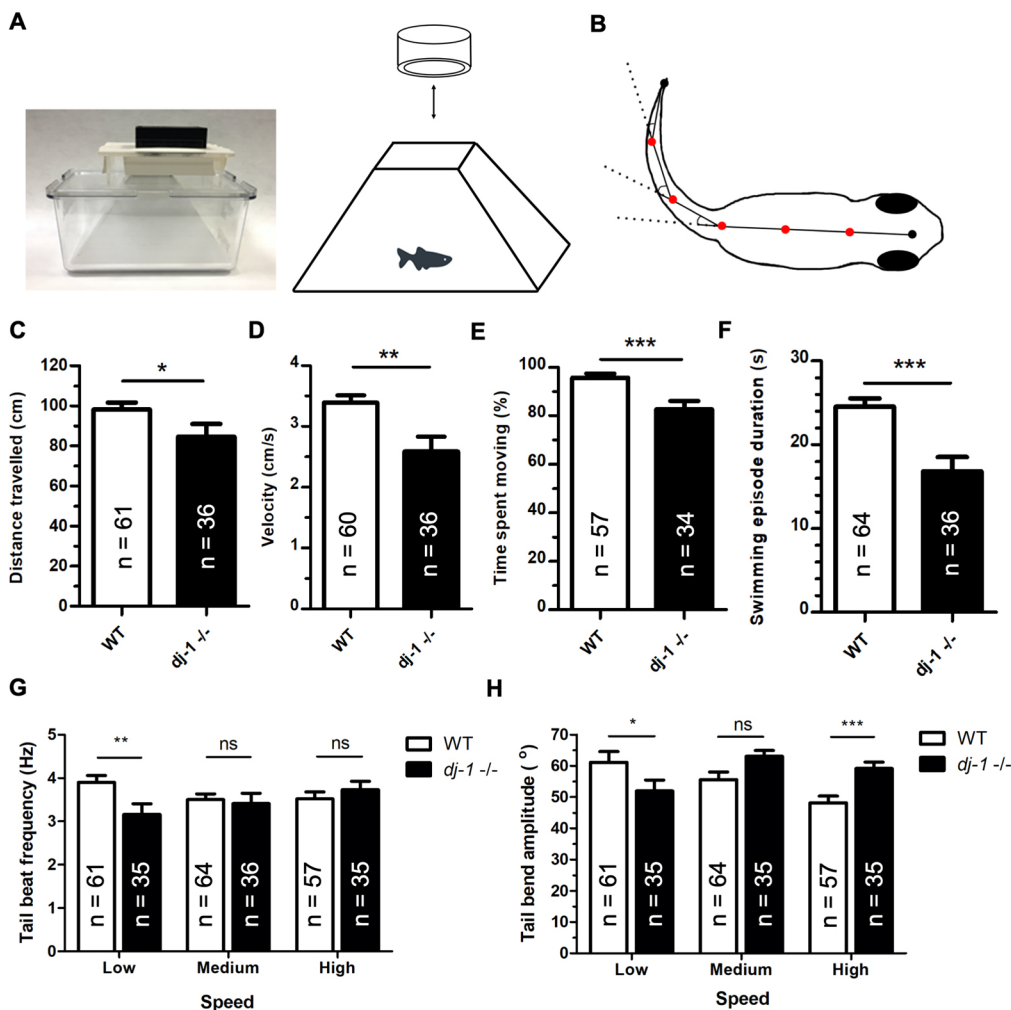
of PD, we used immunofluorescence to detect the protein Tyrosine hydroxylase (Th), an enzyme required to produce catecholamines including dopamine. Our method is informed by the work of Matsui and Sugie (2017) and Rink and Wullmann (2001) to identify and count these neurons in zebrafish. The brains were fixed and embedded in agarose blocks for vibratome sectioning before processing by immunohistochemistry for Th immunoreactivity and imaging by confocal microscopy. Cell bodies in the pT positive for Th that have a pear-shaped appearance and a periventricular position were identified as pT DA neurons and counted in each of the sections (Fig. 2B–G, I–K). Fig. 2L indicates the mean number of pT DA neurons identified for each genotype. In the wild-type brain, a range of four to five of these cells was identified, consistent with the retrograde tracing study by Rink and Wullmann (2001), while the number of pT DA neurons identified in the *dj-1*<sup>-/-</sup> and *pink1*<sup>-/-</sup> brains ranged from one to three and one to five, respectively, suggesting a reduction of these neurons in these genotypes, as shown previously for *pink1*<sup>-/-</sup> (Flinn et al., 2013).

### A simple and effective method for measuring micro-movement in zebrafish

A progressive loss of DA neurons is characteristic of PD in humans (Damier et al., 1999); however, to establish whether our *dj-1*<sup>-/-</sup> null zebrafish provide a bone fide animal model of PD we wanted to test whether the fish share a dyskinesia phenotype. To do this, we developed a new and simple method to measure features of

movement. In contrast to other fish-tracking programs, our analysis extracts movement data from the adult tail bending along its whole axis while accelerating, decelerating and turning. These data are collected as a continuous feed and converted by bespoke software into data suitable for use with novel computational methods to test whether our zebrafish model shows micro-movement features identified in human patients with PD.

We established filming methods using a specially designed tank and camera setup that allows consistent video recordings free of reflections and shadows (Fig. 3A). We also wrote a computer program (ShadowFish) to generate a minimal set of data to characterise fish movement; this greatly reduces the dimensionality of the data, which is essential for input into computational analyses. ShadowFish software and tank design are freely available ([github.com/ghughesyork/ShadowFish](https://github.com/ghughesyork/ShadowFish)). The position of the fish spine is approximated in each frame of an input video clip, and then the flexion or bending angle is measured at five equidistant positions along the spine (Fig. 3B). Seven *x,y* coordinates were also measured along the spine for each frame. Simplifying the video data in this way facilitates our subsequent classification analyses using EAs. Forty-six *dj-1*<sup>-/-</sup> mutant zebrafish (12 weeks old) and the same number of age-matched wild-type controls were recorded swimming for 5 min following an acclimation time of 10 min after moving into the ShadowFish tank. Using *dj-1*<sup>-/-</sup> and wild-type fish from multiple different breedings reduced the chances of evolving a classifier that recognised a pattern



**Fig. 3. Analysis of extracted features reveals distinct movement in *dj-1*<sup>-/-</sup> zebrafish.** (A) A photograph of the frustum insert designed to fit an Aquatics Habitat mating tank with a GoPro camera attached (left), providing a simple system for accurate, high-resolution video capture of adult zebrafish movement. A diagram of the fish inside the frustum insert, recorded from above using the GoPro camera (right). (B) A diagram of the angles measured along the zebrafish trace, at the five vertices (red dots), when there is a bend in the tail. *x* and *y* coordinates were measured for the vertices and endpoints. Measurements were recorded at 100 frames/s from the video input, allowing analysis of selected features of movement. (C–H) The features of movement compared between *dj-1*<sup>-/-</sup> and wild type (WT) at 12 wpf including distance travelled (C), velocity (D), percentage of time spent moving (E), mean duration of a swimming episode (F), tail beat frequency at low, medium and high swimming speeds (G), and tail bend amplitude at low, medium and high swimming speeds (H) (single replicate). The number of replicates (*n*) is shown for each graph. Student's *t*-tests (two-tailed, unpaired) were used to compare features that followed a normal distribution; the Mann–Whitney *U*-test (two-tailed) was used to compare non-parametric features. Data are mean ± s.e.m., \**P*<0.05, \*\**P*<0.01, \*\*\**P*<0.001; ns, not significant.

occurring in a single progeny. The increased generality of the evolved classifier improves classification accuracy when used to discriminate future *dj-1<sup>-/-</sup>* mutants. Then, 30-s video clips, filmed at 100 frames/s, were processed from recordings. Two 30-s clips from each recording were processed using the ShadowFish software; these were visually assessed to ensure that we removed any video with a reflection or a shadow or where the fish was not in view. The 30-s clips were also processed using the ShadowFish software to collect movement data in the same way from two other genotypes, *pink1<sup>-/-</sup>* and *dmd<sup>1a222a/+</sup>*.

### Extracting features of movement

Initial analyses were undertaken using MATLAB to calculate selected features of movement including distance travelled (Fig. 3C), velocity (Fig. 3D), time spent moving (Fig. 3E) and swimming episode duration (Fig. 3F). Over a 30-s period, *dj-1<sup>-/-</sup>* zebrafish covered significantly less distance at a slower velocity compared to wild-type fish. While wild-type fish spent 96% of their time moving, *dj-1<sup>-/-</sup>* mutants were moving only 83% of the time. Additionally, the mean swimming episode duration by a *dj-1<sup>-/-</sup>* was 17 s, compared to 25 s for the wild type. Tail beat frequency (Fig. 3G) and mean tail bend amplitude (Fig. 3H) were analysed at low, medium and high speeds of movement (Keatinge et al., 2015). It was hypothesised that mutant fish might need to work harder to achieve higher speeds through more tail beats; however, only the tail beat frequency at low speed was significantly affected and reduced in the *dj-1<sup>-/-</sup>* mutant. A trend was seen in the wild-type zebrafish for the tail bend amplitude to decrease with an increase in swimming speed, suggesting that momentum allows fish to coast once higher speeds have been reached, reducing tail bend amplitude and energy expenditure (McHenry and Lauder, 2005). In contrast, the mean tail bend amplitude is significantly less in the *dj-1<sup>-/-</sup>* zebrafish at low speeds but significantly greater at high speeds, and a greater tail bend amplitude was observed in the *Dj-1*-deficient zebrafish at high speeds compared to the wild type (Fig. 3H). As the *dj-1<sup>-/-</sup>* mutant fish stop more frequently (Fig. 3F), the resulting inertia might impede them from gaining the momentum required for coasting. Alternatively, the *dj-1<sup>-/-</sup>* 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 et al., 2012). In contrast to the differences in movement seen at 12 weeks post-fertilisation (wpf), we found that, in younger zebrafish (8 wpf), most features of movement of *dj-1* mutants were not significantly different from those of age-matched controls (Fig. S1). This progressive impairment of movement in *dj-1<sup>-/-</sup>* mutant fish is consistent with the progressive loss of motor control seen in PD patients (Greffard, et al., 2006).

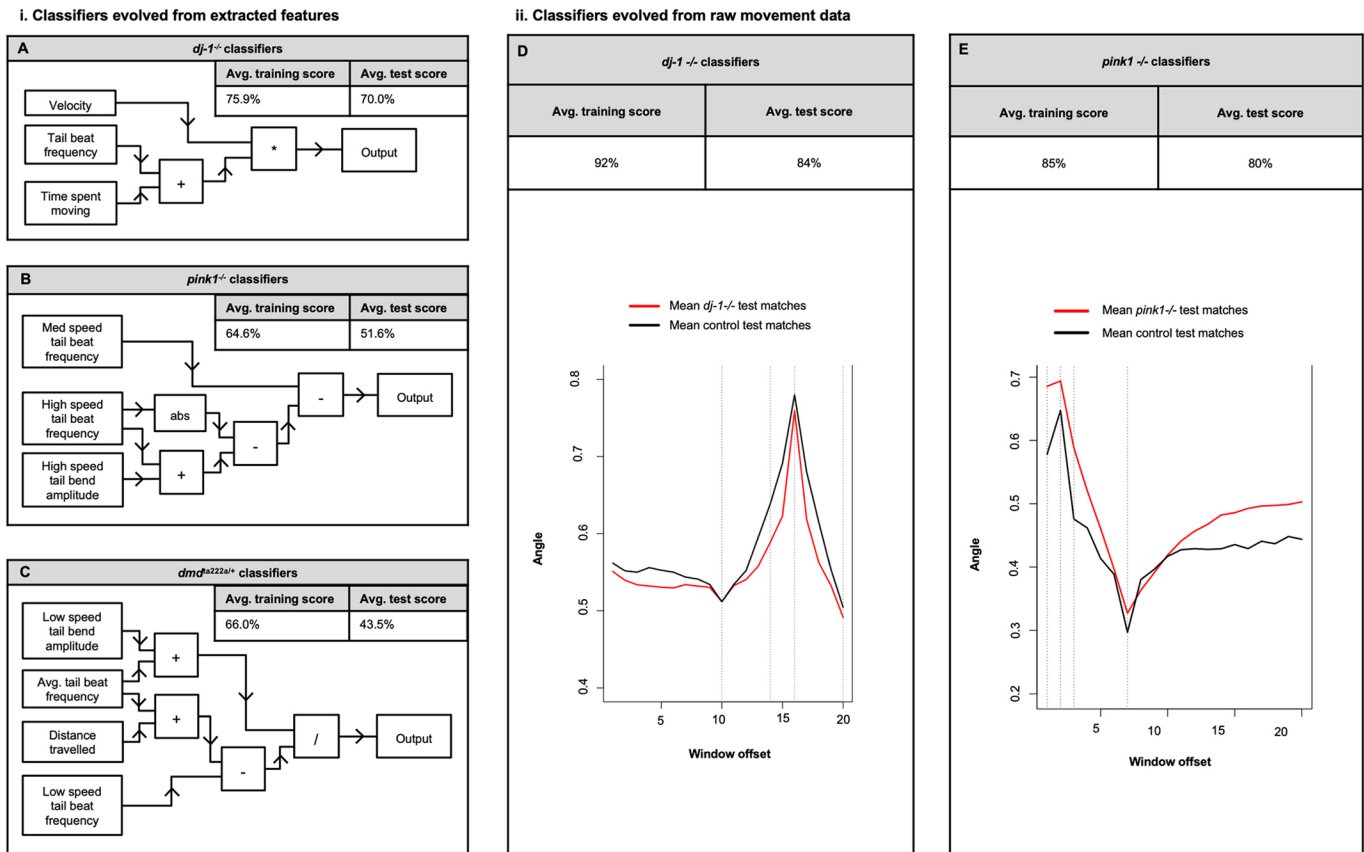
### Evolving classifiers using extracted features of movement

Machine learning has proved successful in analysing highly complex, non-linear data in human PD patients (Lones et al., 2013), and we have adapted the methods from our previous study to this zebrafish model of PD. Automated analyses of the digitised swimming data were used to measure selected features of movement in the mutant and wild-type fish, with the aim to distinguish any combination of features that characterises the movement of the *dj-1<sup>-/-</sup>* zebrafish. In addition, we included data extracted from another parkinsonian model, the *pink1<sup>-/-</sup>* mutant zebrafish, and from zebrafish heterozygous for *sapje* (*dmd<sup>1a222a/+</sup>*) (Bassett et al., 2003). *sapje* is a recessive mutation that is lethal when homozygous

and a model for muscular dystrophy, another human disease in which movement is disrupted. The *dmd<sup>1a222a</sup>* allele is recessive so in a heterozygote condition there is no expected phenotype or expected change in movement, thus providing a negative control for evolving classifiers.

We describe here an objective method, using machine learning, to identify any movement phenotype that can distinguish mutants from the wild type with a defined level of accuracy. Machine learning can automatically analyse large datasets and ‘learn’ to recognise differences between the classes of data (Alpaydin, 2016). We use a form of ‘supervised learning’, where a supervisor provides the desired output for every input when generating 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 is the class of fish, either mutant or wild type. An EA is a population-based process designed to optimise the solution to a defined problem; by evaluating a population of candidate solutions using a ‘fitness function’ (an objective mathematical measure that determines the most effective solution) and subsequently spawning a new population of candidate solutions 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 (Smith et al., 2015). Moreover, EAs are one of the few ‘white box’ forms of machine learning for which the functions that make up the classifier can be inspected (Lones et al., 2013); this is a useful characteristic to allow discovery of any movement phenotype characteristic of a zebrafish model of PD.

Using selected features of movement, an EA was used to train classifiers labelling fish as either *dj-1<sup>-/-</sup>* or wild type. Forty-six *dj-1<sup>-/-</sup>* and 46 age-matched control zebrafish were recorded swimming for a period of 5 min. Two separate 30-s clips were processed from each 5-min recording, generating 92 comma-separated value (CSV) files containing raw movement data for each class of fish. After visual quality checks, 30 recordings of *dj-1<sup>-/-</sup>* and an equal number from age-matched wild-type controls were used to evolve classifiers. The type of EA used was a Cartesian genetic programming (CGP) algorithm, in which programs are represented as directed graphs with two-dimensional grids of computational nodes (Miller, 2011). CGP was selected as it is a white box machine learning algorithm that is suitable for relatively small sample sizes and can provide insight into the contribution of respective features in the resulting classifier, as demonstrated in previous work in analysing human movement disorders (Lones et al., 2013). Here, we use 30 recordings, and 60 data points were obtained for each class, which, according to the methods of Bland and Altman (2007), provides a 95% confidence interval of  $\pm 0.44$  standard deviations. The CGP network for classifying *dj-1<sup>-/-</sup>* fish that achieved the highest test score on a fold of datasets is shown in Fig. 4A. Velocity, mean tail beat frequency and time spent moving were the extracted features of movement used in the highest-scoring *dj-1<sup>-/-</sup>* classifier. These data reveal these features of movement as the most useful for discriminating the *dj-1<sup>-/-</sup>* mutants from wild type. The mean test score of classifiers evolved using the 20 folds of datasets was 70%, indicating that an equation could distinguish *dj-1<sup>-/-</sup>* fish by their features of movement 70% of the time. Similar analyses were carried out to classify extracted features of movement from *pink1<sup>-/-</sup>* (another PD model) and *dmd<sup>1a222a/+</sup>* (a negative control) and are presented in Fig. 4B,C. The EA was unable to compute an equation to distinguish these genotypes from controls using extracted features of movement (test scores were 51.6% and 43.5%, respectively).



**Fig. 4. Machine learning evolves classifiers from movement data and discriminates *dj-1<sup>-/-</sup>* zebrafish as distinct from controls.** (i)(A-C) An analysis using an evolutionary algorithm to discriminate a classifier by analysing extracted features of movement. (A) The mean training and test scores for classifiers evolved to recognise *dj-1<sup>-/-</sup>* zebrafish ( $n=30$ ) at 12 wpf over 20 folds of datasets containing extracted features of movement. The CGP network of the highest scoring *dj-1<sup>-/-</sup>* classifier evolved using the extracted features is depicted as a flow diagram. (B) The mean training and test scores for classifiers evolved to recognise *pink1<sup>-/-</sup>* zebrafish ( $n=37$ ) at 14 wpf over 20 folds of dataset containing extracted features of movement. The CGP network of the highest-scoring *pink1<sup>-/-</sup>* classifier evolved using the extracted features is depicted as a flow diagram. (C) The mean training and test scores for classifiers evolved to recognise *dmd<sup>ta222a/+</sup>* zebrafish ( $n=25$ ) at 12 wpf over 20 folds of dataset containing extracted features of movement. The CGP network of the highest-scoring *dmd<sup>ta222a/+</sup>* classifier evolved using the extracted features is depicted as a flow diagram. (ii)(D,E) A separate analysis of the raw movement data using sliding window classifiers trained with the PC2 time series data to generate symbolic mathematical expressions that describe discriminatory local patterns in the data. (D) The training and test accuracies for the classifier evolved to recognise *dj-1<sup>-/-</sup>* zebrafish at 12 wpf. Mean plots of the PC2 time series data in the 20 windows most useful for discriminating *dj-1<sup>-/-</sup>* mutants ( $n=36$ ) (red) and age-matched controls ( $n=64$ ) (black) in the test dataset. (E) The training and test accuracies for the classifier evolved to recognise *pink1<sup>-/-</sup>* zebrafish at 14 wpf. Mean plots of the PC2 time series data in the 20 windows most useful for discriminating *pink1<sup>-/-</sup>* mutants ( $n=39$ ) (red) and age-matched controls ( $n=44$ ) (black) in the test dataset.

### Evolving sliding window classifiers from raw movement data

For an unbiased classification we used the raw data extracted from the videos, without selection of any particular features of movement, for input into an EA. To do this, we processed the movement data using principal component analysis (PCA) to condense the data into principal components (PCs) by linear combination of the tail bend angles in each frame (Fig. S2). This reduced the dimensionality of the tail bend angles, transforming the multivariate time series into a univariate time series, while retaining the inherent variation of the data. PC2 was used to evolve classifiers as it retained the high frequency movement and subtle variation important for characterising a movement phenotype (Lones et al., 2013).

Sliding window classifiers, encoded using CGP, were evolved to classify the mutants using the PC2 time series as input data. The mathematical equations generated describe local patterns in the movement data that can be used for discrimination (Lones et al., 2013). The window in a sliding window classifier contains the data from a specific range in a time series, which become the input data

for the evolved algorithm. The raw movement data extracted from 36 clips of *dj-1<sup>-/-</sup>* mutants and 64 clips of age-matched control zebrafish with visually assessed, high-quality tracking were used to evolve sliding window classifiers. Here, 20 PC2 values (extracted from 20 frames of a recording) were used as the input data for a window. The algorithm applies pre-defined functions to the input data and produces a continuous value between 0 and 1 (Lones et al., 2013). Subsequently, 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 final classifier output (the mean of all the output values) labels the data as belonging to one of two classes depending on a threshold value. Similar analyses were carried out on the *pink1<sup>-/-</sup>* homozygous mutants and *dmd<sup>ta222a/+</sup>* heterozygote zebrafish.

One of the aims of this work was to confidently classify zebrafish using a protocol that is effective and practical, by acquiring short 5-min videos for analysis. Two 30-s clips were determined



empirically to provide sufficient data to identify differences in movement that allow classification of both the *dj-1<sup>-/-</sup>* and *pink1<sup>-/-</sup>* mutants from control fish. For the *dj-1<sup>-/-</sup>* and *pink1<sup>-/-</sup>* datasets, the sliding window classifiers were able to discriminate mutant fish around 80% of the time (Fig. 4D,E). The classifiers evolved using the raw movement data were more effective than classifiers evolved using select features. This is likely due to the objectivity that is possible using a continuous stream of data as well as the increased amount of data analysed from each recording. In addition, each selected feature was calculated to a single value summarising the movement in a video clip and reducing the data available to the EA. To visualise local patterns in the movement data characteristic of each class of fish [mutant (red) or wild type (black)] a mean plot of the 20 windows that were most useful for discriminating mutants is shown for *dj-1<sup>-/-</sup>* (Fig. 4D) and *pink1<sup>-/-</sup>* (Fig. 4E) mutant zebrafish. This same set of computational analyses was undertaken using heterozygous *dmd<sup>ta222a/+</sup>* zebrafish and a useful classifier could not be evolved. Fig. S3 reports the training and test areas under the curves (AUCs) for 20 runs, showing a mean of 0.78 for train and 0.43 for test. This shows that classifiers are evolved that over-fit the training data, but these models cannot generalise or classify any meaningful differences between the movement of *dmd<sup>ta222a/+</sup>* and the wild-type zebrafish. Together, these data indicate that machine learning can discriminate movement phenotypes when present, as is the case with the zebrafish models of PD, but not in the heterozygous carrier of the muscular dystrophy gene *dmd<sup>ta222a/+</sup>* that has no phenotype; this was expected given the recessive inheritance of the disease. One limitation of our study is that we did not treat adult animals with drugs; for instance, MPTP treatment mimics PD, which would be a useful addition to our protocols. Moreover, the ability of artificial intelligence to identify PD fish presents the possibility (indeed the next step) of treating the fish with known or potential therapeutics and then re-assessing the same fish. We are currently working with collaborators to develop these further protocols.

### Molecular signatures consistent with parkinsonian pathology characterise *dj-1<sup>-/-</sup>* brains

Animal models can provide important insight into the molecular basis of genetic disease. To investigate any global changes in gene expression associated with a loss of Dj-1, RNA-seq was carried out on RNA extracted from adult zebrafish brains using *dj-1<sup>-/-</sup>* mutants and three wild-type siblings at 16 wpf. An overview of these effects is shown as a volcano plot (Fig. 5A) and a short list of affected genes was curated using the criteria of fold change  $\geq 1.2$  (up- or downregulated) and false discovery rate (FDR)-adjusted *P*-value  $\leq 0.05$ . The list is depicted as a heatmap illustrating the relative expression of 22 transcripts that were found to have significantly altered expression in the *dj-1<sup>-/-</sup>* mutant brains (Fig. 5B). Validation of a set of these targets by qRT-PCR analysis confirmed the changes in expression for some genes of interest (Fig. 5C). These data have been deposited in NCBI's Gene Expression Omnibus (GEO) (Edgar et al., 2002), under accession number GSE135271 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE135271>). A list of all differential transcripts with  $q < 0.05$  is provided in Table S2A,B.

Genes significantly affected by loss of DJ-1 include those with cell functions known to be disrupted in PD. *VAMP3*, part of a SNARE complex, found on the vesicle membrane and required for vesicular transport from early, recycling endosomes to the trans Golgi network was found downregulated in HeLa cells (Mallard et al., 2001). DA neurons in the substantia nigra pars compacta (SNc) have increased vulnerability to dysfunctional intracellular

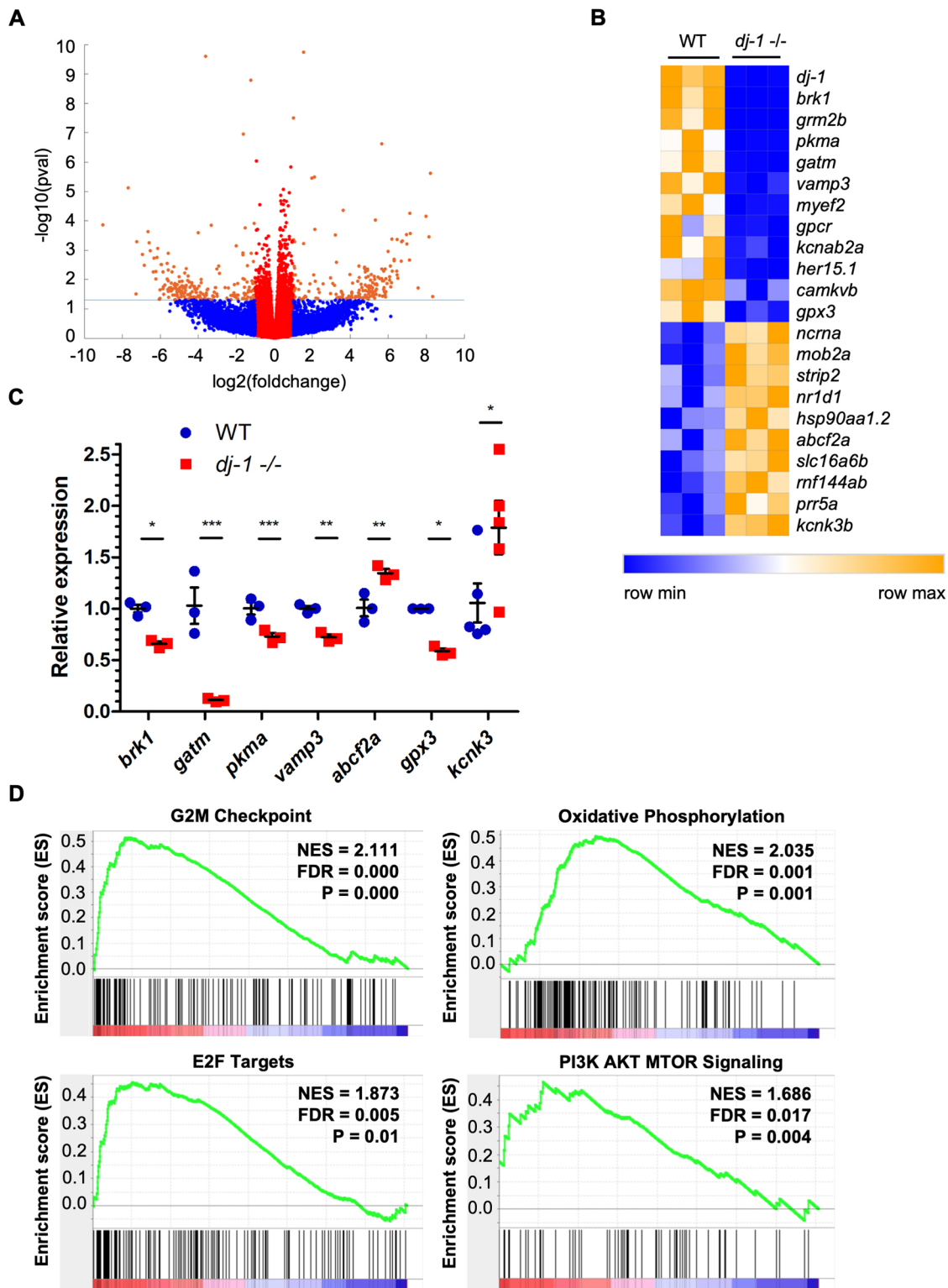
trafficking, as vesicles transport cargo great lengths and to a high number of synapses due to their extensive axonal arborisation (Bolam and Pissadaki, 2012; Giguère et al., 2019). Mitochondrial dysfunction is accepted as central to PD pathophysiology (Hu and Wang, 2016). Here, we find that *glycine amidinotransferase (gatl)*, encoding a mitochondrial enzyme involved in creatine synthesis (Sandell et al., 2003), is significantly downregulated in *dj-1<sup>-/-</sup>* brains. As well as an energy store, creatine has an antioxidative function, scavenging for reactive oxygen species (ROS) (Allen, 2012; Gabriel et al., 2013). In addition, *glutathione peroxidase 3 (gpx3)*, encoding an enzyme that reduces hydrogen peroxide, thereby protecting the cell from oxidative stress, also has reduced expression in the brains of *dj-1<sup>-/-</sup>* zebrafish. Interestingly, the human *GPX3* gene contains a peroxisome proliferator response element (PPRE) that is activated by PPAR $\gamma$  (Chung et al., 2009), a signalling pathway found activated downstream of DJ-1 (Han et al., 2018). *pyruvate kinase M1/2a (pkma)*, a key regulator of aerobic glycolysis (Sun et al., 2011), is also downregulated, suggesting dysregulated metabolism in *dj-1<sup>-/-</sup>* brains.

In order to take a broader view of the classes of genes sensitive to loss of DJ-1, we used gene set enrichment analysis (GSEA) to analyse our RNA-seq data (Subramanian et al., 2007, 2005) and found enrichment profiles consistent with disrupted metabolism and cell cycle regulation. This method uses pre-defined sets of genes that have been categorised based on interactions in a common pathway or a related biological function. The data from the RNA-seq analysis were processed and a normalised enrichment score (NES) calculated to determine whether genes affected are over-represented in any molecular signature database. The NES reflects the degree to which the genes in a gene set are over-represented at the top or bottom of a gene list ordered by association with a genotype. Our results revealed that *dj-1<sup>-/-</sup>* brains have over-representation of the gene sets presented in Table S3. As would be predicted from several previous reports (Kim et al., 2005; Lan et al., 2017), we find enrichment of gene sets associated with PI3K/AKT/mTOR and mTORC1 signalling (Fig. 5D; Table S3). This is consistent with increased levels of mTOR in mouse models of PD (Wills et al., 2012) and the finding that DJ-1 negatively regulates the tumour suppressor PTEN, which is part of the PI3K/AKT signal transduction pathway (Kim et al., 2005).

Our GSEA also indicates an upregulation of genes associated with oxidative phosphorylation in the absence of Dj-1. Metabolic regulation in neurons is crucial given the high and dynamic energy demands of the vertebrate brain (Yellen, 2018); moreover, SNc DA neurons in particular have high metabolic needs (Giguère et al., 2019). Our results show that some genes coding for players in glycolysis (Fig. 5C) are downregulated, while the gene set for oxidative phosphorylation is over-represented (Fig. 5D; Table S3). We also find the enrichment of the G2 to M transition gene set and targets of E2F transcription factors (Fig. 5D), as well as other cell cycle-related genes sets (Table S3) in the brains of *dj-1<sup>-/-</sup>* zebrafish. Interestingly, entering the cell cycle has been linked to cell death in neurons (Herrup et al., 2004; Herrup and Yang, 2007), and, moreover, Höglinger et al. (2007) found that the pRb/E2F pathway caused neuronal cell death in a chemical model of PD, consistent with the over-representation of mitotic gene sets in *dj-1<sup>-/-</sup>* brains, reflecting their neurodegenerative pathology.

### DISCUSSION

Age-dependent neurodegeneration is difficult to model in organisms with short lifespans, and, moreover, motor disorders are not always apparent and can be difficult to measure (Lee et al.,



**Fig. 5. Differential gene expression in the *dj-1*<sup>-/-</sup> mutant brain.** (A) A volcano plot representing the global changes in gene expression observed by RNA-seq analysis, comparing RNA from the brains of *dj-1*<sup>-/-</sup> mutants ( $n=3$ , biological replicates) and their wild-type siblings ( $n=3$ ) at 16 wpf. The fold change ( $\log_2$ ) of each transcript was plotted against the  $P$ -value ( $-\log_{10}$ ). Transcripts in red are down/upregulated less than 2-fold. Transcripts in orange that appear above the dotted line are differentially expressed with a  $P$ -value of  $<0.05$  and down/upregulated more than 2-fold. Transcripts in blue have a  $P$ -value of  $\geq 0.05$ . (B) A heatmap showing select genes differentially expressed with an FDR  $q$ -value of  $<0.05$ . Relative expression is represented on a colour scale from blue (low) to orange (high). (C) qRT-PCR analysis validating the results seen in the RNA-seq using RNA extracted from the brains of *dj-1*<sup>-/-</sup> and wild-type siblings at 16 wpf ( $n=3$  for all but  $n=5$  for *kcnk* analysis, biological replicates). Student's  $t$ -tests (two-tailed, unpaired) were used to compare the dCt values for *dj-1*<sup>-/-</sup> and wild-type samples. Data are mean  $\pm$  s.e.m., \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ . (D) GSEA enrichment plots for Hallmark gene sets: G2M checkpoint, oxidative phosphorylation, E2F transcription factors and PI3K/AKT/mTOR signalling. FDR, false discovery rate-adjusted  $P$ -value; NES, normalised enrichment score;  $P$ , nominal  $P$ -value. Data have been deposited in GEO, accession number GSE135271.



2017; Rousseaux et al., 2012). There has been some success modelling PD using neurotoxin-based models in the mouse; however, these present a rapid loss of DA neurons, which does not align with the progressive neurodegeneration of PD in humans (Jackson-Lewis and Przedborski, 2007). Genetic approaches for modelling PD in mice have had less success in reproducing the neurodegeneration. The loss of *Pink1*, *Prkn*, *Dj-1* or *Fbxo7* in the mouse does not reduce 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*, *Prkn* and *Dj-1* in the mouse still failed to produce a phenotype with recognisable aspects of PD (Kitada et al., 2009). In contrast, 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 (Ramonet et al., 2011; Janezic et al., 2013). In addition, aggregates of  $\alpha$ -synuclein forming in the motor neurons of transgenic mice overexpressing  $\alpha$ -synuclein has been reported, which could impact the movement phenotype (van der Putten et al., 2000). Rats may prove a better mammalian model for PD; Dave et al. (2014) reported a loss of DA neurons and locomotor defects in rats lacking *Dj-1* and *Pink1*, but not *Prkn*. 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). Nonetheless, compared to a mouse or rat models, there are advantages to using zebrafish in drug screens. Zebrafish are smaller, have lower maintenance costs and produce much larger broods for generating large numbers of animals for analyses; there is also the simplicity of administering candidate drugs in their water. Although the method described here uses adults and is not suitable for the high-throughput drug screens used on zebrafish larvae (Gallardo et al., 2015), it may provide a valuable addition to current drug screening protocols targeted at genetic diseases that have no larval phenotype but present a later movement disorder, as part of a multifaceted approach (Volpatti et al., 2020).

### An unbiased method to discriminate movement disorders in zebrafish

Here, we have used an EA to classify data obtained from recordings of zebrafish movement by measuring the bending of the spine. Previous tracking software has analysed location coordinates as fish swim; this is done in larvae and adults and informs only on the location or locomotor activity of individuals or groups of fish (Stewart et al., 2015; Cronin and Grealy, 2017). In contrast, we have measured the tail bend angle along the back of individual adults, to inform on muscle movement and assess dyskinesia. Using this approach, machine learning discriminates a parkinsonian phenotype in zebrafish deficient in *dj-1/park7*, a gene that causes a recessive form of early onset PD in humans. Off-target effects are a worry of any experimental program that relies on gene targeting. Back crossing to the wild-type strain for several generations can remove any potential modifiers linked to the targeted gene, and extending our study in this way would ensure that the movement phenotypes distinguished in our analyses tracked with the mutant allele. It is also important that the fish are genetically similar, except for the allele of interest, making it difficult to compare across different generations; therefore, the fish compared were always cousins of the same generation. In addition, raising fish in separate tanks and the emergence of traits from unrelated homozygous mutations resulting from in-crossing wild-type fish are potential caveats to our interpretations. Nonetheless, machine learning was able to

distinguish *dj1*<sup>-/-</sup> and *pink1*<sup>-/-</sup> mutants from wild-type cousins, but could not evolve an equation to discriminate between our control line, *dmd*<sup>gta222a/+</sup> heterozygotes, from wild-type fish, which were also raised in separate tanks.

Machine learning has also been used to classify mouse models of PD generated by injection of neurotoxins directly into the brain; the pronounced movement phenotype is measured using the well-established Noldus Catwalk gait analysis and discriminated with an accuracy of 96% (Frohlich et al., 2018). Similarly, the extraction of movement data and evolution of classifiers to diagnose PD in humans has been refined to reach accuracies of up to 95% (Lones et al., 2013). Our novel ShadowFish tracking software extracts data suitable for analysis by EAs and results in classifiers with ~80% accuracy. The ability of our new method to evolve algorithms that discriminate movement of zebrafish carrying two different PD, but not that of zebrafish carrying a heterozygous mutation in *dmd*, indicates that our method has the potential to reach higher classification accuracies and provide a valuable new method to detect movement disorders in a non-mammalian vertebrate model.

Clinical diagnosis of PD is challenging, with an accuracy of ~80% (Hughes et al., 1992; Rizzo et al., 2016). More accurate, less subjective diagnosis of PD using EAs has improved diagnostics and is able to describe the bradykinesia characteristic of PD in a finger-tapping exercise (Lones et al., 2013), allowing early diagnosis of PD distinct from other neurodegenerative diseases. Our adaptation of this technology to evolve classifiers that identify mutant zebrafish based on movement data will provide a method to screen for drugs that improve the movement phenotype. We have found that *Dj-1*-deficient zebrafish display an overall loss of movement measured as a reduction in distance travelled, velocity, time spent moving and duration of a swimming episode; comparable to the bradykinesia observed in PD patients (Jankovic, 2008). Furthermore, the shorter duration of swimming episodes indicated more frequent periods of inactivity, akin to the freezing episodes observed in PD (Chee et al., 2009). There is, however, a great deal of subjectivity when studying features selected based on conventional movement measures. In comparison, the sliding window classifier is much more objective, automatically extracting features based on minute differences in movement that are undiscernible to the human eye, rather than the general features of movement. On examination of the input video data, we find that the discriminating classifiers often use data from when the fish turns. This suggests that the movement phenotype in parkinsonian zebrafish is similar to that found in PD patients, in whom bradykinesia is particularly evident when turning around (e.g. Chou and Lee, 2013). A further consideration to take into account is that a muscle disorder might be contributing to the movement phenotype, as loss of *dj-1* has been found to affect metabolic respiration in skeletal muscle cells and result in overall reduced body mass (Edson et al., 2019).

### Molecular signature of a PD brain: increased oxidative phosphorylation and cell cycle re-entry

Confirmation of a valid parkinsonian model was provided by GSEA (Subramanian et al., 2005), which revealed molecular signatures described in other genetic models of PD. The DJ-1 protein is sensitive to ROS (Canet-Avilés et al., 2004; Mitumoto et al., 2001), acting as a neuroprotector by quenching excess ROS at the mitochondrial membrane (Taira et al., 2004) and is also required for the subcellular localisation of hexokinase (HK1) to the mitochondria (Hauser et al., 2017). This links DJ-1 activity directly to mitochondrial function, alongside other notable autosomal recessive PD genes *PINK1* and *PRKN* that monitor and

maintain mitochondrial function. Mitochondrial dysfunction is accepted to be a major factor in the pathology of PD (reviewed in Cookson, 2012), although it is puzzling how defective mitochondria, organelles that are essential in all cells, leads to the specific depletion of SNc DA neurons. One explanation is that the high number of synapses associated with these neurons results in extraordinarily high energy demands (reviewed in Bolam and Pissadaki, 2012); indeed, there is a higher basal rate of mitochondrial oxidative phosphorylation in SNc DAs compared to DA neurons of the ventral tegmental area (Pacelli et al., 2015). Recently, good evidence for this hypothesis has been provided by the Trudeau laboratory, when they showed that a mouse engineered with increased axonal arborisation of SNc DA neurons is more vulnerable to PD-causative neurotoxins (Giguère et al., 2019). Our findings that genes associated with oxidative phosphorylation are altered in the *dj-1*<sup>-/-</sup> zebrafish is consistent with the notion that disruption of bioenergetics is a key feature of a neurodegenerative brain.

Upregulation of genes associated with oxidative phosphorylation in a mouse *Prkn*<sup>-/-</sup> model (Giguère et al., 2018) and a zebrafish *dj-1*<sup>-/-</sup> model of PD has been described (Edson et al., 2019), while dysregulation of genes associated with metabolism is consistent with a neurodegenerative state in PD (Giguère et al., 2018; Shi et al., 2015) as well as Alzheimer's disease (Demetrius et al., 2015). One interpretation (Yellen, 2018) is that the high energy demands in neurons requires the uncoupling of glycolysis and oxidative phosphorylation, and it may be that DJ-1 is required for this switch. Another interpretation is that neurodegeneration results from an 'inverse Warburg effect'; this has been proposed as a cause of sporadic Alzheimer's disease, in which oxidative phosphorylation is upregulated in ageing or otherwise impaired neurons as a compensatory response to meet energy needs (Demetrius et al., 2015), an idea supported by our data that oxidative phosphorylation genes are enriched in the absence of Dj-1.

GSEA also reveals that in zebrafish brains lacking Dj-1, there is an enrichment of genes associated with mitosis, such as those regulating the G2/M checkpoint, E2F target genes, as well as the cancer-related gene sets associated with epithelial mesenchymal transition, Myc targets and the mitotic spindles. This might seem surprising, because DJ-1 was first identified by its interaction with Myc and indeed can transform 3T3 cells when co-expressed with Myc or Ras (Nagakubo et al., 1997). However, PD genes have been implicated in cell cycle control (reviewed in West et al., 2005) and the re-activation of cell cycle proteins (including the E2F factors) has been detected in tissue from PD patients (Höglinger et al., 2007). Post-mitotic neurons have mechanisms in place that prevent cell cycle progression, and re-entering the cell cycle can lead to apoptosis; this may be a mechanism underlying neurodegeneration (reviewed in Folch et al., 2012; Herrup et al., 2004; Herrup and Yang, 2007). We conclude that the over-representation of mitotic gene sets, together with those indicative of altered metabolism, reflects the neurodegenerative state of *dj-1*<sup>-/-</sup> zebrafish brains.

### Machine learning can contribute to drug development

To date, most zebrafish models of PD have been transient models in larvae, generated using neurotoxins such as MPTP, which cause a rapid loss of the DA neurons (Bretaud et al., 2004; Feng et al., 2014), or by morpholino knockdown of PD-associated genes (Bretaud et al., 2007; Flinn et al., 2009; Zhao et al., 2011); in addition, an adult *pink1*<sup>-/-</sup> mutant zebrafish has been described (Flinn et al., 2013). A few studies have tested potential therapies on zebrafish models of PD 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). Given the progressive loss of DA neurons, an adult model of PD is intuitively more suitable, and evolving classifiers using tail bend data provides a more comprehensive, unbiased assessment of the movement phenotype. The novel use of machine learning described here has the potential to accelerate the drug discovery pipeline (Lam and Peterson, 2019) by allowing more comprehensive screening using non-mammalian vertebrate models of movement disorders.

## MATERIALS AND METHODS

### gRNA design

A synthetic gRNA was designed to target Cas9 cleavage in *dj-1* upstream of the codon for residue C106, in order to produce an early stop codon and loss of C106 from the translated protein. This cysteine is critical for the oxidative stress response of DJ-1 (Wilson, 2011). The online CRISPR design tool ChopChop (<https://chopchop.cbu.uib.no>) was used to identify a target sequence near the start of the *dj-1* coding sequence. A 20 bp target sequence (5'-GCCGGTTCAGTGCAGCCGTG-3') in exon 2 of *dj-1* was selected and incorporated into the forward primer for gRNA production. 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). This resulted in the following forward primer sequence: 5'-GCAGCTAATACGACTCACTATAGGCCGGTTCAGTGCAGCCGTGGTTT-TAGAGCTAGAAATAGCAAG-3'. *pink1* was disrupted using the same approach, with the modification of two gRNAs designed 100 bp apart: *pink1*(1) 5'-GCAGCTAATACGACTCACTATAGGGTGAAGCAGAAA-GTCGAAGGTTTTAGAGCTAGAAATAGCAAG-3' and *pink1*(2) 5'-GCAGCTAATACGACTCACTATAGCGCAGCTG-TTTATGAGGCTG-GTTTTAGAGCTAGAAATAGCAAG-3'. The reverse primer used for gRNA synthesis is common to all gRNAs: 5'-AAAAGCACCGACTCG-GTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAACT-TGCTATTTCTAGCTCTAAAAC-3' (Nakayama et al., 2014). For genotyping, the primers to amplify the *dj-1* genomic region were as follows: For, 5'-GGCTCTGGCCATCATTACTAT-3'; Rev, 5'-GTAAA-GTCAGACCTGTTTGTGTG-3'. Primers to amplify the *pink1* genomic region were as follows: For, 5'-GGCTGTATTTAGAAAAGAAGAAGTT-TCAG-3'; Rev, 5'-GCAGCACAGTACAATTGTCAACTATAAAA-3'.

### Generating mutant lines

The strain of zebrafish used was London Wild Type (LWT) and both males and females between 8 wpf and 14 wpf were analysed for movement phenotypes. Transcriptomic analyses were carried out on the brains of zebrafish at 16 wpf and the immunohistochemical analysis was carried out on the brains of zebrafish at 12 mpf. This study was carried out using procedures authorised by the UK Home Office in accordance with the Animals Scientific Procedures Act (1986) and approved by the Animal Welfare and Ethical Review Body at the University of York and the UK Home Office.

Cas9 protein was co-injected with *dj-1*-targeting single-guide RNA into single-cell embryos from the LWT zebrafish strain to produce the F0 fish that were outcrossed with wild type to generate heterozygous F1 mutants. Genotyping of the F1s revealed a male and female carrying the same mutation in *dj-1*; these heterozygotes were in-crossed to produce the F2 generation. F2 zebrafish homozygous for *dj-1* were out-crossed with wild type to generate a stock of heterozygous fish (F3); these were in-crossed and the offspring raised together (F4), genotyped, and used for molecular analysis and generating stocks of homozygous and wild-type fish for movement analysis. Genotyping the *dj-1* locus in zebrafish was carried out by PCR of the target region followed by BbvI restriction digest assay; digestion of wild-type *dj-1* amplicon results in products of 214 bp and 55 bp in length. The BbvI restriction site is lost in the mutated *dj-1* so the PCR product remains 286 bp in length. *pink1* was disrupted using two gRNAs designed 100 bp apart, resulting in a 101 bp deletion detectable by gel

electrophoresis following PCR of the target region. The breeding strategy was the same as described above.

### Western blotting

Zebrafish brains were dissected from adults and flash frozen in liquid nitrogen before storing at  $-80^{\circ}\text{C}$ . Frozen brains were homogenised in 100  $\mu\text{l}$  lysis buffer consisting of 50 mM Tris-HCl (pH 7.4), 10 mM sodium glycerophosphate, 10 mM sodium pyrophosphate, 150 mM NaCl, 5 mM  $\text{MgCl}_2$ , 1 mM EDTA, 1 mM DTT, 1% Triton X-100, 10% glycerol, 1 mM sodium orthovanadate and cOmplete<sup>TM</sup>Mini EDTA-free Protease Inhibitor Cocktail (Roche). The lysate was centrifuged at 21,000  $g$  for 10 min prior to running on a 12% SDS-PAGE gel. After transfer to a PVDF membrane, immunodetection used the anti-Dj-1 polyclonal antibody (PA5-72638, Invitrogen) and anti-Gapdh monoclonal antibody (G8795, Sigma-Aldrich), both at a dilution of 1:50,000. Horseradish peroxidase (HRP)-linked anti-rabbit IgG (7074, Cell Signaling Technology) and HRP-linked anti-mouse IgG (62-650, Invitrogen) were used at concentrations of 1:2000 and 1:4000, respectively.

### qRT-PCR

Zebrafish brains were dissected out of adults before homogenising in 1 ml TRIzol reagent (Sigma-Aldrich) for RNA extraction. Following centrifugation at  $4^{\circ}\text{C}$  for 10 min at 21,000  $g$ , RNA was purified from the upper aqueous phase using the RNA Clean & Concentrator Kit (ZYMO) following the manufacturer's instructions. Complementary DNA (cDNA) was synthesised from 1  $\mu\text{g}$  RNA by reverse transcription using Superscript IV Reverse Transcriptase (Thermo Fisher Scientific) following the manufacturer's instructions with random hexamer primers. Primers were designed to amplify a product of 50-150 bp with one primer in each pair crossing an exon junction (Table S1). Each quantitative PCR analysis was carried out with a minimum of three biological replicates and three technical repeats. The mean cycle threshold (Ct) value was normalised to the Ct of reference gene *efla* (also known as *eeftall1*), 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  $2^{(-\text{ddCt})}$ . Student's *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 s.e.m.

### Immunofluorescence

Adult zebrafish were euthanised at 12 mpf, and whole brains were extracted and fixed overnight at  $4^{\circ}\text{C}$  in a 4% paraformaldehyde/PBS solution. Brains were washed twice in PBS for 10 min. Brains were then sectioned transversely in 3% agarose/PBS by vibratome at a thickness of 100  $\mu\text{m}$ . Based on the protocol by Matsui and Sugie (2017), brains were incubated in 10 mM sodium citrate buffer (pH 8.5) for 2 h at  $80^{\circ}\text{C}$ . Sections were washed twice in PBS with 1% Triton X-100 (PBS/1% Tx) for 15 min. Sections were then blocked in 2% bovine serum albumin (BSA)/PBS/1% Tx for 30 min. After blocking, sections were incubated overnight at  $4^{\circ}\text{C}$  with a 1:500 dilution of mouse anti-Th (22941, Immunostar) in 2% BSA/PBS/1% Tx. Sections were then washed four times in PBS/1% Tx for 30 min. Sections were incubated overnight at  $4^{\circ}\text{C}$  with a 1:500 dilution of goat anti-mouse IgG1 conjugated to Alexa Fluor 488 (A21121, Invitrogen) in 2% BSA/PBS/1% Tx. Sections were then washed four times in PBS/1% Tx for 30 min. Staining for nuclei was carried out by incubating the sections in 1  $\mu\text{g}/\text{ml}$  Hoechst 33342 in PBS/1% Tx for 10 min. Sections were then washed four times in PBS/1% Tx for 30 min. Sections were mounted on 15 mm cavity microscope slides before imaging.

### RNA-seq

RNA was extracted and purified as described above from the brains of three *dj-1*<sup>-/-</sup> mutants and three wild-type siblings at 16 wpf. A NanoDrop ND-1000 Spectrophotometer was used to quantify each RNA sample and the RNA integrity was measured using an Agilent 2100 Bioanalyzer. The NEBNext<sup>®</sup> Poly(A) mRNA Magnetic Isolation Module was used to isolate mRNAs from the total RNA before preparing cDNA libraries using the

NEBNext<sup>®</sup> Ultra RNA Library Prep Kit for Illumina<sup>®</sup> 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 $\times$ 150 bp paired end reads on one lane of an Illumina<sup>®</sup> HiSeq 3000 system. Cutadapt 1.16 was used to trim the adaptor sequences from each read. The fastq files generated were then aligned to the genome assembly GRCz11, the transcriptome was annotated using RefSeq (NCBI) and transcript abundance was quantified using Salmon 0.10.2. 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. A volcano plot of  $-\log_{10}(\text{pval})$  against  $\log_2(\text{fold change})$  was generated using MATLAB, and a heatmap representing the changes in expression of a curated set of these transcripts was created by uploading the transcripts per million (TPM) values to <https://software.broadinstitute.org/morpheus/>.

### GSEA

A GSEA was performed using the GSEA 4.0.0 (Broad Institute) software on the RNA-seq data. All of the genes from the RNA-seq data were used in the GSEA to identify gene sets enriched in a certain phenotypic class. Permutations of the phenotype or genotype labels are used to calculate the statistical significance of a gene set enrichment score (Subramanian et al., 2005). Following instructions on the Broad Institute website ([software.broadinstitute.org/gsea](https://software.broadinstitute.org/gsea)) an expression dataset file (.gct) was created containing the TPM values for each transcript, and a phenotype labels file (.cls) was created to label the phenotype of each sample. The Zebrafish.chip file was used to translate zebrafish transcript names to Human Genome Organisation (HUGO) gene symbols. GSEA was then performed to detect enriched HALLMARK gene sets from the Molecular Signature Database (Broad Institute). 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](https://software.broadinstitute.org/gsea)). Gene sets with a normalised enrichment score (NES) $\geq 1.5$ , FDR q-value  $\leq 0.05$  and nominal *P*-value  $\leq 0.05$  were considered statistically significant.

### Video capture

An Aquatics Habitat breeding tank was modified to record zebrafish swimming while keeping the fish in frame and eliminating any reflection and shadow. To achieve this, a white plastic insert of 145 mm $\times$ 90 mm was designed to fit the bottom of the tank, and a hollow square frustum with openings at the top and bottom was created with a camera fitted above (Fig. 3A). The insert allowed for background subtraction and the frustum kept the fish in frame and reduced reflections and shadows. A raised platform on the lid positions the camera so that the fish remain in frame at all times. Following an acclimation time of 10 min, zebrafish were recorded for 5 min. Recordings were carried out between the hours of 14:00 and 17:00 using a GoPro Hero 3 (GoPro, San Mateo, CA, USA) at 100 frames/s at a resolution of 1280 $\times$ 720 pixels.

### Extracting data

The computational processing resources required to analyse a full 5-min recording are prohibitive, and previous work on assessing movement disorders in humans found that 30-s samples were sufficient (Lones et al., 2013). Consequently, two 30-s clips were selected for processing from the first and second halves of the 5-min recording using bespoke ShadowFish tracking software. Inspired by methods used for larval fish (Budick and O'Malley, 2000), this software was written to identify a fish using background subtraction and trace its midline, dividing it into six parts of equal length (Fig. 3B). For each frame of recording, the absolute *x* and *y* coordinates of the two end points and five vertices along the spine are extracted, in addition to the angle at each vertex. This produces five bend angles and seven *x,y* coordinates from each frame. ShadowFish automatically determines the lateral orientation of fish by the direction of movement and therefore the normalised bending angle can be determined accordingly. Each processed video clip was visually assessed before being used to calculate the features of movement or



evolve a classifier. Any clip in which the fish was out of the field of view, or the ShadowFish tracking software misidentified a reflection or shadow of the zebrafish for the fish itself, was removed. The extracted movement data are written to a CSV file.

### Calculating features of movement

MATLAB (MathWorks, Natick, MA, USA) was used to calculate features of movement, inspired by previous studies (Table S4), from the extracted raw movement data. The distance travelled by a zebrafish in a recording was determined by taking the absolute coordinates of the fish every 100 frames and using the Pythagoras theorem to work out the distance covered in between. It was necessary to establish when a fish was stationary so data points at those times were not included in certain calculations of features. When a fish travelled less than 5 mm over a 1-s period it was classed as stationary for that second. The time spent moving was the number of seconds in a clip that the fish travelled over 5 mm out of the total 30 s, displayed as a percentage. The mean velocity was calculated by dividing the distance travelled by the number of seconds spent moving. Swimming episode duration was how long a fish was classed as moving before a stationary period. To study the tail beat frequency of a zebrafish in a recording, the sum of the five angles along the spine was plotted over time. A five-frame moving average was also used to help reduce any noise in the angles from the raw data. The points of maximum tail bend were identified as the peaks on the plot. Only peaks with a minimum prominence of five degrees and a minimum separation of five frames were identified, to leave out spikes in the data that were unlikely to be full tail bends. Tail bend amplitudes were the angles identified at the peaks of the plot. Tail beat frequencies and tail bend amplitudes were also calculated at different speeds, in case a difference in movement phenotype was more apparent at a certain speed. Low-speed movement was defined as  $0.5 \leq X < 2$  cm/s. Medium-speed movement was defined as  $2 \leq X < 4$  cm/s and high-speed movement was  $\geq 4$  cm/s.

### Methodology and statistics of movement data analyses

Multiple features of movement were extracted from the mutant and wild-type fish and all parameters that were measured have been reported. In order to raise and film the numbers of fish required for movement analyses, two or more in-crossings of F4 homozygous mutants or homozygous wild-type siblings were carried out. Statistical analysis of the comparison of movement in wild-type and mutant fish produced the *P*-values quoted, derived from either a Student's *t*-test, if the data were found to be normally distributed, or a Mann–Whitney *U*-test, if the data were non-parametric. Statistical analyses were carried out using GraphPad Prism5 to assess differences in the features of movement for mutant zebrafish and age-matched controls. Any features extracted that contained no numerical value were removed prior to analysis. For example, if a fish only travelled at low and medium speed in a clip then the high-speed tail beat frequency extracted would contain no numerical value and would therefore be left out of future analyses. The features containing numerical values were then 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  $1.5 \times$  IQR below the lower quartile or greater than the upper quartile were identified as outliers and removed prior to statistical analysis. This criterion for excluding outliers was established to remove fish that failed to swim in a clip; without swimming a fish contributes no real movement data to the analysis. Features were compared between classes using a Student's *t*-test if the data were normally distributed or a Mann–Whitney *U*-test if the data were non-parametric. The investigator was not blinded to the genotype of zebrafish during video capture; however, when evolving classifiers, the extracted features, or PC values, fish were randomly organised into the three datasets – training, validation and test – and so the investigator was blinded to the genotype of fish when evolving a classifier.

### Training classification models

PCA was used to transform the multivariate time series of bend angles into a univariate time series suitable for classification. A type of EA called implicit context representation CGP has been effective at classifying PD movement

in humans (Lones et al., 2017) by finding mathematical expressions that describe patterns of movement indicative of dyskinesia. This approach was used to train classifiers to discriminate the features of movement from PC2 time series data of *dj-1<sup>-/-</sup>* fish from those of wild-type controls. Sliding window classifiers trained using the PC2 time series data generate symbolic mathematical expressions that describe discriminatory local patterns of movement within the data. The classifiers were trained using a stochastic optimisation algorithm and this was repeated multiple times (20) to generate multiple models. The features of movement and PC2 time series data from each recording were organised into three datasets: training, validation and test. To train classifiers with the features of movement, 60% of the recordings went into the training set, used to evaluate the effectiveness of classifiers during training. The remaining 40% of recordings were split equally between the validation and test sets. This split of the dataset proved the most effective to classify extracted features of movement. With a lower number of overall recordings used to train the CGP classifier, it made more sense to have a higher number in the training dataset. Aiming for a high number of clips to train the classifier allows it to recognise a common movement pattern. Whereas classifiers trained with the PC2 time series data had 50% of the recordings in the training set, 25% in the validation set and 25% in the test set, the sliding window classifier was evolved with a higher number of clips so more of the clips could go in the validation and test sets. This split of data was found best for evolving the most effective sliding window classifier. The validation set was used both for early stopping (i.e. to stop training when a model began to over-fit the data) and to select the most general classifier from the multiple models generated during training. The test set was used to get an unbiased measure of the selected model's discriminatory power. The tracking data were visually assessed to make sure they were of a high quality and the number of clips used in each class are shown in Tables S5 and S6. The sliding window classifiers were evolved using the AUC as a measure of fitness that is not sensitive to class imbalances.

### Acknowledgements

We thank the Department of Electronic Engineering Workshop for tank design; Sally James, Lesley Gilbert, Katherine Newling and John Davey in the Bioscience Technology Facility for support with RNA-seq analysis; and Sean Sweeney, Ian Hitchcock and Steve Johnson for helpful discussions.

### Competing interests

The authors declare no competing or financial interests.

### Author contributions

Conceptualization: S.L.S., M.E.P.; Methodology: M.A.L.; Software: M.B.; Formal analysis: S.L.S.; Investigation: G.L.H., M.A.L.; Writing - original draft: G.L.H., M.E.P.; Writing - review & editing: S.L.S., M.E.P.; Supervision: P.D.C., S.L.S., M.E.P.; Project administration: M.E.P.; Funding acquisition: P.D.C., S.L.S., M.E.P.

### Funding

This work was supported by a PhD studentship to G.L.H. funded by the Engineering and Physical Sciences Research Council (EPSRC) and a York Cross-disciplinary Centre for Systems Analysis (YCSSA) summer studentship to M.B. M.E.P., P.D.C. and S.L.S. received funding from the Wellcome Trust-supported Centre for Chronic Diseases and Disorders and Centre for Future Health at the University of York.

### Data availability

The data from the RNA-seq analysis have been deposited in NCBI's GEO (Edgar et al., 2002) under accession number GSE135271 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE135271>). The ShadowFish tracking software and MATLAB code to extract the features of movement can be accessed online at [github.com/ghughes/york/ShadowFish](https://github.com/ghughes/york/ShadowFish).

### Supplementary information

Supplementary information available online at <https://dmm.biologists.org/lookup/doi/10.1242/dmm.045815.supplemental>

### References

Allen, P. J. (2012). Creatine metabolism and psychiatric disorders: does creatine supplementation have therapeutic value? *Neurosci. Biobehav. Rev.* **36**, 1442–1462. doi:10.1016/j.neubiorev.2012.03.005

- Alpaydin, E. (2016). *Machine Learning: The New AI*. Cambridge, Massachusetts: MIT Press.
- 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. doi:10.1523/JNEUROSCI.0979-08.2008
- Ayodele, T. O. (2010). *Types of Machine Learning Algorithms, New Advances in Machine Learning*. London: IntechOpen.
- 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. doi:10.1016/j.brainres.2006.07.057
- Bandyopadhyay, S. and Cookson, M. R. (2004). Evolutionary and functional relationships within the DJ-1 superfamily. *BMC Evol. Biol.* **4**, 1-9. doi:10.1186/1471-2148-4-6
- 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. doi:10.1242/dev.00799
- Bland, J. M. and Altman, D. G. (2007). Agreement between methods of measurement with multiple observations per individual. *J. Biopharm. Stat.* **17**, 571-582. doi:10.1080/10543400701329422
- 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. doi:10.1002/mds.25135
- 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. doi:10.1126/science.1077209
- Bretaud, S., Lee, S. and Guo, S. (2004). Sensitivity of zebrafish to environmental toxins implicated in Parkinson's disease. *Neurotoxicol. Teratol.* **26**, 857-864. doi:10.1016/j.ntt.2004.06.014
- 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. doi:10.1111/j.1471-4159.2006.04291.x
- 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.
- 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-sulfenic acid-driven mitochondrial localization. *Proc. Natl. Acad. Sci. USA* **101**, 9103-9108. doi:10.1073/pnas.0402959101
- 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. doi:10.1093/brain/awp053
- Chou, P.-Y. and Lee, S.-C. (2013). Turning deficits in people with Parkinson's disease. *Tzu Chi Med. J.* **25**, 200-202. doi:10.1016/j.tcmj.2013.06.003
- 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  $\gamma$  in human skeletal muscle cells. *Mol. Cell. Biol.* **29**, 20-30. doi:10.1128/MCB.00544-08
- 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. doi:10.1038/nature04779
- Clements, C. M., McNally, R. S., Conti, B. J., Mak, T. W. and Ting, J. P.-Y. (2006). DJ-1, a cancer- and Parkinson's disease-associated protein, stabilizes the antioxidant transcriptional master regulator Nrf2. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 15091-15096. doi:10.1073/pnas.0607260103
- Cookson, M. R. (2003). Crystallizing ideas about Parkinson's disease. *Proc. Natl. Acad. Sci. USA* **100**, 9111-9113. doi:10.1073/pnas.1633722100
- 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. doi:10.1101/cshperspect.a009415
- 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. doi:10.1016/j.neuroscience.2017.10.018
- 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. doi:10.1093/brain/122.8.1437
- 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. doi:10.1016/j.nbd.2014.06.009
- Demetrius, L. A., Magistretti, P. J. and Pellerin, L. (2015). Alzheimer's disease: the amyloid hypothesis and the Inverse Warburg effect. *Front. Physiol.* **5**, 1-20. doi:10.3389/fphys.2014.00522
- 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. doi:10.1523/JNEUROSCI.5092-10.2011
- Edgar, R., Domrachev, M. and Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207-210. doi:10.1093/nar/30.1.207
- 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**, 8306-8322. doi:10.1007/s12035-019-01667-w
- 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.-D. 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. doi:10.1089/zeb.2013.0950
- Filippi, A., Dürr, K., Ryu, S., Willaredt, M., Holzschuh, J. and Driever, W. (2007). Expression and function of nr4a2, lmx1b, and pitx3 in zebrafish dopaminergic and noradrenergic neuronal development. *BMC Dev. Biol.* **7**, 1-21. doi:10.1186/1471-213X-7-135
- Flinn, L., Mortiboys, H., Volkmann, K., Koster, 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. doi:10.1093/brain/awp108
- 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. doi:10.1002/ana.23999
- Folch, J., Junyent, F., Verdagué, E., Auladell, C., Pizarro, J. G., Beas-Zarate, C., Pallàs, M. and Camins, A. (2012). Role of cell cycle re-entry in neurons: a common apoptotic mechanism of neuronal cell death. *Neurotox. Res.* **22**, 195-207. doi:10.1007/s12640-011-9277-4
- Frohlich, H., Claes, K., De Wolf, C., Van Damme, X. and Michel, A. (2018). A machine learning approach to automated gait analysis for the Noldus Catwalk System. *IEEE Trans. Biomed. Eng.* **65**, 1133-1139.
- 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. doi:10.1523/JNEUROSCI.3284-13.2013
- Gallardo, V. E., Varshney, G. K., Lee, M., Bupp, S., Xu, L., Shinn, P., Crawford, N. P., Inglese, J. and Burgess, S. M. (2015). Phenotype-driven chemical screening in zebrafish for compounds that inhibit collective cell migration identifies multiple pathways potentially involved in metastatic invasion. *Dis. Model. Mech.* **8**, 565-576. doi:10.1242/dmm.018689
- Gao, C., Smith, S., Lones, M., Jamieson, S., Alty, J., Cosgrove, J., Zhang, P., Liu, J., Chen, Y., Du, J. et al. (2018). Objective assessment of bradykinesia in Parkinson's disease using evolutionary algorithms: clinical validation. *Transl. Neurodegener.* **7**, 18. doi:10.1186/s40035-018-0124-x
- Giguère, N., Pacelli, C., Saumure, C., Bourque, M.-J., Matheoud, D., Levesque, D., Slack, R. S., Park, D. S. and Trudeau, L.-É. (2018). Comparative analysis of Parkinson's disease-associated genes in mice reveals altered survival and bioenergetics of Parkin-deficient dopamine neurons. *J. Biol. Chem.* **293**, 9580-9593. doi:10.1074/jbc.RA117.000499
- 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. *PLoS Genet.* **15**, e1008352. doi:10.1371/journal.pgen.1008352
- 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. doi:10.1371/journal.pone.0005777
- Greffard, S., Verny, M., Bonnet, A.-M., Beinis, J.-Y., Gallinari, C., Meaume, S., Piette, F., Hauw, J.-J. and Duyckaerts, C. (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. doi:10.1001/archneur.63.4.584
- Han, T., Liu, M. and Yang, S. (2018). DJ-1 alleviates angiotensin II-induced endothelial progenitor cell damage by activating the PPAR $\gamma$ /HO-1 pathway. *J. Cell. Biochem.* **400**, 392-400. doi:10.1002/jcb.26191
- 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**, 70. doi:10.1186/s13024-017-0212-x
- Hernandez, D. G., Reed, X. and Singleton, A. B. (2016). Genetics in Parkinson disease: Mendelian versus non-Mendelian inheritance. *J. Neurochem.* **139**, 59-74. doi:10.1111/jnc.13593
- Herrup, K. and Yang, Y. (2007). Cell cycle regulation in the postmitotic neuron: oxymoron or new biology? *Nat. Rev. Neurosci.* **8**, 368-378. doi:10.1038/nrn2124



- 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. doi:10.1523/JNEUROSCI.3347-04.2004
- Höglinger, G. U., Breunig, J. J., Depboylu, C., Rouaux, C., Michel, P. P., Alvarez-fischer, D., Bouillier, A.-L., DeGregori, J., Oertel, W. H., Rákic, P. et al. (2007). The pRb/E2F cell-cycle pathway mediates cell death in Parkinson's disease. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 3585-3590. doi:10.1073/pnas.0611671104
- 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. doi:10.1016/S0925-4773(01)00287-8
- Hu, Q. and Wang, G. (2016). Mitochondrial dysfunction in Parkinson's disease. *Transl. Neurodegener.* **5**, 14. doi:10.1186/s40035-016-0060-6
- 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. doi:10.1136/jnnp.55.3.181
- Jackson-Lewis, V. and Przedborski, S. (2007). Protocol for the MPTP mouse model of Parkinson's disease. *Nat. Protoc.* **2**, 141-151. doi:10.1038/nprot.2006.342
- 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. doi:10.1073/pnas.1309143110
- Jankovic, J. (2008). Parkinson's disease: clinical features and diagnosis. *J. Neurol. Neurosurg. Physiatry* **79**, 368-376. doi:10.1136/jnnp.2007.131045
- 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. doi:10.1093/hmg/ddv369
- Kim, R. H., Peters, M., Jang, Y. J., Shi, W., Pintilie, M., Fletcher, G. C., DeLuca, C., Liepa, J., Zhou, L., Snow, B. et al. (2005). DJ-1, a novel regulator of the tumor suppressor PTEN. *Cancer Cell* **7**, 263-273. doi:10.1016/j.ccr.2005.02.010
- 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. doi:10.1111/j.1471-4159.2009.06350.x
- Lam, P.-Y. and Peterson, R. T. (2019). Developing zebrafish disease models for in vivo small molecule screens. *Curr. Opin. Chem. Biol.* **50**, 37-44. doi:10.1016/j.cbpa.2019.02.005
- Lan, A.-P., Chen, J., Zhao, Y., Chai, Z. and Hu, Y. (2017). mTOR signaling in Parkinson's disease. *Neuromolecular Med.* **19**, 1-10. doi:10.1007/s12017-016-8417-7
- Lee, Y., Stevens, D. A., Kang, S.-U., Jiang, H., Lee, Y. I., Ko, H. S., Scarffe, L. A., Umanah, G. E., Kang, H., Ham, S. et al. (2017). PINK1 primes parkin-mediated ubiquitination of PARIS in dopaminergic neuronal survival. *Cell Rep.* **18**, 918-932. doi:10.1016/j.celrep.2016.12.090
- Lo Bianco, C., Ridet, J.-L., Schneider, B. L., Déglon, N. and Aebischer, P. (2002).  $\alpha$ -synucleinopathy and selective dopaminergic neuron loss in a rat lentiviral-based model of Parkinson's disease. *Proc. Natl. Acad. Sci. USA* **99**, 10813-10818. doi:10.1073/pnas.152339799
- Lones, M. A., Smith, S. L., Alty, J. E., Lacy, S. E., Possin, K. L., Jamieson, D. R. S. and Tyrrell, A. (2013). evolving classifiers to recognize the movement characteristics of Parkinson's disease patients. *IEEE Trans. Evol. Comput.* **18**, 559-576. doi:10.1109/TEVC.2013.2281532
- Lones, M. A., Alty, J. E., Cosgrove, J., Duggan-Carter, P., Jamieson, S., Naylor, R. F., Turner, A. J. and Smith, S. L. (2017). A new evolutionary algorithm-based home monitoring device for Parkinson's dyskinesia. *J. Med. Syst.* **41**, 176. doi:10.1007/s10916-017-0811-7
- Lu, B. and Vogel, H. (2009). *Drosophila* models of neurodegenerative diseases. *Annu. Rev. Pathol. Mech. Dis.* **4**, 315-342. doi:10.1146/annurev.pathol.3.121806.151529
- 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. doi:10.1083/jcb.200110081
- Matsui, H. and Sugie, A. (2017). An optimized method for counting dopaminergic neurons in zebrafish. *PLoS ONE* **12**, e0184363. doi:10.1371/journal.pone.0184363
- Mazzoni, P., Shabbott, B. and Cortés, J. C. (2012). Motor control abnormalities in Parkinson's disease. *Cold Spring Harb. Perspect. Med.* **2**, a009282. doi:10.1101/cshperspect.a009282
- McHenry, M. J. and Lauder, G. V. (2005). The mechanical scaling of coasting in zebrafish (*Danio rerio*). *J. Exp. Biol.* **208**, 2289-2301. doi:10.1242/jeb.01642
- Miller, J. F. (2011). *Cartesian Genetic Programming*, in: *Cartesian Genetic Programming*. Natural Computing Series, pp. 17-34. Berlin, Heidelberg: Springer.
- Mitsumoto, A., Nakagawa, Y., Takeuchi, A., Okawa, K., Iwamatsu, A. and Takanezawa, Y. (2001). Oxidized forms of peroxiredoxins and DJ-1 on two-dimensional gels increased in response to sublethal levels of paraquat. *Free Radic. Res.* **35**, 301-310. doi:10.1080/10715760100300831
- Nagakubo, D., Taira, T., Kitaura, H., Ikeda, M., Tamai, K., Iguchi-Arigo, S. M. 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. doi:10.1006/bbrc.1997.6132
- 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. doi:10.1016/B978-0-12-801185-0.00017-9
- Oswald, M. C. W., Brooks, P. S., Zwart, M. F., Mukherjee, A., West, R. J. H., Giachello, C. N. G., Morarach, K., Baines, R. A., Sweeney, S. T. and Landgraf, M. (2018). Reactive oxygen species regulate activity-dependent neuronal plasticity in *Drosophila*. *eLife* **7**, e39393. doi:10.7554/eLife.39393
- Pacelli, C., Giguère, N., Bourque, M.-J., Lévesque, M., Slack, R. S. and Trudeau, L.-É. (2015). Elevated mitochondrial bioenergetics and axonal arborization size are key contributors to the vulnerability of dopamine neurons. *Curr. Biol.* **25**, 2349-2360. doi:10.1016/j.cub.2015.07.050
- 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. doi:10.1073/pnas.0409598102
- Postuma, R. B., Berg, D., Stern, M., Poewe, W., Olanow, C. W., Oertel, W., Obeso, J., Marek, K., Litvan, I., Lang, A. E. et al. (2015). MDS clinical diagnostic criteria for Parkinson's disease. *Mov. Disord.* **30**, 1591-1601. doi:10.1002/mds.26424
- 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. doi:10.1371/journal.pone.0018568
- 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. doi:10.1016/S0006-8993(00)03174-7
- 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. doi:10.1016/S0165-3806(02)00354-1
- 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. doi:10.1212/WNL.0000000000002350
- Rousseaux, M. W. C., Marcogliese, P. C., Qu, D., Hewitt, S. J., Seang, S., Kim, R. H., Slack, R. S., Schlossmacher, M. G., Lagace, D. C., Mak, T. W. et al. (2012). Progressive dopaminergic cell loss with unilateral-to-bilateral progression in a genetic model of Parkinson disease. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 15918-15923. doi:10.1073/pnas.1205102109
- Sandell, L. L., Guan, X.-J., Ingram, R. and Tilghman, S. M. (2003). Gata, a creatine synthase enzyme, is imprinted in mouse placenta. *Proc. Natl. Acad. Sci. USA* **100**, 4622-4627. doi:10.1073/pnas.0230424100
- 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. doi:10.1371/journal.pgen.1000914
- Shi, S. Y., Lu, S.-Y., Sivasubramaniyam, T., Revelo, X. S., Cai, E. P., Luk, C. T., Schroer, S. A., Patel, P., Kim, R. H., Bombardier, E. et al. (2015). DJ-1 links muscle ROS production with metabolic reprogramming and systemic energy homeostasis in mice. *Nat. Commun.* **6**, 7415. doi:10.1038/ncomms8415
- 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. doi:10.1038/77060
- 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. doi:10.1049/iet-syb.2015.0030
- Stewart, A. M., Gerlai, R. and Kalueff, A. V. (2015). Developing highER-throughput zebrafish screens for in-vivo CNS drug discovery. *Front. Behav. Neurosci.* **9**, 14. doi:10.3389/fnbeh.2015.00014
- 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. U.S.A.* **102**, 15545-15550. doi:10.1073/pnas.0506580102
- Subramanian, A., Kuehn, H., Gould, J., Tamayo, P. and Mesirov, J. P. (2007). GSEA-P: a desktop application for gene set enrichment analysis. *Bioinformatics* **23**, 3251-3253. doi:10.1093/bioinformatics/btm369
- 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. U.S.A.* **108**, 4129-4134. doi:10.1073/pnas.1014769108



- Taira, T., Saito, Y., Niki, T., Iguchi-Arigo, S. M. M., Takahashi, K. and Ariga, H. (2004). DJ-1 has a role in antioxidative stress to prevent cell death. *EMBO Rep.* **5**, 213-218. doi:10.1038/sj.embor.7400074
- Thomas, K. J., Mccoy, M. K., Blackinton, J., Beilina, A., van der Brug, M., Sandebring, A., Miller, D., Maric, D., Cedazo-minguez, A. and Cookson, M. R. (2011). DJ-1 acts in parallel to the PINK1/parkin pathway to control mitochondrial function and autophagy. *Hum. Mol. Genet.* **20**, 40-50. doi:10.1093/hmg/ddq430
- Tysnes, O.-B. and Storstein, A. (2017). Epidemiology of Parkinson's disease. *J. Neural Transm.* **124**, 901-905. doi:10.1007/s00702-017-1686-y
- 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  $\alpha$ -synuclein. *J. Neurosci.* **20**, 6021-6029. doi:10.1523/JNEUROSCI.20-16-06021.2000
- Ved, R., Saha, S., Westlund, B., Perier, C., Burnam, L., Sluder, A., Hoener, M., Rodrigues, C. M. P., Alfonso, A., Steer, C. et al. (2005). Similar patterns of mitochondrial vulnerability and rescue induced by genetic modification of  $\alpha$ -synuclein, parkin, and DJ-1 in *Caenorhabditis elegans*. *J. Biol. Chem.* **280**, 42655-42668. doi:10.1074/jbc.M505910200
- 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. doi:10.15252/embj.201593585
- Volpatti, J. R., Endo, Y., Knox, J., Groom, L., Brennan, S., Noche, R., Zuercher, W., Roy, P., Dirksen, R. T. and Dowling, J. J. (2020). Identification of drug modifiers for RYR1-related myopathy using a multi-species discovery pipeline. *eLife* **9**, e52946. doi:10.7554/eLife.52946
- West, A. B., Dawson, V. L. and Dawson, T. M. (2005). To die or grow: Parkinson's disease and cancer. *Trends Neurosci.* **28**, 348-352. doi:10.1016/j.tins.2005.05.002
- 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. doi:10.1371/journal.pone.0030745
- Wilson, M. A. (2011). The role of cysteine oxidation in DJ-1 function and dysfunction. *Antioxid. Redox Signal.* **15**, 111-122. doi:10.1089/ars.2010.3481
- Winklhofer, K. F. and Haass, C. (2010). Mitochondrial dysfunction in Parkinson's disease. *BBA Mol. Basis Dis.* **1802**, 29-44. doi:10.1016/j.bbadis.2009.08.013
- Wullmann, M. F., Rupp, B. and Reichert, H. (1996). Neuroanatomy of the Zebrafish Brain: A Topological Atlas. Basel: Birkhäuser Verlag.
- 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. doi:10.1093/hmg/ddi134
- Yellen, G. (2018). Fueling thought: management of glycolysis and oxidative phosphorylation in neuronal metabolism. *J. Cell Biol.* **217**, 2235-2246. doi:10.1083/jcb.201803152
- Zhao, T., de Graaff, E., Breedveld, G. J., Loda, A., Severijnen, L.-A., Wouters, C. H., Verheijen, F. W., Dekker, M. C. J., Montagna, P., Willemsen, R. et al. (2011). Loss of nuclear activity of the FBXO7 protein in patients with parkinsonian-pyramidal syndrome (PARK15). *PLoS ONE* **6**, e16983. doi:10.1371/journal.pone.0016983
- 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. doi:10.1371/journal.pone.0048911
- Zondler, L., Miller-Fleming, L., Repici, M., Gonçalves, S., Tenreiro, S., Rosado-Ramos, R., Betzer, C., Straatman, K. R., Jensen, P. H., Giorgini, F. et al. (2014). DJ-1 interactions with  $\alpha$ -synuclein attenuate aggregation and cellular toxicity in models of Parkinson's disease. *Cell Death Dis.* **5**, e1350. doi:10.1038/cddis.2014.307

## SUPPLEMENTARY DATA

Table S1

## Primer sequences for qRT-PCR analysis (5' – 3').

Primer pairs that were designed based on previous publications: *dat*, *th* (Chen et al., 2012), *pitx3* (Sanchez-Simon et al., 2010) and *ef1a* (Fan et al., 2010).

Name	Forward primer	Reverse primer
<i>dj-1</i>	CGCACACAAACAGGGTCCATA	AAGCAAACCTCCAGGCAGAA
<i>dat</i>	CGTCACCAACGGTGAATCTA	TGCCGATGGCCTCAATTAGTA
<i>th</i>	GACGGAAGATGATCGGAGACA	CCGCCATGTTCCGATTTCT
<i>pitx3</i>	GACAACAGTGACACAGAGAAGT	TGTCGGGATAACGGTTTCT
<i>syn2a</i>	CAAGGTTGTGAGGTCATTTAAGC	GTAAGTGGAGGCCGATGATTAAG
<i>ef1a</i>	TTGAGAAGAAAATCGGTGGTGCTG	GGAACGGTGTGATTGAGGGAAATTC
<i>pkma</i>	CCGCACACAGCACATGATTG	ACTGCATCTGAAGGGTCACG
<i>vamp3</i>	GAGCTGGATGATCGTGCTGA	ATGGCCCACATCTTGACGTT
<i>brk</i>	GCAGGTCTCGTTTAGCCACT	TTGTGACTCTGGCTTCGATGT
<i>gpx3</i>	ACCATAAACGGGACGCAGTT	AGTGCATTCAGTTCCACGTACT
<i>gatm</i>	TCCAGAGGAGTCAGGTTACGA	ATGTTGAAGGTGGCGTCGAT
<i>abcf2a</i>	CGGAACAGGCAAATCCATGC	GCTCTGGAATTGGCACCTCT
<i>kcnk3b</i>	AGGTCGTGTTGCTCCTGAAG	CATGGCCATAACCTATGGTTGTTA

**Table S2a****Genes significantly up-regulated in *dj-1*<sup>-/-</sup> zebrafish brains**

Genes that were up-regulated in the brains of *dj-1*<sup>-/-</sup> mutants (n = 3) when compared to their wild-type siblings (n = 3) at 16 wpf. Differentially expressed transcripts from the RNA-Seq analysis with a Q-value of <0.05 were considered significant. tpm = transcripts per million.

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

**Table S2b****Genes significantly down-regulated in *dj1*<sup>-/-</sup> zebrafish brains**

Genes that were down-regulated in the brains of *dj-1*<sup>-/-</sup> mutants (n = 3) when compared to their wild-type siblings (n = 3) at 16 wpf. Differentially expressed transcripts from the RNA-Seq analysis with a Q-value of <0.05 were considered significant. tpm = transcripts per million.

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



**Table S3**  
**Gene set enrichment analysis (GSEA)**

**Hallmark gene sets enriched in the dj-1<sup>-/-</sup> zebrafish brain.** A list of Hallmark gene sets found significantly enriched in the dj-1<sup>-/-</sup> 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.

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 S4****Selection and calculations of extracted features of movement**

The features of movement being extracted and the previous research, investigating movement phenotypes in zebrafish, that inspired them. The principles used in this work are described in addition to the principles used in previous works.

Feature	Based on	Stage in previous work	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). Pythagoras theorem used to calculate distance travelled each second.
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 $\geq 5\text{mm/s}$
Low/medium/high speed swimming	Keatinge et al., 2015	Adult	Low speed: $<5\text{cm/s}$ Medium speed: $5 < x < 7\text{cm/s}$ High speed: $>7\text{cm/s}$	Low speed: $0.5\text{cm} < x < 2\text{cm/s}$ Medium speed: $2 < x < 4\text{cm/s}$ High speed: $>4\text{cm/s}$
Mean swimming episode duration	Lambert et al., 2012	Larval	Speed of $\geq 3\text{mm/s}$ for swimming episode	Speed of $\geq 5\text{mm/s}$ for swimming episode
Tail bend amplitude	Budick and O'Malley, 2000	Larval	Angle between rostral and caudal tangents subtracted from $180^\circ$	Sum of angles along zebrafish trace at maximum tail bend (low/medium/high speed).
Tail beat frequency	Budick and O'Malley, 2000	Larval	Frequency of tail beat cycles in a second	Frequency of maximum tail bends per second (low/medium/high speed).

**Table S5****Numbers used to evolve classifiers with features of movement**

The number of mutant and age matched control clips used to evolve classifiers with features of movement for each mutant line. An equal number of clips needed to be present in each class as accuracy was used as the measure of fitness. Each clip also had to have a value present for the features calculated (e.g. low/medium/high speed features).

Mutants classified	No. of mutant clips	No. of age matched control clips
<i>dj-1<sup>-/-</sup></i>	30	30
<i>pink1<sup>-/-</sup></i>	37	37
<i>dmd<sup>ta222a/+</sup></i>	25	25

**Table S6****Numbers used to evolve sliding window classifiers**

The number of mutant and age matched control clips used to evolve sliding window classifiers for each mutant line. The sliding window classifiers used the area under the curve as a measure of fitness which is insensitive to class imbalances. This allowed each class to have a different number of clips.

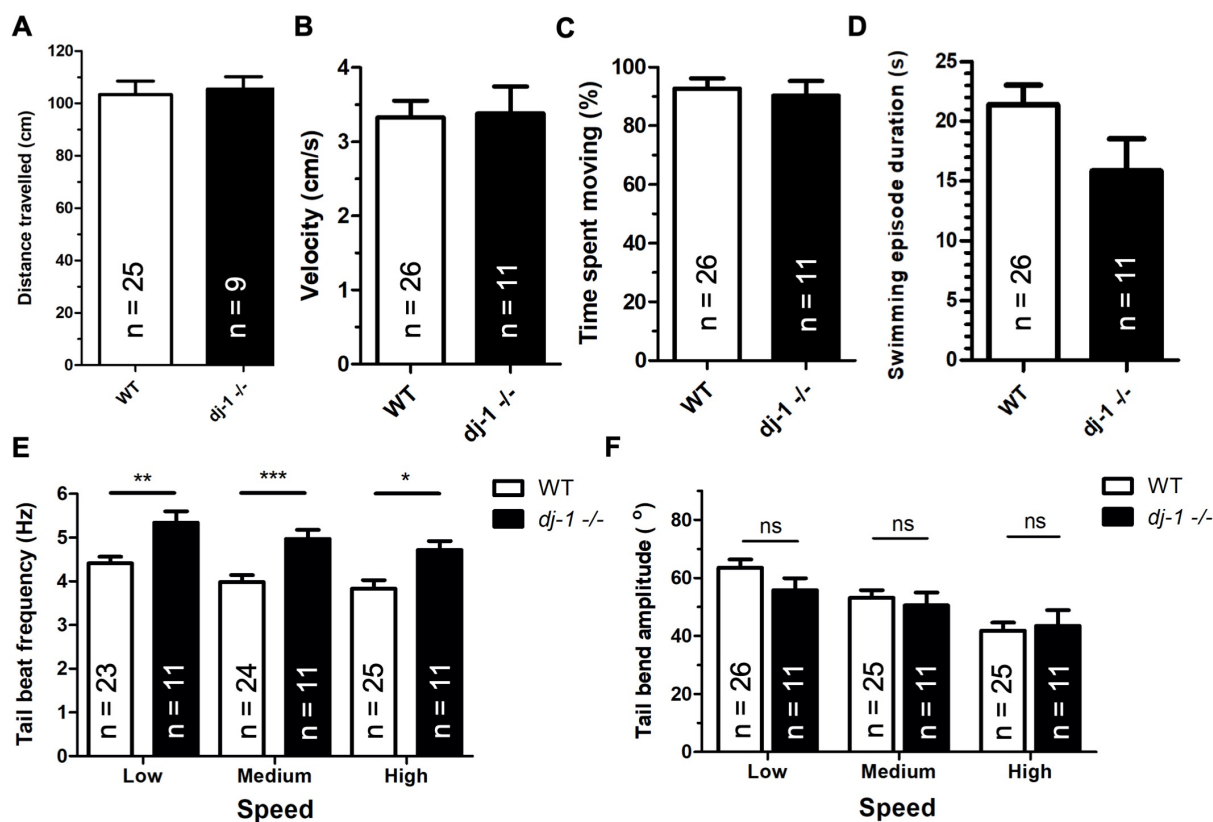
Mutants classified	No. of mutant clips	No. of age matched control clips
<i>dj-1<sup>-/-</sup></i>	36	64
<i>pink1<sup>-/-</sup></i>	39	44
<i>dmd<sup>ta222a/+</sup></i>	46	42



**Fig. S1**

**Analysis of extracted features of movement in *dj-1*<sup>-/-</sup> zebrafish at 8 wpf**

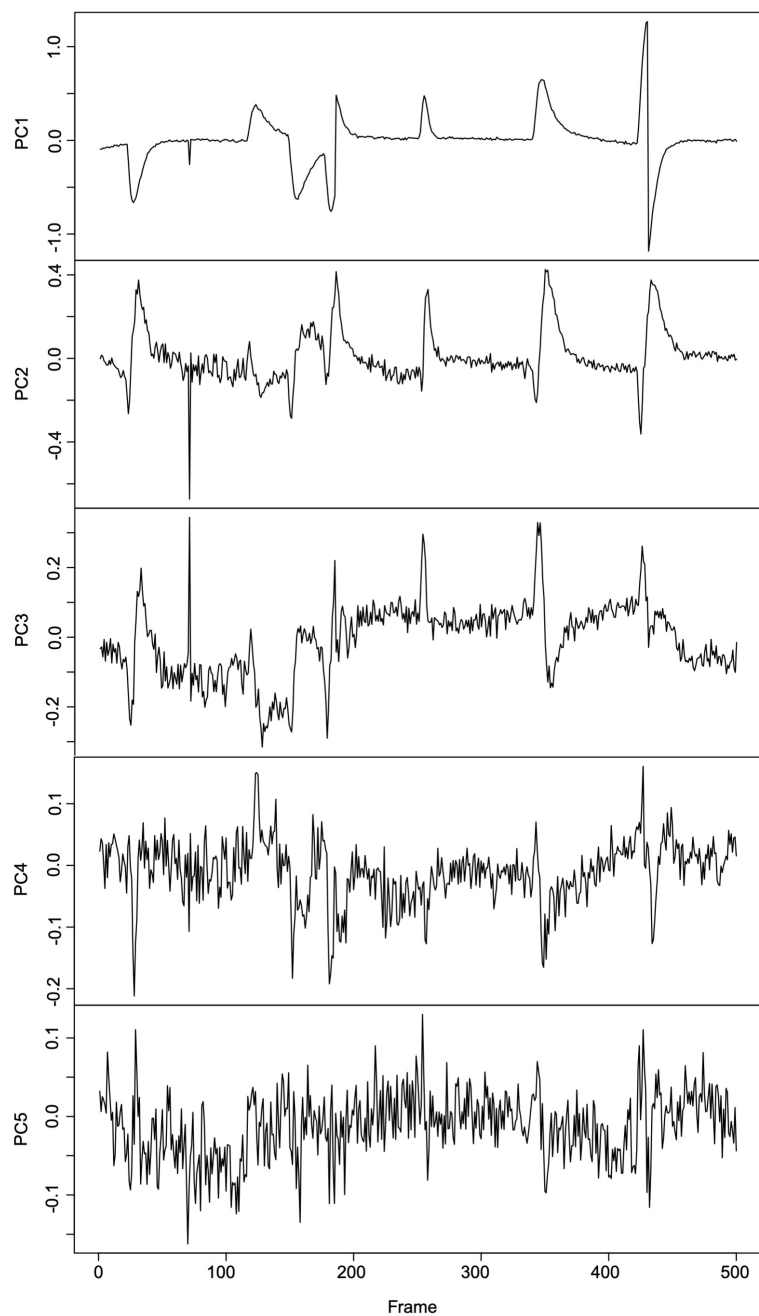
The features of movement compared between *dj-1*<sup>-/-</sup> and wild-type (WT) at 8 wpf including (A) distance travelled, (B) velocity, (C) percentage of time spent moving, (D) mean duration of a swimming episode, (E) tail beat frequency at low, medium and high swimming speeds, (F) tail bend amplitude at low, medium and high swimming speeds. \* = P < 0.05, \*\* = P < 0.01, \*\*\* = P < 0.001, ns = not significant.



**Fig. S2**

**Principle component analysis**

The 5 principal components (PCs), generated from linear combinations of the 5 angles along the zebrafish spine, plot against time. The first PC (PC1) explains the most variation in the data, followed by PC2 and with each subsequent PC explaining less of the variation. Together all of the PCs explain 100% of the variation within the tail bend angle data. PC1 captured only broad movements whilst PC2 retained some high frequency movements.



**Fig. S3****Scores of the *dmd*<sup>ta222a/+</sup> sliding window classifiers**

Training and test scores (AUCs) from the 20 sliding window classifiers evolved using the PC2 time series data from the 20 folds of data set.

Run	Training score	Test score
1	0.7318	0.3833
2	0.7227	0.3166
3	0.8636	0.45
4	0.8137	0.6334
5	0.8591	0.3834
6	0.8046	0.5334
7	0.7591	0.5334
8	0.8409	0.5833
9	0.8137	0.5167
10	0.7773	0.3167
11	0.7636	0.4166
12	0.7228	0.35
13	0.7409	0.4833
14	0.7727	0.25
15	0.7772	0.5166
16	0.7454	0.4166
17	0.7455	0.3167
18	0.7818	0.45
19	0.7682	0.3834
20	0.7182	0.3334
<b>Mean</b>	0.77614	0.42834