INVESTIGATING CELL TYPE SPECIFIC DEFECTS IN PROTEIN QUALITY CONTROL ACROSS THE GENETIC LANDSCAPE OF PARKINSON'S DISEASE

A Data Management Plan created using DMPonline.be

Creator: Eliana Nachman

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: 1282123N

ID: 197735

Start date: 30-09-2022

End date: 29-09-2025

Project abstract:

Parkinson's disease (PD) is a progressive neurodegenerative disease without cure. Today >20 genes have been linked to familial PD of which most are involved in various aspects of protein quality control. But it is unclear how these defects ultimately manifest in overlapping pathology. Intriguingly, several different PD fly models show similar transcriptome changes in defined single cell in the brains of young animals. The deregulated genes mainly affect protein homeostasis and neuronal dysfunction is rescued by increasing protein turnover.

I therefore speculate that different PD mutations converge on a common hub that impairs proteostasis, driving early neuronal dysfunction.

To investigate my hypothesis, I will source a unique Drosophila collection containing all known PD mutations. I will assess how PD mutations affect the proteostasis globally and in synapses. I will determine common vulnerability pathways by combining fly genetics and biochemical assays with single cell RNA sequencing. Finally, I will validate these central hubs in iPSC derived neurons from patients with the corresponding PD mutations.

I will thus identify the common hubs in cell type specific vulnerability pathways that are deregulated early in disease across several PD mutations.

My project will deliver 3 important results: 1. obtain innovative therapeutic targets to modify disease progression, 2. reveal how proteostasis maintains synaptic function and 3. answer why this process is affected early in PD.

Last modified: 19-07-2023

INVESTIGATING CELL TYPE SPECIFIC DEFECTS IN PROTEIN QUALITY CONTROL ACROSS THE GENETIC LANDSCAPE OF PARKINSON'S DISEASE
FWO DMP (Flemish Standard DMP)

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .cvs,.pdf, .txt, .rtf, .dwg, .gml, ... • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Data set 1	Drosophila melanogaster lines	Generate new data Reuse of existing data	Physical	Experimental	NA	NA	20 lines
Data set 2	Nucleic acid sequences	Generate new data Reuse existing data (see below)	Digital	Experimental	- Next generation sequencing raw data from fly brains: binary base call format (.bcl), .fastq(.gz) - Sanger sequencing for validation of fly lines and plasmids: Nucleotide sequences: raw sequence data trace (.ab1), text-based format (.fasta/.fa) -Text files: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Mark-up Language (.xml), Adobe Portable Document Format (.pdf), LaTex (.tex) format; -Quantitative tabular data: comma-separated value files (.csv), tab-delimited file (.tab), delimited text (.txt), MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb);	< 1 TB	NA
Data set 3	Algorithms and scripts	Generate new data, Reuse existing data (see below)	Digital	Other	NA	< 1 GB	NA
Data set 4	Digital images (microscopy images and videos, gel images)	Generate new data	Digital	Experimental	-Digital images in raster formats: uncompressed TIFF (.tif/.tiff), JPEG (.jpg), JPEG 2000 (.jp2), Adobe Portable Document Format (.pdf), bitmap (.bmp), .gif; - Digital video data: MPEG-4 High Profile (.mp4), motion JPEG 2000 (.mjp2), Audio Video Interleave (.avi);	< 20 TB	
Data set 5	Vectors for fly strain generation	Generate new data	Physical	Experimental	NA	NA	1 Box
Data set 6	iPSC lines edited with CRISP/Cas9	Generate new data Reuse of existing data	Physical	Experimental	NA	NA	1 Box
Data set 7	Plate reader measurements	Generate new data	Digital	Experimental	-Quantitative tabular data: comma-separated value files (.csv), tab-delimited file (.tab), delimited text (.txt), MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb);	< 1 GB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

Reuse of existing single cell RNA sequencing data from Drosophila brains published in Pech et al. (2023), BioRix, https://doi.org/10.1101/2023.03.11.532176

Reuse of synuclein biosensor HEK cell lines obtained from Marc Diamond (Holmes et al., 2014, PNAS,https://doi.org/10.1073/pnas.1411649111) under an MTA.
Reuse of fly and cell lines already available in lab.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes, human subject data

For the work with human iPSCs we refer to ethical committee approval:
- for work on human biological material/human data: Ethics Committee Research UZ / KU Leuven S62339.
Human biomaterial will be obtained following the three ethical principles (voluntary donation, informed consent and protection of privacy). This material will be used following our Center's Standard Operating Procedure for the handling of human biomaterial, and in accordance with European and national regulations and guidelines.
Environmental permits and bio-safety requirements to generate and propagate genetically modified Drosophila are met and authorized by the Flemish authorities (environmental permit number: 20171017-0004 and biosafety approval number: AMV/12062020/SBB219.2020/0433).

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes

We will collect information from the donors (including PD patients and controls) of the cells. The following information will be used/processed: age at disease onset (in the case of patients), age at donation, Parkinson's disease (PD) status, PD causative mutation status, PD polygenic risk score (genetic data), sex, and ethnicity.
For personal and sensitive data, we will abide by the Belgian law on the protection of individuals with regard to the processing of personal data (30th July 2018) and the General Data Protection Regulation 2016/679. The Privacy Team of KU Leuven has been notified before the start of the data processing activities (upon registering the study with the Clinical Trial Center and requesting ethical approval - see below) and the Data Steward therefore:
- has designated the categories of persons who have access to the sensitive data, with a precise description of their capacity in relation to the processing of these data; - keeps the list of the designated categories of persons at the disposal of the competent supervisory authority (Data Protection Authority);
- has ensured that the designated persons are obliged by a legal or statutory obligation, or by an equivalent contractual provision, to observe the confidential nature of the data concerned.
The reference of the file in the KU Leuven privacy register is: G-2021-3547.

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- Yes

We do not exclude that the proposed work could result in research data with potential for tech transfer and valorization. Ownership of the data generated belongs to KU Leuven and VIB in accordance with the framework agreement of both institutes. VIB has a policy to actively monitor research data for such potential. If there is substantial potential, the invention will be thoroughly assessed, and in a number of cases the invention will be IP protected (mostly patent protection or copyright protection). As such the IP protection does not withhold the research data from being made public. In the case a decision is taken to file a patent application it will be planned so that publications need not be delayed.
The use of Drosophila melanogaster lines will be subjected to the terms described in their respective MTAs.
Potential biomarkers/-targets related to neurodegenerative diseases will be claimed if this opportunity arises.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

Data will be generated following standardized protocols. Metadata will be documented by the research and technical staff at the time of data collection and analysis, by taking careful notes in the electronic laboratory notebook (E-notebook) that refer to specific datasets.
Cryotubes of biological samples stored at -80°C will be labelled with a reference number that links to an entry in or strain database.
Drosophila lines will be stored in a dedicated room and managed using a specific database for storage of the corresponding information (including genotype, origin, number of vials and date of transfer, crossing schemes) and vial tracking via unique QR codes.
All datasets will be accompanied by a README.txt file containing all the associated metadata (see more details below).
The data will be generated following standardized protocols. Clear and detailed descriptions of these protocols will be stored in our lab protocol database, and published along with the results.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- No

Metadata will include the following elements:
• Title: free text
• Creator: Last name, first name, organization
• Date and time reference
• Subject: Choice of keywords and classifications
• Description: Text explaining the content of the data set and other contextual information needed for the correct interpretation of the data, the software(s) (including version number) used to produce and to read the data, the purpose of the experiment, etc.
• Format: Details of the file format,
• Resource Type: data set, image, audio, etc.
• Identifier: DOI (when applicable)
• Access rights: closed access, embargoed access, restricted access, open access. Additionally, we will closely monitor MIBBI (Minimum Information for Biological and Biomedical Investigations) for metadata standards more specific to our data type.
For specific datasets, additional metadata will be associated with the data file as appropriate. Give details as needed for the project.
Specific examples (adjust as required):
- SOPs for biological data generation are kept on a dedicated KU Leuven shared drive. A central excel file is stored on that same drive, detailing for examples: (1) sample ID; (2) SOP with which data generation was performed; (3) abnormalities or deviations from SOP in data generation; (4) experimental QC values (e.g. DNA concentrations); (5) location of the source sample in the freezer.
- For bioinformatics processing, a data analysis log will be kept that details: (1) sequencing run ID; (2) the bioinformatics SOPs/scripts that were applied; (3) location of source files; (4) abnormalities or deviations.
The final dataset will be accompanied by this information under the form of a README.txt document. This file will be located in the top level directory of the dataset and will also list the contents of the other files and outline the file-naming convention used. This will allow the data to be understood by other members of the laboratory and add contextual value to the dataset for future reuse.

3. Data storage & back-up during the research project

Where will the data be stored?

Digital files will be stored on KU Leuven servers, except for private data that will be stored on KU Leuven secure server (digital vault).
- Tissue samples: Tissues will be stored locally in the laboratory. All human tissue samples will be registered with a Belgian biobank, in compliance with the Belgian law on human body material (dd 19-12-2008).
- Omics data (RNA single cell seq data): omics data generated during the project will either be stored on KU Leuven servers or on The Flemish Supercomputer Centre (VSC), initially in the staging area and later in the archive area.

- Vectors: As a general rule at least two independently obtained clones will be preserved for each vector, both under the form of purified DNA (in -20°C freezer) and as a bacteria glycerol stock (-80°C). All published vectors and the associated sequences will be sent to the non-profit plasmid repository Addgene, which will take care of vector storage and shipping upon request.

- Cell lines: Newly created human cell lines will be stored locally in the laboratory in liquid nitrogen storage and will be deposited in the UZ Leuven-KU Leuven Biobank. Other human cell lines will be stored locally in liquid nitrogen cryostorage of the laboratory when actively used for experiments.

- Genetically modified organisms: Drosophila lines will be stored in lab of Patrik Verstreken Drosophila lines will be stored in a dedicated room and managed using a specific database for storage of the corresponding information (including genotype, origin, number of vials and date of transfer, crossing schemes) and vial tracking via unique QR codes.

- Other biological and chemical samples: storage at 4°C and/or as frozen samples in cryovials as appropriate.

- Algorithms, scripts and softwares: All the relevant algorithms, scripts and software code driving the project will be stored in a private online git repository from the GitHub account of the department (https://github.com/vibcbd).

- Nucleic acid sequences: All nucleic acid sequences generated during the project will be stored on KU Leuven servers. Upon publication, all sequences supporting a manuscript will be made publicly available via repositories such as the GenBank database or the European Nucleotide Archive (nucleotide sequences from primers / new genes / new genomes), NCBI Gene Expression Omnibus (RNA-seq data).

How will the data be backed up?

KU Leuven drives are backed-up according to the following scheme:
- data stored on the “L-drive” is backed up daily using snapshot technology, where all incremental changes in respect of the previous version are kept online; the last 14 backups are kept.
- data stored on the “J-drive” is backed up hourly, daily (every day at midnight) and weekly (at midnight between Saturday and Sunday); in each case the last 6 backups are kept.
- data stored on the digital vault is backed up using snapshot technology, where all incremental changes in respect of the previous version are kept online. As standard, 10% of the requested storage is reserved for backups using the following backup regime: an hourly backup (at 8 a.m., 12 p.m., 4 p.m. and 8 p.m.), the last 6 of which are kept; a daily backup (every day) at midnight, the last 6 of which are kept; and a weekly backup (every week) at midnight between Saturday and Sunday, the last 2 of which are kept.
- All omics data stored on the Flemish Supercomputer Centre (VSC) will be transferred on a weekly basis to the archive area which is backed up.
Incremental backups are done daily from one 20 TB QNAP NAS to a second 20 TB QNAP NAS.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

There is sufficient storage and back-up capacity on all KU Leuven servers:

- the “L-drive” is an easily scalable system, built from General Parallel File System (GPFS) cluster with NetApp eseries storage systems, and a CTDB samba cluster in the front-end.
- the “J-drive” is based on a cluster of NetApp FAS8040 controlers with an Ontap 9.1P9 operating system.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

Both the “L-drive” and “J-drive” servers are accessible only by laboratory members, and are mirrored in the second ICTS datacenter for business continuity and disaster recovery so that a copy of the data can be recovered within an hour.
Access to the digital vault is possible only through using a KU Leuven user-id and password, and user rights only grant access to the data in their own vault. Sensitive data transfer will be performed according to the best practices for “Copying data to the secure environment” defined by KU Leuven. The operating system of the vault is maintained on a monthly basis, including the application of upgrades and security patches. The server in the vault is managed by ICTS, and only ICTS personnel (bound by the ICT code of conduct for staff) have administrator/root rights. A security service monitors the technical installations continuously, even outside working hours.
All private data will be rendered anonymous before processing outside the digital vault. Only the PI will be granted access to the server to deposit private data. The PI will be the only responsible for linking patient information, survey data and/or tissue samples, and will strictly respect confidentiality. All de-identified data will be exported from the database by the PI, and stored on KU Leuven servers from where it can be accessed by the research and technical staff from the laboratory.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

The total estimated cost of data storage during the project is 15.000 €. This estimation is based on the following costs:
-The costs of digital data (ca. 22 TB per year) storage are as follows: 173,78€/TB/Year for the “L-drive” and 519€/TB/Year for the “J-drive”.
-Maintaining a Drosophila line alive costs about 4 euro per year including the fly food, vials and personnel’s cost. The PI and the lab manager if applicable are responsible for the preservation of fly lines as well as administrative and experimental data. All published lines will be preserved for the remainder of the PI’s research career. All unpublished lines will be preserved for a minimum of 5 years after the end of the project.
Electricity costs for the -80° freezers present in the labs are included in general lab costs.
Data storage and backup costs are included in general lab costs.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

The minimum preservation term of 5 years after the end of the project will be applied to all datasets. All datasets will be stored on the university's central servers with automatic back-up procedures for at least 5 years, conform the KU Leuven RDM policy. The costs (€156 per TB per year for “Large volume-storage”) will be covered by the PI.

Where will these data be archived (stored and curated for the long-term)?

As a general rule, datasets will be made openly accessible, whenever possible via existing platforms that support FAIR data sharing (www.fairsharing.org), at the latest at the time of publication.
For all other datasets, long term storage will be ensured as follows:
-Digital datasets: files will be stored on the “L-drive”.
-Tissue samples (samples from Drosophila tissues and cell lines): Tissues will be stored locally in the laboratory.
-Omics data (RNA sequencing data): datasets will be stored on the “L-drive” or, for larger datasets, on the Vlaams Supercomputer Centrum.
-Vectors: As a general rule at least two independently obtained clones will be preserved for each vector, both under the form of purified DNA (in -20°C freezer) and as a bacteria glycerol stock (-80°C).
-Cell lines: human cell lines will be stored in the UZ Leuven Biobank (-80°C). Human pluripotent stem cell lines generated during this project will be deposited in hPSCreg.
-Genetically modified organisms: Drosophila lines will be housed locally.
-Other biological and chemical samples: storage at 4°C and/or as frozen samples in cryovials as appropriate.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

The total estimated cost of data storage during 5 years after the end of the project is 20.000 €. This estimation is based on the following costs:
-The costs of digital data storage are as follows: 173,78€/TB/Year for the “L-drive” and 519€/TB/Year for the “J-drive”.
-Maintaining a Drosophila line alive costs about 4 euro per year including the fly food, vials and personnel’s cost. The PI and the lab manager are responsible for the preservation of fly lines as well as administrative and experimental data. All published lines will be preserved for the remainder of the PI’s research career. All unpublished lines will be preserved for a minimum of 5 years after the end of the project.
Electricity costs for the -80° freezers present in the labs are included in general lab costs.
Data storage and backup costs are included in general lab costs.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in an Open Access repository
- Other, please specify:

Data sets will be published in peer-reviewed open-access journals.

If access is restricted, please specify who will be able to access the data and under what conditions.

Upon publication generated data, such as fly lines, cell lines, and vectors will also available upon request by e-mail to the PI.
Highly-requested materials will be made available to the community after publication via existing platforms (see below).

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- No

Where will the data be made available? If already known, please provide a repository per dataset or data type.

As a general rule, datasets will be made openly accessible via existing platforms that support FAIR data sharing (www.fairsharing.org). Sharing policies for specific research outputs are detailed below:

- Vectors:
Upon publication, all vectors supporting a manuscript will be made publicly available via the non-profit plasmid repository Addgene, along with the corresponding DNA sequences. Addgene in turns performs quality control on the DNA, curates the plasmids online with all relevant information (maps, sequences), and for a minimal cost (typically $65) ships the vectors upon simple request and signature of a material transfer agreement. The MTA will be prepared before depositing the vectors with the help of our organization's Tech Transfer office. For transfer between nonprofit or academic institutions, Addgene typically uses the Uniform Biological MTA. All non-published vectors and the associated documentation will be shared by the PI upon request and after signature of a MTA, at no cost except the cost of shipment.
- Genetically modified organisms:
All genetically modified organisms (Drosophila lines) used in publications will be made available to researchers upon request at the time of publication.
- Manuscripts:
All scientific publications will be shared openly. Manuscripts submitted for publication will be deposited in a pre-print server such as bioRxiv. At the time of publication, research results will be summarized on the PI's website (https://verstrekenlab.sites.vib.be/en) and post-print pdf versions of publications will be made available there if allowed by copyright agreements. Publications will also be automatically listed in our institutional repository, Lirias 2.0, based on the authors name and ORCID.
- Nucleic acid sequences:
Upon publication, all sequences supporting a manuscript will be made publicly available via repositories such as the GenBank database (nucleotide sequences from primers / new genes / new genomes), NCBI Gene Expression Omnibus (RNA-seq data).

- Data that do not support publication will be either deposited in an open access repository or made available upon request by email.

When will the data be made available?

The data will be made available upon publication of the research results.
As a general rule all research outputs will be made openly accessible at the latest at the time of publication.

Which data usage licenses are you going to provide? If none, please explain why.

Whenever possible, datasets and the appropriate metadata will be made publicly available through repositories that support FAIR data sharing.
As detailed above, metadata will contain sufficient information to support data interpretation and reuse, and will be conform to community norms. These repositories clearly describe their conditions of use (typically under a Creative Commons CC0 1.0 Universal (CC0 1.0) Public Domain Dedication, a Creative Commons Attribution (CC-BY) or an ODC Public Domain Dedication and Licence, with a material transfer agreement when applicable). Interested parties will thereby be allowed to access data directly, and they will give credit to the authors for the data used by citing the corresponding DOI.
For data shared directly by the PI, a material transfer agreement (and a non-disclosure agreement if applicable) will be concluded with the beneficiaries in order to clearly describe the types of reuse that are permitted.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

What are the expected costs for data sharing? How will these costs be covered?

It is the intention to minimize data management costs by implementing standard procedures e.g. for metadata collection and file storage and organization from the start of the project, and by using free-to-use data repositories and dissemination facilities whenever possible.
Data management costs will be covered by the laboratory budget. A budget for publication costs has been requested in this project.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

Metadata will be documented by the research and technical staff involved in the project at the time of data collection and analysis, by taking careful notes in the electronic laboratory notebook (E-notebook) that refer to specific datasets.

Who will manage data storage and backup during the research project?

The research and technical staff will ensure data storage and back up, with support from René Custers and Alexander Botzki for the electronic laboratory notebook (ELN) and from Raf De Coster for the KU Leuven drives.

Who will manage data preservation and sharing?

The PI is responsible for data preservation and sharing, with support from the research and technical staff involved in the project, from René Custers and Alexander Botzki for the electronic laboratory notebook (ELN) and from Raf De Coster for the KU Leuven drives.

Who will update and implement this DMP?

The PI and Eliana Nachman are ultimately responsible for all data management during and after data collection, including implementing and updating the DMP.