DMP title

Project Name My plan (FWO DMP) - DMP title

Project Identifier 11N2222N

Grant Title 11N2222N

Principal Investigator / Researcher Alessio Silva

Description This project aims to elucidate the role of early alterations in lipid metabolism during Schwann cell development in Charcot-Marie-Tooth type 1A (CMT1A), and aims to discover new therapeutic targets for CMT1A that could favor cholesterol redistribution inside the cell and promote proper Schwann cell differentiation.

Institution KU Leuven

1. General Information

Name applicant

Alessio Silva

FWO Project Number & Title

11N2222N

Investigating the pathological mechanisms underlying the demyelinating form of Charcot-Marie-Tooth disease using patient-derived induced pluripotent stem cells

Affiliation

- KU Leuven

2. Data description

Will you generate/collect new data and/or make use of existing data?

- Generate new data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

Type of Data	Format	Volume	How Created
Analysis of confocal images	.xls, .pzfx	50-200MB	Quantification of confocal image data performed using ImageJ, in Microsoft Excel and GraphPad Prism 9.
Analysis of western blot images	.xls, .pzfx	50-200MB	Numerical data. Quantification of western blot images, performed using ImageJ, in Microsoft Excel and GraphPad Prism 9.

Confocal images	.LIF , .ometiff, .tif	100GB-500GB	Confocal microscopy images of stained iPSCs, iPSC-derived Schwann cells, iPSC-derived motoneurons precursors taken using a Leica SP8 confocal microscope
FACs sort reports	Numerical and multimedia (Flowjo files, .TIF, PDF)	200MB-1GB	Acquisition on flow cytometers of flow cytometry core (VIB- KU Leuven) by myself or operator from the core
Figures of data for publication	.ai, .ppt	10-100GB	Figures of amalgamated data produced during the project, created using adobe illustrator or Microsoft PowerPoint.
Lab books	Electronic lab notebook and paper notes	5 lab books + 2GB	Dated written notes associated with carrying out experimental procedures
Lipidomics and transcriptomics data	Numerical (.xls)	10GB	Lipidomics and transcriptomics analysis will be performed by the VIB Lipidomics and Genomics Cores.
Microscopy slides	Glass microscopy slides	250-1000	Microscopy slides used during imaging, consisting of cells immunostained or dyed using chemicals.
Plotted graphs	.pzfx	200-500MB	Graphs of data from .xls files produced using GraphPad Prism 9.

Protein samples	tubes of liquid containing protein	200-1000	Samples of denatured or undenatured proteins stored at -20°C or -80°C extracted from tissue or cells using detergents.
qRT-PCR data analysed files and statistical analysis	.xls, .pzfx	100-500MB	Analysis of qRT-PCR data performed using Quant Studio software (Thermofisher), statistical analysis performed in GraphPad Prism 9.
qRT-PCR data raw files	.eds, .xls	100-500MB	Raw qRT-PCR data collected using Quant studio 3 thermocycler and associated quant studio software (Thermo Fisher)
Risk assessments (RAs)	.docx, .pdf	50-200MB	Written risk assessments associated with standard operating procedures for experimental procedures performed within the lab
RNA samples	tubes of liquid containing RNA	10-100	RNA samples extracted from tissue or cells
Standard operating procedures (SOPs)	.docx, .pdf	50-200MB	Written protocols for experimental procedures performed in the lab
Text manuscript for publication	.docx	1-20GB	Text files associated with submitted publications
Western blot images	.gel, .ometiff, .tiff	750-1000MB	Tif images of chemiluminiscent signal taken using ImageQuant LAS 4000 instrument.
Cell imaging data	Multimedia (.tif) + numerical (.xsl)	200-300GB	Imaging on Operetta CLS High-Content Analysis System and analysis performed by me.

3. Legal and ethical issues

Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- No

Privacy Registry Reference:

Short description of the kind of personal data that will be used:

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, add the reference to the formal approval by the relevant ethical review committee(s)

- Yes

Patient-derived cell lines: S50354 (Ethical Committee of the University Hospital Leuven)

Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

- No

Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- No

4. Documentation and metadata

What documentation will be provided to enable reuse of the data collected/generated in this project?

Data will be generated following standardized protocols. Metadata will be documented by me at the time of data collection and analysis, by taking careful notes in the electronic laboratory notebook (E-notebook) and/or in hard copy lab notebooks that refer to specific datasets.

Biological samples stored at -80°C will be labelled with all information, referring to the E-notebook where extra documentation is noted.

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? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/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. More specifically:
- 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.
- For microscopy a file with the following information will be stored (1) microscope used; (2) channel (lasers+ filters); (3) objective; (4) date of imaging; (5) laserpower

and aquisition time; (6) resolution; (7) microscope settings.
- For qPCR data a file with the following information will be stored (1) aquisition method; (2) machine used; (3) date and time of analysis.
- For western blot data a file with the following information will be stored (1) imaging method (chemiluminscence/ fluorescence/bright field); (2) machine used; (3) aquisition time; (4) aquisition date
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.

5. Data storage and backup during the FWO 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.
- Omics 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.
- Cell lines: Human cell lines will be stored locally in liquid nitrogen cryostorage of the laboratory when actively used for experiments. Animal cell lines will be stored in liquid nitrogen cryostorage of the laboratory.
- Genetically modified organisms: Mice will be maintained in facilities of the Laboratory Animal Center of KU Leuven, which applies Standard Operation Procedures concerning housing, feeding, health monitoring to assure consistent care in accordance with European and national regulations and guidelines. All animals will be registered in the Leuven Animal Information System (LAIS) database, along with corresponding genotyping information, ethical approval documents and animal provider receipts. 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).

How is backup of the data provided?

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.
- For all cell lines used : A backup of all lines will be stored in -80°C freezers at the center for brain and disease (VIB-KU Leuven)

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.

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

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”.
The total estimated cost of data storage during the project is therefore 3000€. This cost will be covered by the lab.

Data security: 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.

6. Data preservation after the FWO project

Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).

All data will be retained for the expected 5 year period after the end of the project.

Where will the data be archived (= stored for the longer term)?

The data will be stored on the university's central servers (with automatic back-up procedures) for at least 10 years, conform the KU Leuven RDM policy.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

The costs of digital data storage 18000€ are (based on the costs in section 5). These costs are covered by the lab.

7. Data sharing and reuse

Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

- No

Which data will be made available after the end of the project?

Participants to the present project are committed to publish research results to communicate them to peers and to a wide audience. All research outputs supporting publications will be made openly accessible. Depending on their nature, some data may be made available prior to publication, either on an individual basis to interested researchers and/or potential new collaborators, or publicly via repositories (e.g. negative data).

Where/how will the data be made available for reuse?

- In an Open Access repository
- Upon request by mail

When will the data 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. No embargo will be foreseen unless imposed e.g. by pending publications or ongoing projects requiring confidential data. In those cases, datasets will be made publicly available as soon as the embargo date is reached.

Who will be able to access the data and under what conditions?

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 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.

What are the expected costs for data sharing? How will the 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.

8. Responsibilities

Who will be responsible for data documentation & metadata?

Metadata will be documented by the research (mainly me) 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.

Who will be responsible for data storage & back up during the 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 and from Raf De Coster for the KU Leuven drives.

Who will be responsible for ensuring data preservation and reuse ?

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 and from Raf De Coster for the KU Leuven drives.

Who bears the end responsibility for updating & implementing this DMP?

Alessio Silva (researcher)
Nicole Hersmus (Lab manager of lab of Neurobiology)
Ludo Van Den Bosch (PI)