Understanding the role of cilia dysfunction in amyotrophic lateral sclerosis.
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
Imaging data	Confocal images; Time-lapse movies; Obtained from ICC or labeling with fluorescent cell-permeable dyes; Images from western blots	Generate new data	Digital	Experimental	.tiff, .jpg, .png, .czi, .pdf, .pptx	<5TB	
Omics data	Proteomics of primary cilia using proximity labeling	Generate new data	Digital	Other	seq	<1TB	
Samples	Frozen cell pellets, fixed samples from MNs, RNA samples, DNA samples, protein cell lysates, ...	Generate new data	Physical	NA	NA	NA	
Viral vectors	Viral vectors and plasmids	Generate new data + Reuse existing data	Physical	NA	NA	NA	
Lab notebooks, text manuscripts for publication, SOPs and RA for procedures	Electronic lab notebook and paper notes, text manuscript and associated images, standard operating procedures files and risk assesment	Generate new data + Reuse existing data	Digital and Physical	Compiled/aggregated data and other	.docx, .pdf and other	<1GB	

qRT-PCR data	Both raw qRT-PCR data collected using Quant studio 3 thermocycler and associated quant studio software (Thermo Fisher) and analysed data performed using Quant Studio software (Thermo fisher), as well as statistical analysis performed in graph pad prism.	Generate new data	Digital	Experimental	.xls, .pzfx, .eds	<1GB	

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:

We will reuse data generated previously in https://doi.org/10.1093/brain/awae331

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

Regarding the work related to induced pluripotent stem cellls (iPSCs) ethical approval was granted by the Ethische Commissie Onderzoek UZ/KU Leuven.
(S50354)
If post-mortem material will be used, ethical approval was granted to Prof. Dr. Dietmar thal who will be responsible for the samples.

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.

- No

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.

- No

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.

- Yes

Yes, resource sharing with external partners was always accompanied by the generation of a material transfer agreement (MTA) between the labs and clearly mentions the restrains.
To our knowledge, the only restraint concerns the isogenic iPSC lines that may not by used for drug screening applications.

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).

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. All datasets will be accompanied by a README.txt file containing all the associated metadata (see more details below). These metadata include the experimental design and protocols used, data analysis and relevant scripts, raw data, analyzed data, location of raw digital and physical data

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.

- Yes

While specific data types might require particular metadata, as a general rule the metadata will be based on a generalized metadata schema according to the Dublin Core, including 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.
For specific datasets, additional metadata will be associated with the data file as appropriate such as experimental procedures to generate omics data.
The final dataset will be accompanied by this information under the form of a README.txt or .docx 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 (including the digital ELN, the shared network drive (J-drive) and the large volume storage drive (L-drive)), with hourly on-site backup and mirroring. 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 be stored on KU Leuven servers
- Cell lines: Human cell lines will be stored locally in liquid nitrogen cryostorage of the laboratory when actively used for experiments.
- 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 will the data be backed up?

Standard back-up provided by KU Leuven ICTS for my storage solution

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.

Regarding the cell lines and plasmids, a back up vial is always stored at another location (ON4 instead of ON5)

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

Data storage is arranged at a departmental level and can be increased if needed.
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.
- Onedrive provides 2TB of storage data which should be sufficient for (most of) the data generated in the project.

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

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.

What are the expected costs for data storage and backup during the research 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 519EUR/TB/Year for the “J-drive”.)

These costs will be covered by the FWO grant.

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...).

All data will be preserved for 10 years according to KU Leuven RDM policy

We expect that we will make the data publicly available on data repositories upon publication of the manuscripts.

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

Digital data will be stored at the Archive (K:) server from KU Leuven ICTS.
Physical samples will be stored in the freezers from the Research Group of Neurobiology.
Proteomics data: will be deposited in public repositories.
If appropriate, 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.

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

We expect the costs to gradually increase up to 3000 euro/year. After the project, data preservation costs will be covered by the lab and other grants.

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

The data will be made available after publication via the required link in the publications or upon request after an embargo period after publication (f.i. phenotype files, genetic data). The same holds true for unpublished data, they can be made available upon request but only after an embargo period (3 years; exceptionally 5 years after the project).

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

All team members have access as long as they are affiliated to KU Leuven. Once all files are released, anyone can use these data to generate new results, referring to the original publication and not for commercial use. Other data will be only released upon request and after an embargo period after publication. Data will be released under a CC-BY 4.0 reuse license.

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.

If appropriate, 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.
Proteomics data: will be deposited in public repositories such as the PRIDE archive.

When will the data be made available?

Upon publication of research results.

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

CC-BY 4.0 (data).

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 depend on the data repository selected and will be covered by the laboratory budget and project funding.

6. Responsibilities

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

With respect to E-Notebook: René Custers and Alexander Botzki With respect to other datatypes: The PI (Philip Van damme) and day-to-day managers of the FWO-project; currently: Nicole hersmus and Jimmy Beckers.

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

With respect to E-Notebook: René Custers and Alexander Botzki With respect to other datatypes: The PI (Philip Van damme) and day-to-day managers of the FWO-project; currently: Nicole hersmus and Jimmy Beckers.

Who will manage data preservation and sharing?

With respect to E-Notebook: René Custers and Alexander Botzki With respect to other datatypes: The PI (Philip Van damme) and day-to-day managers of the FWO-project; currently: Nicole hersmus and Jimmy Beckers.

Who will update and implement this DMP?

The PI (Philip Van damme) and day-to-day managers of the FWO-project; currently: Nicole hersmus and Jimmy Beckers.