DMP - KU Leuven Internal Funds - IDN/22/012

1. General Information	
Name PI	Ludo Van Den Bosch - ludo.vandenbosch@vib.be
Project Number & Title	IDN/22/012 Walking the lines of engineering, biology and medicine: Creating 3D neuromuscular junctions to evaluate novel therapies for neurodegenerative diseases

2. Data description	
Will you generate/collect new data and/or make use of existing data?	New data, Existing data

Describe the origin, type and format of the data (per dataset) and its (estimated) volume	Experimental data Digital images Imaging motor neurons, muscle cells and myelination: uncompressed TIFF (.tif/.tiff), JPEG (.jpg), JPEG 2000 (.jp2), Purpose is to visualise the differentiation of the different cell types. We are collaborating with our microscopy core facility (LiMONE, VIB) and are currently using NIS software (Nikon) to perform these analyses. Estimated size 1 gigaBite. Video and audio files In WP1, calcium measurements will be measured over time in the myotubes (using Fluo-4 as a calcium indicator) after depolarizing the motor neurons present in the other compartment in order to test whether the NMJs formed are functional. Inhibition using tubocurarine, an acetyl choline receptor (AChR) antagonist, will be used to confirm that the calcium transients were the result of the stimulation of the myotubes at the NMJs. File format MPEG-4 High Profile (.mp4) and excel (.xls). Estimate size: 500 megaBites. Cytometry data We are optimising new protocols to optain improved Schwann cell differentiation. One step in this protocol is the purification of Schwann cells: Flow Cytometry Standard (.fcs) and excel (.xls). Estimated size:100 megaBites Electrophysiology To test the functionality of the 3D muscle fiber bundle (WP3), we will induce muscle contraction by electrical and chemical stimulation. Electrical stimulation is performed with platinum electrodes positioned on both sides of the muscle compartment. Chemical stimulation is induced by adding acetylcholine to the culture medium. Measurements of muscle contraction will be in excel (.xls) format: Estimated size: 500 megaBItes Omics data

	The effect of ALS-causing FUS mutations will be studied on the formation of the neuromuscular junctions (WP4). We will perform an RNAseq analysis to investigate which mRNAs and pathways are affected.Transcriptomic RNAseq experiments: binary base call format (.bcl) or fastq (.gz) and excel (.xls). Estimated size: one gigaByte Cell lines Commercially obtained control iPSC lines as well as previously established iPSC-lines from patients with FUS myutations: stored in -80°C . In addition to human satellite cell-derived myoblasts and mesoangioblasts to differentiate myotubes (WP1). Also these cells will be stored in -80°C. Antibodies Use of commercially available antibodies to stain the different cell types as well as the neuromuscular junctions (all WPs). Using imunocytochemistry, we will compare the relative number of NMJs formed on the myotubes starting from mesoangioblasts in comparison to myoblasts. Starting from the z-stacks using a confocal microscope (DMI8 Leica), we will visualize the myotubes (stained with MF20; red; Alexa555), motor end plates (stained with bungarotoxin, far red; Alexa 647) and motor neuron axons (stained with SMI-32 and alpha-synuclein; green; Alexa 488). Subsequently, the myotube surface will be determined and all values will be expressed relative to the surface size of each myotube. Next, the size of the motor end-plate and of the axon signal overlapping with the individual myotubes will be determined. The relative overlap of the axonal signal with the motor end plate signal will be compared to the overlap of the rest of the axonal signal with the myotube, to compensate for potential differences in number of axons.

	Simulation data Derived and compiled data Research documentation Regular reports will be written in MS Word (.doc/.docx) format. In addition, tables will be generated using MS Excel (.xls/.xlsx). Laboratory notes will be generated in MS Word (.doc/.docx). Protocols will be written in MS Word (.doc/.docx) and saved as Adobe Portable Document Format (.pdf). Digital images will be generated shared in uncompressed TIFF (.tif/.tiff), JPEG (.jpg) or JPEG 2000 (.jp2), Estimated size 100 megaBites. Manuscripts Manuscripts will be written in MS Word (.doc/.docx) format and submitted in Adobe Portable Document Format (.pdf) format. Estimated size 10 megaBites. Canonical data These datasets represent an important source of information for the laboratory of the PI (including future staff), for scientists, journalists and higher education teachers working in the field of neurodegeneration, but also for non-profit organizations and industries active in the field of neurodegeneration.

3. Ethical and legal issues	
Will you use personal data? If so, shortly describe the kind of personal data you will use AND add the reference to your file in your host institution's privacy register.	No
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 The ethical approval to use human derived cells in the context of our research was already obtained before the start of this project. The use of patient fibroblasts for the generation of hiPSCs was approved by the ethics committee of University Hospital Leuven (n° S50354 and S63792), while the use of myoblasts was approved by the ethical commission (n° NH019-2020–04-02).
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?	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 iPSCs, differentiation protocols and 3D printing protocols will be subjected to the terms described in their respective MTAs.

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 understanding and reuse of the data collected/generated in this project?	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) and/or in hard copy lab notebooks that refer to specific datasets. Cryotubes of biological samples (bacterial and yeast strains) stored at -80°C will be labelled with a reference number that links to an entry in or strain database. 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? If so, describe in detail which standard will be used. If not, state in detail which metadata will be created to make the data easy/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 such as Dublin Core or DataCite. We will closely monitor MIBBI (Minimum Information for Biological and Biomedical Investigations) for metadata standards that are more specific to our data. 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.

5. Data storage & backup during the 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). - Cell samples: Cell samples will be stored locally in the laboratory. All human cells will be registered with a Belgian biobank, in compliance with the Belgian law on human body material (dd 19-12-2008). - 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 iPSCs will be stored locally in liquid nitrogen cryostorage of the laboratory when actively used for experiments.

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.

What are the expected costs for data storage and backup during the project? How will these costs be covered?	The total estimated cost of data storage during the project is 200 €. 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”. - 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.
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. Only 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. Databases are encrypted, password protected and within KU Leuven firewalls.

6. Data preservation after the end of the project KU Leuven expects that data generated during the project are retained for a period of minimally 10 years after the end of the project, in as far as legal and contractual agreements allow.	
Which data will be retained for the expected 10 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, ...).	The minimum preservation term of 10 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 10 years, conform the KU Leuven RDM policy. The costs (€156 per TB per year for “Large volume-storage”) will be covered by the budget of the lab.
Where will these data be archived (= stored 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”. -Cell samples: Homogenised or fixed cell samples will be stored locally in the laboratory. -Omics data: datasets will be stored on the “L-drive” or, for larger datasets, on the Vlaams Supercomputer Centrum. - Cell lines: human cell lines will be stored in the UZ Leuven Biobank (-80°C). - Other biological and chemical samples: storage at 4°C and/or as frozen samples in cryovials as appropriate. - Following publication, the results associated with each study will also be deposited in the Dryad repository, where they will be preserved indefinitely.

What are the expected costs for data preservation during these 10 years? How will the costs be covered?	The total estimated cost of data storage during 10 years after the end of the project is 2500€. 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”. - 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.

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)?	Yes Differentiation protocols and the protocols to print the 3D devices will be shared upon simple request following publication, unless we identify valuable IP. In this case, we will first protect commercial exploitation, either through patenting or via an MTA that restricts the material from commercial use. We aim at communicating our results in top journals that require full disclosure of all included data. iPSC lines will be shared upon simple request following publication, unless we identify valuable IP, in which case we will first protect commercial exploitation, either through patenting or via an MTA that restricts the material from commercial use
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). We aim at communicating our results in top journals that require full disclosure upon publication of all included data, either in the main text, in supplementary material or in a data repository if requested by the journal and following deposit advice given by the journal. Depending on the journal, accessibility restrictions may apply. Biological material will be distributed to other parties if requested
Where/how will the data be made available for reuse?	In a restricted access repository
When will the data be made available?	Upon publication of the research results

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, 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.
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 is available in the respective labs.

8. Responsibilities	
Who will be responsible for the data documentation & metadata?	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.
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 (ELN) and from Raf De Coster for the KU Leuven drives.
Who will be responsible for ensuring 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 bears the end responsibility for updating & implementing this DMP?	The PI is ultimately responsible for all data management during and after data collection, including implementing and updating the DMP.