DMP - KU Leuven Internal Funds - C14/22/132

1. General Information	
Name PI	Ludo Van Den Bosch - ludo.vandenbosch@vib.be
Project Number & Title	C14/22/132 Mechanisms of motor neuron degeneration in amyotrophic lateral sclerosis (ALS): cell death pathways and/or aberrant communication between cell body and neuromuscular junctions

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

Describe the origin, type and format of the data (per dataset) and its (estimated) volume	Observational data Tissue samples In order to determine the activated cell death processes that contribute to motor neuron loss in sporadic ALS, we will use precentral and cortex samples and spinal cord of 20 sporadic ALS patients and 20 control cases (WP2). Those cell death pathways that are identified in sporadic ALS cases will be analyzed for its upregulation in genetic forms of ALS driven by the C9orf72 and SOD1 mutations. Both fresh frozen as well as fixed postmortem material will be used collected by Dr. D. Thal in the UZ Gasthuisberg. Fresh frozen and fixed tissue will also be isolated from transgenic and control mice. Brain and spinal cord tissue will be isolated from FUS(WT), TDP-43(A315T) mice, and SOD1(G93A) mice. Last but not least, we will also isolate tissue (fresh frozen and fixed) from zebrafish embryos. Frozen samples will be stored at -20°C or -80°C and will be stored in the freezer of the different laboratories involved in this project. Digital images in uncompressed TIFF (.tif/.tiff), JPEG (.jpg) or JPEG 2000 (.jp2) format will be obtained after staining of slices obtained of these different tissues. Estimated size: 1 gigaByte

	Experimental data Digital images Postmortem tissue, tissue obtained from transgenic mice and zebrafish as well as fixed iPSC-derived cells will be stained with different antibodies and pictures will be taken using confocal microscopy. The format of the digital images will be uncompressed TIFF (.tif/.tiff), JPEG (.jpg) or JPEG 2000 (.jp2). Western blot images (WP1 & WP5) will be obtained and quantified using our Western blot documentation system. Files will be saved as uncompressed TIFF (.tif/.tiff) files Estimated size: 5 gigaBytes. Video and audio files We will systematically investigate the axonal transport of different cargos in differentiated neuronal cultures (obtained from iPSCs). The proportions of antero- and retrogradely moving and stationary organelles, as well as their transport velocity will be quantified using kymographs obtained after time-lapse imaging. This is done using ImageJ and the files contain RGB images (both TIFF and raw) in interleaved format. Image width is the number of pixel in each row of image data and Image Height is the number of rows in the image. Offset to First Image is the number of bytes in the file before the first byte of image data. Using similar methods, we will measure live axonal transport using in vivo time-lapse imaging of fluorophore-labeled cargo in zebrafish embryos. We will investigate to which types of defects the expression of the different mutant and control mRNAs lead (effect on speed, number of cargoes moving, directionality: reduction, pausing

	frequency and time). We have an assay available to follow axonal transport in live zebrafish neurons by the use of constructs expressing fusion proteins of known vesicle markers and organelle proteins. File format: MPEG-4 High Profile (.mp4), motion JPEG 2000 (.mjp2) and/or Audio Video Interleave (.avi). Estimated size: 1 gigaByte. Spectroscopy data Liquid Chromatography Coupled to Tandem Mass Spectrometry (LC-MS/MS) will be used The identity of the positive and negative factors incluencing the properties of the iPSC-derived models will be determined using mass spectrometry analysis of the medium (WP2). The file format will be mass spectral files (.msp) and converted to MS Excel (.xls/.xlsx) files. Omics data For proteomics (WP2), we will use brain homogenates from the precentral cortex, precentral white matter and the spinal cord to extract proteins from the same cases used for immunohistochemistry and single cell transcriptomics. To allow distinction between soluble, dispersible (endosomal, exosomal, synaptosomal proteins and insoluble proteins in mixture with the intra- or extracellular fluid), membrane-associated SDS-soluble proteins, and aggregated formic-acid (FA) soluble proteins, we will separate four fractions of the brain/ spinal cord lysates: soluble, dispersible, SDS, and FA fraction. Proteomic analysis will be performed by Liquid Chromatography Coupled to Tandem Mass Spectrometry (LC-MS/MS). The file format of these data will be mass spectral files (.msp) and converted to MS Excel (.xls/.xlsx) files. The collected proteomics data sets will be analyzed a) in a supervised approach comparing the four groups of sporadic ALS, C9orf72 ALS, SOD1 ALS and control cases and b) in an unsupervised approach correlating the neuropathological measures for cell death

	pathways, TDP-43, SOD1 and dipeptide repeat pathology by weighted correlation network analysis (WGCNA). RNA seq. In order to unravel the molecular mechanisms linking axonal dysfunction to cell death pathways, we will perform the axon versus soma RNA sequencing at different time points (WP 4). In addition, single cell RNAseq will be performed on fresh frozen human tissue (WP2). RNAseq will also performed on the astrocytes to determine the pathways (WP3) activated in the mutant astrocytes. The data files will be raw sequence data trace (.ab1), text-based format (.fasta/.fa) and accompanying QUAL file (.qual) files. Estimated file size: 5 gigaBytes Cell lines We already obtained 10 well characterized iPSC lines from Answer ALS (https://www.answerals.org/) as well as age- and sex-matched control iPSC lines and will investigate whether there is a difference in axonal transport in motor neurons differentiated from these iPSC lines. For KIF5A, the patient derived iPSC lines as well as the isogenic control were obtained from Dr. B. Traynor (NIH, Bathesda, USA). Our plan is to create iPSCs with mutations in TUBA4A using CRISPR/Cas9 technology (3.1.4) Genetically modified organisms Transgenic FUS(WT) mice, TDP-43(A315T) mice, and SOD1(G93A) mice will be used. Antibodies Pyroptosis will be visualized by the immunohistochemically detected exppression of cleaved gasdermin D (GSDMD-NT). Its activation via the canonical inflammasome pathway will be documented by the presence of caspase 1 and NLRP-3. The activation

	via the non-canonical athway will be identified by detecting caspase 4/5. In addition, IL-18 will be stained. The activation of the apoptotic pathway will be visualized by the immunohistochemical detection of activated caspase 3 and caspase 8. The contribution of ferroptosis to ALS-related motor neuron death will be analyzed with antibodies against glutathione peroxidase 4 (GPX4), p53, SQSTM1/p62, and ATG5. The activation of autophagy will be visualised using antibodies against SQSTM1/p62, ATG5, ATG9A, ATG16L1, and LC3 for immunohistochemistry. 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, Alexa 647) and motor neuron axons (stained with SMI-32 and alpha-synuclein). 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

	megaBytes. 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 For the use of postmortem material, approval was obtained on September 21th of the Ethics Committee Research of the University Hospital Leuven (UZ Leuven) for a modification of the project with number S60803 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). Upon collection of primary tissue from human donors, donors were fully informed using an informed consent that their cell material may be used for the production of iPSCs and that these iPSCs may be shared internationally to support research. For the ALS animal related research approval of the experiments was obtained by the Ethical Committee Animal Experimentation (ECD) (P020/2020) and changes in these protocols will always be submitted to the ECD before starting the respective experiments.. These animals are housed 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. Animal administrative, husbandry and animal welfare data are sensitive data and are stored in the LAIS database according to security procedure of KU 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?	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 material obtained within this project 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 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). - 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: 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. - Cells: 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. 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. - 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 and protein sequences: All nucleic acid and protein 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 (microarray data / RNA-seq data / CHIPseq data), the Protein Database (for protein sequences), the EBI European Genome-phenome Archive (EGA) for personally identifiable (epi)genome and transcriptome sequences.
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 5000 €. 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 mouse colony alive costs about 1,200 euro per year (for 6 cages), excluding the costs of genotyping. When no experiment is planned with a particular mouse strain, and in compliance with the 3R’s rule (https://www.nc3rs.org.uk), cryopreservation will thus be used to safeguard the strain, prevent genetic drift, loss of transgene and potential infections or breeding problems. Cryopreservation of sperm/embryos costs about 500 to 700 euro per genotype, plus a minimal annual storage fee (25 euro per strain for 250 to 500 embryos). Frozen specimen are kept in two separate liquid nitrogen tanks at two different sites on campus. When necessary, the costs of revitalization from cryopreserved sperm/embryos are about 1,100/600 euro. -Maintaining zebrafish costs about 10,000€ per year. When no experiment is planned with a particular line, and in compliance with the 3R’s rule (https://www.nc3rs.org.uk), cryopreservation will thus be used to safeguard the line, prevent genetic drift, loss of transgene and potential infections or breeding problems. Cryopreservation of sperm costs about 500 to 700 € per genotype, plus a minimal annual storage fee of 25€ per strain. Frozen specimen are kept in two separate liquid nitrogen tanks at two different sites on campus. When necessary, the costs of revitalization from cryopreserved sperm are about <euro>. 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. 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.

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 project.
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”. -Tissue samples: Tissues 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). Human pluripotent stem cell lines generated during this project will be deposited in hPSCreg. Animal cell lines will be stored in liquid nitrogen cryostorage of the laboratory. -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 15,000 €. 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 mouse colony alive costs about 1,200 euro per year (for 6 cages), excluding the costs of genotyping. When no experiment is planned with a particular mouse strain, and in compliance with the 3R’s rule (https://www.nc3rs.org.uk), cryopreservation will thus be used to safeguard the strain, prevent genetic drift, loss of transgene and potential infections or breeding problems. Cryopreservation of sperm/embryos costs about 500 to 700 euro per genotype, plus a minimal annual storage fee (25 euro per strain for 250 to 500 embryos). Frozen specimen are kept in two separate liquid nitrogen tanks at two different sites on campus. When necessary, the costs of revitalization from cryopreserved sperm/embryos are about 1,100/600 euro. -Maintaining a zebrafish line alive costs about 500€ per year. When no experiment is planned with a particular line, and in compliance with the 3R’s rule (https://www.nc3rs.org.uk), cryopreservation will thus be used to safeguard the line, prevent genetic drift, loss of transgene and potential infections or breeding problems. Cryopreservation of sperm costs about <euro> per genotype, plus a minimal annual storage fee of <euro>. Frozen specimen are kept in two separate liquid nitrogen tanks at two different sites on campus. When necessary, the costs of revitalization from cryopreserved sperm are about 300€. 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 We aim at communicating our results in top journals that require full disclosure of all included data. Biological material 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?	Immediately after the end of the project , 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 has been requested in this project.

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.