DMP - KU Leuven Internal Funds - ENA-E0835-METH/21/05

1. General Information	
Name PI	Bart De Strooper - bart.destrooper@vib.be
Project Number & Title	ENA-E0835-METH/21/05 Resilience and vulnerability in the cellular phase of Alzheimer's Disease

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 Data from study participants: Informed consent documents (text files in .doc/.docx format and .pdf format), ethical approval documents (text files in .doc/.docx format and .pdf format), patient information data, neuropsychological evaluation data and neuroimaging data (electronic health record files). iPS cell lines: 6 with high PRS, 6 with low PRS, 6 superresilient centenarians: genomic scans of the patients cell lines were derived. Experimental data Digital images Microscopy pictures (.tif/.tiff and .jpg format), gel scans (.jpg format), graphs (.xls/.xlsx and .pzf/.pzfx format) Omics data Bulk, single cell transcriptomics and spatial transcriptomics data: nucleotide and protein sequences (.ab1, .fasta/.fa format and accompanying .qual finle), next generation sequencing raw data (.bc) and .fastq format), sequence alignment data: (.sam and .bam format), coverage data (.bed format), read/UMI count data (.tsv format). Vectors CRISPR-Cas9 viral vectors to engineer cell lines (stored frozen) Lentiviral vectors to overexpress or knockdown genes of interest Lentiviral vector containing guideRNA library

	AAV vectors to express APOE variants Cell lines Human induced pluripotent stem cells (iPS) and embryonic pluripotent stem cells (ePS) and iPS cell lines from 6 patients with high PRS, 6 patients with low PRS and 6 superresilient centenarians Genetically modified organisms Transgenic mouse strains (APP N-LG-F strain, Rag2-/- ɣc-/- CSF1h/h strain, THY-Tau22 strain, APOE-/- strain and animals resulting from crosses) (live animals in KU Leuven housing). Antibodies Antibodies against classical microglia markers (e.g. IBA-1), activated microglia (e.g. CD11b), neurons (e.g. MAP2) and astrocytes (e.g. GFAP) (stored at 4°C or frozen) Synthetic compounds Amyloid-staining dyes (e.g. Thioflavin-S); Csfr1 inhibitor PLX3397 Simulation data Derived and compiled data

	Research documentation Research documentation generated by the research and technical staff or collected from online sources and from collaborators, including ethical approval documents, laboratory notes, protocols, animal husbandry data (text in .doc/.docx format, tabular data in .xls/.xlsx or .csv format, images and text in pdf format). Manuscripts Text files (.doc/.docx and .pdf format) and associated figures (.jpg, .svg and .pdf format), authors declaration (.pdf format) Algorithms and scripts Existing and custom-made algorithms and scripts Canonical data Nucleic acid sequences DNA and RNA sequences (.ab1, .fasta format and accompanying .qual file) 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 neurodegenerative diseases and brain science in general, and for journalists and higher education teachers working in the field of neuroscience.

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.	Yes We will use patient data (age, sex, diagnosis, clinical data). All the data processing activities have been approved by the Ethics Committee of UZ Leuven / KU Leuven (see below), after review by the Privacy and Ethics team. 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.

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 work received approval from the following ethical committees: for work on human subjects/human biological material/human data: Ethics Committee Research UZ / KU Leuven (S64349, S63259, S62888, S62734, S61187 and S61188) for work with laboratory animals: Ethical Committee Animal Experimentation (ECD) (P081/2021, P021/2021, P154/2020, P072/2019, P053/2021, P105/2021, P005/2022, P082/2022, P014/2022, P177/2022, P125/2022, P127/2022, ¨132/2022, P153/2022, P150/2022) The work on human biomaterial samples strictly complies with the ethics principles of free donation, informed consent and protection of privacy, in particular and the General Data Protection Regulation (GDPR) 2016/679. Genetically modified organisms: 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. 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.

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. 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. 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. Animal cell lines will be stored in liquid nitrogen cryostorage of the laboratory. Genetically modified 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 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 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 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. 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. Datasets collected in the context of clinical research, which fall under the scope of the Belgian Law of 7 May 2004, will be archived for 25 years, in agreement with UZ Leuven policy and the European Regulation 536/2014 on clinical trials of medicinal products for human use.
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. 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). Genetically modified organisms: strains that are no actively used for experiments will be cryopreserved. 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 these 10 years? How will the 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”. 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. 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 Personal data will only be published after de-identification and identifiers will not be published. If despite all efforts it is not possible to protect the identities of subjects even after removing all identifiers, personal data will not be made public. In order to respect the patient's privacy, clinical samples will only be available to the research and technical staff involved in the project, not to other groups, studies or purposes. 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 an Open Access repository, 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 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 the KU Leuven IT team 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, 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.