Human-mouse neuronal xenografts to study incipient Alzheimer’s Disease.

A Data Management Plan created using DMPonline.be

Creator: Sriram Balusu

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Data Manager: Eliane Cherretté

Grant number / URL: G064023N

ID: 200061

Start date: 01-01-2023

End date: 31-12-2026

Project abstract:

Alzheimer’s disease is a major health problem. Progress in the field has been hampered seriously by the lack of a good animal model to study the early steps of the disease. We have developed a new xenograft mouse model for Alzheimer's disease. We transplant human neuronal precursor cells in mouse brain. Remarkably these neurons integrate well and survive >18 months. However when we transplant the neurons in a mouse model that develops the amyloid plaques characteristic for Alzheimer's disease, the neurons become sick after 3 months and develop all the characteristics (including neuronal tangles and neurodegeneration) that are seen in Alzheimer patient brains. This pathology is already fully present at 6 months of age. Thus we are able for the first time to see how a healthy neuron evolves to a diseased neuron over time. In the current project we propose to use cutting edge technology, such as single nuclei sequencing, spatial transcriptomics and advanced proteomics to characterize all the molecular and cellular changes that are occurring over these 4 month time period. We propose also a way to validate any candidate drug target coming from those experiments.

Last modified: 22-06-2023

Human-mouse neuronal xenografts to study incipient Alzheimer’s Disease. 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
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .cvs,.pdf, .txt, .rtf, .dwg, .gml, ... • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Microscopy images xls, .pzfx, .NIS	Experimental immunohistochemistry, histology and Immunofluorescence images taken in brightfield, confocal microscopes or electron microscopy	Generate new data	Digital	Experimental	.NIS .tif .jpeg	<100GB	Tissue slides will be stored at 4 degrees, and they will take approximately ten boxes.
Single-cell transcriptomic datasets	Gene expression level of individual cells by simultaneously measuring the mRNA concentration of thousands of genes	Generate new data	Digital	Experimental	.fastq	<1TB	
FACS data	Flow Cytometry and fluorescence-activated cell sorting (FACS) data	Generate new data	Digital	Experimental	.fcs	<100GB	
vectors	Bacterial vectors, viral vectors	Generate new data	Digital and physical	Experimental			Viral vectors will be stored at -80 degrees, and they will take approximately 2-3 boxes.
cell lines	new stable stem cells line generated by transducing the lentiviral vectors	Generate new data	physical	Experimental			Will be stored in Liquid nitrogen, and they will take approximately 5-10 boxes.
Tissue samples	Either brain sections, protein samples or RNA samples derived from the tissue samples.	Generate new data	physical	Experimental			Will be stored at -80 degrees, and they will take approximately 10-15 boxes.
Westernblots .xls, .pzfx	Numerical data. Quantification of western blot images, performed using imageJ, in microsoft excel and graph pad prism.	Generate new data	Generate new dataDigital	Experimental			

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:

Not applicable.

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, animal data

All experiments involving research on small laboratory animals are approved by the ethics committee of KU Leuven (ECD application number P177/2022) and will be executed in compliance with the ethical regulation for animal research.

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.

- Yes

We do not exclude that the proposed work could result in research data with the potential for tech transfer and valorization. VIB has a policy to monitor research data for such potential actively. If there is substantial potential, the invention will be thoroughly assessed, and in several 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 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.

- No

No third-party agreement restricts dissemination or exploitation of the data from this project. In particular, existing agreements between VIB and KU Leuven do not restrict publication of data.

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).

Experiments are tracked using ELN, a cloud-based software platform providing an electronic lab notebook, provided by VIB.

Digital files will be named following a standard procedure so that all the names of all files in a given dataset will be in the same format: All names will start with the date (and time if applicable), followed by the project acronym, a short but specific descriptive name and a version number (containing leading zeros as needed) if applicable whenever possible names will be kept under 32 characters. Names will only contain letters, numbers and underscores. Dots will only be used for version control indicators (minor revisions indicated by decimal numbers and major revisions by whole numbers): YYYYMMDD_HHmm_Project_Experiment_ version.format

All changes in the files will be recorded. Data files will be stored in suitably labelled and organized folders and sub-folders, accompanied by a README.txt file containing all the associated metadata (see section 3). File names and locations will be recorded in the E-notebook to allow electronic records to be linked to the raw data.
Specific naming and search procedures will be applied to:
- Omics and sequencing data: Digital files, raw sequencing data files mainly (.fastq, .fastq.gz), will be named following an in-house procedure, so that all the name of all files in a given dataset will be in the same format. All names will start with a project 3 letter code (Project Code) and a random 6-alphanumeric character code (unique to each sample given the Project Code), followed by a specific descriptive name of the sample and the technology used (e.g.: 10x). Names will only contain letters, numbers and underscores.

- Human pluripotent cell lines: these cell lines will be named according to the nomenclature rules established by the hPSCreg.eu database (https://hpscreg.eu/about/naming-tool). Names consists of a maximum of 15 characters and have the following structure: up to 6 characters for the institution, an i or e indicating the cell line type, 3 characters for the unique donor ID and finally a letter that is incremented depending on how many cell lines are generated by the same donor. Subclones are indicated by a hyphen followed by the subclone number. Cell lines can be searched for in the hPSCreg.eu database based on cell type, country of origin, associated disease and/or date of derivation.

- Manuscripts: Metadata information will be submitted alongside the final version of the manuscript, including the names, titles, email addresses, ORCIDs and affiliations of all authors. Upon publication, this metadata information will also be submitted to bibliographic databases such as Medline. All manuscripts will be assigned a unique Digital Object Identifer (DOI) by the publisher. Manuscripts will be given a descriptive title, and will be accompanied by keywords provided by the authors in order to maximize their findability.

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

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, ELN) that refer to specific datasets. All datasets will be accompanied by a README.txt file containing all the associated metadata (see more details below).
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.
For specific datasets, additional metadata will be associated with the data file as appropriate.
When depositing data in a repository, 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 (see section 7 below). 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, 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 staging 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: cell lines will be stored locally in the laboratory.
- Genetically modified organisms: Mice will be maintained in facilities of the Laboratory Animal Center of KU Leuven, which applies Standard Operation Procedures concerning housing, feeding, health monitoring to assure consistent care in accordance with European and national regulations and guidelines. All animals will be registered in the Leuven Animal Information System (LAIS) database, along with corresponding genotyping information, ethical approval documents and animal provider receipts. Other biological and chemical samples: storage at 4°C and/or as frozen samples in cryovials as appropriate.

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.

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 controllers with an Ontap 9.1P9 operating system.

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

Digital data is securely stored cloud drives and prevents unauthorized access or modification, KU Leuven ICT protocol employs robust security measures such as controlled access, and user authentication. Both the “L-drive” and “J-drive” servers are accessible only by laboratory members and are mirrored in the second ICTS data center 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 one 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.

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

The total estimated cost of data storage during the project is about 4000€ (for a final volume of 5 TB). 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.

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...).

The minimum preservation term of 5 years after the end of the project will be applied to all datasets.

Where will these data be archived (stored and curated 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: cell lines will be stored locally in the laboratory (-80°C).
- Genetically modified organisms: Drosophila lines will be housed locally. All other lines 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 the expected retention period? How will these costs be covered?

The total estimated cost of data storage during the 5 years after the end of the project is about 4500€ (for 5 TB). This estimation is based on the different costs described above.

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
- Other, please specify:

As a general rule all research outputs will be made openly accessible at the latest at the time of publication. No embargo will be foreseen unless imposed e.g. by pending publications, potential IP requirements – note that patent application filing will be planned so that publications need not be delayed - or ongoing projects requiring confidential data. In those cases, datasets will be made publicly available as soon as the embargo date is reached.

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

NA

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.

As a general rule, datasets will be made openly accessible via existing platforms that support FAIR data sharing (www.fairsharing.org). Sharing policies for specific

research outputs are detailed below:

- Omics datasets will be deposited in open access repositories such as the PRIDE Archive for proteomics data, the EMBL-EBI platform for genomics and epigenomics data, the LIPID MAPS Lipidomics Gateway for lipidomics data, the Metabolomics Workbench Data Repository for metabolomics data, or the NCBI Gene Expression Omnibus (GEO) or the EBI ArrayExpress databases for functional genomics data.
- Vectors: Upon publication, all vectors supporting a manuscript will be made publicly available via the non-profit plasmid repository Addgene, along with the corresponding DNA sequences. Addgene in turns performs quality control on the DNA, curates the plasmids online with all relevant information (maps, sequences), and for a minimal cost (typically $65) ships the vectors upon simple request and signature of a material transfer agreement. The MTA will be prepared before depositing the vectors with the help of our organization's Tech Transfer office. For transfer between nonprofit or academic institutions, Addgene typically uses the Uniform Biological Material Transfer Agreement (https://www.addgene.org/terms/1047/). All non-published vectors and the associated documentation will be shared by the PI upon request and after signature of a material transfer agreement, at no cost except the cost of shipment.
- Cell lines: All human pluripotent cell lines supporting publications will be registered in hPSCreg, the European human embryonic stem cell registry supported by the European Commission (https://hpscreg.eu/). Information about the deposited lines (including donor information, derivation method, availability and characterization) will also be made accessible. Registration of cell lines in hPSCreg will provide visibility, confirm ethical procurement and facilitate comparison with other hPSC lines. The PI will remain the distributor of the pluripotent cell lines. All other cell lines supporting publications will be deposited in the American Type Culture Collection (ATCC) database (https://www.atcc.org/), which is a private, non-profit biological resource center. This will provide a secure back-up for this material. Investigators can purchase cell lines from the ATCC database upon signature of a material transfer agreement (https://www.lgcstandardsatcc.org/~/media/PDFs/MTA_2.ashx)
- Research documentation: All protocols used to generate published data will be described in the corresponding manuscript(s), and the related documentation will be included as supplementary information. These data and all other documents (daily logs, raw data) deposited in the E-Notebook are accessible to the PI and the research staff, and will be made available upon request.
- Manuscripts: All scientific publications will be shared openly. Manuscripts submitted for publication will be deposited in a pre-print server such as bioRxiv, arXiv, Nature Precedings or ASAPbio). At the time of publication, research results will be summarized on the PI's website (add website address) and post-print pdf versions of publications will be made available there if allowed by copyright agreements, possibly after an embargo as determined by the publisher. Before the end of the embargo or in cases where sharing the post-print is not allowed due to copyright agreements, a pre-print version of the manuscript will be made available. Publications will also be automatically added to our institutional repository, Lirias 2.0, based on the authors name and ORCID ID.
- Data that do not support publication will be either deposited in an open access repository or made available upon request by email.

When will the data be made available?

Upon publication of the research results.
As a general rule all research outputs will be made openly accessible at the latest at the time of publication. No embargo will be foreseen unless imposed e.g. by pending publications, potential IP requirements – note that patent application filing will be planned so that publications need not be delayed - or ongoing projects requiring confidential data. In those cases, datasets will be made publicly available as soon as the embargo date is reached.

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

As detailed above, metadata will contain sufficient information to support data interpretation and reuse and will be conform to community norms. These repositories clearly describe their conditions of use (typically under a Creative Commons CC0 1.0 Universal (CC0 1.0) Public Domain Dedication or an ODC Public Domain Dedication and Licence, with a material transfer agreement when applicable). Interested parties will thereby be allowed to access data directly, and they will give credit to the authors for the data used by citing the corresponding DOI. For data shared directly by the PI, a material transfer agreement (and a non-disclosure agreement if applicable) will be concluded with the beneficiaries in order to clearly describe the types of reuse that are permitted.

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

NA

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.

6. Responsibilities

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

Metadata will be documented by the research and technical staff at the time of data collection and analysis

Who will manage data storage and backup during the research 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 manage 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 and from Raf De Coster for the KU Leuven drives.

Who will update and implement this DMP?

The PI is ultimately responsible for all data management during and after data collection, including implementing and updating the DMP.

Human-mouse neuronal xenografts to study incipient Alzheimer’s Disease. Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

Question not answered.

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

Question not answered.

What’s the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

Question not answered.

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

Question not answered.

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

Human-mouse neuronal xenografts to study incipient Alzheimer’s Disease. DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- Not applicable

Human-mouse neuronal xenografts to study incipient Alzheimer’s Disease. GDPR

GDPR

Have you registered personal data processing activities for this project?

- No