DMP title

Project Name Dissecting the role of cell-surface molecules in establishing longrange connectivity of a hippocampal interneuron subtype - DMP title

Project Identifier DMP_1189322N

Grant Title 1189322N

Principal Investigator / Researcher Natalia Martinez Vizcaino

Project Data Contact natalia.martinezvizcaino@kuleuven.be

Description I focus on a population of CA1 hippocampal interneurons that reside in the border between stratum radiatum and stratum lacunosum-moleculare. A subtype of these interneurons project long-range and the data that I will collect will address the differences between the long-range and local projecting subtypes. These differences include molecular markers, axonal development, synapse composition...

Institution KU Leuven

1. General Information

Name applicant

Natalia Martínez Vizcaíno

FWO Project Number & Title

1189322N Dissecting the role of cell-surface molecules in establishing long-range connectivity of a hippocampal interneuron subtype

Affiliation

- KU Leuven

2. Data description

Will you generate/collect new data and/or make use of existing data?

- Generate new data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

Aim	Type of data	Format	Volume	How created
1.1	Microscopy images	.tif	~100 GB	Confocal images of hippocampus and cortex at different postnatal ages.
1.1	Neurolucida neuronal reconstruction	.dat	~1 MB	Neuronal reconstruction from microscopy images
1.1	Sample quantification	.xls	~1 MB	SypGFP puncta quantification from microscopy images
1.1	Quantification graphs and analysis	.pzfx	~100KB	Analysis of SypGFP puncta quantification
1.2	Microscopy images	.tif	~100 GB	Confocal images of brain slices after rabies tracing.
1.2	Neuroinfo 3D reconstruction	.xml, mp4	~1 MB	3D brain reconstruction from rabies injected mice

2.1	Spatial transcriptomics data	.csv	~300 MB	List of all detected transcripts and their spatial locations in three dimensions
2.1	Spatial transcriptomics data	.tiff	~100 GB	Mosaic images to visualize transcripts and their location in the brain slice.
2.1	Spatial transcriptomics data	.csv	~300 MB	Output from the cell segmentation analysis: the transcripts per cell matrix
2.1	Spatial transcriptomics data	.csv	~100 MB	Output from the cell segmentation analysis: cell metadata.
2.1	Spatial transcriptomics data	.hdf5	~100 MB	Output from the cell segmentation analysis: cell boundaries.
2.1	Comparative analysis	.png	~10 MB	R and python generated graphs for spatial transcriptomics data
2.2	Microscopy images	.tif	~100 GB	Confocal images of brain slices from animals injected with TurboID AAV.
2.2	Blots images	.tif	~1 MG	Western blot images of brain lysates from animals injected with TurboID AAV.
2.2	Mass spectrometry raw data	.raw, index	100-300GB	Peptide detection of biotinylated proteins from TurboID injected brain lysates.
2.2	DIA-NN search file	Mainly .txt, tsb	20-100GB	Peak detection and quantification of raw mass spectrometry data
2.2	Perseus analysis	.sps	200MB-1GB	Analysis and statistics of DIA-NN output data.
2.3	smFISH probes database	.xls	~1 MB	List of genes targeted by smFISH probes, including gene region targeted by each probe.

2.3	Microscopy images	.tif	~100 GB	Confocal images of brain sections labeled with RNA scope or antibodies (validation of candidates from previous data)
3	Microscopy images	.tif	~100 GB	Confocal images of brain sections from candidate knock-out animals.
3	Neurolucida neuronal reconstruction	.dat	~1 MB	Neuronal reconstruction from microscopy images
3	Sample quantification	.xls	~1 MB	SypGFP puncta quantification from microscopy images
General	Benchling lab notebook	.html	~1GB	Protocols, experiment description, dates, animals used, etc
General	Written lab notebook	Hard copy	NA	Daily lab notes, agenda, etc
General	Nucleotide sequences	.ape	~100 MB	Maps and sequencing results vectors used
General	Manuscript/thesis figures	.tif	~10 GB	Figures created with Adobe Illustrator for publication and thesis
General	Manuscript/thesis figures	.docx	~100 MB	Written manuscript and written PhD thesis

3. Legal and ethical issues

Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- No

Privacy Registry Reference:
Short description of the kind of personal data that will be used:

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

Project licence number: 175/2021
Project licence name: Ontwikkeling ve hippocampaal interneuron subtype met lang connecties
Project licence system ID: 4082

Project licence number: Creation De Wit/2020

Project licence name: Creation project De Wit
Project licence system ID: 2024

Project licence number: 000/(GS1/GS2)Breeding-De Wit
Project licence name: 000-Breeding NO EXPERIMENTS
Project licence system ID: 1479

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?

- No

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 reuse of the data collected/generated in this project?

Data will be generated following standardized protocols. Detailed descriptions of these protocols will be stored in our lab protocol database and published along with the results. Metadata will be documented by the research and technical staff at the time of data collection and analysis in their personal laboratory notebook (either benchling and/or hard copy), and/or in hard copy lab notebooks that refer to specific datasets.

Cryotubes containing bacterial and yeast strains stored at -80°C will be labelled with a reference number that links to an entry in or strain database.

24well-plates containing brain slices will be stored at -20°C, allowing them to be reused. Labelling will include experiment name, animal number and date to allow the identification of the sample within the experiments detailed in the personal laboratory notebook (hard copy and benchling).

All datasets will be accompanied by a README.txt file containing all the associated metadata (see more details below).

Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- Yes

Metadata will include the following elements:
Title: free text
Creator: Last name, first name, organization
Date
Subject: Choice of keywords and classifications
Description: Text explaining the content of the data set and other contextual information required for the correct interpretation of the data, the purpose of the experiment, the software(s) (version number included) used to produce and to read the data, etc.
Format: File format details
Resource Type: image, data set, audio, etc.
Identifier: DOI (when applicable)
Access rights: closed access, open access, restricted access, embargoed access.

Moreover, 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.

SOPs for biological data generation are kept on a dedicated KU Leuven shared drive. A central

excel file is stored on Microsoft Teams, documenting details as sample ID, location of the source sample in the freezer/fridge, SOP with which data generation was performed, deviations from SOP in data generation, experimental QC values (e.g. DNA concentrations, antibodies dilution factor), etc

For bioinformatics and microscopy image processing, a data analysis log will be kept detailing sequencing run ID; location of source files, the bioinformatics SOPs/scripts that were applied including any possible deviations from them, etc. 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 outline the file-naming rules used and list the contents of the other files. This will allow the data to be understood and reused if necessary by other laboratory members.

5. Data storage and backup during the FWO project

Where will the data be stored?

- **Digital files** will be stored on KU Leuven servers ("L-drive") and/or KU Leuven Microsoft teams.
- **Transcriptomics and proteomics data** generated during the project will be stored on KU Leuven servers.
- **Nucleic acid and protein sequences** generated during the project will be stored on KU Leuven servers. Upon publication, all sequences mentioned in the 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), the Protein Database (for protein sequences), etc.
- **Softwares, scripts and algorithm** used for the project will be stored in a private online git repository from the GitHub account of the department (https://github.com/vibcbd).
- **Vectors and associated sequences** supporting the manuscript will be sent to the non-profit plasmid repository Addgene, which will take care of vector storage and shipping upon request. As a general rule in the lab, 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).
- **Bacterial and yeast strains** will be stored in a -80°C freezer in the host laboratory.
- **Genetically modified organisms**: mouse lines 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. Animals will be listed in the Leuven Animal Information System (LAIS) database, along with corresponding genotyping information, age, ethical compliance documents, etc.
- **Tissue and other biological or chemical samples** will be stored at 4°C or -20°C, as appropiate, in the host laboratory.

How is backup of the data provided?

The data will be stored in the KU Leuven servers, concretely the L drive, where the data is backed up daily using snapshot technology. All incremental changes in respect of the previous version are kept online and the last 14 backups are kept.

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

The data will be stored in the KU Leuven servers, concretely the L drive, where the space can be increased upon request if necessary.

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

Electricity costs for freezers and fridges used for data and reactive storage are included in general lab costs as well as data storage and backup costs (173,78€/TB/Year to use the L drive from KU Leuven servers).

Maintaining a mouse colony alive can cost around 1,600 euro per year (for 8 cages), excluding the costs of genotyping. When the mouse strain is no longer used for experiments, cryopreservation of sperm/embryos will be utilized to safeguard the strain, prevent genetic drift, loss of transgene and potential breeding problems. This costs around 500 to 700 euro per genotype, plus a minimal annual storage fee (25 euro per strain for 250 to 500 embryos). Frozen specimens are kept in two separate liquid nitrogen tanks at two different sites on campus. When necessary, revitalization can be performed, which costs around 1,100/600 euro.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The “L-drive” server is accessible only by laboratory members and require KU Leuven credentials. Our passwords need to be changed once a year to strenghen security.

6. Data preservation after the FWO project

Which data will be retained for the expected 5 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 5 years after the end of the project will be applied to all datasets. As mentioned before, these datasets will be stored in KU Leuven servers and costs will be covered by the host laboratory.

Where will the data be archived (= stored for the longer term)?

All digital data (digital files, vector sequences, omics data, scripts, etc) will stored in KU Leuven servers for at least 10 years, conform the KU Leuven RDM policy. In addition, upon publication, datasets will be made openly accessible, whenever possible via existing platforms that support FAIR data sharing (www.fairsharing.org). Finally, all datasets supporting the publication will also be deposited in the Dryad repository, where they will be preserved indefinitely.

Regarding bacterial and yeast strains and tissue samples, they will be stored as described during the project, in fridges/freezers in the host laboratory.

Lastly, mouse lines used will either maintained in facilities of the Laboratory Animal Center of KU Leuven, if other members of the laboratory are using them, or cryopreserved to safeguard the strain.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

As mentioned before; electricity costs for freezers and fridges used for data and reactive storage are included in general lab costs as well as data storage and backup costs (173,78€/TB/Year to use the L drive from KU Leuven servers). At the moment, I don't expect to generate more than 1TB of data.

The costs will also depend on the decision of maintaining the mouse line that I use alive or cryopreserve sperm or embryos. Maintaining a mouse colony alive can cost around 1,600 euro per year (for 8 cages) while cryopreservation costs around 500 to 700 euro per genotype, plus a minimal annual storage fee (25 euro per strain for 250 to 500 embryos). When necessary, revitalization can be performed, which costs around 1,100/600 euro.

All these costs can be assumed by the lab.

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)?

- No

Which data will be made available after the end of the project?

All research data supporting publications will be made openly accessible. We are committed to publish research results to communicate them to peers and to a wide audience. Depending on the datasets nature, some data may be made available before publication, either individually to interested researchers and/or potential new collaborators, or publicly via repositories (e.g.negative data). Also, biological material will be distributed to other parties if requested.

We aim at communicating our results in top journals that require full disclosure upon publication

of all included data, following their advice. However, accessibility restrictions may be applied depending on the journal.

Where/how will the data be made available for reuse?

- In an Open Access repository
- Upon request by mail

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?

For data shared directly by the PI, a material transfer agreement (and a non-disclosure agreement if necessary) will be redacted in order to clearly describe the types of reuse that are allowed.

Whenever possible, datasets and the associated metadata will be made publicly available through repositories that support FAIR data sharing. As mentioned before, metadata will contain enough information to support data interpretation and reuse, and will be conform to community norms. These repositories thoroughly 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). Thus, interested researchers will be allowed to access data directly, and they will give credit to the authors by citing the corresponding DOI.

What are the expected costs for data sharing? How will the costs be covered?

Our intention is to minimize data management costs by implementing standard procedures as using free-to-use data repositories and dissemination facilities whenever possible. Data management costs will be assumed by the host laboratory.
A budget for publication costs will be requested for this project.

8. Responsibilities

Who will be responsible for data documentation & metadata?

Data and metadata will be documented by the reseacher responsible of this project as well as and technical staff by taking careful notes in their personal laboratory notebooks (benchling and/or hard copy).

Who will be responsible for data storage & back up during the project?

The reseacher responsible of this project as well as and technical staff will ensure data storage and back up, with support from Raf De Coster (ICT department) for the KU Leuven drives.

Who will be responsible for ensuring data preservation and reuse ?

The PI is responsible for data preservation and sharing, supported by Raf De Coster (ICT department) for the KU Leuven drives.

Who bears the end responsibility for updating & implementing this DMP?

The PI bears the end responsability for all data management during and after data collection, including implementing and updating the DMP.