

## DMP title

**Project Name** IDENTIFICATION OF METABOLIC STATES THAT STIMULATE THE DIVERSIFICATION OF INTERNEURONS AND THE ASSEMBLY OF CORTICAL CIRCUITS. - DMP title

**Project Identifier** G0E9121N

**Grant Title** G0E9121N

**Principal Investigator / Researcher** Lynette Lim

**Description** Brains are composed of a multitude of neuronal cell types that assemble together into specific functional circuits. Elucidating how a small number of neural progenitors generates the vast diversity of neuronal cell types in the brain remains a major goal in neuroscience, since the different types of neuron " and their specific connectivity " are the basis for information processing. In this grant, we are collecting single cel RNAseq data and mass spectrometry data to uncover the metabolites driving cellular diversification.

**Institution** KU Leuven

### 1. General Information

#### Name applicant

Lynette Lim

#### FWO Project Number & Title

G0E9121N, IDENTIFICATION OF METABOLIC STATES THAT STIMULATE THE DIVERSIFICATION OF INTERNEURONS AND THE ASSEMBLY OF CORTICAL CIRCUITS.

#### Affiliation

- KU Leuven

KUL and VIB-CBD

### 2. Data description

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

- Generate new data
- Reuse existing 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).**

Type of data	Format	Volume	How created
Mass spec	.csv	10-50GB	Mass spectrometry
microscopy images	.tif/ ome.tiff	100-500GB	confocal and light microscopy images of mouse brain sections
single cell RNA seq	.bam/ fastq	500GB to 2TB	sequencing files
Spatial transcriptomics	.tif/ome.tiff	100-500GB	widefield images of probes
in vivo recordings	.csv	500GB to 2TB	electrical probe measurement

### 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)**

- No

**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?**

1. Microscopy images the following information will be collected in the metadata of the software which includes: dimensions, image type, bit-depth, pixel sizes and microscope settings. The methodology and protocol will be described in detail in our lab enotebook system which is an online based platform.
2. Mass spec data, the information on the instrument settings will be collected in the metadata which includes: ionization energy, mass spectrum collection range. The extraction and sample collection detail will also be recorded in the our lab enotebook system which is an online based platform.
3. Single cell sequencing data. The alignment to mouse genome, QC procedures and cell identity/ clustering will be recording in a R markdown file. The sample collection and library prep protocol will also be recording in our lab enotebook.
4. Spatial transcriptomics data. The genes and probe barcode used will be recorded in a csv file as metadata. the settings of imaging will also be recording in the imaging software as a metadata and this will include the dimension, laser power, and rounds of washes during the image acquisition. The sample preparation will also be recording our enotebook system.
5. In vivo recording data. The electrical activity recorded will include the details of animals used and genotype information as well as age in our enotebook system.

**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

#### **5. Data storage and backup during the FWO project**

**Where will the data be stored?**

1. The raw data, with time-stamped master copy of the data will be kept on our research unit central storage facility.
2. We will collaborate with researchers from other research units and groups for some data analysis work. Thus, we will use both dropbox and OneDrive for active use of the data during the project.

**How is backup of the data provided?**

The data will be stored on the university's central servers with automatic daily back-up procedures. Copies can be made and kept on personal devices and dropboxes and one-drive as backup on the cloud.

**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

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

25,000 euro

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

All our data will be stored in the university's secure environment. Dropbox and ONEdrive login are also secured by authorized user login name and password.

## **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, ...).**

All data will be store for a minimun of 5 year period.

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

The data 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.

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

The database of images, sequencing datasets, massspec and in vivo recording, that will be compiled to the project, will be hosted on the servers of KU Leuven as well as dropbox/ onedrive. In view of the expected size of the database (> 20 TB), estimated cost will be 10,000 euro to set up the database and additional storage space and an annual fee of 5,000 euro for support, totalling abotu 25,000 euro. Our lab is part of the VIB-CBD which has dotation from VIB to cover this cost.

## **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)?**

During the research, all electronic files (images, sequences, algorithms, etc...): will be stored on KUL servers, with hourly on-site backup and mirroring. All samples will be stored as appropriate: - 80°C for nucleic acids, protein samples and cell lines; animal house for living organisms (sperm and/or embryos will be cryo-preserved).

Upon publication, all raw data will be deposited on public repositories (sequencing data in NCBIGEO). Manuscripts: will be published and archived in public repositories. Vectors: will be deposited in public repositories such as Addgene.

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

The full dataset will be uploaded in a cvs format in Zenodo under a CC-BY license.

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

- In an Open Access repository

1. The full dataset with documentation will be uploaded in a cvs format in Zenodo.
2. Vectors/ plasmids generated will be deposited in Addgene.
3. mice generated will be deposited on Jax.
4. Upon publication, all sequencing data in NCBI-GEO.
5. The source code will be released on GitHub.
6. all other data will available on request after signing a data sharing agreement.

**When will the data be made available?**

- Upon publication of the research results

The full dataset will be uploaded in various repository, depending on data type (such as physical material such as vectors and mice), immediately upon publication of research results. Most will be uploaded on csv format in Zenodo.

**Who will be able to access the data and under what conditions?**

The full dataset will be uploaded in a csv format in Zenodo as an open access dataset under a CC-BY license. Therefore, it will be available to anyone for any purpose, provided that they give appropriate credit to the creators.

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

the cost of datasharing will be set by the repository. For raw data/ csv, the cost will be downloading / storage cost for the user. For plasmids and animals, these are shipping and cryopreservation cost which will be covered by the request users.

**8. Responsibilities****Who will be responsible for data documentation & metadata?**

Designation of responsible person: Lynette Lim

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

Designation of responsible person: Lynette Lim & Rani Moors (Lim lab manager)

**Who will be responsible for ensuring data preservation and reuse ?**

Designation of responsible person: Lynette Lim

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

The PI bears the end responsibility of updating & implementing this DMP.