MULTISCALE DYNAMICS OF MAMMARY GLAND REMODELING

A Data Management Plan created using DMPonline.be

Creator: Colinda Scheele

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: G003923N

ID: 200848

Start date: 01-01-2023

End date: 01-01-2027

Project abstract:
The structure of the breast epithelium looks like a tree, and the
branches in this tree continuously remodel. In the adult female, this remodeling is driven by hormones, including estrogen and
progesterone, which together drive cyclic changes of branch growth and regression in the breast tree during each menstrual cycle to optimally prepare the breast for lactation in case the female gets pregnant. Although this cyclic remodeling of the breast is well described, it is unknown how the breast epithelial cells are able to make these branches every menstrual cycle. For example, it is unknown whether it is always the same cell that is driving the growth of a branch, or whether each round different cells are responsible for the branch formation. We want to understand how the different breast cells cooperate to drive the cycles of growth and regression in the breast. We will use a microscopy technique, intravital microscopy, which enables us to follow the cycles of growth and regression live at a cellular resolution in a living mouse. This will give us detailed 4- dimensional information (3 dimensions over time) of this remodeling process of the breast. We will analyze the behavior of the breast cells in these images, and with this information we will build a theoretical model of breast remodeling. This model will teach us how breast remodeling works, and will give us insights into situation when the remodeling is out of control, for example during tumor formation.

Last modified: 29-06-2023

MULTISCALE DYNAMICS OF MAMMARY GLAND REMODELING

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.

Dataset Name	Description	New or reused	Digital or	Only for digital data Digital Data Type	Only for digital data Digital Data	Only for digital data Digital data volume	Only for physical data Physical volume

			Physical		format	(MB/GB/TB)
Intravital microscopy images	Images derived of mammary glands of different genetic mouse models	Generate new data	Digital	Experimental	.lif/.tiff	>50TB
3D images	Whole mammary gland images	New data	Digital	Experimental	.lif/.tiff	>20TB
Analyses data	Analyses of branching patterns/quantification of clonal dynamics	New data	Digital	Experimental	.xlsx	>100mb
Biophysical models	Biophysical models of branching dynamics	New data	Digital	Simulation data, Models	Code (R-code, python), .xlsx	>2TB
Final manuscripts	Figures and text	New data	Digital	Aggregated data	.pdf, .ai	10GB

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:

-

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

Mouse experiments related to this project will be performed under the following animal licenses approved by the ethical committee (ECD) of KULeuven:
P177/2021
P093/2022

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.

● No

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

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).

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 in hard copy lab notebooks that refer to specific datasets.

All protocols and necessary details related to data collection as well as methods of analysis will be recorded in licensed E-lab journal containing **word (.docx), endnote (.enl)**, and **excel (.xlsx) files** stored at a shared KU Leuven Large Volume Storage drive, which is backed up by KU Leuven IT services. The raw files will be segregated in separate folders according to the Work Packages and experiments within the Work Packages itself.

All standard operating procedures, protocols, lists of materials, lists of cell lines (either commercially available or generated for the project) will be stored in a shared folder on the KU Leuven Large Volume Storage server. The names of files will include date of the experiment, experiment number, type of experiment and different experimental conditions to make the data findable. All biological material will be labelled and stored according to good scientific practice. Mouse data will be kept in the LAIS mouse database, including all the data on the procedures and surgeries that were performed for each mouse.

All data generated in this project will be available to the wider scientific community upon publication. Raw data will be deposited in data repositories, code will be made available on github.

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.

- No

Each folder containing a separate experiment will also contain information in a Word (.docx) and Excel (.xlsx) file explaining data methods and all relevant metadata, which include but are not limited to experimental conditions, genetic models used, all sample identification numbers and computational analysis pipelines. Metadata files with detailed explanations will be stored in a shared folder on the KU Leuven Large Volume Storage server. This will ensure the reusability of the data and the reproducibility of any further data generation.

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.

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.

3. DATA STORAGE & BACK-UP DURING THE RESEARCH PROJECT

Where will the data be stored?

All data (except for the large imaging files) will be stored on the L-drive (Large Volume Data Storage) in a dedicated folder for this project. Only the PI's and the researchers involved in this project will have access to these folders. This project will generate extremely large image files. All processed image files will be stored on the L-drive. To accommodate the raw imaging data, we have purchased an additional 180TB of network storage hosted by the KU Leuven ICT. Upon publication the data will be moved to the data archive (K-drive), which is designed for long-term storage of archived data. The data on this drive cannot be moved, modified, or deleted by the researchers, nor the PI's (only ICT service can modify these data).
- 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.
- Algorithms, scripts and softwares: All the relevant algorithms, scripts and software code driving the project will be stored in a private online git repository.

How will the data be backed up?

Data stored on the KU Leuven L-Drive is managed, maintained, and backed up by KU Leuven IT services. Specifically, mirror copies of the stored data are made immediately upon upload, for safety backup purposes. Raw imaging data stored on the extra network storage will be backed up once every few weeks, and only 1 backup will be kept.
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.

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:
- 180 TB of storage is already foreseen for raw image data storage
- The "L-drive" is an easily scalable system and is expandable in blocks of 5TB, built from General Parallel File System (GPFS) cluster with NetApp eseries storage systems, and a CTDB samba cluster in the front-end.

- The archive storage is done on the "K-drive" and this drive is expandable in blocks of 100GB.

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

Access to the servers is only possible through a KU Leuven user-ID and password, and will only grant access to data made accessible to the specific user-ID. KU Leuven works with a multi-factor authentication mechanism to increase the security. 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. Together, these measures ensure that non-authorized persons can't access or modify the data.

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

The estimated cost for the KU Leuven Large Volume Storage (L-)drive per 5TB per year is 569,20 euro. Total estimated size of the generated data on the L-drive within this project is 100TB, which reflects 11.384 euro per year. The performance of mirror copies of the stored data for safety backup purposes is included in the prize. These costs will be jointly covered by the project leaders at KU Leuven.
An additional 180TB is already available, paid by Prof. Scheele.

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. All datasets will be stored on the university's central servers with automatic back-up procedures for at least 5 years, conform the KU Leuven RDM policy. The costs (€156 per TB per year for "Large volume-storage") will be covered by the lab.

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.

-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).
-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 the expected retention period? How will these costs be covered?

Each year €128 will be charged from our ICT service for the use of 1 TB on the k-drive (long term storage), back-up service is included in the price. These costs were foreseen in the budget request of the application and if more, the lab budget will be used to cover these expenses.

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

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

Question not answered.

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.

In an Open Access repository
Other (specify): 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.

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

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.

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

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.

6. RESPONSIBILITIES

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

PI and involved researchers

Who will manage data storage and backup during the research project?

ICT managers of KUL and VIB

Who will manage data preservation and sharing?

PI, assisted by ICT managers

Who will update and implement this DMP?

PI

MULTISCALE DYNAMICS OF MAMMARY GLAND REMODELING

GDPR

GDPR

Have you registered personal data processing activities for this project?

- Not applicable

MULTISCALE DYNAMICS OF MAMMARY GLAND REMODELING

DPIA

DPIA

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

- Not applicable