TO CO-OPT OR DESTROY: UNDERSTANDING THE MECHANISMS DRIVING REPLACEMENT OR DESCTRUCTIVE HISTOLOGICAL GROWTH PATTERNS IN CANCER

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)

Principal Investigator: Colinda Scheele

Grant number / URL: G073223N

ID: 200835

Start date: 01-01-2023

End date: 01-01-2027

Project abstract:
Blending in with your surroundings is a tactic that many organisms use to disguise their appearance. Strikingly, also metastatic tumors are very efficient at hiding within the existing tissue structures. One particular way is to adapt to the existing tissue structure by replacing the surrounding healthy epithelial cells and subsequently making use of the healthy tissue infrastructure, such as supply of nutrients by blood vessels, to drive tumor growth. This type of tumor growth is called the replacement growth pattern, and these tumors are often resistant to treatment. In contrast, tumors that do not have the ability to co-opt the existing tissue structure, are classified as destructive. The destructive growth pattern does not leave the healthy tissue architecture intact, but instead destroys the surrounding healthy tissue, and is more sensitive to treatment. Until now it is unknown what the trigger is for the different growth patterns, and how these patterns arise. We think that the replacement and destructive tumor cells have a different way of communicating with the neighboring healthy tissue. In this project, we aim to understand how replacement and destructive tumor cells grow within the healthy tissue and how
they communicate with the neighboring healthy cells. We will use these results to interfere with the identified communication
mechanisms to make tumors with replacement growth pattern
sensitive to treatment again.

Last modified: 29-06-2023

TO CO-OPT OR DESTROY: UNDERSTANDING THE MECHANISMS DRIVING REPLACEMENT OR DESCTRUCTIVE HISTOLOGICAL GROWTH PATTERNS IN CANCER

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.

New data will be generated. We will not be working with personal data.
All human tissue samples were obtained through the UPTIDER tissue donation program without any personal identifyers. These samples will be used to generate organoid models and patient-derived xenograft models. All experiments will be carried out with the experimental models.

Type of data	Format of data	Volume	How created
Flow cytometry	.fcs files	25 GB	Using diverse flow cytometers/sorters
Microscopy images	.lif, .tiff, .zem	150 TB	Using diverse microscopes
Sequencing files	.raw, .fastq	2 TB	Using 10x genomics or other platforms
Spatial transcriptomics files	.fastq, .gem, gef	15 TB	Stereo-seq platfom or equivalent
Data analysis files	.xls	50 GB	Using image analysis software, FIJI, Matlab
Statistical data	.pzfx, .xls, .txt	5 GB	Graphpad, Rstudio
Tissue samples/organoid lines	Barcoded vials stored in liquid nitrogen	1000 vials	Excel sheet on L-drive with barcodes and details
Plasmids and vectors	.DNA	5 GB	On L-drive
R/Python scripts	.txt, .py	1 GB	Seurat, GSEA, ScVelo
Electronic lab notebooks	.enl	10 GB	On L-drive
Publication reports	.pdf, .docx	3 GB	On L-drive
Data presentation	.pptx	15 GB	On L-drive
Computational data	.dat, .txt	10 GB	On L-drive
Mouse strains	LAIS database	/	/
Photoshop files of data	.psd	20 GB	On L-drive
Illustrator files of data figures	.ai	20 GB	On L-drive
Experimental readouts	.pda, .txt, .xpt, .asyr, .xlsx	1 GB	Flex station, plate reader, Seahorse XFe24 analyzer

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:

Information of the patient samples collected through the UPTIDER program (led by Prof Christine Desmedt) will be obtained from the lab of Prof. Desmedt through an existing MTA/DTA with UZLeuven under study number S64410 UZ/KU Leuven Program for post-mortem Tissue Donation to Enhance Research. No personal identifiers will be shared.

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:
Ethical approval to generate the PDX models was obtained through number P158-2021

Human tissue donation:
All human biomaterial will be obtained through our clinical collaborators following the ethical principles under informed consent, protection of privacy, and upon voluntary donation. Metastasis from breast cancer patients will be obtained through the post-mortem tissue donation program UPTIDER (study number S64410/NCT04531696).
All patient-derived material will be used following the standard operating procedures for handling human biomaterial in our center, and in accordance with the European and national regulations.

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

All employees have signed a contract, and the IP rights are implemented in this employment contract. The proposed work could result in data with potential for technology transfer or valorization. All data generated within this project belongs to KU Leuven and VIB, in accordance with the agreement between both institutes.
The project will be actively monitored by the tech-transfer offices of both institutes to scout for research data with valorization potential. Each invention will be thoroughly assessed and if desired, the invention will be IP protected (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 in such a way that publication of the data is not 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.

- Yes

This project can result in data with the potential for technology transfer or valorization. The need for 3rd party agreements will be evaluated case by case in consultation with KU Leuven and VIB. We do not exclude that the work could become restricted due to 3rd party agreements.

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, sequencing raw data will be made publicly available using GEO or equivalent, data quantification will be shared using .xslx files.

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 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).

- Omics and single cell RNAseq data: omics and scRNAseq 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 subsequently in the archive area.

- Cell lines: Newly created human cell line/organoids 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 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 from the GitHub account of the department (https://github.com/vibcbd).

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.
- 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 single cells/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

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

Data will only be accessible by authorized members, i.e., PI and researchers working on the project. Folders will be managed by the ICT service of KU Leuven, and all data will be stored on drives and servers managed by KU Leuven. 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 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 for image storage. The total size of the imaging data within this project will be around 150TB.
For the K-drive (data archive) storage space of 1 TB is foreseen and will cost €128 each year, this is also expandable in blocks of 100 GB. The lab budget will be used to cover these expenses.

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.
-Omics and scRNASeq data: datasets will be stored on the "L-drive" or, for larger datasets, on the Vlaams Supercomputer Centrum.
-Organoid lines will be stored in the UZ Leuven Biobank (liquid nitrogen).
-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.

- Yes, Intellectual Property Rights

This project can result in data with the potential for technology transfer or valorization. The need for 3rd party agreements will be evaluated case by case in consultation with KU Leuven and VIB. We do not exclude that the work could become restricted due to 3rd party agreements.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

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.

When will the data be made available?

In an Open Access repository
Other (specify): Upon request by email

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

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.

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?

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. The data will be reviewed by the principal investigator.

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

Data storage and back-up: VIB IT-manager (Urbain Schepereel) and the KULeuven ICTS-IT department (Raf De Coster).

Who will manage data preservation and sharing?

The PI is responsible for data preservation and sharing, supported by the research/technical staff involved in the project, the VIB IT manager (Urbain Schepereel) and the KUL IT department.

Who will update and implement this DMP?

The PI

TO CO-OPT OR DESTROY: UNDERSTANDING THE MECHANISMS DRIVING REPLACEMENT OR DESCTRUCTIVE HISTOLOGICAL GROWTH PATTERNS IN CANCER

GDPR

GDPR

Have you registered personal data processing activities for this project?

- No

TO CO-OPT OR DESTROY: UNDERSTANDING THE MECHANISMS DRIVING REPLACEMENT OR DESCTRUCTIVE HISTOLOGICAL GROWTH PATTERNS IN CANCER

DPIA

DPIA

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

- No