1278223N

A Data Management Plan created using DMPonline.be

Creators: Jeroen Declercq, n.n. n.n.

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: n.n. n.n.

Project Administrator: Jeroen Declercq

Grant number / **URL:** 1278223N

ID: 198333

Start date: 01-10-2022

End date: 30-09-2025

Project abstract:

Being the deadliest form of skin cancer, cutaneous melanoma represents a major clinical and societal challenge. Melanomas are made up of millions of both malignant and stromal cells, including immune cells, fibroblasts and blood vessels. Recently, the host laboratory has identified a malignant stem-like cell population that fuels tumor growth. These melanoma stem cells (MSCs) are predominantly located around blood vessels. Based on preliminary data I generated, I hypothesize that extrinsic cues emanating from the perivascular niche promote tumor fueling competences. Accordingly, endothelial cells promote the MSC phenotype in melanoma cells in vitro. Herein, I propose to draft dynamics maps of the cellular interaction networks occurring at the perivascular niche in primary and metastatic melanoma lesions making use of spatial multi-omics single-cell technologies, and, subsequently, to build in vitro model(s) that recapitulate key features of these networks. State-ofthe-art bioinformatic approaches will be used to identify molecular signals by which stromal cells drive the phenotypic reprogramming of cancer cells into MSCs. Finally, I will test a series of potential therapeutic targets aiming to intercept the communication between melanoma cells and their niche using in vitro and in vivo models. Together, this project represents a unique training opportunity, deploying cutting edge technologies to address a very exciting question in the urgent field of melanoma biology.

Last modified: 07-04-2023

1278223N
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	
10X visium		Reuse	Digital	experimental	.fastq .TIFF	<100GB	
Resolve		Reuse	Digital	experimental	.TIFF	<100GB	
Stereo-seq		Reuse	Digital	experimental	.fastq .TIFF	<1TB	
OPAL		new	Digital	experimental	.TIFF	<100GB	
Genetically Engineered Murine Model (GEMM)		Reuse	physical				1/2 rack
Non-genetically engineered mice		new	physical	experimental			1/2 rack
paraffin blocks		Reuse	physical				boxes with blocks
human samples		reuse	physical				box with slides
Engineered cell-lines		new	physical				box in liquid nitrogen
imaging		new	digital	experimental	.tiff	<1TB	
primary cells		new	physical				box in liquid nitrogen
FACS		new	digital	experimental	.fcs	<100GB	
bioinformatic scripts		reuse	digital	software	.r .py	<100MB	
RNA-Seq		new	digital	experimental	.fastq	<100GB	

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:

For the bioinformatic analysis we will reuse scripts that have previously used for other projects and are available on Gitlab: Marine Lab · GitLab (kuleuven.be)
We will reuse tumor blocks (cryo + paraffin) from mouse and human tissue that are available in the lab.
We will reanalyse spatial omics data:

- 10X visium and stereoseq are in the GEO repository: GSE207592
- Resolve data are in Zenodo: A cellular hierarchy in melanoma uncouples growth and metastasis | Zenodo (DOI: 10.5281/zenodo.6856193)

We will use mouse cell line previously generated from GEMMs used in our lab.

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, human subject data
- Yes, animal data

ECDs related to this project: 183/2022, 097/2020
for WP 3.1: the ECD still has to be written
Ethical committee approval: S62275 (SPECIAL), S67149 (MelanoMAP)

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 potential for tech transfer and valorization. VIB has the policy to actively monitor research data for such potential. If there is substantial potential, the invention will be thoroughly assessed, and in number of 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 publication need not be delayed.
The use of the mouse models will be subjected to the terms descried in their respective MTAs.
Specific examples:

- WP3 might identify druggable targets for NRAS-mutated melanomas
- new fluidic chips concepts will be claimed (in collaboration with J Lammertyn lab)

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

We have MTAs with addgene for several plasmids. Those plasmids can only be used for teaching and academic research purposes. We can not distribute the plasmids to other users without written consent of the provider. The full MTA can be found here: Addgene UBMTA.
We have MTA with Pr A Regev for Watermelon barcoding system which we will use only for academic research purposes and we will not share with anyone else.

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).

The data will be generated follow standard operating procedures (SOPs). Clear and detailed descriptions of these protocols will be stored in our lab protocol database and published along the results.
Metadata will be documented by the research and technical staff at the time of the data collection and analysis. Those metadata will be stored in the same excel file together with the results and the SOP which is kept in the J-drive.
Large data sets will be kept on the L-drive.
For bioinformatic data the scripts associated with this project will be stored in Gitlab.
Several databases are used to keep track of the storage of physical data, eg. -80°C freezers, liquid nitrogen storage, antibodies, plasmids, primers, cryoblocks, paraffin blocks,...

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

The data are always kept together with the SOP and the relevant metadata.
For the following type of experiments international metadata standards will be used:

- qPCR: MIQE compliant RDML files will be exported from qbase+ (https://www.ncbi.nlm.nih.gov/pubmed/19246619)
- FACS & flow cytometry: MiFlowCyt compliant metadata files will be exported from Fortessa & Sony and Aria Fusion (https://www.ncbi.nlm.nih.gov/pubmed/18752282)
- Image metadata will be exported as OME-XML of TIFF-XML files using QuPath software.

If no metadata standards are available to following metadata will be stored:

- Investigator
- unique identifier
- link to related identifiers
- Project
- Keywords
- Type of experiment (in vitro, in vivo, bioinformatics)
- source of data (animal, cell lines, database, patients)
- type of material generated (DNA, RNA, protein, tumor fragment, digital)
- location of the results
- location of physical material
- device + program run
- compounds
- clearance

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. Cryoblocks of tumor tissue will be stored at -20°C, paraffin blocks of tumor tissue will be stored at room temperature. All human

tissue samples will be registered with a Belgian biobank, in compliance with the Belgian lawon 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 archive 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: Newly created human cell lines 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.
Bacterial and yeast strains will be stored in a -80°C freezer in the labs. Costs are covered by general lab expenses.
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 or ticket@lab, 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 in Gitlab
Nucleic acid and protein sequences: All nucleic acid and protein sequences generated during the project will be stored on KU Leuven servers. Upon publication all sequences supporting a manuscript will be made publicly available via repositorie such as the GenBank database or the European Nucleotide Archive (nucleotide sequences from primers / new genes / new genomes), NCBI Gene Expression Omnibus (microarray data / RNA-seq data / CHIPseq data), the Protein Database (for protein sequences), the EBI European Genome-phenome Archive (EGA) for personally identifiable (epi)genome and transcriptome sequences.

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) and a backup is stored on the L-drive.

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 controlers with an Ontap 9.1P9 operating system.
- the sequencing data is maintained on VSC storage and backed up on the L-drive

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

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 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 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. Data on VSC is only accessible by VSC username, password, and MFA, and permissions on data folders are set for specific groups in case of GDPR.

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 10,000 euro. 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”.
- the Staging on VSC has been renewed in 2021 with 3 Pb of space; and the gbiomed NAS is scaleable and will start with 150 Tb space. Cloud storage is scalable.
-Maintaining a mouse colony alive costs about 2,400 euro per year (for 12 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 specimens 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.
Electricity costs for the -80° freezers present in the labs are included in general lab costs.
Data storage and backup costs are included in general lab costs.

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 (€114 per TB per year for “Large volume-storage”) will be covered by the Marine labs.
Datasets collected in the context of clinical research, which fall under the scope of the Belgian Law of 7 May 2004, will be archived for 25 years, in agreement with UZ Leuven policy and the European Regulation 536/2014 on clinical trials of medicinal products for human use.

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: human cell lines will be stored in the UZ Leuven Biobank (-80°C). Animal cell lines will be stored in liquid nitrogen cryostorage of the laboratory.
-Genetically modified organisms: will be kept in the animal facility
-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 Data repositories, 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?

It is the intention to minimize data management costs by implementing standard procedures (internally generated operational procedures (https://en.wikipedia.org/wiki/Standard_operating_procedure) 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.

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

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.

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

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.

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.

GEO will be used to share mouse sequencing data
EGA will be used to share human sequencing data
RDR will be used to share other data sets such as eg. spatial omics data

When will the data be made available?

The data will be made available upon acceptance of the publication of the results except if data sharing has to be postponed for a certain time, for example to protect IP during patent application.

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

We are going to use CC-BY-NC-SA-4.0

- Free to share and adapt.
- Give appropriate credit, indicate if changes were made.
- Do not use the material for commercial purposes.
- Distribute your contributions under the same license as the original.

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

A permanent identifier is added to your data upon deposit in a repository such as EGA, GEO, RDR. Whenever possible we will use those DOI link in our publications.

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

Deposition of smaller datasets in data repositories is usually covered by the repository. For example in RDR every researcher can use 50 GB/year for free. If we will deposit larger datasets the costs will be covered by the lab of Chris Marine
The costs for sharing physical data will be paid by the researcher requesting the materials.

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 Pieter Joris and Urbain Scherpereel 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.

Who will update and implement this DMP?

The PIs are ultimately responsible for all data management during and after data collection, including implementing and updating the DMP, with the support of the labmanager.