Mapping the multi-modal single-cell landscape of metastatic and non-metastatic thin melanoma

A Data Management Plan created using DMPonline.be

Creators: Asier Antoranz Martinez, 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: Asier Antoranz Martinez

Grant number / URL: 12ATN24N

ID: 206732

Start date: 01-10-2023

End date: 30-09-2026

Project abstract:

Cutaneous melanoma is the most aggressive form of skin cancer accounting for 72% of skin cancer-related deaths. The main prognostic factor in primary melanoma is Breslow thickness which is used in the clinics to predict the metastatic potential of each patient (high risk if Breslow thickness > 1mm). Under 1 mm (“thin melanomas”, tCM), melanomas have such a low risk of metastasis that patients do not get further treatment nor follow-up. However, thanks to prevention policies 70% of newly diagnosed melanomas nowadays are tCM and 3-5% of these cases will develop metastasis. We have collected an international cohort of metastatic (M+) and non-metastatic (M-) tCM cases that I will use to build a spatial multi-omics single-cell landscape integrating different imaging modalities and decipher the cellular mechanisms and risk factors driving metastasis in tCM. Specifically, I will focus on characterizing the differences between M+ and M- tCM in terms of: (i) anatomical areas, cell populations, and cellular states present in the tumor and immune components, (ii) tissue architecture in terms of spatial proximity of cells and ligand-receptor colocalization, (iii) cellular and acellular features critical for tumor staging such as mitotic density and solar elastosis. These questions will be addressed by developing white-box oriented artificial intelligence algorithms setting the ground for future biomarker development that could be used to predict thin melanomas with risk of metastasis.

Last modified: 25-04-2024

Mapping the multi-modal single-cell landscape of metastatic and non-metastatic thin melanoma
DPIA

DPIA

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

- Not applicable

Mapping the multi-modal single-cell landscape of metastatic and non-metastatic thin melanoma
GDPR

GDPR

Have you registered personal data processing activities for this project?

- Not applicable

Mapping the multi-modal single-cell landscape of metastatic and non-metastatic thin melanoma
Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

The main datatypes generated throughout the project are images. Specifically:
- Spatially resolved mIHC images (czi and ome tiff format).
- Spatially resolved transcriptomic images (ome tiff format).
- H&E images (czi and ome tiff format).
- Histopathological parameters of the patients (csv format).

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

1. Designation of responsible person: Asier Antoranz (main applicant).
2. Storage capacity/repository: during and after the research, the data will be stored on the centrally managed KU servers which represents a secure environment with automatic backup procedures. The data will be stored for at least 5 years, and conform to the KU Leuven RDM policy which will provide to the responsible person. Additionally, the hosting laboratory has recently purchased a network-attached storage (NAS) system with 500Tb of storage (mirrored) where an extra copy will be stored. This NAS system is attached to KU Leuven's network so the same safety measurements apply.

What’s the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

Not Applicable.

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

All data come from patients. They will be coded (i.e. pseudonymized). There continues to be a link between the data and the individual who provided it. The subjects' identifiers will however be stored separately (site file) from their research data and replaced with a unique code to create a new identity for the subject. This code is stored on the UZ Leuven
server which is password protected, but which also allows to consult the electronic medical chart of the patient stored on UZ Leuven Hospital servers, only if necessary.
In addition, we will store all data on the central servers of the KU and UZ Leuven, which are protected against unauthorized access by firewalls. The same approach is applied to patients derived from external centers.

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

KU Leuven has recently joined the iRODS consortium (https://irods.org/), an open-source data management software that will be fully adopted as DMP.
All datasets will be made available upon publication of the research results, pending potential embargo periods due to IPR-related confidentiality clauses (no more than 2 years). Research results will be further discussed with the Leuven Research and Development (LRD) department before being made available.
All research results will be uploaded in csv format in Zenodo and GitHub 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.

Mapping the multi-modal single-cell landscape of metastatic and non-metastatic thin melanoma
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	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
MILAN	Multiplex immunohistochemistry images	New	Digital	Experimental	.czi and .ome.tiff	<50TB	NA
ST	Spatial Transcriptomics	New	Digital	Experimental	.ome.tiff	<10TB	NA
HE	hematoxylin and eosin images	New	Digital	Experimental	.ome.tiff and .czi	<10TB	NA
META	Database	Reuse existing data collected in the hospital	Digital	Observational	.csv	<1GB	NA

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:

Not applicable.

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

The clinical data, samples, and studies included in this project are approved by the Ethical Review Committee of the University Hospitals Leuven (S61610, amend 0002).

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.

- Yes

For this project, clinical data will be used from patients to associate features of interest with clinical parameters. All data will be pseudonymized. All datasets.

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 datasets. One of the endpoints of the projects is to identify biomarkers that are able to differentiate between metastatic and non-metastatic thin melanomas.

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

Yes, there are two 3rd parties. There are MTAs and DTAs in place for the transfer of the material and of the pseudonimized clinicopathological data for the cases we received from Florence and Poland.

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).

All data and accompanying information will be stored exclusively on KU Leuven servers, Onedrive. All data will be accompanied with a README file or tab that outlines the exact data collection procedure, especially important for experimental data. All experimental work is prepared by extensive preparations, each step is logged. Standard operating
procedures are written out in the lab and safely stored together with the experimental data in the same folders, to allow easy recovery of the metadata. All team members have access to these metadata.

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

All metadata for the experimental work are maintained in the lab's repository. These follow a standard format and vocabulary.

3. Data storage & back-up during the research project

Where will the data be stored?

All data are stored at the KU Leuven Onedrive, the internal NAS in our department, or the UZ Leuven servers (for clinical data). No data will be stored on local computers, hard drives, etc.

How will the data be backed up?

All data stored in the NAS is backed up automatically with version control and logging. Additionally, the NAS is mirrored to prevent catastrophic disaster.

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

We have 460 TB of data capacity in the NAS from which 100Tb have been allocated for this project.

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

The data are stored on the KU Leuven servers, only accessible with double authentication.

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

All costs have been covered by the departmental group.

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...).

All data will be retained for at least 5 years.

Where will these data be archived (stored and curated for the long-term)?

The same repositories as mentioned above will be used for long-term storage.

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

There are no added costs for data preservation. Unforeseen costs will be covered by the departmental group.

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 a restricted access repository (after approval, institutional access only, …)

All imaging data will be made available to selected users via our DISSCOvery software. This software provides a safe (password-protected) repository which directly feeds from the NAS.

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

Initially, data will be accessible by the team. In later stages, accessibility will be evaluated on a user-basis with the corresponding DTAs.

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.

As indicated above, data will be made available via our DISSCOvery platform.

When will the data be made available?

Data will be made available after publication of the research results.

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

Data usage will be provided together with a user/password and reviewed on a user-basis. The corresponding DTAs will be placed.

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.

- No

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

Usually, there is no charge for sharing data with third parties. For DTA of privacy-sensitive data, a quid pro quo in the form of co-authorship is usually requested.

6. Responsibilities

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

The team at the MILAN drylab group of the department of imaging and pathology (Asier Antoranz).

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

The IT Gbiomed team of the KU Leuven

Who will manage data preservation and sharing?

The team at the MILAN drylab group of the department of imaging and pathology (Asier Antoranz).

Who will update and implement this DMP?

The team at the MILAN drylab group of the department of imaging and pathology (Asier Antoranz).