11PI824N_PhilipGeorgDemaerel_FWO

A Data Management Plan created using DMPonline.be

Creators: Philip Georg Demaerel https://orcid.org/0000-0002-0187-272X, Jeroen Declercq

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Data Manager: Jeroen Declercq

Grant number / URL: 11PI824N

ID: 204305

Start date: 01-11-2023

End date: 31-10-2027

Project abstract:

Metastasis is responsible for the vast majority of cancer-related deaths. Preventing early metastatic dissemination is therefore a promising clinical strategy to improve patient outcome. However, an incomplete understanding of the cellular origin and mechanisms that drive metastasis has been a major barrier to the rational development of prognostic tools and effective therapeutics that intercept the disease before its dissemination to vital organs. The host laboratory has recently identified a population of cells that is at the origin of metastases in melanoma, a disease with a very high metastatic propensity, and developed tools to fate map and isolate these metastasis-initiating cells (MICs) and their disseminating descendants with high efficiency. With this project, I propose to extensively phenotype these MICs (and their progeny) and firmly establish their clinical relevance. I will study their spatial distribution within the entire melanoma ecosystem, define their cellular niche(s) and interaction partners among components of the stromal and immune cell compartments, with a specific focus on the perivascular niche. I will search for extracellular cues that contribute to their emergence, and investigate the mechanisms exploited by these cells to hide from the immune system. A long-term goal is to develop therapeutic strategies that either target MIC-inducing environmental cue(s) and/or increase MIC immunogenicity.

Last modified: 15-04-2024

11PI824N_PhilipGeorgDemaerel_FWO

FWO DMP (Flemish Standard DMP)

				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, .csv,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
10X Xenium		New	Digital	Experimental	.fastq .TIFF	<100GB	
OPAL		New	Digital	experimental	.TIFF	<100GB	
Genetically engineered Murine Model (GEMM)		Reuse	Physical				1/2 rack
Non-genetically engineered mice		Reuse	Phyisical	Experimental			1/2 rack
Paraffine blocks		Reuse	physical				boxes with blocks
Human samples		Reuse and new	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 and new	digital	software	.r .py	<100MB	
RNA-seq		new	digital	experimental	.fastq	<100GB	

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 and biobank.

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

- Ethical Committee Research S-numbers: S62275 (SPECIAL1.0), S67149 (MelanoMAP), S68578 (submitted 12-2023)
- ECDs P-numbers related to this project: 006/2023, 183/2022.

As part of S62275 (SPECIAL1.0), S67149 (MelanoMAP) and S68578 (submitted 12-2023), clinically relevant information will be collected and pseudo-anonymized by the UZLeuven. It includes the following data:

- demographic data, such as age, gender, ethnicity
- clinical data: dates of biopsies, response data, specifications of primary lesion and metastases, staging, previous treatments, response data, newly commenced treatments, mutational profiling

We acknowledge the potential of the proposed work to yield research data suitable for technology transfer and valorization. VIB maintains an active policy to assess research data for such possibilities. Should significant potential be identified, thorough evaluation will occur, often resulting in intellectual property (IP) protection, primarily through patents or copyrights. Importantly, IP protection does not preclude the public release of research data. If a decision is made to pursue patent protection, arrangements will be made to ensure publication is not delayed.

We currently hold Material Transfer Agreements (MTAs) with Addgene for several plasmids. These plasmids are designated solely for educational and academic research endeavors. Distribution of these plasmids to other parties requires explicit written consent from the provider. The comprehensive MTA is accessible via the Addgene UBMTA.

The use of mouse models will adhere to the terms outlined in their respective Material Transfer Agreements (MTAs).

Additionally, we have an MTA with Professor A. Regev for the Watermelon barcoding system. This system will be exclusively utilized for academic research purposes, with no plans for sharing with any other entities.

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 and ManGO (KULeuven).

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,...

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 and link to related identifiers
- Project and 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

- Digital files will be stored on KU Leuven servers and ManGO, except for private data that will be stored on KU Leuven secure server

(digital vault).
    - Omics data: omics data generated during the project will be stored on ManGO or on The Flemish Supercomputer Centre (VSC), initially in the staging area and later in the archive area.
    - 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 Mango. Upon publication all sequences supporting a manuscript will be made publicly available via repository 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.
- Tissue samples: Animal-derived issues 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 law on human body material (dd 19-12-2008), and stored at the local Pathology Department.
- 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.

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 ManGO.

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
- The ManGO data is stored securely in the data centers of KU Leuven. Of each file, two copies are stored: one in the datacenter of Heverlee, and one in the datacenter of 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 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.

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.
- ManGO has a total of 6 PB of storage space, with a fixed usage price of €35 per TB per year.
- 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.

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.

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 aforementioned “L-drive”.
- Tissue samples: Tissues will be stored locally in the laboratory.
- Omics data: datasets will be stored on the aforementioned “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.

We aim to reduce expenses related to data management by establishing standardized procedures (https://en.wikipedia.org/wiki/Standard_operating_procedure), such as internally developed operational protocols for tasks like metadata collection and file storage organization, right from the project's inception. Additionally, we'll leverage cost-free data repositories and dissemination platforms whenever feasible. The laboratory budget will cover these data management expenses.

Participants in this project are dedicated to disseminating research findings to both peers and a broader audience. All research outputs that support publications will be openly accessible. Depending on their nature, certain data may be accessible prior to publication, either on an individual basis to interested researchers or potential collaborators, or publicly through repositories (e.g., negative data).

Our objective is to share our results in leading journals that mandate comprehensive data disclosure upon publication, whether within the main text, supplementary materials, or deposited in a data repository as requested by the journal and in accordance with their deposit guidelines. Access restrictions may apply based on the journal's policies. Requests for biological materials will be accommodated as needed.

Whenever feasible, datasets along with the appropriate metadata will be openly accessible through repositories that adhere to FAIR data principles. As previously outlined, metadata will be sufficiently detailed to facilitate data interpretation and reuse, conforming to community standards.

These repositories articulate their usage terms, typically under licenses such as Creative Commons CC0 1.0 Universal (CC0 1.0) Public Domain Dedication, Creative Commons Attribution (CC-BY), or an ODC Public Domain Dedication and License, supplemented by a material transfer agreement when necessary. This approach enables interested parties to access data directly while acknowledging the authors through citation of the corresponding DOI. In cases where the Principal Investigator shares data directly, a material transfer agreement (and if applicable, a non-disclosure agreement) will be established with recipients to define permissible types of reuse clearly.

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

The data will be accessible upon acceptance of the publication of the results unless data sharing needs to be delayed for a specific period, such as to safeguard intellectual property during the patent application process.

We'll utilize the CC-BY-NC-SA-4.0 license:

- Free for sharing and adaptation.
- Provide proper attribution, noting any modifications made.
- Do not utilize the material for commercial endeavors.
- Distribute your contributions under the identical license as the original.

The data receives a permanent identifier upon deposition in a repository like EGA, GEO, or RDR. Whenever feasible, we'll incorporate these DOI links into our publications.

The deposit of smaller datasets in data repositories is typically supported by the repository itself. For instance, in RDR, each researcher can utilize 50 GB/year free of charge. Should we need to deposit larger datasets, the expenses will be covered by Chris Marine's laboratory.

The expenses associated with sharing physical data will be the responsibility of the researcher requesting the materials.

The research and technical staff will document metadata during the data collection and analysis phase.

The research and technical personnel will oversee data storage and backup, receiving assistance from Pieter Joris and Urbain Scherpereel for the KU Leuven drives.

The Principal Investigator oversees data preservation and sharing, with assistance from the project's research and technical staff.

The Principal Investigators bear ultimate responsibility for all data management throughout and after data collection, which includes the implementation and updating of the Data Management Plan (DMP), supported by the lab manager.