DMP title

Project Name Spatio-temporal fate mapping of a putative melanoma stem cell population (FWO) - DMP title

Grant Title 11M3822N

Principal Investigator / Researcher Chris Marine

Project Data Contact flavie.luciani@kuleuven.be

Description Melanoma is notorious for its high degree of heterogeneity, a feature associated with poor outcomes. Although cellular heterogeneity can have a genetic origin, there is growing evidence that cancer cells can acquire diverse phenotypic properties, including stem-like features, without any genomic alterations, through â€œphenotype switchingâ€. Combining single-cell RNA sequencing approaches with in vivo lineage tracing and mathematical modeling, the host laboratory identified a dedifferentiated/stem-like cell state that exhibits a high differentiation potential. Preliminary evidence indicated that this population may exhibit cancer stem cell properties. I propose herein to establish their functional contribution to primary and metastatic growth using genetic lineage tracing and ablation experiments. Importantly, this population is spatially distributed in close proximity to blood vessels indicating that this phenotype may be induced and/or maintained by a perivascular niche. I therefore propose to identify intrinsic and extrinsic mechanisms of tumor-stromal cell interactions that regulate their cellular identity, cell reprogramming ability (i.e. plasticity) and phenotype. I will then test whether pharmacological interventions that target key niche-dependent CSC specification mechanisms can restrict melanoma growth. This project will pave the way for the development of innovative therapies that contend with, or even exploit, the â€œchameleonicâ€ behavior of cancer cells.

Institution KU Leuven

1. General Information

Name applicant

Cecilia Pazzi

FWO Project Number & Title

11M3822N

Spatio-temporal fate mapping of a putative melanoma stem cell population

Affiliation

- KU Leuven
- Other

Dual affiliation KU Leuven and VIB

2. Data description

Will you generate/collect new data and/or make use of existing data?

- Generate new data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

Genetically Engineered Mouse Models (GEMM), allografts models (C57BL/6J and NU/J) & Patient Derived Xenografts (PDX) (WP1.1,WP1.2, WP1.3, WP2.1,WP3.1 WP3.2):

GEMMs of different genotypes, allografts (C57BL/6J and NU/J) and PDX models with implanted tumors will be bred in the 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. Treatment and follow-up of the animals tumor development and general health status will be detailed in an excel-file (<5MB). Sperm will be cryopreserved of newly imported genotypes.

Tumor derived materials (WP1.1, WP1.2, WP2.2, WP3.1, WP3.2, WP3.3):

From the mouse models in both GEMM, allografts and PDX models, we expect to extract gDNA, RNA and proteins. Additionally, we aim also to generate primary cell lines (primary melanoma, fibroblasts, pericytes, endothelial, immune cell lines) deriving from the primary tumors and we will extract the same type of samples (gDNA, RNA and proteins). This either out of bulk tumor pieces or cell suspension of FACS-sorted specific cell populations. Tumor pieces, organs & cells will be harvested for (m)IHC, (m)IF and (m)FISH staining experiments, either flash frozen for cryosections or embedded for paraffin sections, and fixed cells.

Crispr/Cas9 experiments (WP1.1):

A digital vector map and a short schematic representation of the construction of every vector is provided as a .gbk and pptx-file, using CLC and SnapGene programs. Databases with the DNA vectors and primers are kept as .xlsx files and schematic representations as .pptx files. All files have a size <10MB.

Sequencing files (WP1.1, WP1.2, WP1.3):

Sequencing data are returned from the sequencing facilities as fastq-files. Average size of these files ranges from 500MB per cell for single cell sequencing up to 3GB for bulk sequencing and 20GB for 10X sequencing.

Scripts & codes (WP1.1, WP1.2, WP1.3 WP3.1):

Scripts that are generated to process the sequencing or proteomics data are saved as .r-files or .sh-files (<1MB). Output files for scripts are .pdf (100-500KB), table files in .txt-format (500KB), Seurat objects .rds (1-2GB), R mark down reports .rmd (10-50KB), and SpatialIDE objects .gz, .jpg, .png. .json, .csv (|1-2GB).

Imaging (WP1.2, WP1.3, WP2.1, WP3.2):

Acquired images & processed images of tumor sections after (m)IHC, (m)IF and (m)FISH staining experiments will be saved in different formats depending on the microscope that will be used for the image acquisition and software processing. (.tiff, .jpeg, .czi. or .nd2-files). The size of the files ranges from 1GB-10GB for single plane images of from 10GB-50GB for multiple plane images.

Flow cytometry & FACS sorting (WP1.1, WP1.2, WP1.3, WP3.1):

Flow cytometry data will be generated .fcs-files of an average file size of 10-50 MB and stored. The data will be processed using FlowJo software and exported as .tiff image files or .xlsx-tables (<5MB). FACS-sorted cells will be collected to extract RNA or proteins.

Proliferation, self-renewel, and drug perturbation experiments (WP3.2):

Co-culture, survival drug perturbation experiments will be performed on the Incucyte. Data is exported as an comma-separated .txt-file with a size <1MB. Fotographs and videos are exported as .tiff (2-5MB) and .mp4 (5-10MB) files respectively.

3. Legal and ethical issues

Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- No

Privacy Registry Reference:

Short description of the kind of personal data that will be used:

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, add the reference to the formal approval by the relevant ethical review committee(s)

- Yes

GEMMs, PDX and allografts ECDs are currently in writing process before approval by the animal facility of KU-Leuven. This will allow us to perform the in vivo lineage tracing experiments delineated in WP1.1, WP1.2, WP1.3, WP2.1,WP3.1 WP3.2.

Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

- Yes

We do not exclude that the proposed work could result in research data with potential for tech transfer and valorization. As the promoter of this grant is a member of VIB, VIB has a policy to actively monitor research data for such potential. If there is substantial potential, the invention will be thoroughly assessed, and in a 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 publications need not be delayed.

Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- No

4. Documentation and metadata

What documentation will be provided to enable reuse of the data collected/generated in this project?

All experiments will be performed from an experimental design document (.docx or .xlsx) that details:

- The hypothesis
- The different conditions/experimental endpoints
- The metadata linked to the experiment
- The link to the Standard Operating Procedure (SOP) or published methodology used as protocol. Any experiment-specific changes or optimization will be
- Any other supporting files (schemes, tables, images, …) or intermediate data output will be stored alongside the data result files
- Software type, version and analysis type when applicable (e.g. FlowJo for flow cytometry data, R for scripts, qBase+ for (RT-)qPCR analysis, CLC and SnapGene for cloning analysis, etc…)
- Device settings when applicable and not recorded in the data output (e.g. PCR protocol for (RT-)qPCR)
- Storage conditions (and location) of physical samples

All types of ethical clearance for animal experiments will be stored in dedicated folders on the J-drive.

Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- Yes

For following type of experiments, the below 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 files using Fiji and QuPath software

For all other types of experiments no specific metadata standard will be used. Following metadata (non-limiting list) will be recorded, where applicable:

- Image metadata will be exported as OME-XML files using Fiji and QuPath softwares
- Date
- Investigator
- Unique identifier
- Link to related identifier
- Type of experiment: In vivo/In vitro/Bio-informatics
- Source of data: Animal/Cell lines/Database
- Type of material generated: DNA/RNA/Protein/digital
- Location of physical material (digital database)
- Device & details of programs ran
- Compounds, formulation & application
- Applicable ethical form/study protocol/informed consent

Metadata for sequencing data sets will be defined by the public database they will eventually be released in.

5. Data storage and backup during the FWO project

Where will the data be stored?

The data will be stored on the university's central servers with automatic daily back-up procedures.

How is backup of the data provided?

The data will be stored on the university's central servers with automatic daily back-up procedures.

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

Storage space on the J-drive and L-drive are unlimited and can be expanded upon simple request.

Upon closure of the project (or intermediately if required), all data will be transferred to the archive on the K-drive for long term storage. Data placed on the K-drive is more strictly secured with only very specific members of the lab having the authority to place data on the K-drive and only the head of lab has the authority to have data removed by the IT department.

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

L-drive: 128,38 euro/TB/year in blocks of 5TB K-drive: 128,38 euro/TB/year in blocks of 5TB J-drive: 519euro/TB/year

All expenses for storage are shared among all funded projects in the lab.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

All lab members have access to the university's central servers from their KUL IT-configured computers, either directly on the KUL network or with remote access via Pulse Secure. Access to the drives requires login with personal credentials. All folders on the J-drive and L-drive are accessible lab members only. In case transportation of the files for home use on non-IT-configured computers or to share with external collaborators is required, the lab is equipped with encrypted hard drives. For sharing of data with KULeuven internal members, the ICTS supported program BOX will be used as an intermediate.

6. Data preservation after the FWO project

Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).

All digital data will be stored for the expected 10year period. Of the physical data only mouse sperm, (modified) cell lines & DNA vectors (as DNA or glycerol stocks) can be cryopreserved in the lab or at the KULeuven biobank for long term storage. All other physical data will be removed over the course of 1-5years after completion of the project (gDNA, RNA, proteins) due to limited preservation possibilities.

Where will the data be archived (= stored for the longer term)?

Upon closure of the project, all digital data will be transferred to the archive on the K-drive for long term storage. Data placed on the K-drive is more strictly secured with only very specific members of the lab having the authority to place data on the K-drive and only the head of lab has the authority to have data removed by the IT department. The physical data in the form of mouse sperm, (modified) cell lines & DNA vectors will be cryopreserved in the lab or at the KULeuven biobank.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

Digital storage on the K-drive is expected to cost 1200 euro per year (at current rates) for 10TB.

Cost of physical storage at the KULeuven Biobank is currently still unclear. Costs of storage will be carried by the labs future grants and VIB dotation

7. Data sharing and reuse

Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

- No

No restricting factors foreseen but sharing of data might be temporarily postponed due to IP potential.

Which data will be made available after the end of the project?

All omics data and code & scripts will be made publicly available. Other data can be made available upon request.

Where/how will the data be made available for reuse?

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

1. Omics data sets will be made available on the NCBI Gene Expression Omnibus (GEO) or the EBI ArrayExpress databases.
2. Code & scripts will be made available on GitHub.
3. Other data can be made available upon request by mail after signing an MTA.

When will the data be made available?

- Upon publication of the research results

Who will be able to access the data and under what conditions?

The full dataset of sequencing experiments, proteomics experiments and the codes & scripts used for their analysis will be uploaded in their respective databases 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. All other data can be made available upon request by mail after signing an MTA.

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

Costs of data sharing are minimal (<100euro/year) and will be financed by the lab.

8. Responsibilities

Who will be responsible for data documentation & metadata?

The implementation of data documentation & metadata is the responsibility of the promoter, Professor Jean-Christophe Marine.

Who will be responsible for data storage & back up during the project?

All decisions on data storage & back up will be made by the promoter, Professor Jean-Christophe Marine.

Who will be responsible for ensuring data preservation and reuse ?

All decisions on data preservation & sharing will be made by the promoter, Professor Jean-Christophe Marine.

Who bears the end responsibility for updating & implementing this DMP?

The end responsibility for updating and implementing the DMP is with the promoter, Professor Jean-Christophe Marine.