DMP title

Project Name (Re)defining the origin and cellular niche of melanoma stem cells - DMP title

Project Identifier FWO G070622N

Grant Title G070622N

Principal Investigator / Researcher Chris Marine

Project Data Contact flavie.luciani@kuleuven.be

Description It has been suggested that only a restricted number of cancer cells, known as cancer stem cells (CSCs), are fated to fuel tumour growth. As a result, these specialised cells have become the targets of intense drug discovery efforts. In melanoma, however, whether CSCs exist has remained unclear. Combining single-cell RNA sequencing with in vivo lineage tracing and mathematical modelling, we identified a population of melanoma cells that may function as CSCs. We propose herein to firmly establish their functional contribution to primary and metastatic growth using genetic lineage tracing and ablation experiments. Importantly, based on (our own) recent findings, we hypothesise, that most, if not all, melanoma cells can acquire these â€œdangerousâ€ CSC properties when exposed to appropriate cues. We therefore also propose to define the spatial distribution of this newly identified stem cell population, map their cellular niche and identify intrinsic and extrinsic mechanisms of tumour-stromal cell interactions that regulate their identity and cell reprogramming ability (i.e. plasticity). We will then test whether pharmacological interventions that target key niche-dependent CSC specification mechanisms can restrict tumour growth. This project will pave the way for the development of innovative therapies that contend with, or even exploit, the â€œchameleonicâ€ behaviour of cancer cells. We propose herein to: (i) determine the functional contribution to these putative melanoma CSCs to primary and metastatic growth using genetic lineage tracing and perturbation experiments (i.e. lineage ablation); (ii) dissect the intrinsic and extrinsic mechanisms underpinning their function(s) and potential reprogramming capacity (i.e. plasticity); (iii) define their spatial localization, map their cellular niche, identify mechanisms of stromal-tumor cell interactions that involve either direct cell-cell contact and/or secreted ligand-receptor interactions, which impact on the activity of downstream transcription factor(s) (TFs) that, in turn, regulate their identity and plasticity; (iv)investigate whether pharmacological interventions that target CSC niche-dependent specification mechanisms can restrict primary and/or metastatic tumor growth.

Institution KU Leuven

1. General Information

Name applicant

Pr. Jean-Christophe MARINE

u0063154

jeanchristophe.marine@kuleuven.be

FWO Project Number & Title

G070622N

(Re)defining the origin and cellular niche of melanoma stem cells

Affiliation

- KU Leuven

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) (Aim 1, 2, 3):

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 (Aim 1, 2, 3):

From the mouse models in both GEMM, allografts and PDX models, we expect to extract gDNA, RNA and proteins. This either out of bulk tumor pieces or cell suspension of FACS-sorted specific cell populations. RNA and proteins will be stored at -80°C, gDNA at 4°C (short term) or at -20°C (long term). Physical materials of human origin will also be deposited at the KULeuven BioBank.

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. Cryosections will be stored at -20°C and paraffin embedded samples can be stored at RT.

Imaging (Aim 1, 2, 3):

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.

We will generate graphs and illustrations to visualize the omics data using R or R-studio (exported as .pdf, 0.1-10MB) or Graphpad (exported as .pdf, 0.1-1MB). Data illustrations will be further processed for publication using Inkscape (exported as .svg or .pdf, 1-15MB) or Adobe Illustrator (exported as .ai or .pdf, 1-15MB).

Digital images will be stored in the following formats:

- Digital images in raster formats: uncompressed TIFF (.tif/.tiff), JPEG (.jpg), Adobe Portable Document Format (.pdf), bitmap (.bmp), .gif
- Digital images in vector formats: scalable vector graphics (.svg), Adobe Illustrator (.ai), Adobe Portable Document Format (.pdf)

Sequencing files (Aim 1, 2, 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.

Genomic and transcriptomic data (single-cell RNA-sequencing (scRNA-seq; raw data (.fastq file; 10 GB per sample), single-cell T cell receptor-sequencing (scTCR-seq; raw data (.fastq file; 20 GB per sample), full or single-nucleus genome-and-transcriptome-sequencing ((sn)G&T-seq; file type, size) and bulk whole-exome sequencing (WES); .fastq file 10 GB + .bam file 10 GB: ~20 GB per sample) from tumour samples will be generated locally using Illumina HiSeq4000 and NovaSeq6000 machines, producing files in .fastq or .bcl format.

Data will be stored in the following formats:

- Next generation sequencing raw data: binary base call format (.bcl), .fastq
- Sequence alignment data: .bam
- Structural variations data: .vcf, .bcf
- Read/UMI count data: .tsv, .rds

Scripts & code (Aim 1, 2, 3) :

While the project is ongoing, scripts and results are saved in ‘Staging’ HPC cluster storage and locally on personal computers with copies on the J-drive. At the end of a project, only the scripts, important matrices and plots associated with the project will be and all intermediate files will be removed, but can easily be recreated from the saved scripts. Scripts 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) and R mark down reports .rmd (10-50KB).

Research methods and practices (SOPs) will be fully documented and published along with the results in one manuscript (.pdf, 20MB). When the wet lab techniques, scripts, algorithms and software tools are finalized, they will be additionally described in manuscripts and/or on GitHub, accompanied by a README (.txt, 10KB) file containing all the associated information.
Data will be generated in the following formats:

- Text files: Plain text data (Unicode, .txt), MS Word (.doc/.docx), Adobe Portable Document Format (.pdf) format

Co-culture & survival experiments (Aim 3):
Co-culture and survival 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.

Flow cytometry & FACS sorting (Aim 3):
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.

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

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

Utilisation of some GEMMs will be done on ECD already approved P125/2021.
PDX and allografts ECDs are currently in writing process before approval by the animal facility of KU-Leuven.
Harvesting human samples by TRACE to establish PDX models (prof. Leucci): UZLeuven S63799
Establishing PDX models by implanting tumor fragments (prof. Leucci): KULeuven P164/2019
Any further experimental work on animals or human samples are awaiting approval or an approval request will be submitted before we begin the experiments.

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?

Storage of small to intermediate size data is done on the KULeuven J-drive. The J-drive has the

benefit of securing the data with restricted access where necessary and a daily backup system with recorded history that allows the restoration of data in case of accidental removal/modification.

Large size data (sequencing data, proteomics data, imaging files) is stored on the KULeuven L-drive. The L-drive has the same characteristics as the J-drive but is especially designed to handle large data files like sequencing files or imaging files. Total combined storage required for this project is estimated at 50TB.

While the project is ongoing, scripts used to process sequencing files and their output results are saved in 'Staging' HPC cluster storage and locally on personal computers with copies on the J-drive. At the end of a project, only the scripts, important matrices and plots associated with the project will be and all intermediate files will be removed, but can easily be recreated from the saved scripts.

The sequencing data generated during the project will either be stored on VIB-KU Leuven servers or on the Flemish Supercomputer Centre (VSC), initially in the staging and archive area, and later only in the archive area (mirrored).

Raw and processed data will be submitted to a public repository (e.g. EGA) with appropriate access control if required, to enable sharing and long-term validity of the data.

Fresh tissue, single-cell suspensions, nucleic acid samples resulting from (single-cell) nucleic acid isolation and sequence library preparations will be stored in in labeled tubes or SBS plates in boxes labelled per clinical trial, in -20°C or -80°C freezers purchased by our own funding. Cryoblocks of tumor tissue will be stored at -20°C, paraffin blocks of tumor tissue will be stored at room temperature.

How is backup of the data provided?

KU Leuven drives are automatically (daily) backed up using KU Leuven services 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.
- Incremental backups are done daily from one 20 TB QNAP NAS to a second 20 TB QNAP NAS.

All omics data stored on the Flemish Supercomputer Centre (VSC) will be transferred on a regular basis to the archive area which is backed up.

Data is stored on EGA for the purpose of data sharing.

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?

OneDrive	J-drive (100Gb)	K-drive (100Gb)	L-drive (5Tb)	KUL Ent. Box (100Gb)	
Price	Free*	€51,9	€11,384	€569,2	€10
Period		Yearly	Yearly	Yearly	Yearly
Conditions		Only stored data is charged	All space is charged	All space is charged	All space is charged

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 and can be further restricted in case of sensitive/personal data.

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.

Sequencing data will not be shared with any internal or external collaborators.

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 10years period.

Of the physical data only PDX models, (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)?

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 or preprint deposition.

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 PDX models, (modified) cell lines & mouse tissues 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

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

All sequencing 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
- Other (specify):

Whenever possible, datasets and appropriate metadata will be made publicly available through repositories that support FAIR data sharing. Personal data will be double coded and no reference to subject name will be made. Data will be reused by transfer via Belnet Filesender or secure copy. Sharing policies for specific research outputs are detailed below:

- Double-coded raw sequencing data (linked to double-coded clinical patient data) will be deposited in open access repositories with restricted access control such as the EBI European Genome-phenome Archive (EGA). The EGA is a repository for personally identifiable genetic and phenotypic data. Sequencing data at EGA will only be available upon reasonable request via our institutional data access committee and if necessary a material transfer agreement will be concluded with the beneficiaries in order to describe the types of reuse that are permitted.
- All protocols used to generate published data will be described in the corresponding manuscript(s), and the related documentation will be included as supplementary information. These data and all other documents deposited in the (electronic) notebook are accessible to the researchers, and will be made available upon request.
- All scientific publications will be shared openly. Manuscripts submitted for publication will be deposited in a pre-print server such as bioRxiv. At the time of publication, research results will be summarized on the institutional website (https://www.vibcancer.be) and post-print pdf versions of publications will be made available there if allowed by copyright agreements, possibly after an embargo as determined by the publisher. Before the end of the embargo or in cases where sharing the post-print is not allowed due to copyright agreements, a pre-print version of the manuscript will be made available. (Pre-print) publications will also be automatically added to our institutional repository, Lirias 2.0, based on the authors name and ORCID ID.
- All the relevant algorithms, scripts and software tools driving the project will be described in manuscripts and/or on GitHub (https://github.com) accompanied by a README.txt file containing all the associated information and/or on our interactive webserver (http://blueprint.lambrechtslab.org).
- Extra data that do not support publication will be either deposited in an open access repository or made available upon request by email.

When will the data be made available?

- 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 (or preprint deposition). 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 do not need to 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.

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

The full dataset of sequencing 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 PI, Pr. Jean-Christophe Marine, bears the end responsibility of updating & implementing this DMP.