DMP title

Project Name Understanding and tackling T cell dysfunctions towards improving colorectal cancer therapies.(FWO DMP) - DMP title

Grant Title 1SF0422N

Principal Investigator / Researcher Allyson Peddle

Project Data Contact Allyson Peddle

Description Although recent advances in checkpoint-based immunotherapies have revolutionized treatments for many cancers, it has failed in microsatellite stable (MSS) colon cancer (CRC), which accounts for 90% of all colorectal tumors. Single cell RNA sequencing data by the Tejpar group led to identification of defects both in CD4 and CD8 T cells in MSS CRC which may indicate that ineffective priming and activation of T cells by dendritic cells underlies the lack of cytotoxic CD8 T cells and response to immunotherapy. The Tejpar group showed heterogeneity across tumor subsets where most MSS lack Th1 and CD8 T effector cells, but for instance CMS2 and CMS4 are enriched in Th17 and Treg, respectively, pointing to imbalances in CD4 and CD8 T cells at different levels per subtype. This heterogeneity should not be ignored when attempting to drive T cell balances to favorable Th1 and cytotoxic CD8 responses. In addition, successful activation of cytotoxic T cells can also be hampered by the basal tolerogenic interactions in the colon during homeostasis, implying the importance of studying the disease orthotopically. Therefore, I aim to characterize and manipulate CD4 and CD8 T cell (dys)functions in orthotopic patient-matched CRC models, which can suggest novel therapeutic interventions. Moreover, will provide clinically relevant colorectal tumor models with a robust interrogation toolset for phenotyping and tracking of tumor-specific immune responses.

Institution KU Leuven

1. General Information

Name applicant

Allyson Peddle

FWO Project Number & Title

Understanding and tackling T cell dysfunctions towards improving colorectal cancer therapies.

1SF0422N

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).

In this project, I plan to collect genomics/transcriptomics (single cell and bulk), peptidomics and immunophenotypic data for MC38, KPN, BMT, AKPT cell lines in CRC CMS specific models. Moreover, we will identify novel and CMS model specific epitopes serving as very valuable tools in future. From these, we will
build correlations and aim to generate appropriate manuscripts, describing the insights on the native MC38 as well as other cell line models, the effect of treatments on this model and in collaboration with Prof Sansom, we will describe unique CMS models. For translation of mouse to human research, this project has
access to the group's extensive bulk and scRNAseq database.

Type of data	Format	Volume	How Created
Single cell genomics data	fastq, bam	3-10TB	10X Genomics technology & Illumina sequencing
Multi-region whole exome sequencing (Illumina)	fastq, bam	3-10TB	Illumina sequencing
bulk RNA sequencing	fastq, bam	3-10TB	Illumina sequencing
Immunopeptidomics	fastq, bam	3-10TB	Proteins will be islodate from fresh-frozen tissue & MHC1/MHC2

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

ECD Approval has been granted: P194/2021

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 we are in close
contact with the group of O. Samson, Janssen Pharmaceutics and the SIB (swiss institute for bio-informatics) for this project.

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?

- Yes

A data access agreement is in place. This agreement relates to the primary sequencing data and any secondary analysis results. Any public
disclosure on new data on this project needs to follow a disclosure process.

4. Documentation and metadata

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

For the survey data a codebook will be generated containing study design, sampling

methodology, variable-level information (lael, question
text, codes, frequencies). All of the metadata of the different cohorts will be stored and encoded in an electronic labbook / excel.
For further analysis: Software, algorithms and scripts will be commented and extensively documented, e.g. using Jupyter Notebooks and R
Markdown. Software will be version-controlled and tracked using git via version numbers, tags and commit IDs (hashes). When scripts,
algorithms and software packages are finalized, they will be made publicly available (e.g. via GitHub) and additionally described in their
documentation and in manuscripts.

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.

I plan to use a metadata standard, basing myself on different available sources, such as http://www.dcc.ac.uk/resources/metadatastandards;
http://rd-alliance.github.io/metadata-directory/standards/; https://fairsharing.org/
Further, I will store metadata from experiments in a searchable database format. These will include the following search parameters: data,
initials of the investigator, type of the study: which cohort, phenotyped mice, compound used.

5. Data storage and backup during the FWO project

Where will the data be stored?

All necessary measures will be taken to archive all biological material according to good scientific practices. The researcher data and
remarks are stored in personal (electronical) notebooks. All the raw sequencing data as well as secondary analysis results are stored both
university's central servers and the vlaamse supercomputer (VSC). This includes automatic daily back- up procedures. After publication, all
supporting data is deposited into an access-controlled public repository such as the European Genome-Phenome Archive (EGA).

How is backup of the data provided?

All the raw sequencing data as well as secondary analysis results are stored both university's central servers and the vlaamse
supercomputer (VSC).

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

The KUL offers a high performance primary storage, along with a backup system. These are expandable in blocks, the backup capacity is
technically not an issue.

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

The L-drive (big storage capacity drive of the KUL) charges 570 euros for every 5TB. These costs are covered by credits of the DIO-lab.

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

Anonymized data is stored on the VSC protected by their security policy. Measurements are taking to ensure the security of the data storage
and to prevent modification by unauthorized persons.

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 raw and processed data are stored for a minimum of 5 years after the project.

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

Metadata and smaller data files 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. Large raw data and processed sequencing data objects are stored longterm on the Servers of the KU
Leuven (L-drive).

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

Digital data storage costs are included in general lab costs. The total storage cost is estimated to be at 600€/TB and 50€/TB backup.

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)?

- Yes. Specify:

Prior to publication, data sharing is restricted to members of our research group. After publication, pseudonymised data will be deposited in
the databases indicated above and access for research purposes can be granted.

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

All research outputs supporting publications will be made 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.

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

- In an Open Access repository
- In a restricted access repository

After publication, raw sequencing data and analysis results are deposited into an access controlled public repository such as the European
Genome-Phenome Archive (EGA). Developed software, algorithms and scripts will be made available and stored on a public repository
such as GitHub.

When will the data be made available?

- Upon publication of the research results

All research outputs will be made accessible at the latest at the time of publication. No embargo is foreseen except when imposed e.g. by
publishers for papers in press, potential IP requirements, or ongoing projects requiring data confidentiality. In those cases, data will be made
available as soon as the embargo is lifted.

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

Third parties can apply for access to the whole or subsets of the pseudonymised data submitted to the European Genome-phenome Archive
(EGA) by applying. Data transfer will be covered by a Data Access Agreement (DAA), specifying its conditions of use.

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

Costs for data sharing will be minimised by leveraging free-to-use data repositories and dissemmination facilities where possible. The
European Genome-phenome Archive (EGA) is part of the European Molecular Biology Laboratory (EMBL), and as such, is funded largely
by the governments of EMBL's member states.

8. Responsibilities

Who will be responsible for data documentation & metadata?

(Meta)data will be documented by the researcher at the time of analysis and compiled at regular intervals as well as when preparing for
publication.

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

The researcher and technical staff ensure secure and backed-up data storage.

Who will be responsible for ensuring data preservation and reuse ?

The (co-)supervisors and the researchers are responsible for data preservation and sharing, with support from scientific computing.

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

The PI bears the end responsibility of updating & implementing this DMP.