DMP title

Project Name C1 DMP - DMP title
Project Identifier C14/21/103
Grant Title C14/21/103
Principal Investigator / Researcher BjÃ¶rn Meijers
Project Data Contact bjorn.meijers@uzleuven.be
Description In this translational project, we study the gut barrier as master regulator of microbiome-host interactions in IBD and CKD patients. We will examine the structure and function of paracellular and transcellular transport in the large intestine.
Institution KU Leuven

1. General Information

Name of the project lead (PI)

Prof. Dr. B. Meijers

Internal Funds Project number & title

Project number:

C14/21/103

Project title:

The gut barrier as master regulator of microbiome-host interactions in IBD and CKD patients: exploring the structure and function of paracellular and transcellular transport in the large intestine.

2. Data description

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

- Generate new data
- Reuse existing data

2.2. What data will you collect, generate or reuse? Describe the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a numbered list or table and per objective of the project.

Type of data	Format	Volume	How created
RNA sequencing	.count .xls .tif	20 GB	RNA sequencing data of colon cells/organoids after stimulation with selected compounds
Collection of biopsies and/or feces	.xls	1 GB	Collection, processing and storage of biopsies and or feces
RT-qPCR data	.xls .eds	1 GB	RT-qPCR data of organoids after stimulation with selected compounds
16S rRNA sequencing	R file TSV file	1 GB	16S rRNA sequencing of microbiota samples
Western blot data	.tiff	1 GB	Pictures of western blot and analysis
Barrier measurements (Transepithelial electrical resistance)	.xls	1 GB	Measurements of TEER of epithelial cells
Clinical metadata	.xls	1 GB	Collection of clinically relevant metadata encoded in a pseudonymised manner
Metabolomics data	.raw .xls	5 GB	LC/MS-MS

3. Ethical and legal issues

3.1. Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to the file in KU Leuven's Record of Processing Activities. Be aware that registering the fact that you process personal data is a legal obligation.

Personal data will be used. We will use the 'klinisch werkstation' (KWS, Nexuzhealth) as source data. Study participants will have their study number recorded in the KWS. For new studies, we will use REDCAP as eCRF. The study ID will be stored separately from pseudonimized clinical data in Redcap. We will store the following pseudonimized data: demographics, markers of kidney function, drug therapy, coded past medical history, smoking status, bowel disorders.

3.2. 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).

We will use data and tissue obtained from patients. Two protocols have already received ethics approval (S53684 for organoid experiments and sampling of colonic tissue/faeces from IBD patients; S62454 for colon biopsies from patients having CKD).

3.3. Does your research possibly result in research data with potential for tech

3.3. Does your research 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?

It is not anticipated that IPR will be claimed on the data generated. In the event that there surges any
potential for IPR, rapid interaction with LRD will be initiated.

3.4. 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 regarding reuse and sharing are in place?

There are no third party agreements in place restricting dissemination or exploitation of the data.

4. Documentation and metadata

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

A readme file will be provided with each of the datasets (clinical and experimental data on clinical
samples).
Raw expression data files, and image files will be collected per sample. A metadata file will be provided
with the clear description of what the raw data files represent and how they were generated. This metadata file will be kept in the same folder as the expression data.
Research methods will be fully documented as word files.

4.2. Will a metadata standard be used? If so, describe in detail which standard will be used. If not, state in detail which metadata will be created to make the data easy/easier to find and reuse.

Currently the use of shared metadata standard(s) has not been implemented in the daily routine of the different research groups. Standard metadata formats will be explored during the project, e.g. using the Dublin core (http://www.dcc.ac.uk/resources/metadatastandards/dublin-core). In case we reach common ground on metadata standards during the course of the project, this will be reported at the end of the project.

5. Data storage and backup during the project

5.1. Where will the data be stored?

The master copy of the data will be kept on our research unit central storage facility. Copies of part of the data can be made for detailed analysis, but will not be allowed to stored on personal devices in case human data are involved. We will use the KU Leuven enterprise box to share ongoing research topics, allowing for multiple persons to work and collaborate around the same datasets.

5.2. How will the data be backed up?

The data will be stored on the centrally managed KU and UZ Leuven servers with automatic daily back-up
procedures and version tracking.

5.3. 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.

The current storage and backup capacity is sufficient for the clinical, genotype/transcriptomic and metabolomic/metagenomic data storage.

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

We anticipate a back-up cost per Tb (KU Leuven ICTS) of 295€/year (5 Tb anticipated).
In principle, we don't anticipate needing the KU Leuven ICTS digital vault, as all data with high privacy
risks (i.e. the codes to decode the pseudonymization) will be stored behind the firewall of UZ Leuven. The costs will be shared between the participating labs, and are part of the running costs of the labs.

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

All data 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.

6. Data preservation after the end of the project

6.1. Which data will be retained for the expected 10 year period after the end of the project? If only a selection of the data can/will be preserved, clearly state why this is the case (legal or contractual restrictions, physical preservation issues, ...).

The pseudonymized data will be preserved for at least 20 year period after end of this project. The
generated data will be stored on designated KU and UZ Leuven servers.

6.2. Where will these data be archived (= stored for the long term)?

The generated data will be stored on designated KU and UZ Leuven servers (KU Leuven Enterprise Box,
KU Leuven Large Volume Storage, UZ Leuven server).

6.3. What are the expected costs for data preservation during these 10 years? How will the costs be covered?

Long-term data storage in the UZ Leuven environment is guaranteed by UZ Leuven IT department. The long-term cost is shared between collaborators.

7. Data sharing and re-use

7.1. 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 or because of IP potential)?

- Pseudonymized (coded) data will only be shared with third parties after agreement and signing a
dedicated Data Transfer Agreement or Material Transfer Agreement.
- Fully anonymized and aggregated data can be shared in Open Access repositories, if necessary upon
publication, or upon request by third parties.

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

- Pseudonymized (coded) data will not be shared, unless a proper MTA/DTA is in place. This implies that
pseudonymized data will not be made public, also not after the end of the project.
- Publicly relevant anonymized or aggregated datasets could be made available during or after the end of
the project.

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

- Pseudonymized (coded) data can only be made available after a signed DTA/MTA is in place.
- Anonymized data can be made available in Open Access repositories (e.g. Gene Expression Omnibus
for gene expression data; other formats relevant for other type of datasets).

7.4. When will the data be made available?

- Upon publication of the research results

Data will be made available upon request, after fulfilling all necessary obligations.

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

Pseudonymized data will only be made available to third parties after a signed DTA/MTA is in place.
Relevant anonymized or aggregated data will be made available publicly, upon publication or upon
individual request. Public data can then be accessed by anyone.

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

Usually, data repository in Open Access repositories like Gene Expression Omnibus or ArrayExpress is free of cost. If
there are any costs associated with data sharing to third parties, the costs of this data transfer will be
negotiated in the DTA/MTA.

8. Responsibilities

8.1. Who will be responsible for the data documentation & metadata?

The principal investigator and two co-investigators will share responsible for data documentation and metadata, each for the part that is generated under their supervision.

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

The PI of the project will be the final responsible for data storage and back up during the project.

8.3. Who will be responsible for ensuring data preservation and sharing?

The PI of the project will ensure data preservation and reuse. All requests for data sharing and reuse
should be directed to the PI of the study.

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

The end responsibility for updating and implementing the DMP is with the supervisor (promotor).