Plan Overview

A Data Management Plan created using DMPonline.be

Title: Host-microbiome interactions through IBD-on-a-chip

Creator: Kaline Arnauts

Principal Investigator: First Name Surname

Affiliation: KU Leuven (KUL)

Funder: Internal funding

Template: KU Leuven BOF-IOF

Principal Investigator: First Name Surname

Project abstract:

Ex vivo research models mimicking the complexity of multifactorial diseases, such as inflammatory bowel disease (IBD), are lacking. Alterations in immune cells, dysbiosis of the microbiota and disruption of the epithelial barrier form a dynamic interplay within IBD that requires further understanding. The use of non-patient and disease specific research models, hampers the latter and translation of findings into patients. The 3R principle, denoting Replacement, Reduction, and Refinement of animal models in an utmost important framework for scientific research. Intestinal organoids, 3D ex vivo models of the intestinal epithelium, are gamechangers by mimicking patientand disease-specific characteristics but lack physiological fluid flow and peristalsis. This issue can be overcome by the use of organ-on-a-chip models, in which organoids can be incorporated. Yet, IBD is a multifactorial disease including multiple factors such as immune cells, microbiota and epithelial cells and requires a recapitulation of this complex interplay. The aim of this research project is to establish a patient-specific IBD gut-on-a-chip model including immune cells, microbiota and epithelial cells on the Emulate® platform to further study their interplay. A deeper understanding of these interactions, and the establishment of this disease specific model will be a major step forward to dissect disease driving mechanisms.

ID: 211775

Start date: 01-10-2024

End date: 30-09-2025

Last modified: 18-12-2024

Host-microbiome interactions through IBD-on-a-chip

Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
		*Indicate: **N**(ew data) or **E**(xisting data)*	Indicate: **D**(igital) or **P**(hysical)	Indicate: **A**udiovisual **I**mages **S**ound **N**umerical **T**extual **M**odel **SO**ftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
Clinical data	Data retrieved from the patient records system (KWS platform UZ Leuven), stored in a pseudonymised manner.	E	D	T		<1GB	
Storage files -serum samples -mucosal biopsy/organoid - feces	Samples are being collected in a pseudonymised manner, and stored physically in UZ Leuven biobank using 2D barcode labels. Registration and storage information are being recorded in a standardized manner (ie. one file per sample type: serum, mucosal biopsy, derived organoids, feces). The samples for this project in particular are thus part of our general sample storage files.	E	D	T		<1GB	

Raw data	Transcriptomic analysis of epithelial cells, sequenced generated by the Genomics core (UZ/KU Leuven) secretomic analysis of epithelial cells, sequenced generated by MSD panels Immunofluorescent staining of chips Papp (barrier integrity measurements) Microbiome 16S sequencing	N N N N N	D D D D D	T T,I I T T		<100GB <100GB <100GB <100GB <100GB	
Processed data	Transcriptomic analysis of epithelial cells, sequenced generated by the Genomics core (UZ/KU Leuven) secretomic analysis of epithelial cells, sequenced generated by MSD panels Immunofluorescent staining of chips Papp (barrier integrity measurements) Microbiome 16S sequencing	N N N N N	D D D D D	T T,I I T T		<100GB <100GB <100GB <100GB <100GB	
Metadata	Overview file with a clear description of what the data represent, how they were generated, qualtiy control etc.	N	D	T		<1GB	
Script	Code that will transform raw data into processed data. Code that will transform processed data into results	N	D	T		<1GB	
Results	The outcome of the project including tables, figurers and text explaining those	N	D	T, I		<100 GB	
Lab Note books	Data written notes associated with carrying out experimental procedures	N	P	T		/	
Standard operating procedures	Written protocols for experimental procedures	N	D	T		<1GB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

Reuse existing patient data from patient files including age, gender, disease duration and therapy.
Personal data relating to the study participants including name and date-of-birth will be collected for identifier purposes. These personal data will only be available to researchers directly involved in the recruitment phase. For the remainder of the study, all derivative clinical parameters such as age, gender, disease duration and therapy will be coded, and thus pseudonymised. The file linking the code and personal identifiers will only be accessible to authorized individuals and stored in a restricted access, secure environment managed by the KU Leuven/UZ Leuven ICT facility.
Personal data collection is covered by the Ethical approval of S53684 (Ethical approval committee UZ Leuven)

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, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- Yes, human subject data (Provide SMEC or EC approval number below)

Personal data relating to the study participants including name and date-of-birth will be collected for identifier purposes. These personal data will only be available to researchers directly involved in the recruitment phase. For the remainder of the study, all derivative clinical parameters such as age, gender, disease duration and therapy will be coded, and thus pseudonymised. The file linking the code and personal identifiers will only be accessible to authorized individuals and stored in a restricted access, secure environment managed by the KU Leuven/UZ Leuven ICT facility.
Personal data collection is covered by the Ethical approval of S53684 (Ethical approval committee UZ Leuven)

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- No

The file linking the code and personal identifiers will only be accessible to authorized individuals and stored in a restricted access, secure environment managed by the KU Leuven/UZ Leuven ICT facility.

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- Yes

The project includes the development of a IBD-on-a-chip system. In a later phase, the developed model might be of interest for commercial valorization. We are currently having meetings with LRD regarding this topic and are discussing any potential for research valorisation.
After finishing the project and before dissemination of the results, we will look carefully if any valorisable material is present. Additional IP discussions might start from that point on in agreement with LRD.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or Data transfer agreements, Research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- Yes

In agreement with LRD, we are currently having meetings with Emulate (company from Gut-on-a-chip platform we are using) to discuss dissemination of data and research collaborations.

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- Yes

In agreement with LRD, we are currently having meetings with Emulate (company from Gut-on-a-chip platform we are using) to discuss dissemination of data and research collaborations.

Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, codebook.tsv etc. where this information is recorded).

Standard experimental procedures (SOPs) and practices are/will be fully documented as PDF and saved on the KUL Shared J- drive assigned to our group.The methodology and protocol will be described in detail in the physical lab book that will be stored at the lab at all times.
An accompanying key file (.xls) to decipher which result files match which protocol (which will reference the pages in the lab book) will also be available on a shared drive. Only members from the team will have access to these folders.
Data folders containing the raw data are being stored on our KUL Archive K drive. Data folders containing pseudonymized clinical data, processed data, metadata and scripts* are stored on our KUL Shared J- drive. Data folder names will always contain the date, type of experiment, and the name of the study cohort.
*Scripts will be commented and extensively documented, e.g. using Jupyter Notebooks and R Markdown. Used software will be version-controlled and tracked via version numbers in the scripts.

Will a metadata standard be used to make it easier to find and reuse the data?
If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- No

Metadata standard:

The biological samples registered in the biobank will contain metadata required by the royal decree of biobanking (9JAN2018), standardized metadata to trace the pre-analytical factors of the sample which are most likely to impact research results.

Where no metadata standard exists ,we do have a minimal set of requirements that will be followed in order to ensure standardization and possibility to reinterpret and reuse the data when necessary and permitted. All collected data will be labelled with (1) title, (2) author, (3) data type, (4) data created and date modified, (5) file size, (6) equipment reference (such as manufacturer and model identification). Depending on the nature of data additional metadata are collected.
Text documents and Excel files stored within each experiment folder will respectively contain guidelines describing data collection/analysis methods and all relevant metadata (including experimental conditions, quality control metrics, computational analysis pipelines and their parameters) to ensure the reusability of the data and the reproducibility of any further data generation
RNA,16s rRNA Sequencing:
All experimental data will be noted every day in the lab book including all detailed experimental data. Once the sequencing data will be available, all files and folders will be labeled in a clearly structured way. The explanation of the labeling and the performed analysis will also be written down in the lab book. For sequencing, sequencing depth and analysis cut-offs will also be noted.
Microscopy:
The labeling of every individual picture taken will contain the patient culture number, the passage number, the experimental stimulation and the magnification.
Papp measurements
Every sampling will be noted every day and contain all experimental information (patient culture number, passage number, experimental stimulation). Analysis will be performed in a platereader and stored in excel files. The excel file will also contain all experimental data and refer to the according pages in the labbook.

Biobank

All patient information (age, gender, disease) will be registered in an psuedonymized way in the file containing all collected samples. Every patient will receive an identification nummer which can only be decoded by the responsible data manager.

MSD: all information on the samples (plate layouts that indicate which samples will be used) and the used kits will be registered in the lab book and electronically in .xls (which will go into the shared drive). The protocol will be stored in the lab book as well. The data analysis will be done in R for which the script (that also specifies every step) will be made available in the shared folder.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- OneDrive (KU Leuven)
- Shared network drive (J-drive)
- Other (specify below)

Our data will be stored on KU Leuven administered drives (Archive K drive storage, Shared J drive and OneDrive KU Leuven). For some data analyses, some raw data will need (temporarily) to be stored on the encrypted local PC hard drive (analyses from a non-local source are too slow and lead to computational failures).
Paper lab notebooks will be kept in locked closets in the labs of the PI.

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution

Since the data are stored on KU Leuven storages drives, the general ICT back-up Policy is applied.

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

Since the data are stored on KU Leuven storages drives, and these drives are expandable in blocks, the backup capacity is technically not an issue.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

Access to KU Leuven administered drives is conditioned by KU Leuven security groups. Data concerning patient information stored in excel files will be password protected and only the responsible researchers will have access. Furthermore, the raw data are stored on the archive K drive with (1) limited access (only a limited set of people have access) and (2) an overwrite and delete protection (based on read-write access) in order to prevent accidental loss of these data.
Hard copies of the Informed Consent forms and paper lab notebooks are kept in locked cabinets in the lab of the PIs.
Access will be controlled by PI determined access rights mediated by password protection and customised read/write permissions

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

1 TB is available on the central archive K drive of our research group (€100/year) – for this project we will need < 0,5 TB for raw data storage
0,5 TB is available on the central Shared J-drive of our research group (€500/year) – for this project we will need < 0,5 TB for pseudonymized clinical data, processed data, metadata, scripts, SOPs, and results.
When needed, these drives are expandable in blocks and funding to cover the costs is available in our group

Data Preservation after the end of the Research Project

Which data will be retained for 10 years (or longer, in agreement with other retention policies that are applicable) after the end of the project?

In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

- All data will be preserved for 10 years according to KU Leuven RDM policy

Where will these data be archived (stored and curated for the long-term)?

- Large Volume Storage (longterm for large volumes)
- Shared network drive (J-drive)
- Other (specify below)

The data will be stored on the university's central servers (with automatic back-up procedures) for at least 10 years, conform the KU Leuven RDM policy.
Hard copies of the Informed Consent forms, and paper lab notebooks are kept in locked cabinets in the lab of the PI.

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

The cost of archival on KU Leuven servers is estimated to be between 50 and 75 EUR for the 5 years after project end. Funding is available in our group to cover these costs.

Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project?
Please explain per dataset or data type which data will be made available.

- Yes, as open data
- Yes, as restricted data (upon approval, or institutional access only)

In open Access repository:
In case of sequencing data, these datasets will be deposited to NCBI. Applied codes can be made available on Github

Upon request by email
Data is stored in the central server of KU Leuven and will be available upon request at least 5 years after the project. The information regarding this data can be found in the publications related to the project and the responsible PI will provide the requested data. All patient-related information is protected by the UZ Leuven.
If the participants have allowed that their data can be reused, other researchers can ask for the

data. The data will be provided using a secure medium, e.g. the filesender of Belnet.

If access is restricted, please specify who will be able to access the data and under what conditions.

Whenever possible, datasets and the appropriate metadata will be made publicly available through repositories. As detailed above, metadata will contain sufficient information to support data interpretation and reuse. These repositories clearly describe their conditions of use.
For data shared upon request, a data transfer agreement will be concluded with the involved parties in order to clearly describe the types of reuse that are permitted.

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

Please explain per dataset or data type where appropriate.

- Yes, intellectual property rights

We are currently discussing potential IP with LRD and will make a decision in agreement with them before publishing or submitting any data.

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- Other data repository (specify below)
- Other (specify below)

All data obtained can be available after publication (upon agreement from LRD regarding IP aspects).

In open Access repository:
In case of sequencing data, these datasets will be deposited to NCBI. Applied codes can be made available on Github

Upon request by email
Data is stored in the central server of KU Leuven and will be available upon request at least 5 years after the project. The information regarding this data can be found in the publications related to the project and the responsible PI will provide the requested data. All patient-related information is protected by the UZ Leuven.
If the participants have allowed that their data can be reused, other researchers can ask for the data. The data will be provided using a secure medium, e.g. the filesender of Belnet.

When will the data be made available?

- Upon publication of research results
- Other (specify below)

Upon approval from LRD.

Which data usage licenses are you going to provide?

If none, please explain why.

- Other (specify below)

Data usage licences will be discussed with LRD before any licences are granted. Similarly, when DTAs or MTAs are discussed, this will always be after consulting and collaborating with LRD.

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- Yes, a PID will be added upon deposit in a data repository

Depending on the data repository and the type of data that would be made available, a unique identifier will be added to the data set.

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

None, the filesender of Belnet is for free.

Responsibilities

Who will manage data documentation and metadata during the research project?

The researcher (Kaline Arnauts) and PI (Séverine Vermeire) are responsible for data documentation & metadata

Who will manage data storage and backup during the research project?

The researcher (Kaline Arnauts) and PI (Séverine Vermeire) are responsible for data storage & back up during the project

Who will manage data preservation and sharing?

The PI (Séverine Vermeire) is responsible for ensuring data preservation and reuse

Who will update and implement this DMP?

The PI (Séverine Vermeire) bears the end responsibility for updating & implementing this DMP. The DMP will be evaluated at regular meetings between the researcher and the PI during the project