DMP title

Project Name LIPIDOMIC, MICROBIAL AND FUNCTIONAL CHARACTERIZATION OF TRIGGERS OF POST-OPERATIVE RECURRENCE IN CROHNâ€™S DISEASE - DMP title

Grant Title 1296122N

Principal Investigator / Researcher SÃ©verine Vermeire

Project Data Contact Sare Verstockt

Description Crohnâ€™s disease (CD) is a chronic, relapsing inflammatory bowel disease which is caused by a complex interaction of genetic, immune, microbial and environmental factors. Surgical resection is needed in up to 70% of CD patients. At the luminal side of the intestine, new CD lesions recur within weeks to months in two-thirds of patients undergoing surgery, but only in the presence of the fecal stream. At the external side of the intestine, a pathognomic phenomenon of CD is observed, ie.creeping fat, which is a local hyperplasia of mesenteric adipose tissue migrating to the sites of intestinal inflammation. However, which luminal components are triggering post-operative recurrence and whether creeping fat has a pathological or protective role in CD, is still not clear. The current project aims to identify triggers of post-operative recurrent CD, by focusing both on the external and the luminal side of the intestine. First, we will profile the lipidomic landscape of creeping and paired mesenteric fat at the time of surgery, and assess if these lipid signatures are predictive for post-operative recurrence. Next, we will characterize fecal components predicting recurrence by combining shotgun metagenomics with metabolomics. Furthermore, to get a better functional understanding, these fecal contents will also be exposed to their ileal epithelial cells using in vitro 2D transwells derived from organoids, followed by epithelial RNA sequencing as a readout. Overall, the clinical value of this project lies in a better understanding of creeping fat being pathognomonic for CD, and in identifying predictive markers for post-operative recurrence, both external and luminal. The findings from the different analyses will enable to identify pointers for new therapies or repurposing of existing drugs. Moreover, the impact for patients suffering from CD is that early strategies can be developed to identify and to treat patients at risk for disease recurrence to ultimately reduce surgery rates.

Institution KU Leuven

1. General Information

Name applicant

Sare Verstockt

FWO Project Number & Title

1296122N

LIPIDOMIC, MICROBIAL AND FUNCTIONAL CHARACTERIZATION OF TRIGGERS OF POST-OPERATIVE RECURRENCE IN CROHN’S DISEASE

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).

Type of data	Format	Volume	How created

Clinical data	.xls	1 file, <1 MB	Data retrieved from the patient records system (KWS platform UZ Leuven), stored in a pseudonymised manner.
Storage files -fat samples -mucosal biopsy and derived organoid -feces	.xls	1 file, <5 MB (ie. for samples being part of this project)	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: fat, mucosal biopsy, derived organoids, feces). The samples for this project in particular are thus part of our general sample storage files.
	[Lipidomics and Mediator lipidomics]	/	[raw HILIC LC-MS/MS data and raw C18 LC-MS/MS data from fat samples will be archived at Lipometrix (KU Leuven) – data are available upon request. Normalized data are usually transferred to the clients)]
	Microbiome: .fastq	50 GB	Shotgun metagenomics of fecal samples: sequences generated via the Raes lab and the Nucleomics core

			(VIB/KU Leuven)
Raw data	Metabolomics: .mzML .xls	7 GB	Raw LC-MS data of fecal samples to study the bile acid profiles, generated via the Metabolomics core (VIB/KU Leuven)
	Electrical resistance .xls	2 GB	TEER measurements to determine epithelial barrier integrity in transwells (pilot study + functional experiment)
	Microscopy .tif	50 MB	H&E staining
	Microbiome: .fastq	500 MB	16S rRNA sequencing of the microbiome (pilot study transwells)
	RNAsequencing .fastq	50 GB	Genome-wide gene expression of functional transwell experiment, sequenced generated by the Genomics core (UZ/KU Leuven)

Processed data	Lipidomics and mediator lipidomics: Normalized data .csv	1 file, <1 MB	A combination of preprocessing steps (eg. normalization, alignment) and filtering steps to a format that is easier to process and/or analyze.
	Microbiome (shotgun + 16s rRNA): .counts .xls	50 MB	
	Metabolomics: .xls	1 file, < 1MB	
	RNA sequencing .counts .xls	25 MB	
metadata	.txt .xls	2 MB	An overview file with a clear description of what the data represent, how they were generated, related quality control metrics etc. Metadata standards are used when applicable

Scripts	Code .py, .ipynb .R, .Rmarkdown or similar	5 GB	Code that will transform Raw data into the Processed data Code that will transform Processed data into Results Codes will be commented and extensively documented using notebooks such Jupyter Notebooks and R markdowns
Results	.pdf, .tif, .svg, .xls, or similar	10 GB	The outcome of this projects. Results can be tables, figures and text explaining those.
Lab note books			Dated written notes associated with carrying out experimental procedures
Standard operating procedures	.pdf	2 GB	Written protocols for experimental procedures performed in the lab

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.

- Yes

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 smoking status, medication, type of surgery 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, add the reference to the formal approval by the relevant ethical review committee(s)

- Yes

This project involves sample and data collection from human participants. It has formal approval by the UZ Leuven ethical committee (S53684).

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 cannot exclude that this project might result in research data with potential for tech transfer and valorisation. In case of any potential, the invention will be evaluated and may be IP protected with the support of the Intellectual Property Unit of KU Leuven Research & Development (LRD). As such the IP protection does not withhold the research data from being made public. Furthermore, we regularly interact with various stakeholders, including pharma and patients, which are highly relevant when it comes to the subsequent steps of valorisation of the findings in the current 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?

- No

4. Documentation and metadata

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

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

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? 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
- 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.

Minimum Information about a high-throughput nucleotide SEQuencing Experiment (MINSEQE) will be used for the shotgun metagenomics data, 16S rRNA seq data and RNAseq data

- Where no metadata standard exists

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

5. Data storage and backup during the FWO project

Where will the data be stored?

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 is backup of the data provided?

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 yes, specify concisely. If no or insufficient storage or backup capacities are available then 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.

What are the expected costs for data storage and back up during the 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 security: 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

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 data will be retained for the 5 year period after the project

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

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 retention period of 5 years? How will the 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.

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:

The informed consent contains a section in which the participant can if choose his/her data is shared, including sharing data within non-Belgian countries towards international scientific collaborations, including biotechnological or pharmaceutical companies. We will honour their choice. Furthermore, shared data will always be pseudonymised.

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

All data obtained can be available after publication.

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

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

- In an Open Access repository
- Upon request by mail

- 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 the research results

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.

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

None, the filesender of Belnet is for free.

8. Responsibilities

Who will be responsible for data documentation & metadata?

The researcher (Sare Verstockt) and PI (Séverine Vermeire) are responsible for data documentation & metadata

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

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

Who will be responsible for ensuring data preservation and reuse ?

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

Who bears the end responsibility for updating & implementing 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