Characterization of triggers of post-operative recurrence in Crohn’s disease in both the external and luminal side of the intestine

A Data Management Plan created using DMPonline.be

Creators: Sare Verstockt, n.n. n.n.

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: n.n. n.n.

Data Manager: Sare Verstockt

Grant number / URL: G088723N

ID: 200821

Start date: 01-01-2023

End date: 31-12-2026

Project abstract:

Crohn’s disease (CD) is a chronic, 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 after surgery in two-thirds of patients, but only in the presence of a fecal transit. At the external side of the intestine, the phenomenon “creeping fat” is observed, which is mesenteric fat 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.

This project aims to identify triggers of post-operative recurrent CD, by studying both the external and the luminal side of the intestine at the time of surgery, as well as six-months follow-up samples; and this by using state-of-the-art technologies. At the time of surgery, we will profile the cellular and molecular composition of both creeping fat and the affected mucosa (ie. the inner layer of the intestine), thereby evaluating how host and microbial components interact, and if these can predict post-operative recurrence. In addition, we will characterize the bile acids and microbiome present in the stool and the mucosa six months after surgery, and evaluate how these interact with the gene activity in the mucosa, in particular in epithelial cells.

Last modified: 30-06-2023

Characterization of triggers of post-operative recurrence in Crohn’s disease in both the external and luminal side of the intestine
FWO DMP (Flemish Standard DMP)

1. 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.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
Clinical data	Data retrieved from the patient records (KWS platform UZ Leuven), stored in a pseudonymised manner	new	digital	observational	.xls	<1 MB	NA
Storage files -fat samples -mucosal biopsy and derived 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: fat, mucosal biopsy, derived organoids, feces). The samples for this project in particular are thus part of our general sample storage files.	new	digital	experimental	.xls	<5 MB (ie. for samples being part of this project)	NA

Raw data - Lipidomics and Mediator lipidomics	raw HILIC LCMS/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)	new	digital	experimental	/	/	NA
Raw data - single nuclei RNA sequencing	raw single-nuclei RNA sequencing data from fat and mucosal samples	new	digital	experimental	.bam .fastq	500 GB	NA
Raw data - spatial transcriptomics	coordinates and count files from the multiplexed single molecule RNA fluorescent in situ hybridization (smFISH) in a subset of our fat samples using the Molecular Cartography platform	new	digital	experimental	.txt	100 MB	NA
Raw data - Microbiome	Shotgun metagenomics of fecal samples: sequences generated via the Raes lab and the Nucleomics core (VIB/KU Leuven)	new	digital	experimental	.fastq	50 GB	NA
Raw data - Metabolomics	Raw LC-MS data of fecal samples to study the bile acid profiles, generated via the Metabolomics core (VIB/KU Leuven)	new	digital	experimental	.mzML .xls	7 GB	NA
Raw data - Electrical resistance	TEER measurements to determine epithelial barrier integrity in transwells (pilot study + functional experiment)	new	digital	experimental	.xls	2 GB	NA

Raw data - Microscopy	H&E staining	new	digital	experimental	.tif	50 MB	NA
Raw data - Microbiome	16S rRNA sequencing of the microbiome (pilot study transwells): sequences generated via the Raes lab and the Nucleomics core (VIB/KU Leuven)	new	digital	experimental	.fastq	500 MB	NA
Raw data - bulk RNAsequencing	Genome-wide gene expression of functional transwell experiment, sequences generated by the Genomics core (UZ/KU Leuven)	new	digital	experimental	.fastq	50 GB	NA
Processed data - Lipidomics and mediator lipidomics	A combination of preprocessing steps (eg. normalization, alignment) and filtering steps to a format that is easier to process and/or analyze.	new	digital	experimental	.csv	<1 MB	NA
Processed data: Microbiome (shotgun + 16s rRNA)	A combination of preprocessing steps (eg. normalization, alignment) and filtering steps to a format that is easier to process and/or analyze.	new	digital	experimental	.counts .xls	50 MB	NA
Processed data - Metabolomics	A combination of preprocessing steps (eg. normalization, alignment) and filtering steps to a format that is easier to process and/or analyze.	new	digital	experimental	.xls	<1MB	NA

Processed data - bulk RNAsequencing	A combination of preprocessing steps (eg. normalization, alignment) and filtering steps to a format that is easier to process and/or analyze.	new	digital	experimental	.counts .xls	25 MB	NA
Processed data - single nuclei RNAsequencing	with Bioinformatic softwares implemented in R/pvthon	new	digital	experimental	.mtx,.rds,.tif,.csv,.R,.pv	500 GB	NA
Processed data - spatial transcriptomics	images and segmentation results as Seurat objects from multiplexed single molecule RNA fluorescent in situ hybridization (smFISH) using the Molecular Cartography platform (Resolve BioSciences)	new	digital	experimental	.txt .tiff	10 GB	NA
Metadata	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.	new	digital	experimental	.txt .xls	2 MB	NA
Scripts	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.	new	digital	experimental	Code .py, .ipynb .R, .Rmarkdown or similar	5 GB	NA

Results	The outcome of this projects. Results can be tables, figures and text explaining those.	new	digital	experimental	.pdf, .tif, .svg, .xls, or similar	10 GB	NA
Lab note books	Dated written notes associated with carrying out experimental procedures.	new	physical	NA	NA	NA	2-3 books
Standard operating procedures	Written protocols for experimental procedures performed in the lab.	new	physical	experimental	.pdf	2 GB	NA

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:

NA

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes, human subject data

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

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- 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)

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

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 exploitation or dissemination of the data you (re)use (e.g. Material/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.

- No

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

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.

2. 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.
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 (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- Yes

-Metadata standard:
The biological samples registered in the biobank will contain metadata required by the royaldecree of biobanking (9JAN2018), standardized metadata to trace the pre-analytical factors of the sample which are most likely to impact research results. The required minimum metadata for each dataset will be provided according to the corresponding repository (EMBL-EBI data repositories).

-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.

3. Data storage & back-up during the research 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 will the data be backed up?

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.

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 andcustomised read/write permissions.

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

4 TB is available on the central archive K drive of our research group (€250/year) – for this project we will need < 5 TB for raw data storage.

2 TB is available on the central Shared J-drive of our research group (€1000/year) – for this project we will need < 2 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.

4. Data preservation after the end of the research project

Which data will be retained for at least five 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 retained for the 5 year period after the project.

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

The data will be stored on the university's central servers (with automatic backup 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 ~6000 EUR for the 5 years after project end. Funding is available in our group to cover these costs.

5. Data sharing and reuse

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

- Yes, in an Open Access repository

- Human transcriptomics and metagenomics: European Genome-Phenome Archive
- Human lipidomics: European Genome-Phenome Archive

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 in the comment section per dataset or data type where appropriate.

- Yes, Privacy aspects

The informed consent contains a section in which the participant can choose if 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.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

- In open Access repository: In case of omics data, these datasets will be deposited to European Genome-Phenome Archive. 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.

Which data usage licenses are you going to provide? If none, please explain why.

Data will be deposited in a standard format that is commonly accepted and which can be accessible for everyone. The FAIR (Findable, Accessible, Interoperable, and Reusable) principles will be followed to make data findable using persistent identifiers, which will ensure data accessibility and reusability with clear and standardized metadata, through open licenses and data formats.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

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

None, the filesender of Belnet is for free.

6. Responsibilities

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

PI (Séverine Vermeire), and CoPIs Sare Verstockt and Bram Verstockt

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

PI (Séverine Vermeire), and CoPIs Sare Verstockt and Bram Verstockt

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