

Data Management Plan

FWO project title: Transcriptional bet hedging as a resilience strategy in the human gut microbiome (Gut-BET)

General information

- **Grant number:** G0AAV24N
- **Principal investigator:** Jeroen Raes
- **Project Data contact:** Jeroen Raes (jeroen.raes@kuleuven.be), Lindsey De Commer (lindsey.decommer@kuleuven.be), Lindsay Devolder (Lindsay.devolder@kuleuven.be)
- **Affiliation:** KU Leuven, VIB-Center for Microbiology

Description:

The gut microbiota is essential for intestinal homeostasis, yet how it keeps its resilience from external perturbations is unclear. Previous work has shown that various drugs can lead to dysbiosis in patients with gastrointestinal pathologies, but the mechanisms by which the microbiota recovers from these insults is not understood. Here, we hypothesise that the microbiota uses so-called 'bet hedging' approaches to this aim, in which a subset of microbial cells within a clonal population expresses functionalities that reduce its own fitness but which will be useful in case of a future insult - a process well-known in so-called persister cells in the area of antibiotics resistance by human pathogens. To assess whether this process is more generally applicable to gut microbiota resilience, we will implement bacterial single-cell RNAseq to study transcriptional heterogeneity in gut microbial strains, simple communities and fecal populations. Bet hedging will be studied using in vitro analyses subjecting gut bacteria to antibiotics and/or drugs with suspected antibiotic side effects (otilonium bromide/Spasmomen, used in irritable bowel syndrome), and validated in vivo through analysis of samples from intervention studies in IBS patients under Spasmomen treatment. Together these experiments will elucidate microbiota resilience mechanisms and help understand gastrointestinal drug mode-of-actions.

Data description:

1. **Will you generate/collect new data and/or make use of existing data:**
 - Generate new data in work package 3 of this project
2. **Describe the origin, type and format of the data (per dataset) and its (estimated) volume, ideally per objective or WP of the project. You might consider using a table in the guidance:**

WP	Origin	Type	Format	Volume (Gb)
WP1	Bacterial isolates	Experimental data: scRNA-seq	csv files	100
WP2	Bacterial isolates	Experimental data: scRNA-	csv files	100

		seq		
WP3	Stool	Experimental data: scRNA-seq		100
	Surveys/questionnaires	Observational data	csv files	1

Legal and ethical issues

- 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.**
 - Yes.
- 2. Privacy Registry Reference:**
 - Not applicable.
- 3. Short description of the kind of personal data that will be used:**
 - Participant demographics and health data, dietary information (FFQ), stool water content, Bristol stool Scale, 16S QMP gut microbiome data, Rome IV IBS questionnaire data, fecal calprotectin and standard blood analysis data was collected, next to treatment response and symptom data (bloating, gas production, pain, ...).
- 4. 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. Ethical approval was obtained for the Fermbrella study (S65767).
- 5. 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, the project is basic science, and while it is not the main goal of the project, it is possible that some of the project's findings would have commercial value. This potential tech transfer and valorization might come from: 1) The method developed for scRNAseq, 2) New targets or approaches based on drugs or diets to treat irritable bowel syndrome (IBS) and 3) Novel bacterial genes of interest as targets for IBS.
 - IP will be claimed under the VIB-KU Leuven framework agreement in collaboration with VIB TT and LRD. Data will be provided to academics under controlled access upon the signature of an MTA. Commercial licenses will be negotiated by VIB/LRD.
- 6. 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

Documentation and metadata

1. **What documentation will be provided to enable reuse of the data collected/generated in this project?**
 - Metadata will be documented by the research and technical staff at the time of data collection and analysis. Updates of the data record will be documented in the metadata sheet, linking to all the different file formats for the specific data types generated. Additionally, comprehensive documentation will be provided, including data dictionaries, codebooks, and user guides, to ensure clarity on data structure, variables, and methodologies used. This documentation will be regularly reviewed and updated to reflect any changes or additions to the dataset, facilitating ease of understanding and reuse by future researchers and stakeholders.
2. **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. Metadata standard templates will be used (e.g. MlXS – Minimum information about any (x) sequence, MIGS - Minimum Information about a Genome Sequence, MIMS – Minimum Information about a Metagenome Sequence, MIMARKS – Minimum Information about a MARKer gene Sequence). Final datasets will have a README.txt file with metadata standard information and references to supporting documents. The file will be detailed enough to add contextual value to the dataset for future reuse.

Data storage & back up during the FWO project

1. **Where will the data be stored?**
 - Data will be stored on the secured storage services offered by KU Leuven and the Raes Lab (VIB) production server (physically located at the Plant Systems Biology (PSB) Facility (Zwijnaarde, Gent), with the latter being continually maintained and backed up by VIB.
2. **How is back up of the data provided?**
 - VIB PSB IT service regularly backs up the data stored (daily, monthly, and yearly basis). Fail safes are in place to ensure that data is retrievable
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.**
 - Yes. The Raes secure server has sufficient storage capacity (Oracle Storage DE2-24C (with 24x 8 Tb disks)) and will be expanded if need be.
4. **What are the expected costs for data storage and back up during the project? How will these costs be covered?**
 - For this project, no extra costs for storage and backup are foreseen. If need be, these will be covered on the FWO project and/or lab internal funding.
5. **Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?**
 - In order to access the server, an individual must be a member of the Raes lab and have been granted appropriate access to the server. The individual will be provided with a username and passcode once they are approved to work in the lab. No personal data or identifiable data will be made available to the general members of the Raes lab.

Data preservation after the FWO project

- 1. 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.
- 2. Where will the data be archived (= stored for the longer term)?**
 - The secured, backed up Raes lab server will serve as an adequate storage space for the datasets in perpetuity (at minimum 30 years). Data cleaning, metadata descriptors, and contextual information will all be stored in server folder. Personalized/identifiable data will be stored in a separate secured database accessible to only the Principal investigators, and study coordinator.
- 3. What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?**
 - The Raes lab is the owner of the data storage infrastructure on its servers. Maintenance and system administrator cost is ±5000 euros/y, paid on internal lab funding.

Data sharing and reuse

- 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)?**
 - The general public or other institution members will not be able to access the project data. Upon publication, data will be made available under restricted access. There are no privacy or ownership issues between the participating members of the project. The nonidentifiable project data is viewable to all members of the Raes lab provided that they have an account and passcode to the server with appropriate credentials.
- 2. Which data will be made available after the end of the project?**
 - At the time of publication, anonymized data necessary to replicate results will be made available under restricted access (EBI EGA). This might consist of sequencing data and/or a subset of metadata.
- 3. Where/how will the data be made available for reuse?**
 - In a restricted access repository. Anonymized data will be made available at the European Genome-phenome Archive (EGA, <https://www.ebi.ac.uk/ega/>). The source code will be released on GitHub.
- 4. When will the data be made available?**
 - Upon publication of the research results
- 5. Who will be able to access the data and under what conditions?**
 - Data released, including research outputs supporting publications, might be used by anyone for any purpose, provided that appropriate credit is given. Bonafide academic researchers will be able to access the data upon approval of the project PI, and upon signature of a data access agreement with VIB. Commercial reuse is to be negotiated with VIB, but generally, raw data is not released for commercial use.
- 6. What are the expected costs for data sharing? How will the costs be covered?**
 - No cost.

Responsibilities

1. **Who will be responsible for data documentation & metadata?**
 - Jeroen Raes, PI
2. **Who will be responsible for data storage & back up during the project?**
 - VIB gent system administrators. Jeroen Raes, PI
3. **Who will be responsible for ensuring data preservation and reuse ?**
 - Jeroen Raes, PI
4. **Who bears the end responsibility for updating & implementing this DMP?**
 - The PI bears the end responsibility of updating & implementing this DMP.