DMP for FWO senior postdoc fellowship - Getting a gRiPP on early BRCA1-mutant breast cancer detection and intervention

Application DMP

Questionnaire

The questions in this section should only be answered if you are currently applying for FWO funding.
Are you preparing an application for funding?

Question not answered.

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

Question not answered.

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

Question not answered.

What's the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

Question not answered.

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

Question not answered.

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

Question not answered.

For whom might your data be useful outside of the research project, e.g. researchers or other stakeholders? How will you share this data?

Question not answered.

DMP for FWO senior postdoc fellowship - Getting a gRiPP on early BRCA1-mutant breast cancer detection and intervention

DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

Question not answered.

DMP for FWO senior postdoc fellowship - Getting a gRiPP on early BRCA1-mutant breast cancer detection and intervention

GDPR

GDPR

Have you registered personal data processing activities for this project?

Question not answered.

DMP for FWO senior postdoc fellowship - Getting a gRiPP on early BRCA1-mutant breast cancer detection and intervention

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
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .csv,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
L. lactis strains capable of sensing and reacting to breast cancer	Engineered L. lactis strains capable of sensing and reacting to breast cancer	Generate new data and reuse existing data	Physical (strains) and digital (inventory)	Metadata: Excel file with inventory and description of plasmids	.xlsx	<1GB	Stored in duplicate as glycerol stocks in a 80°C freezer
L. lactis cell factory that can produce antibacterial and anticancer RiPPs	Engineered L. lactis strains capable of producing antibacterial and anticancer RiPPs	Generate new data and reuse existing data	Physical (strains) and digital (inventory)	Metadata: Excel file with inventory and description of plasmids	.xlsx	<1GB	Stored in duplicate as glycerol stocks in a 80°C freezer
DNA libraries of thioalbamide	DNA engineering experiments	Generate new data and reuse existing data	Physical (DNA libraries) and digital (inventory)	Metadata: Excel file with inventory and description of plasmids	.xlsx	<100GB	Stored in duplicate in a -20°C freezer
DNA agarose electrophoresis gel	DNA engineering and deletion experiments	Generate new data	Digital	Experimental	TIFF	<1GB	

Plasmid constructs	Protein purification and gene deletions experiments	Generate new data	Physical (strains) and digital (inventory)	Experimental	.xlsx	<1GB	Stored in duplicate in a -20°C freezer
DNA sequencing of plasmids	Confirm DNA engineering experiments by Sanger sequencing	Generate new data	Digital	- Other: DNA sequencing from an external company - Metadata: excel file with inventory and description of sequencing data	Original: ABI, Inventory and descriptions of results: .xlsx	<100GB	
Complete plasmid and genome sequencing	Confirm sequence of new plasmid and bacteria	Generate new data and reuse existing data	Digital	Experimental	FASTQ	<1TB	
SDS PAGE gels	Expression and purification of engineered thioalbamide variants	Generate new data	Digital	Experimental	TIFF	<1GB	
LC-MS data	LC-MS analyses of bacterial metabolites and enzymatic reactions	Generate new data	Digital	Experimental	FID	<5TB	
HPLC data	HPLC purification of engineered RiPP variants	Generate new data	Digital	Experimental	PDF	<5TB	
NMR data	NMR data of engineered thioalbamide variants	Generate new data	Digital	Experimental	FID	<100GB	
MIC and MBC data	Antibacterial spectrum and potency of engineered thioalbamide variants	Generate new data	Digital	Experimental	TIFF and inventory and descriptions of results: .xlsx	<100GB	
Anticancer data	Anticancer potency of engineered thioalbamide variants in cell-lines and organoids	Generate new data	Digital	Experimental	TIFF and inventory and descriptions of results: .xlsx	<1GB	
Design of microfluidic chips	Design of microfluidic chips for the high-throughput screening of sactipeptide variants	Generate new data	Digital	Others: designs	.CAD		

Validation and running of microfluidic chips	Validation and running of microfluidic chips for high-throughput screening of sactipeptide variants	Generate new data	Digital	Others: Pictures and spectroscopic data (e.g. brightfield, fluorescence)	.TIFF .csv .matlab		
Xenograft mice models	Engineered L. lactis strains capable of sensing and reacting to breast cancer	Generate new data	Digital	Experimental	TIFF and inventory and descriptions of results: .xlsx	<100GB	
Lab books	All data will be transcribed into digital format and will be stored in an online lab management system with an integrated Electronic Lab Notebook (ELN)	Generate new data	Digital	Experimental	Text (Word) and tables (xlsx)		

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:

The reused data from public databases will comprise DNA and protein sequences.
NCBI: https://www.ncbi.nlm.nih.gov/
antiSMASH and antiSMASH database: https://antismash.secondarymetabolites.org/#!/start and https://antismash-db.secondarymetabolites.org/
RiPPMiner: http://www.nii.ac.in/~priyesh/lantipepDB/new_predictions/index.php
RiPPER: https://github.com/streptomyces/ripper
EFI-EST: https://efi.igb.illinois.edu/efi-est/

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, animal data

Task 4.2: I will validate the therapeutic L. lactis strain in a preclinical model of BRCA1-mutant TNBC development. To mimic TNBC development, human BRCA1-mutant cells derived from TNBC will be intraductally injected into the mammary glands of immunocompromised female recipient mice (n = 4 injected glands per mouse, 3 replicates with 2 different TNBC lines and 2 control lines). Mutant cells will engraft within the ductal network of the mammary gland and form pre-invasive fields of mutant cells representing pre-invasive lesion development, followed by invasive transformation later. Upon establishment of mutant fields of cells (2-3 months after injection) therapeutic L. lactis bacteria will be injected intraductally and mammary glands will be harvested 1, 2, 4, and 6 months after injection.
Expected start date: 3/11/2025
Ethics committee category Research centres: Research centres
Ethics committee: Ethical Committee for Animal Experimentation (ECD) - KU Leuven/UZ Leuven

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.

- No

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

Potential engineered thioalbamide variants with enhanced anticancer properties, as well as *L. lactis* strains proven to function as biosensors for BRCA-related breast cancer development, will be protected through patent applications. This intellectual property strategy will facilitate future exploitation and commercial development in collaboration with LRD (the University's intellectual property and technology transfer office) and VIB Ventures.

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.

- No

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).

Raw data will be collected per test, including a .txt file with a clear description of what the data represents and how they were generated. An inventory of all data and a description will be collected in an Excel file. All work performed in the lab will be saved as an Electronic Lab Notebook (ELN). This notebook will contain all information on methodology, protocols, results and conclusions. The ELN will be ordered chronologically, with a title for every date and a subtitle for every experiment. Every month, a time-stamped pdf copy of the ELN of that month will be made and stored on the KU Leuven servers. General protocols and standard operating procedures will be collected in a dedicated folder on a Shared network drive (J:) at KU Leuven.
Raw LC-MS, HPLC and NMR data will be collected alongside a .txt file (README file) with a description of the experimental conditions.
All the images with their explanation will be saved in the ELN. Additionally, the images will be saved in separate folders. The name of this folder consists of the date and the title of the experiment.
Various tools will be used to process data (e.g. SnapGene for sequencing data ...). The input files (raw data) will be kept in the same folder as the processed files. The name of the folder will contain the date and information about the experiment.
Background information will be stored in an Excel file and the Electronic Lab Notebook.

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.

- No

No metadata standard is available for the type of data that will be generated. The metadata for all NMR, LC-MS and UV spectroscopic data will be available within the file format they will be stored in. The metadata for all other experiments will be generated in a descriptive format which can be interpreted easily by other people in the future.

3. Data storage & back-up during the research project

Where will the data be stored?

All data will be transcribed into digital format and will be stored in an online lab management system with an integrated Electronic Lab Notebook together with accompanying information. This Electronic Lab Notebook (monthly time-stamped copies) and digital data (e.g. images, spectroscopic data, sequencing data, metadata ...) will be stored in a personal folder on a Shared network drive (J:) which is backed up by the ICTS service of the KU Leuven. Additional copies will be made and kept on personal devices. Once a researcher leaves the lab, their data will be transferred to a large-volume network archive drive. Besides, the LC-MS data will be stored in duplicate on dedicated external hard drives. Bacteria and DNA will be stored in duplicate in -80 and -20°C freezers, respectively.

How will the data be backed up?

The data will be stored on the university's central servers with automatic daily back-up procedures. Copies of the LC-MS data will be stored on dedicated external hard drives.

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

KU Leuven provides sufficient storage and backup capacity during and after the project. The initial storage capacity of the server of KULeuven is 5 GB. However, it can be extended without extra cost to 10 GB. Besides, dedicated external hard drives of 5 and 10 TB are available for storing the LC-MS data.

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

The Shared network drive (J:) of the KU Leuven is only accessible to group members. Their access is determined by their KU Leuven personnel number. The drive has a high level of security. Furthermore, the data on the external hard drives will be protected by a password.

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

The KU Leuven shared network drive costs € 503,66/ TB / year. Additionally, a large volume storage drive from the KU Leuven costs € 104,42 / TB / year. This drive will be used to store all the large files. These costs, along with the costs for the external hard drives (approx. € 200) will be covered by the project budget.

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...).

After the research project, the principal investigator (Prof. Joleen Masschelein) will take responsibility for data preservation. All the data will be stored in an online lab management system with an integrated Electronic Lab Notebook. The 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)?

The data will be stored indefinitely on the university's large volume network archive drive (with automatic backup procedures). The LC-MS data will be stored in duplicate on external hard drives.

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

The KU Leuven large-volume network archive drive costs € 104,42 / TB / year. These costs and the costs of the external hard drives (approximately € 200) will be covered by the project budget.

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 a restricted access repository (after approval, institutional access only, …)

During the project, all data will be stored on the shared network drive of the KU Leuven. Lab members and people participating in the project can get access to the data stored on this drive based on their personnel number.
If novel engineered sactipeptide variants with improved properties are found or generated, a patent application will be filed. This may temporarily restrict the sharing of data.
After the end of the project, all published data will be made available.

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

Only researchers participating in the project and lab members will be able to access the data before publishing. Upon publication, everyone will be able to access the data.

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, Intellectual Property Rights

If novel engineered thioalbamide variants with improved properties are found or if the L. lactis biosensor works, a patent application will be filed. This may temporarily restrict the sharing of data.

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

After publishing, the data will be made available in KU Leuven RDR, KU Leuven's institutional research data repository for the publication of research data. Genome data will be made available in GenBank (NCBI).

When will the data be made available?

The data will be made available upon publication of the research results.

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

If novel engineered thioalbamide variants with improved properties are found or if the L. lactis biosensor works, a patent application will be filed. Data can be accessed and reused upon request by email.

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

A unique identifier will be added to the published data. Besides, genome data deposited in GenBank will also get a unique accession number.

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

There are no expected costs for data sharing.

6. Responsibilities

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

During the research project, the fellowship holder (Hans Gerstmans), the Principal Investigator (Prof. Joleen Masschelein) and the co-promoter (Dr. Prof Colinda Scheele) will manage the data documentation and metadata.

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

During the research project, the fellowship holder (Hans Gerstmans), the Principal Investigator (Prof. Joleen Masschelein) and the co-promoter (Dr. Prof Colinda Scheele) will manage the data documentation and metadata.

Who will manage data preservation and sharing?

During the research project, the fellowship holder (Hans Gerstmans), the Principal Investigator (Prof. Joleen Masschelein) and the co-promoter (Dr. Prof Colinda Scheele) will manage the data documentation and metadata.

Who will update and implement this DMP?

This data management plan will be updated and implemented by the fellowship holder (Hans Gerstmans). The Principal Investigator (Prof. Joleen Masschelein) bears the end responsibility of updating and implementing this DMP.