DMP title

Project Name VOS FWO Junior Project 2021 DMP (FWO DMP) - DMP title

Project Identifier G060322N

Grant Title G060322N

Principal Investigator / Researcher Robin Vos

Project Data Contact Prof. Bart Vanaudenaerde, BREATHE, KU Leuven, ON Ibis Herestraat 49 - bus 706, 3000 Leuven, tel. +32 16 37 74 77 tel. +32 16 33 01 94 email: bart.vanaudenaerde@med.kuleuven.be

Description Title: Transcriptome-based management of transplant-related pulmonary fibrosis in chronic pulmonary rejection after lung or allogeneic hematopoietic stem cell transplantation Type of Project: Junior research project fundamental research - G060322N Duration: 2021-2024 Brief Summary: 1. Aims: The primary objective of this translational research project is to investigate the pathways and mechanisms involved in transplantrelated pulmonary fibrosis in obstructive and restrictive phenotypes of CLAD and pulmonary cGvHD, in order to identify novel diagnostic markers and to design improved therapeutics, allowing to better manage rejection-mediated fibrotic lung remodeling. 2. Methods: A multi-modal approach, using human lung tissue, bronchoalveolar lavage and blood samples, is used to assess the morphometric structural changes, immunopathologic organization, and gene signature in CLAD and pulmonary cGvHD.

Institution KU Leuven

1. General Information

Name applicant

Robin Vos (prof. dr., MD, PhD)

FWO Project Number & Title

Junior research project fundamental research - G060322N

Title of the application: Transcriptome-based management of transplant-related pulmonary fibrosis in chronic pulmonary rejection after lung or allogeneic hematopoietic stem cell transplantation

Affiliation

- KU Leuven

2. Data description

Will you generate/collect new data and/or make use of existing data?

- Generate new data
- Reuse existing 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).

The data comprise:

- Anonymized **clinical** (patient-related) data from electronic patient records (including general demographic information, specific (para-)medical information and therapies) (WP1-4).

- Anonymized **non-clinical** data including data regarding lung collection, processing of samples, microCT scanning and digital processing of lung casts; processing of BAL fluid, blood and lung tissue to sections for immunostaining; immunostaining and histology; single cells and RNA isolation, processing and profiling (WP1-4).

Morphological organization of airways in the human lung (WP1-2). MicroCT Scans of whole lungs (8-14GB) and small cores (5-6GB) of lungs tissue exist of large datasets in .tiff and .mha format. Depending on the application software packages include RadiAnt, ITK-SNAP, NeuronStudio, ImageJ. For the conversion from DICOM to .tiff and/or .mha ImageJ is used.

Histology and single nuclear RNA profiling in human lung tissue (WP2-4). The physical data include human tissue, sections, cells and RNA. Histology digital images are saved as .tiff or .jpeg format. For tissue and single cell/nucleus RNA profiling datasets are in .txt files. Datasets have a size of about 50MB and sub analytical file 2-5GB.

For all WPs extracted analyzed data are converted to MicroSoft Excel (.xls), Graphpad prism (.pzf), SAS (.sas) and R-statistical software (.R) for statistical analysis and summarized in Word (.word), PDF (.pdf) and PowerPoint files (.ppt), or als pictures (.tiff or .jpg files).

Physical data are stored in the **UZ/KUL Biobank** (histology rooms, fridges, freezers and cryotanks) and **digital data** (format: .xls, .word, .pdf, .xml, .txt, .ppt, .prism, .sas, .R, .mha, .tiff, or .jpeg, -files) are stored on **access-restricted KU Leuven secured data-servers** (GBW-0076_LTx and GBW-0017_LTx), as per Institution's regulations and per regulations of Dept. Chrometa, in accordance with the existing data management plan of BREATHE: DMP C1 small airways, C16/19/005 (B. Vanaudenaerde). The volume of all data is estimated at 1TB (1000 GB) maximum.

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

Privacy Registry Reference:

CTC UZ Leuven and PRET KU Leuven approval will be sought for each new study/experiment including personal data, if not yet covered under ongoining Biobank and Ethics approval numbers of BREATHE which apply to the current project: : S61653, S51577, S52174, S57742, S54739, S6397.

Short description of the kind of personal data that will be used:

Personal data collected are general clinical (patient-related) and demographic data, such as age, gender, length, body weight,..., but also specific medical information on the lung condition (lung function, radiological evaluation, pathology of the lung) and general health status and therapy (CRP, smoking behavior, steroid use, immunosuppression…).

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

CTC UZ Leuven and PRET KU Leuven approval will be sought for each new study/experiment including data not yet covered under ongoining Biobank and Ethics approval numbers of BREATHE which apply to the current project: S61653, S51577, S52174, S57742, S54739, S6397.

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

This is unclear at the current moment in time, and will depend on the results/progression of the plannen research in 2021-2024. If needed, this will be discussed in due time with the Intellectual Propery Department of KU Leuven Research & Development (LRD).

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

Currently no 3rd party agreements restrict dissemination or exploitation of the foreseseen data toe be obtained form this project.

4. Documentation and metadata

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

For each study/experiment, standard experimental procedures (SOPs) and practices will be documented in.word (and .PDF) files and saved on the GBW-0076_LTx data server. The methodology and protocols will also be described in detail in the lab book.

SOPs include: management of clinical data, lung collection and processing, microCT scanning and digital processing of lungs, processing of lung tissue to sections for immunostaining, imunostaining and histology procedures, single cell and RNA isolation, processing and profiling, handling of analytical parameters.

Raw experimental data from the analytical SOPs are collected per experimental test, and will include a MS .word file with a clear description of what the data represent and how they were generated. This description will be documented in notebooks (with page numbers), as well as in electronic format (.word files).

The name of the folder always contains the date, name of the experiment, and the name of the person who performed the experiment.

Each individual file with experimental data contains information on the study design, the origin of the samples, and all necessary information for an independent analyst to use or reuse the data accurately and efficiently.

Raw data are stored at the slower long-term storage GBW-0017_LTx which includes an accompanying meta-datafile of the experimental settings. The format is method and software dependent and includes .xml, .Dicon, .lif, .xml, .fcs, .vsi,... or excel. Other metadata from experiments in a searchable database format (in an excel file) include data, initials of the investigator, type of the study (clinical, pre-clinical, in vitro, compound used). Project specific analyzed data per fellow, PI will include an excel file, statistical file (Graphpad Prism, R software, SAS), powerpoint file and word file allowing to reconstruct and reanalyze each manuscript. Temporary files per project will be stored on the faster GBW-0076_LTx server per fellow/PI during the project but once finished and published they will be stored on the larger server to back-up the study for a period of minimal 5 years.

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.

- No

Temporary data files per project will be stored on the GBW-0076_LTx server per fellow/PI during the project but once finished these data files will be stored on the long-term storage data server GBW-0017_LTx for a period of minimal 5 years, which includes an accompanying meta datafile of the experimental settings.The format of included files is method and software dependent, bur are stored in a searchable database form.

5. Data storage and backup during the FWO project

Where will the data be stored?

please also see answers above regarding our dataservers:

GBW-0076_LTx (0.5Tb, to store temporary files)

GBW-0017_LTx ((15Tb, for long term storage of all (raw) data files)

How is backup of the data provided?

Backups of project data on KU Leuven Faculty network shares are made using “snapshot” technology, which is the online storage of incremental data changes. The standard backup regime is as follows:

- An hourly backup (at 8 AM, 12 PM, 4 PM and 8 PM) the last 6 of which are stored on our servers
- A daily backup, at midnight, the last 6 of which are stored on our servers
- A weekly backup, Saturday night at midnight, the last 12 of which are stored on our servers

The end user can use his own Windows PC to restore files to an older version using the “previous versions” function. According to the above backup scheme, it is possible to go back in time up to 12 weeks (~3 months).

For the purpose of “business continuity” or “disaster recovery”, a mirror (exact copy) of all data is created in a second datacenter. The data are copied every hour to the second datacenter. In the event that the primary storage unit is corrupted, the ICTS team can get this copy online within the hour.

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

Yes, please also see answers above regarding our dataservers are:
GBW-0076_LTx (0.5Tb, to store temporary files)
GBW-0017_LTx ((15Tb, for long term storage of all (raw) data files)

What are the expected costs for data storage and back up during the project? How will these costs be covered?
For GBW-0076_LTx the annual cost is 519 euro/terrabyte. BREATHE is currently use 377 Gigabyte of the 500 Gigabyte reserved, resulting in an annual cost of 195 euro/year currently.
For GBW-0017_LTx the annual costs are 156,6 euro. BREATHE currently uses 10,7 terrabytes of the 15 terrabytes reserved resulting is a yearly cost of 2350 euro currently.
We estimate te current project will create an additional 1 TB of data (max). The overall currently yearly cost for both servers is 2550 euro. These costs are already covered for many years by the running projects in BREATHE and will be included in the budget of the laboratory by existing budget and newly obtained projects.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?
The data are secured on both servers by the university ICT. All ICT solutions at KU Leuven are subject to the university-wide ICT information security standards. The group ICT service organizes the raw network storage, it procures from central ICT services in such a way that access permissions are limited, fixed, delegated to and audited by data managers who do not need to have an IT background. Access to these servers are restricted to the research group members. The lab manager (Karen Maes) and the PI of the project control who has access to the servers.

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 obtained from studies/experiments during the current project will be saved as described above, for at least a 5 year period after the end of the project. Currently our data servers allow for indefinite storage.

Where will the data be archived (= stored for the longer term)?
The data will be stored on a university's central server GBW-0076_LTx (with automatic back-up procedures as describe above) for at least 10 years, conform the KU Leuven RDM policy.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?
The currently yearly cost we expect for data preservation in the current project (1 TB) are estimated at 1250 euro, or a total of 6 250 euro (2021-2025). The department CHROMETA reserves for each separate group per year a small budget which is enough to cover these annual (and total) cost of basic storage. Additional funding for long-term data storage can be obtained from existing reserves and newly obtained projects in the future.

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)?

- No

For data sharing, a formal MTA is arranged and at the collaborating centers data will be destroyed at the end of the project. At these collaborating centers, a DMP is also installed compliant with this DMP. No 3rd parties are currently involved preventing/restricting data sharing.

Which data will be made available after the end of the project?
For most part of the project, raw data will not be made open access available. The only expected

open access will be for RNA profiling (anonymized) at the time of publishing in international peer reviewed papers, as this is international obligatory.

If required, raw data can be accessed through the PI and made available in an Open Access repository or by direct contact (mail). The full (anonymized) dataset will be uploaded in a .cvs format in Zenodo upon acceptance of the manuscript and this for everybody without restrains.

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

- In an Open Access repository
- Upon request by mail

he full (anonymized) dataset will be uploaded in a .cvs format in Zenodo upon acceptance of the manuscript and this for everybody without restrains.

When will the data be made available?

- After an embargo period. Specify the length of the embargo and why this is necessary
- Upon publication of the research results

For most other data, no release of raw data is requested and considered, unless asked during manuscript revision or by request of the medical journal Editorial Office. IP - if applicable - will request specific embargo's and will only be considered at the time needed. RNA profiling will be made available upon acceptance of the data for publication/manuscript, without restrains.

Who will be able to access the data and under what conditions?

For most other data, no release of raw data is requested and considered, unless asked during manuscript revision or by request of the medical journal Editorial Office. IP - if applicable - will request specific embargo's and will only be considered at the time needed. RNA profiling will be made available upon acceptance of the data for publication/manuscript, without restrains.

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

No cost expected.

8. Responsibilities

Who will be responsible for data documentation & metadata?

Data documentation and metadata will be organized by the PIs and fellows organizing the laboratory and project, namely Bart VAnaudenaerde (Leading Lab research professor for lung transplantation), Karen Maes (BREATHE lab manager) and Celine Aelbrecht (lab technician).

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

Gert Goos (KU Leuven ICT) - responsible for digital data storage, back-up and protection on data-servers Prof. Dr. Kristel Van Landuyt (UZ/KUL Biobank) - responsible for physical data storage in Biobank

Prof. Bart Vanaudenaerde (Leading BREATHE Lab research professor for lung transplantation) and Karen Maes (BREATHE lab manager) oversee and coordinate data storage and back-up.

Who will be responsible for ensuring data preservation and reuse ?

Prof. Bart Vanaudenaerde (Leading BREATHE Lab research professor for lung transplantation) and Karen Maes (BREATHE lab manager) are responsible for ensuring data preservation and reuse.

Who bears the end responsibility for updating & implementing this DMP?

The PI (Robin Vos) bears the end responsibility of updating & implementing this DMP.