Elucidating the Active Role of Traction bronchiectasis and Honeycombing in IPF's persistent fibrogenesis (EARTH-IPF)

A Data Management Plan created using DMPonline.be

Creator: Laurens De Sadeleer

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Grant number / **URL:** G034124N

ID: 208163

Start date: 01-01-2024

End date: 31-12-2027

Project abstract:

Idiopathic Pulmonary fibrosis (IPF) is a fibrotic lung disease affecting more than 300,000 patients in Europe. Until now, most research in the field focused on the deposition and accumulation of extracellular matrix (ECM) proteins. However, as this is highly intertwined with physiological repair upon injury, the quest for intervening herein has been unfruitful. In this project, we will focus on an entirely different disease aspect: the vast increase in conductive airway-like structures, called traction bronchiectasis (TBx) and honeycombing (HC). While initially thought of as passive phenomena due to traction appearing in end-stage disease, these opinions are not supported by literature and our own preliminary findings.

We hypothesize that TBx/HC are the result of an active process induced by an aberrant regenerative signaling response which should be specifically targeted to stop further disease progression, and which is not addressed with current therapeutic options. The global aim of this research project is to understand the morphological and molecular dynamics of this distinct disease aspect and to pave the way for entirely novel therapeutic options targeting these HC/TBx phenomena.

This research project consists of two main aims, both focusing on computational analysis. The first goal of the project is to assess 3D structure of HC/TBx leveraging our unique inflated explant lung biobank and state-of-the-art whole-lung ultra-high-resolution CT-scan and microCT of smaller samples of the lung. Second goal is to molecularly characterize these enigmatic disease aspects, using multi-omics and spatially resolved transcriptomics datasets which are readily available.

Last modified: 27-06-2024

Elucidating the Active Role of Traction bronchiectasis and Honeycombing in IPF's persistent fibrogenesis (EARTH-IPF)

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	
Lung cores	Samples used in the project.	Reused	Physical				n=15x6+15x3=135
Slices for staining	From the FFPE part of the core	New	Physical				n= 10x135 = 1350
BAL samples	WP5, validation	New	Physical				n=250
ex vivoCT images	after explantation, all lungs are scanned using ultra-high resolution CT	Reused	Digital	Experimental	.dicom	150GB	
microCT images	after processing the lungs into cores, the cores included in the project are microCT scanned	Reused	Digital	Experimental	.tiff	650GB	
in vivoCT	in vivoCT data for T2.3	Reused	Digital	Observational	.dicom	50GB	
CALIPER analysis	analysis of in vivo CT scans	New	Digital	Experimental	.dicom	75GB	
airflow analyses	T2.4	New	Digital	Simulation	.tiff	50GB	
digitalised microscopy: H&E stainings	from the FFPE part of all cores, an initial H&E section is made	Reused	Digital	Experimental	.czi	15GB	
digitalised microscopy: Movat/pentachromestaining	from the FFPE part, Movat/pentachrome stainings will be additionally made	new	Digital	Experimental	.czi	30GB	
digitalised microscopy: 4i panels	4i stainings from a selected group of cores	Reused	Digital	Experimental	.czi	150GB	
digitalised microscopy: validation stainings	digitalised stainings	New	Digital	Experimental	.czi	10GB	

snRNAseq raw sequences	raw data after RNA sequencing	reused	Digital	Experimental	.fastq	2TB	
snRNAseq raw counts	gene counts calculated from the RNAseq data	reused	Digital	Experimental	.mtx	500GB	
Ann data objects of RNAseq data	objects of RNAseq data, annotated with metadata and dimensionality reduction info	New	Digital	Experimental	.adata	500GB	
GeoMX raw counts	raw data from the GeoMX experiments	reused	Digital	Experimental	.txt	5GB	
LCMD-SP: raw peptide sequences	raw peptide sequence data from the LCMD-SP layer	reused	Digital	Experimental	.raw	500GB	
LCMD-SP: protein counts	protein counts calculated from the LCMD-SP raw data	reused	Digital	Experimental	.txt	1 GB	
BAL bulkRNAseq raw data	raw data from BAL bulkRNAseq experiments	New	Digital	Experimental	.txt	10GB	
BAL bulkRNAseq count data	count data from BAL bulkRNAseq experiments	New	Digital	Experimental	.txt	5GB	
BAL Luminex data	data from the Luminex experiment	New	Digital	Experimental	.txt	100MB	

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:

We will reuse data generated in another project from the researcher (under the same ethical approval), which is still confidential (so no DOI available)

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

Ethical approval was granted for the project and informed consent was obtained from all included patients

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

We will use clinical pseudonymized data to assess experimental findings with clinical outcome of the patients. We will also use survival data as this is an important outcome parameter.

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.

- No

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

The initial dataset is also used in a project which involves a third party. However, the dataset itself is fully owned by the researchers. Exploitation of the results of this specific project is not restricted.

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

The initial dataset is also used in a project which involves a third party. However, the dataset itself is fully owned by the researchers

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).

TYPE OF DATA	METADATA STORAGE MODALITY	METADATA STORED
clinical metadata	excel file	clinical variables per pseudonymized sample
ex vivo CT scan	included in DICOM format	technical specs of scanning protocol
microCT scan	separate .txt file	technical specs of scanning protocol
stainings/4i panel	included in.czi format	technical specs of scanning protocol
stainings/4i panel	separate word document	standing process (e.g., antibdoy concentration, etc.)
snRNAseq data	separate word file	technical specs of nuclei isolation
snRNAseq data	separate .rmd file	technical specs of library prep and sequencing
GEoMx raw counts	separate .rmd file	technical specs of sequencing and ROI determination
LCMD-SP	separate word file	LCMD process, process of niche delineation
LCMD-SP	sparate .jpg and.palm files	niche delineation
LCMD-SP	separate .txt file	clinical metadata
BAL samples	excel file	clinical metadata
BAL samples	separate word file	RNAseq and luminex processing

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

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

Where will the data be stored?

The data will be stored on KU Leuven servers.
GBW-0017_LTx (15TB) will be used for rw data storage, including metadata en SOP. GBW-0076_LTx (0.5TB) will be used for storage of data which is used in day-to-day work.

How will the data be backed up?

Alle KU Leuven servers are protected with standard back-up provided by KU Leuven itself.
Spare -80°C freezesr are available in the KU Leuven biobank to prevent loss of frozen samples in case of dysfunction of freezers

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

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

KU Leuven servers are are protected with two-factor authorization (as all KU Leuven ICT systems)

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

- long term storage (GBW-0017_LTX): € 104,42 / TB / year; covered by the BREATHE lab
- day-to-day data storage (GBW-0076_LTx): € 503,66 / TB / year, covered by the BREATHE lab

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 digital data will be preserved on the long-term storage server GBW-0017_LTX. All FFPE blocks will be stored in the lab, frozen material in -80°C freezers in the biobank.

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

All digital data will be preserved on the long-term storage server GBW-0017_LTX. All FFPE blocks will be stored in the lab, frozen material in -80°C freezers in the biobank. I

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

All costs for digital preservation as well as -80°C costs are covered by the BREATHE lab.

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, …)

The raw and analyzed RNAseq and spatial omics datasets wil be made available

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

all data include clinical metadata of the samples included. Hence, we feel restricted access is most appropriate.

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
- Yes, Ethical aspects

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

KU Leuven RDR

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.

to be considered

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?

use of the RDR repository of KU Leuven is free for KU Leuven researchers

6. Responsibilities

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

In the project, budget for two PhD students is included. Data documentation and metadata will be managed by these students

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

In the project, budget for two PhD students is included. Data storage and backup will be managed by these students

Who will manage data preservation and sharing?

Laurens De Sadeleer

Who will update and implement this DMP?

Laurens De Sadeleer