Unraveling the regulatory role of transcription factor Prdm16 on the cellular and molecular mechanisms involved in the pathogenesis of pulmonary arterial hypertension at a single-cell resolution.

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	
Single-nuclei RNA sequencing raw data sets	Acquired from single-cell core facility	Generate new data	Digital	Experimental	.fastq files, .gz files, .bam files	<100 GB	
Single-nuclei RNA sequencing processed data sets	Processed data allowing downstream analysis of single-cell results	Generate new data	Digital	Experimental	.xls files, .jpeg files, .pdf files, .txt files, .csv files	<1TB	
Raw data set hemodynamic measurements	Acquired via PV-loop measurements in the RV	Generate new data	Digital	Experimental	.adicht	<100 GB	
Hypoxia cage monitoring data	Continuous monitoring hypoxia	Generate new data	Digital	Experimental	(generated by Anawin software)	< 100GB	
Vaso-reactivity data	Performed on pulmonary artery isolated from mouse	Generate new data	Digital	Experimental	.xls	<1 GB	
Microscopic images of immunofluorescence and histochemical staining	Images from both mouse and human tissues	Generate new data	Digital	Experimental	.jpeg files, Zen .zvi files, .czi files, .tiff files	<100GB	
Mouse genotyping gel pictures	Genotyping results from gel electrophoresis	Generate new data	Digital	Experimental	.jpeg files	<1GB	
Electron microscopic images	Images from mouse tissues	Generate new data	Digital	Experimental	.jpeg files, .tiff files	<100GB	

Western-blot gel pictures	Images of Western-blot gels	Generate new data	Digital	Experimental	.jpeg files	<10GB	
Statistical analysis		Generate new data	Digital	Experimental	Prism .pzfx files, R package .r files	<100GB	
Composition figures, digital images	Figures for abstracts, posters and publications	Generate new data	Digital	Experimental	.eps files, Acrobat .pdf files, Adobe Indesign .indd files, Adobe Illustrator .ai files, .TIFF files, .jpeg files, .PNG files	<100GB	
SOPs	Staining, qPCR, genotyping, nuclei isolation	Generate new data Reuse existing data	Digital	Experimental	Word.doc files	<1GB	
quantitative (q)RT-PCR data	On mouse or human tissues or cells	Generate new data	Digital	Experimental	.cvs files	<1GB	
Sample inventories		Generate new data Reuse existing data	Digital	Experimental	.xlsx files	<1GB	
Single nuclei suspensions	Single nuclei isolated from murine lung	Generate new data	Physical			NA	2 boxes of samples stored in -80 freezer
Tissues	Paraffin embedded (human and animal origin)	Generate new data	Physical			NA	Approximately 300 tissue blocks
Paraffin sections	Human and animal origin	Generate new data	Physical			NA	30 drawers of 100 slides
Snap frozen tissues	Human and animal origin, stored at -80 degrees Celsius	Generate new data	Physical			NA	2 boxes of samples stored in -80 freezer
Cryo embedded tissues	Human and animal origin, stored at -80° Celsius	Generate new data	Physical			NA	5 boxes of samples stored in -80 freezer
Cryo sections	Human and animal origin, stored at -20° Celsius	Generate new data	Physical			NA	10 boxes of slides stored in -20 freezer
Pulmonary arterial endothelial cells (PAECs)	PAECs isolated from Swan-Ganz catheter, stored at -80° degrees Celsius	Generate new data	Physical			NA	1 box of samples stored in -80° freezer
Protein/RNA/cDNA	Human and mouse origin	Generate new data	Physical			NA	3 boxes of samples stored in -80°C freezer

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
- Yes, animal data

Human data: S57114 and S63978 (UZ Leuven, BELGIUM) and N° AC-2020-4282 and n°DC-2021-4467 (INSERM, FRANCE).
Animal data: ECDN° Creation Luttun/2023, ECDN° P094/2023

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

- N° AC-2020-4282 and n°DC-2021-4467 (INSERM, FRANCE)
- S57114 and S63978 (UZ Leuven, BELGIUM)
- Privacy Registry Reference: UZ Leuven patient registry
- Short description of the kind of personal data that will be used: Personal data include demographic information (age, gender, length, body weight, body mass index), medical information (hemodynamics, exercise capacity (6-min walk distance or 6MWD), echocardiography, aetiology, mortality/survival, NYHA functional class) and general health status (smoking behaviour) and therapy (medical treatment). Data will be collected from patients with pulmonary arterial hypertension who come to the Pneumology Unit at UZ Leuven (headed by Prof. Dr. D. Lieven) for diagnosis or follow-up Swan-Ganz right heart catheterization (UZ Leuven, BELGIUM) and from end-stage patients with pulmonary arterial hypertension who undergo a lung transplantation or from patients that undergo a lung tumor resection whom serve as controls (UZ Leuven, BELGIUM and INSERM, FRANCE).

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.

- Yes

- Collaboration with non-KU Leuven partner (INSERM, FRANCE): agreements related to exploitation and dissemination have been written down in a collaboration agreement signed on March 22nd, 2023 by all parties involved in the project.
- Dissemination and exploitation of patient data to 3rd parties obtained in collaboration with Prof. M. Delcroix is restricted, and we will ensure privacy of the patients by pseudonymisation. A link between the patient and the dataset will be encoded. This code will not contain information that could lead to the identification of the patient and will be stored in an encrypted way on a different location than the pseudonymised data. The unique 'eenmalig administratief dossiernummer' (EAD number) in the UZ Leuven patient database will not be used. The code will only be accessible to Marion Delcroix (collaboration) and Rozenn Quarck (co-promoter), the clinical data manager of BREATHE.
- Human data from French PAH patients will not be sent to other participants of the project. Generated raw data from pseudonymised samples will be sent to Prof. David Montani who will depseudonymise and analyse the data. Prof. Montani, partner of the project with Dr. Frédéric Perros, is Professor of Respiratory Medicine at the French National Referral Centre for Pulmonary Hypertension in the Department of Respiratory Medicine, at the Hôpital Bicêtre, Le Kremlin-Bicêtre, Paris-Sud University, France and team leader in the INSERM U999 laboratory « Pulmonary Hypertension: Pathophysiology and Novel Therapies», Hôpital Marie Lannelongue, Le Plessis Robinson.

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

Since this project encompasses a collaboration with a non-KU Leuven partner (Inserm, France), agreements related to IP rights and ownership have been written down in a collaboration agreement signed on March 22nd, 2023 by all parties involved in the project.

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).

Main results and methods will be published in peer-reviewed journals (open access as required by FWO regulations) and all publications will be archived in Lirias, the digital KU Leuven document repository.
All **digital data** generated in the project for each WP and associated metadata will be archived digitally and a searchable database format (Excel or Access) will be implemented. Electronic lab note books will be used and templates have been designed for writing protocols/SOPs, for excel spreadsheets for raw data and (statistical) analysis. When raw data are uploaded on repositories, keywords will be affixed along with readme files containing the needed information for reuse. In the final stage of the project, a master index file with the combined metadata for each WP will be generated and archived on a non-editable drive of the host institution KU Leuven ('K drive').
All **physical data** collected during the course of the project will be stored at designated storage places (at room temperature or frozen) and location and preservation method of the biological samples (tissues, tissue sections, genetic material,...) will be documented digitally (.xlsx files).

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 will be a combination of machine-generated and manually generated metadata. Metadata of raw images (file size, pixel number, acquisition date, settings, etc.) and qRT-PCR are captured automatically and saved on the server together with the corresponding data files. Other metadata (on quantification procedures, biochemical analysis, *etc.*) are mostly captured manually and logged in lab notebooks or in searchable Excel/Access databases. For these metadata, we will progressively design own metadata standards using http://dublincore.rog/. We will also consider archiving our data using general data repositories (https://figshare.com/ and https://zenodo.org/). RNAseq data will be uploaded to the GEO repository which uses the MIAME standard.

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

Where will the data be stored?

Shared network drive (J-drive)
OneDrive (KU Leuven)
During the project **digital data** will be stored at different locations, depending on the type of data and accessibility. Non-personal data will be stored on the researchers' computers, on the KU Leuven network editable drives where the principal investigator has reserved dedicated space (the J drive for data that needs to be accessible daily and is exchangeable between research lab members or (later during the project) the L drive for longer-term storage of large data files that do not need to be frequently accessed). Personal patient-related data will be stored on the UZ Leuven central server. Both UZ Leuven and KU Leuven servers are compatible with GDPR regulations.

All **physical data** collected during the course of the project will be stored at designated storage places. An inventory of each storage place is available. Human-derived material where possible will be registered and stored at the UZ Leuven Biobank.

How will the data be backed up?

Standard back-up provided by KU Leuven ICTS for my storage solution
Personal back-ups I make: All of the data and documents on the researcher's computer are automatically synchronized to the KU Leuven Onedrive cloud with a capacity of 100GB per user. In addition, data will be stored in a designated folder on an external hardware disk (8TB).

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

Digital data: I have access to 100 Gb storage space on the KU Leuven Enterprise Box. This should be sufficient, except for large volume datasets (e.g., microscopy images, sequencing datasets). For large volume datasets, space is reserved on the editable KU Leuven network drives (J or L). Also, space is reserved on the read-only K drive for storage after the end of the project or after publication of manuscripts. Storage on the J and K KU Leuven network drives is extendable by blocks of 100 Gb), storage on the L drive is extendable per 5 Tb, hence, by acquiring additional storage space based on the project's requirements, sufficient storage can be made available.
For storage of **physical data**, sufficient storage space is available.

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

Digital data. Both UZ Leuven and KU Leuven network servers are compatible with GDPR regulations and allow for secure storage of personal data. KU Leuven ICTS services provide the option to control data access for authorized persons only (in this case, research lab members involved in this project). As mentioned above for personal data stored on the UZ Leuven central server, access will be restricted with access right management only to clinical data manager Rozenn Quarck (co-promoter) and Prof. M. Delcroix (collaborator).
All **physical data**, printed forms and notebooks are stored in the labs in locked cabinet. Access to the lab is secured and badge controlled. In case samples are stored at the Biobank, secure storage is guaranteed by controlling access to the storage location. Access to this location will be limited to one person and one back-up person per research group.

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

Costs for **digital data** storage and back-up during the project have been included in the research budget of the project. The current cost rate for the KU Leuven network drives are: 45.08€/y/100Gb block (J) and 9.514€/y/100Gb block (K), 475.7€/y/5Tb block (L-drive).

Costs for **physical data** are only applying for biological samples of human origin. The UZ Leuven Biobank is currently still in the process of calculating the yearly storage cost per sample.

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** and metadata will be retained for 5 years after the project (per the requirements of FWO). The same term will be applied to **physical data**. Long-term storage of personal data additionally requires GDPR clearance, which has been obtained upon approval from the Ethics Committee of UZ Leuven.

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

Digital data will be archived on the KU Leuven K drive for storage of read-only data.

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

Cost rate for storage on the K drive is 9.514€/year/100Gb. To store a total of 5 Tb for 5 years, the estimated cost hence is 2,378.500 €. Costs will be allocated to 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 an Open Access repository

Main findings of the research with all supporting processed data will be made available through publication in peer-reviewed journals with open access policies (as required by FWO). All manuscripts will also be deposited in the KU Leuven Lirias digital repository. Raw RNAseq data will be made available publicly upon acceptance of the manuscript. Other raw data related to published manuscripts may be available upon specific request as will be stated in a data availability statement included in the published manuscripts.

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

NA

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

Personal data will be shared only with certain third parties (as will be specified in the GDPR addendum to the informed consent form) if needed, thereby always ensuring the privacy of the donors.

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

Main findings of the research with all supporting processed data will be made available through publication in peer-reviewed journals with open access policies (as required by FWO). Raw RNAseq data will be made available publicly through the GEO repository upon acceptance of the manuscript.

When will the data be made available?

Upon publication of research results. Other data will be made available upon request, where considered appropriate, following publication.

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

CC-BY 4.0 (data)

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

Yes, a PID will be added upon deposit in a data repository.
For raw snRNAseq data, a GEO accession number will be provided.

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

For sharing **digital data**, no sharing costs are foreseen. For sharing **physical data**, Material Transfer Agreements will have to be put in place which will be mutually signed. Shipping costs would be covered by either party (through the FWO budget in case of the provider) as long as the costs are low, however, significant sharing costs will be expected to be borne by the requestor.

6. Responsibilities

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

All researchers involved in this project.

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

All researchers involved in this project.

Who will manage data preservation and sharing?

The supervisor (A. Luttun)

Who will update and implement this DMP?

The supervisor (A. Luttun)/co-promotor (R. Quarck)/the grant holder (A. Mahy)