Monogenic immune disorders in adults: from Mendelian inheritance to somatic mosaicism.

A Data Management Plan created using DMPonline.be

Creators: Willem Roosens, n.n. n.n.

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: n.n. n.n.

Data Manager: Jonathan Cremer

Project Administrator: Willem Roosens

Grant number / **URL:** 11PDZ24N

ID: 202623

Start date: 01-11-2023

End date: 31-10-2027

Project abstract:

Disease manifestations of monogenic immune disorders, such as immunodeficiency, autoinflammation or autoimmunity, can be delayed until adulthood and this is mediated through various genetic mechanisms, extending beyond Mendelian inheritance. Despite the discovery of multiple genetic causes in the last decade, most adult patients lack a molecular diagnosis preventing personalized care. Furthermore, while the role of pathogenic somatic variants is becoming more widely recognized, the underlying molecular mechanisms remain unknown. Using state-of-the art sequencing technology and in vitro functional validation tools, we aim to define new monogenic inborn errors of immunity and unravel disease mechanisms in selected adult patients presenting with symptoms suggestive of immune dysregulation. Thorough functional validation of candidate variants is pivotal for individual patient care (diagnosis, rational design/choice of therapy, prognosis, and familial genetic counseling) and holds the promise to provide essential insights in human immunology through analysis of experiments by nature, as this research track has demonstrated in the recent past. By deciphering the molecular determinants of carefully selected adult-onset immune disorders, this project will advance the area of immunogenetics research, explore uncharted territory, and pave the way for precision treatment.

Last modified: 18-04-2024

Monogenic immune disorders in adults: from Mendelian inheritance to somatic mosaicism. Application DMP

Questionnaire

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

(A) Patient metadata: pseudonymised and encoded databank (MS Excel and REDCap)
(B) Cell culture parameters: manual entry in lab book + microscopy images (.jpg).
(C) Reagantia, oligonucleotide primers, antibodies will be stored at required conditions in our lab and databased in shared spreadsheet (.xlsx)
(D) Raw experimental data: incl. data from FACS (.fcs), immunofluorescence (.tiff), qPCR (.xlsx), Western blot (.tiff), TEER/FD4 permeability data (.xlsx), DNA/RNAseq (.fastq, .bam) and metadata file (.doc)
(E) Analysed data: incl. textual data (.doc, .pdf), presentations slides (.ppt), figure files (.png), data analysed by Prism (.pzfx), FlowJo (.wsp), SPSS (.dat/.sps) and R (.RData/.rds).

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)

Personal research data will be stored at the KU Leuven LUNA server in the GBW Allergy and Clinical Immunology group. Data will be saved on the shared J-drive, accessible for all lab members. Each lab member has their own folder. Databases, protocols, etc. are centralized within this group making the data accessible for all lab members. Data are backed-up by LUNA server on a daily basis. A fixed laboratory staff member will preserve the data at least 5y after the end of the research. Data will be stored during and after the research at our research group.

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)

There are no other issues related to data management.

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)

There are no other issues related to data management.

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

There are no other issues related to data management.

Monogenic immune disorders in adults: from Mendelian inheritance to somatic mosaicism.
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	
INFORMED CONSENT	Informed Consent s58466 and s67807	Generate new data	Physical	Observational	NA	NA	10 binders
PATIENT METADATA	Pseudonymised and encoded database in EXCEL (for study s58466) and REDCAP (for study s67807)	Generate new data Reuse existing data	Digital	Observational	.xls .xlsx .csv	<1GB	
PATIENT AND HEALTHY CONTROL BIOSAMPLES	Secondary use of stored human body material (tissue, cells or extracted genomic DNA) and prospective collection of blood samples (including isolated PBMCs), bone marrow samples, buccal swabs, tissue and skin biopsies, nail clippings, hair roots, skin biopsies and derivatives (including genomic DNA, RNA, copy DNA, cell lysates)	Generate new data Reuse existing data	Physical				Approximately 100x 1,8 mL cryovials 300x 1.5 mL Eppendorf tubes 50x sterile 40mL containers
IN VITRO PATIENT-DERIVED AND COMMERCIAL CELL LINES	In vitro culture of patient-derived cells including generation of immortalized cell lines and storage of commercially available cell lines	Generate new data Reuse existing data	Physical				Approximately 100X 1,8 mL cryovials
ANTIBODIES, REAGENTIA, OLIGONUCLEOTIDE PRIMERS, VIRAL VECTORS		Generate new data Reuse existing data	Physical				Approximately 10X 9x9 microtube freezer storage boxes

LAB INVENTORY	Shared spreadsheet containing orders, reagentia, oligonucleotide primers, antibodies etc.	Generate new data Reuse existing data	Digital	Observational	.xls .xlsx .csv .tab	<100 MB	
LAB NOTEBOOK	Primary record of research with documentation on hypothesis, experiments, initial analysis or interpretation of results	Generate new data	Physical	Observational			2 to 3 paper lab notebooks
RAW EXPERIMENTAL DATA	Source (primary) data from experiments including flowcytometry, immunofluorescence, qPCR, Western Blot (SDS PAGE analysis), ELISA, multiplex cytokine analysis, FD4 permeability data, DNA & RNA quantification, other microplate-based colorimetric/fluorometric assays	Generate new data Reuse existing data	Digital	Experimental	.fcs .wsp .tiff .xlsx .csv .tab .gif .jpeg	< 100 GB	
PROCESSED EXPERIMENTAL DATA	Processed experimental data to ease analysis (f.e. cleaned up .fcs files, excluding debris, doublets,..)	Generate new data	Digital	Experimental	.fcs .wsp .tiff .xlsx .csv .tab	< 100 GB	
ANALYSED EXPERIMENTAL DATA	Results of the analysed data: incl. textual data, tables, presentations slides, figure files, data analysed by Prism, FlowJo, SPSS and R.	Generate new data	Digital	Experimental	.doc .pdf .ppt .png .pzfx .wsp .dat/.sps .RData .rds	<100 GB	
SCIENTIFIC OUTPUT	Manuscripts, scientific presentations, scientific posters	Generate new data	Digital	Compiled/aggregated data	.pdf .pptx .docx .psd	<100 GB	
RAW DNA/RNA SEQUENCING DATA	Source data from sequencing platforms (Illumina, Pacbio or Nanopore) and Sanger sequencing	Generate new data Reuse existing data	Digital	Experimental	.bcl .fastq .fasta .ab1 .seq	> 50 TB	
PROCESSED DNA/RNA SEQUENCING DATA	Analysed sequencing data (aligned reads, variant calls, etc.)	Generate new data Reuse existing data	Digital	Compiled/Aggregated data	.sam .bam .vcf	< 1 TB	
R data	R scripts and markdowns	Generate new data	Digital	Compiled/aggregated data	.R .rmd	<100 GB	

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 datasets and biosamples collected in the framework of ongoing clinical cohort studies locally stored in our databases (S58466).

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

The data will contain sensitive personal data about the subject and the subjects' health condition. Human bodily materials will also be collected and processed and may result in genetic data being analysed.
Ethical approval for the collection and analysis of human bodily material is granted by the Ethics Committee Research UZ/KU Leuven for work package 1 (**s58466**) and work package 2 (**s67807**).

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

After signing of the informed consent, relevant **patient metadata** (including demographics, medical history familial history, medication, laboratory results) will be obtained from the electronic medical patient file (Klinisch Werkstation, KWS).
This non-anonymous data will be entered in an pass-protected Excel file on the UZ Leuven server (for WP1, s58466) OR the database program **REDCap** (for WP2, s67807). Here we will be using **pseudonymisation** of the participant's personal data and samples. Identifiers are stored separately from study data and replaced with a unique code to create a new research-specific identify for each participant for the purpose of this study. The key linking the patients name and electronic medical file number to this identity will be kept in an Excel file on a secured UZ Leuven server, separate from the research data and accessible only to the principal investigator (treating physician) and co-investigator (PhD student, Willem Roosens, MD), enabling them to go back to the patient chart, when needed.

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

To transfer patient-derived material and associated metadata to the lab of professor Jonas Demeulemeester (Integrative Cancer Genomics, KUL-VIB) a material transfer agreement (MTA) is executed between
Katholieke Universiteit Leuven, for the purposes of this Agreement represented by University Hospitals Leuven, with its offices at Herestraat 49, 3000 Leuven, Belgium acting on request of Professor **Prof. Dr. Rik Schrijvers** of the **Department of Microbiology, Immunology and Transplantation**, hereinafter referred to as "**Provider**";
and
VIB, with offices in Rijvisschestraat 120, 9052 Zwijnaarde, VAT number BE0456.343.923.

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

In the material transfer agreement (cf. supra), the two parties agreed on a joint ownership regarding patentable results.

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).

All data will be collected according to predetermined standard protocols. If a deviation from the protocol occurs, it will be noted in the appropriate location (patient or sample checklist, laboratory notebook, experiment summary)

README.txt files are created for interpreting data files. R scripts are created with the necessary explanations using # inside the scripts or with R.markdown files.
The codebook of each coworker will contain information on date of puncture, study design, sampling methodology, fieldwork, variable-level detail, and all information necessary for a secondary analyst to use the data accurately and effectively

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

Patient samples will be collected following SPREC guidelines as mandated by UZ Biobank. Samples will be registered in the provided software (via PeopleSoft) and database provided by UZ Biobank.
Flow cytometry experiments will be documented following the MIFlowCyt standard. This standard collects information about the samples, the experiment, the reagents, the instrument and the analysis.
QPCR experiments will be documented using the MIQE standard.

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

Where will the data be stored?

All patient personal information will be stored on the UZ Leuven IT systems and local hospital IT systems, accessible only to a restricted number of individuals (project PI and Drs. Willem Roosens) on a need-to-know basis.
All pseudonymized experimental data, metadata and protocols will be stored on the KU Leuven network drives:
Personal working files and actively used files (by multiple people) can be stored on OneDrive and Sharepoint Online (Teams-site) for ease-of-use purposes but must be archived on the project J:-drive.
For long-term preservation and backup, all electronic files will be subsequently stored on KU Leuven ICTS-managed servers, on the archival network drive (so called K:-drive). The data will be compressed as a .zip archive and uploaded to this network drive.
All samples will be stored according to the host laboratory Standard Operating Procedures (SOPs):

Liquid nitrogen vapor phase	Cryogenic storage of live cells at KUL/UZL cryogenic facility
-80 °C	Nucleic acids, protein samples, tissue samples, project specific reagents
-20 °C	Short term storage of supernatants, DNA samples, project specific reagents
4 °C	Short term storage of biological specimens, project specific reagents
Room temperature	Formalin fixed paraffin embedded tissue samples and sections, histochemically stained slides

How will the data be backed up?

The network J:-drive has a backup regime in the following manner: “hourly” backup at 08.00h, 12.00h, 16.00h and 20.00h of which the last 6 are kept; daily backup at 24.00h of which the last 6 are kept; weekly backup on Saturday at 24.00h of which the last 12 are kept. Furthermore, the primary server in the Heverlee datacenter is mirrored hourly to the secondary data center in a different geographic location for disaster recovery.
The network K:-drive has a backup regime in the following manner: daily snapshots of which the last 14 are kept. Furthermore, immediate mirroring of files from the primary datacenter to the secondary datacenter is in place for disaster recovery.
The Sharepoint Online (Teams-site) is a service provided by the Microsoft corporation as part of the Microsoft 365 Education license. Data is protected from user error by version management (the last 500 major versions are kept). Furthermore, the data is protected from infrastructure calamities by the provider.
OneDrive is a cloud service provided by the the Microsoft corporation as part of the Microsoft 365 Education license. Data on the user’s device (e.g. personal computer) is synchronized to the Microsoft servers. Version history is available, data on the servers is protected from infrastructure calamities.

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

There is sufficient storage and back-up capacity on the KUL OneDrive server and all UZ Leuven and KU Leuven servers: KUL OneDrive provides 2TB storage for non-confidential data, which is estimated to be adequate for the entire research project. In case more storage capacity is needed, this can be enlarged to a maximum of 5 TB upon request.

The KUL/UZL “J-drive” is an easily scalable system, has unlimited maximal size and is expandable in blocks of 100GB.
The network storage provided by KU Leuven ICTS is provisioned for our project’s needs and can be scaled up upon simple request. Our project does not produce data on a scale that would exceed the provisioned reserves of the datacenter - except for raw DNA and RNA sequencing (NGS) data, that will be stored in a Google Cloud Bucket that is hosted by the sequencing provider (Genomics Core Leuven, UZ Leuven). The storage policy can be modified when requirements change. The storage service is billed on a quarterly basis.

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

The network storage provided by KU Leuven ICTS is secured appropriately for strictly confidential data. Access is restricted to members of the research project and system administrators, all members are registered in a logbook when access is given and revoked. The storage can only be accessed through the KU Leuven LUNA domain (either through a KUL configured PC directly connected to the network, or through a PC connected via VPN), or via a password protected web portal.
Access to the Sharepoint Online (Teams-site) is restricted to members of the research project and system administrators, all members are logged when given and revoked access. Confidential and strictly confidential data can only be stored on this medium if KUL Multi-Factor Authentication for the user account is activated (KUL ICTS code of conduct).
Access to OneDrive for Business is tied to the user’s personal KUL account. Access to data stored within a OneDrive can be shared by the user, at their discretion. The KUL ICTS code of conduct applies.
Access to the UZ Leuven server of the Department of Internal Medicine is restricted to members of the research project and system administrators. Access to the data as well as the access level will be limited on a project need and individual basis. All data files are password-protected and benefits from the firewall services provided by the UZ Leuven IT department. Personal data is always pseudonymized. The coding key to patient information of linked pseudonymized data does not carry any personal identifiers and all records containing the identity of each participant will be kept private and confidential.

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

The total estimated cost of data storage during the 4 years of this FWO project is less than 5,000 EUR. This estimation is based on the following costs:
* €519/TB/Year for the “J-drive”.
* €960/TB/2 Years for the Google Cloud Bucket hosted by Genomic Core Leuven
(simulation assumes 3 month of standard data storage (STD), 9 months of coldline (CL) and 12 months of archive (ARC), including 5 downloads: 1 STD, 2 in CL, 2 in ARC)
Costs for Sharepoint Online and OneDrive are covered by the Microsoft 365 Education License provided by KU Leuven ICTS.
The physical samples will be stored within available freezers in our laboratory.
These costs are covered by the budget of the project lead (Rik Schrijvers).

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...).

The minimum preservation term of 10 years after the end of the project mandated by KU Leuven will be applied to all digital datasets (in accordance with the informed consent). Datasets that are eligible for longer preservation for future use will be stored in data repositories.

All remaining physical samples will be transferred to the UZ/KU Leuven Biobank after the study has been concluded. For s67807 (WP2) the Biobank application has been approved (BB-GEN002-FO03).

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

Personal data of study subjects will stay in the hospital clinical information systems.
Experimental data will be archived on KU Leuven network drive K.
Upon final publication of results, datasets may be published in the KU Leuven RDR or domain-specific repositories including the European Genome-Phenome Archive.

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

The total estimated cost of electronic data storage for 10 years after the end of the project is < €100. This estimation is based on 70EUR/Tb/year. All costs for data preservation will be covered by our own funding.

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, …)

Regarding **digital data**, only published data will be available in the form of publications or other dissemination of scientific work. All data will be anonymised when disseminated. More data can be made available or shared after permission of the responsible person (prof. Rik Schrijvers). Non-published data will remain confidential until a final decision on publication of the data has been taken. Clinical data about the patients are confidential, subject to compliance with applicable personal data protection laws, and not publicly available.
Raw .fcs-files will be made available after publication as specified by MIFlowCyt. R script may be made publicly available of GitHub.

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

Personal data will only be available to the head investigator and co-investigators in agreement with the patients' consent.
Afterward, the data will be shared between the PI and co-PI's and within the research unit but always respecting the pseudonymization.
Sequence data will be uploaded in an open access repository and shared upon request.
Data without sharing restrictions will be shared through peer reviewed publications

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, Intellectual Property Rights

Yes, Intellectual Property Rights Some of the analyses will be performed in collaboration with other laboratories (cf. Material Transfer Agreement mentioned above).
With respect to patient data and biomaterial: patient privacy is respected and permission to share data/samples is obtained by the informed consents and stipulates clearly who will have insight in the pseudonymized data. Data will not be shared with other (third) parties, not participating in the study.

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

Since we are working with sensitive data, KU Leuven RDR will be used. Sequence data will be uploaded in an open access repository with restricted access control such as the EBI European Genome-phenome Archive (EGA). Data without sharing restrictions will be shared through peer reviewed publications.

- **Double-coded patient data**
  Upon publication, all double-coded patient details supporting a manuscript will be made publicly available as supplemental information.
- **Research documentation**:
  All protocols used to generate published data will be described in the corresponding manuscript(s), and the related documentation will be included as supplementary information. These data and all other documents (raw data) deposited in the notebooks are accessible to the research staff and will be made available upon request.
- **Manuscripts**: All scientific publications will be shared openly. Manuscripts submitted for publication will be deposited in a pre-print server such as bioRxiv. Before the end of the embargo or in cases where sharing the post-print is not allowed due to copyright agreements, a pre-print version of the manuscript will be made available. (Pre-print) publications will also be automatically added to our institutional repository, Lirias 2.0, based on the authors name and ORCID ID.
- **Algorithms, scripts and software**: All the relevant algorithms, scripts and software tools driving the project will be described in manuscripts and/or on GitHub (https://github.com).
- Data that do not support publication will be either deposited in an open access repository or made available upon request by email. Data will be reused by transfer via Belnet Filesender or secure cop

When will the data be made available?

All research outputs will be made openly accessible, at the latest, at the time of publication (or preprint deposition). No embargo will be foreseen unless imposed e.g. by pending publications.

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

No specific license has been selected yet. Data usage licences will be discussed with involved parties before any licences are granted. This will be done in close collaboration with the RDM Support group at KU Leuven and will likely be a Creative Commons-type license.

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

All datasets published in KU Leuven RDR will have a DOI assigned.

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

Each researcher can currently deposit up to 50 GB per year in KU Leuven RDR at no charge. This will likely be sufficient for the project's datasets. Peers may use the data at no cost under condition of co-authorship. Commercial organizations are not anticipated to share data (or biomaterial) at the moment but if secondary use of material would later on be possible for them (after informing the patient/legal representative of this third party commercial use of this material), they will have to pay a fee that will be determined by LRD-KU Leuven.

6. Responsibilities

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

The PI (Rik Schrijvers), Willem Roosens (doctorandus) and Jonathan Cremer (Lab Technician and Lab DMP reference person) will manage documentation and metadata during this research project.

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

Jonathan Cremer and Sabien Fevery will manage data storage and backup during the project.

Who will manage data preservation and sharing?

The PI (Rik Schrijvers), UZ Leuven co PI (s58466, Isabelle Meyts), KU Leuven co PI's (Stephanie Humblet-Baron, Jonas Demeulemeester) and PhD student (Willem Roosens)

Who will update and implement this DMP?

Rik Schrijvers, Willem Roosens and Jonathan Cremers will update and implement this DMP. The PI bears the end responsibility of updating and implementing this DMP.