Reclassification of kidney transplant rejection and the role of innate allorecognition

A Data Management Plan created using DMPonline.be

Creator: Maarten Naesens

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: G038024N

ID: 207407

Start date: 01-01-2024

End date: 31-12-2027

Project abstract:

Kidney transplant rejection is a complex interplay of the genetic disparity between donors and recipients and the immune system of the host by these mismatched antigens. We and others have demonstrated the role of innate allorecognition, as exemplified by NK cell–mediated microvascular inflammation through missing self, and by monocyte-driven allorecognition. The Banff classification for kidney transplant rejection does not represent the full spectrum of disease phenotypes and the novel allorecognition pathways. In this project, we aim to map the complex interplay of various allorecognition mechanisms and connect this to a spectral view of kidney transplant rejection, moving away from a monocausal and dichotomous approach to rejection classification. In three work packages, we will (1) evaluate the relative contributions of different allorecognition mechanisms in the phenotypic spectrum of rejection, (2) develop a causally informed reclassification system of rejection, including parameters of cell composition and disease stage (activity and chronicity) with biopsy-based transcript analyses, and (3) investigate the potential of deep learning on whole slide images of transplant biopsies with causal information and information on immune cell composition. This will allow better classification of rejection phenotypes according to the underlying cause, specific immune cell infiltration/activation pathways and disease activity/chronicity, paving the way for true precision medicine.

Last modified: 15-05-2024

Reclassification of kidney transplant rejection and the role of innate allorecognition
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)

1. Demographic, clinical and histology data come dircectly from the included patient files. These data are stored in SAS/CSV/Redcap format (demographic and clinical data) or in Microsoft Access format (histology data).
2. Genotyping data are collected in .csv format. They are stored as CSV and SAS files.
   Digital images will be saved as .ZSI and TIFF files (scans), or as .GIF or .JPEG files (microscopic images). The total volume expected is 2 Tb.
3. Genome-wide molecular expression data are collected and stored as COUNT files, DAT files, CHP files and CEL files. The total volume expected is 500 Gb. Aggregated molecular expression files will be collected and stored as CSV files.
4. Single-cell expression data are collected and stored as TSV and MTX files. The total volume anticipated is 1.5 Tb.
5. Spatial transcriptomics data are collected and stored as cell-feature matrix (cell_feature_matrix) in three file formats: the Market Exchange Format (MEX), the Hierarchical Data Format (HDF5), and the Zarr format, with anticipated volume of 8 Tb.

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)

1. Designation of responsible person: Maarten Naesens, PI of this project
2. Storage capacity/repository: during and after the research, the data will be stored on the centrally managed KU and UZ Leuven servers with automatic daily back-up procedures and version tracking. Some of the data sets (e.g. bulk transcriptomics, scRNAseq data etc) are/will not only be stored locally, but be made available through public repositories.

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)

Not applicable

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)

All data come from patients. They will be coded (i.e. pseudonymized). There continues to be a link between the data and the individual who provided it. The subjects' identifiers will however be stored separately (site file) from their research data and replaced with a unique code to create a new identity for the subject. This code is stored on the UZ Leuven server which is password protected, but which also allows to consult the electronic medical chart of the patient stored on UZ Leuven Hospital servers, only if necessary.

In addition, we will store all data on the central servers of the KU and UZ Leuven, which are protected against unauthorized access by firewalls.

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

Not applicable

Reclassification of kidney transplant rejection and the role of innate allorecognition
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.

Dataset Name	Description	New or reused	Digital of Physical	Digital Data Type	Digital Data Format	Digital Data Volume	Physical Volume
TEMPLATE+	Database	Reuse existing data collected in the hospital	Digital	Observational data	.csv forma	<100MB	NA
scRNAseq	Expression data	Reuse	Digital	Experimental	.fastq.gs format	<10TB	NA
Bulk mRNA	Expression data	Reuse	Digital	Experimental	.csv format	<100GB	NA
MILAN	Multiplex immunohistochemistry images	Reuse and new	Digital	Experimental	.tiff	<10TB	NA
WSI	Biopsy whole slide scans	Reuse	Digital	Experimental	.tiff	<10TB	NA
DNA	genotyping data	Reuse and new	Digital	Observational data and experimental data	.csv file	<10TB	NA
Xenium	spatial transcriptomics data	New	Digital	Experimental data	MEX, HDF5, and Zarr format,	<8TB	NA

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:

MILAN and Bulk and scRNAseq: https://doi.org/10.1038/s41467-023-39859-7 WSI: /
TEMPLATE+: /

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 clinical data, samples, and studies included in this project are approved by the Ethical Review Committee of the University Hospitals Leuven (S53364; S64006; S64904).

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

For this project, clinical data will be used from patients transplanted at the UZ Leuven. In addition, samples from the Biobank Renal Transplantation will be used for genotyping, expression analysis and imaging. All data will be pseudonymized.

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

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).

All data and accompanying information will be stored exclusively on KU Leuven servers, Onedrive. All data will be accompanied with a README file or tab that outlines the exact data collection procedure, especially important for experimental data. All experimental work is prepared by extensive preparations, each step is logged. Standard operating procedures are written out in the lab and safely stored together with the experimental data in the same folders, to allow easy recovery of the metadata. All team members have access to these metadata.

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

Yes, all metadata for experimental work are well maintained in the lab repository of procedures. They follow a standard format and vocabulary.

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

Where will the data be stored?

All data are stored at the KU Leuven Onedrive, the KU Leuven Large Volume Storage L-drive, or the UZ Leuven servers (for clinical data). No data will be stored on local computers, hard drives etc.

How will the data be backed up?

All data stored in the central KU Leuven facilities are backed up automatically with version control and logging.

**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

We have 20 TB of data capacity and the KU Leuven Large Volume Storage, which will clearly be sufficient for the proposed project. If not, we will apply for more storage volume.

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

The data are stored on the KU Leuven servers, only accessible with double authentication.

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

All costs are covered by the departmental group, or otherwise through existing grant funding.

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 data will be retained for at least 25 years. As data are stored digitally in secured servers of KU/UZ Leuven.

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

The same repositories as mentioned above will be used for long-term storage.

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

The same provisions are in place for long-term data preservation as for the data storage on shorter term. Smaller datasets are stored on the Onedrive, larger datasets on the KU Leuven L-drives.

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

All data will remain available for other researchers at KU Leuven without restrictions. The scRNAseq and bulk transcriptomic data are already made available publicly. Other datasets can only be made available with third parties with a DTA.

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

As indicated above.

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

All data originate from patients. Privacy regulations and ethical aspects restrict the sharing of these sensitive data.

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

Gene expression data are made available in the Gene Expression Omnibus.
The single-cell RNA-sequencing data have been deposited in BioStudies accession code E-MTAB-12051.

When will the data be made available?

Some data are already available (see above). Other data newly generated in this project will be made available upon publication of research results.

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

Data usage will either be open for public without any license in place, or be restricted and need a dedicated DTA before data sharing.

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, if applicable for the specific dataset (e.g. scRNAseq).

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

Usually, there is no charge for sharing data with third parties. For DTA of privacy-sensitive data, a quid pro quo in the form of co-authorship is usually requested.

6. Responsibilities

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

The team at the Nephrology and Renal Transplantation Research Group, under responsibility of the PI (Maarten Naesens)

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

The research team at the Nephrology and Renal Transplantation Research Group at KU Leuven

Who will manage data preservation and sharing?

The research team at the Nephrology and Renal Transplantation Research Group at KU Leuven

Who will update and implement this DMP?

Maarten Naesens, PI