Harnessing single-cell and spatial transcriptomics to uncover novel insights into the pathological events that trigger the progression of acute-on-chronic liver failure

A Data Management Plan created using DMPonline.be

Creators: Hannelie Korf, 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: Hannelie Korf

Project Administrator: Hannelie Korf

Grant number / URL: G025324N

ID: 208743

Start date: 01-01-2024

End date: 31-12-2027

Project abstract:

ACLF is a clinically important syndrome that develops in patients with decompensated cirrhosis. This distinct disease entity is characterized by a rapid deterioration of liver function, the development of organ failure and an extremely high short-term mortality. Currently, no therapy exists and ACLF is managed by organ support and treatment of the precipitating event. Important to note is that patients with ACLF exhibit an immunological paradox whereby they feature a hyperinflammatory state on the one hand and a co-existing phase of immunoparesis and infection susceptibility, on the other hand. Understanding the molecular mechanisms that tips the balance towards dampening inappropriate systemic inflammation and reinstating effective antibacterial function, would be crucial to extend patient survival. Here we aim to compose a comprehensive transcriptional map of hepatic (immune) cells and their microanatomical location that underlie ACLF immune dysfunction. We further aim to validate- the identified targets in independent patient cohorts or in murine models mimicking the ACLF syndrome. A success in this project will have groundbreaking impact on shedding light on the pathogenesis of ACLF.

Last modified: 27-06-2024

Harnessing single-cell and spatial transcriptomics to uncover novel insights into the pathological events that trigger the progression of acute-on-chronic liver failure

DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- Not applicable

Harnessing single-cell and spatial transcriptomics to uncover novel insights into the pathological events that trigger the progression of acute-on-chronic liver failure

GDPR

GDPR

Have you registered personal data processing activities for this project?

- Yes

Harnessing single-cell and spatial transcriptomics to uncover novel insights into the pathological events that trigger the progression of acute-on-chronic liver failure

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) Sequencing data
2) Imaging data
3) Spatial transcriptomics
4) Samples
5) Scripts
6) Patient clinical data

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)

Question not answered.

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)

Question not answered.

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)

Question not answered.

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

Question not answered.

Harnessing single-cell and spatial transcriptomics to uncover novel insights into the pathological events that trigger the progression of acute-on-chronic liver failure

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
Sequencing data	Single nuclei sequencing data; raw sequencing reads, processed data	☒Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☒ Other: sequences	.txt, .xlsx, .csv, .tsv, .fa, .bam, .rds, .h5ad, .loom, .R, .ipynb, .html, .pdf, .tif, .png, .py, .pbs	☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☒ > 5 TB ☐ NA	
Imaging data	Confocal images; Lightsheet images; Time-lapse movies; Obtained from single molecule FISH, IHC, as well as labelled living tissue imaging	☒Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	tif, .jpg, .png, .czi, sis, .mp4, .avi, .ai, .pdf, .pptx	☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☒ > 5 TB ☐ NA	
Spatial transcriptomics	Sequencing reads, processed data	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.tif, .png, .jpg, .loom, .csv, .hdf5, .h5ad	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☒ > 5 TB ☐ NA	

Samples	Tissue samples, Human or mice tissues, fixed samples, frozen samples	☒ Generate new data ☐ Reuse existing data	☐ Digital ☒ Physical	na	na	☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☒ NA	Frozen samples: Tubes stored at -80°C. Tissue for histology: fixed and stored at 4 °C
Scripts	Code written for analysis pipelines	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☒ Software ☐ Other:	R script	☒ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Patient clinical data	Biochemical and clinical characteristics describing the disease state of the patients	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	.xlsx, SLIMS software	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ 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:

We generate new data but some connections will be made by reusing data generated previously in DOI: 10.1097/HEP.0000000000000907

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

Most data sets will be generated on human samples (EC approval nr for human studies: S54588)
In the last year of the project a mouse model will be set up (ECD still to be applied for).

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

Patient clinical characteristics will be integrated within the data as approved in S54588

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.

- Yes

There is a potential that the data (transcriptomics & spatial transcriptomics) may generate novel concepts that could be further valorized. However, this is realistically not within the scope of the next 4 years.

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

We do have a MTA agreement with VIB (Prof. Diether Lambrechts) regarding the transcriptomic data, however this is not restrictive to exploitation or dissemination of the data since we have a full collaboration and shared controllership between contributing parties.

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).

We will maintain a record of the following for every WP (where applicable):
- Experimental design and protocol (.docx file)
- Abbreviations used (.docx file)
- Structure of the data (.docx file)
- Steps involved in data analysis and relevant analysis scripts (R, MATLAB, Python and ImageJ scripts)
- Raw data (specific file format according to data type)
- Analysed data (specific file format according to data type)
- Index file/read me file (.txt file) for every WP, linking the name, location (folder and subfolder on /server) and description of above-mentioned files.

Physical data:
Samples taken from experiments will be documented and stored. Storage will be in fixative, in paraffin, at 4°C or in freezers depending on the kind of sample.

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 standards will be used for genomics (http://www.dcc.ac.uk/resources/metadata-standards/genome-metadata). For all other data, metadata will be created using the Dublin core (http://www.dcc.ac.uk/resources/metadatastandards/dublin-core).

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

Where will the data be stored?

☒ Shared network drive (J-drive)
☒ Personal network drive (I-drive)
☒ OneDrive (KU Leuven)
☐ Sharepoint online
☐ Sharepoint on-premis
☒ Large Volume Storage
☐ Digital Vault
☒ Other: K-drive (large raw data files) and SLIMS software (Patient data and sample info)

How will the data be backed up?

☒ Standard back-up provided by KU Leuven ICTS for my storage solution
☒ Personal back-ups I make (specify): KU Leuven OneDrive
☐ Other (specify)

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?

Researchers involved in the project can control who they give access to the files on their personal OneDrive. To access the KU Leuven servers, access is provided and controlled by the research manager, Hannelie Korf. The KU Leuven ICTS data center hosts the network storage, with a mirror available in the second ICTS center. This ensures additional back-up capacity, recovery of lost data and long term data availability. The access is controlled by KU Leuven security groups and it is password protected.

Data related to Patients and controls are pseudonymized, which is managed by the research manager, Hannelie Korf. The list linking the anonymized code with patient records (name, EMDNR number, clinical data) is stored on the hepatology common J-drive, and is locked under password protection. Password access if provided for read-only access to the students and technicians who are working directly with the samples. Hannelie Korf retains the sole ability to modify this file under alternative password protection. All samples are processed and stored as per their anonymized code only.

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

Personal and Shared network drive (I, J) 503,66 euro/TB per year. All large primary data will be moved to LVS. LVS storage costs per 5 Tb (KU Leuven ICTS): 104,42euro/year. Expected amount of data (50TB). The costs have been budgeted on the grant.

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 preserved for 10 years according to KU Leuven RDM policy

☐ All data will be preserved for 25 years according to CTC recommendations for clinical trials with medicinal products for human use and for clinical experiments on humans
☐ Certain data cannot be kept for 10 years (explain)

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

☐ KU Leuven RDR
☒ Large Volume Storage (longterm for large volumes)
☐ Shared network drive (J-drive)
☒ Other (specifiy):

Digital data will be stored at the Archive (K:) server from KU Leuven ICTS. HCR probes and physical samples will be stored in the freezers from the laboratory of Hepatology. Code scripts will be stored on Github.

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

We expect the costs to gradually increase up to 3000 euro/year. After the project, data preservation costs will be covered by other grants.

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 data will be made available after publication via the required link in the publications or upon request after an embargo period after publication (f.i. phenotype files, genetic data). The same holds true for unpublished data, they can be made available upon request but only after an embargo period (3 years; exceptionally 5 years after the project).

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

All team members have access as long as they are affiliated to KU Leuven. Once all files are released, anyone can use these data to generate new results, referring to the original publication and not for commercial use. Other data will be only released upon request and after an embargo period after publication.

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.

- No

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

☐ KU Leuven RDR
☐ Other data repository (specify)

☒ Other (specify)
Experimental data will be made available through a data repository such as EGA, ncbi, github, Genbank, FigShare (https://figshare.com/), Dryad (https://datadryad.org/) or https://zenodo.org/depending on the type of data. We will explore the possibilities via online repositories and will use the website www.re3data.org

When will the data be made available?

☒ Upon publication of research results
☐ Specific date (specify)
☐ Other (specify)

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

☒ CC-BY 4.0 (data)
☐ Data Transfer Agreement (restricted data)
☐ MIT licence (code)
☐ GNU GPL-3.0 (code)
☐ Other (specify)

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
☐ My dataset already has a PID
☐ No

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

The transfer costs depend on the data repository selected. Costs will be covered by the project funding.

6. Responsibilities

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

The research manager, Hannelie Korf

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

The research manager, Hannelie Korf

Who will manage data preservation and sharing?

The research manager, Hannelie Korf

Who will update and implement this DMP?

The research manager, Hannelie Korf