Plan Overview

A Data Management Plan created using DMPonline.be

Title: Unraveling the EBV negative post transplant diffuse large B-cell lymphoma genomic landscape: pioneering early detection in cell-free DNA

Creator: Ana Lucia Rocha Iraizos

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Project abstract:

Post-transplant diffuse large B-cell lymphoma (PT-DLBCL) is a hematologic malignancy affecting up to 20% of solid organ transplant patients. PT-DLBCL is characterized by uncontrolled B cell proliferation, often caused by immunosuppressive treatment. Patients receive rituximab, which is effective in 34% of patients. The remaining 66% proceeds to chemotherapy, a particularly burdensome treatment for these fragile patients. Therefore, there is an urgent need for targeted therapies to treat PT-DLBCL. An additional problem is that we cannot predict which transplant patients are at risk of developing PT-DLBCL. As a result, PT-DLBCL is usually diagnosed at full blown disease, complicating the treatment and increasing mortality. Early detection could enable timely interventions, potentially preventing (advanced) PT-DLBCL development. I will decipher the genomic basis of PT-DLBCL pathogenesis by identifying mutations triggering PTDLBCL development. I will then introduce these mutations in cell lines to investigate how they drive oncogenesis. In the long term, this information could contribute to the development of targeted therapies for PT-DLBCL. I will also assess which of these mutations can be detected in the plasma of transplant patients before PT-DLBCL development. Mutations that enable non-invasive prediction of PT-DLBCL are ideal candidates for future biomarkers. Early detection could lead to timely interventions and improved life expectancy of these vulnerable patients

ID: 211152

Start date: 01-11-2024

End date: 31-10-2028

Last modified: 17-03-2025

Unraveling the EBV negative post transplant diffuse large B-cell lymphoma genomic landscape: pioneering early detection in cell-free DNA

DPIA

DPIA

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

Question not answered.

Unraveling the EBV negative post transplant diffuse large B-cell lymphoma genomic landscape: pioneering early detection in cell-free DNA

GDPR

GDPR

Have you registered personal data processing activities for this project?

Question not answered.

Unraveling the EBV negative post transplant diffuse large B-cell lymphoma genomic landscape: pioneering early detection in cell-free DNA

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)

Question not answered.

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.

Unraveling the EBV negative post transplant diffuse large B-cell lymphoma genomic landscape: pioneering early detection in cell-free DNA

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	
Demographic and clinical data of EBV negative post transplant DLBCL patients	Includes gender, date of organ transplantation, date and place of birth and response to certain therapies.	New data	Digital	Other (genetic data)	In REDCap eCRF exported as .csv	<100MB	NA
Genomic and (singel cell) transcriptomics data	Data files obtained from NGS and (single nuclei) RNA sequencing	New data	Digital	Other (transcriptomic data)	In High Performance computing Cluster (HPC) VSC: .bam, .sam, .fastq, .csv, .r, .count, .txt	<50 TB	NA
Cell-free DNA data	Data files obtained from NGS on cell-free DNA	New data	Digital	Other (genetic data)	In High Performance computing Cluster (HPC) VSC: .bam, .sam, .fastq, .csv, .r, .count, .txt	<50 TB	NA

genomic DNA (gDNA)	Genomic DNA extracted from tumor biopsies, blood (buffy coat mononuclear cells) and buccal swabs	New data	Physical	Physical	NA	NA	1 gDNA sample per sample-type (tumor, blood or buccal swab). Up to 6 DNA samples per patient (diagnosis and relapse).
Cell-free DNA (cfDNA)	extracted from blood plasma	New data	Physical	NA	NA	NA	Up to 10 cfDNA samples per patient (at different timepoints after transplantation)
RNA	mRNA extracted from tumor biopsies	New data	Physical	NA	NA	NA	Two mRNA samples per patient (diagnosis and relapse).
Cells	Cells extracted from tumor, blood (plasma, and buccal swabs.	New data	Physical	NA	NA	NA	1 sample per type, up to 6 samples per patient (3 sample types x 2 timepoints per patient (diagnosis and relapse))

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:

No existing data is used

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

In this project, genomic, transcriptomic and cfDNA data obtained from human blood, bucal mucosa, and tumor samples will be generated. In order to collect these human body materials and the corresponding clinical and demographic data of the patients, two ethical approvals are needed. The first ethical approval (S-61200) concerns collecting patient data from blood, bucal mucosa, and tumor samples, as well as the corresponding clinical and demographics data from lymphoma patients. The second study is submitted to the ethical committee (S-69378), and includes the collection of plasma samples as well as the corresponding clinical and demographics data from patients who developed lymphoma after a kidney transplantation.

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

Clinical and demographics data of lymphoma patients. This data is held by UZ Leuven hospital. This data includes gender, date of organ transplantation, date and place of birth and response to certain therapies amongst others. All personal data is stored on KWS and LWS, to which only Dr. MD. Daan Dierickx (supervisor) has access to. The researcher will only have access to pseudonimised data. This pseudonimised data will be stored in REDCap eCRF.

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

The genomics and transcriptomics data will be analyzed in collaboration with VIB (lab of prof. Jan Cools), for this, an MTA (of S-69378) has been submitted to the ethical committee.

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 MTA that applies to this project (S-69378) and that has been submitted to the ethical commitee, the following paragraphs are written:
The Material and Data are and remain the property of Provider (Prof. Dr. Daan Dierickx, supervisor of the project). Recipient (VIB) agrees not to apply for any patent or any other proprietary title which would claim the Material or its use. Provider furthermore retains ownership of any form or part of the Material included in the Modifications or of the Data included in the Research Data. JOINT OWNERSHIP. The Parties agree that all rights, title and interests in the Results will vest jointly in Provider and Recipient. In any event, both Parties agree to perform reasonable efforts to notify the other Party in a timely manner prior to the filing of a patent application, including by providing the application text itself, that covers Data and/or Material and/or Results in all cases related to the Research. Regarding patentable Results, Parties agree to negotiate a joint ownership agreement which shall specify rights and obligations of each Party related to the use, exploitation and protection of the patentable joint Results at least within twelve (12) months of the filing of the first patent application. Unless otherwise agreed in the joint ownership agreement, patent applications with regards to Inventions will be filed only with mutual consent of and in the joint name of Provider and Recipient. Regarding non-patentable Results, Parties agree to negotiate a joint ownership agreement which shall specify the rights and obligations of each Party related to the use, exploitation and protection of the non-patentable joint Results prior to any use for Commercial Purposes. Prior to such joint ownership agreement, neither party will be entitled to use for Commercial Purposes the jointly owned Result without the prior written consent of the other Party. Neither Party will grant any rights or transfer its rights in the joint Result to any third party without the prior written consent of the other Party. Each Party will in any case retain a royalty free, non-exclusive
license to use any Results for its internal non-commercial research and teaching
purposes, subject to applicable laws. The Parties will agree in good faith on the sharing of revenues generated from the exploitation of the joint Result based on the contribution of each Party in the generation of the Result.

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).

- Physical data: all the information and the methodology and protocols used to collect the data, as well as the exact location of physical data will be described in the electronic notebook (eLabJournal), to which all the lab memebers and lab manager have access to.
- Genomic and (singel cell) transcriptomics data and cell free DNA data: all the codes used to generate the data will be stored on a shared VSC staging of the lab, to which the lab staff scientist will have access to. The codes contain comments explaining each of the different steps, and a README.txt file will be added to elaborate on certain decisions taken. This file

will be added to the same directory of the corresponding code.

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

All the sequencing data will be stored on the secure systems provided by KU Leuven, i.e. in the storage of a high performance computer (HPC) and on a secure and redundant server system (ManGO). REDCap will be used to stored patient demografic data.
The deposit and sharing of research data will be done via KU Leuven RDR DataCite.

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

Where will the data be stored?

All the sequencing data will be stored on the secure systems provided by KU Leuven, i.e. in the storage of a HPC and on a secure and redundant server system (ManGO).
All the materials derived from patient samples will be stored in the UZ Leuven Biobank (cryotheek)

How will the data be backed up?

"L-drive" and "J-drive": Standard back-up provided by KU Leuven ICTS
Genomic data will be stored and on a secure and redundant server system (ManGO).

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

The lab has sufficient resources and funding to acquire the necessary storage capacity on the HPC-VSC.
The estimated volume of data is 5 TB for the single-nuclei RNA-seq experiments. For all the genomic data, including the cf-DNA data will take up to 15 TB.
Furthermore, there is sufficient storage and back-up capacity on all KU Leuven servers:
- the “L-drive” is an easily scalable system, built from General Parallel File System (GPFS) cluster with NetApp eseries storage systems, and a CTDB samba cluster in the front-end.
- the “J-drive” is based on a cluster of NetApp FAS8040 controlers with an Ontap 9.1P9 operating system
- Sharepoint will be used for EC- and biobank- related documents, such as study protocols, biobank application forms and GDPR-questionnaires.

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

The VSC-HPC staging of the lab is restricted to only lab members, and the folders containing genomic data are only available for the lab bioinformatician, the researcher and the supervisor.
All the biological samples are stored in the UZ Leuven Biobank (cryotheek), which can only be accessed with badge access.

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

The expected cost for storing data on the VSC is 20 euros per TB per year, we expect to generate unitl 20 TB of data. All these costs will be covered by the 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 the data will be preservesa for 10 years according to the KU Leuven RDM policy

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

All the sequencing data will be stored on the secure systems provided by KU Leuven, i.e. in the storage of a high performance computer and on a secure and redundant server system (ManGO). Furthermore, all data that is used for publications will be submitted to an appropriate public system, i.e. EGA since we are working with sensitive patient data.

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

The expected cost for storing data on the VSC is 20 euros per TB per year, we expect to generate unitl 20 TB of data. All these costs will be covered by the 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 an Open Access repository

Restricted genomic and transcriptomic in combination with related metadata (e.g. age, gender, sequencing platform/library... etc.) data will be submitted to the European-Phenome Archive (EGA), after publication.

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

As patient data is very sensitive data, this access will be restricted. Restrictions will be specified in the DAC application form.

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, Ethical aspects

All data derived from patients is sensitive data, therefore, the sharing of this data is restricted.

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

Genomic data will be made available through EGA. RDR will be used to share protocols, and Github will be used to share codes used for the analysis of genomic and transcirptomic data. The different datasets will be registered on Lirias

When will the data be made available?

The data will be made available upon publication of the research results

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

Our genomic data will be submitted to EGA, data submitted to EGA is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) 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

A DOI and PID will be added upon deposit in a data repository

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

Not applicale

6. Responsibilities

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

The documentation will be managed by the researcher (Ana-Lucía Rocha) as well as the supervisor of the project (Dr. Daan Dierickx), the post-doc of the project (Dr. Flore Sneyers) and the bioinformatician and staff scientist of the laboratory (Dr. Sofie Demeyer)

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

The data storage and backup will be managed by the researcher (Ana-Lucía Rocha), the post-doc of the project (Dr. Flore Sneyers) as well as the supervisor of the project (Dr. Daan Dierickx) and the bioinformatician and staff scientist of the laboratory (Dr. Sofie Demeyer)

Who will manage data preservation and sharing?

The data preservation and sharing will be managed by the researcher (Ana-Lucía Rocha) as well as the supervisor of the project (Dr. Daan Dierickx) and the bioinformatician and staff scientist of the laboratory (Dr. Sofie Demeyer)

Who will update and implement this DMP?

This DMP will be updatad and implemented by the researcher (Ana-Lucía Rocha), as well as the supervisor of the project (Dr. Daan Dierickx) and the bioinformatician and staff scientist of the laboratory (Dr. Sofie Demeyer)