Ultra-sensitive minimal residual disease assessment by blood-based clonotypic mass spectrometry approach in plasma cell dyscrasias (MASSIAS)

A Data Management Plan created using DMPonline.be

Creators: Louis Nevejan, 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: Louis Nevejan

Grant number / URL: T005323N

ID: 204434

Start date: 01-10-2023

End date: 30-09-2027

Project abstract:

With nearly 1000 new diagnoses in Belgium in 2020, multiple myeloma accounted for ~1.5% of all new registered cancer diagnoses in that year. Current response criteria in treated multiple myloma patients recommend determining minimal residual disease (MRD) detection with either next-generation flowcytometry (NGF) or next-generation sequencing (NGS) on bone marrow aspirates as the cornerstone of disease management in patients in complete remission. In this doctoral project we aim to develop a mass spectrometry (MS)-based assayo detect MRD in the peripheral blood of myeloma patients and patients with other plasma cell dyscrasias.

This approach, called ‘clonotypic peptide MS’ will detect patient unique clonotypic peptides in the monoclonal immunoglobulins at diagnosis and quantify them as personalized biomarkers in follow-up samples. Clinically relevant proof-of-concept is available by other research groups and confirmed by a preliminary experiment of our research group showing the feasibility to detect MRD with this approach. We aim to further develop the clonotypic peptide MS assay and translate its utility in clinical applications by initiating a single-center retrospective trial and by joining a multi-center prospective trial. By the establishment of a consortium of applicants, we will be able to compare our clonotypic peptide MS assay with i) conventional blood-based diagnostic tools, with ii) an alternate blood-based MS approach, with iii) bone-marrow based NGF and iv) with imaging techniques to assess MRD.

With the results of these trials, we aim to prove the feasibility of the clonotypic peptide MS approach which will have a major positive medical impact for the majority of patients with plasma cell dyscrasias (as repeated painful bone marrow examinations in follow-up could be omitted) and a significant cost reduction for the Flemish healthcare system (as the blood-based clonotypic peptide MS approach will be more cost-effective compared to bone marrow-based NGF and NGS for MRD assessment).

Last modified: 02-02-2024

Ultra-sensitive minimal residual disease assessment by blood-based clonotypic mass spectrometry approach in plasma cell dyscrasias (MASSIAS)

DPIA

DPIA

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

- Not applicable

Ultra-sensitive minimal residual disease assessment by blood-based clonotypic mass spectrometry approach in plasma cell dyscrasias (MASSIAS)

GDPR

GDPR

Have you registered personal data processing activities for this project?

- Yes

Ultra-sensitive minimal residual disease assessment by blood-based clonotypic mass spectrometry approach in plasma cell dyscrasias (MASSIAS)

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)

Following datatypes will be collected during this TBM project:
- mass spectrometry run raw data (.RAW format - Xevo G3 QTOF)
- Progenesis data (Progenesis LC MS file, MZNLD format – Progenesis)
- PEAKS data (Peaks project format - PEAKS DB, de novo, PTM, spider search)
- MASCOT data (Mgf, xml format - Mascot search)
- Statistical processed data (.Excel format – Excel) and R
- Figures and graphs processed data (.JPG format - Excel & GraphPad Prism)
- Protocols (.doc(x) format - Word)
- Patient data: REDCAP data file
- NGS data (.txt format - collected in REDCAP data file)
- Serum samples: biobank of UZ Leuven

All mass spectrometry data is newly collected data. Regarding the patient data: this concerns partially newly collected data (work package 2), partially reused data (work package 3).

The estimated total volume of data is ~5 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 (If already designated, please fill in his/her name.)

Promotor: prof. dr. Xavier Bossuyt (contact: xavier.bossuyt@uzleuven.be)

2. Storage capacity/repository

All data will be stored during the research & at least 5 years after the research on a UZ Leuven or KU Leuven server in protected folders which are only accessible for the involved staff of the research group. Sufficient storage capacity is guaranteed. All raw mass spectrometry data that is used in a publication will be shared online via ProteomeXchange.

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)

Personal data of patients will collected in this research. To ensure that these data are protected, we will use the REDCap platform (GBW [Medical Science Faculty] REDCap environment). The pseudonimization code (link between personal data and specific study number) will be stored in a separated section on the platform and will solely be available to the promotor of the project and to the doctoral student of the project.

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

Not applicable

Ultra-sensitive minimal residual disease assessment by blood-based clonotypic mass spectrometry approach in plasma cell dyscrasias (MASSIAS)

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
Mass spectrometry run data	Raw data from Xevo G3 QTOF analyzer	Generate new data	Digital	Experimental	.raw	<5 TB	N.A.
Progenesis data	Software tool for relative quantification of data	Generate new data	Digital	Software	.mznld	<1 TB	N.A.
PEAKS data	Software tool for de novo sequencing	Generate new data	Digital	Software	PEAKS project format	<1 TB	N.A.
MASCOT data	Software tool for identification of peptides against Uniprot Homo sapiens database	Generate new data	Digital	Software	.mgf	<100 GB	N.A.
Statistical procession data	processing of mass spectrometry data and clinical data	Generate new data	Digital	Software	.excel .R	<100 GB	N.A.
Figures and graphs processed data	processing of mass spectrometry data and clinical data into scientific illustrations	Generate new data	Digital	Software	.jpeg	<1 GB	N.A.
Study protocol	Study protocol	Generate new data	Digital	N.A.	.word	<100 MB	N.A.
Patient data work package 2	Clinical data retrieved from medical files (KWS)	Generate new data	Digital	Observational	REDCAP	<1 GB	N.A.
Patient data work package 3	Clinical data retrieved from patients participating in the IMMPROVED trial (collected in UAntwerpen)	Reused data	Digital	Oberservational	REDCAP	< 1GB	N.A.
NGS data work package 1	For 40 samples, UZ Brussel will perform NGS on bone marrow aspirates; the data of these experiments will be transfered to KU/UZ Leuven	Generate new data	Digital	Experimental	.txt (collection in REDCAP)	< 100 MB	N.B.
Serum samples UZ Leuven patients	All samples that are used for mass spectrometry analyses (work package 2 + 3)	Generate new data	Physical	Experimental	N.A.	N.A.	750 blood samples
Blood samples UZ Brussel patients	All samples for work package 1	Generate new data	Physical	Experimental	N.A.	N.A.	40 blood samples
Blood samples UZ Antwerpen patients	Samples for work package 3	Generate new data	Physical	Experimental	N.A.	N.A.	250 blood samples
Blood samples UZ Gent patients	Samples for work package 3	Generate new data	Physical	Experimental	N.A.	N.A.	250 blood samples
Blood samples UMC Amsterdam patients	Samples for work package 3	Generate new data	Physical	Experimental	N.A.	N.A.	250 blood samples

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:

Not applicable

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

Residual blood samples of patients will be used in this research project. In additional, pseudonymized clinical data of these patients will be used to interpret the clinical relevance of the acquired data. Approval of the ethics comité of UZ Leuven has been given on January 31, 2024 for this research project (reference S68376).

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

We will only use pseudonymised data in our project. In the REDCAP data collection site, a separate file will include the patient identification number (EAD number) and a study specific number (e.g. MASSIAS 1). This separate file will only be available to the protomor and to the doctoral student of the project. All other files in the REDCAP data collection site will only mention the study specific number of the included patients. In here, we will collect all clinical and laboratory data of the patient that is related to his/her hematological neoplasm.

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

A collaboration agreement is signed between the consortium partners of this project to share all relevant data and patient samples (partners: KU/UZ Leuven, VUB/UZ Brussel, U/UZ Antwerpen, AZ Sint-Jan Brugge).
Specific for work package 3, samples will be analyzed from patients in UZ Leuven, UZ Antwerpen, UZ Gent and UMC Amsterdam. For the latter two hospitals, an additional material transfer agreement will be drafted.

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 mass spectrometry data and accompanying information will be stored exclusively on UZ Leuven servers and/or on KU Leuven servers. Each folder will be accompanied with a README.txt file that outline the data collection procedure. The experimental work will be written out in standard operating procedures 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. Scripts for data analysis are all written in R.

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 raw mass spectrometry data that is used in a publication will be shared online via ProteomeXchange.

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

Where will the data be stored?

All data is stored electronically either at UZ Leuven or at the personal KU Leuven One Drive for daily back-up. They will be moved to the Large Volume Storage of the KU Leuven ICTS on a monthly basis.

How will the data be backed up?

Daily back-up of newly generated data is guaranteed by UZ Leuven and within the KU Leuven One Drive service. the Large Volume Storage of the KU Leuven ICTS secures daily backup of the large files/datasets directly uploaded there.

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?

Data stored at UZ Leuven is actively logged; data stored in the KU Leuven One Drive is not accessible by any other person except the reseracher. Data files in the Large Volume Storage of the KU Leuven ICTS are only accessible by lab members via a two-step authentication system.

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

The KU Leuven OneDrive comes without charge, and will be enough for storing data for the entire duration of the project. The cost of the Large Volume Storage of the KU Leuven ICTS is covered by the FWO/TBM 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 above mentioned data (mass spectrometry data, clinical data collected in REDCAP file) will be retained for 10 years (according to KU Leuven RDM policy). If too much space is needed to store all the data, the raw data generated by the Xevo G3 Q-TOF mass spectrometer will not be maintained for 10 years. All mass spectrometry data collected at AZ Sint-Jan Brugge (EXENT mass spectrometer) will be kept there.

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

Large Volume Storage of the KU Leuven ICTS

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

For the first 6 years the costs will be covered by the FWO-TBM grant. As the grant money will only be available for 6 years, the budget to cover the cost for storage after 6 years is not clear at the moment.

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 papers from this project will be publised as 'open access' and/or will be published via Lirias. All raw mass spectrometry data that is used in these publications will be shared online via ProteomeXchange.

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

Not applicable

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.

All raw mass spectrometry data that is used in these publications will be shared online via ProteomeXchange.

When will the data be made available?

Upon acceptance of the publication or at the end of the project.

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

there is no license associated to the use of raw mass spectrometry 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

Not yet available, only after publication

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

There are no expected costs for dara sharing.

6. Responsibilities

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

Doctoral student (Louis Nevejan)

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

Doctoral student (Louis Nevejan)

Who will manage data preservation and sharing?

Doctoral student (Louis Nevejan)

Who will update and implement this DMP?

Doctoral student (Louis Nevejan)