Elucidating the role of Epstein Barr virus in multiple sclerosis pathophysiology

A Data Management Plan created using DMPonline.be

Creator: 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: Benedicte Dubois https://orcid.org/0000-0003-0604-8702, n.n. n.n.

Grant number / URL: 11P8K24N

ID: 206193

Start date: 01-11-2023

End date: 31-10-2027

Project abstract:

Multiple sclerosis (MS) is a chronic autoimmune disease of the central nervous system affecting 2.8 million people worldwide. It is the most common neurological disorder affecting young adults. Despite many genome wide association and epidemiological studies the etiology of MS is still unknown. Recent insights have put Epstein-Barr virus (EBV) and its host (the B cell) at the forefront of MS pathophysiology. However, the mechanisms linking the virus to CNS inflammation remain elusive. Understanding this is essential to determine the next steps in management and novel therapeutic developments in MS. Until now the state-of-the art has lacked a comprehensive view on the EBV-B cell-T cell axis and especially its link with clinical disease progression. In this project, I will investigate the molecular mimicry hypothesis in the location of interest, the cerebrospinal fluid, and correlate this reactivity with disease progression. I will combine state of the art single cell RNA-seq technologies with our in-house targeted sequencing approach to achieve maximal sensitivity for low abundant EBV transcripts, without losing information on general gene expression profile. Lastly I will examine chromatin remodeling on a single cell level in CSF. At the end of the project, I am to obtain a fully characterized pre-clinical package on EBV- B cell - T cell interaction that can form the basis for future therapeutic strategies.

Last modified: 26-04-2024

Elucidating the role of Epstein Barr virus in multiple sclerosis pathophysiology
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)

- Primary and secondary quantitative and qualitative: demographic and clinical data
- Primary experimental: experimental from human biological material
- Primary derived: data processing, statistical and bioinformatics analyses
- Secondary reference: standard genetic references databases that are publicly available
- Primary quantitative and qualitative: summaries, presentations, publications,...
- Biological material: human biological material from MS patients and controls (blood, DNA, RNA, PBMCs, plasma, serum, cerebrospinal fluid)

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)

a) Responsible persons:
- Demographic and clinical data, patient identifiers, ethics approval and implementation of UZ Leuven guidelines: Prof. Bénédicte Dubois
- Pseudonymized, experimental and derived data and implementation of KU Leuven guidelines: Prof. An Goris and prof. Bénédicte Dubois

b) Storage:
- During research: on KU Leuven file shares (I, J, L), removable storage media for large data exceeding the capacity of file shares, KU Leuven approved laptops and on the supercomputer server for calculations, biobank.
- After research: KU Leuven file shares (Archive K-drive and Large-Data L drive), data repositories for genetic data if allowed, long-term storage biobank

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)

NA

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)

KU Leuven Ethics and GDPR guidelines will be followed

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

NA

Elucidating the role of Epstein Barr virus in multiple sclerosis pathophysiology
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	
Patient data	Demographic and clinical data collected by Prof. Bénédicte Dubois (neurologist). The researcher only receives the sample number and characteristics important for the analysis (pseudonymization). Approval from EC was already obtained.	New data and existing data.	Digital	Observational	Microsoft Access Database .csv	<100 MB	/
MRI patient data	Radiological data (MRI) collected by collaborating radiologist (Dr. Sarah Cappelle). Patient ID is pseudonymized similar to demographic and clinical data. Approval from EC was already obtained.	New data	Digital	Observational	.csv	<100 MB	/

Biological material - LNI	Preservation of plasma, CSF supernatant, DNA and RNA at -80°C. Preservation of PBMCs and CSF cells in liquid nitrogen. Collected by researcher and other members of team.	New and reused	Physical	/	/	/	1 mL per aliquot
Biological material - LAG	Preserved serum and CSF supernatant at -20°C. Collected by the laboratory for clinical biology UZ Leuven.	Reused	Physical	/	/	/	0.5-3 mL per sample
Biological material - EBV+ cell lines	Commercially available EBV+ cell lines, used for technical validation of the assays. Obtained by collaborator (Rega Institute Virology), cultured and preserved in liquid nitrogen by researcher.	Reused	Physical	/	/	/	1mL per aliquot
ELISA data	Readout data of conducted ELISA experiments, on CSF supernatans, serum and plasma. • Raw ELISA readout files • Standard curve plotting output files • Scripts for analysis (R) • Output figures (Graphpad Prism, R)	New	Digital	Experimental	• .csv • .xlsm • .R, .rds • .prism, .png, .pdf	<1GB	/
Sequencing data	Sequencing data from scRNA-sequencing and scATAC sequencing experiments. This includes: - Raw sequencing data files - Scripts for analysis (supercomputer + R) - Output figures	New	Digital	Experimental	- .fastq, .bam, .mtx, .tsv - .bash, .R, .rds - .xlsx, .png, .pdf	<5TB	/

Documentation	- Experimental protocols (stored digitally + on paper in laboratory notebooks) - Manuscripts - Standard genetic reference databases (publicly available)	New and reused	Digital and Physical	Experimental	.docx, .pdf, .xlsx + stored on paper in laboratory notebooks	<100GB	1 A4 laboratory notebook

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:

- Patient data collected as part of current or previous research projects within our laboratory will be reused. This data was generated for the projects of Stijn Swinnen, Margaux David or Jarne Beliën.

- Biological samples collected by the clinical laboratory of UZ Leuven will be reused. These samples are collected as part of the UZ Leuven 'neuro-biobank' (study number S59940)

- Biological samples of commercially available EBV+ B cell lines will be reused. These samples are were collected for research by the group of prof. G. Andrei (Virology, Rega institute).

- Publicly available genetic reference databases such as Genome Browser, Gene Cards, Human Protein Atlas, Ensembl. Publicly available databases of chromatin accessibility, such as ATACdb. Publicly available ChiPseq data.

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 Ethics Committee Research UZ Leuven/KU Leuven approved the use of samples (blood and CSF) of patients and healthy controls for the purposes of research (S60222 and S50354).
The Ethics Committee Research UZ Leuven/KU Leuven approved the use of the remaining volume of samples (serum and CSF) which were obtained for clinical purposes in the department of Neurology UZ Leuven (S59940), for the purposes of research.

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 use personal data of patients and controls included in the study. The collected patient data consists of general characteristics such as age and sex, as well as disease-specific information (e.g. types of treatment received, dates during which patient was treated, MRI derived parameters...) and the result of the ELISA, single-cell RNA sequencing, single-cell ATAC sequencing and genotype information. The data of the controls included in the study consists of general characteristics, clinical data and results of the ELISA, single-cell RNA sequencing, single-cell ATAC sequencing and genotype information.

Clinical data is collected by Prof. Bénédicte Dubois. The researchers will receive the data pseudonymized to prevent patient name tracing.

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).

Documentation including experimental protocols and parameters (concentrations or titers measured, number of cells counted,...) will be recorded in physical laboratory books and stored into Word or Excel files, which also contain the necessary metadata (user, date,...) regarding the experiment.

Data folders containing the raw and processed data will be hierarchically organized and labeled based on the date of data generation and source of the data. Each main data folder will contain a ReadMe.txt file containing all the necessary information to keep the data understandable and usable.

All of the R scripts will be made in Rmarkdown format and knitted to HTML format so that comments can be added in between the chunks of code, thereby providing explanations to keep the analysis understandable. R markdown files automatically contain metadata (user, date,...)

All of these files will be stored in the KU Leuven share storage space (J-drive) and will be backed up regularly to the KU Leuven large storage space (L-drive) and an external hard drive.

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.

- No

See above.

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

Where will the data be stored?

Electronical data will be stored in conformity with KU Leuven and FWO RDM policy. All datatypes including protocols, raw data and analysed data will be stored on KU Leuven file shares (J, K, L). A copy of all data will also be stored on an external hard drive kept by the researchers in the laboratory. In case additional storage is required, the capacity of the KU Leuven shares can be increased.
Hard copy laboratory notebooks with raw data and records of the experiments will be physically stored in our laboratory.

How will the data be backed up?

The KU Leuven file shares (central ICTS data centre) provides hourly, daily and weekly automatic back ups. These back-ups are saved on their servers.
Moreover, the data will also be backed-up on external hard drives kept by the researchers in the laboratory.

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

Yes, enough storage and back-up capacity is available on the KU Leuven shares. In case additional storage is required, the capacity of the KU Leuven shares can be increased.

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

To ensure our data is protected, the KU Leuven shares are secured with a login and password connected to our KU Leuven accounts. Only members of our laboratory have access to these drives.

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

Data storage and costs will be managed by the principal investigator responsible for this project, Prof. Bénédicte Dubois. The costs for data storage is € 503.66/TB/year, which comes to approximately € 2000 for the duration of this FWO fellowship.
The costs will be covered by previous funding obtained by the laboratory and/or by the FWO fellowship bench fee.

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...).

Physical samples from patients and controls will be kept for at least 10years (for liquid nitrogen samples in the KU Leuven biobank, for plasma and CSF supernatans samples in -80°C in the research laboratory).
Clinical data from patients and controls will be kept for at least 30 years, according to recommendations for clinical trials.

According to KU Leuven RDM policy, all other research data will be kept for at least 10 years.

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

During and after the research project, data will be stored on KU Leuven shares. The same data will also be backed-up on an external hard drive.
The registry containing all clinical data of patients and controls will be stored on UZ Leuven drives, with adequate back-up provided by UZ Leuven.

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

Data that need to be easily accessed will be stored on the KU Leuven share J-drive, for which the costs are € 503.66/TB/year. On the longer term, data can be preserved on the K-drive, specifically dedicated to archive storage. This costs € 5.69/100Gb/year.
Long-term data storage and costs will be managed and evaluated by the principal investigator responsible for this project, Prof. Bénédicte Dubois.

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
- No (closed access)
- Yes, in a restricted access repository (after approval, institutional access only, …)

Raw experimental data of ELISA experiments will be shared in open access (as a supplementary table to the manuscript). Scripts for analysis will be shared in open access (on the Github page of the laboratory for Neuroimmunology). Other data (e.g. individual clinical data) will not be made available to third parties.
Counts matrix (or raw data) of single-cell RNA sequencing data and single-cell ATAC sequencing will be made available in an open access repository such as GEO (Gene Expression Omnibus). Scripts will be shared upon request.The rest of the data will not be made available to third parties.

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

/

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.

Raw experimental data of ELISA experiments will be shared in open access (as a supplementary table to the manuscript) if this is a requirement of the journal.
The raw single-cell sequencing data (or counts matrix) will be made available in an open access repository such as GEO (Gene Expression Omnibus) if this is a requirement from the journal.
The scripts for data analysis will be shared through a GitHub repository.

When will the data be made available?

The data that will be shared, namely the raw ELISA data, counts matrix from single-cell sequencing data and the scripts, will be made available upon publication.

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

As advised by the KU Leuven RDM helpdesk and the license selector tool, we will use the Creative Commons Attribution-ShareAlike (CC BY-SA-4.0) license. This license allows others to use, copy, distribute, and modify the data for any purpose, as long as the original source is properly attributed and any modifications are distributed under the same license terms.

For scripts shared on Github, we will use the MIT license. As advised by the license selector tool. This permissive license allows others to copy, distribute and modify the software as long as they include the original copyright and 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

Data in the GEO repository will be granted a GEO accession number.

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

Sharing single-cell RNA sequencing data ion GEO is usually free of charge for smaller datasets.
The basic GitHub accesses are also free for individuals and organization.

6. Responsibilities

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

The principal investigator (Prof. Bénédicte Dubois) and the researcher (Dries De Wit) bear the responsibility for the data documentation during the research project.

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

The principal investigator (Prof. Bénédicte Dubois) and the researcher (Dries De Wit) bear the responsibility for the data storage and backup during the research project.

Who will manage data preservation and sharing?

The principal investigator (Prof. Bénédicte Dubois) and the researcher (Dries De Wit) bear the responsibility for the data preservation and sharing during the research project.

Who will update and implement this DMP?

The principal investigator (Prof. Bénédicte Dubois) and the researcher (Dries De Wit) will update and implement this DMP.