Mapping the microglial heterogeneity in C9orf72-ALS and sporadic ALS using single cell profiling

A Data Management Plan created using DMPonline.be

Creator: Pegah Masrori

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Grant number / URL: PDMT2/23/078

ID: 204125

Start date: 01-11-2023

End date: 31-10-2026

Project abstract:

Amyotrophic lateral sclerosis (ALS) is an adult-onset progressive neurodegenerative disease and is characterized by the accumulation of neurotoxic protein depositions, consisting of TAR DNA binding protein (TDP-43) in >95% of ALS cases. The neuroanatomical distribution of neuropathological lesions varies and correlates with the clinical presentation. In ALS, the motor system is primarily affected with degeneration of motor neurons in the motor cortex, brainstem and spinal cord. The clinical phenotype of ALS is characterized by a combination of upper and lower motor involvement leading to muscle weakness, hyperreflexia, spasticity, as well as muscle atrophy and fasciculations. The disease typically has a focal onset, most commonly with unilateral distal weakness in one limb or tongue weakness, but has a tendency to spread to other body regions. The most common cause of ALS is GGGGCC hexanucleotide repeat expansions (HRE) in the 5' non-coding region of chromosome 9 open reading frame 72 gene (C9orf72). While it is only partially understood how C9orf72 HRE causes neurodegeneration, several disease mechanisms have been put forward.

An important pathway estimated to be essential in the pathogenesis of ALS is that of inflammation. This process is deemed vital in all neurodegenerative diseases, with an inflammatory reaction as a response to neuronal damage. However, a prolonged activation of inflammatory pathways is detrimental to the remaining neurons and will thereby worsen the disease process. Therefore, studying microglia in ALS is crucial because microglia are a type of immune cell within the central nervous system that play a significant role in maintaining brain health, responding to injuries, and regulating inflammation.

The primary aim of this stuy is to better understand and characterize the interaction between microglia and how this affects motor neurons. First, formalin-fixed paraffin-embedded (FFPE) postmortem samples of ALS patients (available at our lab) will be analyzed using spatial single-cell analysis. Second, we will generate and explore in vitro co-culture models of available induced pluripotent stem cell (iPSC)-derived motor neurons and microglial cells from C9orf72 mutation carriers, sporadic ALS patients (sALS) and controls. Next, we will do pathway mapping to study the mechanisms and elucidate interesting therapeutic targets. By using multiplexed immunohistochemitry-based single-cell analysis, we will assess the spatial heterogeneity of the various microglia cell types in tissue samples, allowing usto investigate which cell subtypes specifically. Fourth, we will translate these findings into diagnostic and prognostic biomarkers of microglial activation in ALS patients.

Last modified: 26-01-2024

Mapping the microglial heterogeneity in C9orf72-ALS and sporadic ALS using single cell profiling

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 / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
snRNAseq_ALS	single nuclei RNA sequencing - postmortem	E	D	N	loom files	>5TB	
scRNAseq_ALS	single cell RNA sequencing - PBMC	E	D	N	loom files	>5TB	
spatial_ALS	spatial transcriptomcics - postmortem	N	D	N	loom files	>5TB	
protemics_ALS	protemics - CSF	N	D	N	loom files	>5TB	

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:

snRNAseq_ALS and scRNAseq_ALS:

https://ega-archive.org/datasets/EGAD00001009686

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- Yes, human subject data (Provide SMEC or EC approval number below)

University Hospitals Leuven, Multi-OMics approach on patient materials from patients with ALS and controls (s65125)

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- Yes (Provide PRET G-number or EC S-number below)

s65125

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 or 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

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).

Prior to publication, sample processing steps and sample metadata will be stored within Electronic Lab Notebooks. Data analysis steps will stored within Digital Notebooks (Jupyter) and version controlled within an in-house Git server. Analysis files will be maintained in an organized project-based file structure detailing the ownership and generation date of all data, and raw data will be stored in an institutional, iRODs backup solution alongside any relevant metadata

Upon publication, data will be submitted to publicly available databases (such as EGA) alongside all information regarding data generation and sample metadata.

Will a metadata standard be used to make it easier to find and reuse the data?
If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- Yes

Data analyzed within the Single Cell Bioinformatics Expertise unit will have metadata generated and stored using standards generated within the unit. The location of all data will be standardized, and metadata will be stored within the unit's metadata database, including sample names, conditions and origins, tissue types, sequencing conditions.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- ManGO

Ongoing analyses will be stored at the Flemish Supercomputing Center (VSC), raw and final data will be stored within the KULeuven ManGO system. Upon publication, data and associated metadata will be stored in public repositories.

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution

The KULeuven ManGO system includes backups, with data stored in multiple location. Notebooks and metadata are stored in databases with version control.

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, 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 both on the VSC and ManGO will be access controlled using standard Linux permissions and Access Control Lists, providing permission only to authorized persons.

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

Storage at the VSC for ongoing analyses are expected to cost approximately 200 EUR/year.
ManGO costs are expected to be approximately 750 EUR/year.

Costs will be covered by Van Den Bosch lab, VIB-KU Leuven Center for Brain & Disease Research.

Data Preservation after the end of the Research Project

Which data will be retained for 10 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

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

- Large Volume Storage (longterm for large volumes)
- Other (specify below)

ManGO.

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

LVS – 1000 EUR/year
ManGO – 750 EUR/year
Costs will be covered by Van Den Bosch lab, VIB-KU Leuven Center for Brain & Disease Research.

Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project?

Please explain per dataset or data type which data will be made available.

- Yes, as restricted data (upon approval, or institutional access only)

The parts of the datasets that were used for a publication, will be made available in full.

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

Requests for accessing raw sequencing data will be reviewed by the VIB data access committee (dac@vib.be). Any data shared will be released via a Data Transfer Agreement that will include the necessary conditions to guarantee protection of personal data (according to European GDPR law). Further information and requests for resources and reagents should be directed to and will be fulfilled by Pegah Masrori (pegah.masrori@kuleuven.be) and Philip Van Damme (philip.vandamme@kuleuven.be).

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 per dataset or data type where appropriate.

- Yes, aspects of dual-use

Phd and postdoc's from my lab will also be working on parts of the datasets. These will not be shared with third parties before we published our findings.

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- Other data repository (specify below)

Raw data: Publicly available repositories, such as EGA.
Processed Data: datasets will be uploaded to SCope which is a fast, user-friendly visualization tool for large-scale scRNA-seq datasets.

When will the data be made available?

- Upon publication of research results

Which data usage licenses are you going to provide?

If none, please explain why.

- Data Transfer Agreement (restricted data)

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- No

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

Not applicable.

Responsibilities

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

Metadata will be documented by the research and technical staff at the time of data collection and analysis, by taking careful notes in the laboratory notebook that refer to specific datasets, and additionally compiling applicable metadata along with the data in the manner described above.

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

The research and technical staff will ensure data storage and back up, with support from René Custers and Kristofer Davie.

Who will manage data preservation and sharing?

The PI is responsible for data preservation and sharing, with support from the research and technical staff involved in the project, from René Custers and Kristofer Davie.

Who will update and implement this DMP?

The PI bears the end responsibility of updating & implementing this DMP.