Plan Overview

A Data Management Plan created using DMPonline.be

Title: Unleashing the Healing Potential of Microglia in neurorepair in the fast-aging killifish

Creator: Juhi Jain

Principal Investigator: Lut Arckens

Data Manager: Lut Arckens, Ayana Rajagopal

Project Administrator: Lut Arckens

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Grant Number/URL: 1S41025N

Project abstract:

The African turquoise killifish is the shortest-living vertebrate, displaying the same cellular and molecular aging hallmarks as humans. The host lab established that young fish are capable of full-blown recovery from traumatic brain injury (TBI) driven by a set of a-typical progenitor populations while aged fish lose this ability. Aged killifish even phenocopy chronic inflammation and glial scarring as seen in injured mammalian brain. A strong age-related change in number, morphology and gene expression of progenitor cells and microglia/macrophages (MG/MF) is observed.
My goal is to map the changes in cell/state diversity of MG/MF and progenitor populations and their cell-cell communication upon aging and injury via snRNA sequencing. I will apply cross-species (killifish-human) cell-cell comparison methods to prioritize and boost the translational value of potential drug targets using computational based drug mapping techniques. Pre-clinical validation will include spatial in situ sequencing and morphometric analyses of MG/MF types that typify the (lack of) recovery response. In sum, I propose a dedicated discovery pipeline to unveil the exact molecular and cellular interactions between acute/chronic neuroinflammation and (lack of) neurorepair using the killifish as bio-gerontology model. Once validated, new hits for drug-based therapy to prevent or overcome MG/MF associated inflammaging will instigate neuroregenerative abilities in the brain of TBI patients.

ID: 212031

Start date: 01-11-2024

End date: 31-10-2028

Last modified: 17-01-2025

Unleashing the Healing Potential of Microglia in neurorepair in the fast-aging killifish
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	
Single cell RNA/Single-nuclei sequencing	RNA sequencing of killifish telencephalon	Generate new data	Digital	Software	.fastq, seurat object, csv, txt, pdf	<10TB	
Spatial sequencing	Spatial sequencing of killifish telencephalon	Generate new data	Digital	• Observational • Experimental • Software	avi,mov,tiff,jpg, ipynb,csv	1-2 TB	
SAMAP	Comparative species analysis	Reuse existing data	Digital	Software	Seurat object, SAMAP object, png, csv, jpg, csv	<50GB	

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:

https://doi.org/10.1016/j.celrep.2023.113395
https://doi.org/10.1038/s41593-022-01199-y

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, animal data

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.

- No

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

By the single cell analysis, the outcomes can be further used in drug discovery platform or in discovering another genetic marker for the cell types/states.

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).

For genomic sequence data (Single cell/Single-nuclei RNA),
NCBI metadata standards for data submission will be used. This includes:-

- Sample ID (NCBI accession and lab ID)
- Sample origin- Date of sampling- Location of sampling
- Tissue- Life stage
- Sequence type and sequence platform

Data processing pipelines stored on GitHub will be accompanied with readme.txt files.
An electronic lab book in which all metadata is shared and curated to common standards will be kept through a lab group on MANGo platform.

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

1. Raw sequencing reads for: NCBI metadata standards for data submission.
2. Processed sequenced data: All data backed-up according to '3-2-1-rule': 3 copies made & stored on 2 types of media (Hardware: 2 PC and external drives + Network: KU Leuven LUNA domain (I: (Personal), L: (Lab)) + MANGo). For specific high-performance jobs, space on VSC HPC cluster
3. Bioinformatic analyses pipeline: readme.txt files with descriptions of the pipelines
4. Omics datasets will be publicly available on GEO/SRA and a user-friendly interface.

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

Where will the data be stored?

Raw sequencing reads:

1. KU Leuven LUNA domain (I: (Personal), L:(Lab)) + MANGo
2. External hard-drive

Processed data:

1. KU Leuven LUNA domain (I: (Personal), L:(Lab)) + MANGo
2. External hard-drive

Bioinformatic analyses pipeline:

1. Personal computer
2. GitHub
3. MANGo
4. HPC storage

How will the data be backed up?

Personal computer
MANGo
KU Leuven LUNA domain (I: (Personal), L:(Shared)) + MANGo
HPC storage

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

We have Lab storage space in MANGo of 10TB storage capacity. My personal One-drive space of 250GB. For further storage need to buy more storage space in MANGo.

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

KU Leuven LUNA domain (I: (Personal), L:(Lab)): Only restricted to me for my personal data and other is shared in the lab protected drive in L Drive server shared between only Arckens lab people.
Personal computer will be well protected with the password.
MANGo is secured with a password and double authentication method.

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

1500€ per year.

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 sequencing and experimental data will be publicly stored >10 years.

1. Raw sequencing reads for:- Single cell RNA (.fastQ)
2. Processed sequenced data:- read alignments to genomes (.bam)- gene expression counts (.counts)- open chromatin regions (.bw .bed)- genome alignments (.maf)
3. Single-cell/nuclei analyses pipeline(Seurat object)

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

1. Raw sequencing reads for: KU Leuven LUNA domain, external hard drive, MANGo
2. Processed sequenced data: KU Leuven LUNA domain, external hard drive, MANGo
3. Single-cell/nuclei analyses pipeline: GitHub

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

600€ per year for 5 more years.

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
- Yes, in a restricted access repository (after approval, institutional access only, …)

1. Raw sequencing reads for:- Single cell RNA (.fastQ)
2. Processed sequenced data:- read alignments to genomes (.bam)- gene expression counts (.counts)- open chromatin regions (.bw .bed)- genome alignments (.maf)
3. Single-cell/nuclei analyses pipeline(Seurat object)

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

Processed sequence data will be stored in at least two location using MANGo and KU Leuven LUNA domain. These data can be reproduced from raw reads and the available processing pipelines (shared on GitHub). After publication, these data will also be freely available upon request and shared through web transfers.

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.

1. Raw sequencing reads for: KU Leuven LUNA domain, external hard drive, MANGo
2. Processed sequenced data: KU Leuven LUNA domain, external hard drive, MANGo
3. Single-cell/nuclei analyses pipeline: GitHub

When will the data be made available?

Upon publication.

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

CC-BY-NC-ND-4.0

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

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

NA

6. Responsibilities

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

Ayana Rajagopal (IT responsible), Prof. Lut Arckens (Supervisor)

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

Ayana Rajagopal (IT responsible), Prof. Lut Arckens (Supervisor)

Who will manage data preservation and sharing?

Ayana Rajagopal (IT responsible), Prof. Lut Arckens (Supervisor)

Who will update and implement this DMP?

Ayana Rajagopal (IT responsible), Prof. Lut Arckens (Supervisor)