Plan Overview

A Data Management Plan created using DMPonline.be

Title: InReverse: Investigating genes that can reverse sarcopenia-associated phenotypes using CRISPR screens in stem cell models.

Creator: Atilgan Yilmaz

Affiliation: KU Leuven (KUL)

Funder: Bijzonder Onderzoeksfonds

Template: KU Leuven BOF-IOF

Project abstract:
Sarcopenia is an age-related disorder defined by the progressive loss of skeletal muscle mass and function, posing a significant health risk to rapidly aging societies. Although several in vivo and in vitro models of this disorder have been established, there are no well-established experimental tools to study the gene networks involved in this disorder in a high-throughput way. Stem cell biology and CRISPR screens offer unique tools that we have been using to study early human development and disease modeling. In this project, we will leverage the cellular models of sarcopenia together with CRISPR screening to systematically search for genes that can reverse the cellular phenotypes of this disorder. These studies will shed light on the gene networks regulating these cellular phenotypes and will potentially reveal putative drug targets.

ID: 214598

Start date: 01-10-2024

End date: 30-09-2028

Last modified: 14-04-2025

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
		Indicate: **N***(ew data) or* **E***(xisting data)*	Indicate: **D**(igital) or **P**(hysical)	Indicate: **A**udiovisual **I**mages **S**ound **N**umerical **T**extual **M**odel **SO**ftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
Nucleic acid samples	H2O2-treated immortalized human myoblasts	New	P	RNA	Physical data	NA	One 15ml tube including a dry pellet of the treated cells will be frozen for each replicate.
Nucleic acid samples	Ceramide-treated immortalized human myoblasts	New	P	RNA	Physical data	NA	One 15ml tube including a dry pellet of the treated cells will be frozen for each replicate
Nucleic acid samples	TNFa-treated immortalized human myoblasts	New	P	RNA	Physical data	NA	One 15ml tube including a dry pellet of the treated cells will be frozen for each replicate
Cell culture samples	CRISPR/Cas9-mediated knock-in line for a TNFa reporter gene in immortalized human myoblasts	New	P	cells	Physical data	NA	A maximum of 10 cryovials containing the edited immortalized myoblasts will be frozen.
Cell culture samples	CRISPR/Cas9-mediated genome-wide loss-of-function mutant library of immortalized human myoblasts	New	P	cells	Physical data	NA	A maximum of 15 cryovials containing the mutant pools of immortalized myoblasts will be frozen.
Cell culture samples	3D-printed skeletal muscle cultures	New	P	3D cell aggregates	Physical data	NA	These 3D cell aggregates will be cryosectioned and imaged by fluorescence microscopy. They will not be stored by cryopreservation.
Sequencing data	RNA sequencing	New	D	T	.sam, .bam, .tsv (.gz), .mtx, .loom, .rds (.gz), .bcl, .fastq (.gz), .csv, .txt, .xlsx, .R, .py	> 5 TB	
Digital images	Phase contrast and fluorescence microscope images, gel scans, graphs, schematic illustrations	New	D	I	.png, .tiff, .jpg, .mp4, .avi, .pdf, .ai, .pptx,	< 5 TB	
Algorithms and scripts	Codes written and reused from existing pipelines for NGS analyses	New	D	SO	R and python scripts	< 1 TB	

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:

Bulk RNA sequencing data: GSE167186, GSE111006, GSE111010 and GSE111016.

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)

Experiments will be performed on immortalized human myoblasts. The research will be performed under normal laboratory safety rules. All necessary laboratory safety measures will be taken. The use of human-derived cell lines in the present study was approved by the Ethics Committee of the University Hospital Leuven (S66794).

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).

- 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.

- 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.

- Yes

The immortalized human myoblasts are obtained from Sorbonne University through an MTA, which restricts the use of these cells solely and exclusively for the purpose of research. However, there are no restrictions to the dissemination of the data obtained throughout the project for academic purposes.

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

Under the MTA conditions, the immortalized human myoblasts used in this study:

(a) will be used solely for purpose of Research described here before;
(b) will not be used in human subjects, in clinical trials, or for diagnostic purposes involving human subjects;
(c) will be used only in compliance with all laws and regulations applicable to the Material; and
(d) will be used only at my group's Site of Investigation at KU Leuven and by scientists working in my laboratory or by scientists or other personnel of the permitted third party only providing research services under my direct responsibility solely for purpose of Research.

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).

To keep data understandable and usable, we will document and store the data generated throughout the project in an electronic laboratory notebook using our Benchling account. We are also utilizing our Teams and KU Leuven OneDrive accounts with a project-based folder structure to store the data. In addition, we will deposit the data in a J drive dedicated for our research group. We will deposit the data in these folders with a file naming strategy that includes a project-based unique identifier code, the name of the scienties, the date of the analysis/experiment, and a code for the instrument used for the measurement. For previously published NGS datasets, we will generate a shared folder in our KU Leuven OneDrive storage space, where we organize the URLs of the source data, the names, dates, authors of the associated manuscripts, the type of NGS datasets and the species of the samples used in the study. We will keep a record of experimental protocols, the digital map of the freezers where experimental materials are kept in numbered boxes, data files (raw and processed), analysis scripts (R and Python -as Jupyter Notebooks), observations during the experiments. NGS samples will be stored for up to 5 years after the end of the project.

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

Next Generation Sequencing data is required to be deposited to public repositories such as GEO, SRA, ArrayExpress or ENA at the time of the publication. The data that has not yet been submitted to these databases will be standardized using the metadata schemes such as Dublin Core, DataCite or Genome Metadata. This standardization will include a title, the name and the affiliation of the creator, date and time references, the subject and a text describing the contents of the dataset and its analysis pipeline, the format of the file, the category of the data such as NGS datasets, images or audio/video files, an identifier and access rights. In any data deposition case, a readme.txt document will be added to the initial directory to include all of the above-mentioned information to allow easy sharing, reusing and interpretation of the data in future.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Shared network drive (J-drive)
- Other (specify below)

In addition to the J-drive, we will also utilize the KU Leuven OneDrive and Teams accounts, as well as our electronic notebooks on a Benchling account for our laboratory. For temporary storage during the analysis of next generation sequencing datasets, we also utilize our allocated space in our Flemish Supercomputer (VSC) account.

How will the data be backed up?

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

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?

The J drive will only be available to the members of the laboratory. Researchers working on the project have the control over who can access the files in their personal or shared OneDrive folders. The VSC storage is reached by personal accounts in a space accessible to group members. The ICTS data center at KU Leuven ensures a mirror network storage as a back-up strategy and easy recovery of the data.

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

The current cost of J Drive is 45.08 Euros/100Gb/Year. OneDrive and Teams accounts are free of charge to KU Leuven personnel. VSC staging costs 130 Euros/TB/Year.
The costs will be covered by part of the allocated project budget.

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...).

- Certain data cannot be kept for 10 years (explain below)

NGS data will be made publicly available upon the publication of the manuscripts. Unpublished NGS data will be made available upon request 5 years after the end of the project. Since the sample quality may drop in time, physical data including the NGS samples and cell pellets will be stored until 5 years after the end of the project.

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

- Shared network drive (J-drive)
- Other (specify below)

All published NGS datasets will be deposited in dedicated publicly available data repositories (i.e. GEO and EBI ArrayExpress). In addition, we will store these sequencing files in VSC archive. All digital data will be stored in our KU Leuven OneDrive folders and J-drive, while the experimental procedures and notes will also be kept in our electronic lab notebook, together with physical notebooks. K-drive will be used for long term storage of published datasets. Algorithms that deviate from the standard analysis pipelines and developed within our group will be stored on Github.

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

For a total of ~8TB over 4 years, we anticipate a cost of around 2500 Euros. These costs will be covered by this grant and other grants.

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 embargoed data (temporary restriction)

Sequencing data will be submitted to public databases (EBI-ENA/NCBI-SRA), where they will be permanently archived to preserve access to the public. Written progress reports and theses will be stored for internal purposes and can be accessed by KU Leuven researcher upon request. The data will be embargoed until the publication of the manuscript and will be made fully available to the public thereafter.

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

As soon as the embargo is lifted on the data, it will be open to public for re-use, conforming the community norms and the rules listed in the relevant public repository. In such cases of re-use, the article associated with the datasets generated under this project will be cited by the third parties re-using the data. For datas shared directly by the PI upon request from third parties, a material transfer agreement will be made to describe the extent and the types of the re-use. Data will also be shared under a CC-BY 4.0 reuse license.

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, intellectual property rights

The MTA that is referred to above imposes restrictions on the use of material for non-academic purposes and requires the consent of the donor for distributing the material for academic purposes.

Where will the data be made available?

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

- Other data repository (specify below)
- Other (specify below)

NGS datasets will be deposited in open access repositories such as the NCBI Gene Expression Omnibus (GEO) and the EBI ArrayExpress databases. Experimental protocols will be documented in the published manuscripts. Manuscripts submitted to scientific journals for publication will be made publicly available in pre-print servers such as BioRxiv. Publications will be added to KU Leuven institutional repository, Lirias.

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.

- CC-BY 4.0 (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.

- Yes, a PID will be added upon deposit in a data repository

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

There are no expected costs for data sharing.

Responsibilities

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

The supervisor, Prof. Atilgan Yilmaz and the PhD student, Angela Moreno Anguita, will manage data documentation and metadata during the research project.

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

PhD student, Angela Moreno Anguita, will manage data storage and backup during the research project. The supervisor, Prof. Atilgan Yilmaz, will ensure that the PhD student is aware of all the storage and backup options listed above.

Who will manage data preservation and sharing?

The supervisor, Prof. Atilgan Yilmaz and the PhD student, Angela Moreno Anguita, will manage data preservation and sharing.

Who will update and implement this DMP?

The supervisor, Prof. Atilgan Yilmaz, will update and implement this DMP.