C14/23/133 Metabolic dysfunction of skeletal cells during osteoporotic bone loss

A Data Management Plan created using DMPonline.be

Creator: Steve Stegen

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Principal Investigator: Steve Stegen

Grant number / **URL:** C14/23/133

ID: 205324

Start date: 01-10-2023

End date: 30-09-2027

Project abstract:

Metabolic diseases such as obesity are often associated with skeletal cell dysfunction and osteoporotic bone loss. The resulting fractures represent an important medical problem that is expected to increase due to our ageing population. The detrimental skeletal effects are not only caused by disrupted bone remodeling, but also by the increase in bone marrow adipocytes. The expansion of marrow adiposity occurs at the expense of bone-forming osteoblasts, as both cell types derive from a common progenitor, but the mechanisms that drive adipocyte accumulation remain largely unknown. In this project, based on our preliminary data and using preclinical bone loss models together with in-depth metabolic analyses, we will test the hypothesis that disturbed cell metabolism alters lineage specification of skeletal progenitors, resulting in increased adipogenic differentiation and impaired bone properties.

Last modified: 07-03-2024

C14/23/133 Metabolic dysfunction of skeletal cells during osteoporotic bone loss

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	
Flow cytometry data	Flow cytometry files (raw flow cytometry files and analysis files)	N	D	I, N, other (experimental)	.fcs .jpg .pdf .xlsx	<1GB	
Single-cell transcriptomics data	Skeletal cell transcriptomes of mice that are fed a control or high-fat diet	N	D	N, other (genetic data)	.var .rds .xlsx	<5TB	
Histological data	Imaging files from microscopes (fluorescent/confocal and brightfield)	N	D	I	.jpg	<1TB	
	Tissue sections	N	P	Other (glass slides)	NA	NA	5 drawers in cupboard, 2 polystyrene foam storage boxes (StarBox)
Mass spectrometry-based metabolomics	Raw files from mass spectrometry, and analyzed data files using Xcalibur/ElMaven	N	D	N, other (metabolic spectra)	.raw .mzML .emDB .sld .pmd	<1TB	
MicroCT data	Raw projection images and image reconstruction	N	D	I	.jpg	<1TB	
Protein expression data	Western blot analysis	N	D	I	.jpg	<100GB	
Body composition	Raw files from EchoMRI analysis	N	D	I, N	.jpg .xlsx	<100GB	
Cell culture	Raw files from cell counts, microscopic pictures	N	D	I, N	.jpg .xlsx .docx	<1GB	
Data analysis and manuscript preparation	Analysis of obtained raw data, and manuscripts describing research findings	N	D	I, T	.xlsx .txt .docx .jpg .tiff .pdf	<100GB	

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:

We will not reuse existing data

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, animal data (Provide ECD reference number below)

We have already obtained an ECD for this project: P214/2023

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

Since this is a fundamental research project, we do not foresee immediate valorization options.
However, if necessary and depending on the outcome of the studies, IP restrictions will be claimed in collaboration with KU Leuven Research & Development before making new data public.

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).

We have implemented the following procedures in the lab:

1. Specific folder structure: each project will be divided into different subfolders, with a logical hierarchy and clearly labeled according the following format: Date (YYYYMMDD)_experiment type_description_researcher initials_number of experiment.
2. Per subfolder: a read-me file (.txt), with a short description of the experiments, and where further information can be found in the lab book of the researcher.
3. Lab book: all necessary information to repeat the experiment is provided (including cell type, animal age, treatment: vehicle buffer/concentration, etc.), together with a specific protocol identifier.
4. Protocols: general research methods and protocols, including those for in vivo (animal) and in vitro (cell culture) experiments, are fully documented as standard operating procedures with a specific numeric identifier and are available for all researchers in the lab.

Additional information for the most important datasets is provided below:
For flow cytometry: information on gating strategy will be saved in electronic files with details on antibody concentrations. Additional information on the specific protocols (cell type/staining procedure) will be found in the lab book of the responsible researcher.

For single-cell transcriptomics: generated sequencing data will be uploaded to GEO database, together with the necessary metadata (e.g. sample type, conditions, sequencing platform/library...) as described in Füllgrabe A (Nat Biotech 2020).

For histology: images and settings for image acquisition will be saved in electronic files (.czi). Details on staining techniques and antibody/dye concentrations will be described in detail in lab books. Histological sections (glass slides) will be labeled to identify sample number and the specific experiment.

For metabolomics: all metabolomics experiments will be labeled with a specific identifier, generated by the VIB Metabolomics Core. All samples are labeled, and a reference will be provided in the lab book (explaining cell type, treatment, condition, etc.). Mass spectrometry settings (used column, buffers, quality controls, etc.) are provided by the Metabolomics Core for each experiment.

For microCT: all generated images will be saved in a folder that is labeled with the specific identifier of the investigated animal. Imaging settings are saved as read-me files in the specific folders.

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.

- No

The metadata standards of the GEO database will be followed when depositing the single-cell transcriptomics datasets.
For now, no existing metadata standard will be used for the other types of experiments, but this issue will be taken up as a regular point during the meetings of the project to see when and if we do need to implement a metadata standard.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Shared network drive (J-drive)
- Personal network drive (I-drive)
- OneDrive (KU Leuven)
- Large Volume Storage

The main data storage drive is the KU Leuven Shared network drive (J-drive), where all data from all researchers will be stored. This specific network drive allows reading and writing data, and thus is the most flexible. It is safely accessible outside KU Leuven premises via two-factor authentication.
In addition, mirror backups will be stored on the Personal network drive (I-drive) and via Cloud service (OneDrive - KU Leuven). Finally, long-term storage of published data will be done at the Large Volume Storage drive (K-drive; see also below for more information on long-term storage policy), but the large transcriptomics/microCT/histology datasets will be additionally backed up on external hard drives.
Specifically for hard copy data (including histological sections), these will be stored in closed cabinets in the researchers' offices.

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution
- Personal back-ups I make (specify below)

As mentioned above, data will be stored on the J-drive and mirrored on the Personal I-drive, as well as on the OneDrive. Additional backups of large transcriptomics/microCT/histology datasets will be done on external hard drives. In total, we will thus have several backup scenarios.
As per lab policy, these mirror backups will be performed once a month.

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

If needed, we will request additional storage space to accommodate new data.

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

Access to the KU Leuven drives/servers is restricted to unauthorized persons, as a two-factor authentication (personnel number/password + additional

log-in) is required.

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

The costs for data storage (+/- €500/year) have been taken into account in the budget calculation and are thus covered by the project.

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)

Sufficient storage and backup capacity is available. We anticipate that the single-cell transcriptomics and microCT data will account for the largest amount of used storage. Specifically for microCT datasets, reconstructed CT data will be discarded 5 years after publication to reduce the total storage cost. Raw CT data will be kept for the required 10 years, and image reconstruction can be rapidly performed based on the documented settings. Raw transcriptomics data are stored at the Genomics Core of the KU Leuven via a payed cloud storage service. An additional mirror backup will be stored at the KU Leuven Archive storage (K-drive).

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

Permanent storage is foreseen at the KU Leuven Archive storage (K-drive), costs are €5,69 per 100 GB per year. These costs will be covered by follow-up projects.

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 open data

In accordance with most publishers' requirements, all data (including specific reagents, transgenic mouse lines, cell lines, etc.) from an accepted manuscript will be made available upon request. Single-cell transcriptomics data will be made available via public repositories (see below).

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 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.

- KU Leuven RDR (Research Data Repository)
- Other data repository (specify below)

Single-cell transcriptomics data will be deposited at Gene Expression Omnibus (GEO).

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.

- 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?

Not applicable

Responsibilities

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

Every lab member who is involved in the project is responsible for the documentation and metadata of his/her own experiments. Guidelines on data acquisition and storage have been implemented in the lab, and are expected to be respected by the researchers.

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

The PI bears the end responsibility of managing data storage and backup, and will provide sufficient and correct data storage space.
Every lab member is responsible for archiving his/her own data on the 'Long Term Storage folder', thereby respecting the guidelines that have been put forth. This aspect also includes 'self-management' of file sharing, keeping track on the volume of data and informing the PI in case of problems. If necessary, we will liaise with the KU Leuven ICT department.

Who will manage data preservation and sharing?

The PI bears the end responsibility of managing data preservation and sharing.
All researchers involved in the project will have unrestricted access to all data, and will be responsible for managing his/her own personal folder according to the laboratory guidelines.
Specific requests for data sharing with 3rd parties will be immediately discussed with the PI.

Who will update and implement this DMP?

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