Plan Overview

A Data Management Plan created using DMPonline.be

Title: (C14/24/151) Skeletal stem/progenitor cell (SSPC) quiescence and activation in health and disease.

Creator: Christa Maes

Principal Investigator: Christa Maes, n.n.

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Principal Investigator: Christa Maes, n.n. n.n.

Project abstract:

Ageing and common metabolic disorders are detrimental for bone integrity and strength, leading to a high fracture risk. New osteogenic drugs and cell therapy procedures are urgently needed. The cells responsible for bone formation, osteoblasts, derive from skeletal stem/progenitor cells (SSPCs); preserving the stemness and quiescence of SSPCs is consequently critical for preserving lifelong bone-forming capacity. Loss of quiescence and aberrant activation and differentiation of SSPCs is also thought to be the primary cause of giant cell tumour of bone (GCTB), a debilitating bone cancer. Here, we will combine these different contexts and employ complementary approaches to bridge biology, regeneration and disease settings, fundamental and translational research, and mouse and human sample input, to unravel the regulatory cues and epigenetic control of SSPC quiescence and osteogenic activation, and the mechanisms by which SSPCs can initiate GCTB. This project will provide insights towards osteo-anabolic drugs for bone diseases and age-related bone loss, bone regenerative medicine, and cancer therapy.

ID: 213707

Start date: 01-10-2024

End date: 30-09-2028

Last modified: 02-04-2025

Grant number / URL: C14/24/151

(C14/24/151) Skeletal stem/progenitor cell (SSPC) quiescence and activation in health and disease.

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	
MicroCT	Raw and processed microCT scans of mouse bones and their analysis	N or E	D	I and N	various formats including .png, .tif., .jpg, .bmp, .xlsx	<5TB	Mouse bones are processed and stored (fridge) or used further for histological processing and analysis.
Histology samples	Biological samples, including embedded mouse bones and cut tissue sections thereof, for general histology, immunohistochemistry, and other staining procedures.	N or E	P	N/A	N/A	N/A	Tissue samples/biopsies, cells and cell lysates/products are stored in dedicated, closed and locked containers at room temperature, 4°C, -20°C, -80°C, or liquid nitrogen (N2).
Histology data	Microscopy images and analysed histomorphometric data of histological samples and sections. Includes raw and processed images, from 2D and 3D microscopy, and numerical data files.	N	D	A, I and N	various formats including .nd2, .vsi, .png, .tif., .jpg, .xlsx	<1TB	N/A
Gene expression samples	Biological samples, including stored or processed biopsies collected from mice (serum, tissues) and cultured cells, and derivates thereof (RNA, DNA, protein), for gene expression analysis during the project or in future research of the lab.	N or E	P	N/A	N/A	N/A	Tissue samples/biopsies, cells and cell lysates/products are stored in dedicated, closed and locked containers at room temperature, 4°C, -20°C, -80°C, or liquid nitrogen (N2).

Gene expression data	Molecular biology data, obtained by qRT-PCR, ELISA, western blot, (bulk/single-cell) RNA-sequencing, multiomics analyses, etc.	N	D	mostly N (some I)	various formats, foremost .eds, .xlsx, .nd2	<1TB	N/A
Flow cytometry data	Raw and analysed flow cytometry data	N	D	N	.fcs, .xlsx	<100GB	N/A
Metabolic data	Metabolic data include results from glucose tolerance tests, insulin tolerance tests, bodyweight, organ/tissue weight, behaviour (feeding, physical activity) data, indirect calorimetry data, etc.	N	D	N	.xlsx	<1GB	N/A
Bulk chromatin profiling data	Bulk CUT&TAG, ACT-Seq or DiMeLo-Seq data	N	D	Sequencing data	FASTQ	<10TB	
Single-cell transcriptome and epigenome data	Single-cell RNA and CUT&TAG data from GCTB patients	N	D	Sequencing data	FASTQ	<10TB	N/A
Cells	Cells derived from mouse tissues and from human samples	N or E	P	N/A	N/A	N/A	Cells and cell lysates/products are stored in dedicated, closed and locked containers in liquid nitrogen (N2) or at -80°C, respectively.

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:

Biological samples and data generated in previous projects under supervision of the PIs of the current project (Christa Maes and Pavlo Lutsik) are planned to be reused.
The project is expected to also rely upon rich data resources provided by public data repositories, collaborators, and consortia:
Single-cell transcriptional skeletal cell atlas (mouse): https://doi.org/10.1101/2022.03.14.484345
Single-cell multiome atlas of embryonic skeletal development: https://doi.org/10.1038/s41586-024-08189-z

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

Animals (mice): Preceding and running studies related to this project have been approved under active ethical dossiers #P016/2022, #P083/2022, #P216/2023, and #P155/2024. Breeding of the mouse lines required for the work runs under Breeding License "000/(GS1/GS2) Breeding-Maes Christa-Breeding non harmful phenotypes".
Whenever needed, additions or amendments to these animal experimentation licenses will be applied for prior to the start of the animal experiments.
Human: A new application dedicated to the human studies on this project is being prepared for submission to the UZ Leuven / KU Leuven Ethics Committee.

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

Pending. This research project plans to involve data from human subjects, but the dossier is still in preparation.
Ethically sensitive data from human subjects implemented or obtained in this project will be handled according to the highest standards of ethical conduct, protection, and privacy of the persons involved. All patient information and genomic data will be stored in pseudonymized form with the key available only to the clinical partner.

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

No commercial valorisation is anticipated at this early stage. If this changes, it will be declared accordingly.

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

Material transfer agreements may apply regarding the use of genetically modified mouse cell lines for the studies in this project and for use of GCTB-related materials generated previously by PL when he was employed at the German Cancer Research Center (DKFZ) in Heidelberg. The MTAs are currently in preparation, pending the ethical approval in Leuven.

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

GCTB-related materials from previous work are the ownership of the Orthopaedic University Hospital Heidelberg and has to be managed in accordance with the earlier acquired approval by the Ethics Committee of the University Heidelberg Medical Faculty (S-341/2021). Already generated and published GCTB-related data from previous work is available in public repositories and can be reused without restrictions.

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).

- Metadata will be collected regarding the mice used for experiments (age, strain, genotype, gender, date of sacrifice, tissue storage and purpose, weight and other relevant parameters); .xlsx and .docx files in dedicated folders for each experiment will contain this information. Moreover, a general “log list” will contain a full summary of all the experiments with information from each individual animal, including type of sample, storage location and processing status.
- Protocols (sampling, processing, histology staining, flow cytometry gating and antibody staining, etc…) will be kept in dedicated folders for each experiment and a printed copy in the personal lab book will be stored in the lab. Furthermore, these protocols will be put on the internal shared lab drive (J) drive, so that they are easily accessible to all current and future lab members.

- Metadata regarding data acquisition (e.g. for image acquisition: objective, camera settings,…) will be stored together with the raw data and processed data, both on the personal hard drives as well as the shared KU Leuven drives.
- The newly generated sequencing data will be supplied with extensive documentation and metadata, including machine level data acquisition and QC reports (e.g. FASTQC files, parameter logs etc)
- We will set up a local project-specific database (as a collection of online spreadsheets), to keep track of all sequencing datasets

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

Both on the personal hard drive and on the shared KU Leuven drives, a logical folder organization will be used such that current and future lab members can easily access all raw and processed data. An easy-to-follow hierarchy will be used, and folder names will be descriptive. Main folders are kept for each researcher or project. Subfolders will be used, with broader topics at higher levels. Next, subfolders will be created according to type of data (e.g. histology, flow cytometry, qPCR,)… Within each type of data folder, descriptive subfolders will be created (depending on type of data) in which the raw and processed data will be stored. As mentioned above, mouse sampling log lists, protocols and metadata regarding data acquisition will be stored on the shared drive, such that they are easily accessible for all lab members. For sequencing data we will explore the applicability of the RDF format to describe the metadata.

Data Storage & Back-up during the Research Project

Where will the data be stored?

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

Mouse data generated in the Maes lab will be stored on the shared network drives, including the KU Leuven shared Large Volume Storage (L-drive); selected raw datasets will be transferred to the KU Leuven data Archive Repository (K-drive). Additionally, throughout the project the PhD students and other personnel involved in this project will keep a copy of the data on personal external hard drives (backup of computer), which will be handled over to the PI Christa Maes after finalization of the project.
For all computational data processing and analysis work (mostly performed in the Lutsik lab) we will make use of the state-of-the-art computational infrastructure available at KU Leuven and Flemisch Supercomputing Center (VSC). Current services include a high-performance HPC, OnDemand cloud computing environment with flexible resource quotas, high-speed network storage with various levels of staging. Basic computing resources are available to Pavlo Lutsik at considerably reduced in-house prices. Primary sequencing data generated with collaborators internally at KU Leuven will be automatically imported to the VSC infrastructure. Local data storage systems are ISO-certified with respect to data protection requirements. Data generated by external collaborators or downloaded from portals of international consortia will be transferred to the VSC infrastructure and handled consistently with the internal data, all under stringent security protocols.

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)

Generally, during and for at least 10 years after the project, all data remains stored on secured, backed-up internal servers provided by KU Leuven IT Service, including Large Volume storage and Archive Storage repositories.

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

Currently there is sufficient storage in the Long-term KU Leuven server drive and in the back up external hard drives. Additional server capacity can be reserved at any time, and additional working hard drives and back up hard drives will be purchased when needed.

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

Shared drives KU Leuven and VSC storage: access is only given to authorized researchers associated to the lab.
Personal hard drives: Hard drives are stored in the lab, which is accessible only to lab members.

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

We expect an approximate 4000 euro per year for this project for data storage on the KU Leuven servers (electronic files). These costs will be covered by the PIs.
Expenses for temperature-controlled storage of biological samples (fridges, freezers, ultra-freezers -80°C, liquid N2) is anticipated to 1000 euro annually. Purchase of freezers has been done before from preceding grants to the PI; costs for repair are occasional.

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

All data will be retained for at least the expected 10-year period after finalization of the project. After that, the PIs Christa Maes and Pavlo Lutsik will decide on what data will be retained. In principle, all key raw and analyzed data are preserved for >10 years or undetermined time after the project. Data or samples deemed uninformative or of no further value for research purposes (e.g., quality loss from long-term stored biological samples) may be discarded after the 10-year period.

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

- Large Volume Storage (longterm for large volumes)
- KU Leuven RDR
- Shared network drive (J-drive)

Data will be stored on the shared network drives, including the KU Leuven shared Large Volume Storage (L-drive); selected raw datasets will be transferred to the KU Leuven data Archive Repository (K-drive). Additionally, external hard drives containing data backups are stored on campus by the PIs after finalization of the project.

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

We expect an approximate 4000 euro per year for this project for data storage on the KU Leuven servers (electronic files) and

expenses for temperature-controlled storage of biological samples (fridges, freezers, ultrafreezers -80°C, liquid N2). These costs will need to be covered by the PIs from grant funding.

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)
- No (closed access)
- Yes, as open data
- Yes, as embargoed data (temporary restriction)

The principle route of sharing the data is via the publication as peer-reviewed, original research papers. Published results will be available, either as open-access research papers, embargoed data, or restricted data, depending on the journal's specifics. Published papers are also deposited on the KU Leuven Lirias repository.

Datasets that are relevant to the community, such as genome-wide transcriptome profiles of cells or tissues (RNA-Seq datasets), will be shared via public repositories (e.g., ENA (the European Nucleotide Archive, from the European Molecular Biology Laboratory (EMBL)). These can be openly accessed by anyone for reuse.

Remaining samples and datasets that are stored and preserved after the project, can be reused internally, by other or new team members, in consultation with the PI.

Workflows implementing in-house developed Common Workflow Language (CWL) and other workflows are shared publicly via the group's Github page: https://github.com/CompEpigen and/or available at e.g.: https://github.com/nf-core.

Any other reuse of data or samples requested by external parties, will be considered on a case-by-case basis by the PI. In principle, data can be made available upon reasonable request after publishing the key results of the project. Remaining samples from the project that are not critical for the own research, can be made available to others in collaboration. No special licenses or transfer agreements will be required when working in this context with academic partners that have research/educational goals and no commercial interest.

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

Results of the study will be published as research papers, preferentially open-access. Datasets that are relevant to the community, such as genome-wide transcriptome profiles of cells or tissues (RNA-Seq datasets), will be shared via public repositories. Any other data can be made available upon reasonable request after publishing the key results of the project. Uses for research and educational purposes will in principle be allowed, if in line with the further plans and goals of the lab. Commercial reuse will in general be excluded. For all internal use of data and samples, authorized researchers associated to the lab have access to the shared KU Leuven drives, to preceding team members' protocols, data, lab notebooks, and stored samples.

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
- Yes, ethical aspects
- Yes, other
- Yes, privacy aspects

Some samples cannot be shared because of prior material transfer agreements that preclude sharing with third parties (e.g., genetically modified mice and their derivates including tissue samples etc.).

Some data and samples would preferably not be shared based on reasons of competition or further plans by the lab for internal reuse and subsequent studies building on the current project.

Some data cannot be fully shared because of privacy aspects (e.g., human patient data).

In general, we will consider only uses for research and educational purposes; commercial reuse or any reuse with ethical issues

will be excluded.

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)

Results of the study will be published as research papers, preferentially open-access, and deposited via the KU Leuven Lirias repository.
Datasets that are relevant to the community, such as genome-wide transcriptome and epigenome profiles of cells or tissues (RNA-Seq and multiomics datasets), will be shared via flagship public repositories (e.g., ENA, the European Nucleotide Archive, from the European Molecular Biology Laboratory (EMBL), the German Genome-Phenome Archive, Array Express, and/or Gene Expression Omnibus). These platforms take responsibility for long-term data availability and exchange.
Any other data can be made available upon reasonable request after publishing the key results of the project.

When will the data be made available?

- Upon publication of research results

The data will likely be made available after publishing the key data and findings of the study.

Which data usage licenses are you going to provide?

If none, please explain why.

- Other (specify below)

None. Data will be made available only after publishing the key results of the project, either as open data (datasets of broad interest) or upon request by email. When relevant, advice will be sought about appropriate usage licenses.
Upon publication, primary and processed data will be deposited in flagship repositories (e.g., European and German Genome-Phenome Archive, Array Express, Gene Expression Omnibus) responsible for long-term data availability and exchange.

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

Published research papers will have a PID/DOI.
Upon publication, primary and processed data will be deposited in flagship repositories (e.g., European and German Genome-Phenome Archive, Array Express, Gene Expression Omnibus) responsible for long-term data availability and exchange.

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

Data sharing is generally not associated with costs. Data are shared either via public repositories (free for depositing and users) or via email or free-of-costs data sharing services (e.g., Belnet, OneDrive). Upon publication, primary and processed data will be deposited in flagship repositories (e.g., European and German Genome-Phenome Archive, Array Express, Gene Expression Omnibus) responsible for long-term data availability and exchange which is free of charge.

Responsibilities

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

The researchers who generate the data are responsible for documentation, metadata, storage and backup.
The supervisors of the project have the end responsibility and manage long-term preservation and sharing.

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

The researchers who generate the data are responsible for documentation, metadata, storage and backup.

Who will manage data preservation and sharing?

The supervisors of the project, Christa Maes and Pavlo Lutsik, have the end responsibility and manage data preservation and sharing.

Who will update and implement this DMP?

The PIs (Christa Maes and Pavlo Lutsik) will update and implement this DMP.