Nuclear metabolism in the control of cancer cells methylome under hypoxia.

A Data Management Plan created using DMPonline.be

Creators: Roxane Verdikt, n.n. n.n.

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: n.n. n.n.

Data Manager: n.n. n.n., Roxane Verdikt

Project Administrator: n.n. n.n., Roxane Verdikt

Grant number / URL: 1266525N

ID: 203535

Start date: 02-10-2023

End date: 01-10-2024

Project abstract:

Hypoxia induces both metabolic and epigenetic remodelling in cancer cells. Several biochemical reactions that require oxygen are directly inhibited, including oxidative phosphorylation and (di)oxygenation, with consequences on DNA methylation profiles. Indeed, TET enzymes are dioxygenases and their decreased activity controls the dynamics of DNA methylation in cancer cells upon hypoxia. However, it remains to be determined how hypoxic-induced metabolic changes concertedly translate to the specific DNA methylation signatures observed. Here, we propose that nuclear nanoreactors, consisting of complexes of transcription factors, metabolic and epigenetic enzymes, control the balance and genomic distribution of key metabolites involved in DNA methylation under hypoxia in cancer cells. We further extend beyond our molecular characterization and show how understanding the fundamental mechanisms of hypoxia allows the development of new anti-cancer pharmacological interventions.

Last modified: 14-12-2023

Nuclear metabolism in the control of cancer cells methylome under hypoxia. DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- No

Nuclear metabolism in the control of cancer cells methylome under hypoxia. GDPR

GDPR

Have you registered personal data processing activities for this project?

- No

Nuclear metabolism in the control of cancer cells methylome under hypoxia. Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

We will perform epigenomics and metabolomics studies to uncover the role of nuclear metabolism in cancer hypoxia. To do so, we will generate the following datatypes:
- confocal microscopy images, about 5-10GB per image, expected 200 images per year - to study the subcellular localization of metabolic enzymes under varying oxygen concentrations. Total estimated volume=2TB.
- high-throughput sequencing data for transcriptomics (RNA-seq), epigenomics (Whole Genome Bisulfite Sequencing, WGBS and chomatin immunoprecipitation, ChIP-seq), and metabolomics (untargeted and targeted metabolic profiling). Total estimated volume=2TB.

All high-throughput sequencing data generated (library QC, raw sequences, as well as the large quantity of processed data) will be processed and stored on password-protected and double-authenticated KULeuven IT infrastructures, in accordance with the KULeuven SOPs, the principles of GDPR 2016/679, and the Belgian privacy law. Upon publication, processed data will be uploaded to a public repository with appropriate access control.

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

1. Designation of responsible person (If already designated, please fill in his/her name.)

   The responsible person for the project will be the lead PI (Bernard Thienpont). To ensure continuity, we will designate a collaborator of the PI with a permanent position at the KU Leuven to ensure all information is properly stored and retrievable.
2. Storage capacity/repository
   - during the research
   - after the research

     Digital files will be stored on local computers and hard drives and/or on KU Leuven servers with backup capacities (KU Leuven large volume storage, KU Leuven Enterprise box, Google cloud storage, the VSC (Flemish Supercomputer Centre) and UZ Leuven server). Physical samples obtained and used in this project will be kept according to the EC licenses and agreements. The EGA, dbGAP, GEO databases will be used.

What’s the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

There are various reasons why some data types might not be kept for 5 years. For instance, if the
file sizes are too large and the raw data can be generated at a fraction of the cost of storage.
Additionally, intermediate files produced are often large and useless to keep, given they can be regenerated easily running the same pipeline (frozen pipeline version needs to be logged).

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

None.

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

None.

Nuclear metabolism in the control of cancer cells methylome under hypoxia. 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
Confocal images	Subcellular localizations of metabolic enzymes under varying oxygen concentrations	Generate new data	Digital	Experimental	.czi, .tiff	<5TB	
High-throughput sequencing data	RNA-seq, ChIP-seq, ChIP-SPACE, WGBS, profiling	Generate new data	Digital	Experimental	.fastqc, .png, .xls	<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:

NA.

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.

- No

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.

- No

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).

From the onset of the project, data documentation will be tailored for their ultimate deposition in public repositories, with spreadsheet headers corresponding to fields required by these public repositories. Technical and analytical methods used to generate the data will be documented in sufficient detail to allow for independent reproduction. These will include analysis package version numbers, sequencing instrument metadata, antibody and analysis kit catalogue and lot number, reagent sources, treatment type and duration, cell passage number, organism, strain, genome build, ….

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.

- No

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

Where will the data be stored?

All data generated will be processed and stored on password-protected KULeuven IT infrastructure, in accordance with the KULeuven SOPs, the principles of GDPR 2016/679, and the Belgian privacy law. Upon publication, processed data will be uploaded to a public repository with appropriate access control.

How will the data be backed up?

Data will initially be stored and processed on the local computing cluster, which has local back-up as well as geo-replication in place. Quality is checked by regular testing of backups.

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

Digital files will be stored on local computers and hard drives and/or on KU Leuven servers with backup capacities (KU Leuven large volume storage, KU Leuven Enterprise box, Google cloud storage, the VSC (Flemish Supercomputer Centre) and UZ Leuven server).

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

All servers and hard drives are password-protected.

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

These costs are minimal since this is an internal server for our university, no fee needs to be paid.

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 pre-mentioned data will be retained in their raw form for at least five years.
Intermediate files that are often large and useless to keep, given they can be regenerated easily running the same pipeline (frozen pipeline version needs to be logged), will be discarded after 5 years.

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

Digital files will be stored on local computers and hard drives and/or on KU Leuven servers with backup capacities (KU Leuven large volume storage, KU Leuven Enterprise box, Google cloud storage, the VSC (Flemish Supercomputer Centre) and UZ Leuven server.

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

These costs are minimal since this is an internal server for our university, no fee needs to be paid.

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

Upon publication, processed data will be uploaded to a public repository with appropriate access control.

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

NA.

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.

- Yes, Privacy aspects
- Yes, Ethical aspects

Where will the data be made available? If already known, please provide a repository per dataset or data type.

The EGA, dbGAP, GEO databases will be used.

When will the data be made available?

Upon publication.

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

As is standard practice with high-throughput sequencing, all data (anonymized and without any sensitive personal data) are made publicly available in open databases for the entire scientific community, without restriction.

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

The doi and GEO accession numbers will be linked with publications.

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

Depositories for high-throughput sequencing data are free.

6. Responsibilities

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

Roxane Verdikt

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

Roxane Verdikt

Who will manage data preservation and sharing?

Roxane Verdikt

Who will update and implement this DMP?

Roxane Verdikt