DEREGULATED MRNA TRANSCRIPTION AND PROTEIN TRANSLATION IN THE TRANSFORMATION OF NORMAL T-CELLS TO MALIGNANT CELLS.

A Data Management Plan created using DMPonline.be

Creators: Jan Cools, First Name Surname

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Grant number / **URL:** IBOF/23/014

ID: 196906

Start date: 01-01-2023

End date: 31-12-2027

Project abstract:

T-cell acute lymphoblastic leukemia (T-ALL) is a cancer that arises in normal blood cells and hijacks normal T-cell development pathways to cause malignant disease. Normal T-cell development is well studied and the majority of genomic changes in T-ALL are known. From these data, there is clear evidence that chromatin structure, transcription and translation are major processes that are deregulated in T-ALL, but how specific mutations cooperate to change chromatin, gene expression and protein translation remains poorly studied. In this project, we aim to deconvolute T-ALL development at single-cell resolution in accurate mouse models of T-ALL by mapping chromatin, transcriptional and in particular protein changes from the early stages of leukemia initiation to the progression towards acute disease. Moreover, we aim to map the interactions of developing leukemia cells with normal cells and determine how transcriptional/translational inhibitors affect leukemia cell function.

Last modified: 10-05-2023

DEREGULATED MRNA TRANSCRIPTION AND PROTEIN TRANSLATION IN THE TRANSFORMATION OF NORMAL T-CELLS TO MALIGNANT CELLS.

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	
RNA-seq mouse proT cells	RNA-seq data (bulk) of proT cells expressing various oncogenic transcription factors	N	D	N		<30GB	
Proteomics mouse proT cells	proteomics data (bulk) of proT cells expressing various oncogenic transcription factors	N	D	N		50GB	
Single-cell RNA-seq data mouse cells	single-cell RNA-seq data of cells harvested from bone marrow, spleen, blood, thymus of mice developing leukemia	N	D	N		<30GB	
Spatial transcriptomics mouse	spatial RNA-seq data of mouse hematopoietic tissues.	N	D	I		100GB	
ribo-seq proT	ribo-seq data of proT cells expressing various oncogenes	N	D	N		<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:

not applicable

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)

P104/2020; P187/2022
an additional new ethical approval was obtained in 2023: 30/2023

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).

- Yes (Provide PRET G-number or EC S-number below)

S66882

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

Potential for patent applications based on expression data or proteomics data.

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).

All datasets have a description of the samples used for analysis. The samples will be described with information on the cell type, the mouse strain, the oncogenes expressed.

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

Data Storage & Back-up during the Research Project

Where will the data be stored?

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

VSC (Vlaams Supercomputer Center).

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)

external hard drive; VSC (Vlaams Supercomputer Center)

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?

secure login (2 factor authorization login)

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

70 Euro per TB per year. These costs can be covered by our consumable costs.

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

Not applicable

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

- Large Volume Storage (longterm for large volumes)
- KU Leuven RDR
- Other (specify below)

- GEO (Gene Expression Omnibus NCBI: https://www.ncbi.nlm.nih.gov/geo/) for RNA & DNA data
- PRIDE (https://www.ebi.ac.uk/pride/) for proteomics data

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

70 Euro per TB per year. These costs can be covered by our consumable costs. There are no costs for GEO.

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

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

not applicable

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)

- GEO (Gene Expression Omnibus NCBI: https://www.ncbi.nlm.nih.gov/geo/) for RNA & DNA data
- PRIDE (https://www.ebi.ac.uk/pride/) for proteomics data

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

all datasets will have an accession number.

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

There are no costs expected for data sharing.

Responsibilities

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

Sofie Demeyer, Marino Caruso

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

Sofie Demeyer, Marino Caruso, Jan Cools, Kim De Keersmaecker

Who will manage data preservation and sharing?

Jan Cools, Kim De Keersmaecker

Who will update and implement this DMP?

Jan Cools