TLX TRANSCRIPTIONAL COMPLEXES IN T-CELL ACUTE LYMPHOBLASTIC LEUKEMIA

A Data Management Plan created using DMPonline.be

Creators: Jan Cools, 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: Jan Cools

Data Manager: n.n. n.n.

Grant number / **URL:** G085823N

ID: 199478

Start date: 01-01-2023

End date: 31-12-2026

Project abstract:

TLX1 and TLX3 are related transcription factors that are upregulated by chromosomal translocations in about 35% of T-cell acute lymphoblastic leukemia cases. We have previously shown that TLX1 can cooperate with the oncogenic kinase NUP214-ABL1 to drive leukemia development and we have now demonstrated a similar cooperation between TLX3 and mutant FLT3 kinase. However, it remains unclear how TLX1 and TLX3 deregulate transcription in T-ALL. In the current project we will use single-cell CRISPR screens and proximity biotinylation of the TLX1/3 transcriptional complexes to obtain more insight in the transcriptional co-factors for TLX1/3. We will validate the importance of the identified co-factors in luciferase reporter assays and chromatin immunoprecipitation followed by sequencing (ChIP-seq) will be used to study where these co-factors co-bind with TLX1/3 in the genome. These results will provide insight in the mechanisms by which TLX1 and TLX3 deregulate transcription in T-ALL and can identify new targets for therapy.

Last modified: 10-05-2023

TLX TRANSCRIPTIONAL COMPLEXES IN T-CELL ACUTE LYMPHOBLASTIC LEUKEMIA
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)

- single-cell RNA-seq data from mouse pro-T cells transformed by TLX factors
- proteomics datasets from mouse pro-T cells for TLX interacting proteins

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: Jan Cools
2. Storage capacity/repository
    - during the research: VSC (Vlaams Supercomputer Center), hard drives, network drives
    - after the research: hard drives, public repositories

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)

no deviation.

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)

Primary mouse cells are used to establish pro-T cell cultures that are used to establish single-cell RNA-seq data and proteomics data. We have ethical approval for the use of mice in our research and for the isolation of bone marrow cells, which are the source for the pro-T cells. The mice are only used to isolate the tissues/cells and are not housed for this project. Based on these ex vivo cultures, which are used for several parts of this project, we can reduce the use of laboratory animals in our research.

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

none

TLX TRANSCRIPTIONAL COMPLEXES IN T-CELL ACUTE LYMPHOBLASTIC LEUKEMIA DPIA

DPIA

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

- Not applicable

TLX TRANSCRIPTIONAL COMPLEXES IN T-CELL ACUTE LYMPHOBLASTIC LEUKEMIA GDPR

GDPR

Have you registered personal data processing activities for this project?

- Not applicable

TLX TRANSCRIPTIONAL COMPLEXES IN T-CELL ACUTE LYMPHOBLASTIC LEUKEMIA
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
single-cell RNA-seq CRISPR screen	single-cell RNA-seq dataset from mouse pro-T cells upon CRISPR inactivation of a set of genes	Generate New data	Digital	Experimental	*next-generation sequence data (Illumina)*	<100GB	
protein interaction data	proteomics data from interaction partners of TLX transcription factors	Generate New data	Digital	Experimental	proteomics dataset (MassSpec analysis)	<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:

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.

- Yes, animal data

The experiments for this project will be conducted on primary mouse pro-T cells for which we only need very few animals. We will isolate bone marrow cells from the animals and culture the cells ex vivo. Thus, we only need post-mortem animals, there is no suffering, and there is no housing of the animals needed. Total number of animals (mice) will be below maximum 10 mice for the entire project.
We have ethical approval to isolate bone marrow cells from mice: 030/2023 "Studie van oncogenen betrokken in de pathogenese van leukemie via beenmergtransplantatie modellen"

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.

- Yes

Both the RNA-seq dataset and the proteomics dataset can lead to the identification of new targets for therapy. We will discuss the data and potential commercial valorization with the Tech transfer office at VIB and LRD at KU Leuven to determine the possibilities for tech transfer. We will work with them to determine a publication plan to ensure that publication does not affect the tech transfer possibilities.

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).

Each dataset will be accompanied by a detailed excel file and text file explaining how the experiment was performed (cells used, oncogenes used, cell culture conditions, amounts of cells used, RNA/protein isolation methods, purification methods, meaning of the different labels used in the dataset).
For the RNA-seq data: this will be the result of a CRISPR screen, so each of the gRNA sequences used in the screen will be described.
For the proteomics data: this will be the result of protein-protein interactions that are mapped, so the exact method to isolate the protein-protein interactions will be described.

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.

- Yes

RNA-seq data is next-generation sequencing data for which specific standards are used to deposit the data. We will follow the recommendations of the KU Leuven genomics core facility. MIAME guidelines will be followed: https://www.ncbi.nlm.nih.gov/geo/info/MIAME.html
Proteomics data has specific standards for data deposits. We will follow the recommendations of the VIB proteomics facility. We will follow the guidelines of the PRIDE data repository: https://www.ebi.ac.uk/pride/markdownpage/specificsoftwareformats

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

Where will the data be stored?

KU Leuven network drives, external hard drive (as additional backup); VSC (Vlaams Supercomputer Center)

How will the data be backed up?

KU Leuven network drives and VSC have automatic back-ups ; the data will also be stored on additional external hard drives as back-up.

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

We pay yearly for storage space at VSC and KU Leuven.

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.

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 data is kept for 10 years according to KU Leuven RDM policy.
Upon completion of the project, data is deposited at GEO (RNA-seq) or PRIDE (proteomics).

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

Upon completion of the project, data is deposited at GEO (RNA-seq) or PRIDE (proteomics).

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

Data storage at GEO and PRIDE is free.

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

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.

- No

We will make sure that IP rights are handled before the data is deposited.

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

Upon completion of the project, data is deposited at GEO (RNA-seq) or PRIDE (proteomics).

When will the data be made available?

Upon completion of the project, or earlier at each publication requiring the data deposit.

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

The dataset generated in this project are from mouse cells and will be made publicly available without license.

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

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

No costs for datasharing.

6. Responsibilities

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

Sofie Demeyer

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

Sofie Demeyer

Who will manage data preservation and sharing?

Jan Cools

Who will update and implement this DMP?

Jan Cools