Analysis of JAK-inhibitor treatment for STAT1 gain-of-function and development of a CRISPR-based gene therapy approach

A Data Management Plan created using DMPonline.be

Creators: Cecilia Iglesias Herrero, n.n. n.n., 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., n.n. n.n.

Project Administrator: Cecilia Iglesias Herrero

Grant number / **URL:** 1S07023N

ID: 196832

Start date: 01-11-2022

End date: 31-10-2026

Project abstract:

Inborn errors of immunity (IEI) are rare inherited disorders affecting the immune system presenting with increased susceptibility to pathogens and often associated with severe non-infectious comorbidities. Their poor understanding makes them difficult to diagnose and thus, challenging to treat or cure. The monogenic nature of several IEI converts them in the perfect target for precision medicine. In this project, I focus in autosomal dominant (AD) signal transducer and activator of
transcription 1 (STAT1) gain-of-function (GOF), which presents with a range of mutations and a broad range of mostly unexplained phenotypes. STAT1 is a pivotal transcription factor in the immune response. In follow-up of our recent publications I propose to (i) assess the underlying molecular mechanisms behind the responsiveness to JAK inhibitors treatment in the different mutations to improve their use in the clinics; (ii) study whether the overexpression of STAT1 wild-type results in a
GOF phenotype; and ultimately (iii) develop a proof-of-concept of a safer curative gene therapeutical approach for most, if not all, STAT1 GOFs using the last advances in viral vector technology.
Altogether, I aim to improve the quality of life of patients providing insights into the disease, optimize use of existing drugs and proof a potential gene therapeutical strategy for GOF disorders.

Last modified: 29-03-2023

Analysis of JAK-inhibitor treatment for STAT1 gain-of-function and development of a CRISPR-based gene therapy approach
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**
JAKi *in vitro* – THP1 cell lines	Existing STAT1 mutation-specific cell lines to be used for JAKi compounds testing 7 cell lines	☐Generate new data ☒Reuse existing data	☐Digital ☒Physical	☐Observational ☐Experimental ☐Compiled/ aggregated data ☐Simulation data ☐Software ☐Other ☒NA	☐.fcs ☐.tiff ☐.tab ☐.cvs ☐.pdf ☐.txt ☐.fastq ☐.bam ☒ NA	☐<100MB ☐<1GB ☐<100GB ☐<1TB ☐<5TB ☐<10TB ☐<50TB ☐>50TB ☒NA	THP1 cells lines are stored in liquid nitrogen (LN2) in MolMed, KUL
JAKi *in vitro* – pSTAT1 data	Flow cytometry and western blot data for the pSTAT1 and STAT1 measurements on cell lines	☒Generate new data ☐Reuse existing data	☒Digital ☐Physical	☐Observational ☒Experimental ☐Compiled/ aggregated data ☐Simulation data ☐Software ☐Other ☐NA	☒.fcs ☒.tiff ☐.tab ☐.cvs ☐.pdf ☐.txt ☐.fastq ☐.bam ☐NA	☐<100MB ☐<1GB ☒<100GB ☐<1TB ☐<5TB ☐<10TB ☐<50TB ☐>50TB ☐NA	None
JAKi *in vitro* – Sequencing	Bulk RNAseq sequencing data to study effect of JAKi in IFNy pathway 7 cell lines x 4 conditions x 5 replicates	☒Generate new data ☐Reuse existing data	☒Digital ☒Physical	☐Observational ☒Experimental ☐Compiled/ aggregated data ☐Simulation data ☐Software ☐Other ☐NA	☐.fcs ☐.tiff ☐.tab ☒.cvs ☐.pdf ☒.txt ☒.fastq ☒.bam ☐NA	☐<100MB ☐<1GB ☐<100GB ☒<1TB ☐<5TB ☐<10TB ☐<50TB ☐>50TB ☐NA	RNA samples will be stored in LN2 at MolMed, KULeuven
JAKi *in vivo* – PBMCs	Healthy and patients' PBMCs obtained in UZLeuven	☐Generate new data ☒Reuse existing data	☐Digital ☒Physical	☐Observational ☐Experimental ☐Compiled/ aggregated data ☐Simulation data ☐Software ☐Other ☒NA	☐.fcs ☐.tiff ☐.tab ☐.cvs ☐.pdf ☐.txt ☐.fastq ☐.bam ☒NA	☐<100MB ☐<1GB ☐<100GB ☐<1TB ☐<5TB ☐<10TB ☐<50TB ☐>50TB ☒NA	PBMCs are stored at the Biobank in KULeuven
JAKi *in vivo* – pSTAT1 data	Flow cytometry and western blot data for the pSTAT1 and STAT1 measurements on PBMCs	☒Generate new data ☐Reuse existing data	☒Digital ☐Physical	☐Observational ☒Experimental ☐Compiled/ aggregated data ☐Simulation data ☐Software ☐Other ☐NA	☒.fcs ☒.tiff ☐.tab ☐.cvs ☐.pdf ☐.txt ☐.fastq ☐.bam ☐NA	☐<100MB ☐<1GB ☒<100GB ☐<1TB ☐<5TB ☐<10TB ☐<50TB ☐>50TB ☐NA	None
JAKi *in vivo* – Sequencing	Bulk RNAseq sequencing data to study effect of JAKi in IFNy pathway # samples x 4 conditions x 5 replicates	☒Generate new data ☐Reuse existing data	☒Digital ☐Physical	☐Observational ☒Experimental ☐Compiled/ aggregated data ☐Simulation data ☐Software ☐Other ☐NA	☐.fcs ☐.tiff ☐.tab ☒.cvs ☐.pdf ☐.txt ☒.fastq ☒.bam ☐NA	☐<100MB ☐<1GB ☐<100GB ☒<1TB ☐<5TB ☐<10TB ☐<50TB ☐>50TB ☐NA	RNA samples will be stored in LN2 at MolMed, KULeuven

Overexpression – THP1 cell lines	Existing STAT1 mutation-specific cell lines to be used for JAKi compounds testing 7 cell lines (Same as for JAKi)	☐Generate new data ☒Reuse existing data	☐Digital ☒Physical	☐Observational ☐Experimental ☐Compiled/ aggregated data ☐Simulation data ☐Software ☐Other ☒NA	☐.fcs ☐.xml ☐.tab ☐.cvs ☐.pdf ☐.txt ☐.fastq ☐.bam ☒NA	☐<100MB ☐<1GB ☐<100GB ☐<1TB ☐<5TB ☐<10TB ☐<50TB ☐>50TB ☒NA	THP1 cells lines are stored in liquid nitrogen (LN2) in MolMed, KUL
Overexpression – Primary Cells	EBV-B and CD4+ primary cells from healthy donors obtained in UZLeuven	☐Generate new data ☒Reuse existing data	☐Digital ☒Physical	☐Observational ☐Experimental ☐Compiled/ aggregated data ☐Simulation data ☐Software ☐Other ☒NA	☐.fcs ☐.tiff ☐.tab ☐.cvs ☐.pdf ☐.txt ☐.fastq ☐.bam ☒NA	☐<100MB ☐<1GB ☐<100GB ☐<1TB ☐<5TB ☐<10TB ☐<50TB ☐>50TB ☒NA	PBMCs are stored at the Biobank in KULeuven
Overexpression – pSTAT1 data	Flow cytometry and western blot data for the pSTAT1 and STAT1 measurements on PBMCs	☒Generate new data ☐Reuse existing data	☒Digital ☐Physical	☐Observational ☒Experimental ☐Compiled/ aggregated data ☐Simulation data ☐Software ☐Other ☐NA	☒.fcs ☒.tiff ☐.tab ☐.cvs ☐.pdf ☐.txt ☐.fastq ☐.bam ☐NA	☐<100MB ☐<1GB ☒<100GB ☐<1TB ☐<5TB ☐<10TB ☐<50TB ☐>50TB ☐NA	None
Overexpression – Sequencing	Bulk RNAseq sequencing data to study effect of JAKi in IFNγ pathway # samples x 2 conditions x 5 replicates (From JAKi dataset)	☐Generate new data ☒Reuse existing data	☒Digital ☐Physical	☐Observational ☒Experimental ☐Compiled/ aggregated data ☐Simulation data ☐Software ☐Other ☐NA	☐.fcs ☐.tiff ☐.tab ☒.cvs ☐.pdf ☐.txt ☒.fastq ☒.bam ☐NA	☐<100MB ☐<1GB ☐<100GB ☒<1TB ☐<5TB ☐<10TB ☐<50TB ☐>50TB ☐NA	None
JAKi *in vivo* – scRNAseq	Single-cell RNAseq data to study the *in vivo* effect of JAKi in patients currently in treatment 2 patients x 3 time -points	☒Generate new data ☐Reuse existing data	☒Digital ☐Physical	☐Observational ☒Experimental ☐Compiled/ aggregated data ☐Simulation data ☐Software ☐Other ☐NA	☐.fcs ☐.tiff ☐.tab ☐.cvs ☐.pdf ☒.txt ☒.fastq ☒.bam ☐NA	☐<100MB ☐<1GB ☐<100GB ☒<1TB ☐<5TB ☐<10TB ☐<50TB ☐>50TB ☐NA	None
Gene Therapy – KO/Addback – THP1 cell lines	THP1 cell lines from previous gene therapy experiments. Cells are treated with Nanoblades (NBs) and Lentiviral vectors (LVs)	☐Generate new data ☒Reuse existing data	☐Digital ☒Physical	☐Observational ☐Experimental ☐Compiled/ aggregated data ☐Simulation data ☐Software ☐Other ☒NA	☐.fcs ☐.tiff ☐.tab ☐.cvs ☐.pdf ☐.txt ☐.fastq ☐.bam ☒NA	☐<100MB ☐<1GB ☐<100GB ☐<1TB ☐<5TB ☐<10TB ☐<50TB ☐>50TB ☒NA	Cell lines are stored in LN2 at MolMed, KUL. NBs and LVs are stored at -80ºC at LVVC, KULeuven
Gene Therapy – KO/Addback – Sequencing	Bulk RNAseq sequencing data to study effect of gene therapy IFNγ pathway 3 cell lines x 5 replicates	☒Generate new data ☐Reuse existing data	☒Digital ☐Physical	☐Observational ☒Experimental ☐Compiled/ aggregated data ☐Simulation data ☐Software ☐Other ☐NA	☐.fcs ☐.tiff ☐.tab ☒.cvs ☐.pdf ☐.txt ☒.fastq ☒.bam ☐NA	☐<100MB ☐<1GB ☐<100GB ☒<1TB ☐<5TB ☐<10TB ☐<50TB ☐>50TB ☐NA	RNA samples will be stored in LN2 at MolMed, KULeuven
Gene Therapy – HDR – THP1 cell lines	Cell lines generated via transduction using NBs and rAAV viral vectors	☒Generate new data ☐Reuse existing data	☐Digital ☒Physical	☐Observational ☐Experimental ☐Compiled/ aggregated data ☐Simulation data ☐Software ☐Other ☒NA	☐.fcs ☐.tiff ☐.tab ☐.cvs ☐.pdf ☐.txt ☐.fastq ☐.bam ☒NA	☐<100MB ☐<1GB ☐<100GB ☐<1TB ☐<5TB ☐<10TB ☐<50TB ☐>50TB ☒NA	Cell lines will be stored in LN2 at MolMed, KUL. NBs and LVs will be stored at -80ºC at LVVC, KULeuven
Gene Therapy – HDR – Primary cells	Primary cells from healthy donors and patients obtained in UZLeuven to be transduced by NBs and rAAV viral vectors	☒Generate new data ☒Reuse existing data	☐Digital ☒Physical	☐Observational ☐Experimental ☐Compiled/ aggregated data ☐Simulation data ☐Software ☐Other ☒NA	☐.fcs ☐.tiff ☐.tab ☐.cvs ☐.pdf ☐.txt ☐.fastq ☐.bam ☒NA	☐<100MB ☐<1GB ☐<100GB ☐<1TB ☐<5TB ☐<10TB ☐<50TB ☐>50TB ☐NA	Cell lines will be stored in LN2 at MolMed, KUL. NBs and LVs will be stored at -80ºC at LVVC, KULeuven

Gene Therapy – HDR – HSCs	Hematopoietic stem cells from healthy donors will be provided by the Verhoeyens' lab	☐Generate new data ☒Reuse existing data	☐Digital ☒Physical	☐Observational ☐Experimental ☐Compiled/ aggregated data ☐Simulation data ☐Software ☐Other ☒NA	☐.fcs ☐.tiff ☐.tab ☐.cvs ☐.pdf ☐.txt ☐.fastq ☐.bam ☒NA	☐<100MB ☐<1GB ☐<100GB ☐<1TB ☐<5TB ☐<10TB ☐<50TB ☐>50TB ☒NA	HSCs will be stored in the Verhoeyen's lab at INSERM, Lyon (France) following their DMP procedure
Gene Therapy – HDR – Characterization	Flow cytometry and western blot data for the pSTAT1 and STAT1 measurements and immune cells populations	☒Generate new data ☐Reuse existing data	☒Digital ☐Physical	☐Observational ☒Experimental ☐Compiled/ aggregated data ☐Simulation data ☐Software ☐Other ☐NA	☒.fcs ☒.tiff ☐.tab ☐.cvs ☒.pdf ☒.txt ☒.fastq ☐.bam ☐NA	☐<100MB ☐<1GB ☒<100GB ☐<1TB ☐<5TB ☐<10TB ☐<50TB ☐>50TB ☐NA	None

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:

Transcriptomics data from Giovannozzi *et al.,* 2021 might be reused for corroboration of the mutation-specific transcriptomic profile. DOI: https://doi.org/10.3389/fimmu.2021.632997

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, human subject data

The study protocol for the related studies (ethical committee approval S58466) have been made conformed to recent GDPR-regulations. Assays on hematopoietic stem cells (HSCs) from healthy donors will be obtained via obtained via the collaboration at INSERM, Lyon; all related assays will be performed (and remain there). HSCs are obtained after informed consent of donors and approval was obtained by the ethics committees of the hospitals according to the Helsinki Declaration.

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.

- Yes

Primary cells data are held by Prof. Rik Schrijvers (promotor of the project), the cell vials contain a symbol identifiers that uniquely matches a specific patient. Relevant clinical information such as symptomatology and clinical measurements will be used when necessary, always ensuring the privacy of the patient.

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

The final proof-of-concept for the gene therapy strategy might have potential for commercial valorization in terms of tech transfer.

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.

- Yes

Hematopoietic stem cells' (HSCs) data that will come out of the collaboration with the Verhoeyen's group at INSERM Lyon (France). IPR will be owned by both parts.

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

Nanoblades, also known as, Cas9-engineered virus-like particles are patented by Philippe Mangeot and Emiliano Ricci; application number: WO 2017/068077 Al; patent status: published, 27th April 2017.

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).

All information obtained in the Laboratory for Allergy and Clinical Immunology (LAKI) lab and Laboratory for Molecular virology and Gene Therapy (LMVGT) (division Molecular Medicine (MolMed)), as well as the information obtained from the collaboration, will be stored in synchronized electronic lab notebooks (OneNote) backed up on the university drives.

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

Bulk RNAseq and scRNAseq data will be made available at the following repository: Gene Expression Omnibus (GEO Database) from NCBI.
Metadata standards will be used for all data types, including: date, who, machine, type of experiment, treatment, ...

Dataset	Technique	DMP Standard
JAKi *in vitro* – THP1 cell lines Overexpression – THP1 cell lines Overexpression – Primary Cells Gene Therapy – KO/Addback – THP1 cell lines Gene Therapy – HDR – THP1 cell lines Gene Therapy – HDR – Patient-derived cell lines	Viral vector generation Cell lines generation	Both produced using and in-house DMP standard from Leuven Viral Vector Core (directed by the co-promotor of this project
JAKi in vitro – pSTAT1 data JAKi *in vivo* – pSTAT1 data Overexpression – pSTAT1 data Gene Therapy – HDR – Characterization	Flow cytometry data	DOI: 10.25504/FAIRsharing.kcnjj2
JAKi *in vitro* – Sequencing JAKi *in vivo* – Sequencing Overexpression – Sequencing Gene Therapy – KO/Addback – Sequencing	Next generation sequencing	DOI: 10.25504/FAIRsharing.a55z32
JAKi *in vivo* – PBMCs	Patient sampling	Performed by physicians following UZ Leuven Stardards
Gene Therapy – HDR – HSCs	HSCs obtention and transduction	DMP Standard present at INSERM, Lyon

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

Where will the data be stored?

Data will be stored in the servers of KU Leuven for both laboratories. Ultimately, data will be stored in the archive drive in KU Leuven servers to ensure read-only mode.

How will the data be backed up?

Data will be be backed up in the J and K drive of MolMed and LAKI labs at the KU Leuven servers, respectively.

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

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

VSC stored data is only available for modification by the owner and authorized people of the project: Cecilia Iglesias Herrero, Prof. Rik Schrijvers and Prof. Rik Gijsbers. VSC access is protected under password.
Shared and Archive Drives in KU Leuven servers' data are managed only by allowed users. K drive is read-only. Servers are protected under password and double authentication.
Biobanked samples: locked fridges in a locked lab.
LN2 and -80ºC Samples and viral vectors (rAAV and LV-based): are managed only by staff of the Leuven Viral Vector Core (LVVC) at Molecular Medicine lab.

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

6 Euro per 100Gb/year. Project funding FWO (G054022N)

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 will be retained for at least five years after the end of the project.

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

Ultimately, data will be archived in the K drive in KU Leuven servers. This drive is read-only and is protected under password and double authentication.
Protocols and DMP standards will be stored in electronic lab books.

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

6 Euro per 100Gb/year. Project funding FWO (G054022N)

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

Bulk RNAseq data from cell lines will be made available at the following repository: Gene Expression Omnibus (GEO Database) upon publication.
Patient-derived RNAseq data will be available only under request to ensure patient privacy.
Raw data from the studies will be available upon request. This willl be specified in the publications.

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

Restricted data can be used by all the scientists when privacy can be ensured according to GDPR regulations

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

Privacy aspects are involved being patients with very rare cases of STAT1 gain-of-function a disease.

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

GEO Database for cell lines RNAseq.
Datasets that will be publicly available in this repository:

- JAKi *in vitro* – Sequencing
- Overexpression – Sequencing
- Gene Therapy – KO/Addback – Sequencing

When will the data be made available?

Data will be made available upon publication.

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

NA

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

Not available yet.

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

No costs will be involved.

6. Responsibilities

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

Cecilia Iglesias Herrero and the promotor (Prof. Rik Schrijvers) of the project. Additionally, Sabien Fevery is responsible for Data Management at the Laboratory of Allergy and Clinical Immunology

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

Myself, Cecilia Iglesias Herrero

Who will manage data preservation and sharing?

The data manager and the promotor of the project.

Who will update and implement this DMP?

Myself, Cecilia Iglesias Herrero and the Promotor of this project: Prof. Rik Schrijvers