C14/23/121 - Omics en cel-gebaseerde modellen voor het ontdekken van erfelijke bloedingsziekten.

A Data Management Plan created using DMPonline.be

Creators: Kathleen Freson, n.n. n.n.

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Principal Investigator: Kathleen Freson, n.n. n.n.

Grant number / URL: C14/23/121

ID: 202182

Start date: 30-10-2023

End date: 30-09-2027

Project abstract:

Het samenspel tussen bloedplaatjes en coagulatie is belangrijk om bloedingen te voorkomen en DNA variatie in genen die deze processen regelen kan aanleiding geven tot een erfelijk bloedingsprobleem. Screening met een genpanel test kan een diagnose stellen bij 48%, 26%, en 63% van de patiënten met een defect in bloedplaatjesvorming en -functie en in coagulatie, respectievelijk, terwijl bij slechts 3% van de patiënten met bloedingen van ongekende etiologie een diagnose gesteld werd. Als er een genetisch defect gevonden wordt, duurt het vaak nog jaren alvorens de ziekte echt begrepen is omdat er geen cel-gebaseerde modellen voorhanden zijn. Onze studie zal zich richten op deze tekortkomingen door gebruik te maken van genomen die al beschikbaar zijn voor patiënten zonder diagnose uit de 4 groepen vermeld. Deze data zullen aangevuld worden met genexpressie profielen uit bloedplaatjes, neutrofielen, monocyten, T cellen en Blood Outgrowth Endothelial Cells (BOEC) en het plasma proteoom van deze patiënten. Genexpressie en splicing zal bestudeerd worden met behulp van machine-learning technieken om de detectie van (niet)coderende varianten te bevorderen. Functionele studies in BOEC-culturen en het plasma proteoom zullen voor het eerst gebruikt worden voor de karakterisatie van patiënten met bloedingen van ongekende etiologie. BOEC en stamcel afgeleide modellen zullen ook gebruikt worden om nieuwe genen te valideren. Dit project gaat bijdragen tot nieuwe inzichten voor bloedingen en de modellen kunnen gebruikt worden voor therapeutische ontdekkingen.

Last modified: 10-10-2023

C14/23/121 - Omics en cel-gebaseerde modellen voor het ontdekken van erfelijke bloedingsziekten.

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	
Whole genome sequencing (WGS)	WGS data of patients with an inherited bleeding disorder	E and N	D	Other (genetic data)	.bam .fastq .count .txt .csv	>5TB	
RNAsequencing (bulk)	blood cell and endothelial cell transcriptomes of patients with an inherited bleeding disorder	E and N	D	Other (genetic data)	.bam .fastq .count .txt .csv	>5TB	
Flow cytometry data	Flow Cytometry and FACS sort files (FlowJo and equipment specific files)	N	D	I Other (experimental)	.fcs .jpg .pdf .tiff	<1GB	
Cytation5 imager	Imaging and quantification files from BioTek Cytation5 imager	N	D	I	.xml .jpg .pdf .tiff .mp4 .exp .imm	<100GB	
Fluorescent and Confocal microscopy	Imaging files from microscopes	N	D	I	.jpg .pdf .tiff .czi	<100GB	
Capillaroscopy (nailfold)	images of capillaries in nailfold from patients using an Dermlite equipped iPhone	N	D	I	.tiff .jpg	<1GB	
Cell culture	Cultures of imMKCL and BOEC cell lines	N E	P	Other (cells)	.xml	NA	Storage in N2 tank (< 50 different lines)
Data analysis and manuscript preparation		N	D	I T A	.xml .txt .docx .jpg .pdf	<1GB	

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:

This project will use data previously obtained in other projects. These data are stored in the high performance computer facility in Cambridge and on KULeuven Server. The group from K Freson has access to these datasets on both servers.
- WES en WGS data (S50025)

This data set is also published (Turro et al, Nature 2020) and has been submitted to the European Genome-phenome Archive (EGA): EGAD00001004519

- Blood cell RNAsequencing data (S63666)

These data have been obtained from UZLeuven patients.

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

- WES en WGS data (S50025)
- Bloedcel RNAsequencing data (S63666)
- New data collection: WGS, BOEC isolation and RNAseq and capillary microscopy (Ethical approval pending)

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)

S50025 and S63666
New data (ethical approval pending)
Whole genome sequencing, RNA sequencing and personal metadata (age, gender, clinical condition, laboratory data (hemostatic tests results), blood counts) will be collected. Patient data are pseudo-anonymised.

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

If data will be obtained of interest for valorization, IP restriction will be claimed. It is not clear from the start what novel genetic targets relevant for megakaryopoeisis and platelet formation/function or endothelial function can be identified.

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

The megakaryocyte model itself "imMKCL" cannot be protected by IP restrictions as this model was developed by our collaborators and we have signed an MTA (with dr K Eto, Kyoto university). If this model is needed for the overall IP restriction, a joined application is a possibility (as stated in the MTA). This MTA was recently renewed (08/2023).

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).

Generated sequencing data will be uploaded to EGA in combination with related metadata (e.g. age, gender, case/control status, sequencing platform/library... etc.) to be accessible to the public.

Flow cytometry and sorting: information on gating strategy for cell identification and sorting will be saved in electronic files with details on antibody concentrations and protocols for cell preparation and staining will be described in detail in electronic lab books.

Imaging (confocal and cytation5 imager): images and settings will be saved in electronic files. Details on staining techniques and antibody or dye concentrations and protocols for cell preparation will be described in detail in lab books.

Datafiles and the imaging protocols from the nailfold capillaroscopy will be stored on KULeuven servers.

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.

- Yes

The metadata standards of EGA will be used for submission of sequencing data, as can be consulted on https://ega-archive.org/submission/sequence/unaligned

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Shared network drive (J-drive)
- Personal network drive (I-drive)
- OneDrive (KU Leuven)

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution

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?

The KU Leuven servers are protected from unauthorized persons. The genetic data on the Cambridge server are only accessible for scientists working on the same project as determined by the dAC.

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

The costs are covered by the project and are in the range of 500 euro/year.

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

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

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

KULeuven Archive and lvs drives

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

The costs are covered by project funding
Permanent storage is foreseen at the KULeuven Archive storage K-drive: €5,69 per 100GB per year.

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)
- Yes, as open data

Data access (genetic data) is only possible via the DAC from UZLeuven.
Some data will become available when published.
Other data can be obtained by researchers after request and approval by the PI (Kathleen Freson).

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

Researchers can access the data after request.

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, privacy aspects
- Yes, intellectual property rights

IP only when applicable

Genetic data can be shared but not DNA samples

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- Other data repository (specify below)

The WGS and RNAseq data will be made available on the EGA, where access to data is granted based on applications to a data access committee that oversees the dataset.

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.

- Data Transfer Agreement (restricted 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

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

not applicable

Responsibilities

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

Kathleen Freson

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

Kathleen freson and all researcher within the group that work on the project

Who will manage data preservation and sharing?

Kathleen Freson

Who will update and implement this DMP?

Kathleen Freson