DMP title

Project Name Novel Diagnostics through Methylation Profiling of Cell-free DNA - DMP title
Grant Title G0B4822N
Principal Investigator / Researcher Bernard Thienpont
Project Data Contact +32 16 37 32 08; bernard.thienpont@kuleuven.be
Description We will attempt to: 1.) Leverage cell-free DNA methylation data for development of a disease prediction model in preeclampsia 2.) Apply our optimized model for the development of a highly targeted cfDNA methylation panel 3.) Demonstrate the performance of the classifier in an independent patient cohort. 4.) Extend this strategy to 2 other pathologies (i.e. uveal melanoma and IgA nephropathy). Cell-free DNA will be collected from liquid biopsy samples (blood, urine, aqueous and vitreous humour) taken from both patients and controls.
Institution KU Leuven

1. General Information

Name applicant

Bernard Thienpont

FWO Project Number & Title

FWO number: G0B4822N

Title: Novel Diagnostics through Methylation Profiling of Cell-free DNA

Affiliation

- KU Leuven

2. Data description

Will you generate/collect new data and/or make use of existing data?

- Generate new data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

Type of data	Format	Volume	How created
Raw sequencing output files	.fastq	1000-4000 GB	Output files of bisulfite sequencing experiments.
Processed data	.csv	200 GB	Processed sequencing files after alignment and methylation calling.
Code	.py, .R, .ipynb, .sh	1 GB	Code written by the researcher that was used for processing of data and statistics
Results	.pdf, .png, .svg	5 GB	Results include produced figures and tables along with explaining text.

3. Legal and ethical issues

Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- Yes

Privacy Registry Reference:
Short description of the kind of personal data that will be used: name, age, sex, personal and pregnancy related characteristics, relevant medical history.

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, add the reference to the formal approval by the relevant ethical review committee(s)

- Yes

We received favorable assessment from the local ethics committee for the described research on ocular tumours (S66089) and on preeclampsia (S61883). Approval for the research on kidney samples was granted in ZOL, where these samples will be collected (Z-20211101). Data will be pseudonimyzed by the collaborating clinicans.

Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

- Yes

Some of the data to be produced are dependent on analyses that are the subject of an ongoing patent application. Also other datasets and analyses to be generated in this project could be ameanable to patent protection.

Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- No

4. Documentation and metadata

What documentation will be provided to enable reuse of the data collected/generated in this project?

1. Methodology and protocol of wet lab work will be described in the lab book. Additionally, sequencing protocols will be written down.
2. Bioinformatic analyses, processing of sequencing data and statistics along with figure production will be saved as code. Generated files during processing will be archived.

Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- Yes

Metadata will be saved according to the MINSEQE standards.

5. Data storage and backup during the FWO project

Where will the data be stored?

Data will mainly be securely stored on the Vlaamse Super Computer. During processing and analyses, data will also temporarily be stored on the KU Leuven laptop of the researcher, with access protection through a password with dual authentication.

How is backup of the data provided?

Data is stored on VSC, which is georeplicated and backed 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

34TB of storage is available on the VSC for the host lab, which is easily sufficient to contain the generated data.

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

Current storage options for actively analysing data on VSC cost 630€ per year.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

Data on the VSC is placed under restricted access for everyone except the researcher.

6. Data preservation after the FWO project

Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).

All data described above (raw and processed sequencing data, scripts and results) will be retained for the minimum preservation term of 5 years after the end of the project.

Where will the data be archived (= stored for the longer term)?

After publication, raw and processed sequencing data will be stored in the EGA database under access control by a local DAC. Additionally, data will also be preserved in the archival storage of the VSC. Developed software, algorithms and scripts will be made available and stored on a public repository such as GitHub.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

Costs for data archiving are 1080€ per year. These costs are covered by the funded FWO project.

7. Data sharing and reuse

Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

- No

Which data will be made available after the end of the project?

All pseudonymised raw and processed sequencing data along with clinical characteristics will be made available.

Where/how will the data be made available for reuse?

- In an Open Access repository

When will the data be made available?

- Upon publication of the research results

Who will be able to access the data and under what conditions?

Pseudonymised data will be made available through AGA under access control by the local DAC.

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

Data sharing will be performed using Belnet filesender which is free-to-use.

8. Responsibilities

Who will be responsible for data documentation & metadata?

The researcher will produce data documentation and metadata.

Who will be responsible for data storage & back up during the project?

The researcher will be responsible for data storage & back up during the project.

Who will be responsible for ensuring data preservation and reuse ?
The researcher will be responsible for data preservation and reuse.

Who bears the end responsibility for updating & implementing this DMP?
The researcher bears the end responsibility of updating & implementing this DMP.