DMP title

Project Name My plan (FWO DMP) - DMP title

Project Identifier S62388

Grant Title 11K6822N

Principal Investigator / Researcher Prof. Dr. Frederic Amant/Ilana Struys

Project Data Contact +32 478 73 44 60, ilana.struys@kuleuven.be

Description Though rare, cancer in pregnancy is becoming more frequent due to the trend to postpone pregnancies until a later age. In large-scale clinical studies, we showed that chemotherapy given during pregnancy has no adverse effect on the short-term health of the infants. Yet, these children appear to be more often small for their gestational age, have a lower verbal IQ and increased chance of wearing glasses at the age of 6. Long-term health effects of prenatal chemotherapy exposure are still unknown. Given the known genotoxic effects of chemotherapy and their ability to cross the placenta, it could be speculated that chemotherapy exerts DNA damage in the fetus. This could lead to potential deleterious effects later in life, such as cancer, fertility problems or other diseases. Based on preliminary data, pointing towards genetic DNA damage in newborns born to mothers treated with antenatal chemotherapy, we here propose, at the single cell level, an in-depth evaluation of the genome-wide presence of chromosomal alterations and mutations in cord blood of pregnant cancer patients treated with chemotherapy. Together with the ongoing clinical assessments, the results from this innovative study will represent an important milestone in the large research context of studying the safety profile of chemotherapy treatment in pregnant cancer patients.

Institution KU Leuven

1. General Information

Name applicant

Ilana Struys

FWO Project Number & Title

FWO Project Number: 11K6822N

Title: Single-cell evaluation of DNA damage in offspring after prenatal exposure to chemotherapy

Affiliation

- KU Leuven

2. Data description

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

- Generate new data
- Reuse existing 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).

In WP1, the following clinicopathological data of 30 pregnant breast cancer patients treated with chemotherapy during pregnancy and 30 pregnant breast cancer patients not treated with chemotherapy during pregnancy will be collected from Belgian and Dutch hospital databases and the International Network on Cancer, Infertility and Pregnancy (INCIP) registry (established and chaired by Prof. Dr. F. Amant): date of birth, date of primary diagnosis, date of partus, hospital were partus took place, tumor characteristics (stage, grade and molecular subtype), type of therapy during pregnancy, date of first chemotherapy cycle during pregnancy, date of last chemotherapy cycle before delivery, gestational age at start treatment and gestational age at birth. Additionally, the following information will be collected from 30 healthy pregnant women from Belgian hospital databases: date of partus, hospital were partus took place and gestational age at birth. Furthermore, a questionnaire will be filled in by the participants in which smoking, alcohol use, medication use, sickness and stress during pregnancy is questioned, as well as education, nationality, age, length, weight of the parents and whether there are previous children, sex of these children, way of delivery, season of delivery and length of pregnancy. Clinical data will be pseudonymized (coded) and compiled in an encrypted Excel database. Total

volume of this file is estimated to be around 3MB.

In WP2, the cytokinesis-block micronucleus assay will be used on a human monocytic cell line (THP-1 cells) treated with different schemes of chemotherapy to generate the following data: the frequency of micronucleated cells, the frequency of cells showing a nucleoplasmic bridge, the frequency of cells showing a nuclear bud, the degree of apoptotic/necrotic cells and the nuclear division index. These data will be compiled in an Excel file of which the total estimated volume is around 120kB. From the patients and control subjects listed above, the mononuclear cells isolated from cord blood and maternal blood are also subjected to the CBMN assay. These data are compiled in an encrypted Excel database, of which the total estimated volume is around 100kB. Furthermore, microscopic images (brightfield, 40x) are stored as .tif files and are estimated to have a final volume of 500GB.

In WP3, again the mononuclear cells isolated from cord blood and maternal blood from 10 subjects per group, will be used for genetic analyses. Using G&T-sequencing and clonal expansion of single hematopoietic stem cells, whole genome and transcriptome sequencing data is generated. Raw sequencing data will be stored as FASTQ and BAM files on the UZ Leuven cumulus drive linked to the UZ Leuven account of Ilana Struys. These newly generated sequencing data will be aligned with the publicly available reference genome hg38 (Bioconductor). The BAM files will be converted into VCF files, which are also stored on the UZ Leuven cumulus drive. We estimate that the total volume of these files will be around 10TB. The BAM and VCF files are analysed using R or other web-based platforms, to generate information on the presence, types and location of unbalanced structural variation, copy number variants, single nucleotide variants, insertion and deletions and telomere lengths. These data will be converted into graphs and figures which will be saved as .PDF, .JPG and .PNG and will contain on estimate a volume of 10MB.

In WP4, genotoxicity in mice blood and organs are analysed using the following assays (i) whole-formaline fixed pups are used for H&E stainings and immunohistochemistry stainings. Microscopic photographs of these slides will be saved on the web viewer Slide Score as high resolution scans of microscopy images. The results of the analyses will be compiled in a PDF, Excel and Powerpoint file, of which the total estimated volume will be around 3MB. Furthermore, microscopic images are stored as .JPG and will contain a total estimated volume of 5GB. DNA isolations from the blood and tissue samples will be used for a fragment analyses. The data generated from the fragment analyses consists of mean fragment sizes and DNA concentrations, which are graphically presented in .JPG files and summarized in an Excel file 5MB.

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: Ethics Committee Research UZ/KU Leuven reference S62388

Short description of the kind of personal data that will be used:

WP1. Clinicopathological data: date of birth, date of primary diagnosis, date of partus, hospital were partus took place, tumor characteristics (stage, grade and molecular subtype), type of therapy during pregnancy, date of first chemotherapy cycle during pregnancy, date of last chemotherapy cycle before delivery, gestational age at start treatment and gestational age at birth.

WP3. Genetic data: Whole genome sequencing and transcriptome data.

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

The project is a research study that was approved by the Ethics Committee Research UZ/KU Leuven (S62388).

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?

- No

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?

Data will be reused only under permission from the Ethics Committee Research UZ/KU Leuven.

A codebook will be created which will contain information on sampling methodology, variable-level detail, and all information necessary for a secondary analyst to use the data accurately and effectively. It will be stored in a separate, restricted-access folder on the UZ server.

Details about patients selection and study design are reported in the study protocol. The informed consent files will be fully documented as word files, as well as a blank copy of the informed consent form.

For the collection of biological samples, a laboratory manual is available, detailing the procedures of blood sampling, storage and shipment of samples to UZ Leuven.

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

Our metadata standards will be based on the Minimum Information for Biological and Biomedical Investigations (MIBBI) standards (source: https://fairsharing.org/MIBBI).

The MIBBI registries that we will base our template files on are:

- Minimal Information about a high troughput SEQencing Experiment (MINSEQE) (Source: https://fairsharing.org/FAIRsharing.a55z32)
- Minimum Information about Flow Cytometry (MIFlowCyt)

(Source: https://fairsharing.org/FAIRsharing.kcnjj2)

- Minimum Information Specification For In Situ Hybridization and Immunohistochemistry Experiments (MISFISHIE)

(Source: https://fairsharing.org/FAIRsharing.ht22t4)

A general overview of the most important parameters that we will retrieve are listed below.

MINSEQE:

Description of

- the biological system (e.g. organism, strain)
- Samples (e.g. cell type, type of treatment)
- Experimental variables being studied (e.g. length of treatment, timing of treatment, exposure duration)
- Sequence read data (FASTQ format)
- The final processed (or summary) data for the set of assays in the study
- General information about the experiment and sample-data relationships (e.g. summary of experiment, goals, associated publications)
- Essential experimental and data processing protocols (e.g. method of nucleic acid extraction and processing, library prep strategy, instrumentation)

MIFLowCyt:

Description of

- Experiment overview
  - Purpose
  - Keywords
  - Experiment variables
  - Organization (name, address)
  - Primary contact (name, email address)
  - Date
  - Quality control measures
- Flow sample/specimen details
  - Sample description
  - Sample source
  - Sample organism (taxonomy, phenotype, genotype, treatment)
  - Sample treatment
  - Fluorescence reagents (analyte, detector, reporter, reagent manufacturer, reagent catalogue number)
- Instrument details
  - Manufacturer
  - Model
  - Configuration and settings (optical filters, optical detectors)
- Data analysis details
  - List-mode data file
  - Compensation details
  - Data transformation details
  - Gating (data filtering) details (gate description, gate statistics, gate boundaries)

MISFISHIE:
Description of

- Experimental design
  - Experiment description (aims)
  - Assay types (IHC, CBMN)
  - Experiment design type (normal vs treated tissue/cells, type of tissue/cells)
  - Experimental factors (tested conditions)
  - Total number of assays performed in the experiment (replicates, reruns)
  - URL of any website or database accession numbers pertinent to the experiment
  - Contact information for communicating with the experimenters
- Biomaterials (specimens) and treatments (section or whole-mount preparations)
  - Origin of the biological specimen (organism species, sex, age, developmental stage, genotype, phenotype)
  - Physiological state of the individual(s) (normal vs. diseased)
  - Relevant exogenous factors (e.g. treatment)
  - Anatomic source of the tissue or cell sample
  - Provider of the specimen
  - Nature of the specimen (e.g. whole tissue, tissue sections, thickness of sections, whole cells)
  - Manner in which the specimens were prepared for the experiment (e.g. fixation with type and duration of fixation, frozen samples)
  - Protocols used
- Reporters (probes or antibodies)
  - Unambiguous genomic identification of each reporter (e.g. protein identifier)
  - Clone number of each antibody
  - Protocol(s) for how the reporters were designed and produced or the source from which they were obtained (e.g. company name, address, catalogue number, lot number)
  - Additional attributes of the reporter (type of primary antibody, immunoglobulin isotype, organism in which the antibody was generated)
- Staining
  - Number of detectable reporters in the hybridization or stain and specific details about the detection method (detection reagents, source of detection system)
  - Protocol used to produce the hybridization (method of mounting on slide, treatments, buffer, temperature, post-wash conditions, steps to decrease non-specific reaction products, antigen or gene product retrieval method)
  - Assay controls (same level of information as for sample itself)

- Imaging data
  - File format
  - Observation method
- Image characterizations
  - Ontology entries and references to the ontology, terms, accession numbers
  - Intensity scale
  - Staining intensity
  - Other optional annotations/characterizations of the structural unit (e.g. feature density, qualitative characteristics, spatial distribution)
  - Protocol for the characterization and information about the basic technique for characterizing the assays (e.g. how many observers, observations made from images or visually through the instrument)

5. Data storage and backup during the FWO project

Where will the data be stored?

Since we will be working with sensitive personal data that will be pseudonymized, clinicopathological data and genomic data will be stored at UZ Leuven's secure environment and the REDCap database. Microscopical image data (non persons data; CBMN, IHC, H&E) will also be stored to the UZ Leuven central servers.

How is backup of the data provided?

The data will be stored on UZ Leuven's central servers with automatic daily back-up procedures. Furthermore, the genomic data will be stored on the UZ Leuven cumulus drive.

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

Yes, data will be stored on the central server of UZ Leuven. Furthermore, the genomic data will be stored on the UZ Leuven cumulus drive which has no storage limit.

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

Costs for storage on the central servers of UZ Leuven are taken by the Department of Oncology. UZ cumulus drive is free for persons with an UZ Leuven account.

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

Data handling and statistical analysis will be done in a coded fashion by the investigator, with the subject identification code list only available to the local investigator (and research nurse if applicable) working in the local centre. The coded data will be stored at the UZ Leuven secure environment for private data.

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 and physical samples (i.e. DNA extractions, cryopreserved cells, microscopic slides …) will be retained for at least 5 years after the end of the project, according to current regulations and policies.

After research: clinicopathological patient data (encrypted) are transferred to the online REDCap database, for use in other projects. All other data (imaging, experimental, genomic) will be retained on the central servers of UZ Leuven.

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

Clinical data (encrypted) will be stored according to legal recommendations for 20 years on the UZ Leuven server and the data will also be transferred to the online REDCap database.

Genomic will be stored according to legal recommendations for 20 years on the UZ Leuven server.

Imaging data: H&E, IHC and CBMN colored slides will be stored at the laboratory for at least 5 years after the end of the research project.

Biological samples (i.e. cryopreserved cells, DNA extractions …) will be stored at the laboratory for at least 5 years after the end of the research project.

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

All data will remain on the central servers of UZ Leuven, which is free. No additional cost is foreseen for storage of biological samples.

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)?

- Yes. Specify:

Participants of the study can indicate whether they agree with potential usage of data for subsequent other research purposes in the informed consent form. Data can be shared upon request, with the exception of personal data, which will never be made available, under any circumstance.

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

Data can be shared upon request, with the exception of personal data, which will never be made available, under any circumstance.

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

- Other (specify):

Data can be shared only in coded form.

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?

Given that we deal with sensitive data, anyone can request data, but a Data Access Committee (UZ Leuven) will go over the request and decide whether data can be shared with a third user.

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

Costs of data sharing would be negotiated upon request.

8. Responsibilities

Who will be responsible for data documentation & metadata?

The PI (Prof. Dr. F. Amant) and senior members of his laboratory will be responsible for verifying data is accurate and records are up to date.

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

UZ Leuven IT

Who will be responsible for ensuring data preservation and reuse ?

The PI (Prof. Dr. F. Amant) is responsible for ensuring data preservation and reuse.

Who bears the end responsibility for updating & implementing this DMP?

The PI (Prof. Dr. F. Amant) bears the end responsibility of updating & implementing this DMP.