C1 - CURIOUS - DMP

DMP TITLE

ADMIN DETAILS

Project Name: C1 - CURIOUS - DMP - DMP title
Grant Title: C14/21/113
Principal Investigator / Researcher: Frédéric Amant
Project Data Contact: liesbeth.lenaerts@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 of the project lead (PI)
Prof. Dr. Frédéric Amant

Internal Funds Project number & title
C14/21/113

2. DATA DESCRIPTION

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

- Generate new data

- Reuse existing data

2.2. What data will you collect, generate or reuse? Describe the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a numbered list or table and per objective of the project.

1. 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.
2. In WP2, from the patients and control subjects listed above, the mononuclear cells isolated from cord blood and maternal blood will be subjected to the cytokinesis-block micronucleus (CBMN) assay. These data will be compiled in an encrypted Excel database, of which the total estimated volume is around 100kB. Furthermore, microscopic images (brightfield, 40x) will be stored as .tif files and are estimated to have a final volume of 500GB.
3. 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. For confirmatory analyses, we will resort to FISH or droplet digital PCR (ddPCR) on the viably frozen bulk cord blood samples. Raw sequencing data will be stored as FASTQ and BAM files on the UZ Leuven cumulus drive linked to the UZ Leuven account of one of the researchers linked to this project. 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.

3. ETHICAL AND LEGAL ISSUES

3.1. Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to the file in KU Leuven's Record of Processing Activities. 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
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.

3.2. 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 has been approved by the Ethics Committee Research UZ/KU Leuven (S62388).

3.3. Does your research 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

3.4. 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 regarding reuse and sharing are in place?
No

4. DOCUMENTATION AND METADATA

4.1. What documentation will be provided to enable understanding and 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.

4.2. Will a metadata standard be used? If so, describe in detail which standard will be used. If not, 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)
- Minimum Information Specification For In Situ Hybridization and Immunohistochemistry Experiments (MISFISHIE) (Source: https://fairsharing.org/FAIRsharing.kcnjj2)
- Minimal Information about a high troughput SEQencing Experiment (MINSEQE) (Source: https://fairsharing.org/FAIRsharing.a55z32)
- Minimum Information about Flow Cytometry (MIFlowCyt)
- Minimum Information Specification For In Situ Hybridization and Immunohistochemistry Experiments (MISFISHIE) (Source: https://fairsharing.org/FAIRsharing.ht22t4)

5. DATA STORAGE AND BACKUP DURING THE PROJECT

5.1. 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, linked to the cancer-in-pregnancy project. Microscopical image data (non persons data) will also be stored to the UZ Leuven central servers.

5.2. How will the data be backed up?
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.

5.3. 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, 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.

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

Costs for storage of research data on the central servers of UZ Leuven are taken by the Department of Oncology. Cost for storing genomic data are estimated to be 5000 euro and are covered by the internal C1 grant.

5.5. 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 END OF THE PROJECT

6.1. Which data will be retained for the expected 10 year period after the end of the project? If only a selection of the data can/will be preserved, clearly state why this is the case (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 10 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.

6.2. Where will these data be archived (= stored for the long 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: colored slides will be stored at the laboratory for at least 10 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 10 years after the end of the research project.

6.3. What are the expected costs for data preservation during these 10 years? How will the costs be covered?
All digital data will remain on the central servers of UZ Leuven. Biological samples will be stored in lab storage containers (closet, freezers, liquid nitrogen). Part of the costs are covered by the internal C1 grant.

7. DATA SHARING AND RE-USE

7.1. 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 or because of IP potential)?

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.

7.2. 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.

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

- Other (specify):

Data can be shared only in coded form.

7.4. When will the data be made available?

- Upon publication of the research results

7.5. Who will be able to access the data and under what conditions?
Given that we deal with sensitive data, other researchers 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.

7.6. What are the expected costs for data sharing? How will these costs be covered?
Costs of data sharing would be negotiated upon request.

8. RESPONSIBILITIES

8.1. Who will be responsible for the 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.

8.2. Who will be responsible for data storage & back up during the project?
UZ Leuven/KU Leuven IT

8.3. Who will be responsible for ensuring data preservation and sharing?
The PI (Prof. Dr. F. Amant) is responsible for ensuring data preservation and reuse.

8.4. Who bears the end responsibility for updating & implementing this DMP?
The end responsibility for updating and implementing the DMP is with the supervisor (promotor).