DMP title

Project Name My plan (Internal Funds DMP) - Disruptive endometriosis: unraveling its pathogenesis and concomitant infertility with cutting-edge organoid models and tissue mimics - DMP title

Project Identifier C14/21/1169

Grant Title C14/21/1169 (ZKE0469)

Principal Investigator / Researcher Hugo Vankelecom / Nina Maenhoudt

Project Data Contact nina.maenhoudt@kuleuven.be; 016373656

Description The impressive capacity of cutting-edge 3D organoid models to reproduce organ biology and pathology is revolutionizing human research. Endometriosis (growth of endometrial tissue outside the uterus) is a highly disruptive gynecological disorder. Little is known on its pathogenesis and concomitant infertility. We succeeded in establishing organoid models from human endometrium and endometriosis lesions, faithfully reproducing the tissue. Here, we employ these (as well as further advanced) organoid systems to delve into the mechanisms underlying endometriosis pathogenesis by applying single-cell transcriptomics and unraveling two important pathways, further tested using organoids as drug screening platform. To investigate the endometriosis-linked infertility, we scrutinize organoid-based tissue mimics derived from endometrium of endometriosis patients and compare embryo-endometrium interaction with fertile women using an innovative in vitro implantation model.

Institution KU Leuven

1. General Information

Name of the project lead (PI)

Hugo Vankelecom

Internal Funds Project number & title

C14/21/1169

Disruptive endometriosis: unraveling its pathogenesis and concomitant infertility with cutting-edge organoid models and tissue mimics

2. Data description

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

- Generate new 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.

Physical samples:

- Endometrial biopsies of healthy women and endometriosis patients. Obtained via UZ Leuven, Gynecology and Obstetrics (ethical dossier: S59006).
- Cryopreserved biopsies and derived organoids (in liquid nitrogen). Organoids are obtained as published in Boretto et al. 2017 (PMID: 28442471).
- Paraformaldehyde (PFA)-fixed endometrial biopsies and derived organoids (stored in designated closets in the lab).
- Microtome sections obtained from PFA-fixed samples (stored at 4°C).
- RNA from biopsies and derived organoids (at -80°C).
- cDNA and/or gDNA from biopsies and derived organoids (at -20°C).
- Lab log books.

Digital data:

- PCR results: gel electrophoresis (gel images) obtained via Image Lab software: .tif files (150MB)
- Light, epifluorescence and confocal images from sections of organoids and primary tissue (biopsies): .lif, .lsm and .tiff files. (25GB)
- RNA/DNA concentration/quality (obtained via measurement with Nanodrop): excel file (.xlsx) manually created. (<1MB)
- qPCR data/graphs (created via QuantStudio Real Time pcr software): .eds, .xlsx files. (0.5-1GB)

- MSD readouts (obtained via measurement with MSD reader): .txt, .xlsx files. (<1MB)
- Drug screening readouts (obtained via measurement with fluorescent plate reader): .xlsx files. (<1MB)
- Ca2+ microfluorimetry: nd2 files (Nikon software); post-analysis .tiff, .jpeg, or .png images
- Bulk and single-cell RNA-sequencing (scRNA-seq) data of biopsies and derived organoids: bam and fastq files. (200GB uncompressed)
- Statistic/bioinformatic software and output: R; Seurat; pyScenic, … (5-10 GB)
- Experimental analysis data and manuscripts: docx, .pptx, .xlsx files. (10-100MB)
- Biopsies and organoid biobank database and clinical data: .xlsx files. (10-100MB)

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.

Relevant clinical data (patient age, pathology, medication, fertility) will be retrieved from ‘UZ Leuven clinical work station’, done by the applicant’s UZ collaborators’ group (profs. drs. Christel Meuleman, Carla Tomassetti, Karen Peeraer). Ead number and inclusion criteria (e.g. medical history, medication use, physical examination parameters, laboratory parameters) will be kept in a separate encrypted database (access authorization for the applicant’s UZ collaborators’ group, with audit trail). Patients' samples are assigned a unique identifier, so that if any patient-specific scientific data will become relevant to improve clinical care, this could a posteriori be provided for. Data are collected under ethical approval by The Ethics Committee Research UZ/KU Leuven (EC Research) (S59006).

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).

Permission for the described research using healthy endometrium and endometriosis biopsies and derived organoids has been obtained from The Ethics Committee Research UZ/KU Leuven (EC Research), i.e. dossier S59006.

Permission for the study with human embryos was also obtained (in collaboration with the group of Prof. Vriens, KU Leuven and Karen Peeraer, UZ Leuven), i.e. S62765.

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?

We anticipate that some of the research results (e.g. new drug targets for treating endometriosis and its linked infertility) may have potential for tech transfer and valorisation. IP restrictions will be claimed, and patents filed. All this will be evaluated and executed case by case through Leuven Research and Development (LRD).

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?

- Daily lab activities are recorded in detail in the lab log book.
- For documentation of microscopy images (of organoid cultures), the following information will be noted down: date, experimental condition, passage of organoid culture, number of days in culture, imaging magnification used. Images are saved on the shared drive of the lab (KU Leuven) and a copy is stored on the KU Leuven OneDrive in a designated folder of the particular experiment type. Both drives run automatic backups. Within the experiment folder, additional folders are labeled in a clearly structured manner (according to different experimental conditions or different timepoints within the experiment). The setup of an experiment is written down in the lab log book. A metadata file, generated by the microscope program, is saved automatically together with the image.
- For RNA and cDNA concentration and quality measurements using Nanodrop: 260/230 and 260/280 ratios (quality measure) and concentrations are written down in the lab log book and then transferred to an excel file where all RNA/cDNA measurements are stored. Date of

measurement together with name of the sample and experimentator is included.
- For qPCR data: Excel file containing sample setup, raw data, results and melt curve data are labeled with the name: "date, qPCR_experiment name". The qPCR data is saved in a "qPCR folder" within the folder of the specific experiment, together with the template of the particular qPCR setup. Name of the template file: "date, qPCR_experiment name_layout". Graphs from the data are made using Graphpad Prism (.pzfx file). File is named: "date, Graphs_expriment name", and saved in the same folder.
- Methodology and protocols for RNA extraction, cDNA preparation, immunohistochemistry, organoid culture, medium preparation, ... are all included in the lab log book and on the shared lab drive and OneDrive. In the table of contents of the lab log book, the page number of each protocol included in the book can be found.
- For bulk and scRNA-seq data: raw data files contain dates + sample names. All additional files contain date + name of experiment and sample. Databases in MS Excel/OriginPro are labelled according to the experiment.
- MSD data: file names contain date of experiment + sample name. Folder structure is organized according to type of experiment.

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.

At present, metadata standards are not (yet) implemented in the research group. The microscope program automatically generates a metadata file for every image which is saved together with the image. In general, to make the data easy searchable, a personal folder on the shared drive of the lab and OneDrive is constructed and is further subdivided a clearly structured manner (e.g. specific folders for different experiments). In the lab log book, a description of every experiment can be found, including all the experimental conditions.

5. Data storage and backup during the project

5.1. Where will the data be stored?

- Digital data are stored on the shared drive of the lab (KU Leuven; with automatic backup) and a copy is stored on the KU Leuven OneDrive (also with automatic backup).
- Physical samples and biospecimens are stored in the restricted-access cool room/fridges (4°C, for PFA-fixed patient biopsies) or freezers (-20°C for cDNA and gDNA or -80°C for RNA), or liquid nitrogen containers (for cryopreserved biopsies and organoids) of the research group. All data stored in the lab log books remain available in the host lab even after departure of the applicants.

5.2. How will the data be backed up?

Digital data are stored on the shared server of the lab (KU Leuven) and on the KU Leuven OneDrive, which both run automatic backups. Moreover, the large bulk/scRNA-seq data are stored on the storage space of the Flemish Super Computer (VSC) in our dedicated account. After completing the study, all data are uploaded to a repository to be determined (e.g. archive space of the VSC) and placed under embargo for five years. Then, the data that are not under ethical restriction, are released.

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. The KU Leuven OneDrive provides 2 TB of storage which will be sufficient for the project. Moreover, the large bulk/scRNA-seq data are stored on the storage space of the Flemish Super Computer (VSC) in our dedicated account.

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

The costs of the 2TB of storage on KU Leuven OneDrive are covered by KU Leuven. The storage space of the Flemish Super Computer (VSC) costs 250 €/year, covered by the applicant’s research funds.

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

Access to the shared drive of the lab and KU Leuven OneDrive are secured by login with personal u-number and password. An extra layer of protection is implied by the multifactor Authenticator

app of the KU Leuven. Moreover, each applicant's computer is only accessible through a password. Access to patient data in the 'UZ Leuven clinical work station' is restricted to the applicant's UZ collaborators' group of profs. drs. Christel Meuleman, Carla Tomassetti, Karen Peeraer.

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 (digital and physical) will be retained for at least 10 years after the end of the project. After completing the study, the large bulk/scRNA-seq data will be uploaded to a repository to be determined (e.g. archive space of the VSC) and placed under embargo for 10 years.

RNA degenerates within a period of 5 years making the samples unusable after that period. Therefore, physical samples containing RNA are only kept for 5 years after the end of the project.

6.2. Where will these data be archived (= stored for the long term)?

After the end of the project, all digital data are stored on the shared server of the lab (KU Leuven) and on the KU Leuven OneDrive, which both run automatic backups. The large bulk/scRNA-seq data will be uploaded to a repository to be determined (e.g. archive space of the VSC) and placed under embargo for 10 years.

6.3. What are the expected costs for data preservation during these 10 years? How will the costs be covered?

As long as the data do not exceed the 2TB of storage space of our KU Leuven OneDrive,

no additional costs for data preservation are expected. The costs for storage on VSC (to be determined) will be covered by other research grants of the applicants.

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)?

No.

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

The obtained data in the project will be made available through publications and PhD theses. scRNA-seq data are made available on ArrayExpress after each respective publication.

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

- Other (specify):

Data can be requested for reuse by others after accomplishing a data transfer agreement (DTA). Public availability, after publishing the data, will also depend on the journal's policy (postpublication data repository). scRNA-seq data are made publicly available on ArrayExpress after each respective publication.

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?

Public availability, after publishing the data, will also depend on the journal's policy (postpublication data repository). scRNA-seq data are made publicly available on ArrayExpress after each respective publication.

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

There are no expected costs for data sharing.

8. Responsibilities

8.1. Who will be responsible for the data documentation & metadata?

Nina Maenhoudt

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

Nina Maenhoudt

8.3. Who will be responsible for ensuring data preservation and sharing?

Prof. dr. Hugo Vankelecom

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) prof. dr. Hugo Vankelecom