FONDS VOOR WETENSCHAPPELIJK ONDERZOEK - RESEARCH FOUNDATION FLANDERS (FWO): FWO DMP (FLEMISH STANDARD DMP) - APPLICATION DMP

QUESTIONNAIRE

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

My research will generate new data; these will be digital microscopy images [immunofluorescence images and embryo images (TIFF)], sequencing data (bisulfite, RNA and CUT&RUN sequencing data including fastq, SAM and bigwig files), metabolomics data (.csv), new scripts (written in bash and R) and quantitative data coming from the analysis of qPCR, sequencing and metabolomics data. The latter two will also lead to the visualization of these data using appropriate visual representations including tables, heatmaps, bar graphs and plots, among others. In addition, presentation, word and PDF files will be generated, prior to oral presentations and manuscript submissions. No personal data will be used.

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

Experimental data, including experimental protocols, microscopy images, sequencing, metabolomics and qPCR data will be saved on the KU Leuven large volume storage. This is intended for the long-term storage of large volumes of research data with minimum cost and is paid by my PI, Prof. Kian Koh. A record of the experimental details, protocols, samples etc. along with dates when experiments were carried out, will be recorded in a hard copy laboratory notebook that will remain in the lab for at least 5 years after the end of my research. Physical data (cell lines and oligonucleotides) will be labelled and stored at -20 and -80^{o}C, respectively, and the records will be stored in a database in KU Leuven large volume storage.

What's the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

All data will be stored for a minimum of 5 years.

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

Data from mouse embryos (images and sequencing data) will be collected. These data will be stored in the KU Leuven large volume storage, which provides high security and is suitable for the storage of confidential data.

FONDS VOOR WETENSCHAPPELIJK ONDERZOEK - RESEARCH FOUNDATION FLANDERS (FWO): FWO DMP (FLEMISH STANDARD DMP) - FWO DMP (FLEMISH STANDARD DMP)

1. 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.

Data Type	Description	New/Re-	Digital/	File extension	Maximum

		used	Physical		Volume
Digital Images	Images derived from microscope, stereoscope and scanned images	New	Digital	.TIFF	8 GB
Sequencing data	Sequencing raw data from bisulfite, RNA, CUT&RUN and whole genome sequencing	New	Digital	Sequence fastq files: fq.gz; alignment: sam, bam; coverage: bed, .bg, .bw, bigwig; Tabulation: .csv, .txt, .xls	10 TB
Metabolomics data	Pre-processed metabolomics data	New	Digital	Tabulated: .csv, .txt, .xls	2 GB
Scripts	Scripts written in bash and R	New	Digital	.txt, ipynb	100 MB
Quantitative data	Data from downstream analysis of qPCR, sequencing and metabolomics data and visual representations including tables, heatmaps, bar charts, graphs and plots	New	Digital	Tabulated:.csv, .xls Statistical analysis: pzfx Visualization: .TIFF, .PDF, . ppt, .ai	4 GB
Cell Lines	Cell lines generated using CRISPR	New	Physical	-	2-3 cell lines
Oligonucleotides	Primers, sgRNAs, ssODNs	New	Physical	-	~10 oligos

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

Yes, animal data. I will perform single-cell methylation and transcriptome sequencing (scNMT-seq) in E6.5 mouse embryos. scNMT-seq involves RNA sequencing as well as treatment of DNA with a methyltransferase and sodium bisulfite, to determine methylation and chromatin accessibility data. These data will be stored in the KU Leuven large volume storage, which provides high security and is suitable for the storage of confidential data.

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

No

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, data from my research could result in a better understanding of lineage bifurcation and epiblast/trophoblast epigenetics. These are based on the sequencing data (sequencing raw data from bisulfite, RNA, CUT&RUN and whole genome sequencing) and the quantitative data (data from downstream analysis of qPCR, sequencing and metabolomics data and visual representations including tables, heatmaps, bar charts, graphs and plots)

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/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.

- No

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

2. 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).

Experiments are recorded in a hard copy laboratory notebook, according to work package (WP) and date. Similarly, digital data are stored in the KU Leuven large volume storage, in different folders according to WP, name of experiment and date. Once a project is completed, an index file containing all the project information is compiled that lists all the WPs and the final disc location paths in the KU Leuven large data storage drive. Concerning physical data, cell lines are stored in the laboratory's liquid nitrogen tank and an electronic inventory (.xls database) containing the cell line box/grid position and cell line characteristics (description, passage number and freezing date) is maintained. Similarly, oligonucleotides are stored in -20^{o}C, while an .xls database contains the box/grid position and the oligonucleotide description, gene name and user initials.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

Next generation sequencing data will be deposited in public repositories, i.e., GEO and SRA. These repositories have specific metadata requirements that need to be filled upon data submission. These metadata include how samples were collected/prepared, technical details on how sequencing was performed etc., so that data can be reused in future analyses. In the case of digital images, including microscopy images, metadata are embedded into the image file and contain details including the instrument used to take the image, its settings, and image details including pixels and magnification, among others.

3. DATA STORAGE & BACK-UP DURING THE RESEARCH PROJECT

Where will the data be stored?

The data will be stored in the KU Leuven large storage drive.

How will the data be backed up?

Data stored in the KU Leuven large storage drive are stored in a KU Leuven datacenter and backed up regularly. The system is proactively monitored 24 hours per day, 7 days a week (24/7) and there is a mirror in a second KU Leuven datacenter. Technical equipment is constantly monitored with an emergency on-call service. Concerning physical data, cell lines are stored in the laboratory's liquid nitrogen tank which is connected to an alarm system in case of any temperature rise; this system is monitored 24/7 through the use of a laboratory phone. The same applies to -80 and 20^{o}C freezers.

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 large storage drive can easily scale into several dozen petabytes, essentially providing unlimited strorage. New storage can be purchased in blocks of 5 TB.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The large storage drive has high security and is suitable for confidential and strictly confidential data. Different rights can be given to different users, allowing access to data only to specific individuals or groups.

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

The cost of the KU Leuven large volume storage is €104,42 per TB per year. This cost is covered by the laboratory funds provided by Prof. Kian Koh.

4. DATA PRESERVATION AFTER THE END OF THE RESEARCH PROJECT

Which data will be retained for at least five 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 stored in the KU Leuven large storage drive will be retained for at least five years. The hard copy laboratory notebook will also be stored for at least 5 years in the laboratory, after the end of my research.

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

Data will be stored in the KU Leuven large storage drive. Published sequencing, metabolomics, image and quantitative data will ultimately be deposited into repositories and used in publications.

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

The costs are covered by the laboratory funds provided by Prof. Kian Koh. The expected cost of the KU Leuven large volume storage for my data (~10 TB) will be € 417.68 for the 2-year period of my FWO fellowship (€ 208,84 per year).

5. DATA SHARING AND REUSE

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

Sequencing and metabolomics data will be deposited in public open access repositories, after publication in a scientific journal. Also, significant images and quantitative data will be published in a scientific journal.

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

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 in the comment section per dataset or data type where appropriate.

- No

Where will the data be made available? If already known, please provide a repository per dataset or data type.

The sequencing data will be made available in repositories like the GEO and SRA. Significant images and quantitative data will be published in a scientific journal.

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.

Sequencing data will be publicly accessible, downloadable and usable, according to the repository rules. Journal articles will be published under an open access license that will allow the data to be accessed, redistributed, translated and reused in other works.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

Sequencing data can be stored indefinitely in the GEO and SRA repositories and given an accession number. Significant images and quantitative data will be published in a scientific journal, and will be given a PMID and DOI.

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

No costs are associated with public repositories for sequencing data deposition. For publication of journal articles, the costs depend on the journal. These costs will be covered using the FWO bench fees or other laboratory funds.

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

I will

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

I will, in collaboration with my PI, Prof. Kian Koh.

Who will manage data preservation and sharing?

I will, in collaboration with my PI, Prof. Kian Koh.

Who will update and implement this DMP?

I will

FONDS VOOR WETENSCHAPPELIJK ONDERZOEK - RESEARCH FOUNDATION FLANDERS (FWO): FWO DMP (FLEMISH STANDARD DMP) - GDPR

GDPR

Have you registered personal data processing activities for this project?

- Not applicable

FONDS VOOR WETENSCHAPPELIJK ONDERZOEK - RESEARCH FOUNDATION FLANDERS (FWO): FWO DMP (FLEMISH STANDARD DMP) - DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- Not applicable