FWO DMP Template

Project supervisors (from application round 2018 onwards) and fellows (from application round 2020 onwards) will, upon being awarded their project or fellowship, be invited to develop their answers to the data management related questions into a DMP. The FWO expects a **completed DMP no later than 6 months after the official start date** of the project or fellowship. The DMP should not be submitted to FWO but to the research co-ordination office of the host institute; FWO may request the DMP in a random check.

At the end of the project, the **final version of the DMP** has to be added to the final report of the project; this should be submitted to FWO by the supervisor-spokesperson through FWO's e-portal. This DMP may of course have been updated since its first version. The DMP is an element in the final evaluation of the project by the relevant expert panel. Both the DMP submitted within the first 6 months after the start date and the final DMP may use this template.

1. General Information	
Name applicant	Marta Ciwinska
FWO Project Number & Title	Project title: Unravelling the effect of tissue morphology on cell identity in complex branched structures. Project number: **11L7222N**
Affiliation	☒ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☒ Other: KU Leuven Department of Oncology VIB–KU Leuven Center for Cancer Biology
2. Data description	

Will you generate/collect new data and/or make use of existing data?	☒ Generate new data ☐ Reuse existing data
Describe the origin, type and format of the data (per dataset) and its (estimated) volume *If you **reuse** existing data, specify the **source** of these data.* *Distinguish data **types** (the kind of content) from data **formats** (the technical format).*	1 **WP1.1-1.3 Mammary gland organoid culture and modulation of organoid geometry:** Mammary gland organoids will be derived from two different genetically modified mouse models. Healthy organoids will be grown from well-established R26-Confetti reporter mouse crossed with Rosa26-CreERT2 mice and organoids in which the cell identity will be changed, from novel oncogenic models Red2Onco overexpressing one of oncogenes (PIK3CA-H1047R, KRAS-G12D, or Notch1) upon induction. The modulation of geometry will be done mechanically by different matrix composition and chemically by growth factors. The data regarding *in vitro* organoid culture being important timepoints as seeding, induction dates, age and estrous cycle of the mice used to obtain mammary glands, intermediate media replacement/replating time points, staining information, will be stored in **Excel file (.xlsx)** organized per each experiment. The raw data regarding imaging will be stored in the **Leica microscopy original format (.lif)** and representative images will be stored in **.tiff** format. The data analysis like branching and clonal dynamics analysis of imaged organoids will be stored in **Excel (.xlsx).** The presentation of data in the graph form will be stored in **GraphPad Prism file (.pzfx)** and final versions of graphs will be exported to **.tiff** format. In order to follow in real time the clonal dynamics of the oncogenic (RFP) clones and wild-type (CFP, YFP, GFP) clones, the 3D confocal microscopy over time (4D analysis) will be performed. The raw data from this timelapses will be stored in **Leica microscopy original format (.lif)**. The representative images will be stored in **.tiff** format and representative timelapse movies in **.avi** format. All data analysis like clone sizes, distances between clones and cell velocity will be performed in Las X, ImageJ and stored in **Excel (.xlsx)**. **Estimated volume: 1 TB** 2 **WP1.4 Isolation of clones by micro dissection and single cell sequencing to determine correlation between cell identity and tissue geometry:** At each step of the organoid growth (sphere, elongated, branched, hyperbranched) from the two models (*R26CreERT2;R26-Confetti* and *R2-Onco;R26-Confetti)* parts of the organoids will **micro-dissect** and the identity

of the cells will be determined using sc-mRNA sequencing. The single cell sequencing data will be analysed in Python and stored in **.fastq** format.

Estimated volume: 50 GB

3 **WP2 Studying the impact of tissue geometry on cellular identity in the complex branched structure of the mammary gland:** In order to assess the effect of tissue geometry on cell identity as well as tumorigenic potential of mutant clones in the context of the complex branched organ being mammary gland we will use healthy reporter mouse strain Rosa26-Confetti 2) mouse strain and ***Red2-Onco*** overexpressing inducible oncogene (**PIK3CA-H1047R, KRAS-G12D, Notch1**) together with reporter fluorescent proteins. Mice will be induced at pubertal stage to answer the question how branching affects the identity of cells in differently curved parts of the organ and in adulthood where we are able to study how constant local tissue remodelling (mainly side branching) driven by successive rounds of the oestrous cycle combined with oncogenic expression will impact tumour development. All data related to timepoints of induction and sacrifice will be stored in **Excel (.xlsx)**. At different timepoints after lineage tracing initiation I will use **3D whole-gland *ex vivo* imaging** to compare clone size, clonal morphology in relation to location and local geometry of the mammary gland. Raw imaging data will be stored in **Leica microscopy original format (.lif)** and representative images will be exported to **.tiff** format. Clonal analysis will be performed using **LasX** and **ImageJ,** quantified data will be stored in **Excel (.xlsx) file** divided by timepoints and oncogene. Data will be tested for statistical significance and represented in the graph format using **GraphPad Prism** and stored in **(.pzfx) format.** Final graph versions will be stored in **.tiff** format. After determining the optimal time point for clonal outgrowth, the real-time behaviour of labelled cells will be followed by implantation of an optical mammary imaging window and time-lapse imaging. Time-lapse imaging data will be stored in **Leica microscopy original format (.lif)**, the movies representing the cellular dynamics will be exported in **.avi** format. Time-lapse data will be analyzed using **LasX** and **ImageJ** and stored in **Excel (.xlsx)**.For both puberty and adulthood, we will micro-dissect the clones from ducts, tips, and TEBs (only in puberty), and perform **sc-mRNA sequencing** right after the imaging session, which will allow

	me to correlate the spatial information and cellular behaviour to the identity of the clones. Sc-mRNA sequencing data will be analysed using Python and stored in **.fastq** format. Data will be tested for statistical significance and represented in the graph format using **GraphPad Prism** and stored in **(.pzfx).** **Estimated volume: 45TB (15 TB per oncogene)** 4 **WP3.1 : In vivo manipulation of cellular identity or tissue morphology in the branched mammary gland. Manipulation of the tissue architecture *in vivo:*** As we suspect that tissue curvature and geometry indeed instruct cellular identity, therefore we will change the geometry *in vivo* chemically by locally implanting beads coated with factors that influence branching dynamics, giving local control of the inhibition or induction of the branching. In order to determine whether induced remodelling influence clonal dynamics, the *ex vivo* whole gland imaging and intravital imaging. The imaging data will be stored in **.lif** format, representative images exported in **.tiff** and representative movies in **.avi** format. Additionally, to confirm that molecular identity of the cells changed, we will micro-dissect the clones and perform **sc-mRNA sequencing. Sc-mRNA data** will be analysed using Python and stored in **.fastq format.** Data will be presented in the graph format using Graphpad **(.pzfx)** and final graph versions exported in **.tiff** format. **Estimated volume: 15TB (5TB per oncogene)** 5 **WP3. Manipulation of cellular identity by overexpression of oncogenes:** Finally, I will use our Red2-Onco models crossed with newly developed light inducible Cre mouse (PA-Cre) model which allows us to induce recombination at any desired location within the mammary gland, to induce recombination at desired part of the mammary gland tree. By combining this inducible Cre model with IVM, I can exactly pinpoint the effect of the local tissue architecture on tumour initiation dynamics, and determine whether the curvature of the tissue is impacting tumour initiation. The IVM data will be stored in **.lif** format and representative images will be exported in **.tiff** format. All quantitative data will be stored in **Excel (.xlsx)**. Next, I will micro-

	dissect clones from ducts, tips and interesting malignant clones that are initiating tumor growth, and perform single-cell mRNA sequencing. Sc-mRNA sequencing data will be analyzed in Python and stored in **.fastq** format. **Estimated volume: 2TB** 6 **Protocols, presentations, metadata, and manuscripts:** Experimental protocols, details related to data processing and methods of data analysis will be stored in **Word (.docx)** and **Excel (.xlsx)**. Presentations and posters will be stored in **Power Point (.pptx)** and final versions will be exported to **PDF format** to preserve their integrity. All the figures needed for the manuscripts will be produced in Adobe Illustrator and exported to **.tiff** format. **Estimated volume: <2GB**

7 Ethical and legal issues	
Will you use personal data? If so, shortly describe the kind of personal data you will use AND add the reference to your file in your host institution's privacy register. *In case your host institution does not (yet) have a privacy register, a reference is not yet required of course; please add the reference once the privacy register is in place in your host institution.*	☐ Yes ☒ No If yes: - Privacy Registry Reference: - Short description of the kind of personal data that will be used:

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 ☐ No If yes: - Reference to ethical committee approval: Animal experiments will be performed as part of this project. All animal experiments will be performed in the laboratory of Prof. Colinda Scheele and have been approved by the Ethical committee for Animal Experimentation (ECD) at KU Leuven, and are outlined in ECD project **P155/2020.**
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 ☒ No If yes, please comment:
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?	☐ Yes ☒ No If yes, please comment:

8 Documentation and metadata	
What documentation will be provided to enable understanding and reuse of the data collected/generated in this project?	All protocols and necessary details related to data collection as well as methods of analysis will be recorded in licenced E-lab journal containing **Word (.docx)** and **Excel (.xlsx) files** stored at Ku Leuven Large Volume Storage drive, which is backed up by Ku Leuven IT services. The raw files will be segregated in separate folders according to Work Packages and experiments within the Work Packages itself. The names of files will include date of the experiment, number of experiment, type of experiment and different experimental conditions.

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 ☒ No If no, please specify: Each folder containing a separate experiment will also contain an information in the form of Word (.docx) and Excel (.xlsx) file explaining data methods and all relevant metadata being experimental conditions, genetic models used, all sample identification numbers and computational analysis pipelines. The files with detailed explanation stored at Large Volume Storage drivee will ensure the reusability of the data and the reproducibility of any further data generation.

9 Data storage & backup during the FWO project	
Where will the data be stored?	The data will be temporarily stored on the expansion drives and a copy of the data will be immediately uploaded to the KU Leuven Large Volume Storage space (L-Drive) for long-term preservation and backup.
How will the data be backed up?	Data stored on the KU Leuven L-Drive is managed, maintained, and backed up by KU Leuven IT services. Specifically, mirror copies of the stored data are made immediately upon upload, for safety backup purposes.
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 ☐ No If no, please specify: Yes, the data will be stored in the Ku Leuven Large Volume Storage drive as well as in the Ku Leuven Archive drive (K drive). This ensures enough of storage space during the project as well as afterwards for preserving the data. Ku Leuven drives also provide automatic generation of a backup.

What are the expected costs for data storage and backup during the project? How will these costs be covered? *Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of* ***the allocated project budget*** *to be used to cover the cost incurred.*	As data will be stored in the Ku Leuven drives, the estimated cost for the Ku Leuven Large Volume Storage drive per 5TB per year is 569,2 euro. Total estimated size of the generated data within this project is 63TB. During the four years of the duration of the project, the cost is estimated on 28 677 euro. Additionally, the cost and capacity include the performance of mirror copies of the stored data, for safety backup purposes. The cost of data storage will be covered by the budget of the project lead - Prof. Colinda Scheele.
Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?	In order to preserve my data, I will use a licensed E-lab journal to record all my results and protocols in a findable way, as well as the backup files regarding the results will be stored in the L-drive. The raw files acquired through imaging, sequencing and analysis will be saved on the institutional server being Ku Leuven Large Volume Storage drive (backed up every 12h), being at the same time accessible and reusable by staff members granted server access.

10 Data preservation after the end of the FWO project FWO expects that data generated during the project are retained for a period of minimally 5 years after the end of the project, in as far as legal and contractual agreements allow.	
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, ...).	The raw imaging data (.lif files) and analysis (.tiff and .xlsx) will be preserved for a least 5 years on the Ku Leuven Archive drive. The processed images (.tiff) will be stored on two external hard drives, which will be kept for safety at two separate places. Generated Red2Onco organoids will be preserved in the -80 °C. Whole organ staining and multiplexed staining as well as RNA and DNA samples cannot be stored for 5 years due to sample instability, but the data (.lif and .fastq files and processed analysis) obtained from these samples will be stored on the data server.
Where will these data be archived (= stored for the long term)?	The data will be stored at the Ku Leuven L-drive and K-drive (Archive drive) and already processed images on two external drives.

What are the expected costs for data preservation during these 5 years? How will the costs be covered? *Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of* ***the allocated project budget*** *to be used to cover the cost incurred.*	The data during preservation period will be stored on Ku Leuven Archive drive and already processed images on two external drives. The cost of the external drives as well as the Archive drive will be covered by the budget of the project lead Prof. Scheele. The cost of Archive drive per year is estimated on 5,69 euro per 100 GB, the cost for the 5 years for 63TB will be 17 923 euro accordingly.

11 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 ☒ No There are currently no legal restriction to share the data in this project. There is no 3rd party involved. If yes, please specify:
Which data will be made available after the end of the project?	The key findings as well as data interpretation will be accessible through publication of journal articles in established, peer-reviewed (non-predatory) academic journals. The raw data will be made available through upload to well-established open-access data repositories.
Where/how will the data be made available for reuse?	☒ In an Open Access repository ☐ In a restricted access repository ☒ Upon request by mail ☐ Other (specify):
When will the data be made available?	Published data will be made available at the time of publication in case of open access or upon request for other publications. Additional, not-published data will be made available for external users upon request during the post-project trajectory.

Who will be able to access the data and under what conditions?	The raw data as well as unpublished protocols, additional experiments and imaging data will be in principle only accessible to members of Prof. Scheele lab members. The staff and students within VIB–KU Leuven Center for Cancer Biology, as well as within the KU Leuven Department of Oncology will be able to access these data upon reasonable request. Any other user can also request data upon reasonable request.
What are the expected costs for data sharing? How will these costs be covered? *Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of* ***the allocated project budget*** *to be used to cover the cost incurred.*	There is no expected costs related to data sharing.

12 Responsibilities	
Who will be responsible for the data documentation & metadata?	The applicant (Marta Ciwinska) and the project lead (Prof. Colinda Scheele) will share the responsibility for data documentation and metadata generation/preservation.
Who will be responsible for data storage & back up during the project?	The applicant (Marta Ciwinska) will be primarily responsible for data collection, generation and storage. The applicant will also take responsibility for documentation and uploading the data onto the L-Drive storage space. The KU Leuven IT department will be responsible for maintenance and back up of the L-Drive data storage space.
Who will be responsible for ensuring data preservation and sharing?	The applicant (Marta Ciwinska) and the project lead (Prof. Colinda Scheele) will share the responsibility for ensuring data preservation and reuse.
Who bears the end responsibility for updating & implementing this DMP? *Default response: The PI bears the overall responsibility for updating & implementing this DMP*	The PI bears the overall responsibility for updating and implementing this DMP.