# FWO JR RESEARCH- EXPLOITING A NOVEL LINK BETWEEN METABOLISM AND ACETYLATION TO PREDICT AND TO TREAT METASTASIS FORMATION

## ADMIN DETAILS

**Project Name:** FWO Jr Research- Exploiting a novel link between metabolism and acetylation to predict and to treat metastasis formation

**Grant Title:** G0B4122N

**Principal Investigator / Researcher:** Sarah-Maria Fendt

**Project Data Contact:** Sarah-Maria Fendt

**Description:** Metastasis formation is the leading cause of death in breast cancer patients with nearly 30% of patients developing metastases. Unfortunately, we currently cannot predict which patient will develop metastases due to our limited understanding of how cancer cells gain metastatic capacity. This greatly reduces the application of personalized follow up screening and preventative measures. We have discovered that increased concentrations of the metabolite acetyl-CoA and consequently histone acetylation greatly potentiates the ability of breast cancer cells to metastasize. Thus, we hypothesize that elevated acetyl-CoA and histone acetylation in primary cancers predict metastasis risk and can be targeted for therapy once metastases arise. Thus, we will define 1) To which extent acetyl-CoA and histone acetylation are predictive of metastasis formation, 2) How histone acetylation drives metastatic capacity and 3) Which organ-specific metabolic factors sustain acetyl-CoA-mediated histone acetylation. To address these questions, we will exploit ChIP-, ATAC-, and RNA-sequencing, spatial/bulk metabolomics, multiplex IHC and pharmacologic interventions in breast cancer cell lines, mouse models and human breast cancer samples. In conclusion, we will deliver a mechanistic understanding of how breast cancer cells gain metastatic capacity. We expect that this project will contribute in the long-term to an improved prediction of metastasis risk and to therapeutic strategies against metastases.

**Institution:** KU Leuven

## 1. GENERAL INFORMATION

**Name applicant**
Sarah-Maria Fendt

**FWO Project Number & Title**
G0B4122N-Exploiting a novel link between metabolism and acetylation to predict and to treat metastasis formation

**Affiliation**

- KU Leuven

KULeuven-VIB Center of Cancer Biology

## 2. DATA DESCRIPTION

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

- Generate new 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).**

1. ***In vitro* cell culture measurements:** Mouse breast cancer cell lines (4T1 and EMT6.5) and human breast cancer cell lines (MCF7, T47D and SUM149) will be used as *in vitro* models for cell culture data collection. Frozen stocks will be stored in liquid nitrogen tanks and excel files (.xlsx) with a detailed description of the frozen stocks and their location will be updated regularly. Cell counting will be performed using a cell counter and resulting data will be manually recorded in an excel file (.xlsx). Microscopy images of the cell cultures will be stored as image files (.jpg). All above mentioned files will be stored on the lab server (KU Leuven L-drive). *Estimated volume: <1GB*
2. ***In vivo* mouse data:** We will use syngeneic (mouse breast cancer cell lines: 4T1 and EMT6.5), xenograft (human breast cancer cell lines: BT474, T47D, MCF7, MDA-MB-231, SUM149, SUM159) and patient-derived xenograft (PDX) mouse models. The PDX mouse model includes previously used BCM-3107, BCM-42072, BCM-3561 and BCM-2665 PDX mice from collaboration with Prof. M. Lewis and newly developed ER/HER2+ PDX mouse models of brain and lung metastasis of the UPTIDER consortium (S64410). Body weight, blood/organ metabolite measurements and tumor burden will be recorded manually in Excel files (.xlsx). Images of mouse tissues/metastasis (lung and brain) will be taken with a digital camera and saved as .jpg image files. *Estimated volume: 5GB*
3. **Human sample data:** We will use samples from triple negative, HER2 positive and ER positive breast cancer patients with detailed clinical follow up (1-20 years depending on the tumor type). For this we will collaborate with the pathologist Prof. Thomas Grünewald (DKFZ, DE) and the breast cancer oncologist Prof. Hans Wildiers (UZ Leuven, BE). Clinical data from patients will be accessed

through the UZ Leuven (Prof. Wildiers) or DKFZ (Prof. Grünewald) biobanks and will contain no information that allows the identification of the patient's identity. *Estimated volume: <10MB*

4. **Pharmacological treatments:** Chemicals used for pharmacological inhibition will be stored in appropriate conditions and a detailed information excel file (.xlsx) will be maintained in the shared folders on the lab server (KU Leuven L-drive). *Estimated volume: <1MB*
5. **Genome engineering:** Plasmids, oligonucleotides and lentiviral vectors to generate knockout, knockdown or overexpression cell lines will be stored at the appropriate temperatures (-20 or -80 degrees). They will be linked to a unique identifying number in the lab's inventory excel lists. Excel files (.xlsx) for plasmids/oligonucleotides/lentiviral vectors will contain detailed information about the targeted genes, species, base pair sequence and storage location. They will be maintained in the shared folder on the lab server (KU Leuven L-drive). *Estimated volume: <10MB*
6. **Gene expression analysis:** RT-qPCR data from RNA, extracted from cells/tissues, will be exported to excel files (.xlsx). Oligonucleotide primers for qPCR will be stored at -20 degrees, and will be listed in the laboratory primer inventory excel files (.xlsx) with unique identifying numbers, base pair sequences, the target gene/species and storage location. These files will be updated on a regular basis in the shared folder on the KU Leuven L-drive. *Estimated volume: <1GB*
7. **Protein expression analysis:** Protein expression analysis of mouse and patient cells/tissues will be performed using western blot, immunohistochemistry/immunofluorescent staining and FACS. Images from western blot experiments will be taken using the LAS4000 imager, leading to the generation of .gel files. Images will subsequently be exported as .tiff or .jpg image files. Protein quantifications will done using a plate reader and the data will be exported as an excel file (.xlsx). Microscopy images of (multiplex) immunohistochemistry and immunofluorescence experiments will be exported as .jpg or .tiff files. Raw flow cytometric data will be stored as FCS files (.fcs) and processed using FlowJo software leading, leading to the generation of FlowJo workspace files (.wsp). All data will be stored on the lab server (KU Leuven L-drive). *Estimated volume: 100GB*
8. **Mass spectrometry data (metabolite and acetylation measurements, 13C-tracing):** Tissue and plasma samples will be measured using the GC-MS and LC-MS systems, resulting in the generation of raw data files (.ms and .d files, respectively). The .ms and .d files will be integrated using the Agilent ChemStation (compatible with the .ms files) and Agilent MassHunter software ( compatible with the .d files). The integrated data will subsequently be exported as excel files (.xlsx) and further analyzed in MATLAB, leading to the generation of data analysis scripts (.m) and data files (.mat). The output of the data analysis will be exported in excel files (.xlsx) and figure files (.png or .pdf). All data will be stored on the lab server (KU Leuven L-drive). *Estimated volume: 5GB*
9. **Spatial metabolomics:** Mass spectrometry imaging data will be obtained from the mouse primary breast tumors tissues with Matrix-Assisted Laser Desorption Ionization (MALDI) and processed using MetaboScape from Bruker to perform non-targeted analyses. The raw data files (.d) are stored on the acquisition computer and on the lab server after data processing. Additional files created during the acquisition are .txt files, containing the sample information and metadata and .mis file for data visualization in the Scils lab software from Bruker. A metaboscape compound annotations .mca file is created with MetaboScape and allows the visualization of the identified metabolites in Scils Lab, leading to the generation of .slx files, that can be exported to .png images. All data will be stored on the lab server (KU Leuven L-drive). *Estimated volume: 250GB*

10. **ChIP/ATAC/RNA sequencing:** ChIP, ATAC and RNA sequencing will be performed. Sequencing data file (.fastq), count matrix (.mtx) and genome-aligned (.bam) files will be generated. Data analysis will be conducted in R, generating R script files (.R) with analysis output (.RData, and .txt). Gene set enrichment analysis (GSEA) will be conducted, requiring libraries of pathways (.gmt) and normalized transcriptome reads in a ranked format (.rnk). Peak Calling will be performed for chIP and ATAC sequencing analysis to identify the regions which are enriched with aligned reads as a consequence of protein binding or chromatin accessibility, leading to the generation of .narrowPeak or .broadPeak files. The data will be will visualized in .bedGraph and .tdf files. Output files are produced in various formats (.html, .xls, .png images). All data will be stored on the lab server (KU Leuven L-drive). *Estimated volume: 100GB*
11. **Bioluminescence:** We will inject human (SUM149 and MDA-MB-231) and mouse (4T1 and EMT6.5) breast cancer cell lines, expressing luciferase, into the mammary pad or intracardially in mice. Bioluminescence imaging will be performed to measure the luciferin flux using the IVIS Spectrum system. Bioluminescence will be analyzed using the LivingImage processing software. Generated files are .txt, .xlsx, .tiff and .png. All data will be stored on the lab server (KU Leuven L-drive). *Estimated volume < 10MB*
12. **Statistical analysis and graphical figures:** GraphPad Prism9 will be used for all statistical analyses and graphical designs, leading to the generation of program-specific files (.pzfx). Final figures will be exported in the following file types: .pdf, .png, .wmf, .svg. All data will be stored on the lab server (KU Leuven L-drive). *Estimated volume: 5GB*
13. **Protocols, documentation, metadata, and papers:** Experimental and computational protocols, as well as details related to collection and processing of data (both documentation and metadata) will be stored in Word (.docx), Excel (.xlsx), or text (.txt) files. Manuscripts originating from the project will be stored as Word documents (.docx), and final versions will be exported to PDF format (.pdf). All data will be stored on the lab server (KU Leuven L-drive). *Estimated volume: <1GB*

## 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

Clinical data from patients will be accessed through the UZ Leuven (Prof. Wildiers) or DKFZ (Prof. Grünewald) biobanks and will only contain information on the age of the patient, initials and a personal identification number as well as health data (e.g. description of characteristics of physical features of the body, medical history and medical test information (such as blood samples results from scans and biopsies)). GDPR approval was obtained before applying for ethical approval at the UZ Leuven EC by Prof. Wildiers.

**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

Animal experiments will be performed during this project. All experiments have been approved by the Ethical Committee for Animal Experimentation (ECD) at KU Leuven and are available and outlined in ECD #P007-2020.
Human samples will be used during this project. The collection and use of samples obtained through Prof. Wildiers has been approved by the Ethical Committee Research UZ/KU Leuven. The approval is available and outlined in EC S63773, S65809 and S64410. The collection and use of samples obtained through Prof. Grünewald has been approved for research on selected proteins by their local ethical committee.

**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

Intellectual Property arising from this work is managed as per the framework agreement between the VIB (VIB Tech Transfer) and the KU Leuven, the two participating institutes in this study.

**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

Yes, as above, dissemination or exploitation of the data is managed according to the framework agreement between the VIB and the KU Leuven.

## 4. DOCUMENTATION AND METADATA

**What documentation will be provided to enable reuse of the data collected/generated in this project?**

Protocols and details related to data collection and processing will be recorded in physical and electronic lab books and transcribed to Word or Excel files by the lead post-doctoral scientist (Dr. Patricia Altea-Manzano), PhD student (Anke Vandekeere) and lab managers/technicians (Dorien Broekaert and Ines Vermeire). Data folders containing raw and processed data will be hierarchically organized and labeled based on the source of the data, the type of experiment, the date of data generation, and the different experimental conditions analyzed. Data analysis methods and particularities (including metadata) will be described in text documents and Excel files included in these folders. All files will be stored in the KU Leuven Large Volume Storage space (L-Drive), with sharing possibilities via Box Sync and One Drive (managed by the KU Leuven IT department).

**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.**

- No

Text documents and Excel files stored within each experiment folder in the L-Drive will contain guidelines describing data collection/analysis methods and all relevant metadata (including experimental conditions, sample keys, computational analysis pipelines and their parameters) to ensure the reusability of the data and the reproducibility of any further data generation.

## 5. DATA STORAGE AND BACKUP DURING THE FWO PROJECT

**Where will the data be stored?**
Upon data collection/preprocessing, temporary copies of the data will primarily be stored in the KU Leuven-managed personal computer of the lead post-doctoral scientist (Dr. Patricia Altea-Manzano) and the PhD student (Anke Vandekeere). 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 is backup of the data provided?**
Data stored on the KU Leuven L-Drive is managed, maintained, and backed up by KU Leuven IT services.

**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 L-drive has sufficient storage capacity for the outlined project.

**What are the expected costs for data storage and back up during the project? How will these costs be covered?**
The annual cost of KU Leuven L-Drive storage is 569.2 euro/5TB/year. This cost and capacity include the performance of mirror copies of the stored data, for safety backup purposes. We expect that 5TB will be sufficient to store all data generated as part of the project. These costs will be covered by the budget of the project lead (Prof. Sarah-Maria Fendt).

**Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?**
Data stored intermediately on personal computers are protected via password access to the computers as set up by KU Leuven IT. Off-site access to L-drive data is available by KU Leuven data access points and is password protected. Access to modify these files are limited to lab members with access to the Fendt Lab L-Drive folders.

## 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 raw data will be retained for 5 years on the KU Leuven L-Drive. Publication data is further organized and catalogued by figure for future reference to raw datasets used for figure generation.

**Where will the data be archived (= stored for the longer term)?**
Long term data archival is maintained in specific archive folders on the KU Leuven L-Drive.

**What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?**
The annual cost of L-Drive storage is 569.2 € per 5TB of storage space per year. We expect that 5 TB will be sufficient for long-term storage of all data generated as part of the project. These costs will be covered by the budget of the project lead (Prof. Sarah-Maria Fendt).

## 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:

We will use tumor samples collected from patients at UZ Leuven. Therefore, we have a Material and Data Transfer Agreement (MTA) between the legal entities of KU Leuven/UZ Leuven and VIB in place/filled. These 3rd party agreements will not restrict dissemination of the data that will be generated within this project.
The results of this project will be communicated in established, peer-reviewed (non-predatory) academic journals, that require disclosure of all included data. Patient data will therefore be made available in restricted access repositories (e.g. EGA (sequencing) or in specialized open access repositories with appropriate access control such as OMERO (images)). Requests for access to the patient data will be assessed on a case-by-case basis by the Data Access Committee, KU/UZ Leuven and the project lead (Prof. Sarah-Maria Fendt). Personal data will only be published after de-identification to protect the identity of the patients. The identifiers will not be published.
In order to protect the privacy of the patients, their tumor samples will only be made available to lab members involved in the project, not to other groups or studies unless ethical approval has been obtained.

All sequencing data generated from non-patient material will be made available on the open access repository, GEO. Other data will be shared externally upon request of collaborating scientist, which will be reviewed and approved on case-by-case basis by Prof. Fendt.

**Which data will be made available after the end of the project?**
The key findings of the project and their interpretation will be made available to peers and the wide audience through publication in established, peer-reviewed (non-predatory) academic journals. Additionally, the sequencing data of non-patient-related samples will be made available on the GEO website. Data related to patient samples will be made available in repositories with appropriate access control (e.g. EGA or OMERO). Access to these data will be made available to any individuals making a specific request and this

request will be handled by the institutional data access committee (DAC) in agreement with UZ/KU Leuven and the project lead, Prof. Sarah-Maria Fendt. Any data shared will only be released prior to a Data Transfer Agreement that will have to include the necessary conditions to guarantee protection of personal data (according to European GDPR law). The personal data will be de-identified and the identifiers will not be published to protect the privacy of the patients.

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

- In an Open Access repository
- In a restricted access repository
- Upon request by mail

Datasets will be made publicly available, if possible, through repositories to support FAIR data sharing. Mouse sequencing data will be made available on the GEO website. All patient data will be made available in repositories with appropriate access control (e.g. EGA or OMERO) upon de-identification of the personal data. Access to these data will be made available to any individuals making a specific request and this request will be handled by the institutional data access committee (DAC) in consultation with UZ/KU Leuven and the project lead, Prof. Sarah-Maria Fendt. The remaining data can be approved for reuse upon request and will be assessed for approval on a case-by-case basis by the project lead (Prof. Sarah-Maria Fendt).

**When will the data be made available?**

- Upon publication of the research results

Summaries of key findings of the project and their interpretation will be made available through the publication of journal articles in reputable academic journals. Upon publication, the mouse and human sequencing data will be made available in open access repository GEO and restricted access repositories EGA, respectively. Patient data will be de-identified to protect the privacy of the patients.

**Who will be able to access the data and under what conditions?**

Unpublished data will only be accessible to lab members involved in the project. Upon publication of the data in reputable journals, the published data will be made publicly available through data repositories to support FAIR data sharing. Any user can place reasonable data requests for non-commercial purposes, and these requests will be assessed on a case-by-case basis by the project lead (Prof. Sarah-Maria Fendt), and in case of patient data, assessed in consultation with the Data Access Committee and UZ/KU Leuven. Commercial-based requests will be navigated in coordination between Prof. Sarah-Maria Fendt, if applicable: UZ/KU Leuven, and the VIB/VIB Tech Transfer team.

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

Costs for data sharing will be discussed with collaborators on a case-by-case basis. To minimize data management costs, free-to-use data repositories will be used when possible. Data management will be covered by own funding

## 8. RESPONSIBILITIES

**Who will be responsible for data documentation & metadata?**

Prof. Sarah-Maria Fendt accepts all responsibility for data documentation and metadata. Patricia Altea-Manzano and Anke Vandekeere will be responsible for experimental data. Melanie Planque will be responsible for mass spectrometry data.

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

The lead postdoctoral scientist (Dr. Particia Altea-Manzano) and PhD student (Anke Vandekeere) will be primarily responsible for collecting/generating data, and for correct documentation and upload 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 reuse ?**

Prof. Sarah-Maria Fendt will be responsible for ensuring data preservation and reuse.

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

Prof. Sarah-Maria Fendt bears the end responsibility of updating & implementing this DMP.