Targeting AKR1Cs to reverse chemoresistance in ovarian cancer

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.

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
		*Indicate: **N**(ew data) or **E**(xisting data)*	Indicate: **D**(igital) or **P**(hysical)	Indicate: **A**udiovisual **I**mages **S**ound **N**umerical **T**extual **M**odel **SO**ftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
DS1. Clinical data	Data retrieved from the patient records (KWS platform UZ Leuven), stored in a pseudonymised manner	E	D	I N T	.xls .doc	<1GB	N/A
DS2. Personal data	Age, clinical history	E	D	N T	.xls .doc	<1GB	N/A
DS3. Genetic data	mRNA expression data (qPCR and RNASeq)	N	D	N T	.fastq files (raw data) and .csv files (analyzed data)	<10TB	N/A
DS4. Patient-derived explant models	Fresh tumor resection collected and biobanked under the S67067 protocol will be treated with AKR1C inhibitors	N	D P	I N T			FFPE biological samples RNA/DNA Extracted proteins

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

Clinical and personal data will be retrieved from the UZ Leuven Klinikal Werk Station (KWS) system by authorized personnel in the Gynecological Oncology Lab

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, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- Yes, human subject data (Provide SMEC or EC approval number below)

S67067

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- Yes (Provide PRET G-number or EC S-number below)

S67067

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.

- No

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

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).

All digital data and metadata accompanying physical data will be stored on the data management platform from KU Leuven ManGO. This platform allows researchers to store and manage their data via different clients such as ManGO portal, or SFTP clients.

• Storage
The data is stored securely in the data centers of KU Leuven. Of each file, two copies are stored: one in the datacenter in Heverlee, and one other in the datacentre of Leuven.

• Metadata
Data can be described in the platform by adding metadata (at the file or folder level) to provide context to the data and make the data findable via the search interface. Each dataset will include a metadata file which will contain:
o Title and description of the dataset
o Names and contact information of the data creators
o Date of data collection
o Anonymized identifiers of patients from which samples were derived
o File formats and structure
o Instructions for accessing and using the data and relationship between different data elements

The structure of the ManGO platform ensures data adheres to the FAIR data principles.

Will a metadata standard be used to make it easier to find and reuse the data?
If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- Yes

The ManGO platform allows users to design metadata schemas. Using these schemas, users of the ManGO portal have to fill in specific information accompanying their data-upload. Specific schemes will be designed for each datatype.
RNA seq data
- o Dataset title
- o Dataset description
- o File format
  - ■ Raw data: FASTQ
  - ■ Processed data: BAM/SAM, HDF5, CSV
  - ■ Metadata: CSV, JSON
- o Sample ID
- o Donor ID
- o Sample collection data
- o Sequencing platform
- o Read length
- o Sequencing depth
- o Library preparation protocol
- o Raw data processing software
- o Alignment reference genome

For all physical datasets (e.g. FFPE and frozen tumor material) containers will be labelled with:
- o Project ID
- o Sample ID
- o Donor ID
- o Date of collection
- o Material type
- o Sample processing method
  - • Fixation
  - • Freezing
- o These physical samples will be accompanied by a metadata file in ManGO detailing the same information as the label on the physical container, with the addition of location where container is stored
  - • Liquid nitrogen tank ID (flash frozen/viable frozen tissue)
    - ■ Rack number
    - ■ Box number
    - ■ Box position
  - • Storage container ID (FFPE material)

Data Storage & Back-up during the Research Project

Where will the data be stored?

- ManGO
- OneDrive (KU Leuven)

An electronic lab notebook will be used to register and manage sample collection data and molecular analyses data

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution
- Other (specify below)

The data are stored securely in the data centers of KU Leuven. Of each file, two copies are stored: one in the datacenter in Heverlee, and one other in the datacentre of Leuven.

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

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

ManGo can only be accessed via KU Leuven accounts (after being granted permission) via the ManGO portal or through SFTP clients. The KU Leuven RDM-ICTS support team manages access and permission to the ManGO platform.

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

The expected costs for data storage are EUR 330.00/year. These costs will be covered by the budget granted for the CELSA application.

Data Preservation after the end of the Research Project

Which data will be retained for 10 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 will be preserved for 10 years according to KU Leuven RDM policy

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

- Other (specify below)

All digital data and metadata accompanying physical data will be stored on the data management platform from KU Leuven ManGO Archive.

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

The expected costs for data storage are EUR 330.00/year. The costs of storage during the active years of the project are covered by the CELSA grant. The following years, these costs will be covered as part of the Gynecological Oncology Lab running costs from running credits.

Data Sharing and Reuse

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

- Yes, as restricted data (upon approval, or institutional access only)

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

Sequencing data will be submitted to the European Genome-Phenome Archive (EGA) and can be accessed after submission of a controlled data access (CDA) application

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 per dataset or data type where appropriate.

- Yes, privacy aspects
- Yes, ethical aspects

To protect patient privacy, sequencing data of patient-derived tumor material will be deposited in an access-controlled data repository.

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- Other data repository (specify below)

European Genome-Phenome Archive

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.

- Data Transfer Agreement (restricted data)

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- Yes, a PID will be added upon deposit in a data repository

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

No costs are foreseen for sharing the data on EGA (free)

Responsibilities

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

Prof. Dr. Frédéric Amant
Dr. Daniela Annibali

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

Prof. Dr. Frédéric Amant
Dr. Daniela Annibali

Who will manage data preservation and sharing?

Prof. Dr. Frédéric Amant
Dr. Daniela Annibali

Who will update and implement this DMP?

Prof. Dr. Frédéric Amant
Dr. Daniela Annibali