DMP title

Project Name FWO FC 2022 - DMP title

Project Identifier AR Dim

Grant Title G057422N

Principal Investigator / Researcher Frank Claessens

Description Study basic aspects of the action mechanism of the male hormones (androgens) and their receptors (AR). The AR is a transcription factor that according to the dogma has to dimerise in order to be active. There are three ways that AR can dimerise. Research questions: what is the molecular role of each of the three dimerization functions of the AR. This will be studied in transgenic mouse models and in cell lines. What is the effect of mutations in the dimerization surfaces on the in vivo androgen responses in mice? Data that will be collected is general phenotyping of mutant mice (organ weights, endocrinology), gene expression in different androgen-responsive organs (RNAseq and qPCR), chromatin binding by the AR and changes in chromatin landscape by AR and other ChIPseq (chromatin immunoprecipitation followed by high throughput sequencing).

Institution KU Leuven

1. General Information

Name applicant

Frank Claessens

FWO Project Number & Title

G057422N: New therapeutic targets in the dimerization modes that drive androgen receptor functioning

Affiliation

- KU Leuven

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).

	Type of data	Format	Volume	How created
on mouse models	phenotyping rodent models	.xls	3 files, max 1 GB	breeding success of male ARDmon mice, organ weighing on a scale, hormone levels in serum by Laboratorium geneeskunde GHB
	microscope images (TissueFax)	.tif; .AQPROJ; .dat	35 GB per experiment (max 140 GB in total)	histology in collaboration with pathology lab GHB
	RNAseq	.BAM; .BAI	3 GB per group (containing 5 biol replicates) (max 24 GB in total)	by genomics core KU Leuven

	AR ChIPseq	.BAM; .BAI	3 GB per group (containing 5 biol replicates) (max 24 GB in total)	immunoprecipitation followed by NGS sequencing at genomics core KU Leuven
	Type of data	Format	Volume	How created
datatypes on cell lines	cell proliferation	.xls	max 1GB	Incucyte ZOOM or S3
	Western blots	.gel; .xls	max 1 GB	cell extracts; ImageQuant LAS4000
	gene reporter assays	.xls	max .1 GB	transfecting cell lines; Luminoscan
	RNAseq	.BAM; .BAI	3 GB per group (containing 5 biol replicates) (max 24 GB in total)	by genomics core KU Leuven
	AR ChIPseq	.BAM; .BAI	3 GB per group (containing 5 biol replicates) (max 24 GB in total)	immunoprecipitation followed by NGS sequencing at genomics core KU Leuven
	Type of data	Format	Volume	How created
purified protein fragments	BLI	.xls	max 1 GB	Biolayer inferometer
	EMSA	.tif and .xls	max 1 GB	radiography after gel electrophoresis
	crystal diffractions	pdb database	max 1 GB	model of structure (diffraction patterns are kept by the Laboratory of Biomolecular Modelling and Design (PI Arnout Voet)

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.

- No

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)

- No

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?

- No

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?

- No

4. Documentation and metadata

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

All methodology and protocols will be described in detail in the lab book.

RNAseq and ChIPseq datasets will be deposited in the GEO EMBL databank, together with requested metadata.

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

5. Data storage and backup during the FWO project

Where will the data be stored?

Master copies of the data will be kept purely for archiving on our research unit central storage facility (K-Drive). Additional copies will be kept on J-Drive (GBW-0085_Molendo) and on personal devices for data analysis.

How is backup of the data provided?

The data will be stored on the university's central servers (K-DRive and J-Drive) with automatic daily back-up procedures.

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

Within the dimensions of our project, the capacity of K-Drive and J-Drive can easily be adjusted according to our needs.

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

For keeping 300 GB (overestimate based on Table in section 2):

J-drive costs (@52€/100GB/year): 156€/y

K-drive costs (@11,38€/100GB/year): 34€/y

3000€ per year are mentioned in the application for data storage and bioinformatics.

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

We will have transcriptomics data on mouse models and human cell lines. These will be stored at the university's secure drives untill accepted for publication and at least 5 more years after publication. To have access to the data, Frank Claessens will have to give autorisation to the K and J Drive to every member of the laboratory.

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 data mentioned in section 2 will be kept for 5 years on the archive drive of KU Leuven (K-Drive).
Lab notebooks will be kept for 10 years.

Where will the data be archived (= stored for the longer term)?
The RNAseq and ChIPseq data and histology data will be stored on the university's central servers (K-Drive with automatic back-up procedures) for at least 5 years. If during these 5 years, data would become of interest, this can be prolongued to 10 years. These data are on mouse and cellular models.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?
Data storage was a separate cost mentioned in the FWO application (3000€/y):
J-Drive costs are (@52€/100GB/Y) 880€ for 5 years;
K-Drive costs (@11,38€/100GB/y) 170€ for 5 years.

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)?

- No

Which data will be made available after the end of the project?
Full RNAseq and ChIIPseq datasets as .bam files.

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

- In an Open Access repository

GEO database.

When will the data be made available?

- Upon publication of the research results

As mandatory by most journals in which our work is published, the datasets will become available upon publication.

Who will be able to access the data and under what conditions?
Data are from muse models and cell lines. They will be publicly available.

What are the expected costs for data sharing? How will the costs be covered?
No costs. The databases are free.

8. Responsibilities

Who will be responsible for data documentation & metadata?
Frank Claessens is PI of the project, Kaat Peperstraete is the PhD student who will generate and analyse the data.

Who will be responsible for data storage & back up during the project?
Frank Claessens is PI of this project, Kaat Peperstraete will store the data and submit them to the GEO database.

Who will be responsible for ensuring data preservation and reuse ?
Frank Claessens

Who bears the end responsibility for updating & implementing this DMP?
The PI (Frank Claessens) bears the end responsibility of updating & implementing this DMP.