How do androgens and anti-androgens affect the different cell types that populate localized, high risk prostate cancer
DPIA

DPIA

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

- Not applicable

How do androgens and anti-androgens affect the different cell types that populate localized, high risk prostate cancer
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)

Sequencing data
Imaging data
Single cell multi-ome: ATAC-seq and RNA-seq on clinical samples

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)

Frank Claessens and Steven Joniau will be the end responsible persons for data management.
Storage: Digital data will be stored on the J- and L-drives of the Mol Endo Laboratory (automatic back up, password- and authentication-controlled access). Long term storage of large datasets will be on the archive K-Drive.
Physical samples are stored at the Mol Endo Laboratory and the UZ Leuven biobank.
After the research, the sequencing data will be kept on the K-Drive. At publication, the data will be uploaded to repositories according to publication strategies of the journals (if applicable, under embargo for 5 years, after which they are released).

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)

We do not wish to deviate.

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)

Clinical samples were collected by Steven Joniau at UZ Leuven. They were pseudo-anonymized by the clinician. The key data will not be available to the people that do the experiments described in the application.

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

Bioinformatic pipelines that will be developed will be kept safe by our collaborator Matti Nykter, as per guidelines of the Tampere University (MTA is in place).

How do androgens and anti-androgens affect the different cell types that populate localized, high risk prostate cancer

GDPR

GDPR

Have you registered personal data processing activities for this project?

- Yes

How do androgens and anti-androgens affect the different cell types that populate localized, high risk prostate cancer
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.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
Sequencing data	SUM-seq: single nuclei sequencing data on RNA and ATAC-seq; raw sequencing reads, processed data bulk RNA-seq and ATAC-seq data	New data existing data	Digital	*Please choose from the following options:* Other	*Please choose from the following options:* .txt, .xlsx, .csv, .tsv, .fq, .bam, .R, .html, .pdf	*Please choose from the following options:* >5TB	
imaging data	microscopic images, IHC	New data	digital	images	.tif, .jpg, .png, .pdf	<1TB	
physical samples	tissue samples, fixed samples, frozen samples	Existing data	physical	na	na	na	150 Frozen samples frozen at -80°C. 150 Fixed samples are kept at pathology department.
patient clinical data		reuse existing (recurrence, pathology, clinical) data + generate new (recurrence) data	digital	.xls	.xls	>1 GB	

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:

Metadata on the ARNEO study as described in doi: 10.1016/j.eururo.2022.09.009.

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, human subject data

From a punch biopsy from the resected prostate cancer, single cell RNA and chromatin structure data will be collected (SUM-

seq). The generated molecular data will be correlated with clinical data. Ethical committee of UZ Leuven has given approval for the clinical intervention and the omics analyses (S58827). An MTA with Tampere University is in place.

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.

- Yes

Anonimized clinical data on prostate cancer development for the two treatment arms of the clinical ARNEO trial and a comparator group of patients will be used.

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

There is a potential that the data may generate novel concepts that could be valorized e.g. in the development of prognostic or predictive markers or inspiration of next generation clinical trial(s). However, this is realistically not within the scope of this research project.

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.

- Yes

We have an MTA with Tampere University (Matti Nykter) and University College London (Gerhardt Attard) for the deeper analyses of the data and combination with data generated in the respective institutes (published in doi: 10.1038/s41467-024-54364-1).

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

There is a potential that the data may lead to new concepts that could be valorized e.g. in the development of prognostic or predictive markers or inspiration of next generation clinical trial(s). However, this is not within the scope of this research project.

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).

We will maintain a record of the following for every WP (where applicable):
- Experimental design and protocol (.docx file)
- Abbreviations used (.docx file)
- Structure of the data (.docx file)
- Steps involved in data analysis and relevant analysis scripts
- Raw data (specific file format according to data type)
- Analyzed data (specific file format according to data type)
- Index file/read me file (.txt file) for every WP, linking the name, location (folder and subfolder on /server) and description of above-mentioned files.

Physical samples will be stored (frozen at -80*C or formalin fixed at room temperature) at UZ Leuven (pathology) or at the Molecular Endocrinology Laboratory, depending on the kind of sample.

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.

- Yes

Metadata standards will be used for genomics (http://www.dcc.ac.uk/resources/metadata-standards/genomemetadata). For all other data, metadata will be created using the Dublin core (http://www.dcc.ac.uk/resources/metadatastandards/dublin-core).

3. Data storage & back-up during the research project

Where will the data be stored?

Data will be stored on a shared network drive (J-Drive), a personal network drive by the analysing collaborators (I-drive), OneDrive (KU Leuven) and large volume archival storage (K-Drive) at KU Leuven.

How will the data be backed up?

Standard back-up is provided by KU Leuven ICTS, Biomedical Sciences Group.
Personal backups will be made to KU Leuven OneDrive.

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

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

Researchers involved in the project can control who they give access to the files on their personal OneDrive. To access the KU Leuven servers, access is provided and controlled by Frank Claessens. The KU Leuven ICTS data center hosts the network storage, with a mirror available in the second ICTS center. This ensures additional back-up capacity, recovery of lost data and long term data availability. The access is controlled by KU Leuven security groups and it is password and authenticator-protected.
Data related to Patients and controls are pseudonymized, which is managed by Steven Joniau.
All samples are processed and stored as per their password-protected anonymized code only.

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

Personal and Shared network drive (I, J) 503,66 euro/TB per year. All large primary data will be moved to LVS. LVS storage costs per 5 Tb (KU Leuven ICTS): 104,42euro/year. The costs have been budgeted on the grant.

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 will be preserved for 5 years after the end of this project, according to KU Leuven RDM policy.

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

Large Volume storage (K-Drive) from KU Leuven ICTS. Physical samples will be stored at the Molecular Endocrinology Laboratory at KU Leuven (PI Frank Claessens).

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

Estimated costs of data storage of approximately 3000€ per year will have to be covered by other grants of either Frank Claessens or Steven Joniau.

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.

- Yes, in a restricted access repository (after approval, institutional access only, …)

The data will be made available after publication via the required link in the publications or upon request after an embargo period after publication. The same holds true for unpublished data, they can be made available upon request but only under an MTA and after an embargo period (3 years; exceptionally 5 years after the project).

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

All team members have access as long as they are affiliated to KU Leuven or to the group of Matti Nykter (Tampere University). Once all files are released, anyone can use these data to generate new results, referring to the original publication and not for commercial use. Other data will be only released upon request and after an embargo period after publication.

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.

Experimental data will be made available through a data repository such as EGA, ncbi, github, Genbank, FigShare (https://figshare.com/), depending on the type of data.

When will the data be made available?

Upon publication in a peer reviewed scientific journal.

Which data usage licenses are you going to provide? If none, please explain why.

CC-BY 4.0, or DTA for unpublished data.

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.

- Yes

not available yet. DOIs of accepted papers in peer-reviewed journals.

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

The transfer costs depend on the data repository selected. Costs will be covered by the project funding or by the recipient.

6. Responsibilities

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

Frank Claessens and Sofie De Block

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

Frank Claessens and Sofie De Block

Who will manage data preservation and sharing?

Frank Claessens and Sofie De Block

Who will update and implement this DMP?

Frank Claessens and Sofie De Block