Activating negative regulators of KRAS signaling as a novel anticancer paradigm

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	
WP1.1	Reporter plasmids	Library of (mutant) RAS stability reporters, 24-mer-peptide reporters, and site-directed mutants of these	N	P	Other: new tools (DNA)	NA	NA	<50µl/plasmid
	guide RNAs	for LZTR1 knockout	E	P	Other: existing tools (gRNA)	NA	NA	<50µl/gRNA
	plasmid sequencing data		N	D	T			NA
	Fluorescence measurements		N	D	N, I			NA
	mass spectrometry data	targeted proteomics (external collaborator)	N	D	N+T			NA
WP1.2 + WP2.1	Plasmids for E. coli expression	full-length LZTR1 or isolated LZTR1 subdomains; different RAS isoforms and truncations; LZTR1 truncations	N	P	Other: new tools (DNA)	NA	NA	<50µl/plasmid
	purified recombinant proteins	RAS, LZTR1 and/or subdomains of each	N	P	Other: new tools (protein)	NA	NA	<10ml/protein
	Protein purification data:chromatography profiles,SDS-PAGE gels and immunoblots		N	D				NA
	ITC data	Data from Molecular Biophysics facility KU Leuven	N	D				NA
	SPR data	Data from Rega institute facility	N	D				NA
	SEC-MALS data	Data from Molecular Biophysics facility KU Leuven	N	D				NA

WP1.3 + WP2.2	Screening of crystallization conditions	Crystallization facility Biocrystallography Lab -- Formulatrix Imagers	N	D				NA
	initial quality data of obtained crystals	X-ray diffraction -- Biocrystallography lab	N	D				NA
	raw full X-ray diffraction data of crystals	X-ray diffraction -- ESRF synchrotron	N	D				NA
	raw SEC-SAXS data	ESRF synchrotron	N	D				NA
	raw Cryo-EM data	200kV JEOL microscope; JEOL cryoARM300 microscope	N	D				NA
	Processed data: structures	of LZTR1-KRAS complex and/or 'domain variants' of that complex; of LZTR1 multimers	N	D				NA
WP1.4 + WP2.3	Raw and processed data from FBS-X screen	XChem facility Oxford	N	D				NA
	differential scanning fluorimetry (DSF) data	Data from Molecular Biophysics facility KU Leuven	N	D				NA
	In silico predictions of extended fragments (molecular dynamics data, AI-predictions)		N	D				NA
	New synthesized compounds	Medicinal Chemistry lab	N	P	Other: new tools (chemicals)	NA	NA	<20ml/compound
	ITC data		N	D				NA
	SPR data		N	D				NA
	initial quality data of obtained co-crystals		N	D				NA
	raw full X-ray diffraction data of co-crystals		N	D				NA
	Processed data: co-structures	LZTR1/KRAS/compound; LZTR1/LZTR1/compound	N	D				NA
WP1.5 + WP2.4	Hibit-tagged RAS constructs; split luciferase constructs for LZTR1		N	P				<50µl/plasmid
	luciferase reporter assay data		N	D				NA
	fluorescence measurements	GPS screen	N	D				NA
WP3.1	PP2A-C expressing baculovirus		N	P	Other: new tool (virus)	NA	NA	<20ml
	purified recombinant proteins	A subunit (E.coli), C subunit (bacculo); AC-dimer; ABC trimers	N	P	Other: new tools (protein)	NA	NA	<10ml/protein
	Protein purification data:chromatography profiles,SDS-PAGE gels and immunoblots		N	D	I	jpeg	<100 GB	NA
	chemical compounds	old and newest ITH compounds	R	P	Other: chemical	NA	NA	<20ml/compound
	ITC data		N	D				NA

	SPR data		N	D				NA
	guide RNAs	CRISPR Paint tagging	N	P	Other: new tool (gRNA)	NA	NA	<50µl/gRNA
	CRISPR-edited HEK293 cell lines		N	P	Other: new tool (cell lines)	NA	NA	min. 5 cryovials (1ml/vial)
	Screening of crystallization conditions		N	D				NA
	initial quality data of obtained crystals		N	D				NA
	raw full X-ray diffraction data of crystals		N	D				NA
	raw Cryo-EM data		N	D				NA
	Processed data: atomic structures	PP2A AC dimer + ITH compound complexes; PP2A ABC trimer + ITH	N	D				NA
WP3.2	purified recombinant proteins	PME-1 (E.coli)	N	P	Other: new tools (protein)	NA	NA	<10ml protein
	In silico predictions of extended fragments (molecular dynamics data, AI-predictions)		N	D				NA
	New synthesized compounds	done in Madrid (URJC)	N	P	Other: new tools (chemicals)	NA	NA	<20ml/compound
	SPR data	competition with PME-1?	N	D				NA
	initial quality data of obtained co-crystals	with new compounds	N	D				NA
	raw full X-ray diffraction data of co-crystals	with new compounds	N	D				NA
	Processed data: atomic co-structures	PP2A dimer/new compound	N	D				NA
WP3.3	GFP-trapping and IP: immunoblots	PME-1-GFP; C binding; C methylation, signaling	N	D	I, N	Jpeg, Excel	<100 GB	NA
	PP2A activity assays		N	D	N	Excel	<10 GB	NA
	mass spec data	C interactome, signaling	N	D	N	Excel	<100 GB	NA
WP3.4	Toxicity assays: necropsy data, mouse weights, ELISA	histology; AST/ALT expression	N	D, P	I, N	Jpeg, Excel	<10 GB	Tissue slides, paraffin blocks
	2D-clonogenic growth assays	colony number/area	N	D	N	Graphpad	<1 GB	NA
	tumor spheroids growth assays	histology, growth/apoptosis	N	D, P	N, I	Jpeg	<10 GB	Tissue slides
	xenografts	IHC, IB, growth	N	D	N, I	Jpeg, Graphpad, Excel	<10 GB	Tissue slides, paraffin blocks

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:

Existing crystallography data available in the PDB database (https://www.rcsb.org/); these will be used for model building (using the alphafold algorithm) and molecular dynamics studies to predict binding poses of compounds; these data are all publicly available, and can be freely accessed and used (upon proper acknowledgment and citation);
Existing plasmids and compounds: produced by the applicants before, and stored in the applicants' labs (freely available; all

published)

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

Effects of compounds on tumor growth will be tested in mouse xenograft models; ECD approval has been obtained to perform these studies (15/01/2023-15/01/2028: Muis xenograft modellen voor evaluatie van bestaande of nieuwe kankergeneesmiddelen - P007/2023).
For use of commercial human cancer cell lines, we have obtained department-wide ethical approval (S63808) that is annually updated in case additional cell lines need to be incorporated.

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).

- No

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.

- Yes

In case our project will create new protein or protein-complex structures, or new compounds that could have a valorization potential, we will get in touch with LRD and/or VIB Tech transfer offices to determine how to proceed to protect IP, before any publication or other form of data presentation to the wider scientific community will be considered.

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).

Each researcher involved in the project provides detailed descriptions of data acquisition in paper or electronic lab notebooks, according to good laboratory practices. *E.g.* all data belonging to a certain sub-part of the project are properly linked to each other

and referred to each other in the lab notebooks, and detailed protocols are provided (.doc files). The complete list of sub-projects is summarized as a final list (.doc file); the detailed protocols are also digitally archived on the lab's Dropbox or J-drive. In the notes, clear links are provided to (1) digital data (pictures, excel files, movies, others), and where to find these (Dropbox, J-drive, L-drive, etc); file names, types and dates have to be provided; and to (2) physical data (new cell lines, plasmids, gRNA, recombinant proteins, compounds) and where/how they are stored. Concerning the physical data, the dedicated lab lists/archives need to be updated with the new tools and their storage information; for plasmids this involves registration of the unique identifier, all relevant info on how the plasmid was made (cloning sites, primers, resistance markers), the accompanying DNA sequencing information, and its storage site.

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.

- No

No real metadata standard will be used. Each researcher provides a clear overview (summary) of the data generated or processed, per subproject (see above). Lists of new tools will be digitally archived on the Dropbox, the J-drive and the L-drive; processed data will be provided as digital info on the J-drive, and eventually on the K-drive (e.g. labmeetings, posters, intermediate reports, papers). On the K-drive, the data are structured/ordered per paper, and per researcher. The data of one researcher are structured/ordered per sub-project, and subsequently, per date and per type (for raw data).

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Other (specify below)
- Shared network drive (J-drive)
- Large Volume Storage

Paper lab notebooks are kept in locked cabinets in the labs of the PIs concerned. Digital data files are stored on local KU Leuven PCs or shared KU Leuven J-drive (connected to specific equipment: scanner, microscope, IncuCyte,…). Raw MS data are stored on storage drives backed up by KU Leuven, as well as on the L-drive (for a limited time) . After that time period, data are kept on external SSDs (1 TB). Processed MS data are stored on KU Leuven local PCs and transferred to the SSDs.
Crystallization images will be stored using dedicated 10Tb NAS storage. Raw and processed crystallography and cryo-EM data will be stored at the KU Leuven using dedicated NAS storage (180Tb).
Regarding the data pertaining to publications, it is our policy that the first author is responsible for storing ALL raw and processed data of the paper concerned in a subfolder on the K-Archive drive (<100 GB). This measurement was taken, in part, to prevent that relevant data get lost or can no longer be properly retrieved later on; in part, to prevent scientific integrity issues. Thus, the first author of the paper is responsible for making all the data concerning his/her paper fully available and tractable, by archiving them on the K-drive. We also plan to upload manuscripts before publication to bioRxiv (https://www.biorxiv.org/) whenever appropriate. In addition, after publication, all proteomics data will be deposited to the ProteomeXchange Consortium via the PRIDE partner repository; all structural data to PDB.
The physical data for this project will be stored in approriate storage conditions, i.e. liquid nitrogen cell containers (Cryothek at Biomedical campus), -80°C or -20°C freezers, in a sufficient amount of aliquots (at least 5 per cell line, 2 per plasmid or gRNA, 10 per compound). In part, and after publication, newly generated plasmids with accompanying descriptions and sequencing data, will be submitted to Addgene; new cell lines may be submitted to Cell Bank.

How will the data be backed up?

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

Besides regularly provided automated backups by ICTS (of J-drive, L-drive, K-drive), the data stored on personal PCs and J-drive

will be weekly backed up on the personal KU Leuven I-drive, and the own Dropbox of the labs concerned (password-protected). Every two months, additional backups will be made on external hard disks, as an extra security measurement. MS data older than three years and stored on SSDs will also be backed up by copying to another SSD. External SSD hard-disks (up to 1-2 TB storage capacity) keep the storage costs feasible since cost of an external hard drive is cheaper than ICTS storage costs for such big data volumes. Whenever a scientific collaborator leaves the lab, all his/her data will be stored on the K-drive in a dedicated subfolder under his/her name.
Crystallography and cryo-EM data will be backed up initially after collection (ESRF/Diamond) and also locally at the KU 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

K-drive: up to 100 GB per PI; J-drive: up to 1 TB per PI; L-drive: up to 10 TB per PI. Extensions of this volume can be asked for at ICTS at all times (for an additional cost).
For physical storage, there is currently enough capacity within the PIs' labs.
2 dedicated NAS systems with 180Tb total capacity.

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

For the K-drive: only the PI has writing access on this drive, i.e. can add data, but not modify them; this drive is read-only for lab members only (password-protected)
Lab Dropbox: access password-protected
Paper lab notebooks are kept in locked cabinets; digital lab notebooks: password-protected.

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

A small data management budget was applied for and granted in the project to manage extra costs; if this would appear not sufficient (e.g. in case of repair costs, or replacement of a physical storage device, or replacement of a PC is required), we will try to cover the costs from the consumables budget (or by matched funding)
Archive drive costs: 20€/year/100 GB; daily back-ups.
Dropbox: free
External hard disks: max. 1500€ (=5 disks of 1-2 TB)

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)?

- Large Volume Storage (longterm for large volumes)
- Other (specify below)
- Shared network drive (J-drive)

- KU Leuven RDR

K-drive
Physical: cryotheek, repositories, -80°C, -20°C

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

K-Archive drive costs: 20€/year/100 GB
External hard disks: max. 3000€ (=10 disks of 1-2 TB)
As no budget can be spent after the current project has ended, these costs will be covered by other funds of the PIs.

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.

- Other (specify below)
- Yes, as restricted data (upon approval, or institutional access only)
- Yes, as open data

Data sharing with peers will exclusively occur through publications, taking into account the Open Access policy of KU Leuven. The planned depositions of physical data in the relevant responsible repositories (as described above) will only occur *after* publication. In principle, we do not plan to share any unpublished data with people who are not directly connected with the labs of the contributing PIs (these data will internally be stored and treated as explained above). We might consider using the RDR database for preliminary (unpublished) data deposition, if appropriate, e.g. to serve for future project applications allowing continuation of the work and eventually, publication of the data.

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

Publications and repositories mentioned: all open access.
For unpublished data: only the PIs involved, or their scientific collaborators who will continue and follow up on the research after the project: they will will be granted access to the J-; L- and K-drives (password-protected)

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, intellectual property rights

We might have to embargo access to certain data, pending prior agreements on how IP rights will be secured or managed, e.g. before publication or before deposition in public repositories

Where will the data be made available?

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

- Other (specify below)
- KU Leuven RDR (Research Data Repository)
- Other data repository (specify below)

Data will foremost be made available upon open-access publication in peer-reviewed journals

When will the data be made available?

- Upon publication of research results
- Other (specify below)

For some data, full sharing may have to be postponed until IP protection has been put in place.

Which data usage licenses are you going to provide?

If none, please explain why.

- CC-BY 4.0 (data)
- Data Transfer Agreement (restricted data)

We anticipate that the project may generate data and results with potential for intellectual property (IP) protection. Therefore, data will not be immediately released under an open license. Instead:
Initially, access to specific datasets will be governed by a Data Transfer Agreement (DTA) to protect potential patent applications, confidentiality, or third-party rights. This allows us to manage the timing of data release and ensure compliance with institutional IP policies.
If patentable inventions arise, we will follow standard procedures for protection prior to publication or public data release. Once protection is secured, data may be made available under CC-BY 4.0, unless further restrictions apply.

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

Yes, a PID, such as a DOI, will be assigned upon deposit of the dataset in a certified data repository.
This will ensure the long-term findability, accessibility, and citability of the data in accordance with FAIR principles. If the dataset is associated with a publication, we will ensure that the DOI is included in the data availability statement.

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

The related costs are expected to be minimal, and can be covered by matched funding of the PIs involved

Responsibilities

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

All researchers involved in data collection, *i.e.* post-docs (Dr. R. Derua, Dr. E. Osipov, Dr. R Sewduth), PhD students (M. Butyrin, E. Van Boxel, Y. Ma) and technicians (F. Withof, S. Beelen, B. Lechat).
The final responsibilities lies with the PIs of the contributing labs.

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

All researchers involved in data collection, *i.e.* post-docs (Rita Derua, Raj Sewduth, E. Osipov), PhD students (Maria Butyrin, Emiel van Boxel, Yan Ma) and technicians (Fabienne Withof, Benoit Lechat, Steven Beelen).
The final responsibilities lies with the PIs of the contributing labs; this particularly applies to archiving data on the K-drive, and

deposition of data in repositories.

Who will manage data preservation and sharing?

All contributing PIs.

Who will update and implement this DMP?

Prof. Veerle Janssens, promotor of the current C1 project