Mapping radiation-induced plasticity in glioblastoma: from fundamental insights to therapeutic opportunities.

A Data Management Plan created using DMPonline.be

Creator: Lou Van Eupen

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Grant number / **URL:** 11B2323N

ID: 193511

Start date: 01-11-2022

End date: 31-10-2026

Project abstract:

Glioblastoma (GBM), the most malignant primary brain tumor in adults, is treated by maximal resection and radiochemotherapy. Despite numerous clinical trials, median overall survival rarely surpasses 2 years. Our inability to treat GBM has been attributed to interpatient and intratumoral heterogeneity, while the stem cell-like behavior and GBM plasticity have been correlated to therapy resistance, including radiotherapy. Investigating the molecular characteristics of individual GBM samples, how they respond to radiotherapy and the mechanisms they employ to confer resistance, could provide important insights to improve therapeutic strategies. In this project, I will, across a cohort of patient-derived models that represent the mainstay of GBM geno- and phenotypes, map the mechanisms of GBM plasticity that can bypass sensitivity to various types of radiotherapy. I will single-cell profile various models to which I will apply both photon (low-LET), and carbon irradiation (high-LET) which has been suggested to be more efficacious to target glioma stem cells. Using CRISPR screening, I will moreover search for targetable pathways that can modulate plasticity, and
this against the heterogeneous background of the GBM patient population, insights that I will eventually validate in PDX mouse models. Overall, this project will draw the landscape of radiation-induced plasticity across the GBM tumor cell and patient population and identify the molecular mechanisms that underlie plasticity.

Last modified: 22-03-2023

Mapping radiation-induced plasticity in glioblastoma: from fundamental insights to therapeutic opportunities.
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)

For this project, we will use existing data that has been collected by the LPCM lab and we will generate new data. For the PDCLs, patient and tumor information is stored in a MySQL database. Raw transcriptomic and proteomic data will be stored as FASTQ files, text and Excel files whereas analyzed data will be stored as VCF files. Generated graphs, statistical results and used codes are stored as Prism files and R scripts. In addition, notes will be kept in an online lab book. Results obtained through this project will be communicated via peer-reviewed scientific papers and more general press releases/presentations.

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)

1. Designation of responsible person (If already designated, please fill in his/her name.)

Asier Antoranz Martinez is the designated person for DMP.

1. Storage capacity/repository
    - during the research
    - after the research

Labcollector is used for data storage. This is a database hosted on a password-protected KUL server, automatically backed up by KUL services. All data is traceable and stored beyond the 5-year requirement. Patient-derived cell lines are stored in cryogenic tanks (liquid nitrogen) and extracted materials in freezers of -20 °C or -80 °C. Storage is only limited by the size of the storage rooms. Generated data are also stored on SCK CEN protected servers (3 Tiers, RAID 6 level with double parity disks).

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)

NA

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)

For the generation and use of patient-derived cell lines an informed consent is used for every patient which includes the duration of data preservation. Personal patient information, however, is destroyed for the use of these cell lines to respect privacy legislation. All data is stored in a strictly secured environment.

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

Mandatory lab notebooks are kept by all staff at KUL and SCK CEN and animal welfare is documented in designated servers (e.g. ANIBIO).

Mapping radiation-induced plasticity in glioblastoma: from fundamental insights to therapeutic opportunities.
DPIA

DPIA

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

Question not answered.

Mapping radiation-induced plasticity in glioblastoma: from fundamental insights to therapeutic opportunities.
GDPR

GDPR

Have you registered personal data processing activities for this project?

- No

Mapping radiation-induced plasticity in glioblastoma: from fundamental insights to therapeutic opportunities.
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
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .cvs,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Clinical database of included patients UZ Leuven	Both retrospective and prospective data, where the data is manually extracted from patient files (KWS). Decoded personal data, clinical parameters, pathology information and genetic information (if available) and longitudinal followup. KWS data is stored for at least 25 years, all patient data has been anonymized (TSSxx number, GBMxx number)	• Reuse existing data	• Digital	• Observational • Experimental	• .cvs, • .txt	• <1GB	
Clinical database of included patients from external hospitals	Comparable clinical characteristics and pathology/genetic information from GBM patients in Maastricht (MUMc) and Genk (ZOL).	• Reuse existing data	• Digital	• Observational • Experimental	• .cvs, • .txt	• <1GB	
generation of patient-derived cell lines (PDCL)	Cell lines derived from fresh tumor biopsies will be generated to perform the described experiments. Cell lines are stored in liquid nitrogen. Most of the cell lines were already generated beforehand and stored in the available biobank (Leuven Living Tissue Bank).	• Reuse existing data	• Physical				Cells are stored in cryovials in liquid nitrogen.
Microscopy	Images of immunofluorescence stained slides. Analysis carried out using Excel, Graphpad, ImageJ (Fiji).	• Generate new data	• Digital	• Experimental	• other: .nd2, .tif	• <5TB	
Bulk RNA-sequencing	RNAseq data for the optimization of experiments and characterization of PDCLs with vs. without treatment. Analysis carried out using RStudio.	• Generate new data	• Digital	• Experimental	• .abi, • .count • .fastq, • .xls, • .pdf • .R	• <1TB	
Single-cell RNA-sequencing	scRNAseq data for the characterization of PDCLs with vs. without treatment. Count tables and associated metadata files will be analyzed in UniAPP, developed by Unicle. Analysis will consist of QC, normalization, clustering, dimensionality reduction, etc.	• Generate new data	• Digital	• Experimental	• .abi, • .count • .fastq, • .xls, • .pdf • .R	• <1TB	
qPCR	Validation of scRNAseq expression data, optimization of experiments. analysis carried out using Excel, Graphpad and/or RStudio	• Generate new data	• Digital	• Experimental	• .cvs, • .txt • .xlsx • .R • .pzfx	• <100GB	

Viability data	Viability data generated using CellTiter-Glo for functional characterization and identification of treatment sensitivities. Analysis carried out using Excel, Graphpad and/or RStudio.	• Generate new data	• Digital	• Experimental	• .cvs, • .txt • .xlsx • .R • .pzfx	• <100GB	
Proteomic data	Proteomic data generated using CyTOF to characterize PDCLs with vs. without treatment + with vs. without gene editing (CRISPR/Cas9.	• Generate new data	• Digital	• Experimental	• .fcs, Flowjo v.10 software	• <100GB	
CRISPR/cas9 editing	Gene-edited cell lines using the CRISPR/cas9 technology. Stable Cas9-expressing cell lines have been generated by the REGA institute. Constructs and libraries will be generated in collaboration with REGA.	• Generate new data	• Digital	• Experimental	• .abi, • .fastq	• <100GB	
Incucyte (real-time imaging)	Cell imaging and analysis platform that enables quantification of cell phenotype over time	• Generate new data	• Digital	• Experimental	• IncuCyte.exe files	• <100GB	
Pathway mapping, GSEA, GO enrichment analysis, gene regulatory network analysis	Based on scRNAseq data, involved pathways will be investigated further using GSEA, SCENIC, GO analysis, ...	• Generate new data	• Digital	• Experimental	• .abi, • .count, • .fastq, • .xls	• <100GB	
PDX mouse models, FFPE tissue blocks, frozen material	Orthotopic and PDCL injections for the generation of PDX mouse models in immunodeficient mice. This will be done in collaboration with the KUL PDX platform (TRACE).	• Generate new data	• Physical				Tissues are stored in liquid nitrogen
Bioluminescence and MRI imaging	Follow-up of tumor growth in PDX mice.	• Generate new data	• Digital	• Experimental	• DICOM	• <100GB	
In vivo treatment repsonse of PDX mice	Kaplan Meier survival analysis of in vivo treated PDX mice. (sc)RNA-seq on ex vivo an in vivo treated PDX tumor samples, CyTOF on ex vivo and in vivo treated PDX tumor samples.	• Generate new data	• Digital	• Experimental	• .csv, • .xls, • Graphpad • .fastq, • .abi, • .count, • .pdf	• <5TB	

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:

Coded patient-related information including: clinical follow-up, survival data, treatment information, genetics and pathological features is collected and samples are anonymized (TSSxx, GBMxx). As part of prior research, extensive genetic profiling has also been done already. All obtained research data points and clinical information will be added in a coded manner to the LabCollector/Glioma2015 database of the Translational Cell and Tissue Research group, located on a KU Leuven hosted and secured server and is password protected. A dedicated, trained person will add all genetic and research information from this project to a local, KULeuven based Bioportal database. Only coded information will be extracted and used for the downstream research analyses. Personal data will only be kept as long as necessary for the research activities, but is available in the clinical files (KWS) for at least 25 years (legal obligation).
Our laboratory has established a library of patient-derived cell lines (PDCLs) derived from fresh glioblastoma tumor samples (collaboration with Prof. S. De Vleeschouwer, UZLeuven). These PDCLs have been characterized genomically and transcriptomically. For the project, only cell lines will be used of already-deceased patients, so that no additional requests for informed consent will be necessary. We do not expect that this will lead to biased data. In this project, new findings will be linked to the available data points of the used models.

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
- Yes, animal data

In this project, we will use PDCLs derived from high-grade brain tumors, which were developed in the framework of study S59804 and S61081, both approved by the Ethische Commissie Onderzoek UZ/KU Leuven. A selection of glioblastoma cell lines will be used in this project for transcriptomic, proteomic and functional characterization. Only cell lines are included from which the patient is deceased, so that no additional requests for informed consent will be necessary.
Ethical approval for the project has been submitted to the EC, CTC and Biobank (S64820) and **pending approval.**
For the PDX mouse models, ethical approval will be submitted to the Ethical Committee for Animal Experimentation (ECD).

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.

- No

All patient information is anonymized. Only cell lines are included from which the patient is deceased, so that no additional requests for informed consent will be necessary.

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

There is potential tech transfer/valorization in both the novel models we develop (genetic, proteomic and heterogeneity profiles with clinical information and drug responsivity) and the putative

assays/biomarkers we measure. Depending of the translatability and strength of the generated models, they could possibly be used for broader research endeavors in the future. We are in touch with LRD regarding these matters.

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

A research collaboration agreement between KU Leuven and SCK CEN is in place. Developed methodologies and their results will be shared. PDCLs generated in LPCM will be transferred to SCK for maintenance, a material transfer agreement is in place.
Collaborations with other parties (VIB, REGA, GSI, TRACE, ...) will not restrict dissemination of data, provided that there is proper acknowledgement. The data will be published before making it partly available. Our goal is to fuel collaborations where our data and materials can be used.

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.

- Yes

Ownership of the PDCLs and their data belongs to LPCM. Methodologies developed by collaborators (VIB, REGA, TRACE, GSI, ...) are considered as their intellectual property.

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).

At SCK CEN:
Protocols, research progress and obtained data, will be collected in the data management software (Alexandria). In Alexandria, each item receives its own identifier which is used throughout the progress reporting. A physical sample inventory will be stored in freezers and fridges. A file with sample details will be saved on Alexandria.
At KUL:
All collected data points (i.e., cell line information, diagnosis, clinical information, genetic
information, proteomic information, etc) are stored in a coded manner in LabCollector (an online labmanagement system with integrated Electronic Lab Notebook (ELN)) which runs on a secured and backed up server of KULeuven (managed by ICT of the Biomedical Sciences Group). All patient information is anonymized. Every patient will receive an identification number which can only be decoded by the responsible data manager. This system also provides a logging system so no data can ever be erased, making that everything will be traceable and stored long-term (> 5-years). Only coded information will be extracted and used for the downstream research analyses. Patient material, patient-derived cell lines and PDX mouse tissues are and will be stored in -80°C freezers or liquid nitrogen, where the exact location is defined in labcollector. Here too, protocols, research progress and obtained data, will be saved on the KUL server.
Digital documentation will be supplemented by a physical lab book where important notes, protocols, experimental details will be noted in such a way that others could understand and apply this information as well. These lab books are owned by the research group and remain in the laboratory during the entire time.

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.

- No

AT KUL:
All data is searchable and includes various levels of metadata. The LabCollector platform is structured according to projects and topics where all relevant information is directly linked to each experiment. The integration of a lab inventory to the ELN system in LabCollector, makes that information can be retrieved both from the experiment point of view or from the sample point of view (e.g., which experiments have been performed with cell line A, or which cell lines were used in experiment B).
RNA/DNA seq and CyTOF data: metadata about raw fastq-files will be maintained in CSV-files. These files will include data like sample ID, date, treatment, timing, name of creator, etc.
These metadata files will be saved in the same folder as the raw data files to facilitate reuse in the future.
qPCR data: we will conform to the MIQE (minimal information on qPCR experiments) guidelines (Bustin et al., Clin. Chem., 2008). In addition, a CSV file containing information on sample ID, date, treatment, timing, name of creator, etc., will be saved in the same folder as the raw data files to facilitate reuse in the future.
At SCK CEN:
The data management software Alexandria utilize three aspects when generating the metadata of each unique identifier:

- Classification: the classification divides documents into particular groups, based on the same characteristics. A classification may be created when documents have certain templates, workflows, policies or attributes which are necessary to identify them into the system.
- Categories: a category is a group of attributes. The attributes are grouped by a specific topic, e.g. all attributes related to events are grouped into the category Event Attributes.
- Attributes: attributes are additional information fields, which gives the document a certain context: e.g. the author of the document.

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

Where will the data be stored?

Biological material will be stored under the Biobank of KU Leuven and Biobank of SCK CEN.
The extracted DNA/RNA/Protein/snap frozen materials are stored in -20°C/-80°C freezers. Fixed materials are stored in designated closets in the lab. Storage locations are included in the labcollector/Alexandria database to keep track of each sample.

How will the data be backed up?

At KU Leuven:
Labcollector is a database hosted on a password protected KUL server, automatically and daily backed up by KUL services. The vast majority of the generated data is stored in 2 different places, e.g., external hard drive and KUL server, external hard drive and NAS. In case of data loss, one can contact IT to retrieve the back-upped data. All data are traceable and stored beyond the 5-year requirement.

At SCK CEN:
All data is hosted on the SCK CEN protected servers (3 Tiers, RAID 6 level with double parity disks). Only SCK CEN personnel with SCK CEN computer can have access to the server (access right are limited depending on the hierarchical position). All data are traceable and stored beyond the 5-year requirement.

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

Yes, we have unlimited space available – payable on yearly basis. Storage of cell lines/biological samples is limited to the size of our storage rooms. However, we are planning to move the less relevant materials to the UZLeuven Biobank from the moment it will be ready to accept materials (will be for free initially).

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

At SCK CEN:
Data storage are secured by the ICTS service of SCK CEN. Access rights are managed by the data management software Alexandria prior the requirements. The project folder can be found in the Unit repertory (accessible by the lab members). The content of the project folder is however limited to the personnel working on the project and the head of the Unit.
at KUL:
Secured KULeuven server with automatic back-up, password access by users and person-based decision on rights to access and modify data. In addition, every modification to data is logged and no data can be erased. In case necessary, Belnet filesender will be used for a secure transfer of files.

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

At SCK CEN:
Data storage and backup costs are managed at the Unit level (not at the project level). Each stored TB cost € 35 per month + € 200 maintenance costs/year. The estimated maximal cost for the 4-year project would therefore be € 5000 (assuming the generation of 1TB data/year) and is covered by the Unit internal funds.
At KUL:
<1000 euro/year for server space and backup. Labcollector has been purchased (only updates required), where general lab budget is foreseen to keep this system up and running. Other costs are currently connected to the pricing model of KULeven for large volume and cold storage. External hard drives for temporary storage are bought as deemed necessary (20 Tb per drive, +- 550 euro).

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 retained for at least five years after the end of the project.

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

1) The digital data will be stored on the university's central servers (with automatic backup procedures) for at least 5 years, conform the KU Leuven RDM policy. SCK CEN operates similarly with the obtained data on their servers.
2) The physical data will be stored in freezers/cryogenic tanks/... for at least 10 years after the project.
3) The accompanying metadata will be stored in Labcollector (KUL) and on Alexandria (SCK CEN).

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

Costs directly depend on the pricing of storage drives of KULeuven for large volume and cold storage. These costs will be covered through the general lab budget (non-related grant applications). As stated before, we except <1000 euro/year for server space and backup, where Labcollector has been purchased (only updates required). The Translational Cell and Tissue Research Unit also decided to invest in this system for our entire group. In addition, the NAS system is up-and-running and does not require an annual fee. At SCK, assuming a total of 4 TB, the total cost will be € 1680 per year. The costs are covered by the Unit internal funds.

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, …)
- Yes, in an Open Access repository

Relevant data will be made publicly available post publication depending on the journals policy (post-publication data repository):
1. Raw sequencing data (fastq files, metadata) will be made available upon publication by upload to the EBI European Genome-phenome Archive (EGA).
2. Raw datapoints originating from e.g. viability experiments, survival analysis, qPCR (Ct values), will be made available upon publication through upload of an Excel file.
3. Scripts will be made available on GitHub upon publication.
Non-published data will remain confidential and only accessible upon request. We will comply with open access regulations of FWO.

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

Upon request, data will be available in a collaborative setting (i.e. any other internal and external research group with whom we may work in the future that could benefit from data and materials gathered in this project):
1. Source data files underlying figures in published papers will be available to anyone for any purpose, provided that they give appropriate credit to the creators. The exact license will depend on the journal where the data will be published (e.g. Creative Commons Attribution (CC-BY)).
2. All raw data (e.g. sequencing data) underlying published study results and uploaded to public repositories, will be available to anyone for any purpose, provided that they give appropriate credit

to the creators. The exact license will depend on the journal where the data will be published (e.g. Creative Commons Attribution (CC-BY)), and the type of repository (e.g. EBI EGA).
3. All other data can be obtained in the context of a collaboration, after signing a data sharing agreement, which will be established with the support of KU Leuven LRD. In such a case, data will be made available on basis of material transfer agreements (MTAs) if the objectives fulfill the aims referred in the informed consent signed by the donors of the biopsies and after approval by the UZLeuven ethical committee, if requested by a third party. The contents of this MTA will be determined according to partner type (academic or private), and will ensure that IP is protected at all times. As such, data that will be shared with third parties will exclude commercial use and will require appropriate credit to the data owners. Detailed data sharing agreements will therefore be implemented.

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.

- Yes, Intellectual Property Rights

A research collaboration agreement between KU Leuven and SCK CEN is in place. Developed methodologies and their results will be shared. PDCLs generated in LPCM will be transferred to SCK for maintenance, a material transfer agreement is in place.
Collaborations with other parties (VIB, REGA, GSI, TRACE, ...) will not restrict dissemination of data, provided that there is proper acknowledgement. The data will be published before making it partly available. Our goal is to fuel collaborations where our data and materials can be used.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

Gene expression data will be deposited at the Gene Expression Omnibus.

When will the data be made available?

Upon publication of research results and/or after signing a data sharing agreement which will be established with the support of KUL R&D after a request.

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

Data from the project that can be shared will be made available under a creative commons attribution license (cc-by 4.0), so that users have to give credit to the original data creators.

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

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

The expected data sharing costs are minimal and covered by university services (at KU Leuven) or by the research Unit (at SCK CEN). Data shared through journal repositories will be covered by publication costs.

6. Responsibilities

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

I will have final responsibility for data documentation, if needed, guided by mentors or administrative personnel.

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

under the supervision of mentors/promotor, I will be responsible for data storage. The ICTS service of KU Leuven and SCK CEN is responsible for the back-up of the network drives at KU Leuven. The folders will be managed by the supervisors.

Who will manage data preservation and sharing?

My PI and I will be responsible for the data preservation and sharing of obtained data.

Who will update and implement this DMP?

The PIs bears the end responsibility of updating & implementing this DMP