Mapping single-cell drug responses in Glioblastoma using a multi-omics approach

Project Name FWO DMP - Mapping single-cell drug responses in Glioblastoma using a multi-omics approach

Project Identifier G0B3722N

Grant Title G0B3722N

Principal Investigator / Researcher Frederik De Smet

Project Data Contact Basiel Cole, basiel.cole@kuleuven.be, +32473641837

Description Glioblastoma (GBM) remains the most malignant primary brain tumor. In spite of intensive treatment, current approaches are insufficiently effective to achieve major clinical benefits. GBM typically exhibits complex genetic aberrations which lead to extensive inter- and intratumoral heterogeneity. The presence of exceptional responders across many failed clinical trials suggests that GBM consists of dozens of small subgroups that likely all require a different therapeutic approach. Because this functional heterogeneity remains largely unexplored, this project aims at mapping single-cell drug responses across GBM tumor types. To achieve this, we will compare control- and drug-treated tumor samples using an ex vivo multi-omics approach combining transcriptomic and protein-based single cell measurements. As such, we will not only be able to define responsive and resistant tumor cell (sub)populations, it will also allow us to identify potential biomarkers for each included therapy. Appropriate experimental conditions and biomarkers will first be identified using a heterogeneous library of patient-derived GBM cell lines. Next, we will use mouse models of GBM to correlate the ex vivo and in vivo response capabilities at single cell level, to finally analyse these features in freshly resected clinical tumor samples. This project will as such describe the landscape of drug responsiveness across GBM at unseen resolution, and provide a fundamental step towards more tailored therapies.

Institution KU Leuven

1. General Information

Name applicant

Frederik De Smet

FWO Project Number & Title

FWO Project Number: G0B3722N

Title: Mapping single-cell drug responses in Glioblastoma using a multi-omics approach

Affiliation

- KU Leuven

2. Data description

Will you generate/collect new data and/or make use of existing data?

- Generate new data
- Reuse existing 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).

Work package	Type of data	Format	Volume	How created
WP1	Dose-response results	.csv, .xls, Graphpad Prism	500 MB	Cell-Titer Glo Viabilty assay, performed in triplicate on n=15-30 PDCL. Fluorescence values will be obtained with a SpectraMax iD3/5 machine. Data will be analyzed in Graphpad Prism.

WP1	Western blot images	.jpeg, .tif	1 GB	Further validation of dose-response results (e.g. by measurement of apoptotic markers).
WP1	ELISA	.xls	250 MB	Further validation of dose-response results (e.g. by measurement of apoptotic markers).
WP1	DNA-seq + RNA-seq	.fastq, .vcf, .count,	5 TB	Bulk DNA- and RNA-seq on PDCL (WP1) and on human tumor samples (WP2) for genomic/transcriptomic characterization.
WP1 + WP2 + WP3	Single-cell RNA-sequencing + barcoding (MULTI-Seq)	.abi, .fastq, .xls, .count, .pdf	>5 TB	Single-cell RNA-seq on n=5 resistant and n=5 susceptible PDCL per drug (n=20-30, WP1), scRNA-seq on human ex vivo tumor samples (WP2), scRNA-seq on ex vivo tumor samples derived from PDX mice (WP3), sequencing on Illumina NovaSeq, data will be demultiplexed, read-count tables will be generated, associated metadata files will be generated.
WP1 + WP2 + WP3	Intermediate and final scRNA-seq analysis files	.xls, .pdf, .jpeg, .svg	50 GB	Count tables and associated metadata files will be analyzed in UniAPP, developed by Unicle. Analysis will consist of QC, normalization, clustering, dimensionality reduction, etc. Intermedite analysis files will be saved in .jpeg, .pdf, .xls files.
WP1 + WP2 + WP3	CyTOF	.fcs	50 GB	CyTOF measurement of phenotypic and drug response markers on resistant and susceptible PDCL (WP1), CyTOF measurement of phenotypic and drug response markers on human ex vivo tumor samples (WP2), CyTOF measurement of phenotypic and drug response markers on ex vivo tumor samples from PDX mice (WP3), Validation of scRNA-seq data (WP1 + WP2 + WP3)
WP1 + WP2 + WP3	qPCR	.csv	1 GB	Validation of scRNA-seq data (measure expression of differentially expressed genes as identified by scRNA-seq)

WP2	Database of primary tumor biopsies and patiend-derived cell lines	.xls, text fields in MySQL database, flat-text in electronic lab notebook (LabCollector)	10 GB	Collection and biobanking of tumor biopsies and derived cell lines: cryopreserved cells/tissue slides will be biobanked in liquid nitrogen or at -80C. Positions, characteristics and other (clinical) metadata will be stored in a relational MySQL database. The collection of samples will be documented in an electronic lab notebook, implemented in LabCollector.
WP3	Multiplex IHC (MILAN), spatial proteomics	.czi, .jpeg, .xls, .tiff	>10 TB	Immunofluorescence measured by a ZeisAxionscan, performed on brain tissue slides obtained from PDX mice, Excel files with calculations for the antibody staining, screenshots for quality control.
WP3	Data analysis MILAN	.tiff, .ometiff, .Rmd, .html	>20 TB	MILAN data analysis pipeline: intermediate data files, R work environments and final data files will be generated.
WP3	In vivo and ex vivo drug exposure analysis in PDX mice	.csv, .xls, Graphpad Prism, .fastq, .abi, .count, .pdf	>5 TB	Kaplan Meier survival analysis of in vivo treated PDX mice, scRNA-seq on ex vivo an in vivo treated PDX tumor samples, CyTOF on ex vivo and in vivo treated PDX tumor samples

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.

- Yes

Privacy Registry Reference:

- G-2021-3789
- G-2021-3895

Short description of the kind of personal data that will be used:

- Clinical data will be collected and coded: age, gender, cancer-related prior history, date of surgery, location of the tumor, clinical imaging parameters (e.g. from MRI), pathological diagnostics, overall survival, disease-free surival, progression free survival, date of tumor progression/relapse, treatment regimen and response, date of decease. All patient information is anonymized (TSSxx, GBMxx). As part of prior research, extensive genetic profiling has also been done already, while further genetic testing will be performed. Importantly, 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 that is password protected. A dedicated, trained person will add all genetic and research information from this project to a local, KULeuven based cBioportal 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).

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)

- Yes

Ethical approval for work packages 1-3 has already been obtained from the Medical Ethical committee from UZLeuven/KULeuven, including informed consent for every patient. No ethical issues concerning research data are to be expected. The goal of the project is to collect data beyond 5 years and data collection, data sharing and long-term preservation has been defined in the informed consent.

Ethical approval numbers

- S59804
- S61081
- S64700

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?

- 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 dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- No

No third-party agreement restrictions are in place as we will perform all assays in-house and data will mostly be analyzed using in-house pipelines.

4. Documentation and metadata

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

1. All collected data points (i.e., cell line information, diagnosis, clinical information, genetic information, proteomic information) are and will be stored in a coded manner in LabCollector. This is a queryable database that runs on a secured and backed-up server of KULeuven (managed by ICT of the Biomedical Sciences Group). LabCollector includes an electronic lab notebook (ELN), which will be used to document all detailed experimental procedures performed throughout the project. Researchers that would like to reuse data generated during the project will be able to find all necessary information regarding experimental conditions, notes and additional metadata by querying the LabCollector database.

2. Standard operating procedures (SOP) and protocols are and will be available on the LabCollector platform. Clear reference to these protocols (each protocol has a distinct ID) will be noted on an experiment-by-experiment basis in the ELN implemented in LabCollector, facilitating reproducibility of the experiments.

3. Raw, intermediate and final data originating from sequencing experiments (e.g. fastq files, BAM files, VCF files, etc.) will be stored on external hard drives for short term/quick access storage, as well as KU Leuven K-drive and L-drive (large storage servers) and a NAS (network-attached storage) coupled to the KUL storage servers for back-up and long-term storage. All file

names will contain reference to a sample ID, date, project and experimental conditions, so that they can unambiguously be linked to associated data (protocol, workfiles, metadata, etc.) stored in LabCollector.

4. Microscopy images (MILAN) will be clearly labeled to identify experiment, sample, and date. The labeling of every individual image taken is standardized and defined in a standard operating procedures (SOP) lab protocol. For the image analysis, the algorithm used for the software analysis will be noted in an Excel database per experiment with the info about the relevant parameter changed.

5. All patient information (age, sex, disease) will be registered in a pseudononymised way in the file containing all collected samples. 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 tracable and stored longterm (way beyond the common 5-year requirement). Only coded information will be extracted and used for the downstream research analyses.

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

1. 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).

2. 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.

3. 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.

4. Metadata concerning the technical specifications of the multiplex IHC images (e.g., channel/gating...) are collected with the .czi files, or else will be noted down with each experiment.

5. Data storage and backup during the FWO project

Where will the data be stored?

1. The data generated during the project will be safely stored in our research unit central storage facility. Copies can be made and kept on personal devices. All obtained research data points will be kept in electronic lab notebooks in the LabCollector database of the Translational Cell and Tissue Research group, located on a KU Leuven hosted and secured server (password protected).

2. Large data sets, such as images from microscopy and sequencing data, will be stored on the KU Leuven K-drive and L- drive (large storage servers). In addition, external hard drives for temporary storage are bought as deemed necessary. On top of the large storage servers from KU Leuven, we regularly expand the longterm storage capacity with a user-friendly NAS (network-attached storage) coupled to the KUL storage servers.

3. Sensitive personal data will be anonymised throughout the project (coded). The data will be stored in the KULeuven secure environment for private data.

How is backup of the data provided?

The KULeuven K- and L-drives are maintained and backed-up daily by the IT department. In addition, to ensure redundance in case of data loss, all data will be stored in at least 2 different locations (e.g. K-drive and NAS, or NAS and physical drive). In case of data loss on the KULeuven server, data can be restored from a back-up by the IT department.

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

Our research group maintains a NAS server with ample storage space to meet the requirements of the project. In addition, physical hard drives (ranging in capacity from 8-32 TB) are readily available and will cover the requirements of the project.

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

As mentioned above, ample storage is readily available. LabCollector has previously been purchased and does not require annual payment. However, additional physical storage can be purchased as deemed necessary (approx. €550 for 20 TB).

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

Controlled access will be ensured through the following measures:

1. Personal data will be anonymised throughout the project (coded). The data will be stored in the KULeuven secure environment for private data, which is password protected. Only authorized persons (physicians) will be able to access non-anonymised personal data through Redcap (as implemented at UZ Leuven). These data will be extracted from Redcap by the authorized persons, anonymised through coding (by addition of an additional data field with a sample ID) and saved in Excel files, which can be accessed by the researchers.

2. All data will be stored on either the KULeuven data servers (K- and L-drive) and/or our NAS system. Both systems are password protected and can only be accessed from within the KULeuven network (or via VPN).

3. Two-factor authentication is required in order to access the KULeuven network.

4. KULeuven servers are backed-up daily, ensuring that data cannot ever be erased by unauthorized persons, or by accident.

5. Personal devices (laptops) are password protected.

6. Where necessary, Belnet filesender will be used for secure file transfer.

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 research data generated will be retained for at least 10 years after the project, thus complying to the data preservation rules of KULeuven. Data will be stored in a safe, secure and sustainable way for purposes of reproducibility, verification and potential reuse.

Where will the data be archived (= stored for the longer term)?

The data will be stored on the university's central servers (with automatic, daily back-up procedures) for at least 10 years, conform the KU Leuven RDM policy. Data is also stored on Labcollector, hosted on a secured KUL server with password-protected access. The server is automatically backed up using KUL services. The system provides a logging system so no data can ever be erased and everything will be tracable and stored longterm (beyond the 5-year requirement).

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

As stated before, we expect <1000 euro/year for server space and backup, as Labcollector has already been purchased and does not require an annual fee. The Translational Cell and Tissue Research Unit also decided to invest in this system for our entire group. In addition, our NAS system is up-and-running and does not require an annual fee.

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

There are no concrete factors restricting or preventing the sharing of produced/collected data as defined with a 3rd party or legal restriction. The informed consent includes coded data sharing with both academic and non-academic parties. The data will be published before making it partly available. Our goal is to fuel collaborations where our data and materials can be used. Considering the use of patient materials, data will be either shared in a coded manner through publications, or, if more specific information is required, following ethical approval. Data sharing can also happen with specific access controls in place, for example only allowing reuse for research purposes.

Which data will be made available after the end of the project?

1. Raw sequencing data (fastq files, metadata) will be made available upon publication by upload to the NCBI Gene Expression Omnibus (GEO).

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.

4. Non-published data will remain confidential until a final decision on publication of the data has been taken. All 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).

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

- In an Open Access repository
- In a restricted access repository

Depending on data type and/or requirements by journals, data will either be uploaded to Open Access or closed access repositories.

Published data will be uploaded to Open Access repositories, according to journal requirements:

1. Raw sequencing data (fastq files, metadata) will be made available upon publication by upload to the NCBI Gene Expression Omnibus (GEO).

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.

All other data can be shared in a collaborative fashion after signing a data sharing agreement that will be established with the support of KUL R&D after a request. Once KU Leuven has established a university managed and owned data repository for sharing of data (or a subset of data) on this repository, we will evaluate data sharing and data reuse depending on the policy and conditions of this repository. The conditions of access will be determined by the need of the third party that wants access (e.g., only data, only materials, industrial or academic party, etc). External users will have access to our database through our publications and after setting up a new collaboration. For academic users the requirement will be to be included as co-authors in publications where our data and materials are used. For industrial users, the conditions will be determined in collaboration with LRD.

When will the data be made available?

- Upon publication of the research results

Published data will be uploaded to Open Access repositories, according to journal requirements:

1. Raw sequencing data (fastq files, metadata) will be made available upon publication by upload to the NCBI Gene Expression Omnibus (GEO).

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.

Who will be able to access the data and under what conditions?

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. Genbank, GEO).

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.

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

The KU Leuven repository will not request any cost contribution for KU Leuven researchers. Data shared through journal repositories will be covered by publication costs. Bilateral agreements for data sharing will be established through the services of KU Leuven R&D. The costs expected for data sharing are thus low and will be reported in the final DMP at the end of the project. They will be covered through funds available in the laboratory.

8. Responsibilities

Who will be responsible for data documentation & metadata?

As a post-doctoral researcher that will work full-time on this project, I (Basiel Cole), will be responsible for the documentation of data and metadata. If needed, help can be requested from technicians and administrative personnel.

Who will be responsible for data storage & back up during the project?

As a post-doctoral researcher that will work full-time on this project, I (Basiel Cole), will make sure that all data will be stored on servers that are backed-up. The responsibility for maintaining the infrastructure for data storage (maintenance of servers) lies in the hands of the ICT department of KU Leuven.

Who will be responsible for ensuring data preservation and reuse ?

Both myself (Basiel Cole) and my PI, Prof. Frederik De Smet, will take responsibility to ensure data preservation, access and reuse.

Who bears the end responsibility for updating & implementing this DMP?

The PI bears the end responsibility of updating & implementing this DMP.