Leveraging droplet microfluidics and nucleic acid engineering for development of a CRISPR-based cell imaging toolbox

A Data Management Plan created using DMPonline.be

Creator: Charlotte Van Tricht

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Grant number / **URL:** 1S56423N

ID: 197334

Start date: 01-11-2022

End date: 01-11-2026

Project abstract:

Fluorescence in situ hybridization (FISH) is the golden standard for DNA imaging in cells, but it involves lengthy experimental workflows and genomic DNA denaturation. CRISPR-Cas9 (stands for: Clustered Regularly Interspaced Short Palindromic Repeats and associated protein 9) technology, alternatively, has enabled specific cell imaging without the need for DNA denaturation. Still, all so far reported CRISPR-Cas9-based approaches require complex engineering and do not all provide signal amplification required for sensitive detection. In this project, we will utilize the power of microfluidics and nucleic acid engineering to develop novel tools for specific, sensitive, and multi-omic cell imaging. To achieve this, we will (1) develop a high-throughput droplet-microfluidic (DM) platform to (2) enable, in a completely innovative way, highly combinatorial engineering of a repertoire of CRISPR-dCas9 tools that (3) will be delivered into fixed cells for proof-of-concept multi-omic cell imaging. This novel CRISPR-dCas9 toolbox will be combined with transcriptome imaging, which together is crucial for answering a multitude of biomedical research questions in embryogenesis, aging, and disease.

Last modified: 15-03-2023

Leveraging droplet microfluidics and nucleic acid engineering for development of a CRISPR-based cell imaging toolbox
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)

The following types of new data will be generated in this project (defined per work package):
WP1 (Developing droplet-based Cas9 nuclease assay)

- Spectrophotometer data (raw data and analysis, .xcel), approximately 245kb/experiment, 4 years estimate = 49 MB
- Gel electrophoresis (raw data in .scn format and .tif format, data analysis in .xcel format), approximately 4Mb/experiment, 4 years estimate = 800 MB
- Cas9 nuclease assay in droplets (images as .tif format and data analysis in .xcel), approximately 2Mb/experiment, 4 years estimate = 410 MB
- Droplet microscopy images and movies or fluoroscence data, etc. (.tiff, .avi, .xlsx), 4-years estimate = 10 GB

WP2 (Developing non-catalytic DNA tools for signal generation)

- Nucleic acid designs (.dna, .txt or .fas), 4 year estimate = 10 MB
- Droplet microscopy images and movies or fluoroscence data, etc. (.tiff, .avi, .xlsx), 4-years estimate = 10 GB

WP3 (Developing catalytic DNA tools for signal generation)

- Nucleic acid designs (.dna, .txt or .fas), 4 year estimate = 10 MB
- Droplet microscopy images and movies or fluoroscence data, etc. (.tiff, .avi, .xlsx), 4-years estimate = 10 GB

WP4-WP5 (CRISPR-based fixed-cell imaging)

- High-resolution microscopy images of fixed cells (Z-stacking image series in .tif or .nd2 format and data analysis in .xcel), 4 years estimate = 10 TB
- High-resolution microscopy images of live cells (Z-stacking and/or time lapse image series in .tif or .nd2 format and data analysis in .xcel), 4 years estimate = 10 TB
- Multi-resolution processed and/or summarized microscopy data (*.tiff, delimited text files), 4 years estimate = 3 TB
- Algorithmically generated in-house gRNA sequence designs (.fasta format), 4 years estimate = 10 GB

We will not reuse existing data

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. The PhD student, Charlotte Van Tricht, working on this project will be responsible to store the data on the appropriate accommodation provided by KU Leuven. The ICTS service of KU Leuven is responsible for the back-up of the network drives at KU Leuven. The folders will be managed by the supervisors.
2. Storage capacity/repository
    - during the research
        - Data will be primarily stored on the PhD's personal KU Leuven OneDrive. This time-stamped digital data will be backed-up to an already created project folder on the shared drive (J:) of KU Leuven. The time-stamped digital metadata will be stored on the server of the electronic labbook (eLABJournal, Bio-ITech). The folder is open for all the staff that will be working on this project and is secured and backed-up by the ICTS service of KU Leuven. Copies can be made and kept on personal devices.
        - The network drive for the project shared folder and the large volume storage folder are secured by the ICTS service of KU Leuven with a mirror copy. Confidential data can and will be protected with a password (available only for PI Jeroen Lammertyn). Visitors, MSc thesis students and internship students in the groups as well as other unauthorized persons will not have access to the data on the shared folder. Data storage in the cloud will be avoided, unless for temporary use only, e.g., to transfer large files between the researchers involved in the project.
        - KU Leuven provides sufficient storage and back-up capacity during and after the project. A dedicated folder is made for the project to store data files.
        - Type 1 server back-end storage with mirror backup for the project shared folder will cost 57 Euro per Tb per year. Costs will be covered by the project consumables budget.
    - after the research
        - The data to be retained for the expected 5 years after the project's end are: dissemination data (source files of publications and presentations) and the most relevant measurement data.
        - The research data will be stored on an external hard drive after the end of the project. Dissemination data, namely files corresponding to papers and presentations, will be stored on the PCs of PI (J Lammertyn), and backed-up daily on the departmental server for long term storage.
        - The volume corresponding to dissemination data is expected to be relatively low (<10 GB), and therefore can be seamlessly embedded in the PI's allocation on the departmental server. The costs (1000 EUR/year) will be covered by other on-going projects at that point in time. For research data, at current archiving costs of 10 Euro/(TB*year), we estimate a cost of 2000 Euro/year. These costs will be covered by funding acquired by the project PI in the context of other research projects.

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)

There is no intention to deviate from the principle of preservation of data.
Cell samples on which experiments are executed will not be stored as they are subject to organic decay.

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)

Yes. We have already acquired The Ethics Committee Research UZ / KU Leuven (EC Research: S64403) approval for technology development on human cell lines. Remaining approvals in this project are for: fetal, adult, and ageing healthy tissues as well as diseased tissues, e.g. cancers; ECD approval for work with animal (e.g. mouse) embryos. We are in the process of acquiring these additional ethical approvals.
We will reuse sequencing data previously generated and will also generate new sequencing and/or spatial data from commercially available human cell lines and mouse embryos. For personal and sensitive data, we will abide by the Belgian law on the protection of individuals with regard to the processing of personal data (30th July 2018) and the General Data Protection Regulation 2016/679. Privacy registrations via the KU Leuven PRET tool has been applied for and approved.
The reference of the file in the KU Leuven privacy register regarding the collection and use of human cell lines for method development, including spatial (multi-)omics methods, is: G-2020- 2313-R2(AMD)

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

There are no 3rd party agreements in place regarding this project.

Regular meetings with KU Leuven LRD will be held to evaluate and protect possible IP generated during the project that could lead to valorization actions. If deemed necessary, data that fall under IP will either not be shared, put under embargo, or a suitable license will be applied to the data when published (e.g. Creative Commons License).

Leveraging droplet microfluidics and nucleic acid engineering for development of a CRISPR-based cell imaging toolbox
DPIA

DPIA

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

- No

Leveraging droplet microfluidics and nucleic acid engineering for development of a CRISPR-based cell imaging toolbox

GDPR

GDPR

Have you registered personal data processing activities for this project?

- No

Leveraging droplet microfluidics and nucleic acid engineering for development of a CRISPR-based cell imaging toolbox
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	
Spectrophotometer data	Raw and analysed data	New	Digital	Experimental	.xlsx	<100MB	
Gel electrophoresis	raw and analysed data	New	Digital	Experimental	.scn, .tiff, .xlsx	<100 MB	
Cas9 nuclease assay in droplets	Raw and analysed data	New	Digital	Experimental	.tiff, .xlsx, .avi	< 1GB	
Droplet microscopy images and movies or fluorescence data	Raw and analysed data	New	Digital	Experimental	.tiff, .avi, .xlsx	<100 GB	
Nucleic acid designs	Nucleic acid sequence data	New	Digital	Other	.dna, .txt, .fas	<100 MB	
Hgih-resolution microscopy images of fixed cells	Raw and analysed data	New	Digital	Experimental	.tiff, .nd2, .xlsx	<50 TB	
Hig resolution microscopy images and videos of live cells	Raw and analysed data	New	Digital	Experimental	.tiff, .nd2, .xlsx	<50 TB	
Multi-resolution processed and/or summarized microscopy data	Analysed data	New	Digital	Compiled/aggregated data	.tiff	<10 TB	
Algorithmically generated in-house gRNA sequence designs	Software	New	Digital	Software	.fasta	<50 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:

We will not reuse already existing data

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

Yes. We have already acquired The Ethics Committee Research UZ / KU Leuven (EC Research: S64403) approval for technology development on human cell lines. Remaining approvals in this project are for: fetal, adult, and ageing healthy tissues as well as diseased tissues, e.g. cancers; ECD approval for work with animal (e.g. mouse) embryos. We are in the process of acquiring these additional ethical approvals.

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

Yes. We will reuse sequencing data previously generated and will also generate new sequencing and/or spatial data from commercially available human cell lines and mouse embryos. For personal and sensitive data, we will abide by the Belgian law on the protection of individuals with regard to the processing of personal data (30th July 2018) and the General Data Protection

Regulation 2016/679. Privacy registrations via the KU Leuven PRET tool has been applied for and approved.
The reference of the file in the KU Leuven privacy register regarding the collection and use of human cell lines for method development, including spatial (multi-)omics methods, is: G-2020- 2313-R2(AMD)

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

Yes. We do expect new IP generation in this project, more specifically to expand our IP portfolio around patent-pending CRISPR-dCas9DNAzyme tool and its applications for genomic imaging in fixed and live cells. Additionally, we can expect that the proposed work will result in research data with the potential for tech transfer and valorization, especially with regards to newly developed spatial (multi-)omics protocols and data analysis pipelines. When there is concrete potential for tech transfer, the IP related to these research data will be protected, with the support of KU Leuven LRD and the IOF managers supporting this project, i.e. IOF manager linked to the Biosensors group (Dr. Francesco Dal Dosso) and the Voet group (Genomics Medicine, Dr. Gregory Maes). The IP protection will not withhold the research data from being made public. In the case a decision is taken to file a patent application, it will be planned as such not to delay publications longer than needed.

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.

- No

There are no 3rd party agreements in place regarding this project.

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

Yes. Personal data (including sequencing and/or spatial omics data obtained from human material) will only be published after pseudonymisation, and identifiers will not be published. Also data sharing restrictions might potentially apply due to generation of IP. Regular meetings with KU Leuven LRD will be held to evaluate and protect possible IP generated during the project that could lead to valorization actions. If deemed necessary, data that fall under IP will either not be shared, put under embargo, or a suitable license will be applied to the data when published (e.g. Creative Commons License).

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).

The Biosensors group uses the electronic lab notebook in which a number of predetermined topics have to be described for each experiment (objective, protocol, results, and conclusion). The electronic lab notebook facilitates searching for particular metadata through a search engine. By mimicking the folder structure of the electronic lab notebook in the server-based folder with the experimental data, linking of the metadata to the actual data will be facilitated.
As a general rule, datasets will be made openly accessible, whenever possible via existing platforms that support FAIR data sharing (www.fairsharing.org). When depositing data in a local or public repository, the final dataset will be accompanied by this information in a README.txt document, following the Dublin Core Metadata standard if no other meta-standard is available yet. This file will be located in the top-level directory of the dataset and will also list the contents of the other files and outline the file-naming convention used. This will allow the data to be understood by other members of the laboratory and add contextual value to the dataset for future reuse.
For each peer-reviewed article, a separate folder will be made on the server, containing the latest Word version and all raw and processed data used in the article. In addition, a separate file will be made in the electronic lab notebook for each article, containing clickable links to all metadata files of data that were used in that particular article, to facilitate tracing back of protocols, results and conclusions.

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

To guarantee reusable aspect of data, sufficient documentation and methods information will be provided, whereas CC-BY license will be attached to data through data repositories.
For more details, please see the previous question.

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

Where will the data be stored?

The time-stamped digital data will be stored in an already created project folder on the shared drive (J:) of KU Leuven. The time-stamped digital metadata will be stored on the server of the electronic labbook (eLABJournal, Bio-ITech). The folder is open for all the staff that will be working on this project and is secured and backed-up by the ICTS service of KU Leuven. Copies can be made and kept on personal devices.

How will the data be backed up?

The digital data will be stored on the university's central servers 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

Yes. KU Leuven provides sufficient storage and back-up capacity during and after the project. A dedicated folder is made for the project to store data files.

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

The network drive for the project shared folder and the large volume storage folder are secured by the ICTS service of KU Leuven with a mirror copy. Confidential data can and will be protected with a password (available only for PI Jeroen Lammertyn). Visitors, MSc thesis students and internship students in the groups as well as other unauthorized persons will not have access to the data on the shared folder. Data storage in the cloud will be avoided, unless for temporary use only, e.g., to transfer large files between the researchers involved in the project.

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

Type 1 server back-end storage with mirror backup for the project shared folder will cost 57 Euro per Tb per year. Costs will be covered by the project consumables budget.

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...).

The data to be retained for the expected 5 years after the project's end are: dissemination data (source files of publications and presentations) and the most relevant measurement data.

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

The research data will be stored on an external hard drive after the end of the project. Dissemination data, namely files corresponding to papers and presentations, will be stored on the PCs of PI (J Lammertyn), and backed-up daily on the departmental server for long term storage.

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

The volume corresponding to dissemination data is expected to be relatively low (<10 GB), and therefore can be seamlessly embedded in the PI's allocation on the departmental server. The costs (1000 EUR/year) will be covered by other on-going projects at that point in time.
For research data, at current archiving costs of 10 Euro/(TB*year), we estimate a cost of 2000 Euro/year. These costs will be covered by funding acquired by the project PI in the context of other research projects. VLIR infrastructure funding for long-term data storage has been awarded to Prof. Voet.

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 an Open Access repository

Only researchers participating in the project will be able to access the data for the duration of the project. As soon as the article associated with the data is ready for publication, the data will be made open through the institutional repositories mentioned in 2.1. The data will be deposited in the institutional repositories: (KU Leuven: Research Data Repository (RDR) Research Data Repository (RDR) - RDR - Research Data Repository (kuleuven.be). Data will be assigned with DOIs to create trustworthy and persistent links for online content.

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

Not applicable.

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

Before making data and other research output from the project (e.g. journal articles, book chapters and conference proceedings) openly available, they will be aligned with the project IP strategy to avoid premature disclosure, which can compromise the patent filing application(s).

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

KU Leuven RDR (Research Data Repository)

When will the data be made available?

As soon as the research results have been published, the data can be made available to other researchers.

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

Data will be assigned with DOIs to create trustworthy and persistent links for online content.
ORCID will be included in all publications and other dissemination material for all the researchers involved in the project.

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

A restricted access repository can be implemented on a free tool, such as Dropbox, up to a certain volume. If this volume does not suffice, time-limited storage will be considered, thus limited to the time needed to download the data. The costs associated with data storage will be covered by the budget foreseen in the project agreement.

6. Responsibilities

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

PhD student working on this project will be responsible for the data collection, documentation and metadata. They will be trained in data management at the beginning of their contract. Supervisors will manage the data storage facilities

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

PhD student working on this project will be responsible to store the data on the appropriate accommodation provided by KU Leuven. The ICTS service of KU Leuven 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?

Jeroen Lammertyn

Who will update and implement this DMP?

Charlotte Van Tricht, Jeroen Lammertyn