DMP title

Project Name The epigenetic reader LEDGF/p75 in health and disease - DMP title
Project Identifier D-2022-1408
Grant Title C14/21/099
Principal Investigator / Researcher Zeger Debyser
Project Data Contact 049127152 zeger.debyser@kuleuven.be
Description Basi c research on the role of LEDGF/p75 in HIV replication and MLL Leukemia. Target validation for use in HIV cure strategy. Drug discovery of epigenetic drugs.
Institution KU Leuven

1. General Information

Name of the project lead (PI)

Zeger Debyser

C1-C2 Project number & title

C14/21/099

The epigenetic reader LEDGF/p75 in health and disease.

2. Data description

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

- Generate new data

2.2. What data will you collect, generate or reuse? Describe the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a numbered list or table and per objective of the project.

1. Analogous data generated : cell lines are stored in cell bank with appropriate registration, proteins expressed and purified are stored in protein database of the lab; plasmids will be engineered and stored in a plasmid database; primers and antibodies will be commercially acquired and stored in respective databases.

2. Throughout WPs analytical data (PCR data, electrophoresis data, western blot data, immunohistochemistry data, microscopy data) will be initially be collected in a variety of file formats mainly Microsoft Excel, MS Word and equipment specific software such as FlowJo for flow cytometry data. These files will also be stored in Open Document Format or as CSV files. Estimated size 200-300GB.

3. Digital images will be saved as TIFF files with an estimated volume of 200 GB;

3. Ethical and legal issues

3.1. Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to the file in KU Leuven's Record of Processing Activities. Be aware that registering the fact that you process personal data is a legal obligation.

No

3.2. 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).

Only related to the use of human cell lines and human PBL from Blood Transfusion.

A S number is requested for.

3.3. Does your research 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?

Hit compounds will be discovered in the project; their chemical nature will remain confidential but only after chemical optimisation the compounds will be protected by IP after consulting with the Tech Transfer office (Ku Leuven LR&D).

3.4. 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 regarding reuse and sharing are in place?

No

4. Documentation and metadata

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

All collected digital data (i.e. cell line information, proteomic information, protocols, etc) are and will be stored in an online labmanagement system with integrated Electronic Lab Notebook OneNote which runs on a secured and backed up server of KULeuven (managed by ICT of the Biomedical Sciences Group). This system also provides a logging system so no data can ever be erased, making that everything will be tracable and stored longterm (well beyond the common 5 year requirement). Certificates of cell lines and antibodies are stored as well with a link to te experiment. All researchers also keep a paper labbook with annotations of dates/experiments/protocols/results/ and references to OneNote labbook and other files/folders (imaging data). Labbooks are archived in personal folders in a secured archive room. For all coworkers it is mandatory to use a tracking system between labbook, electronic note book and digital files. During the project we attempt to select a unified tracking system.

4.2. Will a metadata standard be used? If so, describe in detail which standard will be used. If not, state in detail which metadata will be created to make the data easy/easier to find and reuse.

Yes

Coworkers already store important info together with experimental data in electronic labbook. We are in the process of adapting this info to metastandards in our Data Repository.

We will use standards listed on https://fairsharing.org/ with already available standards for

- Flow cytometry :
- https://www.re3data.org/search?query=statistics&metadataStandards%5B%5D=other
- Microscopy imaging :
- https://www.openmicroscopy.org/
- qRT-PCR:
- http://miqe.gene-quantification.info/

In house metastandards for western blotting, cellular experiments will be generated during the project..

5. Data storage and backup during the C1-C2 project

5.1. Where will the data be stored?

The time-stamped master copy of the data will be kept on our research unit central storage facility. Copies can be made and kept on personal devices. Source data for publications are collected in separate digital or paper folders at the moment of submission and handed over to the PI. These folders are archived in the archive (paper folders) or on the central storoge facility of KU Leuven.

5.2. How will the data be backed up?

The data will be stored on the university's central servers with automatic daily back-up procedures.

5.3. 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 We have a shared drive (restricted acces) on the university's central servers (GBW0408_Molecular_Medicine_Backup, size 700gb) and archive drive (GBW0025_Molecular_Medicine_Storage, size 5TB) for during the project

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

The expected amount of data is 500 gb, the cost for during the project is then estimated on 250 euro per year.

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

All data are stored on university central servers, paper folders (labbooks, paper source data) are

kept in a locked archive room with restricted acces (PI and secretary).

6. Data preservation after the end of the C1-C2 project

6.1. Which data will be retained for the expected 10 year period after the end of the project? If only a selection of the data can/will be preserved, clearly state why this is the case (legal or contractual restrictions, physical preservation issues, ...).

All data are kept for 5 years on the electronic labbook (and/or paper labbook). Data of technically flawed experiments are not kept. Analagous data (plasmids, cell lines) are kept for 5 years. Proteins, vectors are available until exhaustion. Mouse strains are kept in principle. Source data for publications are kept in a separate folder in the archive room for 10 years after publication.

6.2. Where will these data be archived (= stored for the long term)?

Digital data are archived on the central server facility of the KU Leuven.

Publication folders (physical) are stored in the archive room with restricted access.

Cell lines are kept frozen in liquid Nitrogen (cell biobank).

Proteins are kept at -20C or -80C in protein database as long as available..

Plasmids and antibidies are kept at -20°C in databases.

6.3. What are the expected costs for data preservation during these 10 years? How will the costs be covered?

Archiving in the archive room comes with no cost. The database of 3D images, that will be compiled to realise objective 2, will be hosted on the servers of KU Leuven. In view of the expected size of the database (500 gb), estimated cost will be 80 euro per year.

7. Data sharing and re-use

7.1. 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 or because of IP potential)?

Data on hit compounds will not be shared due to confidentiality. Research data will be published in open access papers.

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

Research data will be published in open access papers. Non-published data will be made available upon official request.

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

Upon request by mail Published data and reagents used for data generation are in principle available upon request. Restrictions may be attached for sharing materials (eg MTA). Commercial use will require official agreement negotiated with the help of LR&D. Non-published data (eg big data on genomics and proteomics) can be shared upon request; restrictions (eg for commercial use) can be imposed.

7.4. When will the data be made available?

Upon publication of the research results Published data are available upon publication. Journals may require deposition of data in repository. (eg PDB). Who will be able to access th

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

Members of Molecular Medicine will have access to the research data on the server. Hit compounds will only be accessible to team members working on the project (restricted acces through passwords). After publication, data will be available upon request

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

Limited costs. Transfer of reagents comes at a cost for the requestor (samples will be shipped on Fedex account of person who requests reagents)

8. Responsibilities

8.1. Who will be responsible for the data documentation & metadata?

Each researcher is responsible for own data collection and storage including metadata. The PI is responsible for the data management plan. Special care is taken for source data of publications. Control of policy is done by sampling.

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

Each researcher is responsible for own data collection and storage including metadata. The PI is responsible for the data management plan. Special care is taken for source data of publications. Control of policy is done by sampling.

8.3. Who will be responsible for ensuring data preservation and sharing?

Each researcher is responsible for own data collection and storage including metadata. The PI is responsible for the data management plan. Special care is taken for source data of publications. Control of policy is done by sampling.

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

The end responsibility for updating and implementing the DMP is with the supervisor (promotor).