Congenital Disorders of Glycosylation: new defects and mechanisms

A Data Management Plan created using DMPonline.be

Creators: Matthew Wilson https://orcid.org/0000-0003-2252-8730, n.n. n.n., n.n. n.n., n.n. n.n.

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: n.n. n.n.

Data Manager: Matthew Wilson https://orcid.org/0000-0003-2252-8730

Project Administrator: n.n. n.n., n.n. n.n.

Grant number / **URL:** 1289023N

ID: 199099

Start date: 01-10-2022

End date: 30-09-2025

Project abstract:

Congenital Disorders of Glycosylation (CDG) are a group of >150 inherited metabolic disorders. CDG can be caused by the disruption of many aspects of cell biology, from those caused by defects in enzymes catalyzing glycan transfer to those that affect the function of entire organelles, leading to indirect disruption of glycosylation pathways.

This project details the study of two CDG newly identified in our lab. The first of these, DHRSX deficiency, is thought to disrupt the metabolism of dolichol. This is a lipid used as a sugar/glycan carrier for the attachment and formation of N-glycans, as well as having other functions that are poorly understood, such as in the membranes of various organelles. We expect to identify a completely novel step in dolichol metabolism. DHRSX deficiency will also become the first identified truly pseudo-autosomal recessive disorder in human genetics.

The second disorder, caused by mutations in CAMLG, is a defect of the transmembrane domain recognition complex (TRC) pathway, required for the insertion of tail-anchored (TA) proteins into the endoplasmic reticulum membrane. Researching the mechanism by which mutations in CAMLG lead to a CDG will reveal important information about both the TRC pathway and protein trafficking.

In summary, this project will take two diverse novel disorders and use their study to not only describe new aspects of human disease, but also to reveal fundamental and exciting aspects of cell biology.

Last modified: 28-04-2023

Congenital Disorders of Glycosylation: new defects and mechanisms
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)

Sequences of patient DNA may be collected in order to facilitate the diagnosis of affected individuals as part of the contingency plan described or during the investigation of the disorders described elsewhere in the project proposal (DHRSX deficiency or CAMLG-CDG). Additional small-scale data as to the pathophysiology of the disorders described will also be collected/created. For example, this could be the abundance of proteins or other biomolecules. Clinical data related to the affected individuals will also be collected in order to facilitate publication or to delineate the pathophysiological features of these disorders. The data collected will be published as peer-reviewed manuscripts.

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)

The responsible person for preservation of data will be the project proposal author, Matthew Wilson. It is also Leuven RDP policy to keep data for at least 5 years. The infrastructure for this is provided centrally. In addition, any next generation sequencing data collected will be stored for 30 years according to UZ Leuven guidelines. This is also stored on secure UZ Leuven servers.

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)

N/A

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)

N/A

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

N/A

Congenital Disorders of Glycosylation: new defects and mechanisms
DPIA

DPIA

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

- Not applicable

Congenital Disorders of Glycosylation: new defects and mechanisms
GDPR

GDPR

Have you registered personal data processing activities for this project?

- Not applicable

Congenital Disorders of Glycosylation: new defects and mechanisms
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.

Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
Cells from affected CDG patients	Fibroblasts or EBV-immortalized lymphoblast cell lines creating under the framework of another project	Reuse existing data	Physical				
Model human cell lines	HEK293T, HAP1 cell models	Generate new data	Physical				
Cell lines from other organisms	Chinese Hamster Ovary Cell models, yeast models, bacteria (E. Coli) developed for plasmid and protein production	Reuse existing data	Physical				
Plasmids used for research goals	Plasmids containing cDNA of interest for expression in model cell lines	Generate new data	Physical				
Sanger Sequencing data	Analysis of plasmids/cDNA/genomic DNA	Generate new data	Digital	Experimental	.ab	<1GB	
Imaging data	Immunoblotting images, confocal microscopy of cells	Generate new data	Digital	Experimental	.tif	<1TB	
HPLC and LCMS data	Analysis of cell lines for LLO and isoprenoid species content	Generate new data	Digital	Experimental	.D, .ch, .cdf	<1TB	
RNA-Seq data	RNA-Seq data from generated model cell lines and affected CDG patients	Generate new data	Digital	Experimental	.fastq, .bam, .cram, .csv, .xls	<5TB	
WGS data	Whole genome data from undiagnosed affected CDG patients, including files generated during analysis for the identification of pathogenic variants	Generate new data	Digital	Experimental	.fastq, .bam, .cram, .csv, .xls	<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:

Physical and digital data previously produced within the laboratory will be reanalysed as needed and used for experiments during the project. Digital data is either stored on the KU Leuven OneDrive or ManGO and can be freely accessed by researchers there.

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

The project involves human subjects who are affected by several Congenital Disorders of Glycosylation (CDG). Cells derived from primary biopsies will be studied (dermal fibroblasts and EBV-immortalized lymphoblasts). These samples were collected under previous projects (EUROGLYCAN-omics and CELSA/21/027 that allowed the reuse of these samples for future projects within the same lab.
Consent for data retention and sharing with other researchers has been (or will be) obtained from each of the study participants. This consent is through the EUROGLYCAN-omics or CELSA/21/027 projects informed consent forms which have been approved by the KU Leuven Ethics Commission.
For the publication of manuscripts detailing clinical features, it is also clearly necessary to collect personal (clinical) data from participants, as well as images. This is also detailed in the informed consents mentioned above.

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

Because patient-derived cells and next-generation sequencing data will be studied, it is impossible to fully anonymise this data as genetic information can theoretically always be linked back to the original patient. Personal data will therefore be pseudonymized.
A PRET application will be submitted at KU Leuven and full ethics approval thorugh the UZ/KU Leuven ethics commission is pending.

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.

- No

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

Specifically for the study of DHRSX a material transfer agreement is in place between KU Leuven and The Hospital for Sick Children, Toronto for the study of patient samples (S58358 - R7363).

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

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).

Experimental data will be stored in a clearly labelled filesystem on ManGO. Metadata will be stored in an Excel file, noting the pseudonymised codes (see below) assigned to each individual studied, the datasets as they have been integrated and the structure of the filesystem used, as well as the formats of each individual piece of experimental data.
Patients and their individual data will be pseudonymised using 'GC codes' assigned by the UZ Leuven Genomics Core service, who will perform NGS; these are stored in another secure separate metadata document in the Microsoft Excel format. In addition, for datasets uploaded to RD-Connect, persistent IDs are also generated.
Identifiers will be in the pseudonymised format GCxxxxxx, with the 6 digit xxxxxx code corresponding to a fixed number assigned by Genomics Core. Data for each of these patients will be stored in a folder under this name, with each experiment run using their data clearly stored underneath this master file (e.g. WGS, RNAseq, CHIPSeq). Bioinficatics pipelines used for data analysis will also save experiments according to the date the analysis was performed (e.g. 280423_1613_RNASeq). Combined datasets will be stored according to the date that the experiment was created and a description of the analysis performed (e.g. April2023_RNAseq). Inside this folder, a metadata document in Excel format will contain each of the experiments integrated into this combined analysis dataset. This data will all be stored on the Google Cloud storage system.

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

Data deposited in the KU Leuven RDR will use DataCite as the metadata standard.
Other experimental metadata will be stored in an Excel file, noting the pseudonymised codes (see below) assigned to each individual studied, the datasets as they have been integrated and the structure of the filesystem used, as well as the formats of each individual piece of experimental data.
For data uploaded to RD-Connect, data will also be findable there.

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

Where will the data be stored?

Short-term non-personal data will be stored on the KU Leuven OneDrive assigned to researchers on the project.
Larger data sets (e.g. Next generation sequencing data) will be stored on the KU Leuven ManGO data storage service.
ManGO will also be used as a backup for smaller datasets regarded to be important to share securely (e.g. .tif files created by confocal microscopy or chemiluminescence imaging)

How will the data be backed up?

Both the KU Leuven OneDrive and ManGO storage systems have redundancy that means backups are automatically created or data is stored in multiple secure locations. Some smaller data (NB not larger datasets which are unsuitable for OneDrive storage) will also be kept in both of these storage systems.

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, ManGO has sufficient storage for the datasets produced and these are extended (and paid for) on a pro-rata basis.

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

Both OneDrive for business and ManGO are secured using the KU Leuven data security systems and authentication.

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

OneDrive for business is free for researchers based at KU Leuven.
ManGO costs are estimated at 35 EUR/TB/Year. The research group currently uses approximately 4TB and during this project no more than a total of 5TB additional data for ManGO storage is expected to be produced. This would mean a maximum cost by the end of the project of 315 EUR per year. These costs are relatively minor and can be paid by the fellow's bench fees during the project. After the project has been completed, the costs can be paid by existing grants within the laboratory of Gert Matthijs.

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...).

Experimental NGS data will be stored for 30 years as per UZ Leuven guidelines. Derived experimental data will be stored on the ManGo servers for at least 5 years, conforming to the Leuven RDP policy.

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

ManGO KU Leuven storage volumes and price (35 EUR/TB/year) mean that this is a feasible location for long-term storage of the data. ManGO allows the inclusion of metadata and the preservation of file systems for easy access and curation.

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

ManGO has costs of 35 EUR/TB/year. Since we expect to produce a total of approximately 5TB data during the project, long-term storage costs are not large and can be covered by future research grants within the laboratory.

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
- Yes, in a restricted access repository (after approval, institutional access only, …)

Whole genome and RNA-Seq data will be made available by uploading it to the RD-Connect (https://rd-connect.eu/) platform. This will allow researchers (i.e. other verified users of the RD-Connect platform) that agree to certain terms and conditions free and open access to the data collected. Specifically, Fastq files will be provided to RD-Connect. This restricted access repository is necessary due to the personal nature of the data shared.
Other data directly relevant to publications, and the publications themselves, will be made open access in their respective journals. In addition, data necessary for replication of experiments but not shared directly in a publication will be shared in the KU Leuven Research Data Repository (RDR).

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

Data shared in RD-Connect can only be accessed by academics who have registered and been verified by the platform.

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, Privacy aspects
- Yes, Ethical aspects

Personal data derived from next generation sequencing technologies cannot be shared openly and will only be openly usable within the framework of RD-Connect. This fully anonymises the data and allows access only to registered and trusted scientific professionals.

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

Whole genome and RNA-Seq data will be made available by uploading it to the RD-Connect (https://rd-connect.eu/) platform.

When will the data be made available?

After the end of the project an embargo of 6 months will be applied to the data as laid down in the 3rd party agreement; this is also valid for data shared on RD-Connect. After this it will be made available.

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

Data shared in the KU Leuven RDR will be available under at CC-BY-4.0 license.

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

KU Leuven RDR uses DOIs and one will be assigned to any datasets uploaded there.
Manuscripts produced during the project will receive DOIs upon publication.

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

The small amount of data in RDR will be free. RD-Connect is also free.

6. Responsibilities

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

Matthew Wilson

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

Matthew Wilson

Who will manage data preservation and sharing?

Matthew Wilson

Who will update and implement this DMP?

Matthew Wilson