Unravelling the molecular pathomechanisms in the newly delineated CBFB-related cleidocranial dysplasia

A Data Management Plan created using DMPonline.be

Creators: Gretl Hendrickx, n.n. n.n.

Affiliation: KU Leuven (KUL)

Funder: Bijzonder Onderzoeksfonds

Template: KU Leuven BOF-IOF

Principal Investigator: n.n. n.n., Gretl Hendrickx

Data Manager: n.n. n.n., Gretl Hendrickx

Project Administrator: n.n. n.n.

Grant number / URL: C14/23/140

ID: 200832

Start date: 01-10-2023

End date: 30-09-2027

Project abstract:

Cleidocranial dysplasia is a rare skeletal disorder that was previously related to pathogenic variants in the *RUNX2* gene. We have recently discovered six disease-causing mutations in *CBFB*, leading to a distinct form of this skeletal disorder. *CBFB* encodes the CBFβ protein, a key partner of the RUNX transcription factors in the heterodimeric core-binding factor complex. The aim of this proposal is to characterize the molecular events leading to this unique cleidocranial dysplasia phenotype caused by the *CBFB* mutations. We will perform an in-depth analysis of all identified mutations using *in vitro* and *in vivo* strategies to define genotype-phenotype correlations. This proposal is highly relevant as it will provide novel insights into the pathogenesis of this rare bone disease, and by extension also in the biological role of the core-binding factor complex in skeletal and dental development. On the long term these new insights may open new therapeutic avenues.

Last modified: 17-03-2024

Unravelling the molecular pathomechanisms in the newly delineated CBFB-related cleidocranial dysplasia

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 / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
		*Indicate: **N**(ew data) or **E**(xisting data)*	Indicate: **D**(igital) or **P**(hysical)	Indicate: **A**udiovisual **I**mages **S**ound **N**umerical **T**extual **M**odel **SO**ftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
Informed consent forms (ICFs)	signed ICFs (paper and scanned)	N	P D	T	.pdf	<1GB	1 ICF (5-10 sheets)/participant
Clinical/personal data	Date/year of birth and/or age, EAD number, health data (e.g. medical history)	N	D	T	in REDCap eCRF; exported as .csv or .pdf	<1GB	NA
Genome sequencing data	Raw genome sequencing data	N	D	T	.fastq	<5TB	NA
Processed Genome sequencing data	Bio-informatic analysis output genome data	N	D	T	.vcf .cram .xlsx	<1TB	NA
Sanger sequencing data	Raw Sanger sequencing data	N	D	T	.ab1	<100GB	NA
Gel electrophoresis data	Images of agarose gels	N	D	I	.png .jpeg	<1GB	NA
Data generated from cellular experiments	Raw and processed data of cellular experiments	N	D	N	.xlsx .pfzx .pptx .docx	<1GB	NA
Western blot detection	Images of western blots	N	D	I	.tiff .pdf	<1GB	NA
quantitative PCR raw data	Gene expression data obtained by performing qPCR	N	D	N	.xlsx	<1GB	NA
RNA sequencing raw data	Gene expression data generated by Genomics Core	N	D	N	.fastq	<1TB	NA
RNA sequencing data processed	Analysis of raw RNAseq data	N	D	N	.xls, .csv, .r	<1GB	NA
Images of histology of organoids and zebrafish models	Pictures taken with microscope or camera from organoids or zebrafish	N	D	I	.jpeg	<100GB	NA
Electronic lab books	Electronic reporting of experimental procedures and results in OneNote	N	D	Other	.onepkg	<1GB	NA
Lab books	Paper lab books with documented experimental design, notes and protocols	N	P	Other	NA	NA	<5 lab books
Standard operating procedures	Written protocols and procedures performed in the lab	Reuse existing data	P, D	Other	.docx, .pdf	<1GB	<50 sheets
Composite figures for data presentation	Figures illustrating project results, with schematic overviews	N	D	I	.pptx, .tiff	<100GB	NA
Manuscripts for publication	Text files of publications	Reuse existing data	D	Other	.docx	<1GB	NA

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:

NA

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, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- Yes, human subject data (Provide SMEC or EC approval number below)
- Yes, animal data (Provide ECD reference number below)

For the ethical issues concerning human subject data, approval by the CTC/EC is currently ongoing. EC approval number will be provided once available.
For the ethical issues concerning generation and use of zebrafish models, ethical approval by the ECD is planned for the 3th project year (2026). ECD reference number will be provided then.

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- Yes (Provide PRET G-number or EC S-number below)

We will process personal and genetic data of patients (UZ Leuven study S68246)

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

NA

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or 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 Material transfer agreement and Data processing agreement have been created for study S68246, for our collaboration with the Paediatric Hospital SAMIC Prof. Dr. Juan P. Garrahan in Buenos Aires, Argentina. We will obtain DNA samples and limited personal data from them, and we will generate genetic data to share with them.
Future collaborations with other hospitals/institutes might occur with a similar purpose. If so, these will be added here throughout the 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.

- No

Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keepdata 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).

All performed experiments will be documented in the electronic OneNote lab book of the researchers involved in the project, with all the necessary information to keep the data understandable. Acquired digital files are organized in respective experiment folders with file names specifying date and type of measurement of stored data.

Will a metadata standard be used to make it easier tofind and reuse the data?

If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- No

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Shared network drive (J-drive)
- OneDrive (KU Leuven)
- Other (specify below)

OneDrive (KU Leuven) will be used for temporary storage of files and files that are shared with many lab members and students, whereas the shared network drive (J-drive) will be used for general storage of data and the Archive Drive (K-drive) for the storage of large datasets.

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

We will use OneDrive (KU Leuven), J-drive and K-drive for data storage. If storage at OneDrive is not sufficient, data will be moved to the J-drive/K-drive. If storage capacity at J-/K-drive is not sufficient, we will contact ICTS of KU Leuven to increase the storage capacity according to our needs.

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

Patient-related data will be gathered in a case report form (CRF) and stored in a REDCap database (UZ Leuven). Genetic data will strictly be stored on network drives of KU Leuven, which are not accessible for unauthorized persons. Other research data will mostly be stored on these drives as well throughout the project, and completely at the end of the project. All storage spaces (OneDrive, J drive and K-drive) need researcher authorization for access, as they are secured with KU Leuven credentials and double authentication system.

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

OneDrive storage is provided by KU Leuven. J-drive and K-drive storage cost is paid by the PI, prof. Geert Mortier, in the price of €52/100GB and €5,69/100GB, respectively.

Data Preservation after the end of the Research Project

Which data will be retained for 10 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 preserved for 10 years according to KU Leuven RDM policy

NA

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

- Large Volume Storage (longterm for large volumes)
- Shared network drive (J-drive)
- Other (specify below)

Data will be stored indefinitely on the J- and K-drives of KU Leuven

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

OneDrive storage is provided by KU Leuven. J-drive and K-drive storage cost is paid by the PI, prof. Geert Mortier, in the price of €52/100GB and €5,69/100GB, respectively.

Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project?
Please explain per dataset or data type which data will be made available.

- Yes, as open data

Research results will be published in journals with open access. RNA sequencing results with be put in public repositories. All generated iPSC lines will be registered in the Human Pluripotent Stem Cell Registry (hPSCreg®). The generated zebrafish strains will be deposited to the Zebrafish International Resource Center (ZIRC). Raw data and physical documentation will continue to be available to the members of the laboratory.

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

NA

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 per dataset or data type where appropriate.

- Yes, privacy aspects

The collection, processing and disclosure of personal data, such as patient health and medical information is subject to compliance with applicable personal data protection and the processing of personal data (Regulation (EU) 2016/679 also referred as the General Data Protection Regulation ("GDPR") and the Belgian Law of July 30 2018 on the protection of natural persons with regard to the processing of personal data).
When data are coded, there continues to be a link between the data and the individual who provided it. The research team is obligated to protect the data from disclosure outside the research according to the terms of the research protocol and the informed consent document. The subject's name or other identifiers should be stored separately from their research data and replaced with a unique code to create a new identity for the subject. The data will only be accessible to the members of the research project. Patient-related data will be gathered in a case report form (CRF) and stored in a REDCap database.

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)
- Other (specify below)

All generated iPSC lines will be registered in the Human Pluripotent Stem Cell Registry (hPSCreg®). The generated zebrafish strains will be deposited to the Zebrafish International Resource Center (ZIRC). Raw data and physical documentation will continue to be available to the members of the laboratory.

When will the data be made available?

- Upon publication of research results

NA

Which data usage licenses are you going to provide?

If none, please explain why.

- CC-BY 4.0 (data)

NA

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- Yes, a PID will be added upon deposit in a data repository

DOI will be provided upon acceptance for publication of the related research papers

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

Minimal costs are expected, as most of the data can be put on public repositories without costs.

Responsibilities

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

The PI (prof. Geert Mortier) and co-PIs (dr. Gretl Hendrickx, prof. Przemko Tylzanowski) will be responsible for data documentation & metadata.

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

Data management, storage and backup will be performed by the PI (prof. Geert Mortier) and co-PIs (dr. Gretl Hendrickx, prof. Przemko Tylzanowski) of the project.

Who will manage data preservation and sharing?

The PI of the project, prof. Geert Mortier, will be responsible for data preservation and sharing.

Who will update and implement this DMP?

As co-PI, dr. Gretl Hendrickx will be responsible for updating the DMP throughout the project, whereas prof. Geert Mortier as PI has the end responsibility for updating and implementing the DMP.