FOCAL SEGMENTAL GLOMERULOSCLEROSIS (FSGS) IN FLANDERS: EPIDEMIOLOGY AND SINGLE-CELL TRANSCRIPTOMICS ANALYSIS (FWO DMP)

DMP TITLE

ADMIN DETAILS

Project Name: Focal segmental glomerulosclerosis (FSGS) in Flanders: epidemiology and single-cell transcriptomics analysis (FWO DMP) - DMP title
Project Identifier: 11L5622N
Grant Title: FWO PhD Fellowship fundamental research
Principal Investigator / Researcher: Dries Deleersnijder (Promotor: Prof. Dr. Ben Sprangers)
Project Data Contact: dries.deleersnijder@kuleuven.be
Description: Focal segmental glomerulosclerosis (FSGS) is a histopathological pattern of kidney injury and is subdivided into primary, secondary and genetic subtypes. Primary FSGS is caused by a sudden and generalized injury to the visceral glomerular epithelial cells, is treated with immunosuppression and has a grim prognosis with patients frequently progressing to end-stage kidney disease despite treatment. Secondary FSGS results from glomerular hypertension, is treated with renin–angiotensin-aldosterone system (RAAS) inhibition and has a considerably better renal prognosis. However, in clinical practice, patients initially present with a similar clinical picture and we currently lack biomarkers that can guide correct diagnosis and treatment of the underlying subtype. In this research project, we therefore aim to elucidate the underlying pathophysiology of FSGS and identify candidate diagnostic biomarkers that can help us to more reliably differentiate between the three subtypes. In WP1, we will study the epidemiology of biopsied native (non-transplant) kidney disease in Flanders. By analyzing the Flemish Collaborative Glomerulonephritis Group (FCGG) native kidney biopsy registry, we will identify all FSGS patients in Flanders since 2017. In WP2, we will select these FSGS patients and stratify them according to their presumed subtype, using clinical, histopathological and genetic data (retrospectively collected from the clinical charts). A subset of these patients at UZ Leuven will serve as a validation cohort in WP4. In WP3, we apply single-nucleus RNA-sequencing techniques on kidney core needle biopsies of human patients with primary, secondary and genetic FSGS to identify differential gene expression in the different subtypes and propose candidate diagnostic markers that can help in the differential diagnosis. In WP4, we aim to validate our findings on the RNA and protein level in our validation cohort from WP2.
Institution: KU Leuven

1. GENERAL INFORMATION

Name applicant
Dries Deleersnijder
Promotor: Prof. dr. Ben Sprangers

FWO Project Number & Title
Grant application number: 11L5622N
Focal segmental glomerulosclerosis (FSGS) in Flanders: epidemiology and single-cell transcriptomics analysis

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).

WP1-2: Epidemiological analysis of native kidney diseases in Flanders and profile of the Flemish FSGS cohort

WP1: We will study Flemish patients from the Flemish Collaborative Glomerulonephritis Group (FCGG) database, which is a population-based native kidney biopsy registry that has been including all biopsies performed in Flanders since 2017. The database is managed by the Nederlandstalige Belgische Vereniging voor Nefrologie (NBVN) and a dataset will be extracted for this work package. We aim to study demographic and clinicopathological features of included patients.

WP2: We will study Flemish patients with a diagnosis of FSGS, identified through the FCGG registry. Additional data on clinical course, histopathology, treatment and outcomes will be collected and stored in the FCGG database. Next, a dataset of these FSGS patients will be extracted from the database, to be used in this work package. We aim to classify patients according to updated diagnostic criteria, provide an overview of received treatments and study prognosis and determine predictors of renal outcomes and treatment effect.

Data(set) name	**- Origin of data - Mode of data collection - Content of data**	**Format of data**	**Volume**	**How created? Source?**
Dataset 1.1: Extracts of FCGG database	- secondary - observational - numerical and textual	- Spreadsheet: MS Excel (.xls/.xlsx), comma-separated value files (.csv)	20MB	Extracted from FCGG online database
Dataset 1.2: Statistical analysis in R	- secondary - derived from dataset 1.1 - numerical and textual	- R source file (.R) - Text files: plain text data (Unicode, .txt), MS Word (.doc/.docx), Adobe Portable Document Format (.pdf). - Spreadsheet: MS Excel (.xls/.xlsx), comma-separated value files (.csv)	100MB	Processed from dataset 1.1
Dataset 1.3: Manuscript(s), supplementals, tables, graphs, images	- secondary - derived from dataset 1.1 - numerical, textual, multimedia	- Text files: MS Word (.doc/.docx), Adobe Portable Document Format (.pdf). - Spreadsheet: MS Excel (.xls/.xlsx), comma-separated value files (.csv) - Image file: Adobe Portable Document Format (.pdf), uncompressed TIFF (.tif/.tiff), JPEG (.jpg), PNG (.png), EPS (.EPS). - Graphpad Prism project file (.pzfx)	500MB	Processed from dataset 1.1

WP3-4: Single-nucleus RNA-sequencing on kidney biopsies of FSGS patients and validation of candidate markers in the FSGS cohort

WP3: We will perform single-nucleus RNA-sequencing (snRNA-seq) experiments on native kidney biopsy cores of 24 patients with a diagnosis of FSGS (8 primary FSGS, 8 secondary FSGS, 8 genetic FSGS). These biopsy cores have already been stored in the Leuven Renal Research Biobank (LRRB, managed by the UZL biobank) for research purposes. Transcriptomic data from patient samples will be generated locally using Illumina HiSeq4000 and NovaSeq6000 machines, producing files in .fastq or .bcl format. Sequencing data will be correlated with clinical diagnosis and outcomes. Sequencing data will be processed using various packages in R (e.g. Seurat).

WP4: Candidate diagnostic biomarkers identified in snRNA-seq experiments will be validated on the RNA (RT-PCR, *in situ* hybridization) and/or protein level (immunofluorescence [IF], immunohistochemistry [IHC]) in the FSGS validation cohort (subset of patients in LRRB/UZL biobank in which kidney biopsy samples, urine and blood samples are available).

Data(set) name	**- Origin of data - Mode of data collection - Content of data**	**Format of data**	**Volume**	**How created? Source?**
Dataset 2.1: Transcriptomic raw sequencing data	- primary - experimental - numerical and textual	- Raw data files in binary base call format (.bcl) - text-based raw sequencing data and quality scores (.fastq)	~30 GB per sample * 24 samples = 720 GB	Raw sequencing data output from the sequencing platform
Dataset 2.2 Processed sequencing data	- secondary - derived from dataset 2.1 - numerical, textual, multimedia	- Read/UMI count data: .mtx, .tsv, .rds - Text files: plain text data (Unicode, .txt), MS Word (.doc/.docx), Adobe Portable Document Format (.pdf), LaTex (.tex) format - Spreadsheet: comma-separated value files (.csv), tab-delimited file (.tab), delimited text (.txt), MS Excel (.xls/.xlsx) - Digital images in raster formats: uncompressed TIFF (.tif/.tiff), JPEG (.jpg), PNG (.png), Adobe Portable Document Format (.pdf), bitmap (.bmp), .gif - Digital images in vector formats: scalable vector graphics (.svg), Adobe Illustrator (.ai), Adobe Portable Document Format (.pdf)	25 GB	Processed from dataset 2.1
Dataset 2.3 RNA/protein validation data	- primary - experimental - numerical, textual, multimedia	To be determined (TBD); file formats depend upon chosen validation technique/assay and will be determined later on in the project	TBD	Data output from validation experiments
Dataset 2.4: Manuscript(s),	- secondary - derived from	- Text files: MS Word (.doc/.docx), Adobe Portable Document Format	10GB	Processed from datasets

supplementals, tables, graphs, images	dataset 2.2 - numerical, textual, multimedia	(.pdf). - Spreadsheet: MS Excel (.xls/.xlsx), comma-separated value files (.csv) - Image file: Adobe Portable Document Format (.pdf), uncompressed TIFF (.tif/.tiff), JPEG (.jpg), PNG (.png), EPS (.EPS). - Graphpad Prism project file (.pzfx)		2.2 and 2.3

Biological samples:
Biological samples used in WP3-4 are derived from the Leuven Renal Research Biobank (LRRB, S54095) managed by UZ Leuven Biobank. All biological samples (tissue, blood, urine, single-nuclei suspensions, sequencing library preparations) are stored in labelled tubes or SBS plates in -20°C or -80°C freezers. Electronic laboratory databases in .xls format are used to keep track of the of these samples and their link to the original study sampleID. All biological samples are stored according to the guidelines of the UZ Leuven Biobank.

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: Ethical Committee (EC) at UZ Leuven was notified of the data processing activities (FCGG protocol S59182, LRRB protocol S54095, snRNA-seq protocol S65227).

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

WP1: Epidemiological analysis of native kidney diseases in Flanders
WP1 will be performed at KU Leuven (Nephrology and Renal Transplantation Research group). We will use a pseudonymized data extract from the FCGG native kidney biopsy registry in Flanders (S59182). All patients that undergo native kidney biopsy procedure in Flanders are included in the registry. Personal data include demographic data (e.g. age, sex), histopathological data (e.g. histopathological diagnosis, disease activity and chronicity) and clinical data (e.g. proteinuria, serum albumin, hematuria, nephrological diagnosis). All patient data are pseudonymized with a unique FCGG patient code which does not carry personal identifiers. The decoding key is managed by the participating nephrology centers and NBVN.

WP2: Profile of the Flemish FSGS cohort
WP2 will be performed at KU Leuven (Nephrology and Renal Transplantation Research group). We will use a pseudonymized data extract from the FCGG native kidney biopsy registry in Flanders. Collection

of treatment and outcome data in FCGG is not covered by the original study protocol (S59182) and a new research protocol will be submitted at EC UZL. Data will only be retrospectively collected from the clinical charts. This data extract will cover all patients in Flanders with a biopsy-confirmed diagnosis of FSGS. Personal data include demographics (e.g. age, sex), histopathological data (e.g. histopathological diagnosis, disease activity and chronicity, evaluation by electron microscopy), clinical data (e.g. proteinuria, serum albumin, hematuria, nephrological diagnosis), biochemical data (e.g. results from blood and urine tests, genetic data), outcome data (e.g. disease remission, relapse, end-stage kidney disease). All patient data are pseudonymized with a unique FCGG patient code which does not carry personal identifiers. The decoding key is managed by the participating nephrology centers and NBVN.

WP3-4: Single-nucleus RNA-sequencing on kidney biopsies of FSGS patients and validation of candidate markers in the FSGS cohort

WP3-4 will be performed at KU Leuven (Nephrology and Renal Transplantation Research group) in close collaboration with the VIB – KU Leuven Center for Cancer Biology (Laboratory of Translational Genetics (LTG) headed by Prof. Diether Lambrechts), which will support and perform the wet/dry lab procedures, subsequent sample processing and advanced bioinformatics analyses of this project. All data that will be collected and the strategy to guarantee the privacy of the study participants are specified in the research protocol approved by the ethical committee (S65227). We will generate transcriptomics sequencing data from native kidney biopsies of patients with FSGS that were biopsied at UZ Leuven. Next, we will validate candidate diagnostic markers on the RNA level (bulk RNA-seq, RT-PCR and/or *in situ* hybridization on kidney biopsy tissue and/or serum/urine) and protein level (IF, IHC on kidney biopsy tissue) from patients in the validation cohort (patients included in LRRB). Candidate markers/gene expression signatures generated within this project will be correlated with pseudonymized clinical data (e.g. disease remission/relapse, renal survival). The transcriptomics/proteomics data and associated pseudonymized patient information are defined as sensitive personal data and will be processed in accordance with the institutional SOPs, the principles of the General Data Protection Regulation (GDPR) 2016/679 and the Belgian privacy law. These procedures include procedures for pseudonymization, data storage and data protection.

* Data handling of LRRB samples at the Nephrology and Renal Transplantation Research Group:
All human samples included in LRRB (kidney, urine, blood) are pseudonymized and labelled with the following information: ‘nefro ID’, date of collection, type of biological sample, study and barcode number identical for all the biological samples collected at a same timepoint of a participant. Demographic, clinical and histology data come directly from the included patient files. These data are stored in SAS format or .xls format (demographic and clinical data) or in Microsoft Access format (histology data) on the UZ Leuven server of the Department of Nephrology, UZ Leuven, which benefits from the firewall and back-up services provided by the UZ Leuven IT department. Digital images will be saved and stored on the Large Volume Storage platform of ICTS KU Leuven, which also incorporates daily back-up services and version logging. All clinical data, scan data and metadata included, or generated in this study, will be pseudonymized. Patients in UZ Leuven are informed about the possibility to be included in retrospective studies through the hospital admission brochure, and about their inclusion in this specific study through the Mynexuzhealth-app of UZ Leuven.

* Data handling at Laboratory of Translational Genetics (LTG) (VIB – KUL center of cancer biology Leuven):
An MTA between UZL/KUL and LTG has been completed. All human samples arriving in LTG or processed/stored in LTG are already pseudonymized with the 'nefro ID' and are subsequently labeled with a DILA-ID, keeping the identity of the study participant private and confidential (samples are therefore double coded). This DILA ID is further used in the downstream analyses at LTG. The lab only receives pseudonymized human data (including health data linked to the sample IDs) in .xls format. All electronical data will be processed and (temporarily) stored on secured, password-protected and backed-up servers of VIB-KU Leuven (managed by ICT of the Biomedical Sciences Group) which are protected by a genuine user authentication system relying on username and password. Access to the data as well as the access level will be limited on a project need and individual basis. Only the researchers working on the project have access to these data.

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

Flemish Collaborative Glomerulonephritis Group (FCGG) Database
Sponsor: UZ Leuven
S-number: S59182
Approved version: V3 (12th of December 2016)
Applied in WP: 1, 2

Leuven Renal Research Biobank (LRRB)
Sponsor: UZ Leuven
S-number: S54095
Approved version: amendement 2 (5th of November 2020)
Applied in WP: 3, 4

Integrative single-cell analysis of glomerular disease: the application of single-cell methods on samples of the Leuven Renal Research Biobank (LRRB)
Sponsor: UZ Leuven
S-number: S65227
Approved version: V1 (16th of February 2021)
Applied in WP: 3, 4

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?

- No

WP1-2 consist of epidemiological studies with no potential for valorisation.

WP3 consists of an exploratory fundamental study, while WP4 only consists of a preliminary first validation of candidate biomarkers. We do not expect that our research data will have potential for tech transfer and valorisation, as candidate diagnostic markers will first need to be validated in an independent study cohort, which is beyond the scope of this project.

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

4. DOCUMENTATION AND METADATA

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

WP1-2: Epidemiological analysis of native kidney diseases in Flanders and profile of the Flemish FSGS cohort

An explanatory ReadMe-sheet and log-sheet will be included in the .xlsx-file of the FCGG database extracts. Analytic information on calculations will be kept in separate .xlsx-files. Manuscripts are kept on KUL OneDrive which automatically keeps a version history. Appropriate file naming will be applied on all files (e.g. description_version_date_researcherinitials).

WP3-4: Single-nucleus RNA-sequencing on kidney biopsies of FSGS patients and validation of candidate markers in the FSGS cohort

Documentation will consist of notes in the Electronic Laboratory Notebook that refers to specific datasets to ensure that data files, lab samples, and experimental notes remain properly linked to the same study sample ID. These notes will describe the biological/clinical samples used, experimental setup and protocols used, sequences generated, links to the specific computer location and the specific names of the respective datasets. Metadata sheets are maintained with the connection between lab samples, sample IDs and files in the data storage so that lab samples, data files and experimental notes remain properly linked to the corresponding samples IDs. Research methods and practices (SOPs) are fully documented. When wet lab techniques, scripts, algorithms and software tools are finalized, they are additionally described in manuscripts and/or on GitHub. Raw sequencing data (.fastq and .bcl files; each named with their sample ID) will be collected per sequencing run, including an .xlsx file with the sample sheet information containing the sample IDs sequenced in that run and the sequencing run information per sample ID (Illumina sequencer, lane and index information). The name of the folder will contain the date of the sequencing run and the Illumina sequencer used.

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.

- Yes

WP1-2: Epidemiological analysis of native kidney diseases in Flanders and profile of the Flemish FSGS cohort

No organizational metadata standard is available. Nephrological diagnoses in the FCGG registry are coded with the ERA Primary Renal Disease (PRD) diagnostic coding system. ERA PRD is the standard in end-stage kidney disease registries and is increasingly used for kidney biopsy registries, which facilitates comparison between studies.

WP3-4: Single-nucleus RNA-sequencing on kidney biopsies of FSGS patients and validation of candidate markers in the FSGS cohort

Sequencing data types require specific metadata when submitted to public repositories such as EGA, ArrayExpress, GEO or ENA. Data documentation will be tailored to their ultimate deposition in public repositories, with spreadsheet headers corresponding to fields required by these public repositories. Technical and analytical methods used to generate the data will be documented in sufficient detail to allow for independent reproduction. These will include analysis package version numbers, analysis kit, disease status, treatment type and duration, organism, genome build, ... For single-cell experiments, each droplet barcode will also be retained alongside the associated single-cell quality metrics. When depositing data in a repository, the final dataset will be accompanied by this information in the file format that the repository provides. This will allow the data to be understood by other members of the laboratory and add context to the dataset for future reuse. Metadata used in the validation experiments depends upon the chosen methodologies (bulk RNA-seq *vs*. RT-PCR *vs. in situ* hybridization; IF *vs*. IHC) and will be determined later on in the project.

5. DATA STORAGE AND BACKUP DURING THE FWO PROJECT

Where will the data be stored?

Data will always be stored in accordance with the KUL-ICTS storage indicator (level of confidentiality, https://icts.kuleuven.be/sc/english/storage/storageguide). Patient data are always pseudonymized.

WP1-2: Epidemiological analysis of native kidney diseases in Flanders and profile of the Flemish FSGS cohort

The FCGG database extracts consist of strictly confidential patient data and are kept on the password-protected UZ Leuven network environment. Offline storage will always be done on encrypted media. Less confidential data will be stored on KUL OneDrive and (temporarily) on local hard disk.

WP3-4: Single-nucleus RNA-sequencing on kidney biopsies of FSGS patients and validation of candidate markers in the FSGS cohort

All electronical data collected and generated during the project will be processed and (temporarily) stored on secured, password-protected and backed up servers of KU Leuven (managed by ICT of the Biomedical Sciences Group).

The sequencing data generated during the project will either be stored on KU Leuven servers or on the Flemish Supercomputer Centre (VSC), initially in the staging and archive area, and later only in the archive area (archive is mirrored).
All biological samples (tissue, blood, urine, single-nuclei suspensions, sequencing library preparations) are stored in labelled tubes or SBS plates in -20°C or -80°C freezers purchased by our own funding. The samples will be registered and handled according to the UZ Leuven Biobank guidelines.

How is backup of the data provided?

WP1-2: Epidemiological analysis of native kidney diseases in Flanders and profile of the Flemish FSGS cohort
The UZ Leuven server of the Department of Nephrology benefits from automatic back-up services provided by the UZ Leuven IT department. Digital images will be saved and stored on the Large Volume Storage platform of ICTS KU Leuven, which also incorporates daily back-up services and version logging. KUL OneDrive also provides automatic back-up service with a version history and edit-log.

WP3-4: Single-nucleus RNA-sequencing on kidney biopsies of FSGS patients and validation of candidate markers in the FSGS cohort
The KU Leuven server benefits from automatic back-up services provided by ICT of the Biomedical Sciences Group. All sequencing data stored on the Flemish Supercomputer Centre (VSC) will be regularly transferred to the archive area that is mirrored.

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

There is sufficient storage and back-up capacity on the KUL OneDrive server and all UZ Leuven and KU Leuven servers:

- KUL OneDrive provides 2TB storage for non-confidential data, which is estimated to be adequate for the entire research project.
- The KUL/UZL “L-drive” is an easily scalable system, has unlimited maximal size and is expandable in blocks of 5TB.
- The KUL/UZL “J-drive” is an easily scalable system, has unlimited maximal size and is expandable in blocks of 100GB.
- The Staging and Archive on VSC are also sufficiently scalable (petabyte scale).

What are the expected costs for data storage and back up during the project? How will these costs be covered?
The total estimated cost of data storage during the 4 years of this FWO project is ~3,000 EUR. This estimation is based on the following costs:

- The costs of digital data storage are as follows: €868,9/5 TB/Year for the “L-drive” and €519/TB/Year for the “J-drive”.

- The cost of VSC archive is €70/TB/Year and staging €130/TB/Year.

We expect costs to drop slightly during the coming four years. The costs for data storage will be covered by our own funding.

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

The UZ Leuven server of the Department of Nephrology is password-protected and benefits from the firewall services provided by the UZ Leuven IT department. Access to the data as well as the access level will be limited on a project need and individual basis.

Data stored on the KU Leuven IT infrastructure is protected by a genuine user authentication system relying on username and password. Access to the data as well as the access level will be limited on a project need and individual basis. Only the researchers working on the project have access to these data. Personal data is always pseudonymized. The coding key to patient information of linked pseudonymized data does not carry any personal identifiers and all records containing the identity of each participant will be kept private and confidential.

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, ...).

WP1-2: Epidemiological analysis of native kidney diseases in Flanders and profile of the Flemish FSGS cohort

The source data (in FCGG registry, managed by NBVN), as well as datasets 1.1 (FCGG database extracts), 1.2 (statistical analyses) and 1.3 (manuscripts, supplementals, figures) will all be kept for a minimum of 5 years after the end of the project.

WP3-4: Single-nucleus RNA-sequencing on kidney biopsies of FSGS patients and validation of candidate markers in the FSGS cohort

All raw and processed data collected in the scope of this FWO project will be retained for the expected 5-year period after the end of the project. All (remaining) biological samples are preserved for 50 years, in accordance with the guidelines of Biobank UZ/KU Leuven.

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

As a general rule, datasets will be made openly accessible, whenever possible via existing platforms that support FAIR data sharing (www.fairsharing.org), at the latest at the time of publication or preprint deposition.

For all other datasets, long term storage will be ensured as follows:

- Large sequencing data will be stored on VSC archive.

- Small digital files and FCGG database extracts will be stored on the KUL/UZL “L-drive”.
- Developed algorithms and software will be stored on VSC archive and/or L-drive, as well on public repositories such as Github.com.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?
The total estimated cost of data storage for 5 years after the end of the project is ~ €350. This estimation is based on 1 TB in total, at 70EUR/Tb/year. The storage after the project is much smaller because during the project a large working space is needed, and post-publication data will be made accessible via open access platforms. All costs for data preservation will be covered by our own funding.

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

Which data will be made available after the end of the project?
The promotor (and co-promotors) of this project are committed to publishing scientific research in order to communicate results both to peers and a wider audience. All research outputs supporting publications will be made openly accessible, at the latest, at the time of publication (or preprint deposition) via the required link in the publication or upon reasonable request and after an embargo period after publication.
Data that will be made available include:

- Double-coded raw sequencing data. Personal data will always be double coded and no reference to subject name will be made.
- Scripts, algorithms and software tools.

The results will be published as BioRxiv preprints and as Open Access in peer reviewed journal. Upon reasonable request, data will be reused by transfer through Belnet Filesender or secure copy.

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

- In an Open Access repository
- In a restricted access repository
- Upon request by mail
- Other (specify):

Whenever possible, datasets and appropriate metadata will be made publicly available through repositories that support FAIR data sharing. Personal data will be double coded and no reference to subject name will be made.
Sharing policies for specific research outputs are detailed below:

- Double-coded raw sequencing data (linked to double-coded patient data) will be deposited in open access repositories with restricted access control such as the EBI European Genome-phenome Archive (EGA). The EGA is a repository for personally identifiable genetic and

phenotypic data. Sequencing data at EGA will only be available upon reasonable request via our institutional data access committee and, if necessary, a material transfer agreement will be concluded with the beneficiaries in order to describe the types of reuse that are permitted. The double-coded read count data matrix (linked to double-coded patient data) will also be available on an interactive webserver (http://blueprint.lambrechtslab.org).

- Double-coded patient data: Upon publication, all double-coded patient details supporting a manuscript will be made publicly available as supplemental information.
- Research documentation: All protocols used to generate published data will be described in the corresponding manuscript(s), and the related documentation will be included as supplementary information. These data and all other documents (raw data) deposited in the E-Notebook are accessible to the research staff and will be made available upon request.
- Manuscripts: All scientific publications will be shared openly. Manuscripts submitted for publication will be deposited in a pre-print server such as bioRxiv. At the time of publication, research results will be summarized on the (co-)promoters' websites (https://www.vibcancer.be/diether-lambrechts) and post-print pdf versions of publications will be made available there if allowed by copyright agreements, possibly after an embargo as determined by the publisher. Before the end of the embargo or in cases where sharing the post-print is not allowed due to copyright agreements, a pre-print version of the manuscript will be made available. (Pre-print) publications will also be automatically added to our institutional repository, Lirias 2.0, based on the authors name and ORCID ID.
- Algorithms, scripts and software: All the relevant algorithms, scripts and software tools driving the project will be described in manuscripts and/or on GitHub (https://github.com).
- Data that do not support publication will be either deposited in an open access repository or made available upon request by email. Data will be reused by transfer via Belnet Filesender or secure copy.

When will the data be made available?

- Upon publication of the research results

All research outputs will be made openly accessible, at the latest, at the time of publication (or preprint deposition). No embargo will be foreseen unless imposed e.g. by pending publications.

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

Whenever possible, datasets and the appropriate metadata will be made publicly available through repositories that support FAIR data sharing. As detailed above, metadata will contain sufficient information to support data interpretation and reuse, and will be conform to community norms. These repositories clearly describe their conditions of use (typically under a Creative Commons CC0 1.0 Universal (CC0 1.0) Public Domain Dedication or an ODC Public Domain Dedication and License, with a material transfer agreement when applicable). Interested parties will thereby be allowed to access data directly, and they will give credit to the authors for the data used by citing the corresponding DOI. A CC-BY license will be opted for when possible. For data shared directly by the PI, a material transfer agreement (and a non-disclosure agreement if applicable) will be concluded with the beneficiaries in order to clearly describe the types of reuse that are permitted.

For KU Leuven data submitted to the EBI European Genome-phenome Archive (EGA), which operates under controlled access, the data access/submission requests will be received by the Genomics Data Access Committee (DAC) of KU Leuven (https://homes.esat.kuleuven.be/~bioiuser/dac/) and processed in consultation with the PIs produced data. The DAC will provide general guidance in terms of policies and will be referred to in handling controversial cases.

What are the expected costs for data sharing? How will the costs be covered?
It is the intention to minimize data management costs by implementing standard operating procedures (SOPs) e.g. for metadata collection and file storage and organization from the start of the project, and by using free-to-use data repositories and dissemination facilities whenever possible.
All data management costs will be covered by own funding.

8. RESPONSIBILITIES

Who will be responsible for data documentation & metadata?

WP1-2: Epidemiological analysis of native kidney diseases in Flanders and profile of the Flemish FSGS cohort
(Meta)data will be documented by NBVN (FCGG database) and the research team (FCGG database extracts).

WP3-4: Single-nucleus RNA-sequencing on kidney biopsies of FSGS patients and validation of candidate markers in the FSGS cohort
(Meta)data will be documented by the research and technical staff at the time of data collection and analysis, by taking careful notes in the E-notebook that refer to specific datasets. The research and technical staff include the PhD student(s), technical assistants and bio-informaticians directly involved with this research project.

Who will be responsible for data storage & back up during the project?
The research and technical staff will ensure data storage and back up, with support from ICTS, gbiomed-IT and UZ-IT staff. Final responsibility for data storage and back-up lies with the promotor of this project, supported by ICTS, HPC, gbiomed-IT staff and UZ-IT staff.

Who will be responsible for ensuring data preservation and reuse?
The research and technical staff will ensure data preservation and sharing, with support from ICTS, gbiomed-IT staff and UZ-IT staff. Final responsibility for ensuring data preservation and sharing lies with the promotor of this project, supported by ICTS, HPC, gbiomed-IT staff, and UZ-IT staff.

Who bears the end responsibility for updating & implementing this DMP?
The promotor of this FWO project carries the end responsibility for updating & implementing this DMP, supported by the research team.