# Uncovering the role(s) of ABCA7 transporter in Alzheimer's disease

*A Data Management Plan created using DMPonline.be*

**Creators:** Nicole Wawrzyniak, Lucia Chávez-Gutiérrez, Kristel Sleegers

**Affiliation:** KU Leuven (KUL)

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

**Template:** FWO DMP (Flemish Standard DMP)

**Principal Investigator:** Kristel Sleegers, Lucia Chávez-Gutiérrez

**Grant number / URL:** 11PFW24N

**ID:** 203220

**Start date:** 01-11-2023

**End date:** 31-10-2027

**Project abstract:**

Genome wide association studies pinpointed a strong genetic component of sporadic Alzheimer's disease (AD). Premature termination codon (PTC) variants and rare missense mutations in ATP binding cassette transporter subfamily A member 7 (ABCA7) have recently been shown to confer AD risk, suggesting a central role of lipid homeostasis in AD. However, the mechanism of their pathogenicity remains unclear. Here, I propose a project combining the advances in genetic research with state–of–the-art molecular biology and nanobody (Nb) technology aimed at understanding the role of ABCA7 in AD risk. Accordingly, I will decipher the effects of selected AD-linked and protective variants on ABCA7 biology and investigate the endogenous interactome of wild type and mutant ABCA7 to uncover pathways in which ABCA7 operates to modulate AD risk. To address these points, I will generate immunoprobes – Nbs, synthetic Nbs (Sybodies) and antibodies. I will assess in vitro ABCA7 ATPase, floppase and lipid transport activities of the variants and explore their (sub)cellular expression in post mortem brain samples. I will employ innovative Nb-based proximity biotinylation with Nb-TurboID and proteomic analysis for the discovery of ABCA7 interactome in human brain. In silico and transcriptomics approaches will be used to put interactomics data into biological context and to identify pathways affected by ABCA7 variants in AD. The outcomes will provide key insights into ABCA7-related pathophysiology.

**Last modified:** 11-04-2024

# Uncovering the role(s) of ABCA7 transporter in Alzheimer's disease
# 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:*<br>• Generate new data<br>• Reuse existing data | *Please choose from the following options:*<br>• Digital<br>• Physical | *Please choose from the following options:*<br>• Observational<br>• Experimental<br>• Compiled/aggregated data<br>• Simulation data<br>• Software<br>• Other<br>• NA | *Please choose from the following options:*<br>• .por, .xml, .tab, .csv,.pdf, .txt, .rtf, .dwg, .gml, …<br>• NA | *Please choose from the following options:*<br>• <100MB<br>• <1GB<br>• <100GB<br>• <1TB<br>• <5TB<br>• <10TB<br>• <50TB<br>• >50TB<br>• NA | |
| Observational dataset 1- Human tissue samples | Frozen post-mortem brain tissue samples from humans | Reuse existing data | Physical | Observational | | | 60 grams |
| Experimental dataset 1<br>- Cell lines | Bacteria strains, insect cell lines, mammalian cell lines | Generate new data | Physical | Experimental | | | 100 ml |
| Experimental dataset 2- Vectors | Vectors for protein expression in bacteria, mammalian and insect cells | Generate new data | Physical | Experimental | | | 20 ml |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Experimental dataset 3- Recombinant proteins and compounds | Human proteins (transporters and their parts, newly discovered interactors) produced in insect or human cell lines, nanobodies expressed in bacteria, chemical inhibitors and modulators of transporters | Generate new data | Physical | Experimental | | | 6 liters |
| Experimental dataset 4 - Canonical data | Transcriptomic dataset, bulk and single nuclei RNA sequencing | Generate new data<br>Reuse existing data | Digital | Experimental | - Raw sequence data trace (.ab1)<br>- Text-based format (.fasta/.fa) and accompanying QUAL file (.qual)<br>- Sequence and quality score (.fastq)<br>- Genbank format (.gb/.gbk)<br>- Sequence alignment data: (.sam), (.bam), (.cram)<br>- Gene Transfer Format: (GTF) / Gene Feature Format: (GFF) | < 5 TB | |
| Experimental dataset 5 - Mass spectrometry data | Mass spectrometry datasets | Generate new data | Digital | Experimental | - Raw data (.raw)<br>-Quantitative tabular data: comma-separated value files (.csv)<br>- MS Excel (.xls/.xlsx), | < 5 TB | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Experimental dataset 6 - Digital images | Microscopy pictures, gel scans, graphs, illustrations, figures | Generate new data | Digital | Experimental | - Digital images in raster formats: uncompressed TIFF (.tif/.tiff), JPEG (.jpg), Adobe Portable Document Format (.pdf), bitmap (.bmp), .gif, .lif (Leica), .czi (Zeiss), .gel (ImageQuant)<br>- Digital images in vector formats: scalable vector graphics (.svg), encapsulated postscript (.eps), Scalable Vector Graphics (.svg), Adobe Illustrator (.ai) | < 1 TB | |
| Derived and compiled dataset 1 – Research documentation | Documentation generated by the research and technical staff or collected from online sources and from collaborators, including ethical approval documents, laboratory notes and protocols. | Generate new data | Digital | Compiled/aggregated data | -Text files: Rich Text Format (.rtf), MS Word (.doc/.docx), Adobe Portable Document Format (.pdf)<br>-Quantitative tabular data: comma-separated value files (.csv), MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb) | < 1 TB | |
| Derived and compiled dataset 2 - Manuscripts | Manuscripts resulting from the project | Generate new data | Digital | Compiled/aggregated data | -Text files: Rich Text Format (.rtf), MS Word (.doc/.docx), Adobe Portable Document Format (.pdf) | < 1 TB | |

**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:**

Single nuclei RNA sequencing data of brains of AD patients with and without ABCA7 mutations and controls generated in the laboratory of Kristel Sleegers.
Bulk brain RNASeq data from the AMP-AD consortium AD Knowledge Portal (synapse.org) (this includes n=46 ABCA7 mutation carriers; access granted to the KS lab).
We will reuse part of the physical datasets that were generated in laboratory of Lucia Chavez Gutierrez in previous projects, such as bacteria strains, mammalian cell lines, nanobodies and vectors for expression.

**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

Post-mortem human tissue will be used in this project and it is approved by EC at KU Leuven under the number S60470. Human subject data will be used in this project, particularly in the collection or creation of ABCA7 interactome in human brain dataset.

Data re-use:
We will be re-using sequencing data generated by the laboratory of Kristel Sleegers and others. In doing so, we will adhere to the Data Transfer Agreement (DTA) and
the Data Access Committee (DAC) guidelines of the data provider.

**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

I will reuse personal data (genetic information, age, gender, disease status) of the human brain samples donors.
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.
The university of Antwerp maintains a “processing of personal data” registry for research involving personal data in order to comply with the GDPR. Formal approval by the Data Protection Office (DPO) has been granted: project reference number: Antigoon ID 49828, date of approval March 7, 2024

**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

We do not exclude that the proposed work could result in research data with potential for tech transfer and valorization. Ownership of the data generated belongs to KU Leuven/University of Antwerp and VIB in accordance with the framework agreement of both institutes. VIB has a policy to actively monitor research data for such potential. If there is substantial potential, the invention will be thoroughly assessed, and in a number of cases the invention will be IP protected (mostly patent protection or copyright protection). As such the IP protection does 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 so that publications need not be delayed.

**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

3rd party agreements may restrict dissemination or exploitation of data in the following cases:
Data generated with cell lines obtained from repositories must adhere to restrictions on data dissemination specified in the material transfer agreement.

Data generated with samples obtained from humans are subject to patient consent forms and, if applicable, material transfer agreements, and the therein specified restrictions on data dissemination and/or exploitation.

**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

If the restriction of data sharing is a consequence of securing Intellectual Property (IP), the researcher involved and the IP team of the TechTransfer office shall make the necessary arrangements in order to maintain the embargo on the public access (dissemination) of research data, at least until the essential steps in securing intellectual property (e.g. the filing of a patent application) have been taken. However, reasonable efforts will be made to avoid delays in publication. If the restriction of data sharing is a consequence of contracts with third parties or about classified data, the researcher involved and the Legal Team of the TechTransfer Office will specify the restrictions in this DMP. If the restriction of data sharing is a consequence of ethical issues, see also the Ethical and legal session of this DMP. Personal data will only be published after de-identification and identifiers will not be published. If despite all efforts it is not possible to protect the identities of subjects even after removing all identifiers, personal data will not be made public.

#### 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).**

Data will be generated following standardized protocols. Clear and detailed descriptions of these protocols will be stored in our lab protocol database, and published along with the results. Metadata will be documented by the research and technical staff at the time of data collection and analysis, by taking careful notes in the electronic laboratory notebook (E-notebook) and/or in hard copy lab notebooks that refer to specific datasets.
Cryotubes of biological samples (bacterial strains, mammalian cell lines) stored at -80°C will be labelled with a reference number that links to an entry or database.
All datasets will be accompanied by a README.txt file containing all the associated metadata (see more details below).

**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

While specific data types might require particular metadata, as a general rule the metadata are based on a generalized metadata schema such as Dublin Core or DataCite, including the following elements:
• Title: free text
• Creator: Last name, first name, organization
• Date and time reference
• Subject: Choice of keywords and classifications
• Description: Text explaining the content of the data set and other contextual information needed for the correct interpretation of the data, the software(s) (including version number) used to produce and to read the data, the purpose of the experiment, etc.
• Format: Details of the file format,
• Resource Type: data set, image, audio, etc.
• Identifier: DOI (when applicable)
•Access rights: closed access, embargoed access, restricted access, open access.

For specific datasets, additional metadata will be associated with the data file as appropriate.
Specific naming and search procedures will be applied to:
- Omics and sequencing data: Digital files, raw sequencing data files mainly (.fastq, .fastq.gz), will be named following an in-house procedure, so that all the name of all files in a given dataset will be in the same format. All names will start with a project 3 letter code (Project Code) and

a random 6-alphanumeric character code (unique to each sample given the Project Code), followed by a specific descriptive name of the sample and the technology used (e.g.: 10x). Names will only contain letters, numbers and underscores.

- Manuscripts: Metadata information will be submitted alongside the final version of the manuscript, including the names, titles, email addresses, ORCIDs and affiliations of all authors. Upon publication, this metadata information will also be submitted to bibliographic databases such as Medline. All manuscripts will be assigned a unique Digital Object Identifer (DOI) by the publisher. Manuscripts will be given a descriptive title, and will be accompanied by keywords provided by the authors in order to maximize their findability.

- Mass spectrometry metadata will be generated at the VIB Proteomics Core facility. A copy of the results (data) will be stored at the VIB-KU Leuven Proteolytic Mechanisms mediating Neurodegeneration laboratory.

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

**Where will the data be stored?**

Digital files are stored on KU Leuven servers Shared network drive (J-drive) and Large Volume Storage (L-drive).

Other types of data are stored in different forms as follows;

- Tissue samples: Tissues are stored locally in the laboratory. All human tissue samples are registered in VIB-UA biobank, in compliance with the Belgian law on human body material (dd 19-12-2008).
- Vectors: Vectors will be preserved in a form of purified DNA (in -20°C freezer).
- Cell lines: Cryo preserved cell lines will be stored locally in the laboratory in liquid nitrogen tank. At least two vials per cell line derived from independent freezings will be stored.
- Other biological and chemical samples: storage at 4°C and/or as frozen samples as appropriate.
- Nucleic acid and protein sequences: All nucleic acid and protein sequences generated during the project will be stored on KU Leuven servers or internally on the servers of the VIB CMN. These servers have RAID 6 or equivalent disk setup for protection of the data.
- Proteomics data: Proteomics data generated during the project will be stored on KU Leuven servers.
- All other digital data (research documentation and manuscripts) will be stored on KUL servers.
- The ELN database is managed by VIB Headquarters. VIB uses E-Notebook by PerkinElmer. All ELN data are stored, and remain available for (re)use. Sanger sequencing data is stored on the centralized LIMS system., initially in the staging area and later in the archive area.

**How will the data be backed up?**

KU Leuven drives are backed-up according to the following scheme:

- Data stored on the “L-drive” is backed up daily using snapshot technology, where all incremental changes in respect of the previous version are kept online; the last 14 backups are kept.
- Data stored on the “J-drive” is backed up hourly, daily (every day at midnight) and weekly (at midnight between Saturday and Sunday); in each case the last 6 backups are kept.

The VIB CMN LIMS is submitted to daily back-ups on an offsite server, as well as to daily internal back-ups (RAID6). Desktops at VIB CMN are submitted to daily backup through Retrospect. Local data storage servers at VIB CMN are synchronized for active backup purpose using snapshot technology protecting against cryptolockers.

Raw data generated at VIB CMN are stored following the US-CERT recommended back up practice 3-2-1 (3 copies of the data, 2 different storage media, 1 off site copy), including separate back up on tape for long term archival. All servers have RAID6 (or equivalent) redundancy.

Specifically for omics data, VIB CMN has adopted US-CERT recommended backup practice 3-2-1 (n-d-o) for omics data. We have three copies of the data, two different storage media and one offsite copy. Our servers have a direct in-house online backup, a third copy (n = 3) is stored on a tape (d = 2). At this moment the data remains onsite, but in a different room (o = 1).

The ELN database is managed by VIB Headquarters, and backed-up every 24 hours. All ELN data are also stored on tape.

**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

For the data stored at CBD:

There is sufficient storage and back-up capacity on all KU Leuven servers:

- The “L-drive” is an easily scalable system, built from General Parallel File System (GPFS) cluster with NetApp eseries storage systems, and a CTDB samba cluster in the front-end.

- The “J-drive” is based on a cluster of NetApp FAS8040 controlers with an Ontap 9.1P9 operating system.

For the data stored at CMN:

There is +/- 700TB of online storage available for the use in house. It is increased yearly to the needs of the center. For offline backup, tapes are used that store 2.5TB/tape. These have an expected life span of 15-30 years.

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

For the data stored at CBD:

Security of digital datasets in KU Leuven drives: Both the “L-drive” and “J-drive” servers are accessible only by laboratory members and can only be accessed after login (username + password), and are mirrored in the second ICTS datacenter for business continuity and disaster recovery so that a copy of the data can be recovered within an hour.

Security of the physical datasets that are stored in the laboratory unit are ensured by allowing physical access only to the laboratory members through KU Leuven ID card access.

For the data stored at CMN: physical access to the building (building V – CDE) is employee badge-protected.

At VIB CMN, data is protected by domain controlled authorization and authentication. The data can only be accessed after login (username + password) and can be further restricted per project and per group. Researchers have access only to de-identified information. The building is restricted by badge system so only employees are allowed in and visitors are allowed under supervision after registration. Our LIMS system uses Role Based Access Control (RBAC), maintained and verified by PI and HR. Communication with the LIMS database is encrypted and follows secured https.

Only the PI and medical team members will be granted access to the server to deposit private data. The PI and medical team members will be the only responsible for linking patient information and/or samples, and will strictly respect confidentiality.”

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

The total cost of data storage during the project is projected to be max. 1000 € per year. This includes the following costs:

The costs of digital data storage are as follows: 174 € /TB per year for the “L-drive” and 519 € /TB per year for the “J-drive”.

These costs will be covered by the VIB dotation budget assigned to the laboratory of Lucía Chávez-Gutiérrez.

For the transcriptomic datasets and human brain samples stored at University of Antwerp Biobank, the storage costs for digital data on the departmental server is covered by the central departmental budget. Electricity costs for the -80° freezers present in the labs are included in general lab costs. Data storage and backup costs are included in general lab costs of laboratory of Kristel Sleegers.

**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 preservation term of at least 5 years after the end of the project will be applied to all datasets except for the human brain tissue. The human brain tissue will be used in the project for various protocols described in the project outline, and all the sample volume can be used up during this process.

**Where will these data be archived (stored and curated for the long-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.

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

- Digital datasets: files will be stored on the “L-drive”.
- Tissue samples: tissues will be stored locally in the laboratory (-80°C) or in University of Antwerp biobank.
- Vectors: vectors are preserved in the form of purified DNA (in -20°C freezer).

- Cell lines: cell lines are stored locally in the laboratory (liquid nitrogen).

Other biological and chemical samples: storage either at 4°C and/or as frozen, as appropriate, locally in the laboratory.

For the transcriptomics data generated with the lab of Kristel Sleegers, the data are stored following 3-2-1 principle for the foreseeable future. In case data would be archived, we would duplicate our tape and store this in the optimal environment, which has an estimated lifespan of 15-30 years.

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

The total estimated cost of data storage during 5 years after the end of the project is 10000 euro. This estimation is based on the following costs:

At VIB CMN, cost for storage on tape is 44€/Tb, and 9€/Tb/Year for upkeep. Cost of retrieval from tape is 29€. Costs for biobanking at UZA biobank are 4€/sample. Costs will be covered from the Kristel Sleegers laboratory budget.

Electricity costs for the -80° freezers present in the labs are included in general lab costs.

Data storage and backup costs are included in general lab costs of the Kristel Sleegers lab.

The costs of digital data storage are as follows: 569,2€/5TB/Year for the “L-drive”, 519€/TB/Year for the “J-drive

These costs at VIB CBD will be covered by the VIB dotation budget assigned to the laboratory of Lucía Chávez-Gutiérrez.

**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, …)

All research outputs supporting publications will be made openly accessible after the project.

We aim at communicating our results in top journals that require full disclosure upon publication of all included data, either in the main text, in supplementary material or in a data repository if requested by the journal and following deposit advice given by the journal. Depending on the journal, accessibility restrictions may apply.

Sharing policies for specific research outputs are detailed below:

- Physical experimental datasets described above (cell lines, vectors, recombinant proteins and compounds) will be stored as appropriate in the laboratory. Within availability, they will be reused during and after the project.
- Digital experimental datasets (nucleic acid and protein sequences, images, and spectrometry data): they are stored on the servers and will be made available for reuse.
- Pseudonomyzed personal data will be made available through restricted access repositories such as EGA or dbGap, according to Data Protection by Design and by Default principles. In order to respect the patient’s privacy, clinical samples will only be available to the research and technical staff involved in the project, not to other groups, studies or purposes.
- 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 (daily logs, raw data) deposited in the E-Notebook are accessible to the PI and the research staff, for reuse during/after project.

Manuscripts: All scientific publications will be shared openly. At the time of publication, research results will be summarized on the VIB website and post-print pdf versions of publications or links to them 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.

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

We aim at communicating our results in top journals that require full disclosure upon publication of all included data, either in the main text, in supplementary material or in a data repository if requested by the journal and following deposit advice given by the journal. Depending on the journal, accessibility restrictions may apply.

Any scientist will be able to access the physical data, upon request, after publication, or before publication upon agreement.

In order to respect the patient’s privacy, clinical samples will only be available to the research and technical staff involved in the project, not to other groups, studies or purposes.

Access to restricted access dataset (such as human genetic datasets) is governed by the Data Access Committees of VIB.

**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
- Yes, Ethical aspects

Protein sequences, eg. Nanobody - IP in case of a possible patent application will not be shared before patenting.

In order to respect the patient's privacy, clinical samples will only be available to the research and technical staff involved in the project, not to other groups, studies or purposes.

Human genetic data are considered sensitive personal data, and are are only made available on restricted access repositories such as the European Genome Phenome Archive (EGA) or dbGap. Access to these datasets is under control of a Data Access Committee.

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

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. Sharing policies for specific research outputs are detailed below:

- Vectors: vectors are stored as appropriate in the laboratory. Within availability, they will be shared with interested researchers upon request.
- Cell lines: cell lines are stored as appropriate in the laboratory. Within availability, they will be shared with interested researchers upon request.
- Other digital datasets that support publications (including image files, and mass spectrometry data) are stored on the servers and can be made available upon request.
- Antibodies, synthetic and recombinant compounds: samples are stored as appropriate in the laboratory. Within availability, they will be shared with interested researchers upon request.
- 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 (daily logs, raw data) deposited in the E-Notebook are accessible to the PI and the research staff, and will be made available upon request.
- Manuscripts: All scientific publications will be shared openly. At the time of publication, research results will be summarized on the VIB website and post-print pdf versions of publications or links to them 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. Publications will also be automatically listed in our institutional repository, Lirias 2.0, based on the authors name and ORCID ID.
- Nucleic acid and Protein sequences not patient-derived: they are stored on the KU Leuven servers and can be made available upon request.
- Pseudonomyzed personal data will not be openly published, but will be made available through restricted access repositories such as EGA or dbGap, according to Data Protection by Design and by Default principles. In order to respect the patient's privacy, clinical samples will only be available to the research and technical staff involved in the project, not to other groups, studies or purposes.
- Data that do not support publication will be made available upon request by email.

**When will the data be made available?**

After publication of the results in a peer reviewed journal or before publication upon a material transfer agreement. As a general rule all research outputs will be made openly accessible at the latest at the time of publication. No embargo will be foreseen unless imposed e.g. by pending publications, potential IP requirements – note that patent application filing will be planned so that publications need not be delayed - or ongoing projects requiring confidential data. In those cases, datasets will be made publicly available as soon as the embargo date is reached.

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

Whenever possible, datasets and the appropriate metadata will be made publicly available through repositories that support FAIR data sharing. For restricted-access repositories we will work with a Data Access Committee (DAC) from VIB, which provides guidance and oversight of

controlled-access data and systematically reviews data access requests to the data repository. Only after positive evaluation by the DAC will a user be able to access the deposited data (typically under a Creative Commons CC0 1.0 Universal (CC0 1.0) Public Domain Dedication, a Creative Commons Attribution (CC-BY) 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.

For data shared directly by the PIs, 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.

**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

Yes, upon publication DOI or other accession number will be assigned to the data.

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

It is the intention to minimize data management costs by implementing standard procedures 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. Data management costs will be covered by the laboratory budget.

Additionally, we aim for open access upon publication of the results and the research groups will cover the cost.

**6. Responsibilities**

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

Metadata will be documented by the fellowship holder - Nicole Wawrzyniak at the time of data collection and analysis, by taking careful notes in the electronic laboratory notebook (E-notebook) that refer to specific datasets.

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

The fellowship holder - Nicole Wawrzyniak will ensure data storage and back up supervised by the PIs, with support from René Custers and Alexander Botzki for the electronic laboratory notebook and from Raf De Coster for the KU Leuven drives.

**Who will manage data preservation and sharing?**

The PIs - Lucia Chavez Gutierrez and Kristel Sleegers will be responsible for data preservation and sharing, with support from René Custers and Alexander Botzki for the electronic laboratory notebook and from Raf De Coster for the KU Leuven drives.

**Who will update and implement this DMP?**

The fellowship holder - Nicole Wawrzyniak will be responsible to update and implement the DMP.