DMP - FWO - Fonds Wetenschappelijk Onderzoek - 1SE4523N

The participants to the present project understand the value of FAIR Data Management and Open Access to Scientific Publications and Research Data. They are fully committed to abiding by the related FWO policies.

The present data management plan (DMP) describes the specific outputs of this project and how they will be made available to the community. All participants will be informed of updates in the DMP, and any new participant will receive training to ensure compliance with the consortium’s data management conventions.

1. General Information	
Name applicant	Patrik Verstreken - patrik.verstreken@vib.be
FWO Project Number & Title	1SE4523N THERAPEUTIC TARGETING OF PATHOGENIC PRESYNAPTIC TAU IN HUMAN NEURONS IN A CHIMERIC ALZHEIMER'S DISEASE MOUSE MODEL
Affiliation	VIB-KU Leuven Center for Brain & Disease Research

Responsible: Jacqueline van Vierbergen

2. Data description	
Will you generate/collect new data and/or make use of existing data?	New data, Existing data

Describe the origin, type and format of the data (per dataset) and its (estimated) volume	Observational data Tissue samples Biological and chemical samples: Fixed and fresh frozen brain samples from mice, human iPSC and NPC in cryovails in liquid nitrogen tank Experimental data Digital images Microscopy pictures, gels, WB images, illustrations via Prism and figures via Matlab. Digital images in raster formats: uncompressed TIFF (.tif/.tiff), JPEG (.jpg), JPEG 2000 (.jp2), Adobe Portable Document Format (.pdf), bitmap (.bmp), .gif; Video and audio files microscopy movies Vectors GFP_Synapsin_CAAX Lentivral vector Cell lines H9-GFP- human embryonic cell line H9-GFP 3xHA tagged Synaptophysin - CRIPSR Knockin H9-GFP 3xHA tagged Synaptophysin SYNGR3 KO _ CRISPR Knock out

	Kolf2.1 - ipsc cell lines Kolf2.1 3xHA tagged Synaptophysin Kolf2.1 3xHA tagged Synaptophysin SYNGR3 KO Antibodies Use of commercially available antibodies. Simulation data Derived and compiled data Research documentation Research documentation generated by the research and technical staff or collected from online sources and from collaborators, including ethical approval documents, laboratory notes, protocols, animal husbandry data. Canonical data Nucleic acid sequences

	#231 pLenti-hSyn1-Tagmclover_CAAX_SV40-PuroR #SYNGR3_Gene editing_KO 2nd exon #SYP-3xHA after editing Files are stored in L drive and can be opened with Snapgene to see full sequence These datasets represent an important source of information for the laboratory of the P, for scientists, journalists and higher education teachers working in the field of Alzheimer disease, but also for industries active in the field of ASO development

3. Ethical and legal issues	
Will you use personal data? If so, shortly describe the kind of personal data you will use AND add the reference to your file in your host institution's privacy register.	No
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 Use of human embryonic stem cells, previously established and commonly used H9 cell line and and patient derived ipsc cell line Kolf2.1 03/10/2022 Research centres Ethics committee Research UZ / KU Leuven (ECResearch) - KU Leuven/UZ Leuven. : iPSCs, the iPSCs may be shared internationally to support research Transgenic mice will be used in this project, the mice line is already generated and fully characterized and being bred in the laboratory of my cohost lab 03/10/2022 Research centres Ethical Committee for Animal Experimentation (ECD) – KU Leuven/UZ Leuven. Animals are housed in facilities of the Laboratory Animal Center of KU Leuven, which applies Standard Operation Procedures concerning housing, feeding, health monitoring to assure consistent care in accordance with European and national regulations and guidelines. Animal administrative, husbandry and animal welfare data are sensitive data and are stored in the LAIS database according to security procedure of KU Leuven

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?	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 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. The use of <specify material> will be subjected to the terms described in their respective MTAs. Specific examples (adjust as required): -Potential biomarker/-target on microbiome data will be claimed if this opportunity arises. -For projects that involve the Verstrepen lab: The host lab identifies in an early phase the valorization potential of research lines and has a vast network of industrial contacts to efficiently start the route to commercialization. The lab is supported in this matter by Dr. Stijn Spaepen, IOF innovation manager responsible for research valorization. For research with valorization potential, the host lab actively protects its IP by filling patent applications. Type of data with potential for tech transfer and valorization: yeast strains isolated and generated during the timeframe of this project, sequencing information generated during the timeframe of this project and (analysis) data and models hereof derived. Valorization potential includes strain licensing, linking a specific sequence variant to a phenotype.

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 understanding and reuse of the data collected/generated in this project?	Data will be generated following standardized protocols. 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 and yeast strains) stored at -80°C will be labelled with a reference number that links to an entry in or strain database. All datasets will be accompanied by a README.txt file containing all the associated metadata (see more details below). The 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.

Will a metadata standard be used? If so, describe in detail which standard will be used. If not, state in detail which metadata will be created to make the data easy/easier to find and reuse.	No Metadata will include 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. Additionally, we will closely monitor MIBBI (Minimum Information for Biological and Biomedical Investigations) for metadata standards more specific to our data type. For specific datasets, additional metadata will be associated with the data file as appropriate. Give details as needed for the project. Specific examples (adjust as required): - SOPs for biological data generation are kept on a dedicated KU Leuven shared drive. A central excel file is stored on that same drive, detailing for examples: (1) sample ID; (2) SOP with which data generation was performed; (3) abnormalities or deviations from SOP in data generation; (4) experimental QC values (e.g. DNA concentrations); (5) location of the source sample in the freezer. - For bioinformatics processing, a data analysis log will be kept that details: (1) sequencing run ID; (2) the bioinformatics SOPs/scripts that were applied; (3) location of source files; (4) abnormalities or deviations. The final dataset will be accompanied by this information under the form of a README.txt document. This file will be located in the top level directory of the dataset and will also list the contents of the other files and outline the file-naming convention

	used. This will allow the data to be understood by other members of the laboratory and add contextual value to the dataset for future reuse.

5. Data storage & backup during the FWO project

Where will the data be stored?	Digital files will be stored on KU Leuven servers, except for private data that will be stored on KU Leuven secure server (digital vault). Add information about other data types as required (adjust and add more as needed): Mouse tissues will be stored locally in the laboratory - Vectors: As a general rule at least two independently obtained clones will be preserved for each vector, both under the form of purified DNA (in -20°C freezer) and as a bacteria glycerol stock (-80°C). All published vectors and the associated sequences will be sent to the non-profit plasmid repository Addgene, which will take care of vector storage and shipping upon request. - Newly created human cell lines will be stored locally in the laboratory in liquid nitrogen storage and will be deposited in the UZ Leuven-KU Leuven Biobank. Other human cell lines will be stored locally in liquid nitrogen cryostorage of the laboratory when actively used for experiments. Animal cell lines will be stored in liquid nitrogen cryostorage of the laboratory. -Mice will be maintained in facilities of the Laboratory Animal Center of KU Leuven, which applies Standard Operation Procedures concerning housing, feeding, health monitoring to assure consistent care in accordance with European and national regulations and guidelines. All animals will be registered in the Leuven Animal Information System (LAIS) database, along with corresponding genotyping information, ethical approval documents and animal provider receipts. Drosophila lines will be stored in a dedicated room and managed using a specific database for storage of the corresponding information (including genotype, origin, number of vials and date of transfer, crossing schemes) and vial tracking via unique QR codes. Other biological and chemical samples: storage at 4°C and/or as frozen samples in cryovials as appropriate. - Algorithms, scripts and softwares: All the relevant algorithms, scripts and software code driving the project will be stored in a private online git repository from the GitHub account of the department (https://github.com/vibcbd). - All nucleic acid sequences generated during the project will be stored on KU Leuven servers. Upon publication, all sequences supporting a manuscript will be made publicly available via repositories such as the GenBank database or the European

	Nucleotide Archive (nucleotide sequences from primers / new genes / new genomes),
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. - data stored on the digital vault is backed up using snapshot technology, where all incremental changes in respect of the previous version are kept online. As standard, 10% of the requested storage is reserved for backups using the following backup regime: an hourly backup (at 8 a.m., 12 p.m., 4 p.m. and 8 p.m.), the last 6 of which are kept; a daily backup (every day) at midnight, the last 6 of which are kept; and a weekly backup (every week) at midnight between Saturday and Sunday, the last 2 of which are kept. Incremental backups are done daily from one 20 TB QNAP NAS to a second 20 TB QNAP NAS.
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 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.

What are the expected costs for data storage and backup during the project? How will these costs be covered?	The total estimated cost of data storage during the project is (173*4+ 519*4 + 1200*4*2)=10568 This estimation is based on the following costs: -The costs of digital data storage are as follows: 173,78€/TB/Year for the “L-drive” and 519€/TB/Year for the “J-drive”. -Maintaining a mouse colony alive costs about 1,200 euro per year (for 6 cages), excluding the costs of genotyping. When no experiment is planned with a particular mouse strain, and in compliance with the 3R’s rule (https://www.nc3rs.org.uk), cryopreservation will thus be used to safeguard the strain, prevent genetic drift, loss of transgene and potential infections or breeding problems. Cryopreservation of sperm/embryos costs about 500 to 700 euro per genotype, plus a minimal annual storage fee (25 euro per strain for 250 to 500 embryos). Frozen specimen are kept in two separate liquid nitrogen tanks at two different sites on campus. When necessary, the costs of revitalization from cryopreserved sperm/embryos are about 1,100/600 euro. 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.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?	Both the “L-drive” and “J-drive” servers are accessible only by laboratory members, 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. Access to the digital vault is possible only through using a KU Leuven user-id and password, and user rights only grant access to the data in their own vault. Sensitive data transfer will be performed according to the best practices for “Copying data to the secure environment” defined by KU Leuven. The operating system of the vault is maintained on a monthly basis, including the application of upgrades and security patches. The server in the vault is managed by ICTS, and only ICTS personnel (bound by the ICT code of conduct for staff) have administrator/root rights. A security service monitors the technical installations continuously, even outside working hours. All private data will be rendered anonymous before processing outside the digital vault. Only the PI will be granted access to the server to deposit private data. The PI will be the only responsible for linking patient information, survey data and/or tissue samples, and will strictly respect confidentiality. All de-identified data will be exported from the database by the PI, and stored on KU Leuven servers from where it can be accessed by the research and technical staff from the laboratory. Databases are encrypted, password protected and within KU Leuven firewalls. Identification data will be kept separate from scientific data, and access to this separate data will be restricted to an absolute minimal number of persons. Participants are encoded with identifiable data stored in an independent, secured database to enable recontacting of participants. Participants have fully consented on this procedure.

6. Data preservation after the end of the FWO project FWO expects that data generated during the project are retained for a period of minimally 5 years after the end of the project, in as far as legal and contractual agreements allow.	
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, ...).	The minimum preservation term of 5 years after the end of the project will be applied to all datasets. All datasets will be stored on the university's central servers with automatic back-up procedures for at least 5 years, conform the KU Leuven RDM policy. The costs (€156 per TB per year for “Large volume-storage”) will be covered by the PI. If applicable:
Where will these data be archived (= stored 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. -Vectors: As a general rule at least two independently obtained clones will be preserved for each vector, both under the form of purified DNA (in -20°C freezer) and as a bacteria glycerol stock (-80°C). -Cell lines: human cell lines will be stored in the UZ Leuven Biobank (-80°C). Human pluripotent stem cell lines generated during this project will be deposited in hPSCreg. Animal cell lines will be stored in liquid nitrogen cryostorage of the laboratory. -Genetically modified organisms: Drosophila lines will be housed locally. All other lines that are no actively used for experiments will be cryopreserved. -Other biological and chemical samples: storage at 4°C and/or as frozen samples in cryovials as appropriate. - Following publication, the results associated with each study will also be deposited in the Dryad repository, where they will be preserved indefinitely.

What are the expected costs for data preservation during these 5 years? How will the costs be covered?	The total estimated cost of data storage during 5 years after the end of the project is 10.568. This estimation is based on the following costs: -The costs of digital data storage are as follows: 173,78€/TB/Year for the “L-drive” and 519€/TB/Year for the “J-drive”. -Maintaining a mouse colony alive costs about 1,200 euro per year (for 6 cages), excluding the costs of genotyping. When no experiment is planned with a particular mouse strain, and in compliance with the 3R’s rule (https://www.nc3rs.org.uk), cryopreservation will thus be used to safeguard the strain, prevent genetic drift, loss of transgene and potential infections or breeding problems. Cryopreservation of sperm/embryos costs about 500 to 700 euro per genotype, plus a minimal annual storage fee (25 euro per strain for 250 to 500 embryos). Frozen specimen are kept in two separate liquid nitrogen tanks at two different sites on campus. When necessary, the costs of revitalization from cryopreserved sperm/embryos are about 1,100/600 euro. 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.

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)?	Yes We aim at communicating our results in top journals that require full disclosure of all included data. Biological material will be shared upon simple request following publication, unless we identify valuable IP, in which case we will first protect commercial exploitation, either through patenting or via an MTA that restricts the material from commercial use
Which data will be made available after the end of 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. Biological material will be distributed to other parties if requested

Where/how will the data be made available for reuse?	Upon request by mail, In an Open Access repository, Other As a general rule, datasets will be made openly accessible via existing platforms that support FAIR data sharing (www.fairsharing.org). Sharing policies for specific research outputs are detailed below: -Open-access publications in peer-reviewed journals, including supplemental information. 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. Proper links to datasets will be provided in the corresponding publications. -Vectors: Upon publication, all vectors supporting a manuscript will be made publicly available via the non-profit plasmid repository Addgene, along with the corresponding DNA sequences. Addgene in turns performs quality control on the DNA, curates the plasmids online with all relevant information (maps, sequences), and for a minimal cost (typically ? EUR) ships the vectors upon simple request and signature of a material transfer agreement. The MTA will be prepared before depositing the vectors with the help of our organization’s Tech Transfer office. For transfer between nonprofit or academic institutions, Addgene typically uses the Uniform Biological Material Transfer Agreement (https://www.addgene.org/terms/1047/). All non-published vectors and the associated documentation will be shared by the PI upon request and after signature of a material transfer agreement, at no cost except the cost of shipment. -Cell lines: All human pluripotent cell lines supporting publications will be registered in hPSCreg, the European human embryonic stem cell registry supported by the European Commission (https://hpscreg.eu/). Information about the deposited lines (including donor information, derivation method, availability and characterization) will also be made accessible. Registration of cell lines in hPSCreg will provide visibility, confirm ethical procurement and facilitate comparison with other hPSC lines. The PI will remain the distributor of the pluripotent cell lines. All other cell lines supporting publications will be deposited in the American Type

	Culture Collection (ATCC) database (https://www.atcc.org/), which is a private, non-profit biological resource center. This will provide a secure back-up for this material. Investigators can purchase cell lines from the ATCC database upon signature of a material transfer agreement (https://www.lgcstandards-atcc.org/~/media/PDFs/MTA_2.ashx) and, in some cases, of a Limited Use/Label License (e.g. for CRISPR products or iPSC materials) and/or a Customer Acceptance of Responsibility (for potentially highly pathogenic materials). Information about the cell lines (including organism, cell type, tissue, biosafety level and disease if applicable) will also be made accessible. -Genetically modified organisms: All genetically modified organisms used in publications will be made available to researchers upon request at the time of publication. -Other digital datasets that support publications (including image, video or audio files, electrophysiology data, cytometry data, spectroscopy data and simulation data) will be made publicly available via an open research data platform such as Mendeley Data or Zenodo. -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. Manuscripts submitted for publication will be deposited in a pre-print server such as bioRxiv, arXiv, Nature Precedings or ASAPbio. At the time of publication, research results will be summarized on the PI's website (add website address) 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. Publications will also be automatically listed in our institutional repository, Lirias 2.0, based on the authors name and ORCID ID. -Algorithms, scripts and softwares: As soon as a manuscript is publicly available, the

	online private git repository containing the corresponding algorithms, scripts and software code will be changed to a public repository. -Nucleic acid and protein sequences: Upon publication, all sequences supporting a manuscript will be made publicly available via repositories such as the GenBank database or the European Nucleotide Archive (nucleotide sequences from primers / new genes / new genomes), NCBI Gene Expression Omnibus (microarray data / RNA-seq data / CHIPseq data), the Protein Database (for protein sequences). -Data that do not support publication will be either deposited in an open access repository or made available upon request by email.
When will the data be made available?	Upon publication of the research results
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, a Creative Commons Attribution (CC-BY) or an ODC Public Domain Dedication and Licence, 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 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.

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. A budget for publication costs has been requested in this project.

8. Responsibilities	
Who will be responsible for the data documentation & metadata?	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) that refer to specific datasets.
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 René Custers and Alexander Botzki for the electronic laboratory notebook (ELN) and from Raf De Coster for the KU Leuven drives.
Who will be responsible for ensuring data preservation and sharing?	The PI is responsible for data preservation and sharing, with support from the research and technical staff involved in the project, from René Custers and Alexander Botzki for the electronic laboratory notebook (ELN) and from Raf De Coster for the KU Leuven drives.
Who bears the end responsibility for updating & implementing this DMP?	The PI is ultimately responsible for all data management during and after data collection, including implementing and updating the DMP.