FWO DMP Template

Project supervisors (from application round 2018 onwards) and fellows (from application round 2020 onwards) will, upon being awarded their project or fellowship, be invited to develop their answers to the data management related questions into a DMP. The FWO expects a **completed DMP no later than 6 months after the official start date** of the project or fellowship. The DMP should not be submitted to FWO but to the research co-ordination office of the host institute; FWO may request the DMP in a random check.

At the end of the project, the **final version of the DMP** has to be added to the final report of the project; this should be submitted to FWO by the supervisor-spokesperson through FWO's e-portal. This DMP may of course have been updated since its first version. The DMP is an element in the final evaluation of the project by the relevant expert panel. Both the DMP submitted within the first 6 months after the start date and the final DMP may use this template.

1. General Information	
Name applicant	**Elia Micoli**
FWO Project Number & Title	**Application number: 1192822N** **Title: Specification and functional maturation of cortical long-range inhibitory neurons**
Affiliation	☒ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other:
2. Data description	
Will you generate/collect new data and/or make use of existing data?	☒ Generate new data ☐ Reuse existing data

Describe the origin, type and format of the data (per dataset) and its (estimated) volume *If you **reuse** existing data, specify the **source** of these data.* *Distinguish data **types** (the kind of content) from data **formats** (the technical format).*	Identifying which genes drive neuronal progenitors to generate distinct neuronal cell types is the first step to understand brain circuit assembly. My lab found a transcription factor, *POU3f2*, that is uniquely expressed in mature cortical long-range inhibitory neurons, a cell type that plays crucial roles in sleep, motor coordination, and reward. Developmentally, *POU3f2* is also expressed in newly born long-range inhibitory neurons and progenitors that generate these cells. In my project, I will study how *POU3f2* orchestrates the specification and fate maintenance of this unique cell type. My host lab has generated two conditional mutant mouse lines that lack *POU3f2* in either postmitotic long-range inhibitory neurons or their progenitors, respectively. By characterizing the developmental trajectories of long-range inhibitory neurons in these mouse models in vivo, I will determine how *POU3f2* coordinates the specification and maturation of long-range inhibitory neurons in the mammalian neocortex. Unless stated otherwise, the following datasets will be newly created by this project. **1. Experimental data** **Dataset 1.1. – Digital images** This dataset includes digital images obtained from widefield and confocal microscopy on the different cell and animal models using fluorescently-labelled antibodies and endogenous fluorescent protein; digital images obtained from gel electrophoresis; illustrations and figures derived from experimental data sets. Data formats: -Digital images in raster formats: uncompressed TIFF (.tif/.tiff), JPEG (.jpg), JPEG 2000 (.jp2), Adobe Portable Document Format (.pdf), bitmap (.bmp), camera serial interface (.csi) .gif; -Digital images in vector formats: scalable vector graphics (.svg), Scalable Vector Graphics (.svg), Adobe Illustrator (.ai); -Text files describing digital images, illustrations and figures: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Mark-up Language (.xml), Adobe Portable Document Format (.pdf), LaTex (.tex) format; -Imaging data are quantified and quantifications represented in quantitative tabular data: comma-separated value files (.csv), tab-delimited file (.tab), delimited text (.txt), Digital Audio Tape (.DAT), MS Excel (.xls/.xlsx).

	An estimated 100 GB of data will be stored every year. **Dataset 1.2. – Vectors** Bacterial vectors, mammalian expression vectors and viral vectors will be used to generate molecular tools to mark different types of neurons in cell lines and animal models, and to alter the expression levels of key developmental factors. Data formats: - Biological samples: frozen DNA plasmid (-20° C), viral particles (-80° C) - Nucleotide sequences: raw sequence data trace (.ab1), text-based format (.fasta/.fa) and accompanying QUAL file (.qual), Genbank format (.gb/.gbk); - Biological samples: frozen cell lines (-200°C, liquid nitrogen), bacterial glycerol stocks - Text files describing vectors and inserts: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Mark-up Language (.xml), Adobe Portable Document Format (.pdf); An estimated 30 MB of data will be stored every year. **Dataset 1.3. – Single-cell RNA sequencing data** Single cell RNA sequencing data will be generated from isolated cortical SST expressing cells at different developmental stages and from neuronal progenitors. Data formats: - Biological samples: isolated cells (-80°C, 384 wp). - Digital data: flow cytometry data (.fcs), next generation sequencing raw data (fastaq.gz), MS Word (.doc/.docx),), Adobe Portable Document Format (.pdf), MS Excel (.xls/.xlsx); Aside from the physical storage for samples, an estimated 20 GB of data will be stored every year.

	Dataset 1.4. – (Genetically modified) organisms and tissue samples Frozen/fixed (brain) tissue sample from research animals linked to the project, including wild type and genetically mutated animals. I will use different mouse lines to investigate the role of the transcription factors POU3f2, I will include *SSTcre;POU3f2fl/fl;RCEfl/+* and *Nkx2.1Cre;POU3f2fl/fl;RCEfl/+.* Brain tissue will be collected at different time points, including both embryonic and post-natal stages for both mouse lines The experiments planned in this project, have all been approved by the institutional Ethical Committee for Animal Experimentation and designated as project P006-2020 and P091-2020 The department has the obligatory accreditation of the authorized Belgian Ministry and is registered under license number LA1210579. The animals are housed – according to the Belgian and European laws, guidelines and policies for animal experiments, housing and care - in the Central Animal Facilities of the university. Personnel of the Central Animal Facilities and laboratory staff have to be trained in handling animals and must have the appropriate certificate in Laboratory Animal Science. These training measures are according to the Belgian law of 13 September 2004, concerning the training of people that are involved in animal experimentation. Data formats: - Biological samples: one hemisphere or the entire brain is fixed and stored at -20°C for immunohistochemical analysis. -Text files describing models (including ethical approval documents) and storage information: plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Mark-up Language (.xml), Adobe Portable Document Format (.pdf); Aside from the physical storage for samples, an estimated 1 GB of data will be stored every year.

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. *In case your host institution does not (yet) have a privacy register, a reference is not yet required of course; please add the reference once the privacy register is in place in your host institution.*	☐ Yes ☒ No If yes: - Privacy Registry Reference: - Short description of the kind of personal data that will be used: This study will use only animals models and commonly used in vitro cell lines, no personal data are present.
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 ☐ No If yes: - Reference to ethical committee approval: Animal experiments are in accordance with the Belgian and European laws, guidelines and policies for animal experimentation, housing and care (Belgian Royal Decree of 29 May 2013 and European Directive 2010/63/EU on the protection of animals used for scientific purposes of 20 October 2010). The experiments planned in this project, have all been approved by the institutional Ethical Committee of the KULeuven for Animal Experimentation and designated as project P006-2020 and P091-2020.

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 ☒ No If yes, please comment: Participants to the present project are committed to publish research results to communicate them to peers and to a wide audience. Results will be published in accordance with ethical guidelines set by the International Committee of Medical Journal Editors. Existing agreements between VIB and KU Leuven do not restrict publication of data. We do not exclude that the proposed work could result in research data with potential for tech transfer and valorization. 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 are not delayed.
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?	☐ Yes ☒ No If yes, please comment:

4. Documentation and metadata

What documentation will be provided to enable understanding and reuse of the data collected/generated in this project?	Digital data will be stored on KU Leuven servers and will be made available together with the accompanying metadata at the latest at the time of publication. The principle of preservation of data and the minimum preservation term of 10 years after the end of the project will be applied without restriction to raw data as well as processed data. No embargo will be foreseen unless imposed e.g. by pending publications, potential IP requirements or ongoing projects requiring confidential data. In those cases, datasets will be made publicly available as soon as the embargo date is reached. As detailed below, metadata will contain sufficient information to support data interpretation and reuse, and will be conform to community norms. Whenever possible, datasets and the appropriate metadata will be made publicly available through repositories that support FAIR data sharing. 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 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. In this regard, plasmids can be submitted to addgene (https://www.addgene.org/depositing/start-deposit/). 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. **Derived and compiled data** **Dataset 2.1 – Research documentation** Research documentation generated or collected from online sources (e.g. pubmed) and from collaborators, including publications, tutorials, ethical approval documents, laboratory notes, protocols, animal husbandry data. Data formats: **-**Text files**:** plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Mark-up Language (.xml), Adobe Portable Document Format (.pdf); An estimated 100 MB of data will be stored every year.

	Dataset 2.2 – Manuscripts Includes text files, illustrations and figures derived and compiled from experimental data. Data formats: - Text files: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Mark-up Language (.xml), Adobe Portable Document Format (.pdf); - Quantitative tabular data: 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), JPEG 2000 (.jp2), Adobe Portable Document Format (.pdf), camera serial interface (.csi), .gif; - Digital images in vector formats: scalable vector graphics (.svg), encapsulated postscript (.eps), Scalable Vector Graphics (.svg), Adobe Illustrator (.ai); An estimated 5 GB of data will be stored every year.

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.	☒ Yes ☐ No If yes, please specify: 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.g. Benchling) that refer to specific datasets. While specific data types might require particular metadata, as a general rule the metadata will be based on a generalized metadata schema, 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, etc. • Identifier: DOI (when applicable) • Access rights: closed access, embargoed access, restricted access, open access.

5. Data storage & backup during the FWO project	
Where will the data be stored?	As a rule, digital data will be stored on KU Leuven servers. All data generated during the project will be stored on KU Leuven servers or, for larger datasets, on The Flemish Supercomputer Centre (VSC) in the staging area, at first. Upon publication, all data supporting a manuscript will be made publicly available via open access repositories.

How will the data be backed up?	The operating system of the KULeuven 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. Stored data 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. A security service monitors the technical installations continuously, even outside working hours.
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 ☐ No If no, please specify: We give preference to the use of robust, managed storage with automatic backup. Options include central storage facilities of the research unit, the group or KU Leuven, or a cloud service offered by KU Leuven, all of which have sufficient storage & backup capacity during the project.

What are the expected costs for data storage and backup during the project? How will these costs be covered? *Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of **the allocated project budget** to be used to cover the cost incurred.*	It is the intention to minimize data sharing 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. Unless mentioned otherwise, data management costs will be covered by the laboratory budget. **All digital files** Digital files will be stored on KU Leuven servers: - the "L-drive" costs 173,78€/TB/Year. This server is an easily scalable system, built from General Parallel File System (GPFS) cluster with NetApp e-series storage systems, and a CTDB samba cluster in the front-end. Stored data 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. - The "J-drive" costs 519€/TB/Year. This server is based on a cluster of NetApp FAS8040 controllers with an Ontap 9.1P9 operating system. Stored data is backed up using snapshot technology where all incremental changes in respect of the previous version are kept online. Backups are performed hourly, daily (every day at midnight) and weekly (at midnight between Saturday and Sunday); in each case the last 6 backups are kept. Both 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. We will use free-to-use repositories to share digital files, so that there will be no additional cost required to make the data open access. **Vectors** 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. The associated costs are thus minimal (only shipment costs). All other vectors generated during the project will be shared with researchers upon request (handling by the technical staff of the laboratory, shipping costs supported by the receiver). Management of the vector collection is under the responsibility of the PI and the lab manager. Long-term preservation of this collection is of extremely high value for the laboratory, and as a general rule the vector will be preserved under the form of purified DNA (in -20°C freezer). These will be stored for the remainder of the PI's research career.

	Genetically modified organisms Maintaining a mouse colony alive costs about 2000 euro per year (for 6 cages), excluding the costs of genotyping and plug requests. 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.
Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?	Animal administrative, husbandry and animal welfare data are sensitive data and are stored in the LAIS database according to security procedure of KU Leuven.

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 data will be stored for a minimum of 10 years, i.e. at least 5 years after the end of the project. After this period, the PI will regularly evaluate whether retention of the data is still necessary and, if applicable, delete data.
Where will these data be archived (= stored for the long term)?	Any data set will be archived and stored using the available storage system offered by KU Leuven ("L"/"J" drive. Upon publication, all omics data supporting a manuscript will be made publicly available (and archived) via open access repositories.

What are the expected costs for data preservation during these 5 years? How will the costs be covered? *Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of* ***the allocated project budget*** *to be used to cover the cost incurred.*	Similarly to the data management costs during the project, data preservation after the end of the FWO project will be covered by the laboratory budget.

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 ☒ No If yes, please specify:
Which data will be made available after the end of the project?	Participants to the present project are committed to publish research results to communicate them to peers and to a wide audience. Whenever possible, data will be made available via existing platforms that support FAIR data sharing (www.fairsharing.org). This includes the DNAtraffic platform for the DNA dataset. All vectors will be made publicly available via Addgene. As detailed above, metadata will contain sufficient information to support data interpretation and reuse, and will be conform to community norms. Whenever possible, datasets and the appropriate metadata will be made publicly available through repositories that support FAIR data sharing. 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 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.

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):
When will the data be made available?	No embargo will be foreseen unless imposed e.g. by pending publications, potential IP requirements or ongoing projects requiring confidential data. In those cases, datasets will be made publicly available as soon as the embargo date is reached.
Who will be able to access the data and under what conditions?	As detailed above, metadata will contain sufficient information to support data interpretation and reuse, and will be conform to community norms. Whenever possible, datasets and the appropriate metadata will be made publicly available through repositories that support FAIR data sharing. 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 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? *Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of* ***the allocated project budget*** *to be used to cover the cost incurred.*	It is the intention to minimize data sharing 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.

8. Responsibilities

Who will be responsible for the data documentation & metadata?	The research and technical staff will generate, collect, process, analyse and store the data listed above, as detailed in the project description. All staff members are committed to conduct high quality research. In particular, standard protocols will be followed to collect data, if needed after appropriate training. Data and methods used will be regularly discussed during team and lab meetings to ensure a high level of confidence in the data generated.
Who will be responsible for data storage & back up during the project?	Regarding data security, transfer of sensitive data 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. Stored data 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. A security service monitors the technical installations continuously, even outside working hours.
Who will be responsible for ensuring data preservation and sharing?	The PI is responsible for data management. 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.
Who bears the end responsibility for updating & implementing this DMP? *Default response: The PI bears the overall responsibility for updating & implementing this DMP*	The PI bears the overall responsibility for updating & implementing this DMP.