DMP PLAN HFPEF-DIVA

1. DATA DESCRIPTION AND COLLECTION OR RE-USE OF EXISTING DATA

1A. HOW WILL NEW DATA BE COLLECTED OR PRODUCED AND/OR HOW WILL EXISTING DATA BE RE-USED?

The data generated in this project will include the following types:

- *numeric* (databases of cell expression profiles, spreadsheets of histological and flow cytometric evaluations, biomarkers, and of in vitro cell interaction results...
- *images and video* (cardiac structure and function analyses by echocardiography and histology)
- *textual* (documents reporting on the results and conclusions (.doc, .pdf).

Our project will entail data collection from *in vivo*, *ex vivo* and *in vitro* experiments, and *ex vivo* data from clinical samples. Historical data, produced in the period 2021-2024, will be re-used where necessary to keep in line with the 3Rs guiding experimental animal use notably for the HFpEF mouse models. No constraints are envisaged on such internal reuse of laboratory-generated data.

Data provenance will be documented in the file name such as: *Aquisitiondate*-Echo-4wks-mDIVA.xls; *Aquisitiondate*-IHC-CD68 -20x-mDIVA.xls*; Aquisitiondate*-ELISA-4wks-mDIVA.xls

Clinical data will be handled by each PI according to national legislation on handling and sharing of protected data (see section 4A).

1B. WHAT DATA (FOR EXAMPLE THE KIND, FORMATS, AND VOLUMES), WILL BE COLLECTED OR PRODUCED?

The *in vivo* data will include: echocardiography, exercise tolerance, blood pressure, body weight gain, blood flow cytometry, plasma dosage by ELISA or Multiplex arrays of cytokines or antibodies, and morphometry, as well as bulk RNAseq. The file format used for data compilation and statistical analysis is principally **excel** spreadsheets (.xls) and **prism** (.pzfx). For expression analyses, we will commit to standards accepted by data repositories (fq.gz.) to facilitate sharing and long-term re-use of data.

The *ex vivo* data will include cardiac flow cytometry, cardiac single cell RNAseq or bulk RNAseq, tissue qPCR, tissue homogenate ELISAs, histology (light microscopy), immunohistochemistry (fluorescence microscopy, confocal imaging, and lightsheet imaging), western blot, and biochemical assays. The file format used for data compilation and statistical analysis is principally **excel** spreadsheets (.xls) and **prism** (.pzfx) for numerical data. For imaging data we will use ImageJ/Imaris compatible formats (.ims; .tif , .jpeg, .zvi or .avi), for echocardiography we will use software specific dataoutput which is summarized in excel spreadsheets (DICOM; (.xls). and for flow cytometric evaluations we will use **excel** spreadsheets and FlowJo formats (.xls and wsp.). Transcriptomic data from scRNAseq, will be analyzed in **Seurat** (.R) and visualized in **Cerebro** (.crb; .rds), Files will also be generated for data analysis using excel spreadsheets (.xls) and **prism** (.prism; .pzfx) for processed data outputs. For expression analyses, outputs include cvs.gz and .txt; .rds and cbs. We will commit to standards accepted by data repositories to facilitate sharing and long-term re-use of data.

The *in vitro* data will include read-outs from cell culture or co-culture experiments including flow cytometry, cell proliferation assay, migration assay, and cell activation assays, immunocytochemistry (fluorescence microscopy, confocal imaging), as well as secretome analysis by ELISAs and cell analysis by luminometry assays or western blot assays or qPCR. The file format used for data compilation and statistical analysis is principally **excel** spreadsheets (.xls) and **prism** (.pzfx) for numerical data, and imaging data ImageJ/Imaris compatible formats (.tif or .avi).

Protocols will be shared as **Word** text documents (.docs) or **Adobe Acrobat** (.pdf)

We estimate that the total output or analyzed data volumes in the project will be around 7 Tb:

- scRNAseq and RNAseq data volume 0.3 GB per experiment x 3 => 1 Gb data
- Excel sheets and Prism files: 1 Mb per parameter x 8-10 parameters per experiment x 3-4 in vivo experiments=> 1 Gb data
- RNAseq data volume 0.06 Gb per experiment x 6 => 0.4 Gb data
- Imaging data (.tif or .avi) volume: 35 Gb per experiment x 4 => 140 Gb
- Imaging data (.ims) volume: 1-10 GB per image = 6 Tb data

2. DOCUMENTATION AND DATA QUALITY

2A. WHAT METADATA AND DOCUMENTATION (FOR EXAMPLE THE METHODOLOGY OF DATA COLLECTION AND WAY OF ORGANISING DATA) WILL ACCOMPANY THE DATA?

The metadata that will be provided include word file listing the conditions of the experiment, coupled to excel files listing the unique sample identifiers in each group. We will use the metadata standards developed by RDA MIBBI (Minimum Information for Biological and Biomedical Investigations) Metadata Standards https://rdamsc.bath.ac.uk/msc/m23

The data will be organized in project folders, containing subfolders for different types of experiments, such as *in vivo study 1 / 4 weeks timepoint / Immunohistochemistry analysis / Macrophage levels.* Version control will be ensured by listing v1.; v2. etc in the filename, with the final version annotated FINAL.

The raw data analysis excel sheet will refer to the protocol used for data collection (noted initially in lab notebooks, and later transcribed to *readme files*), collected in a separate folder 'protocols'. Units of measure will be standardized, such as cells/mm²; cells/mg, or conc pg/mL except for gene expression analyses that are reported as fpkm for Bulk Rnaseq or as '*normalized expression level'* for scRNAseq.

2B. WHAT DATA QUALITY CONTROL MEASURES WILL BE USED?

New collected data will be compared to historical laboratory-generated data to ensure that no technical deviation in sensitivity of techniques or changes in operating procedures will affect the readout. For most of the variables analysed, the laboratory has established SOPs that define **machine settings** (exposure times on microscopes, laser settings on flow cytometers, thresholds on echocardiographs,..) and normalization procedures for the data collection (timeline for echocardiography or organ collection, plasma sample handling, cell incubation times, ..). Data reproducibility is monitored by inclusion of sample duplicates or triplicates, depending on the technique, and repetition of the experiments for independent validation and inclusion of necessary *n* for *in vivo* experiments. Data entry validation in excel sheets is performed by an independent actor verifying individual data points for each animal and each group, notably to identify outliers that may indicate data entry or technical errors. Peer review of data is performed at the stage of data publication, with access to individual data points in all measurements.

3. STORAGE AND BACKUP DURING THE RESEARCH PROCESS

3A. HOW WILL DATA AND METADATA BE STORED AND BACKED UP DURING THE RESEARCH?

All research data generated in our laboratories will be recorded in personal laboratory protocol books and updated by each lab member on a daily basis. An electronic copy of all collected data is produced at least once per week on a joint secure server, or on OneDrive several times per day. Additional in-house backups are generated on individual external hard-drives dedicated to each operator. The raw data on institutional

servers is backed up for long-term storage (>10 years) on secure space hard-drives (managed by university).

3B. HOW WILL DATA SECURITY AND PROTECTION OF SENSITIVE DATA BE TAKEN CARE DURING THE RESEARCH

Remote access to data servers ensures access to raw data in case of onsite incident excluding access to laboratory computers. The raw data and analysed data will be freely shared within the laboratory on a daily basis, whereas exchanges with collaborators with exchanges of analysed data, or raw data as needed, will be performed either by email (Outlook) or by secure server filesharing. For sharing of sensitive data between Partners only transfer through adapted secure applications including Microsoft Teams/Sharepoint and Inserm Renater FileSender 2.0 will be used.

4. LEGAL AND ETHICAL REQUIREMENTS, CODE OF CONDUCT

4A. IF PERSONAL DATA ARE PROCESSED, HOW WILL COMPLIANCE WITH LEGISLATION ON PERSONAL DATA AND ON SECURITY BE ENSURED?

Any clinical data will be handled in agreement with national regulations and international legislation and applicable codes of conduct including: *the Charter of the Fundamental Rights of the EU; the Declaration of Helsinki (2008); the Convention of the Council of Europe on Human Rights and Biomedicine; the European Parliament and Council Directive 2016/679;* and the *Universal Declaration on the Human Genome and Human Rights* adopted by UNESCO. Participants will comply with all national legislation and regulations regarding Biomedical Research and Protection of Patient data (GDPR). Approval of the Institutional Ethics Committees will be obtained prior to starting any research and all data will be pseudo-anonymized and coded.

Spain: Spanish legislation regarding biomedical research and the protection of patient data is also complied with; Law 14/2007 on biomedical research; Law 3/2018 on the Protection of Personal Data and Guarantee of Digital Rights (LOPD); Law 41/2002 on rights and obligations in relation to clinical data; RD 1716/2011 that regulates biobanks and the use of samples for research.

Storage & Archiving: Only authorized personnel will have access to the patient's personal data and the unblinding code will be kept in a password-protected file on a secure server. Clinical data will be uploaded though a secure website, and all databases will be password-protected and encrypted. To comply with FAIR principles, data generated in the project will be made accessible for verification and re-use for subsequent research in due time. Curation of datasets pertaining to the results of this proposal will be standardized to optimize the benefits for future studies of data acquired in our project. This standardized curation must also ensure the privacy of all participants and confidentiality of individual data (only pseudo-anonymized analysed data to be shared between Partners, and, following publication, with the scientific community).

Only pseudo-anonymized data will be shared with the partners of the consortium.

4B. HOW WILL OTHER LEGAL ISSUES, SUCH AS INTELLECTUAL PROPERTY RIGHTS AND OWNERSHIP, BE MANAGED? WHAT LEGISLATION IS APPLICABLE?

Concerning the *Intellectual Property Rights* (IPR) regulations related to handling and exploitation of IP emanating from our project, we will share newly created foreground IP by directly involved participants as defined and protected by the legal departments (TTOs) of each participant. The Partners will mutually agree to grant each other access to foreground or background IP, as long as this is necessary for the partner to carry out the project. In this sense, the data will be curated and will be available for re-use in publicly available data sets (if possible from IP protection perspective). The Partners respective TTOs will be responsible for the protection of any exploitable result prior to publication. If patentable results are achieved, worldwide patent search will be made to identify any prior disclosures, to assess novelty and patentability. Confidentiality agreements will be signed before discussing sensitive results with interested third parties

and potential end-users (e.g. pharmaceutical companies, bio-technological SMEs). Molecular tools generated in the project will be freely distributed upon reasonable request to academic investigators for *non-commercial use*. Should IP arise that requires patent protection, we pledge to ensure that the technology or material remains widely available to the research community

4C. WHAT ETHICAL ISSUES AND CODES OF CONDUCT ARE THERE, AND HOW WILL THEY BE TAKEN INTO ACCOUNT?

National codes of conducts and institutional ethical guidelines will be followed in the project to ensure RRI.

5. DATA SHARING AND LONG-TERM PRESERVATION

5A. HOW AND WHEN WILL DATA BE SHARED? ARE THERE POSSIBLE RESTRICTIONS TO DATA SHARING OR EMBARGO REASONS?

Data sharing between Partners will be performed using adapted secure applications including *Inserm Renater FileSender*. Processed and analyzed data will be made freely available, following IP protection and peer-reviewed publication of our findings, on dedicated open access servers. These data collections will be rendered freely *Accessible*, *Reusable*, and *Searchable, if it is not IP or GDPR protected*. In addition, both analysed data and interpretations will be made freely available by deposition of manuscripts on open access forums, such as *bioRxiv,* when appropriate. In parallel, experimental transcriptomic data, and metadata, will be deposited on dedicated specific portals, such as *Differential Expression Atlas* maintained by the EMBL-EBI, following the principles of open science practice and EU RRI guidelines to maximize their potential to be *Reused* and *Interoperable* (RRI), when not covered by IP protection or GDPR.

No personal data of patients will be shared at any point. Any clinical data will be shared only if allowed by GDPR.

Results of the proposed research program will be submitted for publication in peer reviewed journals, and disseminated through active participation in national and international meetings. In addition, results will be conveyed to the lay public through the press pages of both Institutions. To target more specifically communities of HF patients, members of our consortium will participate in outreach activities such as Pint-of-Science and local public hospital information days.

5B. HOW WILL DATA FOR PRESERVATION BE SELECTED, AND WHERE DATA WILL BE PRESERVED LONG-TERM (FOR EXAMPLE A DATA REPOSITORY OR ARCHIVE)?

The French partner will use the Entrepot database (https://recherche.data.gouv.fr/) to generate closed Dataverse collections for each study prior to publication, which will be freely accessible, according to FAIR principles, after publication and IP protection.

The Spanish partner (FIMA) will use the REPISALUD repository.

The Belgium partner will use different international repositories. We will upload flow cytometery data to FlowRepository using the miFLowCyt standard. We will upload the sequencing data (that is not IP protected or GDPR protected) to the GEO repository using the MIAME standard.

5C. WHAT METHODS OR SOFTWARE TOOLS ARE NEEDED TO ACCESS AND USE DATA?

Prior to publication and/or IP protection, raw data, processed (analyzed) data, and metadata will be deposited as **closed Dataverse collections**, e.g. using Inserm data archiving (Entrepot de Données l'Inserm: https://entrepot.recherche.data.gouv.fr/dataverse/inserm;jsessionid=29d1d6f563da69d1c2d57b696a88

5D. HOW WILL THE APPLICATION OF A UNIQUE AND PERSISTENT IDENTIFIER (SUCH AS A DIGITAL OBJECT IDENTIFIER (DOI)) TO EACH DATA SET BE ENSURED?

Persistent identifiers will be applied at data deposition as **closed Dataverse collections**, e.g. using Inserm data archiving (Entrepot de Données l'Inserm: https://entrepot.recherche.data.gouv.fr/dataverse/inserm;jsessionid=29d1d6f563da69d1c2d57b696a88

6. DATA MANAGEMENT RESPONSIBILITIES AND RESOURCES

6A. WHO (FOR EXAMPLE ROLE, POSITION, AND INSTITUTION) WILL BE RESPONSIBLE FOR DATA MANAGEMENT (I.E. THE DATA STEWARD)?

Data management activities will be shared between the local PI and local project members in regard to data capture, metadata production, data quality, storage and backup, data archiving, and data sharing. E Brakenhielm will manage the long-term backup on shared data from the project on institutional servers, and handle the dataverse collections, as well as requests for raw data sharing from individuals interested in the project. She is responsible for updating the DMP on an annual basis.

6B. WHAT RESOURCES (FOR EXAMPLE FINANCIAL AND TIME) WILL BE DEDICATED TO DATA MANAGEMENT AND ENSURING THAT DATA WILL BE FAIR (FINDABLE, ACCESSIBLE, INTEROPERABLE, RE-USABLE)?

The data storage is covered by the overhead costs for our institutes (Rouen University and Inserm) and no extra costs have been identified. The time allocated to data handling is estimated at 2-15h per month, to be shared between the local PI and local project members.