DATA MANAGEMENT AND SHARING PLAN

[redacted] (PI) will be committed to share scientific data generated during the proposed study in order to support reproducibility. Accordingly, we have designed the following plan:

Element 1: Data Type

A. Types and amount of scientific data expected to be generated in the project:

1. Data sets – Data from animal behavioral tests will be collected by video recording for each individual mouse. Data will be analyzed for each individual mouse and then tabulated in an excel sheet along with other mouse details (e.g., strain, treatment, age and sex) (~1-10 megabytes for each video and excel file).
2. Data sets – Raw data files from flow cytometry, immunocytochemistry, and immunohistochemistry studies will be collected in the file format of fcs, tiff/jpeg and excel files, respectively. Data will be analyzed and recorded on an excel sheet (<30 megabytes for each file).
3. Data sets – Raw data files from sequencing studies consisting of microRNA sequencing will be provided as fastq files (approximately 1 terabyte for all combined data sets across cohorts).
4. Data sets – Raw data files from cerebral blood flow will be provided as .pss and .nss files, and data from magnetic resonance imaging will be provided as dicom format (approximately 1 terabyte for all combined data sets across cohorts).
5. Data sets – Raw data files from Western blotting, RT-qPCR, and ELISA will be provided as Tiff/jpeg and/or excel files. (<30 megabits for each file).
6. Information for all the methods, experimental designs and protocols (e.g., reagents, antibodies, primers and kits) will be documented in Word and/or Excel files (<1 megabyte for each file).

B. Scientific data that will be preserved and shared, and the rationale for doing so:

In this project, the cleaned, item-level spreadsheet data for all variables will be shared openly, along with example quantifications and transformations from initial raw data. Final files used to generate specific analyses to answer the Specific Aims and related results will also be shared. Data will be made available to the scientific community through journal publications in the appropriate format, in the main text, methods, protocols, and supplementary information. Also, all the data will be available upon individual request. The rationale for sharing the data described in section A will allow researchers to reproduce our publications and will allow them to collect additional data in a similar way to extend our results.

C. Metadata, other relevant data, and associated documentation:

To facilitate the understand and reuse of the data, a README text file, code book, or data dictionary will be generated and deposited into a repository along with all shared datasets. The README text file, code book, or data dictionary will include method description, instrument settings, RRIDs (Research Resource Identifiers) of resources such as antibodies, model organisms, compounds, viral vectors, primers, and other tools (e.g., software, databases, and services). The data dictionary will define and describe all variables in the dataset.

Element 2: Related Tools, Software and/or Code:

Initial data acquisition and analysis will be performed with Noldus EthoVision (Behavioral tests), Nikon NIS-Elements Imaging software (Immunocytochemistry and immunohistochemistry), Bruker Prairie View Imaging Software (Multi-photon imaging), Itksnap Software (MR imaging) and FlowJo Software (Flow cytometry). RNA sequencing data will be analyzed by R studio package in this project. Fiji ImageJ is required to access the raw image data, which is an open-source software that can be downloaded freely online. Links to this or other open-source viewers will be included with the documentation for the shared dataset. Numeric data will be analyzed by statistical programs such as Prism which present the raw data in the .pzfx file. Code generated in R will be shared in selected repository with associated data sets.

Element 3: Standards:

Consensus standards are not currently available to the data generated by the proposed research, however easy and user-friendly access will be provided as we share our scientific data with the broader research community to increase robust data reproducibility and foster future collaborations.

Element 4: Data Preservation, Access, and Associated Timelines

A. Repository where scientific data and metadata will be archived:
We will use NIH-supported Scientific Data Repository for our data preservation and sharing. We will utilize **NIH/NIA AD Knowledge Portal** for data generated by this project.

B. How scientific data will be findable and identifiable:
The AD Knowledge Portal is a NIH-designated repository and the distribution site for multi-omic data as well as general experimental data from human samples and model systems. The Portal hosts raw data, analysis results, analytical methodology, and research tools generated through Alzheimer's disease and related dementia programs supported by the National Institute on Aging.

C. When and how long the scientific data will be made available:

Data will be deposited into the selected repository for sharing no later than the time of an associated publication, or the end of performance period, whichever comes first. The data will also be shared if requested by other investigators following acceptance for publication. The broad objective of the data generated from this project is to stimulate new advances as quickly as possible and to allow prompt evaluation of the results by the scientific community. The duration of preservation and sharing of the data will be a minimum of 5 years after the funding period.

Element 5: Access, Distribution, or Reuse Considerations

A. Factors affecting subsequent access, distribution, or reuse of scientific data:
Since human subjects data is not part of this project, we do not foresee any limitations to data access, distribution, or reuse.

B. Whether access to scientific data will be controlled:
All scientific data generated in this study will be available publicly to the broader research community.

C. Protections for privacy, rights, and confidentiality of human research participants:
Not applicable as the current study is not considered human subjects research (de-identified biorepository samples or autopsy samples).

Element 6: Oversight of Data Management and Sharing:

All data management and sharing practices will be monitored, reviewed, and discussed by the PI who will assume the primary responsibility for overseeing these practices in the lab. The PI will ensure the formal oversight for DMS plan at minimum annually prior to the progress report (RPPR) each year.

The Office of the Executive Vice President & Chief Academic Officer (EVP/CAO) and The Office of Data Science (ODS) at UTHealth Houston will provide joint institutional oversight for the DMS plan. Datasets resulting from this research will be cataloged with in the institutional DEPUT. DEPUT is the institutional oversight management portal supported by UTHealth Houston for DMS validation and tracking. Project Contact PI will update data status in DEPUT, and the institutional office of Sponsored Projects Administration (SPA) will perform annual validation according to the DMS plan. Validation results will be maintained in DEPUT. Noncompliance with the DMS plan will be identified with appropriate correcting measures implemented. The PI will have overall responsibility for compliance with data collection, storage, and safety protocols.

At the project completion, the final progress report will summarize the data sharing accomplishments and provide links to the shared datasets.