Plan Overview

A Data Management Plan created using DMPonline

Title: Translational Cardiology

Creator: Payam Haftbaradaran Esfahani

Principal Investigator: Magnus Bäck

Data Manager: Payam Haftbaradaran Esfahani, Zihan Qin

Affiliation: Karolinska Institutet

Template: KIs Template for Swedish Research Council DMP

Project abstract:

Aortic valve stenosis (AVS) is a leading cardiovascular disease that causes a progressive narrowing of the aortic valve as a consequence of thickening and calcification. Emerging evidence shows that lipid deposition and the unbalance between pro-inflammatory and inflammation-resolving determine the disease progression. Valvular interstitial cells (VICs), the primary cellular component of the heart valve, have been revealed undergoing phenotype transition during CAVD progression. Moreover, as a lipids-driven chronic inflammatory disease, AVS is also facilitated by the involvement of immune cells. Recently, the pathobiology of AVS has become better understood due to the development of advancing single cell sequencing techniques. However, the effect of lipid oxidation and inflammation-resolving on AVS have not yet been deeply investigated. The aim of this project is to establish a comprehensive database that includes published single-cell data, our own single-cell data, and the single-cell sequencing data from other human samples in the context of AVS. This database will not only map the human aortic valve cell atlas but also offer insights into the complex cytopathological differentiation processes involved.

ID: 158986

Start date: 15-09-2024

End date: 15-09-2025

Last modified: 13-09-2024

Translational Cardiology

Description of data

How will data be collected, created or reused?

- Pulished or existing bioinformatics data will be used for new analyses.
- Bulk RNA sequencing and single-cell RNA sequencing data will be generated from aortic valve samples obtained from patients who have undergone aortic valve transplantation.

What types of data will be created and/or collected, in terms of data format? Include version numbers if applicable.

- Single cell sequencing data will be in .fastq format.

What volumes of data will be created and/or collected?

- < 10 TB

Documentation and data quality

How will the material be documented and described, with associated metadata relating to structure, standards and format for descriptions of the content, collection method, file naming-format-versioning, etc

- Documentation will include a standardized folder structure, codebooks (metadata about the data), logbooks (metadata about data processing), analysis plans, input and output files from databases and statistical software
- All files will be named according to the date of acquisition and experimental condition and put into folders. A "read me" file will be generated, explaining the experimental conditions, tissue and cell types.
- Working files will be clearly labelled with a version suffix, e.g. v2.

How will data quality be safeguarded and documented (for example repeated measurements, validation of data input, etc.)?

- Data will be quality-checked at collection/generation by validation against controls or publicly available databases.
- Sequencing data will be quality controlled in terms of sequence quality, sequencing depth, reads duplication rates (clonal reads), alignment quality, nucleotide composition bias, PCR bias, GC bias, rRNA and mitochondria contamination, coverage uniformity. Only high-quality data will be included in the subsequent analysis.
- The register holder assures data quality in terms of completeness and correctness of registration.
-

Storage and backup

How is storage and backup of data and metadata safeguarded during the research process?

- ELN
- During the analysis of the RNA-sequencing data, fastq and analysis files will be stored at the secure cluster Bianca at Uppmax. All files will be transferred to a server at KI when the analysis is over.

How is data security and controlled access to data safeguarded, in relation to the handling of sensitive data and personal data, for example?

- Human sequencing data from NGI will be processed and temporarily stored in the Bianca server for sensitive data at Uppmax (Uppsala Multidisciplinary Center for Advanced Computational Science), which has several layers of security.

Legal and ethical aspects

How is data handling according to legal requirements safeguarded, e.g. in terms of handling of personal data, confidentiality and intellectual property rights?

- Sensitive personal data will be handled according to GDPR. (https://staff.ki.se/gdpr).
- Data will be pseudonymized and a key will be kept separately from the data.

How is correct data handling according to ethical aspects safeguarded?

- Patient data is pseudonymized by the clinical collaborator and the code is not accessible to researchers in our research group. The material will arrive to KI coded, and the original code will be saved by the collaborators.
- Consent has been acquired from human participants to process/share data.
- Data Transfer/Processing agreements will be signed prior to any data sharing.
- The study will be performed in accordance with the ethical principles of the World Medical Association (WMA) Declaration of Helsinki and aims to follow Good Clinical Practice (GCP) guidelines.

Accessibility and long-term storage

How, when and where will research data or information about data (metadata) be made accessible? Are there any conditions, embargoes, licenses and limitations on the access to and reuse of data?

- Only metadata is published openly, underlying data is made available upon request after ensuring compliance with relevant legislation and KI guidelines.

In what way is long-term storage safeguarded, and by whom? How will the selection of data for long-term storage be made?

- Long-term storage will take place at the server at the Institution. Data will be stored as long as possible. The data will include raw data and the final data analysis file.

Will specific systems, software, code or other types of services be necessary in order to open and use/analyse data in the long term?

- The data can be read by R or pathyon software.

How will unique and persistent identifiers for the research data, such as a Digital Object Identifier (DOI), be obtained?

A DOI will be assigned to the dataset by the data repository (e.g. SND).

Responsibility and resources

Who is responsible for data management while the research project is in progress?

- Data management is performed by a dedicated data manager in the research group, who is an experienced researcher with a PhD.

Who is responsible for data management, long-term storage after the research project has ended?

- The PI is responsible for data management and the archive function will be responsible for long-term storage

What resources (costs, labour or other) will be required for data management (including storage, back-up, provision of access and processing for long-term storage)?

- No specific resources are allocated for data management

What resources will be needed to ensure that data fulfil the FAIR principles?

- No particular additional resources will be required.