DESCRIPTION OF DATA

How will data be collected, created or reused?

Image files will be recorded from a confocal microscope.

Patient data will be acquired from the Swedish Hip Arthroplasty Register.

Survey responses will be acquired using the RedCap survey software.

Respondent data will be acquired in clinical interviews.

RNA sequencing data will be generated from normal and tumor tissues from patients.

Existing will be used for new analysis.

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

Biomarker Data will be received/imported/saved in a .csv format.

Questionnaire data will be saved in SAS format.

Interview responses will be saved in Nvivo .nvp format.

Survey responses will be exported from RedCap to .csv format.

Register data will be received in spreadsheet format and will be converted to .tsv format before analysis.

Sequencing data will be in fastq format.

Flow cytometry data will be saved in .fcs format.

Confocal images will be saved in .jpeg format.

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 softwares.

All files will be renamed, with 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.

Survey responses will be curated into the Psych-DS format.

Patient data will be read into SPSS and working files will be named 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.

RNA seq 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.

The transcribed interview material will be coded independently by two researchers.

Images will be inspected for artifacts and the results will be recorded in a spreadsheet file.

Register data will be quality controlled according to a procedure established in our group (REF).

Data input is validated at the point of entry (RedCap) which does not permit missing data to be entered. Data will be checked to verify that all responses are within the possible range of data values.

STORAGE AND BACKUP

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

We will use a local server at the institution for data storage. Working datasets, and metadata will be stored on the dedicated server. The server name is XXX and the folder where the data is saved is XXX. Link to the server:(if possible)

KI ELN be used for the documentation of all analyses and results.

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?

Access to the documentation stored in ELN is restricted to group members.

Access to the data saved on the server is restricted to group members/authorised personnel.

We only work with pseudonymized data, with the key stored in a safety cabinet located at XXX (please specify location) and to which only XXX have access to (please specify the people that have access to it).

It has been judged that controlled access is not required for these data since the data do not contain personal information.

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 KI:s guidelines (https://staff.ki.se/gdpr).

There are no personal data, nor any other grounds for confidentiality.

If necessary, data transfer or data processing agreement will be performed between our research group and collaborators for data transfer, previously approved by KI's legal department.

Data will be pseudonymized and a key will be kept separately from the data.

IP rights will be managed in accordance with the contract drawn up with our industrial partner organization (specify).

How is correct data handling according to ethical aspects safeguarded?

Survey and clinical data will be anonymized, i.e. all possibility to trace the data back to the study participant has been removed. The data is anonymized when the code key is destroyed and it is no longer possible to connect a person to the data.

Patient data is pseudonymized by the clinical collaborator or the register holder, 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/register holder.

Consent has been acquired form human participants to process/share data.

Data Transfer/Processing agreements will be signed prior to any data sharing.

Ethical approvals/amendments and informed consent forms for the project are registered in the diary.

Results will only be presented on aggregated level without any possibility of backward identification.

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.

Data will be made available upon publication as a supplement to the publication.

Data will be deposited at a repository/database (please provide name) immediately and without embargo, using a license (please specify license type, e.g CC-BY).

Metadata will be deposited at SND and be freely searchable after publication, with links to the underlying data.

Information about data and metadata are available for the register X holder.

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 at least 10 years after publication. Data will be stored in X format. The data will include raw data and the final data analysis file. Intermediate working files will be deleted.

As soon as an e-archive is available centrally at KI the data will be transferred to the e-archive.

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

A software license for SAS, STATA, Nvivo, etc will be required.

The spreadsheet data can be read with any software compatible with .csv files.

Image files can be opened with any software compatible with .jpeg files.

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 the PI / a research assistant / a postdoctoral / a dedicated data manager.

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

The PI is responsible for ensuring that the data is stored safely during and after the completion of the project. The PI is also responsible for contacting the archive at the institution or the central KI archive.

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.

Salary for a data manager in the group is funded X% by this grant.

Access to the departmental server is required. It is expected to cost X SEK and is covered by the project budget.

What resources will be needed to ensure that data fulfil the FAIR (= Findable, Accessible, Interoperable & Reproducible) principles?

We will require assistance from the KI Research Data Office to upload the data to the SND catalogue.

No particular additional resources will be required.