Plan Overview

A Data Management Plan created using DMPonline

Title: NC-CHOICE: A cell-free therapy for intervertebral disc regeneration based on the pivotal cargo of Notochordal cells

Creator:Marianna Tryfonidou

Principal Investigator: Marianna Tryfonidou

Data Manager: Frank Riemers

Affiliation: Utrecht University

Funder: Netherlands Organisation for Scientific Research (NWO)

Template: Data Management Plan NWO (September 2020)

ORCID iD: 0000-0002-2333-7162

Project abstract:

The most common cause of pain restricting daily activity is chronic low back pain (LBP). LBP has a tremendous impact on quality of life and productivity and is since 1990 the leading cause for years lived in disability. In >40% of LBP cases, the pain originates from the degenerate intervertebral disc (IVD) (>40 million patients globally). When conservative treatment fails, surgery is a high-impact last resort. Emerging stem cell therapies may mitigate LBP but fail to repair the affected IVD.

I propose a new cell-free treatment for IVD-related LBP based on the game-changing properties of notochordal cells (NCs). The regenerative potential of NCs relates to their pivotal role in biology: they reside within the notochord, which is the defining embryonic structure in vertebrates. The notochord steers organ patterning and disc development. My work pioneered the use of NCs for IVD regeneration and showed that NCs convey their unique regenerative potential via the secretion of cell-derived extracellular vesicles (EVs) and specialized matrix biomolecules.

NC-CHOICE will identify the regenerative potential encrypted in NC-EVs. The set of pivotal NC-EVs biomolecules will be encapsulated in RNA form in synthetic lipid nanoparticles to bypass the drawbacks of natural EVs in clinical application. A new hydrogel based on the NC-matrix will be used to adsorb tissue-specific matrix biomolecules onto the nanoparticle surface and maximize cellular uptake. Upon intra-discal injection, the NC-matrix hydrogel will replenish the treated IVD with essential tissue building blocks lost during degeneration.

We will demonstrate proof of concept in (a) human disc tissue under conditions that mimic loading in daily life and (b) an experimental dog model that recapitulates IVD disease similar to humans. This project will produce the fundamental knowledge needed to begin developing an off-the-shelf injectable nanomedicine for disc regeneration and enduring pain reduction in LBP patients.

ID: 98516

Start date: 01-09-2022

End date: 31-08-2027

Last modified: 10-05-2022

Grant number / URL: 19251

NC-CHOICE: A cell-free therapy for intervertebral disc regeneration based on the pivotal cargo of Notochordal cells

General Information

Name applicant and project number

Name applicant: M.A. Tryfonidou
Project number: 19251

Name of data management support staff consulted during the preparation of this plan and date of consultation.

Consulted during project application with Frank Riemers, a technician that has aquired his masters in bioinformatics and is supporting my team with his expertise and with data management. Once the DMP was prepared we received feedback from Jacques Flores
Research Data Management Consultant (on 02-05-2022).

1. What data will be collected or produced, and what existing data will be re-used?

1.1 Will you re-use existing data for this research?

If yes: explain which existing data you will re-use and under which terms of use.

- Yes

I expect to re-use data produced by my groups and also by other researchers within the specialized field I am working. This data to be re-used refers to "omics" data, e.g. RNAseq profiles, single cell sequencing data, proteomics data already reported in literature and available in publicly available data basis.

1.2 If new data will be produced: describe the data you expect your research will generate and the format and volumes to be collected or produced.

NC CHOICE will generate research data in a broad range of R&D activities to achieve its objectives within the project. This involves data from in vitro, ex vivo and in vivo studies. No personal data will be collected. The following data types will be generated:

- Text based documents, .docx and .txt file formats
- for continuous and binary data: .xls, .xlsx, .csv
- shared powerpoint presentations, .pptx
- illustrations and graphic design: Graphpad Prism (Format: .pzf, .pzfx), Photoshop (Format: different types possible, mostly .png), and will be made available as .jpg, .psd, .tiff, .png and/or .ai files depending on the purpose of use. PDFs, PIDs and layouts will preferentially use inkscape.org, an open source software for vector graphics. (Format: .svg), and will be made available as .png, .jpg and .pdf files. Furthermore, where applicable we will generate Biorender-based images which will also be stored in the final form as pdf file.
- magnetic resonance imaging: DICOM files - 0.5-2MB (~20-150kB per MR slide and size of the object; ~500 kB per CT slide)
- mass spectrometry (LC-MS/MS): .RAW, .csv - 1GB; 20 samples expected
- confocal imaging: .jpg (100kb/image) for pilot experiments and screening; .tiff (2000kb) - to only be used for final images to be used in publications
- RT-qPCR data: .eds, .xls, .xlsx, .csv, .pcrd; size: depends on design of experiment, max 500KB/plate, thus with 30 platen = 15MB (1500KB, 0.015GB)

- ELISA: .exp, .csv
- Flow cytometry: .fcs
- Digiwest: .csv, DigiWest Viewer Format; 5MB/sample; 100 samples - this analysis is conducted in collaboration with the NMI-RI (Germany)
- Transcriptomics data (NGS):

- Read data: general (CRAM, BAM, BED, Fastq)
- Assembled and annotated sequence data: flat file format (FASTA, XML),
- Multiple Sequence Alignment (MSA) formats.
- Quantitative tabular data with minimal metadata:.csv; .tab; .xls; .xlsx, .txt; .mdb; .accdb; .dbf; .ods
- R data objects: .RDA; .RDS
- Quantitative tabular data with extensive metadata: .por; SPSS, .sav; .dta
- Qualitative data: .xml; .rtf; .txt; .html; .doc; .docx;

Size: 10GB/sample; max 20 samples anticipated

1.3. How much data storage will your project require in total?

- 100 – 1000 GB

Estimated based on the numbers described under section 1.2

Costs yoda: €4,- /TB/month > €48,- /TB/Jaar

we do not expect more than 1TB (1000GB). The "omics" data will be stored at SRA en GEO and will be archived too at the UU location.

Transferring data can be either through yoda or savely sending (large) datasets to partners using SURFfilesender (https://www.surf.nl/surffilesender-veilig-en-versleuteld-bestanden-versturen).

2. What metadata and documentation will accompany the data?

2.1 Indicate what documentation will accompany the data.

We have defined the following practical considerations for naming files:

File and folder names should be descriptive of its contents. Files start with the date of its creation in the form YYYYMMDD (W3C/ISO 8601 date standard), contain underscores instead of spaces and no "special" characters (&#!?&%@). The maximum length of the description excluding the extensions, dates, initials etc should be about 35 characters. Folders contain a data collection, larger collections (experiments, manuscripts in submission) may be divided into sub folders for each part of the collection.

When it is appropriate to include a version number these should be included and have the form of vx. This refers to major revisions. In the event that versioning for “minor” revisions is needed (e.g. manuscript), then the vx. is supplemented with the initials of the authors that provides feedback. (vx.initials). The version information is placed at the end of the filename before the dot of the extension. Versions do NOT contain ambiguous words like: "final", "latest", "last", etc.

Folder: [Project]_[task]_[description];

Example: NC_CHOICE_T3.4_EV

File: YYYYMMDD_AA_[description]_vx.ext ;

Example: 20230714_FB_alginateredifferentiation_setup_v2.0.xlsx

Example: 202320714_FB_manuscriptrunningtitle_V1_MT.docx (in this example MT provided feedback to version V1)

As the project evolves we will generate also overview tables of experiments with the respective read out that will be deemed essential depending on the question. For example for set of samples used in the different experiments will have unique identifying numbers and the team will be able to track down for which assays they have been used etc. Such overviews will be generated once we get to the practical implementation of the project and excel sheet in tabular form will be maintained.

2.2 Indicate which metadata will be provided to help others identify and discover the data.

Data stored in YODA,
As a minimum, the metadata in Yoda will comply with the Datacite V4 standard. In addition, the system facilitates validation by per-research-configurable metadata schemas based on JSONschema to support discipline-specific metadata. The metadata is stored persistently along with the data and in the Yoda repository database. It is findable in Datacite upon publication of data. The metadata in the Yoda repository can also be retrieved via a service that complies with the Open Archives Initiatives Protocol for Metadata Harvesting (OAI-PMH).
Alternatively, for omics data stored in a repository like GEO or SRA the metadata provided will meet the requirements for submitting to this repository, and thus will include the metadata on both experiment and sample levels.

3. How will data and metadata be stored and backed up during the research?

3.1 Describe where the data and metadata will be stored and backed up during the project.

- Institution networked research storage

Data storage will be done on the university network drives (U-/O-drive) and local storage when employing a laptop. U-drive and OneDrive is employed by the personnel for the tasks/activities of the specific person. O-drive and OneDrive is employed for documents that are commonly shared between the member of the research team.
To minimize data size the UU iPSpine group will be working with OneNote for minutes of the groups meetings and share files via Sharepoint.
Data from local storage will be on regular basis updated; based on discussions with the Data management team of the UU it will be decided which storage modus is the best to use (U-/O-drive or on One Drive (cloud); Yoda).
Research data (including finalized protocols and raw research data from in vitro & in vivo data) will be stored in the electronic lab-journal eLabJournal (e-Lab).
DICOM; Mass spectrometry data; confocal imaging ; RT-qPCR; ELISA; Flow cytometry; Digiwest; Transcriptomics data (NGS) will be stored at YODA. This will be done per subproject/peer reviewed manuscript published. Data that is stored at YODA will receive a DOI and become open access upon publication of the respective manuscript. Specifically, transcriptomic data (NGS data) will be submitted to ArrayExpress or GEO and the raw data (FASTQ) to either the ENA or SRA. Proteomics data will be submitted to PRIDE.

3.2 How will data security and protection of sensitive data be taken care of during the research?

- Not applicable (no sensitive data)

4. How will you handle issues regarding the processing of personal information and intellectual property rights and ownership?

4.1 Will you process and/or store personal data during your project?

If yes, how will compliance with legislation and (institutional) regulation on personal data be ensured?

- No

Personal data is not collected. Human samples that are received are anonymized by the provided (UMCU) before delivered to us. The only information provided is: which spinal segment, sex (male/female) and age in years.

4.2 How will ownership of the data and intellectual property rights to the data be managed?

NC_CHOICE UU appointed personnel and involved collaborators must retain any data, documents or other material as confidential during the implementation of the project. The End User Committee agreement describes this obligation to protect results. Awareness of confidentiality will be guaranteed by putting this as a regular agenda point during each End User Committee meetings. If new collaborators are onboarded they will be asked to sign a confidentiality agreement prior to collaboration. The Patient advisory board has been aggregated and all members that have onboarded have signed the confidentiality and impartiality agreement.

5. How and when will data be shared and preserved for the long term?

5.1 How will data be selected for long-term preservation?

- Other (please specify)

Research data from pilot experiments leading for example to a final design of an experiment will be discarded the latest by the end of the project.
If data need to be put together to create a publication and this involves various institutions there will inevitably be one corresponding author or (last author) who is the lead in that particular publication. For this to happen data needs to be transferred and congregated in one institution (or as many of the institutions that work together on that publication) to properly interpret and analyze the data and to write the article. The corresponding author's institution will be responsible for this data and archive all the data that was required to create the publication accordingly within their own institution so that if someone contacts them to gain access he should be able to grant access easily.

5.2 Are there any (legal, IP, privacy related, security related) reasons to restrict access to the data once made publicly available, to limit which data will be made publicly available, or to not make part of the data publicly available?

If yes, please explain.

- Yes

There may be IP-related reasons to restrict access to a part of the data that are for example used for patent application; in that case they will only become public available once the patent is secured and if licensing to third parties is not hampered hereby.

5.3 What data will be made available for re-use?

- Other (please specify)

All data will be made available, except for data that relate to aforementioned IP-related reasons.

5.4 When will the data be available for re-use, and for how long will the data be available?

- Data available upon completion of the project

Part of the data will be made available as soon as the article is published. "Omics" data derived from RNAseq and proteomics data may not become immediately available, depending on additional analysis we wish to do with other data generated over the course of the project that will help us keep the leading position in the field.

5.5 In which repository will the data be archived and made available for re-use, and under which license?

Our research uses the research data management system Yoda. Yoda facilitates collaboration on data during research, archiving of data during and after research, and publication of research data. Data along with its metadata is shared within a closed user group, accessible to authorized users anywhere via an internet connection. Client user workstations can be any platform that supports WebDAV protocol (e.g. Mac, Windows, Linux). Research data integrity is enhanced through the use of the Yoda Vault, data can be deposited in the vault where it becomes read-only. Vaulted data can be published (findable and citable via DOI Datacite identifier) and as such can be made available to the research community at large. All Yoda data is stored in at least two geographically spread locations. The data is stored and transmitted in an encrypted format, Yoda complies with Utrecht University's Information Security policy for data classified as public, internal use, sensitive and critical.

5.6 Describe your strategy for publishing the analysis software that will be generated in this project.

The analysis is done using freely available software like R and python employing methods from generally available packages\libraries the data can be (re-)use the data without the specific scripts developed during this project. Nevertheless, the scripts will be made available through GitHub.

6. Data management costs

6.1 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)?

Frank Riemers will provide in kind time to the project for bioinformatic analysis and data management. This has been included in the project description of the project.

Costs that relate to storage of data in YODA will be covered by the project itself and later one, upon completion of the project these costs will be covered by the overhead of follow up projects.