Plan Overview

A Data Management Plan created using DMPonline

Title: ReCognitION: Recognition and Validation of Druggable Targets from the Response to Cognitive Behaviour Therapy in Myotonic Dystrophy type 1 patients from Integrated -Omics Networks

Creator: Peter A.C. 't Hoen

Principal Investigator: Peter A.C. 't Hoen

Affiliation: Radboud University Medical Center (Radboudumc)

Funder: ZonMw (Nederlands)

Template: Data management ZonMw-template 2019

ORCID iD: my-orcid?orcid=0000-0003-4450-3112

Project abstract:

Myotonic dystrophy type 1 (DM1), the most common adult form of muscular dystrophy, affects virtually all tissues; the noncurable condition carries significant morbidity and mortality impacting patient and family quality of life and socio-economic status. The OPTIMISTIC clinical trial has shown that Cognitive Behaviour Therapy (CBT), a patient-tailored intervention to increase activity and enable patients to deal with their disease, imparts strong benefit on patients' activity and participation (Lancet Neurology, 2018). We now propose a multi-omic approach to identify the molecular signatures of the response to this clinical intervention, taking advantage of the thorough clinical characterization of the enrolled patients and the comprehensive set of serum samples at baseline and two follow-up time points. Our lead hypothesis is that pathways associated with the positive response to CBT can be consolidated or reinforced by conventional drug therapies targeting the same pathways. A network-based bioinformatics approach shall be used to identify drug targets in the molecular signatures. We shall repurpose clinically approved drugs for these targets and measure their impact on molecular profiles of patients' induced pluripotent stem cells, differentiated to multiple DM1-relevant cell types (cortical neurons, motorneurons and myofibers). The effect of the most promising drug candidates will be evaluated in the DMSXL and HSA-LR mouse models, employing cognitive, behavioural and motor readouts that are reminiscent of the clinical readouts in the OPTIMISTIC trial. The systemic and muscle-restricted expression of the transgene in two different mouse models allows for exploration of the brain/muscle axis in the cognitive and behavioural aspects of the disease. Repurposed drugs can be evaluated in isolation or combination with other interventions like CBT in future clinical trials for DM1 and other neurological conditions. The drug repurposing strategy based on the reverse engineering of a positive response to a behavioural intervention may set the scene for future drug development trajectories for rare diseases.

ID: 121465

Start date: 01-04-2019

End date: 31-12-2022

Last modified: 01-04-2023

Grant number / URL: https://www.optimistic-dm.eu/recognition

Copyright information:

ReCognitION: Recognition and Validation of Druggable Targets from the Response to Cognitive Behaviour Therapy in Myotonic Dystrophy type 1 patients from Integrated -Omics Networks

1. General features of the project and data collection

1.1 Project leader contact details

Prof.dr. P.A.C. 't Hoen
Center for Molecular and Biomolecular Informatics
Radboud Institute for Molecular Life Sciences
Radboud university medical center
Route 260
PO Box 9501
6500 HB Nijmegen
The Netherlands
Email: peter-bram.thoen - at -radboudumc.nl

1.2 I have composed my DMP with the assistance of a data stewardship (or management) expert. List his or her name, function, organisation/department, phone number and email address.

- The expert is connected to my department or institution (please explain his/hr expertise related to data stewardship)

The DMP has been drafted by the main applicant, who is a data expert himself.
The Radboud Technology Center for Data Stewardship, in the person of Koen ten Hove, has been consulted on important issues

1.3 In collecting data for my project, I will do the following:

- Use existing data (please specify)
- Generate new data
- Add new data to an existing data set (please specify)

The new data that will be generated within the Dutch branch of the project is RNA-sequencing data and proteomics data on the blood samples of the OPTIMISTIC clinical trial.
The new data will be connected to existing RNA-seq data from the trial as well as all the patient characteristics, clinical outcome measures and trial response. The existing trial and the newly generated data are considered as privacy-sensitive and are available and can be combined within Radboudumc's secure Digital Research Environment, and can also be made available to external users from within this environment.

Other existing data necessary for the interpretation of the newly generated RNA-seq profiles are biological annotations of genes and transcripts. Those are available from public resources such as Reactome (European Bioinformatics Institute), WikiPathways, Uniprot, and an in-house graph database from ReCognitION partner Euretos B.V.

1.4 In my research, I will use:

- A combination of quantitative and qualitative data

Quantitative: RNA-seq data, proteomics data, most clinical outcome measures.
Qualitative: gene / transcript / protein annotations.

1.5 I will be reusing or combining existing data, and I have the owner's permission for that.

- Yes, I have permission to use the data

The OPTIMISTIC board has provided permission for the use of all trial data in the ReCognitION project. Data access to these data will be restricted to those granted permission, given the privacy-sensitive nature of the data. A data access policy has been defined and a data access committee consisting of the four PIs of the OPTIMISTIC study plus the PI of the ReCognitION committee has been established.

1.6 In collecting new data, I will be collaborating with other parties.

- No

The RNA-seq data have been generated by GenomeScan B.V. The proteomics data have been generated by ReCognitION partner Ghent University.

1.7 I am a member of a consortium of 2 or more partners. Clear arrangements have been made regarding data management and intellectual property. (also consider the possible effect of changes within the consortium on issues of data management and intellectual property)

- Yes, clear arrangements have been made regarding data management and intellectual property through a consortium agreement

We have signed a consortium agreement between the partners involved in ReCognitION: Radboudumc, Univeristy of Ghent, INSERM, University of Ottawa, University of Prague, CESC, Euretos B.V., University of Essen and University of Quebec.

1.8 I can give an estimate of the size of the data collection; specifically, the number of participants or subjects (“n=”) in the collection and its size in GB/TB

- Yes (please specify)

RNA-seq:
N=30 participants x 2 time points (baseline and after 10m of intervention)
Average size of raw RNA-seq data per sample: 10 GB
Average size of analysed data per sample: 20 GB
Total: 30x60 GB = 1.8 TB
N=256 participants x 2 time points (baseline and after 10m of intervention)
Average size of raw proteomics data per sample: 10 GB
Average size of analysed data per sample: negligible
Totaal: 512 x 10 GB = 5.2 TB.

1.9 The following end products I will make available for further research and verification (please elaborate briefly)

- Raw data
- (Several versions of) processed data
- Biobank
- Data documentation
- Documentation of the research process, including documentation of all participants
- Syntaxes

Raw and processed RNA-seq data and metadata in European Genome Archive (EGA) (data access committee in place and data transfer agreement in place).
Proteomics data and metadata in PRIDE repository at the European Bioinformatics Institute (to be done)
Biobank samples: via Eurobiobank at two locations (Newcastle and Munich).
Documentation, syntaxes in github.

1.10 During the project, I will have access to sufficient storage capacity and sites, and a backup of my data will be available. (please elaborate briefly)

- Yes, I will make use of my institution's standard facilities for storage and backup of my data

Data will be stored and analyzed in Radboud's Digital Research Environment, with fully backuped data storage, on the Microsoft Azure cloud.

2. Legislation (including privacy)

2.1 I will be doing research involving human subjects, and I am aware of and compliant with laws and regulations concerning privacy sensitive data.

- Yes, I will involve human subjects in my research. I will comply with the Algemene Verordening Gegevensbescherming (AVG)
- The Wet Medisch-Wetenschappelijk Onderzoek met Mensen (WMO, or Medical Research (Human Subjects) Act) applies to my project; I will have it reviewed by a Medical Research Ethics Committee. In addition I will comply with the Kwaliteitsborging Mensgebonden Onderzoek (Quality Assurance for Research Involving Human Subjects)

Use of samples and generation of data in ReCognitION falls under the goals of the OPTIMISTIC clinical trials. The OPTIMISTIC trial has been positively reviewed by Radboudumc's Medical Ethical Review board. Informed consent is available from all participants for the specific purposes of the ReCognitION project.

2.2 I will be doing research involving human subjects, and I have (a form of) informed consent from the participants for collecting their data.

- Yes (please describe the form this consent takes)

Yes, the specific questions on which the participants entered Yes is:
--
lk begrijp dat míjn bloed, urine en DNA materiaal maximaaltot 20 jaar kunnen worden opgeslagen, zullen worden geanonimiseerd (niet-identificeerbaar) en kunnen worden gebruĺkt door andere onderzoekers. Het onderzoek zal onderworpen worden aan een goede wetenschappelijke en ethische beoordelîng.
--
There is no specific question in the consent form on reuse of data.

2.3 I will be doing research involving human subjects, and I will protect my data against misuse.

- Yes, the data will be pseudonymised. (please explain how this will be done, and by which organisation) and

The samples have been pseudonymized within the context of the OPTIMISTIC trials. Within ReCognitION, samples are provided to the researchers with participants codes, and treated as anonymous within our project as, for this research, unsollicited findings will not be fed back to the participants.

2.4 I will stick to the privacy regulations of my organisation

- Yes

3. Making data findable

3.1 The data collection of my project will be findable for subsequent research. E.g., on a catalogue, a web portal, or through the search enginge of the repository (note: this is key item 3, which you should report to ZonMw at the end of your project).

- Yes, it can be found through the search engine of the archive or repository in which it is stored (please specify)

The RNA-seq data has been submitted to European Genome Archive (EGA). EGA has a fully searchable catalogue in place. Data is under controlled access given the sensitivity of the data

https://ega-archive.org/datasets/EGAD00001008383

https://ega-archive.org/datasets/EGAD00001010010

We are also in the process of building a catalogue of all datasets within Radboudumc's Digital Research Environment. Also from there, data will be findable.

3.2 I will use a metadata scheme for the description of my data collection (note: this is key item 7, which you should report to ZonMw at the end of your project).

- Yes, I will use a metadata scheme specific for my field of research (please specify)

The EGA has its own metadata schema. Difficult to change to more accepted standards like ISA or DCAT.

3.3 I will be using a persistent identifier as a permanent link to my data collection (note: this is key item 1, which you should report to ZonMw at the end of your project).

- No, I will not be using a persistent identifier (please explain)

EGA does not provide doi's but links will be valid for at least 10 years.

4. Making data accessible

4.1 Once the project has ended, my data will be accessible for further research and verification.

- Yes, immediately

4.2 Once the project has ended, my data collection will be publicly accessible, without any restrictions (open access).

- No, there will be access restrictions to my data collection (please explain)

The RNA-seq data is considered as privacy sensitive. Access is controlled by the OPTIMISTIC data access committee (in which all OPTIMISTIC board members have a seat) and granted to all researchers with a sound research plan and able to demonstrate they work within a safe and protected data infrastructure.

4.3 I have a set of terms of use available to me, which I will use to define the requirements of access to my data collection once the project has ended (please provide a link or persistent identifier; also note that this is a key item 4, which you should report to ZonMw at the conclusion of your project).

- Yes, my institution has drafted a set of terms of use with the help of a legal advisor

We reused a template defined by BBMRI-NL. It is available here: https://optimistic-dm.eu/recognition-data-access-agreement/

4.4 In the terms of use restricting access to my data, I have included at least the following:

- Conditions related to data security
- The manner in which the data set can be accessed
- A steering committee, programme committee or project leader will be charged with approving data requests
- Collaboration in using the data set, including agreements on publication and authorship
- Whether or not the data set may be linked with another data set (for reasons of privacy)

- The proper acknowledgement and citation of the data.

5. Making data interoperable

5.1 I will select a data format, which will allow other researchers and their computers (machine actionable) to read my data collection (note: this is key item 5, which you should report to ZonMw at the end of your project).

- Yes (please specify)

The standard and internationally accepted data formats for RNA-seq data are: .fastq and .bam
The standard and internationally accepted data formats for proteomics data are: mzML

5.2 I will select a terminology for recording my data (e.g., code, classification, ontology) that allows my dataset to be linked or integrated with other datasets (note: this is key item 6, which you should report to ZonMw at the conclusion of your project).

- Yes, metadata standard (please specify)

Experimental Factor Ontology
EDAM Ontology for bioinformatics operations, types of data, data identifiers, data formats, and topics
Data Use Ontology

5.3 I will be doing research involving human subjects, and I have taken into account the reuse of data and the potential combination with other data sets when taking privacy protection measurements.

- Yes, the participants have given their permission for reuse of the data, and the data have been pseudonymised

Use of the data should be in line with the consent given by the patients.

6. Making data reusable

6.1 I will ensure that the data and their documentation will be of sufficient quality to allow other researchers to interpret and reuse them (in a replication package).

- I will document the research process (please explain)
- I will perform quality checks on the data to ensure that they are complete, correct and consistent (please explain)
- In addition, I will take further quality assurance measures (please specify)
- I will document the software used in the course of the project (please specify)

The research process will be documented in the publications arising from the project. Processing and operations will be documented, where possible, in machine readable formats using the EDAM ontology.
The scripts and complete analysis pipelines are available through a dedicated github repository: https://github.com/cmbi/DM1_blood_RNAseq
Quality checks on data integrity will be done with md5 checksums.
Quality check on the data themselves are specific for the RNA-seq data and will include consistence of gender, genotypes, RNA integrity checks.

6.2 I have a number of selection criteria, which will allow me to determine which part of the data should be preserved once the project has ended. (see also question 1.9 and 6.1)

- No

All RNA-seq data have been preserved. Proteomics data coming as soon as data analysis is complete

6.3 Once the project has ended and the data have been selected, I can make an estimate of the size of the data collection (in GB/TB) to be preserved for long-term storage or

archival.

- Yes (please specify)

around 5 TB (see above)

6.4 I will select an archive or repository for (certified) long-term archiving of my data collection once the project has ended. (note: this is a key item, which you should report to ZonMw at the conclusion of your project)

- Yes, and this archive has a data seal of approval (please specify the archive)

RNA-seq data at European Genome Phenome Archive (EGA).
Proteomics at PRIDE

6.5 Once the project has ended, I will ensure that all data, software codes and research materials, published or unpublished, are managed and securely stored. Please specify the period of storage.

- Yes, in accordance with VNSU guidelines (please specify the number of years)

>= 15 years

6.6 Data management costs during the project and preparations for archival can be included in the project budget. These costs are:

- Unknown (please explain)

Budgets for computation and storage amounted to about 1250 euros (plus the basic costs for DRE usage that are covered by the institute)

6.7 The costs of archiving the data set once the project has ended are covered.

- Not yet (please explain)

Funding scheme does not allow for this. EGA and PRIDE store data for free.