
Analyses of genetic variants contributing to Parkinson's disease

A Data Management Plan created using DMPRoadmap

Creators: Maria Swanberg, Maria Swanberg

Affiliation: Lund University

Template: Lund University

Last modified: 29-06-2022

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Analyses of genetic variants contributing to Parkinson's disease

1. Data description

- using existing data (archival data, data from previous project – your own or other researchers')

The data that will be used is whole genome sequencing data and whole-genome genotyping data from individuals with and without a Parkinson's disease diagnosis. Four different cohorts will be analyzed in this project:

- 1) "The Accelerating Medicine Partnership in Parkinson's Disease" (AMP-PD)
- 2) "The Global Parkinson's Genetics Program" (GP2)
- 3) "Swegen"
- 4) "Epihealth"

The data has already been collected and is available to researchers after approval and after setting up data user agreements.

The material that will be used is solely digital. Data for four different cohorts/sites will be analysed:

- "The Accelerating Medicine Partnership in Parkinson's Disease (AMP-PD)" comprising of whole genome sequencing data from individuals with and without Parkinson's disease (PD).

- "The Global Parkinson's Genetics Program (GP2)" comprising of whole genome genotyping data from individuals with and without PD.

Both these cohorts are available for researchers at the site servers (Terra) upon approval and setting up a data user agreement. More information can be found at <https://amp-pd.org/>

- "Swegen" comprising of whole genome sequencing data from individuals without any known Parkinson's disease (will be used as "controls").

- Epihealth comprising of whole genome genotyping data from individuals without any known Parkinson's disease (will be used as "controls").

Individual level data from Swegen is available for researchers at the site servers (Uppmax Bianca) upon approval and setting up data user agreements.

Individual level data from EpiHealth will be delivered from the biobank at Karolinska Institute to Lund University upon approval.

The file formats will be files storing individual level genotype data such as .VCF, .BAM, and plink binary files (.bim, .bed, .fam). Files containing results from analyses will be .txt-files

- >1 TB

2. Documentation and data quality

Creator and date when analyses are done will always be documented and a README-file for each cohort used will be created containing meta-data on each cohort. To follow the analyses, scripts will be written (either using Jupyter notebook, R script, or similar), including information on software used and versions.

Only data that already has been collected will be used in this project.

3. Storage and backup

All analyses that will be performed on the individual level data in this project will be done on research systems dedicated for analyzing sensitive personal data.

For the data from AMP-PD and GP2, the analyses will be done on Terra, a cloud-native platform for biomedical researchers provided by Broad Institute. The Swegen dataset will be handled and analyzed at Bianca at UPPMAX and the EpiHealth data will be analyzed on L-SENS at LUNARC.

No individual level data will be copied, downloaded, or stored elsewhere than their original location on these systems. A copy of the results, scripts, and metafiles will be backed-up on LUCRIS. LUCRIS will also be used to store the data from the EpiHealth cohort. LUCRIS is a platform at LU for e.g. storing data in a secure way which is compatible with GDPR.

All analyses that will be performed on the individual level data in this project will be done on research systems dedicated for analyzing sensitive personal data. This includes systems that are a part of the SNIC-SENS project (Bianca at UPPMAX and L-SENS at LUNARC) and Terra, a cloud-native platform for biomedical researchers provided by Broad Institute. No individual level data will be copied, downloaded, or stored elsewhere than their original location on these systems. A copy of the results, scripts, and metafiles will be backed-up on LUCRIS. LUCRIS is a platform at LU for e.g. storing data at secure way which is compatible with GDPR. Potential collaborators can only access the individual level data after approval and after setting up a data management agreement with each individual site. We will not share any data with other researchers.

4. Legal and ethical requirements

- Yes – in that case you must report this in Pulu (<https://pulu.adm.lu.se>)

Sensitive data (genetic data) will be analyzed in the project. All study participants in each of the cohorts that will be used have given their written consents to participate and share their data. From our side, we will guarantee that we only handle the individual level data on each site's secure servers that are compatible with GDPR or by the rules implied in the specific country. No individual level sensitive data will be downloaded. All data is pseudoanonymized and the researchers in this project will not have access to the code key at any part.

An ethical review for the project has been written and we are waiting for an approval. We will not at any point attempt to identify any individual and no result will be reported on an individual level. All data handled is pseudoanonymized and we do not have access to the code key.

- No, LU is the sole party.

5. Data sharing and long-term preservation

- Yes, but only metadata

Metadata will be available through peer-reviewed scientific publications reporting the research results and scripts used will be available on Github to enable open science and replication when the results have been published. We will not download or share any individual level data. Potential data users have to send in their own applications to each individual site before taking part of the data.

The metadata (here: results) will obtain a DOI at publication in a journal.

- No data will neither be created nor collected.

6. Responsibilities and resources

Research members in the project (Kajsa Brolin [KB] and Maria Swanberg [MS]) are responsible for implementing, reviewing, and revising the DMP. KB and MS are also responsible for keeping the records of the results and the scripts used in the project. Each individual site for the cohort analyzed in this project are responsible for the data management of the original individual level data.

Each individual site for the cohort analyzed in this project are responsible for the data management of the individual level data. Only meta-data and scripts are needed to be stored by the researchers in this project which will be done at LUSEC. No skills outside the project needs to be hired and no special equipment is needed.