Plan Overview

A Data Management Plan created using DMPonline

Title: Shaping of Human Immune Systems by Environmental Influences Early in Life

Creator: Petter Brodin

Affiliation: Karolinska Institutet

Funder: Swedish Research Council

Template: Swedish Research Council Template

ORCID iD: 0000-0002-8103-0046

Project abstract:

Studies of environmental influence in humans are complicated to perform because of the many simultaneous factors of influence. Still, understanding these environmental influences is essential since they explain most of the variation in immune system composition and function among individuals (Brodin et al., 2015). By studying newborn children facing many environmental factors for the very first time, adaptive changes induced by environmental exposures never seen before, are more interpretable. By simultaneously analyzing frequencies, phenotypes and functional responses of all immune cell populations and across many plasma proteins, we reason that coordinated changes can be uncovered and regulatory relationships inferred. By comparing such immune system changes in children born at term or preterm, by vaginal or cesarean delivery and fed breast milk or formula, we aim to investigate the influences of such broadly different initial conditions on immune development. Since newborn children, particularly those born preterm are at increased risk of both infectious diseases and inflammatory conditions, this study holds potential for clinical impact by enabling earlier detection and treatment of such devastating conditions. In a longer perspective, by understanding the timing and mechanisms of environmental imprinting on human immune systems, we hope to understand immune variation, predict risk of immune mediated disease and improve vaccine effects by optimizing vaccine schedules to the state of newborn immune systems.

ID: 50232

Last modified: 08-02-2020

Grant number / URL: 2019-01495

Shaping of Human Immune Systems by Environmental Influences Early in Life

General Information

Project Title

Shaping of Human Immune Systems by Environmental Influences Early in Life

Project Leader

Petter Brodin

Registration number at the Swedish Research Council

2019-01495

Version

1

Date

2020-02-02

Description of data – reuse of existing data and/or production of new data

How will data be collected, created or reused?

Data will be in the form of biomedical data from immune cells and proteins obtained by various high-throughput methods from blood samples. We will assess cell phenotypes, plasma protein concentrations, mRNA-expression. Also, fecal samples will be obtained and composition and function of gut microbes are analyzed by Shotgun next generation sequencing. Finally, clinical information is gathered into a database to use for correlation analyses between clinical metadata and biomedical data.

What types of data will be created and/or collected, in terms of data format and amount/volume of data?

Single-cell data obtained by Mass cytometry and single-cell mRNA-sequencing. Plasma protein relative abundances by Olink assays. Fecal microbiome composition and function by Shotgun next generation sequencing. Clinical data obtained from electronic health records.

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, etc.?

All data is coded by individual patient without unnecessary identifiable information like names, personal numbers or addresses and

all data is gathered into a relational database. This ensures good quality control, version control and tracability of all data. The Database is located at SciLifeLab (cytof.scilifelab.se) and is owned by department of women's and children's health, Karolinska Institutet.

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

All data are longitudinal in nature. Sample preparation is optimized to be done as soon as posible after collection with minimal technical variation (Ref Brodin et al, Immunity, 2019). Using a relational database instead of excel tables or similar ensures quality. The database does include experimental data, protocol files and analyses scripts are available in GitHub to reproduce the analyses.

Storage and backup

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

Relational database custom written for the project by the Brodin lab an hosted in our lab server within the firewall of KI

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

the database is fully password protected and only known IPs can gain access. The database is hosted in our lab server within the firewall of KI

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?

In accordance with GDPR we have minimized the amount of data collected to the bare minimum needed for analyses. We code all pateints and remove personal numbers, names and other unuecessary, sensitive information before adding data to the databases. The info is fully traceable because there is a locked key to decode study IDs to personal numbers. Only the clinical study coordinator have access to this. All subjects have consented to participation of the study after reciveing both written and oral study information.

How is correct data handling according to ethical aspects safeguarded?

Safe data storage system, with full traceability and all informed consent. All data is removed upon request from study participants.

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 and limitations on the access to and reuse of data to be considered?

The data is accessed for analyses purposes only by Brodin lab members and clinical study coordinators. A minimal number of individuals have access as imposed by GDPR. Longterm storage is ensured by backup of the relative database within the KI firewall.

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

Yes by Brodin lab data manager responsible for data infrastructure as well as the KI IT department in charge of the university infrastructure.

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

All analyses and data handling, storage and management is done by custom code written by the Brodin lab. We have a GitHub repository for our group where all analysis code is stored and version controlled. We share all code to reproduce all analyses in the form of a Docker system with every publication to allow readers and researchers to reproduce all our analyses as is state of the art in our field of research.

How will the use of unique and persistent identifiers, such as a Digital Object Identifier (DOI), be safeguarded?

These are only available to study coordinators within our closed database.

Responsibility and resources

Who is responsible for data management and (possibly) supports the work with this while the research project is in progress? Who is responsible for data management, ongoing management and long-term storage after the research project has ended?

KI Data protection officer is Mats Gustavsson. All management of data is managed by Petter Brodin and his lab members involved in the stuyd, as well as clinical collaborators headed by Kajsa Bohlin.

What resources (costs, labour input or other) will be required for data management (including storage, back-up, provision of access and processing for long-term storage)? What resources will be needed to ensure that data fulfil the FAIR principles?

Full time data infrastructure manager employed by the Brodin lab. A dedicated server is in place within the firewall of KI. All necessary infrastructure is available and suitable to esure FAIR principles.