

Plan Overview

A Data Management Plan created using DMPTool

DMP ID: <https://doi.org/10.48321/D19061>

Title: The University of Chicago Consortium for Food Allergy Research Clinical Research Center

Creator: Christina Ciaccio - **ORCID:** [0000-0002-5059-0550](https://orcid.org/0000-0002-5059-0550)

Affiliation: University of Chicago (uchicago.edu)

Principal Investigator: Christina Ciaccio

Funder: National Institutes of Health (nih.gov)

Funding opportunity number: RFA-AI-22-076

Grant: <https://grants.nih.gov/grants/guide/rfa-files/RFA-AI-22-076.html>

Template: NIH-Default DMSP

Project abstract:

Food Allergy (FA) is a common chronic condition associated with life-threatening allergic reactions in all age groups. As an estimated 8% of children and 11% of adults in the United States are affected by FA, a significant need exists to expand the body of knowledge around prevention, diagnosis and treatment across all age groups. One of the most critical needs in FA is the development of minimally and/or non-invasive diagnostic tests and algorithms that provide both highly sensitive and highly specific results. The most commonly used tests for FA, percutaneous skin prick testing (SPT) and allergen-specific IgE (sIgE), often fail to accurately predict clinical reactivity. The double-blind placebo controlled oral food challenge (DBPCFC) therefore remains the gold standard test for confirming FA, determining reactivity thresholds, and understanding reaction severity. The DBPCFC has several limitations, however, as it requires hours to complete, causes significant morbidity, and has substantial site to site variability despite efforts to standardize protocols. These limitations result in high procedural costs, long wait times, and paralyzing anxiety that make this test impractical in many clinical settings. Without accurate, inexpensive, and widely available diagnostics, FA clinical trials are impeded by the high associated cost and reluctance by potential volunteers to incur the added risk. In addition, the accuracy of epidemiologic studies is limited. Thus, we believe developing novel diagnostic tools that accurately detect the presence of disease, the threshold of reactivity, the severity of reactions, the response to therapy and the early detection of anaphylaxis is an essential step to improving the lives of individuals with FA. The objective of this proposal, therefore, is to develop novel diagnostic methods for IgE-mediated FA through already established collaborations at the UChicago in cooperation with and to support the

Consortium for Food Allergy Research (CoFAR). The central hypothesis is that repurposing established platforms with track records of success developed by scientists at UChicago outside the field of FA will allow for the rapid identification of novel biomarkers of FA and anaphylaxis providing the fundamental first step to the development of improved FA diagnostic platforms. Further, we anticipate the generation of large datasets in our aims and believe the creation of a FA data commons to organize and share these datasets, as well as all datasets generated by CoFAR, is fundamental to efficient discovery in FA. In order to achieve our objective, we will perform highly sensitive 5mC profiling utilizing LABS-seq on cfDNA from a cross-sectional cohort of human subjects with peanut or egg allergy and use non-invasive laser Doppler to measure cutaneous blood flow during SPT and during OFC and determine when changes in blood flow occur with respect to allergen SPT and food doses. Finally, we propose the creation of a FA data commons across CoFAR sites to store all data generated from the Consortium.

Start date: 01-01-2024

End date: 12-31-2030

Last modified: 06-07-2023

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 customize 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

The University of Chicago Consortium for Food Allergy Research Clinical Research Center

Data Type

Types and amount of scientific data expected to be generated in the project: *Summarize the types and estimated amount of scientific data expected to be generated in the project.*

Describe data in general terms that address the type and amount/size of scientific data expected to be collected and used in the project (e.g., 256-channel EEG data and fMRI images from ~50 research participants). Descriptions may indicate the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing that has occurred (i.e., how raw or processed the data will be)

Data source	Number of Subjects	Condition	File Format	File Size (approx)
LABS-seq	130	Peanut Allergy	.fastq and .txt	500 gigabytes
LABS-seq	130	Egg Allergy	.fastq and .txt	500 gigabytes
Doppler ultrasound	250	Food Allergy	.vmr and .txt	40 gigabytes

Scientific data that will be preserved and shared, and the rationale for doing so: *Describe which scientific data from the project will be preserved and shared and provide the rationale for this decision.*

In this proposed project, the cleaned, clinical/demographic spreadsheet data for all variables will be shared openly. The rationale for sharing only cleaned data is to foster ease of data reuse.

All LABS-seq and Doppler data will be preserved and shared.

Metadata, other relevant data, and associated documentation: Briefly list the metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) that will be made accessible to facilitate interpretation of the scientific data.

To facilitate interpretation and reuse of the data, a README file and data dictionary will be generated and deposited into a repository along with all shared datasets. The data dictionary will define and describe all variables in the dataset.

In addition, protocols and data collection instruments will be shared and linked to the relevant datasets.

Related Tools, Software and/or Code

State whether specialized tools, software, and/or code are needed to access or manipulate shared scientific data, and if so, provide the name(s) of the needed tool(s) and software and specify how they can be accessed.

Raw seq (.fastq) is a simple text file and can be opened with a text editor. The following R analytics packages are openly available to reuse the genomic data include, EnrichR, ClusterProfiler, Reactome PA, and MethylKit. Other associated analytic code will be provided through GitHub. Doppler (.txt) data can be analyzed in Excel or MatLAB.

Standards

State what common data standards will be applied to the scientific data and associated metadata to enable interoperability of datasets and resources, and provide the name(s) of the data standards that will be applied and describe how these data standards will be applied to the scientific data generated by the research proposed in this project. If applicable, indicate that no consensus standards exist

Data sharing policies will adhere to the principles of FAIR (Findable, Accessible, Interoperable, and Reusable) data management.

Epigenomic data will follow MINSEQ-E

Doppler data will follow best practices,

Data Preservation, Access, and Associated Timelines

Repository where scientific data and metadata will be archived: Provide the name of the repository(ies) where scientific data and metadata arising from the project will be archived; see [Selecting a Data Repository](#)

Epigenome data will be stored in GEO.

If our proposed NIAID Food Allergy Data Commons is funded, our Doppler data will be permanently housed in this Data Commons.

If our proposed NIAID Food Allergy Data Commons is not funded, our Doppler data will be deposited and shared via Knowledge@UChicago (<https://knowledge.uchicago.edu/>), which provides metadata, ensures long-term access, and registers a digital object identifier (DOI) for each dataset to facilitate discoverability and citation. Additionally, the dataset(s) will be openly licensed and made publicly available as soon as possible or at the time of associated publication. As the University of Chicago's institutional repository, Knowledge@UChicago is supported collaboratively by the University's Library and IT Services. It is built on a cloud-based platform maintained by a service provider named TIND. Knowledge@UChicago uses an open archival information system (OAIS) compliant approach to preservation, which is complemented by fixity checking, redundancy backup, and storage of archival packages on geographically separated servers.

Data will be retained for a minimum of 10 years following project completion, in accordance with the University of Chicago's data retention policy. After this period, data may be archived in institutional repositories or other appropriate long-term storage facilities to ensure continued access and preservation.

How scientific data will be findable and identifiable: Describe how the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools.

The data will be made identifiable in the specified data repositories via persistent unique identifiers and multiple keywords to maximize search results. Dataset DOIs will be included in associated publications.

When and how long the scientific data will be made available: Describe when the scientific data will be made available to other users (i.e., no later than time of an associated publication or end of the performance period, whichever comes first) and for how long data will be available.

Data will be made available at the time of associated publication. The data deposited in GEO and the Food Allergy Data Commons (upon creation) will be available indefinitely. Data stored at the Center for Research Informatics at the University of Chicago will be stored for a minimum of three years.

Access, Distribution, or Reuse Considerations

Factors affecting subsequent access, distribution, or reuse of scientific data: NIH expects that in drafting Plans, researchers maximize the appropriate sharing of scientific data. Describe and justify any applicable factors or data use limitations affecting subsequent access, distribution, or reuse of scientific data related to informed consent, privacy and confidentiality protections, and any other considerations that may limit the extent of data sharing. See [Frequently Asked Questions](#) for examples of justifiable reasons for limiting sharing of data.

There are no anticipated factors or limitations that will affect the access, distribution or reuse of the scientific data generated by the proposal.

Whether access to scientific data will be controlled: State whether access to the scientific data will be controlled (i.e., made available by a data repository only after approval).

Access to shared data will not be controlled.

Protections for privacy, rights, and confidentiality of human research participants:

If generating scientific data derived from humans, describe how the privacy, rights, and confidentiality of human research participants will be protected (e.g., through de-identification, Certificates of Confidentiality, and other protective measures).

All shared data will be deidentified

Oversight of Data Management and Sharing

Describe how compliance with this Plan will be monitored and managed, frequency of oversight, and by whom at your institution (e.g., titles, roles).

The PI, Dr. Christina Ciaccio, in collaboration with coinvestigators, will be responsible for overseeing the data management plan and ensuring compliance during the project. The PI will be responsible for updating and revising the Data Management and Sharing Plan when necessary and reporting compliance annually.

Planned Research Outputs

Dataset - "Doppler Data"

.vmr and .txt file of Doppler data generated during anaphylaxis

Dataset - "Metagenomic data of food allergy patients"

LABS-seq data of peanut allergy and egg allergy patients and controls

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Doppler Data	Dataset	2027-12-29	Open	Knowledge@UChicago	40 GB	Creative Commons Attribution Non Commercial Share Alike 4.0 International	None specified	No	No
Metagenomic data of food allergy patients	Dataset	2027-12-30	Open	Gene Expression Omnibus	500 GB	Creative Commons Attribution Non Commercial Share Alike 4.0 International	Minimum Information about any Sequence (MIxS)	No	No