Plan Overview

A Data Management Plan created using DMP Tool

DMP ID: https://doi.org/10.48321/D14851d9a3

Title: Utility of Capillary Blood for Gene Expression Studies

Creator: Scott Nicholson - **ORCID:** 0000-0002-2201-744X

Affiliation: Civil Aerospace Medical Institute (faa.gov)

Funder: United States Department of Transportation (DOT) (transportation.gov)

Funding opportunity number: na

Grant: na

Template: Federal Aviation Administration (FAA) Data Management Plan (DMP) Template v1.1

Project abstract:

Blood RNA is typically derived from blood tubes collected using standard venous phlebotomy practices. Venipuncture provides ample material for RNA extraction, but presents challenges when considering the time, labor, and acceptability among study participants. This study examines the utility of capillary blood collected through standard fingerstick practices as a replacement for venous blood as a source for blood RNA for RNA sequencing. Volunteer subjects will provide venous and fingerstick blood, and RNA from each sample will be extracted from each sample. Extracted RNA will be sequenced, and sequence data from each collection method will be compared to assess what differences exist between sample collection methods. If capillary blood RNA is determined to be a reliable source of gene expression data with equivalence to venous blood RNA, the methods developed in this study may be applied to future blood sample collections intended for RNA sequencing.

Start date: 09-13-2019

End date: 03-14-2024

Last modified: 07-08-2024

Utility of Capillary Blood for Gene Expression Studies

Persistent Link

Include the persistent identifier (PID) that is associated with the dataset.

Persistent Link:

Report: https://doi.org/10.21949/1529628

Dataset: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs003496.v1.p1

Recommended Citation

The recommendeded data citation to be used when citing the dataset.

Recommended Citation:

"Utility of Capillary Blood" (2023). https://doi.org/10.21949/1529628

Change Log

Document the changes that are made to the DMP, any and all changes should be noted to ensure a more complete documentation.

Change Log:
2023_01_13: DMP created

2024_03_14: Project closed

Table of Contents

Optional table of contents included here, in order to better organize the DMP.

none

0. Dataset and Contact Information

Please provide the following information:

- **Name of the dataset or project for which data is being collected.**
- **Name of the FAA Line-Of-Business/Office for which the associated dataset is being generated.**
- **Email for the FAA Line-Of-Business/Office (key field).**
- **If applicable and as reference, project number, contract number, or other number used to link this DMP.**

AAM-612

scott.nicholson@faa.gov

OC4.3, A11J.AM.47

1. Data Description

Name the data, data collection project, or data producing program. Provide high level narrative.

RNAseq data in the form of fastq files will be produced during this research.

Other data, including aviation accident data, drug detection data, and other data will be in the form of .txt., .xlxs, .csv, or other standard formats.

All data will be de-identified.

Describe the purpose of your research and whether results will be documented in a published document or report. How will it be used?

This research will determine if capillary blood RNA is useful to determine gene expression using RNA sequencing, and in analyzing gene expression differences between sample groups.

Results will be made publicly available through a technical report. (https://rosap.ntl.bts.gov/view/dot/73577)

Data will be made available through a technical report, and by posting data on the National Transportation Library and in the NIH database of Genotypes and Phenotypes (dbGaP, https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs003496.v1.p1).

Describe the data that will be generated in terms of nature and scale (e.g., numerical data, image data, text sequences, video, audio, database, modeling data, source code, etc.).

All data produced during this study will be in a text format, as described above.

Describe methods for creating the data (e.g., simulated; observed; experimental; software; physical collections; sensors; satellite; enforcement activities; researcher-generated databases, tables, and/or spreadsheets; instrument generated digital data output such as images and video; etc).

Data will be collected from observations and RNA sequencing.

Describe the period of time over which the data will be collected and frequency at which it will be updated.

The data was collected between January and June of 2022. Each data collection consisted of approx. 30 minutes. 41 subjects provided samples.

If using existing data, describe the relationship between the data you are collecting and existing data.

None.

Describe potential users of the data and the expected manner in which they may use it.

The data may be used in scientific research, such as medical or addiction research, and also by policymakers, industry researchers, or other unanticipated users.

Discuss the potential value of having the data available not only to your institution but also for the public, e.g., might be renewed interest and value in reanalyzing the data with updated and more universally comparable metrics or recently developed analytical methods.

This data may be useful for medical and academic research.

State clearly if data can be shared publicly or not. If you request permission not to make data publicly accessible, explain rationale for lack of public access.

Sequence Data from this study will be shared through a limited-access government database, and made available for legitimate research purposes.

Other data will be shared publicly through a technical report.

Indicate the party responsible for managing the data.

FAA AAM-600

Describe how you will check for adherence to this data management plan.

NA

2. Standards Employed

List in what format(s) the data will be collected. Indicate if they are open or proprietary.

Sequence data will be in .fastq format

Other data will be in standard .txt, .doc, .docx, or .xlsx format.

All are open formats

If you are using proprietary data formats, discuss your rationale for using those standards and formats.

NA

Describe how versions of data be signified and/or controlled.

Original data will be stored without modification.

Interim data versioning will be reflected by file naming including dates, version numbers, or iteration numbers.

Final data will be marked 'final'.

If the file format(s) you are using is(are) not standard to your field, describe how you will document the alternative you are using.

NA

List what documentation you will be creating in order to make the data understandable by other researchers.

The data generated here will be described in technical reports and using metadata standards consistent with NCBI repository requirements.

Indicate what metadata schema you are using to describe the data. If the metadata schema is not one standard for your field, discuss your rationale for using that scheme.

Standard NCBI metadata schema will be employed.

Describe how will the metadata be managed and stored.

NA

Indicate what tools or software is required to read or view the data.

Standard text editors may be employed to read the raw data.

The raw data is not amenable to direct viewing, summaries of the data and its significance will be made available in technical report(s).

Describe your quality control measures.

There are a number of open-source quality control software packages, such as multiQC and fastQC, that will be used to assess data quality.

3. Access Policies

Describe what data will be publicly shared, how data files will be shared, and how others will access them.

Technical reports including non-identifying raw and summarized data will be made available through Office of Aerospace Medicine reports and other publicly-accessible technical publications.

Raw sequence data will be made available to researchers through restricted-access NIH repositories designed to house that data.

Indicate whether the data contain private or confidential information. If so:

- **Discuss how will you guard against disclosure of identities and/or confidential business information.**
- **List what processes you will follow to provide informed consent to participants.**
- **State the party responsible for protecting the data.**

All data involved in this study was de-identified at the site of collection.

RNA sequence data will be treated as sensitive data, and will be released in the restricted-access dbGaP NIH repository as described previously. https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs003496.v1.p1

If applicable, describe how you will deidentify your data before sharing. If not:

- **Identify what restrictions on access and use you will place on the data.**
- **Discuss additional steps, if any you will use to protect privacy and confidentiality.**

Non-sequence data will not be restricted.

Sequence data will be available to researchers who apply for access through the dbGaP data use committee.

4. Re-Use, Redistribution, and Derivative Products Policies

Name who has the right to manage the data.

Unless otherwise noted, the data described in this DMP is generated and managed by the Federal Aviation Administration. The data are in the public domain, and may be re-used without restriction.

Indicate who holds the intellectual property rights to the data.

Unless otherwise noted (e.g., data is partially proprietary by an external entity, where intellectual property is shared), this data is required to be made available in open, machine-readable formats, while continuing to ensure privacy and security in accordance with the OPEN Government Data Act, which is Title II of the Foundations for Evidence-Based Policymaking Act.

List any copyrights to the data. If so, indicate who owns them.

NA

Discuss any rights that are transferred to a data archive.

There are no rights transferred to the permanent archive or repository to accompany this dataset described in this DMP.

Describe how your data will be licensed for reuse, redistribution, and derivative products.

Unless otherwise noted, there is not a need for the data in this DMP to be licensed for reuse, redistribution, and/or its derivative products.

5. Archiving and Preservation Plans

Discuss how you intend to archive your data and where (include URL).

Unless otherwise noted, the data described in this DMP will be uploaded to the FAA's Enterprise Information Management (EIM) through the FAA Data Governance Center. This is the internal FAA landing page and access point to EIM uploaded datasets and processes. Here the metadata is curated and validated for quality and accuracy. The FAA Data Steward enters metadata and verifies quality and accuracy before publishing to data.faa.gov, which is the FAA's clearinghouse site for publicly available FAA data and managed and hosted by the FAA's, IT Shared Services organization - Chief Data Office, see https://catalog.data.faa.gov/about for more information.

Technical reports and non-sequence data generated in this project will be placed in the National Transportation library

Sequence data generated during this project will be placed in the NIH dbGaP repository at accession number phs003496.v1.p1.

Indicate the approximate time period between data collection and submission to the archive.

Data and all research products (e.g., reports) were submitted within the period-of-performance of the research, which concluded 03/14/2024.

Identify where data will be stored prior to being sent to an archive.

Data will be stored in a secure cloud environment and on secure drives within CAMI.

Describe how back-up, disaster recovery, off-site data storage, and other redundant storage strategies will be used to ensure the data's security and integrity, initially and for the long-term.

Off-site data storage will be provided within a secure cloud environment.

Small data files will be stored in duplicate on separate secure drives.

Describe how data will be protected from accidental or malicious modification or deletion prior to receipt by the archive.

Data will be stored in controlled-access sites, access will not be granted to individuals not involved in the research.

Original data will be stored in multiple secure locations.

Indicate how long the chosen archive will retain the data.

Unless otherwise noted, the long term storage of the data described in this DMP will persist indefinitely in the FAA's Enterprise Information Management (EIM) platform following standard government policies and best practices.

Raw sequence data will be stored on the NIH dbGaP repository (accession phs003496.v1.p1).

Indicate if the chosen archive employs, or allows for the recording of, persistent identifiers linked to the data.

dbGaP employs persistent identifiers.

Discuss how your chosen data repository meets the criteria outlined on the Guidelines for Evaluating Repositories for Conformance with the DOT Public Access Plan page.

NIH dbGaP is a government-maintained repository that meets all federal requirements and serves as the repository of record for human sequence data.

6. Policies Affecting this Data Management Plan

Include policies that the data management plan was created to meet, such as the DOT public access plan.

This data management plan was created to meet the requirements enumerated in the U.S. Department of Transportation's "Plan to Increase Public Access to the Results of Federally-Funded Scientific Research" Version 1.1 << https://doi.org/10.21949/1520559 >> and guidelines suggested by the DOT Public Access website << https://doi.org/10.21949/1503647 >>, in effect and current as of 2023_01_13.

Planned Research Outputs

Dataset - "Utility of Capillary Blood for Gene Expression Studies"

Two methods of capillary blood collection and RNA extraction were compared to standard venipucture collection and extraction methods. Comparisons of global gene expression were used as the basis of comparison.

Found at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs003496.v1.p1

Data paper - "Utility of Capillary Blood for Gene Expression Studies Research Paper"

Office of Aerospace Medicine Report describing the problem, method, data, and data analysis regarding this project.

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?
Utility of Capillary Blood for Gene Expression Stu ...	Dataset	2024-09-29	Restricted	Database of Genotypes and Phenotypes		Creative Commons Attribution 4.0 International	None specified	No	No
Utility of Capillary Blood for Gene Expression Stu ...	Data paper	2024-03-11	Open	None specified		Creative Commons Attribution 4.0 International	None specified	No	No