Plan Overview

A Data Management Plan created using DMPTool

DMP ID: https://doi.org/10.48321/D1401B3ec0

Title: Placental and breastmilk microRNAs in relation to early-life growth and metabolism

Creator: Elizabeth Goehring - **ORCID:** 0000-0003-0074-2840

Affiliation: Emory University (emory.edu)

Funder: National Institutes of Health (nih.gov)

Funding opportunity number: PA-20-188

Grant: K99HD104991

Template: NIH-Default DMSP

Project abstract:

Gestational and early childhood exposures, including maternal metabolic status and breastfeeding, are increasingly recognized as critical in the development of childhood and life-long obesity. The proposed research seeks to comprehensively assess the impact of maternal metabolic status on early childhood growth and metabolism, through the gene regulatory dynamics of placental and breastmilk microRNAs. This research has the potential to advance our understanding of the contribution of microRNAs and early-life exposures on growth and metabolism, and to more broadly contribute the Developmental Origins of Health and Disease.

Start date: 02-01-2024

End date: 02-01-2027

Last modified: 01-24-2024

Copyright information:

Placental and breastmilk microRNAs in relation to early-life growth and metabolism

This project will produce gene expression data and hormone concentration data from RNA-sequencing and immunoassay, respectively. Demographic, health endpoint and survey data as well as biospecimen collection will be performed in the Atlanta Metropolitan area with from lactating mothers and their infants over a 6 to 7 month period. Data for sequencing analysis will be collected from 50 research participants at 3 time points, generating a sequencing data for approximately 150 samples. Data for hormone data will be collected from 50 research participants at 2 time points, generating approximately 700 observations for 100 samples. Raw sequencing data files will be processed to generate CSV files containing gene-level counts for all participants, to enable statistical analysis.

Raw sequencing read files and metadata will be accessible through dbGAP.

Processed count data and metadata will be provided at the Gene Expression Omnibus.

All analysis codes and de-identified files needed to reproduce analyses will be shared through Emory Dataverse. Each dataset created will be accompanied by a R markdown file to document the preprocessing steps, basic statistics, and profile figures such as the intended sample heterogeneity shown in PCA, and correlation heatmaps.

For gene expression and hormone data, metadata (including detailed methodology and procedures used to collect the data, data labels, definitions of variables, and any other information necessary to reproduce and understand the data) will be provided within the Material and Methods section or additional Supplementary information of each peer-reviewed publication. For RNA sequencing data, associated metadata will accompany the data sharing through Emory Dataverse.

To facilitate the interpretation and reuse of the data, a README file and data dictionary will be generated and deposited into Emory Dataverse along with all shared datasets. The README file will include methods descriptions (e.g., software, databases, libraries, services), and Protocol DOIs issued from protocols.io. The data dictionary will define and describe all variables in the dataset.

No new tools, software, or code will be developed during the project.

All RNAseq analyses are performed using publicly available software packages available through CRAN and bioconductor.

As noted in (1) above, we will, when possible, make use of public data repositories for storage of all sequencing data and results, applying onotologies and formats required by those repositories. We will use a standardized ontology in line with those used in other large population studies such as NHANES. For any provided data, detailed data dictionaries, metadata and relevant references will be provide which define all variables.

Raw sequencing data and associated metadata from this study will be accessible through dbGAP.

Processed data in the form of a count matrix and associated metadata, will be provided at the Gene Expression Omnibus.

Findable: All public archives utilize a variety of search features for data to be findable based on metadata criteria.

Identifiable: Data stored in public archives such as dbGAP, GEO, or SRA are given uniquely identifiable accession

numbers. We will include those accession numbers in all publications arising from this research.

RNA sequencing data will be shared within 6 months of data completion and will remain available in perpetuity in public archives.

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

Sequencing data made accessible in dbGAP is controlled and requires that investigators requesting the data have training in the conduct of human subject research and have received appropriate IRB approvals or waivers to work with such data. Researchers requesting access to read-level human data will need to request such access through dbGAP.

As noted above, sequence read-level data on human participants will be made available through dbGAP, given the potential for brief in confidentiality associated with that data. Accompanying exposure and covariate data will provided in such as way that it cannot be linked to individual participants. We will follow the Safe Harbor standards in line with Section 164.514(c) of the HIPAA privacy rule.

Elizabeth Kennedy (PI, ORCID: 0000-0003-0074-2840) will be responsible for the day-to-day oversight of lab data management activities and data sharing. Broader issues of DMS Plan compliance oversight and reporting will be handled by the PI as part of general Emory University stewardship, reporting, and compliance processes.

Planned Research Outputs

Dataset - "RNA-sequencing reads"

RNA sequencing reads and metadata generated from breastmilk exosomes for 50 study participants over three time periods.

Dataset - "Hormone concentration data"

Salivary hormone concentration data from 50 infants at 1 and 6 months of age.

Text - "Code, metadata and de-identified, individual level data"

Information necessary to perform analyses related to this study

Dataset - "Sequence counts"

Processed count matrix and associated metadata for sequencing performed on breastmilk exosomes from 50 participants across 3 time points.

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?
RNA-sequencing reads	Dataset	2026-01-31	Restricted	NCBI dbGaP		None specified	None specified	No	Yes
Hormone concentration data	Dataset	2026-01-31	Open	Emory Dataverse		None specified	None specified	No	No
Code, metadata and de-identified, individual level ...	Text	Unspecified	Open	Emory Dataverse		None specified	None specified	No	No
Sequence counts	Dataset	2026-01-31	Open	Gene Expression Omnibus		None specified	None specified	No	No