



Longitudinal and Integrative sample and data management for large scale cohorts



MONIKA AVINA, DENIS SALINS, SHANA LEOPOLD, BENJAMIN LEOPOLD, MICHAEL P. SNYDER, ERICA WEINSTOCK, GEORGE WEINSTOCK,

SNYDER LAB, STANFORD UNIVERSITY, JACKSON LABORATORY FOR GENOMIC MEDICINE

Abstract

Type 2 diabetes mellitus (T2D) is a significant health problem facing our nation and close to 20 million individuals in the United States have T2D. Currently, there are an estimated 79M prediabetics in the United States who have a lifetime risk of diabetes conversion of 50%. Differences in the gut microbiome have been noted between diabetics and healthy individuals, and direct alteration of the microbiome in mice has been shown to lower glucose levels. To better understand the biological changes that occur during T2D disease acquisition, we plan to perform a detailed analysis of the biological processes that occur in patients at risk for T2D and their microbiomes by longitudinal profiling of patients during healthy periods and periods of stress. Towards this goal, 89 consented individuals at risk for diabetes as well as individuals with or without diabetes were recruited. Samples were collected frequently (at 1-to-4-day intervals) during infected and other stress states and less frequently (every ~3 months) during healthy periods, with an average of 27 timepoints sampled per subject. Samples were collected and processed at Clinical & translational research unit (CTRU). Samples were retrieved after processing and stored and organized at -80°C. Sample management and tracking was done using FreezerWorks, a sample management software. This software allows us to visualize the entire inventory virtually or access a few samples at a time and to share samples between team members and labs.

Participant Recruitment

We recruited generally healthy adult participants primarily through internet, radio, and newspaper advertisements. At the onset of the study, the majority of the cohort was pre-diabetic and potential participants were screened for pre-diabetic status with an oral glucose tolerance test (OGTT). At a later point in the study, inclusion criteria was broadened to include diabetic as well as non-diabetic or pre diabetic participants.

Sample Collection

A Research coordinator contacted participants and scheduled an IPOP visit. Reservations were made at CTRU and specific protocols were provided to the nursing staff which in turn collected different sample types (Blood, tissue, stool and urine) from the participant. Samples were processed by the CTRU staff according to specified protocols and prepared for collection by the research team.

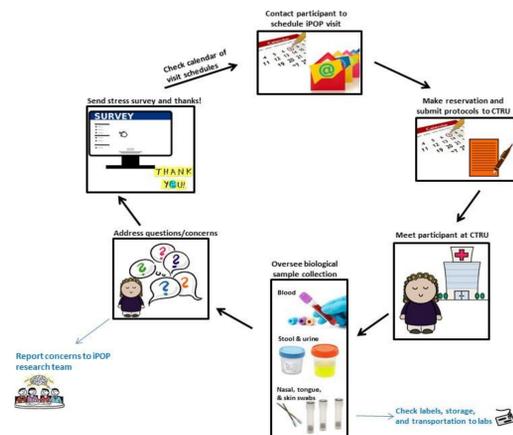


Figure 1. Above figure represents the sample collection workflow which started with a research coordinator contacting the participants and concluded with sample collection by nursing staff at CTRU.

Sample Management

Collected samples were organized and stored at -80°C. For sample management and tracking, FreezerWorks, a configurable sample management software, was utilized. FreezerWorks allowed us to track each and every sample at every stage and to visualize the entire inventory at any time. At the same time, we were able to share samples between different team members and labs.

FreezerWorks workflow

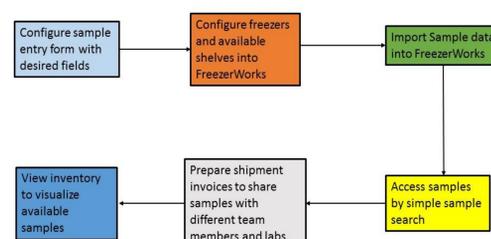


Figure 2. Above figure represents FreezerWorks workflow. Samples can be easily located and shared among different labs and team members through this software.

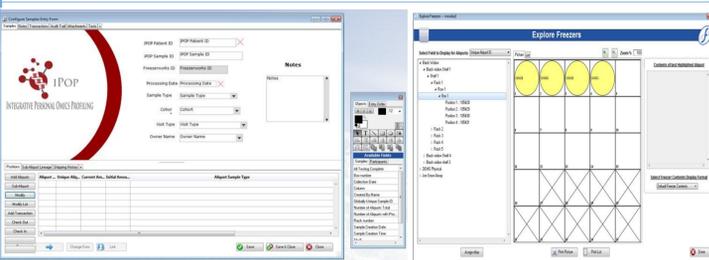


Figure 3. Some of the key features of sample management software that allowed us to configure sample entry forms with desired fields and to explore freezers for available samples.

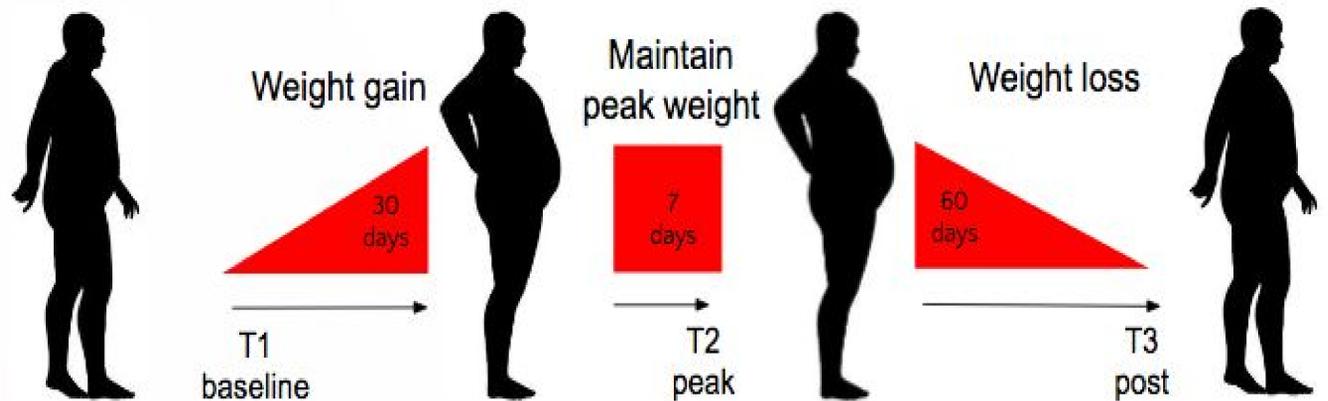


Figure 5. Above figure represents the IPOP overfeeding cohort, where, 24 participants have completed overfeeding through T3 stage. Blood and microbiome samples were collected at the indicated timepoints.

Sample and Data coordination

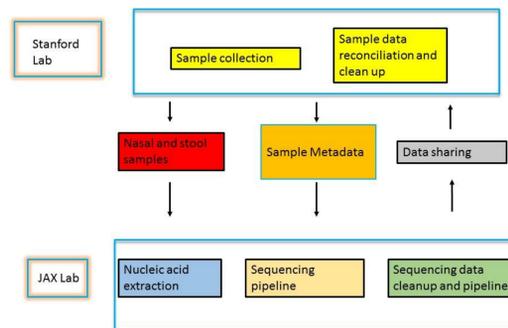


Figure 4. Above figure represents the sample and data flow between Stanford and Jackson (JAX) labs.

Data Management System

The data management platform needed to handle this process starts from the beginning with the enrollment and completes all the way through to final data analysis and sharing. Adding an additional level of complexity is the collaboration across different institutions IT services, (LDAP, oAuth, Firewalls etc.). We also needed to integrate the data from multiple omics for the integration. Similarly to typical systems, we designed the system to handle individual objects and relationship(s), but we needed the ability to integrate the different objects across longitudinal and/or multi-omic.

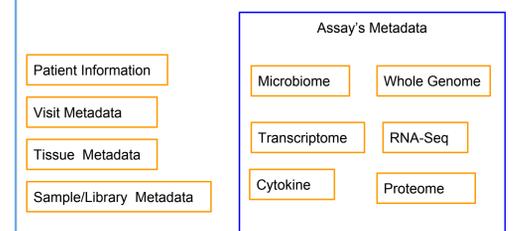
Collaboration Sites

Michael Snyder, Ph.D. Stanford University Palo Alto, CA
George Weinstock Ph.D. Jackson Laboratory for Genomic Medicine Farmington, CT

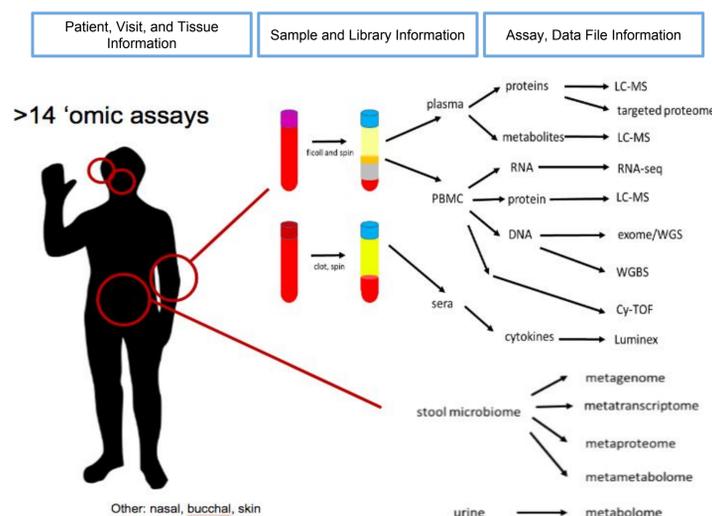


Sample/Data Tracking

Sample and data tracking of the samples, link to their associated data files, but also the data files can track back to the visit information and metadata.



Multi-omic Approach



Data Submission

Data submission is taking all the metadata and data files generated at each project site and transmitting that to the HMP2 DCC, where they will provide data governance and policy. Together all the teams have worked to establish a base set of objects and properties necessary to share the data with the scientific community.

