SINGLE-CELL SEQUENCING CORE DATABASE (SCSC)

Team J:

Carol Muriithi, Conor Shea, Garima Lohani and Robert LeBourdais

Boston University Single Cell Sequencing Core













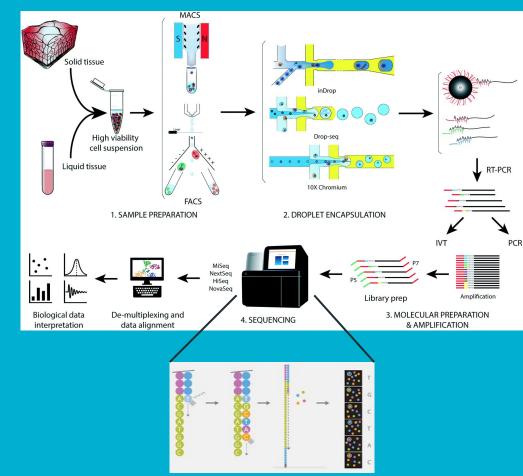




What is Single Cell Sequencing?

- 1. Tissue dissolution into single cells
- 2. (Optional) Flow Sorting
 - a. Enrichment for immune (CD45+), epithelial (EpCAM+) populations

- 3. Loading into wells (CEL-Seq2) or droplets (10x Chromium, Drop-Seq)
- 4. Library Preparation and Sequencing by synthesis
 - 3' v2, 3' v3, 5'
 - B/T Cell
 - CITE-Seq



5. Downstream Analysis

Data Sources

Administrative

- Contact Information
- Billing Information
- Service Requisitions

Library Quality

- Separate library index for multiplexing
- Quality metrics for the PCR
- Barcodes for primer and other sequencing kits

Sequencing Quality

- Produced as .csv files output by the sequencer, after running CellRanger
- Sequencing metrics vary per protocol
- Additional sequencing technologies produce different metrics

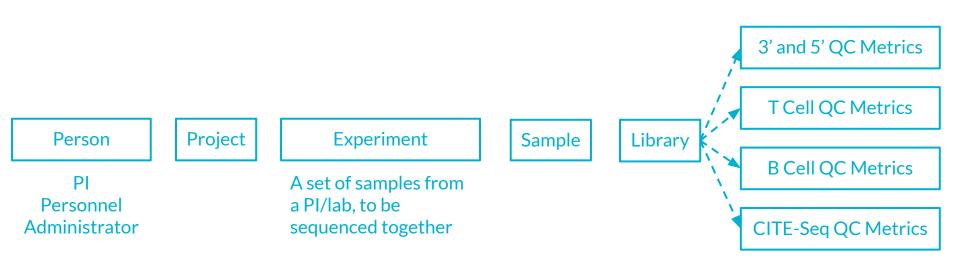
While the technology they use in their sequencing may be "next generation"...



Potential Issues

- Versioning, Limited (and at times incorrect) data validation
- Redundant information storage
- Inefficient data transfer
- Data fragmentation
- Data compilation and retrieval more time intensive

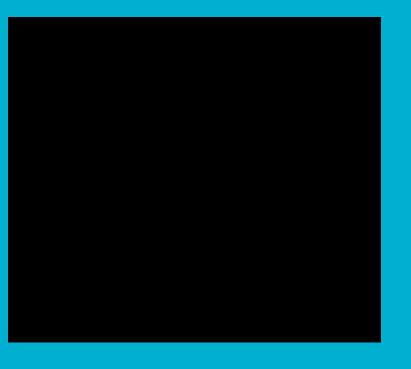
Database Design is Hierarchical



Key Tasks for the Database

- 1. Provide a robust, easy to use mechanism for data upload, with particular consideration paid to corner use cases (e.g. resequenced libraries, experiments split across multiple days)
- 2. Allow for rapid retrieval of important logistic and scientific data for administrative and analytic purposes

Database Home Page





Single Cell Sequencing Core Database

The Single Cell Sequencing Core (SCSC) performs single cell sequencing and preliminary bioinformatics analysis for Boston University and external principal investigators (PIs) on a fee-for-service basis. PIs contact the SCSC via a web form to request services, and then submit information about parameters for the sequencing analysis using this submission form.At this point, SCSC lab technicians perform the library preparation and sequencing on submitted samples. Lastly, depending on the requested services, SCSC computational analysis (may) perform initial bioinformatics analyses (e.g. quality control, cell clustering, and differential expression).Currently, the information in the submission form (Administrative Data), information from the lab technicians (Wet Lab Data), and information from the computational analysts (Dry Lab Data) are all stored in different spreadsheets (Google Sheets) and web forms. Information is sometimes copied and pasted into other spreadsheets to try and maintain a central database, but changes to these spreadsheets are not systematically checked for consistency. The aims for this project are therefore two-fold:

- 1. Compile administrative, wet lab, and dry lab data from previous jobs into a centralized database that allows for systematic query.
- 2. Provide an easy way for administrators, lab technicians, and computational analysts to input data on new jobs into the database.



Data Dashboard

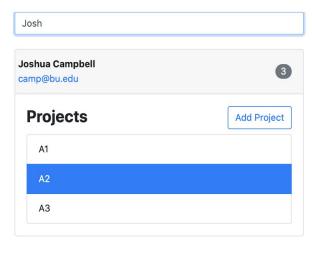
- Search for data
 - Search hierarchy
 - Client
 - Project
 - Experiment
 - Library
 - Filter by client name

 Upload and download library data

Dashboard

Use the Dashboard to view library data organized by principal investigator > project > experiment

Search projects by client



Project A2

Investigator Joshua Campbell

Experime	nt 2020-02-20	‡	New
Library		‡	Upload
Library	HUMAN_LUNG_3prin	ne	
Genome	Human		
Tissue	Lung		
Type	3'		
Type			

User Login

- Login with username and password
 - Secured by HTTPS
 - Password hash stored in Users database
 - Checks credentials and returns secure, temporary session ID
 - Session ID automatically times out
 - Session ID is sent with every server request

• Email-based password reset using python *smtplib*

No interface to register new users

Log In Dashboard Analytics About

Log In

Log into the SCSC data management system

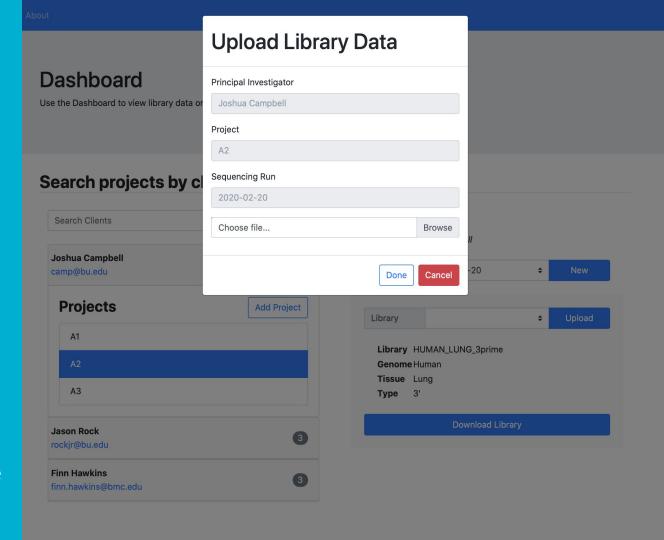
Email	
Email	
Password	
Password	
	Log In
	Forgot your password

Data Upload/ Download

 Upload CSV file output from sequencer

 Download all data for an experiment

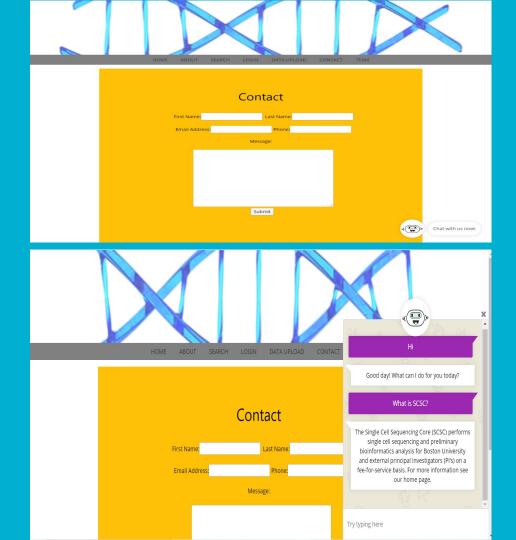
 Auto-detect library type from uploaded file



Contact

 Regular contact form with submit button

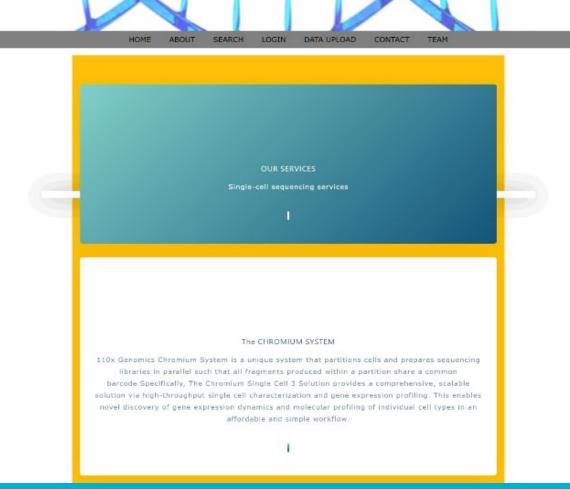
- AI (Artificial Intelligence)
 Chat Bot provides help to clients
- Messages response depend on input from the clients
- Uses machine learning to be trained



About

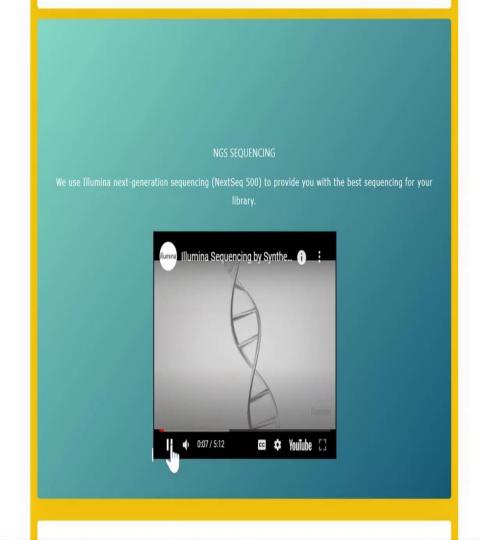
 One of our goals was to provide additional information available for services offered by the core

 The page briefly describes the 6 types of sequencing services and analysis provided by the core



About

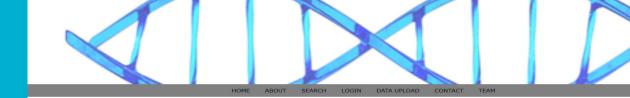
- Features used also include videos/ captivating pictures for each type of sequencing offered
- Our second aim
 was to provide a
 captivating UX
 experience for the
 users by the
 different color
 scheme/template
 while accessing
 information





Team page

- The team page provides a list of the core people in the team
- Images were embedded in the template
- Information describing the core members and their roles
- Easy access 'contact buttons' also available to put you into direct contact with core members



MEET THE TEAM



Yuriy Alekseyev, Ph.D.

Direct

Yuriy earned a ph. D. in Blochemistry from the Department of Chemistry at Wayne State University. He serves in the capacities of a Research Associate Professor, Pathology & Laboratory Medicine.

yurik@bu.edu 617-414-1369

E635

Contact



Joshua D. Campbell, Ph.D.

Josh earned a Ph.D in Bioinformatics at Boston University, He serves in the capacities of Assistant Professor, Division of Computational Biomedicine, Department of Medicine.

camp@bu.edu 617-414-6978

E604B

Contact



W. Evan Johnson, Ph.D.

Co Directo

Evan's earned a Ph.D in Biostatistics at Harvard University. He serves in the capacity of an Associate Professor of Medicine and Biostatistics. His research interest vary from applications in

genomic medicine to metagenomics

wej@bu.edu

617-638-2541

E609

Conta

Pros and Cons of the database

- The UI and UX are user friendly
- Easy search options via the dashboard
- Use of encrypted password for login features
- Pages are fast to load
- Additional client support by use of sophisticated machine learning trained chatbot with 24/7 access
- Easy to use data upload and download functionalities

- The database is currently only limited to internal users
- Limited mobile phone experience
- Data still being compiled by the SCSC Team
- No interface to register new users

CONCLUSION:

- We have created a user friendly, intuitive, and aesthetically pleasing website to facilitate easy data storage and retrieval
- Remaining Challenges:
 - We have made several choices in data storage that the team will likely override, as they consider best use practices for current (and future) services
- In the future, with a consent of the SCSC Management Team, we hope to integrate our website design with their client interface, to provide a unified location for all SCSC operations

QUESTIONS