Gladstone-CIRM Shared Resources Laboratory **Bioinformatics Training Program**

2025 Edition

CONTENTS

Overview

Program Offerings

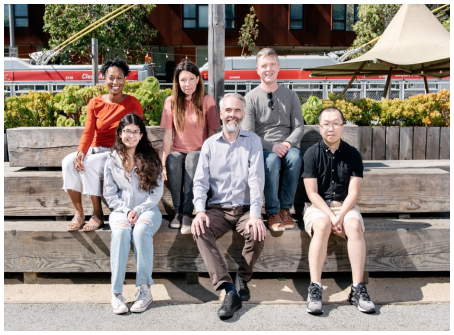
OVERVIEW

Introduction to the Curriculum

Gladstone Institutes is dedicated to training tomorrow's leaders. Gladstone's Bioinformatics Core provide rigorous scientific training, personalized attention and mentoring in a rich research environment.

Initiated in 2025 and in partnership with the California Institute for Regenerative Medicine (CIRM) the Gladstone Bioinformatics Core is now providing a revitalized program of bioinformatics workshops as part of the Gladstone-CIRM Shared Resources Laboratory. Each year we add new topics and update materials to reflect state-of-the-art tools, methodologies, and practices. Our goal is to offer bioinformatics, biostatistics, and data science training to a diversity of researchers, spanning introductory and advanced levels.

In the subsequent pages, you will learn about our team and our curriculum. You will find a summary of our performance in recent years, as well as a detail listing of each workshop at the end of this booklet.



TOP Gladstone Institutes, Mission Bay, San Francisco **BOTTOM** Members of the Gladstone Bioinformatics Core.

GLADSTONE INSTITUTES

CIRM CALIFORNIA'S STEM CELL AGENCY

Organizers

Alexander Pico, PhD

Alex is a Research Investigator and Director of the Bioinformatics Core at the Gladstone Institute of Data Science and Biotechnology. Since 2008, he has been producing training materials for students, postdocs, staff and faculty in a variety of data science areas, spanning programming, systems biology, and data visualization. His research program is coupled to outreach and training through his role as executive director for the National Resource for Network Biology since 2011 and participation as a mentoring organization in Google's Summer of Code since 2007. At Gladstone, Alex's team has developed a curriculum for empowering research scientists with practical bioinformatics, biostatistics, and programming skills.

"Our mission is to empower researchers with the necessary tools to analyze and interpret their hard- earned data."

Alexander Pico, PhD

Zainab Yusuf-Sada, BS

Zainab is the Executive Administrator for the Gladstone Institute of Data Science and Biotechnology. She received a BS in Public Health and Social Issues from SUNY Buffalo State. Zainab has many years of experience providing a high level of administrative support to executive directors. In addition to providing day to day support for the Core, Zainab is also a community group founder of Elevated Voices focused on DEI at Gladstone

"The Core has created an inclusive environment that inspires us to collaborate and lead."

Zainab Yusuf-Sada

PROGRAM OFFERINGS

Overview

The Bioinformatics Training program is a rich curriculum of workshops with didactic and hands-on components designed to help postdoctoral scholars, graduate students, and scientific staff with the skills and resources to design experiments, perform analyses, and communicate results. These are essential skills for our trainees in their careers ahead, though often neglected or unavailable earlier in their training. The curriculum is organized into eight topic areas, spanning novice, intermediate and advanced levels. Popular workshops are offered multiple times per year and altogether total over 80 hours of instruction.

Topic Areas

Statistics for Biologists

- *Experimental Design and Hypothesis Testing*
- *Statistics of Enrichment Analyses Methods*
- *Linear Mixed Effects Modeling*

Computer Skills for Data Science

- *Unix command line - I,II*
- *Working on Wynton*

Machine Learning

- *Introduction to Machine Learning*

Data Analysis Using R

- *Introduction to R for Data Analysis - I,II*
- *Intermediate R for Data Visualization*

RNA-Seq and ATAC-Seq Analysis

- *Introduction to RNA-seq Analysis - I,II*
- *Intermediate RNA-seq Analysis Using R*
- *Single cell RNA-Seq Analysis - I,II,III*
- *Introduction to scATAC-seq Dara Analysis*
- *scATAC-seq and scRNA-seq Data Integration*

Network Analysis and Visualization

- *Introduction to Cytoscape and Network Biology*
- *Intermediate Cytoscape for Omics Data Visualization*
- *Advanced Cytoscape Automation*

Pathways Analysis and Modeling

- *Introduction to Pathway Analysis*

Instructors

Ayushi Agrawal, MS is a bioinformatics analyst and programmer in the Bioinformatics Core of the Gladstone Institute of Data Science and Biotechnology. Her relevant expertise is in omics data analysis, including bulk and single cell RNA-seq, and single cell ATAC-seq, as well as R programming.

Natalie Elphick, MS a bioinformatics analyst and programmer in the Bioinformatics Core of the Gladstone Institute of Data Science and Biotechnology. Her relevant expertise is in omics data analysis, including whole-exome sequencing and bulk/single cell RNA-seq, and has expertise in R, Python, and Unix shell programming.

Kristina Hanspers, MSE is a senior research associate in the Bioinformatics Core of the Gladstone Institute of Data Science and Biotechnology. Her relevant expertise is in network and pathway analysis, and pathway modeling.

Min-Gyoung Shin, PhD is a bioinformatics analyst and programmer in the Bioinformatics Core of the Gladstone Institute of Data Science and Biotechnology. Her relevant expertise is in omics data analysis, including bulk and single cell RNA-seq, and single cell ATAC-seq, as well as machine learning.

Reuben Thomas, PhD is the associate director of the Bioinformatics Core of the Gladstone Institute of Data Science and Biotechnology. His relevant expertise is in omics data analysis, including bulk and single cell RNA-seq, and single cell ATAC-seq, as well as statistical analysis.

Michela Traglia, PhD is a bioinformatics analyst and statistician in the Bioinformatics Core of the Gladstone Institute of Data Science and Biotechnology. Her relevant expertise is in omics data analysis, including bulk and single cell RNA-seq, and single cell ATAC-seq, as well as statistical analysis.

Yihang Xin, MSE is a software engineer in the Bioinformatics Core of the Gladstone Institute of Data Science and Biotechnology. His relevant expertise is in network analysis and automation using R.

Performance Summary

In 2024 we registered a total of 2,674 researchers across 32 events, 84 on average per workshop. We collected survey responses for each event and tallied an overall average rank of 4.5 out of 5 for the workshops (see **Figure 1**).

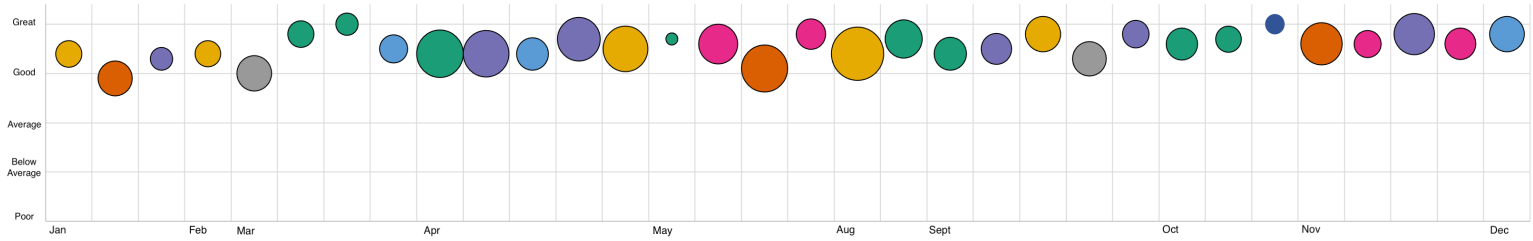
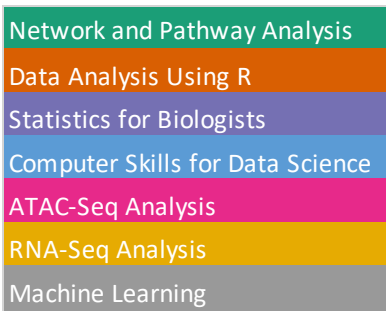


Figure 1. Bubble plot of training events spread throughout the year (x-axis) and the overall ranking presented as a descriptive word, i.e., Poor – Great (y-axis). The size of each event bubble is proportional to the number of attendees, with minimum and maximum labeled. The color of each event bubble corresponds to a topic area (see Legend).

Legend. Color code for topic areas



Workshop Details

Title	Description	Speaker	Duration	Topic Area
Experimental Design and Hypothesis Testing	Why do we perform experiments? What conclusions would we like to be able to draw from these experiments? Who are we trying to convince? How does the "magic" of statistics help us reach conclusions? As we attempt to make discoveries, we test new hypotheses using experimental data and hope that a skeptical and discerning person would start to believe the claims we make. Hypothesis testing, a branch of statistics, is a step we can rely on to bolster our claims. In this 2-day workshop you will gain a fundamental understanding of common hypothesis-testing concepts and terms, such as null hypothesis, alternative hypothesis, type I error, type II error, p-value, and power.	Reuben Thomas, PhD Michela Traglia, PhD	Two 2-hour sessions Twice a year	Statistics for Biologists *Prerequisites: None*
Statistics of Enrichment Analyses Methods	The use of gene set or pathway enrichment methods is ubiquitous for assigning functional or mechanistic significance to results from high-throughput assays like RNA-sequencing. This workshop will focus on the theory behind some of these methods and also on the interpretation of the associated results. Even though different methods will be illustrated using R code, the emphasis will be more on the understanding the background and assumptions of the methods. See our other workshops for performing enrichment using online tools and R packages.	Reuben Thomas, PhD Michela Traglia, PhD	Two 2-hour sessions Twice a year	Statistics for Biologists *Prerequisites: Basic stats, data analysis, and some R.*
Linear Mixed Effects Modeling	We commonly work on experiments involving repeated measures. Examples include responses measured in cellular assays where the repeated measures may occur over multiple experimental batches, multiple plates, multiple wells within a plate etc. Data derived from behavior of mice assayed over consecutive trials would be another example of experiments involving repeated measures. The distinctions between biological and technical replicates become less clear in these scenarios. Standard statistical tests like t-tests, ANOVA are not applicable. Instead it requires the use of more sophisticated approaches, such as Linear Mixed Effect Modeling. In this course, we will go over the motivation for the use of these models, the underlying description and assumptions behind them, example scenarios and code using R.	Reuben Thomas, PhD Michela Traglia, PhD	3 hours Twice a year	Statistics for Biologists *Prerequisites: None*

Workshop Details
(continued)

Title	Description	Speaker	Duration	Topic Area
Introduction to Unix command line - parts I and II	Knowing how to use the command line of your computer can save you a lot of time. In these workshops, you will learn how to navigate the Unix terminal, process files, set permissions, etc. The only thing you need is a Mac or Linux laptop (you can use Windows if you can't access a Linux machine). Part I (Introductory): ➢ What is the command line? ➢ Navigating the file system ➢ Installing software ➢ Files & permissions ➢ Downloading files from the internet ➢ Manipulating and searching text files Part II (Intermediate): ➢ File Compression/decompression ➢ System variables ➢ Shell scripting ➢ Other helpful commands ➢ AWK - advanced text manipulation	Natalie Elphick, MS	Two 2-hour sessions Once a year	Computer Skills for Data Science *Prerequisites:* *None*
Working on Wynton	Are you reaching the limits of your laptop? Are you new to working on clusters and high performance computing (HPC)? New to Wynton? This workshop will start with basic definitions for computing in a cluster environment, including types of nodes, queues and containers. We will cover how to transfer files, submit jobs and run analyses. Analyses can be simple scripts, pipelines or custom-built containers.	Natalie Elphick, MS Alex Pico, PhD	3 hours Twice a year	Computer Skills for Data Science *Prerequisites:* *UCSF account*
Introductory Machine Learning	Machine learning is becoming increasingly important not only in data science, but also in bioinformatics. Machine learning gives you the opportunity to find hidden patterns in your data and draw predictive conclusions. In this workshop, you'll learn key machine learning concepts and categories, and get hands-on practice with simple examples. See how machine learning can be applied to advance your research!	Min-Gyoung Shin, PhD	3 hours Twice a year	Machine Learning *Prerequisites:* *Experience with R*

Workshop Details
(continued)

Title	Description	Speaker	Duration	Topic Area
Introduction to R for Data Analysis - parts I and II	The scripting language R is considered one of the most powerful languages for quantitative analysis, statistics, and graphics. This workshop will help you get started using R to analyze your datasets and create graphs for visualization. Hands-on exercises will demystify data analysis using R.	Natalie Elphick, MS	Two 3-hour sessions Twice a year	Data Analysis Using R *Prerequisites: None*
Introduction to RNA-seq Analysis - parts I and II	Gene expression is central to cell biology. Disease pathways often involve changes in the expression levels of at least some genes. RNA-seq has become one of the most popular experimental methods to quantify gene expression levels. This 2-day hands-on workshop will provide an introduction to a typical bulk RNA-seq protocol and focus on the data analysis steps for recovering actionable insights.	Ayushi Agrawal, MS Michela Traglia, PhD	Two 3-hour sessions Twice a year	RNA-Seq Analysis *Prerequisites: None*
Intermediate RNA-seq Analysis Using R	RNA-Seq is a powerful tool to interrogate cellular functions. In this intermediate workshop you'll learn the skills you need to get the most out of your RNA-Seq data through analysis in R.	Michela Traglia, PhD Min-Gyoung Shin, PhD	3 hours Twice a year	RNA-Seq Analysis *Prerequisites: Intro RNA-seq*
Single cell RNA-Seq Analysis - parts I, II and III	This 3-session workshop will focus on the analysis of single-cell RNA-Seq data, starting from raw count matrices and progressing to network analysis of single-cell datasets. You'll learn about quality control of the raw data, normalization, feature selection, dimensionality reduction, clustering, finding marker genes, and batch correction. You'll analyze data using R. ➢ Day 1:Loading data, quality control, normalization, feature selection, dimensionality reduction ➢ Day 2: Dimensionality reduction (continued), clustering, finding marker genes ➢ Day 3: Advanced discussion on normalization, differential analysis, and batch-correction	Ayushi Agrawal, MS Michela Traglia, PhD Min-Gyoung Shin, PhD Reuben Thomas, PhD	Three 2-hour sessions Twice a year	RNA-Seq Analysis *Prerequisites: Intro and intermediate RNA-seq*

Workshop Details
(continued)

Title	Description	Speaker	Duration	Topic Area
Introduction to scATAC-seq data analysis	This 2-session hands-on workshop will provide an introduction to a typical single-cell ATAC-seq protocol and focus on the data analysis steps to address unique computational challenges before the downstream analysis. You'll learn how to perform quality checks of raw data in the typical format provided by sequencing centers, how to process the data, perform dimensionality reduction for visualization, clustering and cell type annotation, generate a gene score matrix to identify cell types, and perform motif and peak enrichment using ArchR.	Ayushi Agrawal, MS Michela Traglia, PhD Reuben Thomas, PhD	Two 2-hour sessions Twice a year	RNA-Seq Analysis *Prerequisites:* *Intro and intermediate RNA-seq; scRNA-seq recommended*
scATAC-seq and scRNA-seq data integration	This hands-on workshop will provide an introduction to integrating single-cell ATAC-seq data and single-cell RNA-seq data. Required experience: ➢ Prior attendance to scRNA-seq, bulk RNA-seq workshops or experience with analysis of transcriptomics data ➢ Prior attendance at the Introduction to scATAC-seq data analysis workshop, or experience with analysis of scATAC-seq data ➢ Familiarity with R and RStudio (e.g., reading in files, working with lists and dataframes)	Ayushi Agrawal, MS Michela Traglia, PhD Reuben Thomas, PhD	One 2-hour sessions Twice a year	RNA-Seq Analysis *Prerequisites:* *Intro and intermediate RNA-seq; scRNA-seq recommended*

Workshop Details
(continued)

Title	Description	Speaker	Duration	Topic Area
Introduction to Cytoscape and Network Biology	Join this course for an introduction to the basic concepts of biological network analysis and practical instruction on commonly used tools and databases, with a focus on Cytoscape to analyze and visualize biological networks. The course consists of theoretical and practical sessions in which you'll learn how to perform different biological network analyses and how to visualize and interpret the results of your analyses.	Alex Pico, PhD Scooter Morris, PhD Yihang Xin, MSE	4 hours Twice a year	Network Analysis and Visualization *Prerequisites: None*
Intermediate Cytoscape for Omics Data Visualization	This course is a complement to "Introduction to Network Biology and Cytoscape." Following a brief review of the key concepts of network analysis, you'll embark on a deep dive into data visualization and advanced Cytoscape features. Work through three prepared use cases demonstrating various omics data types and strategies. You're also encouraged to bring your own data.	Alex Pico, PhD Scooter Morris, PhD Yihang Xin, MSE	3 hours Twice a year	Network Analysis and Visualization *Prerequisites: Intro to Cytoscape*
Advanced Cytoscape Automation	In this course, you'll learn about the integration of Cytoscape into programmatic workflows and pipelines using Ror Python. Following a brief review of the goals and capabilities of Cytoscape in network analysis, you'll learn how to translate these into scriptable use cases for practical applications. The bulk of the course will be a hands-on demonstration of how to access and control Cytoscape from R/Python to perform a network analysis of tumor expression and variant data. These principles can be extended to other data types and research questions, as well as to other automation environments.	Alex Pico, PhD Yihang Xin, MSE	3 hours Twice a year	Network Analysis and Visualization *Prerequisites: Intro and intermediate Cytoscape, and experience with R or Python*
Introduction to Pathway Analysis	Join this introduction to functional enrichment analysis and effective visualization. You'll learn about different enrichment analysis tools and methods, and how to use them. You'll also learn how to visualize data in pathways and using GO terms resulting from enrichment analysis.	Kristina Hanspers, MSE	3 hours Twice a year	Pathways Analysis and Modeling *Prerequisites: None*

Gladstone Institutes is an independent, not-for-profit research organization affiliated with the University of California San Francisco (UCSF) and a renowned center of excellence focused on applying basic science to disease. Our science contributes to the health and wellbeing of all people through research, education, and outreach in the areas of cardiovascular disease, neurological disease, virology, immunology, and data science.

Gladstone has a vibrant group of labs and core facilities focused on emerging and highly innovative technologies for biomedical research, including stem cells, imaging, single-cell genomics, mass spectrometry, and genome editing.

The Bioinformatics Core provides expertise on experimental design and data analysis of complex data sets, with a specialization on analysis of large-scale data sets acquired from cutting-edge technologies. The core is involved in experimental design, power analysis, sequence analysis, annotation, pathway and network analysis, visualization, data integration, figure generation, and technical writing, in addition to training and workshops on topics in bioinformatics, statistics, and data science.

GLADSTONE
INSTITUTES