PULSE

Postdoc Upskilling in Laboratory & Specialized Expertise

A Professional Development Initiative for Postdoctoral Scholars at UNC Core Facilities

Title of Proposed Project

Multi-Omics Workflow Integration and Preliminary Data Harmonization for the Carolina Endometrial Cancer Study

Name of Core Facility

Lineberger Bioinformatics Core Facility

Core Director Name *Official contact for the program*

David Corcoran

Project Description

This PULSE project will support the development and application of reproducible multi-omics analysis workflows for the Carolina Endometrial Cancer Study, a large population-based investigation that will generate RNA sequencing, microbiome profiling, DNA methylation, variant calls, and copy number variation data across approximately 800 tumor samples. The overall goal of this short-term collaboration is to integrate existing analysis workflows across these data types in order to produce standardized, analysis-ready outputs and to perform early integrative analyses that will enable downstream biological and clinical discovery.

The project is designed to provide focused, hands-on exposure to large-scale cancer genomics within a core facility environment. The participating postdoctoral scholar will work closely with the Lineberger Bioinformatics Core to deploy and test established workflows for each data modality in a secure high-performance computing environment. Emphasis will be placed on workflow execution, data quality assessment, organization of derived data products, and harmonization of outputs across platforms so that results can be readily used by downstream investigators.

A core component of the project will be ensuring that each data type is processed in a consistent, reproducible, and well-documented manner. This includes evaluating data quality metrics, verifying pipeline outputs, organizing files and metadata, and contributing to standardized data structures that facilitate cross-modal integration. As time allows, the postdoc will assist with exploratory multi-omics analyses that examine relationships across transcriptomic, epigenomic, microbiome, and genomic variation data, with the intent of generating early hypotheses for future grant-funded and collaborative research. This project will provide practical exposure to population-scale cancer genomics, reproducible computational research practices, and team-based data science in a regulated research environment. The work will directly support a high-impact translational cancer study while allowing the postdoctoral scholar to broaden computational, organizational, and collaborative research skills.

Expected Skill Development

Hands-on experience harmonizing and standardizing multi-omics data generated from established analysis workflows in a secure high-performance computing environment using bash, Nextflow, R and/or Python. Practical skills in cross-platform quality control, organization of analysis-ready datasets, and consistent formatting of outputs across RNA-seq, methylation, microbiome, and genomic variation data. The project will strengthen their abilities in reproducible research practices, metadata management, and critical evaluation of high-dimensional molecular data. They will also gain experience analyzing and testing formal hypotheses in one or more of these data types in a large population-based cancer study

Scheduling Considerations

The work can be done remotely. The candidate should have a half day a week set aside for the project with an additional 4 hours being more flexible as to how it is split up. Our group meeting is 3-4pm on Tuesdays via Zoom

Estimated Time Commitment

8 hours/week

Expected Project Duration

12 weeks

Designated Mentor(s)

David Corcoran, Core Facility Director. Oversee all aspects of the project
Jennifer Modliszewski, Senior Bioinformatics Scientist. Assist with advancing data processing & analysis knowledge.
Alan Hoyle, Senior Bioinformatics Scientist. Assist with advancing computer science and reproducible workflow knowledge.

How often will the mentor meet with the participating postdoctoral scholar?

Weekly check-ins
As needed / flexible

Required Background or Skills

The ideal candidate will have prior experience working in a Linux-based computing environment and experience using R and/or Python for data analysis. Candidates with a strong interest in computational biology, bioinformatics, or data-driven biomedical research and a willingness to self-learn new technical skills prior to and during the project will be strongly considered.

Onboarding Plan

A 2-hour introductory joint meeting with the 3 designated mentors to get familiar with the core and its computational infrastructure

Training Plan

Depending on the skillset of the individual, up to 2 of the 8 hours per week may be spent taking online courses to advance their analytic or computational skillset. The individual will work closely with both the workflow development team and data analysts that can provide a more guided learning experience

Potential Deliverables/Outcomes

Method development
Contribution to manuscript
Skill Development