RUTGERS HEALTH
School of Health Professions

SHP Student Interns for Research and Scholarly Activities Project Proposal Form

Instructions:

Please fill each box to the right of the required fields, obtain the required signature and return via email to Michele Sisco (mcoral@shp.rutgers.edu) by **March 24, 2025**.

If you are sending attachments, please ensure your contact information is added to all your forms.

Faculty Contact Information:

Date submitted: 3/24/2025

Faculty Name:
Frederick Coffman and Antonina Mitrofanova
Department/Program:
Health Informatics
Telephone number:
973-972-8190
E-mail:
coffmafd@shp.rutgers.edu
amitrofa@shp.rutgers.edu

Project Detail:

Project Title: (56 characters max)	**Machine Learning for predictive modeling of cancer progression using MIR99AHG expression**
Hypothesis:	We hypothesize that MIR99AHG, a long non-coding RNA (lncRNA) implicated in multiple cancers, including acute myeloid leukemia (AML), breast cancer, and lung cancer, could serve as a biomarker of cancer progression. This proposal aims to integrate MIR99AHG expression data into machine learning models for cancer progression classification and survival prediction.
Description: (Include design, methodology, data collection, techniques, data analysis to be employed, evaluation and interpretation methodology for research component)	MIR99AHG is a long non-coding RNA (lncRNA) implicated in multiple cancers, including acute myeloid leukemia (AML), breast cancer, and lung cancer. Understanding its role in cancer progression through computational predictive modeling can open new avenues to novel therapeutic targeting. In this project we will:

1. Extract and preprocess MIR99AHG expression data from publicly available cancer databases.
2. Develop a machine learning framework to assess its predictive power in cancer progression classification.
3. Interpret the biological significance of MIR99AHG using GSEA and other pathway enrichment analysis techniques on phenotypes associated with MIR99AHG overexpression.

Methodology

1. Data Collection and Preprocessing

- **Datasets**: RNA sequencing data will be obtained from dbGap and Gene Expression Omnibus (GEO)
- **Processing**: Data will be mapped to hg38 and normalized using DESeq

2. Machine Learning

- **Model Selection:** Train classifiers such as Random Forest, XGBoost, and Support Vector Machines (SVM) for cancer progression prediction.

3. Evaluation Metrics

- Cox proportional hazards model, Kaplan-Meier analysis , concordance index (C-index), and ROC_AUC will be used to evaluate predictive accuracy.

This is a collaborative project utilizing two HI faculty members – Dr. Frederick Coffman who is an expert in cancer biology and has published papers on regulatory RNAs in cancer, and Dr. Antonina Mitrofanova, who is an expert in cancer informatics and data science.

Specific Student Responsibilities: Student will be responsible for data download and processing, computational analysis and data interpretation.

Start / end date of project:
5/30/25 – 9/1/25

WHAT OTHER EDUCATIONAL OPPORTUNITIES ARE AVAILABLE TO STUDENTS? (e.g., journal club, seminars, clinic, rounds)	Lab meetings, Hot topics seminar series, student will be encouraged to submit abstract to the Annual Student Research and Scholarship Symposium

WHERE DO YOU PLAN TO PRESENT OR PUBLISH THE FINDINGS WITH THE STUDENT? (e.g., national or state meetings, newsletter or journal, SHP poster day)	SHP Annual Student Research and Scholarship Symposium

CHECK ALL APPROPRIATE BOXES BELOW AND PROVIDE REQUESTED INFORMATION.

This project is: ☐ clinical ☐ laboratory ☐ behavioral ☐ survey ☐ educational

☒ Other: please specify______Computational_______________________________

☐ This project involves the use of human subjects (including chart review, retrospective studies and questionnaires).

Pending ☐ Approved ☐ IRB Protocol Number ________________

IRB approval must be obtained by June 2025

______________________________________ _3/24/2025_______

Signature of Department Chair Date