Specific Research Projects Available to REU Students

The 13 primary mentors on the proposed REU Site are listed in the table below. The research interests of these faculty span scales ranging from molecular-scale models (Bourne & Janes) to pathways and cells (Allen, Civelek, Zhang, Fallahi-Sichani, Naegle, & Papin) all the way up to tissues, organs, and populations (Dolatshahi, Peirce-Cottler, Blemker, Gerling, & Barnes). Examples of specific REU independent research projects and student roles for each are listed below. Refer to §D for a description of the student-mentor matching process.

Faculty Participant	**Department/Division**	**Research Focus**
Timothy Allen	Biomedical Engineering (BME)	protein translation, virus evolution
Laura Barnes	Engr. Systems & Environment	health informatics, machine learning
Silvia Blemker	BME / Mechanical Engr.	muscle mechanics & remodeling
Philip Bourne	School of Data Science (SDS) / BME	biomedical data sciences, protein struct.
Mete Civelek	BME / Public Health Genomics	biomed. data sci., diabetes, heart dis.
Sepideh Dolatshahi	BME	systems immunology, cancer
Mohammad Fallahi-S.	BME	data-driven modeling, cancer
Greg Gerling	Systems & Information Engr.	computational neuroscience, design
Kevin Janes	BME	cellular signaling, cancer
Kristen Naegle	BME / Public Health Genomics	proteomics, cell signaling, ML, cancer
Jason Papin	BME	metabolism, infectious disease, cancer
Shayn Peirce-Cottler	BME	tissue patterning, angiogenesis
Aidong Zhang	Computer Science / BME / SDS	bioinformatics, data sci., mach. learning

Molecular-Scale Projects:

Virtual cell modeling of chromosomal separation mitosis: Centromeres create important landing sites for microtubules during mitosis, but the location of centromeres differs among organisms. Humans have “acrocentric” chromosomes, where the centromeres are located within the middle of the chromosome. Rodents, by contrast, have “telocentric” chromosomes, where the centromeres are located at the end of each chromosome. Our lab has built a mathematical model in the open-source software VCell for an important signaling pathway related to chromosome segregation during mitosis. The model is currently constructed for acrocentric chromosomes, and we have not evaluated how results change when telocentric chromosomes are modeled instead. For this project, after being trained in VCell for the first two weeks, the REU student will modify the existing VCell code to reposition the centromere at arbitrary locations along the chromosome axis. The student will run simulations at various locations and evaluate how results change when the centromere is displaced in consultation with faculty, postdoctoral, and graduate-student mentors. This work will critically evaluate the robustness of the working model and address broader questions about the evolution of centromere positioning across organisms. *(Mentor: KJ)*

Using Explainable AI to Decipher an Urfold model of Protein Structure: Proteins are the molecules of life, and their structures can be viewed as curves that trace through three-dimensional space. How can we compare complex curves that look vaguely similar, while allowing for various degrees of variability? What can classifying the curves teach us about the nature and relationships of protein shapes, across the entire protein universe? (Are there implications for molecular evolution?) Decades of research have studied these questions, yet our picture of the protein universe remains murky. Our lab seeks ways to use data science to change this. We recently proposed a new concept, termed the ‘Urfold’, that describes a notion of “architectural identity despite topological variability”. We are now placing this model on a quantitative footing via deep learning, as part of a new ‘DeepUrfold’ framework for flexible, highly scalable structural analysis. While generative models of protein structure, like DeepUrfold, offer powerful ways to map protein relationships, they ignore the underlying causal mechanisms that yield one classification versus another. The REU student will first be trained in the use of explainable AI methods, such as layer-wise relevance propagation (LRP), and learn to sue these to elucidate the classification patterns that we find in highly-populated superfamilies. Armed with pre-existing DeepUrfold models, and aided by a senior scientist, the student will then systematically apply different LRP approaches and formulations (known as ‘rules’) to the models in order to probe their veracity against known protein sequence/structure/function relationships in respective superfamilies (e.g., a small β-barrel urfold). Students will also gain familiarity with the relevant literature, including journal club-style presentations. *(Mentor: PEB)*

Genomics, Molecular Pathways, and Cellular-Scale Projects:

Metabolite biomarkers of hepatocyte drug toxicity: This project will leverage existing models of rat and human hepatocyte metabolic networks and expertise in the Papin lab with high-throughput data integration. The REU student will be trained to make use of established metabolic network analysis tools and how to use the existing metabolic network models. The student will then mine databases for gene expression data that captures the response of hepatocytes to pharmaceutical compounds and will make use of existing computer toolboxes to integrate such data into the rat and/or human hepatocyte metabolic networks to predict metabolites that serve as biomarkers of drug toxicity. The student will then work with a graduate student in the Papin lab to experimentally test the presence of the metabolite in corresponding cell culture models of drug exposure. *(Mentor: JP)*

High-throughput data-driven identification of epigenetic targets that regulate melanoma heterogeneity: Through 10 weeks of mentored research, the REU student will learn and apply quantitative approaches as well as tools of data science for visualizing and analyzing the multiplexed, single-cell data generated through high-throughput fluorescence microscopy experiments performed in the Fallahi-Sichani lab. Through this mentored research experience, the student will be first provided with appropriate hands-on training to develop technical skills in data-driven approaches such as high-content fluorescence microscopy image analysis, data visualization via clustering analysis, and dimensionality reduction using linear methods such as Principal Component Analysis (PCA) and nonlinear methods such as Uniform Manifold Approximation and Projection (UMAP). The student will then apply these techniques to the analysis of single-cell imaging data, exploring the impact of a panel of epigenetic-modifying compounds tested, either alone or in combination with inhibitors of MAPK signaling, on populations of melanoma cells in which differentiation state markers have been measured by fluorescence microscopy. The scientific objective of this summer project is for the student to generate hypothesis about potential epigenetic targets that regulate the heterogeneity of differentiation state in populations of melanoma cells, through the application of data-driven, computational systems biology approaches. *(Mentor: MF-S)*

Systems genetics of gene expression in different sexes: The project goal is to identify sex-specific regulatory gene expression networks that function in subcutaneous and visceral adipose tissue of males and females. The REU student will use gene expression data from nearly 2,000 people who are part of the Genotype-Tissue Expression (GTEx), STARNET, and African American Genetics of Metabolism and Expression cohorts. The student will closely work with a postdoctoral fellow and Dr. Civelek to first familiarize himself/herself with gene expression and genotype datasets in the first week of the program. In the second week, he/she will learn how to perform quality control of the gene expression and genotype datasets to identify outlier samples using pipelines that are already established in the mentor's laboratory. During this time the student will also complete tutorials to learn how to access and use the high-performance computational cluster at UVA before spending the next four weeks performing linear regression analysis between human genotypes and the gene expression values. Lastly the student will compare the identified genetic variants to those which have been associated with metabolic phenotypes that have a distinct pathology in men and women. Throughout the 10-week period, the student will attend the weekly undergraduate meeting and present their results. The student will also read the literature to learn more about adipose physiology and pathology as well as the literature on genome-wide association analysis of metabolic phenotypes. *(Mentor: MC)*

Tissue-, Organ-, and Population-Level Projects:

Statistical Approaches to Understand miRNA Regulation of Glycosylation Genes in Cancer Cell Lines: The REU student will first be trained from the ground up in basic approaches in statistical data analysis. After this initial training, the student will apply these analytical approaches to understand how glycosylation genes are regulated in different types of cancer. Glycosylation is a post-translational modification that involves the addition of oligosaccharides, or groups of sugars known as glycans, to proteins. Perturbations of glycosylation, such as modifications of individual sugars or complexity to glycan structures, are linked to cancer progression and aberrant cancer behavior. Potential regulators of glycosylation are miRNAs, which are generally known to target and inhibit the expression of specific genes. Many miRNAs have been found to be downregulated in cancer, which lead to the upregulation of target genes, resulting in cancer progression. The REU student will determine new miRNA-mRNA pairs in cell lines from the Cancer Cell Line Encyclopedia (CCLE). Throughout the project, the student will read selected papers in cancer biology, cancer modeling, and the role of glycosylation in cancer progression. *(Mentor: SD)*

Agent-based modeling (ABM) to design cell-based therapy: The REU student will adapt an ABM of angiogenesis in the retina to design a cell-based therapy for improving the vasculopathy that ensues during diabetic retinopathy. In the first two weeks, the student will learn about the principles of ABM and construct a simple demonstration ABM in the user-friendly software NetLogo (software that was originally designed for high school students but is actively used in our research). For the third week, the student will shadow a retinal surgeon in the clinic as he cares for patients with diabetic retinopathy and a graduate student in the lab to observe vascular changes in mouse models of retinopathy. Additionally, he/she will read selected papers on the stem cell therapies that have been developed by the mentor and other groups to control angiogenesis. Finally, the student will define new rules for the ABM and run new simulations to evaluate different cell-based therapies (cell type, dose, injection route, etc.) in the revised model. Lastly, the student will compare the predictions to data from mouse model experiments conducted in the mentor's laboratory. *(Mentor: SPC)*

Patient-specific modeling of skeletal muscle disease: The REU student will create a finite-element (FE) model of skeletal muscle based on magnetic resonance imaging (MRI) data of a child with Duchenne muscular dystrophy to investigate the relationship between internal stresses within the muscle and extent of damage observed from the MRI data. For the first two weeks, the student be mentored by a graduate student to learn how to use an established pipeline (within the mentor's lab) for creating FE models from MRI data. The next four weeks will be spent creating a model and running simulations to predict the stress and strain distributions within the muscle during various types of contractions. The student will then analyze the extent and distribution of intramuscular inflammation within the same subject's muscle using a set of image-processing tools developed in the mentor's lab. The student will then correlate the predictions from the FE model with the MRI analysis of inflammation to determine the relationship between mechanical forces and damage in muscles in boys with DMD. Throughout the summer, the student will shadow clinicians at UVA's muscular dystrophy clinic once a month and shadow a post-doc in the lab who is conducting a clinical study on boys with DMD. The student will also read existing literature on modeling muscle physiology, mechanics, and pathophysiology, giving a full presentation of the study to the mentor's lab at the end of the summer. *(Mentor: SSB)*

Biomimetic neural spiking models to efficiently encode force: Man-made force sensors, such as force sensing resistors, are employed at the distal ends of robot arms and neural controlled prosthetics. The typical output of force sensing resistors (FSRs) is analog voltage. While the features of the natural environment that are grasped and palpated with an array of such sensors can be decoded, the analog voltage data streams quickly become computationally intractable. For this project, the REU student will utilize an already setup array of 16 FSRs in an electronics circuit with amplifiers and an Arduino microprocessor. The student will use an existing algorithm on the Arduino, that of the leaky integrate-and-fire neuron, to demonstrate the biologically inspired encoding of instantaneous force into the timing of neural spikes. This method mimics tactile mechanosensitive afferents. Given this platform, she or he will select and set up a series of interaction tasks, such as differentiating into compliant surfaces and curvatures, to demonstrate that changes in force magnitude and rate are encoded by the population of interspike intervals. In forming a new contribution, she or he will determine how to best detect features from within the population of neural spikes, for example the relative latencies of first spikes over a population, and what sort of savings in terms of computational time and bandwidth could be expected as opposed to decoding analog voltage. The REU student's findings and contributions will be validated weekly with the faculty mentor and a graduate student knowledgeable in this area. *(Mentor: GJG)*

Upper EXTremity Examination for Neuromuscular Diseases (U-EXTEND): Objectively Assessing Treatment Efficacy and Accelerating Drug Discovery: The REU student will gain an understanding of the collection, cleaning, and mining of multiscale wearable sensor data for assessing neuromuscular disorders. With emerging therapies in neuromuscular disorders, monitoring progress is vital. For example, Spinal Muscular Atrophy (SMA) is the leading genetic cause of death in infants with a life expectancy of <2 years, which, until recently, had no treatment. However, the development of Nusinersen and other forms of treatment has led to a historic moment in the care of patients with SMA which cannot be evaluated with current assessment tools. Similarly, new treatments are being developed for Duchenne Muscular Dystrophy (DMD), the most common muscular dystrophy, and the most common genetic cause of death in boys. These treatments have been slow to progress through clinical trials and to reach patients, in part due to

lack of sensitive and specific markers of improvement which can be used to monitor effectiveness. As with any medical treatment, the outcomes vary from patient to patient. First, the REU student will assist in the data collection procedure and training on software and data management for handling sensitive medical data. The student will work with a graduate student within the mentor's lab to learn the different sources to analyze the relationship between patient-generated data and multiple clinical outcome measures to assess efficacy of treatments for neuromuscular disorders. *(Mentor: LB)*