<table>
<thead>
<tr>
<th>Research Area</th>
<th>Project Mentors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Host-microbe interactions</td>
<td>Aldridge</td>
</tr>
<tr>
<td></td>
<td>Kumamoto</td>
</tr>
<tr>
<td></td>
<td>Isberg</td>
</tr>
<tr>
<td></td>
<td>Lee</td>
</tr>
<tr>
<td>Biomolecular engineering</td>
<td>Jiang</td>
</tr>
<tr>
<td></td>
<td>Van Deventer</td>
</tr>
<tr>
<td>Computational modeling and design</td>
<td>Hassoun and Nair</td>
</tr>
<tr>
<td></td>
<td>Tzanakakis</td>
</tr>
<tr>
<td>Image processing</td>
<td>Miller</td>
</tr>
</tbody>
</table>
Aldridge Lab
Tuberculosis is caused by infection with *Mycobacterium tuberculosis*, and treatment of TB remains lengthy because some bacteria are tolerant of antibiotic treatment. Mycobacteria grow asymmetrically, giving rise to differences in growth behaviors and antibiotic tolerances. We aim to characterize what is different about mycobacteria that are drug tolerant from those that are killed quickly by antibiotics with the long-term goal of optimizing treatment strategies. Our work combines experimental and computational approaches, and each student project engages both types of work. We use live-cell microscopy, microfluidics, and fluorescent reporter strains of mycobacteria to observe cell growth and drug response. We quantify the contribution of each cellular factor to drug response by building and analyzing computational models.

Photo: Kelvin Ma/Tufts University
Our research focuses on interactions between the opportunistic fungal pathogen, *Candida albicans*, and its environment. This organism is a common commensal of the human GI tract and, in this environment, interacts with the bacterial microbiota, components of the host immune system and the metabolites that make up the intestinal milieu. While colonization by *C. albicans* in a healthy host is benign, organisms colonizing the GI tract represent the reservoir from which *C. albicans* may spread to different parts of the body. In immunocompromised patients, spread by *C. albicans* can result in fatal candidemia. Higher levels of GI colonization are associated with higher rates of candidemia and, thus, the factors that influence levels of *C. albicans* colonization are important to elucidate. To understand how *C. albicans* influences the gut environment, our research employs a murine model. Mice are not naturally colonized by *C. albicans* and we can therefore study the effects of this organism by comparing mice that have been experimentally colonized with *C. albicans* and mice that have not. We and others have demonstrated that colonization with *C. albicans* results in changes in the composition of the bacterial microbiota. We are conducting further studies to understand the biological significance of the changes that have been detected. As part of this project, students will use the murine model to obtain samples, extract DNA and characterize the bacterial groups that are present, and study the composition of the bacterial microbiota using data obtained through Illumina sequencing. Standard data analysis using established methods will be conducted. There will also be opportunities to develop novel data analysis methods specific for the needs of this project.
Isberg Lab
The Isberg laboratory focuses on determining how intracellular bacteria hijack host cell secretory processes to allow formation of a replication compartment. The model system that will be studied by the REU students is the growth of the bacterium \textit{Legionella pneumophila} within either \textit{Drosophila} cells or the amoeba \textit{Dictyostelium discoideum}. Formation of the replication compartments in these cells requires a specialized bacterial protein translocation system that moves at least 275 different proteins into host cells. Although the identity of these proteins is well known, one of the most persistent problems in their analysis is that it is extremely difficult to identify the minimum protein set necessary for intracellular growth. To solve this problem, we have developed a multiplex system called iMAD, in which the phenotypes of more than $10^4$ bacterial insertions can be followed individually in host cells depleted for individual proteins. In our current version of this strategy, the relative fitness of each mutant is determined by high-density sequencing, using abundance of reads as a measure of fitness for each insertion mutant. The data are then analyzed by hierarchical clustering, with each phenotypic cluster of genes assigned to a single functional group. We have shown that combining mutations from different functional groups allows us to construct strains that are defective for intracellular growth, allowing a partial solution to genetic redundancy. Remarkably, proteins that define a single hierarchical cluster appear to perform identical functions in support of intracellular growth. The REU student will take advantage of this strategy, and will first perform bioinformatics analysis of iMAD datasets that they will be provided from data already accumulated in the laboratory. Based on the pathway predictions that the student makes from these large datasets of up to 400,000 phenotypic characterizations, the student will identify pairs of mutants expected to give defects in intracellular growth. The student will then use these predictions to construct double mutants and test the hypothesis that simultaneous loss of these two proteins will result in defective intracellular growth in \textit{Drosophila} cells and \textit{D. discoideum}. 

\begin{center}
\begin{tikzpicture}
  \node {\textit{legA3}};
  \node[below of=legA3] {\textit{lidA}};
  \node[below of=legA3, yshift=-2em] {\textit{sdhA}};
  \node[below of=legA3, yshift=-2em] {\textit{wipB}};
  \node[below of=legA3, yshift=-2em] {\textit{lpg2395}};
  \node[below of=legA3, yshift=-2em] {\textit{vpdB}};
  \node[below of=legA3, yshift=-2em] {\textit{mavP}};
  \node[below of=legA3, yshift=-2em] {\textit{lpg1924}};
  \node[below of=legA3, yshift=-2em] {\textit{lpg1098}};
  \node[below of=legA3, yshift=-2em] {\textit{sidE}};
  \node[below of=legA3, yshift=-2em] {\textit{lpg2182}};
  \node[below of=legA3, yshift=-2em] {\textit{legL3}};
  \node[below of=legA3, yshift=-2em] {\textit{lpg1527}};
  \node[below of=legA3, yshift=-2em] {\textit{ceg8}};
  \node[below of=legA3, yshift=-2em] {\textit{lpg2538}};
  \node[below of=legA3, yshift=-2em] {\textit{legC8}};
  \node[below of=legA3, yshift=-2em] {\textit{lePA}};
\end{tikzpicture}
\end{center}
The adult human gastrointestinal (GI) tract harbors trillions of microorganisms belonging to at least several hundred species. This gut microbiota impacts an array of physiological functions and developmental processes in the host animal. Well-known functions include digestion and immune response to foodborne pathogens. Increasingly, it is becoming apparent that the microbiota plays a critical role in the host’s health. Significant and sustained modification of the microbiota, or dysbiosis, correlates with not only GI diseases, but also other chronic diseases such as diabetes, cancer, and asthma. However, many questions regarding the mechanisms whereby the microbiota impact host physiology remain unanswered. For example, while there is general agreement that microbiota species produce metabolites and other biomolecules that mediate the interactions between the microbiota and host, only a handful of such molecules have been identified.

The goal of our research is to elucidate the physiological roles of metabolites produced by the gut microbiota, and link these molecules to the species responsible for their production. In this way, we seek to develop rationally designed pre-, pro-, and post-biotics beneficial for human health.

The student participating in the summer REU project will investigate how biological and synthetic antimicrobials alter the composition of a microbiota community, and how this alteration in community composition in turn affects the metabolic outputs from the community. Outcomes from this investigation will be used to build a model that can simulate the impact of an external insult, e.g., in the form of a strong antibiotic, on the functions of the gut microbiota. The student researcher will participate in both experimental or computational aspects of the project. The experiments will utilize state-of-the-art mass spectrometry methods to characterize the products of microbial metabolism. The computational aspect of the project deals with developing models and software tools that use information from publicly available genome database to determine the biochemical pathways and microbial organisms responsible for the metabolic products observed through mass spectrometry.
The overall objective of this research is to design and develop biologically derived, electrically conductive extracellular matrices (ECM) as a new class of electronic transducers in chemical sensor development. In particular, we will exploit exoelectrogenic bacteria that are widely available from deep marine sediments as cellular factories to generate metalloprotein-based ECMs, whose electrical resistance can be effectively modulated through the specific interaction with target analytes. The initial project will involve the development of ECM chemiresistor arrays by initiating electroactive biofilm growth on interdigitated micorelectrodes, as well as proof-of-concept experiments and computational modeling to investigate and optimize their performance against a wide range of molecular analytes. Depending on the results, next-phase studies will target the modulation of sensor specificity by chemically modifying the extracellular electron pathways or genetically engineering the matrix conductive properties with synthetic biology tools.
Van Deventer Lab
Genetic code manipulation is a powerful approach for discovering new therapeutics and understanding basic biology. However, the fundamental principles underlying efficient genetic code manipulation are not known. This project seeks to use a combination of experimentation and statistics to identify the best strategies for enhancing genetic code manipulation in yeast.

**Conventional genetic code**

<table>
<thead>
<tr>
<th>Met</th>
<th>Leu</th>
<th>Pro</th>
<th>Leu</th>
<th>STOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUG</td>
<td>CUG</td>
<td>CCC</td>
<td>CUC</td>
<td>UAG</td>
</tr>
</tbody>
</table>

**Versatile, robust expanded genetic code**

<table>
<thead>
<tr>
<th>Met</th>
<th>Leu</th>
<th>Pro</th>
<th>Leu</th>
<th>AzF</th>
<th>Lys</th>
<th>STOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUG</td>
<td>CUG</td>
<td>CCC</td>
<td>CUC</td>
<td>UAG</td>
<td>AAA</td>
<td>UAA</td>
</tr>
</tbody>
</table>
**Identifying and mitigating metabolic incompatibilities in engineered cells.**

PIs: Soha Hassoun (CS), Nikhil U. Nair (ChBE)

Increasing understanding of metabolic and regulatory networks underlying microbial physiology have enabled engineering of progressively more complex cells for production of biofuels, bioplastics, fragrances, and drug molecules. Despite the best efforts by engineers to describe various biological systems from a reductionist perspective, unpredictable outcomes still emerge from unforeseen interactions between heterologous synthesis pathways and the host cell. Such interactions are difficult to predict when designing engineered cells and often manifest during experimental testing as inefficiencies that need to be overcome. Despite advances in tools and methodologies for strain engineering, there currently remains a lack of tools that can systematically and quantitatively identify such interactions.

The objective of this project is to investigate a fundamental yet overlooked interaction mechanism that presents itself at the metabolic level – a phenomenon we dub as “metabolic network disruption”. The premise behind this phenomenon is that high concentrations of heterologous enzymes and metabolites lead to formation of unintended byproducts that disrupt the host metabolic network and decrease biosynthetic pathway productivity. Such disruption is observed experimentally in many studies where careful gene knockout strategies consistently fail to completely silence byproduct formation. Our hypothesis is that substrate promiscuity, where an enzyme transforms other substrates in addition to its natural substrate, is the culprit behind such unexplained byproducts. In this summer project, we will test this hypothesis by developing tools and workflows to computationally identify, quantify, and provide interventions to overcome such metabolic disruptions. This work will contribute to a bigger project that involves the experimental validation of our computational findings.

Metabolic cellular network (green oval) consisting of native metabolites and reactions, and heterologous synthesis pathway (red) consisting of heterologous metabolites and reactions. Heterologous enzymes catalyze their own natural substrates and in some cases, they promiscuously catalyze native host metabolites. The resulting products may be new metabolites that are not part of the metabolic models or existing metabolites. Native enzymes act on their own natural substrates and in some cases on heterologous metabolites, resulting either in native metabolites or new metabolic products.
Tzanakakis Lab

Stem cells hold great promise for the engineering of tissues to replace or reconstitute native tissue after injury or disease. Scalable processes for the expansion of stem cells (SCs) and their differentiation to therapeutically useful phenotypes according to firm specifications are highly desirable. Given the prevailing empirical approaches to SC differentiation in dishes, there is little insight about the underlying biological determinants and their practical manipulation in commercially relevant, large-scale systems. Quantitative models are essential for the rational design and prediction of SC differentiation processes and their translation to biomanufacturing practices. The ‘average’ population behavior described by current models is unsuitable for the inherently heterogeneous SC ensembles. To that end, our laboratory works on quantitative models of SC specification, namely, population balance equations (PBEs) coupled to measurable distributions of cellular traits. We utilize genome-editing methods such as CRISPR/Cas9 to engineer SC lines suitable for lineage tracing experiments. These cells are expanded in stirred-suspension bioreactors and subsequently subjected to directed differentiation along particular lineages such as pancreatic endoderm and cardiac mesoderm. The physiological state functions of the cell populations are extracted through a combination of experiments and computational solution of the relevant inverse PBE problem. This leads to a quantifiable landscape of commitment along domains of variables such as differentiation stimulus concentrations, duration of treatment and bioreactor variables such as agitation, cell density etc. The outcomes can be employed for fine-tuning the yield of therapeutically useful SC-derived progeny. As part of the planned activities, undergraduates will take part in the design of experiments to test and optimize combinations of differentiation factors, their concentrations, and duration of treatment of stem cell-derived pancreatic or cardiac cell precursors. In addition to routine experimental method (e.g. qPCR, immunostaining) undergraduate students in the program write codes for numerical calculations based on the models developed in the laboratory.
Machine-learning based image processing
Light microscopy, coupled with image processing, offers an exciting potential to noninvasively obtain spatially and temporally resolved information on metabolic activity in cell populations. However, it remains a major challenge to determine the precise relationship between image features (geometric and contrast patterns) and metabolic parameters (e.g. pathway activity). Mapping this relationship is a complex, high-dimensional problem as multiple biochemical pathways (e.g. lipid synthesis, degradation, and storage) contribute to multiple image features (lipid droplet size and distribution). We propose a novel approach that draws from graph modeling and machine learning. We will construct a set of expert user-labeled training data comprised of images of (nearly) homogeneous cell cultures where the metabolic state is known. The metabolic state variables of interest are pathway fluxes, determined from isotopic labeling experiments. From these data, we will “learn” a mathematical mapping between image features and metabolic rates of interest. We hypothesize that local morphological features and higher order statistical information in the image – obtained by the local interconnection among the local morphological features – can be directly and robustly correlated to the relevant metabolic parameters.
Miller Lab
The goal of this research is the development of computer vision processing methods for automating the analysis of high throughput microscopy data to address problems such as identification of lipid droplets in fat and liver cells and quantification of their size and shape given images and/or video streams from one or more microscopes. The project will involve close collaboration with faculty and students from the chemical and life sciences. The student researcher is expected to have prior programming experience with Matlab, python, C/C++ or a similar computational tool or language. The initial part of the summer will focus on developing the necessary background in image processing and analysis while the bulk of the times will be devoted to applying the methods to a real-world problem.

(A) Original image (B) Result of Random Forest Classifier (C) Final image showing individual droplets