The 2021 University of Chicago Undergraduate Research Symposium: Online Proceedings

Virtual Poster Session 2:

Biological Sciences Collegiate Division
Identification of -7/del7q Leukemia Specific Genetic Vulnerabilities

Anjali Kotamarthi, 2nd-Year, Biological Sciences, Health and Society
Mentor(s): Prof. Megan McNerney, Pathology; Dr. Madhavi Senagolage, Pathology, McNerney Lab

-7/del7q Acute Myeloid Leukemia (AML) is a high-risk leukemia that accounts for about 8% of all AML patients, often resulting in a poorer prognosis as many generalized therapy methods such as chemotherapy are rendered ineffective. As the prognosis of myeloid leukemia patients remains poor, a need for targeted therapies which prove lethal to a select group of cells is needed in order to create a more effective and less toxic therapy. One approach is to identify essential genes on chromosome 7 which correlate to increased efficacy of particular drugs when expressed at haploinsufficient levels. Using previously published genome wide CRISPR-CAS9 knockout screen data conducted in AML cell lines, we identified 239 essential chromosome 7 genes in AML cell lines and further narrowed this list to 44 genes which can be targeted by a commercially available drug based on data collected from Drug-gene Interaction Database (DGIdb) and DrugBank. After identification, drug screening data performed in AML cell lines from the Genomics of Drug Sensitivity in Cancer (GDSC) database as well as drug screening data from AML patients in the BEAT AML database were utilized to verify if low expression of the target genes correlated with a difference in drug sensitivity through measured AUC values. Of the essential genes on chromosome 7 with a potential drug candidate, Bortezomib, a drug targeting proteasome subunit PSMA2, and Palbociclib which targets cyclin dependent kinase CDK6, hold potential of being more efficient in inhibiting cell proliferation in patients with -7/del7q karyotypes. These findings encourage the expansion of research for these drugs in relation to AML potentially through AML xenograft mouse models, and highlight the importance of characterizing genetics vulnerabilities in specific cancer types.
NLRP3 Activation Controls Prostate Cancer via Induction of Macrophage-mediated Phagocytosis

Anna Argulian, 2nd-Year, Biological Sciences
Mentor(s): Prof. Akash Patnaik, Medicine, Hematology/Oncology; Dr. Kiranj Chaudagar, Patnaik Lab; Dr. Srikrishnan Rameshbabu, Patnaik Lab

Recent advances in immune checkpoint inhibitors (ICI) have revolutionized the cancer therapy landscape, improving survival in subsets of patients across a range of malignancies. However, the responses have been very limited in advanced prostate cancer (PC), due to multiple mechanisms of intrinsic resistance to ICI. One of the central mechanisms for resistance to ICI in PC is a sparse immune infiltrate, with a predominance of myeloid suppressive cells within the tumor microenvironment. We hypothesized that activation of the nucleotide-binding domain (NOD)-like receptor protein 3 (NLRP3) inflammasome can activate innate immunity within myeloid suppressive cells, thereby overcoming intrinsic resistance to ICI. We evaluated the impact of an NLRP3 agonist (BMS-392959) in a murine c-myc-driven prostate cancer mouse model. Our results demonstrate significant tumor control in BMS-392959 treated mice, which is accompanied by a global increase in immune cell infiltration. Mechanistic studies reveal a tumor extrinsic, macrophage mediated anti-tumor response driven by M1 polarization and subsequent tumor phagocytosis. Collectively, our results demonstrate NLRP3 as a promising therapeutic target to overcome resistance to ICI in advanced PC.
Distinct Interferon Signature and Novel Biomarkers of Secondary Progressive Multiple Sclerosis

Avital (Avi) Fogel, 2nd-Year, Biological Chemistry
Mentor(s): Prof. Anthony Reder, Neurology; Dr. Xuan Feng, Reder Lab

Multiple Sclerosis (MS) is a neurodegenerative autoimmune disease in which the immune system attacks neuron myelin sheaths in the brain. MS often leads to physical disability due to lowered communication efficiency between brain and body. During MS attacks, immune control declines, along with dysregulation of interferon-β (IFN-β), an antiviral signaling protein. IFN-β, the first approved MS therapy, reduces attacks, delays disease progression, and corrects abnormal IFN-β signaling in Relapsing-Remitting MS (RRMS), the most common disease form. However, most RRMS patients eventually develop secondary progressive MS (SPMS), and the benefits of IFN-β treatment are then diminished. We studied the IFN signature and therapeutic responses in RRMS and SPMS. We hypothesized that changes in this signature would allow us to differentiate between SPMS and RRMS and find therapeutic targets for SPMS treatment. Blood samples were from 59 patients from 6 groups of treated and untreated MS and controls. We visualized and investigated differential gene expression in immune cells, using a range of bioinformatic tools, including Transcriptome Analysis Console, Ingenuity Pathways Analysis, and Gene Ontology. Serum levels from Luminex multiplex of 78 proteins were correlated and compared. In untreated patients, gene and protein data suggested that SPMS is more ordered or quieter than RRMS, despite being more progressive, in both global gene expression and in specific pathways that regulate immune cell infiltration through the Blood Brain Barrier. We suggest this is due to recovery of aberrant gene signaling from regulatory and repair mechanisms in the transition from RRMS to SPMS. In SPMS, we detected short- and long-term effects of IFN-β therapy and suppression of pro-inflammatory pathways and upregulation of multiple anti-inflammatory, antioxidant genes. These findings suggest new biomarkers for SPMS as well as important pathways for developing future MS therapies.
Mapping Macroevolutionary Change in the Pelvic Fin in Actinopterygians
Elisabeth Incardona, 3rd-Year, Biological Sciences
Mentor(s): Prof. Michael Coates, Organismal Biology and Anatomy; Stephanie Sang, Organismal Biology and Anatomy

The body plan of the ray-finned fishes (Actinopterygii) has undergone considerable change throughout their evolutionary history. The positioning of their paired (pectoral and pelvic) fins are no exception to this: more evolutionarily derived actinopterygians tend to have pectoral fins that are shifted upwards on the body combined with pelvic fins that are shifted forwards and closer to the pectoral fins. Mapping these changes through quantitative anatomical analysis is crucial for understanding the evolutionary significance and functional consequences of these differing fin positions. We investigated a possible mechanism for this aforementioned forward shift in the pelvic fin by determining whether anatomical changes in the bony girdle supporting the pelvic fins correlated with positional shifts in the fin itself across numerous actinopterygian families. We measured pelvic fin position relative to the pectoral fin by using R to place landmarks at key locations on images of fish, which the script would output as XY-coordinates used for computing distances. Different pelvic girdle characteristics (such as girdle plate length relative to body length, etc.) were measured on X-ray images of fish using image measuring software. We then conducted statistical analysis using Spearman’s rank correlation to determine whether these different pelvic girdle measurements were correlated with forward shifts in pelvic fin position. Of the pelvic girdle characteristics chosen, we found that a more raised anterior end and a sharper angle of pelvic girdle inclination relative to the midline of the body were indeed significantly correlated with a more forward-shifted pelvic fin. We suggest that the elevated anterior attachment of the pelvic girdle permitted attachment to the pectoral girdle, thereby stabilizing the forward shift of the pelvic apparatus.
Improving COVID-19 Patient Care by Deep Learning-based CT Scan Assessment

Elise Katsnelson, 2nd-Year, Biological Sciences, Computer Science & Molecular Engineering, Technology, and Innovation
Beatrice Katsnelson, 2nd-Year, Biological Sciences, Computer Science & Molecular Engineering, Technology, and Innovation

Mentor(s): Prof. Maryellen Giger, Radiology; Jordan Fuhrman, Committee on Medical Physics, Radiology

COVID-19 is an infectious disease caused by the novel 2019 coronavirus that usually presents as a respiratory illness with a range of severity and symptoms, such as pneumonia, trouble breathing, acute respiratory distress syndrome, and death. Currently, the main treatment for severe COVID-19 patients is corticosteroid administration. However, corticosteroids should be used sparingly, as they may cause serious detrimental side effects. Thus, this study aims to provide quantitative analysis of CT scans acquired from COVID-19 patients and utilize this understanding to inform treatment decisions and thus improve patient management and clinical outcomes. To accomplish this, the lung and COVID-19 involved tissue were manually segmented CT scans using the MATLAB Image Labeler by a student and experienced radiologist, respectively. The segmentations were then used to train, validate, and test a deep learning model to rapidly and consistently segment CT scans and extract quantitative imaging features such as the ratio of diseased (COVID-19) tissue to total lung tissue and intensity-based radiomic features within the diseased tissue region. The chosen model was a two-headed U-Net architecture, which takes a single slice from a CT scan as input and automatically identifies image pixels that represent the lungs as well as areas of COVID-19 involvement. The segmentation model performance will be evaluated with the Dice similarity coefficient, which compares the overlap of the model’s automatically produced lung and COVID-19 involved tissue segmentations to those provided manually. The quantitative features extracted from the segmentations will be utilized to create a model for determining optimal patient treatment options between multiple corticosteroids (dexamethasone, methylprednisolone) as well as determine COVID-19 severity, monitor response to treatment, and predict long-term dysfunction. Thus, successful application of this approach has significant implications in both clinical care and resource management.
The 2021 University of Chicago Undergraduate Research Symposium: Abstract

Effects of Captivity on the Vertebral Bone Microstructure of Xenarthran Mammals (Title Case)
Ellianna (Ellie) Zack, 4th-Year, Biological Sciences spec. Ecology and Evolution
Mentor(s): Dr. Stephanie M. Smith, The Field Museum of Natural History; Dr. Kenneth D. Angielczyk, Committee on Evolutionary Biology, The Field Museum of Natural History

Captive specimens in natural history collections allow researchers to inspect the morphologies of rare taxa, but the lifestyles, diets, and lifespans of captive animals differ from those of their wild counterparts. We compared bone microstructure of trunk vertebrae in captive and wild xenarthran mammals (sloths, armadillos, and anteaters). Because trabecular bone architecture (TBA) adapts to in-vivo forces, bone microstructure reflects ecology and behavior, but this means that it may differ between captive and wild specimens of the same species. We collected µCT scans of the last six presacral vertebrae in 13 species of fossorial, terrestrial, and suspensorial xenarthrans ranging in body mass from 120g (*Chlamyphorus*) to 35kg (*Myrmecophaga*). For each vertebra, we measured bone volume fraction (BVF); trabecular number (TBN) and mean thickness (TBTH); global compactness (GC); and cross-sectional area (CSA). Wild specimens generally have more robust trabeculae, but this differs based on species, vertebral position, ecology, and pathology. The wild specimens of fossorial taxa (*Dasypus*) have significantly more robust trabeculae than their captive counterparts, but there is no clear difference in TBA of wild and captive specimens in suspensorial taxa (*Bradypus* and *Choloepus*). These data suggest that locomotor ecology affects the level to which captivity affects bone microstructure. Captive specimens of both *Tamandua* and *Myrmecophaga* have higher BVF and TBTH than their wild counterparts, indicating more brittle trabeculae due to bone pathologies caused by captivity. Our results add to the understanding of variation in mammalian bone microstructure and suggest caution when including captive specimens in research on TBA.
The 2021 University of Chicago Undergraduate Research Symposium: Abstract

The Relationship Between Psychotic Disorders and Substance Abuse Comorbidities
Indira (Indi) Khera, 3rd-Year, Biological Sciences, Creative Writing
Mentor(s): Prof. Sarah Keedy, Psychiatry, Cognition and Emotion Neuroscience Laboratory

Individuals with psychotic disorders such as schizophrenia are thought to be disproportionately affected by substance misuse disorders. It has been found that the rate of comorbid substance abuse with schizophrenia is 3x higher than in the general population. The significance of this observed relationship is not comprehensively understood. I was awarded the Hoeft Grant to explore the complicated dynamic between substance abuse and psychotic disorders. The first step was to assess substance abuse prevalence across novel categories of psychosis experimentally created previously on the basis of neurobiological similarity. The novel categories are called “biotypes,” which are novel groupings of psychotic disorders that utilize multimodal biomarkers to define their membership, providing an alternative to current diagnostic groups (schizophrenia, schizoaffective, bipolar). I am currently comparing substance abuse prevalence between diagnostic groups and biotypes to determine whether substance misuse is differentially prevalent among the two subgrouping systems and will move on to analyze cognitive test performance in relation to these findings. I am analyzing a large amount of subject data collected by the Cognition Emotion Neuroscience Laboratory at the University of Chicago, along with four other sites across the US, for the Bipolar Schizophrenia Network on Intermediate Phenotypes study. I am focusing my analysis on variables from a standardized “Diagnostic Form” completed for each subject, which notes primary comorbidities. I am conducting my analyses in SPSS version 26. This research would shed light on the complicated relationship between psychotic disorders and substance abuse, as well as identify differences or similarities in prevalence between traditional non-biologically organized diagnostic groups and biomarker-based biotypes. By understanding how substance abuse interacts with biological markers in psychotic disorders, we can work towards learning more about the relative impact of substance misused on these already complicated illnesses, opening the door to potentially improving outcomes for individuals with these disorders.
The Global Personal Vehicle Fleet: A Case of Systematic Underreporting and Inconsistencies

Jaeda Roberts, Recent Alum, Environmental Science
Mentor(s): Prof. Kavi Bhalla, Public Health Sciences

Global modernization and urbanization have led to an increase in the number of vehicles owned and used worldwide. Developing countries in Sub-Saharan Africa, Asia, and Latin America have seen a dramatic increase in population in the last 30 years and in the number of vehicles on the road, especially in urban areas. Yet, the total number of cars, motorcycles, and bicycles in these countries varies based on data source. Vehicle registration is often not enforced, especially for motorcycles, which creates major discrepancies between the number of working vehicles present in country and the reported vehicles. Countries such as the United States and most of Western Europe have widely utilized vehicle registration systems. Yet, there is often a discrepancy between these numbers and data collected from nationally represented surveys and censuses. In this study, we aimed to collect vehicle registration, census, and nationally representative survey data from every country to compare sources and predict the number of households that own a car, motorcycle, and/or bicycles. Our data suggest that these numbers are underestimated globally. We recommend that new survey and census methods should be implemented on a global scale to accurately count the number of vehicles on the road. Furthermore, this data should be used to develop road traffic policies in developing countries to prevent road injuries and deaths.
Curvature as an Indicator of Successful Aortic Dissection Endovascular Repair
Kameel Khabaz, 2nd-Year, Biological Sciences
Mentor(s): Prof. Luka Pocivavsek, Surgery

Type B aortic dissection (TBAD) is a challenging clinical problem. Historically, TBAD has been treated by medical management or open surgery. Thoracic endovascular repair (TEVAR), the minimally invasive insertion of a stent graft, is a novel treatment method with improved peri-operative outcomes but mixed long-term results. The major challenge in treating TBAD is patient selection, determining the optimal treatment modality on a per-patient basis. Clinicians use computed tomography angiography (CTA) imaging to track dissection evolution and inform their treatment strategy. However, the current paradigm measures linear distances of the aorta’s diameter. This measurement is a poor indicator of aortic remodeling because it fails to capture the aorta’s three-dimensional anatomy. We develop a novel method of characterizing aortic geometry and retrospectively test it to see if it can accurately distinguish successful from failed TEVAR operations. We study 20 TBAD patients that underwent TEVAR, 10 with desired outcomes and 10 with failed outcomes. We also study 10 patients with no aortic pathology. Using CTA scans, pre- and post-operative segmentations of the aorta are performed. For each geometry, we compute surface curvature and calculate probability distribution functions of the Koenderink shape index, a measure of shape, and curvedness, a measure of scale. We then plot the variance of the shape index distribution versus the mean of the curvedness distribution. Finally, we compared our surface curvature approach with a traditional approach of measurements based off of the aortic centerline: the tortuosity index and inverse mean aortic diameter. We observed that our surface curvature approach resulted in much better separation of patients with desired outcomes from patients with poor outcomes when compared to the centerline approach. Our results indicate that surface curvature is a viable indicator of post-TEVAR aortic remodeling. Our ultimate goal is to create a toolset to better inform the surgeon’s decision-making process.
The effective management of aortic diseases, such as dissections and aneurysms, requires the assessment of diseased and non-pathologic (normal) aorta morphology. The current representation of aorta geometry in the literature includes linear measurements of maximum diameter, as well as centerline-based measurements of tortuosity. However, the complexity of aortic pathology warrants the examination of geometric features that go beyond such linearization. The characterization of normal aorta geometry will establish a baseline for comparison with surface curvature of fragile aortas and ultimately improve risk stratification in asymptomatic adults. We use tools from differential geometry to quantify variation in healthy thoracic aortas, specifically the Koenderink curvedness, an intrinsic measure of curvature, and shape index, a measure of local shape that is independent of scale. We studied a cohort of twenty-three trauma patients that presented to the University of Chicago Medicine without aortic pathology. We created models of each patient’s aorta based on their initial computed tomography (CT) scans and subsequent scans. Curvature calculation is performed on the entire thoracic aorta and results confirmed that normal aortas adopt a largely cylindrical geometry with a shape index distribution centered around 0.5. Unlike shape index which remained stable over time, curvedness varied between patients and within patients during one hospitalization. This indicates proportional size changes in the aorta, specifically the arch region. We found that the shape of an aorta can change substantially in a short period of time, which should be taken into account when planning various endovascular interventions. Our next steps are to build a large data set of normal aortic geometries in order to identify deviation in aortic geometry that puts patients at high risk for developing vascular disease and relate variation in aortic surface curvature to clinically measured parameters such as blood pressure and volume status.
Education-Labor Mismatch and Health Outcomes Among Immigrants in the US
Laura Chen, 3rd-Year, Biological Sciences and Public Policy Studies, Statistics
Mentor(s): Prof. Aresha Martinez-Cardoso, Public Health Sciences; Amel Omari, University of Michigan School of Public Health

In the US, immigrants are disproportionately represented among overeducated workers in the labor market, relative to other workers in the same occupations. While prior studies have explored how overeducation (also referred to as overmatch) among immigrant workers shapes their economic and social incorporation in the US, few have investigated its relationship to physical health. Downward occupational mobility, such as overeducation/overmatch, however, has the potential to exact a toll on immigrant health through various stress mechanisms. For this project, we use the 2010 and 2012 waves of the Health and Retirement Study (HRS) to interrogate how education-labor mismatch shapes health outcomes among a representative sample of US and foreign-born older adults. Leveraging labor history as well as enhanced health and biomarker data collected in the HRS, we capture workers’ health and their experiences of education-labor mismatch. After applying sampling weights to account for the complex design of the HRS, we used descriptive and bivariate statistics to better characterize the distribution of mismatch and other characteristics in our sample. Then, we utilized linear and logistic regression models to assess the relationship between mismatch and health, controlling for specific covariates. Overall, we find overmatch to be positively associated with several deleterious health outcomes, including poorer self-rated health, smoking, and stress-sensitive biomarkers such as C-reactive protein and A1c. Further, the relationship between mismatch and health outcomes is stronger for immigrants compared to US-born workers. Our analysis offers a novel contribution to the literature by exploring how a unique manifestation of social inequality in the US, education-labor mismatch, may be especially consequential for the health and wellbeing of immigrants.
Analysis of Synergistic Information in Mouse V1 Neuronal Networks

Leya Luo, 3rd-Year, Mathematics, Neuroscience

Mentor(s): Prof. Jason MacLean, Neurobiology, Grossman Institute; Elizabeth deLaittre, Computational Neuroscience

Analyzing the functional relationships between groups of neurons can help in understanding the dynamics and information processing of the brain. Prior analysis of mouse brain slices has shown that the neuronal triplets with the highest amount of synergistic information have high amounts of recurrent connections, while those with the lowest amount of synergistic information have high amounts of feedback connections. However, it is unclear if the same pattern holds in vivo. To analyze this, I used various spike trains that were inferred from calcium fluorescence traces of V1 in a mouse viewing drifting gratings, and calculated the pairwise and joint transfer entropy for all pairs and triplets of neurons in each dataset. Next, I used the pairwise transfer entropy as an adjacency matrix for a functional network of neurons. In order to see if connection strength impacted the results, I thresholded the transfer entropy matrix at various percentiles between 10 and 99.5, and only counted values above the threshold percentile as connections. After constructing these networks, I categorized all the triplets in each network by their motif and graphed the distribution of the synergy values for each motif. However, in contrast to the previous work, high synergy was positively correlated with the number of feedback connections in a triplet. This pattern held across all the datasets and all the thresholds, and was most prominently present in networks with more neurons. Additionally, when the synergy values were normalized by the entropy of the target neuron, high synergy values were still positively correlated with the number of feedback connections. Therefore, in vivo and in vitro triplet motifs show opposite structural correlations with synergy, suggesting that there exist major differences between the functional relationships between neurons in living brain tissue and organotypic slice culture.
Characterizing the Role of EssH, a Peptidoglycan Hydrolase, in the Type VII Secretion System of *Staphylococcus aureus*
Maria Krunic, 4th-Year, Linguistics, Biological Sciences, Chemistry
Mentor(s): Prof. Dominique Missiakas, Microbiology; Dr. Maksym Bobrovskyy, Missiakas Lab

Nosocomial infections due to *Staphylococcus aureus* pose a significant problem in healthcare today, particularly with the rise of methicillin resistant strains. Protein secretion plays an important role in the pathogenesis of *S. aureus* infections. Recently, the type VII secretion system (T7SS) has been found to be required for virulence of *S. aureus* in the mouse model. The T7SS assembles in the bacterial membrane and promotes secretion of proteins with WXG-motifs and other effectors like the nuclease EssD. Secretion of these substrates relies on the peptidoglycan hydrolase EssH. The CHAP domain in EssH confers hydrolytic activity, however the rest of the protein’s function has not been elucidated yet. To date, the mechanism by which EssH promotes secretion through T7SS is not known. This research aims to investigate how EssH interacts with T7SS and how it promotes the secretion of its effectors. We hypothesize that peptidoglycan hydrolysis might be necessary either for T7SS assembly in the cell envelope or to facilitate substrate passage through the peptidoglycan layer. Previous work suggests that EssH interacts with the T7SS machinery, however the interacting surface of EssH is unknown. To test whether the EssH domain of unknown function (DUF) is responsible for this, we generated *S. aureus* deleted for DUF and tested whether this variant co-purifies with the T7SS machinery. In addition, we probed assembly of the T7SS machine in the cell envelope of the DUF mutant. Finally, we investigated whether T7SS-secreted proteins get trapped in the peptidoglycan layer of the DUF mutant. The EssH protein provides a unique area of study as it does not induce lysis of staphylococci despite exhibiting peptidoglycan hydrolase activities. Since EssH is required for protein secretion via T7SS, understanding more about its function on a molecular level presents a crucial opportunity to develop methods dedicated to decreasing *S. aureus* pathogenicity.
The Cooperation of TCF7 and HEB in β-Catenin Driven Leukemogenesis
Melissa Tracy, 3rd-Year, Biological Sciences, Health and Society
Mentor(s): Prof. Fotini Gounari, Immunology; Stephen Arnovitz, Gounari Lab

The Wnt-signaling pathway is evolutionarily conserved due to its fundamental role in embryonic development, cell proliferation, and differentiation. In the absence of Wnt ligands, β-catenin is degraded in the cytoplasm, however when Wnt signaling is active, the degradation process is interrupted, and β-catenin enters the nucleus. As β-catenin accumulates in the nucleus, it binds the transcription factor TCF7 and activates Wnt regulated genes. TCF7 collaborates with multiple other transcription factors to effect thymocyte development including HEB. Dysregulation of the Wnt pathway results in overexpression of β-catenin which can lead to leukemogenesis, which is commonly observed in human T-cell Acute Lymphoblastic Leukemia (TALL). Stabilization of β-catenin in developing T cells stalls their development at the CD4⁺CD8⁺ (double positive or DP) stage (CAT mouse model). The affected DP T cells are genomically unstable and predisposed to leukemogenesis. It was observed that when TCF-7 was knocked out from T cells with simultaneous β-catenin stabilization (CAT-TCF), the leukemia did not develop, however T cell development was still blocked. In contrast, the potential cooperation of HEB with β-catenin in thymocyte development and leukemogenesis is unknown. Early observations indicate that HEB knock out mice with β-catenin overexpression (CAT-HEB) develop lymphomas that are molecularly distinct from the aberrant β-catenin lymphomas, while the development of T cells is restored. Through ATACseq and RNAseq analysis, we have found that these CAT-HEB lymphomas may be driven in part by dysregulation of several ribosomal proteins and metabolic pathways through a dramatic increase of TCF7 binding sites. Through continued research, we hope to better understand the Wnt signaling pathway as well as the activities of stabilized β-catenin with its partner transcription factors. If we are able to better understand the cooperation of β-catenin with HEB and TCF7 and the mechanisms involved, our work may provide information for future treatment strategies for TALL.
Alignment-free Evolutionary Analysis of Unalignable Protein Sequences
Rosalind Pan, 3rd-Year, Biological Sciences & Computational and Applied Mathematics
Mentor(s): Prof. D. Allan Drummond, Biochemistry and Molecular Biology

Low-complexity regions (LCRs) are important for modulating the formation of many biomolecular condensates, but the sequence-function relationship of LCRs remains unclear. Finding evolutionarily conserved sequence features in related LCRs would inform searches for function. However, many LCRs are largely unalignable, hindering analysis through conventional alignment-based evolutionary approaches. To address this fundamental challenge, here we model the sequences of LCRs using the joint probability distributions of sequence features, which circumvent the need for residue-level homology in sequence comparisons, thereby permitting alignment-free evolutionary analysis. By targeting poly(A)-binding proteins (PABP), a canonical stress-granule marker whose LCR modulates its condensation, we demonstrate that the analyses routinely carried out on aligned sequences can be satisfactorily performed on unalignable LCRs using our model. Importantly, the Jensen-Shannon distances between sequence-based probability distributions can estimate the phylogenetic distances between PABP LCRs and this estimation is comparable to that achieved by quantifying sequence similarity between PABP structured domains using alignment-based methods. Moreover, we show that our model captures the majority of the sequence information in PABP LCRs and can be used to identify evolutionarily conserved sequence features despite poor residue-level conservation. Our future work will focus on broadly applying our promising new approach to other unalignable protein regions.