UCLA Computational and Systems Biology

An Interdepartmental Program

Computational and Systems Biology
Spring 2019 Poster Session

May 16th, 2019
2:00-4:00 PM
Boyer Hall 510 Suites

Comparing Unsupervised and Supervised Methods for scRNA-Seq Cell Type Identification
Julianne Converse, Feiyang Ma, Matteo Pellegrini PhD

One of the most common applications of single cell RNA sequencing (scRNA-Seq) data is cell type identification. A recently developed cell type identification method called Automated Cell Type Identification using Neural Networks (ACTINN) aims to predict cell types automatically by utilizing a reference of predefined cell types. The ACTINN algorithm trains on scRNA-Seq datasets, adding cell type identities to the reference of predefined cell types within each dataset trained. The traditional method of cell type identification involves the grouping of single cells with similar gene expressions into clusters in an unsupervised manner. Cell types can be identified by the canonical genetic markers of each cell cluster. There are limitations to the traditional method, such as variation in cluster formation, batch effects, and is time consuming as it requires a manual search for marker genes in order to identify cell types. ACTINN is designed to overcome these limitations and has proven to identify cell types quickly. The accuracy of ACTINN was tested using human peripheral blood mononuclear cell (PBMC) datasets and it identified cell types accurately. The algorithm was also trained on a mouse brain dataset and was then used to predict cell types of another mouse brain dataset. Several mouse brain cell types were identified correctly, but more training is required in order to expand the reference of predefined cell types. This method of automated cell type identification could have a major impact on single cell approaches and should improve as we increase the amount of training data.

Identifying Gene Expression Profiles of Electroconvulsive Therapy in Major Depressive Disorder
Jingyuan Fu, Daniel Nachun, Eliza Congdon

Electroconvulsive therapy (ECT) is a safe and effective fast-acting treatment for major depressive disorder (MDD) and has been shown to induce changes in peripheral blood biomarkers. However, no study has revealed its effect on peripheral blood gene expression. Therefore, we are interested in studying ECT's effect in gene expression and find biomarkers that could be indicative of risk for MDD. We sampled 40 MDD patients and 25 healthy controls and studied the gene expression patterns across samples, including comparisons between patients and controls, as well as patients before and after ECT treatment. We used differential gene expression analysis, weighted gene co-expression network analysis (WGCNA) and cell-type enrichment to analyze the gene expression profiles. We found that a gene module consisting of 1199 genes (including *SCAMP3, VEGFB*) has a weakly significantly decreased expression in baseline patients than in controls (FDR = 0.0809), while also overrepresented in T-cells and glucocorticoid signaling pathways, which indicates glucocorticoid receptor could be a potential biomarker for depression. However, we found no significant difference in gene and gene module expression within patient group between treatment time points. Further exploration of that gene module is to be considered to understand the mechanism behind glucocorticoid signaling and depression. Other approaches than mRNA-based gene expression profiling would need to be considered for further understanding of ECT's effect.

Structural Analysis of Enantioselective Peptide Nano-assemblies by Micro Electron Diffraction
Ayesha Hamid, Christian Boothby, David Guo, Chih-Te (Ted) Zee, Jose A. Rodriguez

Pseudomonas syringae produces ice nucleating proteins (INPs), one of which is the protein inaZ, which allows water to freeze at elevated temperatures. These bacteria are found ubiquitously in the environment due to their association with the water cycle. The ability of this bacteria to nucleate ice makes plants more susceptible to frost damage and facilitating their invasion of the plants. INPs are thought to facilitate ice crystal formation but the mechanism for how this happens remains unclear. Ice nucleation proteins contain a highly repetitive amino acid sequence which may result in structures with repeated structural motifs that are collectively responsible for the ice nucleation. A beta solenoid structure has been predicted for the ice nucleating domain of inaZ which will be validated through these experiments. To this end, we are synthesizing small peptides representing the observed degenerate repeats found in inaZ using solid phase peptide synthesis. The peptide will then go through purification and the purest samples will be used for crystallization. Any crystals that are formed will be imaged and used to obtain structural data

through micro-electron diffraction. The synthesized peptides will then be tested on their ice nucleation capacity, by measuring the freezing point elevation of water containing the ice nucleating protein. Through crystallization trials we have seen production of fibers and while these are not crystals, this shows the ability of the protein to organize. These fibers should also allow ice nucleation to occur, however, we would prefer the formation of crystals in order to confirm a structure through micro-electron diffraction. Together, the structural and ice nucleating data will serve to validate the initial hypothesis for the structural factors of ice nucleating properties of proteins.

Assembling a DNA Reference Sequence Database for Metagenomic Analysis of Fungi within Microbial Communities

Aaron Karlsberg, Caitlin Loeffler, Serghei Mangul

Recent advances in high throughput gene sequencing technologies have dramatically accelerated the field of metagenomics in which scientists distinguish genetic material contained within environmental and host samples. Complete classification of the microbial community present within a sample depends on the reference genomes available for comparison. Current efforts throughout the globe promise to extend the genomic representation of microbial organisms across the tree of life. However, both the sequencing technologies as well as the conventions used to construct the reference genomes for microbial organisms are rarely uniform between research endeavors. Consequently, existing databases have failed to properly integrate new microbial reference sequences, which has diminished the quality of metagenomic research. Here, we report the development of microDB, a novel method able to integrate disjoint microbial databases into a single database of species and strain genomes. We have applied our method to assemble a comprehensive database of fungi genomes and plan to extend our approach to bacteria, virus and protozoa as well.

Testing Vascular Structure Modelling Principles by Generating Virtual Vascular Networks

Panagiotis Lymperopoulos, Alexander B. Brummer, Van M. Savage

Successful modelling of vascular structure can both further our understanding of metabolism in multicellular organisms and lead to the development of effective diagnostic tools for vascular disease. Testing the fundamental assumptions of vascular modelling is a crucial step toward realizing those aims. Here I evaluate whether space filling, minimal resistance to fluid transport and minimal building cost are valid and sufficient principles to build realistic vascular structure. I generate virtual vascular networks based on these principles, examine their impact on network structure and use the West-Brown-Enquist framework to describe their scaling properties. I then compare the scaling properties of virtual networks with those of mammalian and plant networks and observe that the combination of space-filling, minimization of resistance to fluid transport and minimization of building cost predict high asymmetry in sibling vessel radii. The largely symmetric branching of mammalian vasculature and the mix of symmetry and asymmetry in plant vascular structure support that at least one crucial component that favors network symmetry over asymmetry is absent from the model. Additionally, I find that some assumptions have equivalent effects on network structure despite constraining different geometric features of vessels. Future work can extend our network generator and utilize our insights to understand and describe the missing components to formulate a more complete picture of the forces governing vascular structure.

Validation of Automated Stroke Detection in CT Scans

Vineet Mathew, Lu Li Ph.D, Fabien Scalzo Ph.D

Detection of ischemic lesions at early stages of acute stroke is critical for clinical diagnosis and treatment decisions. Currently, the noncontrast Computer Tomography (CT) scan is the method of choice to detect early lesions thanks to its presence in most modern hospitals, low cost, and fast acquisition time. However, acute lesions are detected visually with very little help from software tools, thus almost only relying on simple thresholds/windowing methods to emphasize the lesion and estimate its location and volume. The development of sophisticated computational methods that could automatically detect acute

ischemic lesions from noncontrast CT scans would provide clinicians valuable decision support and could play a role in clinical diagnosis of acute stroke. Here we present an approach using a deep learning model for the detection of brain lesions on CT scans of patients admitted at the UCLA medical center for acute ischemic stroke. The model was trained using a gold standard obtained from annotated 3D CT scans of the brain where hypoperfused regions have been manually delineated.

Biomimetic Perception for Virtual Humans
Masaki Nakada, Gestina Yassa

In order to develop realistic virtual humans, we must equip them with an artificial visual system that very closely mimics our own. We aim to model a biomimetic human perception system. The model features two human-like eyes with non-uniformly distributed photoreceptors much like the human retina. The non-uniformly distributed photoreceptors capture scene irradiance using the Phong model of illumination and ray tracing. Automatically trained neural networks extract the required information for the virtual human to perform visually guided actions such as tracking and intercepting a moving target. We also explore realistic skin deformation models for the virtual human.

Nest structure influences colony relocation in harvester ants
Artem Pashchinskiy, Noa Pinter-Wollman

Spatial organization of living spaces affects the behavior of their inhabitants. These effects are observed both on the individual and on the group level in many species. Collective behavior of social insect colonies provides a convenient model for studying the details of how spatial constraints affect group dynamics. In this study, we examine nest relocation of harvester ants *Veromessor Andrei* in relation to the spatial architecture of the vacated nests. To identify features of the nest architecture that correlate with changes in collective behavior, we constructed a number of artificial ant nests with a wide range of architectural compositions and measured the efficiency of ant colony relocation in each of them. Our work led to the development of Point Curvature Index (PCI) and Architecture Curvature Index (ACI) which measure the curvature of the optimal path to the nest exit from a given point and from an average point of a given architecture, respectively. Through a combination of empirical work and computational agent-based modeling, we showed that both PCI and ACI are predictive of the colony relocation success. These results indicate that path windiness is an important intermediary between the spatial composition of the space and the behavioral features of its occupants. This observation can be further developed into a novel way of quantifying the complexity of living spaces. Additionally, the computational model of the nest relocation experiment developed in this study can be used to develop and test other hypotheses about the dynamics of nest relocation in harvester ants.

The contribution of sensory-motor cues to spatial selectivity decline in hippocampal place cells
Claire Polizu

Spatial selectivity is a property that plays a major role in spatial navigation for many organisms. Place cells are one class of spatially selective neurons: they selectively fire action potentials as a function of one's location in a space or area. Place cells rely on many cues to become spatially selective. Distal visual cues and sensory-motor cues are especially believed to be involved in spatial selectivity and hence navigation. Previous research has shown that distal visual cues alone cannot generate spatial selectivity and that sensory-motor cues are also essential. However, the specific importance of sensory-motor cues to spatial selectivity is still unclear. To elucidate the cues' importance, activity from the posterior parietal cortex (PPC)-- a part of the neocortex functioning in spatial navigation-- was observed in rats using only distal visual cues to do a task. It was then compared to PPC activity of rats using both distal visual and sensory-motor cues to do the same task. Analysis showed that PPC activity is similar between rats using distal visual cues to navigate, regardless of whether sensory-motor cues are available. It suggests that absence of sensory-motor cues likely causes disruptions in only hippocampal processes, and not cortical processes-- to cause lower spatial selectivity. Autism, in particular, is characterized by abnormal place

cell activity. Insights gained from this project could be applied to future research geared towards highlighting how spatial selectivity is biologically compromised in autistic patients.

Identification of Variants in the Germline Associated with the Development of Genomically Unstable Tumors

Jordan Uyeki, Nikolas Balanis, PhD., Thomas Graeber, PhD.

Numerous past studies have shown the importance of germline Single Nucleotide Polymorphisms (SNPs) in the prediction of cancer incidence. Genomically instable tumors do not properly regulate the fidelity of their genome, often leading to a high frequency of large scale chromosomal alterations known as chromosomal instability (CIN). We hypothesized that pre-existing germline variants may be associated with CIN in cancer, which is associated with a more aggressive course of the disease. Analyzing germline SNP chip data of over 11,000 samples from The Cancer Genome Atlas (TCGA), we performed the first Genome Wide Association Study analyzing the association of SNP patterns with the development of CIN. Running association analysis on SNP chip based genotype data against metrics of genomic instability (Integrated Copy Number Alterations and Ploidy), we were able to identify associated biomarkers for each cancer, at both the SNP level and upon aggregation, at the gene level. Enrichment analysis in Breast Invasive Carcinoma, revealed germline SNPS enriched in the TP53 pathway, and associated with cancer incidence, suggesting a pre-existing genetic basis for the course of the disease.