Jacqueline Penaloza

Title: Multiclass Classifier for Predicting Congenital Heart Disease Subgroups using data from Copy Number Variants

Abstract: Congenital Heart Disease (CHD) is a global health burden that is a major cause of infant mortality. The heart is the first organ to develop, and disruption of this process leads to defects. Copy Number Variants (CNVs) are chromosomal gains or losses. In this study, we leverage our access to data from the Cytogenomics of Cardiovascular Malformations (CCVM) Consortium. This dataset includes information on demographics, diagnosis, and clinical chromosomal microarray analysis of patients with CHD. The CHD subgroups we focused on are (Ventricular/Atrial) Septal Defect (VSD/ASD), Right Ventricular Obstruction (RVOTO), Left Ventricular Outflow Tract Obstruction (LVOTO), Heterotaxy (HTX), Conotruncal Defect (CTD), Atrioventricular Septal Defect (AVSD), and Anomalous Pulmonary Venous Return (APVR). We developed a machine learning classifier to identify patterns, both distinct and shared, between CHD subgroups, among features such as biological pathways, genomic region, developmental biology processes, and other factors.

Austin Allen

Title: An Alternative Method for Molecular Docking

Abstract: As of the year 2021, around 6.2 million Americans aged 65 or older (~36 million worldwide) have neurodegenerative diseases such as Alzheimer’s disease (AD). That’s more than 1 in 9 people over 65—and the prevalence of AD (as well as similar diseases) only increases as age increases.1 Over the past 20 years, AD has increased as a cause of death in America by 145% while other causes of death such as HIV and heart disease have seen significant decreases in lethality. Unfortunately, at the current rate, this is projected to become even worse. It is very clear that there is much work to be done in the field of AD—with great need for drug discovery, structural and mechanistic work, and research into the genetic basis of the disease. This project hopes to contribute to our genetic knowledge about AD progression.

Generally, Alzheimer’s pathology is characterized by the presence of filamentous protein aggregates in the brain. These protein aggregates are considered a biomarker for progressing neurodegeneration. There are two main proteins involved in this aggregation, MAPT and Amyloid-β. Microtubule associated protein tau (MAPT), usually just known as Tau protein, is present in the central nervous system in the form of six different isoforms. Tau’s native function is binding to the neuronal microtubules to stabilize them. However, in an Alzheimer’s brain, Tau protein is hyperphosphorylated and instead forms filamentous aggregates (fibrils) inside neurons.

The field of AD research has become much more structurally focused over the past few years. Many Cryo-EM structures of tau fibrils have been published and characterized, and many other studies have used these structures to attempt docking for small molecule ligands. Although there are a few promising diagnostic ligands to
detect and quantify aggregated tau in the brain (including the recently FDA approved Tauvid, an 18F radiolabeled PET probe), not much is known about their binding mechanisms. Similarly, not much is known about the binding mechanisms or binding locations of ligands known to induce or inhibit tau aggregation. Thus, computational methods have been utilized to perform molecular docking—which is then validated by fitting the results to Cryo-EM densities. However, many of the most commonly used software packages (Autodock Vina in particular) are what I would categorize as “black boxes”. You give the program your input, and it gives you an output—without accounting for a wide range of partial covalent interactions or other binding methods. Many of these methods are also not scalable to potentially perform global docking and are instead limited to local docking experiments. With this project, I propose a potentially better method for molecular docking. Using Rosetta compiled for MPI on a supercomputer cluster allows for high scalability limited only by computational resources. Additionally, use of the Rosetta scoring function “Orbitals” allows docking experiments to be performed which specifically consider partial covalent interactions (pi-pi stacking, cation-pi interactions, etc.).

Birkan Gokbag

Title: Review: Batch Effect Correction Methods for Single Cell Data

Abstract: In the analysis of single-cell datasets batch effect correction methods are vital for data analysis, where possible biases and noises introduced by different sequencing technologies, patients, batch runs, and other factors are eliminated during data integration. Ranging from statistical methods to deep learning frameworks, each method solves batch correction differently and produces different batch effect-free datasets. Therefore, it is important to choose the best approach suited for the datasets at hand. In this review, I will go over several batch effect correction methods, compare their techniques, and discuss their performances from published studies.

Friday, April 22nd, 11:00am-12:00pm

Carmen Zoom