



Undergraduate Research Symposium May 17, 2019 Mary Gates Hall

Online Proceedings

POSTER SESSION 1

Balcony, Easel 115

11:00 AM to 1:00 PM

Inference of Novel Gene-Disease Relationships in Rare and Undiagnosed Diseases

Robert Edward Minneker, Senior, Bioengineering, Computer Science

Amgen Scholar, McNair Scholar, UW Honors Program

Mentor: Sean Mooney, Biomedical Informatics and Medical Education

Mentor: Vikas Pejaver, Biomedical Informatics and Medical Education

Despite being individually uncommon, rare diseases are a heterogeneous set of nearly 7,000 diseases estimated to affect 6-8% of the global population, making them collectively common. Rare diseases are infrequent and their determinants are usually unknown which negatively impact clinicians' ability to make an accurate and timely diagnosis. With the inception of consortia such as the Undiagnosed Diseases Network (UDN) and ontologies such as the Human Phenotype Ontology (HPO) there is now a wealth of information on rare diseases. This rich information has enabled researchers to infer gene-disease relationships and clinicians to assign genetic diagnoses. However, rare diseases that are caused by unique or uncommon gene combinations remain unsolved. In this study, we address this problem by assuming that a disease can be described by a combination of clinical features or phenotypes and, thus, a combination of genes. By developing a pipeline to incorporate known gene-phenotype relationships from publicly available resources such as HPO and using various combinatorial search algorithms, we can prioritize the most likely sets of genes to result in a given set of phenotypes. Integration of our pipeline with individual patient data is expected to further reduce the search space and produce higher quality predictions. We anticipate the results from this study to advance the quality and accuracy of gene-disease relationship predictions for rare and undiagnosed diseases associated with one or more genes. Our pipeline could be used as a tool for the diagnosis and discovery of likely oligogenic, rare and undiagnosed diseases that are normally turned away from consortia like the UDN.

POSTER SESSION 4

Balcony, Easel 122

4:00 PM to 6:00 PM

Statistical De-Identification of Medical Notes

Samir Anthony Yhann, Senior, Physics: Applied Physics

Mentor: Vikas Pejaver, Biomedical Informatics and Medical Education

Mentor: Sean Mooney, Biomedical Informatics and Medical Education

Electronic health records (EHRs) are often used by clinical and data researchers in numerous ways for various scientific investigations. When sharing patient information, certain precautions must be followed as to prevent the risk of a malicious actor being able to extract sensitive information. This study examines an experimental method of removing potentially identifiable information from free text medical notes by finding and removing phrases which are statistically uncommon. Furthermore, this study assesses if this method reduces risk of identification while also maintaining the utility of the data. The method involves analyzing a free-text dataset by first breaking all text up into fixed length phrases. The frequencies of these phrases are then tracked across the entire dataset on a per-patient, per-note, and dataset-wide basis. To benchmark the method, notes are de-identified using the method and privacy and utility are tested under different conditions. The results from using this method on real clinical notes are expected to produce text that will not only be more secure but will also retain information useful for applications such as machine learning, natural language processing, and data analysis. If this method proves to be successful, it could lead to institutions being able to share medical notes with researchers more easily. This in turn would eliminate a major obstacle which medical researchers face, as it would give them access to more data. Finally, when data are shared between institutions for research, the risk of identification can be represented as an objective and quantifiable metric.