individuals of African genetic ancestry. We will approach this goal by completing the following objectives: (i) localize a genetic signal that accounts for the significantly increased risk for primary open-angle glaucoma in African Americans and (ii) utilize electronic health records (EHR) data to expand our understanding of risk to incorporate endophenotypes of glaucoma and other clinically recorded variables that may influence disease risk. METHODS/STUDY POPULATION: We will genotype at least 200 available African American samples with glaucoma on the Illumina Infinium® Expanded Multi-Ethnic Genotyping Array (MEGAEX) and perform admixture mapping. We will then access EHR data to expand our analysis beyond glaucoma to encompass other relevant risk modifiers captured in the clinical record. RESULTS/ANTICIPATED RESULTS: We anticipate localizing a genetic signal or signals that may account for the increased POAG risk in African Americans. Our calculations indicate that we have ~81% power to detect association at a LOD score of 2 and a risk ratio of 2. Thus, we are well-powered to detect a true signal at this modest level of association DISCUSSION/SIGNIFICANCE OF IMPACT: This project will not only help to achieve precision medicine by filling in the gaps in knowledge regarding glaucoma in African Americans, but it will also address health disparities and aid in the realization of the full potential of "big data" so that all of these elements can be incorporated into a better understanding of health disparities.

2140

Estimating microscopic structures of glomeruli in renal pathology

Pinaki Sarder, Rabi Yacoub and John E. Tomaszewski
University at Buffalo, State University of New York, Buffalo, NY, USA

OBJECTIVES/SPECIFIC AIMS: (i) Digitally quantify pathologically relevant glomerular microcompartmental structures in murine renal tissue histopathology images. (ii) Digitally model disease trajectory in a mouse model of diabetic nephropathy (DN). METHODS/STUDY POPULATION: We have developed a computational pipeline for glomerular structural compartmentalization based on Gabor filtering and multiresolution community detection (MCD). The MCD method employs improved, efficient optimization of a Potts model Hamiltonian, adopted from theoretical physics, modeling interacting electron spins. The method is parameter-free and capable of simultaneously selecting relevant structure at all biologically relevant scales. It can segment glomerular compartments from a large image containing hundreds of glomeruli in seconds for quantification—which is not possible manually. We will analyze the performance of our computational pipeline in healthy and streptozotocin induced DN mice using renal tissue images, and model the structural distributions of automatically quantified glomerular features as a function of DN progression. The performance of this structural-disease model will be compared with existing visual quantification methods used by pathologists in the clinic. RESULTS/ANTICIPATED RESULTS: Computational modeling will reveal digital biomarkers for early proteinuria in DN, able to predict disease trajectory with greater precision and accuracy than manual inspection alone. DISCUSSION/SIGNIFICANCE OF IMPACT: Automated detection of microscopic structural changes in renal tissue will eventually lead to objective, standardized diagnosis, reflecting cost savings for DN through discovery of digital biomarkers hidden within numerical structural distributions. This computational study will pave the path for the creation of new digital tools which provide clinicians invaluable quantitative information about expected patient disease trajectory, enabling earlier clinical predictions and development of early therapeutic interventions for kidney diseases.

2166

Semantic characterization of clinical trial descriptions from ClincalTrials.gov and patient notes from MIMIC-III

Jianyin Shao, Ram Gouripeddi and Julio C. Facelli

OBJECTIVES/SPECIFIC AIMS: This poster presents a detailed characterization of the distribution of semantic concepts used in the text describing eligibility criteria of clinical trials reported to ClincalTrials.gov and patient notes from MIMIC-III. The final goal of this study is to find a minimal set of semantic concepts that can describe clinical trials and patients for efficient computational matching of clinical trial descriptions to potential participants at large scale. METHODS/STUDY POPULATION: We downloaded the free text describing the eligibility criteria of all clinical trials reported to ClinicalTrials.gov as of July 28, 2015, ~195,000 trials and ~2,000,000 clinical notes from MIMIC-III. Using MetaMap 2014 we extracted UMLS concepts (CUIs) from the collected text. We calculated the frequency of presence of the semantic concepts in the texts describing the clinical trials eligibility criteria and patient notes. RESULTS/ANTICIPATED RESULTS: The results show a classical power distribution, $Y=2^{10}\ X^{(-2.043)}$, $R^2=0.9599$, for clinical trial eligibility criteria and $Y=5^{13}\ X^{(-2.684)}$, $R^2=0.9477$ for MIMIC patient notes, where Y represents the number of documents in which a concept appears and X is the cardinal order the concept ordered from more to less frequent. From this distribution, it is possible to realize that from the over, 100,000 concepts in UMLS, there are only ~60,000 and 50,000 concepts that appear in less than 10 clinical trial eligibility descriptions and MIMIC-III patient clinical notes, respectively. This indicates that it would be possible to describe clinical trials and patient notes with a relatively small number of concepts, making the search space for matching patients to clinical trials a relatively small sub-space of the overall UMLS search space. DISCUSSION/SIGNIFICANCE OF IMPACT: Our results showing that the concepts used to describe clinical trial eligibility criteria and patient clinical notes follow a power distribution can lead to tractable computational approaches to automatically match patients to clinical trials at large scale by considerably reducing the search space. While automatic patient matching is not the panacea for improving clinical trial recruitment, better low cost computational preselection processes can allow the limited human resources assigned to patient recruitment to be redirected to the most promising targets for recruitment.

2182

Developing a corpus for natural language processing to identify bleeding complications among intensive care unit patients

Rashmee Shah, Benjamin Steinberg, Brian Bucher, Alec Chapman, Donald Lloyd-Jones, Matthew Rondina and Wendy Chapman
The University of Utah School of Medicine, Salt Lake City, UT, USA

OBJECTIVES/SPECIFIC AIMS: An accurate method to identify bleeding in large populations does not exist. Our goal was to explore bleeding representation in clinical text in order to develop a natural language processing (NLP) approach to automatically identify bleeding events from clinical notes. METHODS/STUDY POPULATION: We used publicly available notes for ICU patients at high risk of bleeding (n = 98,586 notes). Two physicians reviewed randomly selected notes and annotated all direct references to bleeding as "bleeding present" (BP) or "bleeding absent" (BA). Annotations were made at the mention level (if 1 specific sentence/phrase indicated BP or BA) and note level (if overall note indicated BP or BA). A third physician adjudicated discordant annotations. RESULTS/ANTICIPATED RESULTS: In 120 randomly selected notes, bleeding was mentioned 406 times with 76 distinct words. Inter-annotator agreement was 89% by the last batch of 30 notes. In total, 10 terms accounted for 65% of all bleeding mentions. We aggregated these results into 16 common stems (eg, "hemorr" for hemorrhagic and hemorrhage), which accounted for 90% of all 406 mentions. Of all 120 notes, 60% were classified as BP. The median number of stems was 5 (IQR 2, 9) in BP Versus 0 (IQR 0, 1) in BA notes. Zero bleeding mentions in a note was associated with BA (OR 28, 95% CI 6.5, 127). With 40 true negatives and 2 false negatives, the negative predictive value (NPV) of zero bleeding mentions was 95%. DISCUSSION/SIGNIFICANCE OF IMPACT: Few bleeding-related terms are used in clinical practice. Absence of these terms has a high NPV for the absence of bleeding. These results suggest that a high throughput, rules-based NLP tool to identify bleeding is feasible.

2204

Evaluations of physiologic perturbations and their relationship with length of stay in neonatal hypoxic-ischemic encephalopathy

Susan Slattery, Lei Liu, Haitao Chai, William Grobman, Jennie Duggan, Doug Downey and Karna Murthy

OBJECTIVES/SPECIFIC AIMS: Neonatal hypoxic-ischemic encephalopathy (HIE) is frequently accompanied with physiologic perturbations and organ dysfunction. Markers of these perturbations and their associations with length of stay (LOS) are uncertain. To estimate the association between changes in selected physiologic and/or laboratory values with LOS in newborns with HIE. METHODS/STUDY POPULATION: Using the Children's Hospitals Neonatal Database (CHND), we identified neonates with HIE at our center born ≥36 weeks' gestation from 2010 to 2016. Those with major congenital anomalies were omitted. Infants uniformly received therapeutic hypothermia for 72 hours unless death occurred sooner. Inpatient vital signs and selected laboratory markers were collected from our institution's health informatics,

https://doi.org/10.1017/cts.2017.59 Published online by Cambridge University Press