ECoG data. The proposed method may be potentially used for identification of seeds in the motor cortex for subsequent network analysis and further studies may delineate its potential use in the operating room.

2296

Functional analysis of the cutaneous microbiome in psoriatic disease

Di Yan, Hsin-Wen Chang, Rasnik Singh, Kevin Lai, Kristina Lee, Ladan Afifi, Xueyan Lu, Derya Ucmak and Susan Lynch
University of California, San Francisco, CA, USA

OBJECTIVES/SPECIFIC AIMS: Psoriasis is one of the most common inflammatory diseases of the skin, affecting about 2%–3% of the US population. Despite its high prevalence, its pathogenesis remains poorly understood. The ability of the microbiome to modify host immunity and metabolism suggests that it may contribute to the development of psoriasis and its cardiometabolic comorbidities. This study aims to characterize the psoriatic skin microbiome and understand the functional role that these bacteria may play. METHODS/STUDY POPULATION: 16s rRNA sequencing of site-matched skin swabs from 8 psoriasis patients and 8 healthy controls was used to identify bacteria and determine their relative abundance and microbial community diversity in the sample. PICRUSt was used to infer the functional roles of the bacteria from 16s rRNA amplicon data. RESULTS/ANTICIPATED RESULTS: Lesional psoriasis skin had lower α diversity ($p = 0.04$), less Actinobacteria ($p = 0.0001$), but higher Firmicutes ($p = 0.009$) compared with controls. At the genus level, lesional skin had more Alloiococcus ($p = 0.01$) and Aerococcus ($p = 0.01$) and demonstrated a trend towards lower Propionibacterium ($p = 0.08$) and higher Gallicola ($p = 0.09$) compared to controls. Interestingly, Alloiococcus ($p = 0.003$) and Gallicola ($p = 0.04$) were also higher in nonlesional skin compared with controls. Furthermore, lesional and nonlesional skin shared an increased abundance of *Acinetobacter* sp., *Staphylococcus pettenkoferi*, and *Streptococcus* sp., relative to controls. Lesional and nonlesional psoriasis skin did not differ significantly in microbiome composition. Predictive functional analysis revealed that both the healthy and psoriatic skin microbiome were enriched with bacteria capable of amino acid and carbohydrate metabolism suggest these functions might have a general role in host-microbe interaction. DISCUSSION/SIGNIFICANCE OF IMPACT: These data reveal intriguing differences in the cutaneous microbiome of psoriatic individuals and healthy controls and suggest that bacterial metabolism may play an important role in host-microbe interaction.

2327

Prescription opioid dependence in Western New York: Using data analytics to find an answer to the opioid epidemic

Shyamashree Sinha, Gale Burstein, Kenneth E. Leonard, Timothy Murphy and Peter Elkin
University at Buffalo, State University of New York, Buffalo, NY, USA

OBJECTIVES/SPECIFIC AIMS: Dependence and abuse of prescription opioid pain medication has substantially increased over the last decade. The consistent rise in opioid dependence contributes to the rising prescription drug overdose deaths over the last decade. The study of the distribution and determinants of opioid dependence among patients who are treated with chronic pain medications prescribed by their healthcare providers would aid in answering some key questions about potential abuse and overdose on opioids. The descriptive epidemiology of opioid dependence would help in identifying the vulnerable age group, race, ethnicity, and type of opioid pain medications that more commonly result in dependence. METHODS/STUDY POPULATION: We implemented an Observational Medical Outcomes Partnership/Observational Health Data Sciences and Informatics (OMOP/OHDSI) database, to hold structured EHR data from our Allscripts patient records. We also created a high-throughput phenotyping, natural language processing system that can parse 7,000,000 clinical notes in 1.5 hours. This runs as a web service and provides a modular component based NLP system. After the full semantic parse, we match the content against any number of ontologies. For each match we tag it as either a positive, negative, or uncertain assertion. We then perform automated compositional expressions. The codes are stored in a Berkley database (BDB) NOSQL database and the compositional expressions are stored in Neo4J (a graph database) and Graph DB (a triple store). This flexibility allows rapid retrieval of complex questions in real time. The High-Throughput Phenotyping (HTP) Natural Language Processing (NLP) Subsystem (HTP-NLP) is software that produces, given biomedical text, semantic annotations of the text. The semantic annotations identify conceptual entities—their attributes, the relations they have with other entities and the events they participate in, as expressed in the input text. The conceptual entities, relations, attributes, and events identified are specified by various knowledge representations (KRs) as documented in Coding Sources. Examples of coding sources are medical terminologies [eg, SNOMED CT, RxNorm, LOINC and open biomedical ontologies (OBO) foundry ontologies, eg, gene ontology (GO), functional model of anatomy, OBI, and others]. The annotation results may be displayed or output in formats suitable for further processing. Entity identified is assigned a truth value from 0 to 1. Values from the text are assigned to entities from ontologies such as SNOMED CT. The retrospective analysis of EHR data from local clinic patients was performed using queries on the problem list, demographic data, and medication list of all the patients in the database. The OMOP/OHDSI database was collected from Allscripts EHRs from 2010 to 2015. This common data model helps in the systematic analysis of disparate observational databases of clinic records from the primary care and family medicine clinics in Western New York region. The database contained 212,343 patient records that were parsed and deidentified. Specific research IDs were assigned to each of the patient records and stored in a secure firewall device for data analytics. The entire 212,343 records were queried for opioid dependence from the ICD-9 and 10 diagnostic codes and SNOMED CT codes mapped to both the clinical notes and the problem list for each patient based on the mapped ICD and SNOMED CT codes. In total, 1356 patients were identified as to having opioid dependence. The records were stratified into 7 age groups from age 18 to 28 and ending with age 79–89 years. RESULTS/ANTICIPATED RESULTS: Of the 212,343 patients in the database 1356 patients revealed opioid dependence on the problem list, ICD9-10 codes and prescription opioid pain medication with or without Buprenorphine and Naloxone (Suboxone) in the medication list. The prevalence of opioid dependence in the clinic population was 0.64% (95% CI: 0.61%–0.67%) over a 5-year period. The 7,000,000 patient records generated 750,000,000 SNOMED CT codes (on average 107 codes per record). The highest numbers of opioid dependence were seen in the 29 to 38 years' age group. That comprised 39.38% (95% CI: 36.78%–41.98%) of the total opioid dependent population but accounted for only 2.03% of whole clinic population in this age group (95% CI: 1.86% to 2.2%). The subjects were then stratified by race and ethnicity. There were 1005 patients with opioid dependence, in the non-Hispanic population (total number 108,402). Among the White non-Hispanic or Latino population with opioid dependence, 41.33% (95% CI: 38.27%–44.39%) were 29–38 years old. The next common age group among the White Non-Hispanic opioid dependent subjects was 19–28 years, comprising of 22.48% (95% CI: 19.88%–25.08%) of the total number of White non-Hispanic or Latino opioid dependent population. Among the total clinic population Hispanics comprise 51.24%, but they comprise only 2.58% (95% CI: 1.74%–3.42%) of the total opioid dependent population. The non-Hispanic population comprise 51.05% of total clinic population while the percent of people who are opioid dependent is 83.26% (95% CI: 83.04%–83.48%) of the total 1356 opioid dependent population. DISCUSSION/SIGNIFICANCE OF IMPACT: The trends of opioid dependence among the clinic population in the study indicate that the prevalence is more in a certain section of the population. The predominance is among the non-Hispanic White population in the 19–38 years of age. The prevalence in younger age implies that the complications related to opioid dependence would be there for a longer duration of time. The prevalence of dependence in this clinic population would be rising if this trend continues. Interventions at curbing prescription opioid dependence is necessary for the vulnerable population. The findings suggest that a broad based approach is necessary to address this problem. The distribution of opioid dependence in this patient population indicate the need for special attention to these specific age group and race ethnicities. The young age of many of the addicted patients demonstrate the risks of legitimate opioid prescriptions in leading this age group towards addiction and implies the need for routine screening for substance abuse. The evidence of complications of opioid overdose among long-term opioid users and risk of abuse with other agents including illicit agents makes the need for an approach that uses real-time interventions in addition to effect long-term improvement in addiction rates. A potentially cost-effective approach to implement monitoring programs and clinical decision support tools would be to develop inter operable linkage from the EHRs to the state Department of Healths' prescription monitoring programs.

2354

Pioneering the pathway with big data to eliminate hepatitis C viral infection (EHCV)

Dawn A Fishbein, Ian Brooks, Emanuel Villa Baca, Ozgur Ozmen, Mallikarjun Shankar, Gil Weigand, Kristina Thiagarajan, Randy Estes, Alex Geboy, Hala Deeb, Mamta Jain and Lesley Miller
Georgetown - Howard Universities, Washington, DC, USA

OBJECTIVES/SPECIFIC AIMS: Hepatitis C viral (HCV) infections are rising significantly both in young adults and as newly diagnosed cases in "baby boomers." New HCV therapeutics cure over 95% of cases, and a call has been

https://doi.org/10.1017/cts.2017.70 Published online by Cambridge University Press

made for elimination of the epidemic by 2030; yet major HCV cascade of care (CoC) barriers exist. We secured CTSA pilot funding to obtain preliminary data for an innovative clinical trial utilizing big data modeling toward HCV elimination. METHODS/STUDY POPULATION: Our pilot work has developed a coordinated, real-time clinical data management process across 3 major CTSA affiliated hospital systems (MedStar Health, Emory-Grady, and UT-Southwestern), and additional data will be obtained from a pragmatic clinical trial. Electronic medical records data will be mapped to the OHDSI model, securely transmitted to Oak Ridge National Laboratory, Knoxville, TN and exposed to integrated data, analytics, modeling and simulation (IDAMS). RESULTS/ANTICIPATED RESULTS: Our U01 CTSA application proposes that HCV-IDAMS will model modifications to the established HCV CoC at community and population levels and thus simulate future outcomes. As data volume increases, system knowledge will expand and recursive applications of IDAMS will increase the accuracy of our models. This will reveal real-world reactions contingent upon population dynamics and composition, geographies, and local applications of the HCV CoC. DISCUSSION/SIGNIFICANCE OF IMPACT: Only an innovative, integrated approach harnessing pragmatic clinical data, big data and supercomputing power can create a realistic model toward HCV elimination.

2356

openSESAME: a "search engine" for discovering drug-disease connections by leveraging publicly available high-throughput experimental data

Adam C. Gower, Avrum Spira and Marc E. Lenburg

OBJECTIVES/SPECIFIC AIMS: Microarray technology has produced large volumes of gene expression data profiling differences in gene expression in a vast array of conditions, much of which is publicly available. Methods to query these data for similarities in patterns of gene regulation are limited to comparisons between preannotated groups. In response, we developed openSESAME to find experiments where a set of genes is similarly coregulated without regard to experimental design. An important application of openSESAME is drug repositioning: if a pattern associated with disease is reversed by a given drug, the drug might target disease-related processes. METHODS/STUDY POPULATION: Experiments from the Gene Expression Omnibus (GEO) were normalized, signature-association (SA) scores computed for each sample, experiments assigned enrichment scores, and ANOVAs used to assign significance to experimental variables automatically extracted from GEO. SA scores were also generated for hundreds of publicly available signatures, and pairwise correlations used to create a relevance network. RESULTS/ANTICIPATED RESULTS: Using signatures of estrogen and p63, we recovered relevant experimental variables, and with the network approach, we recovered previously reported associations between disease states and/or drug treatments. DISCUSSION/SIGNIFICANCE OF IMPACT: openSESAME has the potential to illuminate "dark data" and discover novel relationships between drugs and diseases on the basis of common patterns of differential gene expression.

2378

A scientometric analysis of CTSA collaboration and impact

Kristi Holmes, Ehsan Mohammadi, Karen Gutzman, Pamela Shaw and Donald Lloyd-Jones

OBJECTIVES/SPECIFIC AIMS: Translational science supports the continuum of activities from early-stage bench research to implementation of discoveries for better and faster treatments to more patients. Past studies have attempted to clarify our understanding of the spectrum of translational research by categorizing the activities into stages ranging from T0 to T4 using explanatory definitions. Unfortunately, this approach is often vague and relies on a process of manual classification and binning of research publications into predetermined categories. This study aims to provide a big-picture analysis of clinical and translational science (CTS) based on an in-depth analysis of the entire corpus of publications resulting from research funded by Clinical and Translational Science Awards (CTSA) U54 awards (through 2016). METHODS/STUDY POPULATION: We harvested bibliographic metadata from all papers that cited any of the U54 award numbers since the inception of the CTSA program to the most recent award announcement. Natural language processing techniques were used to create term co-occurrence networks based on English-language textual data. Relevant and nonrelevant terms were distinguished algorithmically and processed accordingly to provide the clustered visualization. RESULTS/ANTICIPATED RESULTS: With this approach, we uncovered 6 natural clustered areas of emphasis of published CTS research, the evolution of specific concepts through time, and gained a better understanding of their relative impact as demonstrated by citations. We performed additional analyses including discipline-specific impact assessment; identification of categories of excellence relating to both productivity and citations; characteristics of collaborative networks such as organizational, industry, and international collaborations and network dynamics; and resulting global impact of the CTSA program. DISCUSSION/SIGNIFICANCE OF IMPACT: Ultimately we gained a clearer understanding of the CTSA program, its evolution through scholarly publications, and key areas of impact of the program using computational, data-driven evaluation methods.

2412

Predicting response to hemodynamic interventions in the ICU using recurrent neural networks

Julian Genkins and Thomas A. Lasko

OBJECTIVES/SPECIFIC AIMS: Our goal is to explore the value of learning algorithms to improve both the efficiency and accuracy of a clinician undertaking the cognitive task of selecting the best resuscitative intervention for a hemodynamically unstable patient in the ICU. Machine learning is an ideal discipline to solve this problem. The ICU is a data rich environment, however there is significant uncertainty regarding the interdependency of this data. Experts consistently struggle to develop deterministic models of the underlying forces driving hemodynamic perturbations and intervention responsiveness. Machine learning, especially deep learning, assumes no correlation between inputs. Deep architectures disentangle these high-level relationships through exposure to abundant, diverse data sets such as those used in this project, obviating the need to manually explore confounding interactions. METHODS/STUDY POPULATION: We are using the "Medical Information Mart for Intensive Care" (MIMIC-III) database for this project. MIMIC-III is a large, single-center database comprising information relating to patients admitted to critical care units at Beth Israel Deaconess Medical Center, a large tertiary care hospital, from 2001 to 2012. It contains data associated with 38,597 distinct adult patients and 53,423 distinct hospital admissions for those patients, with a mean of 4579 charted observations and 380 laboratory measurements available for each hospital admission. Classes of data in the MIMIC-III are varied and include billing, intervention, laboratory, medication, and physiologic data among others. In addition to training an RNN in the task of predicting hemodynamic states, we will also attempt to train 2 additional models on the same data—a multidimensional linear regression and a nonsequence-oriented deep neural network. For each of these models we will measure accuracy using root mean squared error (RMSE) and mean absolute error (MAE) to provide scale-dependent measurements of accuracy. RESULTS/ANTICIPATED RESULTS: Our results will be reported in 2 primary categories: numerical accuracy of the RNN model and applicability, utility, and accuracy in a live clinical setting. The use of RNNs in biomedical informatics, and in general, is a relatively new phenomenon. This means that the body of literature which could provide a basis for our expected results is limited. Because of this we have chosen staged goals in assessing our model. First, we hope to achieve a model that reliably predicts the direction of response. Being able to answer only the question of how a patient will respond—will they move toward or away from our therapeutic goal—is as good as existing prediction methods. It is well established in the literature that, by almost any metric, ~50% of hemodynamically unstable patients respond to a fluid challenge. If we are within 10% of this average (40%–60% respond), then we can be confident in the accuracy of our model in predicting direction. Upon achieving this, we will then measure accurate prediction of response magnitude. To this affect, we hope to achieve an RMSE <10% between our test data and corresponding predicted output before proceeding further. In addition to numeric accuracy, we acknowledge that a plan for practical, clinical validation is needed before utilizing this tool in a clinical environment. Such validation will require 3 separate components. First, numeric accuracy will need to be determined again as compared with prospective data on actual patients in the ICU. This step is critical to prove that no information leakage from target data back to input data occurred during training. Second, there must be a comparison to existing prediction methods, such as the passive leg raise in combination with measurement of cardiac output to predict volume responsiveness. Finally, we must measure the cost to the clinician of implementing our model in an ICU, specifically how it impacts their time to accomplish the task of selecting an intervention for the hemodynamically unstable patient. However, these tasks are beyond the scope of this project and will be left for later investigations. DISCUSSION/SIGNIFICANCE OF IMPACT: If we are successful, this study will provide the first step toward a data-driven model for predicting patient responsiveness to a given hemodynamic intervention or collection of interventions. As compared with current

https://doi.org/10.1017/cts.2017.70 Published online by Cambridge University Press