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 open-SESAME 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 hemodynami-cally 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 hemody-namic intervention or collection of interventions. As compared with current