K3 – Observational Data for Biomedical Discovery

Dr Nicholas TATONETTI, *PhD*
Herbert Irving Assistant Professor of Biomedical Informatics,
Departments of Biomedical Informatics, Systems Biology, and Medicine,
Columbia University, United States

Dr Nicholas Tatonetti is assistant professor of biomedical informatics in the Departments of Biomedical Informatics, Systems Biology, and Medicine and is Director of Clinical Informatics at the Institute for Genomic Medicine at Columbia University. He received his PhD from Stanford University where he focused on the development of novel statistical and computational methods for observational data mining. He applied these methods to drug safety surveillance and the discovery of dangerous drug-drug interactions. His lab at Columbia is focused on expanding upon his previous work in detecting, explaining, and validating drug effects and drug interactions from large-scale observational data. Widely published in both clinical and bioinformatics, Dr Tatonetti is passionate about the integration of hospital data (stored in the electronic health records) and high-dimensional biological data (captured using next-generation sequencing, high-throughput screening, and other "omics" technologies). Dr Tatonetti has been featured by the New York Times, Genome Web, and Science Careers. His work has been picked up by the mainstream and scientific media and generated thousands of news articles.

Observation is the starting point of discovery. Based on observations, scientists form hypotheses that are then tested and evaluated. In the information-age, trillions of observations are being made and recorded every day – from online social interactions to the emergency room visit. With so much data readily available, generating hypotheses using a single scientist's mind is no longer sufficient. Instead, we must turn to computational algorithms to "mine" for new hypotheses and relationships for us. Data mining is an emerging field dedicated to training algorithms to recognize patterns in enormous sets of data to automatically identify new hypotheses.

In this talk, I will discuss how we use data mining algorithms to identify unexpected effects of drugs used singly and in combination with other drugs. Drug-drug interactions (DDIs) are an important and understudied public health concern. DDIs are difficult and expensive to study because of the complex combinatorial nature to their investigation. I have developed new methods for mining clinical data and then discovered and validated two previously unknown novel drug-drug interactions. In the first, published in 2011, I found that paroxetine (selective serotonin reuptake inhibitor) and pravastatin (HMG-CoA reductase inhibitor) together cause hyperglycemia. In the second, published in 2016, I found that ceftriaxone (cephalosporin antibiotic) and lansoprazole (proton-pump inhibitor) are associated with prolonged QT syndrome (LQTS). In both cases, I used a combination of data mining and laboratory experiments to discover and validate the new DDI. First, I mined the FDA's Adverse Event Reporting Systems for signs of drug-drug interactions using supervised machine learning methods. These algorithms were trained to recognize safety signals using single-drug effect and then used to identify when pairs of drugs mimic these effects. I then used electronic health records to corroborate the putative DDIs. In each case, using commonly collected data during clinical practice. For paroxetine and pravastatin we looked at blood glucose values before and after drug combination exposure and, for ceftriaxone and lansoprazole, we used the electrocardiograms recorded post drug combination therapy. I then validated the findings prospectively using model systems (a diabetic mouse model for diabetes and a single cell electrophysiology experiment for LQTS). These studies are the first to use big patient data to discovery a drug interaction and then use prospective experiments to validate the findings.

Using integrative informatics methods, we are able to discover drug-drug interactions that no one considered possible before. In many cases these experiments can be executed in high-throughput and by robotic systems, with the ultimate goal of automating the scientific method.