

Mt. Sinai Researchers Use EMRs to Identify New Subpopulations of Type 2 Diabetes Patients

Oct 29, 2015 | [a GenomeWeb staff reporter](#)

NEW YORK (GenomeWeb) – Researchers from the Icahn School of Medicine at Mt. Sinai have used electronic medical record data from patients with type 2 diabetes to further stratify the population into three distinct subgroups.

The team performed genome-wide association studies using genetic data that had been collected and stored in patients' EMRs. The researchers said the study, published this week in *Science Translational Medicine*, shows the potential for to tailor diagnosis and treatment.

"Our approach demonstrates the potential to unlock clinically meaningful patient population subgroups from the wealth of information that is accumulating in electronic medical record systems," Ronald Tamler, co-author of the study and Director of the Mount Sinai Clinical Diabetes Institute, said in statement. "The unique genetic component of this study yielded high-priority variants for a follow-up study in patients with type 2 diabetes."

First, the researchers used a topology-based approach to computationally sort over 11,000 patients based on clinical similarities across multiple dimensions, such as laboratory tests. This resulted into two groups. One consisted of 3,889 patients that were enriched for endocrine and metabolic diseases, immunity disorders, infectious disease, mental illness, diseases of the circulatory and genitourinary systems, and other symptoms of ill-defined health conditions. The second group of 7,321 patients was enriched for complications of pregnancy, respiratory diseases, and unclassified external causes of injury.

Next, the team applied an algorithm to define the type 2 diabetes phenotype, which identified 2,551 patients. They then rebuilt the topology analysis pipeline with only those patients. The rebuilt model included 73 clinical features, and resulted in three different subgroups consisting of 762, 617, and 1,096 type 2 diabetes patients. To make sure that the model was reproducible the researchers reapplied it on subsamples of patients and came up with an overall accuracy of 96 percent.

Looking at the clinical features from each subtype, the researchers found that the 762 patients in subtype 1 were the youngest and characterized by features commonly associated with type 2 diabetes like high body mass index, diabetic nephropathy, and diabetic retinopathy; the 617 patients in subtype 2 had the lowest average weight and were more likely to have cancer and cardiovascular diseases; and the 1,096 patients in subtype 3 had the strongest associations with cardiovascular diseases, neurological diseases, allergies, and HIV infections.

Doing a genome-wide association study, the researchers identified 1,279, 1,227, and 1,338 genetic variants specific to subtypes 1, 2, and 3, respectively, which corresponded to 425, 322, and 437 unique genes.

Many of those enriched genetic associations correlated strongly with the observed phenotypes. For instance, many of the genetic variants found in subtype 1 are known to be associated with diabetic neuropathy, while the variants in subtype 2 were often associated with cancer, and many of the variants found in subtype 3 were associated with mental disorders.

Overall, the team identified 27 gene-phenotype associations related to subtype 1, 25 related to subtype 2, and 28 related to subtype 3.

"Together, these results suggest that the current clinical definition of T2D subsumes more nuanced subtypes whose definition and recognition might inform important clinical distinctions," the authors wrote.

Although the authors acknowledge that the patient size of just over 2,551 is relatively small for a GWAS, it still demonstrates the power of correlating phenotypic and genetic data stored in EMRs and "can be extended toward the study of other complex, multifactorial diseases," the authors wrote.

"This project demonstrates the very real promise of precision medicine to improve healthcare by tailoring diagnosis and treatment to each patient, as well as by learning from each patient," Joel Dudley, senior author on the paper and director of biomedical informatics at the Icahn School of Medicine at Mount Sinai, said in a statement.