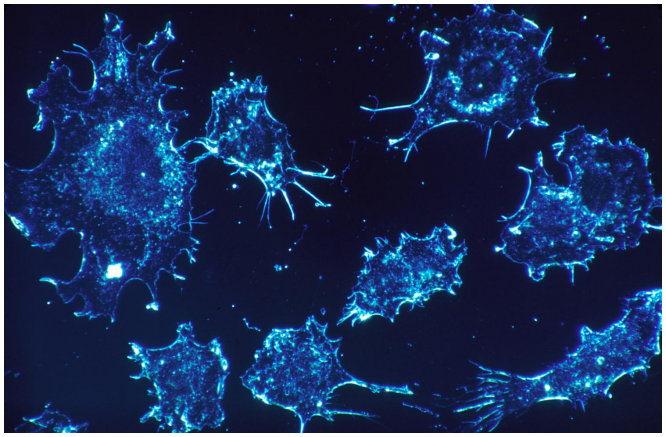
Polygenic scores to classify cancer risk

18 May 2018



Credit: CC0 Public Domain

Polygenic risk scores could be useful to stratify the risk of several cancers among patients in medical centers, allowing for the potential discovery of new associations between genes, disease and secondary effects, according to a University of Michigan study.

Researchers at U-M's School of Public Health conducted a phenome-wide association study in 28,260 unrelated, genotyped patients of recent European ancestry to evaluate whether polygenic risk scores for common cancers are associated with multiple phenotypes. The study used data from participants' electronic health records.

The results demonstrate that polygenic risk scores, a summary score constructed based on results obtained from large population-based genome-wide association studies, can be potentially useful for cancer risk stratification among patients in an academic medical center.

"This is just the tip of the iceberg. It shows the potential and the challenges with using phenome-wide associations," said Bhramar Mukherjee, the John D. Kalbfleisch Collegiate Professor of Biostatistics and professor of epidemiology at U-M's School of Public Health and associate director for Cancer Control and Population Sciences at the Rogel Cancer Center.

"Looking at the data, it was surprising to me how logical the secondary diagnosis associations with the risk score were. When you look at the association plots of the risk scores against the phenome, I expected it to be a lot more noisy. I thought that there would be many spurious associations with random disease codes on the phenome level. It was also striking how results from population-based studies were reproduced using data from electronic health records, a database not ideally designed for specific research questions and is certainly not a population-based sample."

Lead author Lars Fritsche, assistant research scientist in the Department of Biostatistics at U-M's School of Public Health, agreed. For example, he said, a polygenic risk score for squamous cell carcinoma—a common skin cancer form—showed a strong association for various forms of skin cancer, but also with actinic keratosis, a potentially precancerous skin lesion. Studying the sequence of diagnoses of the available patients' electronic health records can help to understand relationships like these.

Mukherjee said investigators will expand the analytic model to other lab tests and biomarkers to see if new associations with diseases across the phenome can be found there.

"What we are trying to do is to understand the pattern of concurrent diagnosis for a patient and determine if a cluster of diagnoses in the past is predictive of a disease state in the future," she said. "The temporal sequence of disease diagnoses and lab results along with genetic data are key to this scientific quest."

Researchers also plan to re-contact the initial subjects to collect data on their behavioral and environmental risk factors.

"There's enormous potential in terms of collecting behavioral data and integrating with molecular data," Mukherjee said. "Family history, stress data, anxiety data, sleep quality data, mental health data, so that you have a much more integrated vision of what is happening in a person's life."

"For future precision health applications," Fritsche said, "it will be important to understand the challenges of such complex big data and to provide reliable approaches that can efficiently explore them, especially because we expect the amount of data to steadily increase over time."

The study, "Association of Polygenic Risk Scores for Multiple Cancers in a Phenome-wide Study: Results from the Michigan Genomics Initiative," will be published in the *American Journal of Human Genetics*.

More information: Lars G. Fritsche et al. Association of Polygenic Risk Scores for Multiple Cancers in a Phenome-wide Study: Results from The Michigan Genomics Initiative, *American Journal of Human Genetics* (2017). DOI: 10.1101/205021