Can our DNA predict our illnesses?

Scientists are working to understand how each person's unique DNA profile suggests which diseases might affect them. Because our DNA profiles provide a recipe for shaping many elements of our biology, such as our height and eye color, scientists have sought to answer the question of how to predict a person's risk of a disease, like lung cancer, from their DNA profile. While current approaches have had limited success, future research will enable clinical and public health applications of genetics data to improve health, efficiently screen for diseases, and promote longevity.

The human genome sequencing project of the 1990s and early 2000s ushered in a scientific effort to identify genes that contribute to different diseases, including heart disease, lung cancer, and schizophrenia. While any two humans share more than 99% of their DNA profiles, the remaining fraction of a percent of DNA is what makes each person unique. This might seem like a small percentage, but, with more than 3 billion letters constituting each person's DNA profile, we differ from other humans at more than 1 million letters.

Large-scale studies have enrolled millions of people around the world to work towards the goal of finding genes that predispose individuals to a disease like lung cancer. Each study collects a tissue sample, such as a cheek swab, saliva collection, or blood draw, from each

volunteer in the study. The collected tissue sample contains human cells, and, from those cells, scientists can isolate a person's DNA, the alphabet that provides the recipe for each person's biology.

A biotechnology called "genotyping" takes as inputs DNA from cells and provides scientists with each volunteer's DNA profile, a sequence of letters that specifies the chemical structure of their DNA.

Scientists then use statistical association tests to identify genes that contribute to a person's chances of getting a disease like lung cancer. The statistical association tests work by comparing the DNA profiles of people with lung cancer to the DNA profiles of people without lung cancer. Each difference in DNA profile between the two groups, 1) people with lung cancer and 2) people without lung cancer, points to a gene that contributes to lung cancer.

By looking across people's 3 billion letters in their DNA profiles, scientists can identify many genes that influence lung cancer development.

It's important to point out that, in nearly all cases, no single DNA profile determines whether a person will get lung cancer; however, a person's chance of getting lung cancer increases if they have a DNA profile that matches those of people who already have lung cancer. In fact, scientists have long recognized that environmental factors, including smoking behaviors and air pollutants, also contribute to lung cancer development. This partially explains why some people who smoke regularly don't get lung cancer, while others who never smoke can still get lung cancer. Ultimately, it is the interplay between DNA profile and environmental exposures that influence lung cancer risk.

A “polygenic risk score” for lung cancer is a number calculated from a person’s DNA profile. It provides a measure of lung cancer risk for that person, with higher polygenic risk scores corresponding to higher lung cancer risks. Unlike traditional measures of lung cancer risk, like smoking history or family history of lung cancer, a polygenic risk score makes uses of only a person’s DNA profile. Thus, it provides a lung cancer risk assessment that can complement that from traditional risk factors.

Scientists now have their sights set on creating lung cancer risk predictors that combine traditional risk factors with polygenic risk scores. Researchers expect these new risk predictors to outperform traditional risk factors alone because they better capture both environmental exposures and genetics, which both affect lung cancer development.

An advance like this new, integrated lung cancer risk predictor would benefit both medicine and public health by providing a numerical summary of each person’s risk of getting lung cancer. With this information, public health interventions, such as routine screening programs for those at highest risk, can be developed to diminish global lung cancer burden.