Teams of microbes are at work in our bodies. Researchers have figured out what they're up to.

December 11 2019

Credit: CC0 Public Domain

In the last decade, scientists have made tremendous progress in

understanding that groups of bacteria and viruses that naturally coexist throughout the human body play an important role in some vital functions like digestion, metabolism and even fighting off diseases. But understanding just how they do it remains a question.

Researchers from Drexel University are hoping to help answer that question through a clever combination of high-throughput genetic sequencing and natural language processing computer algorithms. Their research, which was recently published in the journal *PLOS ONE*, reports a new method of analyzing the codes found in RNA that can delineate human microbial communities and reveal how they operate.

Much of the research on the human microbial environment—or microbiome—has focused on identifying all of the different microbe species. And the nascent development of treatments for microbiota-linked maladies operates under the idea that imbalances or deviations in the microbiome are the source of health problems, such as indigestion or Crohn's disease.

But to properly correct these imbalances it's important for scientists to have a broader understanding of microbial communities as they exist—both in the afflicted areas and throughout the entire body.

"We are really just beginning to scrape the surface of understanding the health effects of microbiota," said Gail Rosen, Ph.D., an associate professor in Drexel's College of Engineering, who was an author of the paper. "In many ways scientists have jumped into this work without having a full picture of what these microbial communities look like, how prevalent they are and how their internal configuration affects their immediate environment within the human body."

Rosen heads Drexel's Center for Biological Discovery from Big Data, a group of researchers that has been applying algorithms and machine

learning to help decipher massive amounts of genetic sequencing information that has become available in the last handful of years. Their work and similar efforts around the world have moved microbiology and genetics research from the wet lab to the data center—creating a computational approach to studying organism interactions and evolution, called metagenomics.

In this type of research, a scan of a genetic material sample—DNA or RNA—can be interpreted to reveal the organisms that are likely present. The method presented by Rosen's group takes that one step farther by analyzing the genetic code to spot recurring patterns, an indication that certain groups of organisms—microbes in this case—are found together so frequently that it's not a coincidence.

"We call this method 'themetagenomics,' because we are looking for recurring themes in microbiomes that are indicators of co-occurring groups of microbes," Rosen said. "There are thousands of species of microbes living in the body, so if you think about all the permutations of groupings that could exist you can imagine what a daunting task it is to determine which of them are living in community with each other. Our method puts a pattern-spotting algorithm to work on the task, which saves a tremendous amount of time and eliminates some guesswork."

Current methods for studying microbiota, gut bacteria for example, take a sample from an area of the body and then look at the genetic material that's present. This process inherently lacks important context, according to the authors.

"It's impossible to really understand what microbe communities are doing if we don't first understand the extent of the community and how frequently and where else they might be occurring in the body," said Steve Woloszynek, Ph.D., and MD trainee in Drexel's College of Medicine and co-author of the paper. "In other words, it's hard to

develop treatments to promote natural microbial coexistence if their 'natural state' is not yet known."

Obtaining a full map of microbial communities, using themetagenomics, allows researchers to observe how they change over time—both in healthy people and those suffering from diseases. And observing the difference between the two provides clues to the function of the community, as well as illuminating the configuration of microbe species that enables it.

"Most metagenomics methods just tell you which microbes are abundant—therefore likely important—but they don't really tell you much about how each species is supporting other community members," Rosen said. "With our method you get a picture of the configuration of the community—for example, it may have E. coli and B. fragilis as the most abundant microbes and in pretty equal numbers—which may indicate that they're cross-feeding. Another community may have B. fragilis as the most abundant microbe, with many other microbes in equal, but lower, numbers—which could indicate that they are feeding off whatever B. fragilis is making, without any cooperation."

One of the ultimate goals of analyzing human microbiota is to use the presence of certain microbe communities as indicators to identify diseases like Crohn's or even specific types of cancer. To test their new method, the Drexel researchers put it up against similar topic modeling procedures that diagnose Crohn's and mouth cancer by measuring the relative abundance of certain genetic sequences.

The themetagenomics method proved to be just as accurate predicting the diseases, but it does it much faster than the other topic modeling methods—minutes versus days—and it also teases out how each microbe species in the indicator community may contribute to the severity of the disease. With this level of granularity, researchers will be able to home

in on particular genetic groupings when developing targeted treatments.

The group has made its [themetagenomics analysis tools publicly available](#) in hopes of speeding progress toward cures and treatments for these maladies.

"It's very early right now, but the more that we understand about how the microbiome functions—even just knowing that groups may be acting together—then we can look into the metabolic pathways of these groups and intervene or control them, thus paving the way for drug development and therapy research," Rosen said.

More information: Stephen Woloszynek et al, Exploring thematic structure and predicted functionality of 16S rRNA amplicon data, *PLOS ONE* (2019). [DOI: 10.1371/journal.pone.0219235](#)

Provided by Drexel University