Northeastern University

Understanding Intrapatient SARS CoV-2 Variants

For assessment of patient susceptibility and disease severity, and for development of informed prevention and treatment strategies

Jonathan Tilly

Dori Woods

Konstantin Khrapko

Background: Without a better understanding of the biology of SARS-CoV-2, and how it evolves and progresses in its human hosts, COVID-19 researchers are caught in a game of chance, and trial and error, as to whether their hard-fought efforts will pay off. With SARS-CoV-2, as with other RNA viruses, many fundamental questions remain unanswered, with three of the most critical being:

1. Why is COVID-19 lethal for some who are infected, while causing no noticeable symptoms in others who are infected?
2. Are patients who have recovered from a SARS-CoV-2 infection protected against future infection?
3. How do we, as a society, prepare for the prospects of a re-emergence of COVID-19 in the future?

To begin answering these questions, it is essential that we learn much more about the behavior of the SARS-CoV-2 virus. For example: To what extent do genetic mutations lead to viral variants (also known as viral *"quasispecies"*) within individually infected patients? Are some quasispecies of SARS-CoV-2 more pathogenic or more contagious than others? What are the patterns of viral genetic variation and progression over time, within a single patient and across the pandemic?

Innovation: At Northeastern University, a team of our biologists has developed a technology known as **long molecule UMI-driven-consensus sequencing** (LUCS), which is positioned to readily answer these questions, and drive forward our understanding of SARS-CoV-2.

The crucial contribution of LUCS is that it enables scientists to accurately sequence, classify and track genetic variants of SARS-CoV-2 that emerge within a given patient. In so doing, the LUCS platform overcomes a significant limitation of commercial approaches to large-scale viral genome sequencing–namely, at the

resolution required for quasispecies analyses, these approaches do not reliably distinguish the true variants in a viral sample from artefact variants (errors) introduced by current third-generation sequencing technologies.

Through LUCS, Northeastern investigators **Jonathan Tilly, Dori Woods, and Konstantin Khrapko** present an ingenious solution to the current error-prone methodologies. By integrating core principles of molecular and computational biology with standard sequencing technologies, the LUCS platform effectively tracks the accumulation of sequencing errors accrued through these methods and then computationally deletes them, in order to identify a true set of variants per viral genome (LUCS is further described in US Patent Application 2018037154).

Impact: Aided by the LUCS platform, our team at Northeastern is uniquely poised to catapult the giant effort needed for accurate sequence analysis of viral variation during the course of infection, on a patient-by-patient basis. They also aim to create a centralized, open-access database of accurate SARS-CoV-2 variant information, in conjunction with corresponding clinical data, thereby providing a vital resource for the larger community of COVID-19 researchers.

Early pilot data from the Northeastern team points to a much less stable viral genome, with a wider range of mutations, than currently believed–underscoring the urgent need to conduct these analyses on a broad scale. Such an effort will help define the transmission and pathogenicity of SARS-CoV-2, as well as inform rational treatment and prevention strategies for vulnerable populations around the world. It will also help scientists and public health experts determine the likelihood of prolonged immunity among recovered patients, as well as the possibilities of COVID-19 resurgence in communities that are experiencing recovery.

By helping researchers characterize how viral quasispecies emerge and infect individuals, and possibly further evolve over the course of infection, LUCS represents a powerful new weapon in the war against COVID-19, as well as health complications caused by other RNA viruses that may emerge in the future.

The crucial contribution of LUCS is that it enables scientists to accurately sequence, classify and track genetic variants of SARS-CoV-2 that emerge within a given patient.