

Free Bioinformatics Tool From Nference, Janssen Uncovers New COVID-19 Insights

Apr 10, 2020 | Neil Versel

CHICAGO – A new research portal from software developer Nference and Johnson & Johnson's Janssen Pharmaceutica has identified several new expression mechanisms for COVID-19 and the novel SARS-CoV-2 coronavirus that causes the deadly disease.

According to a non-peer-reviewed [paper](#) released on prepublication server BioRxiv, Janssen and Nference have been able to learn more about SARS-CoV-2 transmission through ACE2 receptors in the tongue, the olfactory bulb of the nasal cavity, and the airways and lungs. They also have found evidence that the gut is the "dominant expressor" of the viral receptor, as maturation of enterocytes of the small intestine and colon increases the expression of all three known receptors for human coronaviruses, including those linked to the common cold.

Specifically, they found that ACE2 is significantly expressed in keratinocytes in the tongue and in some olfactory epithelial cells, which may help to explain reports that loss of taste and smell could be early indicators of COVID-19. They also identified club cells, ciliated cells, and type II pneumocytes as the likely targets of infection in the respiratory tract.

Additionally, Janssen and Nference researchers have observed more ACE2 RNA in the esophagus of people older than 60 than in younger patients, offering a possible explanation as to why the elderly seem to have higher COVID-19 mortality rates.

The technology portal, called the NferX platform, applies augmented intelligence to help researchers match single-cell RNA sequencing (scRNA-seq) data from more than 1 million individual cells found in public resources with a literature corpus of some 100 million documents, a collection that is growing daily with the rapid prepublication release of so many SARS-CoV-2 papers as researchers worldwide race to tame this pandemic.

Nference, which Cofounder and Chief Scientific Officer Venky Soundararajan said is "disease-agnostic" and "therapeutically agnostic," first started looking at scRNA-seq data in tumor-normal comparisons for cancer research. That approach has translated to the global search for answers on this novel coronavirus.

The partners claim that this platform is the first of its kind to apply AI to aggregated, synthesized, and structured scRNA-seq datasets to address a public health crisis.

NferX helps researchers triangulate knowledge and validate findings, though much of the validation will come in the form of future studies. "But it's a great quick real-time way of integrating so many different types of datasets to generate some hypotheses," said Najat Khan, chief operating officer for Janssen R&D Data Sciences, and global head of R&D strategy and operations for Janssen.

She said that the multiple tools contained in NferX will help Janssen and other researchers investigate both potential treatments and vaccines.

Cambridge, Massachusetts-based Nference is offering free access to the platform for academic researchers through academia.nferx.com. The vendor and Janssen said that NferX would give academia broad, searchable access to machine-computable data to help COVID-19 researchers decode molecular signatures of viral infection and gain insights into the pathways of disease transmission and progression.

The BioRxiv paper includes a how-to guide for using NferX. Nference is exploring ways of making the platform available to commercial enterprises, including sequencing companies and other biotech firms, according to Soundararajan.

In parallel, Nference also has developed a surveillance app as part of its [partnership with Mayo Clinic](#) that the Rochester, Minnesota-based institution is using to identify COVID-19 "hot spots" in its home state. An Nference spokesperson said that Mayo plans to launch the app nationally soon.

Mayo Clinic CEO Gianrico Farrugia appeared on CBS' ["Face the Nation"](#) on March 29 to discuss COVID-19 response and potential treatments. He mentioned the partnership with Nference, noting that the core Nference software provides the health system with real-time updates when someone tests positive for the SARS-CoV-2 coronavirus, and helps Mayo track hospital admissions among those with COVID-19 symptoms. Farrugia said that Mayo has already moved resources within Minnesota in response to potential hotspots.

Nference, founded in 2013, has spent the last seven years building augmented intelligence and deep-learning software for synthesizing biomedical knowledge from scientific, regulatory, and commercial literature to support drug discovery and development, drug life cycle management, and precision medicine.

The vendor applies natural language processing and other extraction techniques to make sense out of unstructured text to build longitudinal phenotypic profiles from electronic health records and clinical reports.

This data extraction is an integral part of NferX.

The NferX portal takes advantage of the neural networks Nference has built to derive insights from medical literature and "triangulate" those learnings with a growing dataset that currently includes 1 million cells from 25 human tissues, Soundararajan said. He said there are hundreds more coronavirus-related studies that Nference is analyzing, since new clinical, omics, and research data is showing up daily.

NferX is Nference's response to the March 16 ["call to action"](#) by the White House Office of Science and Technology Policy, to create a machine-readable dataset for understanding COVID-19.

As part of the call to action, the Allen Institute for AI, the Chan Zuckerberg Initiative, the Georgetown University Center for Security and Emerging Technology, Microsoft, and the US National Library of Medicine last month released a collection of literature called the COVID-19 Open Research Dataset, or CORD-19.

Nference is using CORD-19, but Soundararajan called that dataset merely a small subset of available biomedical knowledge on SARS-CoV-2 and COVID-19. Typical analytic techniques that look for, say, overlap between the novel coronavirus and Alzheimer's disease or heart failure may only search the keywords, which leaves out potentially helpful insights buried in unstructured text.

Soundararajan said that other efforts to understand scRNA-seq have not triangulated the sequencing data with the millions of available pieces of scientific literature that may not be structured in machine-readable form. The neural networks that power NferX help to uncover those insights, Soundararajan said.

"The inherent assumption is that the rest of the literature is not valuable for the specific application that one is focused on," he said. "[But] knowledge is interconnected. Diseases are interconnected, signaling pathways, across cell types, across diseases."

Prepublication platforms like BioRxiv are hosting hundreds of COVID-related articles. Other potential insights might be hidden in publicly available grant applications. New, expedited clinical trials are being announced, he noted.

"Company websites needs to be scraped. You have to scrape the entire published World Wide Web," Soundararajan. "To do that in real time takes an army of neural networks, which is what we are deploying."

Furthermore, there is an explosion of single-cell sequencing and omics information that not every research lab has the bioinformatics capacity to analyze.

It took just two weeks to get NferX up and running after the White House call to action, in part because Nference had an existing data science relationship with Janssen going back to 2018, according to Khan.

Before SARS-CoV-2 came along, the pharma company had been working with Nference mostly in oncology and immunology, but was able to pivot quickly to collaborating in infectious diseases and vaccine research. "The beauty of investing in areas like this, that when there's this terrible pandemic that happens, we can quickly leverage that investment and learning to apply it to the problem," Khan said.

Khan said that data science is central to how Janssen is trying to understand COVID-19 and infectious diseases because the company is bullish on the idea of real-world information to augment omics datasets.

"Applying data science is so important because then it helps us rapidly pivot and leverage those learnings for these urgent medical needs that will continue to come up," she said. "That investment in learning and building those capabilities is really what allowing us to now be able to rapidly pull things together and progress the learning in a very accelerated fashion that we couldn't have done before."