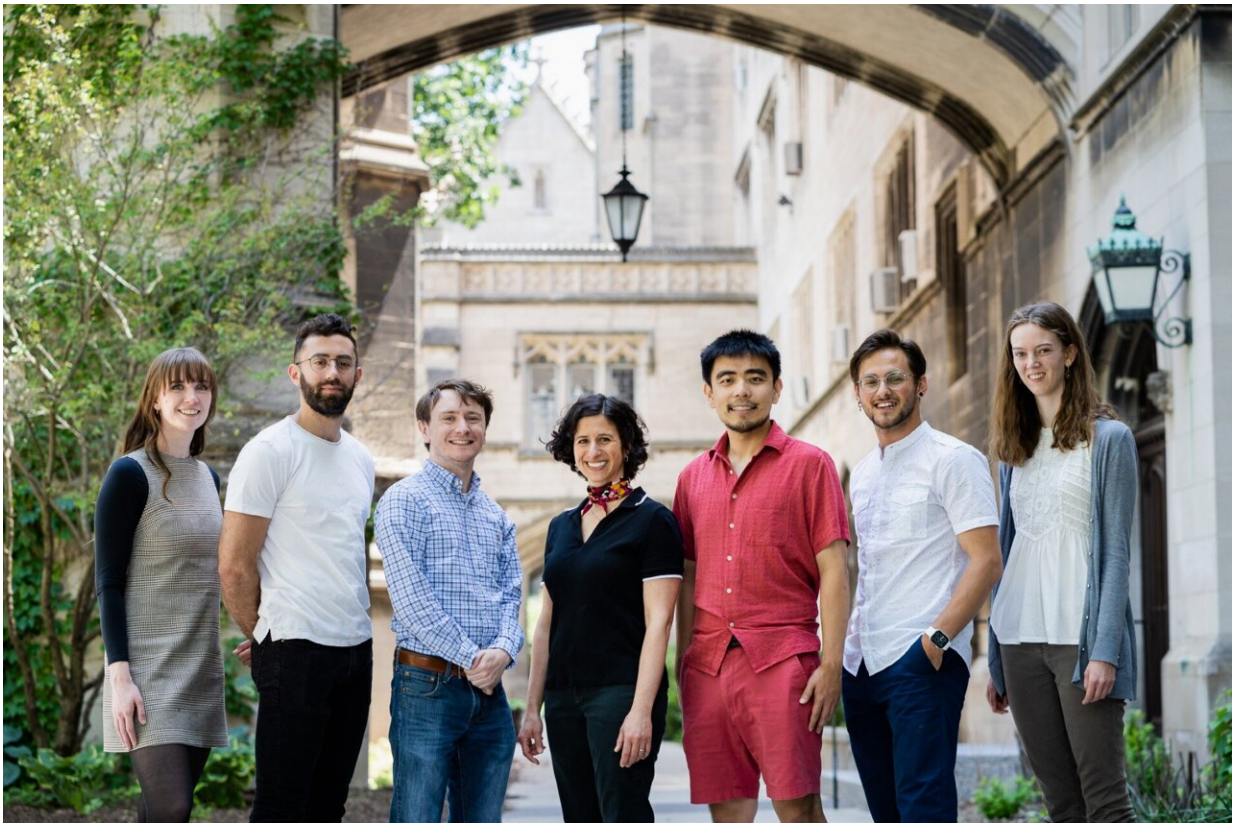# Getting dynamic information from static snapshots

April 27 2024, by Paul Dailing

Part of an interdisciplinary University of Chicago team behind a new method of using static data from single-cell RNA sequencing to study how cells and genes change over time. From left, Biophysics graduate student Hanna Hieromnimon, Pritzker School of Molecular Engineering graduate student Joey Federico, Computer Science graduate student Ryan Robinett, PME Asst. Prof. Samantha Riesenfeld, Chemistry graduate student and paper first author Cheng Frank Gao, Chemistry graduate student Joseph Sifakis and Biophysics graduate student Hope Anderson. Credit: Lorenzo Orecchia

Imagine predicting the exact finishing order of the Kentucky Derby from a still photograph taken 10 seconds into the race.

That challenge pales in comparison to what researchers face when using single-cell RNA-sequencing (scRNA-seq) to study how embryos develop, cells differentiate, cancers form, and the immune system reacts.

In a paper published today in *Proceedings of the National Academy of Sciences*, researchers from the UChicago Pritzker School of Molecular Engineering and the Chemistry Department have created TopicVelo, a powerful new method of using the static snapshots from scRNA-seq to study how cells and genes change over time.

The team took an interdisciplinary, collaborative approach, incorporating concepts from classical machine learning, computational biology, and chemistry.

"In terms of unsupervised machine learning, we use a very simple, well-established idea. And in terms of the transcriptional model we use, it's also a very simple, old idea. But when you put them together, they do something more powerful than you might expect," said PME Assistant Professor of Molecular Engineering and Medicine Samantha Riesenfeld, who wrote the paper with Chemistry Department Prof. Suriyanarayanan Vaikuntanathan and their joint student, UChicago Chemistry Ph.D. candidate Cheng Frank Gao.

The trouble with pseudotime

Researchers use scRNA-seq to get measurements that are powerful and detailed, but by nature are static.

"We developed TopicVelo to infer cell-state transitions from scRNA-seq data," Riesenfeld said. "It's hard to do that from this kind of data because scRNA-seq is destructive. When you measure the cell this way, you destroy the cell."

This leaves researchers a snapshot of the moment the cell was measured/destroyed. While scRNA-seq gives the best available transcriptome-wide snapshot, the information many researchers need, however, is how the cells transition *over time*. They need to know how a cell becomes cancerous or how a particular gene program behaves during an immune response.

To help figure out dynamic processes from a static snapshot, researchers traditionally use what's called "pseudotime." It's impossible to watch an individual cell or gene's expression change and grow in a still image, but that image also captured other cells and genes of the same type that might be a little further on in the same process. If the scientists connect the dots correctly, they can gain powerful insights into how the process looks over time.

Connecting those dots is difficult guesswork, based on the assumption that similar-looking cells are just at different points along the same path. Biology is much more complicated, with false starts, stops, bursts, and multiple chemical forces tugging on each gene.

Instead of traditional pseudotime approaches, which look at the expression similarity among the transcriptional profiles of cells, RNA velocity approaches look at the dynamics of transcription, splicing and degradation of the mRNA within those cells.

It's a promising but early technology.

"The persistent gap between the promise and reality of RNA velocity has

largely restricted its application," the authors wrote in the paper.

To bridge this gap, TopicVelo puts aside deterministic models, embracing—and gleaning insights from—a far more difficult stochastic model that reflects biology's inescapable randomness.

"Cells, when you think about them, are intrinsically random," said Gao, the first author on the paper. "You can have twins or genetically identical cells that will grow up to be very different. TopicVelo introduces the use of a stochastic model. We're able to better capture the underlying biophysics in the transcription processes that are important for mRNA transcription."

Machine learning shows the way

The team also realized that another assumption limits standard RNA velocity. "Most methods assume that all cells are basically expressing the same big gene program, but you can imagine that cells have to do different kinds of processes simultaneously, to varying degrees," Riesenfeld said. Disentangling these processes is a challenge.

Probabilistic topic modeling—a machine learning tool traditionally used to identify themes from written documents—provided the UChicago team with a strategy. TopicVelo groups scRNA-seq data not by the types of cell or gene, but by the processes those cells and genes are involved in. The processes are inferred from the data, rather than imposed by external knowledge.

"If you look at a science magazine, it will be organized along topics like 'physics,' 'chemistry' and 'astrophysics,' these kinds of things," Gao said. "We applied this organizing principle to single-cell RNA-sequencing data. So now, we can organize our data by topics, like 'ribosomal synthesis,' 'differentiation,' 'immune response,' and 'cell cycle'. And we

can fit stochastic transcriptional models specific to each process."

After TopicVelo disentangles this kludge of processes and organizes them by topic, it applies topic weights back onto the cells, to account for what percentage of each cell's transcriptional profile is involved in which activity.

According to Riesenfeld, "This approach helps us look at the dynamics of different processes and understand their importance in different cells. And that's especially useful when there are branch points, or when a cell is pulled in different directions."

The results of combining the stochastic model with the topic model are striking. For example, TopicVelo was able to reconstruct trajectories that previously required special experimental techniques to recover. These improvements greatly broaden potential applications.

Gao compared the paper's findings to the paper itself—the product of many areas of study and expertise.

"At PME, if you have a chemistry project, chances are there's a physics or engineering student working on it," he said. "It's never just chemistry."