



Site
 People



[ABOUT SEAS](#)
[ACADEMICS](#)
[FACULTY & RESEARCH](#)
[NEWS & EVENTS](#)
[OFFICES & SERVICES](#)
[MAKE A GIFT](#)

News & Events

[Calendars & Colloquia](#)

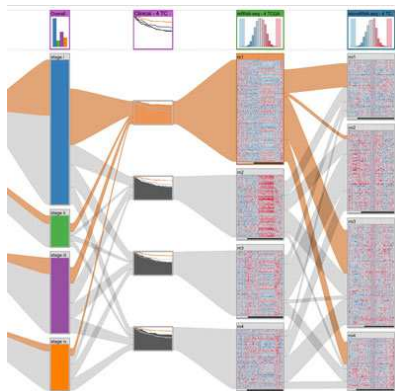
[For the Media](#)

[News & Events](#)
[Recognizing patterns in genomic data](#)

Recognizing patterns in genomic data

NEW VISUALIZATION SOFTWARE UNCOVERS CANCER SUBTYPES FROM A VAST REPOSITORY OF BIOMEDICAL INFORMATION

By Stephanie Dutchen, Harvard Medical School October 7, 2014



Harvard-led computational scientists have created software that makes it easier for nonspecialists to fish out clues from an ocean of numbers. "It's a tool to help you make sense of the data you're collecting and find the right questions to ask," said Nils Gehlenborg. (Image courtesy of Alexander Lex.)

Cambridge, Mass. – October 7, 2014 – Much of biomedical research these days is about big data—collecting and analyzing vast, detailed repositories of information about health and disease. These data sets can be treasure troves for investigators, often uncovering genetic mutations that drive a particular kind of cancer, for example.

Trouble is, it's impossible for humans to browse that much data, let alone make any sense of it.

Computer algorithms and visualization tools help. Still, many biologists and clinicians find themselves having to guess which gene or other variable might be affecting their patients; they have to develop their own custom programs to find possible correlations, analyze the results and then test likely candidates with statistical software. It can be a long, tedious process requiring skills outside their expertise.

Seeing that the toolbox isn't yet complete, computational

specialists in the [lab of Peter Park](#) at Harvard Medical School's Center for Biomedical Informatics and in the lab of [Hanspeter Pfister](#) at the [Harvard School of Engineering and Applied Sciences](#) (SEAS) have teamed up with colleagues at Johannes Kepler University Linz and Graz University of Technology in Austria to produce software that makes it easier for nonspecialists to fish out clues from an ocean of numbers.

"It's a tool to help you make sense of the data you're collecting and find the right questions to ask," said [Nils Gehlenborg](#), research associate in biomedical informatics at HMS and co-senior author of the correspondence in [Nature Methods](#). "It gives you an unbiased view of patterns in the data. Then you can explore whether those patterns are meaningful."

"We meet a lot of biologists who want to test their hypotheses with available data but aren't trained in statistical analysis," said [Peter Park](#), co-senior author and HMS associate professor of pediatrics at Boston Children's Hospital. "We want to give them tools to refine their ideas and come up with new ones without having to rely on a computational person and by reducing the time spent chasing false leads."

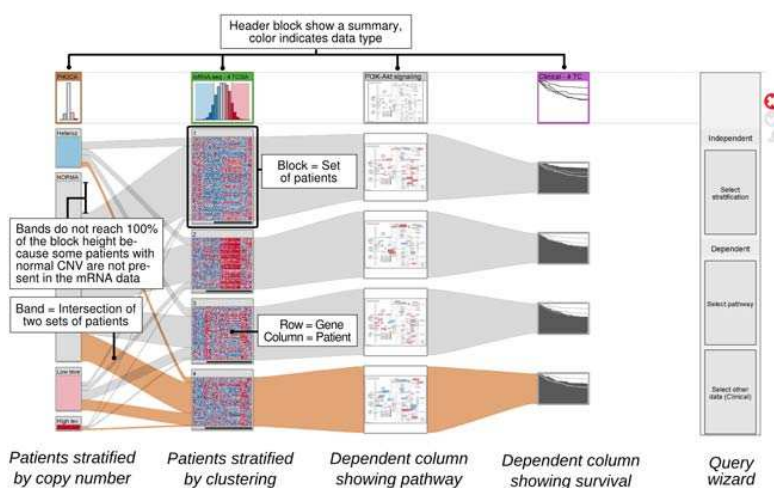
The software, called StratomeX, was developed to help researchers distinguish subtypes of cancer by crunching through the incredible amount of data gathered as part of The [Cancer Genome Atlas](#), a National Institutes of Health-funded initiative. Identifying distinct cancer subtypes can lead to more effective, personalized treatments.

When users input a query, StratomeX compares tumor data at the molecular level that was collected from hundreds of patients and detects patterns that might indicate significant similarities or differences between groups of patients. The software presents those connections in an easy-to-grasp visual format.

"It helps you make meaningful distinctions," said co-first author [Alexander Lex](#), a postdoctoral researcher in the Pfister group.

"You might see that a subset of patients seems to live

longer. Then you can explore whether there's a genetic variant or deletion that affects survival. You don't have to have a suspect in mind when you start," said Gehlenborg.



An example of the StratomeX user interface. (Image courtesy of Alexander Lex.)

"It's an iterative process," added co-first author [Marc Streit](#), assistant professor at the Institute of Computer Graphics at Johannes Kepler University Linz and visiting professor at SEAS. "You can formulate a question, get ranked results and refine the question."

Researchers can then take their refined, informed hypotheses into the clinic for further testing.

StratomeX certainly isn't the only visualization tool out there, but it is the first specifically designed to ferret out cancer subtypes. However, its potential reaches beyond cancer. Researchers can input data sets gathered on any disease and run the same kind of analyses, said Lex.

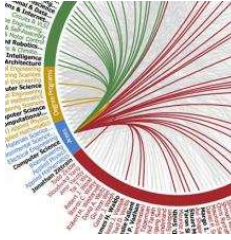
The team has made StratomeX [available for download](#). They plan to make it web-based and hope to enhance it so it can analyze finer differences between patient groups, such as where a particular mutation occurs in a gene rather than simply whether a mutation exists.

"We want it to pick up subtler details that might play an important role in disease," said Streit.

The work was funded by the National Institutes of Health (U24 CA144025, U24 CA143845 and K99 HG007583), Austrian Science Fund (J 3437-N15, P 22902), and Air Force

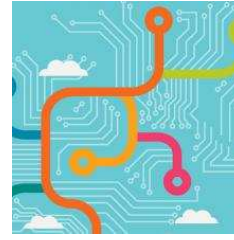
Research Laboratory and Defense Advanced Research
Projects Agency grant FA8750-12-C-0300.

RELATED NEWS



JULY 24, 2014

[Explore connections at SEAS](#)



JULY 22, 2014

[Built for speed: Designing exascale computers](#)



JULY 22, 2014

[The need for speed](#)

[Send Feedback on this page](#)



Harvard SEAS
29 Oxford Street
Cambridge, MA 02138



- [> HARVARD UNIVERSITY](#)
- [> TRADEMARK NOTICE](#)
- [> PRIVACY POLICY](#)

© 2014 President and Fellows of Harvard College

Faculty, Researcher, and Lecturer Positions – Postdoctoral & Fellowship Positions – Staff Positions