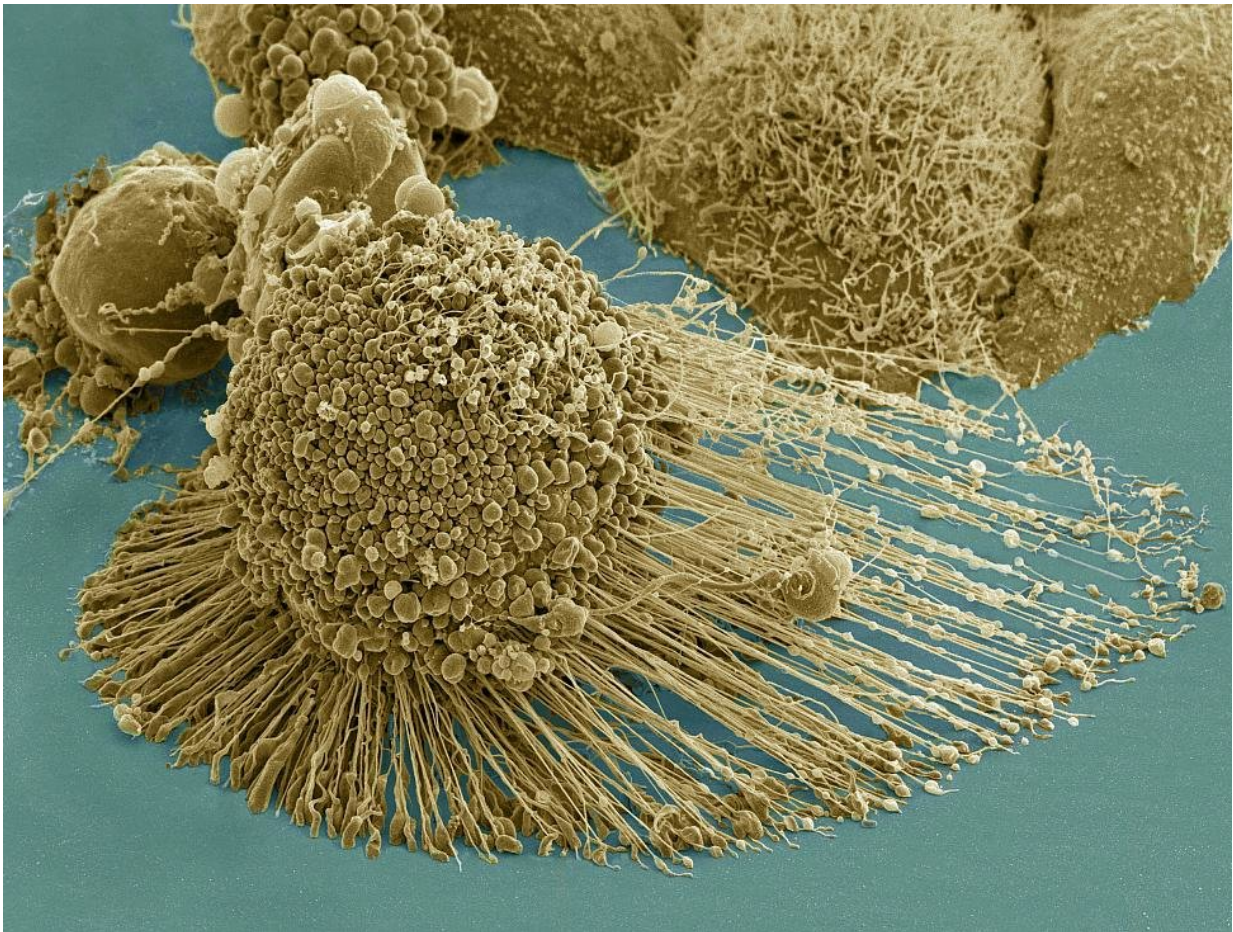


# Rousing masses to fight cancer with open source machine learning

October 30 2017, by Ben Brumfield

---



A dying cancer cell with filopodia stretched out to its right. The protrusions help cancer migrate. Stock NIH NCMIR image. The image does not display a cell treated in the Georgia Tech study. Credit: NIH-funded image of HeLa cell / National Center for Microscopy and Imaging Research / Thomas Deerinck / Mark Ellisman. Use may require permission.

Here's an open invitation to steal. It goes out to cancer fighters and tempts them with a new program that predicts cancer drug effectiveness via machine learning and raw genetic data.

The researchers who built the program at the Georgia Institute of Technology would like cancer fighters to take it for free, or even just swipe parts of their programming code, so they've made it open source. They hope to attract a crowd of researchers who will also share their own cancer and computer expertise and [data](#) to improve upon the program and save more lives together.

The researchers' invitation to take their code is also a gauntlet.

They're challenging others to come beat them at their own game and help hone a formidable software tool for the greater good. Not only the labor but also the fruits will remain openly accessible to benefit the treatment of patients as best possible.

"We don't want to hold the code or data for ourselves or make profits with this," said John McDonald, the director of Georgia Tech's Integrated Cancer Research Center. "We want to keep this wide open so it will spread."

Researchers wanting to participate can [follow this link to a new study published on October 26, 2017, in the journal \*PLOS One\*](#). There they will find links to download the software from GitHub and to access the code.

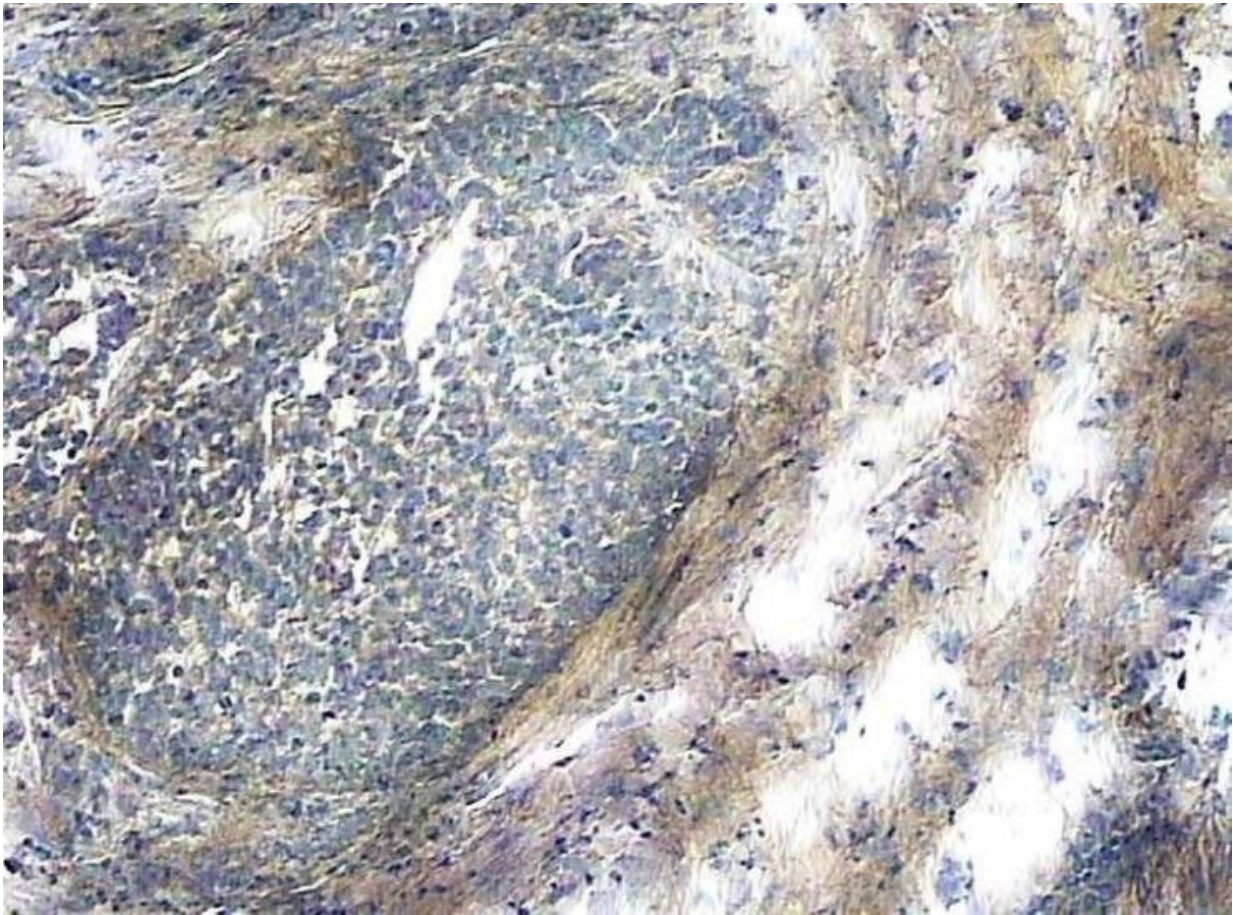
They'll start out with a current program that has been about 85% accurate in assessing treatment effectiveness of nine drugs across the genetic data of 273 [cancer patients](#). The study by McDonald and collaborator Fredrik Vannberg details how and why.

"Nine drugs are in the published study, but we've actually run about 120



drugs through the program all total," said Vannberg, an assistant professor in Georgia Tech's School of Biological Sciences.

The program uses proven machine learning mechanisms and also normalizes data. The latter allows the machine learning to work with data from varying sources by making them compatible.



Ovarian cancer cells stained blue under the microscope. Credit: Georgia Tech / McDonald

And the researchers have reduced human bias about which data are

important for predicting outcomes.

"It's much more effective to put in loads of raw data and let the algorithm sort it out," McDonald said. "It's looking for correlations, not causes, so it's not good to preselect data for what you suspect are most relevant."

One big bias the researchers tossed out was a concentration only on [gene expression data](#) pertaining to the specific type of cancer they were aiming to treat.

"It turns out that it's better to give the program data from a broad diversity of cancers, and that will actually later give a better prediction of drug effectiveness for a specific cancer like [breast cancer](#)," Vannberg said.

"On a molecular level, some breast cancers, for example, are going to be more similar to some ovarian cancers than to other breast cancers," McDonald said. "We just let the algorithm work with about everything we had, and we got high accuracy."

The researchers also want the project to pool large amounts of anonymous patient treatment success and failure data, which will help the program optimize predictions for everyone's benefit. But that doesn't mean some companies can't benefit, too.

"If a company comes along and makes profits while using the program to help patients, that's fine, and there's no obligation to give back to the project," said McDonald, who is also a professor in Georgia Tech's School of Biological Sciences. "Others may just take if they so please."

But hopefully, most players will catch the spirit of kindness.

"With our project, we're advertising that sharing should be what everybody does," Vannberg said. "This can be a win for everybody, but really it's a win for the [cancer](#) patients."

**More information:** Cai Huang et al. Open source machine-learning algorithms for the prediction of optimal cancer drug therapies, *PLOS ONE* (2017). [DOI: 10.1371/journal.pone.0186906](https://doi.org/10.1371/journal.pone.0186906)

Provided by Georgia Institute of Technology

Citation: Rousing masses to fight cancer with open source machine learning (2017, October 30) retrieved 24 April 2024 from

<https://medicalxpress.com/news/2017-10-rousing-masses-cancer-source-machine.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--