A novel network computer model to find co-occurring mutations—researchers improve search for cancer drivers
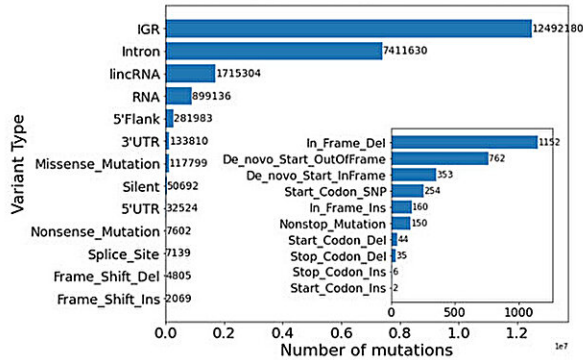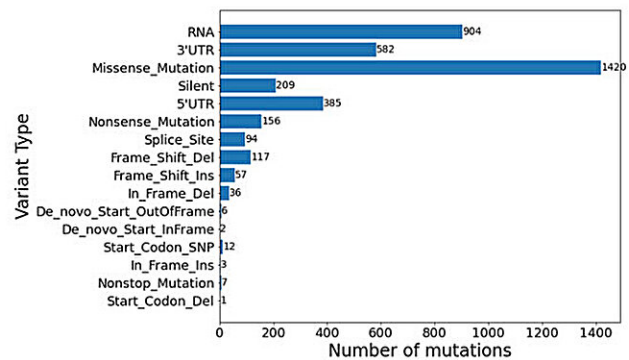
September 11 2024, by Tina Hilding
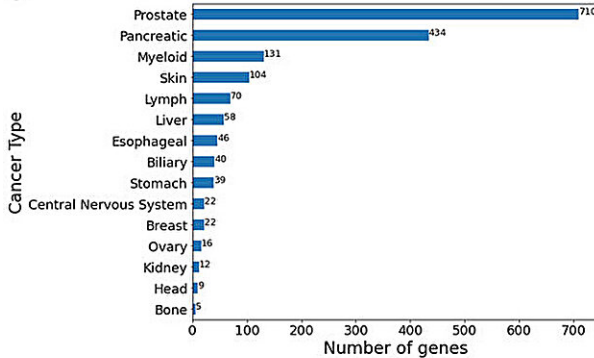

Number of mutations by variant type (A) before filtering and (B) after filtering. (C) Number of genes and their role in cancer. (D) Pie chart showing the proportion of different cancer types in the final dataset comprising 934 samples. The inset in Figure A shows the number of mutations for the variants having the 10 lowest frequencies. Credit: *Frontiers in Bioinformatics* (2024). DOI: 10.3389/fbinf.2024.1365200

A computer algorithm can efficiently find genetic mutations that work together to drive cancer as well as other important genetic clues that researchers might someday use to develop new treatments for a variety of cancers.

Reporting in the journal Frontiers in Bioinformatics, a Washington State University-led team used a novel network computer model to find co-occurring mutations as well as other similarities among DNA sequence elements across several types of cancer. The model allows for easier searches for patterns in huge seas of cancer genetic data.

"This is a study not of one particular cancer, where you dive in and try to understand it, but rather of many cancers together, looking for patterns and things that could someday be used for drug discovery," said corresponding author Assefaw Gebremedhin, associate professor in the WSU School of Electrical Engineering and Computer Science.

Cancer tends to be talked about as one disease, but it's really a spectrum of diseases where different driver mutations dictate the disease's progression and prognosis, said co-author and University of Vermont molecular genetics researcher Steven Roberts.

A better understanding of how common the different mutations and driver genes are across different types of cancer could help to prioritize possible targets for treatment. But cancer researchers have been stymied in their search for clues by the huge amount of computation required to study long genetic sequences and a large number of mutations.

"We couldn't really look at all of the sequences because the computational space would be too large," said Roberts. "If you try to take all the genomic data together and analyze all of it, the mathematical

problem scales exponentially, and you just overwhelm the system."

The network model developed by the WSU team, which is called DiWANN, is sparser and more efficient than existing network models without losing key structural components.

"This model gets to the minimal way to represent things without losing information," said Gebremedhin. "Our model tries to understand relationships between sequences in a more effective way, and by effective, I mean they can be computed quickly. With this minimal representation of the network, you can get more information and make computation scalable."

Since developing their DiWANN model five years ago, the researchers have used it to look at the geographic distribution of tick-borne diseases as well as the spread of COVID-19 during the pandemic.

In this work, the researchers added a data reduction step to further reduce the amount of required computation and a second computer model to gain further insights into co-occurring genes. WSU computer science Ph.D. student Shruti Patil led this effort and is the first author of the study.

The researchers found evidence in their work, for instance, that two mutations in pancreatic cancer nearly always show up together—something that had only been suspected. One of the mutations, tumor protein 53, suppresses tumor growth while the other, known as KRAS, is a driver of proliferation.

The researchers identified cancer types that are closely connected to each other and that could perhaps be susceptible to common drug treatments. Some cancers were homogenous, meaning they all had similar mutations that lead to cancer, whereas other types of cancers had

a huge variety of different mutations.

"Some of the certain cancer types are very, very homogenous, and their driver mutations were almost always the same, but then you get some other cancer types where it's all over the map," said Roberts. "Those are the ones, based on our predictions, that are probably going to be the hardest to treat."

Because it provides sparse information, the WSU network model allows the researchers to expand the number of tumors that they can study.

"This increases our power to actually detect novel interactions and novel aspects of how these tumors behave," said Roberts. "If we can actually screen through large data sets, it's a much faster way to go about it."

The researchers are now working to develop a web-based tool for public health experts, so that researchers in health can more easily use the model to study complex questions in cancer and other diseases.

  More information: Shruti S. Patil et al, Network analysis of driver genes in human cancers, *Frontiers in Bioinformatics* (2024). DOI: 10.3389/fbinf.2024.1365200

Provided by Washington State University