Medical Xpress
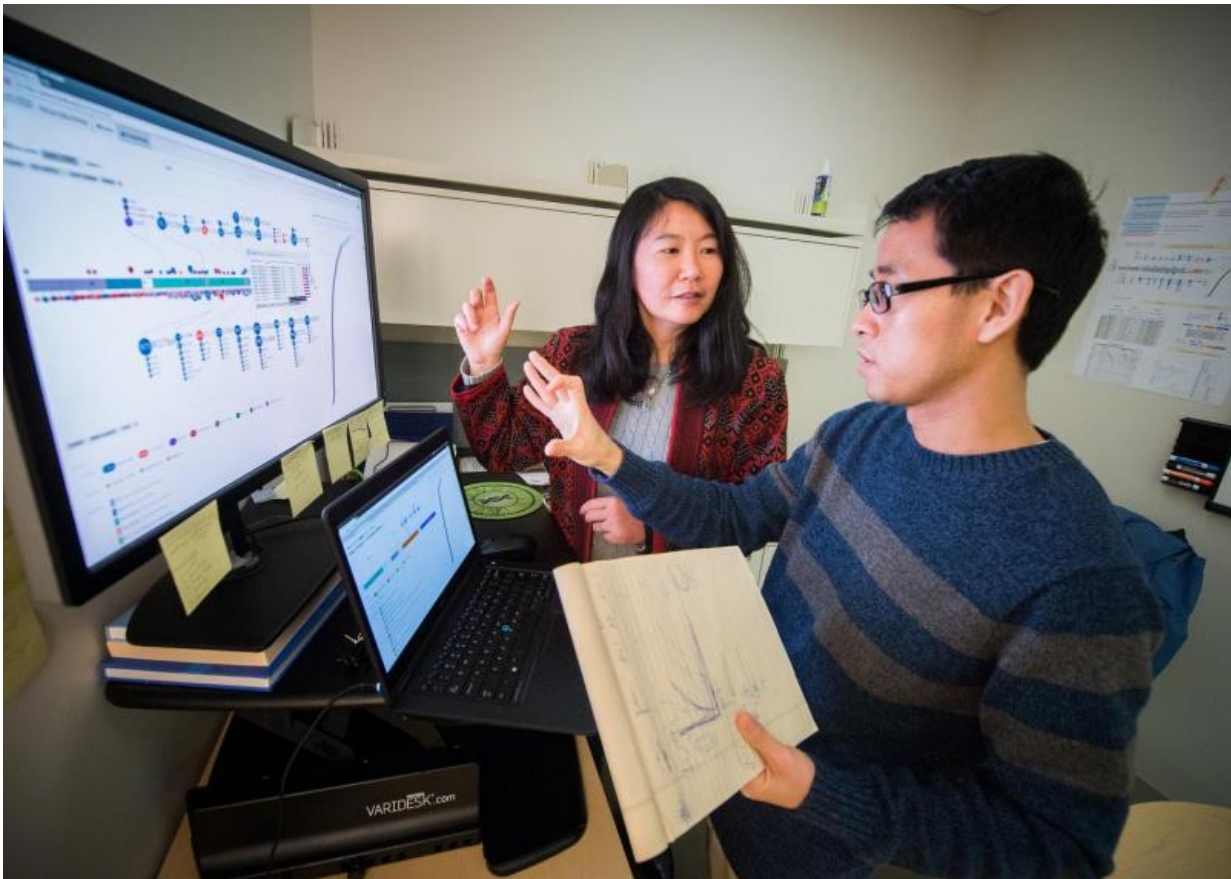
Researchers develop powerful interactive tool to mine data from cancer genome

December 29 2015

Jinghui Zhang, Ph.D. and Xin Zhou, Ph.D. share ideas about web-based application ProteinPaint. Credit: St. Jude Children's Research Hospital / Peter Barta.

St. Jude Children's Research Hospital scientists have developed a web application and data set that gives researchers worldwide a powerful interactive tool to advance understanding of the mutations that lead to and fuel pediatric cancer. The freely available tool, called ProteinPaint, is described in today's issue of the scientific journal *Nature Genetics*.

ProteinPaint provides users with a gene-by-gene snapshot of mutations from pediatric cancer that alters genetic instructions for encoding proteins. The application provides critical information unavailable with existing visualization tools. For example, ProteinPaint shows whether mutations are present at diagnosis or just at relapse, or whether mutations occur in almost every cell (germline) or just cancer cells (somatic).

ProteinPaint's novel interactive infographics also let researchers see at a glance all mutations in individual genes and their corresponding proteins, including detailed information about mutation type, frequency in cancer subtype and location in the protein domain. That information provides clues about how a change might contribute to cancer's start, progression or relapse.

"Each day brings new information about mutations that drive human cancer. Novel tools are essential to help scientists use this wealth of genomic data to advance research and find new cures," said corresponding author Jinghui Zhang, Ph.D., chair of the St. Jude Department of Computational Biology. "We developed ProteinPaint as an intuitive tool any scientist can easily use to explore the vast amount of information now available on cancer genomics."
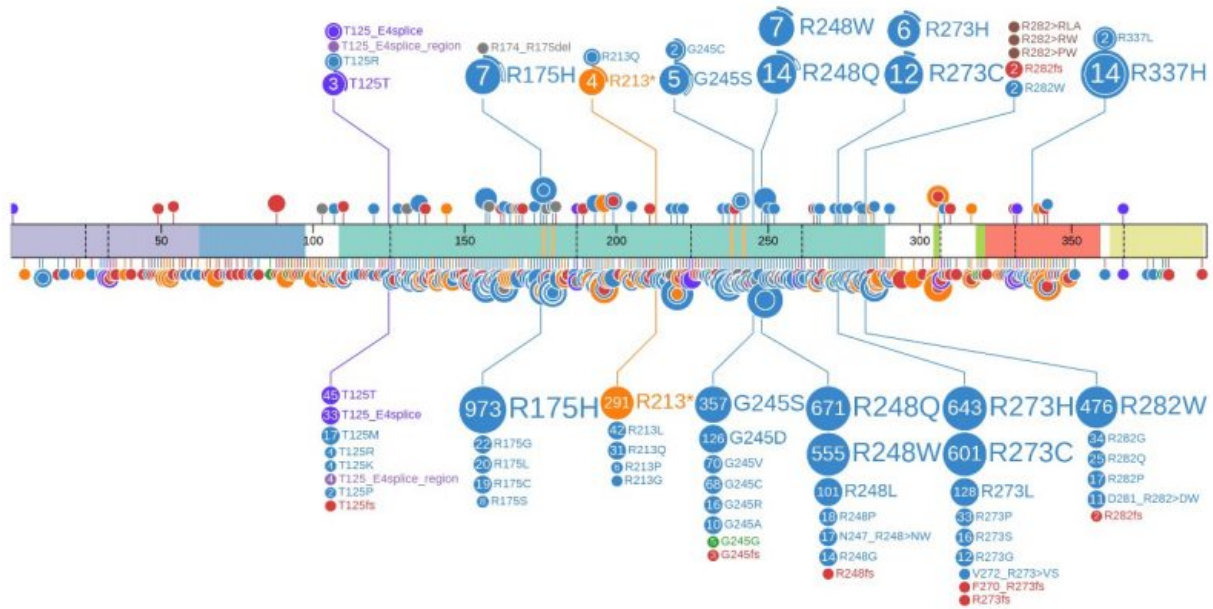
ProteinPaint's novel interactive infographics let researchers see at a glance all mutations in individual genes and their corresponding proteins.

There are multiple types of mutations that disrupt the structure of protein-coding genes and lead to cancer. ProteinPaint integrates mutation information from multiple data sets, which boosts its power as a research tool. The application incorporates findings from the St. Jude Children's Research Hospital—Washington University Pediatric Cancer Genome Project, the National Cancer Institute's Therapeutically Applicable Research to Generate Effective Treatments (TARGET) initiative and other published pediatric cancer studies.

ProteinPaint currently includes information on almost 27,500 mutations discovered in more than 1,000 pediatric patients with 21 cancer

subtypes. The data will be updated as new information is published.

The application's developers use the curated data to "paint" or overlay detailed, annotated information about each mutation on the affected protein. First author Xin Zhou, Ph.D., a St. Jude senior bioinformatics research scientist, developed the infographics to display the range of genomic information in an intuitive and interactive format. A click of the mouse gives users additional details about the mutations, including the pediatric cancer subtype where the change has been validated, and a link to the publication.

The application also 'paints' RNA-sequencing data from 928 pediatric tumors from 36 subtypes to track how mutations affect gene expression. While whole genome sequencing reveals the complete DNA makeup of an organism, RNA sequencing provides a snapshot of how instructions encoded in DNA are transcribed into RNA molecules. The information is essential for developing and delivering individualized cancer therapies.

"ProteinPaint's focus on pediatric cancer and presentation of mutations at the gene level complements existing cancer genome data portals," Zhang said. "For St. Jude, the application is the foundation for developing a global reference database for information about pediatric cancer."

Zhou added that the ProteinPaint software has the potential to help researchers studying other disorders, including sickle cell disease, that involve a mutation that affects protein function.

ProteinPaint is available at no cost to academic researchers who are also free to use the tool to analyze their own data. The application also lets researchers compare information about pediatric and adult cancer genomes by providing a parallel view of data COSMIC, the world's largest database of somatic mutations, primarily from adult cancer. Such

comparisons can help researchers understand and interpret the significance of <u>rare mutations</u>.

The tool was used to study the role played by <u>germline mutations</u> in pediatric cancer. The <u>research appeared</u> in Nov. 18, 2015 in the *New England Journal of Medicine*.

Provided by St. Jude Children's Research Hospital

Citation: Researchers develop powerful interactive tool to mine data from cancer genome (2015, December 29) retrieved 23 April 2024 from <u>https://medicalxpress.com/news/2015-12-powerful-interactive-tool-cancer-genome.html</u>