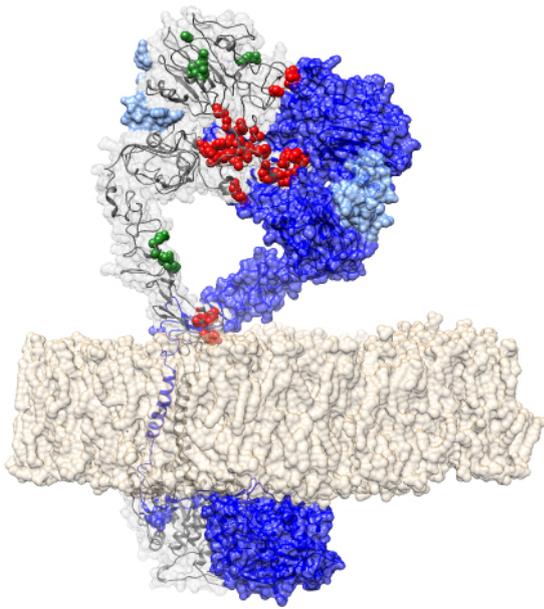


# 'Big Data' used to identify new cancer driver genes

20 October 2015



A structure showing EGFR -- a cancer driver -- in its active dimer conformation. Red indicates mutations that destroy the protein-protein interface Credit: Eduard Porta Pardo

In a collaborative study led by Sanford Burnham Prebys Medical Discovery Institute (SBP), researchers have combined two publicly available 'omics' databases to create a new catalogue of 'cancer drivers'. Cancer drivers are genes that when altered, are responsible for cancer progression. The researchers used cancer mutation and protein structure databases to identify mutations in patient tumors that alter normal protein-protein interaction (PPI) interfaces. The study, published today in *PLoS Computational Biology*, identified more than 100 novel cancer driver genes and helps explain how tumors driven

by the same gene may lead to different patient outcomes.

"This is the first time that three-dimensional [protein](#) features, such as PPIs, have been used to identify driver genes across large cancer datasets," said lead author Eduard Porta-Pardo, Ph.D., a postdoctoral fellow at SBP. "We found 71 interfaces in proteins previously unrecognized as cancer drivers, representing potential new cancer predictive markers and/or drug targets. Our analysis also identified several driver interfaces in known cancer genes, such as TP53, HRAS, PI3KCA and EGFR, proving that our method can find relevant cancer driver genes and that alterations in protein interfaces are a common pathogenic mechanism of cancer."

Cancer is caused by the accumulation of mutations to DNA. Until now, scientists have focused on finding alterations in individual genes and cell pathways that can lead to cancer. But the recent push by the National Institutes of Health (NIH) to encourage data sharing has led to an era of unprecedented ability to systematically analyze large scale genomic, clinical, and molecular data to better explain and predict [patient outcomes](#), as well as finding new drug targets to prevent, treat, and potentially cure cancer.

"For this study we used an extended version of e-Driver, our proprietary computational method of identifying protein regions that drive cancer. We integrated tumor data from almost 6,000 patients in The Cancer Genome Atlas (TCGA) with more than 18,000 three-dimensional protein structures from the Protein Data Bank (PDB)," said Adam Godzik, Ph.D, director of the Bioinformatics and Structural Biology Program at SBP. "The algorithm analyzes whether structural alterations of PPI interfaces are enriched in cancer mutations, and can therefore identify candidate driver genes."

"Genes are not monolithic black boxes. They have

different regions that code for distinct protein domains that are usually responsible for different functions. It's possible that a given protein only acts as a cancer driver when a specific region of the protein is mutated," Godzik explained. "Our method helps identify novel cancer driver genes and propose molecular hypotheses to explain how tumors apparently driven by the same gene have different behaviors, including patient outcomes."

"Interestingly, we identified some potential [cancer](#) drivers that are involved in the immune system. With the growing appreciation of the importance of the immune system in [cancer progression](#), the immunity genes we identified in this study provide new insight regarding which interactions may be most affected," Godzik added.

Provided by Sanford-Burnham Prebys Medical  
Discover Institute

APA citation: 'Big Data' used to identify new cancer driver genes (2015, October 20) retrieved 19 August 2022 from <https://medicalxpress.com/news/2015-10-big-cancer-driver-genes.html>

*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.*