New method identifies genes that can predict prognoses of cancer patients

January 25 2013

In recent years, it has been thought that select sets of genes might reveal cancer patients' prognoses. However, a study published last year examining breast cancer cases found that most of these "prognostic signatures" were no more accurate than random gene sets in determining cancer prognoses. While many saw this as a disappointment, investigators at Beth Israel Deaconess Medical Center (BIDMC), the Dana-Farber Cancer Institute, and the Institut de Recherches Cliniques de Montréal (IRCM) saw this as an opportunity to design a new method to identify gene sets that could yield more significant prognostic value.

Led by Andrew Beck, MD, Director of the Molecular Epidemiology Research Laboratory at BIDMC, the team has developed SAPS (Significance Analysis of Prognostic Signatures), a new algorithm that makes use of three specific criteria to more accurately identify prognostic signatures associated with patient survival.

Their results, the largest analysis of its kind ever performed, are reported in the January 24 on-line issue of the journal PLOS Computational Biology.

"SAPS makes use of three specific criteria," explains Beck, who is also an Assistant Professor of Pathology at Harvard Medical School. "First, the gene set must be enriched for genes that are associated with survival. In addition, the gene set must separate patients into groups that show survival differences. Lastly, it must also perform significantly better than sets of random genes at these tasks."

In the new study, the scientific team applied the SAPS algorithm to [gene expression profiling](#) data from the study's senior author Benjamin Haibe-Kains, PhD, Director of the Bioinformatics and Computational Genomics Laboratory at IRCM and an Assistant Research Professor at the University of Montreal. The first collection of data was obtained from 19 published [breast cancer](#) studies (including approximately 3800 patients), and the second included 12 published gene expression profiling studies in ovarian [cancer](#) (including data from approximately 1700 patients).

When the investigators used SAPS to analyze these previously identified prognostic signatures in breast and ovarian cancer, they found that only a small subset of the signatures that were considered statistically significant by standard measurements also achieved statistical significance when evaluated by SAPS.

"Our work shows that when using prognostic associations to identify biological signatures that drive cancer progression, it is important to not rely solely on a gene set's association with patient survival," says Beck. "A gene set may appear to be important based on its survival association, when in reality it does not perform significantly better than random genes. This can be a serious problem, as it can lead to false conclusions regarding the biological and clinical significance of a gene set."

By using SAPS, Beck and his colleagues found that they could overcome this problem. "The SAPS procedure ensures that a significant prognostic gene set is not only associated with patient survival but also performs significantly better than random gene sets," says Beck. His team revealed new prognostic signatures in subtypes of breast cancer and ovarian cancer and demonstrated a striking similarity between signatures in estrogen receptor negative breast cancer and ovarian cancer, suggesting new shared therapeutic targets for these aggressive malignancies.

The findings also indicate that the prognostic signatures identified with SAPS will not only help predict patient outcomes but might also help in the development of new anti-cancer drugs. "We hope that markers identified in our analysis will provide new insights into the biological pathways driving cancer progression in breast and ovarian cancer subtypes, and will one day lead to improvements in targeted diagnostics and therapeutics," says Beck. "We also hope the method proves widely useful to other researchers." To that end, the team would like to create a web-accessible tool to enable investigators with little knowledge of statistical software and programming to identify gene sets significantly associated with patient outcomes in different diseases.

"We also plan to soon release a software package, which includes all the code and corresponding documentation of our analysis pipeline," adds Haibe-Kains. "This will allow others to fully reproduce our results while enabling the bioinformatics and computational biology communities to take over and potentially adapt and improve our pipeline to address important new issues in biomedicine."

Beck and his collaborators are currently working to further validate the prognostic signatures they identified in breast and ovarian cancers, with the hopes of bringing them closer to the clinic through the development of new diagnostics and treatments. "We are also extending our approach to other common cancers that lack robust prognostic signatures," he notes.

Provided by Beth Israel Deaconess Medical Center