Uploaded to the VFC Website

▶▶ ▶▶ 2016 ◀◀ ◀◀

Riverside County, California

Whole-genome omics data can improve accuracy of breast cancer survival predictions

Published on July 6, 2016 at 12:13 PM

Precise predictions of whether a tumor is likely to spread would help clinicians and patients choose the best course of treatment. But current methods fall short of the precision needed. New research reveals that profiling primary tumor samples using genomic technologies can improve the accuracy of breast cancer survival predictions compared to clinical information alone. The study was published in the journal *GENETICS*, a publication of the Genetics Society of America.

Although this method is not ready for use in the clinic, the proof-of-principle study shows that survival predictions improve when they incorporate comprehensive data on which genes are active in tumor samples compared to non-cancerous tissues from the same patient. This is also true for genome-wide methylation data, which maps the parts of the DNA that carry molecular "tags" that influence gene activation. If developed for use in the clinic, the approach could spare some patients from unneeded chemotherapy.

After surgery, approximately 80% of breast cancer patients are treated with adjuvant therapies, including chemotherapy and radiotherapy. These treatments often have serious long- term side effects, including heart damage, infertility, memory problems, and a higher risk of developing a new, independent cancer. But not every patient necessarily needs adjuvant therapy; breast cancer is estimated to recur or metastasize in only approximately 40% of patients, suggesting that a substantial number of patients suffer side effects needlessly. For now, widespread use of adjuvant therapy remains unavoidable because we can't predict which primary tumors are likely to metastasize and become deadly, and which will stay put.

Today, doctors use a variety of clinical information to help choose the best treatment for an individual patient, including patient age and ethnicity, size of the tumor, the cell type it arose from, how advanced it is (stage), and the presence of various types of receptors and other molecular signatures on the tumor cells (cancer subtype). To help refine treatment choices, several commercial tests estimate the risk of cancer recurrence by measuring the activity (expression) of a set of genes that influence cancer progression. For example, the widely available Oncotype DX panel analyzes expression of 21 genes in tumor samples and is recommended for patients with specific types of breast cancer.

But cancer is a complex disease, and its behavior is likely affected by thousands of genes. Advances in genomic technology mean it is now feasible to measure tumor gene expression across the entire genome. Samples can also be profiled for a variety of other genome-scale measurements, including variation at the DNA level (e.g., deletions or mutations) and methylation. The authors of the new study examined whether such genomic data, whether alone or in combination, could in fact improve predictions of breast cancer survival.

"Rather than pre-select which handful of genes might best predict survival, we used data from all the genes present in the cancer cell --approximately 17,000 in our study--and let our computational model select the informative ones," says study leader Ana I. Vazquez of Michigan State University.

To test their approach, Vazquez and her colleagues used data from The Cancer Genome Atlas, a National Institutes of Health project that profiles several types of genome-scale data in thousands of cancer samples. Samples are matched to normal tissue from the same individual, along with basic clinical information on the patient. The team honed in on primary breast cancer samples from 285 patients who had sufficient clinical follow-up information to allow the team to analyze survival rates.

The authors used this dataset to build computational models that predict a patient's outcome (e.g. survival) using different types of data. They compared the performance of these models using cross-validation. In this method, the data is divided randomly in two: one portion is used to build and tweak a predictive model, and the other portion is used to test how accurately the model performs. This procedure is repeated hundreds of times for new random divisions of the data, and the results are scored to reveal which model makes more reliable predictions.

This showed that whole-genome gene expression data were better predictors of survival than any single source of information currently used by doctors, including cancer stage (how advanced the cancer is) and molecular subtype

(e.g. hormone receptor status). Combining the gene expression data with the clinical data provided better predictions than all the clinical predictors together. The whole-genome gene expression data also outperformed the predictions achieved with genes in the Oncotype DX panel in the subset of patients that met the criteria for the panel. Oncotype DX is a well validated test used in the clinic since 2004.

Methylation data alone was also more predictive than all the standard clinical information, and also improved predictions further when combined with the clinical data. Finally, combining clinical information, whole-genome gene expression, and methylation data provided the most predictive models examined in the study.

"Overall, we can conclude that the predictions keep improving as you add 'omics' data" says Vazquez. "This gives us promising genomic leads for future application in the clinic."

Not all types of genome-scale information were as predictive as gene expression or methylation data, the team found. Prediction accuracy from clinical data was unaffected by adding genomic profiles of microRNAs, small molecules that can influence gene expression. And although accuracy was improved by combining clinical data with genomic profiles of a particular type of DNA change (known as a copy number variant), this improvement was much smaller that the gains provided by gene expression or methylation data.

None of the models can predict survival with certainty. Vazquez says that although the method is promising, a major limitation of the study was the small number of samples available to develop the models. To be applied by clinicians, the method would need to be validated using data from thousands of patients, rather than hundreds. Her team is also investigating how to incorporate other factors into their models, including treatment regimes. Ultimately, this may help doctors and patients match the best course of treatment with the individual characteristics of each tumor.

Source:
Genetics Society of America