Individualizing Care for Ovarian Cancer Patients Using Big Data

Brooke L. Fridley, Devin C. Koeslter, Andrew K. Godwin

Correspondence to: Brooke L. Fridley, PhD, Department of Biostatistics, University of Kansas Medical Center, 3901 Rainbow Blvd, Kansas City, KS 66160 (e-mail: bfridley@kumc.edu).

Over the last 10 years, numerous studies have attempted to determine a prognostic molecular signature for ovarian cancer, with little success. In this issue of the Journal, two articles leverage the wealth of data produced from gene expression microarray array studies to assess gene expression signatures for predicting outcomes in patients with ovarian cancer. Waldron et al. (1) present a framework for assessing and validating previously published prognostic gene signatures for ovarian cancer outcomes, noting that “most previously published models demonstrated lower accuracy in new independent datasets, compared with the validation sets presented in their publication.” Notably, Waldron and colleagues (1) undertook the monumental task of collecting, curating, and combining all publicly available ovarian cancer gene expression studies and deposited the resulting data in the curatedOvarianData database (2), an incredibly valuable resource for future ovarian cancer genomics studies. Using these data and a similar meta-analytic framework, Riester and colleagues developed gene signatures to predict overall survival and debulking status (optimal vs suboptimal) for late-stage ovarian cancer (3).

The prognostic models developed by Riester et al. (3) outperformed previously identified gene expression signatures; additionally their model for debulking status was further validated in an independent study and was able to classify 92.8% of patients correctly into high- and low-risk groups for suboptimal debulking. However, with the number of patients being treated in the neoadjuvant setting increasing, a predictive signature that identifies patients who are unlikely to achieve optimal cytoreduction would likely be of limited clinical benefit. With regard to their “survival signature,” the ability to identify patients at high risk for recurrence would allow clinicians to tailor treatment decisions, such as recommending those patients for clinical trials soon after the completion of the first treatment regimen.

There are several aspects worth noting about these two articles. First, this research represents the largest sample size used in the development and validation of prognostic gene expression signatures for epithelial ovarian cancer. The authors should also be applauded for their rigorous approach in making the data, code, and models used in their research publicly available. As a research community, we need to encourage this level of transparency to ensure reproducibility of research findings and conclusions. The Waldron et al. article (1) provides additional statistical evidence for a handful of previously proposed prognostic models, while providing a framework for the assessment of future expression signatures. They also confirm the importance of accounting for batch effects in the analysis of microarray data, which we and others have also observed in measurements produced from this technology.

This timely research brings us to the crossroad between cancer, genomics, and precision medicine. Is development of a molecular signature that has clinical utility in the treatment of ovarian cancer an obtainable goal in a heterogeneous disease with multiple origins, and if so, how should we go about developing clinically relevant molecular signatures for use in the treatment of ovarian cancer? To improve the likelihood of developing a clinically useful signature that will ultimately change patient care, we believe the following questions need to be carefully considered:

1. What is the clinically relevant phenotype that one should focus on in the development of the molecular signature?
2. What type of molecular information should be included in the development of the signature?
3. In what setting will a given molecular signature have clinical utility?

Regarding the first point, recent screening studies have shown success in diagnosing ovarian cancer at an earlier stage (by simultaneous screening with CA125 and transvaginal ultrasound); however, these approaches have yet to significantly impact ovarian cancer mortality rates (4). One explanation for this finding is that the cancers we detect early are not the ones destined to turn into high-grade serous tumors that contribute to loss of life. Prognostic models that can predict treatment response or time to recurrence could impact treatment decision because most patients eventually experience recurrence and develop platinum-resistant tumors (5). In addition, molecular signatures should be developed for histological subtypes, enabling the reclassification of previously determined high-grade endometrial cancer as serous. The development of a molecular signature able to subclassify tumors based on key genomic events would enable individualized treatment decisions, such as synthetic lethal treatment strategies involving PARP inhibitors (6).

To address the second point, future signatures may need to include multiple levels of molecular information, such as genomic, transcriptome, and epigenomic data, in addition to clinicopathologic features of the patient. Traditionally, molecular profiling has been based on a single data type, primarily gene expression data from microarray technology. However, clinical outcome and response to cancer therapies is most likely not due to a single gene or data type but rather a complex relationship involving somatic mutations, mRNA, miRNA, DNA methylation, copy number variation, and so on. Recent studies conducted by the Cancer Genome Atlas highlight the ability to classify cancer patients into molecular subtypes using an integrative approach involving somatic
mutations in key genes, copy number or structural variants, and epigenetic changes, such as CpG island methylator phenotype, in addition to mRNA gene expression data. This is important because simple nomenclature that groups neoplastic processes by location of origin does little to appreciate the vast number of subtypes that can arise, each bearing discrete -omic signatures.

Finally, in the development of a signature to be used in clinical practice, as stated by Ludwig and Weinstein (7) in the context of biomarkers for staging, the signature needs to be sensitive, specific, cost-effective, fast, generalizable to the population of interest and robust to various sources of variability and must demonstrate clinical value beyond existing prediction tools. As exemplified by the two articles in this issue, the model needs to be assessed with cross-validation methods and tested in an independent study, followed by additional assessment of the ability of the signature to be determined using different technologies, at different labs, and within different institutions. The final step in determining the clinical utility of a prognostic model would entail a comparison of abilities of the model and clinicians at predicting the patient outcome (8) or a comparison of outcomes of patients randomly assigned to use of the signature in treatment decisions within a prospective clinical trial (9). These steps will ensure a reproducible, clinically useful, prognostic signature.

Because each woman and her cancer are unique, successful cures and outcomes will only come from informative biomarkers/signatures and treatments that target specific cells within each person’s tumor. As exemplified by these two articles, we have come a long way toward this goal; however, more research is needed to achieve the ultimate goal of precision medicine for cancer patients.

References

Affiliations of authors: Department of Biostatistics (BLF, DK) and Department of Pathology and Laboratory Medicine (AKG), University of Kansas Medical Center, and University of Kansas Cancer Center (BLF, DK, AKG), Kansas City, KS.