New Approaches but the Same Flaws in the Search for Prognostic Signatures

Ramon Salazar and Josep Tabernero

A six-gene prognostic colorectal cancer hypoxia score was generated from expression data from in vitro experiments and microarray datasets and was validated in two distinct patient cohorts. The approach followed by the authors is original and biologically sound but could be limited by potential biases and other methodologic limitations. Clin Cancer Res; 20(8); 2019–22. ©2014 AACR.

In this issue of Clinical Cancer Research, Dekervel and colleagues (1) describe a novel mechanistic approach to the development of a new prognostic gene signature for stage II and III colorectal cancer in the clinical setting. Colorectal cancer continues to be the third most common cancer worldwide (2). As with most cancers, survival is directly dependent on the stage of the disease. The American Joint Committee on Cancer classification is an invaluable tool, for most neoplastic diseases, to assess a particular patient’s risk based on disease stage and to evaluate the need for adjuvant treatment. Nevertheless, the limitations of the current staging system may hamper our ability to provide the best clinical care to our patients as the clinical decision to administer adjuvant chemotherapy is mainly determined by this clinicopathologic staging tool and does not take molecular biology features into consideration (3). Adjuvant treatment in colorectal cancer is the current standard treatment for stage III but still controversial in stage II. Current prognostic factors used in the clinical setting are insufficient to identify those patients with stage II colorectal cancer at high risk of recurrence or those patients with stage III colorectal cancer at low risk, thus, leading to potential undertreatment or overtreatment with adjuvant chemotherapy. With the aim of solving this clinical conundrum, a large number of gene expression profiling–based assays have been developed over recent years as prognostic signatures using supervised analytical methods. These approaches involve an initial training step in predefined groups of patients, for instance, relapse versus nonrelapse, to identify gene expression signatures with predictive capability. A large number of prognostic gene profiles have been discovered and developed effectively in the clinical setting (4), but only two of them belong to commercial platforms that have succeeded in validating their signature in clinically relevant, multivariate, independent analyses (5, 6).

Dekervel and colleagues report on the development of yet another prognostic gene expression signature that could potentially have clinical implications in the crowded field of colorectal cancer prognostic signatures (1). The mechanistic approach provides a proof-of-concept affirming the importance of the tumor microenvironment and, in particular, hypoxia in the biologic behavior of the tumor. The steps performed that led to the development of the Colon Cancer Hypoxia Score (CCHS) are innovative and may yield promise. Expression data of an in vitro hypoxia experiment were combined with microarray datasets. Subsequently, the six-gene score obtained was validated in two independent patient cohorts. Patients with low CCHS showed significantly better disease-free survival (DFS) at 3 years (77.3%) with respect to high CCHS patients (46.4%; log-rank, \( P = 0.006 \)) in the primary clinical validation patient cohort from the authors’ center. Patients with a favorable CCHS were also more likely to be alive at 3 years (estimated overall survival 81.5% vs. 66.6%; \( P = 0.044 \)) and at 5 years (71.9% vs. 59.1%; \( P = 0.036 \)). In addition, CCHS (high vs. low) could significantly predict DFS within stage II and III separately. This was independently confirmed in an external public dataset of 90 stage II patients in which the CCHS identified two subgroups of stage II patients with distinct DFS at 3 years (86.9% vs. 52.2%; \( P < 0.001 \)) and 5 years (86.9% vs. 52.2%; \( P < 0.001 \)). However, the approach followed by the authors could be limited by potential biases in the selection of the public data used in the training set, and standard rules for clinical validation still need to apply. A few methodologic issues merit further detailed comments. First, differentially expressed hypoxia genes were analyzed in one single colorectal cancer cell line, CaCo-2. This single-cell approach limits gene selection to the individual molecular background of this particular model, which may or may not be representative of the far more heterogeneous clinical setting. The authors are aware of this, because they added TP53 at a later point in the clinical training of their signature based on prior literature interpretation but not on their data (TP53 gene is mutated with undetectable
expression levels in the CaCo-2 cell line). Second, training the model in silico also involves important limitations. Clinical information is limited to what is publicly available, which is usually heterogeneous and biased toward particular populations. Datasets usually differ in patient characteristics, inclusion criteria, and outcome definition. Among the four-gene expression omnibus series used for the training in this article, one has scarce clinical information and even lacks a recurrence outcome report (GSE13294; Table 1 in ref. 1). Nevertheless, this is the only one of the four datasets used that has information on microsatellite instability (MSI) status, which was arguably used as a surrogate variable for clinical outcome. MSI status is unknown in the remaining three series, which makes the authors’ assumptions underlying use of MSI as a surrogate parameter for clinical outcome even more unorthodox. In addition, the majority of the published prognostic signatures have reportedly performed well in two of the four datasets used in the training of the CCHS (GSE17537 and GSE5206), but not in the other public datasets, which casts doubt on the usefulness of these series for such purposes (7). As a corollary, an obvious risk of this in silico training approach is dataset selection bias, which is a critical point in any genomic high-throughput analysis design and subsequent data interpretation. In addition, the inclusion of the p53 expression in the signature training process is based only on the literature and not on data, which is a questionable decision as the real role of p53 expression in prognosis of locoregional stage colorectal cancer is not clear and not recommended in any guidelines or expert consensus publication (8).

With regard to clinical validation of the score, the authors analyzed a series of 71 patients with stage II and 55 patients with stage III colorectal cancer from their own center in Leuven, Belgium. They extracted tumor RNA from formalin-fixed paraffin-embedded (FFPE) samples and analyzed it with the nCounter technique (NanoString Technologies). They then applied a multivariate analysis involving CCHS and relevant clinicopathologic variables but not MSI status, which is a recognized molecular prognostic biomarker used routinely in the clinical setting. Stage (III vs. II), extramural venous invasion (yes vs. no), and CCHS (high vs. low) were the three significant independent prognostic factors with hazard ratios of 2.58, 3.21, and 2.58, respectively. Another major limitation of this analysis is that the authors do not separate this analysis by stage of the disease, and some of the candidate clinicopathologic biomarkers are more relevant to prognosis of patients with stage II than stage III. The AMC–AJCCII-90 dataset (GSE33113), which consisted of 90 patients with stage II colorectal cancer treated at the Academic Medical Center (Amsterdam, the Netherlands), was used as a second independent clinical validation of the CCHS. This public dataset includes information describing clinical outcome such as relapse (yes vs. no), age, gender, and primary location but no other potentially prognostic clinicopathologic features or MSI status. Tumor samples from this series were fresh-frozen and analyzed with the Affymetrix Human Genome U133 Plus 2.0 Array. CCHS did indeed perform well in this series, but a multivariate analysis with other relevant clinicopathologic prognostic factors and MSI status was not feasible and would be required to infer potential future clinical utility of the score. Importantly, CCHS was validated in these two independent groups of patients using optimal non–prespecified cutoff values [receiver operating characteristic (ROC) curves] without internal cross-validation for each dataset. The authors justify this because of the distinct techniques used to measure the gene expression (nCounter vs. Affymetrix microarray), which is a fair argument. However, this downgrades their study to a previous step in the validation process (see Fig. 1), closer to what is defined by Simon as "translation of platforms" or "assay reproducibility" rather than to clinical validation (9).

Despite all the limitations described above, one strength of the work by Dekervel and colleagues is its robustness on different platforms and tumor specimens (fresh-frozen and FFPE samples). In addition, many methodologic issues are shared by the vast majority of prognostic signatures previously published. An article exploring the performance of 31 published gene expression signatures in 11 colorectal cancer gene expression datasets concludes that although most of the published signatures show significant statistical association with prognosis, their ability to accurately classify independent samples into high-risk and low-risk groups is limited. This could be explained both by methodologic weaknesses and intratumor heterogeneity. It is well known that colorectal cancer tumors are composed not only of tumor epithelial cells but also of cancer-associated stromal fibroblasts, endothelial cells, or inflammatory cells, among others. This heterogeneity in tumor cell populations might dilute the prognosis signal of reported gene expression signatures (7).

Finally, the rationale behind this innovative approach to signature discovery is biologically plausible because hypoxia can be one of the drivers of the epithelial–mesenchymal transition (EMT) and it is well known that EMT plays an important role in cancer progression (10, 11). In addition, recent hierarchical clustering of whole-genome expression data has provided new insights into the biologic and prognostic heterogeneity that supports this concept: Subtyping studies on several groups also confirmed that upregulation of the EMT phenotype, matrix remodeling, and TGF-β pathway is clearly associated with increased risk of relapse and/or reduced overall survival. Poor prognosis has also been observed in stem-like subtypes with overexpressed genes associated with mesenchymal and stem cells. In contrast, subtypes associated with an epithelial phenotype showed a better prognosis (12). In conclusion, this article by Dekervel and colleagues reports on a new candidate prognostic score based on a hypoxia-induced phenotype and a novel methodologic approach that requires validation. Despite the originality of the approach, a number of limitations are apparent that need to be addressed in future studies.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.
Authors' Contributions
Conception and design: R. Salazar, J. Tabernero
Development of methodology: R. Salazar, J. Tabernero
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): R. Salazar, J. Tabernero
Writing, review, and/or revision of the manuscript: R. Salazar, J. Tabernero

Acknowledgments
The authors thank Rebeca Sanz-Pamplona for critical reading of the article and Stephen Kelly for English editing.

Received February 18, 2014; accepted February 19, 2014; published OnlineFirst February 28, 2014.

References
New Approaches but the Same Flaws in the Search for Prognostic Signatures

Ramon Salazar and Josep Tabernero


Updated version  Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-14-0219

Cited articles  This article cites 12 articles, 5 of which you can access for free at: http://clincancerres.aacrjournals.org/content/20/8/2019.full.html#ref-list-1

Citing articles  This article has been cited by 2 HighWire-hosted articles. Access the articles at: /content/20/8/2019.full.html#related-urls

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pub@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.