The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 JANUARY 21, 2016 VOL. 374 NO. 3

CDX2 as a Prognostic Biomarker in Stage II and Stage III Colon Cancer

Piero Dalerba, M.D., Debashis Sahoo, Ph.D., Soonmyung Paik, M.D., Xiangqian Guo, Ph.D., Greg Yothers, Ph.D., Nan Song, Ph.D., Nate Wilcox-Fogel, M.S., Erna Forgó, M.D., Pradeep S. Rajendran, B.S., Stephen P. Miranda, B.A., Shigeo Hisamori, M.D., Ph.D., Jacqueline Hutchison, Tomer Kalisky, Ph.D., Dalong Qian, M.D., Norman Wolmark, M.D., George A. Fisher, M.D., Ph.D., Matt van de Rijn, M.D., Ph.D., and Michael F. Clarke, M.D.

ABSTRACT

BACKGROUND

The identification of high-risk stage II colon cancers is key to the selection of patients who require adjuvant treatment after surgery. Microarray-based multigene-expression signatures derived from stem cells and progenitor cells hold promise, but they are difficult to use in clinical practice.

METHODS

We used a new bioinformatics approach to search for biomarkers of colon epithelial differentiation across gene-expression arrays and then ranked candidate genes according to the availability of clinical-grade diagnostic assays. With the use of subgroup analysis involving independent and retrospective cohorts of patients with stage II or stage III colon cancer, the top candidate gene was tested for its association with disease-free survival and a benefit from adjuvant chemotherapy.

RESULTS

The transcription factor CDX2 ranked first in our screening test. A group of 87 of 2115 tumor samples (4.1%) lacked *CDX2* expression. In the discovery data set, which included 466 patients, the rate of 5-year disease-free survival was lower among the 32 patients (6.9%) with *CDX2*-negative colon cancers than among the 434 (93.1%) with *CDX2*-positive colon cancers (hazard ratio for disease recurrence, 3.44; 95% confidence interval [CI], 1.60 to 7.38; P=0.002). In the validation data set, which included 314 patients, the rate of 5-year disease-free survival was lower among the 38 patients (12.1%) with CDX2 protein–negative colon cancers than among the 276 (87.9%) with CDX2 protein–positive colon cancers (hazard ratio, 2.42; 95% CI, 1.36 to 4.29; P=0.003). In both these groups, these findings were independent of the patient's age, sex, and tumor stage and grade. Among patients with stage II cancer, the difference in 5-year disease-free survival was significant both in the discovery data set (49% among 15 patients with *CDX2*-negative tumors vs. 87% among 191 patients with *CDX2*-positive tumors, P=0.003) and in the validation data set (51% among 15 patients with CDX2-negative tumors vs. 80% among 106 patients with CDX2-positive tumors, P=0.004). In a pooled database of all patient cohorts, the rate of 5-year disease-free survival was higher among 23 patients with stage II CDX2-negative tumors who were treated with adjuvant chemotherapy than among 25 who were not treated with adjuvant chemotherapy (91% vs. 56%, P=0.006).

CONCLUSIONS

Lack of CDX2 expression identified a subgroup of patients with high-risk stage II colon cancer who appeared to benefit from adjuvant chemotherapy. (Funded by the National Comprehensive Cancer Network, the National Institutes of Health, and others.)

From the Herbert Irving Comprehensive Cancer Center and the Departments of Pathology and Cell Biology and Medicine, Columbia University, New York (P.D.); Institute for Stem Cell Biology and Regenerative Medicine (P.D., D.S., P.S.R., S.P.M., S.H., J.H., D.Q., M.F.C.) and the Departments of Pathology (X.G., E.F., M.R.), and Medicine, Division of Oncology (N.W.-F., G.A.F., M.F.C.), Stanford University, Stanford, and the Departments of Pediatrics and Computer Science and Engineering, University of California San Diego, San Diego (D.S.) — both in California; Faculty of Engineering, Bar-Ilan University, Ramat Gan, Israel (T.K.); the National Surgical Adjuvant Breast and Bowel Project, NRG Oncology (S.P., G.Y., N.S., N.W.) and the Allegheny Cancer Center at Allegheny General Hospital (N.W.) — both in Pittsburgh; Severance Biomedical Science Institute, Yonsei University College of Medicine, Seoul, South Korea (S.P.); and the Department of Biochemistry and Molecular Biology, Medical School of Henan University, Kaifeng, China (X.G.). Address reprint requests to Dr. Dalerba at the Herbert Irving Comprehensive Cancer Center, Columbia University, Irving Cancer Research Center, 1130 St. Nicholas Ave., Rm. 505A, New York, NY 10032, or at pdd2109@columbia.edu; or to Dr. Clarke at the Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, Lorry I. Lokey Stem Cell Research Bldg., 265 Campus Dr., Rm. G2021A, Mail Code 5461, Stanford, CA 94305, or at mfclarke@stanford.edu.

*Drs. Dalerba and Sahoo contributed equally to this article.

This article was updated on January 21, 2016, at NEJM.org.

N Engl J Med 2016;374:211-22.
DOI: 10.1056/NEJMoa1506597

procedures used in this study is provided in Supplementary Appendix 1. Complete lists of all NCBI-GEO sample number identifiers of individual gene-expression array experiments that were used to perform the various tests are provided in Tables S1 through S5 in Supplementary Appendix 1, Supplementary Appendix 2, Supplementary Appendix 3, Supplementary Appendix 4, and Supplementary Appendix 5, respectively.

IMMUNOHISTOCHEMICAL TESTING

Formalin-fixed, paraffin-embedded tissue sections were stained with 4 mg per milliliter of a mouse antihuman CDX2 monoclonal antibody that was previously validated for diagnostic applications (clone CDX2-88, BioGenex).[28,29] The staining protocol was based on recommendations from the Nordic Immunohistochemical Quality Control organization (www.nordiqc.org), which suggests heat-induced antigen retrieval with Tris buffer and EDTA (pH 9.0) (Epitope Retrieval Solution pH9, Leica).[30] Tissue slides were stained on a Bond-Max automatic stainer (Leica), and antigen detection was visualized with the use of the Bond Polymer Refine Detection kit (Leica).

ANALYSIS OF TISSUE MICROARRAYS

Colon-cancer tissue microarrays, fully annotated with clinical and pathological information, were obtained from three independent sources: 367 patients in the Cancer Diagnosis Program of the National Cancer Institute (NCI-CDP), 1519 patients in the National Surgical Adjuvant Breast and Bowel Project (NSABP) C-07 trial (NSABP C-07), and 321 patients in the Stanford Tissue Microarray Database (Stanford TMAD). A detailed description of the patient cohorts represented in each tissue microarray and of the scoring system used to evaluate CDX2 expression is provided in Figures S11 through S14 in Supplementary Appendix 1.

All tissue microarrays were scored for CDX2 expression in a blinded fashion. In cases in which tissue microarrays contained two tissue cores for a patient (i.e., two samples from distinct areas of the same tumor), the two cores were scored independently and paired at the end. If scores for the two samples were discordant, the final score for the tumor was upgraded to the higher score. All tumors in which the malignant epithelial component showed widespread nuclear expression of CDX2, either in all or a majority of cancer cells, were scored as CDX2-positive. All tumors in which the malignant epithelial component either completely lacked CDX2 expression or showed faint nuclear expression in a minority of malignant epithelial cells were scored as CDX2-negative.

The concordance between the scoring results obtained by two independent investigators was evaluated with the use of contingency tables and by calculation of Cohen's kappa indexes (Fig. S15 in Supplementary Appendix 1). The association between CDX2 expression and survival outcomes was tested by a third investigator who did not participate in the scoring process.

STATISTICAL ANALYSIS

Patient subgroups were compared with respect to survival outcomes with the use of Kaplan–Meier curves, log-rank tests, and multivariate analyses based on the Cox proportional-hazards method. Differences in the frequency of CDX2-negative cancers across different subgroups were compared with the use of Pearson's chi-square test and by computation of odds ratios together with their 95% confidence intervals. Interactions between the biomarker (CDX2 status) and adjuvant chemotherapy were evaluated with the use of the Cox proportional-hazards method in a 2-by-2 factorial design (i.e., by testing for the presence of an interaction factor between the hazard rates of the two variables).

RESULTS

IDENTIFICATION OF CDX2

The first aim of this study was to identify an actionable biomarker of poorly differentiated colon cancers (i.e., tumors depleted of mature colon epithelial cells). An actionable biomarker is one for which a clinical-grade diagnostic test had already been developed. Using a software algorithm designed for the discovery of genes with expression patterns that are linked by Boolean relationships (BooleanNet),[20] we mined a database of 2329 human colon gene-expression array experiments, searching for genes that fulfilled the "X-negative implies ALCAM-positive" Boolean implication (i.e., genes with expression that was, at the same time, absent only in *ALCAM*-positive tumors and always present in *ALCAM*-negative tumors) (Fig. S2 in Supplementary Appendix 1).

The search led to the identification of 16 candidate genes (Fig. S3 in Supplementary Appendix 1). Of these genes, only 1 gene encoded a