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, the Department of Pathology and Cell Biology, and the Department of Medicine, Division of Digestive and Liver Diseases, 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 22, 2016, at NEJM.org.

N Engl J Med 2016;374:211-22.
DOI: 10.1056/NEJMoa1506597
Copyright © 2016 Massachusetts Medical Society.

The New England Journal of Medicine
Downloaded from nejm.org at Yonsei Univ Coll of Medicine on October 12, 2016. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.

During the past decade, disease-free survival among patients with stage III colon cancer has increased significantly owing to the introduction of new adjuvant chemotherapy regimens.[1-3] This therapeutic success, however, has not translated into longer disease-free survival among patients with earlier-stage (stage I or II) cancer.[4] The lack of simple, reliable criteria for the identification of patients with early-stage disease who are at high risk for relapse has made it difficult to identify patients in whom the hazards of multiagent chemotherapy may be offset by benefits with respect to disease-specific survival.[4-9]

A Quick Take is available at NEJM.org

To address this problem, researchers have explored the possibility of stratifying patients with colon cancer according to the gene-expression profile of their tumor tissues, and they have developed multigene-expression signatures that can be used to identify high-risk colon cancers.[10-15] Although gene-expression signatures hold promise, they are difficult to use in clinical practice[16] and are often not predictive of benefit from adjuvant chemotherapy.[17]

Among the gene-expression signatures with the greatest promise are those derived from stem cells and progenitor cells.[18,19] Therefore, we initiated a systematic search for a biomarker that could be used to identify undifferentiated tumors (i.e., tumors depleted of cells with a mature phenotype) by means of immunohistochemical analysis.

To perform this search, we adopted a bioinformatics approach using Boolean logic. This approach, which was designed to discover developmentally regulated genes,[20,21] was used to identify genes with expression in colon cancer that was negatively linked to the activated leukocyte-cell adhesion molecule (ALCAM/CD166). This marker of immature colon epithelial cells is preferentially expressed at the bottom of colon crypts[22,23] and on human colon-cancer cells with enriched tumorigenic capacity in mouse xenotransplantation models.[24]

This screening test led us to identify caudal-type homeobox transcription factor 2 (CDX2) as a candidate biomarker of mature colon epithelial tissues. Using subgroup analysis involving retrospective patient cohorts, we evaluated the association of this biomarker with 5-year disease-free survival and benefit from adjuvant chemotherapy among patients with colon cancer (Fig. 1).

METHODS

BIOINFORMATICS ANALYSIS OF GENE-EXPRESSION ARRAY DATABASES

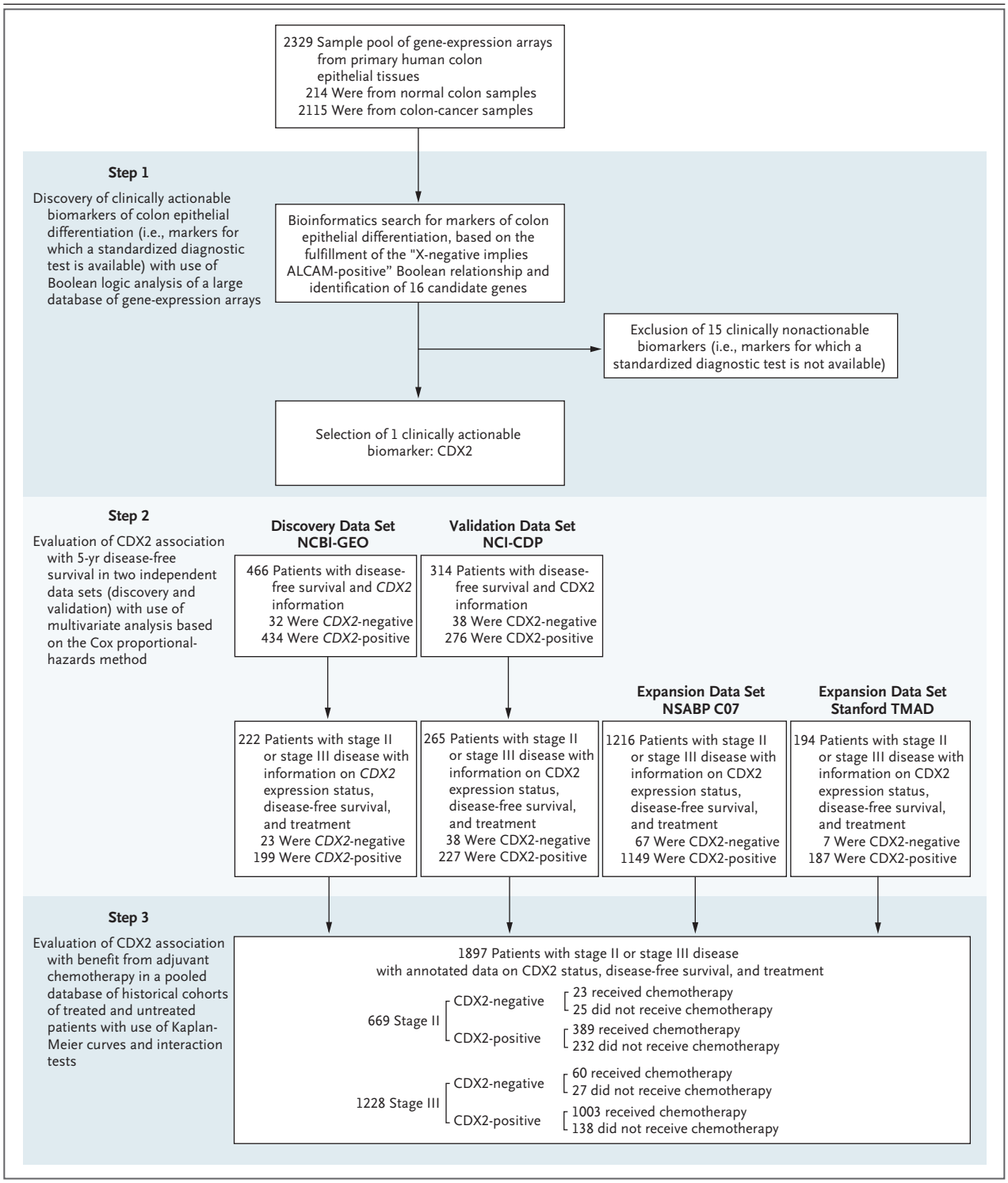
We searched for genes that fulfilled the "X-negative implies ALCAM-positive" Boolean relationship in a collection of 2329 human colon gene-expression array experiments (Fig. S1 in Supplementary Appendix 1, available with the full text of this article at NEJM.org). This collection was downloaded from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) repository (www.ncbi.nlm.nih.gov/geo). The search was conducted with the use of BooleanNet software[20] with a false discovery rate of less than 0.0001 as a cutoff point for positive results (Fig. S2 in Supplementary Appendix 1). Candidate genes were ranked according to the dynamic range of their expression levels (Fig. S3 in Supplementary Appendix 1).

The relationship between *CDX2* expression levels and other molecular features such as microsatellite instability and *TP53* mutations was studied in ad hoc collections annotated with the respective information after tumor samples were stratified into *CDX2*-negative and *CDX2*-positive subgroups with the use of the StepMiner algorithm[25] (Fig. S4 and S5 in Supplementary Appendix 1). The relationship between *CDX2* messenger RNA (mRNA) expression levels or *ALCAM* mRNA expression levels and disease-free survival was tested in a discovery data set of 466 patients. We

Figure 1 (facing page). Study Design.

A database containing 2329 human gene–expression arrays from both 214 normal colon tissue samples and 2115 colorectal-cancer tissue samples was mined to identify genes that fulfilled the "X-negative implies activated leukocyte-cell adhesion molecule (ALCAM)–positive" Boolean implication. The search yielded 16 candidate genes, of which only 1 (*CDX2*) encoded for a clinically actionable biomarker. The association between CDX2 expression and disease-free survival was tested in two independent patient cohorts: a discovery data set (National Center for Biotechnology Information Gene Expression Omnibus [NCBI-GEO]) and a validation data set (Cancer Diagnosis Program of the National Cancer Institute [NCI-CDP]). The association between CDX2 expression and benefit from adjuvant chemotherapy was tested in a pooled database of 669 patients with stage II disease and 1228 patients with stage III disease from four independent data sets (NCBI-GEO, NCI-CDP, National Surgical Adjuvant Breast and Bowel Project [NSABP] C-07 trial [NSABP C-07], and the Stanford Tissue Microarray Database [TMAD]).

The New England Journal of Medicine
Downloaded from nejm.org at Yonsei Univ Coll of Medicine on October 12, 2016. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.



obtained this data set by pooling four NCBI-GEO data sets (GSE14333, GSE17538, GSE31595, and GSE37892) (Fig. S6 in Supplementary Appendix 1).[12,13,26,27] Patients were stratified into negative-to-low (negative) and high (positive) subgroups with regard to *CDX2* and *ALCAM* gene-expression levels with the use of the StepMiner algorithm, implemented within the Hegemon[21] software (Fig. S7 through S10 in Supplementary Appendix 1).

An in-depth description of all bioinformatics

The New England Journal of Medicine
Downloaded from nejm.org at Yonsei Univ Coll of Medicine on October 12, 2016. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.

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).

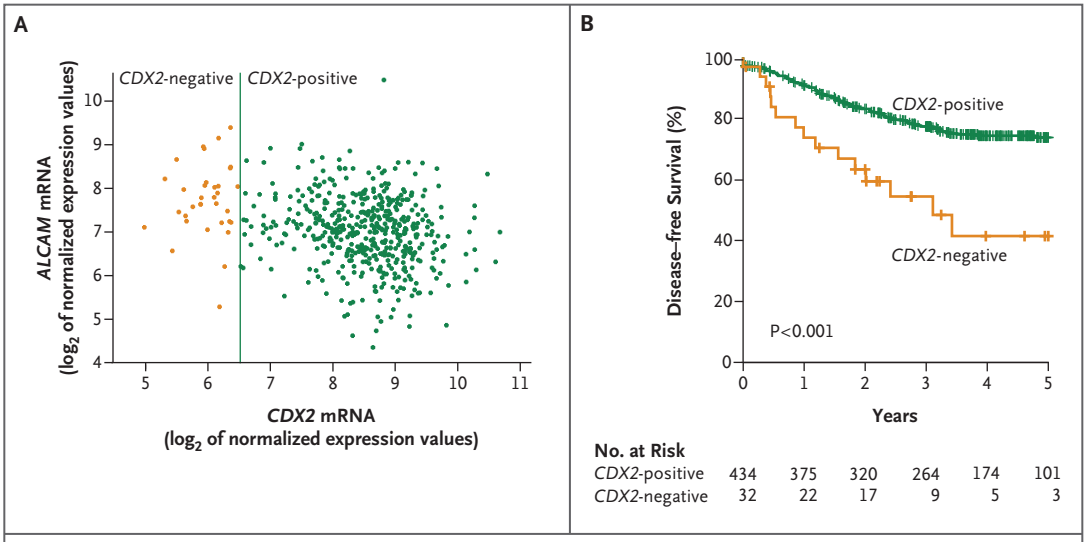
RESULTS

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

The New England Journal of Medicine
Downloaded from nejm.org at Yonsei Univ Coll of Medicine on October 12, 2016. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.



C

Subgroup	Univariate Analysis		Multivariate Analysis	
	Hazard ratio (95% CI)	P value	Hazard ratio (95% CI)	P value
All patients (N=466)				
CDX2-negative	2.71 (1.57–4.67)	<0.001	2.73 (1.58–4.72)	<0.001
Tumor stage, per increase in stage	3.47 (2.62–4.59)	<0.001	3.49 (2.61–4.67)	<0.001
Age, modeled as a continuous variable	0.99 (0.97–1.00)	0.058	0.99 (0.98–1.01)	0.37
Male vs. female sex	1.07 (0.89–1.28)	0.49	1.06 (0.88–1.28)	0.54
Patients with samples annotated with grading information (N=216)				
CDX2-negative	2.83 (1.42–5.64)	0.003	3.44 (1.60–7.38)	0.002
Tumor stage, per increase in stage	3.13 (2.14–4.60)	<0.001	3.28 (2.15–4.99)	<0.001
Tumor grade, per increase in grade	1.63 (0.94–2.82)	0.08	0.99 (0.56–1.74)	0.96
Age, modeled as a continuous variable	0.99 (0.97–1.01)	0.20	0.99 (0.97–1.01)	0.46
Male vs. female sex	1.15 (0.88–1.51)	0.32	1.20 (0.89–1.61)	0.24

Figure 2. Relationship between *CDX2* Expression and Disease-free Survival in the NCBI-GEO Discovery Data Set.
Analysis of *CDX2* messenger RNA (mRNA) expression in the NCBI-GEO discovery data set revealed the presence of a minority subgroup of *CDX2*-negative colon cancers that were characterized by high *ALCAM* mRNA expression levels (Panel A) and that were associated with a lower rate of 5-year disease-free survival than *CDX2*-positive colon cancers (Panel B). In Panel A, each circle in the scatter plot represents one patient sample. The association between *CDX2*-negative cancers and a lower rate of disease-free survival remained significant in a multivariate analysis that excluded tumor stage, tumor grade, age, and sex as confounding variables (Panel C).

protein that could be studied by means of immunohistochemical analysis with the use of a clinical-grade diagnostic test: the homeobox transcription factor CDX2.[28,29,31] CDX2 is a master regulator of intestinal development and oncogenesis,[32,33] and its expression is highly specific to the intestinal epithelium.[29] Colon cancers without CDX2 expression are often associated with an increased likelihood of aggressive features such as advanced stage, poor differentiation, vascular invasion, *BRAF* mutation, and the CpG island methylator phenotype (CIMP).[34-39]

A detailed analysis of the gene-expression relationship between *CDX2* and *ALCAM* confirmed the existence of three gene-expression groups: *CDX2*-negative and *ALCAM*-positive, *CDX2*-positive and *ALCAM*-positive, and *CDX2*-positive and *ALCAM*-negative (Fig. S2 in Supplementary Appendix 1). Lack of *CDX2* expression was restricted to a small subgroup of 87 of 2115 colorectal cancers (4.1%). This subgroup was characterized by high levels of *ALCAM* expression (Fig. S3 in Supplementary Appendix 1) and only partial overlap with tumors defined by microsatellite instability or *TP53* mutations (Fig. S4 and S5 in Supplementary Appendix 1). We thus proceeded to

The New England Journal of Medicine
Downloaded from nejm.org at Yonsei Univ Coll of Medicine on October 12, 2016. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.

evaluate the association between CDX2 expression and disease-free survival in two independent patient data sets: the NCBI-GEO discovery data set and the NCI-CDP validation data set.

CDX2 EXPRESSION AND DISEASE-FREE SURVIVAL IN THE NCBI-GEO DISCOVERY DATA SET

To evaluate the association between *CDX2* expression and disease-free survival among patients in the NCBI-GEO discovery data set, we used the StepMiner algorithm to stratify the population of 466 patients into *CDX2*-negative and *CDX2*-positive subgroups and then used Kaplan–Meier curves to compare the disease-free survival of the two subgroups (Fig. 2). The analysis showed that the rate of 5-year disease-free survival was lower among the 32 patients (6.9%) with *CDX2*-negative tumors than among the 434 (93.1%) with *CDX2*-positive tumors (41% vs. 74%, P<0.001). In a multivariate analysis that excluded age, sex, and tumor stage as confounding variables, the hazard ratio for disease recurrence among patients with *CDX2*-negative versus *CDX2*-positive tumors was 2.73 (95% confidence interval [CI], 1.58 to 4.72; P<0.001).

Within the NCBI-GEO discovery data set, data on only 216 patients were annotated with information on pathological grade (Table S1 in Supplementary Appendix 1). A multivariate analysis that was restricted to these 216 patients showed that *CDX2*-negative tumors were associated with a higher risk of recurrence than *CDX2*-positive ones (hazard ratio, 3.44; 95% CI, 1.60 to 7.38; P=0.002); the hazard ratio associated with the *CDX2* status was higher than that associated with increasing pathological grade (hazard ratio, 0.99; 95% CI, 0.56 to 1.74; P=0.96).

High levels of ALCAM expression had previously been shown to be associated with worse clinical outcomes.[23] Moreover, in the NCBI-GEO discovery data set, the rate of 5-year disease-free survival associated with ALCAM-positive tumors was moderately, but significantly lower than that associated with ALCAM-negative ones (67% vs. 78%, P=0.048) (Fig. S7 in Supplementary Appendix 1). Therefore, we evaluated whether the association between *CDX2*-negative tumors and a lower rate of disease-free survival could be explained by the fact that most *CDX2*-negative tumors were also *ALCAM*-positive. To this end, we used Hegemon software[21] to stratify the NCBI-GEO population into three subgroups (*CDX2*-negative and *ALCAM*-positive, *CDX2*-positive and *ALCAM*-positive, and *CDX2*-positive and *ALCAM*-negative) and then compared their clinical outcomes (Fig. S8 and S9 in Supplementary Appendix 1).

The results showed that *CDX2*-negative and *ALCAM*-positive tumors were associated with a lower rate of 5-year disease-free survival than *CDX2*-positive and *ALCAM*-positive and *CDX2*-positive and *ALCAM*-negative tumors. A similar set of tests also indicated that when compared side by side with the use of multivariate analysis, the hazard ratios for disease recurrence associated with the *CDX2* and *ALCAM* grouping system were higher than those associated with the "intestinal stem-cell" gene-expression signature[19] (Fig. S10 in Supplementary Appendix 1).

CDX2 EXPRESSION AND DISEASE-FREE SURVIVAL IN THE NCI-CDP VALIDATION DATA SET

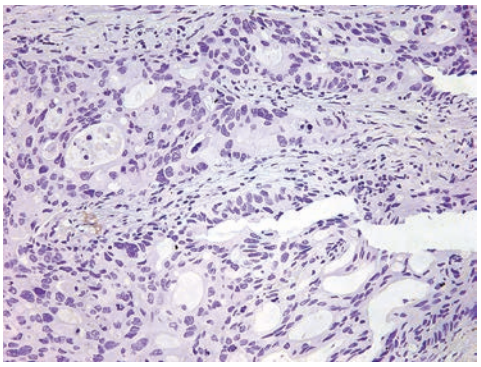
To evaluate the robustness of our findings, we decided to test whether they could be reproduced in an independent data set,[40] and we chose to analyze a human colon-cancer tissue microarray obtained from the NCI-CDP. This microarray was explicitly designed to contain a balanced distribution of patients with and without tumor recurrence, as well as with a relatively homogeneous long-term follow-up, with the aim to maximize the statistical power to find associations between biomarkers and clinical outcomes.

Figure 3 (facing page). Relationship between CDX2 Protein Expression and Disease-free Survival in the NCI-CDP Validation Data Set.

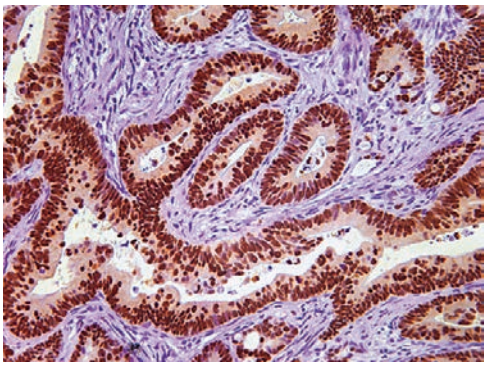
Analysis of CDX2 protein expression in the NCI-CDP validation data set confirmed the existence of a minority subgroup of CDX2-negative cancers (Panel A) that lacked the distinctive CDX2 nuclear expression that is characteristic of epithelial cancer cells in the majority of colon cancers (Panel B). CDX2-negative cancers were associated with a lower rate of 5-year disease-free survival than CDX2-positive cancers (Panel C). The association between the absence of CDX2 expression and a lower rate of 5-year disease-free survival was confirmed by means of a multivariate analysis (based on the Cox proportional-hazards method) that excluded tumor stage, tumor grade, age, and sex as confounding variables (Panel D). CDX2-negative tumors were associated with a lower rate of survival independent of their subclassification with regard to low or intermediate (G1 or G2) or high (G3 or G4) pathological grade (Panel E).

The New England Journal of Medicine
Downloaded from nejm.org at Yonsei Univ Coll of Medicine on October 12, 2016. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.

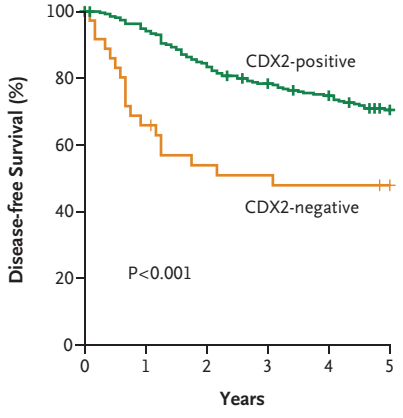
A Specimen without CDX2 Nuclear Expression



B Specimen with CDX2 Nuclear Expression



C Disease-free Survival, According to CDX2 Expression



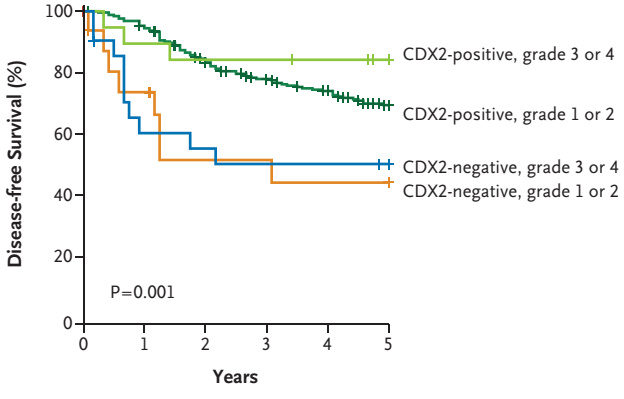
No. at Risk

CDX2-positive	276	258	225	199	182	150
CDX2-negative	38	23	18	17	16	15

D Multivariate Analysis

Subgroup	Hazard Ratio (95% CI)	P Value
CDX2-negative	2.42 (1.36–4.29)	0.003
Tumor stage, according to increase in stage	2.71 (1.92–3.84)	<0.001
Tumor grade, according to increase in grade	0.79 (0.61–1.03)	0.08
Age, modeled as a continuous variable	1.00 (0.99–1.02)	0.68
Male vs. female sex	0.91 (0.61–1.35)	0.63

E Disease-free Survival, According to CDX2 Expression and Tumor Grade



No. at Risk

CDX2-positive, grade 1 or 2	257	241	209	183	167	137
CDX2-positive, grade 3 or 4	19	17	16	16	15	13
CDX2-negative, grade 1 or 2	16	11	7	7	6	6
CDX2-negative, grade 3 or 4	22	12	11	10	10	9

The New England Journal of Medicine
Downloaded from nejm.org at Yonsei Univ Coll of Medicine on October 12, 2016. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.

To evaluate CDX2 protein expression, we used immunohistochemical analysis with an anti-CDX2 monoclonal antibody that had previously been validated for diagnostic purposes.[28,29] Analysis of stained sections confirmed the presence of a minority subgroup of cancers that lacked expression of CDX2 protein in malignant epithelial cells, as compared with the majority of samples that had intense nuclear staining (Fig. 3). On the basis of these results, we stratified the patient cohort into two subgroups: CDX2-negative (48 of 366 patients [13%]) and CDX2-positive (318 of 366 patients [87%]). A description of the scoring system and its performance in terms of interobserver agreement is provided in Figures S14 and S15 in Supplementary Appendix 1.

CDX2-negative tumors were associated with a worse prognosis than were CDX2-positive tumors, with lower rates of 5-year disease-free survival (48% vs. 71%, P<0.001) (Fig. 3), overall survival (33% vs. 59%, P<0.001) (Fig. S16 in Supplementary Appendix 1), and disease-specific survival (45% vs. 72%, P<0.001) (Fig. S16 in Supplementary Appendix 1). The association remained significant in multivariate analyses that excluded age, sex, tumor stage, and tumor grade as confounding variables: in the analysis of disease-free survival, the hazard ratio for disease recurrence associated with CDX2-negative tumors as compared with CDX2-positive tumors was 2.42 (95% CI, 1.36 to 4.29; P=0.003); in the analysis of overall survival, the hazard ratio for death was 1.79 (95% CI, 1.18 to 2.71; P=0.006); and in the analysis of disease-specific death, the hazard ratio for death was 2.09 (95% CI, 1.22 to 3.58; P=0.007).

CDX2-negative status was more common among tumors with a high pathological grade (Fig. S17 in Supplementary Appendix 1). However, CDX2-negative tumors were associated with a lower rate of survival irrespective of their low or intermediate (G1 or G2) or high (G3 or G4) pathological grade — a finding consistent with the results of the multivariate analysis (Fig. 3, and Fig. S17 in Supplementary Appendix 1).

CDX2 Expression and Survival among Patients with Stage II Disease

To evaluate our findings with respect to the prognosis among patients with early-stage colon cancer, we decided to study the association between the CDX2-negative phenotype, assessed at either the mRNA or protein level, and disease-free survival among patients with stage II disease. Stage II CDX2-negative tumors were associated with a lower rate of 5-year disease-free survival than were stage II CDX2-positive tumors in both the NCBI-GEO discovery data set (49% vs. 87%, P=0.003) (Fig. 4) and the NCI-CDP validation data set (51% vs. 80%, P=0.004) (Fig. 4).

We found similar associations with respect to overall survival (40% among patients with CDX2-negative tumors vs. 70% among those with CDX2-positive tumors, P<0.001) (Fig. S18 in Supplementary Appendix 1) and disease-specific survival (66% vs. 89%, P=0.005) (Fig. S18 in Supplementary Appendix 1). These associations were not confounded by risk factors that are known to affect survival rates among patients with stage II colon cancer, such as the depth of invasion of the primary tumor (T3 vs. T4) (Fig. S19 in Supplementary Appendix 1) and the number of lymph nodes resected at surgery (≥12 vs. <12) (Fig. S19 in Supplementary Appendix 1). However, in each of the two data sets, only 15 patients with stage II CDX2-negative disease were identified.

CDX2 Expression and Benefit from Adjuvant Chemotherapy

To evaluate whether patients with CDX2-negative tumors might benefit from adjuvant chemotherapy, we investigated the association between CDX2 status, assessed at either the mRNA or protein level, and disease-free survival among patients who either did or did not receive adjuvant chemotherapy. A preliminary test involving cohorts of patients with stage III disease in both the discovery and validation data sets suggested a strong association between the use of adjuvant chemotherapy and a higher rate of disease-free survival in the CDX2-negative subgroups (Fig. S20 in Supplementary Appendix 1).

We thus decided to validate this observation in an expanded database of 669 patients with stage II colon cancer and 1228 patients with stage III colon cancer. We obtained this database by pooling data from four independent patient cohorts (NCBI-GEO, NCI-CDP, NSABP C-07, and Stanford TMAD); these data were annotated with information about adjuvant chemotherapy (Fig. 1). A detailed description of all patient data sets used for this experiment is provided in Figure S6 and Figures S11, S12, and S13 in Supplementary Appendix 1.

The results confirmed that treatment with

The New England Journal of Medicine
Downloaded from nejm.org at Yonsei Univ Coll of Medicine on October 12, 2016. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.

adjuvant chemotherapy was associated with a higher rate of disease-free survival in both the stage II subgroup (91% with chemotherapy vs. 56% with no chemotherapy, P=0.006) and the stage III subgroup (74% with chemotherapy vs. 37% with no chemotherapy, P<0.001) of the CDX2-negative patient population (Fig. 5). A test for the interaction between the biomarker and the treatment revealed that the benefit observed in CDX2-negative cohorts was superior to that observed in CDX2-positive cohorts in both the stage II subgroup (P=0.02 for the interaction) and the stage III subgroup (P=0.005 for the interaction). The association between CDX2-negative status and benefit from adjuvant chemotherapy was not confounded by risk factors that are known to affect the survival rates among patients with stage II and stage III disease. These risk factors include the depth of invasion of the primary tumor (T3 vs. T4), the number of lymph nodes resected at surgery (≥12 vs. <12), and the number of metastatic lymph nodes (N1 vs. N2) (Figs. S21 through S24 in Supplementary Appendix 1).

DISCUSSION

Prognostic biomarkers are key to the risk stratification of patients with colon cancer and the decision to recommend adjuvant chemotherapy in patients with early-stage disease.[6] Currently, tumor stage, tumor grade, and microsatellite instability remain the most important among a handful of prognostic variables that are considered in the development of algorithms for the treatment of patients with early-stage colon cancer.[5,9] Prognostic variables such as lymphovascular invasion by cancer cells and perineural invasion by cancer cells, though very promising, have proved difficult to standardize because of technical problems inherent in the visual analysis and subjective definition of these features.[6] Microarray-derived gene-expression signatures from stem cells and progenitor cells have also shown promise,[19] but they are often difficult to translate into clinical tests.[16] Overall, it has proved difficult to identify a prognostic biomarker that is also predictive of benefit from adjuvant chemotherapy.[7,8,17]

In this study, we applied a bioinformatics approach to the discovery of prognostic biomarkers in human cancer. We assembled a large database of gene-expression array experiments involving human colorectal cancers and searched for genes with differential expression, defined by a Boolean relationship with respect to a well-established differentiation marker, across the patient population. The concept behind this approach was that genes associated with differentiation processes (e.g., transcription factors involved in the regulation of stem-cell self-renewal, lineage commitment, or both) could be identified as single prognostic biomarkers that could be used to stratify tumors on the basis of a

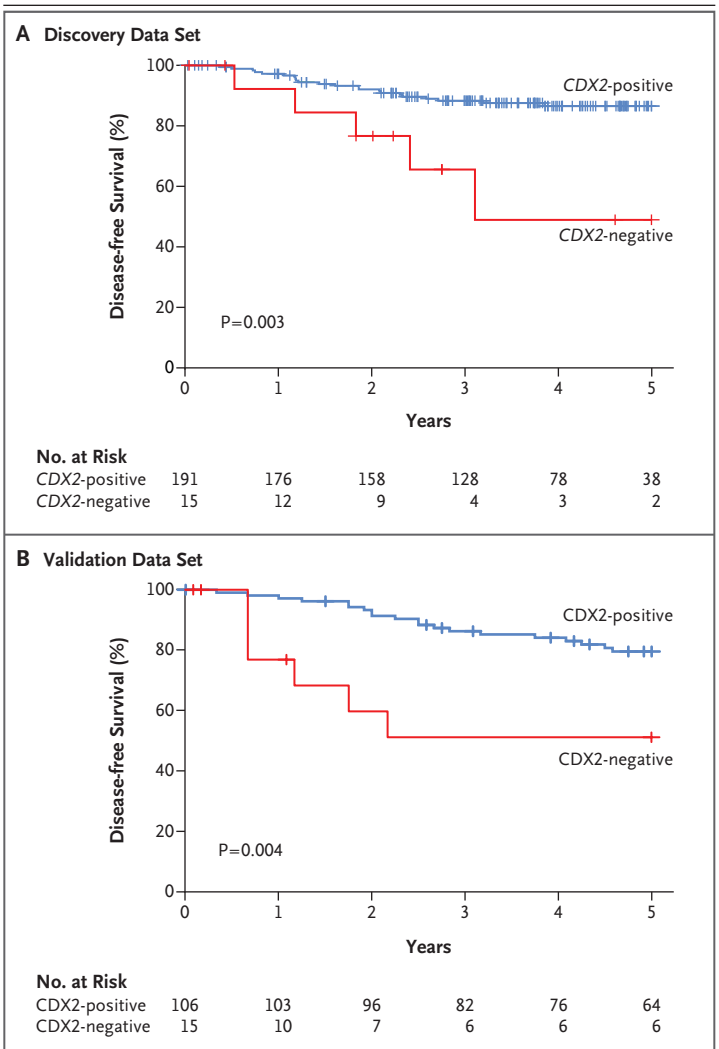


Figure 4. Relationship between CDX2 Expression and Disease-free Survival among Patients with Stage II Disease.

In the NCBI-GEO discovery data set (Panel A), *CDX2*-negative cancers were associated with a rate of 5-year disease-free survival that was lower than the rate associated with *CDX2*-positive cancers. In the NCI-CDP validation data set (Panel B), CDX2-negative cancers were associated with a rate of 5-year disease-free survival that was lower than the rate associated with CDX2-positive cancers.

The New England Journal of Medicine
Downloaded from nejm.org at Yonsei Univ Coll of Medicine on October 12, 2016. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.

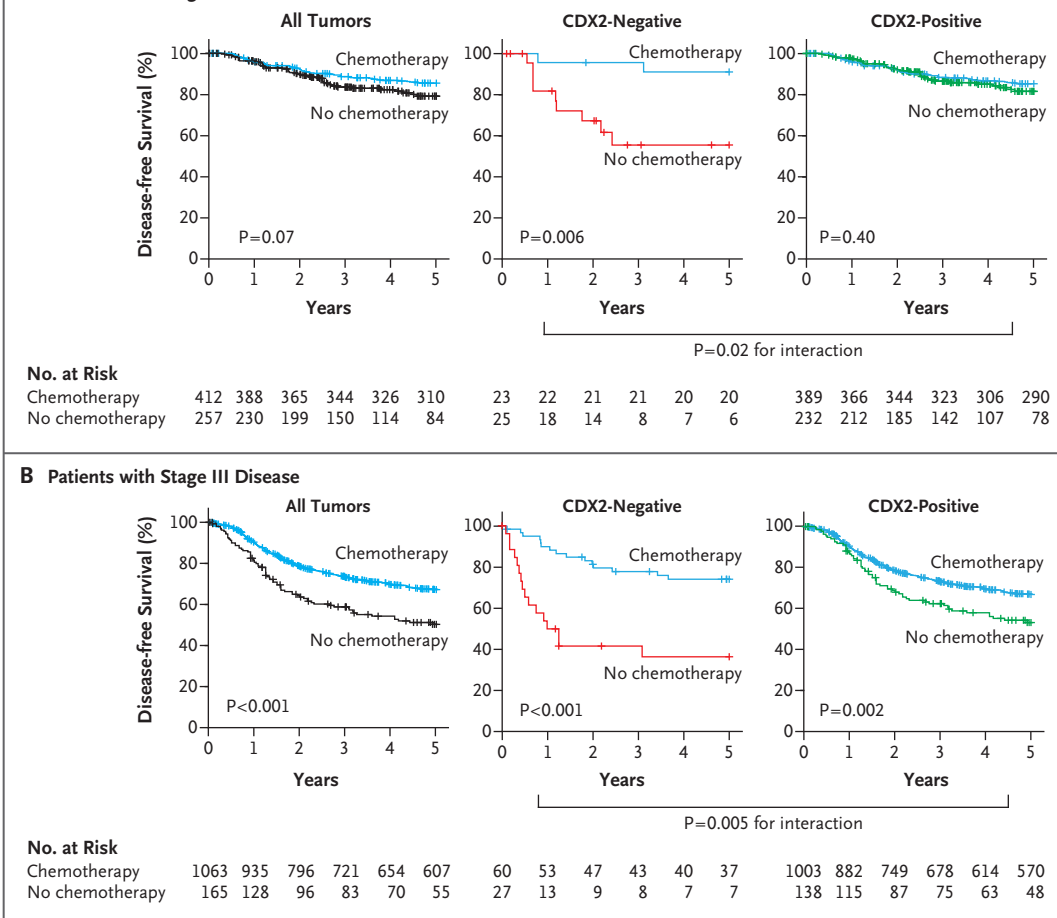


Figure 5. Relationship between CDX2 Expression and Benefit from Adjuvant Chemotherapy.

The relationship between CDX2 expression and benefit from adjuvant chemotherapy was evaluated in a pooled database of 669 patients with stage II disease (Panel A) and 1228 patients with stage III disease (Panel B) from four independent data sets (NCBI-GEO, NCI-CDP, NSABP C-07, and Stanford TMAD). Among all patients with stage II disease in the entire database, treatment with adjuvant chemotherapy was not associated with a higher rate of 5-year disease-free survival. However, treatment with adjuvant chemotherapy was strongly associated with a higher rate of 5-year disease-free survival in the CDX2-negative subgroup, but it was not associated with a higher rate of 5-year disease-free survival in the CDX2-positive subgroup. Among patients with stage III disease, treatment with adjuvant chemotherapy was associated with a higher rate of 5-year disease-free survival in the entire database and in both the CDX2-negative and CDX2-positive subgroups. A test for an interaction between the biomarker and the treatment indicated that in both stage II and stage III disease, the benefit associated with adjuvant chemotherapy was superior among CDX2-negative patients than among CDX2-positive patients.

molecular definition of their differentiation status and to recapitulate the prognostic information contained in complex multigene-expression signatures obtained from populations of stem cells and progenitor cells.

Using this approach, we identified CDX2 as a biomarker with expression that has been found to be absent in a minority subgroup of colon cancers that are characterized by high levels of ALCAM, a molecule that is expressed at the highest levels at the bottom of human colonic crypts[22,23] and on human colon-cancer cells with enriched tumorigenic capacity in mouse xenotransplantation models.[24] We then performed a test to determine whether CDX2-negative cancers might be associated with a worse progno-

The New England Journal of Medicine
Downloaded from nejm.org at Yonsei Univ Coll of Medicine on October 12, 2016. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.

sis. The results revealed that without adjuvant chemotherapy, CDX2-negative tumors were associated with a lower rate of disease-free survival than CDX2-positive tumors across independent data sets. This effect was independent of many known risk factors, including pathological grade.

Previous studies had indicated that CDX2-negative tumors are often associated with several adverse prognostic variables (e.g., advanced stage, poor differentiation, vascular invasion, *BRAF* mutation, and CIMP-positive status).[31,35-38] We hypothesize that the prognostic effect associated with an absence of CDX2 expression could be explained by its aggregate capacity to function as a single biomarker for multiple biologic risk factors, under the common theme of a highly immature progenitor-cell phenotype.

Finally, our results indicate that patients with stage II or stage III CDX2-negative colon cancer might benefit from adjuvant chemotherapy and that adjuvant chemotherapy might be a treatment option for patients with stage II CDX2-negative disease, who are commonly treated with surgery alone. Given the exploratory and retrospective design of our study, these results will need to be further validated. We advocate for these findings to be confirmed within the framework of randomized, clinical trials, in conjunction with genomic DNA sequencing studies.

All opinions, findings, and conclusions expressed in this article are those of the authors and do not necessarily reflect those of the National Comprehensive Cancer Network (NCCN) or the NCCN Foundation.

Supported by an NCCN 2012 Young Investigator Award (to Dr. Dalerba); National Institutes of Health (NIH) grants (U54-CA126524 and P01-CA139490, to Dr. Clarke, and R00-CA151673, to Dr. Sahoo); a grant from the Siebel Stem Cell Institute and the Thomas and Stacey Siebel Foundation (to Drs. Dalerba and Sahoo); a grant from the Virginia and D.K. Ludwig Fund for Cancer Research (to Dr. Clarke); a California Institute for Regenerative Medicine training grant (to Dr. Dalerba); a Department of Defense grant (W81XWH-10-1-0500, to Dr. Sahoo); a Bladder Cancer Advocacy Network 2013 Young Investigator Award (to Dr. Sahoo); and a BD Biosciences 2011 Stem Cell Research Grant (to Dr. Dalerba). The National Surgical Adjuvant Breast and Bowel Project was supported by NIH grants U24-CA114732, U10-CA37377, U10-CA180868, U10-CA180822, UG1-CA189867, and U24-CA196067. Some of the tissue microarrays used in this study were provided by the Cooperative Human Tissue Network and the Cancer Diagnosis Program, which are funded by the National Cancer Institute. Other investigators may have received slides from these same array blocks.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank Edward Gilbert, Chona Enrile, Ivy Mangonon, Marissa Palmor, Coralie Donkers, and Darius M. Johnston, all of Stanford University, for help and support during various phases of the study.

REFERENCES

1. André T, Boni C, Mounedji-Boudiaf L, et al. Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. N Engl J Med 2004;350:2343-51.
2. Meyerhardt JA, Mayer RJ. Systemic therapy for colorectal cancer. N Engl J Med 2005;352:476-87.
3. Saltz LB, Cox JV, Blanke C, et al. Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. N Engl J Med 2000;343:905-14.
4. O'Connor ES, Greenblatt DY, LoConte NK, et al. Adjuvant chemotherapy for stage II colon cancer with poor prognostic features. J Clin Oncol 2011;29:3381-8.
5. Bardia A, Loprinzi C, Grothey A, et al. Adjuvant chemotherapy for resected stage II and III colon cancer: comparison of two widely used prognostic calculators. Semin Oncol 2010;37:39-46.
6. Compton C, Fenoglio-Preiser CM, Pettigrew N, Fielding LP. American Joint Committee on Cancer Prognostic Factors Consensus Conference: Colorectal Working Group. Cancer 2000;88:1739-57.
7. Gill S, Loprinzi CL, Sargent DJ, et al. Pooled analysis of fluorouracil-based adjuvant therapy for stage II and III colon cancer: who benefits and by how much? J Clin Oncol 2004;22:1797-806.
8. Meropol NJ. Ongoing challenge of stage II colon cancer. J Clin Oncol 2011; 29:3346-8.
9. Tournigand C, de Gramont A. Chemotherapy: is adjuvant chemotherapy an option for stage II colon cancer? Nat Rev Clin Oncol 2011;8:574-6.
10. Barrier A, Boelle PY, Roser F, et al. Stage II colon cancer prognosis prediction by tumor gene expression profiling. J Clin Oncol 2006;24:4685-91.
11. Wang Y, Jatkoe T, Zhang Y, et al. Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. J Clin Oncol 2004;22:1564-71.
12. Jorissen RN, Gibbs P, Christie M, et al. Metastasis-associated gene expression changes predict poor outcomes in patients with Dukes stage B and C colorectal cancer. Clin Cancer Res 2009;15:7642-51.
13. Smith JJ, Deane NG, Wu F, et al. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. Gastroenterology 2010;138:958-68.
14. Yothers G, O'Connell MJ, Lee M, et al. Validation of the 12-gene colon cancer recurrence score in NSABP C-07 as a predictor of recurrence in patients with stage II and III colon cancer treated with fluorouracil and leucovorin (FU/LV) and FU/LV plus oxaliplatin. J Clin Oncol 2013;31: 4512-9.
15. Fang SH, Efron JE, Berho ME, Wexner SD. Dilemma of stage II colon cancer and decision making for adjuvant chemotherapy. J Am Coll Surg 2014;219:1056-69.
16. Gröne J, Lenze D, Jurinovic V, et al. Molecular profiles and clinical outcome of stage UICC II colon cancer patients. Int J Colorectal Dis 2011;26:847-58.
17. National Comprehensive Cancer Network. Clinical practice guidelines in oncology — colon cancer, version 3. 2015 (http://www.nccn.org).
18. Liu R, Wang X, Chen GY, et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. N Engl J Med 2007;356:217-26.
19. Merlos-Suárez A, Barriga FM, Jung P, et al. The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. Cell Stem Cell 2011;8:511-24.
20. Sahoo D, Dill DL, Gentles AJ, Tibshirani R, Plevritis SK. Boolean implication networks derived from large scale, whole genome microarray datasets. Genome Biol 2008;9:R157.

The New England Journal of Medicine
Downloaded from nejm.org at Yonsei Univ Coll of Medicine on October 12, 2016. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.

21. Dalerba P, Kalisky T, Sahoo D, et al. Single-cell dissection of transcriptional heterogeneity in human colon tumors. Nat Biotechnol 2011;29:1120-7.
22. Levin TG, Powell AE, Davies PS, et al. Characterization of the intestinal cancer stem cell marker CD166 in the human and mouse gastrointestinal tract. Gastroenterology 2010;139(6):2072-2082.e5.
23. Weichert W, Knösel T, Bellach J, Dietel M, Kristiansen G. ALCAM/CD166 is overexpressed in colorectal carcinoma and correlates with shortened patient survival. J Clin Pathol 2004;57:1160-4.
24. Dalerba P, Dylla SJ, Park IK, et al. Phenotypic characterization of human colorectal cancer stem cells. Proc Natl Acad Sci U S A 2007;104:10158-63.
25. Sahoo D, Dill DL, Tibshirani R, Plevritis SK. Extracting binary signals from microarray time-course data. Nucleic Acids Res 2007;35:3705-12.
26. Thorsteinsson M, Kirkeby LT, Hansen R, et al. Gene expression profiles in stages II and III colon cancers: application of a 128-gene signature. Int J Colorectal Dis 2012;27:1579-86.
27. Laibe S, Lagarde A, Ferrari A, Monges G, Birnbaum D, Olschwang S. A seven-gene signature aggregates a subgroup of stage II colon cancers with stage III. OMICS 2012;16:560-5.
28. Li MK, Folpe AL. CDX-2, a new marker for adenocarcinoma of gastrointestinal origin. Adv Anat Pathol 2004;11:101-5.
29. Werling RW, Yaziji H, Bacchi CE, Gown AM. CDX2, a highly sensitive and specific marker of adenocarcinomas of intestinal origin: an immunohistochemical survey of 476 primary and metastatic carcinomas. Am J Surg Pathol 2003;27:303-10.
30. Borrisholt M, Nielsen S, Vyberg M. Demonstration of CDX2 is highly antibody dependant. Appl Immunohistochem Mol Morphol 2013;21:64-72.
31. Kaimaktchiev V, Terracciano L, Tornillo L, et al. The homeobox intestinal differentiation factor CDX2 is selectively expressed in gastrointestinal adenocarcinomas. Mod Pathol 2004;17:1392-9.
32. Beck F, Stringer EJ. The role of Cdx genes in the gut and in axial development. Biochem Soc Trans 2010;38:353-7.
33. Chawengsaksophak K, James R, Hammond VE, Köntgen F, Beck F. Homeosis and intestinal tumours in Cdx2 mutant mice. Nature 1997;386:84-7.
34. Hinoi T, Tani M, Lucas PC, et al. Loss of CDX2 expression and microsatellite instability are prominent features of large cell minimally differentiated carcinomas of the colon. Am J Pathol 2001;159:2239-48.
35. Lugli A, Tzankov A, Zlobec I, Terracciano LM. Differential diagnostic and functional role of the multi-marker phenotype CDX2/CK20/CK7 in colorectal cancer stratified by mismatch repair status. Mod Pathol 2008;21:1403-12.
36. Baba Y, Nosho K, Shima K, et al. Relationship of CDX2 loss with molecular features and prognosis in colorectal cancer. Clin Cancer Res 2009;15:4665-73.
37. Zlobec I, Bihl MP, Schwarb H, Terracciano L, Lugli A. Clinicopathological and protein characterization of BRAF- and K-RAS-mutated colorectal cancer and implications for prognosis. Int J Cancer 2010;127:367-80.
38. Bae JM, Lee TH, Cho NY, Kim TY, Kang GH. Loss of CDX2 expression is associated with poor prognosis in colorectal cancer patients. World J Gastroenterol 2015;21:1457-67.
39. De Sousa E Melo F, Wang X, Jansen M, et al. Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. Nat Med 2013;19:614-8.
40. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. PLoS Med 2012;9(5):e1001216.
Copyright © 2016 Massachusetts Medical Society.

We are seeking a **Deputy Editor** to join our editorial team at the *New England Journal of Medicine*. The Deputy Editor will review, select, and edit manuscripts for the *Journal* as well as participate in planning the *Journal*'s content. He or she will also ensure that the best available research is submitted to the *Journal*, identify potential *Journal* contributors, and invite submissions as appropriate.

Applicants must have an M.D. degree and at least 5 years of research experience. The successful candidate will be actively involved in patient care and will continue that involvement (approximately 20% of time) while serving in this role.

To apply, visit **https://mmscareers.silkroad.com**.

The New England Journal of Medicine
Downloaded from nejm.org at Yonsei Univ Coll of Medicine on October 12, 2016. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.