Clinical Validation of Colorectal Cancer Biomarkers Identified from Bioinformatics Analysis of Public Expression Data

Yeonjoo Jung\textsuperscript{1,2}, Sanghyuk Lee\textsuperscript{1,2,3}, Hyung-Seok Choi\textsuperscript{1,2}, Soon-Nam Kim\textsuperscript{1,2}, Eunyoung Lee\textsuperscript{1,2}, Youngah Shin\textsuperscript{1,2}, Jihae Seo\textsuperscript{1,2}, Bumjin Kim\textsuperscript{1,2}, Yeonhwa Jung\textsuperscript{1,2}, Wan Kyu Kim\textsuperscript{1,2}, Ho-Kyung Chun\textsuperscript{4}, Woo Yong Lee\textsuperscript{4}, and Jaesang Kim\textsuperscript{1,2}

Abstract

\textbf{Purpose:} Identification of novel biomarkers of cancer is important for improved diagnosis, prognosis, and therapeutic intervention. This study aimed to identify marker genes of colorectal cancer (CRC) by combining bioinformatics analysis of gene expression data and validation experiments using patient samples and to examine the potential connection between validated markers and the established oncogenes such as c-Myc and K-ras.

\textbf{Experimental Design:} Publicly available data from GenBank and Oncomine were meta-analyzed leading to 34 candidate marker genes of CRC. Multiple case-matched normal and tumor tissues were examined by RT-PCR for differential expression, and 9 genes were validated as CRC biomarkers. Statistical analyses for correlation with major clinical parameters were carried out, and RNA interference was used to examine connection with major oncogenes.

\textbf{Results:} We show with high confidence that 9 (ECT2, ETV4, DDX21, RAN, S100A11, RPS4X, HSPD1, CKS2, and C9orf140) of the 34 candidate genes are expressed at significantly elevated levels in CRC tissues compared to normal tissues. Furthermore, high-level expression of RPS4X was associated with non- mucinous cancer cell type and that of ECT2 with lack of lymphatic invasion while upregulation of CKS2 was correlated with early tumor stage and lack of family history of CRC. We also demonstrate that RPS4X and DDX21 are regulatory targets of c-Myc and ETV4 is downstream to K-ras signaling.

\textbf{Conclusions:} We have identified multiple novel biomarkers of CRC. Further analyses of their function and connection to signaling pathways may reveal potential value of these biomarkers in diagnosis, prognosis, and treatment of CRC. \textit{Clin Cancer Res}; 17(4); 700–9. ©2011 AACR.

Introduction

Colorectal cancer is currently one of the most prevalent types of cancer. In the United States, it is the second and third highest ranked cancer in mortality and incidence, respectively (1). The incidence of colorectal cancer is also rising rapidly in other parts of the world including several Asian countries for which westernized dietary lifestyle has been implicated as the major cause (2, 3). Five-year survival is dramatically improved if diagnosis is made at early stages, but limited access to effective colonoscopic screening remains a problem (1). Another difficulty associated with colorectal cancer is that once distant metastasis has occurred, the survival rate is extremely poor (1). Clearly, identification and functional characterization of novel biomarkers for better diagnosis and prognosis and for gaining insights into molecular nature of the disease itself would be beneficial.

One potential source of useful information is the large body of publicly available gene expression data. Expressed sequence tag (EST) libraries and serial analysis of gene expression (SAGE) libraries represent two of such databases. Numerous candidate biomarkers identified from these expression data have been tested positive for utility in risk assessment, diagnosis, prognosis, and monitoring therapeutic efficacy (4, 5). Microarray screening is another powerful high-throughput molecular method providing quantitative expression profiles of large number of genes. Much of the results from microarray screen is also accessible and has been applied for subclassification of diseases and identification of early diagnostic markers (6–8).

We have previously reported a successful development and application of a bioinformatics pipeline to identify...
novel biomarkers for lung cancer (9). Our strategy features
means to overcome the difficulty stemming from biological
sample heterogeneity and varying formats of the expression
data. The process is essentially composed of a serial incor-
poration and filtering of EST and SAGE data followed by
integrative meta-analysis of microarray expression data to
generate a list of candidate genes differentially expressed in
lung cancers. This was followed by a case-matched clinical
validation study with multiple patient samples that ulti-
mately led to confirmation of at least 2 genes as highly
probable novel biomarkers.

One of the strengths of our integrative strategy is that it
can be easily adapted to other types of cancer. In this report,
we present bioinformatics analysis and clinical validation
of multiple candidate biomarkers of colorectal cancer. Thus
identified 9 marker genes (ECT2, ETV4, DDX21, RAN,
S100A11, RPS4X, HSPD1, CKS2, and C9orf140) were sub-
sequently subjected to preliminary examination to assess
their diagnostic value. It was found that at least 3 genes,
RPS4X, ECT2, and CKS2 feature statistically significant
correlation with notable clinical characteristics of CRC.
We also present results from analysis of potential involve-
ment of the identified genes in major oncogenic signaling
pathways including those pivoted by c-Myc and K-ras. We
demonstrate that RPS4X and DDX21 are regulatory targets
of c-Myc and that ETV4 is a downstream gene in the K-ras
pathway.

Materials and Methods

Analysis of the expressed sequence tags and
microarray data

A bioinformatics pipeline was established to analyze all
EST data in the GenBank and cancer microarrays in the
ONCOMINE database (Fig. 1; ref. 9). The initial screening
for biomarker genes was based on the information asso-
ciated with the ESTs themselves. We used the ECgene
database developed in our own lab to isolate EST clusters,
each representing a group of transcripts from a single gene
(10). EST clusters without valid official gene symbol were
discarded. Clusters consisting of less than 10 ESTs from all
sources or less than 5 ESTs from colon tissues were also
excluded. Tissue and cancer type specificities of gene
expression for each cluster were estimated by determining
the ratio of EST numbers from the source of interest and
from the corresponding control tissue. Specifically, the
minimum specificities were set as 5% for colon tissue
(i.e., the ratio of the number of ESTs for a given gene
cluster from colon tissues and the number of ESTs for the
same gene cluster from all tissues should be above
5%) and 30% for colorectal cancer tissues (i.e., the ratio
of the number of ESTs for a given gene cluster from colorectal
cancer tissues and the number of ESTs for the same gene
cluster from all colon tissues should be over 30%). With
these criteria, we obtained 388 EST clusters from nonnor-
malized libraries while counting ESTs from all libraries,
nonnormalized and normalized, gave 722 clusters as initial
candidates. The second step of screening was carried out
using the microarray data in the ONCOMINE database that
included 4 studies on colon cancers comparing normal
versus cancer tissues—by Graudens and colleagues (11
(normal, 18 colorectal carcinoma; ref. 11), by Alon and
colleagues (22 normal, 40 colon adenocarcinoma; ref. 12),
by Zou and colleagues (8 normal, 9 colon carcinoma; ref.
13), and by Notterman and colleagues (18 normal,
18 adenocarcinoma; ref. 14). Lists of overexpressed genes

Translational Relevance

This study was initiated by examining publicly accu-
mulated data on gene expression in colorectal cancer
(CRC) with the purpose of finding biomarkers of clinical
relevance. Accordingly, we examined the high-
expression candidate genes using patient samples and
determined that 3 genes show statistically significant
correlation with major clinical characteristics of CRC.
Therefore, we believe these markers potentially have
diagnostic values pending further confirmation.
Another clinically relevant finding is that ETV4 is a
downstream target of K-ras signaling. Patients with
K-ras mutation show poor response to anti-EGFR ther-
apy as do those with mutations in BRAF, a downstream
gene to K-ras. Such findings indicate that checking for
the activation status of the K-ras signaling pathway may
be valuable in fine-tuning EGFR inhibitor therapy.
Therefore, if ETV4 expression can be used as a marker
for the activated K-ras pathway, another method of
predicting the response to anti-EGFR therapy may be
developed (see the "Discussion" section).

Figure 1. Schematic diagram of the overall procedure.
in carcinoma were available from each study, and only the genes within upper 50% in all 4 studies were selected. This filtering process reduced the candidate genes into 53 from non-normalized libraries and 71 from all libraries summing to 91 in the end. For the experimental validation, further selection based on cancer specificity and differential expression levels was carried out (see the “Results” section).

**Patient samples and RNA isolation from tissues**

Primary colorectal cancer and noncancerous colon samples were obtained from patients who had undergone curative surgical procedure at the Samsung Medical Center in Seoul, Korea. The study was approved by the local institutional review board. In each of the 29 patient cases, cancer tissues and tissues surrounding the cancer free of diseases as judged by histological examination were isolated and used for the case-matched studies. Information regarding major clinical parameters including gender, age, histological tumor subtype, tumor stage, lymphatic invasion, distant metastasis, polyp formation, family history of CRC, smoking history, total cholesterol level, and body mass index (BMI) were obtained via standard diagnostic procedures or pathological examinations. The detailed clinicopathological characteristics are provided in Supplementary Table S1. After harvesting, tissues are immediately submerged in RNAlater RNA stabilization reagent (Qiagen), and were frozen according to the manufacturer’s instructions. Total RNA was extracted from tissues using the Qiagen RNeasy Mini Kit. Concentration of the total RNA was determined using the NanoDrop (Thermo Scientific). The quality of RNA extracted from each specimen was evaluated by the RNA integrity number (RIN) using the Agilent 2100 Bioanalyzer (Agilent Technologies). All RNA preparations had RIN values over 7.

**Cell lines and culture conditions.** Human colorectal cancer cell lines, HCT116 and LoVo, were obtained from the American Type Culture Collection. HCT116 cells were cultured in McCoy’s 5a medium (WelGene), and LoVo cells were cultured in RPMI-1640 (WelGene) supplemented with 10% fetal bovine serum. Total RNA preparations from HCT116 and LoVo cells were extracted using the TRI Reagent (Ambion).

**Real-time RT-PCR analysis.** The quantitative analysis of target mRNA levels was performed by real-time PCR with SYBR Premix Ex Taq II (Takara) using an CFX96 Real-time PCR detection system (Bio-Rad). Single-stranded cDNA was synthesized from 2 μg of total RNA using ImProm-II reverse transcriptase (Promega). For each PCR, amount of cDNA equivalent to 10 ng of total RNA was used. Oligonucleotide primers were designed using Primer3 in conjunction with ProbeFinder version 2.4 software (Roche Applied Science). Primer sequences are available online (Supplementary Table S2). ACTB and HPRT1, 2 internal control genes were used as dual reference genes. Cycling conditions were as follows: 2 minutes at 95°C followed by 40 cycles of 15 seconds at 94°C, 30 seconds at 60°C, and 30 seconds at 72°C. The expression value of candidate genes was calculated for each cDNA sample using the delta Ct method with normalizing to the respective mean value of 2 reference genes with Bio-Rad CFX Manager Software. Statistical significance was evaluated using paired t-test between non-tumor and tumor tissue samples from each patient for each gene.

**Conventional RT-PCR analysis**

For the expression analysis of CKS2 and ETV4 (Fig. 2B), cDNA was amplified by using Platinum Taq DNA polymerase (Invitrogen) and gene-specific primers. The amount of cDNA used for each reaction was equivalent to 20 ng of total RNA. Cycling conditions were as follows: 2 minutes at 95°C followed by 28 or 29 cycles of 15 seconds at 94°C, 15 seconds at 60°C, and 30 seconds at 72°C with a final extension of 10 min at 72°C. ACTB expression was used as an internal control, and PCR products were visualized by agarose gel electrophoresis.

**Statistical analysis**

Paired t-test was used to evaluate the significance of differential mRNA expression levels of candidate genes. The ratios of real-time PCR expression values of case matched normal and cancer tissues were used after normalization by the mean expression value of internal control genes, ACTB and HPRT1. Of the 34 genes put to test, 9 genes with P value less than 0.01 were proposed as the CRC marker genes. The association between gene expression levels and clinical parameters was addressed by analysis of variance (ANOVA). We applied one-way ANOVA to test the null hypothesis that ratios of given gene’s expression levels are the same for all categories of each clinical parameter.

**Transfection of siRNAs**

Specific siRNAs for c-Myc and K-ras were selected either from Invitrogen Stealth Select RNAi or designed using BLOCK-IT™ RNAi Designer (Invitrogen). Target sequences are available on line (Supplementary Table S3). Stealth RNAi siRNA negative control med GC (Invitrogen) was used as a control for nonsequence specific effects. All siRNA duplexes (20 nM) were transfected to 3 × 10^5 cells in 60 mm dishes using Lipofectamine RNAiMAX (Invitrogen) for 48 hours at 37°C in a CO2 incubator according to the manufacturer’s instructions. Gene knockdown effects for c-Myc and K-ras were confirmed by real-time RT-PCR and Western blotting analysis.

**Western blotting analysis**

Cell lysates were prepared by resuspending cell pellets in lysis buffer (50 mM Tris–Cl, pH 7.5, 150 mM NaCl, 2 mM Na2EDTA, 1% NP-40, 0.1% sodium dodecyl sulfate (SDS), 0.5% Na-deoxycholate, and 10 mM NaF) supplemented with a mixture of protease inhibitors (Sigma) and a mixture of phosphatase inhibitors (Sigma). After incubating on ice for 20 minutes, supernatants were isolated by centrifugation at 14,000 rpm for 20 minutes. Protein
concentrations were determined by using the BCA protein assay kit (Thermo Scientific Pierce). SDS-PAGE gel electrophoresis and transfer to polyvinylidene difluoride (PVDF) membrane (Millipore) were carried out using standard techniques. The blots were probed with antibodies against c-Myc (sc-40; Santa Cruz Biotechnology) or K-ras (sc-30; Santa Cruz Biotechnology), and α-tubulin antibody was used to control for loading (AbFrontier). The membranes were incubated with peroxidase-conjugated anti-mouse-IgG or anti-rabbit-IgG antibodies (Thermo Scientific Pierce). Proteins were detected using an enhanced chemiluminescence detection kit (Amersham-Pharmacia Biotech).

Results

Bioinformatics analysis

We started by proposing the following set of conditions which valid biomarker candidate genes for colon cancer must satisfy: (i) they should be differentially expressed between normal and cancer tissues, preferably overexpressed in cancer; (ii) differential expression should be consistent regardless of the data source (e.g., ESTs or microarrays); and (iii) tissue specificity should be above a predetermined cutoff value.

In an effort to identify candidate genes using public gene expression data, we have established a bioinformatics pipeline and have successfully applied the procedure for identification of biomarkers for lung cancers (9). This study also utilizes a similar stepwise filtering process based on an EST cluster analysis followed by an expression level analysis (see the "Materials and Methods" section). The list of 91 candidate biomarker genes for colon cancer resulting from this analysis is provided in the web site http://genome.ewha.ac.kr/CancerBiomarker/ColonCancerTableAll.html. In addition, we applied the same procedure to other cancers as well and built a cancer biomarker database, ACBD (a cancer biomarker database) which can be accessed at http://genome.ewha.ac.kr/CancerBiomarker. The website provides a brief report on biomarkers including gene summary, graphical display of microarray data, summary of statistical tests on differential expression based on microarray and EST data, and literature information obtained from the iHOP text-mining service available at http://www.ihop-net.org.

For detailed validation experiments, we reduced the number of candidates further. Specifically, selecting genes whose cancer specificity was over 60% (see the "Materials and Methods" section) or whose rank of differential expression in microarray data was within upper 20% yielded 34 final candidate genes. The list of these 34 candidate genes is provided (Supplementary Table S4).

Experimental validation of biomarkers

The expression of 34 candidate genes was examined by quantitative real-time RT-PCR. Primary colorectal cancer samples and normal tissues in the surrounding area were
isolated from 29 patients, and 10 randomly selected sample pairs were tested initially leading to 11 out of 34 genes as preliminary marker gene candidates. After expanding the analysis to all 29 patient samples, differential increase in the expression levels was shown to be significant by paired t-test (at P-value < 0.01) for 9 of the 34 genes: ECT2, ETV4, DDX21, RAN, S100A11, RPS4X, HSPD1, CKS2, and C9orf140 (Figs. 1 and 2A; Table 1). We also confirmed the real-time RT-PCR results with conventional RT-PCR followed by gel electrophoresis on selected genes and patient samples (Fig. 2B).

### Correlation with major clinical parameters of colorectal cancer

Some of the genes such as HSPD1 and RAN showed noticeable variations in the differential levels in that several patients showed clear difference between normal and tumor tissues while virtually no change was seen in others. We have thus attempted to find correlation between the differential levels and major clinical parameters (gender, age, histological tumor subtypes, tumor stage, lymphatic invasion status, distant metastasis status, colon polyp presence, family history of CRC, smoking history, total cholesterol level, and BMI (Supplementary Table S1). To examine the correlation between gene expression levels and major clinical parameters, ANOVA statistical test was used. Interestingly, several statistically significant correlations were observed (Fig. 3). Specifically, cases with highest differential expression levels of RPS4X were associated with well- or moderately adenocarcinomas but not with mucinous cell types while such cases of ECT2 were associated with the absence of lymphatic invasion. In addition, high-level expression of CKS2 was associated with early tumor stage and the lack of family history. One of the genes, C9orf140 interestingly was associated with low level of total cholesterol, which indicates that the relevance of this gene with CRC must be re-examined. Collectively, our results are suggestive of potential diagnostic value of some of the discovered genes (see the “Discussion” section).

### Regulation of DDX21 and RPS4X by c-Myc

Among the identified marker genes, DDX21 has been suggested to be a direct transcriptional regulatory target of c-Myc in a study using a prostate cancer cell line (15). Importantly, c-Myc has been shown to be a target of APC signaling which is aberrantly activated in most colorectal cancers (16). We thus questioned if DDX21 and other markers are targeted by c-Myc in colorectal cancers. Small interfering RNAs (siRNAs) specifically targeting c-Myc transcript were transfected into 2 well-established human colorectal cancer cell lines, HCT116 and LoVo. Western blot analysis showed that siRNAs worked satisfactorily as c-Myc protein and mRNA levels showed significant reduction (Fig. 4A, B). We examined the expression of all 9 genes by quantitative real time PCR and found 2 genes, DDX21 and RPS4X, were also downregulated significantly (Fig. 4C, D). We have thus connected 2 of the genes to a major oncogene implicated in colorectal carcinogenesis, which further suggests their potential value as biomarkers.

### Regulation of ETV4 by K-ras

K-ras is known to be mutated in up to 50% of the colorectal cancers (17). Specifically, single missense mutations frequently involving Gly12 or Gly13 generates constitutively active form of K-ras which accounts for its oncogenic activity. HCT116 and LoVo both contain G13D mutations offering the chance to test if any of the biomarkers we discovered is connected with K-ras signaling (17). K-ras-targeting siRNAs were prepared, and the efficacy was confirmed by immunoblot and RT-PCR assays (Fig. 5A and B). Out of the 9 biomarkers, ETV4 was shown to be affected by the downregulation of K-ras which suggests that ETV4 may be a key transcriptional regulator of colorectal carcinogenesis (Fig. 5C).

### Discussion

Identification of biomarkers for specific cancer types can be carried out in multiple fashions. The key feature of our strategy was to take advantage of the large set of existing data. The limited use of public data from cDNA collections

---

**Table 1. List of genes obtained after validation with clinical specimens.** The P-value was obtained by paired t-test with 29 case-matched patient samples.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene full name</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECT2</td>
<td>Epithelial cell transforming sequence 2 oncogene</td>
<td>2.28E-08</td>
</tr>
<tr>
<td>ETV4</td>
<td>Ets variant gene 4 (E1A enhancer binding protein, E1AF)</td>
<td>3.67E-08</td>
</tr>
<tr>
<td>DDX21</td>
<td>DEAD (Asp-Glu-Ala-Asp) box polypeptide 21</td>
<td>1.90E-07</td>
</tr>
<tr>
<td>RAN</td>
<td>Member RAS oncogene family</td>
<td>1.29E-05</td>
</tr>
<tr>
<td>S100A11</td>
<td>S100 calcium binding protein A11 (calgizzarin)</td>
<td>1.60E-05</td>
</tr>
<tr>
<td>RPS4X</td>
<td>Ribosomal protein S4, X-linked</td>
<td>2.06E-05</td>
</tr>
<tr>
<td>HSPD1</td>
<td>Heat shock 60kDa protein 1 (chaperonin)</td>
<td>2.91E-05</td>
</tr>
<tr>
<td>CKS2</td>
<td>CDC28 protein kinase regulatory subunit 2</td>
<td>8.27E-05</td>
</tr>
<tr>
<td>C9orf140</td>
<td>Chromosome 9 open reading frame 140</td>
<td>4.69E-04</td>
</tr>
</tbody>
</table>

---
and microarray assays thus far is in large part due to the difficulty of integrating diverse designs of the studies as well as the diverse formats of the data themselves. In this regard, establishing a working bioinformatics pipeline is the key step in instilling utility into the public data. We have previously reported a successful case study which combined meta-analysis of the public database and validation with clinical samples for identification of potential biomarkers for lung cancers (9). One of the strengths of our study is that this represents a successful utilization of underused public data and a validation of the broad utility of our bioinformatics pipeline. Specifically, we confirmed through this study that the method originally developed for lung cancer can be applied to other types of cancer.

An inherent requirement for identifying cancer biomarkers through bioinformatics analysis is the confirmation by "wet-lab" experiments using patient samples. This in turn necessitated installation of multiple filtering steps to limit the candidate genes to most likely ones. Our strategy includes applying arbitrary cut-off values to come up with a practically manageable number of genes. While this implies that some marker genes may have escaped our screening, it should be noted that cut-off parameters can be easily adjusted for more exhaustive searches using the same acquired data set. We have used case-matched clinical samples which should minimize the effect of interindividual variances and identified 9 genes with significantly elevated levels of expression as potential biomarkers of CRC.

Most of these genes have been either previously implicated in cancer development or shown to be expressed at high levels in certain types of cancers or in cancer cell lines. ETV4, an Ets family transcription factor, and HSPD1, a mitochondrial chaperonin, have been reported to be expressed in colorectal cancers at elevated levels although detailed functional assignments have yet to be made (18, 19). CKS2 is a subunit of cyclin dependent kinases and has been reported to be associated with liver metastasis of colon cancer (20, 21). The expression of RAN, a small GTPase, has been shown to be increased in most types of cancer (22, 23) while S100A11, a member of the S100 family shows either elevated or repressed expression levels depending on cancer types (24). ECT2, a Rho guanine nucleotide exchange factor, is a protooncogene cloned based on its transforming ability in fibroblast and is involved in cytokinesis (25, 26). It is reported to be expressed at elevated levels in various types of cancer including those of esophagus and lung as well as in glioma.

Figure 3. Correlation between differential gene expression levels and clinical characteristics. Scatter plots represent the correlation between the gene expression ratio (T/N: tumor versus nontumor tissues) of 4 genes, RPS4X, CKS2, ECT2, and C9orf140, with specified clinical parameters. In histological subtype, Low means low-grade carcinoma and High means high-grade carcinoma (Low includes moderately differentiated and well differentiated cell types, and High includes mucinous of cell type). In tumor stage, stage II encompasses stages IIA and IIB, and the same applies to stages III and IV. Stage IV patients included in this data set presented with isolated metastatic tumors that could be surgically removed and thus have undergone radical resection of primary tumors despite the metastasis. The numbers within parenthesis indicate the number of patients. The significance of the results was evaluated by the ANOVA test (P value <0.05).
High-level expression of DDX21, a DEAD-box family member, is associated with low survival and rapid relapse of breast cancer (29), and C9orf140, a novel protein with unknown function, show specific expression in primary gastric cancer tissues (30).

Apparently, given the varying status of the genes as biomarkers of diverse cancers including colorectal cancer, mere description of elevated expression levels has rather limited value. We thus examined if any of the markers can be correlated with major clinical parameters and found that 3 of the genes may have certain diagnostic values. Specifically, CKS2, RPS4X, and ECT2 show statistically significant association with certain clinical aspects of CRC. Another gene, C9orf140 seems to be associated with low total cholesterol level, and the high-level of expression of this gene in CRC patients must thus be carefully re-examined. Admittedly, the expression levels of these genes overlap between different categories within given clinical parameters, and a relatively small number of clinical samples clearly limits the interpretation of our study. Thus, the
In further exploring functional implications of these biomarkers, we sought to see if they are mechanistically connected with 2 of the major oncogenes, c-Myc and K-ras. Two genes, RPS4X and DDX21 appear to be regulatory targets of c-Myc. For DDX21, a direct targeting by c-Myc has been proposed in the case of prostate cancer (15). Expansion of this regulatory relationship to colorectal cancer suggests that this relationship may be broadly applicable to diverse types of cancer and that DDX21 may be an important effector of c-Myc function. To the best of our knowledge, RPS4X has not been implicated in cancer development thus far, and its connection to c-Myc warrants further mechanistic studies down the road. ETV4 is a transcriptional regulator suggested to be involved in tumorigenesis and invasion of several cancer types including colorectal cancer (31, 32). Given that a large fraction of colorectal cancers contains constitutively active mutations in K-ras gene, identifying ETV4 as a part of the K-ras pathway may represent an opportunity to elaborate one of the key regulatory axis involved in colorectal carcinogenesis. Such elaboration should include characterizing factors that relay K-ras signal to transcriptional activation of ETV4 and identifying the regulatory target genes of ETV4 itself in colorectal cancers. Additionally, ETV4 may be useful for predicting the efficacy of anti-EGFR therapy, which is currently prescribed for patients with advanced CRC. Interestingly, several recent studies showed that activated K-ras is often associated with limited response to anti-EGFR therapy consistent with that K-ras, a downstream effector of EGFR, would be epistatic to EGFR (33). Although activating mutations in K-ras can be routinely screened for, the known K-ras mutations and EGFR signaling are not likely to be the only ways to activate K-ras signaling pathway. For example, activating mutations in downstream genes to K-ras could be epistatic to both K-ras and EGFR thereby limiting anti-EGFR therapy. In fact, one recent study reported that patients with wild-type K-ras but with mutant BRAF are nonresponsive to EGFR Inhibitor therapy (34). Thus, in addition to screening
for known mutations of K-ras (and possibly BRAF down the road), it would be valuable to be able to test if the K-ras signaling pathway is in activated state. If ETV4 expression can indeed be used as a downstream marker for the activated K-ras signaling pathway, another method of predicting the response to EGFR therapy can be developed. Clearly, a larger clinical study as well as molecular mechanistic analyses should be carried out to further dissect the relationship between the expression of ETV4, activation of K-ras signaling pathway, and efficacy of anti-EGFR therapy before the significance of ETV4 expression is fully evaluated.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References


Grant Support

Supported by the “Systems biology infrastructure establishment grant” provided by Gwangju Institute of Science & Technology in 2008 through the Ewha Research Center for Systems Biology (ERCBS), by grant No. 2008–0061331 from the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST), by a grant (2010K0000803) from Brain Research Center of the 21st Century Frontier Research Program funded by the Korea government (MEST), and by grant No. R15–2006–020 from the National Core Research Center (NCRC) program of the Ministry of Education, Science & Technology (MEST) and the National Research Foundation of Korea (NRF) through the Center for Cell Signaling & Drug Discovery Research at Ewha Womans University.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received May 20, 2010; revised September 28, 2010; accepted November 29, 2010; published OnlineFirst February 8, 2011.


Clinical Cancer Research

Clinical Validation of Colorectal Cancer Biomarkers Identified from Bioinformatics Analysis of Public Expression Data

Yeonjoo Jung, Sanghyuk Lee, Hyung-Seok Choi, et al.

*Clin Cancer Res* Published OnlineFirst February 8, 2011.

<table>
<thead>
<tr>
<th>Updated version</th>
<th>Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-10-1300</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplementary Material</td>
<td>Access the most recent supplemental material at: <a href="http://clincancerres.aacrjournals.org/content/suppl/2011/02/17/1078-0432.CCR-10-1300.DC1">http://clincancerres.aacrjournals.org/content/suppl/2011/02/17/1078-0432.CCR-10-1300.DC1</a></td>
</tr>
</tbody>
</table>

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