Serum Proteomic Profiles Suggest Celecoxib-Modulated Targets and Response Predictors

Zhen Xiao,1 Brian T. Luke,2 Grant Izmirlian,3 Asad Umar,3 Patrick M. Lynch,4 Robin K. S. Phillips,5 Sherri Patterson,4 Thomas P. Conrads,1 Timothy D. Veenstra,1 Peter Greenwald,3 Ernest T. Hawk,3 and Iqbal U. Ali3

1Laboratory of Proteomics and Analytical Technologies and 3Advanced Biomedical Computing Center, SAIC-Frederick Inc., National Cancer Institute at Frederick, Frederick, Maryland; 2Division of Cancer Prevention, National Cancer Institute, Bethesda, Maryland; 4University of Texas M. D. Anderson Cancer Center, Houston, Texas; and 5Imperial Cancer Research Fund, St. Mark’s Hospital, London, United Kingdom

ABSTRACT

Cyclooxygenase-2 is a valid target for cancer prevention and treatment. This has been shown in preclinical and clinical cancer prevention studies by using a cyclooxygenase-2 inhibitor, celecoxib. When used in a randomized cancer prevention clinical trial on patients with the inherited autosomal dominant condition, familial adenomatous polyposis, celecoxib proved efficacious. However, a remarkable heterogeneity in patients’ responses to the chemopreventive effects of celecoxib was observed. Proteomic profiling of sera from these patients identified several markers, the expression of which was specifically modulated after treatment with celecoxib. A decision tree algorithm identified classifiers for response to celecoxib with relatively high sensitivity but moderate to low specificity. In particular, a spectral feature at m/z 16,961.4 was identified as a strong discriminator between response and nonresponse to celecoxib at the highest dose.

INTRODUCTION

The inherited syndrome, familial adenomatous polyposis (FAP) develops as a consequence of a germ-line mutation in the adenomatous polyposis coli tumor suppressor gene, which is also mutated frequently in sporadic colorectal cancer (1). Individuals with FAP have a marked predisposition to colorectal carcinoma and develop numerous adenomatous polyps, considered to be the premalignant precursors to carcinoma, relatively early in life with surgical excision being their only effective option. Among the molecular abnormalities identified in the adenomatous polyps are elevated expression of the inducible form of prostaglandin G/H synthase [also known as cycloxygenase 2 (COX-2)] (2, 3) a crucial enzyme in the arachidonic acid metabolism generating multifunctional prostanoids. COX-2 is implicated in the process of carcinogenesis in general and colorectal carcinogenesis in particular (4) and hence can be targeted in cancer prevention/intervention strategies (5). Similarly, nonsteroidal anti-inflammatory drugs that inhibit both isomers of cyclooxyngeneses, COX-1 and COX-2, nonselectively have also been shown to be effective chemopreventive agents in animal as well as human intervention studies (6–8). Recently, three randomized clinical trials have shown a modest effect of a nonselective nonsteroidal anti-inflammatory drug, aspirin, in reducing the risk of adenomas (9–11). The chemopreventive efficacy of celecoxib, a selective COX-2 inhibitor, was demonstrated in mouse models of FAP and in human clinical trials (12–14).

The cancer prevention trial using celecoxib in FAP patients in a double-blinded, placebo-controlled study (14) demonstrated a significant reduction both in the number and burden of polyps. Even so, the individual response of patients in this trial was remarkably heterogeneous. An extensive profiling of the molecular circuitry affected by celecoxib would be helpful in understanding the heterogeneity of patients’ responses in this trial. Currently, high throughput proteomic technologies are being used to identify diagnostic markers for early cancer detection and plausible therapeutic targets responsive to specific treatments. In particular, surface-enhanced laser desorption/ionization time of flight mass spectrometry (SELDI-TOF MS) is being explored extensively for its use in diagnostic medicine. This technique circumvents sample purification/fractionation procedures in a direct one-step affinity capture of a subset of proteins from highly complex biological fluids, such as serum and plasma (15, 16), which are believed to reflect the physiological as well as pathological molecular events. Recently, SELDI-TOF MS, together with bioinformatics tools, has been successfully used for differential serum proteomic profiling of various cancers. In each case, a panel of markers was identified with the ability to discriminate between individuals with no cancer, cancer, and among various stages of cancer development (17–22). These proteomic profiling approaches can be instrumental in identifying molecular signatures of cancers; furthermore, these can also be used in monitoring molecular alterations that ensue after any intervention with the potential of impacting drug development.

We used SELDI-TOF MS to obtain proteomic profiles of sera from patients of the FAP/celecoxib clinical trial with the assumption that celecoxib-modulated changes may be reflected in the alteration of serum protein components. The objectives of this study were to delineate serum proteomic changes in the FAP patients after receiving celecoxib and to identify serum markers to predict response to treatment with celecoxib.

MATERIALS AND METHODS

Patients and Serum Samples. The baseline characteristics and treatment outcome of the FAP patients from the celecoxib prevention trial have been described previously (14). Serum samples from 31 patients were collected at M. D. Anderson Cancer Center at the time of both pre- and post-treatment. At St. Mark’s Hospital, only pretreatment serum samples were collected from 37 patients. All serum samples were stored at −80°C before analysis. To monitor the sample preparation and mass spectral acquisition in profiling experiments, a normal serum sample was obtained from Sigma (St. Louis, MO) and used as the quality control sample on each chip.

SELDI Proteomic Profiling. Weak cationic exchange (WCX) and strong anionic exchange (SAX) protein chips (Ciphergen Biosystems, Inc., Fremont, CA) were used to obtain protein profiles from serum samples analyzed in duplicate. SAX chips were first pre-equilibrated twice with 200 μL of binding buffer [20 mM Tris (pH 8.0), 0.1% Triton X-100] and agitated for 5 min each time. Equilibration of WCX chips consisted of treating with 100 μL of 10 mM HCl for 5 min on a shaker, rinsing twice with deionized H2O for 5 min, and coating with 150 μL of binding buffer [50 mM sodium acetate (pH 4.5), 0.1% Triton X-100] for 10 min. Serum samples (5 μL of each) were thawed on ice, diluted to 100 μL in respective binding buffers, applied on the chip arrays, and incubated on a shaker for 1 h. The chips were then washed with 200 μL of the
PROTEOMIC PROFILING IN FAP-CELECOXIB PREVENTION TRIAL

Detection of Celecoxib-Modulated Protein Changes. The pre- and post-treatment serum samples of 31 patients (6 in the placebo arm; 13 in the 100-mg bid; and 12 in the 400-mg bid treatment arms) were used to search for proteins, the expression levels of which significantly changed as a result of celecoxib treatment. WCX and SAX protein chips were used to selectively capture subgroups of protein chips before spectral acquisition using SELDI-TOF MS. Specific m/z ranges were used for the mass spectra acquired using the WCX and SAX chip surfaces to maximize the number of peak clusters with useful signals. The WCX chip produced 68, 82, and 82 peaks in the m/z 2,000–20,000 range, and the SAX chip produced 108, 130, and 131 peaks in the m/z 2,000–50,000 range for the placebo, 100-mg bid and 400-mg bid arms, respectively.

Because the serum from each patient was analyzed in duplicate, a linear mixed effects model (23) was used to remove the random variation in each patient’s duplicate spectra from the total variation between pre- and post-treatment spectra thereby obtaining a better estimate of the effect of the treatment. The peak clusters with P < 0.05 were identified from both chip surfaces. The WCX spectra contained seven significant peaks from the placebo arm and 10 and 26

Table 1 Pre- and post-treatment comparison on weak cationic exchange surface

<table>
<thead>
<tr>
<th>Treatment arm [No. of patients/no. of peaks (2–50 kDa)]</th>
<th>100-mg bid (13/82)</th>
<th>400-mg bid (12/82)</th>
</tr>
</thead>
<tbody>
<tr>
<td>m/z (Da)</td>
<td>+/-</td>
<td>P</td>
</tr>
<tr>
<td>----------------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>Peaks (m/z) with discriminating power (P &lt; 0.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2,023.2</td>
<td>++</td>
<td>0.023</td>
</tr>
<tr>
<td>3,158.7</td>
<td>++</td>
<td>0.0218</td>
</tr>
<tr>
<td>3,816.2</td>
<td>+</td>
<td>0.046</td>
</tr>
<tr>
<td>4,300.3</td>
<td>+</td>
<td>0.0492</td>
</tr>
<tr>
<td>11,080.6</td>
<td>-</td>
<td>0.021</td>
</tr>
<tr>
<td>15,526.2</td>
<td>-</td>
<td>0.0345</td>
</tr>
</tbody>
</table>

* +, increased expression pre- vs. post-treatment.
** - , decreased expression pre- vs. post-treatment.

Table 2 Pre- and post-treatment comparison on strong anionic exchange surface

<table>
<thead>
<tr>
<th>Treatment arm [No. of patients/no. of peaks (2–50 kDa)]</th>
<th>100-mg bid (13/130)</th>
<th>400-mg bid (12/131)</th>
</tr>
</thead>
<tbody>
<tr>
<td>m/z (Da)</td>
<td>+/-</td>
<td>P</td>
</tr>
<tr>
<td>----------------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>Peaks (m/z) with discriminating power (P &lt; 0.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5,093.6</td>
<td>-</td>
<td>0.0328</td>
</tr>
<tr>
<td>10,346.9</td>
<td>+</td>
<td>0.0274</td>
</tr>
</tbody>
</table>

* - , decreased expression pre- vs. post-treatment.
** +, increased expression pre- vs. post-treatment.
peaks from the 100-mg bid and 400-mg bid arms, respectively. The spectra from the SAX chip produced three, nine, and seven significant peaks for the respective treatments. The differentially expressed markers in the placebo arm probably represented time-related changes between pre- and post-treatment (data not shown).

Six differentially expressed peak clusters from the WCX chip and two from the SAX chip were common to both 100- and 400-mg arms and absent from the placebo arm. The abundance level of each of the eight clusters followed the same direction in pre-versus post-treatment change with both doses of celecoxib and on both chip surfaces (Tables 1 and 2). Fig. 1, A and B, show examples of two patients each on the WCX and SAX chips, respectively, with the gel views and trace views of increased expression of the marker at \( m/z \) 3,158.7/3,158.8 and reduced expression of the marker at \( m/z \) 5,089.5/5,093.6 before and after treatment with celecoxib.

Identification of Serum Markers as Molecular Predictors of Response to Celecoxib. Fig. 2 displays the heterogeneity of patients' response as determined by percent change from baseline in the number of colorectal polyps in all three treatment arms. Patients were divided into the following three categories based on their response to celecoxib: (a) strong responders whose polyp counts decreased by at least 20% (\( n = 32 \)); (b) weak responders whose polyp counts decreased by <20% (\( n = 13 \)); and (c) nonresponders whose number of polyps stayed constant or increased (\( n = 23 \)). To obtain as distinct a discriminatory pattern as possible for response prediction, the 13 weak responders were excluded from the analysis. The total of 55 patients were then placed into four groups with increasing confidence in the classification (see Table 3). Group 1 consisted of all 55 patients from the three treatment arms, placebo, 100-mg bid, and 400-mg bid. Group 2 contained patients from both treatment arms but excluded the subject from the placebo arm, because it is questionable whether the subjects in the placebo arm can be considered as either responders or nonresponders. Group 3 excluded the nonresponders treated with 100 mg, because it is possible that some of these nonresponders would respond to a 400-mg bid celecoxib treatment, and therefore their categorization as nonresponders would degrade the classification. The 18 strong responders and five nonresponders from the 400-mg bid treatment arm in group 4 represented the set with the highest classification confidence and should best identify one or more markers that signify whether a patient will respond to celecoxib. Each of these groups was used to build a classifier based on the RF algorithm (24–26).

Baseline pretreatment serum samples were used to search for proteomic markers capable of predicting response to celecoxib. Classification algorithms were constructed for groups 1–4 using 110, 92, 70, and 46 spectra, respectively (two pretreatment spectra per patient) with each spectrum containing 72 peak clusters from the WCX chip and 138 peak clusters from the SAX chip. The spectral features from WCX and SAX chips were used in combination because a better classification was obtained by combining the serum spectra acquired from both chip surfaces as opposed to using them separately (data not shown). For each set of spectra, all decision trees that did not use a subject during construction were polled and the "raw votes" for responders and nonresponder classifications were counted. These
votes for each class were then weighted to minimize the number of misclassified subjects.

The classification yielded 94 and 97% sensitivity but only a 52 and 50% specificity for group 1 and group 2, respectively. The low specificity could be attributable to the potentially erroneous classification of subjects in the placebo arm and nonresponders in the 100-mg treatment arm. In group 3, the algorithm classified all 30 responders correctly, but three of the nonresponders were misclassified. This reduced the specificity to 40%, but the sensitivity and negative-predictive value increased to 100%, and the positive-predictive value increased to 91%. The low specificity of 40% may be attributable to the highly imbalanced numbers of responders (n = 30) versus nonresponders (n = 5) producing a 6:1 ratio of responders to nonresponders in this group. This means that the randomly chosen samples used to build each decision tree were very heavily weighted toward responders. Any nonresponders in the out-of-bag samples were then tested in their corresponding responder-biased decision trees, and this is the most likely reason why two of the nonresponders were misclassified in the group 3 run. When the responders and nonresponders of the 400-mg arm alone were examined (group 4), one subject from each set was misclassified. This produced a sensitivity of 94%, a specificity of 80%, and positive- and negative-predictive values of 94 and 80%, respectively.

The experimental results are presented in Table 3.

Table 3. Sensitivity, specificity, positive and negative predictive values (PPV and NPV, respectively), and important features obtained from random forest models.

<table>
<thead>
<tr>
<th>Response classes</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responders</td>
<td>32</td>
<td>30</td>
<td>30</td>
<td>18</td>
</tr>
<tr>
<td>Nonresponders</td>
<td>23</td>
<td>16</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Sensitivitya</td>
<td>94% (30/32)</td>
<td>97% (29/30)</td>
<td>100% (30/30)</td>
<td>94% (17/18)</td>
</tr>
<tr>
<td>Specificitya</td>
<td>52% (12/23)</td>
<td>50% (8/16)</td>
<td>40% (2/5)</td>
<td>80% (4/5)</td>
</tr>
<tr>
<td>PPV</td>
<td>73%</td>
<td>78%</td>
<td>91%</td>
<td>94%</td>
</tr>
<tr>
<td>NPV</td>
<td>86%</td>
<td>89%</td>
<td>100%</td>
<td>80%</td>
</tr>
<tr>
<td>Important featuresb</td>
<td>7,567.48 (1.88)</td>
<td>7,928.24 (1.46)</td>
<td>16,961.4 (2.36)</td>
<td>16,961.4 (5.58)</td>
</tr>
</tbody>
</table>

a The ratio in parentheses is the number of correct classifications over the number of subjects in this class.

b The number in parentheses is the random forest importance score for the feature.

The peaks with sufficiently high importance scores are also listed in Table 3. It is noticeable that the classifiers for groups 1 (m/z 7,567.48 and 7,928.24) and 2 (m/z 7,928.24 and 3,975.82) and groups 3 (m/z 16,961.4) and 4 (m/z 16,961.4 and 4,647.23) are different with one shared peak each between groups 1 and 2 (m/z 7,928.24) and groups 3 and 4 (m/z 16,961.4). The fact that the important spectral features were different for groups 1 and 2 than groups 3 and 4 supports the contention that the algorithm was simply trying to fit less precise data. The degradation in the assigned class of the samples resulted in new features being chosen as the most important ones.

The peak cluster m/z 16,961.4 with a very high importance score is...
an important response discriminator for both groups 3 and 4. Fig. 3A displays the region of the SAX spectra between $m/z$ 16,500 and 17,500 from two strong responders and nonresponders. The spectra exhibit the presence of $m/z$ 16,961.4 in nonresponders and its absence in the strong responders. Although there is a pattern of several differentially expressed markers of much higher intensities in this region, Fig. 3B demonstrates that the peak cluster at $m/z$ 16,961.4 is by far the best discriminator of celecoxib response in group 4 patients treated with a 400-mg bid dose. A separate classification algorithm using a distance-dependent K-nearest neighbor algorithm also demonstrated that the $m/z$ 16,961.4 SAX peak was an important discriminator between responders and nonresponders (data not shown).

**DISCUSSION**

A randomized controlled phase II trial of celecoxib proved the potential of celecoxib for chemoprevention in FAP patients (28% reduction in polyp numbers versus 5% in the placebo arm; $P = 0.003$). On the basis of the results of this landmark trial (14), the Food and Drug Administration approved celecoxib as an oral adjunct to the standard of care for adults with FAP. Celecoxib selectively inhibits the inducible COX-2, which is implicated in inflammation and carcinogenesis and has emerged recently as a molecular target for the prevention and treatment of a variety of human cancers (5). Several lines of evidence indicate that overexpression of COX-2 not only induces cellular proliferation but also impact other vital cellular processes such as apoptosis and angiogenesis (27–30). Depending on the molecular context, many of these processes may be affected by COX-2 at any given time, thereby perturbing cellular signaling networks. Information on such multiple deranged signaling pathways can be obtained by the recent state-of-the-art proteomic expression profiling technologies. In this context, serum proteomic profiling has emerged as a noninvasive method of identifying diagnostic molecular patterns for various pathological conditions including various cancers (17–22).

In this study, we, for the first time, describe serum proteomic profiling to analyze drug response in a clinical trial of cancer prevention. We examined serum proteomic changes in FAP patients in response to celecoxib and attempted to identify predictive molecular patterns for the response or nonresponse of patients. Comparative profiling of sera from pre- and post-treatment samples identified a panel of markers with specific $m/z$ ratios, the expression of which was significantly modulated by celecoxib. More importantly, these proteomic markers were over- or underexpressed in both treatment arms with low and high doses of celecoxib suggesting that at least some of them are relevant to the celecoxib-mediated regression of adenomas. Information on the identities of these celecoxib-modulated proteomic markers is expected to provide an understanding of the complex cellular networks affected by celecoxib, which may act in multiple ways to restrain carcinogenic process via COX-2-dependent as well as COX-2-independent mechanisms.

The second objective of our study was to use pretreatment sera for the identification of markers, which may serve as classifiers for response to celecoxib. The RF algorithm identified two proteomic markers, $m/z$ 16,961.4 from the SAX chip and $m/z$ 4,647.23 from the WCX chip, especially $m/z$ 16,961.4 with a very high importance score, that could distinguish between responders and nonresponders of the treatment arm receiving 400 mg of celecoxib with misclassification of only one subject each of the responders and nonresponders (Table 3). Fig. 4 shows that the features $m/z$ 16,961.4 and 4,647.23 can effectively separate the responders from the nonresponders in groups 3 and 4. The fact that $m/z$ 16,961.4 and 4,647.23 markers from the pretreatment sera are recognized by the RF algorithm as important classifiers for response to celecoxib and are not among the markers in Tables 1 and 2 suggests that they are relevant for the chemopreventive effect of celecoxib without being significantly modulated by it. Identification and characterization of the proteomic markers that are modulated by celecoxib and/or aid in response prediction are clearly warranted.

The low specificity in classification may be due largely to the few samples in various response categories, a limitation dictated by the number of patients in the clinical trial. In the future, analysis of a larger number of patients in various response categories together with combined use of markers from multiple chip surfaces are expected to result in improved performance of the classification algorithm providing highly accurate response prediction.

Our study is a significant first attempt in using proteomic tools to predict response outcome in a cancer prevention trial. It also addresses the issue of inter-patient variability in response resulting in variable benefit of the same drug to individual patients. Identification of molecular predictors would be beneficial in maximizing potential response to specific prevention/treatment agents and simultaneously minimizing undesired side effects. The efficacy of celecoxib is currently being evaluated in a much larger clinical trial of sporadic colorectal adenomas. The use of celecoxib for the prevention and treatment of a variety of other cancers is also being investigated. Serum expression profiling to identify drug-modulated proteomic markers and clinical response predictors may revolutionize treatment strategies based on our ability to predict response of individual patients to various treatments. There is certainly a need for predictive molecular factors if personalized medicine is to become a reality.

**ACKNOWLEDGMENTS**

We thank Robert Gentleman for suggesting the use of the random forest algorithm, Andy Liaw for valuable information, and Carolyn Whistler for excellent assistance in preparing the figures.

**REFERENCES**

Serum Proteomic Profiles Suggest Celecoxib-Modulated Targets and Response Predictors

Zhen Xiao, Brian T. Luke, Grant Izmirlian, et al.

Cancer Res 2004;64:2904-2909.