Differential prognostic values of mRNA expression of CEACAM gene family members in nonsmall cell lung cancer

Haruhiko Nakamura¹,² Hiroki Sakai¹ Tomoyuki Miyazawa¹ Toshiaki Somehara³ Noboru Nakayama² Kiyonaga Fujii² Toshihide Nishimura²
¹Department of Chest Surgery, ²Department of Translational Medicine Informatics, St Marianna University School of Medicine, Kawasaki, Kanagawa, Japan

Abstract: Serum carcinoembryonic antigen (CEA) is widely used as a representative marker of various malignant tumors. CEA-related cell adhesion molecules (CEACAMs), including CEACAM5, are encoded in the human genome by 12 independent genes and can be potential targets for future cancer treatments. In nonsmall cell lung cancer, serum CEA levels have been reported to predict patient survival. However, associations between mRNA expression of CEACAM gene family members in tumor tissues and patient prognosis remain unclear. To clarify this point, we used the Kaplan–Meier plotter global portal site, which collects the results of Affymetrix gene expression microarray analyses from the publicly accessible Gene Expression Omnibus database and combined it with survival data of patients. A total of 1,926 nonsmall cell lung cancer patients were identified from the Gene Expression Omnibus series, Cancer Biomedical Informatics Grid, and The Cancer Genome Atlas databases. We found statistically significant associations between mRNA expression of several CEACAMs and overall survival (OS) in patients with nonsmall cell lung cancer and lung adenocarcinoma (n=720) but not squamous cell carcinoma (n=524). In adenocarcinoma, higher expression levels of CEACAM6 and CEACAM8 were significantly associated with better OS, whereas higher expression levels of CEACAM3, CEACAM4, CEACAM19, and CEACAM21 were associated with worse OS. Conflicting results among multiple probe sets for the same gene were found for CEACAM1, CEACAM5, and CEACAM7. The findings of this study indicated that CEACAMs play important roles in tumor progression and impact OS of patients with adenocarcinoma. As the impact on OS differed based on the gene family members or the probe set used, the individual CEACAMs seem to function through complicated mechanisms. Further studies are necessary to resolve the problems encountered in our present study.

Keywords: mRNA, microarray, survival, nonsmall cell lung cancer, CEACAM, CEA

Introduction
Approximately 50 years ago, carcinoembryonic antigen (CEA) was identified as an oncofetal antigen in colorectal cancer in addition to normal human fetal organs, including the gut, liver, and pancreas.¹ Further studies revealed the presence of CEA and numerous CEA cross-reacting antigens in human sera, normal tissues, and various cancers other than colorectal cancer.²,³ A family of CEA-related cell adhesion molecules (CEACAMs), including CEACAM5, is known to be encoded in the human genome by 12 independent genes on chromosome 19q13.²,⁶,⁷ These CEACAM proteins belong to the immunoglobulin superfamly, and the molecules contain one or two variable-like domains with or without constant 2-like domains.² CEA is widely used as a representative serum tumor marker of various malignant tumors and elevated serum levels of CEA are often associated with poor prognosis.¹,³,⁸,⁹
CEA levels are reported to be frequently associated with a poor clinical outcome in cancer patients presumably through a variety of mechanisms, including the promotion of invasion, dissemination, metastasis, and immune suppression. Serum CEA concentrations are increased in both nonsmall cell lung cancer (NSCLC) and small cell lung cancer. In NSCLC, elevated serum CEA levels have been reported to be associated with histological types, advanced disease stages, and worse prognoses. More recently, vaccination therapy, antibody therapy, and small interfering RNA therapy targeting CEACAMs have been developed as new therapies for several solid tumors, including lung cancer.

Thus, CEA and related molecules are important not only for diagnoses but also for future therapeutic targets in various malignant tumors.

In spite of many reports regarding associations of serum CEA levels and prognosis of NSCLC, information about mRNA expression in tumor tissues and its relationship to patient survival is quite limited. Thus, we studied associations between mRNA expression detected by gene expression microarrays and overall survival (OS) in NSCLC patients by accessing an online public database. To the best of our knowledge, this is the first report focusing on associations between mRNA expression of the CEACAM gene family members and OS in NSCLC patients.

Materials and methods
Gene Expression Omnibus (GEO; www.ncbi.nlm.nih.gov/geo) is a public functional genomics data repository supported by the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). Based on the GEO database for NSCLC, the Kaplan–Meier plotter global portal site (http://kmplot.com/analysis/index.php?p=service&cancer=lung) provides combined data on gene expression group relative to the lower expression group, 95% confidence intervals, and log-rank p-values were automatically calculated on the same webpage.

The OS rates of two groups of patients subdivided by the median value of mRNA expression were calculated, and Kaplan–Meier survival plots were obtained from the webpage. Simultaneously, the hazard ratio of the higher expression group relative to the lower expression group, 95% confidence intervals, and log-rank p-values were automatically calculated on the same webpage.

Results
Differences in OS between groups with higher and lower expression levels of the investigated CEACAM genes are shown according to histological types: NSCLC, adenocarcinoma (AD), and squamous cell carcinoma (SQ; Table 3). Expression of most CEACAM family members was significantly associated with OS in patients with NSCLC (n=1,926). Similar results were obtained for patients with AD (n=720). In contrast, none of the examined CEACAM gene family members were associated with OS in patients with SQ (n=524), suggesting

### Table 1 Datasets of nonsmall cell lung cancer included in the analysis

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Platform</th>
<th>Sample Size</th>
<th>M</th>
<th>F</th>
<th>Stage</th>
<th>Histology</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE4573</td>
<td>GPL96</td>
<td>130</td>
<td>82</td>
<td>48</td>
<td>I/II/III/IV</td>
<td>NSCLC AD/SQ</td>
</tr>
<tr>
<td>GSE14814</td>
<td>GPL96</td>
<td>89</td>
<td>66</td>
<td>23</td>
<td>45/44/0/0</td>
<td>0/130</td>
</tr>
<tr>
<td>GSE19188</td>
<td>GPL570</td>
<td>83</td>
<td>59</td>
<td>24</td>
<td>NA</td>
<td>41/24</td>
</tr>
<tr>
<td>GSE3141</td>
<td>GPL570</td>
<td>111</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>58/53</td>
</tr>
<tr>
<td>GSE31210</td>
<td>GPL570</td>
<td>226</td>
<td>105</td>
<td>121</td>
<td>168/58/0/0</td>
<td>226/0</td>
</tr>
<tr>
<td>caArray</td>
<td>GPL96</td>
<td>468</td>
<td>240</td>
<td>228</td>
<td>NA</td>
<td>0/71</td>
</tr>
<tr>
<td>TCGA</td>
<td>GPL3921</td>
<td>74</td>
<td>49</td>
<td>25</td>
<td>NA</td>
<td>0/71</td>
</tr>
<tr>
<td>GSE29013</td>
<td>GPL570</td>
<td>55</td>
<td>38</td>
<td>17</td>
<td>24/14/17/0</td>
<td>30/25</td>
</tr>
<tr>
<td>GSE37745</td>
<td>GPL570</td>
<td>196</td>
<td>107</td>
<td>89</td>
<td>130/35/27/4</td>
<td>106/66</td>
</tr>
<tr>
<td>GSE30219</td>
<td>GPL570</td>
<td>293</td>
<td>252</td>
<td>41</td>
<td>NA</td>
<td>85/61</td>
</tr>
<tr>
<td>GSE31908</td>
<td>GPL96</td>
<td>20</td>
<td>4</td>
<td>16</td>
<td>10/5/3/0</td>
<td>20/0</td>
</tr>
<tr>
<td>GSE50081</td>
<td>GPL570</td>
<td>181</td>
<td>98</td>
<td>83</td>
<td>127/54/0/0</td>
<td>127/42</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1,926</td>
<td>1,100</td>
<td>715</td>
<td>652/320/70/4</td>
<td>720/524</td>
</tr>
</tbody>
</table>

**Abbreviations:** AD, adenocarcinoma; F, females; M, males; NA, not applicable; SQ, squamous cell carcinoma.
that the significant differences in OS among NSCLC patients were mainly due to OS differences in AD patients.

The prognostic value of mRNA expression of CEACAM genes in AD tissues according to the individual probe sets of the microarrays are summarized in Table 4. Higher expression of CEACAM6 and CEACAM8 was associated with better OS, whereas higher expression of CEACAM3, CEACAM4, CEACAM19, and CEACAM21 was associated with worse OS. Conflicting results among multiple probe sets for the same gene were found in CEACAM1, CEACAM5, and CEACAM7.
Kaplan–Meier plots of OS using the Jetset probes for individual genes are shown in Figures 1–9. Because the online database listed no Jetset probe for CEACAM3, 208052_x_at was used as a representative probe for this gene to construct the OS curves.

**Discussion**

Quantification of serum CEA in lung cancer patients is widely performed to arrive at a diagnosis, evaluate tumor responses to various therapeutic modalities, and predict risks of postsurgical recurrences. However, evidence of prognostic values in lung cancer patients remains unclear. Because of this, there are no official guidelines or recommendations for the use of CEA as a prognostic indicator of lung cancer.

According to a recent review article regarding the prognostic significance of CEA in lung cancer, 18 studies reported statistically significant evidence for the use of CEA as a
prognostic marker in NSCLC patients, while seven studies showed negative results. Among the 25 studies included in this review article, only one examined the relationship between immunohistochemical CEA expression in tumor tissues and prognoses of patients but found no association. A meta-analysis of 16 studies (4,296 NSCLC patients) reported that preoperative high serum CEA levels were associated with poor OS with a combined hazard ratio of 2.28. This meta-analysis concluded that preoperative serum CEA levels can predict OS in patients with NSCLC, although high heterogeneity between included studies and publication biases should be taken into consideration.

Little is known about the functions of CEACAMs, particularly impacts on lung cancer tumorigenesis and
and CEACAM6, lead to undifferentiated cell growth and development. However, the functions and prognostic values of CEACAM1, CEACAM5, and CEACAM6 have been investigated in several tumors. For instance, surface expression of CEACAM1-4L in A549 human lung AD cells is reported to play a critical role in differentiation, contact-inhibited cell growth, and tumor suppressive functions. Disturbances in CEACAM1-4L signaling in A549 cells by CEACAM1-4S and other CEACAMs, such as CEACAM5 and CEACAM6, lead to undifferentiated cell growth and malignant transformation. In contrast, multiple clinical studies reported that CEACAM1 overexpression was associated with worse prognosis in melanoma, gastric cancer, thyroid cancer, and NSCLC, suggesting that CEACAM1 contributes to tumor progression. Thus, there are discrepancies concerning the functions of CEACAM1 among these studies. In colorectal cancer, immunohistochemically detected CEACAM6 overexpression in tumor tissues independently predicted poor OS and shortened disease-free survival, whereas CEACAM1 and CEACAM5 were not significantly related to these outcomes. In epidermal growth factor receptor mutation-negative lung AD patients, an immunohistochemical study of tumor tissues revealed that CEACAM6 expression was associated with worse prognoses, whereas CEACAM3 expression was associated with better prognoses. These studies indicated that CEACAM6 overexpression was a worse prognostic factor for selected lung cancer patients.

In the present study, none of the CEACAM gene family members were predictive of OS in patients with SQ. The most apparent result in this study is that CEACAM expression in lung SQ is not useful to predict OS. In contrast, statistically significant differences in OS were confirmed in NSCLC and AD. Since NSCLC is mainly composed of AD and SQ, differences in OS among NSCLC patients are mostly a reflection of OS differences in AD patients.

Associations with worse OS in patients with AD were confirmed by higher mRNA expression of CEACAM3, CEACAM4, CEACAM19, and CEACAM21, while CEACAM6 and CEACAM8 were associated with better OS. Since CEACAM6 overexpression is reportedly associated with worse prognosis of various cancers, the results of the present study are unique and should be confirmed in further studies. We found no reports examining serum concentration or expression levels of the other CEACAMs and associated impacts on survival of cancer patients.

For some CEACAM gene family members, conflicting results were obtained because of the use of unique probes for each gene. This is not surprising because a given gene may be detected by multiple probe sets on an Affymetrix microarray, which can result in inconsistent or even contradictory findings. The cross-reactivity of probes to other genes and multiple transcripts produced by alternative splicing events are plausible reasons. In order to create simple one-to-one mapping between genes and probe sets, a scoring system using a specific algorithm was proposed for Jetset probes as the most reliable, and these probes are identified on the Kaplan–Meier plotter webpage. The most conflicting results
were found for the CEACAM1 probe sets. Because CEACAM1 has 13 splice variants and each may have a different function in lung cancer progression, different prognostic significance might be due to the different specificity of the probe set reactive with the different variants. Jetset probe 209498_at for CEACAM1 indicated better OS. Similarly, conflicting results were found in CEACAM5 and CEACAM7. No differences in OS based on mRNA expression were found for Jetset probe sets 201884_at for CEACAM5 and 206199_at for CEACAM7.

Since there is quite limited information concerning the relationships between overexpressed CEACAMs in tumor tissues and survival of lung cancer patients, data mining using an accessible public database is both reasonable and useful. A major limitation to this study was the limited clinical information of individual patients, thus it was difficult to perform subgroup analyses. However, the total number of included patients was sufficient to obtain reliable results, if interpreted cautiously.

In conclusion, we found statistically significant associations between mRNA expression of CEACAMs and OS of NSCLC patients. These close associations were confirmed in AD, but not in SQ, suggesting that CEACAMs play important roles in progression of AD. Since the impact of expression of individual CEACAMs on OS differed (better, worse, or neutral) based on the gene family members or used probe sets, each CEACAM seems to function through complicated mechanisms. Hence, further studies are necessary to resolve the problems encountered in the present study.

Disclosure

The authors report no conflicts of interest in this work.

References


