Serial Analysis of Gene Expression Identifies Connective Tissue Growth Factor Expression as a Prognostic Biomarker in Gallbladder Cancer

Hector Alvarez,1,6 Alejandro Corvalan,1 Juan C. Roa,3 Pedram Argani,5 Francisco Murillo,8 Jennifer Edwards,6 Robert Beaty,5 Georg Feldmann,8 Seung-Mo Hong,5 Michael Mullendore,5 Ivan Roa,3 Luis Ibañez,2 Fernando Pimentel,2 Alfonso Diaz,4 Gregory J. Riggins,6,7 and Anirban Maitra5,7,8

Abstract

Background: Gallbladder cancer (GBC) is an uncommon neoplasm in the United States, but one with high mortality rates. This malignancy remains largely understudied at the molecular level such that few targeted therapies or predictive biomarkers exist.

Experimental Design: We built the first series of serial analysis of gene expression (SAGE) libraries from GBC and nonneoplastic gallbladder mucosa, composed of 21-bp long-SAGE tags. SAGE libraries were generated from three stage-matched GBC patients (representing Hispanic/Latino, Native American, and Caucasian ethnicities, respectively) and one histologically atheriasic gallbladder. Real-time quantitative PCR was done on microdissected epithelium from five matched GBC and corresponding nonneoplastic gallbladder mucosa. Immunohistochemical analysis was done on a panel of 182 archival GBC in high-throughput tissue microarray format.

Results: SAGE tags corresponding to connective tissue growth factor (CTGF) transcripts were identified as differentially overexpressed in all pairwise comparisons of GBC (P < 0.001). Real-time quantitative PCR confirmed significant overexpression of CTGF transcripts in microdissected primary GBC (P < 0.05), but not in metastatic GBC, compared with nonneoplastic gallbladder epithelium. By immunohistochemistry, 66 of 182 (36%) GBC had high CTGF antigen labeling, which was significantly associated with better survival on univariate analysis (P = 0.0069, log-rank test).

Conclusions: An unbiased analysis of the GBC transcriptome by SAGE has identified CTGF expression as a predictive biomarker of favorable prognosis in this malignancy. The SAGE libraries from GBC and nonneoplastic gallbladder mucosa are publicly available at the Cancer Genome Anatomy Project web site and should facilitate much needed research into this lethal neoplasm.

Gallbladder carcinoma (GBC) is the second most common malignancy of the hepatobiliary tree (1). The annual incidence of GBC in the United States is less than 5,000 new cases (2). Worldwide, however, GBC continues to have high rates of incidence, as well as mortality, particularly in pockets of South America, Far Eastern and Southeast Asia, and Eastern Europe (3). GBC shows notable ethnic and gender biases. For example, in the United States, incidence rates of GBC are highest in females of Native American Indian background (14.5 per 100,000), followed by Hispanics (6.8 per 100,000) and non-Hispanic Whites (1.4 per 100,000; ref. 4). These findings have been replicated in other populations showing a high prevalence of GBC, especially in Chile, where the Native American Indian (“Mapuche Indian”) have one of the highest incidence and mortality rates in the world (5). Gallstone disease, and attendant chronic cholecystitis, remains the most important recognized risk factor for GBC, with up to 85% of resected gallbladders for GBC harboring stones (5). Most cases of GBC are diagnosed at an advanced stage, and in concert with the observed resistance to most conventional chemotherapeutic modalities available, the prognosis of GBC remains poor, with a 5-year survival of <15% (5). Thus, a better understanding of the molecular pathogenesis of GBC will enable identification of molecular targets that can form the basis for rational early detection and therapeutic strategies.

Serial analysis of gene expression (SAGE) enables unbiased and quantitative analysis of cellular transcriptomes (6). Unlike microarrays, SAGE does not require a priori knowledge of the queried transcripts, and therefore is both a platform for quantification and for gene discovery. SAGE libraries are
composed of short nucleic acid sequences or "tags" that are concatenated for the purposes of sequencing: the tags correspond to unique transcripts or expressed sequences in the transcriptome, and the frequency of tags within the library corresponds to the absolute transcript abundance. Since its first description in 1995 (7), SAGE has been used for comprehensive profiling of a multitude of neoplastic and nonneoplastic cellular states (8). Furthermore, the original SAGE protocol has undergone incremental improvements, including the development of so-called long SAGE (L-SAGE) protocols, which uses tags of greater sequence length (21 bp), permitting a higher degree of specificity in alignment to the human genome and, therefore, of transcript annotation (9).

In this article, we have generated the first panel of publicly available SAGE libraries of GBC from three distinct ethnic backgrounds (Caucasian, Hispanic/Latino, and Mapuche Indian) along with nonneoplastic gallbladder tissue. Our study identifies a plethora of dysregulated genes and signaling pathways in GBC and, in particular, identifies the transforming growth factor β (TGF-β) pathway target gene connective tissue growth factor (CTGF) as abrogated in this malignancy. We confirm the differential expression of CTGF protein in surgical specimens of GBC by immunohistochemistry and show that cancers with retained CTGF expression harbor a favorable prognosis. The online availability of SAGE libraries from GBC should facilitate research into the molecular pathogenesis of this neoplasm and, further, allow researchers to interrogate the potential genetic differences in GBC arising in diverse ethnicities.

Materials and Methods

Procurement of tissue samples for SAGE. The experimental samples used were de-linked from direct patient identifiers and met the exemption criteria for human subject research. SAGE analysis was done on three snap-frozen GBC samples and one nonneoplastic gallbladder specimen (Supplementary Table S1). The three GBC samples were obtained, respectively, from a self-identified Hispanic/Latino female undergoing open cholecystectomy at Pontificia Universidad Católica de Chile Hospital, Chile (sample designated “H”); from a self-identified Native American Indian female (Mapuche Indian, with both parents from the same ethnic background) undergoing open cholecystectomy at Temuco Regional Hospital, Chile (sample designated “M”); and from a non-Hispanic White (Caucasian) male whose snap-frozen GBC was obtained from the Cooperative Human Tissue Network in the United States (sample designated “C”). To reduce confounding variables due to the stage of lesion, the three cases selected were American joint Committee on Cancer stage Ia/Iib adenocarcinomas. Cryo-molds were prepared from all three snap-frozen specimens, and nonneoplastic cellularity of >70% was verified on corresponding H&E sections by two pathologists (A.M. and P.A.) before homogenization for mRNA extraction (see below; Supplementary Fig. S1A). A snap-frozen alithiasic gallbladder was obtained from a self-identified Hispanic/Latino woman undergoing gastrectomy for gastric cancer at Sotero del Rio Hospital, Chile (sample designated “N”). The absence of tumor cells and chronic inflammation in this control sample was corroborated by extensive gallbladder mapping and histology. The integrity of the total RNA was confirmed with the 2100 Bioanalyzer (Agilent) in all samples before library preparation (Supplementary Fig. S1B).

Preparation of L-SAGE libraries and bioinformatics analysis. Total RNA from the four snap-frozen specimens was isolated with the RNAsend kit (Promega) according to the manufacturer’s instructions. L-SAGE libraries were constructed with NlaIII as the anchoring enzyme and MmeI as the tagging enzyme, as previously described (9). At least 3,456 pZero-1 plasmid clones were sequenced for each of the four libraries as part of the Cancer Genome Anatomy Project SAGE project (10). The Cancer Genome Anatomy Project “SAGE Tag Extraction Tool” was used to extract SAGE tags and linker sequences used in library construction. Human SAGE libraries were generated at an approximate resolution of 78,000 SAGE tags per library, and as a “first-pass” filtration, sequences occurring only once with multiple annotations to the genome were removed from further analysis. Before analysis, all SAGE libraries were normalized to 200,000 tags. Statistically significant pairwise comparison was done using the web tool described by Romualdi et al. (11). We first calculated the ratio of tag counts for each transcript in the three individual GBC libraries over the representation of the same individual tags in the normal library and identified differentially expressed SAGE tags at a significance level of P < 0.001 (Audic and Claverie test). As an additional enrichment step following this individual pairwise analysis, we digitally extracted tags that were significantly different across all comparisons (P < 0.001). This subset of highly significantly differentially expressed SAGE tags in GBC versus normal gallbladder represented an enriched candidate list of transcripts for subsequent validation. In addition to pairwise comparison, we also performed two-way (by genes and samples) unsupervised hierarchical clustering to examine the relationship among the three ethnically diverse GBC specimens. SAGE data management, tag to gene matching, as well as additional gene annotations and links to online resources (UniGene, LocusLink, etc.) were done using the SAGE Genie and SAGEmap tools (10, 12). For automated functional annotation and pathway analysis of genes of interest, the web tool Pathway Explorer was used (13). The library information and tag counts are publicly available at the Cancer Genome Anatomy Project SAGE Genie web site.9

Microdissection of gallbladder samples for real-time PCR validation. An additional eight snap-frozen gallbladder specimens were collected from surgically resected specimens at Pontificia Universidad Católica de Chile Hospital with the intent of microdissection and RNA extraction for real-time PCR validation. The eight snap-frozen cases included five matched pairs of primary GBC and associated nonneoplastic gallbladder epithelium and three unmatched metastatic samples; histologic documentation of lesional tissue was confirmed on the cryosections in all instances. Cryosections (8 μm) were placed on 1-mm polyethylene naphthalate membrane–covered slides, and cells of interest were isolated by laser microdissection and laser pressure catapulting (P.A.L.M. Microlaser Technologies AG), as previously described (Supplementary Fig. S2; ref. 14). Total RNA was isolated from 1,500 to 2,500 cells with the Picopure RNA Isolation Kit (Molecular Devices) according to the manufacturer’s instructions. Reverse transcription was done with SuperScript II Reverse Transcriptase (Invitrogen) and quantitative PCR was done on cDNA templates on the 7300 Real-time PCR machine (Applied Biosystems) with SYBR Green PCR Master Mix from Applied Biosystems. Primer pairs used were, for CPD1 (housekeeping gene), sense AGGGGGAGCCAAAAGGG and antisense TGGCCAGCCCCAGCAGCAAG, and for CTGF, sense TGGCAGGCT-GATTCTCTAGT and antisense GTGGCAACATGTAACCTTTTG. All quantitative PCR experiments were done in triplicate. Analysis of relative gene expression changes was done with the 2−ΔΔCT method.

GBC tissue microarray and patient demographics. Archival formalin-fixed samples of resected GBC were retrieved from the surgical pathology archives of the two Chilean institutions, including 37 GBC samples derived from Hispanic/Latino patients operated at Pontificia Universidad Católica de Chile Hospital and 138 GBC samples (88 Hispanic/Latino and 50 Mapuche Indian samples) from Temuco Regional Hospital. The seven genes of interest were chosen from the previous set of 21 genes that were highly differentially expressed in the GBC tissue microarray

9 http://cgap.nci.nih.gov/SAGE
prepared at Johns Hopkins, and these were also included in the analysis. Overall, a total of 182 archival GBC collected over an 8-y period were used, which included tumors obtained from 125 Hispanic, 50 Mapuche Indian, 5 non-Hispanic Whites, and 2 Asian patients. There were 32 men and 150 women, and the mean age was 67 y (range, 30-95 y) and 63 y (range, 31-93 y), respectively. Most of the tumors were “advanced” GBC with invasion of the gallbladder subserosa (129 cases, 71%) and serosa (48 cases, 26%), with only a minority (5 cases, 3%) classified as “early” GBC (mucosa or muscularis propria invasion). By histologic grade, 46 (25%) GBC were well differentiated, 81 (45%) moderately differentiated, and 55 (30%) poorly differentiated. Postoperative follow-up was available on 170 of 182 (93%) cases, and these were classified based on the predominant pattern of expression in the neoplastic cells. Subsequently, the 182 GBC were classified as “absent/low” CTGF expressers (intensity of 0-1) or “high” CTGF expressers (intensity of 2-3).

Statistical analysis. Associations between CTGF expression and other clinicopathologic variables were assessed by Pearson’s χ², Mann-Whitney, and Fisher’s exact tests. Survival analysis was done using the Kaplan-Meier method and compared using the log-rank test. The proportional hazard regression analysis for predictors of survival was assessed with the Cox regression model. Significant values were defined as P < 0.05 (SPSS).

Results

Identification of differentially expressed SAGE tags in GBC. In the four libraries, 312,622 L-SAGE tags (21-bp length; range, 66,988-85,706 tags) were generated, of which 105,927 (33%) tags were unique. The numbers of tags and unique tags for each library are shown in Supplementary Table S2. We compared SAGE tags from each of the three GBC libraries with the normal alithiasic gallbladder tissue (N) by pairwise analysis. The results of these pairwise comparisons (significance level of P < 0.001) are summarized in Table 1. To further enhance the stringency of our analysis and elucidate an enriched subset of differentially expressed transcripts in GBC, we identified SAGE tags that were significantly different in all three pairwise comparisons.

Table 1. Pairwise analysis of GBC L-SAGE libraries

<table>
<thead>
<tr>
<th>Pairwise comparison</th>
<th>Relation</th>
<th>Total tags per comparison</th>
<th>Unique tags</th>
<th>Altered tags</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal vs tumor H (Hispanic/Latino)</td>
<td>N &gt; H</td>
<td>126,032</td>
<td>31,780</td>
<td>268</td>
<td>47.18</td>
</tr>
<tr>
<td></td>
<td>N &lt; H</td>
<td>292</td>
<td></td>
<td></td>
<td>47.18</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>152</td>
<td>140</td>
<td>52.05</td>
<td>47.95</td>
</tr>
<tr>
<td>Normal vs tumor M (Mapuche Indian)</td>
<td>N &gt; M</td>
<td>135,935</td>
<td>34,692</td>
<td>568</td>
<td>52.82</td>
</tr>
<tr>
<td></td>
<td>N &lt; M</td>
<td>292</td>
<td></td>
<td></td>
<td>49.48</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>256,421</td>
<td>49,726</td>
<td>104</td>
<td>50.52</td>
</tr>
<tr>
<td>Normal vs tumor C (Caucasian)</td>
<td>N &gt; C</td>
<td>136,154</td>
<td>33,420</td>
<td>578</td>
<td>49.48</td>
</tr>
<tr>
<td></td>
<td>N &lt; C</td>
<td>292</td>
<td></td>
<td></td>
<td>49.48</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>256,421</td>
<td>49,726</td>
<td>104</td>
<td>50.52</td>
</tr>
<tr>
<td>Normal vs all tumors</td>
<td>N &gt; T</td>
<td>136,154</td>
<td>33,420</td>
<td>578</td>
<td>49.48</td>
</tr>
<tr>
<td></td>
<td>N &lt; T</td>
<td>292</td>
<td></td>
<td></td>
<td>49.48</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>256,421</td>
<td>49,726</td>
<td>104</td>
<td>50.52</td>
</tr>
</tbody>
</table>

NOTE: In pairwise analysis, only SAGE tags that were differentially expressed at the P < 0.001 level were included (Audic and Claverie test).

Table 2. Significantly up-regulated SAGE tags in all three pairwise comparisons of GBC versus normal

<table>
<thead>
<tr>
<th>Tag</th>
<th>Symbol</th>
<th>Gene name</th>
<th>Unigene</th>
<th>Entrez ID</th>
<th>Band</th>
<th>Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>GATAATCGTCATGCAGTAC</td>
<td>HFM1</td>
<td>HFM1, ATP-dependent DNA helicase homologue (S. cerevisiae)</td>
<td>Hs.454818</td>
<td>164045</td>
<td>1p22.2</td>
<td>77.5</td>
</tr>
<tr>
<td>GAAGCAATTAATTCCTT</td>
<td>HLA-DQA2</td>
<td>MHC, class II, DQ α2</td>
<td>Hs.59798</td>
<td>3118</td>
<td>6p21.3</td>
<td>38</td>
</tr>
<tr>
<td>CCGACGGGATTTCCTT</td>
<td>COL3A1</td>
<td>Collagen, type III, α1 (Ehlers-Danlos syndrome type IV)</td>
<td>Hs.443625</td>
<td>1281</td>
<td>2q31</td>
<td>14.6</td>
</tr>
<tr>
<td>AGAACATCCAGAGTGCC</td>
<td>HLA-A</td>
<td>MHC, class I, A</td>
<td>Hs.181244</td>
<td>3105</td>
<td>6p21.3</td>
<td>12.4</td>
</tr>
<tr>
<td>TACCTCTGTAATAAA</td>
<td>S100P</td>
<td>S100 calcium binding protein P</td>
<td>Hs.2962</td>
<td>6286</td>
<td>4p16</td>
<td>8</td>
</tr>
<tr>
<td>GACCCGAGTGGCGGCT</td>
<td>KRT17</td>
<td>Keratin 17</td>
<td>Hs.2785</td>
<td>3872</td>
<td>17q12-q21</td>
<td>4.6</td>
</tr>
<tr>
<td>CTTCCAGCATTAGGGCTC</td>
<td>ANXA2</td>
<td>Annexin A2</td>
<td>Hs.511605</td>
<td>302</td>
<td>15q21-q22</td>
<td>3.6</td>
</tr>
<tr>
<td>GTTCGGGCGTTGAGA</td>
<td>TAGLN2</td>
<td>Transgelin 2</td>
<td>Hs.517168</td>
<td>8407</td>
<td>1q21-q25</td>
<td>2.5</td>
</tr>
<tr>
<td>TTTCGGACCTTCTGATGTT</td>
<td>CTGF</td>
<td>Connective tissue growth factor</td>
<td>Hs.591346</td>
<td>1490</td>
<td>6q23.1</td>
<td>2.3</td>
</tr>
<tr>
<td>GTTTGCGTTAATCTGTT</td>
<td>B2M</td>
<td>β2-microglobulin</td>
<td>Hs.534255</td>
<td>567</td>
<td>15q21-q22.2</td>
<td>1.7</td>
</tr>
</tbody>
</table>
In this stringent analysis, a larger number of SAGE tags were significantly down-regulated \( (n = 93) \) versus up-regulated tags \( (n = 11) \) in GBC compared with nonneoplastic gallbladder, and the results are tabulated in Table 2 and Supplementary Table S3 for up-regulated and down-regulated tags, respectively. Many of the corresponding genes identified as differentially overexpressed in GBC (e.g., S100P, \( \beta_2 \)-microglobulin, annexin A2, keratin 17, and CTGF) have previously been reported to be overexpressed in human cancers (16–21), underscoring the overall validity of our approach; these transcripts represent novel candidates that have a high likelihood of becoming clinically relevant biomarkers for GBC. Of note, the maximal fold change in GBC versus nonneoplastic gallbladder was observed for the SAGE tag corresponding to the human homologue of the yeast MER3 gene (hHFM1), a RecQ-like helicase involved in meiosis (22). The gene product of HFM1 is an evolutionarily conserved helicase predominantly expressed in the gonads, and it is postulated to play a role in the maintenance of genomic stability during meiosis (23). To the best of our knowledge, this is the first demonstrable association between hHFM1 and human cancer.

Using a subset of 956 most highly up-regulated or down-regulated SAGE tags in pairwise comparisons at the \( P < 0.001 \) significance level, we performed unsupervised clustering using the TM4 software (24) to determine the relatedness of the three adenocarcinoma and one nonneoplastic libraries. Remarkably, hierarchical clustering revealed that the H and N samples, both obtained from patients of Hispanic/Latino ethnicity, were most similar to each other, despite the obvious differences in their histology (Fig. 1A). In contrast, the two remaining adenocarcinomas clustered separately, although the Mapuche Indian GBC had a greater degree of relatedness to the Hispanic/Latino expression profile than to the non-Hispanic White (Caucasian) GBC (Fig. 1B). Whereas the sample numbers preclude us from generalizing the results of this analysis, it suggests that, independent of the underlying histology, the ethnic background of the individual likely has a profound influence on global expression profile and, hence, sample relatedness on cluster analysis.

**CTGF is overexpressed in microdissected primary GBC but not in metastatic tissues.** From the enriched list of significantly overexpressed genes in GBC (Table 2), we selected CTGF for further validation by real-time quantitative PCR and immunohistochemistry. Microdissected epithelial cells were isolated from five primary GBC samples, matched uninvolved gallbladder mucosa adjacent to the cancers, and three unmatched metastasis samples. Real-time quantitative PCR analysis confirmed that CTGF is overexpressed in primary GBC compared with nonneoplastic gallbladder epithelium in all five of five cases (Fig. 2A). When the fold changes were averaged, there was significant up-regulation of CTGF \( (P < 0.05) \) in primary GBC versus nonneoplastic epithelium (Fig. 2B). Of note, CTGF transcript levels were not significantly different in the three metastasis samples compared with nonneoplastic gallbladder epithelium, suggesting that the observed CTGF mRNA overexpression is restricted to the primary cancers and is subsequently down-regulated on tumor progression.

**CTGF protein is overexpressed in GBC and correlates with improved survival.** Immunohistochemistry for CTGF protein was done in our series of GBC tissue microarrays. CTGF labeling was absent/low in 6 of 6 (100%) nonneoplastic gallbladder control tissue cores on the tissue microarrays (Fig. 3A), including in areas of pyloric metaplasia (Fig. 3B). Of the 182 GBC cases analyzed, 116 (64%) had low/absent CTGF expression, whereas 66 (36%) showed high CTGF levels (Fig. 3C and D). No correlation was observed between CTGF expression levels and a variety of clinicopathologic variables (Supplementary Table S4) including gender, age, ethnicity, tumor differentiation, degree of infiltration, tumor-node-metastasis classification, or American Joint Committee on Cancer stage. Follow-up survival data was available on 170 of 182 patients whose GBC were assessed for CTGF expression. Patients with absent or low CTGF labeling \( (n = 110) \) had a
median survival of 1.1 years, whereas patients with high CTGF labeling ($n = 60$) had a median survival of 3.5 years. Thus, high CTGF expression in the primary GBC was significantly associated with better survival ($P = 0.0069$, log-rank test; Fig. 4). In univariate analysis, besides CTGF, other significant predictors of survival in GBC included degree of infiltration of the tumor, T classification, metastasis status, and American Joint Committee on Cancer stage (Supplementary Table S5). In the multivariate Cox regression analysis, degree of infiltration and metastases retained their significance ($P = 0.004$ and $P = 0.001$, respectively), but CTGF expression was borderline nonsignificant ($P = 0.068$) as an independent factor in predicting survival (Supplementary Table S6).

## Discussion

SAGE provides an unbiased and quantitative approach toward identifying differentially expressed transcripts in human neoplasms, facilitating discovery of cancer biomarkers, imaging targets, and therapeutic avenues (6, 8). For example, in 2001, our group used SAGE to first identify mesothelin as an up-regulated transcript in pancreatic cancer; the gene product of mesothelin is a glycosylphosphatidylinositol-anchored protein (25). Since that time, mesothelin expression has been used as an adjunct in cytologic diagnosis (26), as a candidate biomarker for early detection of pancreatic cancer (14), as a potential therapeutic target using monoclonal antibodies, and as the target of adoptive immunotherapy in patients receiving adjuvant vaccination against pancreatic cancer (27). A comparable success has been achieved for prostate stem cell antigen, the transcript overexpression of which was also first identified by our group in SAGE libraries of pancreatic cancer (25). Prostate stem cell antigen expression has been used as an adjunct in cytologic diagnosis (26) and as a target of pancreatic cancer immunotherapy (28). In addition to immediate translational effect, SAGE also provides novel insights into basic cancer biology, such as understanding of tumor-stroma relationships and mechanisms of tumor progression (29).

In this study, we used 21-bp L-SAGE to generate libraries from three ethnically diverse GBC and one nonneoplastic gallbladder mucosa. In addition to a GBC derived from a non-Hispanic White (Caucasian) individual, we also selected samples of Native American (Mapuche Indian) and Hispanic/Latino heritage because these populations show a significantly higher propensity for developing GBC (5). Hierarchical clustering of the four libraries using the top 956 maximally dysregulated SAGE tags unexpectedly showed a greater degree of relatedness of the Hispanic/Latino adenocarcinoma to the unmatched Hispanic/Latino normal mucosal sample than to the two remaining adenocarcinomas derived from Mapuche Indian and Caucasian backgrounds. The current prohibitive costs of large-scale automated sequencing (as required for SAGE libraries) preclude us from extrapolating this preliminary observation to additional cancer and normal samples obtained from diverse ethnicities. Nevertheless, there is emerging evidence that underlying ethnicity has profound influence on gene expression profiles (30–32), with as many as a quarter of the genes in the human genome displaying significant differences in expression between populations (30). Therefore, in terms of gene expression relatedness, the commonality of a Hispanic/Latino background in two samples seems to circumvent their differences in histology (normal versus cancer) when compared with two ethnically unrelated cancer samples. Of note, other genetic discrepancies are also present in cancers arising in different ethnic backgrounds (33), and multiethnic studies of the type reported here may be necessary to obtain the complete genetic spectrum for a given neoplasm.

Using highly stringent analysis criteria, we identified SAGE tags corresponding to CTGF as being significantly overexpressed in GBC compared with nonneoplastic gallbladder mucosa. We confirmed the significant overexpression of CTGF transcripts in microdissected GBC compared with nonneoplastic gallbladder epithelium, but we were unable to show overexpression in metastatic samples. Finally, we confirmed immunohistochemical overexpression of CTGF protein at high levels in 36% of archival GBC and showed that cancers with high CTGF had a significantly better outcome than low/absent CTGF expressers on Kaplan-Meier survival analysis ($P = 0.0069$). CTGF, also known as CCN2, is a member of the CCN family, which derives its acronym from the initials of CTGF, Cyr61, and Nov, comprising the three genes represented in this family (34). CTGF is up-regulated by a multitude of extracellular stimuli, including hypoxia and
TGF-β signaling (35, 36). CTGF has been implicated in a broad range of functions in vivo, such as proliferation, migration, and angiogenesis, as well as growth regulation of mesenchymal cells (36). CTGF is overexpressed in several solid cancers including breast, colorectal, lung, esophageal, and pancreatic cancers, melanoma, and gliomas (15, 20, 21, 37-43); to the best of our knowledge, this is the first report of CTGF overexpression in GBC. In the vast majority of published studies, CTGF is expressed in the neoplastic cells themselves, with or without accompanying stromal expression, which is concordant with our own observations in GBC. The functional implications of neoplastic CTGF overexpression seem to be context dependent in terms of the tumor type. For example, in breast cancers, high CTGF levels have been correlated with the presence of osteolytic metastases (20, 37); similarly, CTGF overexpression is associated with tumor progression and/or adverse prognosis in gliomas (43), esophageal adenocarcinomas (41), and pancreatic cancer (21). On the contrary, and in concordance with our own findings in GBC, CTGF overexpression is associated with metastasis inhibition and is a favorable prognostic marker in colorectal cancers (39), non-small cell lung cancers (15, 40, 44), and in some studies on primary breast cancers (38). Recent studies in lung cancer models suggest that the metastasis inhibition by CTGF is mediated through degradation of hypoxia inducible factor 1α and consequent reduced angiogenesis (44). Of note, in our study, CTGF overexpression was observed in primary cancers, but its expression was reduced to levels comparable to that of normal gallbladder epithelium in metastatic lesions. The basis for this dichotomy is unclear, and may represent differing requirements or roles for CTGF during tumor development and during progression. For example, TGF-β is implicated as a tumor suppressor during the early stages of multistep tumor progression, but it is also known to function as a proangiogenic and metastasis-promoting molecule during the later stages of tumorigenesis (45). Functional studies in GBC cell lines are warranted to elucidate how modulation of endogenous CTGF levels affects tumor initiation and progression.

In addition to the “single gene” validation of CTGF, we have preliminarily mined the SAGE libraries for known and putative abrogated signaling pathways in GBC, using a web-based Pathway Explorer tool (13). For example, Kiguchi and colleagues have recently described a unique transgenic model of GBC that develops on aberrant c-erbB2 expression from a bovine keratin BK5 promoter (46). Subsequent studies by this
Can interrogate large-scale databases for expression platforms and model systems. Analysis modules that confirm the consistency of this observation across multiple libraries versus the nonneoplastic gallbladder library, thus allowing further mining of the GBC SAGE libraries for aberrantly FRAP1/mTOR.


In summary, we have generated a panel of multietiologic SAGE libraries of GBC and identified CTGF expression as a prognostic biomarker in this neoplasm. The availability of these libraries on the Genome Cancer Anatomy Project SAGE Genie web site will allow scientists to query the expression data for biomarkers and therapeutic targets in GBC and will facilitate a better understanding of signaling pathways involved in gallbladder carcinogenesis occurring in distinct ethnic backgrounds. Finally, the first normal gallbladder SAGE library we have generated is a potential source for identifying tissue specific transcripts, as done recently in mice and human SAGE data sets (50); such tissue-specific transcripts are the seedbed for developing genetically engineered animal models that recapitulate the cognate human disease.

Acknowledgments

We thank the family of Margaret Lee.

References


Serial Analysis of Gene Expression Identifies Connective Tissue Growth Factor Expression as a Prognostic Biomarker in Gallbladder Cancer

Hector Alvarez, Alejandro Corvalan, Juan C. Roa, et al.


Updated version Access the most recent version of this article at: http://clincancerres.aacrjournals.org/content/14/9/2631

Supplementary Material Access the most recent supplemental material at: http://clincancerres.aacrjournals.org/content/suppl/2008/07/11/14.9.2631.DC1

Cited articles This article cites 50 articles, 23 of which you can access for free at: http://clincancerres.aacrjournals.org/content/14/9/2631.full.html#ref-list-1

Citing articles This article has been cited by 2 HighWire-hosted articles. Access the articles at: /content/14/9/2631.full.html#related-urls

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.