

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

Hector Alvarez,^{1,5} Alejandro Corvalan,¹ Juan C. Roa,³ Pedram Argani,⁵ Francisco Murillo,⁸ Jennifer Edwards,⁶ Robert Beaty,⁵ Georg Feldmann,⁵ 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 Maitra^{5,7,8}

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 alithiasic 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

Authors' Affiliations: Departments of ¹Pathology and ²Surgery, Pontificia Universidad Católica de Chile, Santiago, Chile; ³Department of Pathology, Universidad de la Frontera, Temuco, Chile; ⁴Department of Surgery, Hospital Sótero del Río, Santiago Chile; and Departments of ⁵Pathology, ⁶Neurosurgery, and ⁷Oncology, and ⁸McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland
Received 8/12/07; revised 10/21/07; accepted 1/25/08.

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.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

Requests for reprints: Anirban Maitra, Johns Hopkins University School of Medicine, CRB II, Room 345, 1550 Orleans Street, Baltimore, MD 21231. Phone: 410-955-3511; Fax: 410-614-0671; E-mail: Amaitra1@jhmi.edu or Gregory J. Riggins, Johns Hopkins University School of Medicine, CRB II, Room 257, 1550 Orleans Street, Baltimore, MD 21231. E-mail: Griggin1@jhmi.edu.

© 2008 American Association for Cancer Research.

doi:10.1158/1078-0432.CCR-07-1991

composed of short nucleic acid sequences or "tags" that are concatemerized 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 Catolica 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 IIa/IIb adenocarcinomas. Cryo-molds were prepared from all three snap-frozen specimens, and neoplastic 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 RNAgents kit (Promega) according to the manufacturer's instructions. L-SAGE libraries were constructed with *Nla*III as the anchoring enzyme and *Mme*I 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 project⁹ (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 Catolica 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 μ m) 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 *GPD1* (house-keeping gene), sense AGGGGGGAGCCAAAAGGG and antisense TGCCAGCCCCAGCGTCAAAG, and for *CTGF*, sense TGCCAGGCT-GATTCTAGGT and antisense GGTGCAAACATGTAACCTTTGG. All quantitative PCR experiments were done in triplicate. Analysis of relative gene expression changes was done with the $2^{-\Delta\Delta CT}$ 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. Tissue microarrays for immunohistochemistry were generated from these samples. An additional seven GBC specimens (obtained from two Asian and five non-Hispanic White patients) were arrayed on a previously described biliary cancer tissue microarray

⁹ <http://cgap.nci.nih.gov/SAGE>

Table 1. Pairwise analysis of GBC L-SAGE libraries

Pairwise comparison	Relation	Total tags per comparison	Unique tags	Altered tags	%
Normal vs tumor H (Hispanic/Latino)	N > H	126,032	31,780	152	52.05
	N < H			140	47.95
	Total			292	
Normal vs tumor M (Mapuche Indian)	N > M	135,935	34,692	268	47.18
	N < M			300	52.82
	Total			568	
Normal vs tumor C (Caucasian)	N > C	136,154	33,420	292	50.52
	N < C			286	49.48
	Total			578	
Normal vs all tumors	N > T	256,421	49,726	93	89.42
	N < T			11	10.58
	Total			104	

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

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 used for Kaplan-Meier survival and Cox regression analyses.

Immunohistochemical staining. Unstained 4-mm sections were cut from each tissue microarray and deparaffinized before treatment with 200-mL antigen retrieval solution, pH 6.0, for 15 min at 95°C. After cooling for 30 min, slides were quenched with 3% H₂O₂ for 10 min before incubation with the goat polyclonal anti-CTGF antibody (1:100 dilution; clone L-20, Santa Cruz Biotechnology) for 120 min at room temperature. Labeling was detected with the DakoCytomation DAB system according to the manufacturer's protocol. All sections were counterstained with Harris' hematoxylin. Immunohistochemical labeling was assessed in an outcome-blinded fashion by two of the authors (H.A. and A.M.) on a compound microscope. CTGF is expressed in the cytoplasmic compartment with membranous accentuation (15). Based on intensity of labeling in the neoplastic cells, a four-tier numerical scheme (0, absent; 1, mild; 2, moderate; 3, intense) was used for evaluation; in cases with heterogeneous labeling intensity, the case was

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 χ^2 , 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 2. Significantly up-regulated SAGE tags in all three pairwise comparisons of GBC *versus* normal

Tag	Symbol	Gene name	Unigene	Entrez ID	Band	Fold change
GTAATCCTGCTCAGTAC	<i>HFM1</i>	HFM1, ATP-dependent DNA helicase homologue (<i>S. cerevisiae</i>)	Hs.454818	164045	1p22.2	77.5
GAAGCAATAAATCCCT	<i>HLA-DQA2</i>	MHC, class II, DQ α 2	Hs.591798	3118	6p21.3	38
CCACAGGGGATTCTCCT	<i>COL3A1</i>	Collagen, type III, α 1 (Ehlers-Danlos syndrome type IV)	Hs.443625	1281	2q31	14.6
AGAACCTTCCAGAGTCC	<i>HLA-A</i>	MHC, class I, A	Hs.181244	3105	6p21.3	12.4
TACCTCTGATTAATAAA	<i>S100P</i>	S100 calcium binding protein P	Hs.2962	6286	4p16	8
GACATCAAGTCGCGGCT	<i>KRT17</i>	Keratin 17	Hs.2785	3872	17q12-q21	4.6
CGCCGACGATGCCAGAG	<i>IF16</i>	IFN, α -inducible protein 6	Hs.511731	2537	1p35	4.2
CTTCCAGCTAACAGGTC	<i>ANXA2</i>	Annexin A2	Hs.511605	302	15q21-q22	3.6
GTCTGGGGCTTGAGGAA	<i>TAGLN2</i>	Transgelin 2	Hs.517168	8407	1q21-q25	2.5
TTTGACCTTTCTAGTT	<i>CTGF</i>	Connective tissue growth factor	Hs.591346	1490	6q23.1	2.3
GTTGTGGTTAATCTGGT	<i>B2M</i>	β ₂ -microglobulin	Hs.534255	567	15q21-q22.2	1.7

($P < 0.001$, Audic and Claverie test with Bonferroni correction). 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*, β_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

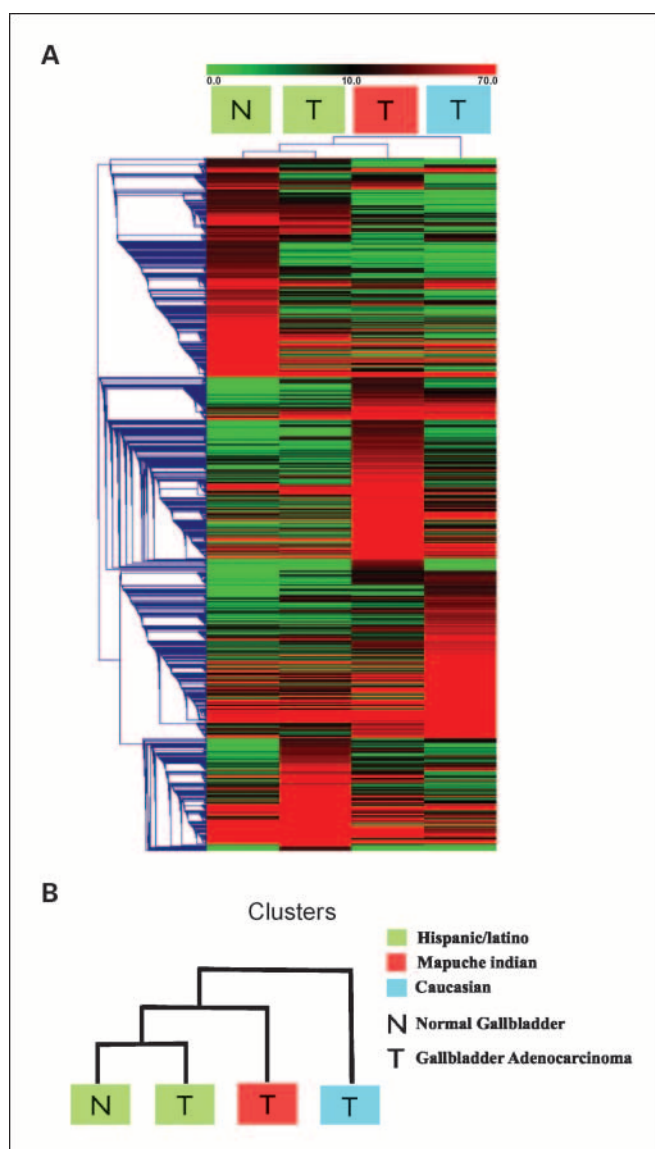


Fig. 1. A, hierarchical clustering of three GBC (Hispanic/Latino, Native American, and non-Hispanic White ethnicities) and nonneoplastic gallbladder mucosa (Hispanic/Latino ethnicity) using 956 most differentially expressed SAGE tags. B, clusters evidence that the two Hispanic/Latino samples are most similar to each other, despite their different underlying histology. The Native American (Mapuche Indian) GBC shows greater relatedness to the Hispanic/Latino samples than to the non-Hispanic White GBC.

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 expressors 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

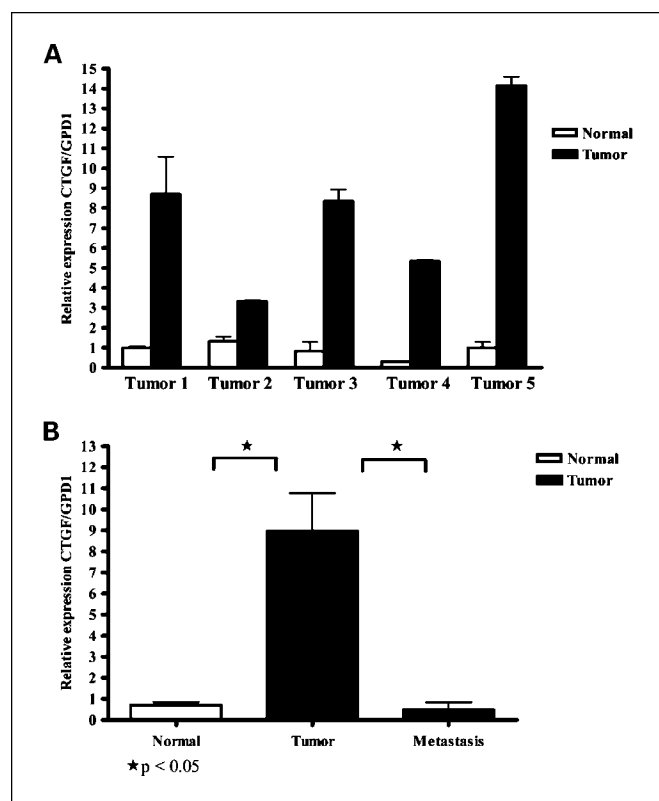


Fig. 2. *A*, *CTGF* real-time quantitative PCR analysis in paired samples of five microdissected GBC and nonneoplastic gallbladder epithelium confirms overexpression in all cases. *GPD1* is used as housekeeping control gene. Columns, mean of triplicate assays for each individual sample; bars, SD. *B*, relative fold *CTGF* transcript expression in the five nonneoplastic gallbladder epithelial samples, five GBC, and three unmatched metastases. Columns, average fold expression for each category. There is significant overexpression of *CTGF* in GBC compared with nonneoplastic gallbladder epithelium ($P < 0.05$), but not in the metastatic samples.

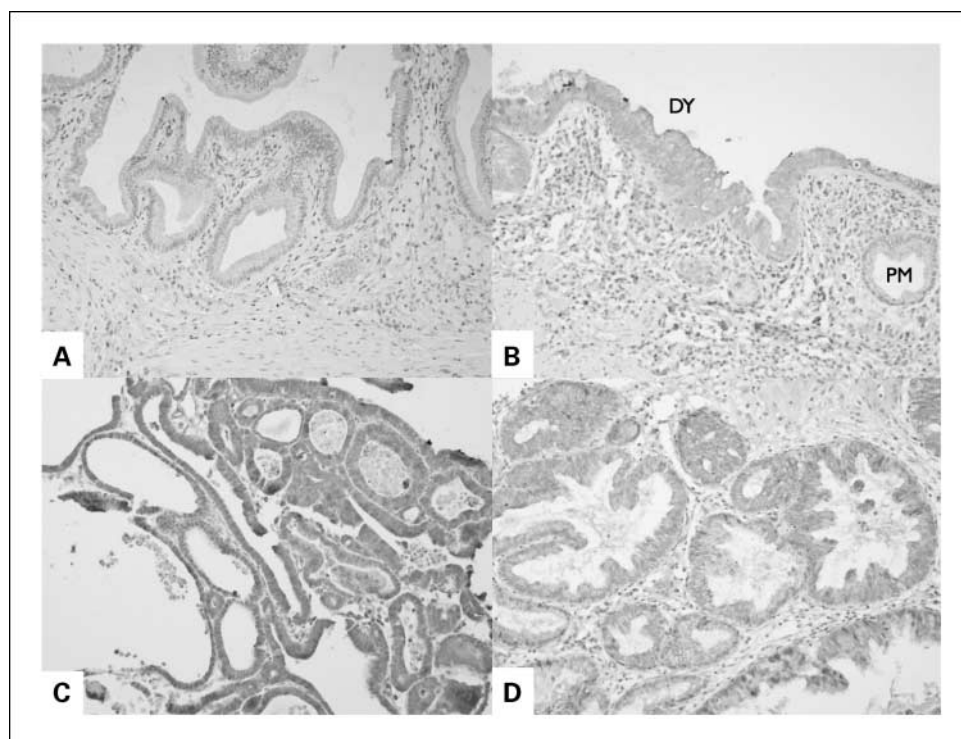


Fig. 3. Immunohistochemistry for CTGF protein expression in GBC and nonneoplastic gallbladder mucosa. *A*, absent/low CTGF expression in normal gallbladder epithelium. *B*, absence of negative labeling in pyloric metaplasia (*PM*), with CTGF expression in associated gallbladder dysplasia (*DY*). *C* and *D*, representative examples of GBC with high CTGF expression.

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 1 α and consequent reduced angiogenesis (44). Of note, in our study, CTGF overexpression 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

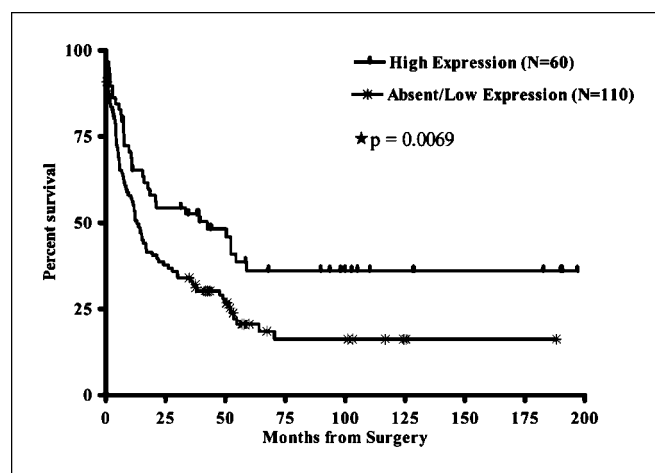


Fig. 4. Kaplan-Meier survival analysis of 170 GBC shows that cancers with high CTGF expression have a significantly better survival than cancers with absent/low CTGF expression ($P = 0.0069$).

group have established that the oncogenic mammalian target of rapamycin (mTOR) pathway is up-regulated in the resulting murine GBC, and that rapamycin, a specific mTOR small-molecule inhibitor, can significantly reduce tumorigenesis in the BK5.erbB2 model (47). Our group has also previously reported activation of the mTOR pathway in human biliary tract cancer cell lines using oligonucleotide-based (Affymetrix) microarrays (48). Therefore, we performed an *in silico* analysis of the mTOR pathway using the GBC SAGE data sets. We found transcriptional up-regulation of multiple mTOR pathway genes, including SAGE tags corresponding to *FRAP1/mTOR* itself, as well as upstream and downstream activating pathway components (*AKT1*, *RPS6KA1*, *EIF4E*, and *RHEB*) in the GBC libraries versus the nonneoplastic gallbladder library, thus confirming the consistency of this observation across multiple expression platforms and model systems. Analysis modules that can interrogate large-scale databases for *in silico* identification of enriched gene sets and “druggable” pathways (49) should allow further mining of the GBC SAGE libraries for aberrantly activated signaling pathways like mTOR. These pathways could be targeted in relevant preclinical models of GBC, such as

xenografts or in BK5.erbB2 mice, as a preamble toward eventual clinical translation.

In summary, we have generated a panel of multiethnic SAGE libraries of GBC and identified CTGF expression as a prognostic biomarker in this neoplasm. The availability of these libraries on the Cancer Genome 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

- Albores-Saavedra J, Henson DE, Sobin LH. The WHO histological classification of tumors of the gallbladder and extrahepatic bile ducts. A commentary on the second edition. *Cancer* 1992;70:410–4.
- Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. Cancer statistics, 2007. *CA Cancer J Clin* 2007;57:43–66.
- Randi G, Franceschi S, La Vecchia C. Gallbladder cancer worldwide: geographical distribution and risk factors. *Int J Cancer* 2006;118:1591–602.
- Barakat J, Dunkelberg JC, MaTY. Changing patterns of gallbladder carcinoma in New Mexico. *Cancer* 2006;106:434–40.
- Wistuba II, Gazdar AF. Gallbladder cancer: lessons from a rare tumour. *Nat Rev Cancer* 2004;4:695–706.
- Hu M, Polyak K. Serial analysis of gene expression. *Nat Protoc* 2006;1:1743–60.
- Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. *Science* 1995;270:484–7.
- Riggins GJ, Strausberg RL. Genome and genetic resources from the Cancer Genome Anatomy Project. *Hum Mol Genet* 2001;10:663–7.
- Saha S, Sparks AB, Rago C, et al. Using the transcriptome to annotate the genome. *Nat Biotechnol* 2002;20:508–12.
- Boon K, Osorio EC, Greenhut SF, et al. An anatomy of normal and malignant gene expression. *Proc Natl Acad Sci U S A* 2002;99:11287–92.
- Romualdi C, Bortoluzzi S, D'Alessi F, Danieli GA. IDEG6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments. *Physiol Genomics* 2003;12:159–62.
- Lash AE, Tolstoshev CM, Wagner L, et al. SAGEmap: a public gene expression resource. *Genome Res* 2000;10:1051–60.
- Mlecnik B, Scheideler M, Hackl H, Hartler J, Sanchez-Cabo F, Trajanoski Z. PathwayExplorer: web service for visualizing high-throughput expression data on biological pathways. *Nucleic Acids Res* 2005;33:W633–7.
- Maitra A, Adsay NV, Argani P, et al. Multicomponent analysis of the pancreatic adenocarcinoma progression model using a pancreatic intraepithelial neoplasia tissue microarray. *Mod Pathol* 2003;16:902–12.
- Chang CC, Shih JY, Jeng YM, et al. Connective tissue growth factor and its role in lung adenocarcinoma invasion and metastasis. *J Natl Cancer Inst* 2004;96:364–75.
- Fuentes MK, Nigavekar SS, Arumugam T, et al. RAGE activation by S100P in colon cancer stimulates growth, migration, and cell signaling pathways. *Dis Colon Rectum* 2007;50:1230–40.
- Nomura T, Huang WC, Zhou HE, et al. β_2 -microglobulin promotes the growth of human renal cell carcinoma through the activation of the protein kinase A, cyclic AMP-responsive element-binding protein, and vascular endothelial growth factor axis. *Clin Cancer Res* 2006;12:7294–305.
- Emoto K, Yamada Y, Sawada H, et al. Annexin II overexpression correlates with stromal tenascin-C overexpression: a prognostic marker in colorectal carcinoma. *Cancer* 2001;92:1419–26.
- Potemski P, Kusinska R, Watala C, Pluciennik E, Bednarek AK, Kordek R. Prognostic relevance of basal cytokeratin expression in operable breast cancer. *Oncology* 2005;69:478–85.
- Kang Y, Siegel PM, Shu W, et al. A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* 2003;3:537–49.
- Dornhofer N, Spong S, Bennewith K, et al. Connective tissue growth factor-specific monoclonal antibody therapy inhibits pancreatic tumor growth and metastasis. *Cancer Res* 2006;66:5816–27.
- Nakagawa T, Kolodner RD. *Saccharomyces cerevisiae* Mer3 is a DNA helicase involved in meiotic crossing over. *Mol Cell Biol* 2002;22:3281–91.
- Tanaka K, Miyamoto N, Shouguchi-Miyata J, Ikeda JE. HFM1, the human homologue of yeast Mer3, encodes a putative DNA helicase expressed specifically in germ-line cells. *DNA Seq* 2006;17:242–6.
- Saeed AI, Sharov V, White J, et al. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 2003;34:374–8.
- Argani P, Iacobuzio-Donahue C, Ryu B, et al. Mesothelin is overexpressed in the vast majority of ductal adenocarcinomas of the pancreas: identification of a new pancreatic cancer marker by serial analysis of gene expression (SAGE). *Clin Cancer Res* 2001;7:3862–8.
- McCarthy DM, Maitra A, Argani P, et al. Novel markers of pancreatic adenocarcinoma in fine-needle aspiration: mesothelin and prostate stem cell antigen labeling increases accuracy in cytologically borderline cases. *Appl Immunohistochem Mol Morphol* 2003;11:238–43.
- Hassan R, Bera T, Pastan I. Mesothelin: a new target for immunotherapy. *Clin Cancer Res* 2004;10:3937–42.
- Gu Z, Yamashiro J, Kono E, Reiter RE. Anti-prostate stem cell antigen monoclonal antibody 1G8 induces cell death *in vitro* and inhibits tumor growth *in vivo* via a Fc-independent mechanism. *Cancer Res* 2005;65:9495–500.
- Porter DA, Krop IE, Nasser S, et al. A SAGE (serial analysis of gene expression) view of breast tumor progression. *Cancer Res* 2001;61:5697–702.
- Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG. Common genetic variants account for differences in gene expression among ethnic groups. *Nat Genet* 2007;39:226–31.
- Storey JD, Madeoy J, Strout JL, Wurfel M, Ronald J, Akey JM. Gene-expression variation within and among human populations. *Am J Hum Genet* 2007;80:502–9.
- Stranger BE, Nica AC, Forrest MS, et al. Population genomics of human gene expression. *Nat Genet* 2007;39:1217–24.
- Calvo E, Baselga J. Ethnic differences in response to epidermal growth factor receptor tyrosine kinase inhibitors. *J Clin Oncol* 2006;24:2158–63.
- Perbal B. CCN proteins: multifunctional signalling regulators. *Lancet* 2004;363:62–4.
- Kondo S, Kubota S, Shimo T, et al. Connective tissue growth factor increased by hypoxia may initiate angiogenesis in collaboration with matrix metalloproteinases. *Carcinogenesis* 2002;23:769–76.
- Leask A, Holmes A, Abraham DJ. Connective tissue growth factor: a new and important player in the pathogenesis of fibrosis. *Curr Rheumatol Rep* 2002;4:136–42.
- Shimo T, Kubota S, Yoshioka N, et al. Pathogenic role of connective tissue growth factor (CTGF/CCN2) in osteolytic metastasis of breast cancer. *J Bone Miner Res* 2006;21:1045–59.
- Jiang WG, Watkins G, Fodstad O, Douglas-Jones A, Mokbel K, Mansel RE. Differential expression of the CCN family members Cyr61, CTGF and Nov in human breast cancer. *Endocr Relat Cancer* 2004;11:781–91.
- Lin BR, Chang CC, Che TF, et al. Connective tissue growth factor inhibits metastasis and acts as an

- independent prognostic marker in colorectal cancer. *Gastroenterology* 2005;128:9–23.
40. Chien W, Yin D, Gui D, et al. Suppression of cell proliferation and signaling transduction by connective tissue growth factor in non-small cell lung cancer cells. *Mol Cancer Res* 2006;4:591–8.
41. Koliopoulos A, Friess H, di Mola FF, et al. Connective tissue growth factor gene expression alters tumor progression in esophageal cancer. *World J Surg* 2002;26:420–7.
42. Kubo M, Kikuchi K, Nashiro K, et al. Expression of fibrogenic cytokines in desmoplastic malignant melanoma. *Br J Dermatol* 1998;139:192–7.
43. Xie D, Yin D, Wang HJ, et al. Levels of expression of CYR61 and CTGF are prognostic for tumor progression and survival of individuals with gliomas. *Clin Cancer Res* 2004;10:2072–81.
44. Chang CC, Lin MT, Lin BR, et al. Effect of connective tissue growth factor on hypoxia-inducible factor 1 α degradation and tumor angiogenesis. *J Natl Cancer Inst* 2006;98:984–95.
45. Zavadil J, Bottinger EP. TGF- β and epithelial-to-mesenchymal transitions. *Oncogene* 2005;24:5764–74.
46. Kiguchi K. Constitutive expression of ErbB-2 in gallbladder epithelium results in development of adenocarcinoma. *Cancer Res* 2001;61:6971–6.
47. Wu Q, Kiguchi K, Kawamoto T, et al. Therapeutic effect of rapamycin on gallbladder cancer in a transgenic mouse model. *Cancer Res* 2007;67:3794–800.
48. Hansel DE, Rahman A, Hidalgo M, et al. Identification of novel cellular targets in biliary tract cancers using global gene expression technology. *Am J Pathol* 2003;163:217–29.
49. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545–50.
50. Kouadjo KE, Nishida Y, Cadrin-Girard JF, Yoshioka M, St-Amand J. Housekeeping and tissue-specific genes in mouse tissues. *BMC Genomics* 2007;8:127.