

# Comparative Profiling of Primary Colorectal Carcinomas and Liver Metastases Identifies LEF1 as a Prognostic Biomarker

Albert Y. Lin<sup>1,2\*</sup>, Mei-Sze Chua<sup>3</sup>, Yoon-La Choi<sup>4,5</sup>, William Yeh<sup>1</sup>, Young H. Kim<sup>4</sup>, Raymond Azzi<sup>6</sup>, Gregg A. Adams<sup>7</sup>, Kristin Sainani<sup>8</sup>, Matt van de Rijn<sup>4</sup>, Samuel K. So<sup>3</sup>, Jonathan R. Pollack<sup>4</sup>

**1** Department of Medicine, Santa Clara Valley Medical Center, San Jose, California, United States of America, **2** Department of Medicine, Stanford University, Stanford, California, United States of America, **3** Department of Surgery and Asian Liver Center, Stanford University, Stanford, California, United States of America, **4** Department of Pathology, Stanford University, Stanford, California, United States of America, **5** Department of Pathology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea, **6** Department of Pathology, Santa Clara Valley Medical Center, San Jose, California, United States of America, **7** Department of Surgery, Santa Clara Valley Medical Center, San Jose, California, United States of America, **8** Division of Epidemiology, Health Research and Policy, Stanford University, Stanford, California, United States of America

## Abstract

**Purpose:** We sought to identify genes of clinical significance to predict survival and the risk for colorectal liver metastasis (CLM), the most common site of metastasis from colorectal cancer (CRC).

**Patients and Methods:** We profiled gene expression in 31 specimens from primary CRC and 32 unmatched specimens of CLM, and performed Significance Analysis of Microarrays (SAM) to identify genes differentially expressed between these two groups. To characterize the clinical relevance of two highly-ranked differentially-expressed genes, we analyzed the expression of secreted phosphoprotein 1 (SPP1 or osteopontin) and lymphoid enhancer factor-1 (LEF1) by immunohistochemistry using a tissue microarray (TMA) representing an independent set of 154 patients with primary CRC.

**Results:** Supervised analysis using SAM identified 963 genes with significantly higher expression in CLM compared to primary CRC, with a false discovery rate of <0.5%. TMA analysis showed SPP1 and LEF1 protein overexpression in 60% and 44% of CRC cases, respectively. Subsequent occurrence of CLM was significantly correlated with the overexpression of LEF1 (chi-square  $p=0.042$ ), but not SPP1 ( $p=0.14$ ). Kaplan Meier analysis revealed significantly worse survival in patients with overexpression of LEF1 ( $p<0.01$ ), but not SPP1 ( $p=0.11$ ). Both univariate and multivariate analyses identified stage ( $p<0.0001$ ) and LEF1 overexpression ( $p<0.05$ ) as important prognostic markers, but not tumor grade or SPP1.

**Conclusion:** Among genes differentially expressed between CLM and primary CRC, we demonstrate overexpression of LEF1 in primary CRC to be a prognostic factor for poor survival and increased risk for liver metastasis.

**Citation:** Lin AY, Chua M-S, Choi Y-L, Yeh W, Kim YH, et al. (2011) Comparative Profiling of Primary Colorectal Carcinomas and Liver Metastases Identifies LEF1 as a Prognostic Biomarker. PLoS ONE 6(2): e16636. doi:10.1371/journal.pone.0016636

**Editor:** John Luk, National University of Singapore, Singapore

**Received:** August 17, 2010; **Accepted:** January 3, 2011; **Published:** February 24, 2011

**Copyright:** © 2011 Lin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** Supported in part by NIH grant CA112016 (JRP); The Valley Foundation (AYL); Lui Hac Minh and C. J. Huang Foundations and the Paul and Tammy Davis Research Fund (MSC and SKS). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: albert.lin@hhs.sccgov.org

## Introduction

Despite considerable progress in the diagnosis and treatment of colorectal cancer (CRC) over the last few decades leading to a significant decline in cancer-related mortality,[1,2] CRC remains a major public health problem throughout the world. In the United States, CRC is the third most common cancer and is also the third leading cause of cancer death in men and women combined.[3]

Worldwide it represents the third most common cancer and second most common cause of cancer-related death.[4] Once metastasis has occurred in CRC, a complete cure of the disease is unlikely. Therefore, there is a need for better understanding of the molecular mechanisms underlying the metastatic phenotype that

may provide information leading to the development of drugs to control or prevent metastatic disease.[1]

Colorectal liver metastasis (CLM), occurring in about 60% of CRC patients during the course of their treatment, is the most common distant metastasis from CRC. Several clinical prognostic factors, such as lymph node status and size of the primary tumor, have been identified for CLM.[5,6] However, little is known about the prognostic significance of molecular markers for CLM.

Recent development and application of human genome and high-throughput technologies, such as DNA microarrays, allows us to simultaneously examine thousands of genes, leading to a much better understanding of carcinogenesis - a great step toward individualized personal medicine.[7] Published studies on CRC gene expression profiling have mainly examined normal vs.

tumor tissues or different stages of CRC,[8,9] or treatment outcomes by the differences in gene expression profiling.[10,11] To identify molecular markers of clinical significance, we used DNA microarrays to compare the gene-expression profiles of specimens from primary CRC and specimens from CLM. We reasoned that genes upregulated in metastasis might also be relatively overexpressed in a subset of clinically-aggressive primary CRC. After identifying upregulated CLM-signature genes, we used tissue microarrays (TMAs) to study the protein expression of selected signature genes in an independent cohort of primary CRC, to correlate their expression with clinical significance and outcome.

## Results

### Identification of gene signatures distinguishing CLM from primary CRC by expression profiling

To survey the differentially expressed genes between CLM and primary CRC (also compared to normal liver, a potential tissue contaminant of CLM), we used cDNA microarrays containing ~19,500 unique genes to profile the gene expression in 31 primary CRC specimens from 30 patients, and 32 unmatched CLM specimens from 31 patients who underwent liver resection. We then performed supervised analysis using SAM (with a Student's t-test metric) and identified 1,186 discriminatory cDNAs (corresponding to 963 unique genes) with significantly higher expression in CLM when compared to primary CRC, and to normal liver (previously profiled,[12]), with a false discovery rate (FDR) of <0.005% (Table S2). The top 35 differentially expressed genes are shown in Fig. 1. The 20 highest-ranking genes were *SPP1*, *CXCR4*, *GPNMB*, *LOX*, *CD53*, *AIF1*, *ARHGDI*, *SLC12A2*, *PRG1*, *SPARC*, *CD3D*, *DZIF1*, *PEG3*, *FYB*, *ITM2A*,

*SLA*, *IGLC2*, *MGP*, *LEF1*, and *MAF*. Similar results were obtained using SAM with a non-parametric, Wilcoxin rank-based analysis (Table S3).

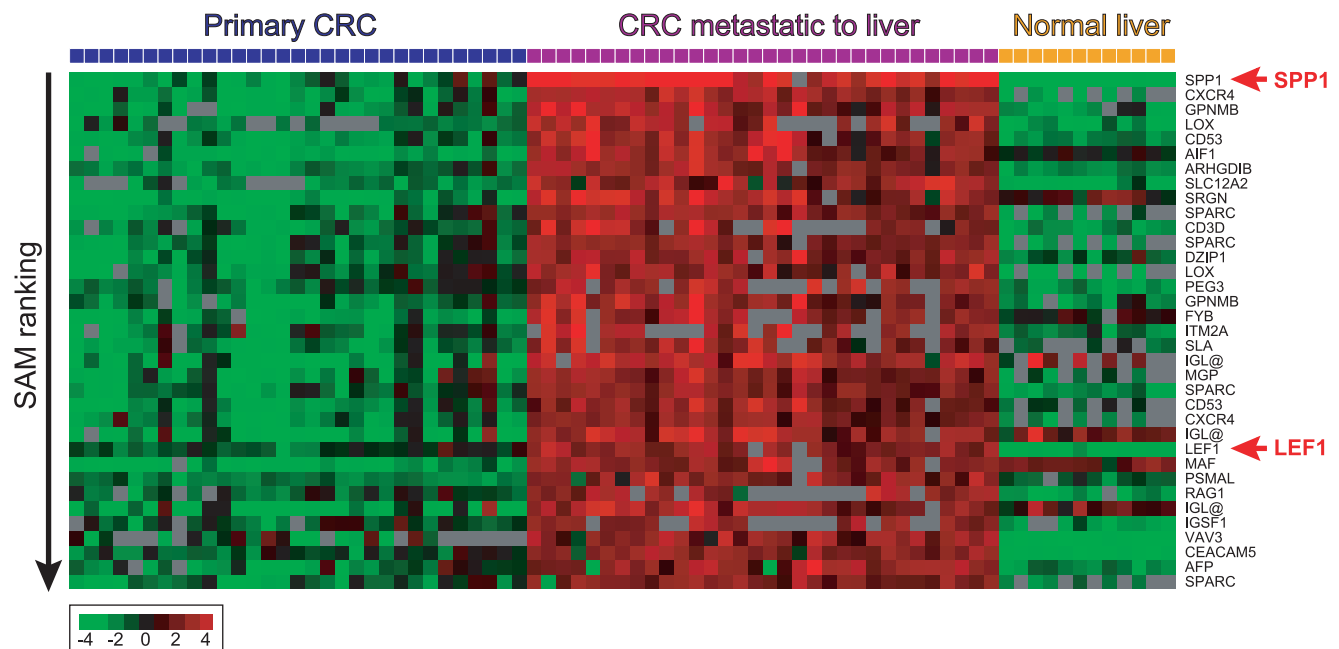
### LEF1 protein overexpression in CRC correlates with CLM and overall survival

Primary CRC specimens with relative increased expression of CLM signature genes might exhibit increased metastatic potential. To further study the significance of such potential biomarkers identified *via* SAM analysis, we examined the protein expression of two highly ranked and biologically plausible signature genes (for which IHC-validated antibodies were also available), *SPP1* (linked to metastasis[13]) and *LEF1* (lymphoid enhancer factor-1; involved in WNT signaling[14]), in an independent TMA cohort of CRC specimens. Examples of antibody staining by intensity are shown in Fig. 2. Correlation of *SPP1* or *LEF1* overexpression (IHC scored +2 or +3) with subsequently occurring CLM is shown in Table 1. *LEF1* overexpression was significantly correlated with CLM (Fisher's exact test,  $p=0.042$ ). No significant correlation was observed between *SPP1* overexpression and subsequent CLM.

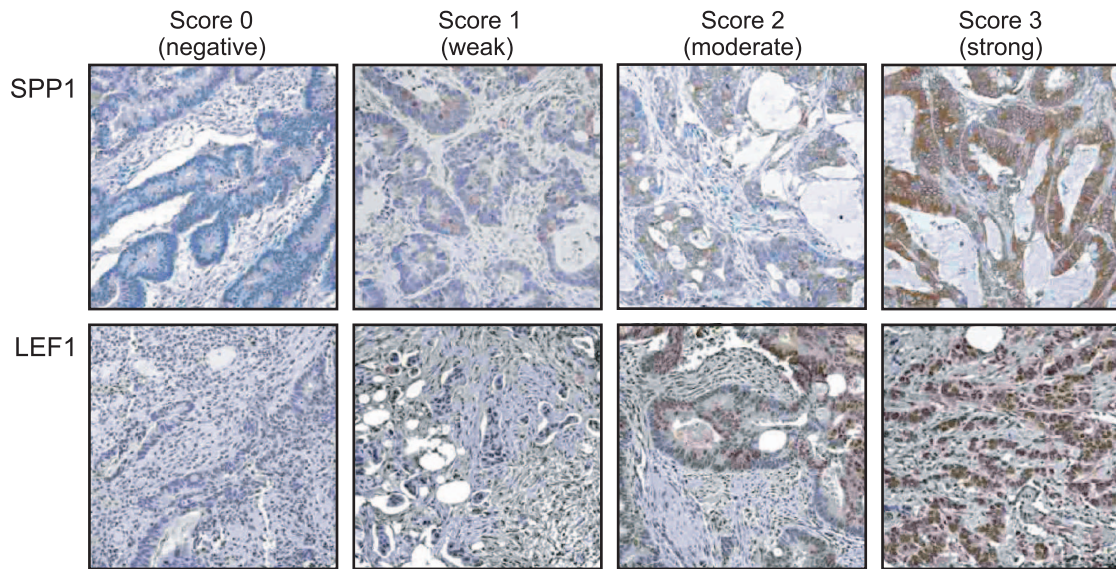
We used Kaplan-Meier analyses to investigate the impact of *SPP1* or *LEF1* overexpression on overall survival. *LEF1* overexpression was significantly associated with worse survival (log-rank  $p<0.01$ ; Fig. 3B). In contrast, *SPP1* overexpression was not significantly correlated with a worse outcome, though there was a trend toward poor survival (log-rank  $p=0.11$ ; Fig. 3A).

### LEF1 is a significant prognostic factor in multivariate analysis

As expected, higher stage was found to be significantly associated with worse overall survival. In the univariate model, the hazard ratios (HRs) for stage 2, 3, and 4 vs. 1 was 1.63, 3.47



**Figure 1. Heatmap showing the top 35 ranking genes, based on SAM analysis, with greatest increased expression in CRC with CLM compared to primary CRC or normal liver.** Rows represent individual genes and columns represent individual tissue samples. In each tissue sample, the  $\log_2$  ratio of abundance of transcripts of each gene relative to its mean abundance across all tissue samples is depicted according to the color score shown at the bottom. Grey indicates missing or excluded data. doi:10.1371/journal.pone.0016636.g001



**Figure 2. Immunohistochemical staining and assessment of SPP1 or LEF1 in representative core sections from the CRC tissue microarray.** Staining of SPP1 and LEF1 were respectively visualized in the cytoplasm and nuclei of cancer cells (X 400 magnification). doi:10.1371/journal.pone.0016636.g002

( $p=0.02$ ) and 12.63 ( $p<0.001$ ), respectively; in the multivariate model, the HRs are 1.74, 3.74 ( $p=0.01$ ), and 13.74 ( $p<0.001$ ), respectively. The HRs were also statistically significant for LEF1 overexpression in both the univariate and multivariate models - 1.78 ( $p<0.05$ ) and 1.66 ( $p<0.05$ ), respectively. Neither SPP1 overexpression nor tumor grade was significantly associated with survival in either analysis. Table 2 summarizes the HRs and 95% confidence intervals for the variables in both models.

## Discussion

In our study, we sought to identify signatures of metastasis embedded in a subset of primary tumors, which might predict clinically-aggressive behavior [15]. Using supervised SAM analysis, we identified 963 unique genes that are significantly overexpressed in CLM vs. primary CRC (and potentially contaminating normal liver tissue). In an independent set of tissue microarrays, we examined two highly-ranked genes (LEF1 and SPP1) as surrogate biomarkers for the CLM signature, and demonstrated that overexpression of LEF1, but not SPP1, in the primary CRC tissues correlates with a statistically significant increased risk of CLM, albeit its sensitivity and specificity in predicting liver metastasis were modest. In addition, independent

of tumor stage, overexpression of LEF1, not SPP1, denotes a poor prognosis for survival.

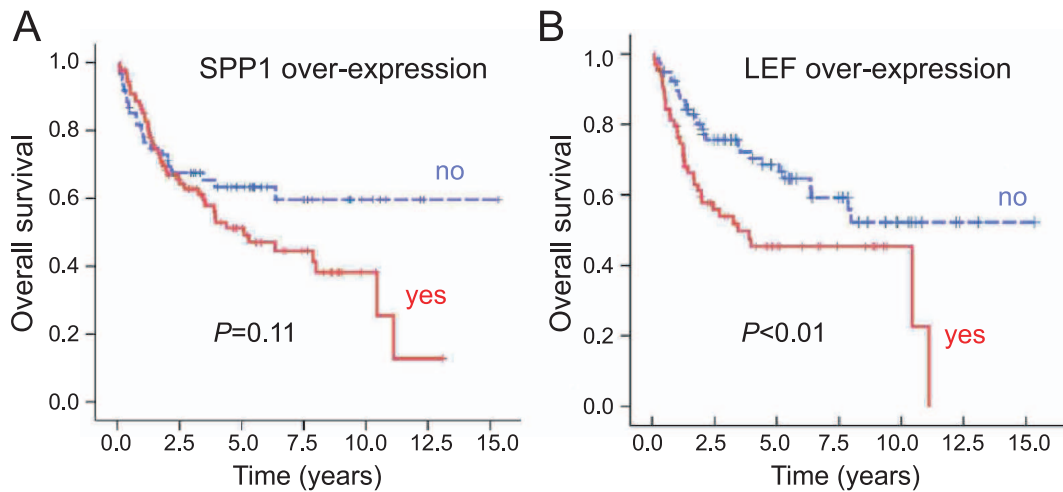
LEF1 was initially identified as a pre-B and T-lymphoid-specific gene encoding a DNA-binding protein of high mobility group (HMG) proteins.[16,17] It is a member of the T-cell factor/lymphoid-enhancing factor (TCF/LEF) family of transcription factors, which acts through the Wnt signaling pathway[18,19,20] to regulate gene expression and coordinate many cellular processes in normal development and tissue homeostasis, and, when deregulated, in colonic tumorigenesis and metastasis. Upon Wnt stimulation, LEF1 or other TCF/LEF-family transcription factors associate with  $\beta$ -catenin, a key cytoplasmic/nuclear mediator of Wnt pathway, and activate Wnt-responsive target genes. In contrast, without Wnt stimulation, glycogen synthase kinase (GSK)-3 (in a complex with APC) constitutively phosphorylates  $\beta$ -catenin, resulting in its proteasome-dependent degradation.[21] Although genetic and epigenetic changes have been documented in several targets throughout the pathway, mutation in either APC or  $\beta$ -catenin appears to be a crucial element in CRC carcinogenesis.[22] The LEF1 gene itself is not normally expressed in the adult intestinal epithelium, but only observed in the embryos while development is in progress. However, its overexpression has been well documented in CRC tumorigenesis[23], and denotes aberrant activation of the

**Table 1. Correlation between liver metastasis and SPP1 or LEF1 overexpression<sup>a</sup>.**

Liver metastasis	SPP1 Overexpression (No. of patients)			LEF1 Overexpression (No. of patients)		
	Yes	No	<i>p</i>	Yes	No	<i>p</i>
Yes	24	10		16	10	
No	66	51		41	63	
	0.14			0.042		

<sup>a</sup>As a predictor for liver metastasis, the sensitivity, specificity, positive predictive value, and negative predictive value for overexpression of SPP1 in the primary colon cancer tissue were 26.7%, 83.6%, 70.6%, and 70.6%, respectively. In contrast, the sensitivity, specificity, positive predictive value, and negative predictive value for overexpression of LEF1 are 28.1%, 86.3%, 61.5%, and 60.6%, respectively. Despite that the sensitivity and specificity were higher in LEF1, both showed modest performance.

doi:10.1371/journal.pone.0016636.t001



**Figure 3. Kaplan-Meier overall survival curves based on (A) SPP1 overexpression (log rank  $p=0.11$ ) or (B) LEF1 overexpression (log rank  $p<0.01$ ).**

doi:10.1371/journal.pone.0016636.g003

Wnt/ $\beta$ -catenin pathway *via* stabilization of LEF1/ $\beta$ -catenin complex. Indeed, LEF1 is a direct transcriptional target of the LEF1/ $\beta$ -catenin complex[23], indicating a positive feedback loop

for Wnt signaling, and suggesting LEF1 might be useful as a prognostic biomarker of Wnt pathway activation. Our findings are consistent with a prior study showing nuclear (active)  $\beta$ -catenin staining to be prognostic in colon cancer,[24] and highlight a role of Wnt signaling in colon cancer progression and liver metastasis.

As the prefix "osteo" suggests, osteopontin (or OPN, also known as secreted phosphoprotein 1 [SPP1], bone sialoprotein I, early T-lymphocyte activation 1) was initially recognized as an important glycosylated, adhesive phosphoprotein in bone.[25] Since then, several lines of evidence have shown its role in controlling tumorigenicity, progression and metastasis *via* its diverse ability as a cell-matrix mediator to interact with a variety of factors such as cell surface receptors (integrins and CD44), secreted proteases (matrix metalloproteinases and urokinase plasminogen activator), and growth factor/receptor pathways (TGF/EGFR and HGF/Met).[26] Overexpression of SPP1 has been reported in several human cancers, including lung, breast and colon cancers. In a gene-expression profiling study, Agrawal *et al* identified SPP1 as a lead marker correlating with CRC progression, and strongly expressed in CLM.[27,28,29] Most recently, Rohde *et al* observed that overexpression of SPP1 is indicative of poor survival in CRC and is significantly correlated with CLM.[30] In addition, overexpression of SPP1 correlates with increased immunohistochemical staining of  $\beta$ -catenin and, in an *in vivo* model, with Wnt activating mutations. These data suggest a crucial role of SPP1 in CRC progression and metastasis likely *via* molecular cross-talk with the Wnt pathway. Our results partially corroborate with published data in that we observe a trend towards poorer survival and CLM with SPP1 overexpression, though it is not statistically significant. This difference may be due to a smaller sample size and/or shorter follow-up time. Nonetheless, our data suggest that the overexpression of LEF1 is a stronger prognostic factor than SPP1 in correlating overall survival and CLM.

Though our analysis focused primarily on LEF1 and SPP1, other highly-ranked signature genes with increased expression in CLM compared to primary CRC also have biological functions consistent with roles in tumor progression, and might have prognostic utility. For example, CXCR4 (Chemokine (C-X-C motif) receptor 4) has been implicated in breast [31] and colon cancer metastasis.[32] LOX (lysyl oxidase) was shown to be associated with hypoxia where it functions in metastasis and predicts poor outcome in breast

**Table 2. Univariate and Multivariate Hazard Ratios From Cox Proportional Hazards Regression Models.**

Variable	P	Hazard Ratio	95% Confidence interval
Univariate analysis			
LEF1	0.03	1.66	1.04–2.63
SPP1	0.37	1.25	0.77–2.02
Stage			
1*			
2	0.36	1.63	0.57–4.63
3	0.02	3.47	1.27–9.50
4	<.0001	12.63	4.86–32.84
Tumor grade			
1*			
2	0.60	0.69	0.17–2.83
3	0.96	1.04	0.24–4.52
Multivariate analysis			
LEF1	0.02	1.78	1.09–2.89
SPP1	0.54	0.85	0.51–1.43
Stage			
1*			
2	0.31	1.74	0.60–5.02
3	0.01	3.74	1.35–10.39
4	<.0001	13.74	5.19–36.36
Tumor grade			
1*			
2	0.48	0.59	0.14–2.54
3	0.82	0.84	0.19–3.76

\*Reference category.

doi:10.1371/journal.pone.0016636.t002

cancer.[33] SPARC (secreted protein, acidic, cysteine-rich; osteonectin) was identified among genes that mark and mediate breast cancer metastasis.[34] Further studies are needed to characterize these and other signature genes in CRC progression.

Much of the published data comparing gene expression profiles from primary CRC vs. CLM have reported the differences in up- or down-regulated genes.[35,36,37,38,39,40,41,42,43,44] To the best of our knowledge, our data are the first to identify the expression of LEF1 as a predictor of overall survival, and an indicator of CLM. Comparisons of primary CRC vs. CLM, or of primary CRC associated with or without CLM, have identified gene signatures with relevance to colorectal cancer progression.[27,30,35,36,37,38,39,40,41,43] However, there are minimal overlaps between our CLM signature genes and the top-ranking CLM/progression-associated genes reported in these studies (Table 3). Several issues may explain the discordance among studies, such as: variability in patient cohorts, technical differences (the composition of the microarrays used, study design, statistical methodologies), and variability in the use of independent cohorts of patients to validate candidate prognostic genes.[45]

In conclusion, our study shows that overexpression of LEF1 in primary CRC correlates with a higher risk of CLM and denotes poor overall survival. It is a stronger predictor than SPP1, a marker reported in previous transcriptome studies. High-throughput gene expression profiling technology has revealed new insights into the molecular heterogeneity of CRC and identified new and better molecular markers for risk stratification. This holds promise for personalized medicine and improved targeted therapy. To achieve these goals, further studies are needed to understand the functional roles and clinical implications of LEF1, SPP1 and other signature genes for CLM.

## Materials and Methods

### Patients and Tumor Specimens

Freshly frozen CRC specimens (from Santa Clara Valley Medical Center) and CLM specimens (from Stanford University Medical Center) were used for gene expression profiling analysis. We used 31 primary CRC specimens from 30 patients (14 males and 16 females; age range: 36 to 80; stage I/II/III/IV = 1/10/11/8, diagnosed between 2000 and 2004), and 32 CLM specimens from 31 patients (16 males and 15 females; age range: 40 to 82). Tissue microarrays were constructed with paraffin blocks from 154 CRC cases (84 males and 70 females; age range: 27 to 92; stage I/II/III/IV = 26/46/38/44) with median follow-up of 3.0 years (range: 0.6–15.3 years). Study protocols were approved by Institutional Review Board (IRB) both at Stanford University Medical Center (IRB # 12473) and Santa Clara Valley Medical Center, a Stanford-affiliated teaching hospital, (IRB #07/28/00-03 and #2/22/2002-04). Individual informed consent was obtained from all participants involved in the gene expression profiling protocols and was waived in the TMA study by IRB due to the use of coded data and retrospective nature of the study. Clinical characteristics of the patient cohorts are summarized in Table S1.

### Gene expression profiling

To confirm that the sample was representative of the case, a frozen section from each specimen was first prepared and examined. Tissue was then homogenized in Trizol reagent (Invitrogen, Carlsbad, CA), and total RNA isolated per the manufacturer's protocol. RNA quality was assessed by gel electrophoresis. Gene expression profiling was performed as described previously.[46] Briefly, using microarrays of comple-

**Table 3.** Overlap of CLM genes with top-ranking CLM/progression genes from other published studies.

Gene Symbol	Overlapping with Other Studies	Gene Name
AGR2	Ki et al.[35], Tackels-Horne et a.[40]	ANTERIOR GRADIENT 2 HOMOLOG (XENOPUS LAEVIS)
CD44*	Ki et al.[35], Lin et al. [38]; Takayama et al.[47],	CD44 ANTIGEN (INDIAN BLOOD GROUP)
CDC2	Ki et al.[35], Takahashi et al.[45]	CELL DIVISION CYCLE 2, G1 TO S AND G2 TO M
CDH17	Ki et al.[35], Tackels-Horne et a.[40]	CADHERIN 17, LI CADHERIN (LIVER-INTESTINE)
CEACAM5	Ki et al.[35], Tackels-Horne et a.[40]	CARCINOEMBRYONIC ANTIGEN-RELATED CELL ADHESION MOLECULE 5
CEACAM6	Ki et al.[35], Tackels-Horne et a.[40]	CARCINOEMBRYONIC ANTIGEN-RELATED CELL ADHESION MOLECULE 6 (NON-SPECIFIC CROSS REACTING ANTIGEN)
CKS2	Li et al.[37], Lin et al. [38]	CDC28 PROTEIN KINASE REGULATORY SUBUNIT 2
EFEMP1	Ki et al.[35]	EGF-CONTAINING FIBULIN-LIKE EXTRACELLULAR MATRIX PROTEIN 1
HNRPA1	Ki et al.[35], Li et al.[37]	HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN A1
MAD2L1	Ki et al.[35], Li et al.[37]	MAD2 MITOTIC ARREST DEFICIENT-LIKE 1 (YEAST)
MMP2	Ki et al.[35], Takayama et al.[43];	MATRIX METALLOPEPTIDASE 2 (GELATINASE A, 72KDA GELATINASE, 72KDA TYPE IV COLLAGENASE)
S100P	Ki et al.[35], Li et al.[37]	S100 CALCIUM BINDING PROTEIN P
SPP1	Agrawal et al.[27], Rohde et al.[30 87]	SECRETED PHOSPHOPROTEIN 1 (OSTEOPONTIN, BONE SIALOPROTEIN I, EARLY T-LYMPHOCYTE ACTIVATION 1)
TIMP1*	Ki et al.[35], Takayama et al.[43]; Takahashi et al.[41]	TIMP METALLOPEPTIDASE INHIBITOR 1
TOP2A	Ki et al.[35], Takahashi et al.[41]	TOPOISOMERASE (DNA) II ALPHA 170KDA
VAV3	Ki et al.[35], Kleivi et al.[36]	VAV 3 ONCOGENE

Gene names are according to the DAVID Bioinformatics Database.

<http://david.abcc.ncifcrf.gov/tools.jsp>.

\*indicates genes present in three studies.

doi:10.1371/journal.pone.0016636.t003

mentary DNA (cDNA) (manufactured by the Stanford Functional Genomics Facility) containing ~40,000 nonredundant cDNA clones, representing ~19,500 unique UniGene clusters (i.e., genes), we hybridized Cy5-labeled total RNA from the tumor specimens, along with Cy3-labeled universal reference mRNA (pooled from 11 different cancer cell lines). We imaged arrays using an Axon GenePix 4000B scanner (Molecular Devices, Sunnyvale, CA), extracted fluorescence ratios (ratio of the specimen value to the reference value) using the GenePix software, and entered the data into the Stanford Microarray Database [47] for subsequent analysis. The microarray data are accessible from the Gene Expression Omnibus (Accession GSE22834). cDNA microarray expression data from normal liver specimens were previously published. [12]

### TMA construction and immunohistochemistry

A tissue arrayer (Beecher Instruments, Sun Prairie, WI) was used to construct a primary CRC tissue microarray as described, [48] comprising of 154 primary colorectal tumors each represented by two 6-mm cores. A 4- $\mu$ m section was cut from the tissue microarray block, deparaffinized in Citrisolv (Fisher Scientific, Hampton, NH) and hydrated in a graded series of alcohol solutions.

For immunohistochemical staining (IHC), anti-osteopontin (SSP1) mouse monoclonal antibody (Novocatra, Newcastle, UK) and anti-LEF1 (lymphoid enhancer factor-1) rabbit polyclonal antibody (Abcam, Cambridge, MA) were used at 1:100 and 1:500 dilutions respectively, and incubated overnight at 4°C. Chromogenic detection was then done using a peroxidase-conjugated secondary antibody and DAB reagents provided with the Envision detection kit (DAKO, Carpinteria, CA). SSP1 expression was scored as positive if distinct cytoplasmic staining was present in more than 10% of tumor cells. LEF1 expression was scored as positive when distinct nuclear staining was present in more than 10% of tumor cells. Weak cytoplasmic staining of LEF1 was not counted. The staining intensity was graded on a semiquantitative score (0, negative; 1+, weak; 2+, moderate; and 3+, strong). Survival analysis was performed in two groups depending on the score: overexpression (2+ and 3+) vs. the remainder (0 or 1+). Immunostains were scored by two pathologists (M.vdR. and YLC) blinded to the clinical data.

### References

- Markowitz SD, Bertagnoli MM (2009) Molecular Basis of Colorectal Cancer. *N Engl J Med* 361: 2449–2460.
- Jemal A, Ward E, Thun M (2010) Declining Death Rates Reflect Progress against Cancer. *PLoS One* 5: e9584.
- Jemal A, Siegel R, Ward E, Hao Y, Xu J, et al. (2009) Cancer Statistics, 2009. *CA Cancer J Clin* 59: 225–249.
- Greenlee RT, Hill-Harmon MB, Murray T, Thun M (2001) Cancer statistics, 2001. *CA Cancer J Clin* 51: 15–36.
- Fong Y, Fortner J, Sun RL, Brennan MF, Blumgart LH (1999) Clinical score for predicting recurrence after hepatic resection for metastatic colorectal cancer: analysis of 1001 consecutive cases. *Ann Surg* 230: 309–318.
- Jarnagin WR, Conlon K, Bodniewicz J, Dougherty E, DeMatteo RP, et al. (2001) A clinical scoring system predicts the yield of diagnostic laparoscopy in patients with potentially resectable hepatic colorectal metastases. *Cancer* 91: 1121–1128.
- Pollack JR (2007) A Perspective on DNA Microarrays in Pathology Research and Practice. *Am J Pathol* 171: 375–385.
- Notterman DA, Alon U, Sierk AJ, Levine AJ (2001) Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. *Cancer Res* 61: 3124–3130.
- Friedrichs J, Rosenberg R, Mages J, Janssen K-P, Macekl C, et al. (2005) Gene expression profiles of different clinical stages of colorectal carcinoma: toward a molecular genetic understanding of tumor progression. *Int J Colorectal Dis* 20: 391–402.
- Inokuchi M, Uetake H, Shiota Y, Yamada H, Tajima M, et al. (2004) Gene expression of 5-fluorouracil metabolic enzymes in primary colorectal cancer and corresponding liver metastasis. *Cancer Chemother Pharmacol* 53: 391–396.
- Matsuyama R, Togo S, Shimizu D, Momiyama N, Ishikawa T, et al. (2006) Predicting 5-fluorouracil chemosensitivity of liver metastases from colorectal cancer using primary tumor specimens: three-gene expression model predicts clinical response. *Int J Cancer* 119: 406–413.
- Chen X, Cheung ST, So S, Fan ST, Barry C, et al. (2002) Gene expression patterns in human liver cancers. *Mol Biol Cell* 13: 1929–1939.
- Rangaswami H, Bulbule A, Kundu GC (2006) Osteopontin: role in cell signaling and cancer progression. *Trends in Cell Biology* 16: 79–87.
- Li TW-H, Ting J-HT, Yokoyama NN, Bernstein A, van de Wetering M, et al. (2006) Wnt Activation and Alternative Promoter Repression of LEF1 in Colon Cancer. *Mol Cell Biol* 26: 5284–5299.
- Ramaswamy S, Ross KN, Lander ES, Golub TR (2003) A molecular signature of metastasis in primary solid tumors. *Nature Genetics* Nature Publishing Group. pp 49.
- Travis A, Amsterdam A, Belanger C, Grosschedl R (1991) LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function [corrected]. *Genes Dev* 5: 880–894.
- Waterman ML, Fischer WH, Jones KA (1991) A thymus-specific member of the HMG protein family regulates the human T cell receptor C alpha enhancer. *Genes Dev* 5: 656–669.

### Statistical Analysis

For cDNA microarray data, ratios were globally normalized by array and median-centered by gene. We included for analysis the 4,824 cDNAs (corresponding to 3,413 unique genes) that were well-measured (intensity/background >2 in either the test or reference channel) in at least 50% of samples, and variably expressed (>4-fold change from the median) in at least 3 samples. Two-class Significance Analysis of Microarrays (SAM) [49] was used to identify genes that were differentially expressed in CLM compared to primary CRC and normal liver (a potential contaminant of CLM), with statistical significance assessed by a false discovery rate (FDR).

For clinicopathological data, a chi-square test or Fisher's exact test (two-tailed) was used to compare differences in categorical variables across patient groups. Kaplan-Meier methods were used to estimate overall survival. We used both univariate and multivariate Cox proportional hazard regression model to assess the prognostic independence of variables (LEF1, SPP1, stage and tumor grade) for survival. Statistical analyses were performed with the SAS System software, release 9.1.3 (SAS Institute Inc., Cary, NC).

### Supporting Information

**Table S1** Summary of clinical characteristics of CRC cohorts. (XLS)

**Table S2** Ranked list of significant (FDR<0.005%) SAM genes (Student's t-test) upregulated in CLM vs. primary CRC and normal liver. (XLS)

**Table S3** Ranked list of significant (FDR<0.005%) SAM genes (Wilcoxin rank) upregulated in CLM vs. primary CRC and normal liver. (XLS)

### Author Contributions

Conceived and designed the experiments: AYL M-SC SKS JRP. Performed the experiments: AYL M-SC Y-LC WY YHK MvdR JRP. Analyzed the data: AYL M-SC Y-LC KS MvdR JRP. Contributed reagents/materials/analysis tools: AYL M-SC Y-LC RA GAA KS MvdR SKS JRP. Wrote the paper: AYL M-SC Y-LC JRP.

18. Cadigan KM, Nusse R (1997) Wnt signaling: a common theme in animal development. *Genes Dev* 11: 3286–3305.
19. Bienz M, Clevers H (2000) Linking colorectal cancer to Wnt signaling. *Cell* 103: 311–320.
20. Nusse R (1999) WNT targets: repression and activation. *Trends in Genetics* 15: 1–3.
21. Fuerer C, Nusse R, Ten Berge D (2008) Wnt signalling in development and disease. Max Delbruck Center for Molecular Medicine meeting on Wnt signaling in Development and Disease. *EMBO Rep* 9: 134–138.
22. Schneikert J, Behrens J (2007) The canonical Wnt signalling pathway and its APC partner in colon cancer development. *Gut* 56: 417–425.
23. Hovanes K, Li TWH, Munguia JE, Truong T, Milovanovic T, et al. (2001)  $\beta$ -catenin-sensitive isoforms of lymphoid enhancer factor-1 are selectively expressed in colon cancer. *Nature Genetics* 28: 53.
24. Wong SC, Lo ES, Lee KC, Chan JK, Hsiao WL (2004) Prognostic and diagnostic significance of beta-catenin nuclear immunostaining in colorectal cancer. *Clin Cancer Res* 10: 1401–1408.
25. Prince C, Oosawa T, Butler W, Tomana M, Bhowan A, et al. (1987) Isolation, characterization, and biosynthesis of a phosphorylated glycoprotein from rat bone. *J Biol Chem* 262: 2900–2907.
26. Tuck AB, Chambers AF, Allan AL (2007) Osteopontin overexpression in breast cancer: Knowledge gained and possible implications for clinical management. *Journal of Cellular Biochemistry* 102: 859–868.
27. Agrawal D, Chen T, Irby R, Quackenbush J, Chambers AF, et al. (2002) Osteopontin identified as lead marker of colon cancer progression, using pooled sample expression profiling. *J Natl Cancer Inst* 94: 513–521.
28. Agrawal D, Chen T, Irby R, Quackenbush J, Chambers AF, et al. (2003) Osteopontin identified as colon cancer tumor progression marker. *C R Biol* 326: 1041–1043.
29. Yeatman TJ, Chambers AF (2003) Osteopontin and colon cancer progression. *Clin Exp Metastasis* 20: 85–90.
30. Rohde F, Rimkus C, Friederichs J, Rosenberg R, Marthen C, et al. (2007) Expression of osteopontin, a target gene of de-regulated Wnt signaling, predicts survival in colon cancer. *Int J Cancer* 121: 1717–1723.
31. Muller A, Homey B, Soto H, Ge N, Catron D, et al. (2001) Involvement of chemokine receptors in breast cancer metastasis. *Nature* 410: 50–56.
32. Zeelenberg IS, Ruuls-Van Stalle L, Roos E (2003) The chemokine receptor CXCR4 is required for outgrowth of colon carcinoma micrometastases. *Cancer Res* 63: 3833–3839.
33. Erler JT, Bennewith KL, Nicolau M, Dornhofer N, Kong C, et al. (2006) Lysyl oxidase is essential for hypoxia-induced metastasis. *Nature* 440: 1222–1226.
34. Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, et al. (2005) Genes that mediate breast cancer metastasis to lung. *Nature* 436: 518–524.
35. Ki DH, Jeung HC, Park CH, Kang SH, Lee GY, et al. (2007) Whole genome analysis for liver metastasis gene signatures in colorectal cancer. *Int J Cancer* 121: 2005–2012.
36. Kleivi K, Lind GE, Diep CB, Meling GI, Brandal LT, et al. (2007) Gene expression profiles of primary colorectal carcinomas, liver metastases, and carcinomatoses. *Mol Cancer* 6: 2.
37. Li M, Lin YM, Hasegawa S, Shimokawa T, Murata K, et al. (2004) Genes associated with liver metastasis of colon cancer, identified by genome-wide cDNA microarray. *Int J Oncol* 24: 305–312.
38. Lin HM, Chatterjee A, Lin YH, Anjomshoa A, Fukuzawa R, et al. (2007) Genome wide expression profiling identifies genes associated with colorectal liver metastasis. *Oncol Rep* 17: 1541–1549.
39. Nakagawa H, Liyanarachchi S, Davuluri RV, Auer H, Martin EW, Jr., et al. (2004) Role of cancer-associated stromal fibroblasts in metastatic colon cancer to the liver and their expression profiles. *Oncogene* 23: 7366–7377.
40. Tackels-Horne D, Goodman MD, Williams AJ, Wilson DJ, Eskandari T, et al. (2001) Identification of differentially expressed genes in hepatocellular carcinoma and metastatic liver tumors by oligonucleotide expression profiling. *Cancer* 92: 395–405.
41. Takahashi Y, Ishii Y, Nishida Y, Ikarashi M, Nagata T, et al. (2006) Detection of aberrations of ubiquitin-conjugating enzyme E2C gene (UBE2C) in advanced colon cancer with liver metastases by DNA microarray and two-color FISH. *Cancer Genet Cytogenet* 168: 30–35.
42. Takata O, Kawamura YJ, Konishi F, Sasaki J, Kai T, et al. (2006) cDNA Array Analysis for Prediction of Hepatic Metastasis of Colorectal Carcinoma. *Surgery Today* 36: 608–614.
43. Takayama T, Miyanishi K, Hayashi T, Sato Y, Niitsu Y (2006) Colorectal cancer: genetics of development and metastasis. *J Gastroenterol* 41: 185–192.
44. Yanagawa R, Furukawa Y, Tsunoda T, Kitahara O, Kameyama M, et al. (2001) Genome-wide screening of genes showing altered expression in liver metastases of human colorectal cancers by cDNA microarray. *Neoplasia* 3: 395.
45. Quackenbush J (2006) Microarray analysis and tumor classification. *N Engl J Med* 354: 2463–2472.
46. Lapointe J, Li C, Higgins JP, van de Rijn M, Bair E, et al. (2004) Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc Natl Acad Sci U S A* 101: 811–816.
47. Demeter J, Beauheim C, Gollub J, Hernandez-Boussard T, Jin H, et al. (2007) The Stanford Microarray Database: implementation of new analysis tools and open source release of software. *Nucleic Acids Res* 35: D766–770.
48. Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, et al. (1998) Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med* 4: 844–847.
49. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 98: 5116–5121.