

# Identification of Tumor Markers in Models of Human Colorectal Cancer Using a 19,200-Element Complementary DNA Microarray<sup>1</sup>

Priti Hegde,<sup>2</sup> Rong Qi, Renee Gaspard, Kristie Abernathy, Sonia Dharap, Julie Earle-Hughes, Cheryl Gay, Nnenna U. Nwokekeh, Tingan Chen, Alexander I. Saeed, Vasily Sharov, Norman H. Lee, Timothy J. Yeatman,<sup>3</sup> and John Quackenbush<sup>3</sup>

The Institute for Genomic Research, Rockville, Maryland 20850 [P. H., R. Q., R. G., K. A., S. D., J. E-H., C. G., N. U. N., A. I. S., V. S., N. H. L., J. Q.], and H. Lee Moffitt Cancer Center, Tampa, Florida 33612 [T. C., T. J. Y.]

## ABSTRACT

Metastasis represents a crucial transition in disease development and progression and has a profound impact on survival for a wide variety of cancers. Cell line models of metastasis have played an important role in developing our understanding of the metastatic process. We used a 19,200-element human cDNA microarray to profile transcription in three paired cell-line models of colorectal tumor metastasis. By correlating expression patterns across these cell lines, we have identified 176 genes that appear to be differentially expressed (greater than 2-fold) in all highly metastatic cell lines relative to their reference. An analysis of these genes reiterates much of our understanding of the metastatic process and suggests additional genes, many of previously uncharacterized function, that may be causatively involved in, or at least prognostic of, metastasis. Northern analysis of a limited number of these genes validates the observed pattern of expression and suggests that further investigation and functional characterization of the identified genes is warranted.

## INTRODUCTION

CRC<sup>4</sup> is the second most common cause of cancer-related deaths in the United States, the majority of which are secondary to liver metastasis. Whereas current treatments have made significant improvements in patient survival, cases with metastatic cancer frequently result in death (1). One of the challenges to effective therapy lies in understanding the complex biology of metastasis in human colon cancer, thus allowing for new therapeutic and genetic intervention. Over the past few decades, a number of *in vivo* and *in vitro* models have been developed to study various aspects of colon carcinomas, including tumor cell heterogeneity (2), drug resistance (3), and metastasis (4, 5). One of the more widely used models for *in vivo* evaluation of human colorectal liver metastasis is the KM12 cell line system (6, 7). The cell lines were established by intrasplenically or s.c. injecting poorly metastatic KM12C cells derived from a Dukes' B stage colon carcinoma into nude mice to isolate variant lines KM12L4A and KM12SM, respectively, both with high liver metastatic potential (7). The advantage of this model system is that the highly metastatic variants are genetically related to the poorly metastatic parental cell line KM12C and, for this reason, presumably share the expression of the majority of genes with the exception of those promoting metastasis. To identify potential markers for metastasis, we used cDNA microarrays to examine global gene expression patterns in

the KM12 cell lines. To further refine the metastatic gene set and to ensure that identified genes were not limited to one cell line, we compared the expression profiles to a separate, genetically matched set of poorly (SW480) and highly (SW620) metastatic CRC cell lines. SW480 was derived from a primary Dukes' C colon cancer, whereas SW620 was derived from a lymph node metastasis in the same patient (8).

cDNA microarrays were used to generate a molecular fingerprint of gene expression patterns with the goal of elucidating changes that may contribute to the metastatic process. Genome sequencing and EST projects have provided a wealth of data to study the biology of tumor progression. Researchers have access to more than 2,000,000 ESTs representing greater than 80% of the estimated human genes. Fewer than 10,000 human genes however, have functionally annotated entries in GenBank. To best use these resources, the microarray technique presents parallel expression analysis of thousands of these genes in a single experiment without prior knowledge of gene function (9, 10).

An array of 19,200 clones selected using EST assemblies that comprise the TIGR HGI,<sup>5</sup> was used to assess patterns of gene expression between low and high metastatic states of colon carcinomas. A comparison of transcriptional levels has identified 176 genes that appear to be differentially expressed (greater than 2-fold) in all highly metastatic cell lines relative to their reference. Included in the gene set are those previously known to possess altered patterns of expression in colon cancer as well as a large number with previously uncharacterized function. The identification of these additional genes is of tremendous importance because they provide potential new insights into cancer biology, they may allow a better understanding of pathways that regulate metastasis, and they yield gene expression patterns that may be diagnostic or prognostic of metastasis in colon cancer.

## MATERIALS AND METHODS

**Clone Set and Array Preparation.** cDNA clones were selected using EST sequences represented within the TIGR HGI.<sup>6</sup> HGI consists of THC sequences, derived from coding and EST sequences deposited in GenBank, using a highly refined process in which sequences are first clustered, then assembled (11, 12). Clones were grown overnight in 96-well microtiter plates; clone inserts were amplified using vector-specific primers and were validated and purified; aliquots of each reaction were combined 1:1 with DMSO for arraying as described previously (13). The success rate for single-band amplification was greater than 87.5%; 6.3% of the reactions yielded multiple or weak bands, and 6.2% failed to amplify. Wherever possible, alternate cDNA clones were selected for those failing amplification. All of the clones were sequenced to verify their identities, with 78% yielding usable sequence (15,028 of 19,200), of which 96% (14,379 of 15,028) allowed positive identification for an overall 75% clone validation success rate. Clone amplicons were printed onto CMT-GAPS amino-silane coated glass microscope slides (Corning) using a high-speed robotic arrayer (Intelligent Automation Systems) at 42–45% relative

Received 6/1/01; accepted 8/28/01.

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.

<sup>1</sup> Supported by grants from the National Cancer Institute [NCI CA85052-01A1 and NCI CA85429-01 (to T. J. Y.) and NCI CA77049-02 and NCI 6120-119-L0-A (to J. Q.)], from the American Cancer Society [ACS RPG-99-099-01-MGO (to T. J. Y.)], the National Institute of Neurological Disorders and Stroke [NINDS NS-35231 (to N. H. L.)], and the National Heart Lung and Blood Institute [NHLBI HL-59781 (to N. H. L.)].

<sup>2</sup> Present address: GlaxoSmithKline, King of Prussia, PA 19406.

<sup>3</sup> To whom requests for reprints should be addressed, at E-mail: yeatman@moffitt.usf.edu; or johnq@tigr.org.

<sup>4</sup> The abbreviations used are: CRC, colorectal cancer; EST, expressed sequence tag; HGI, Human Gene Index; TIGR, The Institute for Genomic Research; THC, tentative human consensus (sequence); IGF-I, insulin-like growth factor I.

<sup>5</sup> Internet address: <http://www.tigr.org/tdb/tgi.shtml>.

<sup>6</sup> Internet address: <http://www.tigr.org/tdb/hgi/hgi.html>.

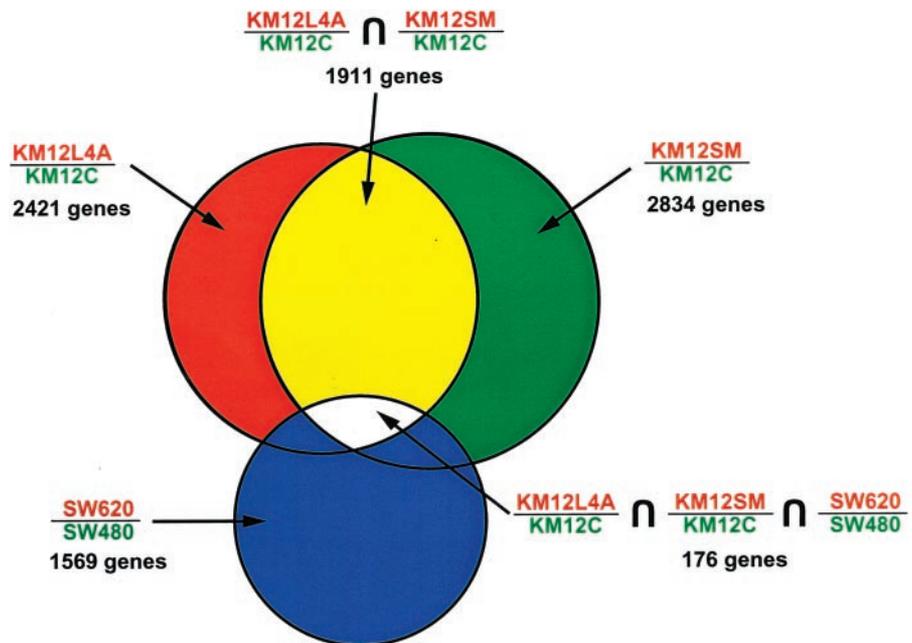


Fig. 1. For each paired potentially high- and low-metastatic cell lines, genes that were consistently differentially expressed across replicates were identified. Gene expression between KM12C (control) and metastatic KM12L4A and KM12SM lines shows that 2421 and 2834 genes, respectively, are differentially expressed. Of these, 1911 are found in common between these cell lines, which suggests possible metastasis genes. This list is significantly refined by comparison with the independently derived SW480 and SW620 cell lines, which share only 176 of its 1569 differentially expressed transcripts with the KM12 lines.

humidity. Printed slides were cross-linked by UV irradiation (60 mJ) and stored at 25°C in a desiccated environment until use.

**Tissue Culture and Probe Preparation.** Tumor cell lines were maintained in RPMI 1640 (Life Technologies, Inc.) modified medium containing L-glutamine and supplemented with 10% fetal bovine serum, 200 units/ml penicillin, and 200  $\mu$ g/ml streptomycin. Cells were maintained in 150-mm culture dishes (Corning) at 37°C in 5% CO<sub>2</sub>. Total RNA was extracted from the cells using Triazol (Life Technologies, Inc.) and was used to prepare direct Cy3- and Cy5-labeled first-strand cDNA probes for hybridization as described previously (13). Before hybridization, slides were incubated in 1% BSA to block nonspecific hybridization to the glass surface; differentially labeled pooled probes were hybridized to microarray slides overnight and washed. Expression was assayed by measuring fluorescence intensities using the Genepix 4000 (Axon) dual-color confocal laser scanner; data were recorded as paired 16-bit TIFF images.

**Image Processing and Data Analysis.** Images were analyzed to determine each spot's background-subtracted integrated fluorescence intensity in both the Cy3 and Cy5 channels using TIGR Spotfinder.<sup>7</sup> This software uses a dynamic thresholding algorithm to identify spots and calculate local background and intensities. For each array, expression ratios were normalized using an iterative mean log<sub>2</sub>(ratio) centering approach. Data from replica arrays were averaged and genes showing consistent differential expression across replicates were tallied for each pair of cell lines. For the KM12-derived lines, the intersection of the KM12SM and KM12L4A gene sets were identified. The intersection of those differentially expressed in the SW620/SW480 lines were also identified. Each of the clones representing differentially expressed genes on the arrays and selected for further evaluation were subjected to an additional round of sequence verification, with 100% validation of clone identity.

## RESULTS

cDNA microarrays containing 19,200 elements were constructed using clones selected by sequence analysis of ESTs comprising the TIGR HGI. In selecting clones, priority was given to known genes with mapping information; the remainder was selected to represent genes of unknown function. Included in this array were 19,159 distinct cDNA clones representing 14,167 separate tentative consensus sequences. This included 9,577 cDNAs corresponding to 6,328 genes of known function, 9,582 of unknown function and 6,467 with map-

ping information (14).<sup>8</sup> Clone inserts were amplified by PCR, purified to remove contaminants, and mechanically spotted at high density onto silane-coated microscope slides using a high-speed robotic system (13).

To identify genes with differential expression profiles in colon cancer metastasis, we used the arrays to compare genetically related colon carcinoma cell lines. For each set of experiments, mRNA was extracted from the parental poorly metastatic KM12C cells, reverse transcribed, and labeled with Cy 3-dUTP. These were used as controls. Gene expression in KM12C was compared with that in the highly metastatic KM12L4A and KM12SM, the RNA of which was reverse transcribed and labeled with Cy 5-dUTP. Similarly, transcript levels of SW480 were compared with SW620. Relative expression was assayed by a two-color hybridization. Four replicates were performed for each experiment. For any pair of cell lines, genes exhibiting a consistent 2-fold up- or down-regulation across three or more replicates were considered significant. Consensus sets from each of the KM12 cell lines were then compared to generate a KM12-specific gene set. Finally, this set was cross-referenced with a consensus set derived from the SW480/SW620 cell lines to produce a final list of candidate metastasis-associated genes.

Of the 19,200 elements analyzed by this method, 176 genes appear to have differential expression patterns between the cellular phenotypes of low and high liver metastasis in the KM12 and SW cells (Fig. 1). Of these, mRNA levels for 121 genes are up-regulated and 55 genes are down-regulated. Included are 108 genes of known function and 68 genes of unknown function. Analysis of each individual set of experiments showed 2421 and 2834 genes showing differential expression patterns in the KM12L4A and KM12SM cell lines respectively. Of these, 1911 genes were differentially expressed in both metastatic variants (Fig. 1). The SW metastatic variant exhibited 1569 genes with altered patterns of expression as compared with its reference (Fig. 1). Overall, most of the genes exhibited an average of ~2- to 4-fold change in relative expression. A small percentage of genes showed >5-fold change in expression levels. Relatively small incre-

<sup>8</sup> Summary information for the array can be found at [http://cancer.tigr.org/data/Hum19k\\_1](http://cancer.tigr.org/data/Hum19k_1).

<sup>7</sup> Internet address: <http://cancer.tigr.org/tools/>.

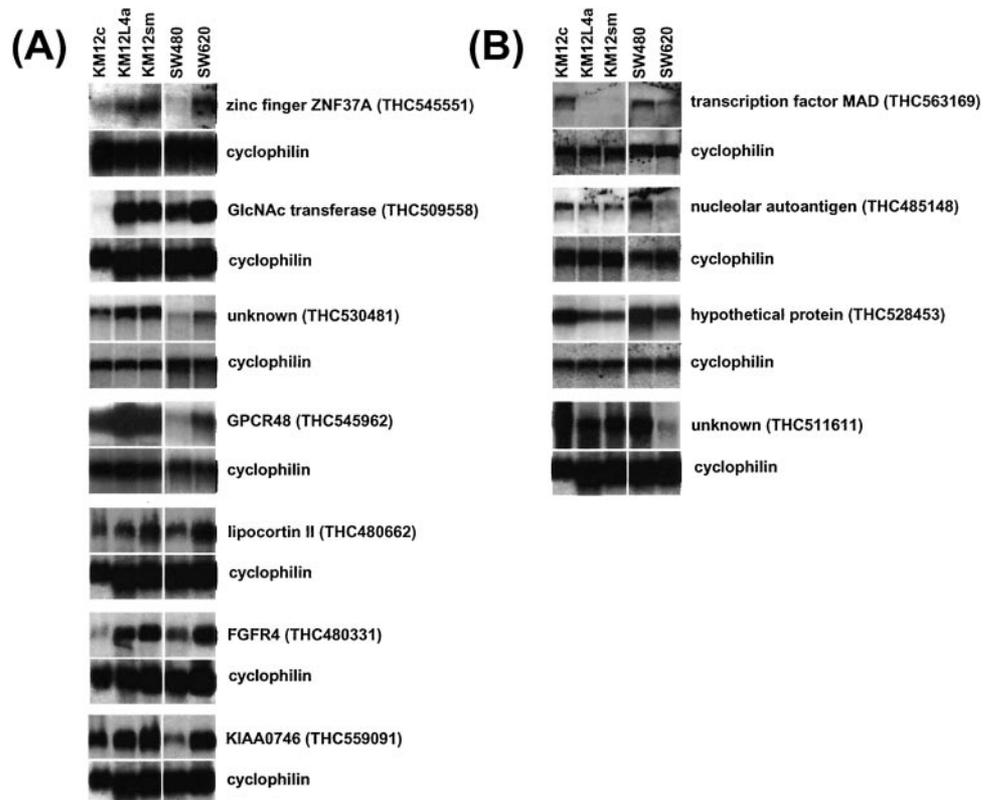


Fig. 2. Up- and down-regulated genes were selected from the 176-gene consensus set derived from a microarray analysis and confirmed by Northern analysis on the KM12 and SW cell lines using cyclophilin as a control. Both (IA) up- and (IB) down-regulated genes confirmed the expression patterns observed on the arrays.

mental changes in gene expression are not surprising because progression to metastasis is the result of the cumulative accumulation of a number of spontaneous molecular changes, some of which may be subtle (15). By comparison, larger-fold differences would be expected when comparing normal mucosa to paired cancerous tissues.<sup>9</sup>

Northern blot analysis was performed on a subset of genes to validate results from the microarrays (Fig. 2). A set of 11 clones was selected from those that were most significantly differentially expressed and that represented both up- and down-regulated genes not previously associated with colon cancer. These were sequence validated, and the inserts were amplified by PCR, purified, labeled, and used to probe blots containing RNA extracted from the cell lines. Cyclophilin was used as the control for all of the Northern blots because its expression was observed to be nearly constant across all microarray assays. Note that glyceraldehydephosphate dehydrogenase (GAPDH), which is commonly used as a control for Northern analyses, was differentially expressed in the highly metastatic cell lines. The zinc finger protein ZNF37A (THC545551), GlcNAc transferase (THC509558), novel G-protein coupled receptor GPCR48 (THC545962), lipocortin II (THC480662), FGFR4 (THC480331), and two genes of unknown function [THC530481 and THC559091 (KIAA0746)], were found to be induced in the highly metastatic cell lines, as had been determined using microarrays. The transcription factor MADS (THC563169), nucleolar autoantigen (THC485148) a hypotetic protein (THC528453), and THC511611 were down-regulated in the metastatic cell lines, similar to the array analysis. As has been reported previously (16, 17), whereas the absolute magnitude of the relative expression level determined by the Northern analysis differed slightly from that measured on the arrays, the direction of change in expression was consistent between the techniques. An additional four cDNAs were selected for confirmation by Northern

analysis; however, these failed to give a detectable signal after hybridization. On analysis of the microarray data, we found these genes to be expressed at very low levels, which suggested that alternative means of validation, such as quantitative reverse transcription-PCR, would be necessary for confirmation.

## DISCUSSION

Metastasis is broadly defined as the formation of secondary foci at a site that is distant from the primary site of origin. This process involves a series of interdependent, sequential events that include initial growth, angiogenesis, invasion, extravasation, and establishment of new growth at the secondary site. Gene expression profiling that uses high-density cDNA microarrays provides a revolutionary approach to studying metastasis. Transcriptional changes of genes associated with CRC cell line systems, and the potential role of these changes in various steps of tumor metastasis, are described below. Differentially expressed genes have been grouped by known or putative functions (Table 1).

**Cell Cycle and Tumorigenesis.** Progression of the cell cycle from its initial growth phase ( $G_1$ ) to its mitotic phase (M) is driven by positive and negative regulators that ultimately direct the fate of a cell either to form two daughter cells or to enter into resting state ( $G_0$ ). Dysregulated cellular proliferation, arising from abnormal expression of genes that control cell cycle checkpoints ( $G_1$ -S and  $G_2$ -M phases), plays a critical role in tumorigenesis. Calpastatin is an endogenous inhibitor of calpain, a cysteine protease implicated in apoptosis via inhibition of the cyclin-dependent kinase inhibitor p27 and cyclin D1 (18). We saw a 3.5-fold increase in expression of calpastatin, which seems to suggest that the gene may be involved in cellular proliferation by inhibiting calpain-induced apoptosis. This is the first report of the association of calpastatin with metastasis. Changes in expression levels of factors controlling transcriptional and translational machinery such as Nap1 also suggest an increased predisposition of these

<sup>9</sup> The list of genes, with average measured levels of expression can be found online at <http://cancer.tigr.org/data>.

Table 1 Representative differential expression in highly metastatic cell lines, defined by a 2-fold or greater change relative to expression in a less metastatic reference, for various functional classes of genes

Fold induction/repression in red and green represent genes that were overexpressed and underexpressed, respectively, in the highly metastatic cells.<sup>a</sup> Each experiment was performed three times. Results are reported for genes with two or more high-quality data points.

Gene Name	THC #	L4a SM SW Fold Induction				THC #			
<b>cell cycle and tumorigenesis</b>					<b>growth factors</b>				
calpastatin	THC395488	3.45	4.30	2.24	Insulin-like growth factor I	THC383910	3.50	3.94	-
Ki-67 antigen	THC349279	3.26	3.41	-	transforming growth factor beta (GB:M60315)	THC529873	2.48	6.20	-
cyclin	THC527078	2.20	2.06	-	transforming growth factor beta-binding protein, latent	THC506202	-4.55	-4.55	2.31
G0/G1 switch regulatory phosphoprotein GOS8	THC485927	-7.69	-14.29	-	bone-derived growth factor	THC486169	-4.00	-6.25	4.55
cell cycle, G1 phase defect	THC483910	-5.88	-8.33	-	fibroblast growth factor receptor 4	THC553705	-5.56	-2.94	-
cdc16	THC509159	-2.38	-2.00	-	insulin-like growth factor binding protein 1	THC529746	-	-	-6.67
apoptosis regulator bcl-x	THC512751	2.86	7.03	2.50	epidermal growth factor receptor	THC484646	2.17	2.21	-
Nck-associated protein 1 (Nap1)	THC512992	2.42	6.14	3.92	Transforming growth factor, alpha	THC498311	2.46	-	4.92
mitotic checkpoint kinase Bub1	THC489479	-2.27	-3.57	-2.63					
mitotic checkpoint protein kinase BUB1B	THC314674	-6.25	-6.67	-2.70	<b>glycolysis</b>				
sphingomyelinase	THC512598	-2.22	-2.70	-	Enolase, alpha	THC479319	4.67	11.94	2.45
14-3-3	THC512636	2.27	3.50	3.10	Pyruvate kinase, M1/M2 type	THC479467	2.55	3.93	2.89
Nmi	THC556245	2.48	5.26	-	lactate dehydrogenase B	THC479476	3.10	4.20	-
					glyceraldehyde -3-phosphate dehydrogenase	THC480123	3.15	4.20	2.03
<b>transcription factors</b>					glucosyltransferase	THC492734	-	3.92	-
transcription factor BTF3b	THC512531	2.59	2.44	-2.44	GlcNAc-1-P transferase	THC503927	1.70	3.84	-
transcription factor CTF-1, CAAT-binding	THC506787	-3.13	-3.70	-	glycogen phosphorylase, isoform 3	THC479658	2.77	3.72	-
transcription factor ETR101	THC554684	2.97	3.31	-	phosphoenolpyruvate carboxykinase	THC547596	2.08	3.34	-
transcription factor ITF-1	THC509691	-4.55	-5.26	-	phosphofructokinase, platelet	THC556562	2.14	3.08	-
transcription factor MADS/MEF2	THC563169	-2.17	-2.27	-3.03	glycogen phosphorylase, isoform 2	THC493599	-	2.95	-
transcription factor NFAT5 isoform b	THC480740	-2.44	-3.23	-	triosephosphate isomerase	THC492605	2.27	2.79	-
transcription factor NF-kappa-B, p65 subunit	THC529952	-7.69	-7.14	-	glycogen phosphorylase, liver	THC547566	2.16	2.70	-
transcription factor S-II	THC493180	2.70	2.76	-	2,3-bisphosphoglycerate mutase	THC485574	-5.88	-10.00	-
transcription factor Sp1	THC527282	-7.14	-4.55	-	phosphoglycerate mutase 2, muscle	THC523404	-5.88	-5.26	-
transcription factor USF2	THC558601	-10.00	-11.11	-	aryl hydrocarbon receptor nuclear translocator	THC553991	3.47	4.39	-
translation elongation factor 1, alpha	THC480091	2.20	2.48	2.05	Aldolase A	THC506082	2.15	2.31	2.77
translation elongation factor 1, beta	THC529810	-2.08	-3.03	-					
translation elongation factor 1, gamma	THC485540	2.17	3.48	6.85	<b>cell-matrix proteins</b>				
translation initiation factor 2 associated p67	THC483408	-12.50	-20.00	-	Epican	THC482446	-	5.82	4.87
Translation initiation factor 2A	THC481279	-3.85	-2.94	3.24	Cofilin	THC484922	4.18	6.53	-
translation initiation factor 4A	THC492573	3.83	5.66	-	Profilin I	THC485330	2.92	5.74	-
translation initiation factor 4AII	THC481077	4.00	4.31	2.27	Profilin II	THC482718	-	-	3.84
translation initiation factor 4B	THC556378	6.12	9.08	2.28	collagen, type XI, alpha 2	THC504818	3.21	3.15	-
translation initiation factor 5	THC547535	-2.63	-3.70	4.74	collagen, Type IX, alpha 1	THC530893	-	-	-3.03
					Vinculin	THC530068	2.71	2.99	-
<b>proto-oncogenes</b>					tubulin, beta d1	THC480951	4.90	6.36	2.53
proto-oncogene c-K-ras	THC481807	2.43	3.73	-	actin, beta	THC563165	10.31	16.05	3.12
proto-oncogene dek	THC506250	2.47	2.52	-	actin-capping protein alpha subunit isoform 1	THC481578	4.60	9.11	3.78
proto-oncogene ets-2	THC559139	1.69	2.84	2.35	Catenin, alpha 2(E)	THC406340	3.03	2.44	2.92
proto-oncogene junD	THC543521	2.35	3.67	-	ATP-dependent metalloprotease YME1L	THC484479	-	-	3.92
proto-oncogene N-ras	THC512835	2.05	2.64	4.21	Annexin IV	THC496004	-	-1.75	-4.00
proto-oncogene rhoA,	THC502870	3.07	5.32	2.28	Caveolin, 22kDa	THC495762	-7.69	-12.50	-3.33
proto-oncogene bcl-3	THC506333	2.84	2.92	-	dynein, light chain 1, cytoplasmic	THC523389	-2.04	-2.08	-3.03
<b>oncogenes</b>					<b>signal transduction</b>				
oncogene E6-AP, papillomavirus	THC482127	3.80	3.40	-	Guanine nucleotide-binding protein G(s), alpha subunit	THC506149	3.33	7.89	3.61
oncogene ret/PTC3	THC536717	3.60	5.09	-	Guanine nucleotide-binding protein ral, ras related	THC493091	-	4.33	2.15
oncogene set	THC526421	3.06	4.02	1.98	serine/threonine kinase	THC519506	2.92	3.22	-
Yamaguchi Sarcoma viral oncogene homolog 1	THC483974	2.90	2.70	3.73	protein phosphatase 2A, alpha catalytic subunit	THC543416	2.44	3.11	2.24
					guanylyl cyclase large subunit, soluble	THC531332	2.38	2.79	-
					Tyrosine kinase	THC544138	-	3.62	5.95
					ERK3	THC485382	2.54	3.16	-
					Tyrosine phosphatase PAC-1	THC484984	-12.50	-25.00	-
					cytosolic phospholipase A2 gamma	THC503323	-7.14	-9.09	-
					cytosolic phospholipase A2 beta; cPLA2beta	THC512549	-2.17	-2.86	-
					<b>markers</b>				
					mutL homolog hPMS2	THC484568	-10.00	-14.29	-
					mutL homolog hPMS1	THC502746	-2.50	-3.85	-
					<b>hepatic carcinomas</b>				
					phosphatidylethanolamine N-methyltransferase	THC523125	-4.55	-4.76	-
					hepatoma-derived growth factor	THC502980	2.73	2.76	-
					oncogene tpr/met, translocated	THC559059	-7.69	-8.33	-
					Hepatocellular carcinoma associated protein	THC481590	9.43	50.38	6.49

<sup>a</sup> -, genes that could not be called as significantly differentially expressed with high confidence.

cells to transformation. Moreover, the nuclear proliferation marker Ki-67 antigen is also overexpressed, which provides further corroboration of cellular proliferation in these cell lines. Cell cycle pathways are critical for the initial steps of tumorigenesis. Our analysis of gene expression changes in late stages of cancer progression continues to show the involvement of cell cycle events in metastasis. SW620 cells also have a higher bromodeoxyuridine labeling index than the SW480 cells (8), which suggests that cell cycle control may be an important factor even in later stages of tumor progression.

Genetic instability is a characteristic feature of CRCs (19). Genes involved in DNA mismatch repair have long been associated with hereditary nonpolyposis CRCs (20). A decrease in expression of the mutL homologues, hPMS1 and hPMS2 suggests that these cell lines exhibit a high level of microsatellite instability that is attributable to a lack of replication error repair activity. It is possible, in fact, that these cell lines were derived from patients with hereditary nonpolyposis CRC, or that this instability is a product of *in vitro* passage. Changes in expression of genes, such as Bub1 and Bub1B, that are involved in chromosomal instability were seen in the KM-derived and SW cell lines, respectively.

**Angiogenesis.** Hypoxia is a common occurrence in tumors and arises from a lack of blood supply to rapidly proliferating cells (21). Once the tumor mass reaches a diameter of ~2 mm, establishment of new vascular system is essential for its survival. Until then, endurance in hypoxic conditions is an important factor for tumor progression. Transformation studies have earlier shown the regulation of individual genes in tumor progression (21). Here we have comprehensively shown the regulation of glycolytic enzymes during metastasis. Enolase ( $\alpha$ ) showed the most dramatic increase in expression, (~6-fold) in all three highly metastatic cell lines and could potentially be used as a marker for metastasis in colon carcinomas. There was a general up-regulation of genes involved in glycolysis in the KM12-derived cell lines, which suggested the need for the glycolytic pathway as an alternate energy source for cell survival during liver metastasis.

The KM12 and SW cell lines showed varied patterns of growth factor expression. High levels of IGF-I have been associated with increased risk of cancer (22). Whereas the KM12 metastatic cell lines showed elevated levels of IGF-I to the order of 3.7-fold, the SW cell lines did not show any significant change in IGF-I expression. Similarly, the growth-enhancing properties of transforming growth factor  $\beta$  during angiogenesis in colon cancers have been well documented. Here too, the SW620 cells did not show any altered patterns of expression. Although IGF-I and transforming growth factor  $\beta$  may be used as prognostic markers in carcinomas (23, 24), their significance as diagnostic or prognostic markers for metastasis would be questionable.

Genes associated with hepatomas were also represented in the expression profile. The gene encoding for the hepatocellular carcinoma-associated protein is overexpressed ~10-fold in all three metastatic cell lines. Differential expression of hepatoma-derived growth factor and hepatocyte nuclear factor 3 $\beta$  are specific only to the liver-metastatic KM12 variants and not to the lymph node-metastatic SW cells, which indicates that these genes may be more specific to liver metastasis whereas the hepatocellular carcinoma-associated protein may be associated with general metastasis of colon carcinomas.

**Tumor Invasion.** Among all of the genes differentially expressed, those involved in cytoskeletal organization and the extracellular matrix formation such as actin- $\beta$ , tubulin  $\beta$ d1, and profilin II were most significant. For a tumor to become invasive, it must pass through the muscularis mucosa and infiltrate the subserosal layer in which terminal lymphatics reside. Subsequently, genes that are involved in breaking the barriers of cellular adhesion play an important role in tumor invasiveness. In general, there was no significant change in expression

of homotypic cell adhesion molecules such as CEACAM7 and the nonspecific cross-reacting antigen (NCA) molecules.

Gross abnormalities were also observed in expression of genes involved in the formation of cytoskeletal architecture and the extracellular matrix. The actin cytoskeleton is the basic machinery that makes cells motile, a characteristic property of invasive cells. Evidence for dynamic actin-based cytoskeletal motility in the metastatic cell lines comes from differential expression of genes that are involved in actin polymerization such as actin-capping proteins, which showed an ~6-fold increase in expression.

**Signal Transduction and Cancer Metastasis.** Cross-talk between different pathways makes intracellular signal transduction a challenging area for cancer research. The molecular components of signal transduction that lead to tumorigenicity are poorly understood. One gene that is consistently overexpressed across all three cell lines is a novel G-protein-coupled receptor, GPCR48 (Fig. 2). The function of this protein is unknown. However, its expression pattern in metastasis could make this a useful marker for colon cancer metastasis. In addition to the protein phosphatases and protein kinases, a large number of oncogenes such as the *set* oncogenes and the *v-Yes* oncogenes show increased expression. Proto-oncogene *c-k-ras*, a highly characterized indicator of carcinomas, showed increased expression in the KM12 cell lines. The SW480 cells contain a point mutation in the *ki-ras* gene that results in activated *ras* gene product (25). No significant change in expression of *c-k-ras* was detected on going from the low metastatic cells to their highly metastatic SW620 variant. Other proto-oncogenes such as *N-ras* and *rhoA* also exhibit elevated levels in all highly metastatic variants. Recently, overexpression of *RhoC* has been demonstrated to stimulate metastasis in melanoma cells (26). Our findings provide evidence that other members of the Rho family of small GTPases may contribute to the process of metastasis.

Apart from the known genes, 39% of the genes that are differentially expressed have no assigned functional role. A hypothetical protein (THC347434) is consistently underrepresented in the highly metastatic variants (Fig. 2). Sequence analysis of the gene using SMART protein prediction tools (27), revealed that it contains multiple zinc finger domains, which suggests it may encode a DNA-binding protein. Although further work will be required to fully characterize this and other genes, their patterns of expression may make them useful as markers of metastasis.

To demonstrate the utility of cell line analysis as a model for understanding clinical prognosis, we selected two previously uncharacterized transcripts, THC511611 and THC559091 (KIAA0746), which were shown to be down- and up-regulated, respectively. These were then used to measure relative expression in paired samples of

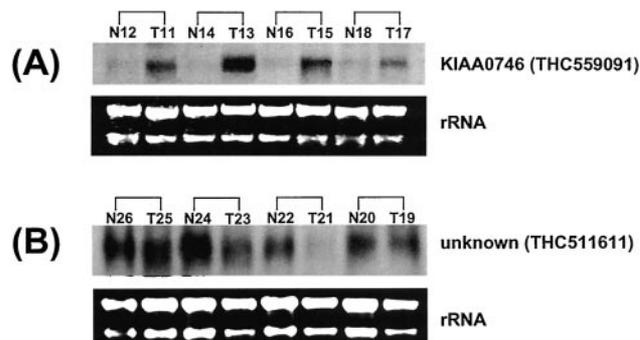


Fig. 3. Of those genes validated by Northern analysis using the cell lines, two genes of unknown function were selected for further investigation using matched CRC and normal tissue samples. A, KIAA0746 (THC559091), up-regulated in the cell lines, also exhibits increased expression in tumor. B, THC511611, down-regulated in the cell lines, also shows decreased expression in tumor relative to surrounding normal tissue.

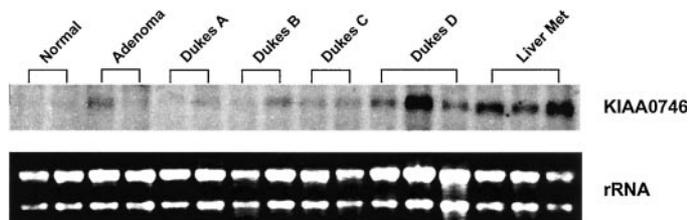


Fig. 4. KIAA0746 (*THC559091*) was further analyzed as a candidate tumor progression gene. This progression blot shows that expression of this gene is significantly increased in advanced CRC tumors, which suggests a possible role in metastasis.

normal and tumor tissue (Fig. 3). Because these analyses again confirmed the results obtained from the arrays, we then used KIAA0746 to probe a tumor progression blot to characterize its expression through tumor progression and to assess its potential utility as a marker for metastasis (Fig. 4).

The use of cell lines as model systems to study colon cancer metastasis may not represent the disease in its entirety. The CRC cellular models used in these experiments, which are based in intrasplenic injections, actually represents a test of the capacity of CRC to adhere to and to colonize the liver (experimental metastasis), which may be different from a model of spontaneous metastasis in which tumor cells are injected s.c. or intracecally and then metastasize distantly. In the former, tumor cells can bypass the invasion processes, whereas in the latter, tumor cells must first have the capacity to invade the primary organ to gain access to the vasculature and lymphatic systems. The model does, however, present a comprehensive picture of critical regulators implicated in the process. We used CRC cell lines to identify genes that may play an important role in metastasis. Many of these genes had not been previously identified and serve as interesting candidates for further investigation. An extension of this analysis would be to use these genes as screening tools in extensive disease states to identify targets that would serve as potential diagnostic or prognostic markers for colon cancer metastasis.

## ACKNOWLEDGMENTS

We thank R. L. Malek and E. Snesrud for assistance in the development of the laboratory protocols; I. E. Holt, J. Li, J. Tsai, and J. White for computational assistance; and C. M. Fraser for her key role in initiating and facilitating this work.

## REFERENCES

- Gutman, M., and Fidler, I. J. Biology of human colon cancer metastasis. *World J. Surg.*, *19*: 226–234, 1995.
- Brattain, M. G., Levine, A. E., Chakrabarty, S., Yeoman, L. C., Willson, J. K., and Long, B. Heterogeneity of human colon carcinoma. *Cancer Metastasis Rev.*, *3*: 177–191, 1984.
- Hill, B. T. *In vitro* human tumour model systems for investigating drug resistance. *Cancer Surv.*, *5*: 129–149, 1986.
- Pocard, M., Tsukui, H., Salmon, R. J., Dutrillaux, B., and Poupon, M. F. Efficiency of orthotopic xenograft models for human colon cancers. *In Vivo (Athens)*, *10*: 463–469, 1996.
- Fidler, I. J. New developments in *in vivo* models of neoplasia. *Cancer Metastasis Rev.*, *10*: 229–243, 1991.
- Morikawa, K., Walker, S. M., Jessup, J. M., and Fidler, I. J. Growth potential of human colorectal carcinomas in nude mice: association with the preoperative serum concentration of carcinoembryonic antigen in patients. *Cancer Res.*, *48*: 1943–1948, 1988.
- Morikawa, K., Walker, S. M., Nakajima, M., Pathak, S., Jessup, J. M., and Fidler, I. J. Influence of organ environment on the growth, selection, and metastasis of human colon carcinoma cells in nude mice. *Cancer Res.*, *48*: 6863–6871, 1988.
- Hewitt, R. E., McMarlin, A., Kleiner, D., Wersto, R., Martin, P., Tsoskas, M., Stamp, G. W., and Stetler-Stevenson, W. G. Validation of a model of colon cancer progression. *J. Pathol.*, *192*: 446–454, 2000.
- Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science (Wash. DC)*, *270*: 467–470, 1995.
- DeRisi, J. L., Iyer, V. R., and Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science (Wash. DC)*, *278*: 680–686, 1997.
- Liang, F., Holt, I., Perlea, G., Karamycheva, S., Salzberg, S. L., and Quackenbush, J. An optimized protocol for analysis of EST sequences. *Nucleic Acids Res.*, *28*: 3657–3665, 2000.
- Quackenbush, J., Cho, J., Lee, Y., Liang, F., Holt, I., Karamycheva, S., Parvizi, B., Perlea, G., Sultana, R., and White, J. The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species. *Nucleic Acids Res.*, *29*: 159–164, 2001.
- Hegde, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Earle-Hughes, J., Snesrud, E., Lee, N. H., and Quackenbush, J. A concise guide to cDNA microarray analysis. *Biotechniques*, *29*: 548–554, 556, 558–560, 2000.
- Quackenbush, J., Liang, F., Holt, I., Perlea, G., and Upton, J. The TIGR gene indices: reconstruction and representation of expressed gene sequences. *Nucleic Acids Res.*, *28*: 141–145, 2000.
- Devereux, T. R., Risinger, J. I., and Barrett, J. C. Mutations and altered expression of the human cancer genes: what they tell us about causes. *In: IARC Sci. Publ. No. 146*, pp. 19–42. Lyon, France: IARC, 1999.
- Yang, G. P., Ross, D. T., Kuang, W. W., Brown, P. O., and Weigel, R. J. Combining SSH and cDNA microarrays for rapid identification of differentially expressed genes. *Nucleic Acids Res.*, *27*: 1517–1523, 1999.
- Maxwell, S. A., and Davis, G. E. Differential gene expression in p53-mediated apoptosis-resistant vs. apoptosis-sensitive tumor cell lines. *Proc. Natl. Acad. Sci. USA*, *97*: 13009–13014, 2000.
- Patel, Y. M., and Lane, M. D. Mitotic clonal expansion during preadipocyte differentiation: calpain-mediated turnover of p27. *J. Biol. Chem.*, *275*: 17653–17660, 2000.
- Hartwell, L. Defects in a cell cycle checkpoint may be responsible for the genomic instability of cancer cells. *Cell*, *71*: 543–546, 1992.
- Aaltonen, L. A., Peltomaki, P., Leach, F. S., Sistonen, P., Pylkkanen, L., Mecklin, J. P., Jarvinen, H., Powell, S. M., Jen, J., Hamilton, S. R., *et al.* Clues to the pathogenesis of familial colorectal cancer. *Science (Wash. DC)*, *260*: 812–816, 1993.
- Dang, C. V., and Semenza, G. L. Oncogenic alterations of metabolism. *Trends Biochem. Sci.*, *24*: 68–72, 1999.
- Pollak, M. Insulin-like growth factor physiology and cancer risk. *Eur. J. Cancer*, *36*: 1224–1228, 2000.
- Fonseca, E., Soares, P., Rossi, S., and Sobrino-Simoes, M. Prognostic factors in thyroid carcinomas. *Verh. Dtsch. Ges. Pathol.*, *81*: 82–96, 1997.
- Nakamura, M., Katano, M., Kuwahara, A., Fujimoto, K., Miyazaki, K., Morisaki, T., and Mori, M. Transforming growth factor  $\beta 1$  (TGF- $\beta 1$ ) is a preoperative prognostic indicator in advanced gastric carcinoma. *Br. J. Cancer*, *78*: 1373–1378, 1998.
- Goyette, M. C., Cho, K., Fasching, C. L., Levy, D. B., Kinzler, K. W., Paraskeva, C., Vogelstein, B., and Stanbridge, E. J. Progression of colorectal cancer is associated with multiple tumor suppressor gene defects but inhibition of tumorigenicity is accomplished by correction of any single defect via chromosome transfer. *Mol. Cell. Biol.*, *12*: 1387–1395, 1992.
- Clark, E. A., Golub, T. R., Lander, E. S., and Hynes, R. O. Genomic analysis of metastasis reveals an essential role for RhoC. *Nature (Lond.)*, *406*: 532–535, 2000.
- Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. USA*, *95*: 5857–5864, 1998.

# Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

## Identification of Tumor Markers in Models of Human Colorectal Cancer Using a 19,200-Element Complementary DNA Microarray

Priti Hegde, Rong Qi, Renee Gaspard, et al.

*Cancer Res* 2001;61:7792-7797.

**Updated version** Access the most recent version of this article at:  
<http://cancerres.aacrjournals.org/content/61/21/7792>

**Cited articles** This article cites 24 articles, 9 of which you can access for free at:  
<http://cancerres.aacrjournals.org/content/61/21/7792.full#ref-list-1>

**Citing articles** This article has been cited by 14 HighWire-hosted articles. Access the articles at:  
<http://cancerres.aacrjournals.org/content/61/21/7792.full#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](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, use this link  
<http://cancerres.aacrjournals.org/content/61/21/7792>.  
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.