Meta-analysis of Colorectal Cancer Gene Expression Profiling Studies Identifies Consistently Reported Candidate Biomarkers

Simon K. Chan,[1,2] Obi L. Griffith,[1,3] Isabella T. Tai,[1,4] and Steven J.M. Jones[1,2,3]

[1]Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Research Centre; and [2]Bioinformatics Graduate Program, [3]Department of Medical Genetics, and [4]Division of Gastroenterology, Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada

Abstract

Purpose: **Elucidation of candidate colorectal cancer biomarkers often begins by comparing the expression profiles of cancerous and normal tissue by performing gene expression profiling. Although many such studies have been done, the resulting lists of differentially expressed genes tend to be inconsistent with each other, suggesting that there are some false positives and false negatives. One solution is to take the intersection of the lists from independent studies. However, often times, the statistical significance of the observed intersection are not assessed.**

Methods: **Recently, we developed a meta-analysis method that ranked differentially expressed genes in thyroid cancer based on the intersection among studies, total sample sizes, average fold change, and direction of differential expression. We applied an improved version of the method to 25 independent colorectal cancer profiling studies that compared cancer versus normal, adenoma versus normal, and cancer versus adenoma to highlight genes that were consistently reported as differentially expressed at a statistically significant frequency.**

Results: **We observed that some genes were consistently reported as differentially expressed with a statistically significant frequency ($P < 0.05$) in cancer versus normal and adenoma versus normal comparisons but not in the cancer versus adenoma comparison.**

Conclusion: **Our meta-analysis method identified genes that were consistently reported as differentially expressed. A review of some of the candidates revealed genes described previously as having diagnostic and/or prognostic value as well as novel candidate biomarkers. The genes presented here will aid in the identification of highly sensitive and specific biomarkers in colorectal cancer.** (Cancer Epidemiol Biomarkers Prev 2008; 17(3):543–52)

Introduction

Colorectal cancer, defined as cancerous growths in the colon, rectum, or appendix, is the third most frequent cancer in both males and females in North America (1). This year, an estimated 20,800 Canadians will be diagnosed with colorectal cancer and ~8,700 will die of it (2). A common area of research interest is the identification of diagnostic biomarkers for early and accurate detection of colorectal cancer (3). Prognostic biomarkers are also being developed to, for example, separate patients who will benefit from adjuvant therapy from those who will not (4) or to determine which patients are at risk for disease recurrence (5). Other studies have focused on understanding cancer progression by identifying differences in gene expression between normal, benign adenoma, and carcinoma stages (6-8).

A common starting point for these studies is the surgical resection of both cancer and normal tissues from patients followed by global expression profiling to determine differentially expressed genes. These studies can result in tens to thousands of such genes, only a small portion of which may actually be of clinical utility. Although an abundance of data comparing the expression profiles of cancerous to normal tissue has been generated, to date, no reliable biomarker has resulted. One explanation for this lack of translational success to the clinic has been the inconsistency in the results of independent studies (1, 9, 10). Explanations for this low overlap include utilization of different tissue resection methods (microdissection, laser capture microdissection, etc.), different expression profiling technologies [cDNA two-channel microarrays, oligonucleotide microarrays, Serial Analysis of Gene Expression (SAGE), etc.], and different analysis methods (multiple correction tests, fold change thresholds, etc.).

Review articles often include lists of genes that have been reported in multiple independent studies. Consistently reported genes are considered to be biologically relevant to colorectal cancer, whereas those reported only sporadically are thought to have resulted from inherent noise or biases in the different platforms and analysis

Received 9/20/07; revised 11/23/07; accepted 12/14/07.

Grant support: BC Cancer Foundation, Canadian Institutes of Health Research (S.K. Chan and O.L. Griffith), and Michael Smith Foundation for Health Research (S.K. Chan, O.L. Griffith, I.T. Tai, and S.J.M. Jones).

Note: Supplementary data for this article are available at http://cebp.aacrjournals.org.

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.

Requests for reprints: Steven J.M. Jones, Genome Sciences Centre, British Columbia Cancer Research Centre, Suite 100, 570 West 10th Avenue, Vancouver, British Columbia, Canada V5Z 4S6. Phone: 604-877-6083; Fax: 604-876-3561. E-mail: sjones@bcgsc.ca

doi:10.1158/1055-9965.EPI-07-2615

methods employed (11). Although these lists are helpful in summarizing the current biomarker candidates, the statistical significance of the level of overlap is usually not considered. One can imagine randomly choosing genes from each expression profiling platform, randomly labeling them as up-regulated or down-regulated, and observing some overlap due to chance alone. Therefore, it would be useful to determine which differentially expressed genes were consistently reported in independent colorectal cancer expression profiling studies with a statistically significant frequency.

To address these challenges, a meta-analysis method was recently developed by our group and applied to published studies of differentially expressed genes in thyroid cancer (12). Such a meta-analysis ignores the differences between studies, such as the expression profiling platform used, and instead focuses on elucidating consistently reported genes. The meta-analysis method involved a vote-counting strategy in which a gene was ranked according to the number of studies reporting its differential expression, the total number of tissue samples used in the studies, and the average fold change. That study resulted in a panel of 12 differentially expressed genes reported at a frequency highly unlikely to have occurred by chance. The panel contained both well-known thyroid cancer markers as well as some uncharacterized genes, showing the ability of the meta-analysis method to highlight novel candidate biomarkers. With these results in mind, the objective of the current study was to apply the meta-analysis method to colorectal cancer to observe whether a statistically significant level of overlap among studies could be observed and to identify promising biomarkers. Also, we improved the meta-analysis method by dividing genes into semiquantitative categories based on the number of tissue samples to highlight genes that may have shown the greatest fold changes but would have been ranked lower by the original method due to fewer tissue samples studied. We curated published lists of differentially expressed genes from 25 independent studies performing global expression profiling to compare colorectal cancer to normal tissue, adenoma to normal tissue, and colorectal cancer to adenoma tissue. Many genes were consistently reported as differentially expressed in multiple studies and this overlap was highly significant. The list of candidate biomarkers we present here will be a valuable resource to the colorectal cancer research community for further studies.

Materials and Methods

Data Collection and Curation. We queried PubMed for colorectal cancer expression profiling studies published between 2000 and 2007. Only studies using tissue samples obtained from surgical resection of cancerous tumors and/or adenomatous polyps were considered. Studies were divided into three comparison types: cancer versus normal, adenoma versus normal, and cancer versus adenoma. We excluded the limited studies that focused on the microsatellite stability of the tissues, specific Dukes stages, or those comparing cancer to cancer samples to determine prognostic biomarkers. In total, differentially expressed genes from 25 independent studies were collected. Twenty-three studies did expression profiling to compare cancer versus normal tissue samples (Table 1), whereas seven and five studies considered adenoma versus normal (Table 2) and cancer versus adenoma (Appendix 6), respectively.

Gene Mapping. In the microarray expression profiling studies, differentially expressed genes were represented

Table 1. Twenty-three colorectal cancer versus normal tissue expression profiling studies included in analysis

Reference	Platform	No. genes/ features	Up-regulated features (Mapped)	Down-regulated features (Mapped)
Habermann et al. (6)	Hs-UniGEM2 human cDNA microarray	9,128	24 (23)	34 (29)
Lin et al. (8)	Custom cDNA microarray	23,040	63 (53)	375 (321)
Buckhaults et al. (19)	SAGE	N/A	153 (106)	246 (201)
Notterman et al. (17)	Affymetrix Human 6500 GeneChip Set	7,457	19 (19)	47 (45)
Galamb et al. (45)	Human Atlas Glass 1.0 cDNA microarray	1,090	83 (83)	17 (17)
Wang et al. (46)	TGS s-4k cDNA microarray	3,800	23 (23)	0
Croner et al. (20)	Affymetrix HG-U133A	22,283	67 (66)	63 (62)
Kwon et al. (47)	Macrogen MAGIC cDNA microarray	4,608	77 (77)	45 (44)
Bertucci et al. (48)	Custom nylon cDNA microarray	8,074	125 (125)	109 (109)
Ohmachi et al. (49)	Agilent cDNA microarray	12,814	84 (82)	0
Mori et al. (50)	Human Atlas Glass 1.0 cDNA microarray	1,090	32 (32)	0
Kim et al. (22)	Oligonucleotide microarray from Compugen/Sigma-Genosys	18,861	272 (271)	216 (216)
Zou et al. (18)	Custom cDNA microarray	8,000	88 (69)	142 (118)
Koehler et al. (51)	Atlas Human Cancer 1.2 Array	1,185	31 (29)	14 (13)
Ichikawa et al. (52)	Custom cDNA microarray	20,784	47 (45)	83 (78)
Jansova et al. (53)	Human 19K microarrays (Clinical Genomic Centre)	19,201	31 (29)	163 (162)
Grade et al. (54)	National Cancer Institute oligonucleotide arrays (Operon V2 oligo set)	21,543	1,057 (994)	36 (36)
Bianchini et al. (55)	Human 19K microarrays (Clinical Genomic Centre)	19,201	76 (76)	12 (12)
Agrawal et al. (21)	Affymetrix Human 6800 GeneChip Set	7,129	257 (253)	82 (78)
Sugiyama et al. (56)	Human Cancer Pathway Finder Gene Arrays (Superarray Bioscience)	96	13 (13)	11 (11)
Kitahara et al. (57)	Custom cDNA microarray	9,216	44 (42)	191 (163)
Williams et al. (58)	Custom cDNA microarray	9,592	203 (192)	85 (76)
Takemasa et al. (59)	Custom cDNA microarray	4,608	22 (22)	36 (36)
Total			3,582 (3,273)	2,955 (2,613)

Table 2. Seven colorectal adenoma versus normal tissue expression profiling studies included in analysis

Reference	Platform	No. genes/features	Up-regulated genes/features (Mapped)	Down-regulated genes/features (Mapped)
Habermann et al. (6)	Hs-UniGEM2 human cDNA microarray	9,128	20 (19)	38 (35)
Lin et al. (8)	Custom cDNA microarray	23,040	63 (53)	375 (321)
Buckhaults et al. (19)	SAGE	N/A	247 (208)	246 (180)
Notterman et al. (17)	Affymetrix Human 6800 GeneChip Set	7,129	20 (20)	0
Galamb et al. (45)	Human Atlas Glass 1.0 cDNA microarray	1,090	12 (12)	33 (33)
Wang et al. (46)	TGS s-4k cDNA microarray	3,800	23 (23)	0
Lechner et al. (60)	Atlas Human Cancer cDNA microarray	588	15 (11)	9 (5)
Total			400 (346)	701 (640)

by an accession ID, HUGO gene name, or Affymetrix probe ID. The sequence identifier was mapped to the National Center for Biotechnology Information Entrez Gene Identifier (Entrez Gene ID; ref. 13) with the aid of custom-developed Perl scripts and the Clone/Gene ID Converter tool (14). For the SAGE study, updated tag to gene mapping data were obtained from SAGE Genie (15).

Total Gene Lists. To estimate the background levels of overlapping studies, we obtained the platform-specific annotation file for each study to identify genes that could potentially be detected as differentially expressed. For commercial platforms, such as Affymetrix and Atlas microarrays, the annotation file was obtained directly from the company Web site. The identifiers in these annotation files were mapped to the corresponding Entrez Gene ID as above to produce a total gene list for each study. Identifiers that could not be mapped to an Entrez Gene ID were ignored. To obtain a total gene list for the SAGE study, all gene names in the tag to gene mapping data from SAGE Genie were mapped to Entrez Gene IDs. For studies that used platforms in which an annotation file could not be obtained, such as the custom cDNA microarrays and some of the oligonucleotide microarrays, an approximation approach was used in which the appropriate number of Entrez Gene IDs was randomly chosen from the combined gene lists from the other platforms.

Assessment of Significance of Study Overlap Using Simulations. To determine if the level of overlap among the studies was significant, we did simulations as described previously (12). Briefly, Perl scripts were created to perform Monte Carlo simulations. In each of the 10,000 permutations, the appropriate number of Entrez Gene IDs from the total gene list of each study was randomly chosen and each ID was randomly labeled as ''UP'' for up-regulated or ''DOWN'' for down-regulated. We used an ''all-or-none'' approach in which the level of overlap for a particular gene was only considered if all the independent studies reporting its differential expression agreed on the direction. The level of overlap among studies in each permutation was counted as in the real analysis. On completion of the permutations, a distribution of overlap results from the simulations was determined and a P value was estimated by comparing the overlap from the simulations to the actual level of overlap in the real data. Significance was defined at $P < 0.05$. Similar to the previous meta-analysis (12), genes were ranked according to three criteria in the following order of importance: (*a*) level of overlap (that is, listing the same gene as differentially expressed with a consistent direction of change), (*b*) total number of samples for overlapping studies, and (*c*) average fold change reported by the studies in agreement. We further subdivided the genes into three categories using a semiquantitative scale: lowest (Q1), moderate (interquartile range), and greatest (values greater than those in Q3) tissue sample sizes. This improvement over the previous published method gives greater importance to the average fold change criteria for ranking genes when total sample numbers are similar.

Results

Of the total 8,176 differentially expressed genes reported in the 25 studies (4,273 up-regulated and 3,903 down-regulated), 7,287 (89.1%) could be mapped to an Entrez Gene ID (3,822 up-regulated and 3,465 down-regulated). In the cancer versus normal and adenoma versus normal comparisons, significant overlap was observed. No such significance was seen in the cancer versus adenoma comparison (Table 3), although each individual study identified differentially expressed genes (refs. 6, 8, 17, 19, 35; see Appendix 6).

We present the results from the cancer versus normal comparison as an example. The simulations showed that the amount of overlap in this comparison was highly significant ($P < 0.0001$), with 573 genes reported as differentially expressed with consistent direction of change in at least two studies (multistudy genes;

Table 3. Summary of comparisons analyzed for overlap

Comparison	Total no. studies	Total no. differentially expressed genes reported (Mapped)	Total no. differentially expressed genes with multistudy confirmation	P
Cancer vs normal	23	6,537 (5,886)	573	<0.0001
Adenoma vs normal	7	1,101 (986)	39	<0.0001
Cancer vs adenoma	5	538 (415)	5	0.08

NOTE: The overlap observed in the cancer versus adenoma comparison was not significant ($P < 0.05$).

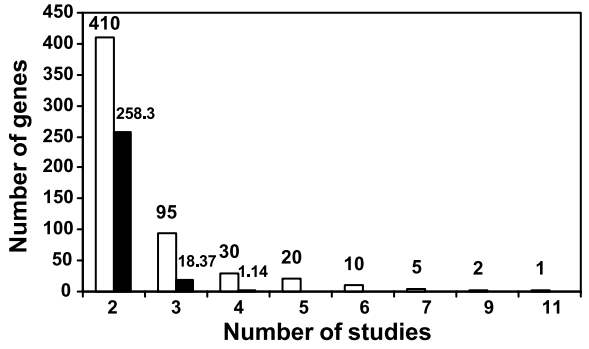

Figure 1. Overlap analysis results for the cancer versus normal comparison. The actual overlap values are bolded (*white columns*), whereas the average overlap values from the Monte Carlo simulations are not (*black columns*). The 95% confidence intervals of the simulations were not included as they were too small to visualize. Note that the average overlap values in the simulations never equaled nor exceeded that which was observed in the actual data.

Fig. 1). There were 175 multistudy genes that were reported with inconsistent direction of differential expression. Thus, the majority of multistudy genes (76.6%) that were reported as differentially expressed agreed on the direction, even for large numbers of studies.

From the Monte Carlo simulations, an average of 258.30 (95% confidence interval, 258.16-258.45) genes would be expected to have an overlap of 2, whereas the actual data contained 410. An average of 18.37 (95% confidence interval, 18.33-18.42) genes would be expected to have an overlap of 3 compared with 95 in the actual data. For an overlap of 4, the simulation produced 1.14 (95% confidence interval, 1.13-1.15) genes, whereas the actual data contained 30 genes. Overlaps of 5, 6, and 7 were observed in the simulations but with averages of less than one hundredth of a gene. In 10,000 permutations, the simulations never produced an overlap greater than 7, whereas two genes had an overlap of 9 and one gene had an overlap of 11 in the real data. Although the total number of genes with an overlap of 2 was still very significant, we present here only the genes reported by three or more studies, as we deemed these to be the most reliable (Appendices 2-5). Additional information on the results appears in the Appendices (online only).

Discussion

A logical solution to the problem of lack of agreement among expression profiling studies in colorectal cancer is to determine the overlap among many studies using different platforms and observe which genes are consistently reported as differentially expressed. These genes likely show biological relevance to the tumorigenesis of colorectal cancer, as opposed to sporadically reported genes, which may be false positives.

Meta-analyses have been done previously to determine differentially expressed genes in colorectal cancer (1, 9, 10). However, these studies and others usually do not consider whether the level of overlap observed is statistically significant. In the newest version (3.0) of the cancer profiling database Oncomine (16), a meta-analysis tool was implemented to compare results from independent studies. However, Oncomine presently contains raw data for eight colorectal cancer profiling studies, only two of which would qualify for our study (17, 18), because they were the only studies that performed at least one of the three comparisons of interest. As discussed previously, our meta-analysis method is useful when raw data are unavailable for consistent reanalysis, which is usually the case (12). However, one limitation of our method is that a measure of confidence cannot be assigned at the gene level, such as from calculating a true combined fold change or *P* value. Thus, in order for more powerful meta-analysis methods to be applied to colorectal cancer profiling studies, researchers should be encouraged to make public their raw data so that they may be included in repositories such as Oncomine.

By applying this method to a near comprehensive collection of colorectal cancer expression profiling studies, we were able to determine the genes that were reported with a statistically significant frequency. As an extension of the meta-analysis method, we categorized some genes according to their total number of tissue samples as lowest (Q1), moderate (interquartile range), and greatest (values greater than those in Q3) instead of using absolute numbers. This allowed the average fold change criterion to have a greater effect on the gene rank in cases where total sample sizes were similar. In the original version of the meta-analysis method, fold change rarely had any effect on rank.

Overlap Significance Observed in Two of Three Comparisons. We observed that for the cancer versus normal and adenoma versus normal comparisons, genes were consistently reported as differentially expressed and that the level of overlap was statistically significant. The results from the cancer versus adenoma comparison were not significant, suggesting that the number of multistudy genes in the five studies could have been observed due to chance. Determining the significance of overlap among studies provides another filtering step to remove false-positive genes from further consideration. When ignoring the significance of the observed overlap, one may be misled by multistudy genes. For example, without knowledge of the statistical significance, one may reason that the multistudy genes in the cancer versus adenoma comparison are biologically relevant, although this decision cannot be reasonably made because the observed level of overlap may be due to chance alone.

Genes Reported with Inconsistent Direction of Differential Expression. In the cancer versus normal comparison, a total of 748 genes were reported as differentially expressed in at least two independent studies. Although the majority of these genes were reported as differentially expressed in the same direction, 175 (23.4%) genes were not. Of these 175 genes, 132 (75.4%) were reported in two studies, 32 (18.3%) were reported in three studies, 8 (4.6%) were reported in four studies, 2 (1.1%) were reported in five studies, and 1 (0.6%) was reported in six studies. There are many potential explanations for these observed inconsistencies. Firstly, one limitation with such meta-analyses is the

overgeneralization of comparisons. Although every effort was made to ensure that each study included in each of the three comparisons were comparable, there are bound to be inconsistencies due to the lack of relevant clinical data being reported in each of the studies. For example, in the cancer versus normal comparison, SLC26A3 was reported as down-regulated in five studies (8, 17, 19-21) but up-regulated in one study (22). The five studies that reported this gene as down-regulated did not specify the microsatellite status of the colorectal cancer tissue samples being used, whereas the one study that reported the up-regulation of this gene used a mixture of microsatellite stable and unstable tissue samples. Other than microsatellite stability, other clinical features, such as the specific portion of the colon where the tissue samples were taken (9), may affect the direction of differential expression. Thus, due to the lack of these clinical data, it is difficult to determine whether the results of each independent study are truly comparable with each other. Conversely, if these clinical data were more readily available, more specific comparisons, such as microsatellite-stable colorectal tissue samples taken from male patients versus paired normal mucosa, could be done.

A related explanation for why some genes were reported as differentially expressed in an inconsistent direction is the heterogeneity in the tissue samples used. The independent studies experimented on tissue samples taken from vastly different populations, each with different genetic and environmental backgrounds that may contribute to differing expression profiles. Furthermore, the tissue samples used by each study themselves will be heterogeneous compared to each other. To have adequate quantities of tissue to work with, most studies do high-throughput expression profiling on pooled tissue samples, which results in a gene expression signal that is ''averaged'' across all cells in the samples (21).

However, the expression of a gene in a single cell may be drastically different from this average. Therefore, depending on how the tissue samples were isolated and which ones were pooled together, the genes may be reported as differentially expressed in an inconsistent direction. One of the studies included in the cancer versus normal comparison (21) investigated the feasibility of pooling tissue samples together by plotting the expression signal of all genes in a pooled sample versus the expression signal of genes from one of the samples in the pool. The authors calculated Pearson correlation coefficients and saw that their values ranged from 0.80 to 0.97, suggesting that the pooling of their specific tissue samples maintained patterns of gene expression representative of each distinct tissue sample. Such an analysis should be done in studies using pooled samples to ensure that the pooled versus unpooled results are comparable.

Finally, poor study design producing inaccurate results may also explain the presence of these genes. In many cases, these genes were ignored because one lone study reported an inconsistent direction of differential expression, which raises suspicions of the validity of the results of the lone disagreeing study. One concern is that some biologically relevant genes may be omitted due to such a study. Therefore, it may be beneficial to include some genes where the majority of the studies agreed on the direction of differential expression instead of the much more stringent ''all-or-none'' approach we have used. However, because the majority of these genes (75.4%) were reported in only two studies, including these genes would not alter the identity of the highest-ranking candidates greatly (Tables 4 and 5).

Despite of these inconsistencies, we remind the reader that the majority of the multistudy genes (76.6%) were consistently reported as differentially expressed in the same direction, which is an encouraging result, given that each independent study used diverse experimental techniques and tissue samples.

Literature Review of Cancer versus Normal Candidates. To further assess our results, we performed a literature review of the genes reported by at least seven studies in the cancer versus normal comparison to determine if any have been shown to have diagnostic and/or prognostic utility in colorectal cancer. The most consistently reported differentially expressed gene in our meta-analysis was carbonic anhydrase II (CA2), which was reported as down-regulated in 11 studies. Along with carbonic anhydrase I, CA2 has been shown to have prognostic significance where the expression of both enzymes was related to the metastatic aggressiveness of colorectal cancer (23). Similarly, the potential diagnostic utility of CA2 was shown in a study in which the average level of fecal CA2 in colorectal cancer patients was shown to be significantly greater than those in the control group (24). Immunohistochemistry has been done on colorectal tumor and healthy mucosa tissue to monitor the protein levels of four carbonic anhydrases, among them CA2 (25). That study showed the level of CA2 protein decreased in cancer relative to healthy tissue, thus confirming the transcript based expression profiling results.

Transforming growth factor-β induced 68 kDa (*TGFβI*) was reported as up-regulated in nine studies. *TGFβI* is a secreted extracellular matrix protein and was discovered through differential expression analysis of a TGF-β1-treated human lung adenocarcinoma cell line (26, 27). This gene has also been shown to be strongly induced by TGF-β1 in many other human cell lines (28, 29). Despite of the consistent overexpression of this gene, as far as we know, no study has focused specifically on its diagnostic and/or prognostic utility or its role in colorectal cancer tumorigenesis. Overexpression at the protein level has yet to be validated with immunohistochemistry.

IFN-induced transmembrane protein 1 (*IFITM1*) was also reported as up-regulated in nine studies. *IFITM1* has been shown to mediate the antiproliferative properties of the IFN cytokines (30) and was observed to be overexpressed in gastric cancer cells, which resulted in tumor cells being more resistant to natural killer cells and produced a more invasive phenotype (31). As far as we know, immunohistochemistry on human colorectal cancer tissue has not been done for IFITM1 protein; however, reverse transcription-PCR (RT-PCR) was conducted previously on adenomas in a murine model as well as a human colorectal carcinoma cell line, HT29, and elevated expression of *IFITM* genes (*IFITM1, IFITM2*, and *IFITM3*) was observed (32). No further studies have considered the diagnostic and/or prognostic potential of *IFITM1* expression in colorectal cancer.

Mal, T-cell differentiation protein-like (*MALL*), reported as down-regulated in seven studies, is a member of the MAL proteolipid family (33) and encodes an integral

Table 4. Up-regulated genes most commonly reported in cancer versus normal expression profiling studies

Gene name	Description	Studies	Studies with fold change	Total sample sizes	Total sample sizes with fold change	Mean fold change	Range	Validation
TGFβI	Transforming growth factor- β induced, 68 kDa	9 (8, 19-22, 47, 51, 54, 57)	8	369	329	8.94	1.11-32.00	RT-PCR (8, 19, 51, 54, 57)
IFITM1	IFN-induced transmembrane protein 1 (9-27)	9 (19, 20, 48, 51-54, 57, 58)	4	351	187	7.52	3.00-12.00	RT-PCR (51, 57)
MYC	V-myc myelocytomatosis viral oncogene homologue (avian)	7 (20, 21, 51, 54, 56, 58, 59)	4	329	243	5.02	1.69-7.50	RT-PCR (6, 51, 54, 58)
SPARC	Secreted protein, acidic, cysteine-rich (osteonectin)	7 (19-22, 51, 58, 59)	5	244	180	6.30	1.27-15.00	Immunohisto-chemistry (39)*
GDF15	Growth differentiation factor 15	7 (8, 18, 19, 21, 22, 51, 58)	5	230	172	7.42	1.58-12.20	RT-PCR (19, 51)
Six studies: greatest sample size								
CXCL1	Chemokine (C-X-C motif) ligand 1 (melanoma growth-stimulating activity, α)	6 (17, 18, 21, 20, 54, 58)	4	287	229	6.54	2.74-10.50	RT-PCR (18, 21)
Six studies: moderate sample size								
CDC25B	Cell division cycle 25 homologue B (Schizosaccharomyces pombe)	6 (17, 20, 21, 51, 57, 58)	4	256	176	4.93	1.81-9.20	RT-PCR (17)
HMBG1	High-mobility group box 1	6 (8, 22, 48, 53, 54, 58)	3	264	161	3.27	2.66-3.91	Western blot, immunohisto-chemistry (61)
Six studies: lowest sample size								
IFITM2	IFN-induced transmembrane protein 2 (1-8D)	6 (8, 19, 20, 52, 57, 59)	3	141	56	7.09	3.00-13.00	RT-PCR (32)
COL1A2	Collagen, type I, α2	6 (19-22, 53, 59)	4	172	130	6.93	2.96-12.00	None found
Five studies: greatest sample size								
CKS2	CDC28 protein kinase regulatory subunit 2	5 (17, 21, 22, 51, 54)	5	285	285	4.21	1.79-7.20	RT-PCR (17, 51)
TOP2A	Topoisomerase (DNA) IIα, 170 kDa	5 (21, 45, 51, 54, 58)	4	277	237	3.61	1.05-5.60	Northern blot, Western blot (62)
UBE2C	Ubiquitin-conjugating enzyme E2C	5 (20-22, 48, 54)	4	274	229	3.03	1.48-5.00	RT-PCR (63)
Five studies: moderate sample size								
CDH3	Cadherin 3, type 1, P-cadherin (placental)	5 (8, 20, 21, 49, 51)	5	194	194	18.16	2.78-74.00	Western blot (64)
INHBA	Inhibin, βA (activin A, activin AB α polypeptide)	5 (20-22, 49, 58)	4	198	158	11.05	1.71-37.00	RT-PCR (65)
SLC12A2	Solute carrier family 12 (sodium/potassium/ chloride transporters), member 2	5 (19, 48, 49, 54, 59)	3	208	139	10.58	3.58-15.15	RT-PCR (54)
MMP11	Matrix metallopeptidase 11 (stromelysin 3)	5 (20-22, 49, 51)	5	208	208	4.22	1.74-5.70	Western blot, immunohisto-chemistry (66)
CSE1L	CSE1 chromosome segregation 1-like (yeast)	5 (17, 20-22, 48)	4	207	162	3.74	1.14-5.00	None found
HNRPA1	Heterogeneous nuclear ribonucleoprotein A1	5 (19, 21, 22, 54, 57)	4	243	203	2.89	1.01-4.50	RT-PCR (67)
Five studies: lowest sample size								
CDK10	Cyclin-dependent kinase (CDC2-like) 10	5 (19-21, 45, 49)	5	150	150	13.85	2.66-17.59	None found
COL3A1	Collagen, type III, α1 (Ehlers-Danlos syndrome type IV, autosomal dominant)	5 (21, 22, 47, 53, 58)	3	178	120	4.31	1.24-9.38	RT-PCR (53)
COL4A1	Collagen, type IV, α1	5 (20-22)	3	168	126	2.70	1.05-4.00	None found

NOTE: The 22 up-regulated genes reported in at least five independent studies with consistent direction are presented here. Genes reported by five and six studies were further subdivided into semiquantitative categories based on the lowest (Q1), moderate (interquartile range), and greatest (values greater than those in Q3) number of tissue samples to give greater importance to the average fold change criteria for ranking genes when total sample numbers were similar. Validation studies that report a gene as differentially expressed in the opposite direction from that of the meta-analysis are marked with an asterisk.

protein located in glycolipid- and cholesterol-enriched membranes. To the best of our knowledge, its expression at the protein level has not been measured by immunohistochemistry, and diagnostic and/or prognostic utilities have not been studied.

Carcinoembryonic antigen-related cell adhesion molecule 1 (*CEACAM1*), reported as down-regulated in seven studies, has been shown to be a tumor suppressor in which expression is lost in adenomas and carcinomas. Moreover, the absence of *CEACAM1* expression was shown to be correlated with reduced rates of apoptosis in polyps (34). However, a retrospective study did immunohistochemistry on *CEACAM1* and showed that 58% of colorectal cancer patients showed an increase in expression (36). It is unclear how the down-regulation of the transcript results in increased *CEACAM1* protein expression. Future studies should focus on the half-life of the cancer *CEACAM1* transcript to determine if it differs

Table 5. Down-regulated genes most commonly reported in cancer versus normal expression profiling studies

Gene name	Description	Studies	Studies with fold change	Total sample sizes	Total sample sizes with fold change	Mean fold change	Range	Validation
CA2	Carbonic anhydrase II	11 (8, 17, 19-22, 53, 55, 57-59)	7	474	352	-15.51	-56.00 to -2.30	RT-PCR (53, 58)
MALL	Mal, T-cell differentiation protein-like	7 (17, 19-21, 51, 57, 59)	5	244	180	-5.34	-10.50 to -1.70	None found
CEACAM1	Carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)	7 (8, 17, 19, 21, 22, 58, 59)	5	222	158	-10.40	-40.00 to -1.38	RT-PCR (17, 58)
Six studies: greatest sample size								
HSD11B2	Hydroxysteroid (11-β) dehydrogenase 2	6 (8, 17, 20-22, 57)	5	224	184	-4.47	-7.60 to -2.23	Northern blot (68)
Six studies: moderate sample size								
SLC26A2	Solute carrier family 26 (sulfate transporter), member 2	6 (8, 18, 20-22, 59)	4	190	148	-6.78	-9.09 to -4.04	None found
FCGBP	Fc fragment of IgG-binding protein	6 (19-22, 48, 57)	4	215	130	-4.88	-7.00 to -1.31	None found
Six studies: lowest sample size								
ACADS	Acyl-coenzyme A dehydrogenase, C-2 to C-3 short chain	6 (8, 17, 19, 20, 22, 58)	5	168	128	-7.11	-20.00 to -2.00	None found
CKB	Creatine kinase, brain	6 (19-22, 53, 57)	4	188	130	-3.11	-5.00 to -1.10	Western blot (69)
Five studies: greatest sample size								
CLU	Clusterin	5 (17, 21, 47, 53, 58)	3	178	120	-3.83	-5.60 to -1.10	Immunohistochemistry (70)
CES2	Carboxylesterase 2 (intestine, liver)	5 (17, 20-22, 59)	4	186	162	-3.58	-6.30 to -1.15	None found
Five studies: moderate sample size								
CA1	Carbonic anhydrase I	5 (17, 19, 20, 22, 57)	4	146	106	-36.90	-59.00 to -5.30	RT-PCR (57)
GPA33	Glycoprotein A33 (transmembrane)	5 (8, 19, 21, 52, 59)	3	131	86	-12.51	-32.50 to -1.70	None found
KRT20	Keratin 20	5 (17, 19, 21, 22, 57)	4	176	136	-8.31	-20.40 to -1.65	None found
SELENBP1	Selenium-binding protein 1	5 (19, 20-22, 59)	4	154	130	-2.80	-3.45 to -1.11	Western blot, immunohistochemistry, mass spectrometry (71)
Five studies: lowest sample size								
CA12	Carbonic anhydrase XII	5 (8, 19, 22, 57, 59)	3	126	62	-4.41	-7.69 to -2.50	Immunohistochemistry (25)*
FABP1	Fatty acid binding protein 1, liver	5 (19, 20, 53, 57, 59)	2	116	34	-4.28	-5.56 to -3.00	RT-PCR (57)

NOTE: The 16 down-regulated genes reported in at least five independent studies with consistent direction are presented here. Genes reported by five and six studies were further subdivided into semiquantitative categories based on the lowest (Q1), moderate (interquartile range), and greatest (values greater than those in Q3) number of tissue samples to give greater importance to the average fold change criteria for ranking genes when total sample numbers were similar. Validation studies that report a gene as differentially expressed in the opposite direction from that of the meta-analysis are marked with an asterisk.

significantly from the normal *CEACAM1* transcript. Furthermore, this study did not observe a relationship between *CEACAM1* protein levels and overall survival or disease-free survival in colorectal cancer patients (36).

Secreted protein, acidic and rich in cysteine (*SPARC*) was reported as up-regulated in seven studies and has been shown by our group to be a putative resistance reversal gene (37). Differentially expressed genes between resistant and sensitive human MIP101 colon cancer cells were determined and *SPARC* was shown to be consistently down-regulated in the resistant cell lines. Their sensitivity was restored by reexpression of *SPARC*, suggesting that expression of *SPARC* has prognostic utility. Immunohistochemistry done on colorectal cancer tissue samples showed increased staining of *SPARC* protein levels (38). However, another immunohistochemistry study (39) showed down-regulation of *SPARC* due to methylation of its promoter. Further studies related to the role of *SPARC* in colorectal tumorigenesis are currently underway in our group.

Growth differentiation factor 15 (*GDF15*), reported as up-regulated in seven studies, is a member of the TGF-β superfamily. Diagnostic and prognostic utility of *GDF15* in colorectal cancer has been suggested by studies that showed increased serum levels of *GDF15* in colorectal cancer patients relative to healthy controls (40). These levels increased during disease progression and may have clinical use in the management of colorectal cancer patients (41).

MYC, reported as up-regulated in seven studies in our meta-analysis, is a transcription factor that regulates various processes, such as cell cycle progression, differentiation, apoptosis, and cell motility (42). Immunohistochemistry on *MYC* has shown that its expression increases during disease progression (43), and when combined with nuclear β-catenin expression, *MYC* expression was shown to have prognostic utility (44).

In conclusion, the results of this meta-analysis identified genes already shown to have diagnostic and/or prognostic potential in colorectal cancer. Perhaps more interesting are the genes, such as *TGFβI* and *IFITM1*, which were consistently reported but have yet to be studied specifically as biomarkers. Also, the genes further down the list (that is, those identified as differentially expressed by four, five, six, etc., independent studies) warrant further investigation. Further studies focused on these genes will help in determining a panel of diagnostic and prognostic colorectal cancer biomarkers with sufficient sensitivity and specificity.

Appendix 1: Mapping Success Rate for the Three Comparisons (Online Only)

The percentage of sequence features that could be mapped to an Entrez Gene ID for each of the three comparison types.

Appendix 2: Up-Regulated Genes Reported in Three or Four Cancer versus Normal Expression Profiling Studies (Online Only)

The 77 up-regulated genes reported by three or four studies were further subdivided into semiquantitative categories based on the lowest (Q1), moderate (interquartile range), and greatest (values greater than those in Q3) number of tissue samples to give greater importance to the average fold change criteria for ranking genes when total sample numbers were similar.

Appendix 3: Down-Regulated Genes Reported in Three or Four Cancer versus Normal Expression Profiling Studies (Online Only)

The 48 down-regulated genes reported by three or four studies were further subdivided into semiquantitative categories based on the lowest (Q1), moderate (interquartile range), and greatest (values greater than those in Q3) number of tissue samples to give greater importance to the average fold change criteria for ranking genes when total sample numbers were similar.

Appendix 4: Up-Regulated Genes Most Commonly Reported in Adenoma versus Normal Expression Profiling Studies (Online Only)

The 23 up-regulated genes reported by two or three studies were further subdivided into semiquantitative categories based on the lowest (Q1), moderate (interquartile range), and greatest (values greater than those in Q3) number of tissue samples to give greater importance to the average fold change criteria for ranking genes when total sample numbers were similar.

Appendix 5: Down-Regulated Genes Most Commonly Reported in Adenoma versus Normal Expression Profiling Studies (Online Only)

The 16 down-regulated genes reported by two or three studies were further subdivided into semiquantitative categories based on the lowest (Q1), moderate (interquartile range), and greatest (values greater than those in Q3) number of tissue samples to give greater importance to the average fold change criteria for ranking genes when total sample numbers were similar.

Appendix 6: Five Cancer versus Adenoma Tissue Expression Profiling Studies Included in Analysis (Online Only)

Appendix 7: Gene Ontology Analysis of Multistudy Genes from the Cancer versus Normal comparison (Online Only)

Of the 573 multistudy genes, 547 were present in the European Bioinformatics Institute Gene Ontology set of 34,242 annotated genes products. A background set of all genes that were represented at least twice among the platforms used in the cancer versus normal expression profiling studies was used. A total of 24 Gene Ontology terms were found to be statistically overrepresented: 5 biological processes (P), 16 cellular components (C), and 3 molecular functions (F). Number Observed and Total Number show the number of genes from the list found associated with each Gene Ontology term over the total number of genes annotated to that term in Gene Ontology.

References

1. Shih W, Chetty R, Tsao MS. Expression profiling by microarrays in colorectal cancer [review]. Oncol Rep 2005;13:517–24.
2. Canadian Cancer Society: colorectal cancer stats. 4/07 update. Available from: http://www.cancer.ca/ccs/internet/standard/0,2283,3172_14447__langId-en,00.html.
3. Srivastava S, Verma M, Henson DE. Biomarkers for early detection of colon cancer. Clin Cancer Res 2001;7:1118–26.
4. Wang Y, Jatkoe T, Zhang Y, et al. Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. J Clin Oncol 2004;22:1564–71.
5. Kim J, Takeuchi H, Lam ST, et al. Chemokine receptor CXCR4 expression in colorectal cancer patients increases the risk for recurrence and for poor survival. J Clin Oncol 2005;23:2744–53.
6. Habermann JK, Paulsen U, Roblick UJ, et al. Stage-specific alterations of the genome, transcriptome, and proteome during colorectal carcinogenesis. Genes Chromosomes Cancer 2007;46:10–26.
7. Kwong KY, Bloom GC, Yang I, et al. Synchronous global assessment of gene and protein expression in colorectal cancer progression. Genomics 2005;86:142–58.
8. Lin YM, Furukawa Y, Tsunoda T, Yue CT, Yang KC, Nakamura Y. Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas. Oncogene 2002;21:4120–8.
9. Cardoso J, Boer J, Morreau H, Fodde R. Expression and genomic profiling of colorectal cancer. Biochim Biophys Acta 2007;1775:103–7.
10. Sagynaliev E, Steinert R, Nestler G, Lippert H, Knoch M, Reymond MA. Web-based data warehouse on gene expression in human colorectal cancer. Proteomics 2005;5:3066–78.
11. Siddiqui AS, Delaney AD, Schnerch A, Griffith OL, Jones SJ, Marra MA. Sequence biases in large scale gene expression profiling data. Nucleic Acids Res 2006;34:e83.
12. Griffith OL, Melck A, Jones SJM, Wiseman SM. Meta-analysis and meta-review of thyroid cancer gene expression profiling studies identifies important diagnostic biomarkers. J Clin Oncol 2006;24:5043–51.
13. Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res 2007;35:D26–31.
14. Alibes A, Yankilevich P, Canada A, Diaz-Uriarte R. IDconverter and IDClight: conversion and annotation of gene and protein IDs. BMC Bioinformatics 2007;8:9.
15. 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.
16. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, et al. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia 2007;9:166–80.
17. Notterman DA, Alon U, Sierk AJ, Levine AJ. Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. Cancer Res 2001;61:3124–30.
18. Zou TT, Selaru FM, Xu Y, et al. Application of cDNA microarrays to generate a molecular taxonomy capable of distinguishing between colon cancer and normal colon. Oncogene 2002;21:4855–62.
19. Buckhaults P, Rago C, St. Croix B, et al. Secreted and cell surface genes expressed in benign and malignant colorectal tumors. Cancer Res 2001;61:6996–7001.
20. Croner RS, Foertsch T, Brueckl WM, et al. Common denominator genes that distinguish colorectal carcinoma from normal mucosa. Int J Colorectal Dis 2005;20:353–62.
21. Agrawal D, Chen T, Irby R, et al. Osteopontin identified as lead marker of colon cancer progression, using pooled sample expression profiling. J Natl Cancer Inst 2002;94:513–21.
22. Kim H, Nam SW, Rhee H, et al. Different gene expression profiles between microsatellite instability-high and microsatellite stable colorectal carcinomas. Oncogene 2004;23:6219–25.
23. Bekku S, Mochizuki H, Yamamoto T, Ureno H, Takayama E, Tadakuma T. Expression of carbonic anhydrase I or II and correlation to clinical aspects of colorectal cancer. Hepatogastroenterology 2000;47:998–1001.
24. Yokoyama S, Shatney CH, Mochizuki H, et al. The potential role of fecal carbonic anhydrase II in screening for colorectal cancer. Am Surg 1997;63:243–6.
25. Kivela AJ, Saarnio J, Karttunen TJ, et al. Differential expression of cytoplasmic carbonic anhydrases, CA I and II, and membrane-associated isozymes, CAIX and XII, normal mucosa of large intestine and in colorectal tumors. Dig Dis Sci 2001;46:2179–86.
26. Skonier J, Neubauer M, Madisen L, Bennett K, Plowman GD, Purchio AF. cDNA cloning and sequence analysis of βig-h3, a novel gene induced in a human adenocarcinoma cell line after treatment with transforming growth factor-β. DNA Cell Biol 1992;11:511–22.
27. Schneider D, Kleeff J, Berberat PO, et al. Induction and expression of βig-h3 in pancreatic cancer cells. Biochim Biophys Acta 2002;1588:1–6.
28. LeBaron RG, Bezverkov KI, Zimber MP, Pavelec R, Skonier J, Purchio AF. βIG-H3, a novel secretory protein inducible by transforming growth factor-β, is present in normal skin and promotes the adhesion and spreading of dermal fibroblasts *in vitro*. J Invest Dermatol 1995;104:844–9.
29. Skonier J, Bennett K, Rothwell V, et al. βig-h3: a transforming growth factor-β-responsive gene encoding a secreted protein that inhibits cell attachment *in vitro* and suppresses the growth of CHO cells in nude mice. DNA Cell Biol 1994;13:571–84.
30. Akyerli CB, Beksac M, Holko M, et al. Expression of IFITM1 in chronic myeloid leukemia patients. Leuk Res 2005;29:283–6.
31. Yang Y, Lee JH, Kim KY, et al. The interferon-inducible 9-27 gene modulates the susceptibility to natural killer cells and the invasiveness of gastric cancer cells. Cancer Lett 2005;221:191–200.
32. Andreu P, Colnot S, Godard C, et al. Identification of the IFITM family as a new molecular marker in human colorectal tumors. Cancer Res 2006;66:1949–55.
33. de Marco MC, Kremer L, Albar JP, et al. BENE, a novel raft-associated protein of the MAL proteolipid family, interacts with caveolin-1 in human endothelial-like ECV304 cells. J Biol Chem 2001;276:23009–17.
34. Nittka S, Gunther J, Ebisch C, Erbersdobler A, Neumaier M. The human tumor suppressor CEACAM1 modulates apoptosis and is implicated in early colorectal tumorigenesis. Oncogene 2004;23:9306–13.
35. Nosho K, Yamamoto H, Adachi Y, Endo T, Hinoda Y, Imai K. Gene expression profiling of colorectal adenomas and early invasive carcinomas by cDNA array analysis. Br J Cancer 2005;92:1193–200.
36. Jantscheff P, Terracciano L, Lowy A, et al. Expression of CEACAM6 in resectable colorectal cancer: a factor of independent prognostic significance. J Clin Oncol 2003;21:3638–46.
37. Tai IT, Dai M, Owen DA, Chen LB. Genome-wide expression analysis of therapy-resistant tumors reveals SPARC as a novel target for cancer therapy. J Clin Invest 2005;115:1492–502.
38. Lussier C, Sodek J, Beaulieu JF. Expression of SPARC/osteonectin/BM4O in the human gut: predominance in the stroma of the remodeling distal intestine. J Cell Biochem 2001;81:463–76.
39. Yang E, Kang HJ, Koh KH, Rhee H, Kim NK, Kim H. Frequent inactivation of SPARC by promoter hypermethylation in colon cancers. Int J Cancer 2007;121:567–75.
40. Welsh JB, Sapinoso LM, Kern SG, et al. Large-scale delineation of secreted protein biomarkers overexpressed in cancer tissue and serum. Proc Natl Acad Sci U S A 2003;100:3410–5.
41. Brown DA, Ward RL, Buckhaults P, et al. MIC-1 serum level and genotype: associations with progress and prognosis of colorectal carcinoma. Clin Cancer Res 2003;9:2642–50.
42. Vita M, Henriksson M. The Myc oncoprotein as a therapeutic target for human cancer. Semin Cancer Biol 2006;16:319–30.
43. Xie D, Sham JS, Zeng WF, et al. Heterogeneous expression and association of β-catenin, p16, and c-myc in multistage colorectal tumorigenesis and progression detected by tissue microarray. Int J Cancer 2003;107:896–902.
44. Bondi J, Bukholm G, Nesland JM, Bukholm IR. Expression of non-membranous β-catenin and γ-catenin, c-Myc and cyclin D1 in relation to patient outcome in human colon adenocarcinomas. APMIS 2004;112:49–56.
45. Galamb O, Sipos F, Dinya E, Spisak S, Tulassay Z, Molnar B. mRNA expression, functional profiling and multivariate classification of colon biopsy specimen by cDNA overall glass microarray. World J Gastroenterol 2006;12:6998–7006.
46. Wang JY, Yeh CS, Tzou WS, et al. Analysis of progressively overexpressed genes in tumorigenesis of colorectal cancers using cDNA microarray. Oncol Rep 2005;14:65–72.
47. Kwon HC, Kim SH, Roh MS, et al. Gene expression profiling in lymph node-positive and lymph node-negative colorectal cancer. Dis Colon Rectum 2004;47:141–52.
48. Bertucci F, Salas S, Eysteries S, et al. Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. Oncogene 2004;23:1377–91.
49. Ohmachi T, Tanaka F, Mimori K, Inoue H, Yanaga K, Mori M. Clinical significance of TROP2 expression in colorectal cancer. Clin Cancer Res 2006;12:3057–63.
50. Mori D, Nakafusa Y, Miyazaki K, Tokunaga O. Differential expression of Janus kinase 3 (JAK3), matrix metalloproteinase 13 (MMP13), heat shock protein 60 (HSP60), and mouse double minute

2 (MDM2) in human colorectal cancer progression using human cancer cDNA microarrays. Pathol Res Pract 2005;2001:777–89.
51. Koehler A, Bataille F, Schmid C, et al. Gene expression profiling of colorectal cancer and metastases divides tumours according to their clinicopathological stage. J Pathol 2004;204:65–74.
52. Ichikawa Y, Ishikawa T, Takahashi S, et al. Identification of genes regulating colorectal carcinogenesis by using the algorithm for diagnosing malignant state method. Biochem Biophys Res Commun 2002;296:497–506.
53. Jansova E, Koutna I, Krontorad P, et al. Comparative transcriptome maps: a new approach to the diagnosis of colorectal carcinoma patients using cDNA microarrays. Clin Genet 2006;69:218–27.
54. Grade M, Hormann P, Becker S, et al. Gene expression profiling reveals a massive, aneuploidy-dependent transcriptional deregulation and distinct differences between lymph node-negative and lymph node-positive colon carcinomas. Cancer Res 2007;67:41–56.
55. Bianchini M, Levy E, Zucchini C, et al. Comparative study of gene expression by cDNA microarray in human colorectal cancer tissues and normal mucosa. Int J Oncol 2006;29:83–94.
56. Sugiyama Y, Farrow B, Murillo C, et al. Analysis of differential gene expression patterns in colon cancer and cancer stroma using microdissected tissues. Gastroenterology 2005;128:480–6.
57. Kitahara O, Furukawa Y, Tanaka T, et al. Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia. Cancer Res 2001;61:3544–9.
58. Williams NS, Gaynor RB, Scoggin S, et al. Identification and validation of genes involved in the pathogenesis of colorectal cancer using cDNA microarrays and RNA interference. Clin Cancer Res 2003;9:931–46.
59. Takemasa I, Higuchi H, Yamamoto H, et al. Construction of preferential cDNA microarray specialized for human colorectal carcinoma: molecular sketch of colorectal cancer. Biochem Biophys Res Commun 2001;285:1244–9.
60. Lechner S, Muller-Ladner U, Renke B, Scholmerich J, Ruschoff J, Kullmann F. Gene expression pattern of laser microdissected colonic crypts of adenomas with low grade dysplasia. Gut 2003;52:1148–53.
61. Volp K, Brezniceanu ML, Bosser S, et al. Increased expression of high mobility group box 1 (HMGB1) is associated with an elevated level of the antiapoptotic c-IAP2 protein in human colon carcinomas. Gut 2006;55:234–42.
62. Shibao K, Takano H, Nakayama Y, et al. Enhanced coexpression of YB-1 and DNA topoisomerase IIα genes in human colorectal carcinomas. Int J Cancer 1999;83:732–7.
63. Takahashi Y, Ishii Y, Nishida Y, et al. 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 2006;168:30–5.
64. Hardy RG, Tselepis C, Hoyland J, et al. Aberrant P-cadherin expression is an early event in hyperplastic and dysplastic transformation in the colon. Gut 2002;50:513–9.
65. Wildi S, Kleeff J, Maruyama H, Maurer CA, Buchler MW, Korc M. Overexpression of activin A in stage IV colorectal cancer. Gut 2001;49:409–17.
66. Madoz-Gurpide J, Lopez-Serra P, Martinez-Torrecuadrada JL, Sanchez L, Lombardia L, Casal JI. Proteomics-based validation of genomic data. Mol Cell Proteomics 2006;5:1471–83.
67. Ushigome M, Ubagai T, Fukuda H, et al. Up-regulation of hnRNP A1 gene in sporadic human colorectal cancers. Int J Oncol 2005;26:635–40.
68. Takahashi K, Sasano H, Fukushima K, et al. 11β-Hydroxysteroid dehydrogenase type II in human colon: a new marker of fetal development and differentiation in neoplasms. Anticancer Res 1998;18:3381–8.
69. Balasubramani M, Day BW, Schoen RE, Getzenberg RH. Altered expression and localization of creatine kinase B, heterogeneous nuclear ribonucleoprotein F, and high mobility group box 1 protein in the nuclear matrix associated with colon cancer. Cancer Res 2006;66:763–9.
70. Chen X, Halberg RB, Ehrhardt WM, Torrealba J, Dove WF. Clusterin as a biomarker in murine and human intestinal neoplasia. Proc Natl Acad Sci U S A 2003;100:9530–5.
71. Kim H, Kang HJ, You KT, et al. Suppression of human selenium-binding protein 1 is a late event in colorectal carcinogenesis and is associated with poor survival. Proteomics 2006;6:3466–76.

Meta-analysis of Colorectal Cancer Gene Expression Profiling Studies Identifies Consistently Reported Candidate Biomarkers

Simon K. Chan, Obi L. Griffith, Isabella T. Tai, et al.

Cancer Epidemiol Biomarkers Prev 2008;17:543-552.