Overlapping Gene Expression in Fetal Mouse Intestine Development and Human Colorectal Cancer

Michael Hu¹ and Ramesh A. Shivdasani¹,²

¹Dana-Farber Cancer Institute and ²Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, Massachusetts

Abstract
Pathways relevant to cancer are well known to overlap with fetal development, as reflected in reactivation of embryonic genes in tumors. However, molecular evidence for this notion has gathered in piecemeal fashion, and systematic approaches have rarely been applied to gauge the extent and global characteristics of the overlap in gene expression between developing tissues and cancer. The fraction of genes that is expressed aberrantly in a given cancer and also developmental in primary function is unknown, and the tissue specificity of recapitulated gene expression remains unexplored. We developed a statistical method to relate expression profiles from human colon cancer and diverse nonintestinal tumors to transcripts that decline in expression with epithelial differentiation in the fetal mouse gut. For genes that are overexpressed in colon cancer, we computed 8% to 19% likelihood that they were expressed transiently during epithelial morphogenesis in intestine development. Among genes dysregulated in other tumors, the corresponding likelihood fell between 1% and 6%. Similarly, low probabilities were obtained when we compared genes not overexpressed in colon cancer with transcriptional profiles in intestine organogenesis. Genes that increase after fetal gut epithelial differentiation were not differentially represented between cancerous and normal colon. Our findings systematically characterize the global extent and tissue specificity of developmental expression programs in colorectal cancer and illustrate the use of such an approach to identify candidate biomarkers and therapeutic targets. (Cancer Res 2005; 65(19): 8715-22)

Introduction
Molecular mechanisms of embryonic development are recognized to correlate with those in cancer, and a growing body of evidence highlights various signaling, transcriptional, and metabolic pathways that are shared between organogenesis and malignant tumors (1–3). Additionally, important properties of tumors including tissue invasion, viability at distant sites, and drug resistance, correlate strongly with the degree of histologic differentiation in resected specimens. Such considerations have popularized the idea that tumor cells represent reversion to a primitive state, although definition of such states is imprecise in both concept and molecular characterization. The supporting evidence is largely anecdotal, and systematic approaches have rarely been applied to gauge the extent and global characteristics of the overlap in gene expression between developing tissues and cancer. Questions such as what fraction of the genes expressed aberrantly in a given cancer reflect reactivation of a developmental program, or whether recapitulated gene expression is characteristic to the affected tissue, remain unanswered. Besides their relevance to tumor biology, these questions have practical implications. First, oncofetal markers can serve as useful tools in cancer diagnosis and in monitoring response to therapy (4, 5). Second, cancer treatments often are limited by severe toxicity that reflects expression of the drug target in unaffected tissues. If tumors depend on the expression of some proteins that are absent in adult tissues, then as therapeutic targets, such proteins might confer a wide therapeutic window.

Computational analysis of suitable mRNA expression data sets could potentially yield systematic approximations of the true underlying statistics on overlapping gene expression between developing and cancerous tissues. One such study (6) interrogated microarray-based gene expression profiles across human medulloblastomas and mouse cerebellar development. The authors used a pattern classification (singular value decomposition) approach to reveal that transcripts increased in expression in human medulloblastoma significantly reflect early mouse cerebellar development, whereas transcripts reduced in this disease correspond to a later, complementary program of gene expression.

The intestinal mucosa is organized into crypts, which house replicating progenitor cells, and villous projections lined by postmitotic epithelial cells with differentiated morphology and functions (7, 8). In a significant developmental transition, the gut endoderm first acquires villous character between 13 and 15 days in mouse gestation. We have assembled over 65,000 serial analysis of gene expression (SAGE; ref. 9) tags (representing over 10,000 unique mRNAs) from the mouse small intestine at 12, 13, and 15 days post-coitus. The complete data set, found at http://genome.dfci.harvard.edu/GutSAGE, shows nearly twice as many significant changes in gene expression between 13 and 15 days post-coitus than between 12 and 13 days post-coitus, revealing notable modulation of gene activity in conjunction with epithelial histogenesis. Genes with reduced expression following the villus transition may serve biological functions that are confined to the period of organogenesis. We recently showed that many such transcripts are absent from the adult gut, and for a small group of developmentally repressed genes, we presented experimental evidence that up to one fifth of the transcripts may be reexpressed in human colorectal tumors (10). Here, we extend these limited findings by applying a computational strategy toward a larger number of transcripts.

We mapped all existing human colon cancer gene expression profiles and a diverse set of expression profiles from nonintestinal tumors onto our data set of mouse intestine development. We applied a conditional probability method to approximate the
to which human colon cancers show tissue-specific recapitulation of developmentally regulated genes. For genes that are overexpressed in colon cancer, our statistical analysis yields 8% to 19% likelihood that they were expressed transiently during gut epithelial morphogenesis in development. Among genes overexpressed in other malignancies, the corresponding probability falls between 1% and 6%. Our findings systematically estimate the extent to which cancer gene dysregulation coincides with the developmental expression program and show the tissue specificity of this process.

Materials and Methods

Treatment of human colon cancer data sets. To capture a comprehensive collection of matched tumor and normal tissue expression data, we queried Oncomine (http://141.214.6.50/oncomine/main/index.jsp) for all results on matched, well-characterized colon cancer compared with normal tissue (11). This search yielded two adequate microarray and two SAGE expression data sets. Starting with SAGE libraries (12) NC_1 versus Tu_98 (14,300 unique SAGE tags) and NC_2 versus Tu_102 (7,400 tags), we first mapped every SAGE tag against all nonredundant human Unigene entries (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene) criteria and retained only those human genes with a reliable murine homologue. Finally, consideration was limited to transcripts with at least one tag entry in our fetal intestine SAGE database. The corresponding mouse genes constitute a set designated CANHOMOL. CANINT contains all such genes with a record in our gut development SAGE database. For the control data set not related to cancer, we scanned the 18 samples for 1,340 transcripts whose mean expression in every case is similar in the tumor and normal tissue (ratio between 0.8 and 1.1) and randomly selected 200 of 1,300 such samples for transcripts with similar mean expression in tumor and normal tissue from three patients for expression corresponding to 14,000 Genbank accession nos. (Affymetrix Hu6800 and Hu35KsubA gene chips). Unlike the above example, the data are available on a raw expression scale, without P for comparisons. We therefore set a more stringent variable and isolated transcripts showing >2-fold mean expression in tumors over the matched normal tissue across all sample pairs; the resulting gene set is denoted mArray_GCM_colon. Filters similar to the above treatment were used to arrive at the corresponding sets CANHOMOL and CANINT. To generate the control, noncancer data set NCINT, we scanned the paired samples for transcripts with similar mean expression in tumor and normal tissues (ratio between 0.9 and 1.1) and randomly selected 200 of 1,300 such transcripts to bring the noncancer and cancer sets to comparable size.

Treatment of data sets from human nonintestinal cancers. The principle of applying successive filters is similar to that described above for colon cancer expression data sets. We first extracted all transcripts that are overexpressed in tumors compared with matched normal tissue. The actual variables differ slightly for each data set, as described below. Where available (liver and plasma cell), we accepted the variables for differential expression from the original reports; for the remaining data sets (prostate and breast), we chose expression cutoffs and P values that are common across the microarray literature and also returned a data set sufficiently large for statistical analysis. Owing to the heterogeneity of these large data sets and our demand of differential expression across >15 or >10 profiled tumors, we chose 1.5- and 1.75-fold cutoffs for the prostate and breast data, respectively. The numerical readouts in microarray analysis are well known

Table 1. Sizes of CANINT and NCINT gene expression sets as they were mapped from human across to homologous mouse genes represented in SAGE profiles of mouse gut development

<table>
<thead>
<tr>
<th>Colon cancer resource</th>
<th>Filter 1 result (no. unique transcripts)</th>
<th>Filter 2 result (no. mouse homologues)</th>
<th>Filter 3 result (no. with SAGE entry)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CAN</td>
<td>NC</td>
<td>CANHOMOL</td>
</tr>
<tr>
<td>Microarray data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mArray_adCa1</td>
<td>160</td>
<td>1,340</td>
<td>69</td>
</tr>
<tr>
<td>mArray_adCa2</td>
<td>150</td>
<td></td>
<td>68</td>
</tr>
<tr>
<td>mArray_adCa3</td>
<td>147</td>
<td></td>
<td>62</td>
</tr>
<tr>
<td>mArray_adCa4</td>
<td>157</td>
<td></td>
<td>70</td>
</tr>
<tr>
<td>mArray_adCa5</td>
<td>176</td>
<td></td>
<td>80</td>
</tr>
<tr>
<td>mArray_GCM_colon</td>
<td>347</td>
<td>3,018</td>
<td>178</td>
</tr>
<tr>
<td>SAGE libraries (tag no.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC1 versus Tu_98 (14,300)</td>
<td>141</td>
<td>619</td>
<td>65</td>
</tr>
<tr>
<td>NC2 versus Tu_102 (7,400)</td>
<td>69</td>
<td>345</td>
<td>30</td>
</tr>
</tbody>
</table>

NOTE: Detailed methods are described in Materials and Methods.
to underestimate, often significantly, the differential expression between cancer and normal tissues (15, 16), and higher arbitrary cutoffs yielded too few genes for further analysis; to reduce chance events, we avoided demanding altered gene expression in fewer tumors.

A second filter eliminated genes that lack a reliable mouse homologue, as specified in LocusLink and annotated in HomoloGene. The final filter retained only genes with a reliable entry in our mouse fetal intestine SAGE database. The resulting groups are denoted as nonintestinal cancer (NIC) gene sets, and Table 2 lists the sizes of the intermediate sets resulting from application of each filter.

**Breast cancer.** We used data from a study in which microarrays containing probes for 12,000 genes were interrogated with RNA derived from 12 invasive ductal carcinomas and three normal breast tissue samples (17). In the first filter, we considered transcripts that are elevated in >10 tumors with mean tumor/normal expression ratio of >1.75.

**Liver tumors.** Data were extracted from a study that characterized mRNA expression using ~17,400-gene chips on 82 hepatocellular carcinoma and 74 nontumorous liver samples (18). For the first filter, we took the reported list of ~1,640 most differentially expressed transcripts between tumor and normal samples (Bonferroni corrected $P < 0.01$; see Supplementary Section of the original report; ref. 18) and considered genes with a mean tumor/normal expression ratio of >2.

**Prostate cancer.** One study (Prostate1) compared gene expression from 52 adenocarcinomas and 30 normal samples using microarrays with probes for ~12,600 genes (19). Owing to heterogeneity of samples in the large cohort, we took transcripts with a "Present" call (expressed at detectable levels) in >15 of the 52 tumor samples and mean tumor/normal expression ratio of >1.5. For the data (Prostate2) from paired SAGE libraries (http://cgap.nci.nih.gov/SAGE), constructed from a microdissected tumor and adjacent normal epithelium, we required a ratio of tumor/normal SAGE tags of >1.5 ($P < 0.01$).

**Plasma cell tumors.** The data derive from a study that used Affymetrix microarrays to compare gene expression profiles from nine patients with multiple myeloma and eight myeloma cell lines with those of nonmalignant plasma cells (20). In our first filter, we took the list of 250 genes that the authors reported as being significantly overexpressed in malignant plasma cells (>2-fold, $P < 0.05$).

### Results

**Statement of the problem and data set resources.** The main question and approach are depicted in Fig. 1. We define the rate of developmental gene recapitulation in cancer as the likelihood that a given gene that is activated in colon cancer was expressed at the highest level during gut epithelial morphogenesis. This is cast as the conditional probability $\Pr(DEV|CAN)$, or the probability of a given gene belonging to the developmental program, conditioned on its overexpression in cancer. The unknown variable, the true likelihood of recapitulation, is approximated by empirically determining the underlying $\Pr(DEV|CAN)$ (the ratio of the size of the subset $x$ to that of the set CAN in Fig. 1). This value is computed for multiple sets of genes (CAN) that are overexpressed in colon cancer, derived from different experiments and profiling methods. We can thus empirically arrive at an approximation for $\Pr(DEV|CAN)$ on each set, resulting in a distribution of likelihood values (right curve in Fig. 1). A similar approximation of developmental gene recapitulation in normal (noncancer, NC) tissues gives the distribution of the complementary probabilities $\Pr(DEV|NC)$ (Fig. 1, left curve). If the cancer and noncancer gene sets are of roughly equal size and the difference between the two empirical distributions (Fig. 1) is statistically significant, one would conclude that developmental gene expression programs are recapitulated at a higher rate in cancer than in the normal tissue. Note that our definitions do not say that more cancer genes are of developmental origin than non cancer genes, nor that more genes overexpressed in cancer are of developmental significance than not.

We considered all transcripts in our mouse intestinal SAGE data set with significantly lower ($P < 0.01$) expression at E15 than at earlier stages (Fig. 2A) and denote the set of 254 transcripts that satisfy this rigorous criterion as $DEV_{INT}$ set 1. Next, we obtained all public expression profiles that compare normal and malignant colorectal tissue by microarray or SAGE as described in Materials and Methods. These cancer data derive from human samples, whereas the developmental data profile mouse tissues; accordingly, the first task is to match human and mouse transcripts. We considered only those human genes with a mouse homologue, as specified through the HomoloGene database.

#### Statistical analysis of overlapping gene expression.

As described in Materials and Methods and summarized in Table 1, we derived sets of genes that are overexpressed in colon cancer and whose mouse homologue is represented in SAGE profiles of the developing gut. The sets designated $CAN_{INT}$ contain transcripts showing >2- or >2.5-fold increase in tumors compared with normal tissue. For one of the two microarray studies (13, 14), we generated five randomly bootstrapped sets ($mArray_{adCa1}$ through $mArray_{adCa5}$), each containing transcripts with >2-fold higher mean expression ($P < 0.001$) in tumors over the normal tissue. For both SAGE and microarray cancer data, the control $NC_{INT}$ sets represent genes whose mean expression was not significantly different between tumor and normal tissues ($P > 0.6$).

We thus generated eight pairs of human gene sets that are overexpressed in colorectal cancer ($CAN_{INT}$) or nearly equally expressed in cancerous and normal intestine ($NC_{INT}$).

To analyze these expression data sets, we adopted the following conditional statistics formulations:

$$\Pr(x \in DEV_{INT} | x \in CAN_{INT})$$  \hspace{1cm} (A)

$$\Pr(x \in DEV_{INT} | x \in NC_{INT})$$  \hspace{1cm} (B)

Eq. A is the conditional probability that any gene $x$, given that it is overexpressed in colon cancer, is also repressed after E12 or E13 in mouse gut development. The second conditional is the distribution on the probability of any gene $x$, given that it is not overexpressed in intestinal cancer, being repressed in gut development. It provides
perturbed our calculations by using a larger developmental gene set, which might be biased because the control necessary to accept or reject the hypothesis that the two conditionals (Eqs. A and B) represent different underlying distributions (Fig. 1).

We computed the conditional probabilities in Eqs. A and B using the eight pairs of CAN_INT and NC_INT gene expression data sets (Table 1) and determined that the distribution on the likelihood of any gene being developmentally regulated in the intestine, given that it is overexpressed in colon cancer, falls between 7.7% and 18.8% (x/n1 in Fig. 2B and C). However, if a gene is not overexpressed in colon tumors, its likelihood of being repressed during gut development (y/n2 in Fig. 2B and C) is at most 6.7% and typically much lower. An unpaired, nonparametric (Mann-Whitney) t test on these data reveals the difference in distribution between the two conditional probabilities to be statistically significant (P < 0.0006). Indeed, the estimated probability of a given gene being silenced during intestine development is, on average, three to four times greater if it is overexpressed in colon cancer than if it is not.

To reduce spurious correlations resulting from data biases, we perturbed our calculations by using a larger developmental gene collection, DEV_INT set 2, with relaxed statistical criteria. This expanded set contains 451 transcripts with lower intestine expression at E15 compared with earlier stages (P < 0.025) and the potential addition of noise creates adversity for the hypothesis under investigation. When DEV_INT set 2 was used to approximate the conditional probabilities given by Eqs. A and B, absolute percentage values differed slightly but the general trends and statistical significance of the comparisons remained similar (Fig. 2C). This similarity in the face of adversity implies that the different distributions on Eqs. A and B represent true underlying biological differences. The results thus suggest a measurable bias for developmentally repressed genes to be reactivated in tumors of the same origin and they support a conclusion we recently reported on a smaller, experimental scale (10). Notably, gene expression in human colorectal cancer mimics only a portion of the transcriptional program that is active during intestine organogenesis.

Assessment of tissue specificity. The apparent reactivation of embryonic gene expression programs is not necessarily specific to the tissue of tumor origin. Gene overexpression in tumors could represent two possibilities. In one case, genes silenced during gut development could reactivate selectively in intestinal tumors, reflecting reversal of tissue-specific epigenetic regulatory mechanisms. Alternatively, reactivation of developmental genes in cancer might represent a non–tissue-specific process, where genes repressed during gut development are appropriated for malignant behaviors in diverse cell types. These scenarios differ both in their implications and in probable underlying mechanisms. Tissue-specific gene reactivation implies that oncogenesis represents some degree of developmental reversal, whereas non–tissue-specific reactivation points simply to shared molecular features between development and cancer. Of note, these implications apply only to the consideration of groups of genes; overexpression of any single gene in fetal and tumor tissues may reflect either coincidence or a cellular process that is common to both processes.

To assess the tissue specificity of developmental gene reexpression, we applied a third statistical formulation and evaluated gene expression data sets from NIC:

$$\Pr(x \in DEV_{INT} | x \in NIC)$$

This conditional is the distribution on the probability that a gene x, given that it is activated in some nonintestinal cancer, is also repressed after E12 or E13 in mouse gut development. It assesses the specificity with which gut developmental gene expression is recapitulated in colon cancer, and we proposed that the conditionals given by Eqs. A and C form different distributions. The conditional probabilities were computed by taking the ratio of the number of genes in the intersection of the appropriate gene sets, as shown in Fig. 3. To compute values for breast cancer, for example, we take the ratio y/n (Fig. 3A), where y is the number of genes in the intersection of the DEV_INT and breast cancer gene sets, and n is the number of genes in the breast NIC set with corresponding entries in fetal gut SAGE libraries (Table 2).

We extracted tumor expression profiles on various nonintestinal human cancers from public data sets. To avoid biases that may result from reliance on a single gene expression platform, we used results from both SAGE and microarray studies on diverse tumors. For five independent NIC gene sets, we computed the conditional probabilities given by Eq. C to fall between 1.3% and 5.6% (y/n in Fig. 3B, column 1). Thus, the estimated probability of a gene being developmentally silenced in the intestine is, on average, 2.5 to 3 times greater if it is overexpressed in colon cancer than if it is dysregulated in other tumors. An unpaired, nonparametric t test between the eight data points for Eq. A and five data points for Eq. C indicates different underlying distributions (P < 0.001), and there was no appreciable change when NIC genes were compared with the less stringent (set 2, P < 0.025) version of DEV_INT (Fig. 3B, column 2).

Tumor expression of differentiation genes. Next, we examined the degree to which colon cancers differ from normal
epithelium in expression of transcripts with the opposite pattern (i.e., increased levels after the developmental villus transition); such genes represent independent controls for the results we have obtained thus far. To define a set of differentiation genes that appear late in mammalian intestine development, we considered all transcripts with significantly higher ($P < 0.01$ and more than eight tags) SAGE representation at E15 than at E12 and/or E13 (Fig. 4A). One hundred seventy-seven transcripts satisfied this requirement and constitute the set we denote as $\text{DIFF}_{\text{INT}}$. To arrive at statistical estimates of the rate at which such late-rising genes are overexpressed in colon cancer, we formulated the following conditional probabilities:

$$
\Pr(x \in \text{DIFF}_{\text{INT}} | x \in \text{CAN}_{\text{INT}}) \quad (D)
$$

$$
\Pr(x \in \text{DIFF}_{\text{INT}} | x \in \text{NC}_{\text{INT}}) \quad (E)
$$

Eq. D is the conditional probability that a gene $x$, given that it is overexpressed in colon cancer, was activated late in fetal gut development. The second conditional is the distribution on the probability of a gene $x$, given that it is not overexpressed in intestinal cancer, displaying such a developmental profile.

We computed the conditional probabilities in Eqs. D and E using the same $\text{CAN}_{\text{INT}}$ and $\text{NC}_{\text{INT}}$ gene expression data sets described above (Table 1) and the methods applied for Eqs. A and B (Fig. 2B). The distribution on the likelihood of any gene increasing in expression after the fetal villous transition, given that it is overexpressed in colon cancer, lies between 0.8% and 11% (Fig. 4B, column 1). If a given gene is not overexpressed in colon tumors (Fig. 4B, column 2), its likelihood of increased expression late in gut development is between 0% and 4%. In an unpaired, nonparametric (Mann-Whitney) $t$ test, the difference between these two conditional distributions is insignificant ($P < 0.53$). Thus, in contrast to the recapitulation of developmentally down-regulated genes in cancer, the rate of expression of differentiation genes is similar between cancerous and normal colon. However, it may be easier in cancer samples to detect developmental transcripts that are reduced or absent in normal tissue than it is to record reduced expression of differentiation markers, which could be contributed by admixed normal mucosa.

**Oncofetal markers.** Our analysis of recapitulation of developmental genes in cancer is limited by at least two factors: the size of currently available gene expression data sets in mouse gut development and human colon cancer and the difficulties in assigning homology between pairs of human and mouse genes. Nevertheless, the statistical basis of our analysis predicts that future studies, which might draw on more comprehensive transcriptional profiles, would yield similar trends. Meanwhile, our current results highlight some developmentally regulated genes that are reactivated in neoplasia (Table 3). Genes with these expression characteristics point to cellular functions that may be common to cancer and developing tissues. The overlap in developmental and malignant gene expression encompasses
Several factors associated with cellular stress response (heat shock proteins 90 and 1B, stress-induced phosphoprotein 1, and chaperonin-containing complexes), protein synthesis (ribosomal proteins L39, L3, and S26 and eukaryotic translation elongation factor 1B2), and cell cycle regulation (CDK4 and the Cdc47 homologue). Insulin-like growth factor II (IGF-II), which is known to enhance tumor growth (21), is also highly expressed in the fetal and cancerous gut. Other genes with similar expression include the paracrine growth factor mkine and an intracellular and cancerous gut. Other genes with similar expression include the paracrine growth factor mkine and an intracellular

**Discussion**

Colorectal carcinomas are heterogeneous in their degree of differentiation and, by definition, lack the normal tissue architecture. Indeed, departure from the normal morphology is a fundamental property of cancer cells that is probably linked to invasion and other malignant behaviors. Underlying these properties is some combination of repression of terminal differentiation genes and reactivation of others associated with development of the target tissue. Previous identification of such oncofetal genes has provided both mechanistic insights in cancer biology and biomarkers that are very useful in managing a variety of human epithelial cancers. Expression profiling reveals many genes with altered expression in tumors relative to the normal tissue (11, 16). We hypothesized that a significant fraction of increased transcript expression between cancers and development of other tissues. The idea that cancers share properties with developing embryos has been discussed by many authors (22, 23) and first gained currency following the embryologic studies of Waddington and Needham in the 1930s, when malignant behaviors were considered in the light of tissue organizers, morphogenetic fields, and cellular hierarchies (24, 25). Recent progress in linking epithelial tumors in general, and colon cancer in particular, to cell signaling pathways that regulate gut development and homeostasis extend these ideas significantly (2). Although many other conceptual and experimental advances have highlighted commonalities between development and cancer, the parallels have for the most part been explored at the level of single or small groups of genes and pathways. Here, we report a systematic and quantitative approach to delineate the degree of overlapping gene expression in colon cancer and development of the mammalian intestine.

We applied computational and statistical strategies to relate expression profiles from human colon cancer and the developing mouse gut. The process applied to generate the target data sets ensures that each human gene considered can be mapped with confidence to a probable murine homolog. Our results reveal that 8% to 19% of genes overexpressed in intestinal tumors had previously shown their highest expression concomitant with fetal villus morphogenesis. The frequency at which such genes are expressed in various nonintestinal tumors or the likelihood of genes not overexpressed in colon cancer having an origin in the gut developmental expression program are both much lower. For each of these trends, the results were similar when we analyzed developmental gene sets defined by different statistical criteria, which suggests that the correlations reflect the true underlying biology. Our analysis thus yields a systematic estimate of the global extent to which developmental gene expression may be recapitulated in a solid tumor and suggests that this process displays a high degree of tissue specificity.

Transcripts expressed in human cancers from various sites have been profiled much more extensively than stages in development of individual organs. This difference limits the degree to which our methods may immediately be extended to explore overlaps in gene expression between cancers and development of other tissues. However, our observations do make a testable prediction: strong correlation between activation of groups of genes in particular tumors and the prior silencing of those genes during fetal development. Therefore, we report a systematic and quantitative approach to delineate the degree of overlapping gene expression in colon cancer and development of the mammalian intestine.
likely contributes to malignant behaviors. Observed in many tumor types, including colon cancer (29), and hypermethylation (28). In contrast, reactivation of developmentally inactivation of tumor suppressor genes, usually by promoter methylation, is important in tumor initiation and experimental evidence indicates that epigenetic alteration, including DNA methylation, is likely amenable to reversal in malignancy. Considerable evidence supports the idea that dysregulated ribosome biogenesis, whether by gain or loss of gene functions, may be tumorogenic (33). We find that developmental transitions reported to a high and unexpected degree in tumors in zebrafish (32). These observations suggest the idea that dysregulated ribosome biogenesis, whether by gain or loss of gene functions, may be tumorigenic (33). We find that developmental transitions may be accompanied by significant modulation in expression of selected genes involved in assembly of the 40S and 60S ribosomal subunits (http://genome.dfci.harvard.edu/GutSAGE). Thus, certain aspects of the ribosome or, alternatively, nonribosomal functions of the genes in question, may be common to developmental and malignant cellular processes. IGF-II, which is known to enhance tumor growth and suppress apoptosis (21), is also highly expressed in the fetal and cancerous gut. Loss of genomic imprinting, with resulting abnormal activation of the normally silent maternal Igf2 allele in all cells, is strongly associated with a

### Table 3. Transcripts overexpressed in human colon cancer, as detected by microarray or SAGE analysis, and also down-regulated during mouse intestine organogenesis

<table>
<thead>
<tr>
<th>LocusLink (human)</th>
<th>Unigene (mouse)</th>
<th>Expression pattern in fetal gut</th>
<th>Gene description</th>
</tr>
</thead>
<tbody>
<tr>
<td>641</td>
<td>12932</td>
<td>E12 – E13 &gt; E15</td>
<td>Bloom syndrome (BLM)</td>
</tr>
<tr>
<td>22948</td>
<td>1813</td>
<td>E12 &gt; E13 &gt; E15</td>
<td>Chaperonin containing TCP1 s5 epsilon (CCT5)</td>
</tr>
<tr>
<td>3320</td>
<td>1843</td>
<td>E12 &gt; E13 – E15</td>
<td>Heat shock 90-kDa protein 1, α</td>
</tr>
<tr>
<td>3192</td>
<td>2115</td>
<td>E12 – E13 &gt; E15</td>
<td>Breakpoint cluster region</td>
</tr>
<tr>
<td>65108</td>
<td>2769</td>
<td>E12 &lt; E13 &gt; E15</td>
<td>MARCKS-like protein</td>
</tr>
<tr>
<td>23204</td>
<td>29924</td>
<td>E12 – E13 &gt; E15</td>
<td>ADP-ribosylation factor-6 interacting protein</td>
</tr>
<tr>
<td>7329</td>
<td>3268</td>
<td>E12 &lt; E13 &gt; E15</td>
<td>Ubiquitin-conjugating enzyme E2I</td>
</tr>
<tr>
<td>10963</td>
<td>4540</td>
<td>E12 &lt; E13 &gt; E15</td>
<td>Stress-induced-phosphoprotein 1</td>
</tr>
<tr>
<td>4869</td>
<td>6343</td>
<td>E12 &gt; E13 &gt; E15</td>
<td>Nuclear phosphoprotein B23</td>
</tr>
<tr>
<td>1019</td>
<td>6839</td>
<td>E12 &lt; E13 &gt; E15</td>
<td>Cyclin-dependent kinase 4 (CDK4)</td>
</tr>
<tr>
<td>3418</td>
<td>246432</td>
<td>E12 – E13 &gt; E15</td>
<td>Isocitrate dehydrogenase 2 (NADP+)</td>
</tr>
<tr>
<td>3608</td>
<td>21334</td>
<td>E12 &gt; E13 – E15</td>
<td>Nuclear factor NF45 mRNA (ILF2)</td>
</tr>
<tr>
<td>4176</td>
<td>18923</td>
<td>E12 – E13 &gt; E15</td>
<td>DNA replication factor CDC47 homologue (MCM7)</td>
</tr>
<tr>
<td>4192</td>
<td>906</td>
<td>E12 &lt; E13 &gt; E15</td>
<td>Midkine (neurite growth-promoting factor 2)</td>
</tr>
<tr>
<td>5902</td>
<td>3752</td>
<td>E12 – E13 &gt; E15</td>
<td>RAN binding protein 1 (RANBP1)</td>
</tr>
<tr>
<td>6428</td>
<td>6787</td>
<td>E12 &gt; E13 &gt; E15</td>
<td>Pre-mRNA splicing factor SRP20</td>
</tr>
<tr>
<td>6749</td>
<td>219793</td>
<td>E12 &lt; E13 &gt; E15</td>
<td>SSRP1 High mobility group box</td>
</tr>
<tr>
<td>7001</td>
<td>42948</td>
<td>E12 &lt; E13 &gt; E15</td>
<td>Thiol-specific antioxidant protein</td>
</tr>
<tr>
<td>7045</td>
<td>14455</td>
<td>E12 &gt; E13 – E15</td>
<td>Transforming growth factor β induced gene (BIGH3)</td>
</tr>
</tbody>
</table>

Cancer gene overexpression detected by SAGE

<table>
<thead>
<tr>
<th>LocusLink (human)</th>
<th>Unigene (mouse)</th>
<th>Expression pattern in fetal gut</th>
<th>Gene description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3326</td>
<td>2180</td>
<td>E12 &gt; E13 &gt; E15</td>
<td>Heat shock protein 1, β</td>
</tr>
<tr>
<td>1933</td>
<td>2718</td>
<td>E12 &gt; E13 &gt; E15</td>
<td>Eukaryotic translation elongation factor 1B2</td>
</tr>
<tr>
<td>6170</td>
<td>30478</td>
<td>E12 &gt; E13 &gt; E15</td>
<td>Ribosomal protein L39</td>
</tr>
<tr>
<td>6122</td>
<td>3486</td>
<td>E12 &lt; E13 &gt; E15</td>
<td>Ribosomal protein L3</td>
</tr>
<tr>
<td>6231</td>
<td>372</td>
<td>E12 &gt; E13 – E15</td>
<td>Ribosomal protein S26</td>
</tr>
<tr>
<td>3481</td>
<td>3862</td>
<td>E12 &gt; E13 &gt; E15</td>
<td>IGF-II</td>
</tr>
</tbody>
</table>

NOTE: Fetal expression patterns, archived at http://genome.dfci.harvard.edu/GutSAGE, were recorded on gestational days E12, E13, and E15. Few transcripts were expressed to varying degrees in the normal adult tissue. Nor do we imply that most transcripts overexpressed in cancer belong to the developmental program. There are undoubtedly many modes of aberrant gene activation in tumors, of which developmental gene reactivation is only one.

Table 3. Transcripts overexpressed in human colon cancer, as detected by microarray or SAGE analysis, and also down-regulated during mouse intestine organogenesis

Tissue differentiation during development is largely under epigenetic control. It follows that there may be two classes of epigenetic modification pertinent to this discussion. One category is represented in genes that are never accessible to the transcription machinery in a particular cell type and accordingly never expressed therein. The second type of modification occurs in genes after their transient embryonic expression and such changes may be especially amenable to reversal in malignancy. Considerable experimental evidence indicates that epigenetic alteration, including DNA methylation, is important in tumor initiation and progression (26, 27); investigation has traditionally emphasized inactivation of tumor suppressor genes, usually by promoter hypermethylation (28). In contrast, reactivation of developmentally regulated genes may result from the aberrant hypomethylation observed in many tumor types, including colon cancer (29), and likely contributes to malignant behaviors.

Historically, oncalfetal genes have been identified individually. Our computational analysis of malignant and developmental expression profiles represents a strategy to identify new candidates and reveals tumor reactivation of fetal genes with a wide range of cellular functions. A few gene classes are prominent, including factors associated with cellular stress response, cell cycle regulation, and protein synthesis. Although increased expression of ribosomal protein genes is recognized as a feature of human cancers (30, 31), the underlying significance is uncertain. Conversely, monoallelic loss-of-function mutations in gene loci for ribosomal proteins were reported to a high and unexpected degree in tumors in zebrafish (32).
risk of developing colon cancer (34). Besides their value in cancer diagnosis and surveillance, oncofetal factors could also be good therapeutic targets. If they are expressed on the tumor cell surface or make a material contribution to the malignant phenotype, then specific antibodies or drugs may be developed in anticipation of limited toxicity as the target is absent from the normal adult tissue.

**Acknowledgments**

Received 3/1/2005; revised 7/7/2005; accepted 7/22/2005.

**Grant support:** Robert Black Charitable Foundation and NIH grant R01DK61139.

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.

We thank Maina Leporezelet and Li Cai for their contributions toward early phases of this work and members of the Shivdasani laboratory for helpful discussions.

**References**

Overlapping Gene Expression in Fetal Mouse Intestine Development and Human Colorectal Cancer

Michael Hu and Ramesh A. Shivdasani


Updated version

Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/65/19/8715

Cited articles

This article cites 33 articles, 17 of which you can access for free at:
http://cancerres.aacrjournals.org/content/65/19/8715.full.html#ref-list-1

Citing articles

This article has been cited by 3 HighWire-hosted articles. Access the articles at:
/content/65/19/8715.full.html#related-urls

E-mail alerts

Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions

To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions

To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.