A Gene Expression Signature of Genetic Instability in Colon Cancer

Craig P. Giacometti,1 Suet Yi Leung,2 Xin Chen,3 Siu Tsan Yuen,3 Young H. Kim,1 Eric Bair,2 and Jonathan R. Pollack1

Departments of Pathology and Statistics, Stanford University, Stanford, California; 1Department of Pathology, The University of Hong Kong, Queen Mary Hospital, Pokfulam, Hong Kong; and 1Department of Biopharmaceutical Sciences, University of California San Francisco, San Francisco, California

Abstract

Genetic instability plays a central role in the development and progression of human cancer. Two major classes of genetic instability, microsatellite instability (MSI) and chromosomal instability (microsatellite stable; MSS), are best understood in the context of colon cancer, where MSI tumors represent ~15% of cases, and compared with MSS tumors, more often arise in the proximal colon and display favorable clinical outcome. To further explore molecular differences, we profiled gene expression in a set of 18 colon cancer cell lines using cDNA microarrays representing ~21,000 different genes. Supervised analysis identified a robust expression signature distinguishing MSI and MSS samples. As few as eight genes predicted with high accuracy the underlying genetic instability in the original and in three independent sample sets, comprising 13 colon cancer cell lines, 61 colorectal tumors, and 87 gastric tumors. Notably, the MSI signature was retained despite genetically correcting the underlying instability, suggesting the signature reflects a legacy of the tumor having arisen from MSI, rather than sensing the ongoing state of MSI. Our findings support a model in which MSI and MSS preferentially target different genes and pathways in cancer. Further, among the MSI signature genes, our findings implicate a role of elevated metallothionein expression in the clinical behavior of MSI cancers. (Cancer Res 2005; 65(20): 9200-5)

Introduction

Cancer is the end result of a multistep process in which cells progressively accumulate mutations in key genes that control cell growth (1). Because the mutation rate of normal somatic cells is prohibitively low, it is widely believed that genetic instability with increased mutation rates necessarily underlies the development of cancer (2). Two major classes of genetic instability, microsatellite instability (MSI) and chromosomal instability, have been studied extensively and are best understood in the context of colon cancer (3).

Colon cancer with MSI results from defects in DNA mismatch repair caused by mutation or hypermethylation of MLH1, MSH2, or, less frequently, MSH6 (4). Such cancers are near-diploid, but accumulate subtle sequence changes, most notably small insertions or deletions within microsatellite sequences (i.e., short mononucleotide or dinucleotide repeat sequences), causing frameshift mutations thereby disrupting cancer genes. MSI is observed in ~15% of sporadic colon cancers and underlies the autosomal dominant syndrome hereditary nonpolyposis colon cancer (5).

Non-MSI colon cancers (microsatellite stable or MSS) exhibit chromosomal instability, with unbalanced chromosome translocations and gains and losses of whole chromosomes, resulting in the altered expression of cancer genes. Transient chromosome instability occurs early in cancer development when telomeres become critically shortened and recombinogenic (“telomere crisis”), thereby initiating fusion-bridge-breakage cycles (6). Ongoing chromosome instability has also been documented in colon cancer cell lines (7) and may arise from defects in DNA double-strand break repair, kinetochore function, or chromatid segregation (8). MSS comprises ~85% of sporadic colon cancers and is observed in tumors arising in the autosomal dominant syndrome familial adenomatous polyposis (5).

Interestingly, MSI and MSS colorectal tumors display some distinct clinicopathologic features. Compared with their MSS counterparts, MSI tumors more often arise in the proximal (right) colon, present with poor differentiation, and are associated with a more favorable prognosis (9, 10). These differing features suggest the underlying genetic instabilities might target distinctly susceptible cancer genes and/or pathways. Indeed, although the WNT signaling pathway is activated in both classes, activation in MSS tumors results from frequent mutation/loss of the APC tumor suppressor locus, and in MSI tumors through frameshift mutation of CTNNB1 (11, 12). Likewise, inhibition of transforming growth factor β growth suppression in MSI tumors results from the frequent frameshift mutation of TGFBR2, and in MSS tumors by mutation/loss of SMAD4 (13, 14). Frameshift mutation in the antiapoptotic gene BAX is frequent in MSI tumors whereas mutation/loss of TP53 is common in MSS samples (12, 15). Additional informative molecular differences are likely to be discernable in global patterns of gene expression, which here we have assayed using DNA microarrays.

Materials and Methods

Colon cancer cell lines. Twenty-eight colon cancer cell lines were obtained from the American Type Culture Collection (Manassas, VA) or the European Collection of Cell Cultures (Salisbury, United Kingdom). RKO was obtained from Dr. Michael Kastan (Department of Hematology-Oncology, St. Jude Children's Research Hospital, Memphis, TN), LIM2125 and LIM2412 from Dr. Richard Simpson (Ludwig Institute for Cancer Research, Royal Melbourne Hospital, Parkville, Victoria, Australia), and HCT116+ch2 and HCT116+ch3 were kindly provided by Dr. Thomas Kunkel (National Institute of Environmental Health Sciences, Research Triangle Park, NC). The MSI status of cell lines, determined by assaying ubiquitous somatic mutations, specific microsatellite loci, and/or functional repair activity of cell extracts, was ascertained from the literature (Table S1; refs. 16–20). Cell lines were propagated in RPMI 1640 supplemented with 10% fetal bovine serum (HyClone, Logan, UT). The HCT116+ch2 and HCT116+ch3 cultures were supplemented with 400 µg/mL of G418 (to maintain the extra chromosome).
Cells were harvested at 80% confluency and total RNA was isolated using the Trizol (Invitrogen, Carlsbad, CA) method.

**Colorectal and gastric cancer specimens.** The clinicopathologic features of the 61 colorectal cancer specimens (39 primary tumors and two metastases) and 87 primary gastric cancer specimens, of which microarray data were analyzed in the current study, are summarized in Tables S2 and S3. For the determination of MSI status, DNA was extracted from paired tumor and normal tissue using either frozen or paraffin blocks. In each case, sections were cut for histologic evaluation before DNA extraction. Only blocks containing >70% tumor areas for colorectal cancer and >50% tumor areas for gastric cancer were used (because of the inherent diffuse infiltrative nature of some gastric cancers). Analysis of MSI was done using the following loci: BAT25, BAT26, BAT40, D2S123, DSS346, D17S250, D18S58, DSS82, TP53, and DCC. At least five loci were analyzed in each case, including both dinucleotide and mononucleotide loci. Tumors were designated MSI-H when at least 40% of loci displayed altered electrophoretic mobility relative to the corresponding normal tissue (21). Tumors with no or <40% altered loci were considered together as an MSS group; no MSI-L group was distinguished in this analysis. Among the 13 MSI colorectal and 21 MSI gastric cancers studied, all but 1 colorectal and 1 gastric cancer displayed loss of MSI2, MLH1, or MSH6 protein by immunohistochemistry (MSI2 loss in 6 colorectal and 1 gastric cancer; MLH1 loss in 5 colorectal and 19 gastric cancers; and MSH6 loss in 1 colorectal cancer). We have previously reported the MSI status, mismatch repair protein expression, germ-line mutations, and promoter methylation status for some of these cases (22–26).

For the determination of MSI status, DNA was extracted using GenePix Pro software, and the data uploaded into the Stanford Microarray Database (29) for storage, retrieval, and analysis. The complete microarray data sets are available at http://smd.stanford.edu, or from the Gene Expression Omnibus at http://www.ncbi.nlm.nih.gov/geo (accession number GSE2591).

**Data analysis.** Fluorescence ratios were normalized for each array, and then well-measured genes (fluorescence intensities for the Cy5 or Cy3 channel at least 2-fold above background) were subsequently “mean centered” (i.e., reported for each gene relative to the mean ratio across all samples). To identify genes differentially expressed in MSI and MSS cell lines, we used the significance analysis of microarrays (SAM) method (30), which is based on a modified t statistic and uses random permutations to estimate a false discovery rate, comparable to a P value. The input for this analysis was the subset of 8,311 cDNAs of which expression was well measured in at least 50% of samples and of which expression varied at least 3-fold from the mean in at least one sample. To classify specimens in the independent sample sets, we used the prediction analysis of microarrays (PAM) method, based on nearest shrunken centroids (31). The microarray data set of 87 gastric tumors has been described (32). The microarray data were analyzed in the current study, are summarized in Tables S2 and S3. To identify genes differentially expressed in MSI and MSS cell lines, we used the significance analysis of microarrays (SAM) method (30), we identified 217 genes (false discovery rate < 10%) of which expression was significantly different in MSI and MSS samples (Fig. 1; Table S4).

**Expression profiling.** cDNA microarrays were obtained from the Stanford Functional Genomics Facility and contained 39,632 different human cDNAs, representing 21,411 human genes (UniGene clusters; ref. 27). We did expression profiling exactly as described (28). Briefly, 50 μg of total RNA from each cell line and 1.5 μg of reference mRNA (derived from 11 different established human cell lines) were differentially labeled with Cy5 and Cy3, respectively, and cohybridized to cDNA microarrays. Following overnight hybridization and washing, arrays were imaged using a GenePix 4000B scanner (Molecular Devices, Sunnyvale, CA). Fluorescence ratios were extracted using GenePix Pro software, and the data uploaded into the Stanford Microarray Database (29) for storage, retrieval, and analysis. The complete microarray data sets are available at http://smd.stanford.edu, or from the Gene Expression Omnibus at http://www.ncbi.nlm.nih.gov/geo (accession number GSE2591).

**Expression profiling.** cDNA microarrays were obtained from the Stanford Functional Genomics Facility and contained 39,632 different human cDNAs, representing 21,411 human genes (UniGene clusters; ref. 27). We did expression profiling exactly as described (28). Briefly, 50 μg of total RNA from each cell line and 1.5 μg of reference mRNA (derived from 11 different established human cell lines) were differentially labeled with Cy5 and Cy3, respectively, and cohybridized to cDNA microarrays. Following overnight hybridization and washing, arrays were imaged using a GenePix 4000B scanner (Molecular Devices, Sunnyvale, CA). Fluorescence ratios were extracted using GenePix Pro software, and the data uploaded into the Stanford Microarray Database (29) for storage, retrieval, and analysis. The complete microarray data sets are available at http://smd.stanford.edu, or from the Gene Expression Omnibus at http://www.ncbi.nlm.nih.gov/geo (accession number GSE2591).

**Data analysis.** Fluorescence ratios were normalized for each array, and then well-measured genes (fluorescence intensities for the Cy5 or Cy3 channel at least 2-fold above background) were subsequently “mean centered” (i.e., reported for each gene relative to the mean ratio across all samples). To identify genes differentially expressed in MSI and MSS cell lines, we used the significance analysis of microarrays (SAM) method (30), which is based on a modified t statistic and uses random permutations to estimate a false discovery rate, comparable to a P value. The input for this analysis was the subset of 8,311 cDNAs of which expression was well measured in at least 50% of samples and of which expression varied at least 3-fold from the mean in at least one sample. To classify specimens in the independent sample sets, we used the prediction analysis of microarrays (PAM) method, based on nearest shrunken centroids (31). The microarray data set of 87 gastric tumors has been described (32). The microarray data were analyzed in the current study, are summarized in Tables S2 and S3. To identify genes differentially expressed in MSI and MSS cell lines, we used the significance analysis of microarrays (SAM) method (30), we identified 217 genes (false discovery rate < 10%) of which expression was significantly different in MSI and MSS samples (Fig. 1; Table S4).

**Results**

To explore the molecular variation in MSI and MSS tumors, we profiled gene expression in a set of 18 colon cancer cell lines, consisting of 8 MSI and 10 MSS samples, of which MSI status had been previously ascertained (Table S1), using cDNA microarrays representing ~21,000 different human genes. By supervised analysis using the SAM method (30), we identified 217 genes (false discovery rate < 10%) of which expression was significantly different in MSI and MSS samples (Fig. 1; Table S4).
We next asked whether the differences in expression were sufficiently robust to classify cancer samples based on their underlying instability. Using the PAM method (31), we built an optimal gene expression predictor for the MSI versus MSS distinction. In 10-fold (i.e., leave 1/10th out) cross-validation, an eight-gene predictor was 100% accurate in classifying MSI status in our training set of 18 colon cancer cell lines (Fig. 2A; Table S5). Notably, the same predictor also displayed high predictive accuracy in three independent test sets, with 100% predictive accuracy in a set of 13 subsequently profiled colon cancer cell lines (B); 61 colorectal carcinomas (C); 87 gastric carcinomas (D); and MSI cell lines HCT116 and HCT116+ch3, and mismatch repair–corrected cell line HCT116+ch3 (E). Mean-centered gene expression ratios are depicted by a log 2 pseudocolor scale (ratio fold change indicated). Predicted MSI status is indicated by text; observed MSI status is indicated by blue (MSI) or magenta (MSS) squares. Predictive strengths (i.e., class probabilities, ranging from 0 to 1) for MSI and MSS are displayed graphically.

The identified expression signatures might reflect a cellular response to ongoing genetic instability. Alternatively, the signatures might reflect the history of altered genes and pathways arising from genetic instability. To distinguish between these two possibilities, we profiled expression of a colon cancer cell line, HCT116+ch3, in which the underlying mismatch repair deficiency and MSI phenotype had been functionally corrected by microcell-mediated transfer of chromosome 3, harboring wild-type MLH1 (17). The expression of MLH1 in these cells was reconfirmed by Western blot analysis (Fig. S1). The MSI cell line HCT116+ch2, which carries an irrelevant transferred chromosome of comparable size, served as a control. Interestingly, our eight-gene predictor classified the HCT116+ch3 cell line as MSI, despite the genetic correction of the instability (Fig. 2E).

An unaltered MSI expression pattern was also clearly evident in the larger SAM set of differentially expressed genes (Fig. S2), notwithstanding the expected increased expression of chromosome 3 genes (Fig. S3). We conclude that the MSI expression signature reflects the history, rather than the ongoing state, of genetic instability.

Discussion

The main objective of our study was to explore the molecular variation underlying MSI and MSS cancer. Using DNA microarrays,
we identified gene expression signatures that distinguish MSI and MSS colon cancer cell lines. Notably, the expression of as few as eight genes could predict with high accuracy the underlying genetic instability in colon cancer cell lines and in colorectal and gastric tumors. That the signatures were apparent in primary tumor specimens validates both the relevance of our cell culture findings and the utility of cultured cells in the study of genetic instability. That the signatures were evident in gastric tumors suggests that our findings reflect a more general feature of genetic instability, not restricted to colon cancer.

Most generally, the signatures might reflect the cellular response to an ongoing state of genetic instability, perhaps, for example, sensed as DNA damage. Alternatively, the signatures might reflect the history of accumulated mutations and altered pathways having arisen as a consequence of one or the other genetic instability. The latter is supported by our finding that HCT116+ch3 retains an MSI signature despite its MSI phenotype having been genetically corrected.

Given this result, one possibility is that the signatures reflect different underlying cell types from which the cancers arose, cell types that are perhaps differentially susceptible to one or the other instability. Indeed, as noted earlier, MSI tumors tend to arise more often in the proximal colon, so perhaps these signatures represent differences in the underlying cells of the proximal and distal colon. This possibility seems unlikely, however, given our finding these signatures in gastric tumors as well. Further, we find no appreciable overlap between our signatures of genetic instability and the patterns of gene expression reported by Glebov et al. (33) that distinguish normal tissue from the proximal and distal colon. We also identify no overlap with a signature of intestinal metaplasia (32), a precursor of some gastric cancers in which metallothionein expression has been noted (ref. 34; see Supplementary data).

More likely then, the different signatures in MSI and MSS tumors reflect distinct sets of genes and pathways that are preferentially susceptible to alteration by one or the other instability. Some of the genes within the signatures might represent direct targets of genetic instability. For example, genes with reduced expression in MSI tumors might harbor frameshift mutations in microsatellite sequences leading to nonsense-mediated decay (35). Likewise, the elevated or reduced expression of genes in MSS tumors might reflect dosage changes accompanying chromosomal gain or loss, respectively. Other genes within these signatures likely represent genes and pathways secondarily affected by the primary targets of genetic instability, a subset of which might nonetheless contribute to tumor development or progression.

Within the signature genes, several are notable for their potential roles in tumorigenesis. For example, four of the eight PAM-derived classifier genes were metallothionein isoforms, displaying elevated expression in MSI samples. In recently published microarray studies of colon cancer cell lines and tumors, we also found metallothioneins listed among the genes exhibiting elevated expression in MSI samples (36–38), and we have confirmed our own results by quantitative RT-PCR. Metallothioneins are a family of small, cysteine-rich intracellular metal-binding proteins, of which expression has been associated with protection against DNA damage, oxidative stress, and apoptosis (39). Metallothionein expression confers resistance to a range of antineoplastic compounds, including platinum-based and alkylating agents (39). Interestingly, despite that MSI has been associated with a more favorable prognosis in colon cancer (10), and one study indicating more favorable response to

<table>
<thead>
<tr>
<th>Test set (n)</th>
<th>MSI predictive accuracy</th>
<th>MSS predictive accuracy</th>
<th>Overall predictive accuracy</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colon cancer cell lines (13)</td>
<td>100% (7 of 7)</td>
<td>100% (6 of 6)</td>
<td>100% (13 of 13)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Primary colon tumors (61)</td>
<td>85% (11 of 13)</td>
<td>88% (42 of 48)</td>
<td>87% (53 of 61)</td>
<td>0.003</td>
</tr>
<tr>
<td>Primary gastric tumors (87)</td>
<td>71% (15 of 21)</td>
<td>88% (58 of 66)</td>
<td>84% (73 of 87)</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Figure 3. Validation of elevated metallothionein expression in MSI samples by quantitative RT-PCR. A, correlation of MT1H expression measurements by DNA microarray and by quantitative real-time RT-PCR; R value is shown. B, differential expression of metallothioneins (MT1H, MT1X, and MT2A) in MSI (blue diamond) and MSS (magenta triangle) samples; bars, class averages. Quantitative RT-PCR analysis was done for eight (four MSI, four MSS) representative samples (indicated) from the original 18 cell lines. Quantitative RT-PCR ratios shown are normalized to GAPDH and mean centered.
flourouracil therapy (40). MSI tumors of the colon and other areas are often reported to exhibit relative resistance to chemotherapeutic agents both in vitro and in vivo (41–44). Chemoresistance has been attributed to the inability of the defective mismatch repair pathway to sense DNA damage and induce apoptosis (45). However, our data suggest an alternative possibility that the elevated expression of metalloproteinases in MSI tumors confers chemoresistance. In support of this hypothesis, although alkylating agent-treated mismatch repair-proficient HCT116-ch3 cells, in comparison with HCT116 parental cells, display impaired growth, they do not undergo apoptosis as do MSS cells (17). We have shown that elevated metalloproteinase expression is retained in HCT116-ch3 and may therefore play a protective role in conferring relative chemoresistance. If so, our findings would suggest a strategy for improving chemotherapeutic response in MSI tumors through inhibition of metalloproteinase synthesis or activity (46).

In addition, and represented twice within the PAM-derived eight-gene signature, SPEC1 (small protein effector 1 of Cdc42) displayed elevated expression in MSI samples. SPEC1 interacts with the Rho GTPase, Cdc42, coordinating its control of filopodia formation (47), involved in cell migration and tumor invasion. Within the list of SAM-derived differentially expressed genes (Fig. 1; Table S4), MDM2 was expressed at higher levels in MSI samples. MDM2 is a negative regulator of the tumor suppressor TP53 (48) and its overexpression was observed in MSS cases (49). Among the genes more highly expressed in MSS samples was matrix metalloproteinase 7 (MMP7; Fig. S2; Table S6). Metalloproteinases promote cancer invasion and metastasis (50), and MMP7 (matrixisin) expression in particular has been linked to metastasis in colorectal cancers (51); our results support the potential of metalloproteinase inhibitors (52) in the treatment of MSS colorectal tumors.

The gene expression signatures may also provide clinical utility in the diagnosis of MSI colon cancer, indicated for the identification of familial cases and for prognostication. PCR-based microsatellite sequence length determination, the current gold standard for MSI diagnosis, is labor-intensive, although immunostaining of mismatch repair proteins has shown promise in screening for MSI cases and suggests the underlying mutation (53). Although classification by microarray is not yet sufficiently accurate for clinical diagnosis, the gene expression signatures may suggest new ancillary immunostain markers for improved clinical diagnosis.

In summary, using DNA microarrays, we have identified distinct gene expression signatures in MSI and MSS colon cancer. As few as eight predictive genes could classify colorectal and gastric tumors with high accuracy, and the signatures were retained despite correcting for the underlying genetic instability. Taken together, our results support a model in which MSI and MSS instabilities target distinct genes and pathways in the development of cancer, and the signatures we have identified are likely legacies reflecting the distinct spectra of alterations, and their downstream effects on patterns of gene expression. These signatures provide new insights into the role of genetic instability in cancer development and progression, and may suggest new strategies for diagnosis and therapeutic intervention.

Acknowledgments

Received 11/22/2004; revised 7/5/2005; accepted 7/26/2005.

Grant support: Stanford Cancer Council (J.R. Pollack).

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 Mike Fero and the staff of the Stanford Functional Genomics Facility for providing high-quality cDNA microarrays, and Gavin Sherlock and Catherine Ball of the Stanford Microarray Database group for providing outstanding database support.
Expression Signature of Genetic Instability


