Elucidation of Epigenetic Inactivation of SMAD8 in Cancer Using Targeted Expressed Gene Display

Kuang-hung Cheng,1,2 Jose F. Ponte,1 and Sam Thiagalingam1,2,3

Departments of 1 Medicine (Genetics Program, Pulmonary Center and Cancer Research Center), 2 Pathology and Laboratory Medicine, and 3 Genetics and Genomics, Boston University School of Medicine, Boston, Massachusetts

ABSTRACT

To address the challenge of identifying related members of a large family of genes, their variants and their patterns of expression, we have developed a novel technique known as targeted expressed gene display. Here, we demonstrate the general application of this technique by analyzing the SMAD genes and report that the loss of SMAD8 expression is associated with multiple types of cancers, including 31% of both breast and colon cancers. Epigenetic silencing of SMAD8 expression by DNA hypermethylation in cancers directly correlates with loss of SMAD8 expression. The SMAD8 alteration in a third of breast and colon cancers makes it a significant novel tumor marker as well as a potential therapeutic target. The utility of targeted expressed gene display for the analysis of highly homologous gene families as demonstrated by its application to the SMAD genes suggests that it is an efficient tool for the identification of novel members, simultaneous analysis of differential expression patterns, and initial discovery of alterations of expressed genes.

INTRODUCTION

Methods such as reverse transcription-PCR (RT-PCR), cDNA subtraction, differential display, representational difference analysis, serial analysis of gene expression, and microarrays have been widely used in the identification of novel transcripts as well as in the assessment of their levels of expression in development and various cellular processes and diseases, including cancer. Despite the usefulness of these techniques in the overall assessment of genes that are highly divergent at the DNA sequence, accurate and high-throughput evaluation and discovery of related members of a gene family have remained a challenge. These methods have generally been unable to discriminate between different members of the gene families with consistency because of the inherent redundancy in DNA sequence among these unique genes and transcripts. A novel method described here, targeted expressed gene display (TEGD), validated using the SMAD family of genes as the prototype enabling one to overcome this dilemma when gene family members contain at least two regions of homology separated by a divergent region of variable length.

The discovery of the SMAD family of signal transducer proteins as mediators of tumor growth factor (TGF-β) and TGF-β-like cytokine-mediated signaling from the cell membrane to the nucleus has revolutionized the understanding of the molecular basis of the signaling and inactivation of TGF-β/BMP pathways in cancer (1). To date, eight human homologues of the SMAD genes have been identified and are classified into three distinct classes based on their structures and biological functions (1, 2). The first category consists of pathway-restricted or receptor-regulated Smads (R-Smads): Smad1, Smad5, and Smad8, which are involved in bone morphogenetic protein (BMP) signaling and Smad2 and Smad3, which are TGF-β/actin pathway restricted. The SMAD8 gene, which displays a high degree of homology to the SMAD1 and SMAD5 genes, was originally described as MADH6 in human, often referred to as SMAD9 and currently listed as MADH9 (HUGO ID) in EnsEMBL at the genomic location 35220321 to 35292902 bp on chromosome 13 (1–3). The R-Smads are activated directly via phosphorylation by type I (RI) receptors after the formation of a complex consisting of the ligand-bound heteromeric RI/RII receptors. Phosphorylated R-Smads interact with the second class of Smads known as the common mediator Smad (Co-Smad) to form a heteromeric complex (4). Smad4 is the only member of this class of Smads known in mammals. The third class of Smads includes Smad6 and Smad7, which were identified as anti-Smads or inhibitory Smads (I-Smads) because of their ability to act as inhibitors of the signaling pathway (5–7).

Because the signaling pathways mediated by the members of the TGF-β family are implicated in a number of biological processes, including cell differentiation, cell proliferation, determination of cell fate during embryogenesis, cell adhesion, cell death, angiogenesis, metastasis, and immunosuppression, it is conceivable that genetic or epigenetic anomalies leading to altered expression patterns of various Smad molecules could contribute to different aspects of neoplastic progression (2, 8–11). Although there has been significant progress in elucidating the association between genetic alterations in the SMAD4 gene and cancer, the nature of defects involving the other Smads has been elusive (12–17). A study aimed at analyzing the expression of Smads in 14 colon tumors compared with normal tissue at the protein level using immunohistochemistry suggested that the levels were increased for Smad2, Smad3, or Smad5, only occasionally increased for Smad1 or Smad8, and there were no apparent differences in Smad4 (18). The apparent lack of genetic alterations in the majority of SMAD genes analyzed thus far in cancer provides compelling support for the potential role of epigenetic alterations, whereby abnormalities in signaling could occur at the level of regulation of gene expression or processing of the transcripts (19–21). Our analysis of the SMAD genes provides evidence for the exploitation of the novel TEGD method described in this article in the initial determination of the potential for an epigenetic mode of inactivation of the SMAD genes in cancer. On the basis of these observations, we predict that the effective utilization of the method described here will find wide use not only in the discovery of novel members of a family of genes and splice variants of a specific gene but also for the simultaneous analysis of the transcript levels of individual genes or their spliced variants in various diseases and during development.

MATERIALS AND METHODS

Cell Culture, RNA Isolation, and cDNA Synthesis. Cancer cell lines were purchased from American Type Culture Collection or the Coriell Cell Repository and culture conditions were followed as suggested by the provider. Tumor samples, some of the cell lines, and their derivatives or nucleic acids isolated from the samples used in this study were obtained from Subra Kugathasan (Medical College of Wisconsin), Peter Thomas (Boston University School of Medicine), Douglas Faller (Boston University School of Medicine), Ramon Parsons (Columbia University), and Kornelia Polyak (Dana-Farber Cancer Institute).
ELUCIDATION OF SMAD8 SILENCING IN CANCER USING TEGD

Cancer Institute). RNA isolation and cDNA synthesis from the cell lines and tumor samples were carried out using previously described procedures (22). Degenerate RT-PCR of the SMAD Genes. On the basis of the amino acid sequences of the human Smads 1–8, regions that are identical and conserved (MH1 and MH2) among the Smads were mapped out (1, 2). The residues targeted for primer design were localized to the MH1 and MH2 domains, and the intervening linker regions were highly divergent enabling the generation of PCR products that are of unique size to correspond to specific Smad homologues. The forward and reverse primers were designed based on the maintenance of codon degeneracy and the representation of the various amino acids at a given position among the known Smad family members as determined from the sequence alignment of the various homologues. All primers were obtained from Integrated DNA Technologies (Corvalle, IA).

The SMAD family-specific degenerate primers used for TEGD are as follows: SmadXF2 (3’-primer), CTCAVCCYTTSRCAAATCTBT (codes for mixing of bases to generate degeneracy: R = A,G; Y = C,T; M = A,C; K = G,T; S = C,G; W = A,T; H = A,C,T; V = A,C,G; D = A,G,T; and N = A,C,G,T). The SmadXF2 primer is inclusive of amino acid residues with a consensus RWPD within the all of the Smads. The nucleotide positions for the RWPDL in the different Smads are in parentheses: SMAD1 (271–292); SMAD2 (414–435); SMAD3 (276–297); SMAD4 (294–315); SMAD5 (279–300); SMAD6 (690–711); SMAD7 (486–507); and SMAD8 (285–306). The SmadXR1 primer is inclusive of amino acid residues with a consensus SFVKG within all of the Smads. The nucleotide positions for the SFVKG in the different Smads are in parentheses: SMAD1 (1242–1253); SMAD2 (1248–1259); SMAD3 (1133–1154); SMAD4 (1509–1530); SMAD5 (1242–1263); SMAD6 (1398–1419); SMAD7 (1194–1215); and SMAD8 (1137–1158).

A 20-µl PCR reaction mixture contained 67 mM Tris-HCl (pH 8.8), 16.6 mM ammonium sulfate, 7 mM magnesium chloride, 1 mM β-mercaptoethanol, 6% DMSO, 100 mM each of dATP, dGTP, dCTP, dTTP, 100 mM ammonium sulfate, 6.7 mM magnesium chloride, 1 mM β-mercaptoethanol, 6% DMSO, 100 mM each of dATP, dGTP, dCTP, dTTP, 100 mM ammonium sulfate, 6.7 mM magnesium chloride, 1 mM β-mercaptoethanol, 6% DMSO, 100 mM each of dATP, dGTP, dCTP, dTTP, and radioactive dCTP [0.25 µl of α[32P]-dCTP (10 mCi/µl); Amersham] for labeling, 20 µM each of the primers, 50 ng of cDNA template, and 2.5 units of Platinum Taq (Invitrogen). An initial denaturation at 94°C for 2 min was followed by 30 cycles, each carried out at 94°C for 30 s, 57°C for 30 s, and 72°C for 10 min. A final extension cycle at 72°C for 10 min.

TEGD Gel Electrophoresis and Recovery of DNA Bands. The samples from the degenerate RT-PCR of the SMAD genes were loaded onto a 4.5% denaturing polyacrylamide gel after a 2 min denaturation step at 95°C. Electrophoresis was performed in a GenomyxLR analyzer (Beckman Coulter) for 20 h. After treatment, the modified DNA was purified using a SV100 kit (Stratagene). The purified DNA was sequenced using the ABI Prism 377 DNA Sequencer (Perkin Elmer Applied Biosystems). The sequence data were analyzed using the Genetyx program (Beckman Coulter). The cycle sequencing procedure used in these studies used 32P-dideoxyxynucleotide triphosphates (dNTPs) (Amersham) along with the Thermosequanaise kit (USB, Cincinnati, OH) as described previously (22).

Genomic DNA Isolation. Genomic DNA from cell lines and tumors was isolated using the DNeasy Tissue kit (Qiagen) according to the manufacturer’s instructions.

Homozygous Deletion Analysis of SMAD8. Radiolabeled microsatellite markers, D13S927 and D13S928, that are localized to the beginning and end, respectively, of the SMAD8 gene in its genomic contig and gene-specific primers localized to the 5′-untranslated region (SMAD8 UTR-F: 5′-GAAA- CATGTTGAGAACACGACGC-3′; and SMAD8 UTR-R: 5′-CGAGACAGCG-GTGCACCGACGC-3′) and encompassing the first exon (SMAD8 EX-1F: 5′-GCTGTGTTCTTGGTCTAGGCTG-3′; and SMAD8 EX-1R: 5′-GTGT-TCCGTGTTGGCATTCAGGC-3′) of the SMAD8 gene were used in PCR amplifications and gel electrophoretic analysis to determine deletion of this genomic region (13).

Analysis of Gene Expression Using Semiquantitative RT-PCR. Total RNA prepared from samples was used for cDNA synthesis, and PCR amplification was done essentially as described previously (22). The gene specific primer pairs used in the analysis of the indicated specific SMAD genes and the β-actin gene were used to standardize for normalization of the abundance of the various transcripts analyzed are described in Table 1.

The relative abundance of the various SMAD gene-specific PCR products was normalized to β-actin or other unaffected SMADs by comparative abundance of the products using densitometry.

Processing of Genomic DNA for the Evaluation of Methylation Status. For bisulfite sequencing and the methylation-specific PCR (MSP) assay, genomic DNA was isolated from cell lines and primary tumors using the Qiagen DNeasy Tissue kit. Genomic DNA was subjected to a denaturation reaction by incubation with sodium bisulfite essentially as described previously (23). In brief, 0.5–2 µg of genomic DNA were denatured with 2 M NaOH for 10 min, followed by bisulfite modification by treatment with freshly prepared 10 mM hydroquinone and 3 M sodium bisulfite (pH 5.0; Sigma), which converts unmethylated cytosines to uracil but does not change methylated cytosines. Each reaction was overlayed with mineral oil and incubated at 50°C for 16–20 h. After treatment, the modified DNA was purified using a Wizard DNA purification kit (Promega, Madison, WI), followed by desulfonation by treating with 3 M NaOH. The ethanol precipitated purified DNA pellet was dissolved in 30 µl of distilled water.

Bisulfite Sequencing. The promoter region of the SMAD8 gene containing CpG islands was first PCR amplified from bisulfite modified DNA (50–100 ng) using gene-specific primers (5′-GAATATGTGGAGGAATGTTTAGGTAGTG-3′ and 5′-CACTCATCCTTCCTCCCTACCACAAATC-3′), and the product was gel purified. Genomic sequencing of the SMAD8 gene-specific PCR products was then performed to obtain the methylation status of the promoter region.

DNA Sequencing. DNA sequence analysis was performed using the GenomyxLR analyzer (Beckman Coulter). The cycle sequencing procedure used in these studies used 32P-dideoxyxynucleotide triphosphates (dNTPs) (Amersham) along with the Thermosequanaise kit (USB, Cincinnati, OH) as described previously (22).

Table 1 Primers used for semiquantitative reverse transcription-PCR

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer sequence</th>
<th>Annealing temperature</th>
<th>Major product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMAD1</td>
<td>5′-CCACTGGAATGTGGTAGTTTC-3′</td>
<td>58°C</td>
<td>950</td>
</tr>
<tr>
<td>SMAD2</td>
<td>5′-GTAGAAGACATGCTCTCACAATCC-3′</td>
<td>58°C</td>
<td>950</td>
</tr>
<tr>
<td>SMAD3</td>
<td>5′-GGGGCCATGGAGCTGTTGAGTTGC-3′</td>
<td>62°C</td>
<td>700</td>
</tr>
<tr>
<td>SMAD4</td>
<td>5′-GGCAGCCATTTGCTTACGAATC-3′</td>
<td>57°C</td>
<td>1720</td>
</tr>
<tr>
<td>SMAD5</td>
<td>5′-GGTAGAGACCATGCTCTCACAATCC-3′</td>
<td>55°C</td>
<td>1010</td>
</tr>
<tr>
<td>SMAD6</td>
<td>5′-GGACTATGCTCATCCTATGCTAGA-3′</td>
<td>60°C</td>
<td>1130</td>
</tr>
<tr>
<td>SMAD7</td>
<td>5′-GGTCGAGGCTTGGCCAAATGTACC-3′</td>
<td>58°C</td>
<td>910</td>
</tr>
<tr>
<td>SMAD8</td>
<td>5′-GCTATGCTCATCCTACCCACCCCATC-3′</td>
<td>58°C</td>
<td>980</td>
</tr>
<tr>
<td>β-Actin</td>
<td>5′-ACACTGTCATCCATCAGAGG-3′</td>
<td>60°C</td>
<td>600</td>
</tr>
</tbody>
</table>

F, forward; R, reverse.
product was accomplished by using the DNA sequencing primer, 5'-GTAAGTAGGGTTTTTTGT-3', along with 32P dNTPs and the ThermoSequenase kit (USB, Cincinnati, OH) as described previously (22).

**MSP.** The methylation status of the SMAD8 promoter region was also analyzed by MSP with the use of primers designed for the amplification of defined CpG islands containing DNA sequences of either unmethylated or methylated DNA using bisulfite-treated genomic DNA as the template (23). Sequences of the forward (F) and reverse (R) MSP primers to distinguish between the methylated (M) and unmethylated (U) genomic DNA used in this study were as follows: 5'-GATGAGGGTTTTTGTAGT-3' (SMAD8U-F) and 5'-CACAACACTACACACTCATC-3' (SMAD8U-R), and 5'-GACCGAGGCGGTTTACG-3' (SMAD8M-F) and 5'-CGACCCCGGACTCCGCG-3' (SMAD8M-R). PCR conditions were as follows: 94°C for 2 min, 35 cycles of 94°C for 30 s, 58°C for 1 min, 70°C for 1 min, followed by a final extension at 70°C for 10 min. A 10-μl sample of each PCR product was mixed with 1X loading buffer and analyzed by electrophoresis on a nondenaturing 8% polyacrylamide gel and visualized by staining with ethidium bromide.

**5'-Aza-2’ Deoxycytidine and Trichostatin A (TSA) Treatment.** HTB129, MDA MB468, MDAMB231, CaCo2, H441, CCL230, and HT29 cells were incubated in culture medium with and without 5'-Aza-2’ deoxycytidine (Sigma) at a concentration of 1–5 μM for 7 days or with 300 nM TSA for 24 h. To assess the effect of a combination of 5'-Aza-2’ deoxycytidine and TSA, cells were exposed sequentially for 7 days to 5'-Aza-2’ deoxycytidine and then to TSA for an additional 24 h. Total RNA was isolated and SMAD8 expression was determined by RT-PCR using the primers SMAD8-1F: 5'-CAAGGTCA-3' and SMAD8-1R: 5'-GAGGAAGCTGGAG-3'.

**RESULTS**

**TEGD and Signature Banding Pattern (SBP) of the SMADs.** Members of the SMAD family of genes have highly homologous amino acid sequences at their N- and COOH-terminal regions (MH1 and MH2 domains, respectively), which are separated by a highly divergent linker region rich in proline, serine, and threonine (1, 2). These regions may have arisen from divergence because of functional specificities from an ancestral unit of activity that has maintained some degree of evolutionary conservation at the protein level. The examination of the MH domains from various SMAD genes indicated that there are either identical or conserved amino acid residues at defined positions, which is consistent with critical structural features required for the function of these proteins. The sequence conservation at the amino acid level is also reflected at the DNA level. Although the delineation of the alterations in SMADs is essential for the comprehension of the molecular basis of various defective processes, the analysis of defects in individual members in this type of family of genes poses a formidable task for efficient detection in a high-throughput platform. Success in identifying alterations in SMAD genes could be expected to provide critical information necessary for deciphering the molecular basis of their functions. The fact that the SMAD genes contain two distinct highly conserved regions separated by a highly variable intervening linker region allowed us to develop a novel screening strategy to simultaneously analyze all of the known members of this family (Fig. 1A).

![Figure 1. Targeted expressed gene display (TEGD) and tissue-wide expression of SMAD genes. A, schematic representation of TEGD for the SMAD family of genes. MH1 and MH2 indicate highly homologous regions in the amino acid as well as the DNA sequence among the various Smad gene family members. The forward and reverse primers for PCR amplification of the cDNA were designed in the conserved regions as indicated. The radiolabeled PCR products were analyzed by denaturing acrylamide gel electrophoresis. A typical signature banding pattern of the various SMADs is indicated in the bottom panel. B, TEGD analysis of the SMAD family of genes in various tissue types. PCR products for SMADs using degenerate primers were analyzed by TEGD. Lanes 1–17 correspond to PCR products generated using cDNA templates from brain, lung, stomach, heart, liver, spleen, kidney, colon, bone marrow, small intestine, trachea, prostate, uterus, thymus, testis, skeletal muscle, and mammary gland, respectively. The lines on the right hand panel point to distinct PCR products. The approximate size of PCR products in bp is indicated on the left panel. The positions of various SMAD genes and their variants as identified from sequence analysis are indicated on the right hand panel. C, reverse transcription-PCR analysis of SMAD genes. Semiquantitative reverse transcription-PCR analysis of the indicated SMAD genes and SMADs 1 and 2 (data not shown) was carried out as described under “Materials and Methods.” The cDNA template was derived from total RNA from normal tissues of brain, lung, heart, liver, bone marrow, kidney, spleen, thymus, prostate, testis, uterus, small intestine, mammary gland, skeletal muscle, stomach, and colon, Lanes 1–16, respectively.
We have designed degenerate oligonucleotide primers corresponding to the conserved regions of the SMAD family of genes based on the preservation of codon degeneracy and conserved amino acids at a given position among the known Smads for PCR amplification of the cDNA templates. PCR amplification in the presence of radiolabeled nucleotides and the subsequent analysis of the products using denaturing PAGE revealed distinct bands on the gel (Fig. 1, A and B). We recovered the distinct bands corresponding to the PCR products generated using SMAD-specific degenerate primers and sequenced them. The bands corresponding to the 1200-, 960-, 840-, 680-, and 570-bp PCR products were found to be identical to the cDNA sequences for SMAD4, SMAD1 and SMAD5, SMAD2, SMAD3 and SMAD8, and SMAD6 and SMAD7, respectively, as predicted from their estimated sizes and sequences (1, 2; Fig. 1B). These results suggested to us that once the SBP of the TEGD is optimized and established such as in this case with the SMAD family of genes, repeat analysis of gene expression in tissues, or other samples of unknown origin could be easily adopted for a routine high-throughput analysis. Although we generated and analyzed radiolabeled nucleotide incorporated PCR products in these initial studies, one could also achieve the same results using fluorescently or radioactively end-labeled primers for PCR amplification.

**Validation of SMAD Expression Patterns Determined from TEGD.** We confirmed the presence or absence of SMAD expression determined from TEGD using gene specific primers by semiquantitative RT-PCR (Fig. 1C). The expression patterns of the various SMADs detected by TEGD remained consistent with the semiquantitative RT-PCR results. Almost all of the SMADs were expressed in all of the tissue types that we have analyzed. However, SMAD8 expression was lost in the liver and was decreased to barely detectable levels in the bone marrow and uterus (Fig. 1C). These results indicated to us that TEGD could be used as a tool for initial diagnostic high-throughput evaluations to determine SMAD gene expression patterns simultaneously and with a high degree of efficiency. Thus, TEGD can be regarded as a highly improved alternate method that may substitute for the traditional multiplex PCR technique not only because of the increased level of sensitivity with TEGD and ability to discriminate between genes that are closely related at their DNA sequence but also because of the low level of cDNA template required for the analysis.

**Differentially Spliced Variants of the SMADs.** TEGD also enabled us to identify the various differentially spliced forms of the SMAD2, SMAD3, SMAD5, and SMAD8 genes (Fig. 1, B and C; data not shown). Alternatively spliced variants of SMAD2 with a deletion of exon3 (SMAD2Δexon3; SMAD2β), SMAD3 with deletions of both exons 3 and 7 (SMAD3Δexon3 Δexon7; SMAD3β), SMAD5 with a deletion of exon3 (SMAD5Δexon3; SMAD5β), and SMAD8 with deletions of either exon 3 (SMAD8Δexon3; SMAD8β) or both exons 2 and 3 (SMAD8Δexon2Δexon3; SMAD8γ) were detected in our analysis. Although one of these variants (SMAD2Δexon3; SMAD2β) has been previously reported, the experimental evidence for the existence of the others is presented for the first time in this study (24). However, our study did not verify the existence of two previously reported alternatively spliced forms with deletions at the 3′-ends, potentially because of the placement of the TEGD primers inside the affected sequence of these alternatively spliced forms (25, 26). The encoded proteins of Smad2, Smad5, and Smad8 resulting from full-length and variant transcripts that have been described also exhibit differences in their biochemical properties (23–25). The overall significance of the described and predicted novel spliced forms of SMAD8 reported both here and elsewhere in disease phenotypes, including cancer, requires

### Table 2. Altered expression of SMAD8 in cancers

<table>
<thead>
<tr>
<th>Cancer</th>
<th>Total no. of samples</th>
<th>Samples with loss of SMAD8 expression (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast</td>
<td>35</td>
<td>11/35 (31)</td>
</tr>
<tr>
<td>Colon</td>
<td>41</td>
<td>13/41 (31)</td>
</tr>
<tr>
<td>Esophagus</td>
<td>4</td>
<td>0/4 (0)</td>
</tr>
<tr>
<td>Head &amp; Neck</td>
<td>4</td>
<td>2/4 (50)</td>
</tr>
<tr>
<td>Lung</td>
<td>19</td>
<td>1/19 (5)</td>
</tr>
<tr>
<td>Ovary</td>
<td>2</td>
<td>1/2 (50)</td>
</tr>
<tr>
<td>Pancreas</td>
<td>3</td>
<td>2/3 (65)</td>
</tr>
<tr>
<td>Prostate</td>
<td>4</td>
<td>3/4 (75)</td>
</tr>
<tr>
<td>Stomach</td>
<td>4</td>
<td>2/4 (50)</td>
</tr>
</tbody>
</table>
ELUCIDATION OF SMAD8 SILENCING IN CANCER USING TEGD

Fig. 3. Epigenetic gene silencing of SMAD8 because of altered DNA methylation. A, schematic drawing of the landscape of CpG island methylation patterns in the region of genomic DNA from upstream promoter region of the SMAD8 gene in the 5'-untranslated region (5'-UTR). Exon I corresponds to the first exon of the SMAD8 gene [MADVH-002 (Vega_transcript ID)]. The flag represents the ATG corresponding to the first methionine of the predicted peptide. Vertical lines indicate CpG islands in the DNA sequence. ○ represents unmethylated cytosines, whereas ● represents methylated cytosines as determined by bisulfite sequencing. The circles above the horizontal line indicate the methylation pattern observed in the CpG islands of the cell lines that express SMAD8. The circles below the line indicate the methylation pattern of the CpG islands from samples, which lacked SMAD8 expression. The nucleotide sequence of the DNA within the dotted lines is shown with the asterisks, indicating CpG islands. B, bisulfite sequence analysis of the indicated CpG islands of the promoter region of the SMAD8 gene in the cell lines that are either proficient (+) or deficient (−) in SMAD8 expression. Cell lines proficient for SMAD8 expression have no bands in the C lane indicative of conversion of unmethylated cytosines to uracil upon bisulfite treatment. C, methylation-specific PCR (MSP) analysis of the various cancers that have lost or retained SMAD8 expression. The methylation-specific PCR products in lanes U and M indicate the presence of unmethylated and methylated templates, respectively. Plasmid DNA and in vitro methylated DNA serve as positive and negative controls, respectively.

further study (Refs. 24–26; University of California at Santa Cruz, Santa Cruz, CA genome browser).4

TEGD in the Analysis of SMADs in Cancer. The ability to simultaneously probe multiple members of a gene family using TEGD prompted us to apply this technique to analyze differential expression patterns of the various SMADs in cancer to validate its use for diagnostic screening (Fig. 2A). We were able to use the SBPs established within the normal tissues to determine the retention or loss of specific DNA bands corresponding to the defined full-length and variant transcripts (Figs. 1B and 2A). The TEGD analysis of the SMAD genes in cancers leads us to conclude that there is a significant level of loss in the expression of SMAD3 and SMAD8 in colon cancer and of SMAD8 in breast and other cancers (Table 2). These initial observations were additionally validated by analyzing the expression patterns of the SMAD8 gene more carefully using gene specific primers and semiquantitative RT-PCR (Fig. 2B). These results confirmed the TEGD data and provided the first clues to suggest that the SMAD8 gene is a critical target for loss of function due to downregulation of gene expression in 31% of breast and colon cancers (Table 2). The analysis of establishing the significance of the loss of SMAD3 expression in colon cancer will be dealt with in greater detail elsewhere (Cheng and Thiagalingam, unpublished results).

Molecular Mechanism for the Silencing of SMAD8 Expression. From our analysis, loss of expression of the SMAD8 gene was estimated to occur in nearly one-third of both breast and colon cancers, which are two of the leading causes of cancer deaths in women and in general, respectively (Fig. 2B; Table 2). Hence, we investigated potential mechanisms for the loss of SMAD8 gene expression in cancer because of the prevalence of this alteration, which is similar to or even more frequent than known tumor markers. We examined whether genetic alterations such as chromosomal deletions affecting the SMAD8 gene could lead to the loss of its expression by homozygous deletion analyses. We used microsatellite markers corresponding to the SMAD8 gene based on the genomic contig as well as by genomic PCR using primers that amplified the genomic region corresponding to the 5'-untranslated region and the first exon of the SMAD8 gene. These experiments indicated that gross genomic deletions are apparently not the major mechanism of SMAD8 inactivation in the affected cancers (data not shown). Therefore, we considered epigenetic silencing as an alternate mechanism for SMAD8 gene silencing.

The genomic sequence of the translated SMAD8 gene (EnsEMBL

Internet address: http://genome.ucsc.edu.
Gene ID: ENSG00000120693; Vega_transcript ID; MADH9-002; exons: 6; transcript length: 2036 bp; chromosome 13: nucleotides 35220321–35251966) and the upstream region was inspected for the presence of CpG islands that may be the targets of DNA hypermethylation and associated chromatin modification effects for their involvement in the silencing of SMAD8 gene expression. Several CpG islands in the upstream promoter of the SMAD8 gene were tested as likely candidate regions that could be critical for differential DNA methylation patterns coinciding with the loss of SMAD8 expression (data not shown, Fig. 3). DNA sequence analysis of the bisulfite-treated genomic DNA revealed that CpG islands localized to nucleotides 35292332 to 35292369 of the promoter region (exon 1 of the untranslated hypothetical transcript MADH9-001 (Vega_transcript ID: OTTHUMT00013001062); chromosome 13q12-14 on the reverse strand between RB and BRCA2) of the SMAD8 gene are only methylated in cancers that exhibit loss of expression (UCSC genome browser; Fig. 5A). DNA sequence analysis of the bisulfite-treated genomic DNA revealed that CpG islands localized to nucleotides 35292323 to 35292369 of the promoter region (exon 1 of the untranslated hypothetical transcript MADH9-001 (Vega_transcript ID: OTTHUMT00013001062); chromosome 13q12-14 on the reverse strand between RB and BRCA2) of the SMAD8 gene are only methylated in cancers that exhibit loss of expression (UCSC genome browser; Fig. 3). DNA sequence analysis of the bisulfite-treated genomic DNA revealed that CpG islands localized to nucleotides 35292332 to 35292369 of the promoter region (exon 1 of the untranslated hypothetical transcript MADH9-001 (Vega_transcript ID: OTTHUMT00013001062); chromosome 13q12-14 on the reverse strand between RB and BRCA2) of the SMAD8 gene are only methylated in cancers that exhibit loss of expression (UCSC genome browser; Fig. 3). DNA sequence analysis of the bisulfite-treated genomic DNA revealed that CpG islands localized to nucleotides 35292332 to 35292369 of the promoter region (exon 1 of the untranslated hypothetical transcript MADH9-001 (Vega_transcript ID: OTTHUMT00013001062); chromosome 13q12-14 on the reverse strand between RB and BRCA2) of the SMAD8 gene are only methylated in cancers that exhibit loss of expression (UCSC genome browser; Fig. 3). DNA sequence analysis of the bisulfite-treated genomic DNA revealed that CpG islands localized to nucleotides 35292332 to 35292369 of the promoter region (exon 1 of the untranslated hypothetical transcript MADH9-001 (Vega_transcript ID: OTTHUMT00013001062); chromosome 13q12-14 on the reverse strand between RB and BRCA2) of the SMAD8 gene are only methylated in cancers that exhibit loss of expression (UCSC genome browser; Fig. 3). DNA sequence analysis of the bisulfite-treated genomic DNA revealed that CpG islands localized to nucleotides 35292332 to 35292369 of the promoter region (exon 1 of the untranslated hypothetical transcript MADH9-001 (Vega_transcript ID: OTTHUMT00013001062); chromosome 13q12-14 on the reverse strand between RB and BRCA2) of the SMAD8 gene are only methylated in cancers that exhibit loss of expression (UCSC genome browser; Fig. 3).

DNA Hypermethylation and SMAD8 Expression in Cancers. To directly determine the physiological significance, the role(s) of apparent epigenetic DNA methylation by itself or in combination with histone acetylation/deacetylation on differential regulation of SMAD8 expression in cancers was examined. We chose six cell lines derived from breast, colon, and lung cancers (HTB129, HT29, CaCo2, CCL253, MDAMB468, and H441) that exhibited loss of SMAD8 expression and one cell line (MDAMB231) that retained SMAD8 expression as a control and examined the effects of 5’-aza-2’-deoxycytidine (a DNA demethylating agent) and/or TSA (an inhibitor of histone deacetylation) on SMAD8 expression. A substantial increase in SMAD8 expression was observed with 5’-aza-2’-deoxycytidine treatment in all of the cell lines, which were previously determined to exhibit DNA hypermethylation-mediated gene silencing of SMAD8 (Fig. 4A). TSA by itself caused only a slight increase in the levels of the transcript in two of the tested cell lines (CaCo2 and CCL253) but had no effect on the majority of the tested cell lines. However, there was a slight up-regulation of SMAD8 expression in the presence of both drugs (Fig. 4A). MSP analysis and bisulfite-modified DNA sequence analysis of the target CpG islands in the promoter region of the SMAD8 gene that were differentially methylated in affected and control cell lines revealed that demethylation due to 5’-aza-2’-deoxycytidine treatment accompanies a corresponding increase in SMAD8 expression (Figs. 4, A and B; data not shown). These observations strongly support the notion that the loss of SMAD8 expression in cancers is primarily mediated by hypermethylation of cis-regulatory CpG islands of the gene.

DISCUSSION

The analysis of highly homologous members of a family of genes to detect and establish differential gene expression patterns as well as the genetic or epigenetic alterations responsible for cancer and other diseases with limited amounts of clinical sample has remained a formidable task. Efficient methods to simultaneously analyze the closely related yet functionally divergent genes belonging to families would not only be important in accurate diagnosis and prognostic evaluation of a disease but could also be exploited for the identification of pharmacogenetic targets to customize therapy. We propose that the TEGD technique described in this article can be effectively used to analyze families of genes that contain at least two stretches of conserved regions, which are separated by a divergent linker region of variable length. TEGD provides a distinct advantage over techniques such as differential display, a comparable methodology that has been adopted for the simultaneous analysis of multiple genes because of the latter’s inability to detect differential gene expression patterns of targeted and defined genes. Furthermore, even an improved version of differential display designed to analyze related genes (e.g., kinases) fell short of efficiently establishing distinct expression patterns of the related genes and failed to identify novel but related genes with different functional roles (27–30).

On the other hand, with TEGD, once a SBP of the TEGD is optimized and established with an array of different normal tissues such as in this case with the SMAD family of genes, repeat analysis of gene expression of samples of unknown origin could be easily carried out in a routine high throughput manner (Figs. 1 and 2). We believe that the TEGD technique should sufficiently address the dilemma of efficient simultaneous expression pattern analysis of related genes with relatively minute amounts of samples in clinical and investigational research settings. The development of an algorithm to predict the suitability of the applications of TEGD based on the presence of two distinct homologous regions separated by an intervening variable region that would enable the establishment of SBPs from the available sequences of already identified genes or expressed tag sequences is in progress. We believe that TEGD has the potential to advance the ability to probe gene families for genetic and epigenetic defects to a new level of sophistication and could be adopted for routine use in the future. The application of the TEGD technique to simultaneously analyze multiple members of the SMAD family of genes has not only...
Fig. 5. A model for the Smad8 connection to cancer. Bone morphogenetic protein (BMP) signaling is initiated by the association between the BMPs and type I (RI) and type II (RII) heteromeric tetrameric receptors. It is followed by the phosphorylation of the type I receptor (RI) kinase that in turn phosphorylates the receptor-regulated Smads (R-Smad), such as Smad8 and initiates the signaling events. The phosphorylated Smad8 forms a heteromeric complex with the common-mediator Smad, Smad4, and is translocated into the nucleus. In the nucleus, the Smad8/Smad4 hetero-oligomer either by itself or by associating with heterologous Smad-interacting DNA binding proteins or other cofactors, could mediate specific transcriptional activation or repression responses. The inhibitory Smads (I-Smad) such as Smad6 and Smad7 are able to compete with the R-Smads by stably binding the RI kinase or by preventing association of R-Smads with the common-mediator Smad, effectively blocking the signaling cascade. There are numerous other signaling pathways such as the Ras-MEK pathway that could also modulate the end-effects by establishing cross-talk among the different pathway members. BMP signaling is implicated in tumor suppression, bone homeostasis, angiogenesis, and metastasis.

validated the enormous advantage of the technique as an initial diagnostic tool but also has illustrated an efficient way to identify novel genes that are closely related at the level of their nucleotide sequence, to identify splice variants of a gene as well as to detect their altered expression patterns.

Our survey of the various SMAD genes using the novel TEGD technique described in this article enabled us to obtain the first clues in identifying the SMAD8 gene as an important target for loss of expression in multiple types of cancers, including ~31% of breast and colon cancers (Table 2). This level of alteration is even more frequent than that of the SMAD4 gene, the most frequent target for genetic inactivation of the known Smad signaling genes in colon cancer. It is also more frequent than the HER/neu gene amplification, the most celebrated tumor marker for breast cancer, which occurs in ~20–30% of breast cancer cases (31). The failure of a previous study to provide evidence for an association between the levels of Smad8 expression and colon cancer could be largely because of the unavailability of an antibody with a high level of specificity to Smad8 at the time of the study (18).

The data presented in this article also provides the first direct evidence that silencing of gene expression via DNA hypermethylation of the SMAD8 gene could be an important event in tumorigenesis of several cancers, including one-third of breast and colon cancers. It is interesting to note that Smad8 is apparently the major target for loss of function among the SMAD genes in breast cancer and is an R-Smad, which becomes phosphorylated during BMP signaling events and modulates BMP-responsive genes, including those that may affect bone homeostasis (Refs. 32–36; Fig. 5). Additionally, Smad signaling events via the BMP cytokines are also implicated in other signaling events that regulate biological processes, including cell differentiation, proliferation, determination of cell fate during embryogenesis, cell adhesion, cell death, angiogenesis, metastasis, and immunosuppression (1, 2; Fig. 5). Although it is intriguing that metastasis to bone is often associated with advanced stage breast and other cancers, additional studies would be required to understand whether metastatic breast cancer cells defective in Smad8 signaling could be responsible for causing an imbalance in normal bone homeostasis by enhancing osteoclastic bone resorption, leading to osteolytic lesions within the bone (37–40).

Although inactivation of the SMAD2 and SMAD4 genes due to intragenic mutations and homozygous deletions has been reported in ~20% of colorectal cancers, evidence for genetic or epigenetic inactivation of other SMAD gene targets at significant levels had remained elusive until this study (2, 22). The loss of expression of SMAD8 in ~31% of colon cancers is more significant than any other Smad alteration known to date. Determination of whether the affected cells play a critical role in tumorigenesis by a mechanism similar to that in breast cancer requires further study. Interestingly, the presence of germ-line mutations in the BMP receptor 1A in juvenile polyposis, which increase the risk of developing gastrointestinal cancers, suggests that inactivation of BMP signaling may play a critical role in colon cancer (41, 42). Although the elucidation of BMP-mediated signaling pathways in which Smad8 is a critical mediator is still in its infancy, these studies clearly provide the incentive for additional investigations that may help gain a better understanding of the effects of Smad8 inactivation in cancer and could pave the way for the exploration of its potential use in diagnosis, prognosis, and for targeting in the development of therapeutic modalities.

ACKNOWLEDGMENTS

We thank our colleagues, Subra Kugathasan (Medical College of Wisconsin), Peter Thomas (Boston University School of Medicine), Douglas Faller (Boston University School of Medicine), Ramon Parsons (Columbia University), and Kornelia Polyak (Dana-Farber Cancer Institute) for generously donating reagents and cell lines and Andrea Russo and the anonymous reviewers for critical reading of the manuscript.

REFERENCES

Elucidation of Epigenetic Inactivation of *SMAD8* in Cancer Using Targeted Expressed Gene Display

Kuang-hung Cheng, Jose F. Ponte and Sam Thiagalingam

*Cancer Res* 2004;64:1639-1646.

Updated version  Access the most recent version of this article at:  
http://cancerres.aacrjournals.org/content/64/5/1639

Cited articles  This article cites 42 articles, 14 of which you can access for free at:  
http://cancerres.aacrjournals.org/content/64/5/1639.full#ref-list-1

Citing articles  This article has been cited by 8 HighWire-hosted articles. Access the articles at:  
http://cancerres.aacrjournals.org/content/64/5/1639.full#related-urls

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

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

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