SMAD2, SMAD3 and SMAD4 Mutations in Colorectal Cancer

Nicholas I. Fleming¹, Robert N. Jorissen¹, Dmitri Mouradov¹, Michael Christie¹,⁴, Anuratha Sakthianandeswaren¹, Michelle Palmeri¹, Fiona Day¹,⁴, Shan Li¹, Cary Tsui¹, Lara Lipton¹,⁴,⁶, Jayesh Desai¹,⁴, Ian T. Jones¹, Stephen McLaughlin⁶, Robyn L. Ward⁶, Nicholas J. Hawkins¹⁰, Andrew R. Ruszkiewicz¹¹, James Moore¹², Hong-Jian Zhu⁵, John M. Mariadason², Antony W. Burgess³, Dana Busam¹³, Qi Zhao¹⁴, Robert L. Strausberg¹⁴,¹⁵, Peter Gibbs¹,⁴,⁶, and Oliver M. Sieber¹,⁴

Introduction

The TGF-β family of cytokines are growth inhibitors of normal epithelial cells, and loss of sensitivity to these factors promotes tumorigenesis (1). Canonical TGF-β signaling is mediated by the TGF-β type I and type II receptors (TGFBR1, TGFBR2), which upon binding TGF-β ligands, phosphorylate the receptor-activated SMADs (R-SMADs), SMAD2 and SMAD3, at conserved C-terminal Ser-Ser-X-Ser motifs (2, 3). The activated R-SMADs bind the common mediator SMAD (co-SMAD) SMAD4 (4–6) and the resulting complexes relocate into the nucleus. There the proteins participate in transcriptional regulation of target genes in conjunction with a wide range of coregulator proteins (7).

Mutation of TGF-β pathway members is common in multiple types of human malignancies including colorectal cancer (CRC; ref. 1). Germline mutations in SMAD4 cause juvenile polyposis syndrome (JPS), an autosomal dominant predisposition to multiple gastrointestinal polyps and cancer (8). In sporadic CRCs, SMAD4 and TGFBR2 mutations are found in approximately 10% and 15% of patients, respectively (9–12). TGFBR2 mutations are particularly prevalent in microsatellite unstable (MSI) tumors, with approximately 80% of such cases harboring frameshift mutations at a poly-adenosine tract in exon 4 (13).

The SMAD proteins are highly homologous and characterized by 2 conserved regions, an N-terminal Mad homology domain-1 (MH1) and a C-terminal Mad homology domain-2 (MH2), joined by a linker domain (7). The MH1 domain

Abstract

Activation of the canonical TGF-β signaling pathway provides growth inhibitory signals in the normal intestinal epithelium. Colorectal cancers (CRCs) frequently harbor somatic mutations in the pathway members TGFBR2 and SMAD4, but to what extent mutations in SMAD2 or SMAD3 contribute to tumorigenesis is unclear. A cohort of 744 primary CRCs and 36 CRC cell lines were sequenced for SMAD2, SMAD4, and SMAD3 and analyzed for allelic loss by single-nucleotide polymorphism (SNP) microarray analysis. Mutation spectra were compared between the genes, the pathogenicity of mutations was assessed, and relationships with clinicopathologic features were examined. The prevalence of SMAD4, SMAD2, and SMAD3 mutations in sporadic CRCs was 8.6% (64 of 744), 3.4% (25 of 744), and 4.3% (32 of 744), respectively. A significant overrepresentation of two genetic hits was detected for SMAD4 and SMAD3, consistent with these genes acting as tumor suppressors. SMAD4 mutations were associated with mucinous histology. The mutation spectra of SMAD2 and SMAD3 were highly similar to that of SMAD4, both in mutation type and location within the encoded proteins. In silico analyses suggested the majority of the mutations were pathogenic, with most missense changes predicted to reduce protein stability or hinder SMAD complex formation. The latter altered interface residues or disrupted the phosphorylation-regulated Ser-Ser-X-Ser motifs within SMAD2 and SMAD3. Functional analyses of selected mutations showed reductions in SMAD3 transcriptional activity and SMAD2–SMAD4 complex formation. Joint biallelic hits in SMAD2 and SMAD3 were overrepresented and mutually exclusive to SMAD4 mutation, underlining the critical roles of these three proteins within the TGF-β signaling pathway. Cancer Res; 73(2); 725–35. ©2012 AACR.
facilitates nuclear import, DNA and transcriptional coregulator binding, and negatively regulates the MH2 domain (14, 15). The MH2 domain is involved in SMAD protein homo- and hetero-oligomerization (4–6), cytoplasmic anchoring (16), and transcription (reviewed in ref. 7). When inactive, the SMAD4 and R-SMAD proteins reside in the cytoplasm and when activated by TGF-β signaling, it is proposed that the proteins form heterotrimers through direct interactions containing 1 SMAD4 and 2 R-SMAD proteins (4–6). Cancer-associated mutations in SMAD4 show a characteristic distribution with respect to its domain structure. The majority of changes cluster in the MH2 domain and often alter residues that are close to the protein interface mediating SMAD4 hetero-oligomerization with the R-SMADs (11, 12). Mutations in the MH1 domain have been reported to alter protein stability, alter DNA binding, prevent nuclear translocation, and enhance interactions with the MH2 domain (14, 15, 17).

Despite the central role of SMAD2 and SMAD3 as direct mediators of TGF-β signaling and binding partners for SMAD4, evidence for mutations affecting these proteins in CRCs is limited. The mutation prevalence has been broadly estimated at 2% to 6%, but with the exception of a recent study from The Cancer Genome Atlas (TCGA) program (18), the sample cohorts used to determine this have been small with only very few identified mutations (19, 20). Importantly, a series of studies using mouse models of intestinal cancer have indicated tumor suppressor roles for Smad2 and Smad3 (21–23). It remains unclear therefore to what extent SMAD2 and SMAD3 mutations are alternative functional genetic hits in the TGF-β pathway and consequently contribute to the development of CRCs.

In this study, we determined the prevalence, spectra, and LOH status of SMAD4, SMAD2, and SMAD3 somatic mutations in 744 sporadic and 36 CRC cell lines. The distribution and nature of the SMAD2 and SMAD3 mutations were compared with those of SMAD4. The pathogenicity of missense mutations was estimated in silico using homology and crystal structure information, and for selected cases, was tested in vitro using assays for transcriptional function and heteromeric complex formation. Relationships with patient clinical characteristics were examined. Our results reveal that the mutation spectra of SMAD2 and SMAD3 are highly similar to that of SMAD4 and suggest that joint inactivation of SMAD2 and SMAD3 constitutes a novel mode of TGF-β pathway inactivation in CRCs.

Materials and Methods

Patients

Fresh-frozen tumor and normal specimens were analyzed from 744 patients with sporadic CRCs treated at St Vincent’s Hospital, Sydney (134 patients), the Royal Melbourne and Western Hospitals, Melbourne (369 patients), and the Royal Adelaide Hospital, Adelaide (241 patients), Australia. Informed consent was given by all participants according to local ethics regulations. Information on patient clinical characteristics was retrieved from hospital databases. The primary cancers comprised 70 stage I, 228 stage II, 347 stage III, and 99 stage IV cases. Three hundred and thirteen cancers were from the proximal colon, 241 from the distal colon, and 189 from the rectum; tumor location data were unavailable for one patient. The median age at cancer diagnosis was 69.2 years and ranged from 25 to 99 years; 414 patients were male and 330 were female (Table 1).

CRC cell lines

A total of 36 colon cancer cell lines were studied: CACO2, COLO201, COLO320, COLO741, DLD1, Gp5d, HCA7, HCC2998, HCT116, HDC135, HDC143, HDC57, HDCC114, HT29, HT55, KM12, LIM1215, LIM1863, LIM1899, LIM2099, LIM2405, LIM2550, LIM2551, LOVO, LS174T, RKO, RW2982, SKCO1, SW1116, SW1222, SW403, SW48, SW480, SW837, SW948, and T84. The cell lines were authenticated in 2010 by short tandem repeat (STR) analysis. Cells were cultured with Dulbecco’s modified Eagle’s medium (DMEM) and 10% FBS at 37°C and 5% CO2. Literature references for the cell lines are provided in Supplementary Table S1.

Mutation detection

For cancer samples, hematoxylin and eosin–stained sections were reviewed by an anatomical pathologist and macrodissected in areas comprising greater than 60% neoplastic cells. Genomic DNA was extracted using the AllPrep DNA/RNA Mini Kit (Qiagen). Coding regions and exon–intron boundaries of the entire SMAD4, SMAD2, and SMAD3 genes were amplified using the PCR, and direct DNA sequencing was conducted using BigDye Terminator v3.1 Ready Reaction Mix (Applied Biosystems). Primer sequences are available from the authors. Reaction products were run on 3730xl DNA Analyzers (Applied Biosystems) using Biomek FX robots (Beckman-Coulter) and Pixsys 4200 nanoliter liquid handling systems (Cartesian Technologies). Mutations detected were confirmed as somatic by bidirectional resequencing of new PCR products from tumor and matched normal DNA or cell line DNA.

LOH and DNA copy number analysis

LOH and DNA copy number status at the SMAD4, SMAD2, and SMAD3 loci was determined from single-nucleotide polymorphism (SNP) analyses (Human610-Quad BeadChip arrays, Illumina) for cell lines, tumor, and matched normal DNA samples using OncoSNP software as described previously (24). SNP call rates for normal samples were greater than 98% (median, 99.8%; range, 98.3%–99.9%); the median call rate for tumor and cell line samples was 97.3% (range: 83.7%–99.8%) and 95.1% (range: 87.9%–99.8%), respectively. Correct pairing of tumor and normal samples was verified from SNP genotype matching. For the cell lines, where normal reference DNA was unavailable, LOH was assigned as present for regions where contiguous SNP homozygosity extended for more than 2 Mb. For assignment of CIN status, copy number for individual autosomes was estimated by calculating the mode of absolute DNA copy number states across SNPs. Samples for which more than 3 autosomes showed deviations from 2n were classified as CIN positive.
Expression constructs for N-terminal HA-tagged SMAD2, untagged SMAD3 and N-terminal FLAG-tagged SMAD4 were generated in pcDNA3.1+ (Invitrogen) by PCR subcloning from the Mammalian Genome Collection clones MGC34440, MGC60396, and MGC8602, respectively (Dana-Farber/Harvard Cancer Center DNA Resource Core, Boston, MA). C-terminal HA-tagged TGFBR1 was generated from a previously published FLAG-tagged version (25). Site-directed mutagenesis was conducted using the QuikChange II XL kit (Agilent Technologies), and the following missense mutations were produced: S276L, D300N, P305L, T413N, D450E, and S647P for SMAD2; D258N, R268C, P336S, R373Q, D408Y, and S425C for SMAD3; and T204D for HA-TGFBR1.

Table 1. Characteristics of 744 CRC patients and associations with SMAD4, SMAD2, or SMAD3 mutation status

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>All patients</th>
<th>SMAD4</th>
<th>SMAD2</th>
<th>SMAD3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Wild-type (%)</td>
<td>Mutant (%)</td>
<td>P</td>
</tr>
<tr>
<td>Total N</td>
<td>744</td>
<td>680 (91.4)</td>
<td>64 (8.6)</td>
<td></td>
</tr>
<tr>
<td>Age, y</td>
<td></td>
<td>69.1 ± 11.3</td>
<td>69.3 ± 11.5</td>
<td>67.5 ± 9.3</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>414</td>
<td>384 (92.8)</td>
<td>30 (7.2)</td>
<td>0.149</td>
</tr>
<tr>
<td>Female</td>
<td>330</td>
<td>296 (89.7)</td>
<td>34 (10.3)</td>
<td></td>
</tr>
<tr>
<td>Tumor location</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right colon</td>
<td>313</td>
<td>282 (90.1)</td>
<td>31 (9.9)</td>
<td>0.274</td>
</tr>
<tr>
<td>Left colon</td>
<td>241</td>
<td>219 (90.9)</td>
<td>22 (9.1)</td>
<td></td>
</tr>
<tr>
<td>Rectum</td>
<td>189</td>
<td>178 (94.2)</td>
<td>11 (5.8)</td>
<td></td>
</tr>
<tr>
<td>AJCC stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>70</td>
<td>66 (94.3)</td>
<td>4 (5.7)</td>
<td>0.384</td>
</tr>
<tr>
<td>II</td>
<td>228</td>
<td>206 (90.4)</td>
<td>22 (9.6)</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>347</td>
<td>321 (92.5)</td>
<td>26 (7.5)</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>99</td>
<td>87 (87.9)</td>
<td>12 (12.1)</td>
<td></td>
</tr>
<tr>
<td>N stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>326</td>
<td>296 (90.8)</td>
<td>30 (9.2)</td>
<td>0.732</td>
</tr>
<tr>
<td>1</td>
<td>255</td>
<td>236 (92.5)</td>
<td>19 (7.5)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>163</td>
<td>148 (90.8)</td>
<td>15 (9.2)</td>
<td></td>
</tr>
<tr>
<td>Lymphovascular invasion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>289</td>
<td>260 (90)</td>
<td>29 (10)</td>
<td>0.338</td>
</tr>
<tr>
<td>Present</td>
<td>206</td>
<td>191 (92.7)</td>
<td>15 (7.3)</td>
<td></td>
</tr>
<tr>
<td>Not available</td>
<td>249</td>
<td>229</td>
<td>20</td>
<td>242</td>
</tr>
<tr>
<td>Differentiation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well/moderate</td>
<td>546</td>
<td>498 (91.2)</td>
<td>48 (8.8)</td>
<td>0.754</td>
</tr>
<tr>
<td>Poor</td>
<td>171</td>
<td>158 (89.4)</td>
<td>13 (7.6)</td>
<td></td>
</tr>
<tr>
<td>Not available</td>
<td>27</td>
<td>24</td>
<td>3</td>
<td>26</td>
</tr>
<tr>
<td>Mucinous</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>578</td>
<td>540 (93.4)</td>
<td>38 (6.6)</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Yes</td>
<td>157</td>
<td>133 (84.7)</td>
<td>24 (15.3)</td>
<td></td>
</tr>
<tr>
<td>Not available</td>
<td>9</td>
<td>7</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>MSI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stable</td>
<td>646</td>
<td>588 (91.0)</td>
<td>58 (9)</td>
<td>0.441</td>
</tr>
<tr>
<td>Unstable</td>
<td>96</td>
<td>92 (93.9)</td>
<td>6 (6.1)</td>
<td></td>
</tr>
<tr>
<td>CIN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>167</td>
<td>150 (98.8)</td>
<td>17 (10.2)</td>
<td>0.759</td>
</tr>
<tr>
<td>Positive</td>
<td>464</td>
<td>421 (90.7)</td>
<td>43 (9.3)</td>
<td></td>
</tr>
<tr>
<td>Not available</td>
<td>113</td>
<td>109</td>
<td>4</td>
<td>109</td>
</tr>
</tbody>
</table>

*P < 0.05; comparisons were made with the Fisher exact and Kruskal–Wallis tests.
primer sequences used are provided in Supplementary Table S2.

**SMAD3 transcription luciferase reporter assay**

Expression constructs for constitutively active TGFBR1 (ca-TGFBR1; substitution T204D in ref. 26), either wild-type or mutated SMAD3, a luciferase reporter for SMAD3 mediated transcription (pCAGA12; ref. 27) and an expression construct for Renilla luciferase to indicate transfection efficiency (28) were cotransfected into HEK293T cells. After 24 hours, luciferase expression was analyzed by use of the Dual-Glo luciferase reporter assay (Promega) on a LumiStar Galaxy luminometer (BMG Labtech).

**Coimmunoprecipitation**

Expression constructs for ca-TGFBR1, FLAG-SMAD4, and either wild-type or mutated HA-SMAD2 were cotransfected into HEK293T cells. After 24 hours, lysates were prepared with lysis buffer containing 1% Triton-X 100 (Sigma-Aldrich) and incubated with anti-FLAG M2 Affinity Gel beads (Sigma-Aldrich). The beads were washed 3 times with lysis buffer, and bound proteins were eluted by boiling in nondenaturing sample loading buffer. The proteins were separated by PAGE, transferred to nitrocellulose membranes, and immunoblotting was conducted using antibodies for FLAG (M2; Sigma-Aldrich) and SMAD2 (D43B4; Cell Signaling Technology).

**Pathogenicity prediction**

*Four in silico* algorithms were applied to predict the pathogenicity of missense mutations in SMAD4, SMAD2, and SMAD3: SIFT-Blink (29), PolyPhen-2 (30), MAPP (31), and I-Mutant-3.0 (ref. 32; Supplementary Methods). The 3-dimensional structure information used for these predictions is summarized in Supplementary Table S3. Pathogenicity predictions were conducted for putative splice site mutations using MaxEntScan (ref. 33; Supplementary Methods).

**Statistical analysis**

Statistical analyses were conducted using the R statistical computing software (34). Differences between groups were assessed using the Fisher exact test for categorical variables and the Kruskal–Wallis test for continuous variables. When multiple continuous variables were compared with a control variable, ANOVA was conducted followed by Dunnett post hoc comparison analysis. All statistical analyses were 2-sided and considered significant when \( P < 0.05 \).

**Results**

**Prevalence of SMAD4, SMAD2, and SMAD3 mutations in CRC**

Primary CRCs from 744 patients and 36 CRC cell lines were screened for somatic mutations in the entire coding regions and exon–intron boundaries of SMAD4, SMAD2, and SMAD3 by direct DNA sequencing (Supplementary Table S4). LOH was determined from SNP array data for 631 of the primary cancers and all of the cell lines. For the primary cancers, the prevalence of somatic truncating, frameshift, missense, and splice site mutations in SMAD4, SMAD2, and SMAD3 was 8.6% (64 of 744), 3.4% (25 of 744), and 4.3% (32 of 744), respectively, and the combined prevalence was 14.8% (110 of 744). Ten further changes were silent mutations that were excluded from subsequent analyses. LOH at SMAD4, SMAD2, and SMAD3 was detected in 54.0% (341 of 631), 53.4% (337 of 631), and 27.1% (171 of 631) of primary cases, respectively, with the loss at SMAD4 and SMAD2 being highly correlated because of their close proximity on chromosome 18q (Pearson correlation coefficient = 0.93). The CRC cell lines showed higher frequencies of SMAD4 and SMAD2 mutations than the primary cancers of 22.2% (8 of 36) and 13.9% (5 of 36), respectively, but the mutation frequency for SMAD3 (5.6%, 2 of 36) was similar. LOH at the SMAD4, SMAD2, and SMAD3 loci was present in 50.0% (18 of 36), 50.0% (18 of 36), and 16.7% (6 of 36) of the cell lines, consistent with the primary cancers.

**Spectrum of somatic SMAD4 mutations**

For SMAD4, the most frequent types of somatic mutation detected in primary cancers and cell lines were missense mutation (66.7%, 52 of 78), followed by nonsense (19.2%, 15 of 78), splice site (5.1%, 4 of 78), frameshift (3.8%, 3 of 78), and in-frame insertion/deletion mutations (3.8%, 3 of 78; Supplementary Fig. S1A). In one case, a nucleotide substitution altered the stop codon, presumably resulting in a protein extension. Sixty-one SMAD4 mutations were unique and 9 were recurrent changes, with 68.6% (48 of 70) to our knowledge representing novel somatic mutations not previously reported in CRCs (Supplementary Table S4). The nonsense and truncating mutations occurred throughout the gene, but 78.8% (41 of 52) of the missense mutations clustered in the MH2 domain, which represents only 41.5% of the coding sequence (\( P = 0.023 \), Fisher exact test; Fig. 1A). In line with previous observations (10–12), a hotspot region for missense mutations was apparent and we could refine its definition to include the 6 codons spanning from Asp351 to Pro356 and the nearby Arg361 (Figs. 1A–3). Missense changes at these residues accounted for 43.9% (18 of 41) of the MH2 domain missense changes detected. In addition, 2 of the 3 detected in-frame insertion/deletion mutations also mapped to these codons (D351del and P356delinsQK).

When the SMAD4 missense mutations were mapped onto the crystal structures of SMAD4:SMAD2:SMAD2 and SMAD4:SMAD3:MH2 domain heterotrimer (4–6), 46.3% (19 of 41) of the mutations occurred within the conserved R-SMAD binding surface. Furthermore, the identified mutation hotspot region (Asp351-Pro356, Arg361) mapped to a defined protein loop (L1 loop) directly involved in binding the R-SMADs (ref. 4; Fig. 3). Notably, the L1 loop is conserved across the 3 SMAD proteins and is used in the formation of both homo- and hetero-oligomeric complexes. At the base of the loop, a salt bridge connects Asp351 to Arg361, which in turn forms a further salt bridge with another conserved aspartate residue in the neighboring subunit of the trimeric SMAD complex (4). In SMAD4, the opposing aspartate (Asp357; that would be close proximity on chromosome 18q) was also altered by missense mutations in our cases (Fig. 3A).
A further 3 SMAD4 MH2 domain missense mutations (A406T, K428T, R515T) altered residues that are separate from the primary mutation hotspot region and are involved in binding the C-terminal phosphorylated Ser-Ser-X-Ser motifs of SMAD2 and SMAD3 (refs. 4–6). Of the remaining 16 MH2 missense mutations, 13 were predicted to affect protein stability by in silico pathogenicity prediction and 3 were deemed benign or ambiguous (Supplementary Table S4).

A total of 19 SMAD4 mutations were outside the MH2 domain, with 14 mapping within the MH1 domain and 5 occurring in the linker domain. Nine of the MH1 mutations were missense changes, including 4 examples of A118V, 2 changes altering Arg100 (R100G, R100T), and the single changes K45N, G65R, and L98F. All of these except K45N were predicted to alter protein stability by our in silico analyses (Supplementary Table S4). K45N and L98F occurred near the secondary structure and predicted to affect protein stability by our in silico analyses with 4 notable exceptions: The mutation A354T in SMAD2 was homologous to the A406T change identified in SMAD4, and was previously identified in an analogous manner. The change S276L in SMAD2 was directly involved in oligomerization or at amino acids supporting those residues (Fig. 1A–C). Overall for SMAD2 and SMAD3, 28 and 34 mutations were unique and 3 and 2 were recurrent changes, respectively. A further 3 SMAD4 MH2 domain missense mutations (A406T, K428T, R515T) altered residues that are separate from the principal mutation hotspot region in SMAD4 (Asp351-Pro356, Arg361; Figs. 1A–C). Notably, 35.0% (7 of 20) of SMAD4 mutations occurred at residues homologous to those of the principal mutation hotspot region in SMAD4 (ASP531-Pro536, Arg536; Figs. 1A–C). The Ser-Ser-X-Ser motif of SMAD2 was directly altered by 2 missense mutations (S464L, S467P), a recurring nonsense mutation (S464X, 2 cases), and probably by a detected stop codon change predicted to extend the protein as seen for SMAD4 (Figs. 1A and B). For SMAD3, the Ser-Ser-X-Ser motif was compromised by a missense mutation (S464L, S467P), a recurring nonsense mutation (S464X, 2 cases), and probably by a detected stop codon change predicted to extend the protein as seen for SMAD4 (Figs. 1A and B). For SMAD2, the Ser-Ser-X-Ser motif was compromised by a missense mutation (S464L, S467P), a recurring nonsense mutation (S464X, 2 cases), and probably by a detected stop codon change predicted to extend the protein as seen for SMAD4 (Figs. 1A and B). For SMAD3, the Ser-Ser-X-Ser motif was compromised by a missense mutation (S464L, S467P), a recurring nonsense mutation (S464X, 2 cases), and probably by a detected stop codon change predicted to extend the protein as seen for SMAD4 (Figs. 1A and B). For SMAD2, the Ser-Ser-X-Ser motif was compromised by a missense mutation (S464L, S467P), a recurring nonsense mutation (S464X, 2 cases), and probably by a detected stop codon change predicted to extend the protein as seen for SMAD4 (Figs. 1A and B). For SMAD3, the Ser-Ser-X-Ser motif was compromised by a missense mutation (S464L, S467P), a recurring nonsense mutation (S464X, 2 cases), and probably by a detected stop codon change predicted to extend the protein as seen for SMAD4 (Figs. 1A and B). For SMAD2, the Ser-Ser-X-Ser motif was compromised by a missense mutation (S464L, S467P), a recurring nonsense mutation (S464X, 2 cases), and probably by a detected stop codon change predicted to extend the protein as seen for SMAD4 (Figs. 1A and B). For SMAD3, the Ser-Ser-X-Ser motif was compromised by a missense mutation (S464L, S467P), a recurring nonsense mutation (S464X, 2 cases), and probably by a detected stop codon change predicted to extend the protein as seen for SMAD4 (Figs. 1A and B). For SMAD2, the Ser-Ser-X-Ser motif was compromised by a missense mutation (S464L, S467P), a recurring nonsense mutation (S464X, 2 cases), and probably by a detected stop codon change predicted to extend the protein as seen for SMAD4 (Figs. 1A and B). For SMAD3, the Ser-Ser-X-Ser motif was compromised by a missense mutation (S464L, S467P), a recurring nonsense mutation (S464X, 2 cases), and probably by a detected stop codon change predicted to extend the protein as seen for SMAD4 (Figs. 1A and B). For SMAD2, the Ser-Ser-X-Ser motif was compromised by a missense mutation (S464L, S467P), a recurring nonsense mutation (S464X, 2 cases), and probably by a detected stop codon change predicted to extend the protein as seen for SMAD4 (Figs. 1A and B). For SMAD3, the Ser-Ser-X-Ser motif was compromised by a missense mutation (S464L, S467P), a recurring nonsense mutation (S464X, 2 cases), and probably by a detected stop codon change predicted to extend the protein as seen for SMAD4 (Figs. 1A and B). For SMAD2, the Ser-Ser-X-Ser motif was compromised by a missense mutation (S464L, S467P), a recurring nonsense mutation (S464X, 2 cases), and probably by a detected stop codon change predicted to extend the protein as seen for SMAD4 (Figs. 1A and B). For SMAD3, the Ser-Ser-X-Ser motif was compromised by a missense mutation (S464L, S467P), a recurring nonsense mutation (S464X, 2 cases), and probably by a detected stop codon change predicted to extend the protein as seen for SMAD4 (Figs. 1A and B).
S266L, has been previously implicated in binding of the transcriptional coregulators Ski and Sno (38).

Within the MH1 domain of SMAD2, 2 of the 3 detected missense mutations were predicted to affect protein stability by in silico analysis and the third, A22V, was deemed benign. Of the 6 SMAD3 MH1 domain missense mutations, only one, K43R, had a solvent exposed wild-type residue and it was predicted to disrupt the conserved NLS (35) similar to the K45N change detected in SMAD4. Of the 5 remaining changes, 4 were predicted to affect protein stability, including R93N that altered the homologous residue to Arg100 in SMAD4. There was no evidence that MH1 mutations in SMAD2 and SMAD3 had altered DNA binding. Finally, for the linker domains, all 5 observed missense changes in the proteins were predicted to be benign, again mirroring the results for SMAD4.

**SMAD3 and SMAD4 tend to acquire two genetic hits**

To determine whether the SMAD4, SMAD2 and SMAD3 genes had a tendency to acquire 2 genetic hits in line with their proposed tumor suppressor roles, mutation data were integrated with LOH data for the primary tumors. As SMAD2 is localized within approximately 3.1 Mb of SMAD4 on chromosome 18q, analyses for SMAD2 were restricted to SMAD4 wild-type cases and analyses for SMAD4 were restricted to SMAD2 wild-type cases. For SMAD4, there was a significant overrepresentation of cases with 2 genetic hits in the form of either 2 mutations or 1 mutation plus LOH ($P < 0.001$, Fisher exact test; Table 2). A similar overrepresentation was observed for SMAD3 ($P = 0.001$), but in the case of SMAD2, there was no evidence for an excess of

---

**Table 2. Relationship between somatic mutations and allelic loss status at the SMAD4, SMAD2, and SMAD3 loci in 631 primary CRCs**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Number of mutations</th>
<th>No loss</th>
<th>LOH</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMAD4a</td>
<td>0</td>
<td>264 (47.9)</td>
<td>287 (52.1)</td>
<td>&lt;0.001b</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>11 (20.4)</td>
<td>43 (79.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4 (80.0)</td>
<td>1 (20.0)</td>
<td></td>
</tr>
<tr>
<td>SMAD2a</td>
<td>0</td>
<td>265 (48.1)</td>
<td>286 (51.9)</td>
<td>0.904</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>9 (47.4)</td>
<td>10 (52.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1 (100.0)</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td>SMAD3</td>
<td>0</td>
<td>448 (74.2)</td>
<td>156 (25.8)</td>
<td>&lt;0.001b</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>11 (42.3)</td>
<td>15 (57.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1 (100.0)</td>
<td>0 (0.0)</td>
<td></td>
</tr>
</tbody>
</table>

aGiven the close proximity of SMAD2 and SMAD4 on chromosome 18q, analyses for SMAD2 were restricted to SMAD4 wild-type cases and analyses for SMAD4 were restricted to SMAD2 wild-type cases.
b$P < 0.05$; comparisons were made with the Fisher exact test.
cases with 2 hits ($P = 0.904$). These relationships remained when the cell line data were included (Supplementary Table S5). There was no association between presence of $\text{SMAD4}$, $\text{SMAD2}$, or $\text{SMAD3}$ mutation and locus-specific DNA copy number ($P > 0.05$ for all comparisons, Supplementary Table S6).

**Relationships between $\text{SMAD4}$, $\text{SMAD2}$, and $\text{SMAD3}$ gene mutation status**

Overall, 1.5% (12 of 780) of cases showed mutations in more than one of the Co- or R-SMAD genes: 1 case had joint mutations in $\text{SMAD4}$ and $\text{SMAD2}$, 4 cases showed joint mutation of $\text{SMAD4}$ and $\text{SMAD3}$, and 7 cases, including 6 primary CRCs, had mutations in $\text{SMAD2}$ and $\text{SMAD3}$. Within the primary CRC cohort, cooccurrence of $\text{SMAD4}$ and R-SMAD gene mutations was consistent with random association ($P = 0.794$, Fisher exact test), but combined $\text{SMAD2}$ and $\text{SMAD3}$ mutation was significantly overrepresented ($P < 0.001$; Table 3). These relationships remained when cell line data were included (Supplementary Table S7). Strikingly, the 6 primary CRCs with mutations in both R-SMAD genes had biallelic genetic hits for both, and furthermore, all the somatic mutations involved mapped to the MH2 domains and were either nonsense or frameshift changes or missense mutations clearly predicted to be pathogenic (Supplementary Table S8). Finally, in all 6 cases, $\text{SMAD4}$ was wild-type, suggesting that joint inactivation of the R-SMAD genes represents an alternative to $\text{SMAD4}$ biallelic compromise in CRCs.

**Pathogenicity validation for novel $\text{SMAD2}$ and $\text{SMAD3}$ mutations**

To test the predicted pathogenicity of the detected missense mutations in $\text{SMAD2}$ and $\text{SMAD3}$, selected changes were introduced experimentally into the corresponding proteins by site-directed mutagenesis. For $\text{SMAD2}$, these included D300N, P305L, and D450E at the oligomerization

---

**Table 3. Relationships between somatic mutations in $\text{SMAD4}$, $\text{SMAD2}$, and $\text{SMAD3}$ for 744 primary CRCs**

<table>
<thead>
<tr>
<th></th>
<th>Wild-type</th>
<th>Mutant</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{SMAD2}$</td>
<td>Wild-type</td>
<td>656 (91.2)</td>
<td>63 (8.8)</td>
</tr>
<tr>
<td></td>
<td>Mutant</td>
<td>24 (96)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>$\text{SMAD4}$</td>
<td>Wild-type</td>
<td>652 (81.6)</td>
<td>60 (18.4)</td>
</tr>
<tr>
<td></td>
<td>Mutant</td>
<td>28 (87.5)</td>
<td>4 (12.5)</td>
</tr>
<tr>
<td>$\text{SMAD3}$</td>
<td>Wild-type</td>
<td>634 (81.5)</td>
<td>59 (18.5)</td>
</tr>
<tr>
<td></td>
<td>Mutant</td>
<td>46 (90.2)</td>
<td>5 (9.8)</td>
</tr>
<tr>
<td>$\text{SMAD2}$ and $\text{SMAD3}$</td>
<td>Wild-type</td>
<td>693 (96.4)</td>
<td>26 (3.6) &lt;0.001*</td>
</tr>
<tr>
<td></td>
<td>Mutant</td>
<td>19 (76.0)</td>
<td>6 (24.0)</td>
</tr>
</tbody>
</table>

*$P < 0.05$; comparisons were made with the Fisher exact test.
interface (Fig. 3A), S467P within the Ser-Ser-X-Ser motif (Fig. 3B), the partially exposed S276L reported to cause a hypomorphic development phenotype in mice (37), and T413N predicted to alter protein stability. For SMAD3, we selected D258N, R268C, and D408Y at the oligomerization interface (Fig. 3A), S425C targeting the Ser-Ser-X-Ser motif (Fig. 3B), P336S that alters a proline facilitating protein backbone bending, and the partially exposed R373Q that was detected twice in this study and also previously in a CRC cell line, SNU-769A (39).

SMAD3 directly binds DNA and regulates gene transcription (27). To test whether the observed mutations in SMAD3 altered this activity, expression constructs for the mutant SMAD3 proteins or the wild-type protein were analyzed in a well-established model of TGF-β signaling where HEK293T cells are activated by a constitutively active version of TGFBR1 (T204D; ca-TGFBR1; ref. 26). SMAD3 transcriptional activity was monitored using a specific luciferase reporter pCAGA12 based on the promoter of plasminogen activator inhibitor-type 1 (PAI1/SERPINE1; ref. 27). Transfection with wild-type SMAD3 or ca-TGFBR1 alone increased the luminescent signal 4-fold compared with cells transfected with vector control. Cotransfection with wild-type SMAD3 and ca-TGFBR1 increased the signal to 10-fold (Fig. 4A). When the various mutant SMAD3 constructs were cotransfected with ca-TGFBR1, the reporter signals were significantly...
lower than the wild-type SMAD3/ca-TGFBR1 signal (P < 0.001 for all comparisons).

Full-length SMAD2 is not able to directly bind DNA due to inclusion of an alternatively spliced exon not present in the highly homologous SMAD3 sequence (40), but it still acts as a transcriptional activator through complex formation with SMAD4 (5). As SMAD2 reporter assays remain poorly defined, we tested selected mutations for altered complex formation with SMAD4 in coimmunoprecipitation experiments. HEK293T cells were transfected with constructs encoding N-terminal FLAG-tagged SMAD4, wild-type or mutant SMAD2, and ca-TGFBR1. Lower amounts of the mutant SMAD2 proteins were pulled down by FLAG-SMAD4 than wild-type SMAD2 (P < 0.01, for all comparisons), suggesting that the selected mutations all reduce the ability of SMAD2 to form a complex with SMAD4 (Fig. 4B, Supplementary Fig. S2).

Relationship between SMAD4, SMAD2, and SMAD3 mutations and patient clinical characteristics

Mutations in SMAD4, SMAD2, and SMAD3 were analyzed for association with clinicopathologic features (Table 1), including age at diagnosis, gender, tumor location, American Joint Committee on Cancer (AJCC) stage, tumor (T) stage, node (N) stage, lymphovascular invasion, differentiation, mucinous histology, MSI, and CIN status. SMAD4 mutations were associated with mucinous histology (P < 0.001, Fisher exact test), but no further correlations were apparent. Mutation of SMAD2 was associated with CIN negative status (P = 0.004), whereas mutation of SMAD3 was associated with female gender (P = 0.045) and poor differentiation (P = 0.042). When the combined occurrence of mutation in the 3 genes was evaluated, the relationships with CIN negative status (P = 0.047) and mucinous histology (P = 0.005) remained significant, but no further relationships were detected.

Discussion

This study presents the first substantial characterization of somatic mutations in SMAD2 and SMAD3 in sporadic CRCs and is the most comprehensive survey to date of the somatic mutation spectrum of SMAD4. Our data show that somatic mutations in SMAD2 and SMAD3 make an important contribution to the mutational load in the TGF-β pathway, occurring in 3.4% and 4.3% of primary cancers, respectively. The R-SMAD genes have similar mutation spectra to that of SMAD4, with pathogenic mutations acting through analogous mechanisms. Furthermore, joint biallelic inactivation of the R-SMAD genes appears to represent an alternative mechanism to SMAD4 inactivation for TGF-β pathway compromise.

The prevalence and spectrum of detected SMAD4 mutations in primary cancers and cell lines was consistent with previous smaller studies (10–12) and the recent TCGA study (ref. 18; Supplementary Fig. S3A). The majority of the detected changes were missense mutations and these tended to cluster at MH2 residues involved in the SMAD complex interface. Here, we defined a mutation hotspot region (Asp351-Pro356, Arg361) mapping to the L1 loop of SMAD4 and a second group of targeted residues ( Ala406, Arg515, and Lys428) involved in binding the Ser-Ser-X-Ser motifs of the R-SMAD proteins. Importantly, functional studies of the L1 residues D351 and R361, and of Arg515 within the second set, have previously shown that these amino acids are important for complex formation (5, 15, 41). Within the MH1 domain of SMAD4, there was no evidence for clustering of missense mutations along the DNA-binding interface, and instead most changes were predicted to affect protein stability. For the changes at Gly65 and Arg100, this is supported by previous functional studies showing reduced stability for proteins containing relevant site-directed changes (15, 17, 42). Residue A118 showed recurrent mutation (A118V) and pathogenicity of this alteration is probable as it has been detected in pancreatic cancer (12, 43). The main consequence of SMAD4 mutations in CRCs therefore appears to be the prevention of SMAD4 homo- and/or hetero-oligomerization with the R-SMAD proteins or the reduction of SMAD4 stability.

Previous studies have only identified a small number of SMAD2 and SMAD3 mutations in primary CRCs, leaving open the question as to whether these genes contribute to colorectal tumorigenesis (19, 20). Our identification of 32 SMAD2 and 36 SMAD3 mutations reveal the somatic mutation spectra for these genes in CRCs. Both spectra showed a high similarity to the mutation spectrum of SMAD4, strongly suggesting that the detected changes were similarly pathogenic. Similar trends are apparent in the recent TCGA study (ref. 18; Supplementary Fig. S3B and S3C). The majority of SMAD2 and SMAD3 changes were missense mutations clustering in the MH2 domain and, in particular, occurred at residues involved in SMAD complex formation. Mutations were observed at residues homologous to the mutation hotspot region in the L1 loop of SMAD4 and within the Ser-Ser-X-Ser motifs critical for R-SMAD activation and subsequent binding to SMAD4. Functional analyses of selected interface, buried, and Ser-Ser-X-Ser motif mutations in SMAD2 and SMAD3 confirmed their pathogenicity, with SMAD3 changes reducing activity in luciferase reporter assays and SMAD2 changes reducing the amount of the SMAD2 protein pulled down with SMAD4 in coimmunoprecipitation. It is likely that these losses of function reflect a combination of effects on protein stability, complex formation, and/or DNA binding, which are likely to be interdependent. Our results are in agreement with previous functional studies that showed that changes at D258 and D408 in SMAD3, and alterations within the Ser-Ser-X-Ser motifs of both proteins, reduced complex formation with SMAD4 (2, 3, 5).

For the MH1 domains of SMAD2 and SMAD3, the missense mutations identified were primarily predicted to affect protein stability rather than alter protein–protein interactions, and similarly to the situation in SMAD4, there was no evidence to suggest targeting of DNA binding function. For the linker domains, all 5 relevant missense changes in SMAD2 and SMAD3 were predicted to be benign mirroring the results for SMAD4. A number of reports have implicated inappropriate phosphorylation of serine/threonine residues within the R-SMAD linker domains in the pathology of CRCs (44, 45), but we found only one change (T184A) in the linker domain of SMAD2 that could represent a relevant change. Taken together, our data suggest that SMAD4, SMAD2, and SMAD3 are mutated in a similar manner with homologous MH1 and MH2 domain
changes acting through analogous mechanisms to prevent the formation of activated complexes.

Somatic \textit{SMAD4} mutations have been reported to be more common in advanced stages of CRCs (11, 46), and LOH at the \textit{SMAD4} locus has been associated with poor prognosis (10). In our cohort of primary CRCs, presence of \textit{SMAD4} mutation showed no relationship to AJCC stage, T stage, N stage, or lymphovascular invasion, and the only significant association identified was with tumor mucinous histology. Mucinous histology has been suggested to be a poor prognostic factor in some studies (47, 48) and it may be in this context that \textit{SMAD4} LOH is relevant to prognosis (10). Consistent with our findings, chromosome 18q LOH has also been previously associated with mucinous histology in CRCs (49). The power to detect clinico-pathologic associations for \textit{SMAD2} and \textit{SMAD3} was limited, but \textit{SMAD2} mutations may be more prevalent in CIN negative tumors and \textit{SMAD3} mutations may be more frequent in tumors from females and or those with poorer differentiation.

A significant overrepresentation of 2 genetic hits was observed for \textit{SMAD4} and \textit{SMAD3}, consistent with these genes acting as tumor suppressors. This tendency was not detected for \textit{SMAD2}, although interpretation of LOH at this locus is confounded by its close proximity to \textit{SMAD4}. It is also remains possible that missense mutations within the \textit{SMAD} proteins confer dominant-negative effects in the context of SMAD complexes (50).

Despite the unclear overall relationship between \textit{SMAD2} mutation and LOH, we found strong evidence for selection of biallelic mutation in both \textit{SMAD2} and \textit{SMAD3}. Markedly, all of the mutations in these double-mutant cases were either truncating mutations or MH2 missense changes of unambiguous pathogenicity, and all of these tumors were \textit{SMAD4} wild-type. Taken together, these data strongly support that dual biallelic mutation of the R-SMADs is an alternative pathogenic mechanism to biallelic \textit{SMAD4} mutation. However, our results do not exclude the possibility that single-gene \textit{SMAD2} or \textit{SMAD3} compromise has significant functional effects.

In conclusion, we have shown a combined prevalence of \textit{SMAD4}, \textit{SMAD2}, and \textit{SMAD3} mutation of 14.8% in primary sporadic CRCs. \textit{SMAD4} mutations were the most common alterations, with missense changes predicted to disrupt complex formation and/or protein stability and in particular targeting the L1 loop of the protein. Our data suggest that \textit{SMAD2} and \textit{SMAD3} mutations are bona fide contributors to the mutation burden in CRCs. The mutation spectra of the R-SMAD genes mirrored that of \textit{SMAD4}, with homologous MH1 and MH2 domain changes shown to act through analogous pathogenic mechanisms. Biallelic \textit{SMAD4} mutation and joint biallelic \textit{SMAD2} and \textit{SMAD3} mutation appear to be functional alternatives for TGF-β pathway inactivation in CRCs.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors’ Contributions


Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): N.I. Fleming, R.N. Jorissen, D. Mouradov, M. Christie, C. Tsui, H.-J. Zhu, J.M. Mariadason, Q. Zhao, R.L. Strausberg, O.M. Sieber


Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): N.I. Fleming, A. Sakthianandeswaren, F. Day, S. Li, I.T. Jones, N.J. Hawkins, A.R. Ruszkiewicz, D. Busam

Study supervision: O.M. Sieber

Acknowledgments

This work was conducted under the auspices of the Hilton Ludwig Cancer Metastasis Initiative. The authors thank the patients for participating in this study, the Victorian Cancer Biobank and Biogrid Australia for the provision of specimens and access to the clinical data, respectively, and Prof. Manfred Schwab at the DKFZ for the provision of cell lines.

Grant Support

The study was supported by the CSIRO Preventative Health Flagship through an Australian Cancer Grid Project Grant (L. Lipton, P. Gibbs, O.M. Sieber), the NHMRC through a Project Grant (Application ID 894814; L. Lipton, P. Gibbs, O.M. Sieber, R.L. Ward), a Program Grant (Application ID 487922; A.W. Burgess), the Hilton Ludwig Cancer Metastasis Initiative (L. Lipton, P. Gibbs, O.M. Sieber), the Victorian Government through a Victorian Cancer Agency Translation Cancer Research Grant (L. Lipton, P. Gibbs, O.M. Sieber), and the Operational Infrastructure Support Program. M. Christie and F. Day are supported by the Cancer Council Victoria through Postgraduate Cancer Research Scholarships and L. Lipton by the CSIRO Preventative Health Flagship through a Clinical Researcher Fellowship.

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.

Received July 13, 2012; revised September 27, 2012; accepted October 17, 2012; published OnlineFirst November 8, 2012.

References


734 Cancer Res; 73(2) January 15, 2013 Cancer Research

Downloaded from cancerres.aacrjournals.org on April 19, 2017. © 2013 American Association for Cancer Research.


SMAD2, SMAD3 and SMAD4 Mutations in Colorectal Cancer

Nicholas I. Fleming, Robert N. Jorissen, Dmitri Mouradov, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-12-2706

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2012/11/08/0008-5472.CAN-12-2706.DC1

Cited articles
This article cites 50 articles, 21 of which you can access for free at:
http://cancerres.aacrjournals.org/content/73/2/725.full.html#ref-list-1

Citing articles
This article has been cited by 11 HighWire-hosted articles. Access the articles at:
/content/73/2/725.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.