SMAD2, SMAD3 and SMAD4 mutations in colorectal cancer

Nicholas Fleming1$, Robert N. Jorissen1$, Dmitri Mouradov1, Michael Christie1,2, Anuratha Sakthianandeswaren1, Michelle Palmieri1, Fiona Day1,2, Shan Li1, Cary Tsui1, Lara Lipton1,2,3, Jayesh Desai1,3, Ian T. Jones4, Stephen McLaughlin5, Robyn L. Ward6, Nicholas J. Hawkins7, Andrew R. Ruszkiewicz8, James Moore9, Hong-Jian Zhu10, John M. Mariadason11, Antony W. Burgess12, Dana Busam13, Qi Zhao14, Robert L. Strausberg14,15, Peter Gibbs1,2,3 and Oliver M. Sieber1,2

$Authors contributed equally to this work

1Ludwig Colon Cancer Initiative Laboratory, Ludwig Institute for Cancer Research, Parkville, VIC, Australia

2Faculty of Medicine, Dentistry and Health Sciences, Department of Surgery, University of Melbourne, Parkville, VIC, Australia

3Department of Medical Oncology, Royal Melbourne Hospital, Parkville, VIC, Australia

4Department of Colorectal Surgery, Royal Melbourne Hospital, Parkville, VIC, Australia

5Department of Colorectal Surgery, Western Hospital, Footscray, VIC, Australia

6Lowy Cancer Research Centre, Prince of Wales Clinical School, University of New South Wales, Sydney, NSW, Australia
School of Medical Sciences, University of New South Wales, Sydney, NSW, Australia

Pathology Department, Institute of Medical and Veterinary Science, Adelaide, SA, Australia

Department of Colorectal Surgery, Royal Adelaide Hospital, Adelaide, SA, Australia

Department of Surgery, University of Melbourne, Parkville, VIC, Australia

Oncogenic Transcription Laboratory, Ludwig Institute for Cancer Research, Austin, VIC, Australia

Epithelial Biology Laboratory, Ludwig Institute for Cancer Research, Parkville, VIC, Australia

J. Craig Venter Institute, Rockville, MD

Ludwig Collaborative Laboratory for Cancer Biology and Therapy, Department of Neurosurgery, Johns Hopkins University School of Medicine, Baltimore, MD

Ludwig Institute for Cancer Research Ltd., New York, NY

Running title: Co- and R-SMAD mutations in colorectal cancer

Key words: SMAD2, SMAD3, SMAD4, mutation, colorectal cancer

Financial support: Supported by the CSIRO Preventative Health Flagship through an Australian Cancer Grid Project Grant (to L.L., P.G., O.M.S), the NHMRC through a Project Grant (Application ID 489418; to L.L, O.M.S, P.G,
R.L.W) and a Program Grant (Application ID 487922; to A.W.B.), the Hilton Ludwig Cancer Metastasis Initiative (to L.L., P.G., O.M.S.) and the Victorian Government through a Victorian Cancer Agency Translation Cancer Research Grant (to L.L., P.G., O.M.S.) and the Operational Infrastructure Support Program. M.C. and F.D. are supported by the Cancer Council Victoria through Postgraduate Cancer Research Scholarships and L.L. by the CSIRO Preventative Health Flagship through a Clinical Researcher Fellowship.

Request for reprints: Oliver Sieber, Ludwig Colon Cancer Initiative Laboratory, Ludwig Institute for Cancer Research, PO Box 2008, Royal Melbourne Hospital, Victoria, VIC 3050, Australia. Phone: 61-3-93413168; Fax: 61-3-93413104; E-mail: Oliver.Sieber@ludwig.edu.au.

Conflicts of Interest: The authors have no conflicts of interest to declare.

Acknowledgments: This work was conducted under the auspices of the Hilton Ludwig Cancer Metastasis Initiative. We are grateful to 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. We thank Prof. Manfred Schwab at the DKFZ for the provision of cell lines.

Notes about the manuscript: This manuscript has 4888 words, 3 tables and 4 figures.
Abstract

Activation of the canonical transforming growth factor-β (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 SMAD4, SMAD2 and SMAD3, and analyzed for allelic loss by SNP microarray analysis. Mutation spectra were compared between the genes, the pathogenicity of mutations was assessed and relationships with clinicopathological features were examined. The prevalence of SMAD4, SMAD2 and SMAD3 mutations in sporadic CRC was 8.6% (64/744), 3.4% (25/744) and 4.3% (32/744), 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 demonstrated 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.
**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) (1). Germline mutations in *SMAD4* cause juvenile polyposis syndrome (JPS), an autosomal dominant predisposition to multiple gastrointestinal polyps and cancer (8). In sporadic CRC, *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 two 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 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 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 hetero-trimeric transcriptional complexes containing one SMAD4 and two 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 CRC is limited. The mutation prevalence has been broadly estimated at 2-6%, 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 employing 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 CRC.

In this study, we determined the prevalence, spectra and loss of heterozygosity (LOH) status of SMAD4, SMAD2 and SMAD3 somatic mutations in 744 sporadic CRC 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 CRC.

Materials and Methods

Patients

Fresh frozen tumor and normal specimens were analyzed from 744 patients with sporadic CRC 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 of 70 stage I, 228 stage II, 347 stage III and 99 stage IV cases. 313 cancers were from the proximal colon, 241 from the distal colon and 189 from the rectum; tumor location data was 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).

Colorectal cancer 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, H2T29, 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 lines are provided in Supplementary Table S1.
Mutation detection

For cancer samples, hematoxylin and eosin-stained sections were reviewed by an anatomical pathologist and macro-dissected in areas comprising greater than 60% neoplastic cells. Genomic DNA was extracted using the AllPrep DNA/RNA Mini Kit (Qiagen, Germantown, MD). Coding regions and exon-intron boundaries of the entire \(SMAD4\), \(SMAD2\) and \(SMAD3\) genes were amplified using the polymerase chain reaction (PCR), and direct DNA sequencing was performed using BigDye Terminator v3.1 Ready Reaction Mix (Applied Biosystems, Foster City, CA). Primer sequences are available from the authors. Reaction products were run on 3730xl DNA Analyzers (Applied Biosystems) employing Biomek FX robots (Beckman-Coulter, Brea, CA) and Pixsys 4200 nanoliter liquid handling systems (Cartesian Technologies, Holliston, MA). Mutations detected were confirmed as somatic by bidirectional re-sequencing 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, San Diego, CA) 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 2Mb. 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 >3 autosomes showed deviations from 2n were classified as CIN+ve.

Expression constructs and site-directed mutagenesis

Expression constructs for N-terminal HA-tagged SMAD2, untagged SMAD3 and N-terminal FLAG-tagged SMAD4 were generated in pcDNA3.1+ (Invitrogen, Carlsbad, CA) 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 performed using the QuikChange II XL kit (Agilent Technologies, Santa Clara, CA) 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. The 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 reference 26), either wild-type or mutated SMAD3, a luciferase reporter for SMAD3 mediated transcription (pCAGA12) (27) and an expression construct for *Renilla* luciferase to indicate transfection efficiency (28) were cotransfected into HEK293T cells. After 24hr, luciferase expression was analyzed by use of the Dual-Glo luciferase reporter assay (Promega, San Luis Obispo, CA) on a LumiStar Galaxy luminometer (BMG Labtech, Ortenberg, Germany).

**Co-immunoprecipitation (Co-IP)**

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, St Louis, MO) and incubated with anti-FLAG M2 Affinity Gel beads (Sigma-Aldrich). The beads were washed three times with lysis buffer and bound proteins were eluted by boiling in non-denaturing sample loading buffer. The proteins were separated by polyacrylamide gel electrophoresis, transferred to
nitrocellulose membranes and immunoblotting was performed using antibodies for FLAG (M2; Sigma-Aldrich) and SMAD2 (D43B4; Cell Signaling Technology, Danvers, MA).

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 (32) (*Supplementary Methods*). The three-dimensional structure information used for these predictions is summarized in *Supplementary Table S3*. Pathogenicity predictions were performed for putative splice site mutations using MaxEntScan (33) (*Supplementary Methods*).

Statistical analysis

Statistical analyses were conducted using the R statistical computing software (34). Differences between groups were assessed using Fisher’s exact test for categorical variables and the Kruskal-Wallis test for continuous variables. When multiple continuous variables were compared to a control variable, ANOVA was performed followed by Dunnett *post hoc* comparison analysis. All statistical analyses were two-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% (64/744), 3.4% (25/744) and 4.3% (32/744), respectively, and the combined prevalence was 14.8% (110/744). Ten further changes were silent mutations that were excluded from subsequent analyses. LOH at SMAD4, SMAD2 and SMAD3 was detected in 54.0% (341/631), 53.4% (337/631) and 27.1% (171/631) of primary cases, respectively, with the loss at SMAD4 and SMAD2 being highly correlated due to their close proximity on chromosome 18q (Pearson correlation coefficient = 0.93). The CRC cell lines showed higher frequencies of SMAD4 and SMAD2 mutations compared to the primary cancers, of 22.2% (8/36) and 13.9% (5/36), respectively, but the mutation frequency for SMAD3 (5.6%, 2/36) was similar. LOH at the SMAD4, SMAD2 and SMAD3 loci was present in 50.0% (18/36), 50.0% (18/36) and 16.7% (6/36) 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%, 52/78), followed by nonsense (19.2%, 15/78), splice site (5.1%, 4/78), frameshift (3.8%, 3/78) and in-frame insertion/deletion mutations (3.8%, 3/78) (Supplementary Fig. S1A). In one case, a nucleotide substitution altered the stop codon, presumably resulting in a protein extension. 61 SMAD4 mutations were unique and 9 were recurrent changes, with 68.6% (48/70) to our knowledge representing novel somatic mutations not previously reported in CRC (Supplementary Table S4). The nonsense and truncating mutations occurred throughout the gene, but 78.8% (41/52) of the missense mutations clustered in the MH2 domain which represents only 41.5% of the coding sequence (P=0.023, Fisher’s 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 six codons spanning from Asp351 to Pro356 and the nearby Arg361 (Fig. 1, 2 and 3). Missense changes at these residues accounted for 43.9% (18/41) of the MH2 domain missense changes detected. In addition, two of the three 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:SMAD3 MH2 domain
heterotrimers (4-6), 46.3% (19/41) of the mutations occurred within the conserved R-SMAD binding surface. Further, the identified mutation hotspot region (Asp351-Pro356, Arg361) mapped to a defined protein loop (L1 loop) directly involved in binding the R-SMADs (4) (Fig. 3). Notably, the L1 loop is conserved across the three SMAD proteins and is employed 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 (Asp537) (that would be bound by the L1 loop of a neighboring SMAD protein) as well as an associated leucine residue (Leu540) were also altered by missense mutations in our cases (Fig. 3A).

A further three 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 (4-6) (Fig. 3B). 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 four examples of A118V, two 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 bound DNA interface (Fig. 2), but the pathogenicity of K45N may be due to disruption of a conserved nuclear localization signal (NLS) (35) rather than an alteration of DNA binding. Hence, we found limited evidence to suggest that MH1 domain mutations were selected to target the DNA binding interface of SMAD4. Finally, the two missense mutations in the SMAD4 linker domain, L172M and T197I, were predicted to be benign in our analyses. However, phosphorylation of the SMAD4 linker domain has been implicated in the regulation of nuclear import/export (36) and it is possible that T197I removes a regulatory phosphorylation site.

Spectra of somatic SMAD2 and SMAD3 mutations

The somatic mutation spectra of SMAD2 and SMAD3 were highly similar to the mutation spectrum of SMAD4 both in mutation type and location. The most common changes were missense mutations (23/32, 71.9% and 23/36, 63.9%, respectively), then nonsense mutations (6/32, 18.8% and 5/36, 13.9%) and then other mutation types (3/32, 9.4% and 8/36, 22.2%) (Supplementary Fig. S1B and C). As observed in SMAD4, the majority of the mutations in both R-SMADs mapped to the protein MH2 domains and were overrepresented relative to the
number of amino acids present in the different regions of the proteins (20/32, 62.5%, \( P=0.026 \) and 24/36, 66.7%, \( P=0.038 \), respectively, Fisher's exact test) (Fig. 1B and C). Overall for SMAD2 and SMAD3, 28 and 34 mutations were unique and 3 and 2 were recurrent changes, with 77.4% (24/31) and 88.9% (32/36) to our knowledge representing novel somatic mutations not previously reported in CRC (Supplementary Table S4).

Within the MH2 domains of SMAD2 and SMAD3, the distribution of missense mutations revealed further similarities to SMAD4. 58.5% (10/17) of SMAD2 and 33.3% (5/15) of SMAD3 MH2 domain missense changes clustered at residues directly involved in oligomerization or at amino-acids supporting those residues (Fig. 1, 2 and 3). Notably, 35.0% (7/20) of SMAD2 and 12.5% (3/24) of SMAD3 mutations occurred at residues homologous to those of the principal mutation hotspot region in SMAD4 (Asp351-Pro356, Arg361) (Fig. 1, 2 and 3A). The Ser-Ser-X-Ser motif of SMAD2 was directly altered by two 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 (Fig. 1 and 3B). For SMAD3, the Ser-Ser-X-Ser motif was compromised by a missense mutation (S425C), an in-frame insertion (S422_S423insRN) and a frameshift mutation (S423VfsX13) (Fig. 1 and 3B). As also seen for SMAD4, the majority of the remaining missense mutations in the MH2 domains of SMAD2 and SMAD3 were buried within the protein structure and predicted to affect protein stability by \textit{in silico} analysis with four notable
exceptions: The mutation A354T in SMAD2 was homologous to the A406T change identified in SMAD4, and is therefore likely to compromise SMAD oligomerization in an analogous manner. The change S276L in SMAD2, which occurred in three cases, is partially surface exposed and was previously identified in an N-ethyl-N-nitrosourea mutagenesis screen where it caused a hypomorphic developmental phenotype in mice by an unknown mechanism (37). The wild-type residue of E228K in SMAD3 is solvent-exposed, but of unknown function. Finally, Ser266 in SMAD3, altered by S266L, has been previously implicated in binding of the transcriptional coregulators Ski and Sno (38).

Within the MH1 domain of SMAD2, two of the three detected missense mutations were predicted to affect protein stability by in silico analysis and the third, A22V, was deemed benign. Of the six 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 five remaining changes, four 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 five 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 two 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 ~3.1mb 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 two genetic hits in the form of either two mutations or one mutation plus LOH ($P<0.001$, Fisher’s 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 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 SMAD4, SMAD2 or SMAD3 mutation and locus-specific DNA copy number ($P>0.05$ for all comparisons, **Supplementary Table S6**).

**Relationships between SMAD4, SMAD2, and SMAD3 gene mutation status**

Overall, 1.5% (12/780) of cases showed mutations in more than one of the Co- or R-SMAD genes: one case had joint mutations in SMAD4 and SMAD2, four cases showed joint mutation of SMAD4 and SMAD3 and seven cases, including six primary CRCs, had mutations in SMAD2 and SMAD3. Within the primary CRC
cohort, co-occurrence of SMAD4 and R-SMAD gene mutations was consistent with random association (P=0.794, Fisher's exact test), but combined SMAD2 and SMAD3 mutation was significantly overrepresented (P<0.001) (Table 3). These relationships remained when cell line data were included (Supplementary Table S7). Strikingly, the six primary CRCs with mutations in both R-SMAD genes had biallelic genetic hits for both, and further, 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 six cases SMAD4 was wild-type suggesting that joint inactivation of the R-SMAD genes represents an alternative to SMAD4 biallelic compromise in CRC.

**Pathogenicity validation for novel SMAD2 and SMAD3 mutations**

To test the predicted pathogenicity of the detected missense mutations in SMAD2 and SMAD3, selected changes were introduced experimentally into the corresponding proteins by site-directed mutagenesis. For SMAD2, these included D300N, P305L and D450E at the oligomerization 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) (26). SMAD3 transcriptional activity was monitored using a specific luciferase reporter pCAGA12 based on the promoter of plasminogen activator inhibitor-type 1 (PAI1/SERPINE1) (27). Transfection with wild-type SMAD3 or ca-TGFBR1 alone increased the luminescent signal 4-fold compared to 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 Co-IP 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 clinicopathological features (Table 1), including age at diagnosis, gender, tumor location, AJCC stage, T stage, N stage, lymphovascular invasion, differentiation, mucinous histology, MSI and CIN status. $SMAD4$ mutations were associated with mucinous histology ($P<0.001$, Fisher’s exact test) but no further correlations were apparent. Mutation of $SMAD2$ was associated with CIN-ve status ($P=0.004$) while 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 three genes was evaluated, the relationships with CIN-ve 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 \textit{SMAD2} and \textit{SMAD3} in sporadic CRC and is the most comprehensive survey to date of the somatic mutation spectrum of \textit{SMAD4}. Our data demonstrate that somatic mutations in \textit{SMAD2} and \textit{SMAD3} make an important contribution to the mutational load in the TGF-\(\beta\) pathway, occurring in 3.4\% and 4.3\% of primary cancers, respectively. The R-SMAD genes have similar mutation spectra to that of \textit{SMAD4}, with pathogenic mutations acting through analogous mechanisms. Further, joint biallelic inactivation of the R-SMAD genes appears to represent an alternative mechanism to \textit{SMAD4} inactivation for TGF-\(\beta\) pathway compromise.

The prevalence and spectrum of detected \textit{SMAD4} mutations in primary cancers and cell lines was consistent with previous smaller studies (10-12) and the recent TCGA study (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 demonstrated 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 demonstrating 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 CRC 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 CRC, 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 CRC. 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 (18) (Supplementary Fig. S3B and C). 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 Co-IP. 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 demonstrated 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 five 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 CRC (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 CRC (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 CRC (49). The power to detect clinicopathologic associations for \textit{SMAD2} and \textit{SMAD3} was limited, but \textit{SMAD2} mutations may be more prevalent in CIN-ve tumors and \textit{SMAD3} mutations may be more frequent in tumors from females and or those with poorer differentiation.

A significant overrepresentation of two 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 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 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 SMAD4 wild-type. Taken together, these data strongly suggest that dual biallelic mutation of the R-SMADs is an alternative pathogenic mechanism to biallelic SMAD4 mutation. However, our results do not exclude the possibility that single-gene SMAD2 or SMAD3 compromise has significant functional effects.

In conclusion, we have demonstrated a combined prevalence of SMAD4, SMAD2 and SMAD3 mutation of 14.8% in primary sporadic CRC. 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 SMAD2 and SMAD3 mutations are bone fide contributors to the mutation burden in CRC. The mutation spectra of the R-SMAD genes mirrored that of SMAD4, with homologous MH1 and MH2 domain changes shown to act through analogous pathogenic mechanisms. Biallelic SMAD4 mutation and joint biallelic SMAD2 and SMAD3 mutation appear to be functional alternatives for TGF-β pathway inactivation in CRC.
References


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>P-value</th>
<th>SMAD2</th>
<th>P-value</th>
<th>SMAD3</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Wild-type</td>
<td>(%)</td>
<td>Mutant</td>
<td>(%)</td>
<td>Wild-type</td>
<td>(%)</td>
</tr>
<tr>
<td>Total N</td>
<td>744</td>
<td>680 (91.4)</td>
<td>64 (8.6)</td>
<td>719 (96.6)</td>
<td>25 (3.4)</td>
<td>712 (95.7)</td>
<td>32 (4.3)</td>
</tr>
<tr>
<td>Age, years</td>
<td></td>
<td>Mean ± SD</td>
<td>69.1 ± 11.3</td>
<td>69.3 ± 11.5</td>
<td>67.5 ± 9.3</td>
<td>0.113</td>
<td>69.1 ± 11.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Range</td>
<td>25.0 - 99.2</td>
<td>25.0 - 99.2</td>
<td>45.7 - 85.9</td>
<td></td>
<td>25.0 - 99.2</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td>Male</td>
<td>414</td>
<td>384 (92.8)</td>
<td>30 (7.2)</td>
<td>0.149</td>
<td>400 (96.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>330</td>
<td>296 (89.7)</td>
<td>34 (10.3)</td>
<td></td>
<td>319 (96.7)</td>
</tr>
<tr>
<td>Tumor location</td>
<td></td>
<td>Right colon</td>
<td>313</td>
<td>282 (90.1)</td>
<td>31 (9.9)</td>
<td>0.274</td>
<td>300 (95.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Left colon</td>
<td>241</td>
<td>219 (90.9)</td>
<td>22 (9.1)</td>
<td></td>
<td>236 (97.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rectum</td>
<td>189</td>
<td>178 (94.2)</td>
<td>11 (5.8)</td>
<td></td>
<td>182 (96.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not available</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>AJCC stage</td>
<td></td>
<td>I</td>
<td>70</td>
<td>66 (94.3)</td>
<td>4 (5.7)</td>
<td>0.384</td>
<td>68 (97.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>II</td>
<td>228</td>
<td>206 (90.4)</td>
<td>22 (9.6)</td>
<td></td>
<td>218 (95.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>III</td>
<td>347</td>
<td>321 (92.5)</td>
<td>26 (7.5)</td>
<td></td>
<td>339 (97.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IV</td>
<td>99</td>
<td>87 (87.9)</td>
<td>12 (12.1)</td>
<td></td>
<td>94 (94.9)</td>
</tr>
<tr>
<td>T stage</td>
<td></td>
<td>1</td>
<td>21</td>
<td>21 (100)</td>
<td>0 (0)</td>
<td>0.388</td>
<td>19 (90.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>91</td>
<td>85 (93.4)</td>
<td>6 (6.6)</td>
<td></td>
<td>87 (95.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>516</td>
<td>471 (91.3)</td>
<td>45 (8.7)</td>
<td></td>
<td>500 (96.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>116</td>
<td>103 (88.8)</td>
<td>13 (11.2)</td>
<td></td>
<td>113 (97.4)</td>
</tr>
<tr>
<td>N stage</td>
<td></td>
<td>0</td>
<td>326</td>
<td>296 (90.8)</td>
<td>30 (9.2)</td>
<td>0.732</td>
<td>312 (95.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>255</td>
<td>236 (92.5)</td>
<td>19 (7.5)</td>
<td></td>
<td>247 (96.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>163</td>
<td>148 (90.8)</td>
<td>15 (9.2)</td>
<td></td>
<td>160 (98.2)</td>
</tr>
<tr>
<td>Lymphovascular invasion</td>
<td></td>
<td>Absent</td>
<td>289</td>
<td>260 (90)</td>
<td>29 (10)</td>
<td>0.338</td>
<td>279 (96.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Present</td>
<td>206</td>
<td>191 (92.7)</td>
<td>15 (7.3)</td>
<td></td>
<td>198 (96.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not available</td>
<td>249</td>
<td>229</td>
<td>20</td>
<td></td>
<td>242</td>
</tr>
<tr>
<td>Differentiation</td>
<td></td>
<td>Well/Moderate</td>
<td>546</td>
<td>498 (91.2)</td>
<td>48 (8.8)</td>
<td>0.754</td>
<td>526 (96.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Poor</td>
<td>171</td>
<td>158 (92.4)</td>
<td>13 (7.6)</td>
<td></td>
<td>167 (97.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not available</td>
<td>27</td>
<td>24</td>
<td>3</td>
<td></td>
<td>26</td>
</tr>
<tr>
<td>Mucinous</td>
<td></td>
<td>No</td>
<td>578</td>
<td>540 (93.4)</td>
<td>38 (6.6)</td>
<td>&lt;0.001*</td>
<td>562 (97.2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yes</td>
<td>157</td>
<td>133 (84.7)</td>
<td>24 (15.3)</td>
<td></td>
<td>148 (94.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not available</td>
<td>9</td>
<td>7</td>
<td>2</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>MSI</td>
<td></td>
<td>Stable</td>
<td>646</td>
<td>588 (91.0)</td>
<td>58 (9)</td>
<td>0.441</td>
<td>626 (96.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unstable</td>
<td>98</td>
<td>92 (93.9)</td>
<td>6 (6.1)</td>
<td></td>
<td>93 (94.9)</td>
</tr>
<tr>
<td>CIN</td>
<td></td>
<td>Negative</td>
<td>167</td>
<td>150 (89.8)</td>
<td>17 (10.2)</td>
<td>0.759</td>
<td>155 (92.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Positive</td>
<td>464</td>
<td>421 (90.7)</td>
<td>43 (9.3)</td>
<td></td>
<td>456 (98.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not available</td>
<td>113</td>
<td>109</td>
<td>4</td>
<td></td>
<td>109</td>
</tr>
</tbody>
</table>

*P<0.05; comparisons were made with the Fisher’s exact and Kruskal-Wallis tests.
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-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMAD4</td>
<td>0</td>
<td>264 (47.9)</td>
<td>287 (52.1)</td>
<td>&lt; 0.001&lt;sup&gt;b&lt;/sup&gt;</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></td>
<td>SMAD2</td>
<td>0</td>
<td>265 (48.1)</td>
<td>286 (51.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>9 (47.4)</td>
<td>10 (52.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>1 (100.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>SMAD3</td>
<td>0</td>
<td>448 (74.2)</td>
<td>156 (25.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>11 (42.3)</td>
<td>15 (57.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>1 (100.0)</td>
<td>0 (0.0)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Given 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.

<sup>b</sup>P<0.05; comparisons were made with the Fisher’s exact test.
Table 3. Relationships between somatic mutations in SMAD4, SMAD2 and SMAD3 for 744 primary CRCs.

<table>
<thead>
<tr>
<th></th>
<th>SMAD4</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wild-type</td>
<td>Mutant</td>
<td>P-value</td>
<td>Wild-type</td>
<td>Mutant</td>
</tr>
<tr>
<td>SMAD2</td>
<td>Wild-type</td>
<td>656 (91.2)</td>
<td>63 (8.8)</td>
<td>0.715</td>
<td>Mutant</td>
</tr>
<tr>
<td></td>
<td>SMAD4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMAD3</td>
<td>Wild-type</td>
<td>652 (91.6)</td>
<td>60 (8.4)</td>
<td>0.346</td>
<td>Mutant</td>
</tr>
<tr>
<td></td>
<td>SMAD4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wild-type</td>
<td>634 (91.5)</td>
<td>59 (8.5)</td>
<td>0.794</td>
<td>Mutant</td>
</tr>
<tr>
<td>SMAD2</td>
<td>and/or</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SMAD3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wild-type</td>
<td>693 (96.4)</td>
<td>26 (3.6)</td>
<td>&lt;0.001&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Mutant</td>
</tr>
</tbody>
</table>

<sup>a</sup><sup>P</sup><0.05; Comparisons were made with the Fisher’s exact test.
**Figure legends**

**Figure 1.** Somatic mutation spectra for (A.) **SMAD4**, (B.) **SMAD2** and (C.) **SMAD3** for 744 primary CRCs and 36 CRC cell lines.

Blue line = missense mutation, red line = nonsense mutation, blue triangle = in-frame insertion/deletion, red triangle = out-of-frame insertion/deletion, blue square = combined deletion and insertion, black triangle = splice site mutation (arrow direction indicates whether it occurs before or after the exon), black star = start/stop site mutation. Height units indicate count of mutations.

**Figure 2.** Distribution of missense mutations within the MH1 and MH2 domains of SMAD4, SMAD2 and SMAD3.

Structures for MH1 domains are shown for complexes with DNA. Structures for MH2 domains are shown for heterotrimeric complexes of SMAD4 with the R-SMADs. Orange = mutation at interface, cyan = mutation adjacent to interface, red = mutation at serine in R-SMAD Ser-Ser-X-Ser motif, blue = amino-acid deletion, green = other mutation (mainly internal) and pink = zinc ion. The SMAD2 MH1 domain structure is based on the SMAD3 MH1 domain structure and represents the DNA binding spliceoform.

**Figure 3.** Mutation hotspot regions in SMAD4 compared to homologous regions in SMAD2 and SMAD3.
(A.) SMAD4 hotspot region spanning Asp351-Pro356 and including Arg361 which forms a key interaction interface with SMAD2 or SMAD3 (B.) SMAD4 hotspot region comprising residues Ala406, Arg515 and Lys428 mediating interaction with the C-terminal Ser-Ser-X-Ser motif of SMAD2 or SMAD3. Purple = SMAD2 or SMAD3, yellow = SMAD4. Frequencies of reoccurring changes indicated in brackets.

**Figure 4.** Functional impact of novel somatic mutations in SMAD2 and SMAD3.

(A.) SMAD binding motif mediated transcription by wild-type and mutant SMAD3 proteins. Luciferase assays were performed in HEK293T cells 24hr post-transfection with 5ng of the pCAGA12 firefly luciferase reporter (27) and 90ng of the indicated SMAD3 and ca-TGFBR1 expression constructs. 5ng of a *Renilla* luciferase expression construct was included to indicate transfection efficiency. The two luciferase signals were quantitated and mean signal ratios of three experiments were generated. Error bars are SEM. Indications of significance are * = *P*<0.001, were tested by ANOVA followed by Dunnett post-hoc analysis and relate to comparisons with the wild-type SMAD3/ca-TGFBR1 cells (bar 5). (B.) Co-IP of wild-type and mutant SMAD2 proteins with FLAG-SMAD4. Representative of three experiments. Indications of significance are * = *P*<0.01 and were tested by ANOVA followed by Dunnett post-hoc analysis and relate to comparisons with the FLAG-SMAD4/wild-type SMAD2/ca-TGFBR1 pulldown (lane 4).
### Figure 4

#### A.

![Graph showing luciferase/renilla (fold expression) for various constructs](image)

#### B.

<table>
<thead>
<tr>
<th>FLAG-SMAD4</th>
<th>L</th>
<th>+</th>
<th>+</th>
<th>+</th>
<th>+</th>
<th>+</th>
<th>+</th>
<th>+</th>
</tr>
</thead>
<tbody>
<tr>
<td>ca-TGFBRI</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SMAD2</td>
<td></td>
<td>WT</td>
<td>D300N</td>
<td>P305L</td>
<td>D450E</td>
<td>T413N</td>
<td>S276L</td>
<td>S467P</td>
</tr>
</tbody>
</table>

![Image of western blot analysis](image)
SMAD2, SMAD3 and SMAD4 mutations in colorectal cancer

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

Cancer Res  Published OnlineFirst November 8, 2012.

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

Author Manuscript  Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

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.

Downloaded from cancerres.aacrjournals.org on July 23, 2017. © 2012 American Association for Cancer Research.