[CANCER RESEARCH 60, 3872–3879, July 15, 2000]

The Human T-Cell Transcription Factor-4 Gene: Structure, Extensive Characterization of Alternative Splicings, and Mutational Analysis in Colorectal Cancer Cell Lines[1]

Alex Duval, Sandra Rolland, Emmanuel Tubacher, Hung Bui, Gilles Thomas, and Richard Hamelin[2]

Institut National de la Santé et de la Recherche Médicale U434, Centre d'Etude du Polymorphisme Humain, 75010 Paris, France

ABSTRACT

The human T cell transcription factor-4 (hTCF-4) interacts functionally with β-catenin in the Wnt signaling pathway, which regulates many developmental processes. Moreover, inappropriate reactivation of this pathway attributable to *APC* or *β-catenin* mutations has been described in colorectal cancers. Because only the human *TCF-4* cDNA sequence was known, we determined its genomic structure. A total of 17 exons, of which 5 were alternative, were identified. Moreover, four alternative splice sites were observed either experimentally or *in silico* by a BLAST approach in expressed sequence tag databases. The alternative use of three consecutive exons localized in the 3′ part of the *hTCF-4* gene changes the reading frames used in the last exon, leading to the synthesis of a number of hTCF-4 isoforms with short, medium, or long-size COOH-terminal ends. We next screened the entire *hTCF-4* gene for mutations in a series of 24 colorectal cancer cell lines by denaturing gradient gel electrophoresis and/or direct sequencing. Besides an already described deletion of an A in an (A)9 coding repeat in four cases, we found DNA variants in eight cases for a total of 12 variants, of which 8 were coding. These include one frameshift mutation in the β-catenin binding domain (exon 1), and one missense mutation in exon 4. In the remaining six cases, nonsense or frameshift mutations were localized in the 3′ part of the gene. These latter alterations have as a common consequence to decrease the proportion of the long COOH-terminal hTCF-4 isoform, which contains two binding domains for c-terminal binding protein, a protein implicated in the repression of the TCF family transcriptional activity. Thus, loss of the TCF-4 capacity to interact with COOH-terminal binding protein could be an important event during colorectal carcinogenesis by modifying Wnt signaling.

INTRODUCTION

Recent molecular and cell biology research in the field of development and cancer has shown that the Wnt signaling pathway, also called APC[3]/β-catenin/TCF pathway, is one of the key developmental and growth regulatory mechanisms of the cell (1–3). TCF-4 knock-out mice die shortly after birth and show absence of a proliferative compartment in the prospective crypt regions between the villi (4). The level of cytosolic β-catenin is normally down-regulated by the APC tumor suppressor gene product, which promotes its degradation (5–7). In contrast, activation of the Wnt pathway causes accumulation of free β-catenin in the cytoplasm, which can then bind TCF/LEF factors (8, 9), and lead to transactivation of several target genes. This occurs during embryonic development by binding of the Wnt protein to its receptor on the cell surface, as well as during the progression of some cancers. On the other hand, recent evidence suggests that in the absence of β-catenin, TCF/LEF factors could also act as transcriptional repressors of Wnt-responsive genes (10). Other proteins were shown to act as corepressors in this process (11, 12), particularly CtBP that binds to the COOH-terminal end of the xTCF-3 protein by two binding domains well conserved throughout evolution and present in hTCF-4 (13).

Involvement of the Wnt signaling pathway has been demonstrated notably in colorectal cancers. TCF-4 (also known as TCF7L2) is the most highly expressed member of the TCF/LEF family in colonic epithelium (9). Inactivating or activating mutations of the *APC* or *β-catenin* (*CTNNB1*) genes, respectively, have been shown to result in aberrant accumulation of transactivating β-catenin/TCF-4 complexes (14), causing up-regulation of the expression of a number of genes including c-*MYC* (15). Moreover, we have recently reported frequent truncating mutations within a coding repeat sequence of *hTCF-4* in a subset of colorectal cancers characterized by microsatellite instability (MSI-H tumors; Ref. 16).

TCF-4 belongs to a family of transcription factors with homology to a region of the HMG I proteins, termed the HMG box (17, 18). To modulate transcription of target genes, TCF/LEF proteins are thought to be factors that facilitate assembly of multiprotein enhancer complexes. The HMG box of these different transcription factors is the DNA binding domain, which binds to the A/T A/T CAAAG consensus (19). The β-catenin binding domain is localized upstream, at the beginning of the coding sequence (8, 20). *TCF/LEF* mRNAs are subject to alternative splicings at their 3′ ends, which has been proposed to be important in regulating transactivational properties of the corresponding protein isoforms (21, 22). Except for the closely related *TCF-1* gene, only the mRNA sequences for the other TCF/LEF factors are available.

To better understand its multiple roles in development and in human carcinogenesis, as well as to allow more accurate screening for mutations, we analyzed the genomic structure of *hTCF-4*. Sequencing of two BACs and RT-PCR experiments has allowed us to detect 17 exon-intron boundaries in this gene. Because alternative splicing mechanisms could be essential in regulating its transactivating properties, we characterized the different *hTCF-4* mRNA species and deduced alternative exons by RT-PCR experiments. Moreover, the complete *hTCF-4* coding sequence was screened for mutations in a series of 24 colorectal cancer cell lines.

MATERIALS AND METHODS

Genomic Sequencing of the *hTCF-4* Gene. The human genomic BAC library from the CEPH was screened by PCR. Two primer pairs localized within the *hTCF-4* coding sequence were chosen, and a BAC able to amplify with these two different primer sets (*hTCF-4* BAC clone I) was isolated. Genomic insert BAC size was determined by pulsed field gel electrophoresis. This BAC was subcloned in a pUC9 plasmidic vector, and 700 subclones containing human DNA inserts were directly sequenced after PCR with universal M13 primers. Big dye terminator cycle sequencing was performed with the ABI PRISM kit and AmpliTaq DNA polymerase and analyzed with an ABI

Received 12/3/99; accepted 5/15/00.

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.

[1] This work was supported in part by Grant 9624 from Association pour la Recherche sur le Cancer. A. D. was recipient of a fellowship from the Ligue Nationale Contre le Cancer.

[2] To whom requests for reprints should be addressed, at INSERM U434-CEPH, 27 rue Juliette Dodu, 75010 Paris, France. Phone: 33-1-53-72-51-09; Fax: 33-1-53-72-51-58; E-mail: richard.hamelin@cephb.fr.

[3] The abbreviations used are: APC, adenomatous polyposis coli; TCF, T-cell transcription factor; hTCF-4, human TCF-4; LEF, lymphocyte enhancer factor; CtBP, c-terminal binding protein; HMG, high mobility group; BAC, bacterial artificial chromosome; RT-PCR, reverse transcription-PCR; CEPH, Center d'Etude du Polymorphisme Humain; DGGE, denaturing gradient gel electrophoresis; EST, expressed sequence tag.

PRISM 377 DNA sequencer (Applied Biosystems). Contigs of shotgun sequences were obtained using the Phred/Phrap/Consed system (23–25), and exon-intron boundaries were deduced by comparison with the *hTCF4* mRNA sequences (GenBank accession no. Y11306). Because the 5′ *hTCF-4* exons were not included in the *hTCF-4* BAC clone I, another BAC (*hTCF-4* BAC clone II) was isolated from the CEPH BAC library after the PCR with primer sets localized to the beginning of the *hTCF-4* coding sequence. Exon-intron junctions of exons contained within this second BAC clone were determined by "primer walking" on both strands using direct sequencing.

Identification of Different Alternative *hTCF-4* mRNAs by RT-PCR Analysis. Total RNA from a series of 24 colorectal cancer cell lines was extracted with Trizol LS (Life Technology) according to manufacturer instructions and reverse transcribed using random hexamers and a Perkin-Elmer kit (Roche Molecular Systems). PCR with a number of primer pairs chosen for almost all of the *hTCF-4* exons were performed, and the resulting products were revealed by ethidium bromide staining in agarose gel. Unique and alternative bands from each RT-PCR products were extracted from the agarose gels and sequenced.

Identification of Alternative Splicings in the 3′ Region of *hTCF-4* between Exons 12 and 17. RT-PCR products obtained with a sense primer in exon 12 (5′-AGGGACAAGCAGCCGGGAGA-3′) and an antisense primer in exon 17 (5′-TGGGCTGAGGCAGCTGCCTT-3′) of the *hTCF-4* gene were separated on a 7 M urea/32% formamide/7% polyacrylamide gel, transferred overnight onto a Hybond N+ nylon membrane (Amersham), and sequentially hybridized with ^{32}P-labeled probes specific for exons 12 to 17. Between each round of hybridization, the membrane was stripped by washing with 0.5 M NaOH and checked for the complete probe removal by autoradiography.

Screening for *hTCF-4* Mutations by DGGE and Sequencing. We looked for *hTCF-4* mutations in a series of 24 colorectal cancer cell lines that were already characterized for other genetic alterations (26, 27). The 17 exons, together with most intron-exon junctions, were amplified by PCR using oligonucleotides as shown in Table 1 and then either directly sequenced or analyzed for mutations by DGGE. DGGE running conditions were defined using the Lerman and Silverstein program (Ref. 28; Table 1) and performed as described (29). Samples showing altered DGGE profiles were sequenced using an ABI Prism 377 DNA sequencer.

RESULTS

Sequencing of the *hTCF-4* Gene. To determine *hTCF-4* genomic structure, we first sequenced the *hTCF-4* BAC clone I that contains a human genomic insert of 120 kb in size. A human genomic DNA plasmid library derived from this BAC was constructed containing 700 pUC9 subclones with genomic inserts from 2 to 4 kb in size (10 genome equivalents). These subclone inserts were sequenced by PCR with M13 universal primers. Four contigs were obtained containing nine *hTCF-4* coding exons, as deduced from known *hTCF-4* cDNA sequences (9). These contigs were interrupted, probably because of the presence of repetitive elements in genomic DNA. One internal region of the *hTCF-4* cDNA was missing from these contigs and was shown to contain a single exon by "primer walking" on both strands using direct sequencing of the BAC clone I. *hTCF-4* cDNA sequences localized to the 5′ part of the gene were also missing in BAC clone I. Another BAC clone containing these sequences (*hTCF-4* BAC clone II) was isolated from the CEPH human BAC library using additional probes, and intron-exon junctions of four additional *hTCF-4* exons were determined by primer walking.

A total of 14 exons were thus identified, each with canonical acceptor and donor sites, as shown in Fig. 1. These exon sequences matched perfectly to the previously published *hTCF-4E* mRNA sequence (9). Additional intron sequences are deposited in the EMBL database (accession nos. AJ270770 to AJ270778).

Characterization of Alternative Exons and Splice Sites. Using different primer sets chosen in *hTCF-4* exons as shown in Fig. 2, we amplified a number of mRNAs by RT-PCR in a series of colorectal

Table 1 *Primers used for the mutational screening of the entire TCF-4 coding sequence*

TCF-4 exon numbering	5′ and 3′ primer sequences[a]	Annealing	PCR length (bp)	Mutational screening
1.	5′-AATTGCTGCTGGTGGGTGA-3′			
	5′-CCCGAGGGGCTTTTCCTA-3′	58°C	236	PCR sequence
2–3.	5′-AATCGTCTCTAGTCAATGAA-3′			
	5′-ACTGCAATTTGGTGTTCAAT-3′	54°C	313	RT-PCR sequence
4.	5′-(a)GAACGCTTTGATTTGGTTTC-3′			DGGE
	5′-GCTTCAGAATCTCTTGCGT-3′	56°C	184	30–80% 3h
5.	5′-ATTTTCAGTCCGGCAGCA-3′			
	5′-TGCGGAGGCCGTGGGATT-3′	56°C	319	RT-PCR sequence
6.	5′-(a)GTGTCTTTCTCTGTTCTCC-3′			DGGE
	5′-AGGACCAGCACTCTACCTT-3′	56°C	262	30–80% 4h
7.	5′-TCTCCCTCCTTTGTACCTA-3′			DGGE
	5′-(b)AACATGCCCATGTAAGAAAC-3′	56°C	240	30–80% 3h
8.	5′-(a)CAAGTCTCTTACTTTGTACC-3′			DGGE
	5′-AGAACTTCTTGAGAAGTCAC-3′	56°C	238	30–80% 3h
9.	5′-(a)GATTCTGACGATTTACACAG-3′			DGGE
	5′-GCTACGAAGAAGGTGAGAA-3′	56°C	256	30–80% 3h
10A.	5′-TTTATTGGGTTGCGTCTGTT-3′			
	5′-ACCTTTGCTCTCATTTCCTT-3′	57°C	140	PCR sequence
10B.	5′-CCTCTTAATGCATTCATGTTG-3′			
	5′-ACTCTGGTCCCTCACCAC-3′	58°C	160	PCR sequence
11.	5′-TCTCTTTGTTCATGTCTTTCT-3′			DGGE
	5′-(b)TCTACAGACTCCAATCCTAA-3′	56°C	240	30–80% 3h
12.	5′-(a)TTCTTGCAGTAATTCCCTGTG-3′			DGGE
	5′-TCCTCTAACCTGATGATTGTC-3′	60°C	175	30–80% 3h
13.	5′-(a)GAAGGCTTTGTATAATTTGTTC-3′			DGGE
	5′-CTATTTTAATGACTCAAAATGAC-3′	58°C	175	20–70% 3h
14.	5′-ACGTTCTCCTCCTCTGCT-3′			DGGE
	5′-(b)ACCAGCACAGCGAATCTG-3′	56°C	198	20–70% 3h
15.	5′-(a)CGACCCACCATTGTGTTGTA-3′			DGGE
	5′-AAAGGCCTCGCAGTGGTAAT-3′	60°C	204	30–80% 2h
16.	5′-AGCTCCTGTCTCATTTCCTT-3′			
	5′-GACGAGGGAGGCACTACAAA-3′	58°C	156	PCR sequence
17.	5′-GCCTCTATTCACAGATAACTC-3′			
	5′-CAGCGGGGTTCACGACGCT-3′	58°C	478	PCR sequence

[a] GC clamps: (a), CCCCGCCCGGCCCGCCCCGCCCCCCGCCCCCCCTCCCGGCCCGCCCCCCTGGCGCCCCGC; (b), CCCCACGCCACCCGACGCCCCAGCCCGACCCCCCCGCGCCCGGCGCCCCCGC.

AAAATTGCTGCTGGTGGGTGAAAAAAAAATGCCG...Exon 1..(189 bp in cds).....**TCCGAG**gtaggaaaagcccctcgggctggtgg
gccctccccacctccacccctctgggaag**GCGGAA**..Exon 2..(67 bp).....**AAGAAG**gtgagtacgccccgcgcgccccgcag
gctgtcccctcgccgccccgccatgttag**CGGCCA**..Exon 3..(125 bp).....**CGAACC**gtaagtgcctccgcgcccggcccccg
gaacgctttgatttggtttctttctacag**CTCCAT**..***Exon 4..(69 bp)*.....**GTTCAG**gtaggaaacgcaagagattctgaagc
ttcatttctcttttgtttcctctcgccag**TATCTC**..Exon 5..(102 bp).....**ATTGTC**gtaagtaacctcccagagatgatggc
ggtgtctttctctgttctcctccccacag**TCTAAC**..Exon 6..(133 bp).....**AAACAG**gtaggctgtgggctacggagccaagg
tgtacctaaaatgctgcttcttccctcag**GAATCC**..Exon 7..(*91*/103 bp).....**AGGATGGTTAGTACCACA**gtaaggagttccatttttttaatttcc
gtctcttactttgtacccctttctttgtag**GCAAGG**..Exon 8..(87 bp).....**GTCCAG**gtgagttccaagaaccggggccttca
gacgatttacacag*CTTTCTGTCTTCTAG***GTTCCC**..Exon 9..(126/*141 bp*).....**TAGTTC**gtaagtgttgctgtttttctcaccttt
tgggttgcgtctgttttgtctatctccag**AAAGCA**..Exon 10..(160 bp).....**CGGAGG**gtaggtgacgccctttctcagggagaa
ttctcatctgtaccccacgtcccctccag**TGGCAT**..Exon 11.(108 bp).....**AACTAT**gtaggtggatcattttcgttaggatt
aattccctgtgtttcatctcttcatctag**GGAAAG**..Exon 12.(49 bp).....**CCAATG**gtaagtgacaatcatcaggttagagg
gctttgtataatttgttctttttttttcag**AACACA**..*Exon 13.(51 bp)*.....**TTACAG**gtgctaatgtcattttgagtcattaa
ctgctcgcttctctcttgaactcattcag**ACCTCA**..*Exon 14.(73 bp)*.....**TTGCAG**gtgtgtatagttttccagattcgctg
tatttttgtgtttaccttatgctaacag**ATGCAA**..***Exon 15.(73 bp)*.....**GTGCAG**gtgcaggtatattaccactgcgaggc
gtgttgtttctttgttctgatacaag*CAG***TCTTTG**..***Exon 16.(25/28 bp)*.....**ACAATG**gtaggtatttcaacaccctaaacacg
tcacag*ATAACTCTCTCCCCTGTTTCTAG***GAGAAA**..Exon 17.(418/*441 bp* in cds).....**GAATAG**ATTCGCCTCCCCCCTCCCCGAACCT

***Exon 4* (complete CDS)**

CTCCATTTTCAGTCCGGCAGCACACATTACTCTGCGTACAAAACGATTGAACACCAGATTGCAGTTCAG

***Exon 15* (complete CDS)**

ATGCAAATACTCCAAAGAAGTGTCGGGCACTGTTCGGGCTTGACCGACAGACTTTATGGTGCAAACCGTGCAG

***Exon 16* (complete CDS)**

CAGTCTTTGAATTTGGAATATTACAATG

Fig. 1. Exon/intron boundary sequences of the *hTCF4* gene. Intronic and exonic sequences are written in small and capital letters, respectively. Exons are numbered 1 to 17, in *italics* when alternative, and in *bold* characters when not reported in previously published works. In the latter case, complete sequences of these exons are indicated. The ag/gt dinucleotides at intron/exon boundaries are *underlined*, and those normally used are *vertically aligned*. Sequences that are added or removed from the coding sequence because of the use of alternative donor and acceptor splice sites are written in *italics*.

cancer cell lines and analyzed the different products on agarose gels. In most cases, a unique PCR product with the predicted size was observed. In cases indicated by framing on Fig. 2, longer fragments were also reproducibly obtained in addition to the predicted DNA fragments. These longer DNA fragments were excised from agarose gels and sequenced. One of these sequences was localized in the 5′ part of the gene and was extended by primer walking on the BAC clone II. Another was found to be present in one of the genomic DNA contigs. In both cases, these sequences were flanked by consensus splicing sequences, and hence we deduced that they corresponded to alternative exons 4 and 15 expressed in all of the cell lines tested. A third additional cDNA sequence was present in one of the genomic contigs just upstream of exon 9, determined by the use of an alternative splice acceptor site in this exon.

Finally, an additional cDNA fragment was shown to be present in the COLO 320 cell line and absent in all other cell lines, except when more sensitive techniques were used (data not shown). The corresponding sequence was present in one of the contigs, flanked by canonical splice sites, and named exon 16.

Fig. 3 is a summary of the *hTCF-4* genomic structure, which is thus composed of 17 exons (with exon 16 mainly expressed in the COLO 320 cell line). In previously published work (16, 30), when the genomic structure of the *hTCF-4* gene was only partially known, we numbered some selected putative exons as deduced from the nomenclature of the closely related *hTCF-1* gene. We now indicate in Fig. 3 the relation between this previous numbering and the proposed new one.

Characterization of *hTCF-4* 3′ Alternative Splicing Products. Because the profiles of RT-PCR products between exons 12 and 17 showed a complicated pattern on agarose gel, we decided to analyze

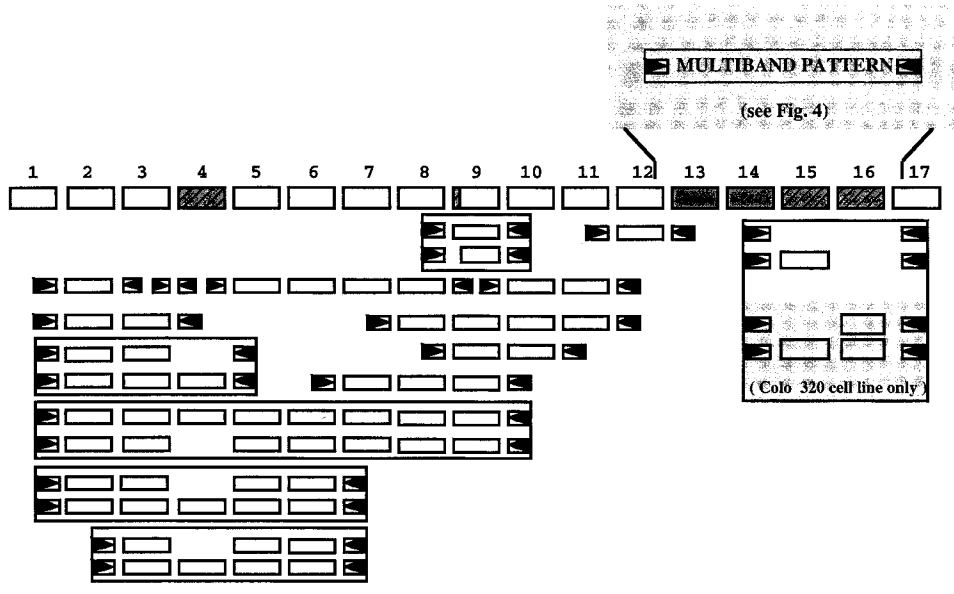

Fig. 2. RT-PCR analysis of the *hTCF-4* gene. Sense and antisense primers were chosen in the *hTCF-4* exons as indicated by *arrows*. PCR experiments giving more than a single DNA fragment are framed. Alternative exons are shown in *gray*, whereas exons corresponding to previously unknown *hTCF-4* sequences are *striped*. An alternative splice acceptor site in exon 9 is also indicated.

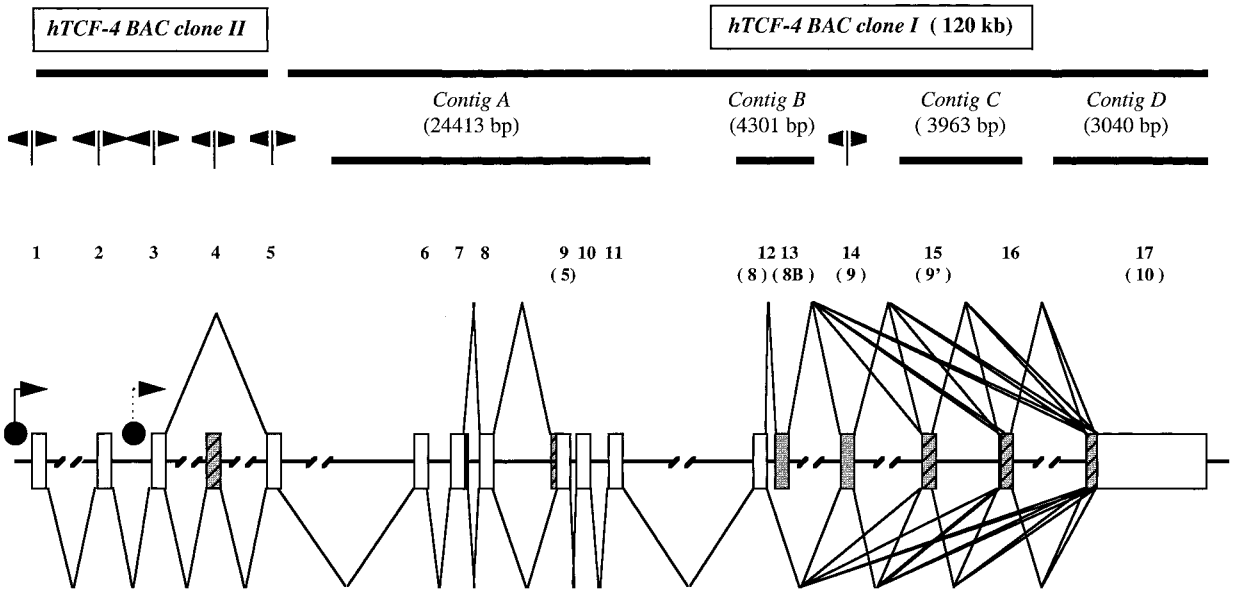

Fig. 3. Summary of the *hTCF-4* genomic organization and its alternative splicings. The extent of BAC clones and plasmid contigs as compared with the genomic sequence is represented at the *top* of the figure. Double direction *arrows* indicate exon/intron boundaries sequenced by primer walking. Alternative exons are *gray*, and exons corresponding to unknown *hTCF-4* sequences are *striped*. Previously used exon numbering is indicated in *parentheses*. The main splicing events occurring between exons 12 and 17 are represented differently in Fig. 6. The known *hTCF-4* promoter and the putative second one (*dotted-lined*) are depicted by *black dots* with *arrows*.

this further by the higher resolution approach of electrophoresis on polyacrylamide denaturing gels, Southern blot transfer, and successive hybridization with probes specific for each exon as shown in Fig. 4. This experiment showed that exons 13, 14, and 15 were three consecutive alternative exons with all eight possible combinations observed (12-17, 12-13-17, 12-14-17, 12-15-17, 12-13-14-17, 12-13-15-17, 12-14-15-17, and 12-13-14-15-17). Exon 16 was detected mainly in the COLO 320 cell line (bands marked by an *asterisk*), and weakly in the other cell lines with an exon 16-specific probe (data not shown). Moreover, an alternative splice acceptor site located 3 bp upstream of the main site was also occasionally used in COLO 320, as shown by the *arrows* in Fig. 4. Finally, in all cell lines we also observed faint additional bands, which were attributable to the use of an alternative splice acceptor site that added 23 bp to exon 17, as revealed by a specific probe (data not shown).

ESTern E-Blot Determination of an Additional *hTCF-4* Splice Site. We then performed a BLAST search at the Swiss Institute of Bioinformatics for homology in EST databases using the complete *hTCF-4* coding sequence (including the newly characterized exons) as a query. We propose to name this kind of electronic Northern blot as an ESTern e-blot. The 25 ESTs that were detected do not cover the entire *hTCF-4* gene, but a high redundancy was observed in the central part of the molecule (Fig. 5). A close comparison of the sequences of these ESTs, together with genomic and cDNA *hTCF-4* sequences, revealed that the new experimentally detected exon 4 was found to be linked to exons 5–9 in human EST AA417665. The alternative acceptor site in exon 9 was also detected *in silico*, because it was used in 3 of 15 ESTs covering this region. Finally, a new alternative splice donor site in exon 7 (12 bp) was evidenced, as shown by EST AA609985.

Screening for *hTCF-4* Mutations in a Series of Colorectal Cancer Cell Lines. Because we had identified the complete sequence of *hTCF-4* exons and intron-exon boundaries, we screened the entire gene for mutations in a series of 24 colorectal cancer cell lines by direct sequencing for exons 1, 2, 3, 5, 10, 16, and 17 and by DGGE for other exons. Because of its differential expression, exon 16 was only sequenced in the COLO 320 cell line and in nine other cell lines.

A total of 12 DNA variants was observed. As shown previously (16), four cell lines with microsatellite instability (LS 174T, LoVo, TC71, and TC7) had a 1-bp deletion in an (A)9 repeat localized at the beginning of exon 17 (exon 10 according to the previous *TCF-1*-like numbering). In the cell line LS 1034, we observed a 2-bp deletion in the GAGA nucleotides at coding position 112–115 within exon 1. The FET cell line contained a C to A transversion at nucleotide 62 of exon 15, leading to a stop codon in one of the used reading frames. We were unable to find a mutation in the flanking intronic sequences of exon 16 of COLO 320, which could explain the abnormally high level of this alternative splicing in this cell line. A C to T transition at the nucleotide 35 of exon 4 was present in the SW 48 cell line, leading to an alanine to valine change. Finally, four noncoding alterations were observed: a T to G transversion and a deletion of one T in a 7T-run in the polypyrimidine tracts of the splice acceptor sites of exon 13 in HCT-15 and exon 14 in Co-115 cell lines, respectively, and a 1-bp deletion in the (A)9 repeat preceding (and including) the initiation ATG codon in two MSI-H cell lines (TC71 and HCT-116). In all cases but two (LoVo and TC71 in the exon 17 coding repeat), only one allele was altered.

DISCUSSION

A link was recently established between TCFs and Wnt signaling, a pathway that plays a crucial role in many developmental processes as well as in human carcinogenesis (1, 2). In the present work, we describe the genomic structure of an important member of the *TCF/LEF* gene family, the human *TCF-4* gene that we have recently mapped on chromosome 10q25.3 by FISH analysis (31). This study also highlights the complexity of *hTCF-4* mRNA transcripts. Indeed, we report here a large number of alternative splicings in the 3′ part of this gene, which generate several TCF-4 isoforms differing at their COOH-terminal ends. This mechanism is expected to be of functional significance because it has been shown that TCF factors were able to interact functionally by COOH-terminal binding domains with CtBP, a corepressor protein required to mediate transcriptional repression of the TCF family activity. Finally, because the APC/β-catenin/TCF

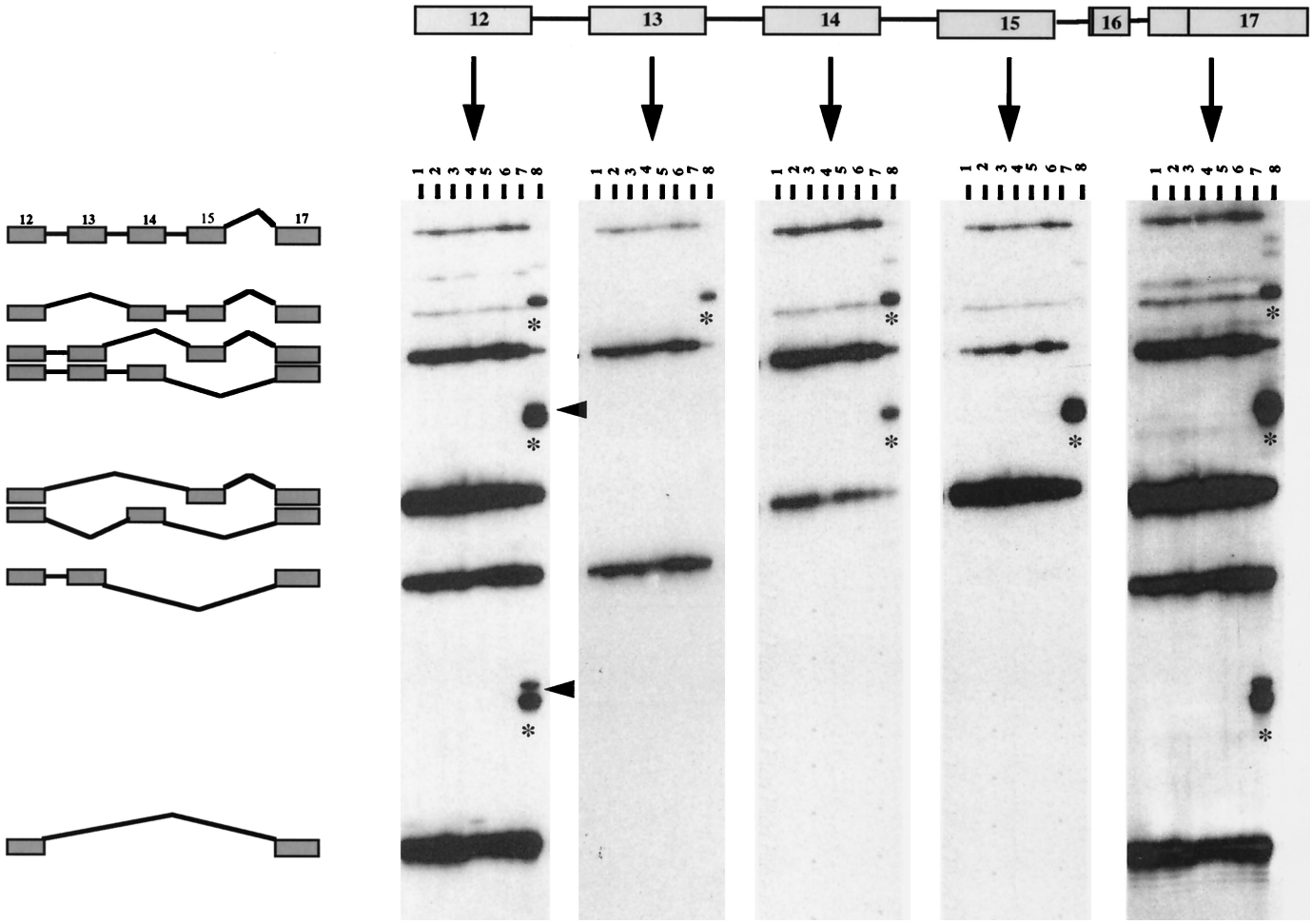

Fig. 4. Analysis of alternative splicing between exons 12 and 17. RT-PCR products obtained from eight colorectal cancer cell lines (*Lane 1,* GLY; *Lane 2,* LS174T; *Lane 3,* LoVo; *Lane 4,* TC71; *Lane 5,* LS1034; *Lane 6,* HT29; *Lane 7,* V9P; *Lane 8,* COLO 320) using a sense primer in exon 12 and an antisense primer in exon 17 were run on a denaturing polyacrylamide gel, transferred onto a nylon membrane, and sequentially hybridized with ^{32}P-labeled, exon-specific probes as indicated. *Left,* the deduced splicing profiles of each mRNA species (as confirmed by sequencing). ✻, COLO 320-specific mRNA species containing exon 16. All other cell lines showed the same profiles (not shown).

pathway is known to play an important role in colorectal carcinogenesis, we screened the entire *hTCF-4* gene for mutations in a series of 24 colorectal cancer cell lines and found DNA coding alterations in 33% of the cases ($n = 8$).

We have obtained experimental evidence for a total of 17 exons within *hTCF-4* of human cell lines. Two different human *hTCF-4* mRNA transcripts, named *hTCF-4b* and *hTCF-4e,* have already been described (9). According to our proposed exon nomenclature, they contain exons 1,2,3,5–12,17 and 1,2,3,5–14,17, respectively. We therefore describe three additional exons (4, 15, and 16), although a polypeptide potentially encoded by exon 15 has been reported previously with no further details (32). It has to be noted that exons 14 and 15 have the same size (73 bp) and show 66% homology at both the nucleotide and amino acid levels, suggesting duplication during evolution.

Exons 4, 13, 14, 15, and 16 are alternative exons, whereas three alternative splice acceptor sites are used in exons 9, 16, and 17. This creates 256 theoretically different splice products, making *hTCF-4* expression more complicated than *hTCF-1* with its 96 potential different mRNA species (20). This number may even be increased to >500 different forms if the *in silico*-detected alternative splice site in exon 7 doubles the number of different mRNA transcripts. The use of exon 4 and the alternative exon 9 acceptor site appear as frequent as their skipping; however, we did not examine whether they were exclusive of each other and whether they were conserved with alternative splicings downstream of exon 12. Moreover, their presence or absence does not change the reading frame.

Alternative mRNAs between exons 12 and 17 were more extensively analyzed by RT-PCR experiments. Eight mRNAs were easily detected and sequenced. They contain: exons 12,17 (*hTCF-4b*); 12,13,17; 12,14,17; 12,15,17; 12,13,14,17 (*hTCF-4e)*, 12,13,15,17; 12,14,15,17 and 12,13,14,15,17 (Fig. 4 and 6). As a consequence of this baroque splicing mechanism, exon 17 is either noncoding ($n = 2$; because of the presence of an upstream in-frame stop codon when exons 14 and 15 are spliced together) or encodes for medium (25 amino acids; $n = 2$) or long (138 amino acids; $n = 4$) hTCF-4 COOH-terminal isoforms, depending of which reading frame is used.

TCF-4 and other members of the TCF/LEF family have conserved sequences corresponding to functional domains (2). The β-catenin binding domain (in exon 1, according to the present exon numbering) and DNA-binding HMG box (exons 10 and 11) of *hTCF-4* are highly conserved with other TCF/LEF family members in human as well as other species including Drosophila (Pangolin), Xenopus (XTcf-3), chicken (chTCF), and mouse (mTCFs). Outside of these domains, a high conservation is observed for the final exon 17. For this exon, there is: (*a*) a complicated splicing pattern that has been conserved in murine *TCF-4* as well as in the related human *TCF-1* gene; (*b*) two reading frames, long or medium in size, that are also well conserved in Pangolin (Drosophila) and POP-1 (*Caenorhabditis elegans*) or in chicken TCF, respectively (2); (*c*) two CtBP binding motifs only

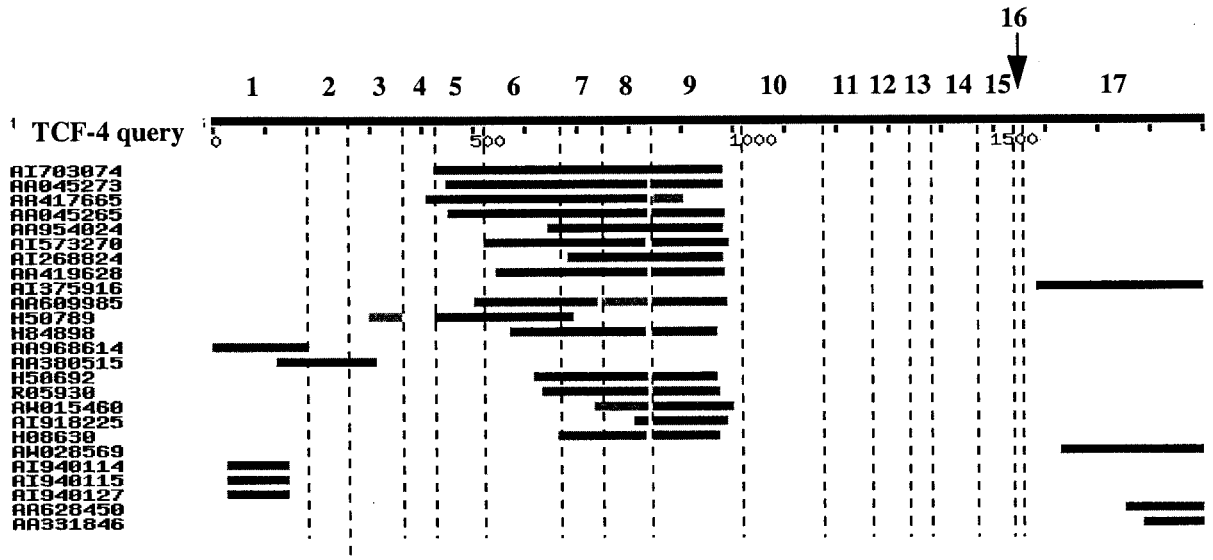

Fig. 5. ESTern e-blot analysis of the *hTCF-4* gene. The result of a BLAST search on EST databases is represented, and accession numbers of the ESTs giving a score >200 are shown. The query was the longer *hTCF-4* mRNA sequence, including all alternative exons and splice sites as deduced from our experiments. An interruption of the line representing an EST indicates a sequence present in the query sequence and absent in the EST. Sequence alignments were also checked for sequences present internally in ESTs and absent in the query sequence; however, none were observed.

present in the long-size COOH-terminal hTCF-4 isoforms that were conserved throughout evolution in the TCF family and may play a role in down-regulating TCF transactivating properties. Together, these observations suggest that alternative use of different reading frames in the 3′ part of *hTCF-4* should generate different protein isoforms with agonist or antagonist transactivating activities.

We detected 12 *hTCF-4* DNA variants in a series of 24 colorectal cancer cell lines. Four are noncoding with unknown consequences, if any, on the TCF-4 protein function. The other eight variants were in coding sequences. A missense alanine to valine alteration in the alternative exon 4 was observed in the SW48 cell line. It is of unclear significance and could be a coding polymorphism. Another variant is a frameshifting, 2-bp deletion within the β-catenin binding domain in exon 1 in the LS 1034 cell line. It is known that the closely related *hTCF-1* gene has two different initiation sites, leading to the synthesis of long and short NH_2-terminal TCF-1 isoforms, and it has been reported that the short NH_2-terminal isoforms act as negative regulators of transcription (20). In the *hTCF-4* gene, no initiation at internal ATG has been demonstrated, although a ΔN-TCF4 truncated construct has been shown to have a dominant-negative effect over the full-length hTCF-4 product bound to β-catenin (9). We noticed that another in-frame ATG fulfilling the minimum Kosak criteria of an A at −3 is present in exon 3 of the *hTCF-4* gene. Preliminary observations indicate that the *hTCF-4* gene could also have an alternative transcriptional initiation site. Indeed, the sequence corresponding to intron 2 of the already known *hTCF-4* mRNAs is 208 bp long and has a GC content of 82% with a CpG/GpC = 1 and a putative TATA box (see AJ270771). Moreover, we were able to amplify and sequence a *hTCF-4* cDNA product containing exons 3 to 6 with a sense primer localized between the TATA box and the beginning of exon 3 in the 5′ untranslated region of the putative alternative transcript (data not shown). In that context, the LS1034 frameshift mutation should be interpreted as a selected event that favors the expression of hTCF-4 isoforms lacking a β-catenin binding domain with an unexpected dominant-negative effect.

The other *hTCF-4* mutations are a 1-bp deletion in an (A)9 repeat within exon 17 in four cell lines, which have been already reported in a large series of primary MSI-H colorectal cancers (16, 30), a nonsense mutation in exon 15 in FET, and a high usage of alternative exon 16 in COLO 320. They are all localized in the 3′ part of the *hTCF-4* gene and are truncating mutations. The common result of all these mutations is to change the reading frames used in exon 17 in the

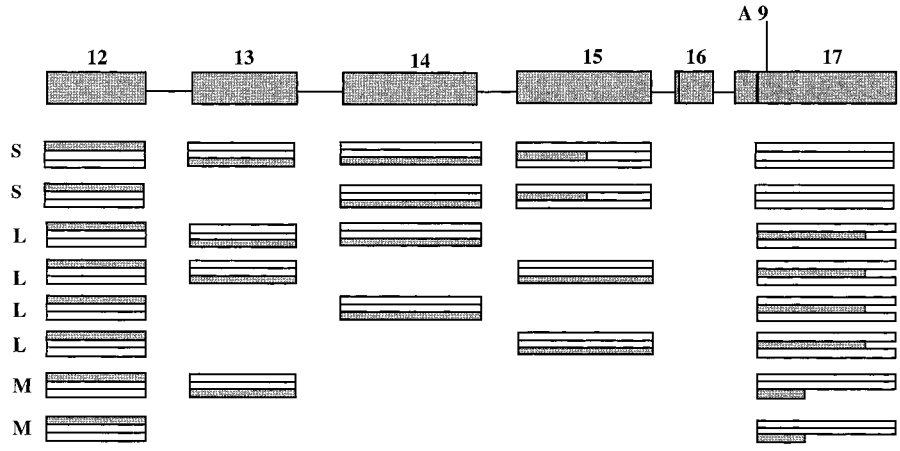

Fig. 6. Reading frames used in alternative exons between exons 12 and 17. The reading frames used in each exon are indicated in *gray*. Exons 13, 14, 15, and 16 have 51, 73, 73, and 25 bp, respectively. The alternative splicing events of exons 14 and 15 change the reading frames used in exon 17 as indicated. For the different COOH-terminal ends, encoded by the corresponding alternative mRNA species: *S*, short; *M*, medium; *L*, long.

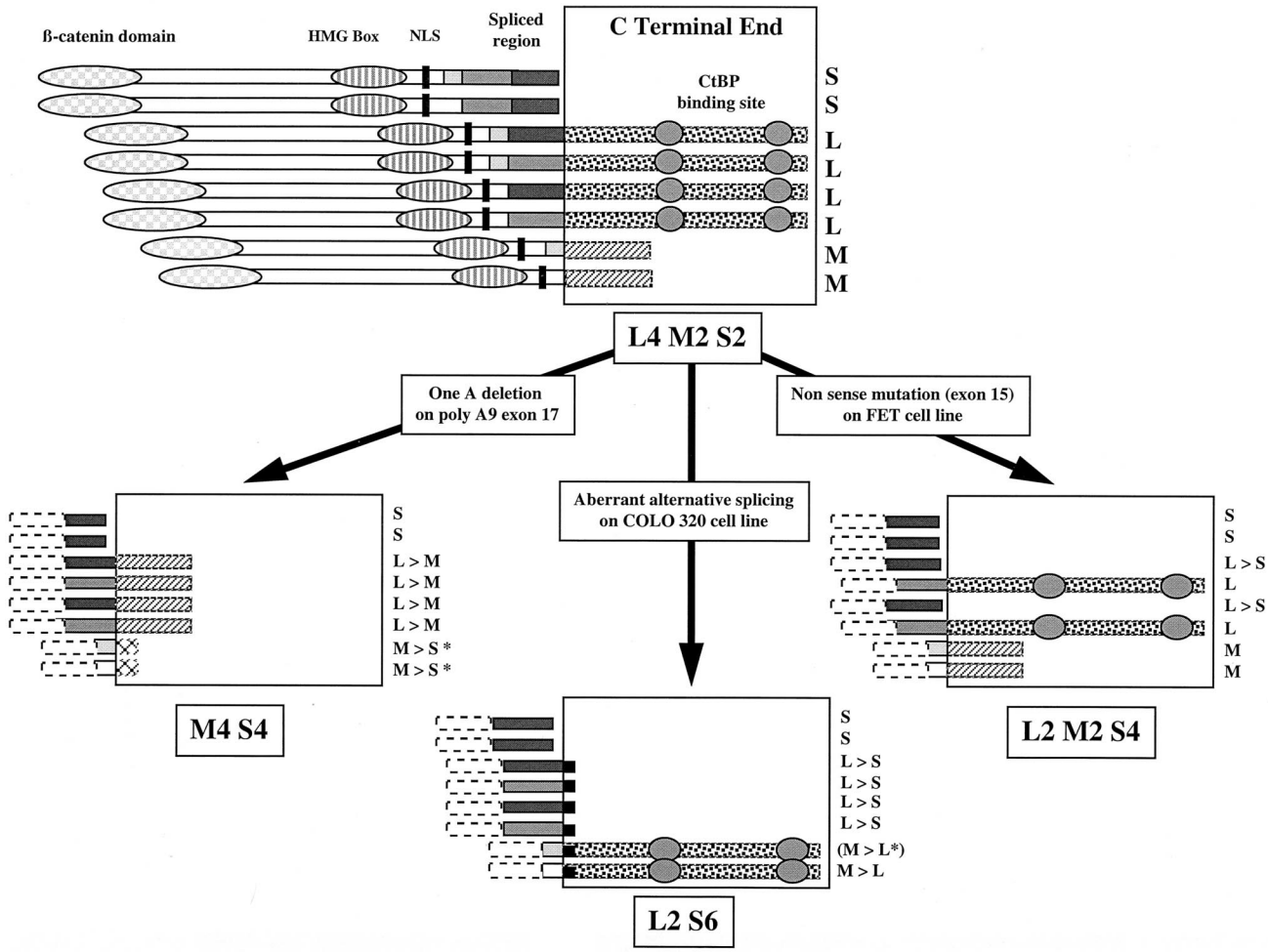

Fig. 7. Consequences of *hTCF-4* mutations on the eight alternative TCF-4 isoforms encoded by the different mRNA species. The main functional domains of hTCF-4 are indicated. COOH-terminal ends of the eight isoforms encoded by the eight differently spliced mRNA species are framed. The *S* (short), *M* (medium), and *L* (long) COOH-terminal ends of each isoform are indicated. The two CtBP binding sites present in long isoforms are represented by *gray circles*. The total content of short, medium, and long isoforms in nonmutated and mutated cell lines are indicated in *boxes* with the same lettering system. The mRNA encoding the *L** form in the COLO 320 cell line has not been experimentally detected.

corresponding cell lines. As a consequence, hTCF-4 isoforms with either a noncoding, a short, or a medium COOH terminus are enriched, whereas the long hTCF-4 isoforms either disappear or are reduced in their expression level (Fig. 7). If the relative proportion of these different COOH-terminal hTCF-4 isoforms plays a role in regulating hTCF-4 activity, such changes may alter the resulting hTCF-4 activity in cell lines harboring these mutations. Notably, all of these alterations favor the synthesis of hTCF-4 isoforms lacking their capacity to interact with CtBP, a protein implicated in the transcriptional repression mediated by the TCF COOH-terminal domain. It should be noted that all cell lines containing *hTCF-4* mutations in coding sequences were also mutated in either the *APC* or the *β-catenin* genes.[4] Thus, *hTCF-4* alterations could constitute a second mutational event frequently selected in these cell lines to reactivate Wnt signaling. Further analysis will be necessary to confirm this hypothesis and to functionally characterize the consequences of each of these alterations on TCF/β-catenin target gene expression.

ACKNOWLEDGMENTS

We thank Drs. Barry Iacopetta and Elizabeth Newcomb for critical reading of the manuscript.

[4] R. Hamelin, J. Gayet, X. P. Zhou, A. Duval, J. M. Hoang, P. Cottu, and G. Thomas. Extensive characterisation of genetic alterations in a series of human colorectal cancer cell lines, manuscript in preparation.

Note Added in Proof

During the review process of this manuscript, an additional chromosome 10q genomic sequence was reported by the Sanger Center, Hinxton, under GenBank Accession number AL135792. This sequence contains exons 6–17 of *hTCF-4* and shows that intron 5 is at least 60 kb long.

REFERENCES

1. Peifer, M. *β-Catenin* as oncogene: the smoking gun. Science (Washington DC), *275:* 1752–1753, 1997.
2. Clevers, H., and van de Wetering, M. TCF/LEF factors earn their wings. Trends Genet., *13:* 485–489, 1997.
3. Barker, N., Morin, P. J., and Clevers, H. The yin-yang of TCF/β-catenin signaling. Adv. Cancer Res., *77:* 1–24, 2000.
4. Korinek, V., Barker, N., Moerer, P., van Donselaar, E., Huls, G., Peters, P. J., and Clevers, H. Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. Nat. Genet., *19:* 379–383, 1998.
5. Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B., and Polakis, P. Regulation of intracellular β-catenin levels by the adenomatous polyposis coli (APC) tumor-suppressor protein. Proc. Natl. Acad. Sci. USA, *92:* 3046–3050, 1995.
6. Rubinfeld, B., Souza, B., Albert, I., Muller, O., Chamberlain, S. H., Masiarz, F. R., Munemitsu, S., and Polakis, P. Association of the *APC* gene product with β-catenin. Science (Washington DC), *262:* 1731–1734, 1993.
7. Su, L. K., Vogelstein, B., and Kinzler, K. W. Association of the APC tumor suppressor protein with catenins. Science (Washington DC), *5140:* 1734–1737, 1993.
8. Behrens, J., von Kries, J. P., Kuhl, M., Bruhn, L., Wedlich, D., Grosschedl, R., and Birchmeier, W. Functional interaction of β-catenin with the transcription factor LEF-1. Nature (Lond.), *382:* 638–642, 1996.
9. Korinek, V., Barker, N., Morin, P. J., van Wichen, D., de Weger, R., Kinzler, K. W., Vogelstein, B., and Clevers, H. Constitutive transcriptional activation by a β-catenin-

Tcf complex in APC−/− colon carcinoma. Science (Washington DC), *275:* 1784–1787, 1997.
10. van de Wetering, M., Cavallo, R., Dooijes, D., van Beest, M., van Es, J., Loureiro, J., Ypma, A., Hursh, D., Jones, T., Bejsovec, A., Peifer, M., Mortin, M., and Clevers, H. Armadillo coactivates transcription driven by the product of the *Drosophila* segment polarity gene dTCF. Cell, *88:* 789–799, 1997.
11. Roose, J., Molenaar, M., Peterson, J., Hurenkamp, J., Brantjes, H., Moerer, P., van de Wetering, M., Destree, O., and Clevers, H. The *Xenopus* Wnt effector XTcf-3 interacts with Groucho-related transcriptional repressors. Nature (Lond.), *395:* 608–612, 1998.
12. Cavallo, R. A., Cox, R. T., Moline, M. M., Roose, J., Polevoy, G. A., Clevers, H., Peifer, M., and Bejsovec, A. *Drosophila* Tcf and Groucho interact to repress Wingless signaling activity. Nature (Lond.), *6702:* 604–608, 1998.
13. Brannon, M., Brown, J. D., Bates, R., Kimelman, D., and Moon, R. T. XCtBP is a XTcf-3 co-repressor with roles throughout *Xenopus* development. Development (Camb.), *126:* 3159–3170, 1999.
14. Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B., and Kinzler, K. W. Activation of β-catenin-Tcf signaling in colon cancer by mutations in β-catenin or APC. Science (Washington DC), *275:* 1787–1790, 1997.
15. He, T. C., Sparks, A. B., Rago, C., Hermeking, H., Zawel, L., da Costa, L. T., Morin, P. J., Vogelstein, B., and Kinzler, K. W. Identification of c-MYC as a target of the APC pathway. Science (Washington DC), *281:* 1509–1512, 1998.
16. Duval, A., Gayet, J., Zhou, X. P., Iacopetta, B., Thomas, G., and Hamelin, R. Frequent frameshift mutations of the *TCF-4* gene in colorectal cancers with microsatellite instability. Cancer Res., *59:* 4213–4215, 1999.
17. Kelly, M., Burke, J., Smith, M., Klar, A., and Beach, D. Four mating-type genes control sexual differentiation in the fission yeast. EMBO J., *7:* 1537–1547, 1988.
18. Staben, C., and Yanofsky, C. *Neurospora crassa*: a mating-type region. Proc. Natl. Acad. Sci. USA, *87:* 4917–4921, 1990.
19. van de Wetering, M., Oosterwegel, M., Dooijes, D., and Clevers, H. Identification and cloning of TCF-1, a T lymphocyte specific transcription factor containing a sequence-specific HMG box. EMBO J., *10:* 123–132, 1991.
20. van de Wetering, M., Castrop, J., Korinek, V., and Clevers, H. Extensive alternative splicing and dual promoter usage generate Tcf-1 protein isoforms with differential transcription control properties. Mol. Cell. Biol., *16:* 745–752, 1996.
21. van de Wetering, M., Oosterwegel, M., Holstege, F., Dooyes, D., Suijkerbuijk, R., van Kessels, A. G., and Clevers, H. The human T cell transcription factor-1 gene. J. Biol. Chem., *267:* 8530–8536, 1992.
22. Mayer, K., Wolff, E., Clevers, H., and Ballhausen, W. G. The human high mobility group (HMG)-box transcription factor TCF-1: novel isoforms due to alternative splicing and usage of a new exon IXA. Biochem. Biophys. Acta, *1263:* 169–172, 1995.
23. Gordon, D., Abadjian, C., and Green, P. Consed: a graphical tool for sequence finishing. Genome Res., *8:* 195–202, 1998.
24. Ewing, B., Hillier, L., Wendl, M., and Green, P. Basecalling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res., *8:* 175–185, 1998.
25. Ewing, B., and Green, P. Basecalling of automated sequencer traces using phred. II. Error probabilities. Genome Res., *8:* 186–194, 1998.
26. Cottu, P. H., Muzeau, F., Estreicher, A., Fléjou, J. F., Iggo, R., Thomas, G., and Hamelin, R. Inverse correlation between RER+ status and p53 mutation in colorectal cancer cell lines. Oncogene, *13:* 2727–2730, 1996.
27. Hoang, J. M., Cottu, P. H., Thuille, B., Salmon, R. J., Thomas, G., and Hamelin, R. BAT-26, an indicator of the replication error phenotype in colorectal cancers and cell lines. Cancer Res., *57:* 300–303, 1997.
28. Lerman, L. S., and Silverstein, K. Computational simulation of DNA melting and its application to denaturing gradient gel electrophoresis. Methods Enzymol., *155:* 482–501, 1987.
29. Hamelin, R., Jego, N., Puig, P. L., Vidaud, M., and Thomas, G. Efficient screening of p53 mutations by denaturing gradient gel electrophoresis in colorectal tumors. Oncogene, *8:* 2213–2220, 1993.
30. Duval, A., Iacopetta, A., Ranzani, G. N., Lothe, R. A., Thomas, G., and Hamelin, R. Variable mutation frequencies in coding repeats of *TCF-4* and other target genes in colon, gastric and endometrial carcinoma showing microsatellite instability. Oncogene, *18:* 6806–6809, 1999.
31. Duval, A., Busson-Leconiat, M., Berger, R., and Hamelin, R. Assignment of the *TCF-4* (*TCF7L2*) gene to human chromosome band 10q25.3. Cytogenet. Cell Genet., *88:* 264–265, 2000.
32. Sparks, A. B., Morin, P. J., Vogelstein, B., and Kinzler, K. W. Mutational analysis of the APC/β-catenin/Tcf pathway in colorectal cancer. Cancer Res., *58:* 1130–1134, 1998.