Multiple Intron Retention Occurs in Tumor Cell CD44 mRNA Processing

Steven Goodison,* Kazuhiro Yoshida,[†]
Michael Churchman,[†] and David Tarin*

From the UCSD Cancer Center,* University of California, San Diego, La Jolla, California, and the Nuffield Department of Pathology and Bacteriology,[†] University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DU, United Kingdom

Markedly increased overall levels of CD44 transcripts and proteins have been recognized in many tumors and the inappropriate expression and abnormal assembly of the CD44 variable exons has been linked to both tumor growth and metastatic potential. We have also previously observed the aberrant inclusion of intron 9 in CD44 mRNA transcripts in tumor tissues. In this study we assessed whether such retention is specific to certain introns or is a more general phenomenon affecting CD44 gene expression in tumor cells. Intron 18 was cloned and sequenced from genomic DNA and the novel sequences analyzed and used to create intron 18-specific probes. The newly characterized intron was found to have consensus 5' splice site and branchpoint sequences but a suboptimal 3' splice site. The status of CD44 intron 18 retention or excision was assessed in a colon tumor cell line (HT29) and in tissue from 20 colorectal tumors and matched normal mucosa. The intron was shown to be retained in transcripts from 15 of the 20 (75%) carcinomas but in only 3 of the 20 (15%) matched normal samples. These results compare with 80% retention of CD44 intron 9 in colonic carcinoma tissue mRNA and confirm that multiple abnormalities of CD44 mRNA processing occur in tumor cells. (*Am J Pathol* 1998, 153:1221–1228)

Current evidence favors the interpretation that inappropriate or aberrant regulation of gene expression does not necessarily lead to cell death but can lead to various forms of dysfunction including neoplasia. Complex genes whose transcripts undergo extensive alternative splicing are more likely to be incorrectly processed. The CD44 gene is an example of one whose expression has been investigated both as an example of misprocessing and for use as a potential tumor marker.

The human CD44 gene resides at chromosomal locus 11p13¹ and is composed of a segment of DNA of ap-

proximately 60 kb containing at least 20 exons.² Ten exons, namely 1–5 and 16–20, are spliced together to be translated as the standard isoform (CD44s) which is a transmembrane glycoprotein. The remaining exons can be alternatively spliced³ and assembled within the standard ones (Figure 1) to generate a number of variant protein isoforms (CD44v). These can be further modified by post-translational glycosylation to produce a large family of transmembrane glycoproteins with diverse functions. The standard isoform (CD44s) is present in almost all normal cells, whereas the distribution of other more complex variant isoforms occurs in a tissue-specific manner.

Markedly increased expression of CD44 proteins and abnormal assembly of CD44 transcripts have been observed in a variety of carcinomas, including those of the stomach,^{4–6} colon,^{7,10} breast,^{11,12} uterus,¹³ bladder,^{14,15} and lung,¹⁶ as well as in hematopoietic malignancies.¹⁷ This inappropriate expression of the CD44 variable exons has been linked to both tumor progression and metastatic potential. These observations indicate that the CD44 molecule could have an important role in tumor biology, cancer diagnosis, and prognostic evaluation.

Much of the data regarding CD44 gene expression has been gathered using reverse-transcriptase polymerase chain reaction (RT-PCR) techniques and with this form of analysis we have consistently observed that distinctive large molecular weight amplicons can be visualized from samples containing tumor cells. Indeed, some of these appear to be larger than would be predicted even if all possible variant exons were included in the transcript. Such data indicate the presence of abnormal gene transcripts. This characteristic pattern is not seen in other conditions associated with epithelial cell injury and inflammatory cell infiltration, but is reproducibly seen in serially passaged tumor cell lines without inflammatory cells. We reasoned that this highly characteristic abnormality could result from incompletely processed or misprocessed mRNA species, possibly containing retained intronic sequences. This hypothesis was investigated by characterizing intron 9 and looking for its retention in CD44 mRNA transcripts in bladder tumor-related sam-

Supported in part by a research contract between Boehringer Mannheim GmbH and Oxford University.

Accepted for publication June 1, 1998.

Address reprint requests to Professor David Tarin, Director, UCSD Cancer Center, 9500 Gilman Drive, La Jolla, CA 92093-0658.

Figure 1. Schematic diagram of the structure of the CD44 gene showing positions to which the primers used for RT-PCR analysis anneal.

ples.¹⁸ Retention of this intron was confirmed in a bladder tumor cell line, in bladder tumor tissue, and in exfoliated urothelial cells collected from the urine of bladder tumor-bearing patients but not in samples from tumor-free subjects. We have also recently demonstrated the retention of intron 9 in CD44 transcripts in breast tumors¹⁹ and in colon tumors.²⁰

The present study was conducted to ascertain whether abnormal retention of other CD44 introns occurs in tumor tissues. Intron 9 is located between two variable exons (Figure 1) and to make the investigation as informative as possible we chose to study an intron that separates two standard exons. Intron 18 interrupts exons 18 and 19 in the 3' region of the standard molecule and unlike most CD44 introns is less than 1 kb in size (~700 bp as determined by PCR across exon boundaries²). This intron was cloned and the novel sequences were used to analyze whether it is removed efficiently by cellular splicing mechanisms. The studies were performed initially on a human colon tumor cell line and then on tissues from 20 human colonic tumors and their matched normal mucosae. Intron 18 sequences were also examined for exonic/intronic consensus motifs and compared with intron 9 in order to define any sequences common to both which may predispose these introns to incorrect splicing. We provide evidence demonstrating tumor-specific disturbances in alternative splicing that result in the accumulation of immature intron-retaining transcripts in carcinoma cells *in vivo* and *in vitro*.

Materials and Methods

Cloning of Intron 18

Intron 18 nucleotide sequence was obtained by amplification between exons 18 and 19 by PCR using Pfu polymerase (Stratagene, La Jolla, CA), a primer pair Ex18, Aex19 and a human CD44 genomic clone DMPC-HHF 1-c2311¹⁸ as template. During the final extension step, 0.5 U of Taq polymerase (Boehringer Mannheim, Indianapolis, IN) were added (72°C for 10 minutes) of the PCR in order to add overhanging dATP bases to the termini of

the amplicon. A product of the estimated intron size (~700 bp) was gel-purified, ligated into the pCRII vector (Invitrogen, San Diego, CA) and propagated using standard bacterial protocols.

Sequencing of Intron 18

The insert within the pCRII vector was sequenced in both directions using M13 reverse primer and the Ex18 primer (Figure 1). The 72 bp of CD44 exon 17 within the cloned fragment served as reference sequence. Sequencing reactions were performed by chain termination reactions in the presence of [α -³²S-dATP] with an Amplicycle sequencing kit (Perkin Elmer, Branchburg, NJ).

Cell Lines

CD44 gene expression was studied in the colonic carcinoma cell line HT-29. Cells were routinely cultured in RPMI 1640 medium (GIBCO-BRL, Gaithersburg, MD) containing HEPES supplemented with 10% fetal calf serum at 37°C in a humidified atmosphere.

Tissue Procurement

Fresh colorectal carcinoma tissue samples and corresponding normal colonic mucosa from surgical resection specimens were snap-frozen and stored in liquid nitrogen until use. The presence or absence of carcinoma cells was routinely monitored in all tissues taken from colonic mucosae by hematoxylin and eosin staining of cryostat sections before RNA extraction.

RNA Extraction and cDNA Synthesis

Total cellular RNA was extracted by the acid guanidinium phenol-chloroform method and mRNA was purified using Oligo-dT latex beads (R&D, Abingdon, UK). mRNA was purified from nuclear and cytosolic fractions of the HT-29 cell line by pelleting of nuclei after Dounce homogenization as described previously.²⁰ RNase and DNase digestion analyses were performed to monitor for possible DNA contamination of RNA samples. Their 500 ng of mRNA were digested with either 1 μ g of RNase (Boehringer Mannheim) or 0.1 U of RQ1 RNase-free DNase (Promega, Madison, WI) at 37°C for 20 minutes. After incubation, RNase or DNase was inactivated by incubation at 70°C for 10 minutes. The reaction mixtures were extracted with phenol-chloroform and precipitated with ethanol before the reverse transcription reaction. cDNA was synthesized from 500 ng of poly A+ selected mRNA at 42°C with AMV-reverse transcriptase using a cDNA Cycle Kit (Invitrogen, San Diego, CA).

RT-PCR

cDNA was amplified with 2.5 U of Taq DNA polymerase (Boehringer Mannheim) in the presence of 1 μ mol of primer pair P1, P4 for variant exon analyses and with

primer pair Ex16, Aex19 for intron 18 retention analyses. The conditions of PCR were as follows: 94°C for 5 minutes and 85°C for 1 minute, during which time Taq polymerase was added (hot start) followed by 30 cycles of 94°C for 1 minute, 55°C for 1 minute and 72°C for 2.5 minutes. The presence of intact input RNA was checked in all samples by amplification of human β -actin cDNA (Stratagene). The presence of CD44 transcripts in all samples was confirmed by reference to amplification of the standard form of CD44 mRNA which is expressed in all colonic mucosa (see Figure 3).

The CD44-specific primers used were: P1: 5'-GACA-CATATTGCTTCAATGCTTCAGC-3'; P2: 5'-CCTGAAGAA-GATTGTACATCAGTCACAGAC-3'; Aex5: 5'-AGCAGG-GATTCTGTCTGTGCT-3'; P4: 5'-GATGCCAAGATGATC-AGCCATTCTGGAA-3'; E1: 5'-TTGATGAGCACTAGTGC-TACAGCA-3'; E2: 5'-CATTTGTGTTGTTGTGTAAGATG-3'; D1: 5'-GACAGACACCTCAGTTTTCTGGA-3'; AD1: 5'-GGTGTCTGGAGATAAAATCTTCATC-3'; EX11: 5'-TC-CAGGCAACTCCTA-3'; AEX11: 5'-TCCAGGCAACTCCTA-3'; Ex16: 5'-GACCAAGACACATTCCACCCAGT-3'; Ex18: 5'-TGGCTGATCATCTTGGCATCCCTC-3'; Aex19: 5'-TCT-TATGCTATAACCTGAATCTCTTCA-3'; I18 f2: 5'-TAAACCT-CATGCAGACTC-3'; I18 r2: 5'-GGGATGATGATCACTGA-3'.

Southern Blot Hybridization

10 μ l of the 50- μ l PCR reaction mixture was electrophoresed in a 1.2% agarose gel, transferred to a Hybond N+ nylon membrane (Amersham, Little Chalfont, UK) with 0.4 N NaOH solution overnight and hybridized with probes made by PCR using the primers listed above with a CD44 genomic clone (c2311) template. Figure 1 depicts primer annealing sites. The DNA probes were labeled with peroxidase using enhanced chemiluminescence (ECL) direct nucleic acid labeling and hybrids were visualized with the ECL detection system (Amersham). Conditions used for hybridization, washing, and detection were those recommended by the manufacturer.

Intron 18 Sequence Database

Accession Number

The sequence of CD44 intron 18 has been submitted to the EMBL database and has the accession number AJ002286.

Results

Intron 18 Sequence Analysis

The 633 bp of intron 18 sequence is shown in Figure 2. The intron/exon boundaries, (5'-GT-AG-3') are highlighted in bold type. CD44 genomic introns are as yet largely uncharacterized and it is possible that not all exon/intron boundaries are accurately defined. Applying the GT-AG rule, a number of partial 5' boundary motifs exist within the proximal region of the sequence and open reading frame (ORF) analysis revealed that it would be

```

1  GTAAGGGGCT GTCCTGGGG CTTTCAACTT GGAAGGGCAT CATTAAACC TGGGCTTTAT
61  ATCCTTGCTG CTCAGAGCGT AGTCCTGGAA CCAGCAGCCT GGTACCTGGA GTTGTGAATG
121 CACAACCTCAG ACCCAACCCC AGACCTCCTA TGGCAGAGTC TGCACTTTAA CAAGTTTCCA
181 GGGATTCATG TGCATGGTGC AGCTTAAGAC ATCTTTCTCT GCCCGGGTTT TTAGAGACAG
241 GAGACACCTG TGATTGTTTG GCTCATCACT TTTAGGTAGT GGTGAATCC TTTGTGATTT
301 TACATGTTTT TTAGGGTTAA ACCTCATGCA GACTCTGTCT GATGAAGTAA TGAAGTATCT
361 CACCAGGAAG GCACACAAAA GTATATACTC ATTCATTGAT TTGTTAAATA TTTATTAAAT
421 ACATTACAAA TGCACAAATA TCGTATAAAG ATTCACTGAT TCATCATCCC CTGTAGCCTC
481 TCTGAGGCCC AGGGAACCTT TAAGAACTTT TACCAGAGG AAGTATTGGC CTTGTGCTTT
541 TGTTTTCTGT TTGTGTTTTT TATTGGCCTT GTGCTTTTGT GTCTGTTTGT GTTTTGTGTA
601 TGATATCATT TCTATTTTC TGGGAACCTG TAG

```

Figure 2. Sequence of CD44 intron 18. Boxed letters indicate consensus splice sites. Underlined letters indicate consensus branchpoint sequence.

possible for exon 18 coding sequence to extend another 38 amino acids into the intron, in the same reading frame, before reaching a stop codon. This extension would require 114 nucleotides but a 3' terminus at that position does not coincide with any complete or partial consensus splice-site sequences. Many CD44 exons are on the order of 100 bp in size so it is possible that an undefined exon resided within intron 18. However, although not conclusive, our analyses showed that no sequences that lay between putative exon/intron boundary consensus sequences had appropriate open reading frames.

Branchpoint sequences have a consensus pattern of nucleosides of PyNPYPyPuAPy, with a stable lariat structure being created by an intra-intron esterification reaction at the adenosine (A) base (21). This motif usually lies 18–40 bases upstream of the 3' splice site and such a region exists in intron 18, incorporating the adenosine residue at position 602 (31 bp upstream of the 3' terminus). Also following previously described intron structures, a pyrimidine-rich stretch of nucleotides (12 of 13 nucleotides between 608 and 620) resides between the branchpoint and the 3' splice site.

RT-PCR/Southern Hybridization Analysis

CD44s and CD44v Expression in Colonic Mucosae

Twenty colon carcinoma specimens and their corresponding normal tissues were analyzed for expression of CD44 exons by RT-PCR using primers P1 and P4. Subsequently Southern hybridization of the products was performed using a probe for the 5' standard exons (contained in all CD44 transcripts) and for the variant exons 7, 8, and 11. The predicted amplification products resulting from the use of the P1, P4 primers would be 482 bp for the CD44 transcript, with several larger products also present if the sample contained alternatively spliced transcripts incorporating variant exons. As previously described,^{5,17,20} carcinoma samples of colonic mucosa displayed characteristic overexpression of numerous alternatively spliced CD44 mRNA species (Figure 3). Variant exon 7 was present in CD44 transcripts in 13 of 20 (65%) tumors, 17 of 20 (85%) contained exon 8, and 16 of 20 (80%) contained exon 11. Probing with exon 11 discriminated the most clearly between tumor and normal colonic mucosae as in normal tissues the signals were either absent or very weak. Expression of exon 8 was

Figure 3. Southern blot hybridization analysis of CD44 RT-PCR. RNA was extracted from human colon carcinoma tissues (T) and from matched adjacent normal mucosa (N). Case numbers are indicated above the lanes. Amplification of cDNA was performed across the variant region between standard exons 3 to 17 (primers P1, P4). PCR products were electrophoresed and blotted to nylon filters for hybridization. The standard probe used to detect all transcripts was synthesized using primers P2 and Aex5. Probes for individual variant exons were generated by PCR using exon-specific primers (see Figure 1 for annealing positions) with a template consisting of a human CD44 genomic clone. Replicate filters were used for each ECL-labeled specific probe. Molecular weight markers are indicated to the right in base pairs.

seen in the majority of tumor tissues (85%) but it was also detectable in most normal tissues (70%). Exon 8 is often the variant exon that abuts standard exon 5 in colonic mucosae.²²

Intron 18 Retention in HT-29 Tumor Cell Line mRNA

Optimization of RT-PCR and Southern blot hybridization conditions for analysis of the presence of the novel intron sequences in CD44 transcripts was performed on colonic tumor cell line HT-29. As is the case in many tumor cell lines, we have previously shown that this cell line has gross overexpression of CD44 variant exon-containing transcripts.^{15,23} This cell line was chosen for study because of its appropriate histogenetic origin.

Intron 18 lies outside the variant exon insertion region of the gene and would not be amplified by the primers P1, P4 (Figure 1). Extension of the RT-PCR amplicon by use of a primer further into the 3' standard region (Aex19) and subsequent Southern blot hybridization with an internal intron 18-specific probe resulted in observation of a wide range of PCR products of molecular weight greater than 1 kb. This suggested that there were several species of mRNA which contained intron 18 sequences in combination with various exons and possibly with other introns. In order to exclude the possibility that the PCR products might result from amplification of contaminating genomic DNA, RNase and DNase digestions were performed on the RNA fraction of HT-29 human colon carcinoma cells before cDNA synthesis. CD44 amplification was not detected in the RNase-digested mRNA sample (Figure 4, lane R) but it was present in mRNA after DNase treatment (lane D), confirming that intron 18 was present in CD44 transcripts in these cells. To determine whether intron 18 is retained in transcripts of the cytoplasmic mRNA pool, mRNA was purified from isolated nuclei and from the cytosol of HT-29 colonic carcinoma cells. Each fraction was analyzed for intron retention and Southern blot hybridization with the intron 18-specific probe revealed

Figure 4. Origin and localization of intron 18-containing CD44 transcripts in HT-29 human colon carcinoma cells. RNA preparations were subjected to DNase (lane D) and RNase (lane R) digestion before cDNA synthesis. RT-PCR was performed using primers P1 and Aex19. Products were electrophoresed and Southern blot analysis was performed using a hybridization probe specific for intron 18 (synthesized by PCR using primers I18-f2 and I18-r2). Similar analysis revealed intron 18-CD44 transcripts in both nuclear (lane N) and cytoplasmic (lane C) RNA fractions.

the presence of intron-retaining CD44 mRNA in both the nuclear and cytoplasmic fractions (Figure 4, lanes N and C).

Intron 18 Retention in Colonic Matched Specimens

CD44 expression was studied in the fresh human colonic mucosa tissue samples from surgical resection specimens to evaluate the incidence of intron 18 retention in this tissue. However, as can be seen in Figure 4, analysis of intron 18 retention in CD44 transcripts amplified across the variant region with primers P1, Aex19 were difficult to interpret because they could include any number of combinations of variant exons and introns. In order to study the expression of intron 18 more clearly and to accurately size the resulting amplicons, a more defined RT-PCR approach was utilized. The primer pair Ex16, Aex19 was chosen for analysis of intron 18 retention, giving amplification across the region of interest from exon 16 to exon 19. Only amplicons containing intron 18 would be revealed by hybridization with the intron-specific probe with transcripts including exons 16 (63 bp), 17 (72 bp), 18 (75 bp), intron 18 (633 bp), and exon 19 (27 bp of primer Aex19) giving a product of 870 bp. Inclusion of introns 17 (~5k bp) or 19 (2.5k bp) as well as intron 18 would result in products of 3370 bp and 8370 bp respectively.

Intron 18 was shown to be retained in mRNA extracted from the frozen colorectal carcinoma tissue specimens (Figure 5). The retention was seen in CD44 mRNA from 15 (75%) of 20 tumors but in only 3 (15%) of the 20 matched normal samples. Because of the fibrous nature of human tissue specimens and the need to avoid RNase degradation in biopsy material, RNA was extracted from frozen tissues by mechanical homogenization without thawing. It was therefore not possible to separate nucleus and cytoplasmic fractions for analysis. In this study, no association was evident between the presence of intron 18 retention and tumor grade or Dukes' stage of colorectal carcinoma tissue. All cDNA reactions were tested for completion by amplification across an intron/exon bound-

Figure 5. Demonstration of the abnormal retention of CD44 intron 18 in mRNA transcripts from normal tissues (A) and carcinoma tissues (B) from the same resected human colonic mucosa specimens, cases 1–20. RT-PCR was performed using the Ex16, Aex19 primer pair and amplification products were transferred to nylon membranes and analyzed by Southern blot hybridization. Detection was achieved using an ECL-labeled probe specific for intron 18 (synthesized by PCR using primers I18-f2 and I18-r2).

ary of human β -actin cDNA. In none of the human specimens nor in the tumor cell line was any alteration in the splicing pattern of β -actin observed. Retention of intron 17 was also assessed indirectly by monitoring the size of intron 18-probed PCR products amplified between exons 16 and 19. Retention of intron 17 was occasionally found in conjunction with intron 18 (3 of 20 tumor samples, 0 of 20 normal tissues) in such transcripts.

Discussion

There are many potential points of dysfunction in the complex mechanisms involved in the regulation of mRNA production. Alterations of gene expression patterns could result from lesions at any of the regulatory steps and may lead to cell cycle aberrations resulting in cell death, immortalization, or tumorigenesis. Alternatively, inappropriate protein function can result from incorrect RNA processing at the point of alternative splicing.

The complex inter- and intramolecular reactions involved in the maturation of pre-mRNA transcripts occur on formation of a multi-component spliceosome complex. The complicated processes depend on the interplay between specific RNA and protein molecules,²⁴ are prone to some degree of error, and are likely to be relatively inefficient. However, in biological processes downstream defects will become manifest only if surveillance and repair mechanisms fail and the number of surviving aberrant molecules rises above an acceptable level. Although alternative splicing greatly increases the coding efficiency of the genome, it is also a potential source of genetic misinformation resulting in atypical cell behavior and tissue function.

A number of human diseases have been traced to mutations within intron splicing sequences. The intronic lariat branchpoint is mutated in lecithin:cholesterol acyltransferase causing inherited human fish-eye disorder²⁵ and in the low-density lipoprotein receptor gene leading to familial hypercholesterolemia.²⁶ A database cataloguing aberrant splicing of mammalian genetic disorder mutations²⁷ has shown that genomic mutation usually results in exon skipping, with intron retention being relatively rare. Intron sequences have also been shown to have motifs within them which can alter expression of the gene

by influencing transcription rates^{28,29} via protein binding. Some of the larger introns can have open reading frames within them which independently code for reverse transcriptase-like proteins that may have a role in intron mobility and RNA splicing³⁰ and it is clear that the definitions of intron and exon are rather blurred and often overlap.

Examples of aberrant mRNA processing include exon skipping,³¹ abnormal splice site selection,³² and full intron retention.^{33,34} However, intron retention may not always be the consequence of splicing errors. The mechanisms of splice site selection and the complex interactions between RNAs and protein factors in the nucleus are gradually being elucidated and a number of examples have shown that sometimes a biologically appropriate effect can be achieved by intron or intron fragment inclusion in transcripts. Removal or retention of an intron in the *Drosophila* P element mRNA is cell type-specific³⁵ and a ratio between retention and spliced removal of introns can be determined in some cells.^{36–38} Introns may be retained because they code for an independent protein themselves, they extend the coding sequence of an adjoining exon, or they provide alternative translation termination signals. Whatever its role, the intron-containing mRNA must, if it is to be effective, overcome confinement to the nucleus^{39,40} and enter the cytoplasm.

The inclusion of introns in mRNA transcripts will lead to alterations in the resulting translated protein and some recent examples highlight a number of downstream consequences. Frame shifts or early stop codon usage can target proteins to different compartments^{41,42} of the cell or to secretory pathways.⁴³ Conversely, inclusion of in-frame intronic coding or noncoding sequences may lead to truncation and loss of functional protein domains.⁴⁴

From the results presented above and from earlier data it is now beyond reasonable doubt that tumor cells contain far more alternatively spliced CD44 transcripts, some of which retain introns,^{18,20} than their normal counterparts. Whether this is causal in the pathogenesis of neoplasia or is a secondary phenomenon is unknown.

Several studies have provided evidence that CD44 gene expression is abnormal in gastrointestinal tumors^{9,10} and *in situ* RNA hybridization data⁴⁵ have confirmed that the abnormalities are localized to the tumor cells in a sample. Because of its known role in cell-cell and cell-matrix adhesion, mediated by binding to ligands such as fibronectin, collagen, and hyaluronic acid,^{46,47} it has been suggested that expression of certain CD44 protein isoforms could facilitate tumor invasion and metastasis.^{48,49} In agreement with previous studies, analysis of the colon tissues in this study (Figure 3) revealed that although many variants of unusual size are overexpressed in tumor tissues, exon 11 appears to be the best marker for distinguishing between tumor and normal mucosa.

Matsumura et al¹⁸ described for the first time detection of tumor cells in a sample achieved by monitoring intron retention in CD44 mRNA. Retention of the normally removed variant intron 9 was first observed in exfoliated urothelia of bladder cancer patients and was subsequently found to occur in breast carcinoma tissues¹⁹ and

in colon carcinoma.²⁰ In this study retention of the newly characterized intron 18 was found in 15 of 20 (75%) of colonic tumor tissues but retained in only 3 of the 20 (15%) matched normal tissues. In the previous study,²⁰ the abnormal retention of intron 9 in colonic tissue transcripts was found in 16 (80%) out of 20 patients with carcinoma and in 4 of the corresponding normal tissues. Some of the human colon tissue samples used in this study and a previous intron 9 retention study were obtained from the same cases.²⁰ Comparison of the data did not reveal a pattern in the incidence of intron 9 and/or intron 18 in these samples. Transcripts containing intron 9 alone, intron 18 alone, or both introns 9 and 18 were all found in tumor cell mRNA. The retention of CD44 intron 17 in mucosal transcripts containing intron 18 was also assessed indirectly by monitoring the size of intron 18-probed PCR products amplified between exons 16 and 19. Retention of intron 17 was found in conjunction with intron 18 in 3 of 20 tumor samples but in none of 20 matched normal tissues. These results show that aberrant splicing is not intron-specific and intron retention can therefore be recognized, along with inappropriate and disorderly overexpression of variant exons, as a characteristic feature of CD44 gene expression in neoplasia.

The consequences of abnormal intron retention for CD44 function may be profound. If introns upstream of exon 17 were retained, then the encoded isoform would contain only part of the extracellular portion of the protein and could be soluble. The protein could then be retained in the cytosol and/or secreted because of the absence of the short anchoring transmembrane region encoded by exon 18.

As discussed above, the retention of introns in tumor cell mRNA species could be due to either aberrant processing or a functional requirement. A number of consensus sequences involved in efficient splicing have been identified and the novel intron 18 sequence was analyzed for the presence of such motifs. A 5' splice donor site consensus sequence is composed of the sequence G/GTAA and a 3'-splice site sequence is composed of the sequence NCAG/N. The intron 18 boundaries do follow the GT-AG rule but the 3' splice site is incomplete, being NTAG/N. Suboptimal splice sites are thought to be spliced with less efficiency and introns with these sequences may therefore be more likely to escape the nuclear splicing process. However, there is a great deal of redundancy in these DNA motifs and consensus sequences are far from being fully defined. Intron 18 does have a consensus-like branchpoint sequence (NPY-PyPuAPy) at the appropriate position of 31 bp upstream (usually -18 to -40 bp) from the 3'-splice site. The presence of a Py-rich region between the branchpoint and the splice site also adds to the likelihood that the intron boundaries are correct. Conversely, intron 9 exon boundaries obey the consensus rules fully and there are also appropriate branchpoint sequences at 36 bp upstream of the 3' splice site.¹⁸ Dirksen et al³² showed that the retention of an intron in bovine growth hormone pre-mRNA was dependent on a balance of the relative strengths of suboptimal splice sites and the presence of an exonic splicing enhancer (ESE) in an adjoining exon.

In such circumstances a weak splice site configuration may interact with spliceosome factors less efficiently and a percentage of such molecules may thereby escape the splicing pathway and by default be transported to the cytosol. These ESE elements consist of purine stretches,^{50,51} the most effective of which repeat GGAA motifs. Such purine motifs were found in adjoining exons 18 and 19 but only as single motifs or as short runs of purines (7 bp in exon 18, G3A4 and 10 bp in exon 19, G3A7). Of interest, exon 9, adjacent to intron 9, contains only one GGAA motif but exon 10 does appear to have an ESE-like run of purines (GAAGGAAA and GAGGAAGAAGAGA), and this is in an exon of only 117 bp. A functional role for these sequences can only be speculative at this point.

Pre-mRNA molecules are initially transcribed containing exons and introns that are later selectively removed or retained. Therefore, altered CD44 splicing efficiency in tumor cells would affect all regions of the molecule equally. Our study shows this to be the case, so intron retention in CD44 transcripts is not directly related to the alternative splicing of the variable exons of the molecule. Splicing regulation may be compromised in only a subset of the mRNA pool, perhaps in complex alternatively spliced transcripts such as the CD44 molecule. Transcripts that undergo extensive processing may recruit specific proteins as part of the spliceosome complex and this may be the site of misregulation. More fundamental studies are needed to determine the mechanisms involved in the splicing of complex genes such as CD44.

Although a tumor detection rate of 75% using a CD44 intron 18 probe is encouraging, as with any marker under investigation the values should improve as we understand the regulation and processing of the marker molecule better, information which can lead to the design of more specific detection probes. Although for diagnostic purposes the detection of large amounts of intron-containing transcripts in tumor samples may be useful in itself, evaluation of the ratio of misprocessed to correctly processed transcripts would also be of interest.

We have previously shown that CD44 transcripts containing intron 9 are abundant in tumor sample mRNA pools¹⁸⁻²⁰ and it is feasible that intron 9 was retained for some specific functional purpose. This study was undertaken to evaluate whether those examples of intron-retention revealed an intron-specific phenomenon or one that also occurs in other regions of the CD44 transcript. Tumor sample CD44 transcripts have now been shown to include multiple CD44 introns. Added to the evidence for the use of cryptic splice sites and the seemingly random inclusion of variable exons in tumor sample CD44 transcripts, the data strongly suggest that there is misprocessing of CD44 mRNA species in neoplastic lesions.

To evaluate the extent of this phenomenon, further studies will be needed to investigate aberrant CD44 processing in a large panel of tissue and cell types. Of particular interest will be proliferative stem cell populations, lymphocytes, and noncancerous or precancerous lesions. Although the relatively low abundance of aberrant CD44 transcripts may be problematical for study without amplification procedures, they could be evalu-

ated using techniques such as microdissection and *in situ* hybridization.

It is unclear why detectable quantities of intron-containing CD44 transcripts accumulate in tumor cells. In the absence of any specific anomalies in CD44 intron sequences it is possible that there is a general derangement of splicing mechanisms in neoplastic transformation; however, other transcripts appear to be spliced efficiently in the same cells. Alternatively, specific splicing components may be in limited supply and the increase in transcriptional activity of subsets of genes occurring in rapidly dividing cells may compromise the processing machinery, resulting in a drop in the efficiency of the processing of some RNA species. Diagnostically, significant increases in RNA misprocessing above background error levels may be a marker of high rates of cell division and therefore one which would be evident at the earliest stages of cancer.

Acknowledgments

We thank Helene Mellor for preparation of the manuscript and Heather Dorricott for technical assistance.

References

- Goodfellow PN, Banting G, Wiles MV, Tunnacliffe A, Parkar M, Solomon E, Dalchau R, Fabre JW: The gene, MIC4, which controls expression of the antigen defined by monoclonal antibody F10.44.2, is on human chromosome 11. *Eur J Immunol* 1982, 12:659-663
- Screaton GR, Bell MV, Jackson DG, Cornelis FB, Gerth U, Bell JL: Genomic structure of DNA encoding the lymphocyte homing receptor CD44 reveals at least 12 alternatively spliced exons. *Proc Natl Acad Sci USA* 1992, 89:12160-12164
- Tolg C, Hofmann M, Herrlich P, Ponta H: Splicing choice from ten variant exons establishes CD44 variability. *Nucleic Acids Res* 1993, 21:1225-1229
- Heider K-H, Dammrich J, Skroch-Angel P, Muller-Hermelink H-K, Vollmers HP, Herrlich P, Ponta H: Differential expression of CD44 splice variants in intestinal- and diffuse-type human gastric carcinomas and normal gastric mucosa. *Cancer Res* 1993, 53:4197-4203
- Yokozaki H, Ito R, Nakayama H, Kuniyasu H, Taniyama K, Tahara E: Expression of CD44 abnormal transcripts in human gastric carcinomas. *Cancer Lett* 1994, 83:229-234
- Harn H-J, Ho L-I, Chang J-Y, Wu C-W, Jiang S-Y, Lee H-S, Lee W-H: Differential expression of the human metastasis adhesion molecule CD44V in normal and carcinomatous stomach mucosa of Chinese subjects. *Cancer* 1995, 75:1065-1071
- Matsumura Y, Tarin D: Significance of CD44 gene products for cancer diagnosis and disease evaluation. *Lancet* 1992, 340:1053-1058
- Tanabe KK, Ellis LM, Saya H: Expression of CD44R1 adhesion molecule in colon carcinomas and metastases. *Lancet* 1993, 341:725-726
- Mulder J-WR, Krut PM, Sewnath M, Oosting J, Seldenrijk CA, Weidema WF, Offenhaus GJA, Pals ST: Colorectal cancer prognosis and expression of exon-v6-containing CD44 proteins. *Lancet* 1994, 344:1470-1472
- Wielenga VJM, Heider K-H, Offerhaus GJA, Adolf GR, van den Berg FM, Ponta H, Herrlich P, Pals ST: Expression of CD44 variant proteins in human colorectal cancer is related to tumor progression. *Cancer Res* 1993, 53:4754-4756
- Kaufmann M, Heider K-H, Sinn H-P, von Minckwitz G, Ponta H, Herrlich P: CD44 isoforms in prognosis of breast cancer. *Lancet* 1995, 345:615-619
- Dall P, Heider K-H, Sinn H-P, Skroch-Angel P, Adolf G, Kaufmann M, Herrlich P, Ponta H: Comparison of immunohistochemistry and RT-PCR for detection of CD44v-expression, a new prognostic factor in human breast cancer. *Int J Cancer* 1995, 60:471-477
- Dall P, Heider K-H, Hekele A, Minckwitz G, Kaufmann M, Ponta H, Herrlich P: Surface protein expression and messenger RNA-splicing analysis of CD44 in uterine cervical cancer and normal cervical epithelium. *Cancer Res* 1994, 54:3337-3341
- Matsumura Y, Hanbury D, Smith J, Tarin D: Non-invasive detection of malignancy by identification of unusual CD44 gene activity in exfoliated cancer cells. *BMJ* 1994, 308:619-624
- Woodman AW, Sugiyama M, Yoshida K, Sugino T, Borgya A, Goodison S, Matsumura Y, Tarin D: Analysis of anomalous CD44 gene expression in human breast, bladder, and colon cancer and correlation of observed mRNA and protein isoforms. *Am J Pathol* 1996, 149:1519-1530
- Penno MB, August JT, Baylin SB, Mabry M, Linnoila RI, Lee VS, Croteau D, Yang XL, Rosada C: Expression of CD44 in human lung tumors. *Cancer* 1994, 54:1381-1387
- Koopman G, Heider K-H, Horst E, Adolf GR, van den Berg F, Ponta H, Herrlich P, Pals ST: Activated human lymphocytes and aggressive non-Hodgkin's lymphomas express a homologue of the rat metastasis-associated variant of CD44. *J Exp Med* 1993, 177:897-904
- Matsumura Y, Sugiyama M, Matsumura S, Hayle AJ, Robinson P, Smith JC, Tarin D: Unusual retention of introns in CD44 gene transcripts in bladder cancer provides new diagnostic and clinical oncological opportunities. *J Pathol* 1995, 177:11-20
- Bolodeoku J, Yoshida K, Sugino T, Goodison S, Tarin D: Accumulation of immature intron-containing CD44 gene transcripts in breast cancer tissues. *Mol Diagnostics* 1996, 1:175-181
- Yoshida K, Bolodeoku J, Sugino T, Goodison S, Matsumura Y, Warren BF, Toge T, Tahara E, Tarin D: Abnormal retention of intron 9 in CD44 gene transcripts in human gastrointestinal tumors. *Cancer Res* 1995, 55:4273-4277
- Reed R, Maniatis T: The role of the mammalian branchpoint sequence in pre-mRNA splicing. *Genes Dev* 1988, 2:1268-1276
- Goodison S, Yoshida K, Bolodeoku J, Woodman AW, Sugino T, Gorham H, Tarin D: Rapid analysis of distinctive CD44 RNA splicing preferences that characterize colonic tumors. *Cancer Res* 1997, 57:3140-3144
- Sugino T, Yoshida K, Zhao S, Goodison S, Tarin D: Disorderly CD44 expression in human cancer cells can be modulated by growth conditions. *J Pathol* 1998 (in press)
- Michaud S, Reed R: A functional association between the 5' and 3' splice site is established in the earliest prespliceosome complex (E) in mammals. *Genes Dev* 1993, 7:1008-1020
- Kuivenhoven JA, Weibusch H, Pritchard PH, Funke H, Benne R, Assmann G, Kastelein JJ: An intronic mutation in a lariat branchpoint sequence is a direct cause of an inherited human disorder (fish-eye disease). *J Clin Invest* 1996, 98:358-364
- Webb JC, Patel DD, Shoulders CC, Knight BL, Soutar AK: Genetic variation at a splicing branch point in intron 9 of the low density lipoprotein (LDL)-receptor gene: a rare mutation that disrupts mRNA splicing in a patient with familial hypercholesterolaemia and a common polymorphism. *Hum Mol Genet* 1996, 5:1325-1331
- Nakai K, Sakamoto H: Construction of a novel database containing aberrant splicing mutations of mammalian genes. *Gene* 1994, 141:171-177
- Jung M, Dritschilo A, Mark G, Kasid U: Identification of multiple repeat sequences and transcription-related elements within introns 4, 8 and 9 of human Raf-1. *Biochem Biophys Res Comm* 1993, 190:462-469
- Brooks AR, Blackhart BD, Haubold K, Levy-Wilson B: Characterization of tissue-specific enhancer elements in the second intron of the human apolipoprotein B gene. *J Biol Chem* 1991, 266:7848-7859
- Mohr G, Perlman PS, Lambowitz AM: Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. *Nucleic Acids Res* 1993, 21:4991-4997
- Porter S, Mintz B: Multiple alternatively spliced transcripts of the mouse tyrosinase-encoding gene. *Gene* 1991, 97:277-282
- Derwahl M, Seto P, Rapoport B: An abnormal splice donor site in one allele of the thyroid peroxidase gene in FRTL5 rat thyroid cells introduces a premature stop codon: association with the absence of functional enzymatic activity. *Mol Endocrinol* 1990, 4:793-799
- Svoboda M, Ciccarelli E, Tastenoy M, Cauvin A, Stievenart M, Christopher J: Small introns in a hepatic cDNA encoding a new glucagon-

- like peptide 1-type receptor. *Biochem Biophys Res Commun* 1993, 191:479–486
34. Eiden MV, MacArthur I, Okayama H: Suppression of the chemically transformed phenotype of BHK cells by a human cDNA. *Mol Cell Biol* 1991, 11:5321–5329
35. Siebel CW, Rio DC: Regulated splicing of the *Drosophila* P transposable element third intron in vitro: somatic repression. *Science* 1990, 248:120–1208
36. Hampson RK, Rottman FM: Alternative processing of bovine growth hormone mRNA: nonsplicing of the final intron predicts a high molecular weight variant of bovine growth hormone. *Proc Natl Acad Sci USA* 1987, 84:2673–2677
37. Cooke NE, Ray J, Emery JG, Liehaber SA: Two distinct species of human growth hormone-variant mRNA in the human placenta predict the expression of novel growth hormone proteins. *J Biol Chem* 1988, 263:9001–9006
38. Dirksen WP, Sun Q, Rottman FM: Multiple splicing signals control alternative intron retention of bovine growth hormone pre-mRNA. *J Biol Chem* 1995, 270:5346–5352
39. Legrain P, Rosbash M: Some cis- and trans-acting mutants for splicing target pre-mRNA to the cytoplasm. *Cell* 1989, 57:573–583
40. Hamm J, Mattaj JW: Monomethylated cap structures facilitate RNA export from the nucleus. *Cell* 1990, 63:109–118
41. van Leusden MR, Kuikman I, Sonnenberg A: The unique cytoplasmic domain of the human integrin variant $\beta 4E$ is produced by partial retention of intronic sequences. *Biochem Biophys Res Commun* 1997, 235:826–830
42. Altieri DC: Splicing of effector cell protease receptor-1 mRNA is modulated by an unusual retained intron. *Biochemistry* 1994, 33:13848–13855
43. Tabiti K, Cui L, Chatwal VJ, Mochhala S, Ngoi SS, Pallen CJ: Novel alternative splicing predicts a secreted extracellular isoform of the human receptor-like protein tyrosine phosphatase LAR. *Gene* 1996, 175:7–13
44. Ebihara K, Masuhiro Y, Kitamoto T, Suzawa M, Uematsu Y, Yoshizawa T, Ono T, Harada H, Matsuda K, Hasegawa T: Intron retention generates a novel isoform of the murine vitamin D receptor that acts in a dominant negative way on the vitamin D signaling pathway. *Mol Cell Biol* 1996, 16:3393–3400
45. Gorham H, Sugino T, Woodman AW, Tarin D: Cellular distribution of CD44 gene transcripts in colorectal carcinomas and in normal colonic mucosa. *J Clin Pathol* 1996, 49:482–488
46. Lesley J, Hyman R, PW K: CD44 and its interaction with extracellular matrix. *Adv Immunol* 1993, 54:271–335
47. Stamenkovic I, Aruffo A, Amiot M, Seed B: The hematopoietic and epithelial forms of CD44 are distinct polypeptides with different adhesion potentials for hyaluronate-bearing cells. *EMBO J* 1991, 10:343–348
48. Seiter S, Arch R, Reber S, Komitowski D, Hoffmann M, Ponta H, Herrlich P, Matzku S, Zoller M: Prevention of tumor metastasis formation by anti-variant CD44. *J Exp Med* 1993, 177:443–455
49. Gunthert U, Hoffman M, Rudy W, Reber S, Zoller M, Haufmann I, Matzku S, Wenzel A, Ponta H, Herrlich P: A new variant of glycoprotein CD44 confers metastatic potential to rat carcinoma cells. *Cell* 1991, 65:13–24
50. Hampson RK, LaFollette L, Rottman FM: Alternative processing of bovine growth hormone mRNA is influenced by downstream exon sequences. *Mol Cell Biol* 1989, 9:1604–1610
51. Dirksen WP, Hampson RK, Sun Q, Rottman FM: A purine-rich exon sequence enhances alternative splicing of bovine growth hormone pre-mRNA. *J Biol Chem* 1994, 269:6431–6436