ORIGINAL ARTICLE

Haruhito Harada · Hisaki Nagai · Michiko Tsuneizumi
Iwao Mikami · Sumio Sugano · Mitsuru Emi

Identification of *DMC1*, a novel gene in the TOC region on 17q25.1 that shows loss of expression in multiple human cancers

Abstract Frequent allelic losses within chromosomal band 17q25.1 in a variety of human cancers have suggested the presence of one or more tumor suppressor genes in this region. Furthermore, a genetic locus responsible for familial focal nonepidermolytic palmoplantar keratoderma, a condition associated with cancer of the esophagus, lies in the same region. This esophageal-cancer susceptibility locus, TOC (tylosis with oesophageal cancer), might be a target of deletions at 17q25.1 in multiple types of malignancy. Using the reverse transcriptase-polymerase chain reaction (RT-PCR) to examine cancer cell lines for alterations in the expression of transcripts from this portion of 17q, we identified a novel gene that we designated *DMC1* (downregulated in multiple cancer-1). The full-length cDNA is 3293 bp long. Its putative product is an integral membrane protein of 788 amino acids, belonging to the class of so-called "inside-out" membrane proteins; it lacks a signal sequence but contains an N-terminal cytoplasmic domain, a single transmembrane peptide, and a C-terminal extracellular domain. We documented loss of expression of *DMC1* in 2 of 10 breast-cancer cell lines, in 7 of 10 cervical-cancer lines, in 7 of 13 hepatocellular-cancer lines, in 3 of 7 lung-cancer lines, in 3 of 6 thyroid-cancer lines, in 2 of 6 gastric-cancer lines, and in 2 of 4 renal cell-cancer lines. Our results suggest that loss of expression of the *DMC1* gene at 17q25.1 may play an important role in the development of cancers in a broad range of human tissues.

H. Harada · H. Nagai · M. Tsuneizumi · I. Mikami · M. Emi (✉)
Department of Molecular Biology, Institute of Gerontology, Nippon Medical School, 1-396 Kosugi-cho, Nakahara-ku, Kawasaki 211-8533, Japan
Tel. +81-44-733-5230; Fax +81-44-733-5192
e-mail: memi@nms.ac.jp

S. Sugano
Department of Virology, Institute of Medical Science, University of Tokyo, Tokyo, Japan

Introduction

Human cancers derived from breast, esophagus, and ovary frequently show allelic losses on the long arm of chromosome 17 (Fukino et al. 1999; Kalikin et al. 1996; Saito et al. 1993). Putative tumor suppressor genes on 17q are postulated as targets of these events. *BRCA1*, known primarily as the site of inherited mutations that confer predisposition to breast cancers, was initially considered to be the target of 17q deletions in sporadic cases also, because it is located at 17q21, where loss of heterozygosity (LOH) is common. However, inactivating mutations within the *BRCA1* gene do not appear to play a major role in the development of sporadic breast cancers. Therefore, regions of 17q other than the BRCA1 locus may be targets of LOH in those tumors.

A locus linked to hereditary focal nonepidermolytic palmoplantar keratoderma, a condition associated with cancer of the esophagus, lies on 17q25.1 (Kelsell et al. 1996); this esophageal-cancer susceptibility locus, TOC (tylosis with oesophageal cancer), could be a target of 17q deletion in other types of malignancy as well (Kalikin et al. 1996). We ourselves have defined a commonly deleted region in breast cancers within the 1-cM interval between D17S1603 and D17S1839 on chromosome 17q25.1 (Fukino et al. 1999). However, the genetic locus for TOC has recently been mapped to an equally narrow region flanked by D17S1839 and D17S785 (Risk et al. 1999) that is immediately adjacent to the segment we defined in breast cancers. These observations make it highly likely that the 17q25.1 region contains more than one tumor suppressor gene.

Here we describe the identification of a novel cDNA derived from chromosomal segment 17q25.1. Using the reverse transcriptase-polymerase chain reaction (RT-PCR) we found loss, or extreme decrease, of expression of this gene in 26 of 56 cancer cell lines derived from tumors that arose in various tissues. The 3293-bp full-length cDNA encodes a 788-amino-acid product with the characteristics of an integral membrane protein. As the gene

showed loss of expression in multiple types of human cancers, we named it *DMC1* ("downregulated in multiple cancers-1").

Materials and methods

Cloning strategy to identify genes with altered expression

We obtained information concerning expressed-sequence tags (ESTs) in the relevant region (D17S1352 to D17S785 on 17q25) from the human GeneMap 99 and from the Whitehead Institute Genome Database, purchased the appropriate clones from Genome Systems (St. Louis, MO, USA), and verified their sequences. We then examined each transcript for altered expression in our cancer cell lines by means of semi-quantitative RT-PCRs, using published primer sets.

Cell culture

The 56 cancer cell lines analyzed were derived from 10 breast cancers (MDA-MB-453, CRL1500, YMB-1-E, MCF7, HBL100, OCUB-M, BT-20, BT-474, MDA-MB-435S, and SK-BR-3), 10 uterine cancers (SIHA, HT-3, D98-AH-2, HeLaTG, HeLa, CaSki, ME-180, HeLaP3, HEC1-A, and SK-UT-1B), 13 hepatocellular carcinomas (SK-HEP-1, C-HC-4, Hep-KANO CL-2, WRL68, Chang liver, Hep-TABATA, HuH7, Hep G2, HT17, Li7, PLC/PRF/5, Hep3b, and C3A), 7 lung cancers (RERF-LC-AI, LU65, Lu99, PC14, A549, EBC-1, and LK-2), 6 thyroid cancers (WRO, NPA, 8305c, ARO, FRO, and 8505c), 4 renal-cell cancers (OS-RC-2, VMRC-RCW, RCC10RB, and Caki-1), and 6 gastric cancers (HuGC-OOHIRA, AZ521, H-111-TC, SH-10-TC, MKN-7, and NUGC-4). The cell lines were either donated by the Cell Resource Center for Biomedical Research at the Institute of Development, Aging, and Cancer of Tohoku University, or by Dr. Baba at Mie University, or they were purchased from the American Type Culture Collection (ATCC; Rockville, MD, USA). Each cell line was cultured in the optimized conditions recommended by the respective distributor.

DNA and RNA extraction and reverse transcription (RT)

DNA was extracted from the cell lines by phenol-chloroform procedures described previously (Emi et al. 1999). Total RNA was extracted using a detergent reagent, following the procedures described elsewhere (Kitamura et al. 1999). Each mRNA was prepared from 1×10^8 cells by the oligo-dT cellulose method (FastTrack 2.0 Kit; Invitrogen, Carlsbad, CA, USA) following the manufacturer's instructions. Reverse transcription reactions were carried out using the SMART RACE cDNA amplification kit (Clontech, Palo Alto, CA, USA) according to the manufacturer's instructions.

Northern-blot analysis of *DMC1*

PolyA RNA-blotted membranes prepared from various normal human tissues were obtained from Clontech. We followed the technique recommended in the Clontech manual for Northern analysis with EST stSG2564, as well as full-length *DMC1* cDNA as probe.

Polymerase chain reactions and RT-PCR

PCR amplifications were performed for 27 cycles in 30-μl reaction volumes containing 10 pmol of each primer, $1 \times$ PCR buffer with 1.5 mM MgCl2, and 0.5 units of Taq DNA polymerase, conditions that we empirically determined to enable comparison of gene expression levels. PCR products were stained with ethidium bromide after electrophoresis in agarose gels. Thermocycling conditions for amplification were described previously (Tsukamoto et al. 1998; Watanabe et al. 1998). To quantify relative expression levels among the cell lines, we carried out duplex PCRs with G3PDH primers (forward, 5′-ACCACAGTCCATGCCA TCAC-3′; reverse, 5′-TCCACCACCCTGTTGCTGTA-3′).

Single-strand conformational polymorphism (SSCP) analysis by silver staining

The coding region of *DMC1* cDNA was divided into 11 amplicons that overlapped their flanking segments by short sequences. Each segment was amplified by PCR; then, about 10 ng of each PCR product was heat-denatured in the presence of 80% of deionized formamide and electrophoresed in an 8% polyacrylamide gel, with or without 10% glycerol in $0.5 \times$ tris-borate/EDTA (TBE), at 8 V/cm for 10 h at room temperature (Hirayama et al. 1998). DNA fragments were visualized by silver staining, as described previously (Hoshino et al. 1992).

Sequence determination

Nucleotide sequences were determined by the BigDye (PE Applied Biosystems, Foster, CA, USA) Terminator cycle-sequencing method, using an auto-sequencer (ABI PRISM 377; PE Applied Biosystems, Foster, CA, USA).

Results

Transcript of downregulated gene at 17q25.1

Among ESTs from an interval flanked by D17S1352 and D17S785 at 17q25, a region that is commonly deleted in breast cancers (Fukino et al. 1999) and in esophageal cancers (Risk et al. 1999), one transcribed sequence (stSG2564; GenBank accession no. Z395239) showed loss of expression in multiple cell lines derived from various cancer types. To
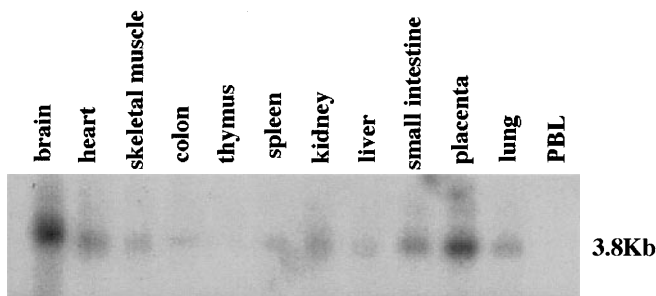
Fig. 1. Northern-blot analysis of *DMC1* in normal adult tissues. Each lane contains approximately 1 μg of polyA RNA (Clontech). The probes of both stSG2564 expressed-sequence tag (EST; shown here) and full-length *DMC1* cDNA (data not shown) displayed almost the same Northern blot profile *PBL*, Peripheral blood lymphocytes

examine the expression pattern and determine the size of the complete transcript, we hybridized this EST probe to a multiple-tissue Northern blot (Clontech). The 3.8-kb transcript (Fig. 1) was expressed ubiquitously, except for thymus and peripheral blood lymphocytes, and was most abundant in placenta and brain. The transcript observed in brain was slightly longer than that in other tissues, suggesting the existence of a splicing variant or the use of an alternative polyadenylation site in brain tissue. Northern hybridization carried out later with the full-length cDNA as probe revealed essentially the same profile (data not shown).

Cloning full-length cDNA for *DMC1*

To clone the full-length cDNA for the novel gene that we named "*DMC1*" (downregulated in multiple cancers-1), we used EST stSG2564 to screen a full-length-enriched cDNA library that had been constructed by means of oligo-capping, using mRNA derived from colonic mucosa (Maruyama and Sugano 1994; Suzuki et al. 1997). Positive clones isolated from this library showed sequence identity with stSG2564 within their 3′-untranslated regions. Sequencing of these clones revealed that the longest representative cDNA, HUV00353, contained an insert 3293 bp long. Its open reading frame (ORF), nucleotides 79 to 2442, encoded a protein of 788 amino acids (Fig. 2). Because the sequence surrounding the initiation codon agrees with the Kozak rule (Kozak 1984), this methionine most likely represents the transcription start site.

Structural features of the *DMC1* protein

The TMpred program (http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html) predicted that the *DMC1* product would be an integral membrane protein of the so-called "inside-out" type; its single transmembrane domain corresponds to amino acid residues 245 to 268 (see Fig. 2). It lacks a signal sequence but possesses an N-terminal cytoplasmic domain and a C-terminal extracellular domain. A basic local alignment search tool (BLAST) search found significant amino-acid homology (26% identical and 45%

positive) with a hypothetical 104.8-kD protein from *Schizosaccharomyces pombe* (GenBank accession no. O13776). Part of the *DMC1* cDNA showed similarity with an archived cDNA, DKFZp434I1120 (GenBank accession no. AL137556), although the latter lacks the 1074-bp of sequence in the 3′ region, and its encoded product contains an insertion of Trp at codon 130 and a substitution of Leu for Pro at codon 352.

Expression of *DMC1* in cancer cell lines and sequence variations

We examined the expression of *DMC1* in 56 cancer cell lines. RT-PCR experiments showed loss of expression of this gene in 2 of 10 breast-cancer lines, in 7 of 10 uterine-cancer lines, in 7 of 13 hepatocellular-cancer lines, in 3 of 7 lung-cancer lines, in 3 of 6 thyroid-cancer lines, in 2 of 6 gastric-cancer lines, and in 2 of 4 renal cell-cancer lines. We confirmed these results by semi-quantitative RT-PCR, using as an internal control the housekeeping gene *G3PDH*. That analysis demonstrated loss or dramatic attenuation of expression of *DMC1* in many cell lines (for examples, see Fig. 3).

To search for sequence variations in *DMC1* we amplified the full-length coding region of *DMC1* mRNA in each of 41 tumor cell lines. The RT-PCR products were screened for variations using SSCP analysis, followed by visualization with silver staining. RT-PCR products with potential variations were sequenced directly with automated instruments. By comparing normal tissues with cancer cell lines, we observed two silent polymorphisms within the coding region of *DMC1*. The first, cytosine/thymine, was located at the third nucleotide of codon 316 (Ala); the allelic frequencies were 0.62 for C and 0.38 for T. The second polymorphism, again cytosine/thymine, involved the third nucleotide of codon 502 (Thr); frequencies were 0.96 (C) and 0.04 (T). As both variations were observed in normal samples, neither represented a somatic mutation.

Discussion

Sporadic cancers of the breast, esophagus, and ovary have often shown allelic losses on the long arm of chromosome 17. Three or four distinct regions of deletion on 17q have been identified in sporadic breast cancers (Cropp et al. 1993; Nagai et al. 1994; Kalikin et al. 1996; Niederacher et al. 1997; Plummer et al. 1997). Mutations of the *BRCA1* gene at 17q21 are responsible for predisposition to a hereditary syndrome of breast/ovarian cancer. Moreover, the TOC (tylosis oesophageal cancer) locus associated with focal nonepidermolytic palmoplantar keratoderma, an autosomal dominant disorder of keratinization that is frequently associated with esophageal cancer, lies between markers AFMb054zf9 (D17S1839) and D17S1603 at 17q25.1 (Risk et al. 1994; Kelsell et al. 1996). We previously examined both of these loci in 178 primary breast cancers (Fukino et al. 1999), and observed allelic losses in 97 (55%) of these

```
  1 GCGGAGCTGGAGCCGGAGCTGAAGCCGGAGCCGGGTTGGAGTCTGGGCGGGGGCCGGGCCGGAGCGGGCTCCAGAGACATGGGGTCGACC  90
  1                                                                                    M  G  S  T   4
 91 GACTCCAAGCTGAACTTCCGGAAGGCGGTGATCCAGCTCACCACCAAGACGCAGCCCGTGGAAGCCACCGATGATGCCTTTTGGGACCAG 180
  5  D  S  K  L  N  F  R  K  A  V  I  Q  L  T  T  K  T  Q  P  V  E  A  T  D  D  A  F  W  D  Q  34
181 TTCTGGGCAGACACAGCCACCTCGGTGCAGGATGTGTTTGCACTGGTGCCGGCAGCAGAGATCCGGGCCGTGCGGGAAGAGTCACCCTCC 270
 35  F  W  A  D  T  A  T  S  V  Q  D  V  F  A  L  V  P  A  A  E  I  R  A  V  R  E  E  S  P  S  64
271 AACTTGGCCACCCTGTGCTACAAGGCCGTTGAGAAGCTGGTGCAGGGAGCTGAGAGTGGCTGCCACTCGGAGAAGGAGAAGCAGATCGTC 360
 65  N  L  A  T  L  C  Y  K  A  V  E  K  L  V  Q  G  A  E  S  G  C  H  S  E  K  E  K  Q  I  V  94
361 CTGAACTGCAGCCGGCTGCTCACCCGCGTGCTGCCCTACATCTTTGAGGACCCCGACTGGAGGGGCTTCTTCTGGTCCACAGTGCCCGGG 450
 95  L  N  C  S  R  L  L  T  R  V  L  P  Y  I  F  E  D  P  D  W  R  G  F  F  W  S  T  V  P  G 124
451 GCAGGGCGAGGAGGGCAGGGAGAAGAGGATGATGAGCATGCCAGGCCCCTGGCCGAGTCCCTGCTCCTGGCCATTGCTGACCTGCTCTTC 540
125  A  G  R  G  G  Q  G  E  E  D  D  E  H  A  R  P  L  A  E  S  L  L  L  A  I  A  D  L  L  F 154
541 TGCCCGGACTTCACGGTTCAGAGCCACCGGAGGAGCACTGTGGACTCGGCAGAGGACGTCCACTCCCTGGACAGCTGTGAATACATCTGG 630
155  C  P  D  F  T  V  Q  S  H  R  R  S  T  V  D  S  A  E  D  V  H  S  L  D  S  C  E  Y  I  W 184
631 GAGGCTGGTGTGGGCTTTGCTCACTCCCCCCAGCCTAACTACATCCACGATATGAACCGGATGGAGCTGCTGAAACTGCTGCTGACATGC 720
185  E  A  G  V  G  F  A  H  S  P  Q  P  N  Y  I  H  D  M  N  R  M  E  L  L  K  L  L  L  T  C 214
721 TTCTCCGAGGCCATGTACCTGCCCCCAGCTCCGGAAAGTGGCAGCACCAACCCATGGGTTCAGTTCTTTTGTTCCACGGAGAACAGACAT 810
215  F  S  E  A  M  Y  L  P  P  A  P  E  S  G  S  T  N  P  W  V  Q  F  F  C  S  T  E  N  R  H 244
811 GCCCTGCCCCTCTTCACCTCCCTCCTCAACACCGTGTGTGCCTATGACCCTGTGGGCTACGGGATCCCCTACAACCACCTGCTCTTCTCT 900
245  A  L  P  L  F  T  S  L  L  N  T  V  C  A  Y  D  P  V  G  Y  G  I  P  Y  N  H  L  L  F  S 274
901 GACTACCGGGAACCCCTGGTGGAGGAGGCTGCCCAGGTGCTCATTGTCACTTTGGACCACGACAGTGCCAGCAGTGCCAGCCCCACTGTG 990
275  D  Y  R  E  P  L  V  E  E  A  A  Q  V  L  I  V  T  L  D  H  D  S  A  S  S  A  S  P  T  V 304
991 GACGGCACCACCACTGGCACCGCCATGGATGATGCCGATCCTCCAGGCCCTGAGAACCTGTTTGTGAACTACCTGTCCCGCATCCATCGT 1080
305  D  G  T  T  T  G  T  A  M  D  D  A  D  P  P  G  P  E  N  L  F  V  N  Y  L  S  R  I  H  R 334
1081 GAGGACGGACTTCCAGTTCATCCTCAAGGGTATAGCCCGGCTGCTGTCCAACCCCCTGCTCCAGACCTACCTGCCTAACTCCACCAAGAAG 1170
335  E  D  F  Q  F  I  L  K  G  I  A  R  L  L  S  N  P  L  L  Q  T  Y  L  P  N  S  T  K  K 364
1171 ATCCAGTTCCACCAGGAGCTGCTAGTTCTCTTCTGGAAGCTCTGCGACTTCAACAAGAAATTCCTCTTCTTCGTGCTGAAGAGCAGCGAC 1260
365  I  Q  F  H  Q  E  L  L  V  L  F  W  K  L  C  D  F  N  K  K  F  L  F  F  V  L  K  S  S  D 394
1261 GTCCTGGACATCCTTGTCCCCATCCTCTTCTTCCTCAACGATGCCCGGGCCGATCAGTCTCGGGTGGGCCTGATGCACATTGGTGTCTTC 1350
395  V  L  D  I  L  V  P  I  L  F  F  L  N  D  A  R  A  D  Q  S  R  V  G  L  M  H  I  G  V  F 424
1351 ATCTTGCTGCTTCTGAGCGGGGAGCGGAACTTCGGGGTGCGGCTGAACAAACCCTACTCAATCCGCGTGCCCATGGACATCCCAGTCTTC 1440
425  I  L  L  L  S  G  E  R  N  F  G  V  R  L  N  K  P  Y  S  I  R  V  P  M  D  I  P  V  F 454
1441 ACAGGGACCCACGCCGACCTGCTCATTGTGGTGTTCCACAAGATCATCACCAGCGGGCACCAGCGGTTGCAGCCCCTCTTCGACTGCCTG 1530
455  T  G  T  H  A  D  L  L  I  V  V  F  H  K  I  I  T  S  G  H  Q  R  L  Q  P  L  F  D  C  L 484
1531 CTCACCATCGTGGTCAACGTGTCCCCCTACCTCAAGAGCCTGTCCATGGTGACCGCCAACAAGTTGCTGCACCTGCTGGAGGCCTTCTCC 1620
485  L  T  I  V  V  N  V  S  P  Y  L  K  S  L  S  M  V  T  A  N  K  L  L  H  L  L  E  A  F  S 514
1621 ACCACCTGGTTCCTCTTCTCTGCCGCCCAGAACCACCACCTGGTCTTCTTCCTCCTGGAGGTCTTCAACAACATCATCCAGTACCAGTTT 1710
515  T  T  W  F  L  F  S  A  A  Q  N  H  H  L  V  F  F  L  L  E  V  F  N  N  I  I  Q  Y  Q  F 544
1711 GATGGCAACTCCAACCTGGTCTACGCCATCATCCGCAAGCGCAGCATCTTCCACCAGCTGGCCAACCTGCCCACGGACCCGCCCACCATT 1800
545  D  G  N  S  N  L  V  Y  A  I  I  R  K  R  S  I  F  H  Q  L  A  N  L  P  T  D  P  P  T  I 574
1801 CACAAGGCCCTGCAGCGGCGCCGGCGGACACCTGAGCCCTTGTCTCGCACCGGCTCCCAGGAGGGCACCTCCATGGAGGGCTCCCGCCCC 1890
575  H  K  A  L  Q  R  R  R  R  R  T  P  E  P  L  S  R  T  G  S  Q  E  G  T  S  M  E  G  S  R  P 604
1891 GCTGCCCCTGCAGAGCCAGGCACCCTCAAGACCAGTCTGGTGGCTACTCCAGGCATTGACAAGCTGACCGAGAAGTCCCAGGTGTCAGAG 1980
605  A  A  P  A  E  P  G  T  L  K  T  S  L  V  A  T  P  G  I  D  K  L  T  E  K  S  Q  V  S  E 634
1981 GATGGCACCTTGCGGTCCCTGGAACCTGAGCCCCAGCAGAGCTTGGAGGATGGCAGCCCAGCTAAGGGGGAGCCCAGCCAGGCGTGGAGG 2070
635  D  G  T  L  R  S  L  E  P  E  P  Q  Q  S  L  E  D  G  S  P  A  K  G  E  P  S  Q  A  W  R 664
2071 GAGCAGCGGCGACCGTCCACCTCATCAGCCAGTGGGCAGTGGAGCCCAACGCCAGAGTGGGTCCTCTCCTGGAAGTCGAAGCTGCCGCTG 2160
665  E  Q  R  R  P  S  T  S  S  A  S  G  Q  W  S  P  T  P  E  W  V  L  S  W  K  S  K  L  P  L 694
2161 CAGACCATCATGAGGCTGCTGCAGGTGCTGGTTCCGCAGGTGGAGAAGATCTGCATCGACAAGGGCCTGACGGATGAGTCTGAGATCCTG 2250
695  Q  T  I  M  R  L  L  Q  V  L  V  P  Q  V  E  K  I  C  I  D  K  G  L  T  D  E  S  E  I  L 724
2251 CGGTTCCTGCAGCATGGCACCCTGGGGCTGGGCCTGCTGCCCGTGCCCCACCCCATCCTCATCCGCAAGTACCAGGCCAACTCGGGCACTGCC 2340
725  R  F  L  Q  H  G  T  L  G  L  G  L  L  P  V  P  H  P  I  L  I  R  K  Y  Q  A  N  S  G  T  A 754
2341 ATGTGGTTCCGCACCTACATGTGGGGCGTCATCTATCTGAGGAATGTGGACCCCCCTGTCTGGTACGACACCGACGTGAAGCTGTTTGAG 2430
755  M  W  F  R  T  Y  M  W  G  V  I  Y  L  R  N  V  D  P  P  V  W  Y  D  T  D  V  K  L  F  E 784
2431 ATACAGCGGGTGTGAGGATGAAGCCGACGAGGGGCTCAGTCTAGGGGAAGGCAGGGCCTTGGTCCCTGAGGCTTCCCCCATCCACCATTC 2520
785  I  Q  R  V  *                                                                             788
2521 TGAGCTTTAAATTACCACGATCAGGGCCTGGAACAGGCAGAGTGGCCCTGAGTGTCATGCCCTAGAGACCCCTGTGGCCAGGACAATGTG 2610
2611 AACTGGCTCAGATCCCCCTCAACCCCTAGGCTGGACTCACAGGAGCCCCATCTCTGGGGCTATGCCCCCACCAGAGACCACTGCCCCCAA 2700
2701 CACTCGGACTCCCTCTTTAAGACCTGGCTCAGTGCTGGCCCCTCAGTGCCCACCCACTCCTGTGCTACCCAGCCCCAGAGGCAGAAGCCA 2790
2791 ATGGGTCACTGTGCCCTAAGGGGTTTGACCAGGGAACCACGGGCTGTCCCTTAAGGTGCCTGGACAGGGTAAGGGGGTGTTTCCAGCCTC 2880
2881 CTAACCCAAAGCCAGTTGTTCCAGGCTCCAGGGGAAAAAGGTGTGGCCAGGCTGCTCCTCGAGGAGGCTGGGAGCTGGCCGACTGCAAAA 2970
2971 GCCAGACTGGGGCACCTCCCGTATCCTTGGGGCATGGTGTGGGGTGGTGAGGGTCTCCTGCTATATTCTCCTGGATCCATGGAAATAGCC 3060
3061 TGGCTCCCTCTTACCCAGTAATGAGGGGCAGGGAAGGGAACTGGGAGGCAGCCGTTTAGTCCTCCCTGCCCTGCCCACTGCCTGGATGGG 3150
3151 GCGATGCCACCCCTCATCCTTCACCCAGCTCTGGCCTCTGGGTCCCACCACCCAGCCCCCCGTGTCAGAACAATCTTTGCTCTGTACAAT 3240
3241 CGGCCTCTTTACAATAAAACCTCCTGCTCCAAAAAAAAAAAAAAAAAAAAAAAAA                                    3293
```

Fig. 2. Nucleotide and predicted amino acid sequences of *DMC1* cDNA. Bold *face* sequence is the coding region. A predicted transmembrane domain is *boxed*. The polyadenylation signal is indicated by *bold italic* characters. The *underlined* 3′-sequence is identical to EST stSG2564
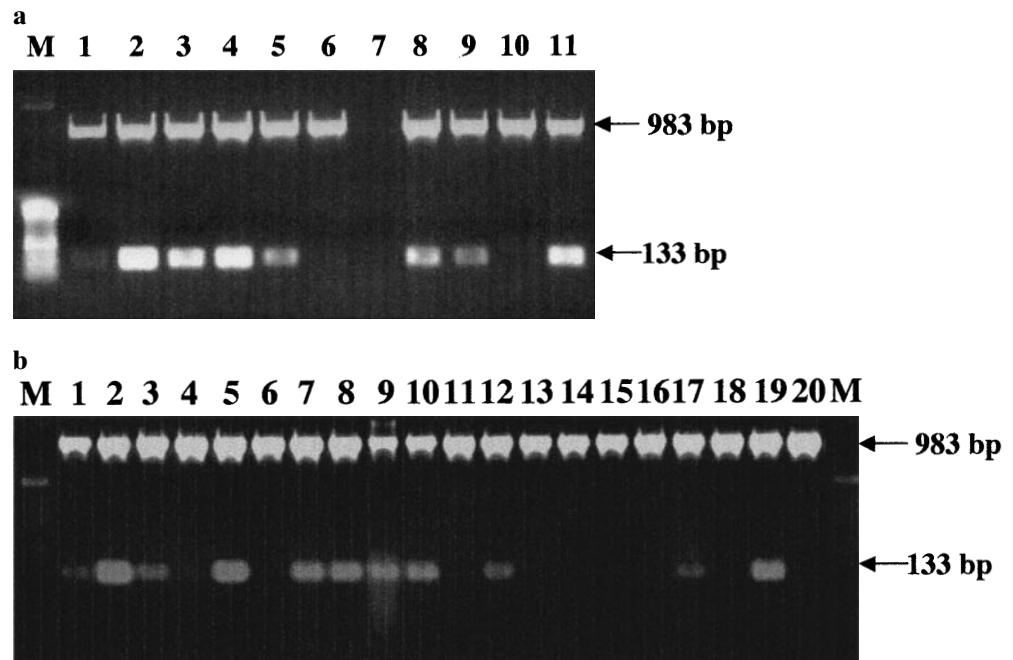
Fig. 3. a Reverse transcriptase-polymerase chain reaction (RT-PCR) analysis of *DMC1* expression in breast-cancer cell lines (*upper panel*). *Lane 1* contains cDNA from normal breast tissue. *Lanes 2–11* are cDNAs from breast-cancer cell lines, as follows: *lane 2*, MDA-MB-453; *lane 3*, CRL-1500; *lane 4*, YMB-1-E; *lane 5*, MCF-7; *lane 6*, HBL-100; *lane 7*, OCUB-M; *lane 8*, BT-20; *lane 9*, BT-474; *lane 10*, MDA-MB-435S; *lane 11*, SK-BR-3. The RT-PCR 983-bp product is G3PDH and the 133-bp product is stSG2564 EST. **b** Expression of *DMC1* in gastric- and uterine-cancer cell lines. *Lanes 1–2 and 9–10* are cDNAs from normal stomach and normal uterus, respectively. *Lanes 3–8* are cDNAs from gastric-cancer cell lines, as follows: *lane 3*, HuGC-OOHIRA; *lane 4*, AZ521; *lane 5*, H-111-TC; *lane 6*, SH-10-TC; *lane 7*, MKN-7; *lane 8*, NUGC-4. *Lanes 11–20* are cDNAs from uterine-cancer cell lines, as follows: *lane 11*, SIHA; *lane 12*, HT-3; *lane 13*, D98-AH-2; *lane 14*, HeLaTG; *lane 15*, HeLa; *lane 16*, CaSki; *lane 17*, ME-180; *lane 18*, HeLa-P3; *lane 19*, HEC1-A; *lane 20*, SK-UT-1B. *M*, Size marker. The RT-PCR 983-bp product is G3PDH and the 133-bp product is stSG2564 EST

tumors. Losses were most frequent at markers around the TOC locus (48% at D17S1839 and 43% at D17S1603), where we identified a distinct commonly deleted region within a 1-cM interval. Another larger but separate commonly deleted region, including the *BRCA1* gene, exhibited allelic loss at D17S934 in 45% of the breast cancers examined. Therefore *BRCA1* and TOC appear to be independent targets of allelic loss on 17q in primary breast cancers, and inactivation of the TOC locus in particular may play an important role in the genesis of sporadic breast tumors. However, large-scale sequencing efforts around the TOC locus within 17q25.1, aided by the Human Genome Project, have not, as yet, resulted in the discovery of a classical (class I) tumor suppressor gene that shows frequent mutations or deletions in human cancers (Zou et al. 1994).

Two known genes in the 17q25.1 region, *EVPL* and *TK1*, can be considered candidates for palmoplantar keratoderma (PPK)/TOC (Petty et al. 1996; Ruhrberg et al. 1996). *EVPL* encodes envoplakin, which belongs to the family of keratin-binding proteins; its alteration may be associated with keratinocyte dysfunction (Petty et al. 1996). *TK1* encodes a soluble thymidine kinase whose abnormal expression has been observed in several malignancies, including breast cancer (Ruhrberg et al. 1996).

The present study has revealed loss of expression of a novel gene from this region, *DMC1*, in a large portion of the human cancer cell lines we examined. Genes that frequently show decreased expression in tumors are characterized as class II tumor suppressors (Zou et al. 1994). Our results suggest that *DMC1* might belong to class II, because of its location in the TOC region, the loss of expression in multiple cancer cell lines, and the absence of detectable mutations of *DMC1* in these specimens. Although no class I tumor suppressor has yet been identified from the 17q25.1 region, it is attractive to speculate that regional inactivation of multiple genes from this chromosomal band contributes to tumor formation (Zou et al. 1994). The 17q25.1 region shows frequent LOH in a number of different cancer types, such as breast, esophageal, and ovarian tumors. However, because our mutational analysis has shown that *DMC1* is not a mutational target, it is apparently inactivated by other mechanisms (e.g., allelic loss, methylation, or even haploinsufficiency). Future studies that include functional analysis of this gene through transfection to cancer cells, or methylation analysis once the promoter region is identified, would help to clarify these issues.

The majority of "inside-out" membrane proteins are cell surface-receptor or ligand proteins. These include CD40L (Hollenbaugh et al. 1992), CD26 (Hegen et al. 1990; Tanaka et al. 1992), CD23 (Ikuta et al. 1987; Ludin et al. 1987), NKR-P1 (Yokoyama and Seaman 1993), and FasL (Suda et al. 1993). DMC1 has the "Pro-Pro-X-Pro" motif near the transmembrane domain. This motif was found to be conserved among receptors for cytokines, such as interleukin-3

receptor (IL-3R)α, granulocyte macrophage-colony stimulating factor receptor (GM-CSFR)α, prolactin receptor, and growth hormone receptor (Miyajima et al. 1992). Thus, we speculate that DMC1 functions as a receptor protein for cytokines or other related peptides.

References

Cropp CS, Champeme MH, Lidereau R, Callahan R (1993) Identification of three regions on chromosome 17q in primary human breast carcinomas which are frequently deleted. Cancer Res 53:5617–5619

Emi M, Keicho N, Tokunaga K, Katsumata H, Souma S, Nakata K, Taguchi K, Ohishi N, Azuma A, Kudoh S (1999) Association of diffuse panbronchiolitis with microsatellite polymorphism at the human interleukin 8 (IL-8) locus. J Hum Genet 44:169–172

Fukino K, Iida A, Teramoto A, Sakamoto G, Kasumi F, Nakamura Y, Emi M (1999) Frequent allelic loss at the TOC locus on 17q25.1 in primary breast cancers. Genes Chromosom Cancer 24:345–350

Hegen M, Niedobitek G, Klein CE, Stein H, Fleischer B (1990) The T cell triggering molecule Tp103 is associated with dipeptidyl aminopeptidase IV activity. J Immunol 144:2908–2914

Hirayama T, Yamaki E, Hara A, Tsuji M, Hashimoto K, Yamamoto M, Emi M (1998) Five familial hypercholesterolemic kindreds in Japan with novel mutation of the LDL receptor gene. J Hum Genet 43:250–254

Hollenbaugh D, Grosmaire LS, Kullas CD, Chalupny NJ, Braesch-Andersen S, Noelle RJ, Stamenkovic I, Ledbetter JA, Aruffo A (1992) The human T cell antigen gp39, a member of the TNF gene family, is a ligand for the CD40 receptor: expression of a soluble form of gp39 with B cell co-stimulatory activity. EMBO J 11:4313–4321

Hoshino S, Kimura A, Fukuda Y, Dohi K, Sasazuki T (1992) Polymerase chain reaction — single-strand conformation polymorphism analysis of polymorphism in *DPA1* and *DPB1* genes: a simple, economical, and rapid method for histocompatibility testing. Hum Immunol 33:98–107

Ikuta K, Takami M, Kim CW, Honjo T, Miyoshi T, Tagaya Y, Kawabe T, Yodoi J (1987) Human lymphocyte Fc receptor for IgE: sequence homology of its cloned cDNA with animal lectins. Proc Natl Acad Sci USA 84:819–823

Kalikin LM, Qu X, Frank TS, Caduff RF, Svoboda SM, Law DJ, Petty EM (1996) Detailed deletion analysis of sporadic breast tumors defines an interstitial region of allelic loss on 17q25. Genes Chromosom Cancer 17:64–68

Kelsell DP, Risk JM, Leigh IM, Stevens HP, Ellis A, Hennies HC, Reis A, Weissenbach J, Bishop DT, Spurr NK, Field JK (1996) Close mapping of the focal non-epidermolytic palmoplantar keratoderma (PPK) locus associated with oesophageal cancer (TOC). Hum Mol Genet 5:857–860

Kitamura Y, Nakata T, Minobe K, Shimizu K, Tanaka S, Fujimori M, Yokoyama Y, Ito K, Onda M, Emi M (1999) Ret/PTC3 is the most frequent form of gene rearrangement in papillary thyroid carcinomas in Japan. J Hum Genet 44:96–102

Kozak M (1984) Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res 12:857–872

Ludin C, Hofstetter H, Sarfati M, Levy CA, Suter U, Alaimo D, Kilchherr E, Frost H, Delespesse G (1987) Cloning and expression of the cDNA coding for a human lymphocyte IgE receptor. EMBO J 6:109–114

Maruyama K, Sugano S (1994) Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. Gene 138:171–174

Miyajima A, Kitamura T, Harada N, Yokota T, Arai K (1992) Cytokine receptors and signal transduction. Annu Rev Immunol 10:295–331

Nagai MA, Yamamoto L, Salaorni S, Pacheco MM, Brentani MM, Barbosa EM, Brentani RR, Mazoyer S, Smith SA, Ponder BA, Mulligan LM (1994) Detailed deletion mapping of chromosome segment 17q12–21 in sporadic breast tumors. Genes Chromosom Cancer 11:58–62

Niederacher D, Picard F, van Roeyen C, An HX, Bender HG, Beckmann MW (1997) Patterns of allelic loss on chromosome 17 in sporadic breast carcinomas detected by fluorescent-labeled microsatellite analysis. Genes Chromosom Cancer 18:181–192

Petty EM, Miller DE, Grant AL, Collins EE, Glover TW, Law DJ (1996) FISH localization of the soluble thymidine kinase gene (TK1) to human 17q25, a region of chromosomal loss in sporadic breast tumors. Cytogenet Cell Genet 72:319–321

Plummer SJ, Paris MJ, Myles J, Tubbs R, Crowe J, Casey G (1997) Four regions of allelic loss on 17q12-qter associated with high-grade breast tumors. Genes Chromosom Cancer 20:354–362

Risk JM, Field EA, Field JK, Whittaker J, Fryer A, Ellis A, Shaw JM, Friedmann PS, Bishop DT, Bodmer J, Leigh IM (1994) Tylosis oesophageal cancer mapped. Nat Genet 8:319–321

Risk JM, Ruhrberg C, Hennies H, Mills HS, Di Colandrea T, Evans KE, Ellis A, Watt FM, Bishop DT, Spurr NK, Stevens HP, Leigh IM, Reis A, Kelsell DP, Field JK (1999) Envoplakin, a possible candidate gene for focal NEPPK/esophageal cancer (*TOC*): the integration of genetic and physical maps of the TOC region on 17q25. Genomics 59:234–242

Ruhrberg C, Williamson JA, Sheer D, Watt FM (1996) Chromosomal localization of the human envoplakin gene (*EVPL*) to the region of the tylosis oesophageal cancer gene (*TOC*) on 17q25. Genomics 37:381–385

Saito H, Inazawa J, Saito S, Kasumi F, Koi S, Sagae S, Kudo R, Saito J, Noda K, Nakamura Y (1993) Detailed deletion mapping of chromosome 17q in ovarian and breast cancers: 2-cM region on 17q21.3 often and commonly deleted in tumors. Cancer Res 53:3382–3385

Suda T, Takahashi T, Golstein P, Nagata S (1993) Molecular cloning and expression of the Fas ligand, a novel member of the tumor necrosis factor family. Cell 75:1169–1178

Suzuki Y, Yoshitomo-Nakagawa K, Maruyama K, Suyama A, Sugano S (1997) Construction and characterization of a full length-enriched and a 5′-end-enriched cDNA library. Gene 200:149–156

Tanaka T, Camerini D, Seed B, Torimoto Y, Dang NH, Kameoka J, Dahlberg HN, Schlossman SF, Morimoto C (1992) Cloning and functional expression of the T cell activation antigen CD26. J Immunol 149:481–486

Tsukamoto K, Haruta K, Shiba T, Emi M (1998) Isolation and mapping of a polymorphic CA repeat sequence at the human interleukin 6 locus. J Hum Genet 43:71–72

Watanabe I, Tsukamoto K, Shiba T, Emi M (1998) Isolation and radiation hybrid mapping of dinucleotide repeat polymorphism at the human matrix Gla protein (MGP) locus. J Hum Genet 43:75–76

Yokoyama WM, Seaman WE (1993) The Ly-49 and NKR-P1 gene families encoding lectin-like receptors on natural killer cells: the NK gene complex. Annu Rev Immunol 11:613–635

Zou Z, Anisowicz A, Hendrix MJ, Thor A, Neveu M, Sheng S, Rafidi K, Seftor E, Sager R (1994) Maspin, a serpin with tumor-suppressing activity in human mammary epithelial cells. Science 263:526–529