Int. J. Cancer: **93,** 12–19 (2001)
© 2001 Wiley-Liss, Inc.
Publication of the International Union Against Cancer

SYSTEMATIC IDENTIFICATION OF GENES WITH CODING MICROSATELLITES MUTATED IN DNA MISMATCH REPAIR-DEFICIENT CANCER CELLS

Stefan M. WOERNER[1], Johannes GEBERT[1], Yan P. YUAN[2], Christian SUTTER[1], Ruediger RIDDER[1], Peer BORK[2,3] and Magnus VON KNEBEL DOEBERITZ[1]*

[1]*Division of Molecular Diagnostics and Therapy, Department of Surgery, University of Heidelberg, Heidelberg, Germany*
[2]*European Molecular Biology Laboratory, Heidelberg, Germany*
[3]*Max-Delbrück-Centrum for Molecular Medicine, Berlin, Germany*

Microsatellite instability (MSI) caused by deficient DNA mismatch-repair functions is a hallmark of cancers associated with the hereditary nonpolyposis colorectal cancer (HNPCC) syndrome but is also found in about 15% of all sporadic tumors. Most affected microsatellites reside in untranslated intergenic or intronic sequences. However, recently few genes with coding microsatellites were also shown to be mutational targets in MSI-positive cancers and might represent important mutation targets in their pathogenesis. The systematic identification of such genes and the analysis of their mutation frequency in MSI-positive cancers might thus reveal major clues to their functional role in MSI-associated carcinogenesis. We therefore initiated a systematic database search in 33,595 distinctly annotated human genes and identified 17,654 potentially coding mononucleotide repeats (cMNRs) and 2,028 coding dinucleotide repeats (cDNRs), which consist of $n \geq 6$ and $n \geq 4$ repeat units, respectively. Expression pattern and mutation frequency of 19 of these genes with the longest repeats were compared between DNA mismatch repair-deficient (MSI^+) and proficient (MSS) cancer cells. Instability frequencies in these coding microsatellite genes ranged from 10% to 100% in MSI-H tumor cells, whereas MSS cancer cells did not show mutations. RT-PCR analysis further showed that most of the affected genes (10/15) were highly expressed in tumor cells. The approach outlined here identified a new set of genes frequently affected by mutations in MSI-positive tumor cells. It will lead to novel and highly specific diagnostic and therapeutic targets for microsatellite unstable cancers.
© 2001 Wiley-Liss, Inc.

Key words: *mismatch repair; coding microsatellites; microsatellite instability; colorectal cancer*

Tumor cells accumulate mutations in components of cellular pathways that are essential for the maintenance of normal growth and differentiation. In human epithelial tumors, 2 types of genetic instability have been identified: chromosomal instability (CIN), which marks structural and numerical chromosomal aberrations in aneuploid neoplastic cells,[1] and microsatellite instability (MSI), which reflects length variations at repetitive DNA sequences in diploid tumor cells.[2,3] The type and spectrum of mutated genes markedly differs among CIN and MSI tumors,[2,4,5] suggesting distinct but not mutually exclusive pathways of carcinogenesis.[6] MSI occurs in about 90% of hereditary nonpolyposis colorectal cancers (HNPCC) as well as in about 15% of sporadic tumors of the colon and other organs[3,7] and is caused by mutational inactivation of different DNA mismatch-repair genes (hMSH2, hMLH1, hMSH6, hPMS1, hPMS2).[8–14] However, hMSH6 mutations appear to be associated with a low level of microsatellite instability mainly affecting mononucleotide repeats.[15,16] MSI tumors show particular clinicohistopathological characteristics[17,18] and have a better prognosis when compared with microsatellite stable colorectal tumors.[3,19–21]

Stratification of tumors according to their MSI status has been proposed to be a useful diagnostic parameter.[22] However, most microsatellites currently used for MSI classification including those recommended by the ICG-HNPCC[22] reside in noncoding intergenic or intronic sequences. Unlike length variations within these repeats, instability of coding microsatellites is expected to result in frameshift mutations of the corresponding genes, inevitably leading to truncated proteins.

Numerous coding microsatellites are expected to exist in the human genome, thereby defining a subset of target genes that might be specifically altered in mismatch repair (MMR)-deficient tumor cells. Thus microsatellite unstable cells might accumulate numerous truncated proteins. So far, only a limited number of coding microsatellite-containing genes frequently affected by microsatellite instability has been identified, encoding proteins either involved in signal transduction (TGFβRII, IGFIIR, PTEN),[23–25] apoptosis (BAX, Caspase 5),[26,27] DNA repair (hMSH3, hMSH6, MBD4),[28,29] transcriptional regulation (TCF-4)[30] or immune surveillance (β2M).[31] However, the frequency of instability of these coding microsatellites is highly variable among MSI tumors of different organs. Frameshift mutations in these genes seem to be selected for in MSI tumor cells, since coding microsatellites of identical length in other tested genes remained unaffected.[29] Moreover, biallelic inactivation of these genes has been reported, strongly suggesting that the encoded proteins display tumor-suppressive functions and thus represent major targets of the mutator pathway-associated carcinogenesis.[27,32]

In an attempt to gain more insight into the molecular mechanisms of the microsatellite mutator pathway, we performed a systematic database search for translated (coding) mono- and dinucleotide microsatellites in human. As a first approach, 15 cMNRs and 4 cDNRs consisting of at least 9 and 5 repeat units, respectively, were examined for genetic instability. The majority of cMNRs but none of the analyzed cDNRs showed a high frequency of genetic instability in MSI-H colorectal tumors and cell lines. Genes affected by frameshift mutations encode protein kinases, hormones, growth factors as well as proteins with different enzymatic or yet unknown functions. Ten of the 15 genes were found to be expressed in all MSI-H and MSS colorectal cancer (CRC) cell lines tested, the remaining 5 genes showing a highly variable

---

Abbreviations: cDNR, coding dinucleotide repeat; cMNR, coding mononucleotide repeat; CRC, colorectal cancer; HNPCC, hereditary nonpolyposis colorectal cancer; MSI, microsatellite instability; MSI-H, high-frequency MSI; MSI-L, low-frequency MSI; MSS, microsatellite stability.

Grant sponsor: Deutsche Krebshilfe. Grant sponsor: DFG; Grant sponsor: BMBF.

The first three authors contributed equally to this work.

Dr. Ridder's current address is: MTM Laboratories AG, INF 519, D-69120 Heidelberg, Germany.

*Correspondence to: Division of Molecular Diagnostics and Therapy, Department of Surgery, University of Heidelberg, Im Neuenheimer Feld 116, D-69120 Heidelberg, Germany. Fax: +49-6221-565981. E-mail: knebel@med.uni-heidelberg.de or mvkd@aol.com

Received 23 February 2001; Accepted 6 March 2001

expression pattern. However, no correlation was observed between mRNA expression level and mutation frequency. These results demonstrate the existence of a broad spectrum of cMNR-containing genes highly susceptible to frameshift mutations in MSI-H tumors. The knowledge of the whole spectrum of such genes, their mutation frequency and type of frameshift mutations not only enables novel diagnostic and therapeutic approaches but also should help to uncover the pattern of genetic alterations contributing to and specifying the microsatellite mutator pathway of tumorigenesis.

MATERIAL AND METHODS

Database analysis

To extract cMNRs and cDNRs, we screened the 109,289 entries with human sequences in the EMBL database (EMBL Rel. 62, March 2000). Using various Perl scripts,[33] we obtained annotated coding sequence (a total of 46,520 entries) and applied a rigorous redundancy check at the 98% level, thus detecting a total of 33,595 coding sequences in these entries. cDNA entries were compared with genomic DNA to identify corresponding mRNA. If both information sources were available for a particular gene, priority was given to the genomic sequence, since repeats in cDNA might span several exons. We finally considered 22,577 cDNAs (mRNAs) and 11,018 coding sequences in genomic entries for further analysis. The list of candidates is routinely updated every 3 months referring to the most recent EMBL database release.

In the coding regions, a minimum repeat length of 6 mononucleotides and 4 dinucleotides was required as the lower limit of the repeats. In regard to the dinucleotide repeats, all 12 different types of dinucleotide repeats were taken into consideration. cDNA and genomic DNA entries were analyzed separately. Using several filters, repeat tracts within pseudogenes, vector sequences as well as homopolymeric nucleotide stretches at the most 5′ or 3′ ends of sequences were excluded. All candidate sequences were stored in a relational database for further analysis. Subsequently, 2 independent BLASTN analyses were performed: Initial database analysis was done using the BLASTN included in the HUSAR program package of the German Cancer Research Center in Heidelberg, followed by an Advanced BLAST analysis of the high throughput genomic sequencing (htgs) database (http://www.ncbi.nlm.nih.gov/). Exon-intron boundaries were determined by MALIGN analysis, matching a candidate cDNA with homologous genomic DNA sequences. Finally, true coding microsatellites were verified by direct sequencing of repeat candidates amplified from genomic DNA.

Cell lines and tumor tissues

Fifteen human colorectal cancer cell lines were obtained from ECACC (http://www.camr.org.uk/frame.htm, SW48, HCT116), the German Cancer Research Center-Tumorbank (HT29, SW707, SW948, Caco 2, CX-2, CXF94, SW480, COLO320, LoVo, LS174T, LS180) or CLS Cell Lines Services, Heidelberg, Germany (COLO-60H, LS-513). The human colorectal adenocarcinoma cell line KM12 subclone KM12L4[34] was originally received from Dr. I.J. Fidler, MD Anderson Cancer Center, Houston, TX. The human colorectal adenocarcinoma cell lines CO115 and RKO were kindly provided by Dr. R. Hamelin, INSERM, Paris, France, and Dr. M. Brattain, University of Texas, Health Science Center, San Antonio, TX, respectively. Cells were grown in RPMI or DMEM supplemented with 10% FCS (Life Technologies, Karlsruhe, Germany), penicillin and streptomycin at 37°C with 5% CO_2. MSI-H (LoVo, KM12L4, HCT116, LS174T, LS180, CO115, RKO, COLO-60H and SW48) and MSS (CXF94, SW948, SW707, Caco2, HT29, Colo320, SW480, CX-2 and LS-513) cell lines have been previously classified with respect to their MSI status using the NCI ICG-HNPCC microsatellite reference marker panel.[22] Formalin-fixed and paraffin-embedded matched normal and tumor tissues of 20 MSI-H colorectal tumors, 1 MSI-H ovary tumor and 2 MSI-L/MSS colorectal tumors had been MSI classified accordingly.[35]

Microsatellite analysis

Genomic DNA was isolated from 5-8 haematoxylin- and eosin-stained 5 μm sections after microdissection using the Qiamp Tissue Kit (Qiagen, Hilden, Germany). Preparation of DNA from cell lines was performed according to standard protocols. PCR primers were designed to closely flank the target sequence, yielding short amplimeres of about 100 bp, thus allowing precise fragment sizing and robust amplification from archival tissues (Table I). PCR reactions were performed in a total volume of 25 μl containing 50 ng genomic DNA, 2.5 μl 10× reaction buffer (Life Technologies), 1.5 mM $MgCl_2$, 200 μM dNTPs, 0.3 μM of each primer and 0.5 U Taq DNA polymerase (Life Technologies) and using the following conditions: initial denaturation at 94°C for 4 min, followed by 35 cycles of denaturation at 94°C for 30 sec, annealing at 58°C for 45 sec and primer extension at 72°C for 30 sec. The final extension step was carried out at 72°C for 6 min. PCR fragments were analyzed on an ALF DNA sequencing device (Amersham Pharmacia Biotech, Freiburg, Germany) using 6.6% polyacrylamide/7 M urea gels. Size, height and profile of microsatellite peaks were analyzed using the AlleleLinks software (Amersham Pharmacia Biotech). Coding microsatellite instability was scored if smaller or larger-sized amplimeres were detected in tumor DNA compared with DNA from nonneoplastic cells. Allele intensities were determined and ratios of wild-type and novel alleles in normal and tumor tissues were calculated, defining a 2-fold difference as threshold for allelic shifts. Similarly, unstable alleles in tumor cell lines were identified by comparison with 36 unmatched normal mucosae. To determine the predicted repeat type and length, amplified coding microsatellites were subjected to Big Dye terminator cycle sequencing (Perkin Elmer, Darmstadt, Germany) and subsequent analysis on an ABI 310 sequencing device.

RT-PCR

Messenger RNA of 14 colon cancer cell lines was prepared as described.[36] An amount of 100 ng poly A+ RNA was reverse transcribed using 0.5 μg oligo $(dT)_{12\text{-}18}$ in a final volume of 20 μl with 200 U M-MLV Reverse Transcriptase (SuperScript, Life Technologies) for 1 hr at 37°C. RT-PCR amplifications were carried out as described above using 1 μl cDNA in 50 μl total reaction, 5 μl 10× reaction buffer (Life Technologies), 1.5 mM $MgCl_2$, 200 μM dNTPs, 0.25 μM of each primer and 0.5 U Taq DNA polymerase (Life Technologies). PCR products were visualized on ethidium bromide-stained 2% agarose gels. As a control for loading and integrity of mRNA, RT-PCR analysis of GAPDH was included. All PCR primers are listed in Table I. Fragment analysis was performed as for genomic analysis using fluorescein-labeled primers. PCR reactions for cDNA MSI analysis were done as described for expression analysis with the exception that one 5′ fluorescein-labeled primer was used.

RESULTS

Database search for coding mono- and dinucleotide microsatellite candidates

Our database searches revealed 17,654 cMNRs and 2,028 cDNRs, consisting of ≥6 and ≥4 repeat units, respectively. The longest cMNR was composed of 32 adenines, whereas the longest cDNR contained 42 nucleotides, *i.e.,* 21 GT repeat units. Note that we did not consider poly A or poly T stretches at or near the termini of annotated coding sequences. More than 90% of the identified cMNRs and cDNRs consisted of less than 9 or 5 repeat units, respectively, indicating an inverse correlation between repeat length and frequency of occurrence in the genome (Fig. 1). cMNRs containing homopolymeric stretches of $n \geq 9$ nucleotides represented only 2.1% (365/17,654) of candidate sequences, in contrast to about 9.2% of cDNRs (187/2,028) consisting of $n \geq 5$ repeat units. In addition to many novel coding region microsatellites affected by MSI, our set of cMNR sequences also included previously described coding microsatellites such as the A_{10}-tract in

TABLE I – PCR PRIMERS USED FOR MICROSATELLITE AND RT-PCR ANALYSIS

Gene	Accession no.	Chromosome	PCR primer for genomic DNA template		PCR primer for cDNA template	
			Sense 5′–3′	Antisense 5′–3′	Sense 5′–3′	Antisense 5′–3′
cMNRs						
FLT3LG	U29874	19q13.3	GGG ATG ACG TGG TGG TG	GTG ATC CAG GGC TTC AGC	CCT ATC TCC TCC TGC TGC TG	GTG ATC CAG GGC TTC AGC
SYCP1	X95654	1p13-12	CCC CTT CAT CTC TAA CAA CCC	CAC TGA TTC TCT GAA ATT AAA CAA ATA AC	CAG TGA AGA CAC CAA CAA AAC C	CAC TGA TTC TCT GAA ATT AAA CAA ATA AC
SLC4A3	U05596	2q36	TGG AGT GGA TGA GGA AGA GG	CTT CTG TGG GGT CCC TGA G	TGG AGT GGA TGA GGA AGA GG	ATC TGT GGG CAC CTG CTG
AC1	D82070	4	CCA GAA GCA AAT TCA CAA GAC	TTT TGC GTG TTC CTT CCT TC	CCA GAA GCA AAT TCA CAA GAC	CAC CCT CTC TCT TCT CCA GTA TTC
PTHL3	M24350	12p12-11	TTT CAC TTT CAG TAC AGC ACT TCT G	GAA GTA ACA GGG GAC TCT TAA ATA ATG	GGA AAC TAA CAA GGT GGA GAC G	GAA GTA ACA GGG GAC TCT TAA ATA ATG
SLC23A1	AF058319	20p13	GAC TAC TAC GCC TGT GCA CG	TGT TTA TTG CGT GGA TGG G	AAA GGA TGG ACT GCG TAC AAG	AAG GAC GAG CCC AAA GAA G
GART	X54199	21q22.11	AGT GTT GAA GAA TGG CTC CC	TGT TCC AGA TAT TAA GAC AGC CAC	GAA CAT CCC CAG AGT CCT CC	TGT TCC AGA TAT TAA GAC AGC CAC
MAC30X	L19183	17	TGT TGC GGA GCC CCT AC	AAC CAC CCT GTA GGC ATC TC	CCT GGT TTA AGT CCT TTC TGT TTT	AAC CAC CCT GTA GGC ATC TC
PRKDC	U63630	8q11	GAC TCA TGG ATG AAT TTA AAA TTG G	TTT GAA AAT AAC ATG TAA ATG CAT CTC	CAG CCC TGG ACC TTC TTA TTA A	GAC AAC CCC TTC AGA CAT CC
ATR	U76308	3q22-24	TCT TCT GTA GGA ACT TGA AAG CC	TGA AAG CAA GTT TTA CTG GAC TAG G	AGC TCC CAT GAA GTA ATC CG	TGA AAG CAA GTT TTA CTG GAC TAG G
MBD4	AF072250	3q21-22	TGA CCA GTG AAG AAA ACA GCC	GTT TAT GAT GCC AGA AGT TTT TTG	TGA CCA GTG AAG AAA ACA GCC	GTC GTG GGG GGC TAA GAG
SEC63	AF100141	6q16-22	AGT AAA GGA CCC AAG AAA ACT GC	TGC TTT TGT TTC TGT TGC TTT G	TGA AAA GGA GCA GTC CAT CTG	TGC TTT TGT TTC TGT TGC TTT G
OGT	U77413	Xq13	TCA CTT TTG GCT GGT CAG AG	GGG AGG GAA AGG AGG TAA AG	TCA CTT TTG GCT GGT CAG AG	TGT CAA AAA TGC GTG CCT C
HPDMPK	Y10936	19q13.3	GCT TGA TCC TGT TGA TTT TCT ACT C	CTG AAT GGA GAA GAA AGT GAG ATG	TCC TAC TGG ATG TGC TGC C	CTG AAT GGA GAA GAA AGT GAG ATG
U79260	U79260	16	TTT GTT ATA TCC CAT TAG GTG CC	AGC CTG GTG ACA GAG TGA GAC	CAT TAA GCA AAG CAG CCA GG	AGC CTG GTG ACA GAG TGA GAC
cDNRs						
CALCA	X15943	11p15	GTG CAG ATG AAG GCC AGT G	TGT ATG CTG CGG GGA GAG	GTG CAG ATG AAG GCC AGT G	GGA AGG AAA GGG AGG AGT TTA G
HPR	X89214	16q22.1	CTT AGT GGG AGG AGT GTG TGT G	AGA GAG AAA ACA AGA AGG AGC G	TGT GAA GCA GGG AGA CCA G	CTC CAA CAG CCT TAT TTA TCC AC
KIAA0040	D25539	1q24-25	CAT CTC AAT ATG GTT CCC AAG TG	CTT GCC CAC GTA CCT GCT AC	CAA GAA GTA ACG TGG AAG GAG G	GTG CAT TAT TTC AGG GGT TCC
API4 (Survivin)	U75285	17q25	CTT GAA AGT GGC ACC AGA GG	CGG GAA TCA AAA CAG CAA AA	GGA CAG AGA AAG AGC CAA GAA C	CGG GAA TCA AAA CAG CAA AA

Genes containing cMNR or cDNR sequences are listed together with the corresponding accession numbers and chromosomal loci.

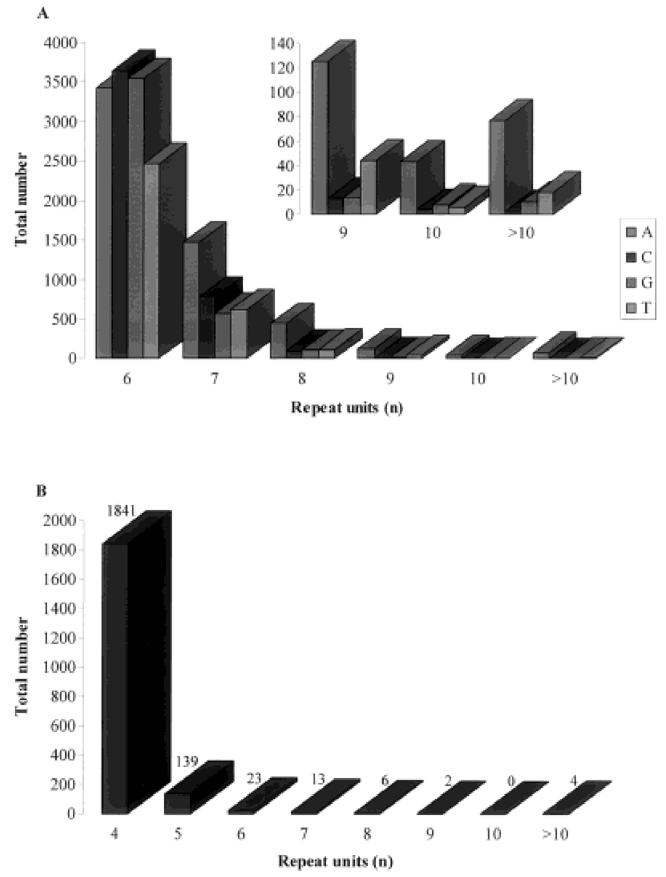

FIGURE 1 – Distribution of cMNR and cDNR candidate sequences in current databases. *(a)* Mononucleotide repeats are depicted according to repeat length and nucleotide type; the inset displays an enlarged view of distribution differences among homopolymeric runs of $n \geq 9$ repeat units. *(b)* Cumulative distribution of dinucleotide repeats.

the TGFβRII gene, the G_8-tracts in IGFIIR and BAX genes, the A_8-tract in the hMSH3 gene, the C_8-tract in the hMSH6 gene, as well as the recently described A_{10}-tract in the Caspase 5 and the A_9-tracts in the TCF-4 and CHK1 genes. These results suggest that our computational approach identified a comprehensive and unbiased subset of candidate coding microsatellites.

Identification of novel coding mono- and dinucleotide microsatellites

Since coding microsatellites of increased repeat length are expected to show a higher probability of instability in MMR-deficient cells than shorter repeat tracts, we chose 31 cMNR candidates [$(A)_{n\geq10}$, $(T)_{n\geq10}$, $(C)_{n\geq9}$, $(G)_{n\geq9}$] and 4 of the longest cDNR candidates [$(XY)_{n\geq5}$] for further analysis relying on EMBL database rel. 58. Subsequent computational comparison of cDNA and genomic sequences revealed that 4/4 cDNR and 7/31 cMNR candidates were present as contiguous repetitive sequences both in cDNA as well as in genomic DNA, 2/31 cMNR candidates occurred in genomic DNA only, whereas 22/31 cMNR candidates were found exclusively in cDNA sequences. When PCR primers flanking these repetitive sequences were designed, amplicons of the expected size could be obtained from genomic template DNA for all 4 cDNR and 18 cMNR candidates, including 9 cMNR candidate sequences originally lacking any information about genomic exon/intron boundaries. For 17/18 cMNR candidates, direct sequencing of the PCR-amplified microsatellites confirmed the presence of the predicted repeat tracts in coding regions of genomic DNA. In addition, 2 of these 17 cMNR candidates were

excluded because the predicted cDNA sequences were incorrect. Overall, 15 novel cMNRs and 4 novel cDNRs (Table I) have been experimentally verified as coding microsatellites in the corresponding genes and were used in subsequent analyses. During the course of this study, 3 of the newly identified cMNR-containing genes (SYCP1, ATR, MBD4/MED1) were reported by other investigators.[27,29,37]

Frequency of coding microsatellite instability in MSI-H tumors and cell lines

We examined instability at these newly identified coding microsatellites in 9 MSI-H and 9 MSS colorectal cancer cell lines. All cMNRs were found to be unstable in at least 1 of the MSI-H cell lines but remained unaltered in almost all MSS tumor cell lines (Table II). To explore whether the observed genetic alterations arose only in cultured cancer cell lines, we also investigated cMNR instabilities in MSI-H and MSI-L/MSS primary tumors and 18 MSS controls. cMNRs within 7 different genes (PTHL3, AC1, HPDMPK, OGT, U79260, SEC63 and SLC23A1) showed a high frequency of repeat-length variations, affecting at least 40% of MSI-H tumors (Table II). The same genes displayed frameshift mutations at similar frequencies in MSI-H CRC cell lines. As expected, the majority of these novel cMNRs (13/15) maintained stable repeat length in almost all matched normal mucosae, in 2 MSI-L/MSS tumors and in 9 MSS CRC cell lines. A control panel of 7 known cMNRs confirmed previously reported mutation frequencies on our set of MSI-H tumors and cell lines, ranging from 9% for hMSH6 to 80% for TGFβRII in the MSI-H tumors and thus excluding any sampling bias. The observed cMNR instabilities in MSI-H tumors and cell lines included both deletions and insertions inevitably shifting the translational reading frame of the affected genes (Fig. 2). However, one T_{14} cMNR within the U79260 mRNA also showed in frame deletions of 3 nucleotides in 7/9 cell lines and 3/20 tumors. No correlation could be observed between the type of frameshift, the frequency of mutations and the nucleotide composition or length of the analyzed coding repeats. In more than 50% of the MSI-H cell lines showing cMNR instability, only 1 allele was affected, while the repeat tract on the second allele remained stable. In most cases of MSI-H tumors showing cMNR instability, the wild-type length was detectable besides the mutated signal.

TABLE II – FREQUENCY OF cMNR INSTABILITY IN MSI-H COLORECTAL TUMORS AND CELL LINES

Gene	Repeat type	Cell Lines MSI+	Cell Lines MSI−	Tumors MSI+	Mucosae MSI−	Controls MSI−
FLT3LG	C_9	6/9[1]	0/9	7/20	0/18	1/16
SYCP1	A_{10}	2/9	0/9	3/21	2/18	0/18
SLC4A3	C_9	4/9	0/9	7/21	0/18	0/18
AC1	T_{10}	6/9	0/9	14/20	3/18	3/18
PTHL3	A_{11}	8/9	0/9	18/20	1/17	0/18
SLC23A1	C_9	7/9	0/9	9/20	0/18	0/18
GART	A_{10}	3/9	0/9	5/21	0/18	0/17
MAC30X	A_{10}	2/9	0/9	7/21	0/18	0/18
PRKDC	A_{10}	1/7	0/9	7/20	0/18	0/16
ATR	A_{10}	2/9	0/9	4/20	0/16	0/18
MBD4	A_{10}	1/9	0/9	2/21	0/18	0/18
SEC63	A_{10}	9/9	0/9	15/21	0/18	0/18
OGT	T_{10}	4/9	0/9	8/20	0/18	0/18
HPDMPK	T_{14}	6/6	0/8	19/20	7/17	5/17
U79260	T_{14}	9/9	3/9	16/20	3/18	1/18
hMSH3	A_8	2/6	0/8	5/10	0/10	n.d.
hMSH6	C_8	5/6	0/8	1/11	0/10	n.d.
BAX	G_8	5/6	0/8	5/11	0/10	n.d.
IGFIIR	G_8	2/6	0/8	4/11	0/10	n.d.
TCF-4	A_9	3/6	0/8	5/10	0/10	n.d.
Caspase 5	A_{10}	5/6	0/8	8/11	0/10	n.d.
TGFβRII	A_{10}	6/6	0/8	8/10	0/10	n.d.

[1]Listed are the numbers of unstable over the total number of samples analyzed.–n.d., not determined.

In contrast to cMNRs, cDNRs appeared to be remarkably stable in MSI-H tumors and cell lines. Only 1 unstable $(AC)_7$ cDNR (KIAA0040 gene) appeared in a single MSI-H tumor (data not shown), suggesting that cDNRs consisting of less than 9 repeat units are less error prone in mismatch repair-deficient cells. These results suggest a high frequency of cMNR instability in MSI-H colorectal tumors and cell lines, while cDNRs rather rarely display MSI.

Expression of genes containing coding microsatellites

When we examined the expression levels of 15 cMNR-containing genes by RT-PCR in 6 MSI-H and 8 MSS colon carcinoma cell lines, major differences were observed. Three of these genes either completely lacked or showed only marginal expression in all cell lines (SYCP1, FLT3LG and SLC4A3; Fig. 3). Two additional cMNR-bearing genes (AC1 and PTHL3) were expressed in about 50% of the cell lines at a rather mixed pattern, ranging from a very low level up to highly abundant transcripts in distinct cell lines. The remaining cMNR-containing genes (10/15) showed consistently high expression levels in all MSI-H and MSS cell lines tested. No relationship between expression level and MSI status could be observed.

MSI analysis of cDNAs

To determine whether somatic cMNR frameshift mutations identified in genomic DNA also occur in stable transcripts, we chose 11 cMNRs identified in our study as well as 6 known cMNR-containing genes for cDNA mutation analysis, excluding SLC4A3, FLT3LG, AC1 and SYCP1, which showed either no or infrequent cDNA expression in CRC cell lines (Table III). One novel (GART) and 1 known cMNR (hMSH6) showed only the wild-type repeat length in the cDNA sequence. Concordance of genomic and transcribed cMNR frameshift mutations was observed for 6 newly characterized (MAC30X, PRKDC, HPDMPK, OGT, U79260, MBD4) and 4 known cMNRs (hMSH3, BAX, TCF-4, TGFβRII). A rather mixed pattern, *e.g.,* genomic DNA/cDNA concordance of cMNR mutations in some cell lines and discordance in others occurred in SLC23A1, SEC63 and IGFIIR. In about 30% of cases, mutant and wild-type alleles of equal intensity were present in cMNR transcripts, and we noticed stable maintenance of 3 different-sized alleles for some cMNRs (SEC63, hMSH6, BAX) in different MSI-H CRC cell lines (Table III). These results demonstrate that about half of the genomic cMNR frameshift mutations are also found at the cDNA level in MSI-H tumor cell lines.

DISCUSSION

In our study, we applied computer-based algorithms to genome-wide screening of coding mono- and dinucleotide microsatellites. Our analysis not only identified a series of novel genes containing cMNRs but also detected all currently described genes with translated mononucleotide repeats, thus making any selection bias of our approach unlikely. It is important to note that not all candidate coding microsatellite sequences obtained in the primary database search represent true coding microsatellites and thus several restrictions apply. The quality of sequence retrieval is limited by the quality of the sequence information available in the database, *e.g.,* about 2/3 of cMNR candidate sequences ($n \geq 9$) were excluded due to sequencing artifacts, incomplete sequence annotation or if repeats were located in genes affected by somatic recombination. Microsatellites identified in EST or cDNA sequences always had to be compared with corresponding genomic DNA in order to verify repeat-length and to exclude those candidate repeats which originate from intron-spaced, bipartite repeats at the genomic DNA level. Thus, incorporation of expert knowledge as well as additional filters are an essential component of the proposed strategy in order to reduce the huge amount of false-positive candidates.

Our results show that cMNRs occurred more frequently than cDNRs with identical repeat length, which is in agreement with a

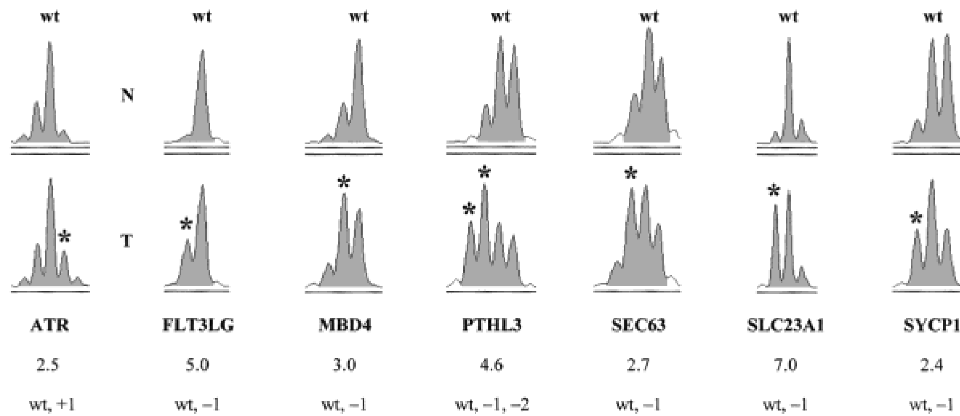

FIGURE 2 – Electropherograms of coding microsatellite instability. Fragment analysis of a representative set of coding mononucleotide repeats in matched normal (N) and tumor (T) tissues is shown. Corresponding wild-type alleles (wt) in normal and tumor tissues are indicated together with tumor-specific allelic shifts marked by asterisks. Calculation of peak intensity ratios is described in Material and Methods. Ratios above 2 are defined as threshold for shifted alleles.

LoVo, KM12, HCT116, LS174T, SW48, CX-2, CXF94, SW948, LS180, SW707, Caco 2, HT29, Colo320, SW480

Gene		Repeat Type
FLT3LG	Fms-related tyrosine kinase 3 ligand	C_9
SYCP1	Synaptonemal complex protein 1	A_{10}
SLC4A3	Anion exchanger 3 brain isoform	C_9
AC1	Differentially expressed in neuroblastoma	T_{10}
PTHL3	Parathyroid hormone-like protein (PTHrP), splice variant 3	A_{11}
SLC23A1	Nucleobase transporter	C_9
GART	Trifunctional ribonucleotide transferase/ synthetase	A_{10}
MAC30X	Differentially expressed in meningeoma	A_{10}
PRKDC	DNA dependent protein kinase, subunit c	A_{10}
ATR	Ataxia teleangiectasia and Rad3-related	A_{10}
MBD4	Methyl-CpG binding domain protein 4	A_{10}
SEC63	ER membrane protein	A_{10}
OGT	O-linked GlcNAc transferase	T_{10}
HPDMPK	Hypothetical protein downstream of DMPK and DMAHP locus	T_{14}
U79260	Clone 23745 mRNA	T_{14}
CALCA	Calcitonin related polypeptide alpha	$(AG)_5$
HPR	Haptoglobin related protein	$(GT)_8$
KIAA0040	Human mRNA for KIAA0040 gene	$(AC)_7$
API4	Apoptosis inhibitor 4 (Survivin)	$(AG)_8$
GAPDH		

FIGURE 3 – RT-PCR expression analysis of cMNR- and cDNR-containing genes in MSI-H and MSS CRC cell lines. Results of gel electrophoresis of RT-PCR for 15 cMNRs and 4 cDNRs are shown. As a control for RNA quality and reverse transcription GAPDH RT-PCR was performed.

recent study on primate libraries.[38] Repeat length among our verified cMNRs or cDNRs did not exceed 14 or 16 nucleotides, reflecting 14- or 8-repeat units, respectively. It is important to note that the most recent database search (EMBL rel. 62) revealed 4 novel cDNRs of more than 10-repeat units. Detailed analysis revealed that they represent noncoding microsatellites. It is conceivable that normal cells encounter problems in replicating, transcribing and/or translating longer runs of repetitive sequences.

TABLE III – ALLELE PATTERN REFLECTING cMNR INSTABILITY IN MSI-H CRC CELL LINES

Gene	Repeat type	Template[1]	LoVo	KM12	HCT116	LS174T	SW48	LS180
SLC23A1	C_9	g	wt	wt, +1	wt, −1	wt, −1	wt, +1	wt, −1
		c	wt	wt	wt	wt	wt, +1	wt
GART	A_{10}	g	wt	wt, +1	wt	wt	wt, +1	wt
		c	wt	wt	wt	wt	wt	wt
MAC30X	A_{10}	g	wt	wt	wt, −1	wt	wt	wt
		c	wt	wt	wt, −1	wt	wt	wt
PRKDC	A_{10}	g	wt	wt, +1	wt	wt	wt	wt
		c	wt	wt, +1	wt	wt	wt	wt
ATR	A_{10}	g	wt	wt	wt	wt	wt	wt
		c	wt	wt	wt	wt	wt	wt
MBD4	A_{10}	g	wt	wt	wt, −1	wt	wt	wt
		c	wt	wt	wt, −1	wt	wt	wt
SEC63	A_{10}	g	wt, −1, −2	wt, −1	wt, −1, −2	wt, −1, −2	wt, −1, +1	wt, −1, −2
		c	wt	wt	wt	wt	wt, −1	wt
OGT	T_{10}	g	−1	−1	wt	−1	wt	−1
		c	−1	−1	wt	−1	wt	−1
PTHL3	A_{11}	g	−1	−1	−2	wt, −1	wt, −1	wt, −1
		c	−1	n.e.	−2	n.e.	n.e.	n.e.
HPDMPK	T_{14}	g	−5, −2	−5, −3	−3	−4, −1	−4, −3	−4, −1
		c	−5, −2	−5	−3	−4, −1	−4	−4, −1
U79260	T_{14}	g	−1	−3	−4, −3	−3, −2	−2	−3, −2, −1
		c	−1	−3	−4	−3, −2	−2	−3, −2
hMSH3	A_8	g	wt	wt−1	−1	wt	wt	wt
		c	wt	wt−1	−1	wt	wt	wt
hMSH6	C_8	g	wt	wt, +1	wt, +1, +2	wt, −1	wt, +1	wt, −1
		c	wt	wt	wt	wt	wt	wt
BAX	G_8	g	wt, −1, +1	wt, −1, +1	wt, −1	−1	wt	−1
		c	wt, −1, +1	wt, −1, +1	wt, −1	−1	wt	−1
IGFIIR	G_8	g	wt	wt, +1	wt	wt	wt−1	wt
		c	wt	wt, +1	wt	wt	wt	wt
TCF-4	A_9	g	wt, −1	wt	wt	wt, −1	wt	wt, −1
		c	wt, −1	wt	wt	wt, −1	wt	wt, −1
TGFβRII	A_{10}	g	wt, −1, −2	wt, −1	−1	−1	−1	wt, −1
		c	−1, −2	wt, −1	−1	−1	−1	−1

[1]Results obtained from genomic DNA (g) or cDNA (c) templates, respectively. Results of 11 novel and 6 previously reported cMNR markers are shown for 6 MSI+ colorectal cancer cell lines.–wt, wild-type repeat length; +1, 1 nucleotide insertion; −1, 1 nucleotide deletion; n.e., no mRNA expression detectable by RT-PCR.

Replicative DNA polymerases preferentially generate single-base deletion and additional errors at homopolymeric repeat sequences and at a rate that increases with the number of repeats in the sequence.[39] This also might explain the high frequency of instability of mono- but not dinucleotide repeats in our study. We suppose that the number of long cDNR units in sequence databases is too low to show a mutation rate comparable with cMNRs. Presumably, evolutionary reasons may account for the observed lack of cDNR sequences extending beyond 11-repeat units. We did not observe any relationship between repeat tract length and mutation rate, which could be due to the small number of repeats investigated.

Most of the previous studies have focused on single genes like TGFβRII, BAX, IGFIIR and others and suggested that specific genes with known relation to carcinogenesis in other cancer models were targets of MSI-associated mutations in MSI tumors. In our study, we undertook for the first time a comprehensive approach to identify genes affected by MSI in mismatch repair-deficient tumors independent of their known role in other cancer models. Therefore, we systematically identified candidate genes containing coding microsatellites and began to analyze those with the longest repeat sequences and thus theoretically highest chances of mutations.[39] Using this approach, an initial subset of 15 experimentally verified coding microsatellites could be identified, which were characterized in our study. Clearly, these studies will be extended and the final goal will be to characterize all potential target sequences with coding repeats, which ultimately allows the establishment of a detailed map of genes and their mutational frequencies in the different progression steps of MSI-associated carcinogenesis. Only large-scale testing of all candidate microsatellites actually will elucidate the exact number of genes contributing to MSI carcinogenesis.

Mismatch repair-deficient tumors accumulate mutations in genes with microsatellites in coding regions. However, genetic instability does not affect all coding microsatellites to the same extent.[29] Survival of cell clones harbouring mutations in coding regions particularly underlie selective pressure. Frameshift mutations at some loci may confer a selective advantage during carcinogenesis. In favor of this hypothesis, a significant variation in the frequency of repeat instability could be observed among different cMNRs, even when comparing repeats of identical type and tract length (A_{10} of SEC63 like TGFβRII vs. A_{10} of PRKDC, MBD4, ATR, MAC30X and SYCP1). Similarly, cMNR mutation frequency did not correlate with the expression status of the corresponding genes. As an example, SYCP1 shows no remarkable mutation rate, although it is not expressed in the colon at all, whereas SEC63 is frequently mutated in spite of its high transcriptional activity. Preliminary data on mutation frequencies of several cMNRs identified in our study in endometrial carcinomas (data not shown) support and extend the previous observation of organ-specific differences.[24,27,30] Finally, some genes harbor multiple coding microsatellites (PRKDC, MBD4, OGT, SEC63, HPDMPK, ATR, SLC4A3, FLT3LG and SLC23A1). Recent studies on other genes containing multiple coding microsatellites indicate that only 1 repeat tract seems to be affected by genetic instability.[23,29,32] Further functional experiments may reveal whether frameshift mutations in specific genes provide growth-promoting properties or circumvent apoptotic pathways.

It is generally assumed that microsatellite instability observed in genomic DNA also manifests at the transcript level. Our results

only partially support this hypothesis. Although we have examined microsatellite length of 11 novel and 6 known cMNRs at the transcript level in a small number of MSI-H CRC cell lines, somatic genomic DNA frameshift mutations could be confirmed in cDNA in about 50% of cases. Lack of mutation in the cDNA was observed in a small subset of cMNRs including SLC23A1, GART, SEC63 and hMSH6. If a given gene is not expressed in a particular tissue like, for example, in colon, somatic mutations in the coding region of this gene even if they occur at high frequency are likely to be functionally irrelevant. However, if transcription occurs, failure to detect a given transcript even by RT-PCR could be due to low transcript levels or increased turnover of mutant transcripts.[40] Transcription and translation of frameshift mutations might be of high importance for the carcinogenic process. We emphasize that deletion/insertion mutations in coding region microsatellites cause frameshift mutations that result in truncated proteins. These mutant proteins by themselves might exert transdominant functions on eventually remaining wild-type proteins, and they might interact with other proteins, thereby representing "gain of function" mutations. Particularly important might be the fact that MSI-induced mutations result in the expression of novel frameshift peptides at the C-terminus of mutant and expressed genes. Expression of such mutant transcripts and encoded frameshift peptides would be a prerequisite for processing and presentation of derived neo-antigens, which would be MSI tumor-specific and might provide a source for new and tumor-specific antigens. Recent results from our laboratory on some of these neo-antigens consistently expressed in MSI-associated colorectal cancers support this hypothesis.[41] Taken together, these data suggest that in particular the expression of mutant genes with coding microsatellites will explain many biological features of MSI cancers and will also provide an ideal source for the identification of new and really tumor-specific tumor antigens for diagnostic and therapeutic applications.

Our genomic DNA and cDNA mutation analyses clearly demonstrated the occurrence of both mutated and wild-type cMNR alleles in primary tumors and cell lines. Despite microdissection, we cannot exclude the contamination of tumor tissue with small amounts of admixed stromal cells and lymphocytes, which might explain the wild-type signal in tumor tissues. However, cell lines are expected to represent a clonal population. Interestingly, we found different alleles in the majority of cell lines with mutated cMNRs. We speculate that subclones may emerge from 1 clonal cell population under high selective pressure and cell culture conditions. This hypothesis can be experimentally tested by limiting dilution experiments of mutated cell lines. The presence of wild-type and mutated alleles does not necessarily exclude biallelic inactivation of MSI target genes, since genomic imprinting or promoter methylation of the wild-type allele might also contribute to inactivation of the second allele. Alternatively, truncated proteins caused by frameshift mutations may confer a dominant negative phenotype, thus conferring a growth advantage to heterozygous cells.[42]

The algorithm developed for the identification of potential MSI target genes also provides a clue to the biologic relationship of the disease and the mutations implicated. In addition to known MSI target genes, whose products are involved in diverse cellular processes like signal transduction, transcription, apoptosis and DNA repair, our study revealed frameshift mutations in proteins including O-linked N-acetylglucosamine transferase (OGT) and an endoplasmic reticulum membrane protein (SEC63) that have not yet been shown to be involved in colon carcinogenesis. OGT attaches GlcNAc residues to serine/threonine residues of nuclear and cytoplasmic proteins in response to cellular signals in much the same way that kinases regulate protein phosphorylation. It has been suggested that OGT plays a direct role in regulating a number of cellular functions including protein synthesis, neurofilament assembly and transcription.[43] This eukaryotic enzyme is highly conserved throughout evolution and consists of an N-terminal regulatory and a C-terminal catalytic domain. Frameshift mutations affecting the T_{10} mononucleotide repeat are expected to truncate the protein in the N-terminal half, thereby eliminating catalytic activity. Since reciprocal O-linked glycosylation and phosphorylation have been reported on serine/threonine residues important in modulating oncoprotein activity,[44] it is tempting to speculate that altered glycosylation might contribute to MSI-mediated carcinogenesis.

Mammalian SEC63 is a ubiquitously expressed component of the ER protein translocation machinery.[45,46] In yeast cells, a mutation in SEC63 abolished peptide export and reduced misfolded protein export from the ER.[47] Since in mammalian cells transporters associated with antigen processing (TAP) efficiently import antigenic peptides from the cytosol into the ER of mammalian cells but only a fraction of these peptides can bind to the MHC class I complex for antigen presentation at the cell surface, the majority of TAP substrates are rapidly exported from the ER to the cytosol,[48] thereby ensuring that class I molecules are loaded preferentially with high-affinity peptides.[49] Impaired SEC63 function in MSI tumor cells might lead to the accumulation of frameshift and low-affinity peptides in the ER, and their aberrant presentation at the cell surface would make these cells vulnerable to cellular immune surveillance. Moreover, SEC63 deficiency generally is expected to interfere with ER protein import and could be deleterious to these cells. Both mechanisms would provide a molecular explanation for the better prognosis of MSI tumors.

Since 4 of these cMNRs (PTHL3, SEC63, HPDMPK and U79260) show mutation rates of 80% or more in MSI-H colon carcinoma samples or cell lines, our approach yields new candidate genes suitable for diagnostic purposes. One might argue that coding microsatellites are not eligible for diagnostic use since they are subject to selection mechanisms. However, our data show that frequently mutated coding microsatellite genes have a high specificity for the MSI status, despite the underlying selection processes.

Our results have several implications for molecular diagnostic and therapy of MSI-H tumors. First, our strategy allows the identification of novel genes affected by MSI potentially involved in the mutator pathway of carcinogenesis. The regulation of these genes together with structural and functional analysis of their encoded proteins might facilitate the design of novel therapeutics specific for the treatment of genetically unstable tumors. Second, monomorphic coding microsatellites represent a powerful tool for MSI typing. Combined analysis of 2 of the most frequently altered coding mononucleotide repeats identified in our study would be sufficient to detect MSI-H CRC tumors with high sensitivity, thus obviating the need for matched normal tissue and significantly reducing time and costs of this prescreening method. Third, variable instability frequencies among individual coding microsatellites in MSI-H tumors of different organs[6,27] (data not shown) indicate a complex pattern of somatic frameshift mutations in these repetitive sequences. Such cMNR instability profiles should reflect different tumor biology and thus might represent a potentially useful parameter for tumor stratification. Fourth, instability in each coding microsatellite causes a limited number of repeat mutations and predictable frameshifts that would facilitate the development of sequence-based and protein-based diagnostic screening assays. Likewise, the resulting predictable translational frameshifts often generate unique tumor antigens suitable for MSI tumor-specific vaccination approaches.

Overall, our results provide a platform of candidate cMNR-containing genes. High-throughput MSI analysis technology will be required to systematically uncover the spectrum of potential target genes and central components of the mutator pathway most frequently altered by somatic frameshift mutations.

ACKNOWLEDGEMENTS

P. Bork is supported by DFG and BMBF.

REFERENCES

1. Lengauer C, Kinzler KW, Vogelstein B. Genetic instability in colorectal cancers. Nature 1997;386:623–7.
2. Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature 1993;363:558–61.
3. Thibodeau SN, Bren G, Schaid D. Microsatellite instability in cancer of the proximal colon. Science 1993;260:816–9.
4. Konishi M, Kikuchi-Yanoshita R, Tanaka K, Muraoka M, Onda A, Okumura Y, et al. Molecular nature of colon tumors in hereditary nonpolyposis colon cancer, familial polyposis, and sporadic colon cancer. Gastroenterology 1996;111:307–17.
5. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. Nature 1998;396:643–9.
6. Perucho M, Peinado MA, Ionov Y, Casares S, Malkhosyan S, Stanbridge E. Defects in replication fidelity of simple repeated sequences reveal a new mutator mechanism for oncogenesis. Cold Spring Harb Symp Quant Biol 1994;59:339–48.
7. Aaltonen LA, Peltomaki P, Mecklin JP, Jarvinen H, Jass JR, Green JS, et al. Replication errors in benign and malignant tumors from hereditary nonpolyposis colorectal cancer patients. Cancer Res 1994;54: 1645–8.
8. Akiyama Y, Sato H, Yamada T, Nagasaki H, Tsuchiya A, Abe R, et al. Germ-line mutation of the hMSH6/GTBP gene in an atypical hereditary nonpolyposis colorectal cancer kindred. Cancer Res 1997; 57:3920–3.
9. Bronner CE, Baker SM, Morrison PT, Warren G, Smith LG, Lescoe MK, et al. Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. Nature 1994;368:258–61.
10. Fishel R, Lescoe MK, Rao MR, Copeland NG, Jenkins NA, Garber J, et al. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell 1993;75:1027–38.
11. Leach FS, Nicolaides NC, Papadopoulos N, Liu B, Jen J, Parsons R, et al. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer, Cell 1993;75:1215–25.
12. Miyaki M, Konishi M, Tanaka K, Kikuchi-Yanoshita R, Muraoka M, Yasuno M, et al. Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. Nat Genet 1997;17:271–2.
13. Nicolaides NC, Papadopoulos N, Liu B, Wei YF, Carter KC, Ruben SM, et al. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. Nature 1994;371:75–80.
14. Papadopoulos N, Nicolaides NC, Wei YF, Ruben SM, Carter KC, Rosen CA, et al. Mutation of a mutL homolog in hereditary colon cancer. Science 1994;263:1625–9.
15. Parc YR, Halling KC, Wang L, Christensen ER, Cunningham JM, French AJ, et al. HMSH6 alterations in patients with microsatellite instability-low colorectal cancer. Cancer Res 2000;60:2225–31.
16. Wu Y, Berends MJ, Mensink RG, Kempinga C, Sijmons RH, Der Zee AG, et al. Association of hereditary nonpolyposis colorectal cancer-related tumors displaying low microsatellite instability with MSH6 germline mutations. Am J Hum Genet 1999;65:1291–8.
17. Vasen HF, Watson P, Mecklin JP, Lynch HT. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. Gastroenterology 1999;116:1453–6.
18. Watson P, Lin KM, Rodriguez-Bigas MA, Smyrk T, Lemon S, Shashidharan M, et al. Colorectal carcinoma survival among hereditary nonpolyposis colorectal carcinoma family members. Cancer 1998;83:259–66.
19. Gryfe R, Kim H, Hsieh ET, Aronson MD, Holowaty EJ, Bull SB, et al. Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer. N Engl J Med 2000;342:69–77.
20. Lothe RA, Peltomaki P, Meling GI, Aaltonen LA, Nystrom-Lahti M, Pylkkanen L, et al. Genomic instability in colorectal cancer: relationship to clinicopathological variables and family history. Cancer Res 1993;53:5849–52.
21. Sankila R, Aaltonen LA, Jarvinen HJ, Mecklin JP. Better survival rates in patients with MLH1-associated hereditary colorectal cancer. Gastroenterology 1996;110:682–7.
22. Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, et al. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res 1998;58:5248–57.
23. Guanti G, Resta N, Simone C, Cariola F, Demma I, Fiorente P, et al. Involvement of PTEN mutations in the genetic pathways of colorectal cancerogenesis. Hum Mol Genet 2000;9:283–7.
24. Markowitz S, Wang J, Myeroff L, Parsons R, Sun L, Lutterbaugh J, et al. Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. Science 1995;268:1336–8.
25. Souza RF, Appel R, Yin J, Wang S, Smolinski KN, Abraham JM, et al. Microsatellite instability in the insulin-like growth factor II receptor gene in gastrointestinal tumours. Nat Genet 1996;14:255–7.
26. Rampino N, Yamamoto H, Ionov Y, Li Y, Sawai H, Reed JC, et al. Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. Science 1997;275:967–9.
27. Schwartz S Jr, Yamamoto H, Navarro M, Maestro M, Reventos J, Perucho M. Frameshift mutations at mononucleotide repeats in caspase-5 and other target genes in endometrial and gastrointestinal cancer of the microsatellite mutator phenotype. Cancer Res 1999;59: 2995–3002.
28. Malkhosyan S, Rampino N, Yamamoto H, Perucho M. Frameshift mutator mutations. Nature 1996;382:499–500.
29. Riccio A, Aaltonen LA, Godwin AK, Loukola A, Percesepe A, Salovaara R, et al. The DNA repair gene MBD4 (MED1) is mutated in human carcinomas with microsatellite instability. Nat Genet 1999; 23:266–8.
30. Duval A, Gayet J, Zhou XP, Iacopetta B, Thomas G, Hamelin R. Frequent frameshift mutations of the TCF-4 gene in colorectal cancers with microsatellite instability. Cancer Res 1999;59:4213–5.
31. Bicknell DC, Kaklamanis L, Hampson R, Bodmer WF, Karran P. Selection for beta 2-microglobulin mutation in mismatch repair-defective colorectal carcinomas. Curr Biol 1996;6:1695–7.
32. Chadwick RB, Jiang GL, Bennington GA, Yuan B, Johnson CK, Stevens MW, et al. Candidate tumor suppressor RIZ is frequently involved in colorectal carcinogenesis. Proc Natl Acad Sci USA 2000; 97:2662–7.
33. Wall L, Christiansen T, Schwartz RL. Programming Perl, 2nd edition. O'Reilly Associates, Cambridge, UK 1996.
34. Morikawa K, Walker SM, Nakajima M, Pathak S, Jessup JM, Fidler IJ. Influence of organ environment on the growth, selection, and metastasis of human colon carcinoma cells in nude mice. Cancer Res 1988;48:6863–71.
35. Sutter C, Gebert J, Bischoff P, Herfarth C, von Knebel Doeberitz M. Molecular screening of potential HNPCC patients using a multiplex microsatellite PCR system. Mol Cell Probes 1999;13:157–65.
36. Vennstrom B, Bishop JM. Isolation and characterization of chicken DNA homologous to the two putative oncogenes of avian erythroblastosis virus. Cell 1982;28:135–43.
37. Mironov N, Jansen LA, Zhu WB, Aguelon AM, Reguer G, Yamasaki H. A novel sensitive method to detect frameshift mutations in exonic repeat sequences of cancer-related genes. Carcinogenesis 1999;20: 2189–92.
38. Metzgar D, Bytof J, Wills C. Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res 2000; 10:72–80.
39. Kunkel TA, Patel SS, Johnson KA. Error-prone replication of repeated DNA sequences by T7 DNA polymerase in the absence of its processivity subunit. Proc Natl Acad Sci USA 1994;91:6830–4.
40. Maquat LE. When cells stop making sense: effects of nonsense codons on RNA metabolism in vertebrate cells. RNA 1995;1:453–65.
41. Linnebacher M, Gebert J, Rudy W, Woerner S, Yuan P, Bork P, et al. Frameshift peptide-derived T-cell epitopes: a source of novel tumor-specific antigens. Int J Cancer 2001; 93:6–11.
42. Dihlmann S, Gebert J, Siermann A, Herfarth C, von Knebel Doeberitz M. Dominant negative effect of the APC1309 mutation: a possible explanation for genotype-phenotype correlations in familial adenomatous polyposis. Cancer Res 1999;59:1857–60.
43. Kreppel LK, Hart GW. Regulation of a cytosolic and nuclear O-GlcNAc transferase. Role of the tetratricopeptide repeats. J Biol Chem 1999;274:32015–22.
44. Chou TY, Hart GW, Dang CV. c-Myc is glycosylated at threonine 58, a known phosphorylation site and a mutational hot spot in lymphomas. J Biol Chem 1995;270:18961–5.
45. Meyer HA, Grau H, Kraft R, Kostka S, Prehn S, Kalies KU, Hartmann E. Mammalian Sec61 is associated with Sec62 and Sec63. J Biol Chem 2000;275:14550–7.
46. Skowronek MH, Rotter M, Haas IG. Molecular characterization of a novel mammalian DnaJ-like Sec63p homolog. Biol Chem 1999;380: 1133–8.
47. Gillece P, Pilon M, Romisch K. The protein translocation channel mediates glycopeptide export across the endoplasmic reticulum membrane. Proc Natl Acad Sci USA 2000:97:4609–14.
48. Roelse J, Gromme M, Momburg F, Hammerling G, Neefjes J. Trimming of TAP-translocated peptides in the endoplasmic reticulum and in the cytosol during recycling. J Exp Med 1994;180:1591–7.
49. Schumacher TN, Kantesaria DV, Heemels MT, Ashton-Rickardt PG, Shepherd JC, Fruh K, et al. Peptide length and sequence specificity of the mouse TAP1/TAP2 translocator. J Exp Med 1994;179:533–40.