



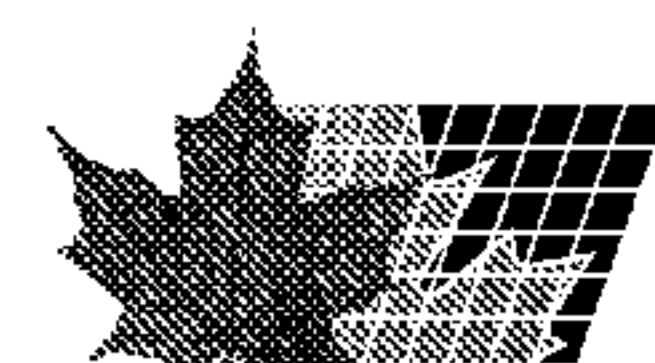
(86) Date de dépôt PCT/PCT Filing Date: 2001/07/04
(87) Date publication PCT/PCT Publication Date: 2003/01/07
(85) Entrée phase nationale/National Entry: 2003/01/07
(86) N° demande PCT/PCT Application No.: DE 2001/002510
(87) N° publication PCT/PCT Publication No.: 2002/004664
(30) Priorité/Priority: 2000/07/07 (100 32 608.0) DE

(51) Cl.Int.⁷/Int.Cl.⁷ C12Q 1/68
(71) Demandeur/Applicant:
KNEBEL DOEBERITZ, VON MAGNUS, DE
(72) Inventeurs/Inventors:
BORK, PEER, DE;
YUAN, YAN PING, DE;
GEBERT, JOHANNES, DE;
WORNER, STEFAN, DE;
LINNEBACHER, MICHAEL, DE
(74) Agent: BERESKIN & PARR

(54) Titre : GENES IMPORTANTS POUR LES TUMEURS A INSTABILITE DES MICROSATELLITES (MSI+) ET LEURS PRODUITS GENIQUES
(54) Title: GENES AND THEIR GENETIC PRODUCTS PERTINENT TO MICROSATELLITE INSTABLE (MSI+) TUMOURS

(57) **Abrégé/Abstract:**

The invention relates to genes comprising coding mononucleotide microsatellites (cMNR) or dinucleotide microsatellites (cDNR). The genes can be isolated from MSI+ tumour cells. Said genes differ from corresponding genes from non-MSI+ (tumour) cells by mutations in the cMNR or cDNR and code for gene products including neopeptides. The invention also relates to the use of the genes and their gene products for the prevention, diagnosis and/or therapy of MSI+ tumours.



(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

BERICHTIGTE FASSUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro(43) Internationales Veröffentlichungsdatum
17. Januar 2002 (17.01.2002)

PCT

(10) Internationale Veröffentlichungsnummer
WO 02/004664 A2(51) Internationale Patentklassifikation⁷: **C12Q 1/68**(74) **Anwalt: HUBER, Bernard**; Huber & Schüssler, Trud-
eringer Strasse 246, 81825 München (DE).

(21) Internationales Aktenzeichen: PCT/DE01/02510

(22) Internationales Anmeldedatum:
4. Juli 2001 (04.07.2001)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
100 32 608.0 7. Juli 2000 (07.07.2000) DE

(71) Anmelder und

(72) **Erfinder: KNEBEL DOEBERITZ, von, Magnus**
[DE/DE]; Chirurgische Universitätsklinik, Sektion
Molekulare Diagnostik und Therapie, Im Neuenheimer
Feld 110, 69120 Heidelberg (DE).(72) **Erfinder; und**(75) **Erfinder/Anmelder (nur für US): BORK, Peer** [DE/DE];
Maxerhofstr. 1, 69118 Heidelberg (DE). **YUAN, Yan,**
Ping [DE/DE]; Mayerhofstr. 1, 69118 Heidelberg (DE).
GEBERT, Johannes [DE/DE]; Chirurgische Universitäts-
sklinik, Im Neuenheimer Feld 110, 69120 Heidelberg
(DE). **WÖRNER, Stefan** [DE/DE]; Chirurgische Univer-
sitätsklinik, Im Neuenheimer Feld 110, 69120 Heidelberg
(DE). **LINNEBACHER, Michael** [DE/DE]; Chirurgische
Universitätsklinik, Im Neuenheimer Feld 110, 69120
Heidelberg (DE).(81) **Bestimmungsstaaten (national):** AE, AG, AL, AM, AT,
AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR,
CU, CZ, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM,
HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK,
LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,
MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL,
TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.(84) **Bestimmungsstaaten (regional):** ARIPO-Patent (GH,
GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW),
eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ,
TM), europäisches Patent (AT, BE, CH, CY, DE, DK,
ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR),
OAPI-Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML,
MR, NE, SN, TD, TG).**Veröffentlicht:**— *ohne internationalen Recherchenbericht und erneut zu
veröffentlichen nach Erhalt des Berichts*(48) **Datum der Veröffentlichung dieser berichtigten
Fassung:** 19. September 2002(15) **Informationen zur Berichtigung:**
siehe PCT Gazette Nr. 38/2002 vom 19. September 2002,
Section II*Zur Erklärung der Zweibuchstaben-Codes und der anderen
Abkürzungen wird auf die Erklärungen ("Guidance Notes on
Codes and Abbreviations") am Anfang jeder regulären Ausgabe
der PCT-Gazette verwiesen.*(54) **Title:** GENES AND THEIR GENETIC PRODUCTS PERTINENT TO MICROSATELLITE INSTABLE (MSI+) TUMOURS(54) **Bezeichnung:** FÜR MIKROSATELLITENINSTABILE (MSI+)-TUMORE RELEVANTE GENE UND IHRE GENPRO-
DUKTE(57) **Abstract:** The invention relates to genes comprising coding mononucleotide microsatellites (cMNR) or dinucleotide microsatel-
lites (cDNR). The genes can be isolated from MSI+ tumour cells. Said genes differ from corresponding genes from non-MSI+ (tu-
mour) cells by mutations in the cMNR or cDNR and code for gene products including neopeptides. The invention also relates to the
use of the genes and their gene products for the prevention, diagnosis and/or therapy of MSI+ tumours.(57) **Zusammenfassung:** Die vorliegende Erfindung betrifft Gene mit kodierenden Mononukleotid-Mikrosatelliten (cMNR) oder
Dinukleotid-Mikrosatelliten (cDNR), wobei die Gene aus MSI+ -Tumorzellen isolierbar sind und sich von den entsprechenden
Genen aus nicht-MSI+(Tumor)zellen durch Mutationen in den cMNR oder cDNR unterscheiden und für Neo-Peptide umfassende
Genprodukte kodieren. Ferner betrifft die Erfindung die Verwendung der Gene bzw. deren Genprodukte zur Prävention, Diagnose
und/Therapie von MSI+ -Tumoren.

WO 02/004664 A2

K 2987

**Genes and their genetic products pertinent to microsatellite
instable (MSI+) tumors**

The present invention relates to genes pertinent to MSI+ tumors and to their gene products. The invention also relates to a method of identifying such genes and to the use of the genes and/or their gene products for the prevention, diagnosis and/or therapy of MSI+ tumors.

Tumor cells accumulate instabilities (mutations) within genes essential for maintaining normal growth and standard differentiation. Two kinds of genetic instability were identified in human tumors: Chromosomal instability (CIN) and microsatellite instability (MSI), the latter being characterized by variations in length of repetitive DNA sequences in diploid tumor cells. There is a wide difference between CIN and MSI+ tumors as regards the type and spectrum of mutated genes, which refers to different pathways of cancerization, which are, however, not mutually exclusive. MSI occurs in about 90 % of hereditary non-polypous-colorectal tumors (HNPCC) and in about 15 % of sporadic tumors of the large intestine and further organs and is caused by mutation-induced inactivation of different DNA mismatch repair genes. MSI+ tumors have special histopathologic characteristics. MSI+ tumors are also classified using as a rule microsatellites in non-coding regions or intron sequences. However, there is information that microsatellites are also subject to instability in coding gene regions. This might be highly significant for the tumor formation of MSI+ tumors.

The present invention is thus based on the technical problem of providing a product which serves for studying MSI+ tumors on a molecular level and which should the occasion arise is suited for the diagnosis and/or therapy of MSI+ tumors.

According to the invention this technical problem is solved by the subject matters defined in the claims.

The present invention is based on Applicant's insight that genes contained in MSI+ tumor-coding mononucleotide microsatellites (cMNR) often have instabilities (mutations) in their cMNRs. To this end, he identified about 17,000 coding mononucleotide microsatellites (cMNR) and about 2,000 coding dinucleotide microsatellites (cDNR), comprising repeat units with $n \geq 6$ or $n \geq 4$, by means of computer-algorithm-assisted database analysis. The genetic instability of 15 cMNRs ($n \geq 9$) and 4 cDNRs ($n \geq 5$) and the expression of the corresponding genes were investigated in 16 MSI+ and 20 non-MSI+ tumors and cell lines, these analyses focusing on long repeat units. The cMNRs and/or cDNRs showed instability (mutation) frequencies covering from 1 - 100 % in MSI+ tumor cells; however, the cMNRs and/or cDNRs were stable in non-MSI+ (tumor) cells. Most cMNR-containing genes (10 of 15 = 66 %) were highly expressed in all of the MSI+ and non-MSI+ (tumor) cells, no significant correlation between expression level and mutation frequency being observable. In addition, he found out that the instable cMNR- and/or cDNR-bearing genes code for neopeptide-comprising gene products and that these gene products are suited for the immunization of an individual against MSI+ tumors and/or the preliminary stages thereof. Reference is made to figures 1-3, Tables 1-3 and the below examples.

The subject matter of the present invention thus relates to genes having coding mononucleotide microsatellites (cMNRs) or dinucleotide microsatellites (cDNRs), wherein the genes can be isolated from MSI+ tumor cells and differ from the corresponding genes from non-MSI+ (tumor) cells by mutations in the cMNRs or cDNRs and code for neopeptide-comprising gene products.

The expression "coding mononucleotide microsatellites" comprises repeat units of at least three equal mononucleotides A, T, G or C ($n \geq 3$), the repeat units being present in coding gene regions.

The term "coding dinucleotide microsatellites" comprises repeat units of at least three equal dinucleotides (AC, AG, AT, CA, CG, CT, GA, GC, GT, TA, TC, TG, $n \geq 3$), preferably at least six ($n \geq 6$) and more preferably at least nine ($n \geq 9$), the repeat units being located within coding gene regions.

The expression "genes having mutated cMNRs or cDNRs", which can be isolated from MSI+ tumor cells, comprises such genes in full length as well as the mutations and parts thereof which contain the sequences coding for the neopeptides.

The expression "MSI+ tumor cells" comprises any tumor cells having a microsatellite instability. Such tumor cells may be available in any form, e.g. in a cell aggregation, in particular in a tumor, or be kept in culture as such. Preferred MSI+ tumor cells comprise the cell lines LoVo, KM12, HCT116, LS174 and SW48.

The expression "non-MSI+ (tumor) cells" comprises any cells having no microsatellite instability. Such cells may be of

any kind and origin, e.g. the cells may be derived from healthy individuals or from tumors, and/or be tumor cell lines.

The terms "mutations" and "neopeptide-comprising gene products" point out that mutations are present in the coding microsatellites (cMNRs or cDNRs) of genes which can be isolated from MSI+ tumor cells as compared to the cMNRs or cDNRs of the corresponding genes from non-MSI+ (tumor) cells, the mutations being such that the genes code for neopeptide-comprising gene products. For example, the mutations are insertions and/or deletions of one or several mononucleotides and/or dinucleotides. The mutations result in reading frame shifts such that the gene products are present in the form of gene products comprising neopeptides, i.e. newly generated peptides.

In a preferred embodiment, the genes according to the invention are those differing from the genes, indicated in figure 1, of non-MSI+ (tumor) cells by mutations in the cMNRs or cDNRs and coding for neopeptide-comprising gene products. Specifically, the genes according to the invention have the mutations indicated in figure 2 in the cMNRs or cDNRs and code for the indicated neopeptide-comprising gene products.

Genes according to the invention can be identified and provided by various methods. It is favorable to use a method in which databases of non-MSI+ (tumor) cells are searched for gene sequences containing coding mononucleotide microsatellites (cMNRs) or dinucleotide microsatellites (cDNRs), these are used for detecting equal genes in MSI+ tumor cells and the latter genes are selected in such a way that they have mutations in the cMNRs or cDNRs as compared

to the gene sequences from non-MSI+ (tumor) cells and code for neopeptide-comprising gene products. In order to detect the genes in the MSI+ tumor cell it is advantageous for the DNA thereof to be subjected to a PCR reaction with primers developed from the cMNR or cDNR-comprising gene sequences. The primers preferably comprise the sequences indicated in Table 1. As to the selection of the genes from the MSI+ tumor cells it is also favorable to carry out the selection for those present in MSI+ tumor cells of different MSI+ tumors of the same type with a frequency of about 1 % - 100 %. In order to isolate the genes from the MSI+ tumor cells, it is also advantageous to use the corresponding gene sequences from the database for the preparation of appropriate primers and amplify with them the genes in the MSI+ tumor cells. Cloning of the amplified genes and the expressions thereof may then be carried out by common methods. For example, pGEMEX, pUC derivatives, pGEX-2T, pET3b and pQE-8 have to be mentioned as vectors for the expression in *E. coli*. For the expression in animal cells e.g. pKCR, pEFBOS, cDM8 and pCEV4 have to be mentioned and the baculovirus expression vector pAcSGHisNT-A is given by way of example for the expression in insect cells. The person skilled in the art is familiar with appropriate cells to express a genes present in the expression vectors. Examples of such cells comprise the *E. coli* strains HB101, DH 1, x1776, JM101, JM109, BL21 and SG13009, the yeast strain *Saccharomyces cerevisiae* and the animal cells L, NIH 3T3, FM3A, CHO, C05, Vero and HeLa as well as the insect cells sf9. The person skilled in the art is also familiar with conditions of culturing transformed and/or transfected cells and isolating and purifying the expressed gene products. Reference is made to Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd edition, Cold Spring Harbor

Laboratory Press, Cold Spring Harbor, NY (1989), by way of reference.

A further subject matter of the present invention relates to gene products which are encoded by the above genes. As to the genes according to the invention reference is made to the information given above. This information applies correspondingly to the gene products according to the invention. In particular, the gene products are those differing from the gene products of the genes indicated in figure 1 by mutations in the regions encoded by the cMNRs or cDNRs and comprising neopeptides. Especially the gene products comprise the mutations caused by cMNRs or cDNRs indicated in figure 2 and have the indicated neopeptides. Common methods can be used to provide the above gene products. Reference is made to the information given above. It may also be favorable to provide the neopeptides as such, in particular by means of peptide synthesis. Reference is made to Sambrook et al., *supra*, by way of supplement.

A further subject matter of the present invention relates to antibodies directed against the above gene products. Reference is made to the information given above as regards the gene products. This information applies correspondingly to the antibodies according to the invention. These antibodies are preferably monoclonal, polyclonal or synthetic antibodies or fragments thereof. In this connection, the term "fragment" refers to all parts of the monoclonal antibody (e.g. Fab, Fv or single chain Fv fragments) which have an epitope specificity the same as that of the complete antibody. A person skilled in the art is familiar with the production of such fragments. The antibodies according to the invention are preferably monoclonal antibodies. The antibodies according to the

invention can be prepared in accordance with standard methods, the above gene products preferably serving as an immunogen. Methods of obtaining monoclonal antibodies are known to the person skilled in the art.

A further subject matter of the present invention also relates to kits which are suited for the study of MSI+ tumors on a molecular level and for the diagnosis thereof. The kits can also be used to identify genes pertinent to MSI+ tumors. Such kits comprise one or several representatives of a gene, gene product, antibody and/or primer pair according to the invention. As regards genes, gene products and antibodies according to the invention reference is made to the information given above. The kits may also contain further substances, such as reverse transcriptase, DNA polymerase, ligase, buffer and reagents, e.g. labelings, dNTPs. In addition, the genes, gene products, antibodies and/or primer pairs according to the invention can be labeled. They may also be freely available or be immobilized by attachment to a solid carrier, e.g. a test tube, a microtration plate, a test rod, etc. The kits can also contain appropriate reagents for the detection of labelings or for labeling positive and negative controls, wash solution, dilution buffers, etc.

A further subject matter of the present invention relates to methods of immunizing an individual against MSI+ tumors and their preliminary stages, in which an individual is given an above gene in an expressible form or a gene product encoded by it. Reference is made to the above information on genes and gene products according to the invention.

For the administration of an above gene, the latter can be available as an RNA or DNA, preferably as a DNA. It may also

be present as such, *i.e.* together with elements suited for its expression, or in combination with a vector. Examples of such elements are promoters and enhancers, such as CMV, SV40, RSV, metallothionein I and polyhedrin promoter and/or CMV and SV40 enhancer. Further sequences suited for expression follow from Goeddel: Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Moreover, it is possible to use as vectors any vectors suited for expression in mammalian cells. These are e.g. pCDNA3, pMSX, pKCR, pEFBOS, cDM8 and pCEV4 as well as vectors derived from pCDNAI/amp, pCDNAI/neo, pRc/CMV, pSV2gpt, pSV2neo, pSV2-dhfr, pTk2, pRSVneo, pMSG, pSVT7, pko-neo and pHyg. Recombinant viruses, e.g. adenovirus, vaccinia virus or adeno-associated virus, can also be used as vectors.

For the administration of an above gene product, the latter may be present as such or in combination with carriers. It is favorable for the carriers not to have an immunogenic effect in an individual. Such carriers may be the individual's own or foreign proteins and/or fragments thereof. Carriers, such as serum albumin, fibrinogen or transferrin or a fragment thereof are preferred.

An individual who may be taken ill with an MSI+ tumor or is already ill therewith can be immunized using an above gene in an expressible form or a gene product encoded by it. Examples of such an individual are humans and animals as well as cells thereof. The immunization can be carried out under common conditions, the amount of gene to be administered or gene product encoded by it being easily determinable. It depends *inter alia* on whether the immunization of the individual rather focuses on an induction of antibodies directed against the gene product or

on a stimulation of cytotoxic T cells directed against the gene product, e.g. CD8⁺ T cells. Both possibilities of immunization can be achieved by the present invention. Besides, the amount depends on whether the immunization aims at a prophylactic or therapeutic treatment. Moreover, the individual's age, sex and weight as well as further clinical parameters, e.g. kidney/liver function, play a role for determining the amount. It is favorable for the individual to be given by injection 100 µg - 1 g of an above gene product or 10^6 - 10^{12} infectious particles of a recombinant virus containing an above expressible gene. The injection can be given intramuscularly, subcutaneously, intradermally or in any other form of application into several individual's sites. It may also be favorable to carry out one or several booster injections having approximately equal amount.

The present invention thus enables the detection of MSI+ tumors by means of diagnosis. These tumors can also be attacked by means of prophylaxis and therapy.

Brief description of the figures:

Figure 1:

Genes having cMNRs or cDNRs from non-MSI+ (tumor) cells

Figure 2:

Mutations in cMNRs or cDNRs from MSI+ tumors and neopeptides resulting therefrom. A comparison with corresponding genes from non-MSI+ (tumor) cells is also shown.

Figure 3:

Study of mRNA expression in large intestine cancer cell lines using RT-PCR

The following examples explain the invention.

Example 1: General method

(A) Database analyses

The EMBL database publication (EMBL Rel. 62, March 2000) was used as a basis for the search for mononucleotide and dinucleotide repeats in human coding sequences. A number of command routines and/or programs, what is called "Perl scripts" (Wall et al., Programming Perl. O'Reilly & Associates, Inc., (1996)) were written, which checked all human 109289-EMBL entries as regards the presence of coding sequences. These coding sequences were checked as regards the presence of mononucleotide and dinucleotide repeats having at least 6 bases (in the case of mononucleotide repeats) and 8 bases (in the case of dinucleotide repeats). In a database entry candidate, only the longest repeat of each nucleotide type was recorded as a mononucleotide repeat. As to the dinucleotide repeats, all of the 12 different kinds of dinucleotide repeats were considered. Entries for cDNA and genomic DNA were dealt with separately; if both DNA types were available for a gene, the genomic sequence was given priority. In addition, various filters were used. For example, entries shown to be pseudogenes were ruled out, since pseudogenes are usually not transcribed. Thus, such repeats are non-coding microsatellites. All of the identified candidate sequences were stored in a relational database for further analyses.

(B) Analysis of cMNR and cDNR candidate lists

In order to avoid the analysis of inappropriate candidate sequences, all candidates known to be pseudogenes or members of the immunoglobulin family were excluded. In addition, all candidates which had the repeats at the outermost 5' or 3'

end of the known sequence as well as all cMNRs or cDNRs which on closer analysis of the primary data could be identified as cloning of sequencing artifacts were ruled out. All cMNRs having more than 9 A and T repeats and all C or G repeats having 9 repeat units were selected, since the probability of micrisatellite instability in tumors with the mutator phenotype increases with the number of repeat units (Strauss *et al.*, *Nucleic Acids Res.* 25 (1997), 806-813). Two independent BLAST analyses were then carried out (Altschul *et al.*, *Nuc. Ac. Res.* 25 (1997), 3389-3402). It was thus tried to identify homologous and genomic sequences to be thus able to identify the exon/intron transitions of the selected cDNAs, which was a confirmation for the fact that the repeat region in the cDNA did not form on account of a splicing process. In some cases, differences between the repeat sequences of the cDNAs and those of other published sequences were obtained in the very repeat region. Hence the genomic DNA had to be sequenced in each repeat region to verify the sequence information. Exon/intron transitions were identified by MALIGN analysis (HUSAR software program package) of a candidate cDNA and a homologous genomic DNA sequence.

(C) Cell lines and tumor samples

14 human large intestine cancer cell lines were studied in each of the above genes as regards microsatellite changes. Five of the 14 large intestine cancer cell lines are classified as MSI+ (LoVo, KM12L4, HCT116, LS174T and SW48), while nine cell lines are classified as MSI-low or MSI-negative (CXF94, SW948, LS180, SW707, CaCo-2, HT29, Colo320DM, SW480 and CX-2). The cell lines SW48 and HCT 116 were obtained from ECACC [<http://www.camr.org.uk/frame.htm>]. The lines HT29, SW707, SW948, CaCo 2, CX-2, CXF94, SW480, COLO320DM, LoVo, LS174T and LS180 were obtained from the

tumor bank of Deutsches Krebsforschungszentrum. KM12L4 cells were provided by Dr. I.J. Fidler, MD Anderson Cancer Center, Houston, U.S.A. 10 MSI+ CRC tumors, an MSI+ ovarioncus (B190 TU) and two MSI-low or MSI-negative CRC tumors (B215 TU and B245 TU2) were also analyzed. The paraffin-embedded tumors were taken from the archived material of Chirurgische Universitätsklinik Heidelberg or provided by Institut für Pathologie Mannheim. Genomic DNA of the tumor samples and the corresponding mucosa samples obtained by microdissection using standard methods were provided by Ch. Sutter (Sutter et al., Mol. Cell Probes, 13 (1999), 157-165). The MSI status was determined by means of the "NCI ICG-HNPCC"-microsatellite marker panel (Boland et al., Cancer Res. 58 (1998), 5248-5257) and additionally by means of amplification of the further microsatellite markers BAT40, ACTC, D13S153, D5S107 and D5S406.

(D) Genomic MSI analyses

Primers were designed by means of the "PRIMER" program contained in the "HUSAR" program package and checked for further binding sites (sequence homologies) with respect to other human sequences by a "FASTA" analysis [HUSAR program package]. The primer positions were chosen such that they were as close as possible to the repeat region so as to obtain a short amplimer having a length of about 100 bp. This showed optimum results for an accurate fragment analysis of the DNA obtained from tissue samples embedded in paraffin. This also proved necessary for analyzing candidates having unknown genomic structure. All of the primers used are listed in below Table 1.

PCR reactions were carried out in a total volume of 25 µl (50 ng genomic DNA, 2.5 µl 10 x reaction buffer (Life Technologies, Eggenstein-Leopoldshafen, Germany), 1.5 mM

MgCl₂, 200 µM dNTPs, 0.25 µM of each primer and 0.5 U Taq DNA polymerase (Life Technologies). One primer was labeled with fluorescein at its 5' end. After an initial denaturation step at 94°C (4 min.), 35 cycles were carried out at 94°C denaturation temperature for 30", different attachment temperatures depending on the primer system at 57°-63°C for 45" and 72°C extension temperature for 30", this was followed by a final elongation step at 72° (6 min.). PCR products were analyzed on a 2 % agarose gel. The amplification products were diluted 1:2 to 1:10 prior to the fragment analysis, and 1 µl of the dilute product was mixed with 5 µl application buffer (0.6 % "blue dextran", 100 % formamide). The samples were denatured at 90° for 3 min., and then the fragments were separated by means of electrophoresis on an "ALF" DNA sequencing device (Amersham Pharmacia Biotech, Freiburg, Germany) using 6.6 % polyacrylamide/7 M urea gels. The size, height and profile of the microsatellite peaks were analyzed by means of the "AlleleLinks" Software (Amersham Pharmacia Biotech).

Table 1

Gene	PCR primers for genomic DNA		PCR primers for cDNA	
MNRs	sense 5'-3'	antisense 5'-3'	sense 5'-3'	antisense 5'-3'
FLT3LG	GGA 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	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	TGG AGT GGA TGA GGA AGA GG	CTT CTS TGG GGT CCC TGA G	TGG AGT GGA TGA GGA AGA GG	ATC TGT GCG CAC CTG CTG
SC1	CCA GGA GCA AAT TCA CAA GAC	TTT TGC GTG TTC CTT CTT TC	CCA GAA GCA AAT TCA CAA GAC	CAC CTT CTC TGT TCT CCA GTA TTC
PTHL3	TTT CAG TTT CAG TAC AGC AGT TCT G	GAA GTA ACA GCG GAC TCT TAA ATA ATG	GGA AAC TAA CAA GGT GGA GAC G	GAA GTA ACA GCG GAC TCT TAA ATA ATG
SLC23A1	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	AGT GTT GAA GAA TGG CTC CC	TGT TCT ACA TAT TAA GAC AGC CAC	GAA CAT CCG CAG AGT CCT CC	TGT TCC AGA TAT TAA GAC AGC CAC
MAC30X	TGT TGC GGA GCC CCG AC	AAC CAC CCT GTA GGC ATC TC	CCT GGT TTA AGT CCT TTC TGT TTT	AAC CAC CCT GTA GGC ATC TC
PRKDC	GAC TCA TGG ATG AAT TTA AAA TTG G	TTT GAA AAT AAC ATG TAA ATG CAG CTC	CAG CCC TGG AGC TTC TTA TTA A	GAC AAC CCC TTC AGA CAT CC
ATR	TCT TCT GTA GGA ACT TCA AAG CC	TGA AAG CAA GTT TTA CTG GAC TAG G	AGC TCC CAG GAA GTA ATC CG	TCA AAG CAA GTT TTA CTG GAC TAG G
MBD4	TGA CCA GTC AAG AAA ACA GGC	GTT TAT GAT GGC AGA AGT TTT TTG	TGA CCA GTG AAG AAA ACA GGC	GTG GTG GCG GGC TAA GAG
SEC63	AGT AAA GGA CCC AAG AAA ACT GC	TGC TTT TGT TTC TGT TCC TTT G	TGA AAA GGA CCA GTC CAT CTG	TGC TTT TGT TTC TGT TCC TTT G
OGT	TGA CTT TTG GGT GAG AG	GCG AGG GAA AGG AGG TAA AG	TGA CTT TTG GGT GAG AG	TGT CAA AAA TGC GTG CCT C
HPDMPK	GCT TGA TCC TGT TGA TTT TCT ACT C	CTG AAT GGA GAA GAA AAT GAG ATG	TCC TAC TGG ATG TGC TGC C	CTG AAT GGA GAA GAA AAT GAG ATG
U79260	TTT GTT AAT TCC CAT TAG CTC CC	AAC CTG GTG ACA GAG TGA GAC	CAT TAA GCA AAG CAG CCA GG	AAC CTG GTG ACA GAG TGA GAC
DNRs	sense 5'-3'	antisense 5'-3'	sense 5'-3'	antisense 5'-3'
KIAA0040	CAT CTC AAT ATG GTT CCC AAG TG	CTT GCG CAC GAA CTT GCT AC	CAA GAA GTA AGT TGG AAG GAG G	GTG CAT TAT TTC AAG GGT TCC

(E) Confirmation of the sequence

All of the coding microsatellites were confirmed by "thermocycle" sequencing. PCR reactions were carried out as described above. PCR products were purified by means of the "QIAquick" PR purification kit (Qiagen, Hilden, Germany) and sequenced with the corresponding primers using the "Big Dye terminator cycle sequencing kit (Perkin Elmer, Darmstadt, Germany).

(F) Expression analyses and cDNA-MSI analyses

Poly(A)RNA of 14 large intestine cancer cell lines were extracted by means of the oligo(dT) cellulose method (Vennstrom and Bishop, Cell 28 (1982), 135-143). The quality of the RNA preparation and the reverse transcription were checked by means of GAPDH amplification (Hsu et al., Int. J. Cancer, 55: 397-401, 1993). Primer pairs permitting a differentiation according to size between cDNA and genomic DNA amplimers possibly contained as a contamination, were considered appropriate for the expression analysis by semi-quantitative RT-PCR. In the case of an unknown exon structure, primers were designed which were localized on the cDNA, and it was checked whether genomic PCR yielded either an amplification product the same as that of the RT-PCR, a longer one or none at all. All of the primers used are listed in Table 1. 100 ng poly(A+)RNA were subjected to reverse transcription by means of 0.5 µg oligo(dT)₁₂₋₁₈ in a final volume of 20 µl with 200 units M-MLV Reverse Transcriptase (SuperScript, Life Technologies) at 37°C for 1 hour. In order to check the RNA integrity and the synthesis of the first cDNA strand, control PCR reactions were carried out by means of GAPDH-specific primers (Hsu et al., Int. J. Cancer 55 (1993), 397-401). PCR reactions were carried out in a total volume of 50 µl (1 µl cDNA, 5 µl 10 x reaction buffer (Life Technologies), 1.5 mM MgCl₂, 200 µM dNTPs, 0.25

μM of each primer and 0.5 units Taq DNA polymerase (Life Technologies) by means of the above amplification protocol described above for the amplification of genomic DNA. The PCR products were separated on 2 % agarose gels and made visible on ethidium bromide staining.

PCR reactions for cDNA-MSI analyses were carried out as described for the expression analyses, except that a primer labeled with fluorescein at its 5' end was used. The fragment analysis was carried out as described for the genomic analysis.

Example 2: Computer-assisted analyses

The computer database sequence analysis resulted in 365 candidates for mononucleotide repeats having a minimum length of 9 bases (total: 17654 mononucleotide repeats having a length of ≥ 6 bases). In addition, 2028 dinucleotide repeats having a minimum length of 8 bases were found. The longest mononucleotide region comprised 32 bases and the longest dinucleotide region comprised 42 bases, i.e. 21 repeat units.

Example 3: Identification of cMNR

All cMNR having 10 or more repeat units and all C or G repeat regions having 9 or more repeat units were analyzed for this purpose, since it was assumed that the mutation rate was increased in the relatively long repeat regions. Moreover, all candidate sequences which satisfied further exclusion criteria were ruled out from an analysis. A total of 43 cMNR candidate sequences were thus obtained, which comprised 12 duplicates so that 31 different candidate sequences were selected.

These candidate sequences had to be verified experimentally as microsatellites in coding regions. Therefore, primers flanking the repeat regions had to be designed for each candidate sequence. The full information about the genomic structure could be obtained by means of sequence comparison of the cDNAs with databases for the genomic sequence. The systematic sequence comparison supplied information about exon/intron transitions and coding regions for 9 of the 31 candidate sequences.

Therefore, primer pairs on the basis of the cDNA sequence had to be designed for the other 22 cDNA sequences. The amplification of the repeat regions in both the cDNA and the corresponding genomic DNA resulted in identical PCR products in further 9 of the 22 candidates, which shows that these 9 sequences contain the mononucleotide region in a coding region. The PCR reaction of the genomic DNA sequences was negative in 13 mononucleotide regions or resulted in amplimers longer than those obtained by means of the amplification of the corresponding cDNA. Another analysis of the genomic structure of the corresponding gene was thus required for each of these candidate sequences. A total of 18 mononucleotide regions were subjected to a sequence analysis.

It was possible to confirm repeat regions by sequence analyses in 17 of the 18 cases. Only one candidate did not show the expected A_{14} repeat. With two candidate sequences, the repeat region was located within the predicted coding sequence of the genomic DNA sequences, yet it was not possible to show an expression by means of RT-PCR analyses. Moreover, no information could be identified as regards ESTs or a partial coding sequence homologous to the repeat region. These studies were carried out by means of the "EST

Clustering software" (Husar program package). Thus, the sequence analysis together with the confirming experiments formed the basis for the identification of 15 cMNRs (cf. figure 1).

Example 4: Detection of cDNR

The same strategies as used for the sequence analysis and the experimental confirmation of the sequence data were applied to the cDNRs, the longest cDNR candidates having been started with. Having identified cDNA sequences of 4 cDNR candidates, the MSI status in MSI+ an MSI large intestine cancer cell lines and MSI+ tumors was analyzed. One of these candidates ((AC)₇) showed a mutation in an MSI+ tumor, but no mutation in the non-MSI+ (tumor) cells (cf. figure 2). It is assumed that the mutation rate is increased in the case of cDNRs having a repeat number of above 9.

Example 5: Studies as to the mutation rates of the cMNRs

In particular, varying mutation rates ranging from 40 to 80 % were observed in the three C₉-cMNRs and rates ranging from 10 to 100 % were detected in the A₁₀-cMNRs, whereas the T₁₀ and N_{≥11}-cMNRs showed constantly higher mutation rates between 75 and 100 %. In three cMNR markers ((SYCP1 (A10), ATR (A10), and MBD4 (A10)) only minor mutation frequencies could be identified in MSI+ cell lines and tumor samples. In MSI+ cell lines and MSI+ tumors, however, high mutation rates were detected for the two cMNR markers (HPDMPK (T14) and U79260 (T14)): All of the 5 MSI+ cell lines and 10 of 11 MSI+ tumors showed a change in sequence as regards HPDMPK. Analogous results were found for the mono-cMNRs in the U79 260 gene which was mutated in all of the 5 MSI+ cell lines and in 9 of 11 MSI+ tumors (cf. below Table 2).

Table 2: Frequency of mutations in cMNR in MSI+ tumor cell lines and MSI+ tumors

Gen	Repeat	LoVo	KM12	HCT116	LS174T	SW48	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T
FLT3LG	C ₉	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
SYCP1	A ₁₁	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
SLC4A3	C ₉	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
aCl	T ₁₀	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
PTHL3	A ₁₁	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
SLC23A1	C ₉	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
GART	A ₁₀	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
MAC30X	A ₁₀	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
PRKDC	A ₁₀	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
ATR	A ₁₀	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
MBD4	A ₁₀	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
SEC63	A ₁₀	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
OGT	T ₁₀	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
HPDMPK	T ₁₄	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
U79260	T ₁₄	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•

○ Mutation non available • mutation available

Example 6: Expression analyses of cMNRs

The expression levels of the above 15 cMNR-containing genes differed widely and varied between non-detectable expression and constantly strong transcription activity in all of the 14 tested large intestine cancer cell lines. The SYCP1 gene involved in meiosis and the gene coding for the hematopoietic growth factor FLT3LG were not expressed in large intestine cancer cell lines. The gene HPDMPK located downstream of the gene locus for two genes associated with myotonic dystrophy (*Dystrophia myotonica*) and coding for a hypothetical protein and the gene coding for the ER membrane protein SEC63 were not expressed very highly, yet were

expressed constantly, in all of the cell lines. The aC1-mRNA and splice variant 3 of the PTHrP gene (PTHl3) were expressed in large intestine cancer cell lines to a different extent. Both genes were expressed in about 50 % of the investigated cell lines. The GART gene coding for the trifunctional ribonucleotide synthetase, the PRKDC gene coding for the DNA-dependent protein kinase and the ATR gene connected with the cell cycle were highly expressed in large intestine cell lines. MAC30X is also highly expressed in large intestine cancer cell lines (cf. figure 3). In summary, it can be pointed out that the expression levels of the corresponding genes do not correlate with the MSI status of the affected cell lines.

Example 7: MSI analysis of cDNAs

A fragment analysis of amplified cDNAs of the above cMNRs of 14 large intestine cancer cells lines was carried out: Three cMNRs showed the wild-type cDNA in affected cell lines (GART) or in most affected cell lines (SEC63). In seven cMNRs there was correspondence between the genomic and transcribed sequences (MAC30X, HPDMPK, U79260, MBD4 and ATR).

Example 8: Stimulation of CD8⁺ T cells against an above gene product and lysis of MSI+ tumor cells expressing this gene product

(a) Stimulation of CD8⁺ T cells against a neopeptide according to the invention

Peripheral blood lymphocytes (PBL) from an HLA-A0201-positive healthy proband were purified by density centrifugation on a Ficoll Paque® gradient. T lymphocytes were obtained by separating the B lymphocytes and/or the

monocytes using antibody-linked magnetobeads (CD11, CD16, CD19, CD36, and CD56) (Pan T-cell isolation Kit®, Milteny, Bergisch Gladbach, Germany). About 2×10^7 T cells were obtained from 30 ml blood. Of these about 2×10^6 T cells were stimulated with autologous cells B cells activated on CD40 (about 5×10^5) which had been loaded with one of the HLA-A0201-restricted neopeptides of below Table 3 (cf. also figure 2), i.e. cocultured in 24 well plates. This stimulation was repeated weakly for a period of five to six weeks.

Table 3: Examples of HLA-A0201-restricted neopeptides encoded by mutated cMNR

	Gene	Peptide length	AS sequence of the identified neopeptide
	PDMPK	9mer	FLSASHFLL
	OGT	10mer	SLYKFSPFPL
	U79260	9mer	TLSPGWSAV

By means of known IFN-gamma ELISpotanalysis, the reactivity over the neopeptides was determined weakly, starting on day 0. On day 28, a reactivity of 1760 specific cells / 1,000,000 cells against peptide #16 (SLYKFSPFPL) was observed, on day 35 there was found one of 1123 specific cells / 1,000,000 cells against peptide #15 (FLSASHFLL) and one of 733 specific cells / 1,000,000 cells against peptide #21 (TLSPGWSAV). The strength of the reaction was thus within ranges usually only achieved by means of viral antigens, the value for the GILGFVFTL peptide, which was derived from a matrix protein of the influenzavirus, was 1170 specific cells / 1,000,000 on day 35. Hence it becomes obvious that it is possible to stimulate activated CD8⁺ T cells against the neopeptides according to the invention.

(b) Lysis of cells loaded with neopeptides according to the invention

After another restimulation, the cytotoxic potential of the activated CD8⁺ T cells was tested for the neopeptide-loaded HLA-A2.1⁺ colon carcinoma cell lines SW480 and HCT 116 as well as T2 cells. Unloaded cells served as a control. 1x10⁶ cells each were labeled radioactively using ⁵¹Cr (100 µCi) at 37°C for 1 h and cocultured with increasing amounts of activated CD8⁺ T cells for 4 h. The specific lysis of the respective cell line was determined by measuring the released radioactivity in the supernatant. It turned out that the HLA-A0201-expressing cell lines can be lysed when they are loaded with neopeptides, unloaded cells are not lysed.

Furthermore, competition experiments were carried out. It was possible to compete with the release of radioactivity and thus the specific cytotoxic activity of the T cells by the addition of an excess (50 "cold" to 1 "hot" neopeptide-loaded cell) of neopeptide-loaded T2 cells which are, however, not labeled radioactively to a reaction batch with radioactively labeled neopeptide-loaded T2 cells and activated CD8⁺ T cells. Thus, it becomes evident that the CD8⁺ T cells directed against the neopeptides specifically detect and lyse the neopeptide-expressing tumor cells.

Claims

1. Genes having coding mononucleotide microsatellites (cMNR) or dinucleotide microsatellites (cDNR), wherein the genes can be isolated from MSI+ tumor cells and differ from the corresponding genes from non-MSI+ (tumor) cells by mutations in the cMNRs or cDNRs and code for gene products which comprise neopeptides.
2. Genes according to claim 1, wherein the corresponding genes from non-MSI+ (tumor) cells are those of figure 1.
3. Genes according to claim 1 or 2, wherein the genes which can be isolated from MSI+ tumor cells have the mutations indicated in figure 2.
4. Gene products, encoded by the genes which can be isolated from MSI+ tumor cells according to any of claims 1 to 3.
5. Antibody directed against the gene products according to claim 4.
6. A method of identifying genes which can be isolated from MSI+ tumor cells with cMNRs or cDNRs according to any of claims 1 to 3, in which databases of non-MSI (tumor) cells are searched for cMNR- or cDNR-containing gene sequences, the latter are used in MSI+ tumor cells for detecting equal genes and the latter genes are selected in such a manner that they have mutations in the cMNRs or cDNRs as compared to the gene sequences from the non-MSI+ (tumor) cells and code for gene products which comprise neopeptides.

7. The method according to claim 6, wherein for detecting the same genes in MSI+ tumor cells a PCR reaction is carried out with primers developed from the cMNR or cDNR-comprising gene sequences.

8. The method according to claim 6 or 7, wherein the primers are those of Table 1.

9. The method according to any of claims 6 to 8, wherein selection is made for those genes from the MSI+ tumor cells, which are present in MSI+ tumor cells of various MSI+ tumors of equal type with a frequency of 1 % - 100 %.

10. Kit comprising one or several representatives of a gene according to any of claims 1 to 3, of a gene product according to claim 4, of an antibody according to claim 5 and/or a primer pair of Table 1.

11. Use of the gene according to any of claims 1 to 3, of the gene products according to claim 4, of the antibodies according to claim 5 or the kit according to claim 10 for the molecular investigation of MSI+ tumors and their preliminary stages as well as the diagnosis thereof.

12. Use of the genes in expressible form according to any of claims 1 to 3 or the gene products according to claim 4 for immunizing an individual against MSI+ tumors and their preliminary stages.

13. Use according to claim 12, wherein the immunization is made on the basis of a prophylactic and/or therapeutic treatment of MSI+ tumors.

Figure 1

Gene	Acc. no.	Chr.
MNRs		
FLT3LG	U29874	19q13.3
SYCP1	X95654	1p13-12
SLC4A3	U05596	2q36
aC1	D82070	4
PTHL3	M24350	12p12-11
SLC23A1	AF058319	20p13
GART	X54199	21q22.1
MAC30X	L19183	17
PRKDC	U63630	8Q11
ATR	U76308	3q22-24
MBD4	AF072250	3q21-24
SEC63	AF100141	6q16-22
OGT	U77413	X?
HPDMPK	Y10936	19q13.3
U79260	U79260	16
DNRs		
KIAA0040	D25539	1q24-25

Designation		description Nucleotide/peptide sequence	Repeat
<u>CMNRs:</u>			
SLC4A3	wt	...CTGAGGCAGAACCTGTGGAGCCCCCCCCCTCAGGGACCCACAGAAAGGC...	C₉
	+1	...CTGAGGCAGAACCTGTGGAGCCCCCCCCCTCAGGGACCCACAGAAAGGC...	
	neo	...EAEPVEPPPLRDPTEGKVLHWK	
FTL3LG	wt	...TCTCCCTCCCTGTCTCCAGCCCCCCCCCAGCTGTCTTCGCTTCGTCCA...	C₉
	-1	...TCTCCCTCCCTGTCTCCAGCCCCCCCCCAGCTGTCTTCGCTTCGTCCA...	
	neo	...VTKCAFQPPPAVFASSRPTSPASCRPPSSWWR	
SLC23A1	wt	...CACGGCTGTCTGTGCCCAACCCCCCCCCCATCCACGCAATAAACAGGGG...	C₉
	-1	...CACGGCTGTCTGTGCCCAACCCCCCCCCCATCCACGCAATAAACAGGGG...	
	neo	...ARLSCAPPPPSIQ	
	+1	...CACGGCTGTCTGTGCCCAACCCCCCCCCCATCCACGCAATAAACAGGGG...	
	neo	...ARLSCAPPPPHPRNKQGNFRGRPLLCSI	
GART	wt	...GACAAATCATTTCTCTTTTGAAAAAAAGGCCAGAGTGGCTGTCTTA...	A₁₀
	-1	...GACAAATCATTTCTCTTTTGAAAAAAAGGCCAGAGTGGCTGTCTTA...	
	neo	...LTNHFSEKRPPEWLS	

MAC30X

wt	...TACAAGTATGAAGAGAAAGAA ^{A10} AAAAAATGAAGGAACAACCACTGG...
-1	...TACAAGTATGAAGAGAAAGAA ^{A10} AAAAAATGAAGGAACAACCACTGG...
neo	...YKYEEKRKKNEGNNHWPRVEMPTGWLLVGYIOEHCSEPTSSAA ^{A10} FETLAA...

...MHKSKMVS^{A10}GTMSNP^{A10}HL^{A10}LP^{A10}FT^{A10}FF

PRKDC

wt	...TCTATGGAGAACTTGCATTGAA ^{A10} AAAAAATACCAGATACAGGTGAGAT...
-1	...TCTATGGAGAACTTGCATTGAA ^{A10} AAAAAATACCAGATACAGGTGAGAT...
neo	...FYGELALKKKYQIQF

ATR

wt	...AGCCATTCCTTTTCCTACTGAA ^{A10} AAAAAATACCTAGTCCAGTAAACT...
-1	...AGCCATTCCTTTTCCTACTGAA ^{A10} AAAAAATACCTAGTCCAGTAAACT...
neo	...KPFLFLKKKYL ^{A10} VQ

SYCP1

wt	...GTAAATTGCTAAATGGATAGAA ^{A10} AAAAAATAAAGAGCTGAAAAGT...
-1	...GTAAATTGCTAAATGGATAGAA ^{A10} AAAAAATAAAGAGCTGAAAAGT...
neo	...AVIAKMDRKK ^{A10} N
+1	...GTAAATTGCTAAATGGATAGAA ^{A10} AAAAAATAAAGAGCTGAAAAGT...
neo	...AVIAKMDRKK ^{A10} KTKRS

MBD4

wt	...AGTGAAGAAACAGCCTTGTAA ^{A10} AAAAAAGAAAGATCATTCAGTTCAG...
-1	...AGTGAAGAAACAGCCTTGTAA ^{A10} AAAAAAGAAAGATCATTCAGTTCAG...
neo	...SVTSEENSLVKKKK ^{A10} KDH

SEC63			A ₁₀
	wt	...	TCAAAAAAAAAAACCTTTAAATAAACCTACCTGTGCTATTAC...
	-1	...	TCAAAAAAAAAAACCTTTAAATAAACCTACCTGTGCTATTAC...
	neo	...	<u>KSKKKPLKKNLHLCYHQSNNRNRKSRMESLGMKLQ</u>
aC1			T ₁₀
	wt	...	AAAGAACCAATTTTACAAATTTTTCCTGTGTCAGTTGAATTTGGG...
	-1	...	AAAGAACCAATTTTACAAATTTTTCCTGTGTCAGTTGAATTTGGG...
	neo	...	<u>SKNQFTIFFSCQLNLGRKEHAKITFFQOLDTMDGNPGELELTLEQTLQKOSONALLPAGPLTOTPY</u>
OGT			T ₁₀
	wt	...	TTCCCCCCTCCCCCAATCTTTTTCCTTCCCTTACAAATTTCCCC...
	-1	...	TTCCCCCCTCCCCCAATCTTTTTCCTTCCCTTACAAATTTCCCC...
	neo	...	<u>SSPPPPNLFESLYKFSFPLPPEPPIFFH</u>
PTH3			A ₁₁
	wt	...	CAGCACTTCTGTGGGTTTGAAAAAAGGAAACACAGAAAC...
	-1	...	CAGCACTTCTGTGGGTTTGAAAAAAGGAAACACAGAAAC...
	neo	...	<u>TALLWGLKKRKITEHHICN</u>
HPDMPK			T ₁₄
	wt	...	TGGACCTCTGAGAGGGTGTTTTATCAGCATCTCACTT...
	-1/-4	...	TGGACCTCTGAGAGGGTGTTTTATCAGCATCTCACTT...
	neo	...	<u>AVDPLRRVFF₇ISISLSSPSPNPLPAMLSTPRTHOOGADGS</u>
	-2/-5	...	TGGACCTCTGAGAGGGTGTTTTATCAGCATCTCACTT...
	neo	...	<u>AVDPLRRVFF₇LSASHELLHSAPISLPFPPOGPTSREQTAANFGIT</u>
	-3	...	TGGACCTCTGAGAGGGTGTTTTATCAGCATCTCACTT...
	shortened	...	AVDPLRRVFFYYQHLTFFSIQPPPPCHAFHPKDPAGSRRQLLVPLKGPPLAPILS...

097610

III

wt -1 neo -2 neo

114

...GCATTTCTTTTCTTTCTTTCTTTTCTTTTCTTTGAGACACAGTCTCAC...
...GCATTTCTTTTCTTTCTTTCTTTTCTTTTCTTTGAGACACAGTCTCAC...
...HFLFSELFEEFLRHSLTLSPGWSAVARSRLTATSASOVVILPQPEWLGLQARAAPS
...GCATTTCTTTTCTTTCTTTCTTTTCTTTTCTTTTCTTTGAGACACAGTCTCAC...
...HFLFSELFEEF

KIAA0040

(ACF)

wt -1 neo

GTGTGTGCACATGCACGTAACACACACACACACAAATTCAGGTAGCAG...
GTGTGTGCACATGCACGTAACACACACACACACAAATTCAGGTAGCAG...
VHMHVTHHTQIQYAGTWA SIFCSSNGHWLCTLCREVHYLQSQKCLMGK...
PCQIQTHIYNFLTSKAPIHHLFHKPLRLQMHAFLLTLHINFYWKYS

Genes		Repeat	Locus
ELT3LG	Ligand of Fms-related tyrosine-kinase 3	C ₉	LoVo
SYCP1	Synaptonemal complex protein 1	A ₁₀	KM12
SLCAA3	Anion exchanger 3 brain isoform	C ₉	HCT116
ACI	Differentially expressed in neuroblastomas	T ₁₀	LS174T
PTHL3	Parathormone-similar protein, splice variant 3	A ₁₁	SW48
SLC23A1	Dissolved transport molecule, family 23, member 1	C ₉	CX-2
GART	Trifunctional ribonucleotide-transferase/-synthetase	A ₁₀	CXJ94
MAC30X	Differentially expressed in meningiomas	A ₁₀	SW948
PRKDC	Subunit of the DNA-dependent protein kinase	A ₁₀	LS180
ATR	Ataxia telangiectasia and Rad3-related	A ₁₀	CaCo 2
MBD4	Methyl-CpG binding domains protein 4	A ₁₀	HT29
SBC63	Endoplasmic reticulum, membrane protein	A ₁₀	Colo320
OGT	O-glycosidic N-acetylglucosamine transferase	T ₁₀	SW480
HPDMPK	Hypoth. protein 3 of the DMPK/DMAHP gene locus	T ₁₄	
U79260	Clone 23745-mRNA	T ₁₄	
KIAA0040	Human mRNA of the KIAA0040 gene	(AC) ₇	
GAPDH			