[CANCER RESEARCH 62, 3641–3645, July 1, 2002]

Advances in Brief

Instabilotyping Reveals Unique Mutational Spectra in Microsatellite-Unstable Gastric Cancers[1]

Yuriko Mori, Fumiaki Sato, Florin M. Selaru, Andreea Olaru, Kellie Perry, Martha C. Kimos, Gen Tamura, Nagahide Matsubara, Suna Wang, Yan Xu, Jing Yin, Tong-Tong Zou, Barbara Leggett, Joanne Young, Toshihiro Nukiwa, O. Colin Stine, John M. Abraham, David Shibata, and Stephen J. Meltzer[2]

Departments of Medicine [Y. M., F. S., F. M. S., A. O., M. C. K., S. W., Y. X., J. Y., T-T. Z., J. M. A., S. J. M.] and Surgery [K. P., D. S.], Division of Gastroenterology and Greenebaum Cancer Center, University of Maryland School of Medicine, Baltimore Veterans Affairs Hospital, Baltimore, Maryland 21201; Department of Epidemiology, University of Maryland School of Medicine, Baltimore, Maryland 21201 [O. C. S.]; Department of Pathology, Yamagata University School of Medicine, Yamagata 990-9585, Japan [G. T.]; Department of Gastroenterological Surgery, School of Medicine, Okayama University, Okayama 700-8558, Japan [N. M.]; Conjoint Gastroenterology Lab, Royal Brisbane Hospital Foundation, Clinical Research Centre, Bancroft Centre, Herston, Queensland 4029, Australia [B. L., J. Y.]; and Respiratory Oncology and Molecular Medicine, Institute of Development, Aging and Cancer, Tohoku University, Sendai 980-8575, Japan [T. N.]

Abstract

Microsatellite instability (MSI) within coding regions causes frameshift mutations (FSMs). This type of mutation may inactivate tumor suppressor genes in cancers with frequent MSI (MSI-H cancers). To identify novel FSMs in gastric carcinogenesis in an unbiased and comprehensive manner, we screened for this type of mutation at 154 coding region repeat loci in 18 MSI-H gastric cancers. We also compared FSM rates and spectra in MSI-H gastric *versus* colorectal cancers. Thirteen novel loci showed FSMs in >20% of gastric tumors. Novel loci with the highest mutation frequencies included the *activin type 2 receptor* gene (44.4%), *DKFZp564K112* (a homologue of the *Drosophila* tumor suppressor gene *multi-sex-combs*; 41.2%), and an endoplasmic reticulum chaperone protein gene *SEC63* (37.5%). The mutational spectra for genes with high mutation frequencies were also significantly different between MSI-H gastric and colorectal cancers.

Introduction

MSs[3] are repetitive DNA sequences consisting of oligonucleotide units, which are distributed widely throughout the human genome. Length mutations in MSs are common among cancers with deficient DNA mismatch repair, such as hereditary nonpolyposis colorectal cancer-associated malignancies and sporadic gastric, colorectal, and endometrial cancers. MSI within coding regions causes FSMs, which result in gene inactivation. FSMs have been reported at several coding region MS loci, including loci within tumor-related genes in cancers with frequent MSI (MSI-H cancers). Known FSMs have been described in *TGFβ type II receptor (TGFBR2)*, *IGF2 receptor (IGF2R)*, *BCL2-associated X protein (BAX)*, *hMSH3*, and *hMSH6* (1–4). However, FSMs are in general rare in MSI-H tumors when coding region MSs are studied without regard to their potential relationship to cancer. These findings suggest that FSMs that occur frequently in MSI-H tumors are the result of clonal selection during tumor development or progression. On the basis of this hypothesis, we previously performed a systematic genome-wide search to discover coding region MSs and FSMs to identify candidate tumor-related genes inactivated by FSMs in MSI-H colorectal cancers (5). We identified several frequently mutated genes, including the *activin type II receptor (ACTRII)* gene (5). In this report, we describe the results of mutational screening of 154 coding MSs in 18 MSI-H primary gastric cancers and compare the resulting FSM spectrum to the one that we had observed in MSI-H colorectal cancers. In this fashion, we identify several newly reported FSMs in candidate tumor-related genes, which provide clues to possible new carcinogenetic pathways and their uniqueness to different organs of origin.

Materials and Methods

Database Analyses and Selection of Coding Region MSs. Coding MSs consisting of homopolymeric tracts of eight nucleotides or longer were identified with a computer script as described in our previous study (5). The input of the script was Unigene Hs.seq.uniq.Z, a database containing only the clone with the longest region of high-quality sequence data among each gene cluster provided by the National Center for Biotechnology Information.[4] We used this database as the input to decrease the likelihood of finding the same MS more than once. As a result, 300 homopolymers of eight or more nucleotides were identified and qualified for the study after manual inspection. PCR primer sets were designed to amplify each of these homopolymers. For this study, we used 154 loci at which PCR amplification using a standard human genomic DNA template was successful. All primer sequences are available on request.

Patients and Sample DNA Preparation. Eighteen MSI-H cancers were identified from a larger group of 126 gastric cancers using five consensus loci (*BAT25, BAT26, D2S123, D5S346,* and *D17S250*) according to criteria from a National Cancer Institute Workshop in 1998 (6). Tumors were classified as MSI-H when MSI was observed at two or more of the five loci. Genomic DNA was extracted from paired normal and cancerous gastric tissues that had been frozen in liquid nitrogen after surgical resection, as described previously (7).

MS Analyses. FSM at each locus was determined by analyses of the length of each PCR-amplified microsatellite. One primer of each pair was labeled with a fluorescent dye, *i.e.,* Hex, Fam, or Tet. PCR reactions were performed in a total volume of 10 μl containing 20 ng of genomic DNA, 0.1 μM of each primer, 1× Taq DNA polymerase buffer (Life Technologies, Inc., Gaithersburg, MD), 0.4 mM of each deoxynucleotide phosphate, 1.5 mM $MgCl_2$, and 0.5 IU of Taq DNA polymerase (Life Technologies, Inc.). Conditions were as follows: an initial denaturation step at 94°C for 4 min, followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 30 s. After these steps, a final extension was performed at 72°C for 4 min. Products were analyzed on an automated DNA sequencer (ABI 377 or 3700; PE Biosystems, Foster City, CA) using the software programs GeneScan and Genotyper (PE Biosystems). We classified a tumor-specific alteration as an FSM only when it caused a change of >50% in peak area in the tumor sample compared with the corresponding normal sample. Biallelic alteration was defined as either two different mutations or a single mutation with absence of the normal allele.

Received 2/11/02; accepted 5/15/02.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] Supported by NIH Grants CA 95323, CA85069, CA77057, CA78843, and DK47717 and by the Medical Research Office, Department of Veterans Affairs. Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org).

[2] To whom requests for reprints should be addressed, at University of Maryland, School of Medicine, Bressler 8-009, 655 West Baltimore Street, Baltimore, MD 21201. E-mail: smeltzer@medicine.umaryland.edu.

[3] The abbreviations used are: MS, microsatellite; MSI, MS instability; FSM, frameshift mutation; ER, endoplasmic reticulum.

[4] Internet addresses: http://www.ncbi.nlm.nih.gov/UniGene/ and ftp://ncbi.nlm.nih.gov/repository/UniGene/.

Statistical Analyses of Mutational Spectra. The χ^2 test for goodness of fit (χ^2) was performed to test the null hypothesis, *i.e.,* that there was no difference in mutation frequency between gastric and colorectal MSI-H cancers. χ^2 value was calculated as follows:

$$\chi^2 = \sum_{i=1}^{k} (Oi - Ei)^2/Ei$$

where Ei is (number of informative samples) × (mean of mutation rates for gastric and colorectal cancers). Whereas Oi represented the observed number of mutated samples, Ei is the expected number of mutated samples. k is the number of genes analyzed. The degree of freedom for this χ^2 value was $(k - 2)$.

Results and Discussion

Instabilotyping of MSI-H Gastric Cancers. To perform comprehensive screening of FSMs in coding region mononucleotide repeat MSs, 154 coding homopolymers consisting of eight or more nucleotides were examined in 18 MSI-H gastric cancers. These 154 loci were selected by using an automated coding region mononucleotide repeat database search as described in "Materials and Methods." It should be emphasized that in order to identify potential involvement of less well-understood genes in carcinogenesis, we included many loci lacking known links to proliferation, differentiation, or cell death. Moreover, as positive controls, genes reported previously to undergo FSM were also included among the 154 loci, *i.e., TGFBR2*, *IGF2R*, *BAX*, *hMS3*, *hMSH6*, *BRCA1*, *BRCA2*, the *retinoblastoma protein-interacting zinc finger protein gene* (*RIZ*), and *methyl-CpG binding protein 4* (*MBD4*; Refs. 1–4, 8–10). Fig. 1 shows examples of electropherograms for FSM-positive and FSM-negative samples.

Among 154 coding region loci tested, 19 loci showed frequent mutation (*i.e.,* in >20% of tumors). Loci mutated in >30% of tumors were *TGFBR2* (11 of 18; 61.1%), *ACTRII* (8 of 18; 44.4%), *DKFZp564K112* (7 of 17; 41.2%), *SEC63* (6 of 16; 37.5%), *MBD4* (6 of 18; 33.3%), *BAX* (4 of 12; 33.3%), and *MAC30* (3 of 9; 33.3%). Among these loci, *ACTRII, SEC63, DKFZp564K112*, and *MAC30* have not been reported previously by other groups as showing mutations or other alterations in human cancer. A BLAST search revealed that *DKFZp564K112* had homology at the amino acid level with the *Drosophila* tumor suppressor gene, *multi-sex-combs (msx).* In addition, loci mutated in >20% of tumors were *proline 4-hydroxylase β polypeptide (P4HB, thyroid hormone binding protein p55, protein disulfide isomerase)*, *TTK*, *KIAA0896 (a human homologue of Drosophila hyperplastic discs, hHYD)*, *RIZ*, *absent in melanoma 2* (*AIM2*), *hMSH3*, *interleukin-1β converting enzyme (ICE/caspase-1)*, *KIAA0905 (SEC31)*, *monocarboxylate transporter member 4 (MCT4)*, *hnRNP H*, *calnexin*, and *hKF-1*. Two separate loci were analyzed for the *tetratricopeptide containing Down syndrome gene 1* (*TPRD1*), and four of 18 samples (22.2%) exhibited FSMs in one of these loci. Forty-two loci with FSM rates of 10% or higher in MSI-H tumors are listed in Table 1.

The process of coding region mutation appeared to be selective, in that high mutation rates were observed in only a small percentage of loci. Mutation frequencies varied widely across the complete set of 154 loci (*i.e.,* from 0 to 61.1%; mean, 7.7%; geometric mean, 3.0%), but the majority of loci did not undergo mutation at all. Sixty-nine loci were not mutated in any of the tumors, and 43 loci were mutated in <10% of tumors. Moreover, many genes showing high FSM rates had known or potential links to carcinogenesis. Longer repeats tended to mutate more frequently, as reported previously (data not shown; Refs. 5, 11). Eight loci were located in close proximity to the 3′ ends of coding regions (*i.e.,* within 50 bases upstream of the last coding region nucleotide). Four of these eight loci (*MAC30, TTK, AIM2,* and *MCT4*) were mutated in >20% of gastric tumors. Mutations near the 3′ ends of genes do not usually result in major truncations of their protein products. However, mutations near the 3′ ends of genes may still cause dysregulation or dysfunction of protein products. Therefore, we did not eliminate them from our analyses.

We also analyzed biallelic alterations that should have resulted in gene inactivation, suggesting involvement of these mutations in carcinogenesis. Biallelic alteration was inferred when a coding region locus exhibited either two different mutations in the same MS or absence of the normal allele. Four of 11 tumors with FSM in *TGFBR2* contained biallelic alterations. Similarly, two biallelic alterations were observed among eight tumors with *ACTRII* mutation, five tumors with *TTK* mutation, and five tumors with *AIM2* mutation, respectively. Other novel loci with biallelic alterations were *DKFZp564K112* (an *msx* homologue), *SEC63*, KIAA0896 (*hHYD*), *MCT4*, and *calnexin*. Moreover, biallelic alterations may have occurred more frequently than we observed; our sensitivity of detecting loss of the normal allele may have been low in some tumors because of normal cell contamination. In addition, it is possible that biallelic alterations in some genes occurred by aberrant promoter methylation or by mutations in regions not analyzed in this study.

Comparison of Mutational Spectra between Gastric and Colorectal Cancers. To address the organ specificity of coding MS mutational spectra, a comparison of these spectra in gastric and colorectal cancers was also performed, based partially on our previous data on colorectal cancers (Fig. 2; Ref. 5). One hundred forty-eight of 154 loci in the current study had mutation frequency data available for both gastric and colorectal cancers (Table 2). Many loci were mutated at similar rates in both types of cancer. Among these were *BAX, TTK, KIAA0896 (hHYD),* and *hMSH3*, as depicted in Table 2, group c, and Fig. 2. In contrast, loci with markedly different mutation frequencies between gastric and colorectal cancers are listed in Table 2, groups a and b. Loci with the highest mutation frequencies are also illustrated in Fig. 2. Although the number of gastric cancers evaluated in this study was relatively small, the following loci were mutated significantly more frequently in gastric than in colorectal cancers: *DKFZ564K112* (an *msx* homo-

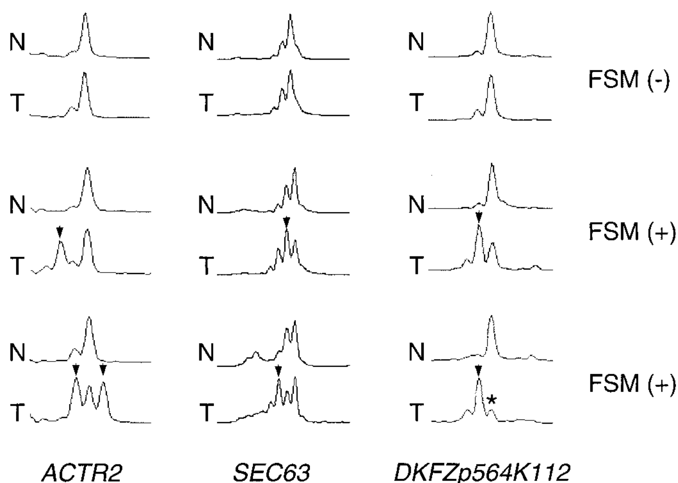

Fig. 1. Typical electrophoresis profiles of FSM-positive and FSM-negative tumors. Coding loci for profiles exhibited in *left column: Activin type II receptor* (*ACTRII*, GI 178049); in *center column: SEC63* (GI 5327053); in *right column: DKFZp564K112* (GI 4884248). *T,* tumor; *N,* normal. The *top row* shows electrophoresis profiles of three FSM-negative tumor and corresponding normal DNAs, whereas the *middle* and *bottom rows* show those of three FSM-positive cases. The *bottom cases* of *ACTRII* and *DKFZp564K112* exhibit biallelic alterations (*ACTRII*, one base insertion and one base deletion on each of the alleles; *DKFZp564K112,* one base deletion and allelic loss on the other allele.) Peaks representing mutant alleles are indicated by *arrowheads.* The *asterisk* shows the loss of normal allele.

Table 1 *List of frequently mutated coding region MS loci and their FSM frequencies*

Descriptions of 42 of 154 target coding homopolymeric repeats mutated in >10% of the gastric tumors studied in this project are given. GI#, gene identification number; Gene description/BLAST search result, GenBank search/BLAST search. BLAST searches are either the identical genes according to BLASTN search (a) or genes with homology at the amino acid level according to BLASTP search (b); Locus, chromosomal localization according to National Center for Biotechnology Information LocusLink (http://www.ncbi.nlm.nih.gov/LocusLink/); Seq, nucleotide sequence of homopolymeric repeat; FSM, number of tumors mutated at each locus divided by number informative at each locus; %, mutation frequency; NA, not available. Descriptions and mutation frequencies for other loci analyzed in this study are available on our website at http://www.microarray.umaryland.edu/YM0320SupplTable.xls.

GI#	*Gene description/BLAST search result*	Locus	Seq	FSM	%
4507468	*Transforming growth factor, β receptor II (TGFBR2)*	3p22	A 10	11/18	61.1
178049	*Activin type-II receptor (ACTRII)*	2q22.2–q23.3	A 8	8/18	44.4
4884248	*DKFZp564K112/multi-sex-combs (Drosophila)*[b]	4	A 8	7/17	41.2
5327053	*SEC63*	6q21	A 10	6/16	37.5
307154	*MAC30 mRNA, 3′ end*	17p11.1	A 10	3/9	33.3
3800808	*Methyl-CpG binding protein 4 (MBD4)*	3q21–q22	A 10	6/18	33.3
4757837	*BCL2-associated X protein (BAX)*	19q13.3–q13.4	G 8	4/12	33.3
339646	*Prolin 4-hydroxylase, β polypeptide/thyroid hormone binding protein p55 (P4HB/p55)*	17q25	A 8	2/7	28.6
435598	*Interleukin-1 β converting enzyme, isoform β/caspase-1 (ICE/CASP1)*	11q23	A 8	5/18	27.8
1405347	*Retinoblastoma protein-binding zinc-finger protein (RIZ)*	1p36	A 9	5/18	27.8
4240280	*KIAA0896/hyperplastic discs homologue (hHYD)*[a]	8q22	A 8	5/18	27.8
181841	*mutS homologue 3 (hMSH3)*	5q11–q12	A 8	5/18	27.8
340010	*TTK protein kinase*	6q13–q21	A 9	5/18	27.8
2558941	*Absent in melanoma 2 (AIM2)*	1q22	A 10	5/18	27.8
2463627	*Monocarboxylate transporter member 4/solute carrier family 16 member 4 (MCT4/SLC16A4)*	1p12	T 9	4/17	23.5
4240298	*KIAA0905/SEC31*[a]	4q21.23	A 9	4/17	23.5
347313	*Heterogeneous nuclear ribonucleoprotein H1 (hnRNP)*	5q35.3	T 8	2/9	22.2
186522	*Calnexin*	5q35	T 8	3/14	21.4
1945614	*hkf-1/zinc finger protein 103 homologue (mouse) (ZFP103)*	2p11.2	A 8	2/10	20.0
415818	*Antigen identified by monoclonal antibody Ki-67 (MKI67)*	10q25–qter	A 8	3/17	17.6
4589529	*KIAA0943/autophagy-related cystein endopeptodase 2 (human)*[b]	2	T 9	3/17	17.6
415586	*Destrin/actin depolymerizing factor*	20p11.23	T 8	3/17	17.6
532055	*Protein-tyrosine-phosphatase D1/protein tyrosine phosphatase, nonreceptor type 21 (PTPD1/PTPN21)*	14q31.3	A 8	3/17	17.6
4504190	*mutS homologue 6 (MSH6)*	2p16	C 8	3/17	17.6
587145	*DEAD family nuclear RNA helicase HLA-B associated transcript (BAT1)*	6p21.3	T 8	1/6	16.7
1632761	*TPR repeat protein D/tetratricopeptide repeat domain 3 (TPRD/TTC3)*	21q22.2	A 8	3/18	16.7
435777	*Retinoblastoma binding protein 2 (RBBP2)*	12p11	A 8	3/18	16.7
3928761	*Ubiquitin specific protease 1 (USP-1/hUB)*	1p32.1–p31.3	A 8	3/18	16.7
4884434	*DKFZp586D0623*	6q23.1–q24.1	A 8	3/18	16.7
4886456	*DKFZp566G163/NADH-ubiquinone oxydodehydrogenase B14.5B subunit*[a]	NA	T 9	3/18	16.7
4205083	*WW domain binding protein-1 (WBP1)*	2p12	C 9	2/15	13.3
940538	*Receptor interacting protein 140/nuclear receptor interacting protein 1 (RIP140/NRIP1)*	21q11.2	A 9	2/16	12.5
2062689	*Solute carrier family 17 member 2/sodium phosphate transporter (SLC17A2/NPT3)*	6p21.3	A 8	2/17	11.8
1869817	*Translocation-associated papillary renal cell carcinoma (PRCC)*	1q21.1	C 8	2/17	11.8
604478	*Transcription factor Dp-2 (Humpdp2/TFDP2)*	3q23	A 8	2/18	11.1
598224	*Protein kinase C iota isoform (PRKCI)*	Xq21.3	A 8	2/18	11.1
296032	*v-myb myeloblastosis viral oncogene homologue-like 1 (AMYB/MYBL1)*	8q22	A 8	2/18	11.1
1827451	*Vaccinia-related kinase 2 (VRK2)*	2p16–p15	A 8	2/18	11.1
2467368	*Synaptonemal complex protein 1 (SYCP1)*	1p13–p12	A 8	2/18	11.1
2224612	*KIAA0336/cutaneous T-cell lymphoma-associated antigen sel-1*[a]	2q12.3	A 8	2/18	11.1
2697004	*Cell cycle protein p38-2G4 homologue/erbB3 binding protein 1 (hG4-1/PA2G4/EBP1)*	12q13	A 8	2/18	11.1
4589597	*KIAA0977/cordone-bleu (Mus musclus)*[b]	2q24.3	T 9	2/18	11.1

logue, 41.2% in gastric cancers *versus* 6.8% in colorectal cancers), *MAC30* (33.3% *versus* 4.9%, respectively), *RIZ* (27.8% *versus* 4.5%), and *ICECASP1* (27.8% *versus* 2.4%); $P < 0.05$, Fig. 2 and Table 2, group a. Conversely, *TGFBR2* (61.1% *versus* 79.1%) and *AIM2* (27.8% *versus* 47.6%) tended to mutate more frequently in colorectal cancers (Table 2, group b). Statistical significance was not achieved at these latter two loci by Fisher's exact test because of the imbalance between the number of gastric ($n = 18$) and colorectal ($n = 46$) cases. The difference between mutational spectra of gastric and colorectal MSI-H cancers was also evaluated using the χ^2 test for goodness of fit, as described in "Materials and Methods." This calculation was performed based on data for 46 genes with mutation rates of 10% or higher in at least one of the two organ sites. Our analysis of colorectal cancers showed that P for the calculated χ^2 value (73.9) was 0.0032 (the number of gastric cancers was too small to reliably use the χ^2 test for goodness of fit). Thus, because colorectal cancers showed a highly significant deviation from the mean mutation frequencies between gastric and colorectal cancers, our data disprove the null hypothesis (that mutation frequencies are the same in both tumor types). This result suggests that the mutational profiles of colorectal and gastric cancers are significantly different, in support of the belief that the pathophysiologies underlying these two types of MSI-H cancers are distinct.

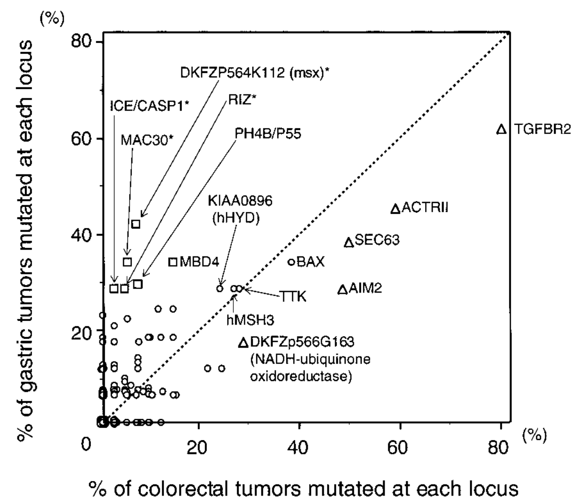

Fig. 2. FSM frequencies for gastric and colorectal tumors at each coding region MS locus. One hundred forty-eight coding MSs with mutational data for both colorectal and gastric cancers are plotted according to their mutation frequencies in 46 colorectal (*X axis*) and 18 gastric (*Y axis*) MSI-H tumors. Along the *diagonal line,* mutation rates for gastric and colorectal cancers are equal. Thirty-eight loci are plotted at the origin (*i.e.,* showing 0% mutation frequencies in both tumor types). □, loci mutated more frequently in gastric than in colorectal cancers; △, loci mutated more frequently in colorectal than in gastric cancers; *, loci with statistically significant differences in mutation rates between gastric and colorectal cancers (Fisher's exact test, $P < 0.05$).

Table 2 *Comparison of mutation frequencies between gastric and colorectal MSI-H cancers*

Group a, 20 loci mutated at least 10% more frequently in 18 gastric cancers than in 46 colorectal cancers. Gene description (BLAST search result), both gene names according to GenBank and results of BLAST search are shown in this column. Gene names in parentheses show either the identical gene according to BLASTN (a) or a gene with homology at the amino acid level according to BLASTP search (b); RAB2L: RAB2, member of Ras oncogene family-like (c). *P*s were calculated by Fisher's exact test. Group b, seven loci mutated at least 10% more frequently in colorectal cancers than in gastric cancers. Group c, 21 loci mutated at similar rates in both gastric and colorectal cancers showing mutation frequencies of >10% in at least one of these cancer types.

GI#	Gene description (BLAST search result)	Gastric		Colorectal		*P*
		FSM	%	FSM	%	
Group a						
4884248	*DKFZp564K112 (multi-sex-combs)*[b]	7/17	41.2	3/44	6.8	0.001
3800808	*MBD4*	6/18	33.3	6/43	14.0	0.082
307154	*MAC30 mRNA, 3′ end*	3/9	33.3	2/41	4.9	0.010
339646	*P4HB/p55*	2/7	28.6	3/42	7.1	0.083
1405347	*RIZ*	5/18	27.8	2/44	4.5	0.008
435598	*ICE/CASP1*	5/18	27.8	1/41	2.4	0.003
2463627	*MCT4/SLC16A4*	4/17	23.5	4/36	11.1	0.238
347313	*hnRNP H*	2/9	22.2	0/41	0.0	0.002
186522	*Calnexin*	3/14	21.4	2/41	4.9	0.063
1945614	*hkf-1/ZFP103*	2/10	20.0	1/43	2.3	0.029
415818	*MKI67*	3/17	17.6	0/45	0.0	0.004
3928761	*USP-1/hUB*	3/18	16.7	1/42	2.4	0.042
1632761	*TPRD/TTC3*	3/18	16.7	1/44	2.3	0.036
435777	*RBBP2*	3/18	16.7	1/43	2.3	0.039
587145	*BAT1*	1/6	16.7	0/33	0.0	0.018
2062689	*SLC17A2/NPT3*	2/17	11.8	0/42	0.0	0.024
604478	*Humpdp2/TFDP2*	2/18	11.1	0/44	0.0	0.025
598224	*PRKCI*	2/18	11.1	0/44	0.0	0.025
296032	*AMYB/MYBL1*	2/18	11.1	0/43	0.0	0.026
1827451	*VRK2*	2/18	11.1	0/41	0.0	0.030
Group b						
4507468	*TGFBR2*	11/18	61.1	34/43	79.1	0.146
178049	*ACTRII*	8/18	44.4	25/43	58.1	0.328
5327053	*SEC63*	6/16	37.5	21/43	48.8	0.437
2558941	*AIM2*	5/18	27.8	20/42	47.6	0.153
4886456	*DKFZp566G163 (NADH-ubiquinone oxydoreductase)*[a]	3/18	16.7	12/43	27.9	0.352
4589597	*KIAA0977 (cordone-bleu)*[b]	2/18	11.1	10/42	23.8	0.260
4886476	*DKFZp564D0782 (RAB2L)*[a,c]	0/17	0.0	5/43	11.6	0.142
Group c						
1869817	*PRCC*	2/17	11.8	1/44	2.3	0.124
4240298	*KIAA0905 (SEC31)*[a]	4/17	23.5	6/43	14.0	0.370
4884434	*DKFZp586D0623*	3/18	16.7	3/41	7.3	0.274
2467368	*SYCP1*	2/18	11.1	1/42	2.4	0.155
4589529	*KIAA0943 (autophagy-related cystein endopeptodase 2)*[b]	3/17	17.6	4/44	9.1	0.347
415586	*Destrin/actin depolymerizing factor*	3/17	17.6	4/42	9.5	0.382
4205083	*WBP1*	2/15	13.3	3/43	7.0	0.450
532055	*PTPD1/PTPN21*	3/17	17.6	5/43	11.6	0.537
940538	*RIP140/NRIP1*	2/16	12.5	3/42	7.1	0.516
4240280	*KIAA0896 (hHYD)*[a]	5/18	27.8	10/43	23.3	0.708
2224612	*KIAA0336 (sel-1)*[a]	2/18	11.1	3/43	7.0	0.591
4504190	*hMSH6*	3/17	17.6	6/43	14.0	0.718
181841	*hMSH3*	5/18	27.8	11/42	26.2	0.899
340010	*TTK*	5/18	27.8	12/44	27.3	0.968
4757837	*BAX*	4/12	33.3	15/40	37.5	0.793
4504610	*IGF2R*	1/16	6.3	5/42	11.9	0.527
2873376	*Exportin t*	1/18	5.6	6/43	14.0	0.348
182817	*Coagulation factor VIII, procoagulant component*	1/18	5.6	6/41	14.6	0.321
2697004	*hG4-1/PA2G4/EBP1*	2/18	11.1	9/43	20.9	0.363

Discussion

Many of the genes in which we found high FSM rates have known potential links to carcinogenesis. For example, *ACTRII* is a member of the *TGF-β receptor* family that is involved in the induction of differentiation, growth suppression, and apoptosis. Mutant *ACTRII* inhibits activin-mediated induction of differentiation (12). *DKFZ564K112* is a human homologue of the *Drosophila* tumor suppressor gene *multi-sex-comb (mxc)*, a member of the homeobox gene transcription repressor family Polycomb group. *Mxc* is ubiquitously expressed, and its loss of function provokes uncontrolled malignant cell growth (13). *ICE/caspase-1,* a human homologue of *Caenorhabditis elegans CED3*, is a mediator of Fas-mediated apoptosis. Down-regulation of *ICE/caspase-1* is observed in various human cancers (14), and overexpression of *ICE/caspase-1* inhibits the growth of renal cell carcinoma cell lines *in vivo* (15). *KIAA0896* (h*HYD*) is a human homologue of the *Drosophila* tumor suppressor gene *hyperplastic discs*. h*HYD* is expressed at moderately high levels in various tissues, and it acts as a ubiquitin ligase that coordinates DNA topoisomerase IIb-binding protein 1 during the DNA damage response (16). *TPRD1* is a candidate gene for Down's syndrome, a syndrome known to predispose to cancer, and its gene product is involved in mediating protein-protein interactions (17). It is expressed ubiquitously during mouse embryogenesis (18), suggesting a role in controlling cell differentiation.

Interestingly, among the loci mutated most frequently in gastric tumors, there was a subgroup of four known ER chaperone protein genes, which included *SEC63*. The other members of this subgroup were *calnexin*, *P4HB (p55),* and *SEC31*. In conjunction with known individual functions of these genes, this result suggests the involvement of chaperone proteins in gastric carcinogenesis. For example, calnexin is involved in ER accumulation of various proteins, including MHC class I antigen and the T cell receptor (19), which may affect the immunogenicity or intercellular interactions of tumor cells. Indeed, reduced expression of calnexin, correlated with anchorage-indepen-

dent growth, was observed in a human colon cancer cell line (20). P4HB (p55) plays a crucial role in collagen synthesis, and the same gene is also known as an inducer of growth arrest, *protein disulfide isomerase* (21). SEC63 is involved in the process of protein folding and translocation, including the nuclear translocation of nucleoproteins (22). Finally, SEC31 has a role in ER-Golgi protein transport (23).

Evaluating the functional significance of coding MS mutations is a somewhat difficult task. Systematic mutation frequency surveys at short homopolymers with eight nucleotides located within the 3′ untranslated region (5) and intronic (24) regions in MSI-H tumors have revealed that some of these loci show mutation rates (up to 50%). Because these mutations are unlikely to have effects on gene function or to be objects of selection pressure during carcinogenesis, the high frequency of mutation in these noncoding tracts may be promoted by surrounding nucleotide sequences or genomic structure. Therefore, frequent coding MS mutations *per se* may not connote direct involvement in carcinogenesis. Additional evidence is needed to prove that disruption of these genes is involved in carcinogenesis. In the current study, we addressed this dilemma by exploring published known or potential functional links to cancer and complete inactivation of genes caused by biallelic alteration. In the future, additional molecular genetic findings supporting involvement of candidate genes in carcinogenesis could include mutations occurring outside of the MS tracts we studied, particularly in non-MSI-H tumors.

Unlike single candidate gene studies of human cancers, instabilotyping offers the potential advantage of removing preconceived biases regarding roles of particular genes in cancers. This unbiased approach identifies some coding region MSs that are not mutated at all. However, the apparent rarity of mutations found by this approach may itself constitute an advantage; frequently mutated loci identified using this approach often have known or potential links to cancer. Moreover, biallelic mutations at these coding region MSs suggest that total inactivation of at least some of these genes is occurring, possible creating a growth or survival advantage for these gastric cancer cells. Thus, our findings suggest that comprehensive instabilotyping is a valid strategy to identify cancer-related genes. The above data imply that instabilotyping is an advantageous approach not only to discover novel mutations among coding MS loci in human cancer but also to identify tissue specificity patterns in mutation rates at these loci. These observed intertissue differences in mutational spectra may reflect known differences in etiology and biology between gastric and colorectal cancers. Finally, these findings imply that our instabilotyping strategy should be applied to other tumors showing frequent MSI, such as endometrial cancers.

References

1. Markowitz, S., Wang, J., Myeroff, L., Parsons, R., Sun, L., Lutterbaugh, J., Fan, R. S., Zborowska, E., Kinzler, K. W., Vogelstein, B., *et al.* Inactivation of the type II TGF-β receptor in colon cancer cells with microsatellite instability. Science (Wash. DC), *268:* 1336–1338, 1995.
2. Souza, R. F., Appel, R., Yin, J., Wang, S., Smolinski, K. N., Abraham, J. M., Zou, T. T., Shi, Y. Q., Lei, J., Cottrell, J., Cymes, K., Biden, K., Simms, L., Leggett, B., Lynch, P. M., Frazier, M., Powell, S. M., Harpaz, N., Sugimura, H., Young, J., and Meltzer, S. J. Microsatellite instability in the insulin-like growth factor II receptor gene in gastrointestinal tumours. Nat. Genet., *14:* 255–257, 1996.
3. Rampino, N., Yamamoto, H., Ionov, Y., Li, Y., Sawai, H., Reed, J. C., and Perucho, M. Somatic frameshift mutations in the *BAX* gene in colon cancers of the microsatellite mutator phenotype. Science (Wash. DC), *275:* 967–969, 1997.
4. Yin, J., Kong, D., Wang, S., Zou, T. T., Souza, R. F., Smolinski, K. N., Lynch, P. M., Hamilton, S. R., Sugimura, H., Powell, S. M., Young, J., Abraham, J. M., and Meltzer, S. J. Mutation of *hMSH3* and *hMSH6* mismatch repair genes in genetically unstable human colorectal and gastric carcinomas. Hum. Mutat., *10:* 474–478, 1997.
5. Mori, Y., Yin, J., Rashid, A., Leggett, B. A., Young, J., Simms, L., Kuehl, P. M., Langenberg, P., Meltzer, S. J., and Stine, O. C. Instabilotyping: comprehensive identification of frameshift mutations caused by coding region microsatellite instability. Cancer Res., *61:* 6046–6049, 2001.
6. Boland, C. R., Thibodeau, S. N., Hamilton, S. R., Sidransky, D., Eshleman, J. R., Burt, R. W., Meltzer, S. J., Rodriguez-Bigas, M. A., Fodde, R., Ranzani, G. N., and Srivastava, S. 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., *58:* 5248–5257, 1998.
7. Meltzer, S. J., Yin, J., Manin, B., Rhyu, M. G., Cottrell, J., Hudson, E., Redd, J. L., Krasna, M. J., Abraham, J. M., and Reid, B. J. Microsatellite instability occurs frequently and in both diploid and aneuploid cell populations of Barrett's-associated esophageal adenocarcinomas. Cancer Res., *54:* 3379–3382, 1994.
8. Yamamoto, H., Sawai, H., and Perucho, M. Frameshift somatic mutations in gastrointestinal cancer of the microsatellite mutator phenotype. Cancer Res., *57:* 4420–4426, 1997.
9. Riccio, A., Aaltonen, L. A., Godwin, A. K., Loukola, A., Percesepe, A., Salovaara, R., Masciullo, V., Genuardi, M., Paravatou-Petsotas, M., Bassi, D. E., Ruggeri, B. A., Klein-Szanto, A. J., Testa, J. R., Neri, G., and Bellacosa, A. The DNA repair gene *MBD4* (*MED1*) is mutated in human carcinomas with microsatellite instability. Nat. Genet., *23:* 266–268, 1999.
10. Chadwick, R. B., Jiang, G. L., Bennington, G. A., Yuan, B., Johnson, C. K., Stevens, M. W., Niemann, T. H., Peltomaki, P., Huang, S., and de la Chapelle, A. Candidate tumor suppressor RIZ is frequently involved in colorectal carcinogenesis. Proc. Natl. Acad. Sci. USA, *97:* 2662–2667, 2000.
11. Duval, A., Rolland, S., Compoint, A., Tubacher, E., Iacopetta, B., Thomas, G., and Hamelin, R. Evolution of instability at coding and non-coding repeat sequences in human MSI-H colorectal cancers. Hum Mol Genet., *10:* 513–518, 2001.
12. Liu, F., Shao, L. E., and Yu, J. Truncated activin type II receptor inhibits erythroid differentiation in K562 cells. J. Cell. Biochem., *78:* 24–33, 2000.
13. Saget, O., Forquignon, F., Santamaria, P., and Randsholt, N. B. Needs and targets for the *multi sex combs* gene product in *Drosophila melanogaster*. Genetics, *149:* 1823–1838, 1998.
14. Winter, R. N., Kramer, A., Borkowski, A., and Kyprianou, N. Loss of caspase-1 and caspase-3 protein expression in human prostate cancer. Cancer Res., *61:* 1227–1232, 2001.
15. Ueki, T., Takeuchi, T., Nishimatsu, H., Kajiwara, T., Moriyama, N., Narita, Y., Kawabe, K., Ueki, K., and Kitamura, T. Silencing of the *caspase-1* gene occurs in murine and human renal cancer cells and causes solid tumor growth *in vivo*. Int. J. Cancer, *91:* 673–679, 2001.
16. Honda, Y., Tojo, M., Matsuzaki, K., Anan, T., Matsumoto, M., Ando, M., Saya, H., and Nakao, M. Cooperation of HECT-domain ubiquitin ligase hHYD and DNA topoisomerase II-binding protein for DNA damage response. J. Biol. Chem., *277:* 3599–3605, 2002.
17. Tsukahara, F., Hattori, M., Muraki, T., and Sakaki, Y. Identification and cloning of a novel cDNA belonging to tetratricopeptide repeat gene family from Down syndrome-critical region 21q22.2. J. Biochem. (Tokyo), *120:* 820–827, 1996.
18. Lopes, C., Rachidi, M., Gassanova, S., Sinet, P. M., and Delabar, J. M. Developmentally regulated expression of *mtprd,* the murine ortholog of *tprd,* a gene from the Down syndrome chromosomal region 1. Mech. Dev., *84:* 189–193, 1999.
19. Hochstenbach, F., David, V., Watkins, S., and Brenner, M. B. Endoplasmic reticulum resident protein of 90 kilodaltons associates with the T- and B-cell antigen receptors and major histocompatibility complex antigens during their assembly. Proc. Natl. Acad. Sci. USA, *89:* 4734–4738, 1992.
20. Yeates, L. C., and Powis, G. The expression of the molecular chaperone calnexin is decreased in cancer cells grown as colonies compared to monolayer. Biochem. Biophys. Res. Commun., *238:* 66–70, 1997.
21. Greene, J. J., and Brophy, C. I. Induction of protein disulfide isomerase during proliferation arrest and differentiation of SH5Y neuroblastoma cells. Cell. Mol. Biol., *41:* 473–480, 1995.
22. Sadler, I., Chiang, A., Kurihara, T., Rothblatt, J., Way, J., and Silver, P. A yeast gene important for protein assembly into the endoplasmic reticulum and the nucleus has homology to DnaJ, an *Escherichia coli* heat shock protein. J. Cell Biol., *109:* 2665–2675, 1989.
23. Salama, N. R., Chuang, J. S., and Schekman, R. W. Sec31 encodes an essential component of the COPII coat required for transport vesicle budding from the endoplasmic reticulum. Mol. Biol. Cell, *8:* 205–217, 1997.
24. Zhang, L., Yu, J., Willson, J. K., Markowitz, S. D., Kinzler, K. W., and Vogelstein, B. Short mononucleotide repeat sequence variability in mismatch repair-deficient cancers. Cancer Res., *61:* 3801–3805, 2001.