DNA AND CELL BIOLOGY
Volume 28, Number 4, 2009

Pp. 1–8
DOI: 10.1089/dna.2008.0827

Variants of *MUC5B* Minisatellites and the Susceptibility of Bladder Cancer

Eun-Kyung Ahn,[1,2] Wun-Jae Kim,[3] Jeong-Ah Kwon,[1] Phil-Jo Choi,[4] Woo Jin Kim,[5] Yangil Sunwoo,[1] Jeonghoon Heo,[2] and Sun-Hee Leem[1]

The human *MUC5B* gene, which is primarily expressed in the tracheobronchial tract, is clustered to chromosome 11p15.5 with three other secreted gel-forming mucins, *MUC6*, *MUC2*, and *MUC5AC*. In this study, we identified seven variable number of tandem repeats (VNTRs; minisatellites) from the entire *MUC5B* region. Six (*MUC5B*-MS1, -MS2, -MS3, -MS4, -MS5, and -MS7) of the seven minisatellites evaluated in this study were novel minisatellites, but the *MUC5B*-MS6 minisatellite was described in a previous study. These minisatellites of *MUC5B* were analyzed in genomic DNA extracted from controls, cancer patients, and multigenerational families. Three (*MUC5B*-MS3, -MS6, and -MS7) of the seven minisatellites were found to be polymorphic and transmitted through meiosis following Mendelian inheritance in seven families; therefore, these minisatellite polymorphisms could be useful as markers for paternity mapping and DNA fingerprinting. In addition, we evaluated allelic variation in these minisatellites to determine if such variation affected the susceptibility to various carcinomas. To accomplish this, we conducted a case–control study in which the genomic DNA of 789 cancer-free controls and cancer patients with five types of cancer were compared. A statistically significant association between the long rare *MUC5B*-MS6 alleles and the occurrence of bladder cancer was identified in the younger group (<60; odds ratio, 4.54; 95% confidence interval, 1.0–20.7; $p = 0.03$). This observation suggests that the long rare *MUC5B*-MS6 alleles evaluated in this study could be used to identify the risk of bladder cancer.

Introduction

Mucus, which is the slimy and viscous matter that covers the delicate epithelial surfaces of the respiratory, gastrointestinal, and reproductive tracts (Gum, 1992; Martínez-Antón *et al.*, 2006), is primarily comprised of mucins (Wickstrom *et al.*, 1998). Mucins are highly glycosylated proteins that contain oligosaccharides that are attached to the hydroxyl group of serine and threonine in the protein backbone through O-linked glycans (Moniaux *et al.*, 2001; James and Robert, 2004). Under normal conditions, mucin genes protect the epithelial surfaces; however, mucus hypersecretion leads to inflammatory diseases such as asthma, as well as to carcinomas (Gendler *et al.*, 1990; Corfield *et al.*, 2001). Mucins have been given the gene symbol *MUC*, followed by a number. To date there are 17 *MUC* genes endorsed by the HUGO gene nomenclature committee (http://www.gene.ucl.ac.uk/nomenclature/). Of these, 10 code for cell-tethered mucins (*MUC1*, *MUC3A*, *MUC3B*, *MUC4*, *MUC12*, *MUC13*, *MUC15*, *MUC16*, *MUC17*, and *MUC20*). The other seven MUC genes code for secreted mucins, which include five for polymeric mucins (*MUC2*, *MUC5AC*, *MUC5B*, *MUC6*, and *MUC19*) and two for nonpolymeric glycoproteins (*MUC7* and *MUC8*) (Thornton *et al.*, 2008).

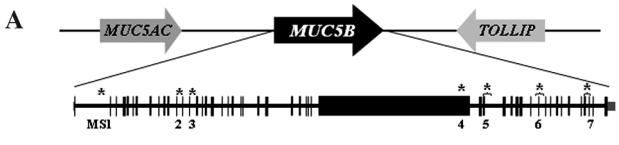
Four secreted gel-forming mucins (*MUC2*, *MUC5AC*, *MUC5B*, and *MUC6*) that have genes clustered on chromosome 11p15.5 contain structural characteristics that are affected by gene expressions (Pigny *et al.*, 1996; Desseyn *et al.*, 1998; Vinall *et al.*, 1998). Among the secreted mucins, the salivary mucin, *MUC5B* (previously MG1, high molecular weight protein), is primarily expressed in the tracheobronchial tract. *MUC5B*, which has been isolated from human submandibular/sublingual secretion (HSMSL), is composed of 14.9% protein, 78.1% carbohydrate, and 7% sulfate (Audie *et al.*, 1995; Troxler *et al.*, 1995; Wickstrom *et al.*, 1998). *MUC5B* is also expressed in many organs that

[1]Department of Biological Science, Dong-A University, Busan, Korea.
[2]Department of Molecular Biology and Immunology, Kosin University College of Medicine, Busan, Korea.
[3]Department of Urology, College of Medicine, Chungbuk National University, Cheongju, Korea.
[4]Department of Chest Surgery, College of Medicine, Dong-A University, Busan, Korea.
[5]Department of Internal Medicine, College of Medicine, Kangwon National University, Chuncheon, Korea.

secrete mucus, including the lungs (Copin *et al.*, 2001), gallbladder (Van Klinken *et al.*, 1998), and breasts (Sonora *et al.*, 2006), as well as in the urogenital tract (Russo *et al.*, 2006).

Most mucin genes contain a central region with a variable number of tandem repeat (VNTR) (Vinall *et al.*, 1998; Desseyn *et al.*, 1999). Tandem repeat sequences have been classified into two groups, minisatellites and microsatellites. Tandem arrays of microsatellites are comprised of short (<10 bp) units, while minisatellites are longer (>10 and <100 bp) (Richard and Paques, 2000; Denoeud *et al.*, 2003). Tandem repeats play a variety of regulatory and evolutionary roles, are used as analytic tools, and have been shown to cause human disease (Nakamura *et al.*, 1987). Several associations between minisatellites and diseases have been reported. For example, it is suspected that the rare alleles of *HRAS1* minisatellites are associated with the risk of a number of different cancers (Krontiris *et al.*, 1993), and rare alleles associated with the *MUC3* gene may confer genetic predisposition to ulcerative colitis (Kyo *et al.*, 1999). We recently demonstrated that the presence of *MUC2*-MS6 minisatellites was associated with an increased likelihood of the development of gastric cancer (Jeong *et al.*, 2007). In addition, a statistically significant association between the rare SLC 6A19-MS7 alleles and the occurrence of hypertension has been identified (Seol *et al.*, 2008).

In this study, we analyzed seven minisatellites from the entire *MUC5B* region and the putative functional significance of allelic variation in these minisatellites with respect to the susceptibility to five types of carcinomas. For genotyping of the *MUC5B* polymorphisms, the genomic DNA from the blood of 789 normal individuals, two generations of five families, and three generations of two families were analyzed. Three (*MUC5B*-MS3, -MS6, and -MS7) of seven minisatellites analyzed were found to be polymorphic in *MUC5B*. Because several diseases are known to be related to VNTR polymorphisms, we evaluated the putative functional significance of allelic variation in these minisatellites with respect to the susceptibility for carcinomas. Genomic DNA samples from 789 cancer-free controls, 436 patients with gastrointestinal cancer, 169 patients with colon cancer, 327 patients with rectal cancer, 269 patients with bladder cancer, and 168 patients with lung cancer were analyzed. Here, we report the possible association of rare *MUC5B* minisatellite variants with the relative risks for carcinomas.



B

Minisatellite	Indices	Location	Size (bp) (PCR product)	Consensus size (bp)	Consensus sequences	GC contents
MS1	1202250 -1202413	Intron 1	164 (336)	21	GCCTGGGGCA GGGGAGGGGC T	81%
MS2	1207606 -1207755	Exon 10 ~ Intron 12	150 (234)	49	AGCTCCTGGT ATTATGACCC ACCAGCCTCC GCCTGGGGTG GGGGTGTGG	65%
MS3	1208476 -1208583	Intron 12	92 (333)	25	GGTCAGGTCC CCCCTCCAGC CCCCA	76%
MS4	1227775 -1228104	Exon 31	286 (388)	86	TCCACCCCCT CCTCCACTCC GGGGACCACC TGGATCCTCA CAGAGCTGAC CACGAAGGCC ACCACAACAG CAGCCACCAG CCCCACG (29aa: STPSSTPGTTWILTE LTTKATTTAATSPT)	67%
MS5	1229657 -1230014	Intron 31	358 (441)	20	TCCCCACTGG CCACACTCGG	70%
MS6	1233357 -1233930	Exon 37 ~ Int 37	566 (901)	59	CCTGTGCGGT GAGTGGGGGC GGCCCCGGGC CCCCCAGACC CCTCGGCCTC TCTGAGTGT	76%
MS7	1236856 -1237162	Exon 44 ~ Intron 44	307 (600)	26	ACCTGCCCCC AGGTGAGACC CGAGGC	73%

FIG. 1. Minisatellites in the *MUC5B* gene. (**A**) Structure of the genomic region around *MUC5B*. It is predicted that 49 exons (black boxes) encode *MUC5B*. (**A**) The approximate positions of minisatellites identified by the Tandem Repeats Finder Program (Benson, 1999) are indicated by asterisks and numbers (MS1, 2, 3, 4, 5, 6, 7). (**B**) The sequences of seven minisatellite repeat units. Positions of indices (1–39,111 bp) were determined using genomic information from the UCSC and NCBI.

Materials and Methods

Analysis of the tandem repeat regions of MUC5B *and primer construction*

All DNA sequence regions analyzed in this study were based on *MUC5B* genomic sequences (length in 1–39,111 bp; Fig. 1) and assembled in UCSC (>hg18_refGene_NM_002458 range=chr11:1200872-1239982) and NCBI (>ref|NC_000011.8|NC_000011:1200872-1239982 *Homo sapiens* chromosome 11, reference assembly, complete sequence). The Tandem Repeats Finder software (Benson, 1999) was used to detect VNTRs and other repeated regions. Repeat units between 10 and 100 bp in length that scored >200 in the program algorithm were selected for further analysis.

All primers used in this study were designed using the Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) (Rozen and Skaletsky, 2000). The following primers were then constructed based on the sequence on the UCSC site: MS1, F-CAGGTGGTCCAGCCCAAGGA & R-ACCCCGACCCACAGCTCACA; MS2, F-CTACAGC CCGGGCACCTCCT & R-GAGGGAGGCCGTGAGCAGAA; MS3, F-GCTGGCTGAGCTGCGGAAGT & R-CACTGGGG ACAGAGCCCACA; MS4, F-ACCACGACAAGGGCCACC AG & R-TGGAGGAGGGAGTGGCTTTGG; MS5, F-TGCC TGTCCTGGGAGCCAGT & R-GCGGTTCCAGGGAAGGG AGT; MS6, F-CCTGCCAGAGCCTGGAGGCTTAC & R-CT GTGGGCACAGAGAGGTGTCAGA; MS7, F-GTCAGGAG GATGCCTGCAACAATA & R-CCCAGCAATGGCTGAGC TAGAGAA.

Preparation of genomic DNA

To assess the degree of minisatellite polymorphism in the *MUC5B* genes, genomic DNAs were isolated from 789 healthy unrelated individuals. A case–control study was then performed using DNA collected from 789 cancer-free controls and 1369 cancer patients (Table 1). Hospital-based cancer case subjects were obtained from three different hospitals in three different cities (Dong-A University Hospital [#IRB-06-10-02 & IRB-07-10-7; Busan, Korea], Kangwon National University Hospital [#IRB-Bioscience: 05-02; Chuncheon, Korea], and Chungbuk National University Hospital [#IRB-2006-1; Cheongju, Korea]). Controls have no history of cancer and were recruited in the same department. For the PCR experiments, genomic DNA was isolated from the peripheral leukocytes, which were isolated from 400 μL of whole blood using a Blood and Cell Culture DNA Mini Kit (Qiagen, Valencia, CA).

Table 1. Age and Sex Distribution of Cases and Controls

Characteristic	*Level*	*Controls, n (%)*	*Gastric cancer*	*Colon cancer*	*Rectal cancer*	*Bladder cancer*	*Lung cancer*
Age (years)	40–49	156 (19.8)	76 (17.4)	27 (16.0)	55 (16.8)	22 (8.2)	11 (6.5)
	50–59	266 (33.7)	122 (28.0)	39 (23.1)	93 (28.4)	50 (18.6)	39 (23.2)
	60–69	216 (27.4)	157 (36.0)	66 (39.1)	117 (35.8)	118 (43.9)	83 (49.4)
	70–79	151 (19.1)	81 (18.6)	37 (21.9)	62 (19.0)	79 (29.4)	35 (20.8)
	<60	422 (53.5)	198 (45.4)	66 (39.1)	148 (45.3)	72 (26.8)	50 (29.8)
	≥60	367 (46.5)	238 (54.6)	103 (60.9)	179 (54.7)	197 (73.2)	118 (70.2)
Sex	Men	513 (65.0)	282 (64.7)	84 (49.7)	202 (61.8)	228 (84.8)	145 (86.3)
	Women	276 (35.0)	154 (35.3)	85 (50.3)	125 (38.2)	41 (15.2)	23 (13.7)
	Total	789	436	169	327	269	168

Analysis of polymorphism of the minisatellites of MUC5B

The primers used to analyze the gene polymorphisms were based on the *MUC5B* genomic sequence. Human genomic DNA was amplified using these primers under the following standard PCR conditions: 50 mM KCl, 10 mM Tris-HCl, pH 9.0, 2.5 mM $MgCl_2$ (MS6 is performed 3.75 mM), and 0.2 mM dTTP, dCTP, dGTP, and dATP in a reaction mixture with a final volume of 40 μL. The thermocycling conditions were as follows: 1 cycle of 2 min of initial denaturation at 94°C, followed by 30 cycles of denaturation at 94°C for 45 s, and annealing at 67°C for 1 min (*MUC5B*-MS1 and -MS6) (1 min/1 kb of DNA), at 68°C for 1 min (*MUC5B*-MS2, -MS3, -MS5, and -MS7), followed by a 7 min extension at 72°C in a 9700 Thermocycler (Perkin-Elmer, Foster City, CA). *MUC5B*-MS4 was amplified under the following conditions: 30 cycles of denaturation at 94°C for 30 s and annealing at 72°C for 30 s. PCR analysis of human DNA samples was performed using Go Taq Flexi DNA polymerase (Promega, Madison, WI) with 100 ng of genomic DNA.

PCR products were analyzed by gel electrophoresis (1 volt/cm) in 1×TAE buffer through a 1.0–1.5% SeaKem LE agarose gel (Cambrex, ME) depending to their repeats sizes. Specifically, *MUC5B*-MS1, -MS2, -MS4, and -MS5 were electrophoresed through a 1% SeaKem LE gel, and *MUC5B*-MS3, -MS6, and -MS7 were electrophoresed through a 1.5% SeaKem LE gel.

All PCR products amplified in the minisatellite regions (Fig. 1) were purified using a Gel Extraction Kit (Qiagen) and then confirmed by DNA sequencing.

Statistical analyses

The degree of polymorphism, which ranges from 0 to 1, generally increases with the number of alleles. To evaluate the probability of two randomly chosen alleles being different (heterozygosity) at a given locus, a measure of genetic diversity was calculated using the method described by Chakravarti and Lynn (2004). Regression analyses were performed to determining the odds ratios (ORs) in association between control and case groups. ORs were estimated using the natural logarithm and its standard error. Where relevant, we used a chi-squared test with one degree of freedom to assess statistical significance. Differences were considered significant for confidence intervals (CIs) of 95%. All tests were two-sided, with $p < 0.05$ being considered statistically significant. Statistical analyses were performed using MS Excel with CHITEST and R statistical software (v2.5.1, www.r-project.org) with chisq.test for the calculation of chi-squared values.

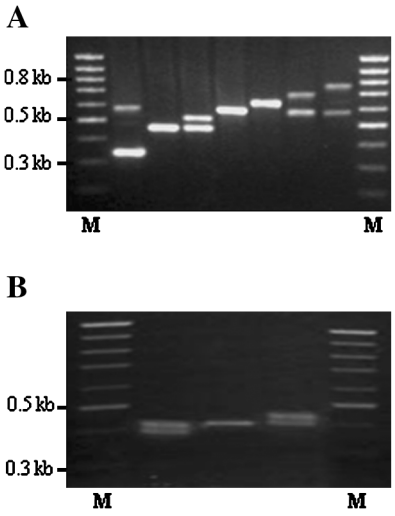
Results

Characterization of the seven minisatellites of the MUC5B *region*

The *MUC5B* DNA sequence region analyzed in this study was based on genomic sequences (length in 1–39,111 bp) obtained from the UCSC and NCBI databases. The *MUC5B* genomic DNA is composed of 49 exons and 48 introns, and sequence analysis of the entire *MUC5B* gene allowed the identification of seven tandem repeats. Six of the seven minisatellites (*MUC5B*-MS1, -MS2, -MS3, -MS4, -MS5, and -MS7) evaluated were novel VNTR loci that were first analyzed in this study, while *MUC5B*-MS6 was located in the same locus as a VNTR region that was previously reported by Desseyn *et al.* (1999). A search of the GenBank database using the BLASTN program revealed that there were no significant similarities between the six novel minisatellites (*MUC5B*-MS1, -MS2, -MS3, -MS4, -MS5, and -MS7) and previously identified regions. The length of these repeats, their locations, and consensus sequences are presented in Figure 1.

The degree of polymorphism within the minisatellites was examined by PCR using diagnostic primers against human genomic DNA samples isolated from control individuals. We characterized all seven minisatellites in the *MUC5B* region (Fig. 1). At first, we analyzed 100 cancer-free controls and 100 gastric cancer patients to determine the polymorphism of each minisatellites. Once polymorphism in minisatellites is identified, we used additional samples to search more variants in alleles. Four minisatellites, *MUC5B*-MS1 (8 copies of a 21 bp repeat unit in intron 1), *MUC5B*-MS2 (3 copies of a 50 bp repeat unit between exon 10 and intron 10), *MUC5B*-MS4 (4 copies of an 86 bp repeat unit in exon 31), and *MUC5B*-MS5 (18 copies of a 20 bp repeat unit in intron 31), were found to have a monomorphic pattern.

MUC5B-MS3 in intron 12 was shown to have a monomorphic pattern comprised of four copies of a 25 bp repeat unit in the controls; however, one additional allele of three copies was found in samples collected from patients with gastric cancer (data not shown). Therefore, we analyzed the *MUC5B*-MS3 region of samples collected from 100 additional



MUC5B-MS6 (h=0.5435)			
Repeats	Size (bp)	N=1578	Frequency
3	338	9	0.0057
5	456	458	0.2902
6	515	8	0.0051
7	574	952	0.6033
8	633	143	0.0906
9	692	1	0.0006
10	751	7	0.0044

MUC5B-MS7 (h=0.0050)			
Repeats	Size (bp)	N=1578	Frequency
11	385	1	0.0006
12	411	1574	0.9975
13	437	3	0.0019

FIG. 2. Polymorphic patterns of *MUC5B* minisatellites. (**A**) Polymorphic patterns of *MUC5B*-MS6 and (**B**) *MUC5B*-MS7. Minisatellites were PCR-amplified from the genomic DNA of control samples using diagnostic primers (Materials and Methods). Haplotype patterns are numbered according to each minisatellite. Size markers (M) are given in kb (100 bp size marker; Invitrogen, Carlsbad, CA). h Shows the heterozygosity of each minisatellite in the controls.

cancer-free controls and 100 gastric cancer patients; however, we did not detect any additional variants for this region. These findings indicate that *MUC5B*-MS3 was monomorphic in 200 controls, but only one of 200 cases with gastric cancer had the minisatellite variant.

The PCR products of seven alleles of the *MUC5B*-MS6 minisatellite from the 789 control samples ranged from 338 to 751 bp in length and contained 3–10 copies of the 59 bp repeat unit, with 7 copies of the most common allele being present (Fig. 2A). The degree of polymorphism within the *MUC5B*-MS6 minisatellite showed the highest heterozygosity ($h = 0.5435$) in the *MUC5B* region (Fig. 2A). In a previous study conducted by Desseyn *et al.* (1999), five alleles in this region that contained 3–8 copies of a 59 bp repeat were observed in 86 unrelated individuals. In the present study, we found two new minisatellite variants of 9 and 10 copies in exon 37–intron 37. Finally, the number of copies in the *MUC5B*-MS7 minisatellite, which was located between exon 44 and intron 44, varied from 11 to 13, with 12 copies being present in the most common allele (99.75% frequency) (Fig. 2B).

Overall, seven minisatellites were identified in *MUC5B*, three of which were polymorphic and four of which were monomorphic.

Mendelian inheritance of polymorphic minisatellites of MUC5B

To perform a segregation analysis of two polymorphic minisatellites, *MUC5B*-MS6 and *MUC5B*-MS7, we selected family groups of two and three generations (four and three families, respectively) (Fig. 3). Blood was collected from the grandparents, parents, and children of each family. Alleles from the *MUC5B*-MS6 and *MUC5B*-MS7 were then identified, and their transmission from parent to child was traced in seven of the families evaluated. Figure 3 illustrates the hereditary segregation of *MUC5B*-MS6 among three generations of one family (Fig. 3A, B) and two generations of

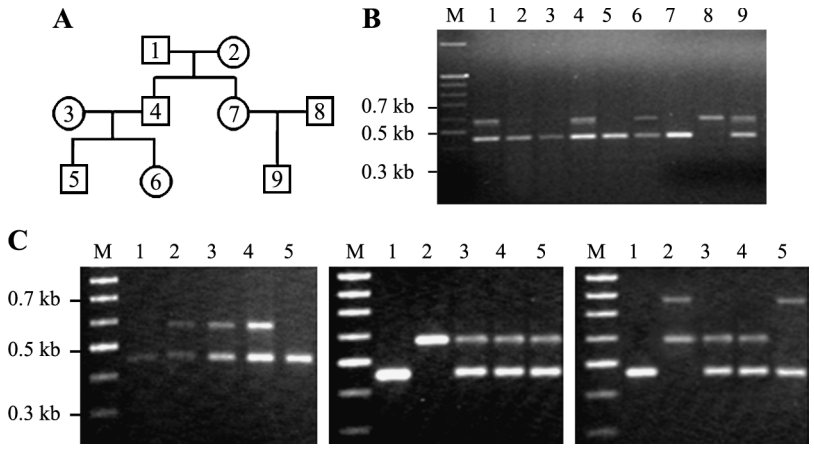


FIG. 3. Meiotic inheritance of *MUC5B*-MS6 minisatellites within families. PCR primers specific to *MUC5B*-MS6 were used to analyze the length of the alleles in the genomic DNA obtained from family members. (**A**) The pedigree demonstrates the relationship of the minisatellites in three generations of a family evaluated in this study (**B**). (**B**) Meiotic inheritance of *MUC5B*-MS6 in a three-generation family: first generation (lanes 1 and 2, grandfather and grandmother, respectively); second generation (lanes 3 and 7, fathers; lanes 4 and 8, mothers); and third generation (lanes 5 and 6, children from parents 3 and 4; lane 9, child from parents 7 and 8). (**C**) Meiotic inheritance of *MUC5B*-MS6 in two-generations of three families: first generation (lanes 1 and 2, father and mother, respectively); second generation (lanes 3–5; children from parents 1 and 2); M corresponds to the size marker.

Table 2. Frequency of Rare *MUC5B*-MS6 Alleles and Risk of Cancer or Asthmas

		Common alleles			*Short rare alleles*					*Long rare alleles*			
	Analyzed alleles	*5*	*7*	*8*	*Total*	*3*	*6*	*Total*	*OR (95% CI);* p	*9*	*10*	*Total*	*OR (95% CI);* p
Control	1578 (%)	458 (29.0)	952 (60.3)	143 (9.1)	1553 (98.4)	9 (0.6)	8 (0.5)	17 (1.1)	1.00 (Reference)	1 (0.1)	7 (0.4)	8 (0.5)	1.00 (Reference)
Gastric cancer	872 (%)	274 (31.4)	510 (58.5)	78 (8.9)	862 (98.9)	2 (0.2)	—	2 (0.2)	0.21 (0.05–0.92); $p = 0.02$	1 (0.1)	7 (0.8)	8 (0.9)	1.82 (0.68–4.86); $p = 0.23$
Colon cancer	338 (%)	93 (27.5)	207 (61.2)	36 (10.7)	336 (99.4)	—	—	—	—	—	2 (0.6)	2 (0.6)	1.17 (0.25–5.53); $p = 0.84$
Rectum cancer	654 (%)	187 (28.6)	399 (61.0)	62 (9.5)	648 (99.1)	3 (0.5)	1 (0.2)	4 (0.6)	0.56 (0.19–1.68); $p = 0.30$	—	2 (0.3)	2 (0.3)	0.60 (0.13–2.84); $p = 0.51$
Bladder cancer	538 (%)	160 (29.7)	316 (58.7)	51 (9.5)	527 (98.0)	2 (0.4)	2 (0.4)	4 (1.1)	0.69 (0.23–2.07); $p = 0.51$	—	7 (1.3)	7 (1.3)	2.58 (0.93–7.14); $p = 0.06$
Lung cancer	336 (%)	107 (31.9)	198 (58.9)	27 (8.0)	332 (98.8)	—	—	—	—	1 (0.3)	3 (0.9)	4 (1.2)	2.34 (0.70–7.81); $p = 0.16$

three families (Fig. 3C). Because of the low heterozygosity ($h = 0.0050$) of *MUC5B*-MS7 (Fig. 2B), we could only detect the 12 copies present in the most common allele in all of the families evaluated. These results demonstrate that these minisatellites are subject to Mendelian inheritance (i.e., children carried one minisatellite allele from each parent), and new minisatellites alleles were not observed during this analysis. Thus, minisatellites in *MUC5B* are meiotically stable and could potentially be used as markers to follow the meiotic segregation of *MUC5B* alleles.

Genetic association with cancer

Because minisatellites are genetically variable, it is possible that they play a role in the activation of *MUC5B* during tumorigenesis. In this study, we tested this possibility by comparing the distribution and frequency of the polymorphic *MUC5B* minisatellite alleles between controls and patients with gastric cancer, colon cancer, rectal cancer, bladder cancer, and lung cancer in a Korean population.

The allelic distribution of the minisatellites of *MUC5B* was compared between cancer-free controls and cancer patients with five types of cancer (Table 1). When the control samples and the cases were compared, no significant differences in the frequency of minisatellite alleles for the *MUC5B*-MS7 locus were observed (data not shown). Table 2 summarizes the frequency of rare and common alleles for the *MUC5B*-MS6 among cases and controls. When the control group and case groups were compared, there was an approximately twofold increase in the odds ratio of bladder cancer for the frequency of rare alleles of *MUC5B*-MS6. However, these results were not statistically significant.

For further analysis, we focused on the association between rare alleles and cancers. To accomplish this, *MUC5B*-MS6 alleles were grouped into two sets (common and rare alleles) according to their frequency in the control population. The expected frequency for rare alleles was considered to be ≤1%. Rare alleles were classified as short (3 and 6 copies) or long (9 and 10 copies) according to the length of their tandem repeats (Table 2). In patients with bladder and lung cancer, the rates of long rare *MUC5B*-MS6 alleles were 1.3% and 1.2%, respectively, whereas the rate was 0.5% in cancer-free controls (Table 2). Table 3 summarizes the frequency of rare *MUC5B*-MS6 alleles based on the age at diagnosis. In the control group, we compared the frequency of long rare alleles between younger individuals (<60 years, 0.95%) and older individuals (≥60 years, 1.09%); however, no significant difference was observed (Table 3). Conversely, when older cancer patients (≥60 years) were compared to younger cancer patients (<60 years), an increased frequency of long rare *MUC5B*-MS6 alleles was observed. Specifically, the following differences were observed among young and old patients: gastric cancer (younger group:older group = 2.15%:1.16%), colon cancer (1.61%:0.89%), rectal cancer (0.71%:0.51%), bladder cancer (4.54%:1.88%), and lung cancer (4.35%:1.56%). Higher frequencies of long rare alleles were found in younger patients with bladder and lung cancers (Table 3). Specifically, when the controls and cases were compared, the differences in the frequency of long rare *MUC5B*-MS6 alleles between younger- and older-bladder cancer patients were as follows: younger, 4.54 (CI, 1.0–20.7; $p = 0.03$) versus older 1.88 (CI, 0.5–7.6; $p = 0.37$) (Table 3).

Table 3. Frequency of Long Rare *MUC5B*-MS6 Alleles and Risk of Cancer or Asthma by Age

Age at diagnosis	*Controls*	*Long rare*	*Gastric cancer*	*Long rare*	*OR (95% CI);* p[a]
Younger (<60)	422	4 (0.95%)	198	4 (2.02%)	2.15 (0.5–8.7); $p = 0.27$
Older (≥60)	367	4 (1.09%)	238	3 (1.26%)	1.16 (0.3–5.2); $p = 0.85$
			Colon cancer		
Younger (<60)	422	4 (0.95%)	66	1 (1.51%)	1.61 (0.2–14.6); $p = 0.67$
Older (≥60)	367	4 (1.09%)	103	1 (0.97%)	0.89 (0.1–8.0); $p = 0.92$
			Rectal cancer		
Younger (<60)	422	4 (0.95%)	148	1 (0.68%)	0.71 (0.08–6.4); $p = 0.76$
Older (≥60)	367	4 (1.09%)	179	1 (0.56%)	0.51 (0.06–4.6); $p = 0.54$
			Bladder cancer		
Younger (<60)	422	4 (0.95%)	72	3 (4.17%)	4.54 (1.0–20.7); $p = 0.03$[b]
Older (≥60)	367	4 (1.09%)	197	4 (2.03%)	1.88 (0.5–7.6); $p = 0.37$
			Lung cancer		
Younger (<60)	422	4 (0.95%)	50	2 (4.00%)	4.35 (0.8–24.4); $p = 0.07$
Older (≥60)	367	4 (1.09%)	118	2 (1.69%)	1.56 (0.3–8.7); $p = 0.61$

[a]Reference (the same age group of control).
[b]Statistically significant ($p < 0.05$).

These results suggest that long rare *MUC5B*-MS6 alleles may be genetically associated. This is the first report in which minisatellites have been characterized in detail for the complete *MUC5B* region, and our observations suggest that the loci of *MUC5B* minisatellites may function as indicators of the risk of bladder cancer.

Discussion

Mucin genes contain a centrally located region of sequence that encodes tandem repeats, and most mucin genes exhibit a high degree of genetically determined polymorphism due to variation in the number of tandem repeats in the TR domain (Gendler and Spicer, 1995; Vinall *et al.*, 1998; Desseyn *et al.*, 1999; James and Robert, 2004). One of the secreted mucins, *MUC5B*, is expressed in the tracheobronchial tract, lung, gallbladder, breast, and the urogenital tract (Van Klinken *et al.*, 1998; Copin *et al.*, 2001; Russo *et al.*, 2006; Sonora *et al.*, 2006). *MUC5B* is expressed in high levels in normal and cancerous endometrial and cervical tissues (Hebbar *et al.*, 2005), and there has been an increasing interest in the use of *MUC5B* for the diagnosis of carcinomas.

VNTRs occur in intronic sequences; however, they also occur at exon–intron junctions and in 5′- or 3′-flanking sequences. A characteristic of these repetitive sequences is their ability to give rise to variants that contain increased or decreased numbers of repeats. Some minisatellite alleles are associated with human disorders and with differential expression of nearby genes (Leem *et al.*, 2002; Seol *et al.*, 2008). In addition, rare alleles of this VNTR are associated with a high risk of various types of cancer (Krontiris *et al.*, 1993; Jeong *et al.*, 2007). *MUC5B* has three polymorphic minisatellites (MS3, MS6, and MS7) and four monomorphic minisatellites (MS1, MS2, MS4, and MS5). A search of the GenBank database using the BLASTN program revealed that there was no significant similarity between these minisatellites and previously identified regions. Therefore, all of the minisatellites examined in this study are unique to *MUC5B*, and the properties they confer may be directly related to *MUC5B* function.

Minisatellites in *MUC5B* were also analyzed in genomic DNA obtained from five two-generation families and two three-generation families. *MUC5B* has three polymorphic (MS3, MS6, and MS7) minisatellites, and their segregation within families indicated that these minisatellites were transmitted through meiosis following Mendelian inheritance. These observations suggest that these novel polymorphic minisatellites could also be useful markers for paternity mapping and DNA fingerprinting. Further, these polymorphisms could be useful as markers for meiotic segregation of *MUC5B* minisatellites in studies evaluating *MUC5B*-related inheritable diseases.

Rare alleles of VNTRs are associated with a high risk of various types of cancer (Krontiris *et al.*, 1993; Jeong *et al.*, 2007). In this study, *MUC5B*-MS1, -MS2, -MS4, and -MS5 were found to have monomorphic patterns, but the *MUC5B*-MS3 minisatellites were found to differ in only one of 200 cases of gastric cancer. Although this variant allele of *MUC5B*-MS3 may be associated with an increased relative risk of gastric cancer, the numbers of cases that it was observed here were insufficient to allow statistical analysis. Two additional polymorphic minisatellites, *MUC5B*-MS6 and *MUC5B*-MS7, were identified from the 789 control samples, and these alleles could be separated as common and rare alleles.

A case–control study was performed using PCR-based methods to score *MUC5B* minisatellite alleles in DNA from controls and cases. To accomplish this, each minisatellite allele was grouped into two sets (common and rare alleles) based on their frequency in the control population. There was no difference in the frequencies of alleles from the *MUC5B*-MS3 and -MS7 loci observed between the controls and cases (data not shown). In addition, there were no significant differences observed in the frequency of minisatellite alleles for the *MUC5B*-MS6 when controls and cases were compared. We also divided the rare alleles into short and long alleles according to the number of tandem repeats. Interestingly, the frequency of long rare alleles was found to be associated with a relative odds ratio of 4.54 (CI, 1.0–20.7; $p = 0.03$) for younger patients with bladder cancer. These

results suggest that rare *MUC5B*-MS6 alleles may be genetically related to bladder cancer. To compare other cancer types of the respiratory tract, we analyzed samples from patients with lung cancer. Interestingly, the results of this analysis revealed that the rare *MUC5B*-MS6 alleles may be associated with an increased risk of lung cancer (4.35 [95% CI, 0.8–24.4]) in younger patients. However, due to the low number of patients evaluated in this study, a statistical analysis could not be conducted.

The functional role that *MUC5B*-MS6 minisatellites play is not clear; however, they may be involved in regulation of the expression of *MUC5B*. One characteristic of these repetitive sequences is their ability to produce variants that contain increased or decreased numbers of repeats. Minisatellites in the 3′-region flanking *H-ras* bind the *rel/NF-κB* family of transcription factors, which contribute to the transcriptional activation of *H-ras* (Trepicchio and Krontiris, 1992). Similarly, the *hTERT* VNTR 2-2 contains CACGT-binding sites for the *MYC* family of oncogenic transcription factors (Leem *et al.*, 2002). Further, the *hTERT* VNTR 2-2 gene contains at least six alternative splicing sites, one of which (β site) produces mRNA lacking exons 7 and 8, which results in the production of a protein that is catalytically inactive and has a truncated C-terminal (Yi *et al.*, 2000). When the minisatellite sequences were analyzed using the Transfac software (MATCH™ public version 1.0; http://www.gene-regulation.com/pub/databases.html), we found the putative binding site (ccccCCAGAcccctcg) for the transcription factor Hand1/E47 in the repeat region of MUC5B-MS6. E47 regulates the expression of genes involved in cell survival, cell cycle progression, lipid metabolism, stress response, and lymphoid maturation (Schwartz *et al.*, 2006). Hand1 has been shown to stimulate the transcription of luciferase reporters harboring degenerate E-boxes in the presence of E12/E47, which suggests that E-proteins are potential dimerization partners in trophoblastic tumor and amnion cells (Knöfler *et al.*, 2002). Therefore, interaction with Hand1/E47 may be related to transcriptional activity. However, this study was conducted using the complete *MUC5B* gene, which includes all of the putative regulatory elements, including introns.

The results of this study suggest that the presence of these long rare *MUC5B*-MS6 alleles indicate susceptibility for bladder cancer. *MUC5B*-MS6 is located between the end of exon 37 and intron 37, and only two amino acids of leucine and cystein are included in the repeats of the exon. The results of the present study suggest that extension of the region of repeats may cause the exon to become to be long, thereby extending the length of the protein. This increased susceptibility for bladder cancer may be due to the presence of these long minisatellites because they may result in the construction of inappropriate O-glycosylation structures. Changes in the length of such variants may result in the production of different glycoforms or mucus hypersecretion, thereby leading to carcinomas. Further, alternatively spliced transcript variants of *MUC5B* that include variations between exon 37 and intron 37 may exist. It is possible that VNTRs in introns may also affect mRNA splicing.

Taken together, the results of this study revealed that an increased incidence of the long rare *MUC5B*-MS6 allele is associated with bladder cancer. This is the first study to characterize minisatellites of the complete *MUC5B* gene in detail. This study should provide a helpful reference for understanding the function and complex genomic properties of mucins.

Acknowledgments

We thank Dr. Jae Woo Kim (Dong-A University Hospital) for his help in collecting the blood samples. This work was supported by a grant (FG06-11-06) from the 21C Frontier Functional Human Genome Project, Ministry of Science & Technology, Korea.

Disclosure Statement

No competing financial interests exist.

References

Audie, J.P., Janin, A., Porchet, N., Copin, M.C., Gosselin, B., and Aubert, J.P. (1995). Expression of human mucin genes in respiratory, digestive and reproductive tracts ascertained by *in situ* hybridization. J Histochem Cytochem **41,** 1479–1485.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res **27,** 573–580.

Chakravarti, A., and Lynn, A. (2004). Meiotic mapping in human. In *Genome Analysis: A Laboratory Manual,* vol 4. B. Birren, E.D. Green, S. Klapholz, R.M. Myers, and Roskams, eds. (Cold Spring Harber Laboratory Press, New York), pp. 33–40.

Copin, M.C., Buisine, M.P., Devisme, L., Leroy, X., Escande, F., Gosselin, B., Aubert, J.P., and Porchet, N. (2001). Normal respiratory mucosa, precursor lesions and lung carcinomas: differential expression of human mucin genes. Front Biosci **6,** D1264–D1275.

Corfield, A.P., Carroll, D., Myerscough, N., and Probert, C.S. (2001). Mucins in the gastrointestinal tract in health and disease. Front Biosci **6,** D1321–D1357.

Denoeud, F., Vergnaud, G., and Benson, G. (2003). Predicting human minisatellite polymorphism. Genome Res **13,** 856–867.

Desseyn, J.L., Buisine, M.P., Porchet, N., Aubert, N., Degand, P., and Laine, A. (1998). Evolutionary history of the 11p15 human mucin gene family. J Mol Evol **46,** 102–106.

Desseyn, J.L., Rousseau, K., and Laine, A. (1999). Fifty-nine bp repeat polymorphism in the uncommon intron 36 of the human mucin gene *MUC5B*. Electrophoresis **20,** 493–496.

Gendler, S.J., Lancaster, C.A., Taylor-Papadimitriou, J., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Lalani, E.N., and Wilson, D. (1990). Molecular cloning and expression of human tumor-associated polymorphic epithelial mucin. J Biol Chem **265,** 15286–15293.

Gendler, S.J., and Spicer, A.P. (1995). Epithelial mucin genes. Annu Rev Physiol **57,** 607–634.

Gum, J.R. (1992). Mucin genes and proteins they encode: structure, diversity, and regulation. Am J Respir Cell Mol Biol **7,** 554–564.

Hebbar, V., Damera, G., and Sachdev, G.P. (2005). Differential expression of MUC genes in endometrial and cervical tissues and tumors. BMC Cancer **5,** 124 (1–12).

James, C.B., and Robert, S.B. (2004). Mucins and mucin binding proteins in colorectal cancer. Cancer Metastasis Rev **23,** 77–99.

Jeong, Y.H., Kim, M.C., Ahn, E.K., Seol, S.Y., Do, E.J., Choi, H.J., Chu, I.S., Kim, W.J., Kim, W.J., Sunwoo, Y.I., and Leem, S.H. (2007). Rare exonic minisatellite alleles in *MUC2* influence susceptibility to gastric carcinoma. PLoS ONE **2,** e1163.

Knöfler, M., Meinhardt, G., Bauer, S., Loregger, T., Vasicek, R., Bloor, D.J., Kimber, S.J., and Husslein, P. (2002). Human Hand1 basic helix-loop-helix (bHLH) protein: extra-embryonic expression pattern, interaction partners and identification of its transcriptional repressor domains. Biochem J **361,** 641–651.

Krontiris, T.G., Devlin, B., Karp, D.D., Robert, N.J., and Risch, N. (1993). An association between the risk of cancer and mutations in the *HRAS1* minisatellite locus. N Engl J Med **329,** 517–523.

Kyo, K., Parkes, M., Takei, Y., Nishimori, H., Vyas, P., Satsangi, J., Simmons, J., Nagawa, H., Baba, S., Jewell, D., Muto, T., Lathrop, G.M., and Nakamura, Y. (1999). Association of ulcerative colitis with rare VNTR alleles of the human intestinal mucin gene, *MUC3*. Hum Mol Genet **8,** 307–311.

Leem, S.H., Londoño-Vallejo, J.A., Kim, J.H., Bui, H., Tubacher, E., Solomon, G., Park, J.E., Horikawa, I., Kouprina, N., Barrett, J.C., and Larionov, V. (2002). The human telomerase gene: complete genomic sequence and analysis of tandem repeat polymorphisms in intronic regions. Oncogene **21,** 769–777.

Martínez-Antón, A., Debolós, C., Garrido, M., Roca-Ferrer, J., Barranco, C., Alobid, I., Xaubet, A., Picado, C., and Mullol, J. (2006). Mucin genes have different expression patterns in healthy and diseased upper airway mucosa. Clin Exp Allergy **36,** 448–457.

Moniaux, N., Escande, F., Porchet, N., Aubert, J.P., and Batra, S.K. (2001). Structural organization and classification of the human mucin genes. Front Biosci **6,** D1192–D1206.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, T., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., *et al.* (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping. Science **235,** 1616–1622.

Pigny, P., Guyonnet-Dupe'rat, V., Hill, A.S., Pratt, W.S., Galiegue-Zouitina, S., d'Hooge, M.C., Laine, A., Van-Seuningen, I., Degand, P., Gum, J.R., Kim, Y.S., Swallow, D.M., Aubert, J.P., and Porchet, J.N. (1996). Human mucin genes assigned to 11p15.5: identification and organization of a cluster of genes. Genomics **38,** 340–352.

Richard, G.F., and Paques, F. (2000). Mini- and microsatellite expansions: the recombination connection. EMBO Rep **1,** 122–126.

Rozen, S., and Skaletsky, H.J. (2000). Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. S. Krawetz and S. Misener, eds. (Humana Press, Totowa, NJ), pp. 365–386.

Russo, C.L., Spurr-Michaud, S., Tisdale, A., Pudney, J., Anderson, D., and Gipson, I.K. (2006). Mucin gene expression in human male urogenital tract epithelia. Hum Reprod **21,** 2783–2793.

Schwartz, R., Engel, I., Fallahi-Sichani, M., Petrie, H.T., and Murre, C. (2006). Gene expression patterns define novel roles for E47 in cell cycle progression, cytokine-mediated signaling, and T lineage development. Proc Natl Acad Sci USA **103,** 9976–9981.

Seol, S.Y., Lee, S.Y., Kim, Y.D., Do, E.J., Kwon, J.A., Kim, S.I., Chu, I.S., and Leem, S.H. (2008). Minisatellite polymorphisms of the *SLC6A19*: susceptibility in hypertension. Biochem Biophys Res Commun **374,** 714–719.

Sonora, C., Mazal, D., Berois, N., Buisine, M.P., Ubillos, L., Varangot, M., Barrios, E., Carzoglio, J., Aubert, J.P., and Osinaga, E. (2006). Immunohistochemical analysis of *MUC5B* apomucin expression in breast cancer and non-malignant breast tissues. J Histochem Cytochem **54,** 289–299.

Thornton, D.J., Rousseau, K., and McGuckin, M.A. (2008). Structure and function of the polymeric mucins in airways mucus. Annu Rev Physiol **70,** 459–486.

Trepicchio, W.L., and Krontiris, T.G. (1992). Members of the rel/NF-kappa B family of transcriptional regulatory proteins bind the HRAS1 minisatellite DNA sequence. Nucleic Acids Res **20,** 2427–2434.

Troxler, R.F., Offner, G.D., Zhang, F., Iontcheva, I., and Oppenheim, F.G. (1995). Molecular cloning of a novel high molecular weight mucin (MG1) from human sublingual gland. Biochem Biophys Res Commun **217,** 1112–1119.

Van Klinken, B.J., Dekker, J., van Gool, S.A., van Marle, J., Büller, H.A., and Einerhand, A.W. (1998). *MUC5B* is the prominent mucin in human gallbladder and is also expressed in a subset of colonic goblet cells. Am J Physiol **274,** G871–G878.

Vinall, L.E., Hill, A.S., Pigny, P., Pratt, W.S., Toribara, N., Gum, J.R., Kim, Y.S., Porchet, N., Aubert, J.P., and Swallow, D.M. (1998). Variable number tandem repeat polymorphism of the mucin genes located in the complex on 11p15.5. Hum Genet **102,** 357–366.

Wickstrom, C., Davies, J.R., Eriksen, G.V., Veerman, E.C., and Carlstedt, I. (1998). *MUC5B* is a major gel-forming, oligomeric mucin from human salivary gland, respiratory tract and endocervix: identification of glycoforms and C-terminal cleavage. Biochem J **334,** 685–693.

Yi, X., White, D.M., Aisner, D.L., Baur, J.A., Wright, W.E., and Shay, J.W. (2000). An alternate splicing variant of the human telomerase catalytic subunit inhibits telomerase activity. Neoplasia **2,** 433–440.

Address reprint requests to:
Sun-Hee Leem, Ph.D.
Department of Biological Science
Dong-A University
Busan 604-714
Korea

E-mail: shleem@dau.ac.kr

Received for publication October 23, 2008; received in revised form December 2, 2008; accepted December 4, 2008.