Genomics & Informatics Vol. 3(3) 74-79, September 2005

Single Nucleotide Polymorphisms of a 16 kb Region on Human Chromosome 11p15.5 that Includes the H19 Gene

Mi-Hyun Park, Hyeon-Jeong Ku, Hye-Ja Lee, Kwang-Joong Kim, Chan Park, Bermseok Oh, Ku-Chan Kimm, and Jong-Young Lee*

National Genome Research Institute, National Institute of Health, 5 Nokbun-dong, Eunpyung-gu, Seoul 122-701, Korea

Abstract

The H19 gene, located at human chromosome 11p15.5, is imprinted in most normal human tissues. However, imprinting is often lost in tumors suggesting H19 is a putative tumor suppressor. We analyzed the single nucleotide polymorphisms (SNPs) of a 16 kb region that includes the H19 gene and its imprinting control region (ICR) in the Korean population. To identify SNPs, we directly sequenced this region in 18 Korean subjects. We identified 64 SNPs, of which 7 were in the exons of H19, 2 were in the introns, 14 were in the 3' intergenic region and 41 were in the 5' intergenic region. Of the 64 SNPs, 21 had not previously been reported and thus appear to be unique to the Korean population. The identified SNPs of H19 in the Korean population may eventually be useful as genetic markers associated with various diseases. In this study, 7 of the 64 identified SNPs were at CTCF binding sites in the ICR and may affect regulation of H19 gene imprinting. Thus, several genetic variations of the H19 gene may be important markers in human diseases that involve genomic imprinting, including cancer.

Keywords: H19, Chr11, SNP, Haplotype, Imprinting, CTCF

Introduction

The human H19 gene consists of five exons separated by four introns. Early studies mapped it to 11p15.5, a chromosomal region with known tumor suppressor activity, and subsequent analysis revealed that H19 expression is often altered in several different types of tumors relative to that in adjacent, non-transformed cells (Chung *et al.*, 1996; Cui *et al.*, 1997). The human sequence has a putative ORF that appears to encode a 26 kDa protein but the gene product appears to be RNA rather than a protein (Hao *et al.*, 1993; Joubel *et al.*, 1996; Pfeifer *et al.*, 1996). The H19 gene is a known imprinted gene. It is expressed exclusively from the maternal chromosome and is linked and co-regulated with the insulin-like growth factor 2 gene, IGF2, which is also imprinted but expressed primarily from the paternal chromosome (Bell *et al.*, 2000).

A 2 kb region of differential methylation located from -2 to -4 kb relative to the H19 transcriptional start site has been proposed to act as the imprinting mark (Davis *et al.*, 1999). A differentially methylated region (DMR) upstream of H19 has been proposed to participate in the imprinting control of human IGF2 and H19 (Thorvaldsen, 1998). The DMR, or imprinting control region (ICR), acts as an epigenetic modifier of allelic expression by recruiting proteins that specifically bind to methylated or unmethylated DNA (Ulaner *et al.*, 2003a). In human, the ICR contains seven specific binding sites for the zinc finger insulator protein CTCF (CCCTC-binding factor) (Hark *et al.*, 2000). CTCF binds to several sites in the unmethylated ICR that are essential for enhancer blocking (Hark *et al.*, 2000). The methylation status of binding sites for CTCF in the H19 promoter is suggested to be involved in the regulation of imprinting of the H19/IGF2 locus (Takai *et al.*, 2001). Only the sixth of the seven CTCF-binding sites has been demonstrated to have allele-specific differential methylation (Takai *et al.*, 2001). This region contains well characterized single nucleotide polymorphisms (SNPs) which make it possible to distinguish the paternal and maternal alleles (Takai *et al.*, 2001). Several studies have used two or three SNPs within the ICR to discriminate between the two alleles to investigate whether tumors have abnormal H19/IGF2 methylation (Nakagawa *et al.*, 2001; Ulaner *et al.*, 2003a; Ulaner *et al.*, 2003b).

In this study, we analyzed SNPs in a 16 kb region that includes the H19 gene and its 2 kb ICR in the Korean population. Our results provide new information about several SNPs that may impact imprinting regulation.

Materials and Methods

Extraction of genomic DNA and PCR

Peripheral blood samples were obtained from 18 healthy adults (8 males and 10 females) aged 25-43 years at the National Genome Research Institute in Korea. Genomic

*Corresponding author: E-mail leejy63@nih.go.kr, Tel + 82-2-380-2259, Fax +82-2-354-1063
Accepted 27 May 2005

Table 1. Primers for sequencing of the H19 region on chromosome 11p15.5

Primer	Primer Sequence and length		Product (bp)
H19-05F	CTTCCTTCGTAGAGTTGGGATTT	23	1843
H19-05R	CATTCATCTTCCTGACCCTACAG	23	
H19-06F	ACATGTAATGGATGGGTTACTTCAT	25	1909
H19-06R	TTCTGTCTCTTTGTTTCTGAGCTTT	25	
H19-07F	AAAGCTCAGAAACAAAGAGACAGAA	25	1724
H19-07R	AGGAGAGTTAGCAAAGGTGACATC	24	
H19-08F	GAAGATGTCACCTTTGCTAACTCTC	25	2330
H19-08R	CTAAGTGGCCAGACATTAACATTCT	25	
H19-9F	AGAATGTTAATGTCTGGCCACTTAG	25	2157
H19-9R	GTGTTGGCTGAGATAATCTGCTAAT	25	
H19-10F	ACATTAGCAGATTATCTCAGCCAAC	25	1227
H19-10R	CAGAGATAGACACTCATGCGACTAA	25	
H19-11F	TTAGTCGCATGAGTGTCTATCTCTG	25	1128
H19-11R	ATCTGTGGCTAATTTGTTAGTCCTG	25	
H19-12F	CAGGACTAACAAATTAGCCACAGAT	25	1841
H19-12R	GAAGAGAAAGCCCCTCTATCTAATG	25	
H19-13F	CCTGGGAGACTGATTTGAGTAATAA	25	1988
H19-13R	AGTTCCTCATGGAGTCTGTTTAGTG	25	
H19-14F	ACTAAACAGACTCCATGAGGAACTG	25	1653
H19-14R	CTGAACATCTTATTTTGCTCCATTT	25	
H19-i05F	TCAAGTCCTCTTCTCTAACTGGATG	25	1107
H19-i05R	AGAAAACCTCTCCTGGGGAAC	21	
H19-i06F	CTTCCTGTCTGTTCTTTGTCAGTC	24	1103
H19-i06R	CCTGGACTCATCATCAATAAACACT	25	
H19-i07F	CAGCATATTCATTTCCAAGCTAGAG	25	1134
H19-i07R	CAATTCATTTAGTAGCAGGCACAG	24	
H19-i08F	CATCCTGGAATTCTCCAAAGAC	22	1166
H19-i08R	GTATTTCCGGAGTATTTCCTGAGTC	25	
H19-i09F	CTCATTGTTCCCCTAGTATCTCCTC	25	1068
H19-i09R	GAATCAGTTGAAGGTATGGAAACAC	25	
H19-i14F	TTATAGGAAGAGGAGAGGAGGACAC	25	985
H19-i14R	GTCTCTTACCACTTTATGCTGCACT	25	

DNA was extracted from blood with a QIAamp DNA Blood Kit (QIAGEN, Valencia, CA) according to the manufacturer's protocol. The primer sets for the PCR walking analysis are shown in Table 1. The PCR reactions contained 20 ng genomic DNA, 0.2 units of AmpliTaq Gold (Perkin-Elmer, Weiterstadt, Germany) or LA taq (Takara Bio, Otsu, Japan) polymerase, 1 mM dNTPs, 1×PCR buffer, 5 pmol sense primer and 5 pmol antisense primer in 20 $\mu\ell$ and were performed in a thermal cycler (Perkin-Elmer) at the following temperatures: 95°C for 5 min, 30 cycles at 95°C for 30 sec. at 60~68°C for 45 sec and at 72°C for 3 min, followed by 72 °C for 10 min. To check the quality of the PCR products before the sequencing reactions, one-tenth of the reaction mixture was separated by electrophoresis on a 1% agarose gel.

Sequencing

The primers used for sequencing are the same as those used for PCR (Table 1). Cycle sequencing reactions were performed using PCR product and an Applied Biosystems (Foster City, CA) Big Dye Terminator (version 2.0) ready reaction kit. The amount of primer in the reaction was 1 pmol, and total reaction volume was 10 $\mu\ell$. Cycling parameters were: 30 cycles of 30 sec at 95 °C, 5 sec at 50 °C, 4 min at 60°C, followed by refrigeration until use. Each reaction mixture was ethanol-precipitated to remove excess dye terminators. The pellets were dissolved in 11 $\mu\ell$ template suppression reagent (TSR, Applied Biosystems), heated at 95 °C for 4 min to denature, quenched on ice for 4 min, mixed, spun briefly, loaded into the autosampler tray of an ABI3100 automated DNA sequencer, and sequenced according to the ABI3100 operator's manual. The sequencer was set up to run using POP6 polymer and a 36 cm capillary with 30 sec injection time and 120 min run time.

Identification of SNPs and haplotype analysis

Polymorphisms were detected by multiple alignments of sequences using the Phred/Phrap/ Consed package (Ewing *et al.* 1998; Gordon *et al.* 1998). Deviation from Hardy-Weinberg expectancy was examined with x^2 or Fisher's exact test.

Haploview software (version 32) was used for the analysis of haplotype and tagging SNP detection (Barrett, 2005).

Results and Discussion

We discovered 64 SNPs in the H19 region (16 kb; NT_009237) on chromosome 11p15.5 in the Korean population (n=18). This region contains the H19 gene (1.4 kb), the H19 promoter region (2 kb), an imprinting control region (ICR) including CTCF binding sites (2 kb), and an intergenic region (~10 kb). This region was amplified from samples from 18 normal Koreans by using PCR primers (Table 1). The PCR products were directly sequenced to analyze the SNPs. The dbSNP database in the National Center for Biotechnology Information (NCBI) has 93 SNPs in the H19 region. Of the 64 SNPs that we identified in this region, 43 were identical to SNPs in the NCBI database. The remaining 21 were thus judged to be novel in the Korean

Table 2. SNPs and allele frequencies of H19 gene and its flanking intergenic region in the Korean population (acc. No. NT_009237.16).

Position in NT_009237.16	Region	Sequence	Allele frequency	NCBI dbSNP acc. No.
778469	3' intergenic region	TCATACACCC C T GGGCCCAGAG	T:C=0.27:0.73	3214201
778470	3' intergenic region	CATACACCCC - T GGGCCCAGAG	T/T= 2(A5,B7): -/-= 16	3214201
778573	3' intergenic region	CGAGGTGGCC C T GCCTTTTCCC	T:C=0.18:0.82	
778586	3' intergenic region	TTTTCCCCC G A GCCCCATTCC	G:A=0.74:0.26	2285935
778649	3' intergenic region	TTCCATGTTC T C CCAGGAGAGG	T:C=0.56:0.44	217228
779192	3' intergenic region	GACCCAAGGG A G CGTGGGGAAC	A:G=0.64:0.36	217229
779447	3' intergenic region	GCGAGGGGCC A G GGAGCTGGGA	A:G=0.33:0.67	217230
779509	3' intergenic region	AAGCTGACCC C G GAACCCATGA	C:G=0.47:0.53	217231
779609	3' intergenic region	GAGCCCGGGA G T CCCCTCACCC	G:T=0.92:0.08	
779668	3' intergenic region	ACCCTCAACT C T CTTCCTCCCT	C:T=0.39:0.61	217232
779754	3' intergenic region	CGGCTCCCCC A C CAACCCCCTG	A:C=0.75:0.25	217233
779775	3' intergenic region	GCTCCTTGTG C T GGGGCCGTCC	C:T=0.64:0.36	217728
780027	3' intergenic region	TGGTTTGGTG A G TCCCCGACCC	A:G=0.19:0.81	
780267	3' intergenic region	CCTCCCCATG C T CCCCCAGCCC	C:T=0.86:0.14	11564741
780559	H19 Exon5	CGGGCCCTGC A G CAGGCACTTG	G:A=0.19:0.81	3741219
780599	H19 Exon5	CGCACACTCG T C ACTGAGACTC	T:C=0.78:0.22	2839704
780602	H19 Exon5	ACACTCGTAC T C GAGACTCAAG	T:C=0.78:0.22	2839703
780848	H19 Exon5	TCAACCGTCC A G CCGCAGGGGG	A:G=0.33:0.67	217727
780864	H19 Exon5	GGGGGTGGCC A G TGAAGATGGA	A:G=0.25:0.75	10840159
780957	H19 Exon5	TGGTTCCTCT A C GCTTCACCTA	A:C=0.78:0.22	2839702
781050	H19 Exon4	GCTCACCTTC C G AGAGCCGATT	C:G=0.78:0.22	2839701
781379	Intron2	AGACAGAGGT C G GGGGCAGTGA	G:C=0.39:0.61	3024270
781404	Intron2	GTGGCCCGGC G A GGGCGGTCTG	G:A=0.86:0.14	3825028
782108	Promoter	CTCCTGGTGA C T GTCCTGCTGC	T:C=0.17:0.83	2067051
782276	Promoter	GTCCCCTGCC A T TGTCCCTGTC	T:A=0.25:0.75	2075745
782664	Promoter	CAGCTCCCCT C T GTCCAACCAG	C:T=0.94:0.06	
782793	Promoter	ATGCCTGGGC A G CCTACTCCAC	A:G=0.25:0.75	2839698
783436	Promoter	ACGGCTCAGA A C CTCACGTTCC	C:A=0.53:0.47	2251375
783558	Promoter	AGCTTGTGCT C G GTCACCGCGG	G:C=0.11:0.89	2251312
783935	ICR	GTCCAATCAC C G CAGCAGGCCC	G:C=0.36:0.64	2158394
784496	ICR	TACCGCCGGC C T GGACGTGATG	T:C=0.03:0.97	
784567	ICR	GGATCCCCAC C A CCGTACCGAG	C:A=0.83:0.17	2071095
784895	ICR	GGAACACACT C G TGATCATCAC	C:G=0.24:0.76	4930098
785015	ICR	GTGCGGCTCC C T ATGAGTGTCC	C:T=0.53:0.47	2107425
785104	ICR	AACCCCAGTT G T GGGCGGGCTC	G:T=0.76:0.24	2171094
785146	ICR	CACTGCCGCC A G CGCGGCCACT	A:G=0.24:0.76	10732516
785920	5' intergenic region	CCAGCTGGGG C T GGGCTCGGGC	T:C=0.31:0.69	2525883
785963	5' intergenic region	CCGCCACGCG G T CCACTTCCGA	T:G=0.39:0.61	10840167
786154	5' intergenic region	ACTTCACTGT C T CCCCAAGTAT	T:C=0.17:0.83	
786997	5' intergenic region	TAATGTCTGG C T CACTTAAGAT	C:T=0.89:0.11	
787206	5' intergenic region	CAGAGTGTGA C T CTGGGGCCAC	C:T=0.94:0.06	
787914	5' intergenic region	TGTGCCTGCC A G GAGACCCTGT	A:G=0.72:0.28	
787947	5' intergenic region	ACGCTTAGGC C T GAGATCCCTA	C:T=0.83:0.17	
788294	5' intergenic region	GGGCCGCCCT C T GGTGAGGCTG	C:T=0.14:0.86	2525885
788391	5' intergenic region	GCTGGAGGCC C T TGTTGGGAGG	C:T=0.61:0.39	2525886
788484	5' intergenic region	TTGCCTCCAA A G TATCCACTGT	A:G=0.44:0.56	4930103
788499	5' intergenic region	CACTGTCCTA C T GTGCTATGGG	C:T=0.94:0.06	
788519	5' intergenic region	GAGCTGTCAT C T GGTGGTGTCC	C:T=0.64:0.36	4929983
788972	5' intergenic region	CAAAAAAAAA A T AAATATATAA	A:T=0.56:0.44	
788974	5' intergenic region	AAAAAAAAAA A T ATATATAACA	A:T=0.56:0.44	
789725	5' intergenic region	CACTGATTCC C T GGAGGAAGAT	C:T=0.92:0.08	
789927	5' intergenic region	GTTTTTTTTG G T TTGTTTTTTT	G:T=0.58:0.42	
791004	5' intergenic region	GCTTTCTCTT C T TCTCTCCCCT	T:C=0.53:0.47	2525887
791029	5' intergenic region	CTGGGACCTA C T CCAGATATAA	T:C=0.06:0.94	
791706	5' intergenic region	CTCCAGAGCC A G GTGTGTGGGC	A:G=0.11:0.89	12292757
791881	5' intergenic region	CAGACACACA C T GGAGGGATGA	C:T=0.92:0.08	
792002	5' intergenic region	AGGACTGGGA A G AGGATATATT	A:G=0.5:0.5	11042194
792041	5' intergenic region	CGGTCCATGG C T GCTTTGTTGC	C:T=0.83:0.17	
792083	5' intergenic region	CGGCTGCAAC A G TGACCAGTGG	A:G=0.36:0.64	3890907
792086	5' intergenic region	CTGCAACATG A G CCAGTGGGTG	A:G=0.92:0.08	
792230	5' intergenic region	TCCTGGGCCA T C GGGTGTCTGT	C:T=0.75:0.25	10840180
792531	5' intergenic region	TCTGTCTTCA C G ACGGCCTCCT	G:C=0.25:0.75	7396803
792969	5' intergenic region	GCCCCAGAGC C T GGGCAGTAGC	C:T=0.89:0.11	
794387	5' intergenic region	CATCTTGGCA A G AAGGTCCGAT	A:G=0.03:0.97	

Table 3. Haplotype frequencies in the H19 gene region.

Haplotype		Frequency
	▼▼ ▼▼ ▼▼▼ ▼ ▼▼▼	
Type 1	ATTAAACCGCACGACGCGTGG	21.50%
Type 2	ATTGAACCGCACGCCCCGCGG	11.40%
Type 3	ATTGAACCGCACGACCCGTGG	11.20%
Type 4	ATTGAACGACACGACCCGTGG	8.30%
Type 5	GCCGGCGGGTTCACCCACCTA	5.60%
Type 6	ATTGAACGGCACGCGCCGCGG	5.60%
Type 7	ATTGAACCGCACGCCGCGCGG	5.30%
Type 8	GCCGGCGGGTTTACCCCCCTA	5.10%

▼ : tagging SNP

Table 4. Common haplotype frequencies in different populations

Haplotype		Korean	Japanese	Chinese	CEPH
Type 1	GACG	47.20%	46.50%	44.40%	30.00%
Type 2	GCCA	19.40%	18.90%	18.90%	1.70%
Type 3	ACAA	16.70%	29.50%	24.40%	54.20%
Type 4	GCCG	8.30%	3.90%	12.20%	14.20%
Type 5	ACCA	8.30%	0%	0%	0%

population. Of the 64 SNPs, 7 were in exons and 2 were in introns (none of which were novel), 6 were in the promoter (1 of which was novel), 7 were in the ICR (1 novel), 28 were in the 5' intergenic region (16 novel), and 14 were in the 3' intergenic region (3 novel) (Table 2).

We investigated the haplotypes of a 5.4 kb region (Position: 780559~785146, Table 2) with 21 SNPs containing the H19 gene, promoter and ICR. The position 784496 SNP (minimal allele frequency < 5%) was omitted. The patterns of haplotype structure and frequencies in this 5.4 kb gene region are shown in Table 3. There were 8 common haplotypes (frequency ≥ 5%) and 11 tagging SNPs were found. We also selected 4 SNPs (rs2839698, rs2251375, rs2071095, rs4930103) that were present in both our data and the HapMap database (The International HapMap Consortium., 2004) and analyzed the haplotype diversity and frequency (Table 4). The frequencies of haplotype 1~4 in the Korean population were similar with those in the Japanese and Chinese populations but different from the frequency in the CEPH population (Utah residents with ancestry from northern and western Europe). Haplotype 5 was the common haplotype in the Korean population, but was not found in the Japanese, Chinese, and CEPH populations (Table 4). Thus we find that several SNPs and haplotypes in the Korean population differ from those in other ethnic groups.

We only found 64 SNPs in this region compared with the 93 in the NCBI database. This may be due to our small sample size (n=18). Detection of 89% of polymorphic sites (minimal allele frequency of 5%) would require n=16, and detection of 99% would require n=48 (Kruglyak *et al.*, 2001). In our case, with n=18, we expect to detect over 99% of SNPs with a minimal allele frequency of 20%. Thus we do not expect to detect low frequency SNPs in our sample. In addition this discrepancy in SNP detection raises the possibility of invalid or extremely low frequency SNPs in the public databases and underlines the need to check the validity and frequency of any potential marker SNPs in an intended study population.

The ICR of H19 is methylated from -2 to -4 kb relative to the start of transcription. CTCF binding was recently shown to play a direct role in inhibiting methylation of the ICR in order to establish and maintain imprinting of the Igf2/H19 region, and CTCF binding sites were shown to be dispensable for initiating imprinting (Szabo *et al.*, 2004). The methylation status of this region was ascertained by bisulfite conversion and methylation-specific PCR (Ulaner *et al.*, 2003b). The paternal allele is methylated and the maternal allele is unmethylated. The SNPs allow distinction of the paternal and maternal alleles. Of the six SNPs in the ICR that were previously reported, two (785015 (C/T) and 785104 (G/T)) were previously analyzed for allelic expression based on the methylation status (Ulaner *et al.*, 2003b). Bisulfite DNA sequencing revealed a C/T polymorphism at base 785015 (an A/G polymorphism at base 6325; accession no. AF087017). The allele containing the A polymorphism was methylated, whereas the allele containing the G polymorphism was unmethylated. In addition, Poon *et al.* (2002) reported that paternally inherited methylated H19 fetal alleles were different from the methylated alleles of the respective mothers using a C/T polymorphism at base 785015. This raised the possibility of using epigenetic markers for the specific detection of fetal DNA in maternal plasma. Furthermore, only the sixth CTCF binding site showed allele-specific methylation (Takai *et al.*, 2001). The 785146 (A/G) site is

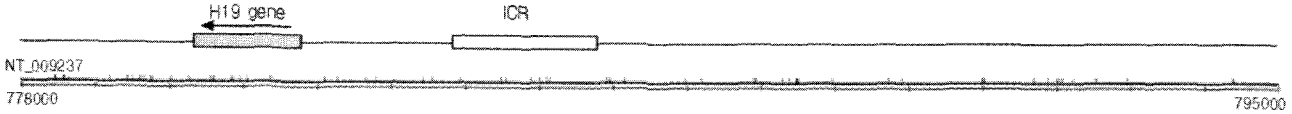

Fig. 1. Map of the human H19 region. It contains the H19 gene (1.4 kb), a promoter region (2 kb), an imprinting control region (ICR, 2 kb), and an intergenic region (~10 kb). On the bottom line the single nucleotide polymorphisms are represented by small vertica lines. The region examined in this report has 64 polymorphic sites (H19 gene; 9 sites, ICR; 7 sites).

located at the sixth CTCF binding site. This SNP site may act as a marker for regulating H19 expression. Nakayashiki *et al.* (2004) recently described three closely located SNPs (g/a g7523a, g/a g7547a, c/t c7591t; accession no. AF125183) and designated them as the H19FR haplotype. They were able to selectively discriminate the parental alleles by enzymatically digesting differentially methylated genomic DNA. This method could be useful for identifying the parental origin of alleles. However, these three SNPs were not present in the Korean population. In any case, SNPs located in the ICR should provide information about the methylation status determined by expression of H19 alleles. In addition, information on the methylation status of the H19 gene may help to understand how imprinting is disrupted in tumors.

Numerous studies have revealed abnormal imprinting of H19 in a wide range of tumors (Nakagawa *et al.*, 2001; Ulaner *et al.*, 2003a; Manoharan *et al.*, 2003; Yin *et al.*, 2004). Loss of imprinting (LOI) of IGF2 correlated strongly with biallelic hypermethylation in a core region of an H19-associated CTCF-binding site (Nakagawa *et al.*, 2001). The presence of this methylation-dependent LOI in both tumors and normal colonic mucosa indicates that hypermethylation may create a field defect predisposing to cancer (Nakagawa *et al.*, 2001). In addition, incomplete gain or loss of methylation at this CTCF- binding site during tumorigenesis can explain the complex and conflicting expression patterns of IGF2 and H19 in a tumor (Ulaner *et al.*, 2003a). Manoharan *et al.* (2003) used a SNP marker in the H19 coding sequence to investigate H19 imprinting. They found monoallelic expression of the maternal gene in fetal liver, but biallelic expression in liver neoplasms, thus demonstrating the basis for the deregulation of the imprinted gene expression during hepatocarcinogenesis. The imprinting of IGF2/H19 is similar to DLK1/GTL2, another reciprocally imprinted gene pair. The imprinting status of DLK1 in brain tumors and lymphomas has also been deduced by analysis of a SNP (Yin *et al.*, 2004). We propose that the polymorphic sites of H19 are good genomic markers for imprinting studies of additional tumor types.

H19 was one of the first imprinted genes to be identified in mice and humans. It is an excellent model for studies of genomic imprinting because it is a representative imprinting gene in human and other animals. In the mouse, monoallelic expression of H19 is regulated by an ICR located at chromosome 7 (Thorvaldsen *et al.*, 1998) and CTCF binding has been shown to have a role at four sites in the IGF2/H19 ICR (Szabo *et al.*, 2004; Schoenherr *et al.*, 2003; Thorvaldsen *et al.*, 2002). A mutation in the mouse CTCF site 4 was sufficient to cause robust activation of the maternal Igf2 allele and to disturb the methylation-free status of the maternal H19 ICR allele (Pant *et al.*, 2004). H19 is also imprinted in cattle, in which the maternal allele was found to be predominantly or exclusively expressed in all tissues examined (Zhang *et al.*, 2004). Identification of a SNP in the bovine H19 gene made it possible to study its imprinting status by following the expression of parental alleles in heterozygous animals. The present results will help to analyze sequences and SNPs of the H19 region in other mammals (e.g., pig and sheep). The SNPs that we identified can be used as markers for the study of developmental abnormalities. The SNPs of the H19 gene discovered in this study should be useful as genomic markers for imprinting studies. Particularly, they should be useful as markers in types of tumors that involve genomic imprinting. We are planning a case-control study of cancer using the SNPs discovered in the Korean population.

Acknowledgements

This work was supported by intramural grants from the Korea National Institute of Health, Korea Center for Disease Control, Republic of Korea (348-6111-221).

References

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263-265.

Bell, A.C. and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* 25, 405, 482-485.

Chung, W.Y., Yuan, L., Feng, L., Hensle, T., and Tycko, B. (1996). Chromosome 11p15.5 regional imprinting: comparative analysis of KIP2 and H19 in human tissues and Wilms' tumors. *Hum. Mol. Genet.* 5, 1101-1108.

Cui, H., Hedborg, F., He, L., Nordenskjold, A., Sandstedt, B., Pfeifer-Ohlsson, S., and Ohlsson, R. (1996). Inactivation of H19, an imprinted and putative tumor repressor gene, is a preneoplastic event during Wilms' tumorigenesis. *Cancer Res.* 15, 4469-4473.

Davis, T.L., Trasler, J.M., Moss, S.B., Yang, G.J., and Bartolomei, M.S. (1999). Acquisition of the H19 methylation imprint occurs differentially on the parental alleles during spermatogenesis. *Genomics* 15, 18-28.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8, 175-185.

Ewing, B. and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 8, 186-194.

Gordon, D., Abajian, C., and Green, P. (1998). Consed: a graphical tool for sequence finishing. *Genome Res.* 8, 195-202.

Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., and Tilghman, S.M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* 25, 405, 486-489.

Hao, Y., Crenshaw, T., Moulton, T., Newcomb, E., and Tycko, B. (1993). Tumour-suppressor activity of H19 RNA. *Nature* 21, 764-767.

Joubel, A., Curgy, J.J., Pelczar, H., Begue, A., Lagrou, C., Stehelin, D., and Coll, J. (1996). The 5' part of the human H19 RNA contains cis-acting elements hampering its translatability. *Cell Mol. Biol.* 42, 1159-1172.

Kruglyak, L. and Nickerson, D.A. (2001). Variation is the spice of life. *Nature Genet.* 27, 234-236.

Manoharan, H., Babcock, K., Willi, J., and Pitot, H.C. (2003). Biallelic expression of the H19 gene during spontaneous hepatocarcinogenesis in the albumin SV40 T antigen transgenic rat. *Mol. Carcinog.* 38, 40-47.

Nakayashiki, N., Kanetake, J., and Aoki, Y. (2004). A parent-of-origin detectable polymorphism in the hypermethylated region upstream of the human H19 gene. *Int. J. Legal Med.* 118, 158-162.

Nakagawa, H., Chadwick, R.B., Peltomaki, P., Plass, C., Nakamura, Y., and de La Chapelle, A. (2001). Loss of imprinting of the insulin-like growth factor II gene occurs by biallelic methylation in a core region of H19-associated CTCF- binding sites in colorectal cancer. *Proc. Natl. Acad Sci. USA* 16, 591-596.

Pant, V., Kurukuti, S., Pugacheva, E., Shamsuddin, S., Mariano, P., Renkawitz, R., Klenova, E., Lobanenkov, V., and Ohlsson, R. (2004). Mutation of a single CTCF target site within the H19 imprinting control region leads to loss of Igf2 imprinting and complex patterns of de novo methylation upon maternal inheritance. *Mol. Cell Biol.* 24, 3497-3504.

Poon, L.L., Leung, T.N., Lau, T.K., Chow, K.C., and Lo, Y.M. (2002). Differential DNA methylation between fetus and mother as a strategy for detecting fetal DNA in maternal plasma. *Clin. Chem.* 48, 35-41.

Pfeifer, K., Leighton, P.A., and Tilghman, S.M. (1996). The structural H19 gene is required for transgene imprinting. *Proc. Natl. Acad. Sci. USA* 26, 13876-13883.

Szabo, P.E., Tang, S.H., Silva, F.J., Tsark, W.M., and Mann, J.R. (2004). Role of CTCF binding sites in the Igf2/H19 imprinting control region. *Mol. Cell Biol.* 24, 4791-4800.

Schoenherr, C.J., Levorse, J.M., and Tilghman, S.M. (2003). CTCF maintains differential methylation at the Igf2/H19 locus. *Nat. Genet.* 33, 66-69.

The International HapMap Consortium (2004). Integrating ethics and science in the International HapMap Project. *Nature Reviews Genetics* 5, 467-475.

Thorvaldsen, J.L., Mann, M.R., Nwoko, O., Duran, K.L., and Bartolomei, M.S. (2002). Analysis of sequence upstream of the endogenous H19 gene reveals elements both essential and dispensable for imprinting. *Mol. Cell Biol.* 22, 2450-2462.

Takai, D., Gonzales, F.A., Tsai, Y.C., Thayer, M.J., and Jones, P.A. (2001). Large scale mapping of methylcytosines in CTCF-binding sites in the human H19 promoter and aberrant hypomethylation in human bladder cancer. *Hum. Mol. Genet.* 1, 10, 2619-2626.

Thorvaldsen, J.L., Duran, K.L., and Bartolomei, M.S. (1998) Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. *Genes Dev.* 1, 3693-3702.

Ulaner, G.A., Vu, T.H., Li, T., Hu, J.F., Yao, X.M., Yang, Y., Gorlick, R., Meyers, P., Healey, J., Ladanyi, M., and Hoffman, A.R. (2003a). Loss of imprinting of IGF2 and H19 in osteosarcoma is accompanied by reciprocal methylation changes of a CTCF-binding site. *Hum. Mol. Genet.* 1, 535-549.

Ulaner, G.A., Yang, Y., Hu, J.F., Li, T., Vu, T.H., and Hoffman, A.R. (2003b). CTCF binding at the insulin-like growth factor-II (IGF2)/H19 imprinting control region is insufficient to regulate IGF2/H19 expression in human tissues. *Endocrinology* 144, 4420-4426.

Yin, D., Xie, D., De Vos, S., Liu, G., Miller, C.W., Black, K.L., and Koeffler, H.P. (2004). Imprinting status of DLK1 gene in brain tumors and lymphomas. *Int. J. Oncol.* 24, 1011-1015.

Zhang, S., Kubota, C., Yang, L., Zhang, Y., Page, R., O'Neill, M., Yang, X., and Tian, X.C. (2004). Genomic imprinting of H19 in naturally reproduced and cloned cattle. *Biol. Reprod.* 71, 1540-1544.