RESEARCH ARTICLE

Effect of splice-site polymorphisms of the *TMPRSS4, NPHP4* and *ORCTL4* genes on their mRNA expression

HIDETAKA YAMADA[1,2], KAZUYA SHINMURA[1], TOSHIHIRO TSUNEYOSHI[2] and HARUHIKO SUGIMURA[1*]

[1]*First Department of Pathology, Hamamatsu University School of Medicine, 1-20-1 Handayama, Hamamatsu, Shizuoka 431-3192, Japan*
[2]*Department of Materials and Life Science, Shizuoka Institute of Science and Technology, 2200-2 Toyosawa, Fukuroi, Shizuoka 437-8555, Japan*

Abstract

Genetic polymorphisms associated with structural changes of their gene product are important in terms of their potential relation with diseases. Therefore, in this study, splice-site variants of the transmembrane serine protease gene *TMPRSS4*, nephronophthisis gene *NPHP4*, and organic-cation transporter gene *ORCTL4*, were selected from the dbSNP single nucleotide polymorphism database as candidates to identify genetic polymorphisms associated with a structural change in their mRNA transcripts. The allele frequencies of the *TMPRSS4* c.4-7A>G, *NPHP4* c.2818-2A>T, and *ORCTL4* c.517-2A>C polymorphisms in a Japanese population were determined to be 0.42, 0.10, and 0.27, respectively, by PCR-SSCP analysis. Next, the effect of these polymorphisms on the mode of pre-mRNA splicing was investigated by RT-PCR analysis followed by sequencing analysis. The *TMPRSS4*, *NPHP4*, and *ORCTL4* polymorphisms were associated with the production of the r.4-6_4-1ins transcript, the r.2818_2823del and r.2818_2859del transcripts, and the r.517-94_517-1ins; r.517-2a>c and r.517_620del transcripts, respectively. Since the proteins encoded by all these transcripts are associated with relatively significant structural changes in the form amino acid insertion/deletion and premature termination, their functional ability may be greatly reduced. Our demonstration of structural changes in mRNA transcripts as a result of splice-site polymorphisms implies that they may be of biological significance in certain pathological conditions.

[Yamada H., Shinmura K., Tsuneyoshi T. and Sugimura H. 2005 Effect of splice-site polymorphisms of the *TMPRSS4*, *NPHP4* and *ORCTL4* genes on their mRNA expression *J. Genet*. **84**, 131–136]

Introduction

RNA splicing removes intron sequences from pre-mRNA and yields mature RNA having a continuous open reading frame that can be translated into protein. Splicing of pre-mRNAs requires accurate selection of the 5′ and 3′ splice-sites, and conserved pre-mRNA *cis*-acting elements that flank exons are important to the splicing (Adams *et al*. 1996). When there is a germline mutation in the splice-site of a gene responsible for a hereditary disease, the patient with the splice-site mutation is susceptible to the disease, as, for example, to hereditary diffuse gastric cancer when a germline *CDH1* (MIM# 192090) mutation is present (Humar *et al*. 2002). The splice-site mutation causes a splicing abnormality associated with a structural change in an mRNA transcript, leading to the production of a truncated protein. The reduction or loss of function of the truncated protein produced on the basis of the mutant allele leads to the onset of the disease with or without inactivation of the wild-type allele. In the same way, it is conceivable that genetic polymorphisms in a splice-site, which in general are recognised when certain variants are present in at least 1% of the popula-

*For correspondence. E-mail: hsugimur@hama-med.ac.jp

Keywords. Splice-site; genetic polymorphism; TMPRSS4; NPHP4; ORCTL4.

tion, also cause the splicing abnormalities associated with structural changes in its mRNA transcript and protein. Actually, we recently showed that the c.457+2T>C polymorphism of the *KLK12* gene (MIM# 605539) is an example of this (Shinmura *et al*. in press). This polymorphism is found in the Japanese population at a high frequency, and it is correlated with the absence of substantial expression of KLK12 serine protease. Similarly, a splice-site polymorphism may be associated with a reduction in the functional capacity of its gene product, and it may therefore act as a low-penetrance susceptibility allele.

Splice-site polymorphisms are registered in the dbSNP single nucleotide polymorphism database of the NCBI Web Site. The validation status consisting of each polymorphism on a four-grade scale is stated in the database, however, some of the variants have been deposited without any annotation on validation status. Moreover, information on the structural changes that result from variants is not available in the database. Therefore, it is important to confirm the existence of splice-site polymorphisms registered in the database and to clarify associations between the splice-site polymorphisms and splicing abnormalities that affect the structure of mature proteins.

In this study we identified the splice-site variants of three genes, the *TMPRSS4* gene (transmembrane protease, serine 4; MIM# 606565), the *NPHP4* gene (nephronophthisis 4; MIM# 606966), and the *ORCTL4* gene (organic-cation transporter-like 4; MIM# 604048), in a Japanese population. In this study, since there are several examples of splice-variants causing gastric cancer, we intentionally tried to look at the genes' status in a gastric cancer population. Gastric cancer is a still one of the biggest issue in Japanese cancers and genetic predispositions are supposed to exist. We then identified the effect of these splice-site polymorphisms on pre-mRNA splicing. The mRNA transcripts resulting from the altered pre-mRNA splicing as a result of each polymorphism were associated with the production of a structurally altered protein.

Materials and methods

***Samples*:** Non-cancerous gastric tissue was obtained from 54 Japanese patients with gastric carcinoma treated at the Hamamatsu University Hospital, Shizuoka, Japan. DNA was extracted by standard SDS-proteinase K digestion followed by phenol/chloroform extraction and ethanol precipitation. Total RNA was extracted with an RNeasy Mini Kit (QIAGEN, Valencia, CA). The research protocol was approved by the Institutional Review Board of Hamamatsu University School of Medicine (12–14).

***Polymerase Chain Reaction (PCR)-Single-Strand Conformation Polymorphism (SSCP) Analysis and DNA Sequencing*:** The region containing each variant was amplified by PCR with HotStarTaq DNA polymerase (QIAGEN). The PCR primers used were: 5′-ACC CCC AAC GTA GAC TCA GG-3′ and 5′-AAA GCC GGA CCT GAA CTT ACC-3′ for the *TMPRSS4* variant (reference sequence: #NT_033899.6), 5′-TGG AGG AGA GTG TGC CTG G-3′ and 5-ACT CAA AGA ACT CGG CGA CC-3′ for the *NPHP4* variant (reference sequence: #NT_021937.16), and 5′-CTT GGC CTA TAG GTC CCA CG-3′ and 5′-TGA GAG AGC CTA TCG GGA GC-3′ for the *ORCTL4* variant (reference sequence: #NT_022517.16). The PCR conditions consisted of initial denaturation at 95°C for 15 min, 35 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 30 s, and extension at 72°C for 1 min, and a final extension at 72°C for 10 min. PCR products were diluted with the loading buffer, denatured, cooled on ice, and loaded on 8% polyacrylamide gel with or without 5% glycerol. SSCP gels were run in 0.5 × TBE for 3–5 h at 200 V at 4°C or at 150 V at room temperature. The bands were visualized by silver staining. PCR products exhibiting abnormally shifted bands in the SSCP analysis were directly sequenced with an ABI 3100 Genetic Analyzer (Applied Biosystems, Tokyo, Japan).

***Reverse Transcription (RT)-PCR*:** RNA samples were converted to first-strand cDNAs with the SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen, Carlsbad, CA). PCR amplification was performed using the following sets of primers: 5′-GGA CTT CTG ACC TGC TGG C-3′ and 5′-TGA TGA TAC TCG CCA GGC TC-3′ for the TMPRSS4 transcripts (reference sequence: #NM_183247.1), 5′-GTA AGC TGG AGC GGA TGA GG-3′ and 5-AAC ATG TCC TCC TCC ACC G-3′ for the NPHP4 transcripts (reference sequence: #NM_015102.2), and 5′-CCT GTG CCT TGG AAT CTG G-3′ and 5′-GAA GAT GAT CAG CCC CAG C-3′ for the ORCTL4 transcripts (reference sequence: #NM_004803.2). The PCR conditions are available on request. PCR products were fractionated by electrophoresis on an agarose gel and stained with ethidium bromide. PCR products exhibiting a single band on gel electrophoresis were directly sequenced, while PCR products exhibiting multiple bands were sequenced after subcloning with pGEM-T Easy vector systems (Promega, Madison, WI).

***Statistical analysis*:** Testing for deviation from the Hardy-Weinberg equilibrium was performed with SNPAlyze (ver 3.2) software (Dynacom, Yokohama, Japan).

Results

We used PCR-SSCP and subsequent sequencing analyses to examine non-cancerous gastric tissue from 54 Japanese patients with primary gastric cancer for the splice-

site polymorphisms of c.4-7A>G (IVS1-7A>G) in the *TMPRSS4* gene, c.2818-2A>T (IVS20-2A>T) in the *NPHP4* gene, and c.517-2A>C (IVS1-2A>C) in the *ORCTL4* gene (figure 1). The allele frequency of the *TMPRSS4* c.4-7A>G, *NPHP4* c.2818-2A>T, and *ORCTL4* c.517-2A>C polymorphism was 0.42, 0.10, and 0.27, respectively (table 1). The allele distribution in each variation was in Hardy-Weinberg equilibrium.

Next, we investigated whether these polymorphisms actually affect the mode of pre-mRNA splicing. cDNAs derived from cases with each genotype were prepared, and the expression of mRNA transcripts was examined by RT-PCR and subsequent sequencing analyses.

In the *TMPRSS4* c.4-7A>G polymorphism, a normal transcript was detected in cases with the *A/A* genotype, whereas only the r.4-6_4-1ins type transcript, which corresponds to the transcript containing the final 6 nucleotides of intron 1 immediately preceding exon 2, was detected in cases with the *G/G* genotype (figure 2a). The sequence before the 6 inserted nucleotides was changed from CTTCAA to CTTCAG by the c.4-7A>G polymorphism, indicating that the c.4-7A>G variation activated the cryptic splice-acceptor site (CTTCAG) in intron 1, resulting in this splicing abnormality. It also appeared that the r.4-6_4-1ins type transcript encoded the p.Met1_Asp2insLeuGln type TMPRSS4 protein.

In the *NPHP4* c.2818-2A>T polymorphism, a normal transcript was detected in cases with the *A/A* genotype, whereas only the type "b" (r.2818_2823del) and "c" (r.2818_2859del) transcripts, corresponding to transcripts missing the first 6 nucleotides of exon 21 and first 42 nucleotides of exon 21, respectively, were detected in cases with the *T/T* genotype (figure 2b). All three of these transcripts were detected in cases with the *A/T* genotype. The sequence of intron 20 splice-acceptor site in the type "b" and "c" genomes was GCGCAG and

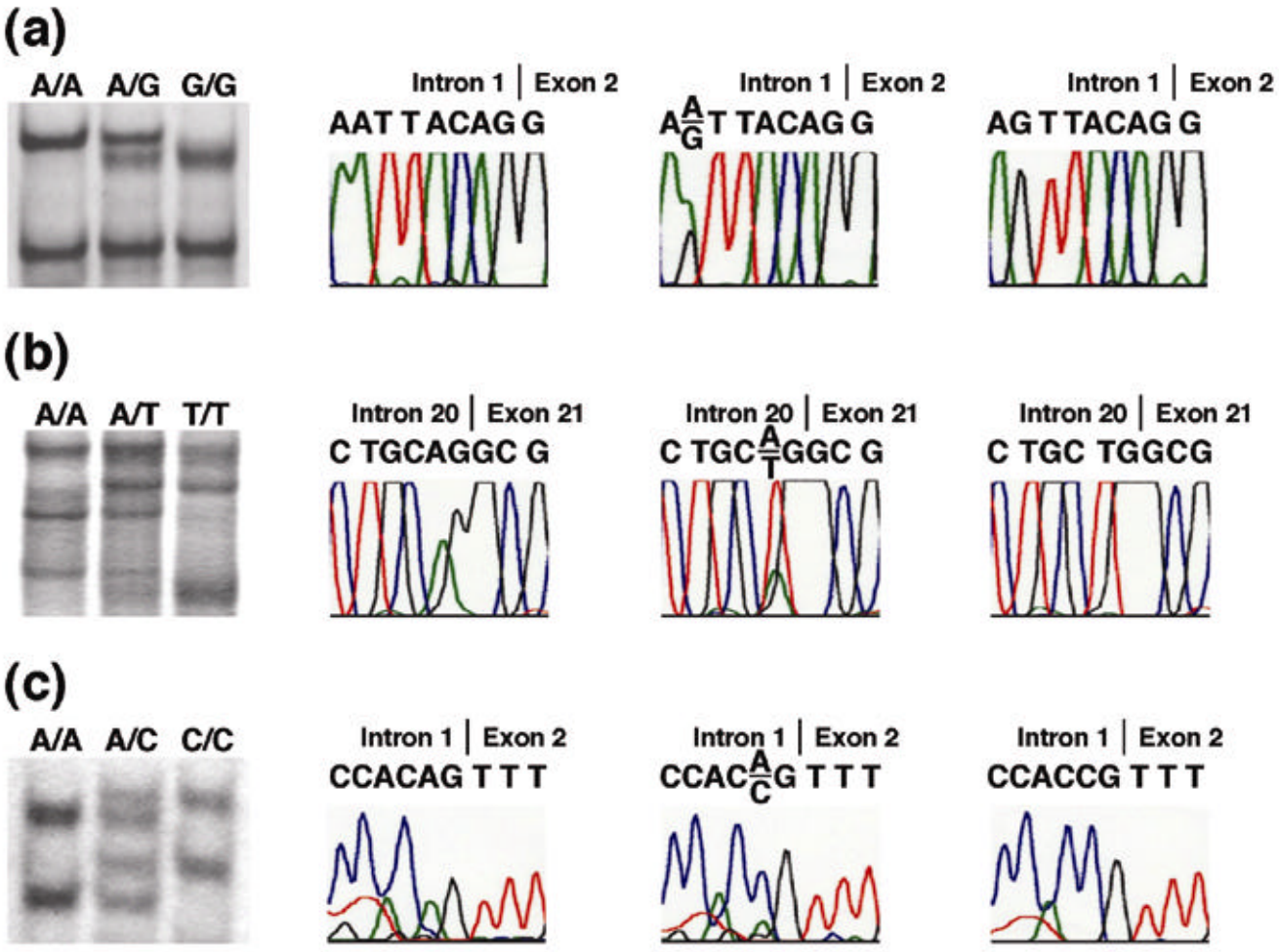

Figure 1. Detection of splice-site polymorphisms in the *TMPRSS4*, *NPHP4*, and *ORCTL4* genes. The panel on the left and the panels on the right show the results of the PCR-SSCP analysis and subsequent sequencing analysis, respectively. (a) The *TMPRSS4* c.4-7A>G polymorphism. (b) The *NPHP4* c.2818-2A>T polymorphism. (c) the *ORCTL4* c.517-2A>C polymorphism.

CTACAG, respectively, indicating that the change in intron 20 splice-acceptor-site sequence from CTGCAG to CTGCTG by the c.2818-2A>T polymorphism leads to activation of the cryptic splice-acceptor site in exon 21, resulting in this splicing abnormality. It also appeared that the structural change in transcripts caused the production of the p.Ala940_Gln941del and p.Ala940_Gln953del type NPHP4 proteins.

In the *ORCTL4* c.517-2A>C polymorphism, a normal transcript was detected in cases with the *A/A* genotype, whereas type "b" (r.517-94_517-1ins; r.517-2a>c) and "c" (r.517_620del) transcripts (figure 2c, lane 5), type "a" and "b" transcripts (lane 6), or a type "a" transcript (lane 7) was detected in cases with the *A/C* genotype (figure 2c). The r.517-94_517-1ins; r.517-2a>c and r.517_620del type transcript corresponded to the transcript containing the final 94 nucleotides of intron 1 before exon 2 and a transcript lacking the entire 104-bp-long exon 2, respectively. The sequence of the intron 1 splice-acceptor site in the type "b" genome was CCCCAG. These results indicated that the change in intron 1 splice-acceptor-site sequence from CCACAG to CCACCG by the c.517-2A>C polymorphism leads to activation of the cryptic splice-acceptor site (CCCCAG) in intron 1 or skipping of the entire exon 2, resulting in this splicing abnormality. It also appears that the structural change in transcript caused the production of truncated ORCTL4 protein (p.Glu172fsX29 and p.Phe173_Lys207delfsX54).

Discussion

Splice-site polymorphisms have been registered in the dbSNP database, however, information on the structural changes resulting from the splice-site variants has not been available in the database. Therefore, in the present study we investigated splice-site polymorphisms of the *TMPRSS4*, *NPHP4*, and *ORCTL4* genes in a Japanese population, and the results showed that these splice-site polymorphisms affect the mode of pre-mRNA splicing. The mRNA transcripts resulting from these splice-site polymorphisms encoded structurally altered proteins.

The results of this study showed that the c.4-7A>G polymorphism in the *TMPRSS4* gene is associated with the production of an mRNA transcript encoding p.Met1_Asp2insLeuGln. The insertion of Leu and Gln residues between Met1 and Asp2 changes the second amino acid from Asp to Leu. According to the N-end rule (Varshavsky 2003), the *in vivo* half-life of proteins is partly dependent on the type of amino acid residue present next to the first Met residue, and thus, this two-amino-acid insertion may alter not only the three-dimensional structure of the protein but its stability. The *TMPRSS4* gene encodes a membrane-bound serine protease, and since it is overexpressed at the mRNA level in pancreatic and gastrointestinal cancers (Wallrapp *et al*. 2000), the c.4-7A>G polymorphism may be associated with susceptibility to the process involved in metastasis formation and tumour invasion.

The results of this study showed that the c.2818-2A>T polymorphism in the *NPHP4* gene is associated with the production of mRNA transcripts encoding p.Ala940_Gln941del and p.Ala940_Gln953del. The existence of the c.2818-2A>T polymorphism was documented in the paper by Otto *et al*. (2002), but its effect on pre-mRNA splicing was not investigated. *NPHP4* has recently been identified as the responsible gene for the juvenile *NPHP* type 4 (Otto *et al*. 2002; Mollet *et al*. 2002). The fact that the NPHP4 protein interacts with NPHP1 (Mollet *et al*. 2002), which is capable of forming complexes with several molecules, including focal adhesion proteins p130Cas and Pyk2 and an actin-binding protein tensin (Benzing *et al*. 2001), suggests that the NPHP4 protein may be involved in cell-cell and cell-matrix adhesions. Thus, the 2- and 14-amino-acid lacking NPHP4 proteins resulting from

Table 1. Distribution of genotype and allele frequencies of the splice-site polymorphisms in the *TMPRSS4*, *NPHP4*, and *ORCTL4* genes.

Gene	Nucleotide change[a]	Genotype frequency			Allele frequency[b]	dbSNP ID
TMPRSS4	c.4-7A>G (IVS1-7A>G)	*A/A* 0.33	*A/G* 0.50	*G/G* 0.17	0.42	rs2276122
NPHP4	c.2818-2A>T (IVS20-2A>T)	*A/A* 0.81	*A/T* 0.17	*T/T* 0.02	0.10	rs1287637
ORCTL4	c.517-2A>C (IVS1-2A>C)	*A/A* 0.52	*A/C* 0.42	*C/C* 0.06	0.27	rs753331

[a]Mutation nomenclature is according to den Dunnen and Antonarakis (2000) and den Dunnen and Paalman (2003). Nucleotide +1 is the A of the ATG-translation initiation codon; the reference sequences are NT_033899.6 (genome) and NM_183247.1 (cDNA) for the *TMPRSS4* gene, NT_021937.16 (genome) and NM_015102.2 (cDNA) for the *NPHP4* gene, and NT_022517.16 (genome) and NM_004803.2 (cDNA) for the *ORCTL4* gene.
[b]Frequency of the minor allele.

the c.2818-2A>T variant may be associated with exhibiting a reduction in ability to perform such a function.

The results of this study showed that the c.517-2A>C polymorphism in the *ORCTL4* gene is associated with the production of mRNA transcripts encoding p.Glu172fsX29 and p.Phe173_Lys207delfsX54. However, unlike the *TMPRSS4* and *NPHP4* polymorphisms, the types and combinations of ORCTL4 transcripts expressed differed among the cases with the *A/C* genotype. This may be due to the difference in the efficiency of transcription or the stability of mRNA among transcripts. ORCTL4 mRNA expression may be controlled by a more complex mechanism. The pathophysiological function of the ORCTL4 protein is unknown, however, it has a structural homology to organic-cation transporters (Nishiwaki *et al*. 1998). Since truncated proteins resulting from the c.517-2A>C polymorphism lack major parts of the sugar transporter and transmembrane domains, presumably they impair capacity to function as a transporter.

In summary, we have demonstrated that the splice-site polymorphisms of three genes change the mode of pre-mRNA splicing and that the mRNA transcript resulting from the altered pre-mRNA splicing encodes a protein displaying a structural change. Since these aberrant proteins may exhibit a reduction in function, our findings may help to explain certain pathological and biological states in which these molecules are involved.

It is very intriguing that the subjects with both variant alleles survive and appear to be normal except having gastric cancer. The contribution of these genotypes may

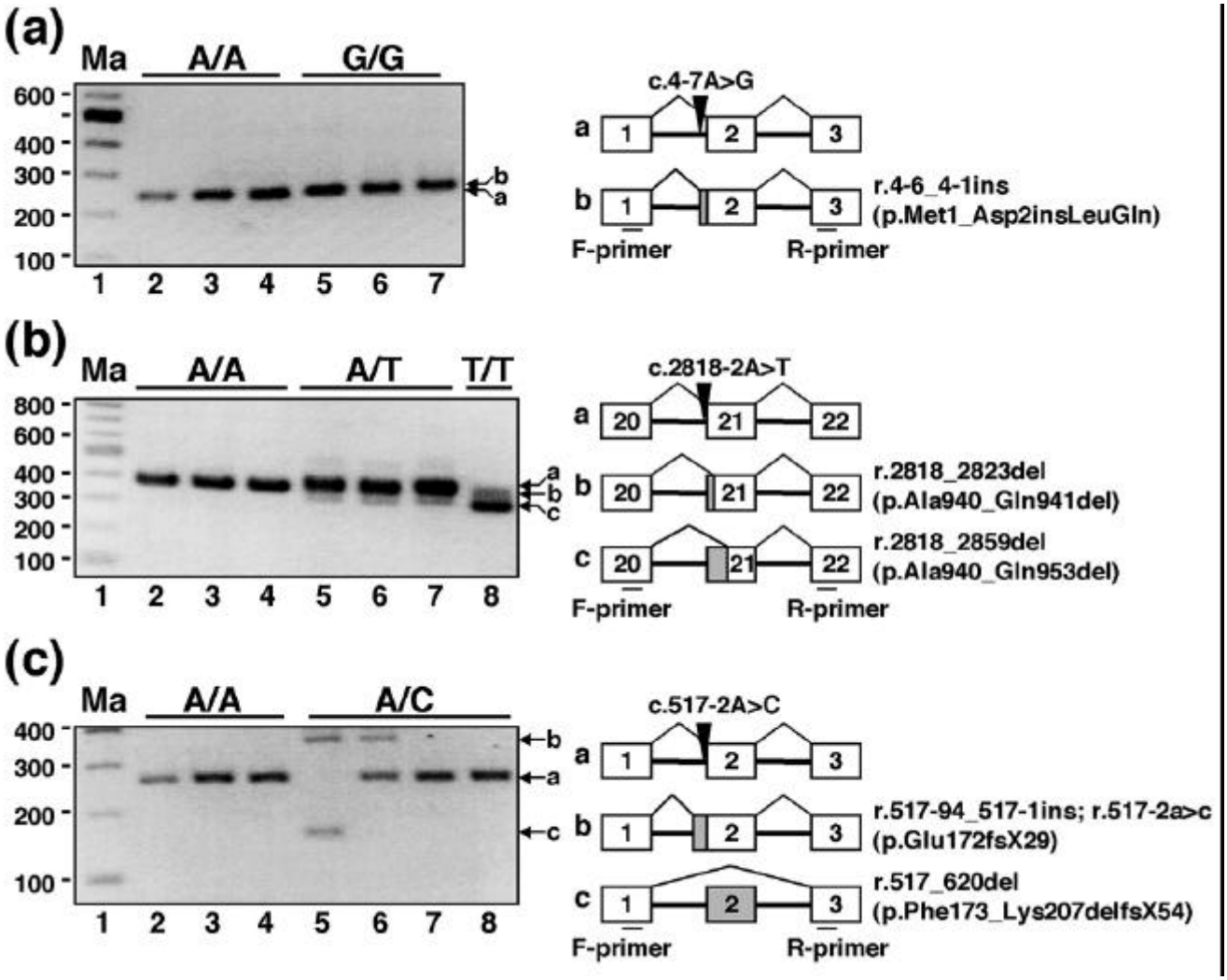

Figure 2. Detection of the mRNA transcripts of the *TMPRSS4* (a), *NPHP4* (b), and *ORCTL4* (c) genes by RT-PCR analysis. The panels on the left show the results of the RT-PCR analysis. Genotypes are indicated above the panels. The diagrams on the right show the genomic structures of the normal (diagram "a") and abnormal (diagram "b" and "c") splicing forms. Inserted and deleted sequences are indicated by closed squares. The location of splice-site polymorphisms is represented by inverted arrowheads. The location of RT-PCR primers is represented by bars. The changes on the RNA level and protein level are indicated on the right side of the diagrams.

be only one of the multiple and sometimes redundant factors in presumed pathological conditions.

Long after submitting our original findings, we noticed updated databases have a little different distribution of these genotypes in 44 Tokyo "control" residents. For the strict evaluation of the distribution of the genotypes in two populations, especially when these are suspects of common disease, we must consider the confounding factors such as age and environmental factors in these populations. We cannot determine whether the apparent difference may be related to the condition of our subjects (gastric cancer) or not, because the information of the control in the database are not known. We are now conducting these genotypes in case-control experiments for further assessment.

Acknowledgements

This work was supported in part by a Grant-in-Aid from the Ministry of Health, Labour and Welfare, from the Ministry of Education, Culture, Sports, Science and Technology of Japan for Scientific Research on Priority Area and the 21st century COE program "Medical Photonics" from the Smoking Research Foundation, and from the Aichi Cancer Research Foundation. The research protocol was approved by the Institutional Review Board of Hamamatsu University School of Medicine (12–14).

References

Adams M. D., Rudner D. Z. and Rio D. C. 1996 Biochemistry and regulation of pre-mRNA splicing. *Curr. Opin. Cell Biol.* **8**, 331–339.

Benzing T., Gerke P., Hopker K., Hildebrandt F., Kim E. and Walz G. 2001 Nephrocystin interacts with Pyk2, p130(Cas), and tensin and triggers phosphorylation of Pyk2. *Proc. Natl. Acad. Sci. USA* **98**, 9784–9789.

den Dunnen J. T. and Antonarakis S. E. 2000 Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum. Mutat.* **15**, 7–12.

den Dunnen J. T. and Paalman M. H. 2003 Standardizing mutation nomenclature: why bother? *Hum. Mutat.* **22**, 181–182.

Humar B., Toro T., Graziano F., Muller H., Dobbie Z., Kwang-Yang H., Eng C., Hampel H., Gilbert D., Winship I., Parry S., Ward R., Findlay M., Christian A., Tucker M., Tucker K., Merriman T. and Guilford P. 2002 Novel germline *CDH1* mutations in hereditary diffuse gastric cancer families. *Hum. Mutat.* **19**, 518–525.

Mollet G., Salomon R., Gribouval O., Silbermann F., Bacq D., Landthaler G., Milford D., Nayir A., Rizzoni G., Antignac C. and Saunier S. 2002 The gene mutated in juvenile nephronophthisis type 4 encodes a novel protein that interacts with nephrocystin. *Nat. Genet.* **32**, 300–305.

Nishiwaki T., Daigo Y., Tamari M., Fujii Y. and Nakamura Y. 1998 Molecular cloning, mapping, and characterization of two novel human genes, ORCTL3 and ORCTL4, bearing homology to organic-cation transporters. *Cytogenet. Cell Genet.* **83**, 251–255.

Otto E., Hoefele J., Ruf R., Mueller A. M., Hiller K. S., Wolf M. T., Schuermann M. J., Becker A., Birkenhager R., Sudbrak R., Hennies H. C., Nurnberg P. and Hildebrandt F. 2002 A gene mutated in nephronophthisis and retinitis pigmentosa encodes a novel protein, nephroretinin, conserved in evolution. *Am. J. Hum. Genet.* **71**, 1161–1167.

Shinmura K., Tao H., Yamada H., Kataoka H., Sanjar R., Wang J., Yoshimura K. and Sugimura H. 2004 Splice-site genetic polymorphism of the human kallikrein 12 (*KLK12*) gene correlates with no substantial expression of KLK12 protein having serine protease activity. *Hum. Mutat.* in press.

Varshavsky A. 2003 The N-end rule and regulation of apoptosis. *Nat. Cell Biol.* **5**, 373–376.

Wallrapp C., Hahnel S., Muller-Pillasch F., Burghardt B., Iwamura T., Ruthenburger M., Lerch M. M., Adler G. and Gress T. M. 2000 A novel transmembrane serine protease (TMPRSS3) overexpressed in pancreatic cancer. *Cancer Res.* **60**, 2602–2606.

Received 16 August 2004; in revised form 9 March 2005