Carcinogenesis vol.34 no.11 pp.2531–2538, 2013
doi:10.1093/carcin/bgt253
Advance Access publication July 19, 2013

Identification and association study with lung cancer for novel insertion polymorphisms of human endogenous retrovirus

Tomoaki Kahyo, Hong Tao, Kazuya Shinmura, Hidetaka Yamada, Hiroki Mori, Kazuhito Funai[1], Nobuya Kurabe, Masaya Suzuki, Masayuki Tanahashi[2], Hiroshi Niwa[2], Hiroshi Ogawa[3], Fumihiko Tanioka[4], Guang Yin[5], Makiko Morita[5], Keitaro Matsuo[5,6], Suminori Kono[5] and Haruhiko Sugimura*

Department of Tumor Pathology and [1]Department of Surgery I, Hamamatsu University School of Medicine, Handayama, Higashi-ku, Hamamatsu 431-3192, Japan, [2]Division of Thoracic Surgery and [3]Division of Pathology, Respiratory Disease Center, Seirei Mikatahara General Hospital, Mikatahara, Kita-ku, Hamamatsu 433-8558, Japan, [4]Department of Pathology, Iwata City Hospital, Ohkubo, Iwata 438-8550, Japan, [5]Department of Preventive Medicine, Graduate School of Medical Sciences, Kyushu University, Maidashi, Higashi-ku, Fukuoka 812-8582, Japan and [6]Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Kanokoden, Chikusa-ku, Nagoya 464-8681, Japan

*To whom correspondence should be addressed. Tel: +81 53 435 2218; Fax: +81 53 435 2225;
Email: hsugimur@hamamed.ac.jp

Sequences of human endogenous retroviruses (HERVs) are members of the long terminal repeat (LTR) retrotransposon family. Although the expression of HERV has long been a topic of investigation, HERV-insertion polymorphisms are not well known, and a genetic association between HERV-insertion polymorphisms and cancer has never been reported. To identify novel HERV loci in the genome from cancer tissues, we carried out the inverse PCR method targeting a conserved LTR region of HML-2, which is the most recently acquired HERV group. Novel two insertions, *HML-2_sLTR(1p13.2)* and *HML-2_sLTR(19q12)*, were identified as insertionally polymorphic solo LTRs. Furthermore, a significant prevalence of *HML-2_sLTR(1p13.2)* homozygosity was detected in female never-smoking patients aged 60 years and over who had lung adenocarcinoma [versus the other genotyping; odds ratio (OR): 1.97; 95% confidence interval (CI): 1.01–3.81]. In another cohort consisting of female never-smoking patients with lung adenocarcinoma, a prevalence of *HML-2_sLTR(1p13.2)* homozygosity tended to be high in patients aged 60 years and over (versus the other genotyping; OR: 2.03; 95% CI: 0.96–4.29), whereas a low prevalence of *HML-2_sLTR(1p13.2)* homozygosity was detected in patients <60 years old (versus the other genotyping; OR: 0.31; 95% CI: 0.11–0.94). Our results suggest that *HML-2_sLTR(1p13.2)* is involved with the susceptibility to lung adenocarcinoma in female never-smokers in an age-dependent manner and that other HERV polymorphisms related to human diseases might remain to be identified in the human genome.

Introduction

Retrotransposons are known as mobile genetic elements that are composed of ~42% of the human genome, and human endogenous retroviruses (HERVs) are members of the long terminal repeat (LTR) retrotransposon family, accounting for ~8% of the human genome (1). Retroviral invasion of the host genome is mediated by the following three processes: infection of a host cell; reverse transcription of retroviral RNA in the cytoplasm and integration of reverse-transcribed DNA into the host genome along with target site duplication (2). When the retroviral invasion occurs in the germ cell lines of human or human ancestors, the integrated DNA (known as a provirus) can be transferred to the next generation and is finally named as an HERV after repeated vertical transfer. A complete HERV sequence contains two LTRs located at both ends and several open reading frames, such as *gag*, *pro*, *pol* and *env*, between the LTRs. A number of solo open reading frames and LTRs, the latter of which is caused by the deletional recombination of two LTRs, are also present in the human genome (3).

Intact LTRs possess transcriptional promoter and enhancer elements, such as a TATA box-like signal, hormone receptor-binding sites and enhancer core (4–6). Indeed, promoter and enhancer activities of an HERV LTR have been shown in human cell cultures (7). Thus, endogenous retroviruses can be involved in the transcriptional regulation of surrounding genes. Regarding their mobility, however, the reintegration of HERV in somatic cells has never been demonstrated in contemporary human tissues. Almost all HERVs seem to lack retrotransposition activity because of the accumulation of mutations within the HERV sequences over the course of millions of years. On the other hand, the *de novo* integration of the retrotranspositions of other species' endogenous retroviruses and other human retrotransposable elements, including Alu and LINE-1, have been reported (8,9). Notably, somatic L1 insertions occurred at high frequencies in human lung cancer genomes (10).

The HERV-K family is the most recently active HERV element. This family is thought to have integrated into the genome of the common ancestor of humans and Old World monkeys around 35 million years ago (11). Although a large number of insertion polymorphisms of Alu and LINE-1 have been identified (12) and have been assigned in the dbRIP database (2569 Alu and 598 LINE-1 loci in hg19 assembly) (13,14), few insertion polymorphisms of HERV-K have been reported (3,15). Only 11 insertion polymorphisms of HERV-K, all of which are either biallelic (preinsertion and provirus, or preinsertion and solo LTR) or triallelic (preinsertion, provirus and solo LTR), are known (Supplementary Table I, available at *Carcinogenesis* Online). Another biallelic pattern (provirus and solo LTR), which is not an insertion polymorphism and should instead be termed a deletion polymorphism, has also been presented.

The HERV-K family has 11 subgroups (HML-1 to HML-11), which are most closely related to mouse mammary tumor virus (3), and HML-2 is thought to be the most recently acquired subgroup (2). The integration of HML-2 appears to have occurred in germ line cells even after the human chimpanzee divergence, since human-specific or polymorphic insertions of HML-2 have been detected in the human genome (16). The youngest HERV provirus *HERV-K113*, a member of the HML-2 subgroup, is estimated to have integrated ~1 million years ago and has intact open reading frames for all viral proteins (17). Although retrotransposition activity of HML-2 has not been reported so far, the HERV elements that were reconstituted from consensus sequence of HML-2 proviruses showed the *de novo* retrotranspositions in culture cells (18,19). Expressions of HML-2 were upregulated in a variety of diseases, including ovarian cancers and melanoma (15). Np9 and Rec proteins, both of which are products from HML-2 proviruses, have been reported to be interacted with a tumor-related protein (20). However, a genetic association between insertional polymorphisms of HML-2 and cancers has not been reported explicitly.

Globally, lung cancer is the leading and second leading cause of death from cancer in men and women, respectively, based on GLOBOCAN 2008 estimates (21). The main histological types of lung cancer are small-cell lung cancer and non-small-cell lung cancer. The major histological subtypes of non-small-cell lung cancer are squamous cell carcinoma, adenocarcinoma and large cell carcinoma. Lung squamous cell carcinoma and adenocarcinoma account for most lung cancers. Colorectal cancer is the fourth most common cause of death from cancer in both sexes and the second most commonly

Abbreviations: CI, confidence interval; EGFR, epidermal growth factor receptor; HERVs, human endogenous retroviruses; LTR, long terminal repeat; OR, odds ratio; SD, standard deviation.

diagnosed cancer in women (21). Pathologically, the location of colorectal cancer (proximal or distal) is associated with various biological effects, such as a preferable incidence of tumors with sporadic chromosomal instability at distal sites (22,23). Thus, genetic association studies for susceptibility to lung squamous cell carcinoma and adenocarcinoma or to site-specific colorectal cancers are drawing the attention of many researchers in the field of cancer.

Here, we report the identification of two novel insertion polymorphisms of an HML-2 solo LTR (HML-2_sLTR) using the inverse PCR method. Furthermore, we present, for the first time, an association between insertion polymorphisms of HERV and a susceptibility to cancer. Our findings suggest that other HERV polymorphisms related to cancers may remain to be found in the human genome.

Materials and methods

Specimens

The lung tissues used in the inverse PCR and genotyping experiments were obtained from eight patients undergoing surgery at Hamamatsu University School of Medicine, University Hospital. In Shizuoka non-concurrent study on lung cancer, the numbers of cases and controls were 213 and 587, respectively. Cases were accrued at three hospitals located in Hamamatsu City or Iwata City (adjacent to each other), and controls were accrued at a hospital in Iwata City. Both the cases and controls were composed of a population aged 60 years and over. In the Aichi cohort study for female never-smoking patients with lung adenocarcinoma, the numbers of age-matched cases and controls were 125 and 248, respectively. Detailed information regarding this cohort study can be obtained from a previous report (24). In the Fukuoka cohort study of colorectal cancer, the numbers of cases and controls were 683 and 778, respectively. Controls were randomly selected from the community using a frequency-matching protocol with respect to sex, age and residence. Detailed information regarding the colorectal cancer study can be obtained from previous reports (25–27). The clinical and demographical information of the cases and control in these studies was shown in Supplementary Tables II and III, available at *Carcinogenesis* Online. In the genotyping experiments, genomic DNA extracted from blood specimens was examined, except in Shizuoka case study for lung cancer, in which genomic DNA was extracted from normal lung tissues. Written informed consent was obtained from each individual patient.

Nested inverse PCR and genotyping PCR

Genomic DNA from tumor or normal lung tissues was extracted using the DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA). Then, 250 ng of genomic DNA was digested with NspI (New England Biolabs, Beverly, MA) at 37°C for >8 h. Following the inactivation of the restriction enzyme by heating at 65°C for 20 min, the DNA fragments were self-ligated with Ligation High Ver.2 (Toyobo, Osaka, Japan) at 4°C for >2 h. When using ApoI (New England Biolabs), the DNA was digested at 50°C, and the enzyme was inactivated at 80°C. An equivalent to 10 ng of genomic DNA was used for the first PCR with PrimeSTAR DNA polymerase (TaKaRa, Otsu, Japan). One two-hundredth (1/200) of the amount of the final product was subjected to a second PCR with HotStarTaq DNA Polymerase (Qiagen). Then, 10–50 ng of genomic DNA was used with Pfu-x (Greiner Bio-One, Frickenhausen, Germany) and HotStarTaq DNA polymerases for the genotyping PCR of *HML-2_sLTR(1p13.2)* and *HML-2_sLTR(19q12)*, respectively. The primers for the nested inverse PCR were designed to target the LTR region between the restriction site and the terminal junction of the flanking region, referring the consensus sequence of HML-2 LTRs that are highly homologous to the other LTR within the same proviral element (Supplementary Figure 1, available at *Carcinogenesis* Online). The primers for the genotyping PCR of *HML-2_sLTR(1p13.2)* and *HML-2_sLTR(19q12)* were designed to target the outside areas of those solo LTRs. The theoretical sizes of the products from the genotyping PCR were 220 bp (preinsertion) and 1195 bp (insertion) for *HML-2_sLTR(1p13.2)*, and 221 bp (preinsertion) and 1187 bp (insertion) for *HML-2_sLTR(19q12)*. The primer sequences are listed in Supplementary Table IV, available at *Carcinogenesis* Online, and the cycle conditions are described in Supplementary Figure 2, available at *Carcinogenesis* Online.

Cloning, sequencing and database search

The inverse PCR products were cloned using the pGEM-T easy vector system (Promega, Madison, WI). The BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) and a capillary sequencer (3100 Model, Applied Biosystems) were used for sequencing. The Basic Local Alignment Search Tool-like alignment tool of the University of California Santa Cruz (http://genome.ucsc.edu/) was used to map the loci of the clones in the human genome (GRCh37/hg19 assembly), and the unmapped clones were further investigated with the Basic Local Alignment Search Tool of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). When the cloned sequence covering the junction of the LTR and the flanking genomic region was not identified within the databases, the finding was considered to indicate a novel HERV insertion. Furthermore, when the novel insertion was also identified in normal tissue within the same individual or in another individual, the finding was regarded as an insertion polymorphism of HERV. If the novel insertion could not found in normal tissue within the same individual, the finding was regarded as a somatic integration of HERV in the tumor tissue, that is, a reintegration that has never been reported in contemporary human tissues. To read the sequences of the LTRs and the surrounding regions, products that were amplified through genotyping PCR were subjected to agarose electrophoresis and directly sequenced. The complete sequences of *HML-2_sLTR(1p13.2)* and *HML-2_sLTR(19q12)* are provided in Supplementary Figure 3, available at *Carcinogenesis* Online.

Statistical analysis

The statistical analysis was performed by estimating the 95% confidence interval (CI). Because of the small number of male never-smokers, female smokers and females with squamous cell carcinomas, these categories were not included in the statistical analysis. Multiple testing was performed using the statistical software SAS (SAS Institute, Cary, NC) for adjustments in the colorectal cancer study, as described in a previous report (25–27). The genotype distributions of *HML-2_sLTR(1p13.2)* and *HML-2_sLTR(19q12)* showed Hardy–Weinberg equilibrium in both control groups in the lung and colorectal cancer studies.

Ethics

The design of this study was approved by the Institutional Review Boards of Hamamatsu University School of Medicine, Kyushu University, Aichi Cancer Center Research Institute and all the participating hospitals except for two hospitals (which had no ethical committees). The survey was performed at the latter two hospitals with the permission of the director of each hospital.

Results

Strategy for identification of novel HERV insertions in human genome

To clone novel HERV insertion sites, if possible polymorphic sites or somatic mutations, from the genomic DNA of tumor tissue, the inverse PCR method, which enables the amplification of regions flanking a primer target site (28), was performed. In this method, the primer target site should be set within the LTR region between the restriction site and the terminal junction of the flanking region (Figure 1). We selected ApoI and NspI as the most suitable sites (Supplementary Figure 1, available at *Carcinogenesis* Online). These two restriction sites are highly preserved among HML-2 LTRs, and the fragments digested at those sites contain a long region sufficient to enable the design of the PCR primers within an LTR (Supplementary Figure 1 and Supplementary Table IV, available at *Carcinogenesis* Online). Furthermore, the average lengths of the genomic fragments digested at the ApoI and NspI sites are relatively small (~460 and 930 bp in the human genome, respectively) (29), and fragments with these lengths do not seem to have difficulty with self-ligation. Finally, NspI was selected because of its sticky end consisting of a GC base pair. The amplified products were TA cloned and sequenced using a capillary sequencer. Then, the sequences were analyzed for the purpose of distinguishing between a known HML-2 site and a novel one. Furthermore, novel HML-2 sites were investigated using normal tissues to determine whether they were polymorphic or somatic (Figure 1).

Identification of novel HERV insertions

Genomic DNA extracted from lung tumor tissues was used for the examination described above, and 178 clones from eight specimens were sequenced using the Sanger dideoxy method (Supplementary Table V and Supplementary Figure 4A, available at *Carcinogenesis* Online). As a result of the analyses, three products were acquired as HERV insertions that were not included in the databases (Figure 2A). Although one of the three products contained a flanking sequence that was too short to identify the genomic region (Figure 2A(c)), the flanked sequence of another product was long enough to be definitely identified at 19q12 (Figure 2A(a)). In the other case, the flanking

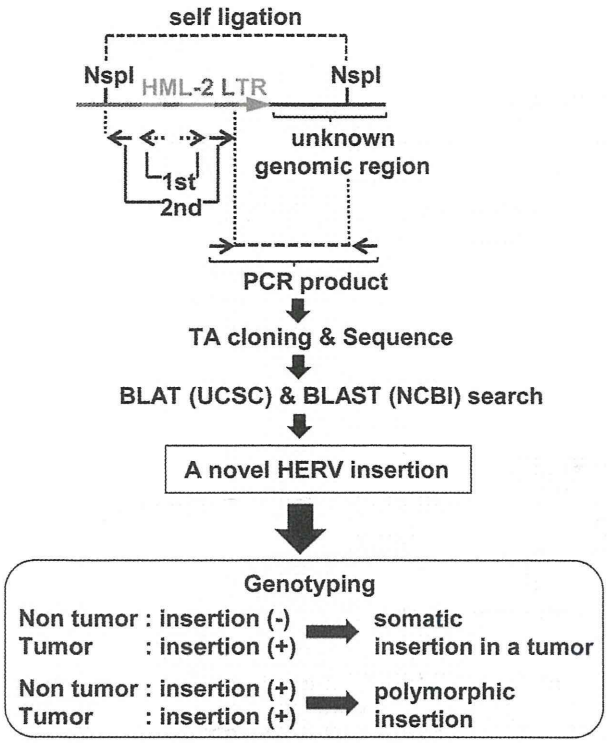

Fig. 1. Strategy for identifying novel HERV insertions in the human genome. Genomic DNA was extracted from lung cancer specimens, digested with NspI and self-ligated, followed by nested inverse PCR. The amplified products covering the junction of the HML-2 LTR and the flanking genomic region were TA cloned and sequenced. The PCR primers were designed for the region between the terminal junction to a flanking genomic region and the NspI site within the HML-2 LTR. The sequenced data were analyzed using BLAT and BLAST searches. The thick gray arrows represent the HML-2 LTR. The dotted and solid arrows indicate the primers used for the first and second PCR, respectively.

region was identified at 1p13.2, but more than a dozen candidates in addition to 1p13.2 were feasible if a one-nucleotide difference was permitted as a sequencing error (Figure 2A(b)). This result arose because the Alu repeat sequence, the copy number of which is >1 000 000 in the human genome (30), was present in the flanking region for product b. As no ApoI sites were identified in the sequence of product b (Supplementary Figure 4A, available at *Carcinogenesis* Online), nested inverse PCR using ApoI digestion and the primer designed across HML-2 LTR and the flanking region of product b was performed to acquire a long sequence sufficient to determine the location (Figure 2B). As expected, a longer sequence was successfully cloned, and the clone was shown, with certainty, to arise from a region at 1p13.2 (Supplementary Figure 4B, available at *Carcinogenesis* Online). This insertion containing HML-2 LTR at 1p13.2 is located within a chitinase gene cluster consisting of *acidic mammalian chitinase* (*CHIA*), *chitinase 3-like 2* (*CHI3L2*), *oviductal glycoprotein 1* (*OVGP1*), *CHIA-like pseudogene* (*RP11-165H20.1*) and *chitinase 1 pseudogene* (*LOC100420342*) (Figure 2C and Supplementary Table VI, available at *Carcinogenesis* Online). This genomic region displays a high degree of synteny among mammals (31). The other insertion containing HML-2 LTR at 19q12, from product a, was located within a non-coding region (*LOC284395*) in the vicinity of a gene cluster with highly correlated gene expression patterns, including *POP4* and *CCNE1*, in addition to the boundary for copy number variations in gastric tumors (32).

Genotyping of HERV insertions at 1p13.2 and 19q12

To determine whether the two HERV insertions described above could also be detected in non-tumor samples from the same individual, genotyping PCR was performed. The primers were constructed for the outside areas of the HML-2 insertions at 1q13.2 and 19q12 (Figure 2C), and the HML-2 insertion and/or preinsertion sites were amplified. For both 1p13.2 and 19q12, the short insertions (~1.2 kb) were also detected in normal tissue within the same individual (Figure 3A and 3B). These results indicate that HML-2 insertions at 1p13.2 and 19q12 are insertionally polymorphic but are not generated by somatic retrotransposition. Furthermore, the results of the genotyping experiments and direct sequencing showed that both of the polymorphic insertions were biallelic (preinsertion and solo LTR) (Figure 3C and D; Supplementary Figure 3, available at *Carcinogenesis* Online). Although the possibility that a provirus may be identified as an allele of these polymorphic HERV loci cannot be excluded, we believe that there were no or few proviral alleles of the polymorphic HERV loci, at least in the genetic pool used in this study, because all the genotypes resulted in patterns composed of preinsertion and/or solo LTR alleles in a large number of samples, as shown below (Tables I and III; Supplementary Table VII). Thus, the polymorphic HML-2 insertions at 1p13.2 and 19q12 were termed *HML-2_sLTR(1p13.2)* (GenBank/EMBL/DDBJ accession number: AB728590) and *HML-2_sLTR(19q12)* (GenBank/EMBL/DDBJ accession number: AB728591), respectively. As a result of sequencing the surrounding regions, 6 bp target site duplications were found in these two HML-2_sLTR loci, similar to other HERV-K polymorphisms (Figure 3C and D; Supplementary Table I, available at *Carcinogenesis* Online). These findings indicate that *HML-2_sLTR(1p13.2)* and *HML-2_sLTR(19q12)* are certainly polymorphic insertions that have been generated in a germ line cell through retrotransposition and subsequent recombination between LTRs.

Epidemiological analysis of HML-2_sLTR(1p13.2) and HML-2_sLTR(19q12) in cancer studies

A significant association between insertion polymorphisms of HERV and cancer has never been reported. Thus, non-concurrent studies involving lung cancer patients were performed to investigate the genetic association between novel insertional polymorphisms and cancers. Although there was no significant prevalence of *HML-2_sLTR(19q12)*, female lung cancer patients exhibited a significantly high prevalence of *HML-2_sLTR(1p13.2)* homozygosity (genotype *LL*) compared with the preinsertion homozygosity (genotype *PP*) [age ≥60 years: odds ratio (OR): 2.16; 95% CI: 1.04–4.45; the numbers of cases and controls were 44 and 101, respectively; see Table I]. The heterozygosity (genotype *PL*) of *HML-2_sLTR(1p13.2)* showed the similar prevalence to the preinsertion homozygosity. Thus, the preinsertion homozygosity and the heterozygosity were regarded as one reference group in the following analyses. Categorized data analyses considering the smoking history and the histological subtypes revealed that female never-smoking patients with lung adenocarcinoma exhibited a significantly high prevalence of *HML-2_sLTR(1p13.2)* homozygosity (versus genotype *PP* + *PL*; age ≥60 years: OR: 1.97; 95% CI: 1.01–3.81; the numbers of cases and controls were 67 and 179, respectively; see Table II), whereas male smokers with lung squamous cell carcinoma exhibited a moderately high prevalence of *HML-2_sLTR(19q12)* homozygosity (versus genotype *PP* + *PL*; age ≥60 years: OR: 1.70; 95% CI: 0.95–3.02; the numbers of cases and controls were 70 and 332, respectively; see Table II). To further validate the association between *HML-2_sLTR(1p13.2)* homozygosity and lung adenocarcinoma, another cohort consisting of female never-smoking patients with lung adenocarcinoma and controls was investigated (Table III). In this study, the means of age of the cases and controls were 59.8 [standard deviation (SD) = ±10.0] and 59.7 (SD = ±9.9) years, respectively, and the median of age of the cases and controls were 61 and 60.5 years, respectively. Thus, analyses in two age subcategories, age <60 years and age ≥60 years, were also performed under consideration of the effect of age. The overall age-adjusted prevalence of *HML-2_sLTR(1p13.2)* homozygosity was similar to that of the reference group (versus genotype *PP* + *PL*; OR: 0.99; 95% CI: 0.56–1.74; the numbers of cases and

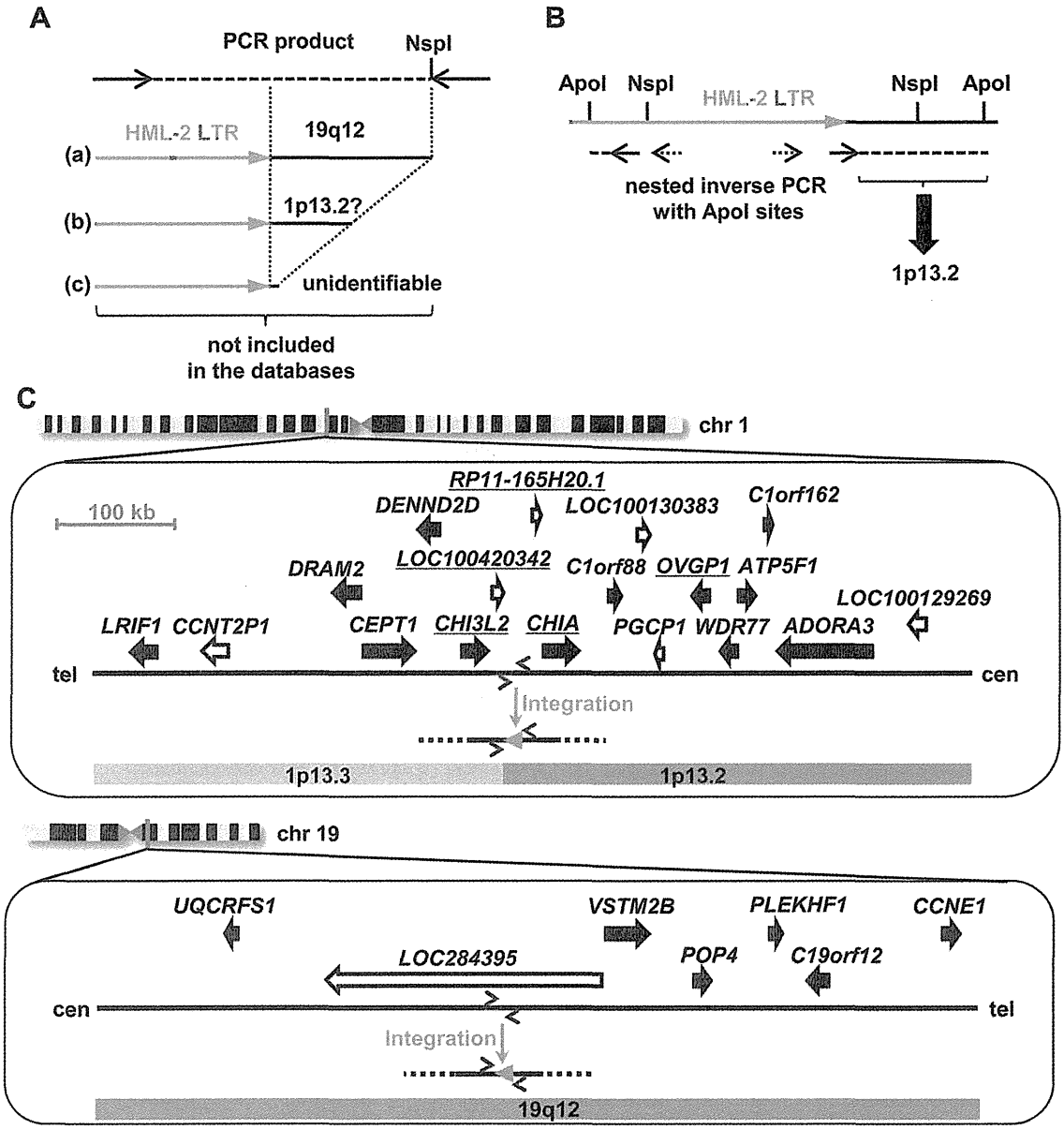

Fig. 2. Novel HERV insertions on chromosome 1 and chromosome 19. (**A**) Three products were not included in the BLAT and BLAST databases (a–c). The black arrows show the second primers used in the nested inverse PCR, and one of the pair was designed to cover the NspI site. The solid line indicates the flanking genomic region. (**B**) Nested inverse PCR with ApoI digestion was performed to clearly identify the flanking region to HML-2 LTR of product b. The amplified products were subjected to agarose electrophoresis and a direct sequence analysis. The primer designed to cover the HML-2 LTR and the flanking region was used for the second PCR. The thick gray arrows represent the HML-2 LTR. The dotted and solid arrows indicate the primers used for the first and second PCR, respectively. (**C**) Schemes for novel polymorphic insertions of HERV. The underlined genes are the chitinase cluster members. The arrowheads indicate the primers used for the genotyping PCR in Figure 3. The black and white arrows represent the protein-coding and non-coding genes, respectively.

controls were 125 and 248, respectively). However, a prevalence of *HML-2_sLTR(1p13.2)* homozygosity tended to be high in an elderly group (versus genotype *PP* + *PL*; age ≥60 years: OR: 2.03; 95% CI: 0.96–4.29; the numbers of cases and controls were 71 and 140, respectively). On the contrary, a significantly low prevalence of *HML-2_sLTR(1p13.2)* homozygosity was detected in a younger group (age <60 years: OR: 0.31; 95% CI: 0.11–0.94; the numbers of cases and controls were 54 and 108, respectively).

In addition to patients with lung cancer, a possible association with *HML-2_sLTR(1p13.2)* homozygosity has also been investigated in patients with colorectal cancer. As a result, no significant association with *HML-2_sLTR(1p13.2)* was detected in any categories of colorectal cancer, whereas a moderately low prevalence of *HML-2_sLTR(1p13.2)* homozygosity was detected in women with proximal colon cancer (versus genotype *PP* + *PL*; overall age group: OR: 0.44; 95% CI: 0.18–1.08; the numbers of cases and controls were 71 and 288, respectively) (Supplementary Table VII, available at *Carcinogenesis* Online). Unlike lung cancer, no apparent association between the *HML-2_sLTR(1p13.2)* genotype and the prevalence of colorectal cancer was seen.

Discussion

Here, inverse PCR and cloning methods were used to identify endogenous retrotransposon loci in the human genome. Inverse PCR enables regions flanking the target site to be amplified, but the sequence of the target site must already be known for primer

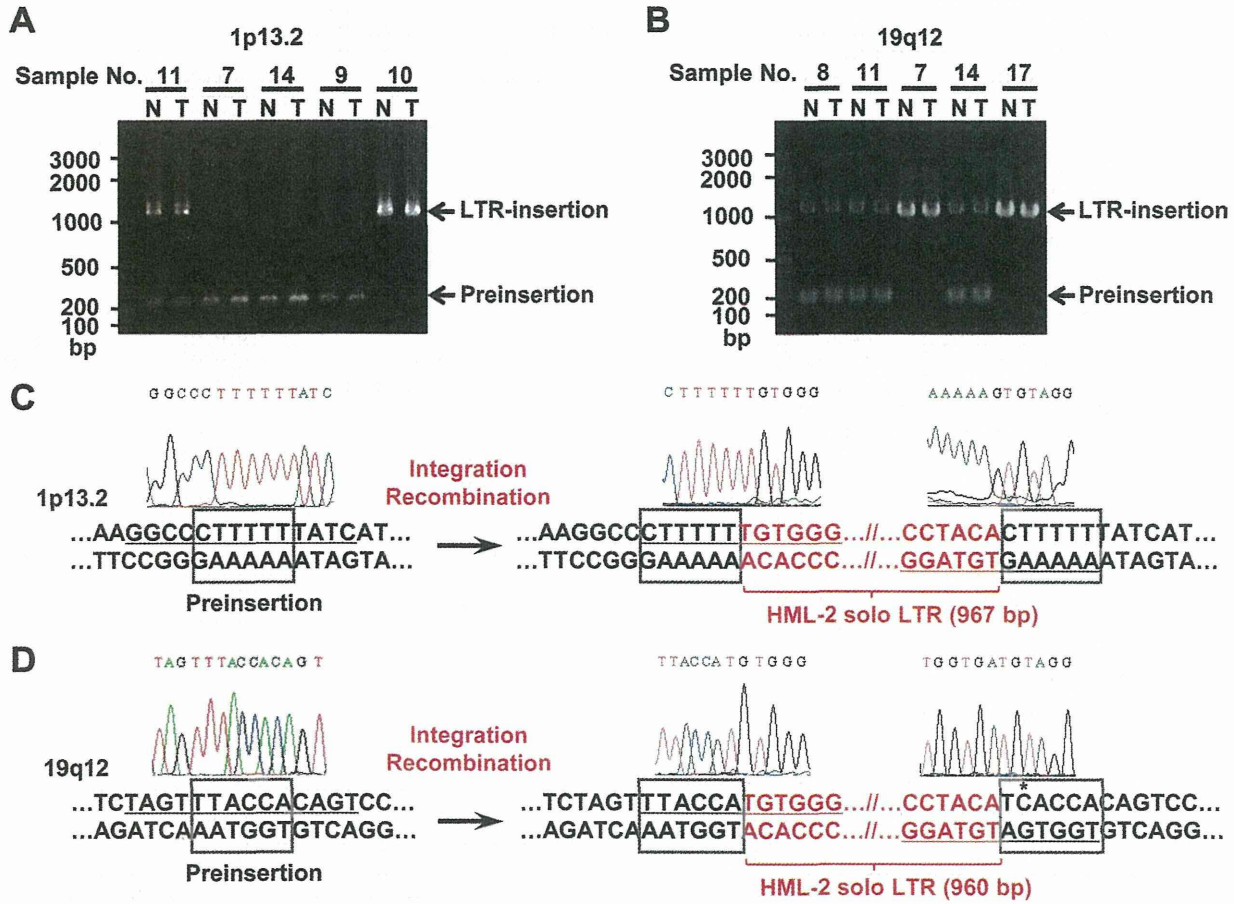

Fig. 3. Genotyping and sequencing of polymorphic HERV insertions. Genotyping PCR for *HML-2_sLTR(1p13.2)* (**A**) and *HML-2_sLTR(19q12)* (**B**) was performed using the primers indicated in Figure 2. The PCR products were subjected to agarose electrophoresis. The junction sequences of HML-2_sLTR and the flanking genomic region at 1p13.2 (**C**) and 19q12 (**D**) are shown. The indicated sequence traces correspond to the underlined sequences. Boxed sequences show the target sites. The asterisk represents a putative mutation site. N, non-tumor, T, tumor.

Table I. Association between lung cancer and genotypes of *HML-2_sLTR (1p13.2)* and *HML-2_sLTR(19q12)*

Genotype	Lung cancer			
	Men		Women	
	N[a]	OR (95% CI)[b]	*N*[a]	OR (95% CI)[b]
1p13.2				
PP	55/142	Reference	22/69	Reference
PL	59/161	0.95 (0.61–1.46)	30/103	0.91 (0.48–1.72)
LL	25/80	0.81 (0.46–1.40)	22/32	2.16 (1.04–4.45)
19q12				
PP	46/111	Reference	26/59	Reference
PL	58/196	0.71 (0.45–1.13)	35/110	0.72 (0.39–1.32)
LL	35/76	1.11 (0.65–1.89)	13/35	0.84 (0.38–1.85)

L, solo LTR; *P*, preinsertion.
[a]Number of cases/controls.
[b]Age ≥60 years; the west area of Shizuoka Prefecture.

construction. Thus, the construction of primers was based on the alignment of HML-2 LTRs that are highly homologous to the other LTR within the same proviral element because retroviral reverse transcription leads to a pair of LTRs with identical sequences and HML-2 is the most recently acquired group (2). In addition, inverse PCR demands a restriction site that meets the following conditions: a sticky end; and the ability to generate digested fragments covering both a long HML-2 LTR region sufficient to allow the design of the PCR primers and a flanking genomic region that is not too long to prevent the subsequent self-ligation and PCR. Our efforts resulted in the identification of two novel insertion polymorphisms of HML-2_sLTR. To read the sequences of PCR products more efficiently, cloning and capillary sequencing methods can be replaced using high-throughput sequencing methods, although some modifications are necessary. Recently, the *de novo* integration of the yeast Ty3 retrotransposon was investigated using a method combining inverse PCR and high-throughput sequencing (33). Thus, more polymorphic loci of HERVs are probably to be identified, and somatic retrotranspositions of HERVs could potentially be detected by combining inverse PCR and high-throughput sequencing. Because retrotranspositions can cause genomic variations, the demonstration of somatic retrotranspositions of HERVs is probably to lead to new insights into human cancer etiology, where genomic instability induced by errors in chromosomal distribution during mitosis and the downregulation of the DNA repair system is considered to be primary internal causes of cancer (34,35).

Lung cancer susceptibility research is vigorously pursuing comprehensive approaches, such as a transcriptome or adductome approach (36). A comprehensive approach targeting insertion polymorphisms of retrotransposons must also contribute to lung cancer risk evaluation. Insertion polymorphisms of HERV-K have been identified mainly using low stringency hybridization (37) or data mining methods involving human genome databases (3,38). In the method of low stringency hybridization, genomic DNA digested with a restriction site was electrophoresed and subjected to hybridization with a DNA probe. Insertional polymorphisms of targeted elements are determined with existence or non-existence of a specific signal on a dried gel or blotted membrane. Therefore, although this method is available in using a number of genomic DNA samples, it would not be

Table II. Association between histology-specific lung cancer and genotypes of *HML-2_sLTR(1p13.2)* and *HML-2_sLTR(19q12)*

Sex	Smoking history	Genotype	Squamous cell carcinoma		Adenocarcinoma	
			N[a]	OR (95% CI)[b]	*N*[a]	OR (95% CI)[b]
Men	1p13.2					
	Never smoking	*PP + PL*	0/39	nd	5/39	nd
		LL	0/12	nd	0/12	nd
	Smoking	*PP + PL*	51/264	Reference	58/264	Reference
		LL	13/68	0.99 (0.50–1.93)	12/68	0.80 (0.40–1.59)
	19q12					
	Never smoking	*PP + PL*	0/42	nd	5/42	nd
		LL	0/9	nd	0/9	nd
	Smoking	*PP + PL*	50/265	Reference	49/265	Reference
		LL	14/67	1.11 (0.57–2.13)	21/67	1.70 (0.95–3.02)
Women	1p13.2					
	Never smoking	*PP + PL*	0/149	nd	48/149	Reference
		LL	1/30	nd	19/30	1.97 (1.01–3.81)
	Smoking	*PP + PL*	2/23	nd	2/23	nd
		LL	1/2	nd	1/2	nd
	19q12					
	Never smoking	*PP + PL*	1/146	nd	55/146	Reference
		LL	0/33	nd	12/33	0.97 (0.46–2.01)
	Smoking	*PP + PL*	2/23	nd	3/23	nd
		LL	1/2	nd	0/2	nd

L, solo LTR; nd, not determined; *P*, preinsertion.
[a]Number of cases/controls.
[b]Age ≥60 years; the west area of Shizuoka Prefecture.

Table III. Association between never-smoking females with lung adenocarcinoma and genotypes of *HML-2_sLTR(1p13.2)* in another cohort study

Age	Genotype	*N*[a]	OR (95% CI)[b]
Overall	*PP + PL*	103/204	Reference
	LL	22/44	0.99 (0.56–1.74)
<60	*PP + PL*	50/85	Reference
	LL	4/23	0.31 (0.11–0.94)
≥60	*PP + PL*	53/119	Reference
	LL	18/21	2.03 (0.96–4.29)

L, solo LTR; *P*, preinsertion.
[a]Number of cases/controls.
[b]Adjusted for age; female never-smokers in Aichi Prefecture.

suitable to distinguish multiple signals with almost same mobility. In the data mining methods, public sequence data of human genome are mined to search targeted elements. If a flanking region of a targeted element is determined, genotyping for the locus can be performed for validation of the data mining method. By filtering the data under a specific condition, such as a human-specific locus, it is possible to adjust the number of genotyped loci. Although this method may be most available from the angle of cost/performance criteria, it is just a data analysis and requires sequence data on databases. The strategy to indentify novel HERV insertions described here includes the sequencing process and also enables to individually analyze the same size of amplicons. As a matter of course, our strategy has also weak points, such as necessity of a specific restriction site near targeted elements. A variety of methods would be necessary for surveying the landscape of repeat elements in the genome.

In the human genome, *HML-2_sLTR(1p13.2)* and *HML-2_sLTR(19q12)* show the highest identity to the 5′-LTR of *HML-2(12q13.2)* (GenBank/EMBL/DDBJ accession number: JN675067; identity 97.9%) and the LTRs of *HML-2(3q13.2)* (GenBank/EMBL/DDBJ accession number: JN675022; identity 99.5%), respectively. *HML-2(12q13.2)* is also insertionally polymorphic (Supplementary Table I, available at *Carcinogenesis* Online). *HML-2(3q13.2)*, also termed *K106*, is deletionally polymorphic and is thought to be one of the youngest HERVs because of the identical LTRs at both of its ends (16). *HML-2_sLTR(1p13.2)* and *HML-2_sLTR(19q12)* may have arisen from the lineages of *HML-2(12q13.2)* and *HML-2(3q13.2)*, respectively, suggesting genomic variations in the ancestral genome. Investigating the haplotype of polymorphic HML-2 elements in various ethnic groups is probably to produce extremely interesting results.

A putative mutation was detected within the duplicated target site of *HML-2_sLTR(19q12)* (Figure 3D). An example of a single-nucleotide mutation within a duplicated target site has been previously reported for HERV-H (39). Because the mutation within the duplicated target site of *HML-2_sLTR(19q12)* was found in all seven samples that were tested, the mutation may have occurred very early after the integration of the provirus into a host genome.

Little evidence of an association between an insertion polymorphism of HERV and human diseases has existed. In previous studies, neither seminoma nor breast cancer was significantly associated with insertion polymorphisms of *HERV-K113* and *HERV-K115*, both of which are known as HML-2 proviruses (40,41). Although multiple sclerosis and Sjögren's syndrome were significantly associated with an insertion polymorphism of *HERV-K113* in a small population study (42), multiple sclerosis was not significantly associated with the insertion polymorphism of *HERV-K113* in a larger population study (43). In these previous studies on disease, the gene frequency, but not the genotype distribution, was calculated for the investigation. Allele frequencies of those HERV loci would be too low to statistically evaluate the genetic association. In fact, each homozygosity of *HERV-K113* and *HERV-K115* was not or rarely detected (41,42). To use larger populations or to target another polymorphic HERV locus with a relatively high allele frequency will be necessary for a statistical evaluation involving the calculation of the genotype distribution. Here, *HML-2_sLTR(1p13.2)* homozygosity was shown to be associated with lung adenocarcinoma in female never-smokers among lung adenocarcinoma patients aged 60 years and over (Table II). Our study is the first report of a genetic association between an insertion polymorphism of HERV and cancer. Remarkably, a never-smoking status, East Asian ethnicity, female sex and adenocarcinoma histology have been shown to be correlated with the presence of somatic active mutations in epidermal growth factor receptor (EGFR), which is associated with a strong response to EGFR tyrosine kinase inhibitor therapy (44). The genetic background of *HML-2_sLTR(1p13.2)* homozygosity and the activated EGFR pathway

may possibly lead to synergic effects on lung adenocarcinoma survival in aged female never-smoker, as suggested below.

Transcriptional regulatory regions that are conserved among LTRs, including a putative TATA box, polyadenylation signal, initiator sequence and enhancer core, are also present in *HML-2_sLTR(1p13.2)* and *HML-2_sLTR(19q12)* (Supplementary Figure 3, available at *Carcinogenesis* Online) (4,5). Since an LTR of HML-2 exhibited an enhancer ability in human cell cultures (7), *HML-2_sLTR(1p13.2)* and *HML-2_sLTR(19q12)* are also expected to affect the expressions of surrounding genes. In the Aichi cohort study, a high prevalence of *HML-2_sLTR(1p13.2)* homozygosity was detected among lung adenocarcinoma patients aged 60 years and over, whereas a low prevalence of this genotype was detected in the younger patient group. A test of heterogeneity between the two age groups indicated a significant difference ($P = 0.006$), showing that the presence of *HML-2_sLTR(1p13.2)* leads to an age-dependent biological fluctuation in the susceptibility to lung adenocarcinoma. The transcriptional activities of LTRs can be regulated by CpG methylation (45), and age-dependent changes in DNA methylation have been reported and discussed (46). The gross methylation level of LTRs reportedly decreased between subjects aged 40–63 years and those aged 64–83 years in a study using peripheral blood mononuclear cells (47). Thus, age-dependent changes in CpG methylation on *HML-2_sLTR(1p13.2)* may also occur in lung tissue and might contribute to the fluctuation in the susceptibility to lung adenocarcinoma. There is the possibility that the hyper- and hypomethylations of *HML-2_sLTR(1p13.2)* in young and elderly subjects affect oncogenic expressions negatively and positively, respectively, although the possibility that menopause may affect the *HML-2_sLTR(1p13.2)*-mediated transcriptional regulation of surrounding genes cannot be excluded in the present study. Interestingly, in a previous study examining schizophrenia, the frequency of *HERV-K115*-positive patients differed significantly between younger (<18 years) and older (>18 years) age groups (48). Thus, other insertionally polymorphic LTRs with biased distributions in young and elderly subject groups are probably to exist in specific patient groups.

CHIA (also known as *AMCase*) is one of the genes in the vicinity of *HML-2_sLTR(1p13.2)*, and its expression might be possibly regulated through the enhance ability of *HML-2_sLTR(1p13.2)*. CHIA is induced in lung epithelial cells and macrophages at sites of Th2 inflammation, and interleukin-13 is necessary and sufficient to stimulate CHIA production (49). CHIA has also been shown to inhibit lung epithelial cell apoptosis through the activation of the PI3/Akt pathway (50). Thus, it is speculated that age-dependent changes in CpG methylation on *HML-2_sLTR(1p13.2)* affect the fluctuation of the CHIA expression and that overproduced CHIA causes the carcinogenesis of lung adenocarcinoma through the antiapoptotic effects in the elderly. These speculations are considered as a model to explain the age-dependent manner of *HML-2_sLTR(1p13.2)* in the carcinogenesis of lung adenocarcinoma. In addition, CHIA is secreted from lung epithelial cells via the EGFR-dependent pathway (51). The secretion of CHIA might be promoted by somatic active mutations of EGFR, resulting in widespread antiapoptotic effects throughout the lung epithelium. A detailed investigation of *HML-2_sLTR(1p13.2)*-surrounding genes, including the quantification of CHIA expression in lung tumor tissues with and without the *HML-2_sLTR(1p13.2)* allele, and genetic studies involving large data sets connected with an EGFR mutation profile will be necessary for further validation.

Modest associations between *HML-2_sLTR(19q12)* homozygosity and lung squamous cell carcinoma and between *HML-2_sLTR(1p13.2)* homozygosity and proximal colon cancer were detected in male smokers (Table II) and women (Supplementary Table VII, available at *Carcinogenesis* Online), respectively. Although these modest associations will be the subjects of future investigations, they suggest that insertionally polymorphic HML-2_sLTR may have more functional effects on the host status, notably the susceptibility to cancer, than previously expected.

With recent information regarding the unexpectedly high variety in copy numbers in the human genome, we have fully realized the complexity of the human genetic nature in terms of disease susceptibility. In fact, a considerable proportion of human beings do not have some of the detoxification enzymes arising from deletion/insertion-type polymorphisms (52). The methodology for the identification and discovery of this type of polymorphism is still largely heuristic, and no effective algorithms are yet available, even in the era of massive parallel sequencing. Our presentation here, for the first time, demonstrates that the intentional isolation of insertion/deletion-type polymorphisms is a promising approach to elucidate genetic susceptibility to diseases in human beings.

Supplementary material

Supplementary Tables I–VII and Figures 1–4 can be found at http://carcin.oxfordjournals.org/

Funding

Grants-in-Aid for Young Scientists (B) from the Japan Society for the Promotion of Science (19790286, 23790396, 25870298); priority areas from the Japanese Ministry of Education, Culture, Sports, Science and Technology (20014007, 22659072, 221S0001); Grants-in-Aid for Cancer Research from the Japanese Ministry of Health (23120201, 10103838); Grants-in-Aid for the United States-Japan Cooperative Medical Science Program; the National Cancer Center Research and Development Fund (23-A-4); the Smoking Research Foundation; the Takeda Science Foundation.

Acknowledgements

We thank Mr Kageyama (Hamamatsu University School of Medicine) for technical support. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Conflict of Interest Statement: None declared.

References

1. Lander,E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
2. Jern,P. *et al.* (2008) Effects of retroviruses on host genome function. *Annu. Rev. Genet.*, **42**, 709–732.
3. Subramanian,R.P. *et al.* (2011) Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology*, **8**, 90.
4. Kovalskaya,E. *et al.* (2006) Functional human endogenous retroviral LTR transcription start sites are located between the R and U5 regions. *Virology*, **346**, 373–378.
5. Cohen,C.J. *et al.* (2009) Endogenous retroviral LTRs as promoters for human genes: a critical assessment. *Gene*, **448**, 105–114.
6. Goering,W. *et al.* (2011) Selective changes of retroelement expression in human prostate cancer. *Carcinogenesis*, **32**, 1484–1492.
7. Ruda,V.M. *et al.* (2004) Tissue specificity of enhancer and promoter activities of a HERV-K(HML-2) LTR. *Virus Res.*, **104**, 11–16.
8. Baillie,J.K. *et al.* (2011) Somatic retrotransposition alters the genetic landscape of the human brain. *Nature*, **479**, 534–537.
9. Maksakova,I.A. *et al.* (2006) Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. *PLoS Genet.*, **2**, e2.
10. Iskow,R.C. *et al.* (2010) Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell*, **141**, 1253–1261.
11. Medstrand,P. *et al.* (1998) Human-specific integrations of the HERV-K endogenous retrovirus family. *J. Virol.*, **72**, 9782–9787.
12. Stewart,C. *et al.* (2011) A comprehensive map of mobile element insertion polymorphisms in humans. *PLoS Genet.*, **7**, e1002236.
13. Wang,J. *et al.* (2006) dbRIP: a highly integrated database of retrotransposon insertion polymorphisms in humans. *Hum. Mutat.*, **27**, 323–329.
14. Brock University. dbRIP. http://dbrip.brocku.ca/ (7 May 2013, date last accessed).
15. Moyes,D. *et al.* (2007) Insertional polymorphisms: a new lease of life for endogenous retroviruses in human disease. *Trends Genet.*, **23**, 326–333.

16. Jha,A.R. *et al.* (2011) Human endogenous retrovirus K106 (HERV-K106) was infectious after the emergence of anatomically modern humans. *PLoS One*, **6**, e20234.
17. Turner,G. *et al.* (2001) Insertional polymorphisms of full-length endogenous retroviruses in humans. *Curr. Biol.*, **11**, 1531–1535.
18. Dewannieux,M. *et al.* (2006) Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. *Genome Res.*, **16**, 1548–1556.
19. Lee,Y.N. *et al.* (2007) Reconstitution of an infectious human endogenous retrovirus. *PLoS Pathog.*, **3**, e10.
20. Denne,M. *et al.* (2007) Physical and functional interactions of human endogenous retrovirus proteins Np9 and rec with the promyelocytic leukemia zinc finger protein. *J. Virol.*, **81**, 5607–5616.
21. Jemal,A. *et al.* (2011) Global cancer statistics. *CA Cancer J. Clin.*, **61**, 69–90.
22. Iacopetta,B. (2002) Are there two sides to colorectal cancer? *Int. J. Cancer*, **101**, 403–408.
23. Derks,S. *et al.* (2008) Integrated analysis of chromosomal, microsatellite and epigenetic instability in colorectal cancer identifies specific associations between promoter methylation of pivotal tumour suppressor and DNA repair genes and specific chromosomal alterations. *Carcinogenesis*, **29**, 434–439.
24. Girard,N. *et al.* (2010) Analysis of genetic variants in never-smokers with lung cancer facilitated by an Internet-based blood collection protocol: a preliminary report. *Clin. Cancer Res.*, **16**, 755–763.
25. Kono,S. *et al.* (2004) A case-control study of colorectal cancer in relation to lifestyle factors and genetic polymorphisms: design and conduct of the Fukuoka colorectal cancer study. *Asian Pac. J. Cancer Prev.*, **5**, 393–400.
26. Yin,G. *et al.* (2007) Alcohol dehydrogenase and aldehyde dehydrogenase polymorphisms and colorectal cancer: the Fukuoka Colorectal Cancer Study. *Cancer Sci.*, **98**, 1248–1253.
27. Tao,H. *et al.* (2008) Association between genetic polymorphisms of the base excision repair gene MUTYH and increased colorectal cancer risk in a Japanese population. *Cancer Sci.*, **99**, 355–360.
28. Ochman,H. *et al.* (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics*, **120**, 621–623.
29. New England Biolabs, Inc. Restriction site densities deviating from expectations. http://tools.neb.com/~vincze/gnsites/ (7 May 2013, date last accessed).
30. Li,W.H. *et al.* (2001) Evolutionary analyses of the human genome. *Nature*, **409**, 847–849.
31. Bussink,A.P. *et al.* (2007) Evolution of mammalian chitinase(-like) members of family 18 glycosyl hydrolases. *Genetics*, **177**, 959–970.
32. Leung,S.Y. *et al.* (2006) Comprehensive analysis of 19q12 amplicon in human gastric cancers. *Mod. Pathol.*, **19**, 854–863.
33. Qi,X. *et al.* (2012) Retrotransposon profiling of RNA polymerase III initiation sites. *Genome Res.*, **22**, 681–692.
34. Kahyo,T. *et al.* (2011) A novel tumor-derived SGOL1 variant causes abnormal mitosis and unstable chromatid cohesion. *Oncogene*, **30**, 4453–4463.
35. Shinmura,K. *et al.* (2011) Reduced expression of MUTYH with suppressive activity against mutations caused by 8-hydroxyguanine is a novel predictor of a poor prognosis in human gastric cancer. *J. Pathol.*, **225**, 414–423.
36. Sugimura,H. *et al.* (2011) Genetic susceptibility to lung cancer. *Front. Biosci. (Schol. Ed).*, **3**, 1463–1477.
37. Hughes,J.F. *et al.* (2004) Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: implications for human and viral evolution. *Proc. Natl Acad. Sci. USA*, **101**, 1668–1672.
38. Belshaw,R. *et al.* (2005) Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K(HML2): implications for present-day activity. *J. Virol.*, **79**, 12507–12514.
39. de Parseval,N. *et al.* (2001) Characterization of the three HERV-H proviruses with an open envelope reading frame encompassing the immunosuppressive domain and evolutionary history in primates. *Virology*, **279**, 558–569.
40. Ruprecht,K. *et al.* (2008) Endogenous retroviruses and cancer. *Cell. Mol. Life Sci.*, **65**, 3366–3382.
41. Burmeister,T. *et al.* (2004) Insertional polymorphisms of endogenous HERV-K113 and HERV-K115 retroviruses in breast cancer patients and age-matched controls. *AIDS Res. Hum. Retroviruses*, **20**, 1223–1229.
42. Moyes,D.L. *et al.* (2005) The distribution of the endogenous retroviruses HERV-K113 and HERV-K115 in health and disease. *Genomics*, **86**, 337–341.
43. Moyes,D.L. *et al.* (2008) HERV-K113 is not associated with multiple sclerosis in a large family-based study. *AIDS Res. Hum. Retroviruses*, **24**, 363–365.
44. Sun,S. *et al.* (2007) Lung cancer in never smokers–a different disease. *Nat. Rev. Cancer*, **7**, 778–790.
45. Lavie,L. *et al.* (2005) CpG methylation directly regulates transcriptional activity of the human endogenous retrovirus family HERV-K(HML-2). *J. Virol.*, **79**, 876–883.
46. Richardson,B. (2003) Impact of aging on DNA methylation. *Ageing Res. Rev.*, **2**, 245–261.
47. Jintaridth,P. *et al.* (2010) Distinctive patterns of age-dependent hypomethylation in interspersed repetitive sequences. *Physiol. Genomics*, **41**, 194–200.
48. Otowa,T. *et al.* (2006) Insertional polymorphism of endogenous retrovirus HERV-K115 in schizophrenia. *Neurosci. Lett.*, **408**, 226–229.
49. Zhu,Z. *et al.* (2004) Acidic mammalian chitinase in asthmatic Th2 inflammation and IL-13 pathway activation. *Science*, **304**, 1678–1682.
50. Hartl,D. *et al.* (2009) Acidic mammalian chitinase regulates epithelial cell apoptosis via a chitinolytic-independent mechanism. *J. Immunol.*, **182**, 5098–5106.
51. Hartl,D. *et al.* (2008) Acidic mammalian chitinase is secreted via an ADAM17/epidermal growth factor receptor-dependent pathway and stimulates chemokine production by pulmonary epithelial cells. *J. Biol. Chem.*, **283**, 33472–33482.
52. Sharp,A.J. *et al.* (2006) Structural variation of the human genome. *Annu. Rev. Genomics Hum. Genet.*, **7**, 407–442.

Received May 14, 2013; revised July 4, 2013; accepted July 16, 2013

CD74-ROS1 fusion transcripts in resected non-small cell lung carcinoma

SHUN MATSUURA[1,2], KAZUYA SHINMURA[1], TAKAHARU KAMO[1], HISAKI IGARASHI[1], KYOKO MARUYAMA[1], MARI TAJIMA[1], HIROSHI OGAWA[4], MASAYUKI TANAHASHI[5], HIROSHI NIWA[5], KAZUHITO FUNAI[3], TAKASHI KOHNO[6], TAKAFUMI SUDA[2] and HARUHIKO SUGIMURA[1]

Departments of [1]Tumor Pathology, [2]Internal Medicine 2 and [3]Surgery 1, Hamamatsu University School of Medicine, Hamamatsu, Shizuoka 431-3192; Divisions of [4]Pathology and [5]Thoracic Surgery, Respiratory Disease Center, Seirei Mikatahara General Hospital, Hamamatsu, Shizuoka 433-8558; [6]Division of Genome Biology, National Cancer Center Research Institute, Chuo-ku, Tokyo 104-0045, Japan

Received April 11, 2013; Accepted June 7, 2013

DOI: 10.3892/or.2013.2630

Abstract. The recent discovery of fusion oncokinases in a subset of non-small cell lung carcinomas (NSCLCs) is of considerable clinical interest, since NSCLCs that express such fusion oncokinases are reportedly sensitive to kinase inhibitors. To better understand the role of recently identified ROS1 and RET fusion oncokinases in pulmonary carcinogenesis, we examined 114 NSCLCs for SLC34A2-ROS1, EZR-ROS1, CD74-ROS1 and KIF5B-RET fusion transcripts using RT-polymerase chain reaction and subsequent sequencing analyses. Although the expression of SLC34A2-ROS1, EZR-ROS1, or KIF5B-RET fusion transcripts was not detected in any of the cases, the expression of CD74-ROS1 fusion transcripts was detected in one (0.9%) of the 114 NSCLCs. The fusion occurred between exon 6 of CD74 and exon 34 of ROS1 and was an in-frame alteration. The mutation was detected in a woman without a history of smoking. Histologically, the carcinoma was an adenocarcinoma with a predominant acinar pattern; notably, a mucinous cribriform pattern and a solid signet-ring cell pattern were also observed in part of the adenocarcinoma. ROS1 protein overexpression was immunohistochemically detected in a cancer-specific manner in both the primary cancer and the lymph node metastatic cancer. No somatic mutations were detected in the mutation cluster regions of the *KRAS*, *EGFR*, *BRAF* and *PIK3CA* genes and the entire coding region of *p53* in the carcinoma, and the expression of ALK fusion was negative. The above results suggest that CD74-ROS1 fusion is involved in the carcinogenesis of a subset of NSCLCs and may contribute to the elucidation of the characteristics of ROS1 fusion-positive NSCLC in the future.

Correspondence to: Dr Kazuya Shinmura, Department of Tumor Pathology, Hamamatsu University School of Medicine, 1-20-1 Handayama, Higashi Ward, Hamamatsu, Shizuoka 431-3192, Japan
E-mail: kzshinmu@hama-med.ac.jp

Key words: fusion gene, non-small cell lung carcinoma, ROS1, CD74, RET, ALK

Introduction

The *ALK* fusion gene is a key oncogenic driver in a subset of patients with non-small cell lung carcinomas (NSCLCs) (1-5). EML4-ALK, consisting of the N-terminal portion of EML4 ligated to the intracellular region of the receptor-type protein tyrosine kinase ALK, has been detected in ~2-7% of NSCLC patients (1-6). EML4-ALK-positive NSCLC is associated with several characteristics, such as early-onset, a never- or light-smoking history, adenocarcinoma and mutual exclusiveness with *EGFR* or *KRAS* mutations (7). Recently, crizotinib, a small molecule inhibitor of ALK, was shown to selectively inhibit the growth of ALK-positive NSCLC (6-8), indicating that a subclass of NSCLC patients are likely to benefit clinically from an ALK inhibitor. Therefore, molecular data regarding oncogenic fusion may have a significant clinical impact.

Both ROS1 and RET receptor tyrosine kinases have recently been identified as oncogenic fusions in NSCLC (4,5,9-17). The expression of these fusion proteins transforms noncancerous cells (4,14). Among some types of such fusion forms, SLC34A2-ROS1, EZR-ROS1, CD74-ROS1 and KIF5B-RET are relatively recurrent (4,5,9-17). However, since the incidence of ROS1 or RET fusions is less than that of ALK fusions in NSCLC, only a small number of NSCLC patients with ROS1 or RET fusions has thus far been identified. In the present study, to contribute to the elucidation of the characteristics of ROS1 or RET fusion-positive NSCLC, we examined 114 NSCLCs derived from Japanese patients for the expression of ROS1 and RET fusion transcripts and pathohistologically and molecularly characterized those fusions that were detected.

Materials and methods

Primary lung carcinoma. Samples of surgical specimens were obtained from 114 NSCLC patients who underwent surgery for

cancer at Mikatahara Seirei General Hospital and Hamamatsu University Hospital. Informed consent to use the resected tissues for genetic analysis was obtained from all the patients and the study was approved by the Institutional Review Boards (IRBs) of Hamamatsu University School of Medicine and Mikatahara Seirei General Hospital. The clinicopathological profiles of the cases are shown in Table I. The staging and histological classification were based on the World Health Organization system.

Detection of ROS1 and RET gene fusion transcripts using reverse transcription (RT)-polymerase chain reaction (PCR). Total RNA was extracted from the lung tissue samples using an RNeasy kit (Qiagen, Valencia, CA, USA) and was converted to first-strand cDNA using a SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen, Carlsbad, CA, USA) according to the supplier's protocol. PCR was performed in 20-μl reaction mixtures containing HotStarTaq DNA Polymerase (Qiagen). The following PCR primers were used: 5'-GGGATTGGGATATTGATTTTAC-3' and 5'-AGCTCA GCCAACTCTTTGTCT-3' for the SLC34A2-ROS1 fusion transcript; 5'-ACCGTGGAGAGAGAGAAAGAG-3' and 5'-AGCTCAGCCAACTCTTTGTCT-3' for the EZR-ROS1 fusion transcript; 5'-ATTGGCTCCTGTTTGAAATG-3' and 5'-TTATAAGCACTGTCACCCCTTC-3' for the CD74-ROS1 fusion transcript; and 5'-TCGGCAACTTTAGCGAGTAT-3' and 5'-TTCTCTTTCAGCATCTTCACG-3' for the KIF5B-RET fusion transcript. The PCR products were fractionated using electrophoresis on an agarose gel and were stained with ethidium bromide. PCR-amplified products were purified with ExoSAP-IT (GE Healthcare Bio-Science, Piscataway, NJ, USA) and were sequenced directly using a BigDye® Terminator Cycle Sequencing Reaction Kit and the ABI 3130 Genetic Analyzer (both from Applied Biosystems, Tokyo, Japan).

Immunohistochemical staining. Sections of formalin-fixed, paraffin-embedded tissue samples were used for immunohistochemical staining performed using the EnVision (Dako ChemMate) kit (Dako, Kyoto, Japan). The primary antibodies were as follows: anti-thyroid transcription factor-1 (TTF-1), anti-napsin A, anti-CK14 (all from Novocastra Laboratories, Newcastle, UK), anti-p63 (Japan Tanner, Osaka, Japan), anti-CD56 (Novocastra Laboratories), anti-chromogranin A (Dako), anti-synaptophysin (Novocastra Laboratories) and anti-ROS1 (clone D4D6; Cell Signaling Technology, Beverly, MA, USA). Hematoxylin and eosin (H&E) staining was also performed.

Mutation search. Genomic DNA was extracted from the lung tissue samples containing ROS1 fusion transcripts using a DNeasy kit (Qiagen) and was examined for somatic mutations in the DNA sequences of mutation cluster regions (hot spots) in the *KRAS*, *EGFR*, *BRAF* and *PIK3CA* genes and in the entire coding sequence of the *p53* gene. PCR amplification was performed in 20-μl reaction mixtures containing HotStarTaq DNA Polymerase (Qiagen). The following PCR primers were used: 5'-AAAGGTACTGGTGGAGTATTTG-3' and 5'-GTCCTGCACCAGTAATATGC-3' for *KRAS* (exon 2); 5'-AATCCAGACTGTGTTTCTCC-3' and 5'-ATATTATATC ATGGCATTAGC-3' for *KRAS* (exon 3); 5'-GCAATATCAGCC TTAGGTGCGGTC-3' and 5'-CATAGAAAGTGAACATTTA GGATGTG-3' for *EGFR* (exon 19); 5'-CTAACGTTCGCCAGC CATAAGTCC-3' and 5'-GCTGCGAGCTCACCCAGAAT GTCTGG-3' for *EGFR* (exon 21); 5'-AATTTTTCTTAAG GGGATCTCTTCC-3' and 5'-GCGAACAGTGAATATTT CCTTTG-3' for *BRAF* (exon 11); 5'-ACCTAAACTCTTCAT AATGCTTGCTC-3' and 5'-CTTCAATGACTTTCTAGTAA CTCAGCAG-3' for *BRAF* (exon 15); 5'-TTAGATTGGTT CTTTCCTGTCTCTG-3' and 5'-TCCAATAGGTATGGTAA AAACATGC-3' for *PIK3CA* (exon 9); 5'-GTGACATT TGAGCAAAGACCTG-3' and 5'-CTGTTCATGGATTGT GCAATTC-3' for *PIK3CA* (exon 20); 5'-TTGGAAGTG TCTCATGCTGG-3' and 5'-AAGAGCAGTCAGAGGACC AGG-3' for *p53* (exons 2 and 3); 5'-ACCTGGTCCTCT GACTGCTC-3' and 5'-TTGAAGTCTCATGGAAGCCAG-3' for *p53* (exon 4); 5'-TTGTGCCCTGACTTTCAACTC-3' and 5'-ACCAGCCCTGTCGTCTCTC-3' for *p53* (exon 5); 5'-TCAGATAGCGATGGTGAGCAG-3' and 5'-GGAGGT CAAATAAGCAGCAGG-3' for *p53* (exon 6); 5'-CTCATC TTGGGCCTGTGTTATC-3' and 5'-GAAGAAATCGGTA AGAGGTGGG-3' for *p53* (exon 7); 5'-CTGCCTCTTGCTT CTCTTTTCC-3' and 5'-ACTTTCCACTTGATAAGAGGT CCC-3' for the *p53* (exons 8 and 9); 5'-ATACTTACTTCT CCCCCTCCTCTG-3' and 5'-GGATGAGAATGGAATCCT ATGG-3' for *p53* (exon 10); and 5'-TGATGTCATCTC TCCTCCCTG-3' and 5'-TTGCAAGCAAGGGTTCAAAG-3' for *p53* (exon 11). Sequencing was performed as described above.

Table I. Summary of the clinicopathological profiles of the patients.

Characteristic	n
No. of patients	114
Age, years (mean ± SD)	68.5±6.0
Gender, n (%)	
Male	87 (76.3)
Female	27 (23.7)
Smoking, n (%)	
Current smokers	49 (43.0)
Ex-smokers	37 (32.5)
Non-smokers	28 (24.5)
Histology, n (%)	
Adenocarcinoma	69 (60.5)
Squamous cell carcinoma	39 (34.2)
Adenosquamous cell carcinoma	4 (3.5)
Others	2 (1.8)
Stage, n (%)	
I	54 (47.4)
II	33 (28.9)
III	27 (23.7)
IV	0 (0.0)

SD, standard deviation.