Human-Specific HERV-K Insertion Causes Genomic Variations in the Human Genome

Wonseok Shin1, Jungnam Lee1, Seung-Yeol Son2, Kung Ahn3, Heui-Soo Kim3, Kyudong Han1*

1 Department of Nanobiomedical Science and WCU Research Center, Dankook University, Cheonan, Republic of Korea, 2 Department of Microbiology, College of Advanced Science, Dankook University, Cheonan, Republic of Korea, 3 Department of Biological Sciences, College of Natural Sciences, Pusan National University, Busan, Republic of Korea

Abstract

Human endogenous retroviruses (HERV) sequences account for about 8% of the human genome. Through comparative genomics and literature mining, we identified a total of 29 human-specific HERV-K insertions. We characterized them focusing on their structure and flanking sequence. The results showed that four of the human-specific HERV-K insertions deleted human genomic sequences via non-classical insertion mechanisms. Interestingly, two of the human-specific HERV-K insertion loci contained two HERV-K internals and three LTR elements, a pattern which could be explained by LTR-LTR ectopic recombination or template switching. In addition, we conducted a polymorphic test and observed that twelve out of the 29 elements are polymorphic in the human population. In conclusion, human-specific HERV-K elements have inserted into human genome since the divergence of human and chimpanzee, causing human genomic changes. Thus, we believe that human-specific HERV-K activity has contributed to the genomic divergence between humans and chimpanzees, as well as within the human population.

Introduction

Repetitive mobile elements are responsible for half of the human genome. Among them, human endogenous retroviruses (HERVs) and related sequences account for ~8% of the human genome [1]. It is thought that HERVs are derived from exogenous retrovirus infections early in the evolution of primates because they have a similar structure to the provirus of an infectious virus [2]. A full-length HERV element is approximately 9.5 kb in length and consists of an internal region of four essential viral genes (gag, pro, pol, and env) and two long terminal repeats (LTRs); gag stands for group-specific antigen which is the retroviral capsid protein, pro encodes for a protease, and pol contains a reverse transcriptase domain [3,4]. HERVs are distinguished from other LTR retrotransposons by the presence of the envelope (env) gene, which codes for viral membrane proteins [5]. The LTRs contain many regulatory elements such as promoters, enhancers, and polyadenylation signals required for retroviral gene expression [6,7].

Since the initial infection of HERV into its host genome, the elements have lost their ability to synthesize mature retroviral particles by accumulating mutations preventing them from infecting other cells [8]. Nonetheless, they have successfully propagated within genomes via retrotransposition and vertical inheritance, reaching ~203,000 copies in the human genome [1]. HERVs fall into three different classes (I-III) based on sequence similarity to different genera of infectious retroviruses, and each class comprises many families with independent origins [1,3]. There are 31 HERV families in the human genome and they are named according to the specificity of the tRNA primer-binding site [3,9]. It was reported that most HERV families underwent radiations in their host genomes after the divergence of Old and New World monkeys [8]. Among the three HERV classes, class II HERVs exist in the lowest frequency in the human genome, but they include the HERV-K family, which is the youngest family and is known to have actively mobilized since the divergence of humans and chimpanzees [1,10]. The HERV-K subfamily could be integrated and endogenized into the human genome by germ-line infection, which was supported by the evidence of purifying selection on the env gene of HERV-K elements [11].

It has been suggested that the HERV-K family is the most biologically active family because it retains the ability to encode functional retroviral proteins and produce retrovirus-like particles [12,13,14]. Due to this, the HERV-K family has been the subject of many studies but to date no functional provirus capable of producing infectious particles has been detected [10]. Although the HERV-K family emerged in the catarrhine lineage prior to the divergence of hominoids and Old World monkeys, some of its members inserted into the human genome after the divergence of humans and chimpanzees [8]. Thus, the HERV-K family may have contributed to the genomic differences between humans and chimpanzees through species-specific insertion and subsequent related genomic rearrangements. In this study, we identified 29 human-specific HERV-K elements in the human genome and examined the human genomic changes caused by these insertions. Our analyses focused on the mechanisms through which the HERV-K insertions caused the observed changes. In addition, we
conducted a polymorphism test of the HERV-K insertions in human populations, the result of which indicates that HERV-K elements may also be contributing to genomic variations within the human species.

Results and Discussion

Identification of Human-specific HERV-K Insertions

To identify human-specific HERV-K elements, we first extracted 2,618 HERV-K elements from the human genome. However, some of these elements contained other internal non-HERV repeat element insertions or internal sequence deletions. In these cases, each HERV-K fragment was counted as a separate element by the tool we used to extract them, rather than counting the un-fragmented element only once. Thus, we manually inspected the HERV-K candidate loci and reassembled all fragmentary elements, resulting in a revised total of 1,390 loci (Table 1). To detect human-specific insertion loci in these 1,390 HERV-K elements, we examined the orthologous loci of each human-derived HERV-K element in the chimpanzee, orangutan, and rhesus macaque genomes. In this way, we identified 26 human-specific HERV-K loci in the human genome. Four previous studies have attempted to identify human-specific HERV-K loci [15,16,17,18]. A comparison of our results showed that our strategy recovered five human-specific HERV-K loci that these previous studies missed. However, three of the human-specific HERV-K loci previously reported in the literature (HERV-K103, 113, and 134) were missing from our dataset. We examined these three loci in detail. Two were solitary LTRs in the human reference genome sequence and since we did not include solitary LTRs in our dataset of human-specific HERV-K loci, it is unsurprising these two loci were missed by our strategy. Close examination of the third missing locus revealed this locus to be polymorphic in human populations. In other words, we were unable to detect the locus because the HERV-K element is absent in the human reference genome sequence. Given this, we assert that our strategy to identify human-specific HERV-K elements in the human reference genome is robust. Thus, as shown in the Figure S1, at least 29 human-specific HERV-K elements are existing in the human genome.

We characterized the human-specific HERV-K elements focusing on their size. A full-length HERV-K element consists of \(\sim 7.5\) kb of internal region and two LTRs, each of which is \(\sim 1\) kb. However, most of the HERV-K elements in the human genome contain internal deletions of variable sizes. In this study, we considered the element whose internal region is \(>7\) kb to be a full-length element. The size of HERV-K internal regions ranged from 97 to 7546 bp, and 17 out of the 29 human-specific HERV-K elements were full-length elements according to our criterion. HERV-K elements have been grouped into two types, type I and type II, according to the presence/absence of a 292 bp sequence at the pol-env boundary of the elements [2]. Only type II elements contain the 292 bp sequence. We further examined the full-length human-specific HERV-K elements. As shown in Table 2, eight and nine elements are identified as type I and type II, respectively, including three previously studied insertions [16,18].

Additionally, we found two interesting human-specific HERV-K loci of non-standard sequence architecture. Each of these consists of two HERV-K internals and three LTRs. One of the two loci, HERV-K108, may have resulted from ectopic homologous recombination between two different LTRs, the mechanism for which was introduced in another study on HERV-K and is depicted in Figure 1A [19]. The three LTRs of HERV-K108 showed a high degree of sequence similarity and were closely related in the phylogenetic tree in Figure 2. The other locus, HERV-K124 also contains three different LTRs. However, it was unclear what mechanism may be responsible for the observed sequence architecture of this locus. If LTR-LTR recombination were to explain this locus, we would expect the three LTRs to have a high degree of sequence similarity to one another, but the 3’ and internal HERV-K124 LTRs are truncated and inverted relative to 5’ HERV-K124 LTR. We therefore speculate that HERV-K124 was generated in two steps: LTR inversion and template switching, as shown in Figure 1B. Although the LTR inversion is a rare event, a possible mechanism responsible for the LTR inversion was suggested in one of previous studies on HERV-K [20].

Genomic Environment of Human-specific HERV-K Insertions

We aligned the human-specific HERV-K elements based on their LTR sequences except for eight loci because those elements contained LTRs that were too short (23–257 bp) resulting in ambiguity in the alignment. Next, we reconstructed the phylogenetic network of these LTRs. It is known that the two LTRs of an HERV element tend to have a high sequence identity to one another. As shown in Figure 2, this expected within-element sequence identity was found in all of our loci except HERV-K115. We suspect that gene conversion may have led to the differences observed between the two LTR sequences of the HERV-K115 [21].

To examine the genomic environment of the human-specific HERV-K insertions, we analyzed the GC content and gene density of genomic regions flanking the elements (Table S1). GC content was calculated for the 20 kb of flanking genomic sequence on each side of each locus. The GC content of these flanking regions averaged 41.6%. This is only slightly higher than the human reference genomic average GC content of 41% [1]. In addition, we analyzed the gene density of the 1 Mb of flanking genomic sequence to each side of the human-specific HERV-K elements and the results are described in Table S1. The gene density of these insertions averaged about 17 genes per Mb, which is substantially higher than the \(\sim 10\) genes per Mb average reported for the human genome [1]. It has been previously reported that HERV-K elements are preferentially integrated into GC-rich regions, and thus gene-rich regions [22], and our findings are consistent to with this assertion.

Polymorphic Distribution of Human-specific HERV-K Insertions

The HERV-K family has been shown to be actively mobilizing in the human genome since the divergence of human and chimpanzee, and thus some of these elements are likely to be polymorphic in the human population. To evaluate the polymorphism levels associated with human-specific HERV-K loci, we

| Classification | No. of loci |
|----------------|------------|
| Computationally predicted HERV-K loci | 1390 |
| Number of human-specific HERV-K insertion events | 29 |
| Full-length human-specific HERV-K insertion | 17 |
| Truncated human-specific HERV-K insertion | 8 |
| Non-classical insertion of HERV-K | 4 |
| doi:10.1371/journal.pone.0060605.t001 |

Table 1. Summary of human-specific HERV-K insertions.
Table 2. The structural characterization of human-specific full-length HERV-K.

| Type | HERV  | Chromosomal position (hg19) | Length (bp) | Comment | Stop codon/Region |
|------|-------|-----------------------------|-------------|---------|------------------|
| I    | K101  | chr 22: 18926187-18935361    | 968/964/7243| In frame | pol broken       |
|      |       |                             |             |         | TGA/pro          |
| I    | K102  | chr 1: 155596457-15605636    | 968/968/7244| In frame | pol-env fusion   |
|      |       |                             |             |         | TGA/gag          |
| I    | K103  | chr 10: 27182399-27183366    | 968/968/7245| In frame | pol-env fusion   |
|      |       |                             |             |         | env broken       |
| I    | K106  | chr 3: 112743124-112752282   | 960/960/7239| In frame | env broken       |
|      |       |                             |             |         |                  |
| I    | K107  | chr 5: 156084717-156093896   | 968/968/7244| In frame | pol-env fusion   |
|      |       |                             |             |         |                  |
| I    | K108a | chr 7: 462057-4631528        | 968/968/7535| Dual internal sequences, triple LTRs | TAG/gag, env |
|      |       |                             |             |         |                  |
| I    | K108b | chr 7: 4630561-4640031       | 968/968/7535| TAG/gag  |                  |
|      |       |                             |             |         |                  |
| I    | K109  | chr 6: 78426662-78436083     | 960/960/7502| –        | TAG, TGA/pol     |
|      |       |                             |             |         |                  |
| I    | K113  | chr 19: 21841536-21841541    | 968/968/7536| –        | –                |
|      |       |                             |             |         |                  |
| I    | K115  | chr 8: 7353397-7364859       | 960/968/7535| –        | –                |
|      |       |                             |             |         |                  |
| I    | K118  | chr 11: 101565794-101575259  | 968/968/7530| –        | TGA/gag, env     |
|      |       |                             |             |         |                  |
| I    | K119  | chr 12: 58721242-58730698    | 968/968/7521| –        | –                |
|      |       |                             |             |         |                  |
| I    | K121  | chr 3: 125609302-125618439   | 804/804/7530| –        | TAG, TGA/gag, pro, pol, env |
|      |       |                             |             |         |                  |
| I    | K132  | chr 19: 21841536-21841541    | 23/995/7869 | Alu insertion within internal/pol broken | –                |

Figure 1. Comparison of human-specific HERV-K108 and HERV-K124 elements. Both of HERV-K108 and HERV-K124 have two HERV-K internal regions (green). However, their sequence architecture is the result of different mechanisms. (A) HERV-K108. After the insertion of the HERV-K element, non-allelic homologous recombination between two different LTRs (yellow chevrons) of the HERV-K element occurred. This resulted in a locus containing two HERV-K internal regions and three LTRs. This locus retains the original TSDs (red chevrons) created upon its initial insertion. (B) HERV-K124. Compared to the HERV-K108, which has two intact internal regions and three intact LTRs, the second internal region of HERV-K124 has largely deleted and its internal and 3’ LTRs inverted and partially deleted. The mechanism(s) responsible for this element’s sequence architecture is not clearly resolved, but we depict here a potential mechanism capable of generating this element. Yellow boxes indicate standard LTRs, pink boxes indicate inverted partial LTRs, and green boxes indicate HERV-K internal regions.
genotyped 25 loci in 80 humans (20 from Asian, 20 from South American, 20 from European and 20 from African American) whose DNAs were purchased from the Coriell Institute for Medical Research. We were not able to amplify the remaining four loci because they reside either in regions of segmental duplication or in centromeric regions. As shown in Figure 3B, there are three possible states for each sister chromatid at a human-specific HERV-K insertion locus: absence of the HERV-K element, presence of the element and presence of a solitary LTR. Among the human-specific HERV-K elements, three loci, HERV-K 109, 118, and 134, exhibit all the three forms in the human populations tested. The polymorphism test found that the polymorphism level of the human-specific HERV-K elements is about 48% (12/25) which is higher than levels reported for other human-specific retrotransposons [23,24,25]. We examined the recombination rate of the genomic regions where the human-specific HERV-K elements reside because a high recombination rate could contribute to the observed increase their polymorphism level. As shown in Table 3, the recombination rates in the genomic regions flanking human-specific HERV-K elements averaged \( \sim 1.2 \) cM per Mb on both long and short arms. We compared the result with the genome-wide average recombination rates, \( \sim 1 \) cM and \( \sim 2 \) cM per Mb on the long and short arms, respectively [1]. Based on the result, we conclude that recombination rate is not a major factor responsible for the higher polymorphism levels observed in human-specific HERV-K elements.

Through the polymorphism test, we found that both type I and II full-length human-specific HERV-K elements are polymorphic in the 80 human individuals. This indicates that both types were capable of retrotransposition after the divergence of human and chimpanzee and increases likelihood that members of these groups are currently able to retrotranspose in the human genome.

**Structural Analysis of Human-specific Full-length HERV-K**

The majority of HERVs in the human genome exist in truncated form and are characterized by multiple stop codons, insertions, and deletions [26,27]. It is suspected that a smaller
subset of human-specific HERV-K elements are capable of retrotransposition and thus contain intact open reading frames (ORFs) because their proteins and particles have been detected in the human genome [28]. We therefore examined whether any of the identified human-specific full-length HERV-Ks contain intact ORFs. As shown in Table 2, five human-specific type I HERV-Ks, HERV-K102, 103, 107, 117, and 133, exhibit fused pol and env genes in the same frame. A search for stop codons in the gene components of the human-specific type I HERV-Ks revealed that HERV-K101, 102, 117, and 133 have stop codons in their pro, gag, env, and pol genes, respectively, and HERV-K133 contains stop codons in all of these genes (Figure 4). In sum, a total of three HERV-K elements have retained intact ORFs in the human genome, indicating that they have a potential to produce the viral particles [29].

As mentioned above, the type of HERV-K element is determined according to the presence/absence of a 292 bp ‘deletion’ at the pol-env boundary. It has been reported that the ancestral precursor of the type I HERV-K lacked the 292 bp sequence and that this deletion must not have been directly related to the precursor’s ability to retrotranspose in the human genome.

This is because the human genome contains at least eight type I full-length HERV-K elements which must be offspring of the precursor [30]. However, we could not rule out other possible origins for Type I insertions. For example, they could result from the recombination between competent Type II viruses and transcripts of preexisting Type I.
| No. | HERV | Genomic location | Features | Rec. rate | size (bp) | Internal 3' TLR | Reference |
|-----|------|-----------------|----------|-----------|-----------|----------------|-----------|
| 1   | K116 | chr1:73542271-73549418 | Inserted into human-specific L1PA2 | 0.7 | 4437 | 968 | 968 | [40] |
| 2   | K107 | chr1:15559645-155605436 | Polymorphic | 0.6 | 968 | 7544 | 968 | [2] |
| 3   | K120 | chr1:15367928-15367960 | Inserted into SD region | 1.1 | 2129 | 964 | 964 | [18] |
| 4   | K106 | chr1:31259093-31259262 | Polymorphic | 0.4 | 960 | 7539 | 960 | [2] |
| 5   | K127 | chr1:31268193-31268439 | Inserted into SD region | 0.8 | 804 | 7530 | 804 | [21] |
| 6   | K123 | chr1:79556554-17955804 | 13 bp L1 sequence in 3' end of ERV, polymorphic | 1.5 | 0 | 143 | 0 | This Study |
| 7   | K124 | chr1:61573938-61582439 | Second HERV-K internal to 206 bp in first 0.9 | 968 | 7117 | 786 | 206 |
| 8   | K125 | chr1:35508941-35509295 | Polymorphic | 0.6 | 968 | 7444 | 968 | [40] |
| 9   | K126 | chr1:35508941-35509295 | Polymorphic | 0.6 | 0 | 142 | 0 | This Study |
| 10  | K127 | chr1:35508941-35509295 | Second HERV-K internal to 206 bp in first 0.9 | 968 | 7117 | 786 | 206 |
| 11  | K128 | chr1:101508593-101509739 | Non-classical insertion, HERV-K9 | 0.9 | 0 | 3021 | 0 | This Study |
| 12  | K129 | chr1:35508941-35509295 | Polymorphic | 0.6 | 968 | 7444 | 968 | [40] |
| 13  | K128 | chr1:101508593-101509739 | Non-classical insertion, HERV-K9 | 0.9 | 0 | 3021 | 0 | This Study |
| 14  | K129 | chr1:35508941-35509295 | Polymorphic | 0.6 | 968 | 7444 | 968 | [40] |
| 15  | K130 | chr1:101508593-101509739 | Polymorphic | 0.6 | 968 | 7444 | 968 | [40] |
| 16  | K130 | chr1:101508593-101509739 | Polymorphic | 0.6 | 968 | 7444 | 968 | [40] |
| 17  | K131 | chr1:101508593-101509739 | Polymorphic | 0.6 | 968 | 7444 | 968 | [40] |
| 18  | K132 | chr1:101508593-101509739 | Polymorphic | 0.6 | 968 | 7444 | 968 | [40] |
| 19  | K133 | chr1:101508593-101509739 | Polymorphic | 0.6 | 968 | 7444 | 968 | [40] |
| 20  | K134 | chr1:101508593-101509739 | Polymorphic | 0.6 | 968 | 7444 | 968 | [40] |
| 21  | K135 | chr1:101508593-101509739 | Polymorphic | 0.6 | 968 | 7444 | 968 | [40] |
| 22  | K136 | chr1:101508593-101509739 | Polymorphic | 0.6 | 968 | 7444 | 968 | [40] |
| 23  | K137 | chr1:101508593-101509739 | Polymorphic | 0.6 | 968 | 7444 | 968 | [40] |
| 24  | K138 | chr1:101508593-101509739 | Polymorphic | 0.6 | 968 | 7444 | 968 | [40] |
| 25  | K139 | chr1:101508593-101509739 | Polymorphic | 0.6 | 968 | 7444 | 968 | [40] |
| 26  | K140 | chr1:101508593-101509739 | Polymorphic | 0.6 | 968 | 7444 | 968 | [40] |
| 27  | K141 | chr1:101508593-101509739 | Polymorphic | 0.6 | 968 | 7444 | 968 | [40] |
| 28  | K142 | chr1:101508593-101509739 | Polymorphic | 0.6 | 968 | 7444 | 968 | [40] |
| 29  | K143 | chr1:101508593-101509739 | Polymorphic | 0.6 | 968 | 7444 | 968 | [40] |

*Full-length human-specific HERV-K locus.*

*Sequence is reversed.*

*Insertion, SD, Segmental Duplication.*

*Polymorphism*
Among the human-specific type II HERV-Ks, HERV-K104, 108, 118, and 121 contained stop codons in multiple genes while HERV-K132 had an Alu insertion within its pol gene. These five HERV-Ks are therefore not functionally and structurally intact in the human genome. However, HERV-K113, 115, and 119 possess intact gene components, which indicates that they have the potential to encode the functional proteins required for their mobilization. HERV-K113 and 115 were previously identified to be full-length and polymorphic (HERV-K presence/absence) in human populations [30]. The result of our polymorphic test on HERV-K119 showed that this element is also polymorphic in the 80 human individuals, but its pattern of polymorphism is different from that of the other elements; the polymorphism at the HERV-K113 and 115 loci takes the form of an absence or presence of the HERV-K element between individuals, but the HERV-K119 locus exists as either a full-length HERV-K or a solitary LTR. We speculate that this architecture is the product of a homologous recombination event between the two LTRs of a full-length HERV-K element. Given this, we suspect that the HERV-K119 locus exists as either a full-length HERV-K or a solitary LTR. We speculate that this architecture is the product of a homologous recombination event between the two LTRs of a full-length HERV-K element. Given this, we suspect that the HERV-K119 element is relatively older than the other two elements (HERV-K113 and 115). These intact full-length HERV-K elements could play a role in human disease. This possibility has been suggested by several reports describing HERV-encoded transcripts and proteins in tumors [31,32] and tissue from patients with autoimmune diseases [27,33,34].

Human-specific HERV-K Insertion-associated Genetic Variations

We found four non-classical HERV-K insertion loci in our dataset (Table 1). Figure 5B depicts one possible mechanism responsible for the non-classical insertion. These elements are 5’ and 3’ truncated, meaning that they also do not have classical TSDs. Additionally, they are involved in target site deletions in the human genome. Through a comparison of the human-specific HERV-K flanking sequence and its corresponding chimpanzee pre-insertion sequence, we calculated the deletion size. However, the chimpanzee orthologous sequence of HERV-K130 insertion contained two unsequenced regions. We amplified one of the regions for sequencing (accession number: JQ811903) and the primer sequences are described in Table S2. We estimated the size of the other region using the orangutan reference genome sequence. The deletion sizes of the target sites of the non-classical HERV-Ks range from 6 bp to 10,207 bp. We further examined their genomic environments and found that three of them occurred in intergenic regions and one occurred in an intronic region. It has been reported that non-classical insertions are associated with double-strand break (DSB) repair, a mechanism proposed to aid in stability of fragile sites in the host genome [35]. Also, it has been suggested that DSBs can be repaired through homologous recombination (HR) or non-homologous end joining (NHEJ) to ensure the maintenance of genome integrity in eukaryotic organisms [36]. As for the four non-classical HERV-K insertions, we examined the microhomology between each HERV-K element and its pre-insertion sequence from the chimpanzee genome. Microhomology, if present, could mediate the insertion of the HERV-K between DSB ends via a NHEJ-associated process. We identified 5’ and 3’ microhomologies for three out of the four loci but were not able to detect microhomology for the HERV-K130 locus, as shown in Table S3. We conclude, therefore, in the cases where microhomology exists at both ends of the HERV-K insertion, the likelihood of DSB repair through NHEJ is increased.

In this study, we identified 29 human-specific HERV-K insertions including previously reported three loci (HERV-K103, 113, and 134) that have integrated into the human genome since the divergence of humans and chimpanzees. During this time,
HERV-K activity contributed to genomic variation between the two species. Through a polymorphism test, we found that the polymorphic rate of these elements is 48%. This indicates that the activity of the HERV-K family has resulted in genomic variations between and within human populations. It is currently unknown whether there are any retrotranspoitionally competent copies of HERV-K element in the human genome. However, based on the results of this study, we assert that HERV-K element activity is a cause of genomic differences between the human and chimpanzee genomes as well as genomic diversity within the human population.

Materials and Methods

Computational Data Mining and Manual Inspection of Human-specific HERV-K Loci

To computationally screen the human genome (hg19; February 2009 freeze) for potential human-specific HERV-K loci, we first extracted all HERV-K loci from the human genome by using UCSC Table Browser utility (http://genome.ucsc.edu/cgi-bin/hgTables?org=Human&db=hg19&hgsid=226995881&hgtadomaMainPage=1). For each HERV-K locus, we next extracted 2 kb flanking sequences, up and downstream. This human sequence was then used as a query against other primate genome sequences (panTro3; October, 2010 freeze, ponAbe2; July 2007 freeze, rheMac2; January 2006 freeze), using UCSC’s BLAT utility (http://genome.ucsc.edu/cgi-bin/hgBlat). For each hit in the BLAT search, we retrieved the human, chimpanzee, orangutan and rhesus macaque sequences. Repeat elements existing in these nonhuman sequences were annotated using the RepeatMasker (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker) tool. Based on these repeat element annotations, we confirmed whether each HERV-K locus was specific to the human genome or not.

PCR Amplification and DNA Sequence Analysis

To experimentally verify the human-specific HERV-K insertion candidates, we conducted PCR analysis with four different DNA templates: Homo sapiens (human; NA10851, Coriell Cell Repository, Camden, NJ), Pan troglodytes (common chimpanzee), Gorilla...
**References**

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409: 860–921.

2. Barbulescu M, Turner G, Seaman MI, Deinard AS, Kidd KK, et al. (1999) Many human endogenous retrovirus K (HERV-K) proviruses are unique to humans. Curr Biol 9: 861–868.

3. Griffiths DJ (2001) Endogenous retroviruses in the human genome sequence. Genome Biol 2: REVIEWS0017.

4. Khodosevich K, Lebedev Y, Sverdlov E (2002) Endogenous retroviruses and human evolution. Comp Funt Genomics 3: 494–498.

5. Balada E, Ordi-Ros J, Viladerri-Tarres M (2009) Molecular mechanisms mediated by human endogenous retroviruses (HERV's) in autoimmune. Rev Med Virol 19: 273–286.

6. Buzdin A, Ustyugova S, Khodosevich K, Mamedov I, Lebedev Y, et al. (2003) Human-specific subfamilies of HERV-K (HML-2) long terminal repeats: three master genes were active simultaneously during branching of hominoid lineages. Genomics 81: 149–156.

7. Dunn CA, van de Lagemaat LN, Baillie GJ, Mager DL (2005) Endogenous primate beta3GAL-T5. Gene 364: 2–12.

8. Sverdlov ED (2000) Retroviruses and primate evolution. Bioessays 22: 161–171.

**Supporting Information**

**Figure S1** The 29 human-specific HERV-K insertion loci in the human genome. Blue and green circles indicate the chromosomal locations of full-length and truncated human-specific HERV-K elements, respectively. Among them, 12 loci were polymorphic and 4 loci were non-classical insertions. The karyotype images were created using the idiographica webtool (http://www.ncrna.org/idiographica/).

**Table S1** GC content and gene density in flanking regions of human-specific HERV-K loci.

**Table S2** PCR primers for the sequences deleted by HERV-K130 insertion.

**Table S3** Additional information on human-specific HERV-K insertions.

**Table S4** PCR primers for human-specific HERV-K loci.

**Acknowledgments**

We would like to thank Dr. Thomas J. Meyer for thoughtful comments on the manuscript.

**Author Contributions**

Performed the computational analysis: WS JL KA. Conceived and designed the experiments: WS JL. Performed the experiments: WS JL. Analyzed the data: WS JL SYS HSK KH. Contributed reagents/materials/analysis tools: KH. Wrote the paper: WS JL KA KH.
9. Katzenrakas A, Rambaut A, Pybus OG (2005) The evolutionary dynamics of endogenous retroviruses. Trends Microbiol 13: 463–468.
10. Dewannieux M, Harper F, Richard A, Letzelter C, Ribet D, et al. (2006) Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retrovirus. Proc Natl Acad Sci U S A 101: 4894–4899.
11. Belshaw R, Pereira V, Katzenrakas A, Talbot G, Paces J, et al. (2004) Long-term reinfection of the human genome by endogenous retroviruses. Proc Natl Acad Sci U S A 101: 17137–17144.

25. Otieno AC, Carter AB, Hedges DJ, Walker JA, Ray DA, et al. (2005) The distribution and evolutionary fates of recently integrated human endogenous retroviruses. J Virol 79: 12507–12514.
26. Macfarlane C, Simmonds P (2004) Allelic variation of HERV-K(HML-2) endogenous retroviral elements in human populations. J Mol Evol 59: 642–656.
27. Belshaw R, Dawson AL, Woolven-Allen J, Redding J, Burt A, et al. (2005) Genome-wide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K(HML-2): implications for present-day activity. J Virol 79: 12507–12514.
28. Shibramanian RP, Wildschutte JH, Ruo C, Coffin JM (2011) Identification, characterization and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. Retrovirology 8: 90.
29. Mayer J, Stuhr T, Reus K, Maldener E, Kitova M, et al. (2005) Haplotype analysis of the human endogenous retrovirus locus HERV-K(HML-2.HOM) and its evolutionary implications. J Mol Evol 61: 706–715.
30. Hughes JF, Coffin JM (2002) A novel endogenous retrovirus-related element in the human genome resembles a DNA transposon: evidence for an evolutionary link? Genomics 80: 453–455.
31. Hughes JF, Coffin JM (2005) Human endogenous retroviral elements as indicators of ecotypic recombination events in the primate genome. Genetics 161: 1183–1194.
32. Brady T, Lee YN, Ronen K, Malani N, Berry CC, et al. (2009) Integration target site selection by a resurrected human endogenous retrovirus. Genes Dev 23: 633–642.
33. Lee J, Cordaux R, Han K, Wang J, Hedges DJ, et al. (2007) Different evolutionary fates of recently integrated human and chimpanzee LINE-1 retrotransposons. Gene 390: 18–27.
34. Carter AB, Salem AH, Hedges DJ, Keegan CN, Kimball B, et al. (2004) Genome-wide analysis of the human Alu Yb-lineage. Hum Genomics 1: 167–178.
35. Otomo NC, Carter AB, Hedges DJ, Walker JA, Ray DA, et al. (2004) Analysis of the Human Alu Ya-lineage. J Mol Biol 342: 109–118.
36. Kim TH, Jeon YJ, Yi JM, Kim DS, Huh JW, et al. (2004) Human endogenous retrovirus glycoprotein-mediated induction of redox reagents causes oligodendrocyte death and demyelination. J Neurosci 24: 1088–1095.
37. Ruprecht K, Ferreira H, Flockerzi V, Wahl S, Sauter M, et al. (2008) Human endogenous retrovirus family HERV-K(HML-2) RNA transcripts are selectively packaged into retroviral particles produced by the human germ cell tumor line Tera-1 and originate mainly from a provirus on chromosome 22q11.21. J Virol 82: 10008–10016.
38. Boller K, Schonfeld K, Lisch S, Fischer N, Hoffmann A, et al. (2008) Human endogenous retrovirus HERV-K113 is capable of producing intact viral particles. J Gen Virol 89: 367–372.
39. Turner G, Barbulescu M, Su M, Jensen-Seaman MJ, Kidd KK, et al. (2001) Insertional polymorphisms of full-length endogenous retroviruses in humans. Curr Biol 11: 1531–1535.
40. Sprafke R, Roche G, Dussart P, Prin L (2002) Expression of a human endogenous retrovirus, HERV-K, in the blood cells of leukemia patients. Leukemia 16: 254–259.
41. Wang J, Johanning F, Frost AR, Johanning GL, Khaizali MB, LoBuglio AF, et al. (2001) Expression of human endogenous retroviruses in human breast cancer. Clin Cancer Res 7: 1553–1560.
42. Clerici M, Fusi ML, Caputo D, Guerini FR, Trabattoni D, et al. (1999) Immunologic responses to antigens of human endogenous retroviruses in patients with acute or stable multiple sclerosis. J Neuroimmunol 99: 173–182.
43. Hishikawa T, Ogasawara H, Kaneko H, Shirasawa T, Matsuura Y, et al. (1997) Detection of antibodies to a recombinant gag protein derived from human endogenous retrovirus clone 4–1 in autoimmune diseases. Viral Immunol 10: 137–147.
44. Srikanta D, Sen SK, Huang CT, Conlin EM, Rhodes RM, et al. (2009) An alternative pathway for Alu retrotransposition suggests a role in DNA double-strand break repair. Genomics 93: 205–212.
45. Chu G (1997) Double strand break repair. J Biol Chem 272: 24097–24100.
46. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series: 95–98.
47. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731–2739.
48. Sperber GO, Airota T, Jern P, Blomberg J (2007) Automated recognition of retroviral sequences in genomic data-RetroTector. Nucleic Acids Res 35: 4964–4976.
49. Hughes JF, Coffin JM (2001) Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution. Nat Genet 29: 487–490.
50. Sugimoto J, Matsamura N, Kinjo Y, Takasu N, Oda T, et al. (2001) Transcriptionally active HERV-K genes: identification, isolation, and chromosomal mapping. Genomics 72: 137–144.
51. Oono M, Kawakami M, Ushikubo H (1987) Stimulation of expression of the human endogenous retrovirus genome by female steroid hormones in human breast cancer cell line T47D. J Virol 61: 2059–2062.
52. Costan J (2001) Evolutionary dynamics of the human endogenous retrovirus family HERV-K inferred from full-length proviral genomes. J Mol Evol 53: 237–243.
53. Medstrand P, Mager DL (1998) Human-specific integrations of the HERV-K human endogenous retrovirus family. J Virol 72: 9782–9787.
54. Tonges RR, Czauderna F, Kurth R (1999) Genome-wide screening, cloning, chromosomal assignment, and expression of full-length human endogenous retrovirus type K. J Virol 73: 9187–9195.
55. Kurdyukov SG, Lebedev VB, Artemovova, II, Gogorodneva TN, Batrak AV, et al. (2001) Full-sized HERV-K (HML-2) human endogenous retroviral LTR sequences on human chromosome 21: map locations and evolutionary history. Gene 273: 51–61.