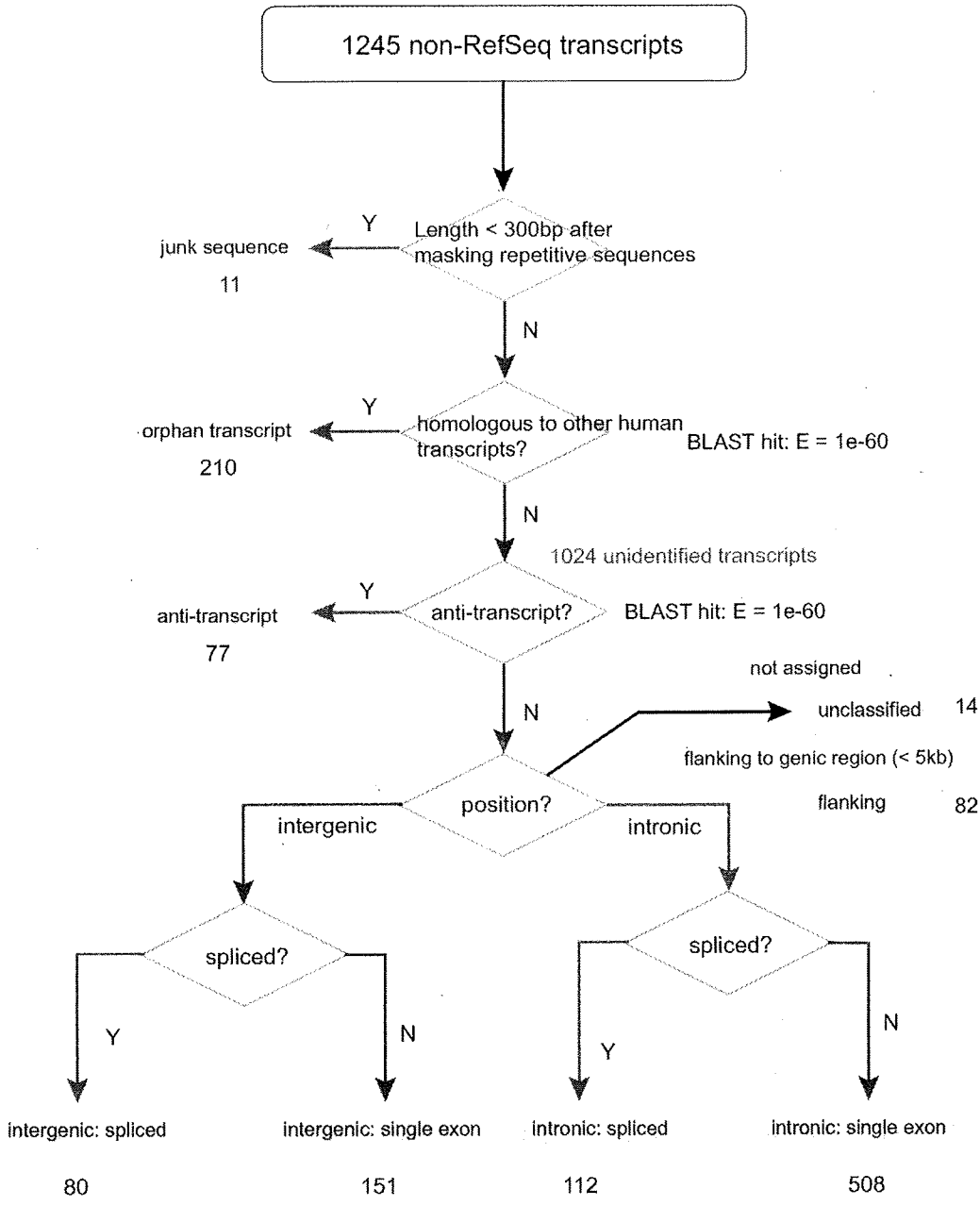

Figure 1
Classification of the 1245 Non-RefSeq transcripts. Transcripts shorter than 300 bp after masking the repetitive sequences were categorized as junk sequences. The remaining sequences were BLAST-searched against all public human cDNA sequences for the forward strand. Homologous sequences to the unannotated human cDNAs were classified as orphan transcripts for the forward strand and anti-transcript for the reverse strand. The remaining 947 clones were mapped on the human genome sequence and arranged according to the annotation from the UCSC genome browser (hg18). The transcripts that overlapped with the genic regions including UTR were classified as intronic transcripts, and the transcripts that were mapped more than 5 kb away from the genic region were classified as intergenic transcripts.

known transcripts, using alternative promoters and/or poly(A) signals in the human genome. These sequences were filtered from the intergenic transcripts and classified as 'flanking' to genic regions. The largest group was the intronic single-exon transcripts. Although they might be acquired from premature mRNA molecules in the cell nucleus, recent studies have revealed the potential abundance of short intronic transcripts in the human genome [32]. Among these classes, anti-transcripts and intergenic spliced transcripts are the most biologically relevant classes, which are unlikely to be derived from contamination by premature mRNAs.

We designed oligonucleotide microarrays (Affymetrix GeneChip) containing probes complementary to the known genes and unidentified transcripts. Hybridizations were performed using the RNA sampled from a 3-year-old macaque cerebrum, cerebellum, liver, and testis with duplications. The significance of expression was determined using Affymetrix MAS5.0 software [33] (see methods). The proportion of the expressed transcripts is presented in Figure 2. In the unidentified transcripts, 544 transcripts were expressed in at least one of the four tissues ($P < 0.05$; Table 2). Because all the unidentified transcripts were isolated from the macaque brain or testis, fewer transcripts were expressed in the liver (14%) than in the cerebrum (31%), cerebellum (41%), and testis (24%). The expressed proportion of the unidentified transcripts was significantly smaller than that of 8428 RefSeq homologs (51% and 81%, respectively; $P < 10\text{-}15$; Fisher's exact test). The orphan transcripts were expressed in an intermediate proportion (72%). The percentages of the expressed unidentified transcripts ranged from 33% to 57% (Fig. 1). A large difference was observed between the intergenic and intronic transcripts; more intronic transcripts displayed significant expression on the microarrays than intergenic transcripts ($P = 0.0005$; Fisher's exact test).

Previous studies have shown that many unannotated transcripts are not conserved at a DNA sequence level in many organisms [34]. In practice, sequence conservation is determined by investigating whether the region is alignable. Here, we directly measure the difference in the DNA sequences between humans and macaques. For protein-coding genes, previous studies have shown large disparities in sequence divergence between brain- and testis-expressed genes, both in the CDS and UTR, owing to the stronger functional constraint on the brain-expressed genes [20,21]. We further inquired whether the trend was observed in the unidentified transcripts. We classified the transcripts into brain-expressed transcripts (expressed in the cerebrum and not in the testis) and testis-expressed transcripts (expressed in the testis and not in the cerebrum). As shown in Figure 3, while the non-synonymous substitution rates of the RefSeq homologs were higher in the testis than in the brain, the DNA sequence divergence

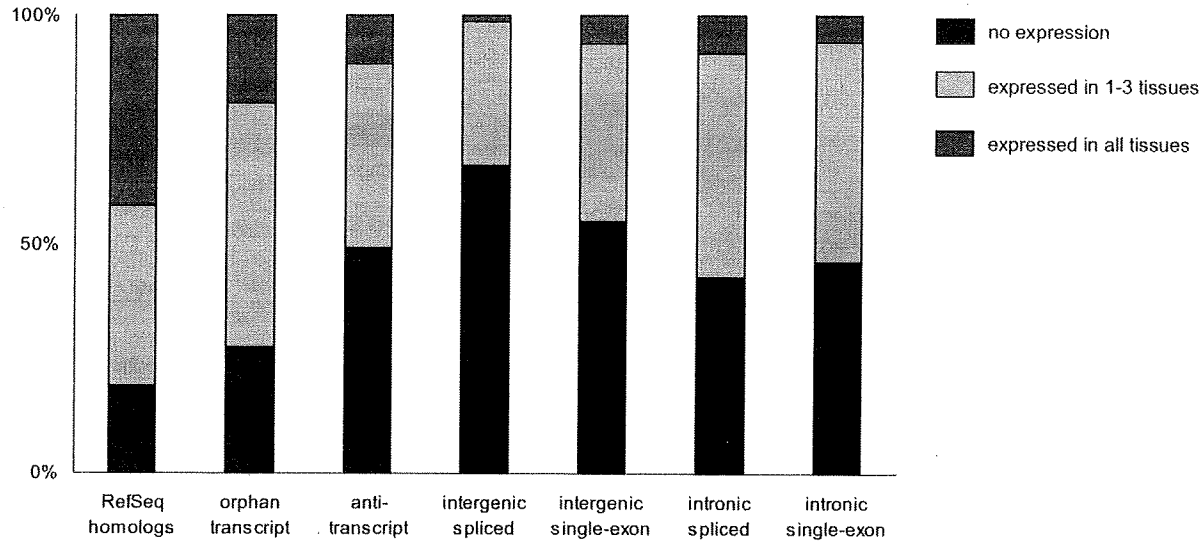

Figure 2
The proportion of the expressed transcripts in the RefSeq homologs (control) and unidentified transcripts. Cerebrum, cerebellum, liver, and testis of a male macaque were used for the microarray experiments with duplicated hybridizations. The transcripts were classified into no expression (blue), expressed in 1–3 tissues (grey), or expressed in all tissues (red).

Table 2: Number of expressed transcripts in the unknown macaque transcripts

	Unidentified transcripts[a]	Intergenic transcripts[b]
Cerebrum	321	54
Cerebellum	417	58
Liver	139	13
Testis	241	52
All tissues	74	10
Any tissue	544	137
Total	1024	231

[a]Transcripts that have no homology to the public human cDNA sequences.
[b]Transcripts that were mapped more than 5 kb away from the annotated genic regions on the human genome (see Fig. 1).

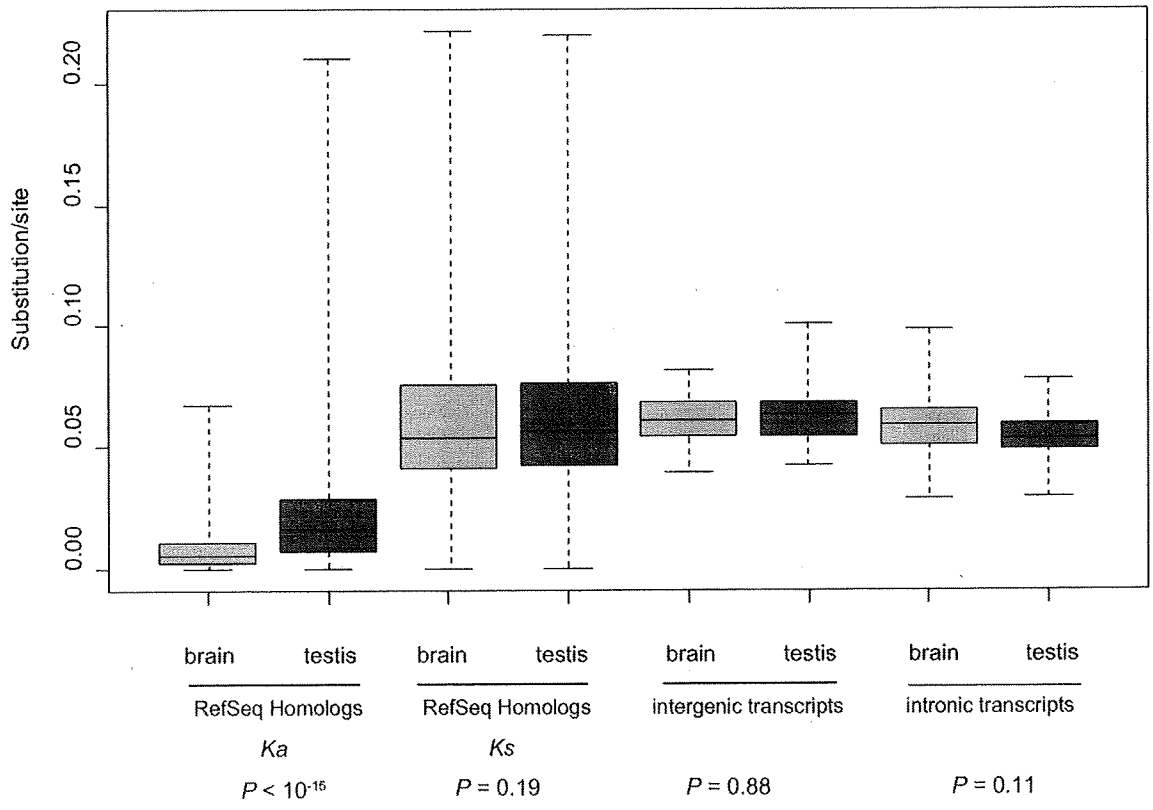

Figure 3
Sequence conservation of the brain-expressed and testis-expressed transcripts between humans and macaques. For the RefSeq homologs (control), the non-synonymous (K_a) and synonymous (K_s) substitution rates were estimated using the Li-Pamilo-Bianchi method [48]. The substitution rates in the intergenic and intronic transcripts were estimated using Kimura's two parameter methods [55]. The heights of the boxes represent the lower and upper quartile points, and the whiskers show the minimum and maximum points.

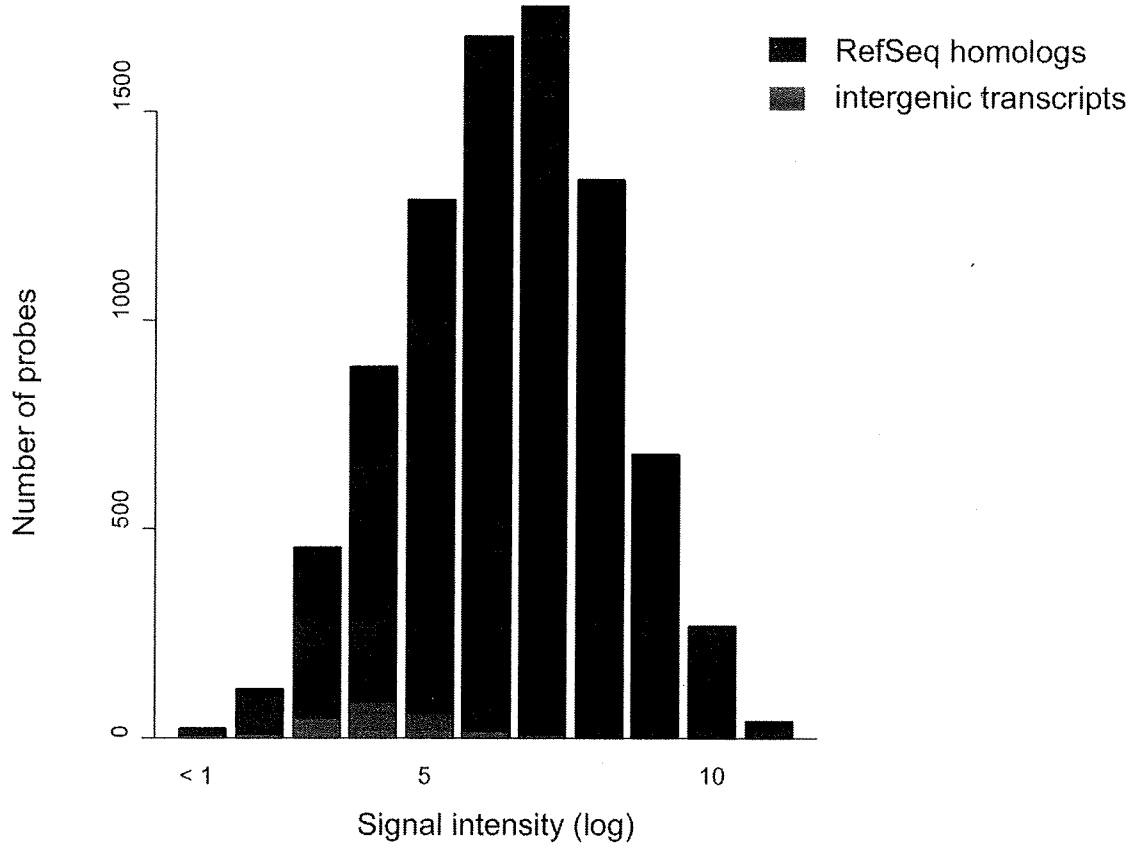

Figure 4
Distribution of transcript expression levels of the RefSeq homologs (blue) and the intergenic transcripts (red). Only the transcripts that were determined as significantly expressed on the microarray are presented in the figure. Log-transformed signal intensity in the tissue with the highest expression was shown. The intergenic transcripts showed significantly lower expression levels than the RefSeq homologs.

of the unidentified transcripts was not associated with the expression pattern. Furthermore, there was no evidence that the unidentified transcripts were more conserved than the synonymous sites of the RefSeq homologs.

We further evaluated the expression level of the 231 intergenic transcripts. We collected the strongest signal intensity of the significantly expressed intergenic transcripts. As shown in Figure 4, even if they were significantly expressed, signal intensities of the intergenic transcripts were significantly weaker than those of the control genes ($P < 10\text{-}13$; Wilcoxon test). Weak expression levels of intergenic sequences have been previously reported [35,36] and these may cause weak detection levels of the intergenic transcripts. To test the reproducibility of the microarray experiments using another method, we selected eight intergenic spliced transcripts and tried to amplify human and macaque transcripts using RT-PCR. We designed the PCR primers that would match both human and macaque sequences and would amplify introns of genomic sequences when the genomic DNA is contaminated. A gel picture of the RT-PCR products is shown in Figure 5. Two transcripts showed positive results, while six showed negative results on the microarray. We confirmed the expression of the two transcripts in the macaque brain using both the microarray and RT-PCR. Furthermore, even though we failed to detect the expression of the six transcripts on the arrays, we recov-

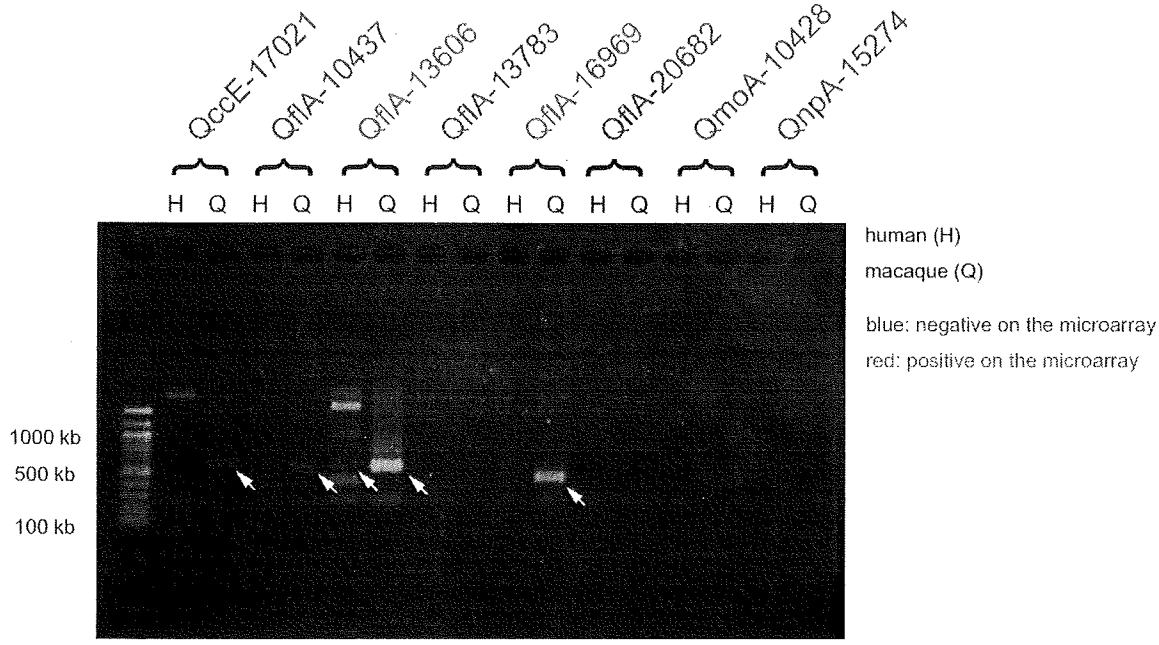

Figure 5
RT-PCR gel images for the expression of the intergenic transcripts in the human (H) and the macaque (Q) brain. Transcript names indicate whether the expression was detected by the microarray experiments (red) or not (blue). Expected PCR products are marked by the white arrows.

ered the expression of the two transcripts by RT-PCR. In these two transcripts, the expression levels detected by RT-PCR resulted in considerably weaker bands on the gel (Fig. 5), indicating that the microarray failed to capture their expression at a very low level. In total, we detected the expression of four transcripts in the macaque brain. Of these four transcripts, two were not detected and one was transcribed in an unspliced form in humans. The other showed multiple extra bands in both humans and macaques. Overall, the expression of the macaque intergenic spliced transcripts was not well conserved between the human and the macaque brain.

Hidden transcript structures in the human genome

Of the 9407 macaque cDNA sequences, 2261 covered the entire CDS of the human RefSeq genes in a single BLAST hit chain. In the 2261 cDNAs, we sought a stretch of UTR sequences (> 50 bp) that did not match any homologous human cDNA sequence. Simple genomic insertion or deletion in the genome was not counted. After filtering the ambiguous entries, in the UTR of macaque cDNAs, we found 262 exons that were not found in the human cDNA data. Out of the 262 unidentified exons, 85 (32%) did not match any human EST sequence. We classified the unidentified UTRs as follows: (A) extended exons and (B) novel exons (Fig. 6). Those unidentified exons were further classified into internal and external exons (Fig. 6). As shown in Figure 1, the distribution of the different types of unidentified exons was not uniform; most of them were external exons.

Because the human transcriptome data is more complex than previously thought, as revealed by genome tiling DNA microarrays [34-36], these unrepresented exons may be expressed at a very low level in human tissues. Moreover, these exons have not been found in the conventional cDNA exploration methods. However, previous studies have suggested a frequent evolutionary turnover of exon sequences [37]. The evolutionary alteration of external exons in the 5'-UTR may be caused by the alternative usage of promoter sequences [38]. The evolutionarily altered exons in the 3'-UTR may be caused by the alternative usage of poly(A) adenylation signals [39]. All the unidentified exons are provided in Additional file 2.

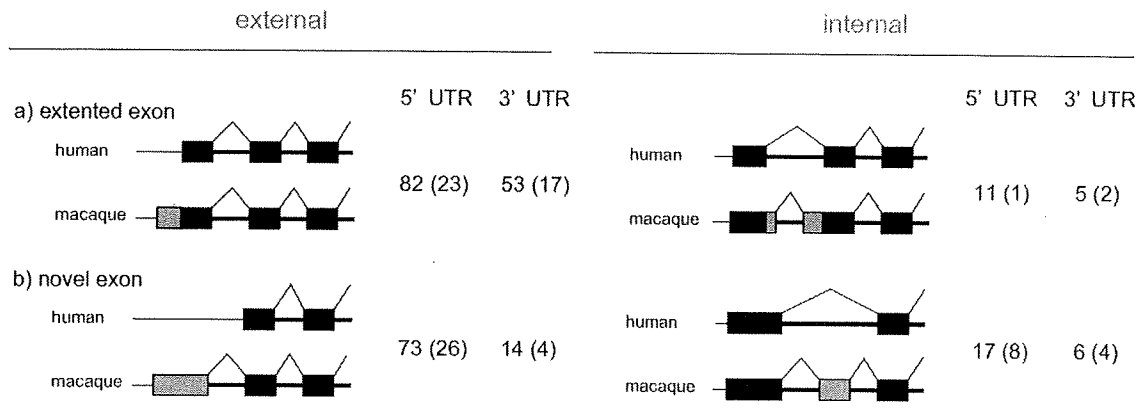

Figure 6
Pattern of the unidentified exons. The closed boxes represent exons in the genomes. Unidentified exons in macaques are presented as blue boxes. Intergenic regions and introns are depicted by thick and thin horizontal lines, respectively. (A) extended exons. (B) novel exons. These exons were further classified into internal (right panel) and external (left panel) exons. The number of genes in each category is shown on the left of each schema. The number of unidentified exons that have not been found even in the EST sequences is shown in parentheses.

Comparison of the human, cynomolgus, and rhesus genes
We compiled 2655 human-rhesus-cynomolgus cDNA alignments (dataset I) using the rhesus macaque genome and the predicted transcript sequences. The phylogenetic relationship among the three species is shown in Figure 7. Because the rhesus and cynomolgus genomes are very similar, we wanted to minimize non-orthologous alignments, which inflate the average and variance of the nucleotide divergence between them. Therefore, the macaque genes showing > 80% homology to more than one locus in the rhesus genome were filtered (dataset II). Although the number of genes analyzed was reduced to 1499 in the second dataset, the subset of the genes would be useful in estimating the divergence among the three species. The results were obtained using dataset II in the following manner. The results using the unfiltered dataset (dataset I), which resulted in the inflation of variance, are provided in Additional file 3. Genes that have evolved under positive selection were searched with the model-based likelihood ratio test [40]. In total, 39, 15, and 22 genes showed evidence of positive selection in the human, cynomolgus, and rhesus lineages, respectively ($P < 0.05$). Thirty-eight genes also showed a positive selection signature between the two macaques and 74 were detected in all the three lineages (Table 3). Note that, in Figure 7, the phylogenetic tree is unrooted. The list of positively selected genes is provided in Additional file 4. Excluding the overlapped genes, we identified 101 out of 1499 genes (6.7%) that underwent positive selection in any lineage at 5% significance level. The number of positively selected genes in each of the two macaque lineages was comparable to that estimated in the human-chimpanzee lineages using the same method [41]. Although these candidates of positively selected genes contain many biologically interesting functions, such as transcriptional regulation (*RELA*, *ZNF263*, and *L3MBTL4*), visual perception (*RGS9*, *GPRC5B*, and *RPGRIP1*), and mitochondrial localization (*PET112L*,*VARS*,*ACAA2*,*YARS2*, *FOXRED1*, and *COQ9*) [19], none of the GO categories were statistically overrepresented probably because of the small sample size.

Genetic divergence between cynomolgus and rhesus macaques
Using the above dataset, we estimated the nucleotide substitution rates of each lineage from the common ancestor of cynomolgus and rhesus macaques (presented in Table 4). Numbers and rates of the non-synonymous and synonymous substitutions for each lineage were estimated using the maximum likelihood method. We assumed that synonymous substitutions are nearly neutral and used

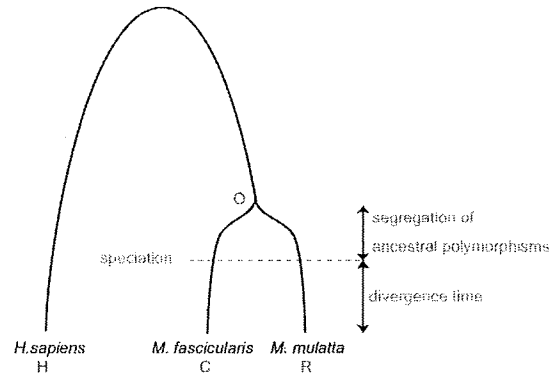

Figure 7
Genealogical relationship (phylogeny of genes) among the humans (H), cynomolgus macaque (C), and rhesus macaque (R). The common ancestor of the two macaques is indicated by the letter O. The time of speciation between the two macaques is shown by the dashed line. Note that the tree is unrooted.

them to estimate the divergence time. If we set the divergence of humans and Old World monkeys to 25–35 million years ago (Mya) [42], genes of the two macaques would be considered to diverge 1.9–2.6 Mya on average. However, since the two species diverged very recently, we have to consider the ancestral polymorphisms segregating at the time of speciation [11,12]. Figure 7 illustrates the impact of the ancestral polymorphisms on the estimation of the divergence time. Suppose that the common ancestor of two macaques had the same population size as the population size of extant chimpanzees, *i.e.*, 1–2 Mya to the most recent common ancestor [43]. In this situation, only one Mya divergence is assigned to the actual divergence time of the two species. Therefore, without considering ancestral polymorphisms, we tend to overestimate the true species divergence. We applied the maximum likelihood method to estimate the divergence time and ancestral population size of the rhesus and cynomolgus monkeys. As a result, we obtained $2tu$ = 0.00213 ± 0.00022 and $4Neu$ = 0.00327 ± 0.00025 where t, Ne, and u represent the divergence time after speciation, ancestral population size at speciation, and mutation rate with standard errors, respectively. We also noticed that a non-negligible number of nucleotide substitutions were erroneously assigned to the cynomolgus macaque lineage owing to PCR errors in the cDNA libraries. Therefore, the actual substitution rate was estimated by correcting the PCR error using a computational method (see methods). After the correction, we obtained $2tu$ = 0.00181 ± 0.00021 and $4Neu$ = 0.00311 ± 0.00024 (Table 4). If we consider u = 10-9 (nucleotide substitution rate per year of Old World monkeys [44]), the divergence time between the two macaques would be estimated as 0.91 ± 0.11 Mya with a standard error. We also estimated the ancestral population size to be 43,200 ± 3300 with a mutation rate per generation of 1.8 × 10-8 in humans [45]. The result suggests that more than a half of the genetic divergence between the two macaques is derived from the ancestral polymorphisms.

Discussion

In summary, the sequencing project of cynomolgus monkey cDNAs yielded 85,721 ESTs and 9407 full-length sequences. Since our project mainly studied the brain and testis, the dataset is deficient in other tissue-specific genes, e.g., the genes related to the immune system that many medical researchers would want to explore [46]. The construction of cDNA libraries from other tissues and EST sequencing is still ongoing to complement the transcriptome of cynomolgus monkeys. The latest sequencing status can be confirmed from the website. Because of the close relationship between the cynomolgus and rhesus macaques, cDNA resources of cynomolgus macaques not only are useful for research using cynomolgus macaques but also complement the relative paucity of the transcriptome data from rhesus macaques. Using macaque tissues to scan the primate transcriptome is advantageous because RNA molecules are unstable and are instantly degraded in the tissues during sampling. This causes serious problems for RNA sampling from human tissues, especially in the brain, where fresh samples are rarely obtainable. Therefore, we hope to uncover rare transcripts

Table 3: Number of genes under positive selection out of 1499 non-duplicated genes determined using the branch-site test of positive selection

Lineage[a]	$P \leq 0.05$	$P \leq 0.01$
H-O	39	14
C-O	15	10
R-O	22	21
Between the macaques (C-O + R-O)	37	32
All lineages (H-O + C-O + R-O)	74	33

[a]H: human; C: cynomolgus macaque; R: rhesus macaque; O: cynomolgus-rhesus ancestor (see Fig. 7).

Table 4: Divergence among the human, cynomolgus, and rhesus genes

Model without ancestral polymorphisms (Raw data)		
Lineage[a]	K_a (± S.E.)	K_s (± S.E.)
H-O	1.06×10^{-2} (3.20×10^{-4})	6.82×10^{-2} (1.07×10^{-3})
C-O	1.02×10^{-3} (4.79×10^{-5})	3.04×10^{-3} (1.20×10^{-4})
R-O	4.98×10^{-4} (3.36×10^{-5})	2.50×10^{-3} (1.15×10^{-4})
Model with ancestral polymorphisms		
	$2tu$[b] (± S.E.)	$4N_eu$[b] (± S.E.)
Raw data	2.13×10^{-3} (2.24×10^{-4})	3.27×10^{-3} (2.52×10^{-4})
PCR error corrected	1.81×10^{-3} (2.12×10^{-4})	3.11×10^{-3} (2.40×10^{-4})

[a]H: human, C: cynomolgus macaque, R: rhesus macaque, O: cynomolgus-rhesus ancestor (see Fig. 7).
[b]t: time after speciation; u: mutation rate; N_e: effective population size of the cynomolgus-rhesus ancestor.

that would be hidden in the human transcriptome data. In this study, we identified 1024 macaque cDNAs that were not represented in the public human cDNA sequences. Although 51% of the cDNA did not show a positive signal on the microarrays, the following RT-PCR experiments recovered the expression in half (3/6) of the transcripts. The results indicate that these unidentified transcripts were expressed at a low level in the tissues even though the microarray could not detect the expression.

The *M. fascicularis* oligonucleotide microarrays contain probes that matched 8316 known genes and 1024 unidentified transcripts. We determined the number of probe sets for the known genes that overlapped among the commercially available microarray (Affymetrix GeneChip) and previously published microarray of rhesus macaque by Wallace et al., which contains the largest number of probe sets among the published microarrays [47]. Of our 8316 probes for the known genes, 1728 (21%) were not represented in the commercial microarray and 1091 (13%) were not found in the published microarray. Combining the three microarrays, 417 probes for the known genes were represented only in the *M. fascicularis* microarrays [see Additional file 5]. In our preliminary study of the polymorphisms within cynomolgus macaques, we found that the level of polymorphisms in cynomolgus macaques was greater than that in rhesus macaques and slightly smaller than the level of divergence between rhesus and cynomolgus macaques (Osada et al., unpublished data). Therefore, even if we should be careful about sequence mismatches within and between species, the information from both macaque transcripts and the rhesus genome can be combined to build more versatile and comprehensive DNA microarrays that can be used for biomedical surveys using laboratory macaques.

Suppose that we identify positively selected genes in the human lineage after the spilt from chimpanzees. Such genes are useful for understanding the human-specific physiology only when those genes have not been under positive selection in other primate lineages. We identified 37 genes under positive selection between the two macaques at 5% significance level. None of these genes were shared with 387 genes under positive selection in the human or chimpanzee lineages previously determined from the whole genome scan [41], providing support that the method has correctly identified positively selected genes in the specific lineages.

For estimating of the divergence time between cynomolgus and rhesus macaques, we assumed that there is no gene flow between the ancestral species throughout their speciation and divergence time (*i.e.*, allopatric model). However, considering the ancestral polymorphisms and the PCR error rate, we estimated the divergence time to be around 0.9 Mya, which is less than the estimation of the age of MRCA of rhesus macaques [10]. Indeed, more than half of the genetic divergence between the two macaques was derived from ancestral polymorphisms. If continuous gene flow is present during speciation, the variance component would be inflated and we would tend to overestimate the amount of ancestral polymorphisms [14].

In this analysis, we used the rhesus macaque genome sequence to represent rhesus macaques. We should note that the rhesus macaque used for genome sequencing was an Indian rhesus macaque; these macaques have genetically differentiated from Chinese rhesus macaques [9]. In addition, our samples of cynomolgus macaques were obtained from different geographic subpopulations. Previous studies using mitochondrial DNA sequences [10] and our preliminary analysis using nuclear DNA sequences (Osada et al., unpublished data) showed that there is a substantial genetic divergence between cynomolgus monkeys of Sundaland (Indonesia and Indochina) and Philippine populations. Therefore, our phylogenetic inference using two sampled sequences has a technical limitation and may be accurate only if there are no complex population structures among the ancestral cynomolgus and rhesus macaque populations. Elucidat-

ing the polymorphisms and divergence among macaque species would provide further insight into the evolutionary history of macaques and benefit biomedical research using macaque monkeys.

In Table 4, without correcting the PCR error rate, both the non-synonymous and synonymous divergences are greater in the cynomolgus lineage. This may be due to shorter generation time and smaller population size of cynomolgus monkeys. However, a more reasonable explanation is that the cDNA sequences of cynomolgus monkeys might incorporate the errors resulting from PCR amplification during the construction of the oligo-capped cDNA libraries. The synonymous substitution rate in the cynomolgus lineage is about 0.0005 points higher than that in the rhesus lineage, and the non-synonymous substitution rate differs in about 0.0004 points. Assuming that the selective constraint and generation time of the two macaque lineages are the same, excess divergence of 0.04%-0.05% in the cynomolgus lineage may be an artifact introduced by PCR amplification, which is fairly close to the estimation from the experiment by Suzuki and Sugano [48]. If we reflect the substitution rate in the rhesus lineage to that in the cynomolgus lineage for correcting the errors, the total divergence of the two macaques will be reduced to about 90% (Table 4).

Conclusion

Transcript data from Old World monkeys provide us with means to determine not only the evolutionary difference between human and non-human primates but also the hidden transcripts in the human genome. Actual cDNA clones of macaques are also indispensable resources for genetic engineering studies. It is considered that the species divergence between rhesus and cynomolgus macaques would be much later than the previous estimates, and the speciation process between them might have been complex. To use laboratory macaques more efficiently, we need to be more aware of the genetic difference within and among macaque monkeys. Increasing the genomic resources and information of macaque monkeys will greatly contribute to the development of evolutionary biology and biomedical sciences.

Methods

Cynomolgus monkey samples

Samples from two cynomolgus monkeys, a 16-year-old female (Philippine origin) and a 15-year-old male (Cambodian-Thai hybrid), were used for the cDNA libraries, except for the liver cDNA library (Qlv). The liver samples were collected from three adult cynomolgus monkeys of unknown origin. The monkeys were cared for and handled according to the guidelines established by the Institutional Animal Care and Use Committee of the National Institute of Infectious Diseases (NIID) of Japan and the standard operating procedures for monkeys at the Tsukuba Primate Center, NIID (present National Institute of Biomedical Innovation), Tsukuba, Ibaraki, Japan. Tissues were excised in accordance with all the guidelines in the Laboratory Biosafety Manual, World Health Organization, at the P3 facility for monkeys of the Tsukuba Primate Center. Immediately after collection, the tissues were frozen in liquid nitrogen and used for RNA extraction. Oligo-capped cDNA libraries were constructed according to the method described previously [48]. The prefix in each clone name represents the location of the source of the tissue: Qnp (brain, parietal lobe), Qfl (brain, frontal lobe), Qtr (brain, temporal lobe), Qor (brain, occipital lobe), Qbs (brain stem), Qmo (medulla oblongata), Qcc (cerebellar cortex), Qlv (liver), and Qts (testis).

Sequencing of cDNA clones

The cDNA clones were sequenced with ABI 3700 and 3730 automated sequencers. The EST sequences were trimmed to avoid the vector sequence of pME18-FL3 [DDBJ/EMBL/GenBank: AB009864]. Entire sequences of the clones were determined by the primer walking method. The repeat sequences at the 5'- and 3'-ends were masked using the Repbase Update database [49] before BLAST search. The BLAST search was performed with an e-60 cut-off value against non-redundant human RefSeq data. The non-redundant data was based on the annotation in the Ensembl Gene database. The longest transcript in the locus was selected as the representative cDNA [24,50]. The macaque cDNA sequences were deposited in the public DNA databases [DDBJ/EMBL/GenBank: CJ430287–CJ493524; BB873801–BB894695; AB303966–AB303967].

Classification of unidentified transcripts

Classification of the Non-RefSeq transcripts was performed as shown in Figure 1. Transcripts shorter than 300 bp after masking the repetitive sequences were categorized as junk sequences. The remaining sequences were BLAST-searched against all public human cDNA sequences (downloaded on Aug 3, 2007) for the forward strand. Homologous sequences to the human cDNAs were classified as orphan transcripts for the forward strand and anti-transcript for the reverse strand. The remaining 947 clones were mapped on the human genome sequence (build 36.1) by BLAST algorithm and arranged according to the annotation from the UCSC genome browser (hg18). The transcripts that overlapped with the genic regions including UTR were classified as intronic transcripts, and the transcripts that were mapped more than 5 kb away from the genic region were classified as intergenic transcripts.

Expression assays

Affymetrix GeneChip was designed using the available cDNA sequences of *M. fascicularis*. The chip loads 10,307

probe sets. RNA samples from the cerebrum, cerebellum, liver, and testis of a 3-year-old male cynomolgus monkey were extracted using TRIZOL (Invitrogen) and hybridized to the GeneChip with duplication in a single experiment. The *M. fascicularis* GeneChip contains at most 11 perfect-match probes (25-mers complete matches to the cDNA sequences) and 11 mismatch probes (containing one mismatched oligo) for each probe set, similar to other GeneChip formats. Normalization, signal detection, and signal intensity calculation of the microarrays were performed using Affymetrix MAS5.0 software. Transcripts were considered as expressed when the probe set of both the duplicates agreed for the significant expression ($P \leq 0.05$) [33]. The raw array data were deposited in Gene Expression Omnibus [GEO: GSM201873–201880]. The array design and the sequences of oligonucleotide probes were deposited in the public database [GEO: GPL5396].

RT-PCR

Templates of the human brain RNA were purchased from Clontech. The macaque brain RNA was obtained from a 21-year-old male cynomolgus monkey. One microgram of total mRNA was amplified using the PrimeSTAR® RT-PCR Kit (TakaraBio). The temperature and time schedules were 30 cycles at 94°C for 20 s, 60°C for 30 s, and 72°C for 1 min. All primer sequences are presented in Additional file 6.

Human-cynomolgus cDNA sequence alignment

Human-macaque orthologous gene pairs were assigned by the reciprocal best BLAST hit with an e-60 cut-off value. We aligned only that part of the coding sequences (CDS) that was homologous to a BLAST search, because representative human and macaque cDNAs do not necessarily have the same splicing isoforms. The sequences of human and macaque cDNAs were aligned using CLUSTAL W [51], and an unaligned macaque nucleotide was marked by the letter X in the database. Alignments shorter than 100 bp (≤ 33 codons) were filtered for further analysis. In the database, the positions including deletion in the human sequence (or insertion in the macaque sequence) were dropped for estimating the substitution rates. The non-synonymous substitution rate per non-synonymous site (*Ka*) and the synonymous substitution rate per synonymous site (*Ks*) were estimated using the Li-Pamilo-Bianchi method [52,53]. *Ka*/*Ks* ratios were set to 100 in the database when the *Ks* value was zero.

Human-rhesus-cynomolgus cDNA sequence alignment

The predicted cDNA sequences of rhesus macaques were downloaded from Ensembl (MMUL1.0) and aligned with the human RefSeq sequences. Orthology between the rhesus and cynomolgus genes was confirmed again using the cynomolgus-rhesus reciprocal BLAST hit, and human-rhesus-cynomolgus cDNA alignments were compiled. Alignments containing any frameshifting indels and those shorter than 100 bp (≤ 33 codons) were filtered, which resulted in 2655 alignments (dataset I). The rhesus cDNA sequences were then mapped on the draft genome sequence of the rhesus macaque (rheMac2). The rhesus macaque genes showing > 80% homology to more than one locus on the rhesus genome were removed from the alignments, which yielded 1499 human-rhesus-cynomolgus cDNA alignments (dataset II). To estimate the divergence among the three species, *Ka*(*dn*) and *Ks*(*ds*) were estimated using the maximum likelihood method implemented in the PAML program package [54]. We estimated the transition/transversion ratios in 4-fold degenerated sites using the concatenated cynomolgus and rhesus alignments in advance, and fixed the value to the observed value. The test of positive selection was conducted using the branch-site test of positive selection described by Zhang et al. [41], applying the critical values of 2.71 and 5.41 at 5% and 1% significance level without a Bonferroni correction, respectively.

Estimation of the divergence time between cynomolgus and rhesus macaques

Maximum likelihood estimation (MLE) of the divergence time and ancestral population size was performed using the method of Takahata and Satta [11,14]. MLE was determined using the Newton-Raphson algorithm with many possible initial values. The standard error was determined from the numerically evaluated Fisher information matrix. In order to correct a PCR error rate, we estimated the PCR error rate to be 5.40 × 10-4, which was derived from the difference in the synonymous substitution rates of the cynomolgus and rhesus lineages. We assumed that the generation time and the effect of selection on the synonymous sites of the two macaques were the same, and that the erroneous nucleotide incorporated by PCR did not skew. Therefore, when a synonymous substitution in the cynomolgus lineage was found, it was considered that the substitution is because of the PCR error with a probability of 0.178 (5.40 × 10-4/3.04 × 10-3). We randomly corrected the number of substitutions in the raw data, generated pseudo data for 1000 times, and estimated the evolutionary parameter for each time.

Abbreviations

CDS, coding sequence; EST, expressed sequence tag; MLE, maximum likelihood estimation; ORF, open reading frame; UTR, untranslated region; MRCA, most recent common ancestor

Authors' contributions

NO contributed to the designing of the research, performed the experiments and data analysis, and wrote the manuscript. KH, KT, and JK designed the research and contributed to the manuscript. M Hirata performed the

computational analysis. YK contributed to the microarray experiments. RT, YU, H, and IT were involved in the cDNA sequencing. M Hida, YS and SS constructed the oligo-capped cDNA libraries. All authors read and approved the final manuscript.

Additional material

Additional file 1

Expression of novel macaque transcripts. Significance of gene expression in the M. fascicularis *oligonucleotide microarray analysis is shown.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-90-S1.xls]

Additional file 2

Unidentified UTR regions in the macaque cDNAs. The regions of macaque cDNAs that did not show homology to human cDNAs are listed.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-90-S2.xls]

Additional file 3

Divergence among the human, cynomolgus, and rhesus genes (dataset I: without a duplication filtering). Estimation of gene divergence using 2655 human-rhesus-cynomolgus alignment is shown.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-90-S3.doc]

Additional file 4

LRT (likelihood ratio test) statistics for the test of positive selection. Candidate genes under positive selection using branch-site test of positive selection are given with log-likelihood ratio.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-90-S4.xls]

Additional file 5

Specific probes for known genes in the M. fascicularis *microarray. These genes are not found in the previously published macaque oligonucleotide microarray.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-90-S5.xls]

Additional file 6

Primer sequences for RT-PCR. The list shows the primer sequences that were used for RT-PCR.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-90-S6.xls]

Acknowledgements

This study was supported by a Health Science Research Grant from the Ministry of Health, Labor and Welfare of Japan and a Grant for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan (19770073).

References

1. Carlsson HE, Schapiro SJ, Farah I, Hau J: **Use of primates in research: a global overview.** *Am J Primatol* 2004, 63:225-237.
2. Melnick DJ, Hoelzer GA, Absher R, Ashley MV: **mtDNA diversity in rhesus monkeys reveals overestimates of divergence time and paraphyly with neighboring species.** *Mol Biol Evol* 1993, 10:282-295.
3. Hayasaka K, Fujii K, Horai S: **Molecular phylogeny of macaques: implications of nucleotide sequences from an 896-base pair region of mitochondrial DNA.** *Mol Biol Evol* 1996, 13:1044-1053.
4. Magness CL, Fellin PC, Thomas MJ, Korth MJ, Agy MB, Proll SC, Fitzgibbon M, Scherer CA, Miner DG, Katze MG, Iadonato SP: **Analysis of the Macaca mulatta transcriptome and the sequence divergence between Macaca and human.** *Genome Biol* 2005, 6:R60.
5. Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, Remington KA, Strausberg RL, Venter JC, Wilson RK, Batzer MA, Bustamante CD, Eichler EE, Hahn MW, Hardison RC, Makova KD, Miller W, Milosavljevic A, Palermo RE, Siepel A, Sikela JM, Attaway T, Bell S, Bernard KE, Buhay CJ, Chandrabose MN, Dao M, Davis C, Delehaunty KD, Ding Y, *et al.*: **Evolutionary and biomedical insights from the rhesus macaque genome.** *Science* 2007, 316:222-234.
6. Dutrillaux B, Biemont MC, Viegas Pequignot E, Laurent C: **Comparison of the karyotypes of four Cercopithecoidae: Papio papio, P. anubis, Macaca mulatta, and M. fascicularis.** *Cytogenet Cell Genet* 1979, 23:77-83.
7. Tosi AJ, Morales JC, Melnick DJ: **Paternal, maternal, and biparental molecular markers provide unique windows onto the evolutionary history of macaque monkeys.** *Evolution Int J Org Evolution* 2003, 57:1419-1435.
8. Ferguson B, Street SL, Wright H, Pearson C, Jia Y, Thompson SL, Allibone P, Dubay CJ, Spindel E, Norgren RB Jr: **Single nucleotide polymorphisms (SNPs) distinguish Indian-origin and Chinese-origin rhesus macaques (*Macaca mulatta*).** *BMC Genomics* 2007, 8:43.
9. Hernandez RD, Hubisz MJ, Wheeler DA, Smith DG, Ferguson B, Rogers J, Nazareth L, Indap A, Bourquin T, McPherson J, Muzny D, Gibbs R, Nielsen R, Bustamante CD: **Demographic histories and patterns of linkage disequilibrium in Chinese and Indian rhesus macaques.** *Science* 2007, 316:240-243.
10. Smith DG, McDonough JW, George DA: **Mitochondrial DNA variation within and among regional populations of longtail macaques (*Macaca fascicularis*) in relation to other species of the fascicularis group of macaques.** *Am J Primatol* 2007, 69:182-198.
11. Edwards SV, Beerli P: **Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies.** *Evolution Int J Org Evolution* 2000, 54:1839-1854.
12. Takahata N, Satta Y: **Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences.** *Proc Natl Acad Sci USA* 1997, 94:4811-4815.
13. Tosi AJ, Morales JC, Melnick DJ: **Y-Chromosome and Mitochondrial Markers in Macaca fascicularis Indicate Introgression with Indochinese M. mulatta and a Biogeographic Barrier in the Isthmus of Kra.** *Int J Primatol* 2002, 23:161-178.
14. Osada N, Wu CI: **Inferring the mode of speciation from genomic data: a study of the great apes.** *Genetics* 2005, 169:259-264.
15. Ota T, Suzuki Y, Nishikawa T, Otsuki T, Sugiyama T, Irie R, Wakamatsu A, Hayashi K, Sato H, Nagai K, Kimura K, Makita H, Sekine M, Obayashi M, Nishi T, Shibahara T, Tanaka T, Ishii S, Yamamoto J, Saito K, Kawai Y, Isono Y, Nakamura Y, Nagahari K, Murakami K, Yasuda T, Iwayanagi T, Wagatsuma M, Shiratori A, Sudo H, *et al.*: **Complete sequencing and characterization of 21,243 full-length human cDNAs.** *Nat Genet* 2004, 36:40-45.
16. Osada N, Hida M, Kusuda J, Tanuma R, Iseki K, Hirata M, Suto Y, Hirai M, Terao K, Suzuki Y, *et al.*: **Assignment of 118 novel cDNAs of cynomolgus monkey brain to human chromosomes.** *Gene* 2001, 275:31-37.
17. Osada N, Hida M, Kusuda J, Tanuma R, Hirata M, Hirai M, Terao K, Suzuki Y, Sugano S, Hashimoto K: **Prediction of unidentified human genes on the basis of sequence similarity to novel**

cDNAs from cynomolgus monkey brain. *Genome Biol* 2002, 3:RESEARCH0006.
18. Osada N, Hida M, Kusuda J, Tanuma R, Hirata M, Suto Y, Hirai M, Terao K, Sugano S, Hashimoto K: **Cynomolgus monkey testicular cDNAs for discovery of novel human genes in the human genome sequence.** *BMC Genomics* 2002, 3:36.
19. Osada N, Kusuda J, Hirata M, Tanuma R, Hida M, Sugano S, Hirai M, Hashimoto K: **Search for genes positively selected during primate evolution by 5'-end-sequence screening of cynomolgus monkey cDNAs.** *Genomics* 2002, 79:657-662.
20. Osada N, Hirata M, Tanuma R, Kusuda J, Hida M, Suzuki Y, Sugano S, Gojobori T, Shen CK, Wu CI, Hashimoto K: **Substitution rate and structural divergence of 5'-UTR evolution: comparative analysis between human and cynomolgus monkey cDNAs.** *Mol Biol Evol* 2005, 22:1976-1982.
21. Wang HY, Chien HC, Osada N, Hashimoto K, Sugano S, Gojobori T, Chou CK, Tsai SF, Wu CI, Shen CK: **Rate of Evolution in Brain-Expressed Genes in Humans and Other Primates.** *PLoS Biol* 2007, 5:e13.
22. Maruyama K, Sugano S: **Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides.** *Gene* 1994, 138:171-174.
23. Maglott D, Ostell J, Pruitt KD, Tatusova T: **Entrez Gene: gene-centered information at NCBI.** *Nucleic Acids Res* 2005, 33:D54-58.
24. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, 25:25-29.
25. **QFbase** [http://genebank.nibio.go.jp/qfbase/]
26. Benson DA, Karsch Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: **GenBank.** *Nucleic Acids Res* 2007, 35:D21-25.
27. Hubbard TJ, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates G, Cunningham F, Cutts T, Down T, Dyer SC, Fitzgerald S, Fernandez Banet J, Graf S, Haider S, Hammond M, Herrero J, Holland R, Howe K, Howe K, Johnson N, Kahari A, Keefe D, Kokocinski F, Kulesha E, Lawson D, Longden I, Melsopp C, Megy K, *et al.*: **Ensembl 2007.** *Nucleic Acids Res* 2007, 35:D610-617.
28. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA: **Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders.** *Nucleic Acids Res* 2005, 33:D514-517.
29. Imanishi T, Itoh T, Suzuki Y, O'Donovan C, Fukuchi S, Koyanagi KO, Barrero RA, Tamura T, Yamaguchi-Kabata Y, Tanino M, Yura K, Miyazaki S, Ikeo K, Homma K, Kasprzyk A, Nishikawa T, Hirakawa M, Thierry-Mieg J, Thierry-Mieg D, Ashurst J, Jia L, Nakao M, Thomas MA, Mulder N, Karavidopoulou Y, Jin L, Kim S, Yasuda T, Lenhard B, Eveno E, *et al.*: **Integrative annotation of 21,037 human genes validated by full-length cDNA clones.** *PLoS Biol* 2004, 2:e162.
30. Kuhn RM, Karolchik D, Zweig AS, Trumbower H, Thomas DJ, Thakkapallayil A, Sugnet CW, Stanke M, Smith KE, Siepel A, Rosenbloom KR, Rhead B, Raney BJ, Pohl A, Pedersen JS, Hsu F, Hinrichs AS, Harte RA, Diekhans M, Clawson H, Bejerano G, Barber GP, Baertsch R, Haussler D, Kent WJ: **The UCSC genome browser database: update 2007.** *Nucleic Acids Res* 2007, 35:D668-673.
31. Li WH: **Molecular Evolution.** Sinauer Associates, Sunderland, MA; 1997.
32. Kapranov P, Drenkow J, Cheng J, Long J, Helt G, Dike S, Gingeras TR: **Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays.** *Genome Res* 2005, 15:987-997.
33. Liu WM, Mei R, Di X, Ryder TB, Hubbell E, Dee S, Webster TA, Harrington CA, Ho MH, Baid J, Smeekens SP: **Analysis of high density expression microarrays with signed-rank call algorithms.** *Bioinformatics* 2002, 18:1593-1599.
34. Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, Zhu X, Rinn JL, Tongprasit W, Samanta M, Weissman S, Gerstein M, Snyder M: **Global identification of human transcribed sequences with genome tiling arrays.** *Science* 2004, 306:2242-2246.
35. Kampa D, Cheng J, Kapranov P, Yamanaka M, Brubaker S, Cawley S, Drenkow J, Piccolboni A, Bekiranov S, Helt G, Tammana H, Gingeras TR: **Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22.** *Genome Res* 2004, 14:331-342.
36. Khaitovich P, Kelso J, Franz H, Visagie J, Giger T, Joerchel S, Petzold E, Green RE, Lachmann M, Paabo S: **Functionality of intergenic transcription: an evolutionary comparison.** *PLoS Genet* 2006, 2:e171.
37. Kondrashov FA, Koonin EV: **Evolution of alternative splicing: deletions, insertions and origin of functional parts of proteins from intron sequences.** *Trends Genet* 2003, 19:115-119.
38. Kimura K, Wakamatsu A, Suzuki Y, Ota T, Nishikawa T, Yamashita R, Yamamoto J, Sekine M, Tsuritani K, Wakaguri H, Ishii S, Sugiyama T, Saito K, Isono Y, Irie R, Kushida N, Yoneyama T, Otsuka R, Kanda K, Yokoi T, Kondo H, Wagatsuma M, Murakawa K, Ishida S, Ishibashi T, Takahashi-Fujii A, Tanase T, Nagai K, Kikuchi H, Nakai K, *et al.*: **Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes.** *Genome Res* 2006, 16:55-65.
39. Xing Y, Lee C: **Alternative splicing and RNA selection pressure – evolutionary consequences for eukaryotic genomes.** *Nat Rev Genet* 2006, 7:499-509.
40. Zhang J, Nielsen R, Yang Z: **Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level.** *Mol Biol Evol* 2005, 22:2472-2479.
41. Bakewell MA, Shi P, Zhang J: **More genes underwent positive selection in chimpanzee evolution than in human evolution.** *Proc Natl Acad Sci USA* 2007, 104:7489-7494.
42. Stewart CB, Disotell TR: **Primate evolution – in and out of Africa.** *Curr Biol* 1998, 8:R582-588.
43. Kaessmann H, Wiebe V, Paabo S: **Extensive nuclear DNA sequence diversity among chimpanzees.** *Science* 1999, 286:1159-1162.
44. Yi S, Ellsworth DL, Li WH: **Slow molecular clocks in Old World monkeys, apes, and humans.** *Mol Biol Evol* 2002, 19:2191-2198.
45. Kondrashov AS: **Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases.** *Hum Mutat* 2003, 21:12-27.
46. Chen WH, Wang XX, Lin W, He XW, Wu ZQ, Lin Y, Hu SN, Wang XN: **Analysis of 10,000 ESTs from lymphocytes of the cynomolgus monkey to improve our understanding of its immune system.** *BMC Genomics* 2006, 7:82.
47. Wallace JC, Korth MJ, Paeper B, Proll SC, Thomas MJ, Magness CL, Iadonato SP, Nelson C, Katze MG: **High-density rhesus macaque oligonucleotide microarray design using early-stage rhesus genome sequence information and human genome annotations.** *BMC Genomics* 2007, 8:28.
48. Suzuki Y, Sugano S: **Construction of a full-length enriched and a 5'-end enriched cDNA library using the oligo-capping method.** *Methods Mol Biol* 2003, 221:73-91.
49. Jurka J: **Repbase update: a database and an electronic journal of repetitive elements.** *Trends Genet* 2000, 16:418-420.
50. Pruitt KD, Tatusova T, Maglott DR: **NCBI Reference Sequence project: update and current status.** *Nucleic Acids Res* 2003, 31:34-37.
51. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, 22:4673-4680.
52. Li WH: **Unbiased estimation of the rates of synonymous and nonsynonymous substitution.** *J Mol Evol* 1993, 36:96-99.
53. Pamilo P, Bianchi NO: **Evolution of the Zfx and Zfy genes: rates and interdependence between the genes.** *Mol Biol Evol* 1993, 10:271-281.
54. Yang Z: **PAML: a program package for phylogenetic analysis by maximum likelihood.** *Comput Appl Biosci* 1997, 13:555-556.
55. Kimura M: **A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences.** *J Mol Evol* 1980, 16:111-120.

ELSEVIER

Journal of Hepatology 48 (2008) 43–50

Journal of Hepatology

www.elsevier.com/locate/jhep

Impact of hepatitis B virus (HBV) X gene integration in liver tissue on hepatocellular carcinoma development in serologically HBV-negative chronic hepatitis C patients☆

Hidenori Toyoda[1], Takashi Kumada[1], Yuji Kaneoka[2], Yoshiki Murakami[3,*,†]

[1]*Department of Gastroenterology, Ogaki Municipal Hospital, 4-86 Minaminokawa, Ogaki, Gifu 503-8502, Japan*
[2]*Department of Surgery, Ogaki Municipal Hospital, 4-86 Minaminokawa, Ogaki, Gifu 503-8502, Japan*
[3]*Laboratory of Human Tumor Virus, Institute for Viral Research, Kyoto University, Shogoin-Kawaharacho, Sakyo-ku, Kyoto 606-8507, Japan*

***Background/Aims*: We analyzed hepatitis B virus (HBV) X gene integration in hepatocytes of HBV-negative, chronic hepatitis C (CH-C) patients with mild fibrosis, and prospectively followed these patients for the development of hepatocellular carcinoma (HCC).**

***Methods*: The study included 39 HBV-negative CH-C patients with mild fibrosis. HBV-X integration was determined by Alu-PCR analysis of liver specimens obtained by fine-needle biopsy.**

***Results*: Integration of HBV-X gene sequence into liver genome occurred in 9 of the 39 patients. Six of the 39 patients developed HCC during the 12-year follow-up period. No significant difference was found in the incidence of HCC between patients with and without HBV-X integration. However, the two patients with HBV-X integration who developed HCC did not have cirrhosis at the time when HCC was diagnosed, whereas the four patients without HBV-X integration who developed HCC did have cirrhosis.**

***Conclusions*: Our findings suggest that HBV-X integration detected at the mild fibrosis stage might not indicate a high risk for HCC. HBV-X integration may be associated with HCC development in the absence of cirrhosis. However, we did not find evidence that HBV-X integration directly plays a role in hepatocarcinogenesis in CH-C patients. Further studies will be needed to clarify this point.**

***Keywords*: HBV-X integration; Chronic hepatitis C; Hepatocellular carcinoma**

1. Introduction

Chronic viral hepatitis is a leading cause of hepatocellular carcinoma (HCC) worldwide [1–4]. Occult hepatitis B virus (HBV) infection, characterized by the absence of circulating HBV surface antigen [HBsAg] but presence of the HBV genome in serum or liver tissue, has been identified in hepatitis C virus (HCV)-infected patients. HBV may affect the clinical course of chronic hepatitis C (CH-C) [5] and increase the risk of hepatocarcinogenesis [6]. Pollicino reported that both integrated and free HBV–DNA sequences were highly prevalent in the liver tissue of CH-C patients with HCC compared to CH-C patients without HCC [7].

Received 28 March 2007; received in revised form 5 August 2007; accepted 8 August 2007; available online 24 October 2007
Associate Editor: K. Koike

☆ The authors declare that they do not have anything to disclose regarding funding or conflict of interest with respect to this manuscript.

* Corresponding author. Tel.: +81 75 751 4034; fax: +81 75 751 3998.
E-mail addresses: ymurakam@virus.kyoto-u.ac.jp, ymurakami@genome.med.kyoto-u.ac.jp (Y. Murakami).
† Present address: Unit of Human Disease Genomics, Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Yoshida-Konoe cho, Sakyo-ku, Kyoto 606-8501, Japan. Tel.: +81 75 753 9313; fax: +81 75 753 9314.

0168-8278/$32.00
doi:10.1016/j.jhep.2007.08.016

These observations support the clinical importance of occult HBV as a carcinogenic factor in HBsAg-negative patients with CH-C. However, it remains controversial whether occult HBV increases the risk of HCC in this population [8].

Several studies have investigated the association between HBV integration and HCC in patients with both chronic HCV infection and HCC [8–10]. However, no study has prospectively evaluated whether HBV integration in liver tissue correlates with HCC development in CH-C patients. In a prospective 12-year study, we attempted to clarify whether HBV integration promotes hepatocarcinogenesis in CH-C patients.

2. Materials and methods

2.1. Patients

A total of 67 HBsAg-negative, CH-C patients underwent ultrasonography (US)-guided fine-needle liver biopsy for histological evaluation between January and December 1994. Of these patients, 39 had chronic hepatitis with mild fibrosis (METAVIR score of F0 or F1) [11] and were included in the study. Clinical characteristics of these patients are summarized in Table 1. The patient group contained 30 men and 9 women with a mean age of 49.0 ± 7.6 years. All patients were negative for both serum HBsAg and HBV–DNA and were shown to have persistent HCV infection by nested reverse transcription-polymerase chain reaction (PCR). Sixteen of thirty-nine patients had a history of blood transfusion. No patient had a history of intravenous drug use, tattooing, or acupuncture. No patient had a history of acute hepatitis B. All patients were followed from the time of liver biopsy until October 2006. They underwent periodic US examination and analysis for HCC tumor markers, including α-fetoprotein and des-γ-carboxy prothrombin every 6 months. When a suspicious liver lesion was detected by US or a tumor marker was elevated, the patient underwent further examination by imaging such as computed tomography (CT), magnetic resonance imaging, or angiography. HCC was diagnosed on the basis of typical imaging findings, which include a mosaic pattern with a halo on B-mode US images, hypervascularity on angiographic images, or a high-density mass on arterial-phase dynamic CT images with a low-density mass on portal-phase dynamic CT images obtained with a helical or multidetector raw CT scanner. All patients who developed HCC underwent a hepatectomy; all tumors were less than 3 cm in diameter when detected under this surveillance. The final diagnosis of HCC was based on histologic examination of the tumor tissue taken from resected specimens.

The study protocol conformed to the ethics guidelines of the Declaration of Helsinki (1975). All patients provided written informed consent for analysis of the biopsy specimens, and the Hospital Ethics Committee approved the study.

Table 1
Clinical characteristics of the study patients (n = 39)

Age (years)	49.0 ± 7.6
Sex (female/male)	9(23.1)/30(76.9)[#]
History of blood transfusion	15 (38.5)
Presumed duration of HCV carriage[*]	19.0 (5–33)[##]
Alanine aminotransferase (IU/L)	60.1 ± 31.4
Aspartate aminotransferase (IU/L)	45.0 ± 23.8
Gamma glutamyl transpeptidase (mg/L)	51.2 ± 55.3
Albumin (g/dL)	4.11 ± 0.33
Total-bilirubin (mg/dL)	0.74 ± 0.33
Platelet count (×10^4/ml)	17.9 ± 6.5
HCV RNA concentration (×10^{3}IU/mL)	570 (3–4900)[##]
HCV genotype	
1b	25(64.1)[#]
2a	11 (28.2)[#]
2b	3 (7.7)[#]
HBV surface antigen	0
HBV surface antibody	6(15.4)[#]
HBV core antibody	25(64.1)[#]
Fibrosis stage[**]	
F0	14 (35.9)[#]
F1	25(64.1)[#]

HBV, hepatitis B virus; HCV, hepatitis C virus.
[*] In patients with a history of blood transfusion.
[**] According to METAVIR score.
[#] Percentages are shown in parentheses.
[##] Median; ranges are shown in parentheses.

2.2. Sample preparation

DNA was extracted from liver tissues obtained at liver biopsy on 1994 with a DNeasy Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. All samples were stored at −80 °C and carefully handled to avoid contamination with nucleic acids.

2.3. Detection of viral–host junctions

A PCR-based technique, Alu-PCR, one of the most effective procedures to detect junctions between integrated HBV–DNA and human DNA, was used to amplify viral–host junctions using 100 ng of genomic DNA [12–14] (Table 2). The sensitivity study for this PCR was performed using human hepatoma cell line Huh-2 cells that contain 1 copy per cell of integrated HBV (kindly provided by Dr. K Koike from Department of Gene Research, Cancer Institute, Tokyo) [15]. Amplified PCR products were analyzed by electrophoresis on 1.0% agarose gel and transferred to a Hybond-N$^+$ nylon membrane (Amersham Pharmacia, Buckinghamshire, UK). About 3.2 kb of the HBV X genome (HBV-X) was amplified according to the method of Günther et al. [16]. HBV-specific bands were then detected by hybridization with a DIG labeled HBV probe (Roche, Mannheim, Germany).

2.4. Direct sequencing

The amplified viral/host junctions were purified with an Easy Trap Kit (Takara, Otsu, Japan) and sequenced using a Prism Taq DyeDeoxy Terminator cycle sequencing kit (Applied Biosystems, Foster City, CA), according to the manufacturer's instructions. Products were precipitated with ethanol and analyzed with a 377 Prism DNA Sequencer (Applied Biosystems Inc.). To identify the HBV-X integration site, we used BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) to compare sequences adjacent to the integrated HBV–DNA with the human genome.

2.5. Other serological and virological tests

HBV surface antigen, surface antibody, and core antibody were measured with ARCHITECT HBsAg QT, ARCHITECT anti-HBs, and ARCHITECT anti-HBc, respectively (Abbott Japan, Tokyo, Japan). Serum HBV–DNA was measured by the Amplicor HBV test (detection limit, 400 copies/mL; Roche Diagnostics, Branchburg, NJ). HCV genotype was determined by PCR with genotype-specific primers [17,18]. HCV RNA concentration was measured by a quantitative PCR assay (detection limit, 5000 copies/mL; Amplicor GT-HCV Monitor, Version 2.0; Roche Molecular Systems, Pleasanton, CA).

2.6. Statistical analyses

Data are expressed as means ± SD or the median and range. Differences in the proportion of patients with and without HBV-X integration were analyzed by χ^2 test. Differences in quantitative values were analyzed by Mann–Whitney U test. For the incidence of HCC

Table 2
Sequences of primers for detection of viral–host junctions

Primer name	Primer sequence	HBV portion and note
UP5	5′-CAGUGCCAAGUGUUUGCUGACGCCAAAGUGCUGGGAUUA-3′	Alu-sense
T3-515	5′-AUUAACCCUCACUAAAGCCUCGAUAGAUYRYRCCAYUGCAC-3′	Alu-antisense
UP6	5′-CAAGTGTTTGCTGACGCCAAAG-3′	Alu-sense (tag)
midT3	5′-ATTAACCCTCACTAAAGCCTCG-3′	Alu-antisense (tag)
pUTP	5′-ACAUGAACCUUUACCCCGUUGC-3′	1131–1152 HB1 (HBV-X)
MD37	5′-TGCCAAGTGTTTGCTGACGC-3′	1174–1193 HB2 (HBV-X)
MD60	5′-CTGCCGATCCATACTGCGGAAC-3′	1258–1279 HB3 (HBV-X)

Numbering of nucleotides is according to Ono et al. [31]. U = dUTP; Y = (C,T); R = (A,G).

development, the date of the initial liver biopsy was defined as time zero. Data pertaining to patients who did not develop HCC were censored. The Kaplan–Meier method was used to calculate the incidence of HCC, and the log-rank test was used to analyze differences. The JMP statistical software package, version 4.0. (SAS Institute. Cary. NC) was used for all statistical analyses. All p values were derived from two-tailed tests, and $p < 0.05$ was considered statistically significant.

3. Results

3.1. Integration of hepatitis B viral genome and patient characteristics

The sensitivity of the PCR amplification was first determined with hepatoma cell line Huh-2 cells. When we made a tenfold serial dilution of Huh-2 cells with normal human PBMC without a history of liver disease, we could detect viral–host junctions at about 100 copies per reaction by the PCR (Fig. 1a).

We amplified virus–host DNA junctions from the liver of CH-C patients and detected several bands on 1.0%-agarose gels (Fig. 1b). Sequencing these PCR products revealed HBV-X integration in 9 of the 39 (23.1%) patients. Nineteen viral–host junctions were detected in these 9 patients. In 4 of these 9 patients, multiple integration sites (range, 2–6) were present. For example, 6 viral–host junctions were detected in patient 15, and the adjacent host sequences were from 6 different chromosomes (red circle, Fig. 2). In the other 5 cases, a single integration site was detected. The sites of HBV-X integration are shown in Fig. 2.

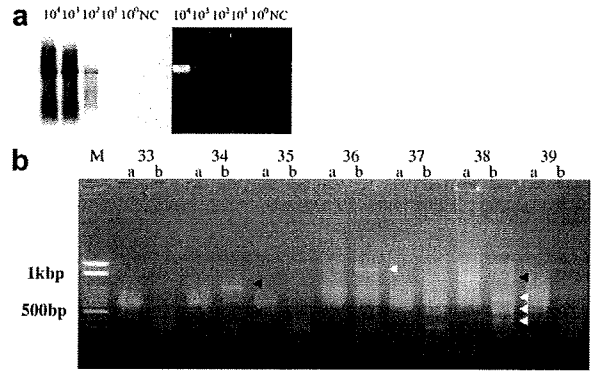

Fig. 1. The detection of HBV-X–host junction by Alu-PCR analysis. (a) The sensitivity study of Alu-PCR method. Serially diluted genomic DNA contained with HBV integrant was amplified by using HBV-X and Alu antisense primer pair. Left is Southern blot analysis from the gel electropheresis (right). (b) The numbers indicate the individual patients, and a and b indicate the primer pair used for amplification (a, HBV-X primer and Alu sense; b, HBV-X primer and Alu antisense). The PCR strategy and the primer sequences used in this study were previously described [12–14]. Arrowheads indicate PCR products with HBV-X–host junctional sequences (white) and without HBV-X–host junctions (black).

Clinical characteristics of patients with and without HBV-X integration are summarized in Table 3. There were no differences in the clinical characteristics. During the observation period, 4 of 9 (44.4%) patients with HBV-X integration and 13 of 30 (43.3%) patients without HBV-X integration received interferon monotherapy. These percentages did not differ significantly.

3.2. Host genome sequences at sites of HBV-X integration

The sites of host integration were divided into two groups: (1) genes already known and/or fully characterized but not previously shown to be involved in carcinogenesis (1 integration site; T cell lymphoma invasion and metastasis 1 [TIAM1] in Patient 8), and (2) unknown open reading frames (ORFs) or genes belonging to a known gene family but not functionally characterized (18 integration sites). The HBV genome ORF was integrated in both the same and opposite orientations of the host gene and both proximal to and into host genes (Table 4).

3.3. Development of HCC

Over the 12-year follow-up period, HCC developed in 6 of the 39 (15.4%) patients. HCC developed in 4 of the 30 (13.3%) patients without HBV-X integration and in 2 of the 9 (22.2%) patients with HBV-X integration (Fig. 3). The difference in the incidence of HCC between patients with and without HBV-X integration was not significant ($p = 0.8041$). Patient age, sex, and histologic data at the time of HBV-X integration analysis and at the time of HCC diagnosis are shown in Table 5. All patients who developed HCC were males. Age at the time HCC developed did not differ between patients

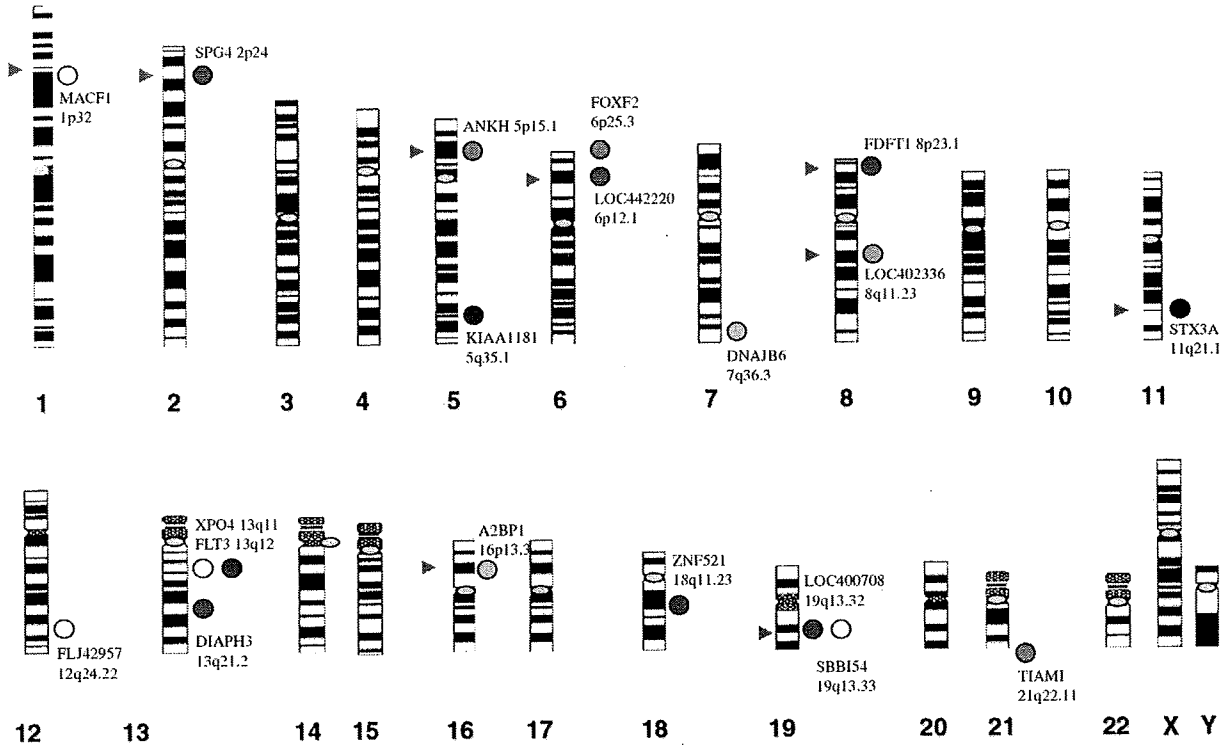

Fig. 2. Chromosomal distribution of HBV-X integration sites. Circles indicate viral integration sites, and the circle color denotes the sample. For example, the three white spots indicate three viral integration sites detected in the same specimen. Gene names and chromosomal localizations are also noted. Red arrowheads indicate DNA fragile sites [32].

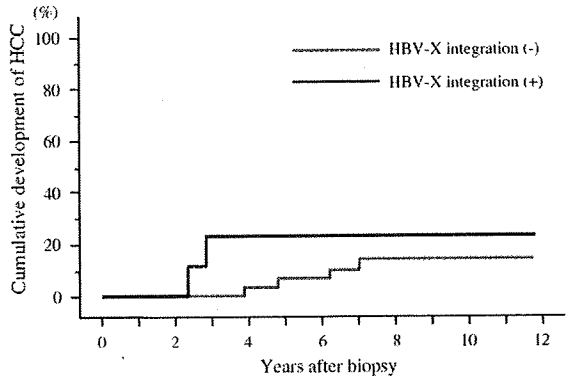

Fig. 3. Kaplan–Meier curves for the incidence of hepatocellular carcinoma (HCC). The blue and red lines represent the incidence of HCC in patients with and without HBV-X integration, respectively. No significant difference was observed between the two groups ($p = 0.8041$).

with and without HBV-X integration. Five of 6 patients who developed HCC (except for patient No. 34) had received interferon therapy, but all of them remained HCV positive. All 4 patients without HBV-X integration who developed HCC had cirrhosis at the time HCC was diagnosed. In contrast, the fibrosis stage was moderate or mild (F1 or F2) in the 2 patients with HBV-X integration who developed HCC. No patient was positive for the circulating low-level HBV–DNA analyzed with a highly sensitive HBV–DNA detection method (detection limit, 35 copies/mL; COBAS TaqMan HBV test, Roche Diagnostics) at the time of HCC diagnosis [19].

We attempted to detect HBV–host junction by the same Alu-PCR method in resected HCC materials that developed in 4 patients (patients #9, 21, 34, and 38) using paraffin-embedded samples. HBV-X integration was detected in HCC materials of none of 4 patients (data not shown).

4. Discussion

This is the first prospective study to analyze HBV integration into the host hepatocyte genome of CH-C patients with mild fibrosis and then to follow these patients over a long period for the development of HCC. Previous studies investigated HBV integration in HCC tissue of patients chronically infected with HCV [8–10] or in HCC tissue of patients without hepatitis virus infection [20]. However, in these studies, HBV integration was analyzed in cancerous and non-cancerous tissue after the development of HCC, and thus the effect of HBV integration on the development of HCC in CH-C patients was not investigated.

Table 3
Characteristics of patients with and without HBV-X–DNA integration

	HBV-X–DNA integration (−) ($n = 30$)	HBV-X–DNA integration (+) ($n = 9$)
Age (years)	48.9 ± 7.6	49.6 ± 7.7
Sex (female/male)	6 (20.0)/24 (80.0)[#]	3 (33.3)/6 (66.7)[#]
History of blood transfusion	11(36.7)[#]	4 (44.4)[#]
Presumed duration of HCV carriage[*]	19.0 (5–33)[##]	22.5 (12–33)[##]
Alanine aminotransferase (IU/L)	60.0 ± 31.8	60.4 ± 31.8
Aspartate aminotransferase (IU/L)	46.2 ± 25.9	41.0 ± 15.8
Gamma glutamyl transpeptidase (mg/L)	49.5 ± 39.0	34.6 ± 29.2
Albumin (g/dL)	4.08 ± 0.37	4.22 ± 0.14
Total-bilirubin (mg/dL)	0.70 ± 0.33	0.84 ± 0.33
Platelet count ($\times 10^4$/ml)	18.0 ± 5.3	17.6 ± 9.9
HCV RNA concentration ($\times 10^3$IU/mL)	790 (3–4900)[##]	320 (3–2100)[##]
HCV genotype		
1b	19(63.3)[#]	6 (66.7)[#]
2a	8 (26.7)[#]	3 (33.3)[#]
2b	3 (10.0)[#]	0
HBs antibody (+)	4(13.3)[#]	2 (22.2)[#]
HBc antibody (+)	20 (66.7)[#]	5 (55.6)[#]
Fibrosis stage[**]		
F0	10 (33.3)[#]	4 (44.4)[#]
F1	20 (66.7)[#]	5 (55.6)[#]

HBV, hepatitis B virus; HCV, hepatitis C virus.
[*] In patients with a history of blood transfusion.
[**] According to METAVIR score.
[#] Percentages are shown in parentheses.
[##] Median; ranges are shown in parentheses.

Table 4
Genes and sequences of HBV-X–DNA integration sites

No.	Supercontig	Position	Orientation	Chromosomal localization	Name	Location	Name/function
8.	NT034880	1375087	Same	6p25.3	FOXF2	39 kb upstream	Forkhead box F2
8.	NT086666	14245273	Opposite	5p15.1	ANKH	177 kb upstream	Ankylosis, progressive homolog
8.	NT011512	18351760	Same	21q22.11	TIAM1	21.5 kb upstream	T-cell lymphoma invasion and metastasis 1
15.	NT_022184	11183657	Same	2p24	SPG4	Intronic seq	Spastic paraplegia 4 (autosomal dominant; spastin)
15.	NT_024524	41428064	Same	13q21.2	DIAPH3	Intronic seq	Diaphanous homolog 3 (*Drosophila*)
15.	NT011109	19337933	Same	19q13.32	LOC400708	3.1 kb upstream	Similar to Serine/threonine protein phosphatase 5 (PP5)
15.	NT_077531	4155242	Opposite	8p23.1	FDFT1	Intronic seq	Farnesyl-diphosphate farnesyltransferase 1
15.	NT_010966	4345775	Opposite	18q11.2	ZNF521	23 kb upstream	Zinc finger protein 521
15.	NT_007592	46424722	Same	6p12.1	LOC442220	5.3 kb upstream	Similar to nitrogen fixation cluster-like
21.	NT_023133	17103986	Opposite	5q35.1	KIAA1181	38 kb downstream	Endoplasmic reticulum-golgi intermediate compartment 32 kDa protein
22.	NT011109	23275592	Same	19q13.33	SBBI54	Intronic seq	Hypothetical transmembrane protein SBBI54
23.	NT008183	6327145	Opposite	8q11.23	LOC402336	16.9 kb upstream	Similar to L21 ribosomal protein
24.	NT_024524	2436145	Opposite	13q11	XPO4	12.6 kb upstream	Exportin 4
27.	NT_007741	2000247	Opposite	7q36.3	DNAJB6	4 kb downstream	DnaJ (Hsp40) homolog, subfamily B, member 6 Homo sapiens
27.	NT086834	6475804	Opposite	16p13.3	A2BP1	31.9 kb upstream	Ataxin 2-binding protein 1
36.	NT_033903	4799121	Opposite	11q21.1	STX3A	29 kb downstream	Syntaxin3A
38.	NT_009775	7468765	Opposite	12q24.22	FLJ42957	71 kb downstream	FLJ42957 protein
38.	NT_024524	9545675	Opposite	13q12	FLT3	20 kb downstream	Fms-related tyrosine kinase 3
38.	NT004511	9911738	Opposite	1p32	MACF1	Intronic seq	Microtubule-actin crosslinking factor 1

In three studies of HCV-related HCC, the rates of HBV integration in tumor tissue are discrepant: 55.6% (10 out of 18 cases) [8], 29.4% (10 out of 34 cases) [10], and 0% (0 out of 21 cases) [9]. Clonal expansion of hepatocytes containing integrated HBV in association with cancer progression may increase the detection rate of HBV integration. Conversely, clonal expansion of cancerous hepatocytes without HBV integration may decrease the

Table 5
Cases of HCC development

No.	Sex	Age at biopsy	Fibrosis at biopsy	Interval between biopsy and HCC development	Age at HCC development	Fibrosis at HCC development[a]	HBV-X–DNA integration
7.	M	61	F1	4y.	65	F4	(−)
9.	M	57	F1	5y.	62	F4	(−)
21.	M	56	F1	3y.	64	F2	(+)
28.	M	56	F1	5y.	61	F4	(−)
34.	M	47	F1	7y.	54	F4	(−)
38.	M	55	F0	2y.	57	F1	(+)

[a] Non-cancerous tissue.

detection of HBV integration. Therefore, hepatocyte clonal expansion may account for discrepancies in the rates of HBV integration between studies. In contrast, clonal expansion of hepatocytes is unlikely in cases of CH-C with mild fibrosis but without HCC. The prevalence of HBV-X integration in our patient population (23.1%), therefore, represents the actual rate of HBV-X integration in CH-C patients. The number of HBV-X -host integration sites in these patients was smaller than patients with chronic hepatitis B and similar to patients with acute hepatitis B in our previous study with the same detection method for HBV integration [13].

HBV integration is detected in approximately 90% of liver tumor samples from patients with HBsAg [21]. HBV insertional mutagenesis is an important step in many cases of hepatocarcinogenesis in patients with chronic HBV infection. Chromosomal inversions, translocations, or micro deletions can occur at the integration sites, causing tumors to develop in some patients [22,23]. Several tumor-associated genes have been identified adjacent to HBV integration sites [24,25]. However, HBV does not integrate in or near a tumor-associated gene in most HBV-infected individuals. Rather, HBV–DNA integrates randomly into host DNA in HBV-related HCC [21,26,27]. This random integration also appears in patients with HCV-related HCC, although one study suggested that HBV–DNA integrates into tumor-associated genes of some HCC patients without HBsAg [8].

In the present study of CH-C patients without HCC, the HBV-X integration sites were distributed across the genome with little similarity and the host sequences adjacent to the viral genome were divergent. These data are consistent with our previous results on HBV-infected patients with the same detection method for HBV-X integration [14]. In the present study, we did not detect HBV-X integration into genes associated with carcinogenesis. Because HBV–DNA integrates randomly into host DNA and the number of HBV-integration sites was smaller in CH-C patients compared to chronic hepatitis B patients [13], the likelihood of HBV integrating into genes associated with carcinogenesis would be considerably low.

We analyzed HBV-X integration in CH-C patients with mild fibrosis and prospectively observed the patient group for 12 years. There was no statistically significant difference in the incidence of HCC between patients with and without HBV-X integration. Taken together with results from clinical observations and genetic analyses, these data suggest that testing HBV-X integration at a mild fibrosis stage may not predict the likelihood of CH-C patients developing hepatocarcinogenesis. However, the lack of statistical significance in the incidence of HCC could be partly because of the small number of study patients. Future studies with a larger patient population may detect patients with HBV integration in tumor-associated genes and a higher incidence of HCC development in CH-C patients with HBV integration.

In the present study, there was no cirrhosis in non-cancerous liver tissue surrounding the tumor at the time of HCC development, and fibrosis was not severe (stage F1 or F2) in patients with HBV-X integration. In contrast, all 4 HCC patients without HBV-X integration had cirrhosis (stage F4). In addition, the interval between the analysis of HBV-X integration and HCC development was shorter in patients with HBV-X integration than those without HBV-X integration. The stage of fibrosis, especially the presence of cirrhosis, is related closely to the incidence of HCV-related HCC [28], and most patients with HCV-related HCC have cirrhosis [10,29]. Our results showed that HCC develops in the absence of cirrhosis in some CH-C patients with HBV-X integration, and this may suggest the possibility that HBV-X integration may play a role in accelerated hepatocarcinogenesis in some cases. However, we did not detect HBV-X integration in paraffin-embedded resected HCC materials of both 2 patients with HBV-X integration at liver biopsy (patients #21 and #38). Although this can be partly due to the use of paraffin-embedded materials for analyses of integration (unfortunately frozen section was not available), we did not find the evidence that HBV-X integration directly played a role in hepatocarcinogenesis in the present study.

There are several limitations of the study. The detection of HBV integration with PCR using Alu repeats may limit the identification of HBV-X sequence integration sites that are far away from the priming site,

therefore, restricting the sensitivity of the assay as the amplicon size increases. In addition, detection of HBV integration only using the X gene-specific primers makes infeasible identification of integration sites of other virus gene sequences. Further, integrated HBV genome can limit or negate entirely the HBV X primer-binding site, because HBV sequences may be deleted upon integration. The Alu-PCR method used in the present study, therefore, may underestimate the integration of HBV in CH-C patients.

In summary, HBV-X integration was detected in 9 of 39 CH-C patients and the number of HBV-X–host integration sites in these patients was similar to patients with acute hepatitis B. They were distributed across the genome with little similarity. In the prospective observation of CH-C patients over 12 years, HBV-X integration detected at the mild fibrosis stage might not indicate a high risk for HCC during the course of CH-C. Although HBV-X integration may be associated with HCC development in the absence of cirrhosis, we did not find evidence that HBV-X integration directly plays a role for hepatocarcinogenesis in this patient population. Further studies with more sensitive and reliable method than Alu-PCR method for the detection of HBV integration are needed to elucidate the association between HBV integration and HCC development in CH-C patients without cirrhosis. Also, the analyses for HBV integration in frozen sections of resected HCC materials from CH-C patients in whom HBV integration was detected at the mild fibrosis stage may provide the evidence for direct association between HBV integration and accelerated hepatocarcinogenesis in this population. In addition, the association between genotype of integrated HBV and hepatocarcinogenesis in this population should also be investigated in the future, because the potential incidence of HCC reportedly differs according to HBV genotype in case of HBV-infected patients [30].

Acknowledgement

The authors thank Prof. Kunitada Shimotohno, Laboratory of Human Tumor Virus, Institute for Viral Research, Kyoto University, for his advice and comments.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jhep.2007.08.016.

References

[1] Beasley RP. Hepatitis B virus. The major etiology of hepatocellular carcinoma. Cancer 1988;61:1942–1956.

[2] Kiyosawa K, Sodeyama T, Tanaka E, Gibo Y, Yoshizawa K, Nakano Y, et al. Interrelationship of blood transfusion, non-A, non-B hepatitis and hepatocellular carcinoma: analysis by detection of antibody to hepatitis C virus. Hepatology 1990;12:671–675.

[3] Di Bisceglie AM, Goodman ZD, Ishak KG, Hoofnagle JH, Melpolder JJ, Alter HJ. Long-term clinical and histopathological follow-up of chronic posttransfusion hepatitis. Hepatology 1991;14:969–974.

[4] Brechot C, Jaffredo F, Lagorce D, Gerken G, Meyer zum Buschenfelde K, Papakonstontinou A, et al. Impact of HBV, HCV and GBV-C/HGV on hepatocellular carcinomas in Europe: results of a European concerted action. J Hepatol 1998;29:173–183.

[5] Cacciola I, Pollicino T, Squadrito G, Cerenzia G, Orlando ME, Raimondo G. Occult hepatitis B virus infection in patients with chronic hepatitis C liver disease. N Engl J Med 1999;341:22–26.

[6] Tamori A, Nishiguchi S, Kubo S, Koh N, Moriyama Y, Fujimoto S, et al. Possible contribution to hepatocarcinogenesis of X transcript of hepatitis B virus in Japanese patients with hepatitis C virus. Hepatology 1999;29:1429–1434.

[7] Pollicino T, Squadrito G, Cerenzia G, Cacciola I, Raffa G, Craxi A, et al. Hepatitis B virus maintains its pro-oncogenic properties in the case of occult HBV infection. Gastroenterology 2004;126:102–110.

[8] Urashima T, Saigo K, Kobayashi S, Imaseki H, Matsubara H, Koide Y, et al. Identification of hepatitis B virus integration in hepatitis C virus-infected hepatocellular carcinoma tissues. J Hepatol 1997;26:771–778.

[9] Kawai S, Yokosuka O, Imazeki F, Maru Y, Saisho H. State of HBV DNA in HBsAg-negative, anti-HCV-positive hepatocellular carcinoma: existence of HBV DNA possibly as nonintegrated form with analysis by Alu-HBV DNA PCR and conventional HBV PCR. J Med Virol 2001;64:410–418.

[10] Tamori A, Nishiguchi S, Kubo S, Enomoto M, Koh N, Takeda T, et al. Sequencing of human–viral DNA junctions in hepatocellular carcinoma from patients with HCV and occult HBV infection. J Med Virol 2003;69:475–481.

[11] Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. The French METAVIR Cooperative Study Group. Hepatology 1994;20:15–20.

[12] Minami M, Poussin K, Brechot C, Paterlini P. A novel PCR technique using Alu specific primers to identify unknown flanking sequences from the human genome. Genomics 1995;29:403–408.

[13] Murakami Y, Minami M, Daimon Y, Okanoue T. Hepatitis B virus DNA in liver, serum, and peripheral blood mononuclear cells after the clearance of serum hepatitis B virus surface antigen. J Med Virol 2004;72:203–214.

[14] Murakami Y, Saigo K, Takashima H, Minami M, Okanoue T, Brechot C, et al. Large scaled analysis of hepatitis B virus (HBV) DNA integration in HBV related hepatocellular carcinomas. Gut 2005;54:1162–1168.

[15] Koike K, Kobayashi M, Mizusawa H, Yoshida E, Yaginuma K, Taira M. Rearrangement of the surface antigen gene of hepatitis B virus integrated in the human hepatoma cell lines. Nucleic Acids Res 1983;25:5391–5402.

[16] Gunther S, Li BC, Miska S, Kruger DH, Meisel H, Will H. A novel method for efficient amplification of whole hepatitis B virus genomes permits rapid functional analysis and reveals deletion mutants in immunosuppressed patients. J Virol 1995;69:5437–5444.

[17] Okamoto H, Kobata S, Tokita H, Inoue T, Woodfield GD, Holland PV, et al. A second-generation method of genotyping hepatitis C virus by the polymerase chain reaction with sense and antisense primers deduced from the core gene. J Virol Methods 1996;57:31–45.

[18] Simmonds P, Alberti A, Alter HJ, Bonino F, Bradley DW, Brechot C, et al. A proposed system for the nomenclature of hepatitis C viral genotypes. Hepatology 1994;19:1321–1324.
[19] Toyoda H, Kumada T, Kiriyama S, Sone Y, Tanikawa M, Hisanaga Y, et al. Prevalence of low-level hepatitis B viremia in patients with HBV surface antigen-negative hepatocellular carcinoma with and without hepatitis C virus infection in Japan: analysis by COBAS TaqMan real-time PCR. Intervirology 2007;50:241–244.
[20] Tamori A, Nishiguchi S, Kubo S, Narimatsu T, Habu D, Takeda T, et al. HBV DNA integration and HBV-transcript expression in non-B, non-C hepatocellular carcinoma in Japan. J Med Virol 2003;71:492–498.
[21] Brechot C, Gozuacik D, Murakami Y, Paterlini-Brechot P. Molecular bases for the development of hepatitis B virus (HBV)-related hepatocellular carcinoma (HCC). Semin Cancer Biol 2000;10:211–231.
[22] Hino O, Shows TB, Rogler CE. Hepatitis B virus integration site in hepatocellular carcinoma at chromosome 17;18 translocation. Proc Natl Acad Sci USA 1986;83:8338–8342.
[23] Nakamura T, Tokino T, Nagaya T, Matsubara K. Microdeletion associated with the integration process of hepatitis B virus DNA. Nucleic Acids Res 1988;16:4865–4873.
[24] Dejean A, Bougueleret L, Grzeschik KH, Tiollais P. Hepatitis B virus DNA integration in a sequence homologous to v-erb-A and steroid receptor genes in a hepatocellular carcinoma. Nature 1986;322:70–72.
[25] Wang J, Chenivesse X, Henglein B, Brechot C. Hepatitis B virus integration in a cyclin A gene in a hepatocellular carcinoma. Nature 1990;343:555–557.
[26] Nagaya T, Nakamura T, Tokino T, Tsurimoto T, Imai M, Mayumi T, et al. The mode of hepatitis B virus DNA integration in chromosomes of human hepatocellular carcinoma. Genes Dev 1987;1:773–782.
[27] Takada S, Gotoh Y, Hayashi S, Yoshida M, Koike K. Structural rearrangement of integrated hepatitis B virus DNA as well as cellular flanking DNA is present in chronically infected hepatic tissues. J Virol 1990;64:822–828.
[28] Yoshida H, Shiratori Y, Moriyama M, Arakawa Y, Ide T, Sata M, et al. Interferon therapy reduces the risk for hepatocellular carcinoma: national surveillance program of cirrhotic and non-cirrhotic patients with chronic hepatitis C in Japan. Ann Int Med 1999;131:174–181.
[29] Tamori A, Nishiguchi S, Shiomi S, Hayashi T, Kobayashi S, Habu D, et al. Hepatitis B virus DNA integration in hepatocellular carcinoma after interferon-induced disappearance of hepatitis C virus. Am J Gastroenterol 2005;100:1748–1753.
[30] Chan HL, Hui AY, Wong ML, Tse AM, Hung LC, Wong VW, et al. Genotype C hepatitis B virus infection is associated with an increased risk of hepatocellular carcinoma. Gut 2004;53:1494–1498.
[31] Ono Y, Onda H, Sasada H, Igarashi K, Sugino Y, Nishioka K. The complete nucleotide sequences of the cloned hepatitis B virus DNA; subtype adr and adw. Nucleic Acid Res 1983;11:1747–1757.
[32] Kusano N, Okita K, Shirahashi H, Harada T, Shiraishi K, Oga A, et al. Chromosomal imbalances detected by comparative genomic hybridization are associated with outcome of patients with hepatocellular carcinoma. Cancer 2002;94:746–751.