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.

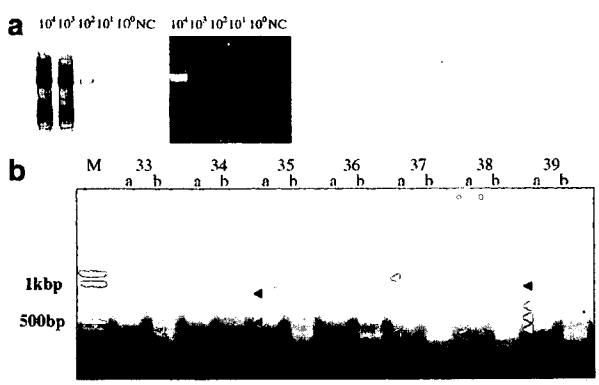


Fig. 1. The detection of HBV-X–host junction by Alu-PCR analysis. (a) The sensitivity study of Alu-PCR method. Serially diluted genomic DNA contained with HBV integrant was amplified by using HBV-X and Alu antisense primer pair. Left is Southern blot analysis from the gel electropheresis (right). (b) 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

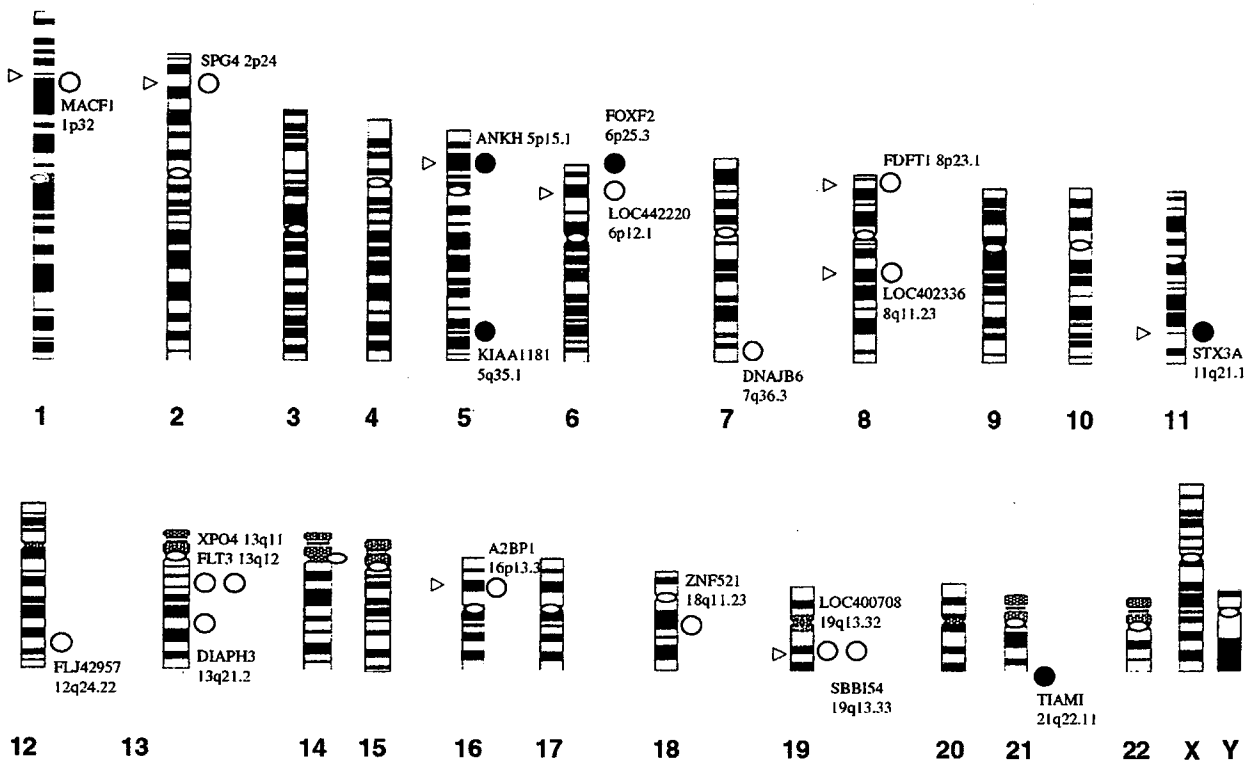


Fig. 2. Chromosomal distribution of HBV-X integration sites. Circles indicate viral integration sites, and the circle color denotes the sample. For example, the three white spots indicate three viral integration sites detected in the same specimen. Gene names and chromosomal localizations are also noted.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

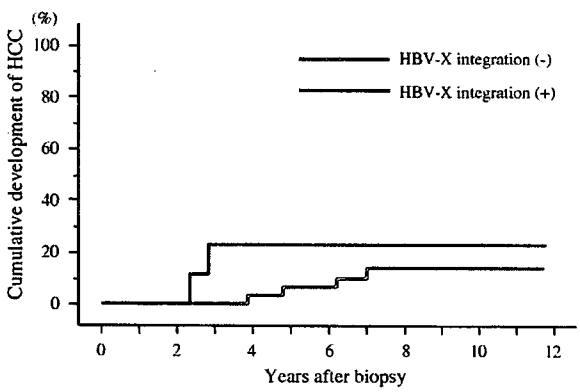


Fig. 3. Kaplan–Meier curves for the incidence of hepatocellular carcinoma (HCC). The blue and red lines represent the incidence of HCC in patients with and without HBV-X integration, respectively. No significant difference was observed between the two groups ($p = 0.8041$).

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.

RESEARCH ARTICLE

Integration of Hepatitis B Virus DNA Into the Myeloid/Lymphoid or Mixed-Lineage Leukemia (*MLL4*) Gene and Rearrangements of *MLL4* in Human Hepatocellular Carcinoma

Kenichi Saigo,[1,2] Kenichi Yoshida,[3]* Ryuji Ikeda,[3] Yoshiko Sakamoto,[3] Yoshiki Murakami,[4] Tetsuro Urashima,[1] Takehide Asano,[5] Takashi Kenmochi,[2] and Ituro Inoue[3,6]

[1]*Second Department of Surgery, School of Medicine, Chiba University, Chiba, Japan;* [2]*Clinical Research Center, Chiba-East National Hospital, Chiba, Japan;* [3]*Division of Genetic Diagnosis, Institute of Medical Science, University of Tokyo, Tokyo, Japan;* [4]*Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan;* [5]*Hepato-Pancreato-Biliary Surgery, School of Medicine, Teikyo University, Tokyo, Japan;* [6]*Division of Molecular Life Science, School of Medicine, Tokai University, Kanagawa, Japan*

Communicated by Dr. Haig H. Kazazian, Jr.

Integration of hepatitis B virus (HBV) DNA into host DNA is detected in about 90% of HBV-related hepatocellular carcinoma (HCC), but the preferential sites of the viral integration etiologically relevant to oncogenesis have been controversial. By using an adaptor-ligation/suppression-PCR, we identified four integrations into the myeloid/lymphoid or mixed-lineage leukemia 4 (*MLL4*) gene from 10 HCC patients with positive HBV surface antigen (HBsAg). Determination of the cellular-virus DNA junction demonstrated that various lengths of the virus were integrated within 300 bp of intron 3 flanked by the Alu element of *MLL4*. Chimeric hepatitis B virus X gene (HBx)/*MLL4* transcripts and the HBx fusion proteins were detected. DNA microarray revealed that HBx/MLL4 fusion proteins suppressed unique genes in HepG2 cells. Finally, chromosomal translocations of intron 3 of *MLL4* to the specific region of chromosome 17p11.2 in 22 out of 32 HCC patients were observed, showing that the intron 3 region of *MLL4* gene would be a target of translocation breakpoint. In conclusion, the present data suggest that the translocation breakpoint of *MLL4* gene is one of the preferential targets for HBV DNA integration into the *MLL4* gene and the HBV DNA integration may be involved in liver oncogenesis. Hum Mutat 0,1–6, 2008. © 2008 Wiley-Liss, Inc.

KEY WORDS: hepatocellular carcinoma; DNA integration; hepatitis B virus; HBx; *MLL4*

INTRODUCTION

Chronic human hepatitis B virus (HBV) infection causes mild to severe liver diseases, such as chronic hepatitis, liver cirrhosis, and hepatocellular carcinoma (HCC) [Block et al., 2003]. Nearly 25% of patients with chronic HBV infections terminate in untreatable liver cancer. HBV DNA frequently integrates into the human host genome, whereby insertional mutagenesis plays a crucial role in oncogenesis [Brechot et al., 2000; Gozuacik et al., 2001]. Although integration of HBV DNA is thought to be involved in oncogenesis of human hepatocytes, preferential HBV DNA integration sites targeting cellular genes were not identified until recently. Two groups have reported that HBV DNA is preferentially integrated into the human telomerase reverse transcriptase (*TERT*) gene (MIM# 187270) in HCC [Ferber et al., 2003; Paterlini-Brechot et al., 2003].

In this study, we investigated the integrated HBV DNA and flanking cellular DNA sequences. In four cases, integrations of HBV DNA into intron 3 of the myeloid/lymphoid or mixed-lineage leukemia 4 (*MLL4*) gene (MIM# 606834) were demonstrated, indicating that *MLL4* serves as a cellular target for HBV in liver oncogenesis. The *MLL4* gene is a human member of the *MLL* gene family locating on chromosome 19q13.1 [FitzGerald and Diaz, 1999], where a frequent rearrangement or amplification has been reported in solid tumors [Mitelman et al., 1997; Curtis et al., 1998]. Subsequently, we detected chromosomal translocation between intron 3 of *MLL4* and a specific region of chromosome17p11.2 in 22 HCC samples. These results indicate that intron 3 of the *MLL4* gene is one of the sites of translocation break point, which serves as a preferential target for HBV DNA integration, and may be implicated in the etiology of liver oncogenesis.

The Supplementary Material referred to in this article can be accessed at http://www.interscience.wiley.com/jpages/1059-7794/suppmat.

Received 18 September 2007; accepted revised manuscript 16 November 2007.

*Correspondence to: Kenichi Yoshida, Department of Life Sciences, Meiji University School of Agriculture, 1-1-1 Higashimita, Tama-ku, Kawasaki, Kanagawa 214-8571, Japan (present address). E-mail: yoshida@isc.meiji.ac.jp

Grant sponsor: CREST of Japan Science and Technology (II).
DOI 10.1002/humu.20701
Published online in Wiley InterScience (www.interscience.wiley.com).

PATIENTS AND METHODS

Patients

We studied 32 Japanese patients with HCC who had undergone hepatic resection without preoperative therapies at The Second Department of Surgery, Chiba University Hospital between 1987 and 2003. Serological tests for HBV were done by EIA kit (Dainabot, Tokyo, Japan) for HBV surface antigen (HBsAg), and RIA kits (Dainabot) for anti-HBs and anti-HBc antibodies. Anti-HCV antibody was tested by a recombinant immunoblot assay (Ortho Diagnostic, Westwood, MA). The study protocol conformed to the ethical guidelines of the Declaration of Helsinki (1975) and was approved by the Institutional Review Board (IRB) of Chiba University, School of Medicine. All patients gave written informed consent.

PCR and Southern Blot

HBV/cellular DNA junctions in the tumor tissues were analyzed by an adaptor-ligation/suppression-PCR [Siebert et al., 1995], according to Genomewalker Kits (Clonetech, Mountain View, CA) (Supplementary Fig. S1, available online at http://www.interscience.wiley.com/jpages/1059-7794/suppmat). Primer sequences used for PCR detection of HBV/*MLL4* and *MLL4*/HBV junctions and chromosome 19/chromosome 17 boundaries are listed in Supplementary Table S1.

Hind III-digested DNA (10 μg) were electrophoresed on 1.0% agarose gel and blotted onto nylon membrane (Hybond N+; GE Healthcare, Buckinghamshire, UK). The membrane was first hybridized with ^{32}P-labeled hepatitis B virus X gene (HBx) probe and the blot was autoradiographed. After dehybridization of the same membrane, a rehybridization was carried out with ^{32}P-labeled *MLL4* probe (the PCR products spanning exon 4 and exon 5) and autoradiographed.

RT-PCR

Total cellular RNA was extracted using Trizol (Invitrogen, Carlsbad, CA). An RT-PCR was performed with SuperScript One-Step RT-PCR system (Invitrogen) with gene-specific primers on exon 5 and exon 6 of *MLL4*. MD26c primer was used as the common sense primer. The PCR products were subjected to sequencing analyses.

Immunoprecipitation and Western Blot

Tumor tissues were lysed in a buffer containing 0.1% SDS, 0.5% deoxycholate, 1% NP-40, 150 mM NaCl, 50 mM Tris-HCl (pH 8.0), protease inhibitors (complete protease inhibitor tablets; Roche, Basel, Switzerland), and centrifuged. The supernatant was incubated with anti-HBx monoclonal antibody, generously provided by Dr. Yosef Shaul (Weizman Institute of Science), and immunoprecipitation/Western blot was performed with a standard protocol. Anti-Flag antibody was purchased from Sigma-Aldrich (St. Louis, MO).

HBx/*MLL4* Expression Plasmid, Transfection, and DNA Microarray

The HBx expression vector, pECE-X, was a gift of Dr. Jinghsiung James Ou (University of Southern California). Human *MLL4* partial cDNA clone KIAA0304 (accession number AB002302.2) was obtained from Kazusa DNA Research Institute (Chiba, Japan). We deleted intron 7 sequence from KIAA0304 and constructed N-terminally Flag-tagged HBx/*MLL4* chimeric sequence in pcDNA3 (Invitrogen). Human hepatoma cell line HepG2 (RCB1648; RIKEN Cell Bank, Tsukuba, Japan) were transfected using Lipofectamine (Invitrogen). After 48 hr of transfection, total RNA was recovered and the microarray analysis including 12,814 unique clones from Incyte UniGene 1 was performed according to the manufacturer's instructions (Agilent Technologies, Santa Clara, CA).

RESULTS

Detection and Sequence Analysis of HBV/Cellular DNA Junctions

A total of 10 tumor specimens from HCC patients with positive HBsAg were examined for HBV DNA (accession number AB033550.1) integrations into cellular genome. The clinical backgrounds of the patients are summarized (Supplementary Table S2). We could detect four integrations into the *MLL4* gene on chromosome 19q13.1 and one into the *TERT* gene (Table 1). Integration sites of *MLL4* (accession number AD000671.1) from the four patients were all in intron 3 of the *MLL4* gene (Fig. 1A; Table 1) within or flanked with the Alu repeat sequence (Fig. 1B). As shown in Fig. 1C, full-length viral integration could be expected in HCC131 (g.17752_17753insAB033550.1:g.1827_1826), while truncated virus integrations were detected in the other three tissues, HCC143 (g.17817_17818insAB033550.1:g.2974_1794), HCC146 (g.17514_17515insAB033550.1:g.?_1807), and HCC002 (g.17542_17543insAB033550.1:g.1051_1762). In all four patients, the viral junctions described above were located in the vicinity of DR1, suggesting that the DR1 region is the preferred viral junction for HBV DNA integration.

On Southern blot analysis, clonally integrated HBV DNA sequences were detected in the tumor tissue of HCC131 and a positive control. We encountered the limitations with the heterogeneity of other samples. Using Southern blot hybridization

TABLE 1. Detection of HBV Integration and the Translocation of *MLL4* in HCC

Case no.	Chromosome	Accession no.	Gene	t(17;19)(p11;q13.1)
HCC131	19q13.1	AD000671.1	*MLL4*	Positive
HCC146	19q13.1	AD000671.1	*MLL4*	Positive
	7p14_15	AC005090.2		
HCC002	19q13.1	AD000671.1	*MLL4*	Positive
HCC003	5p13	AY007685.1	*TERT*	Positive
HCC9907	9q13_21.3	AL133578.1		Negative
HCC155				Positive
H20				Positive
H54	18p11.3	AP000845.4	*NMP p84*[a]	Positive
H120				Positive
HCC143	19q13.1	AD000671.1	*MLL4*	Positive
H49				Positive
H62				Negative
H70				Positive
H72				Positive
H78				Positive
H89				Positive
H76				Positive
H57				Negative
H71				Positive
H85				Positive
H86				Negative
H87				Positive
HCC128				Positive
HCC147				Positive
HCC127				Positive
HCC001				Positive
H148				Negative
H149				Negative
H150				Negative
HCC9833				Negative
HCC9901				Negative
HCC9906				Negative

[a]This integration was already reported. Chromosome locations, GenBank accession numbers, and gene names are indicated for eight viral/cellular junctions from seven HCC samples. Detection of t(17;19)(p11;q13.1) was indicated as positive.

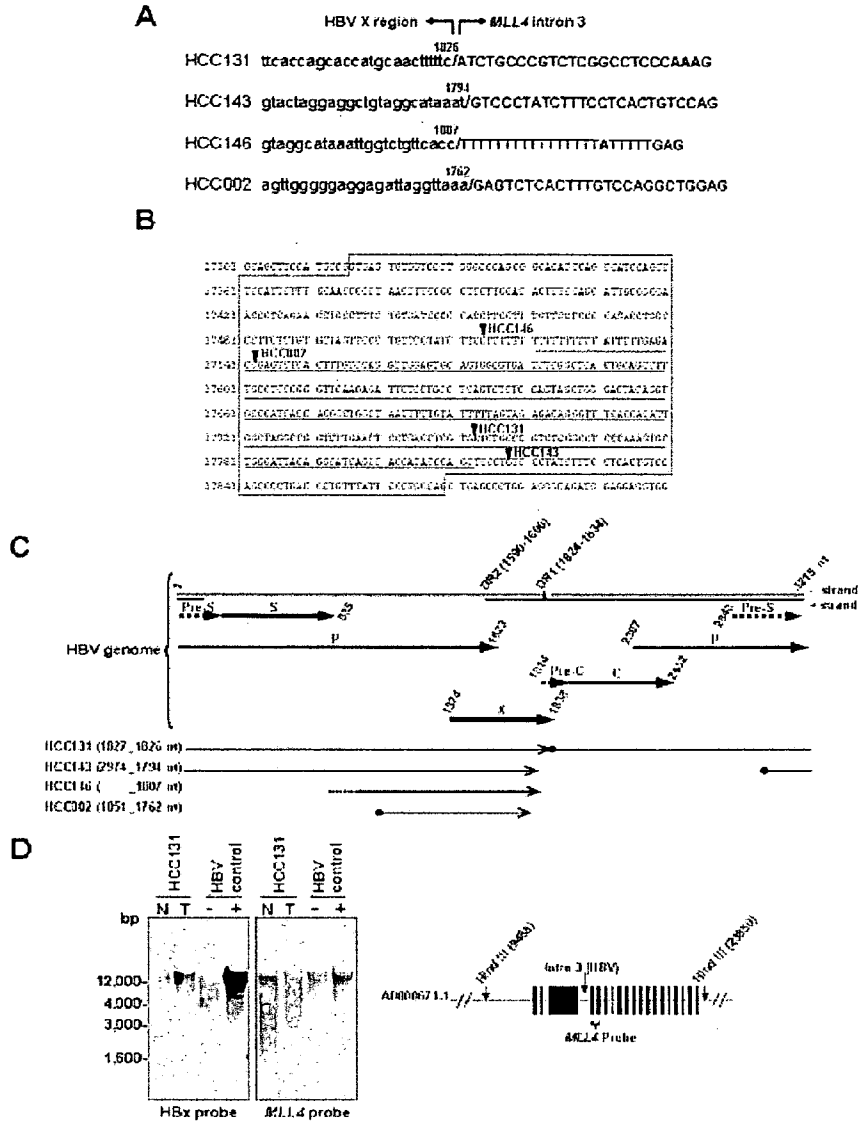
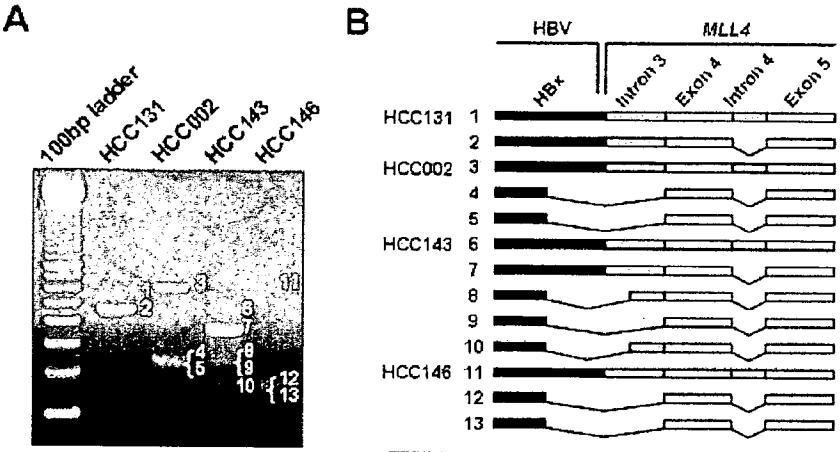


FIGURE 1. **Surrounding sequences of HBV integration sites. A:** Sequences of HBV/cellular DNA junctions in the *MLL4* gene in HCC131, HCC143, HCC146, and HCC002. In each sample, the small letters on the left show sequences of the integrated HBV DNA and capital letters on the right show the flanking *MLL4* gene sequences. Numbers above indicate HBV nucleotides at the HBV/cellular DNA junction (Accession number AB033550.1). **B:** Sequences around intron 3 of the *MLL4* gene and four HBV DNA integration sites are shown. The left side indicates nucleotide positions of *MLL4* gene (accession number AD000671.1). Intron 3 of the *MLL4* gene is indicated by the box (17316_17869 nt). The Alu repeat is shown by underline (17521_17812 nt). **C:** Schematic representation of gene organization of HBV genome and four integrated HBV genomes (HCC002, HCC131, HCC143, and HCC146). Open reading frames and their directions of transcription are represented by an arrow. The numbers above the arrow indicate location of each open reading frame (Accession number AB033550.1). DR1 and DR2 are the 11 basepair direct repeats. • and > indicate the 5′ and 3′ end of integrated HBV DNA sequences (we could not obtain the 5′ end for HCC146). The lengths of the solid lines represent the size and location of the integrated HBV. **D:** Southern blot analysis, using the HBx region as probe (left panel) and the *MLL4* probe (right panel). Hind III-digested DNAs from nontumor tissue of HCC 131 (lane 1), tumor tissue of HCC131 (lane 2), colon cancer tissue as negative control (lane 3), and the HBV integrated HCC tissue as positive control (lane 4). Schematic representation for *MLL4* gene and Hind III site are shown. HBV integration site (intron 3) and *MLL4* probe are indicated. Closed boxes indicate exons of *MLL4* gene.

with an *MLL4* probe of the same membrane, hybridization signals were also detected in the tumor tissue of the patient (Fig. 1D).

HBV Integration Into the *MLL4* Gene Drives Expression of Chimeric Transcripts

In the four HBV/*MLL4* samples, all the integrated viral genome contained HBx promoter and HBx ORF (1374_1838 nucleotides of AB033550.1) except the C-terminus (Fig. 1C). RT-PCR study for detecting fusion transcripts was carried out with HBx primer and reverse primers on various exons of *MLL4* in the four HCC tissues showing various species in each sample (Fig. 2A). In all HCC tissues, in-frame chimeric transcripts that contained exon 4 and exon 5 of *MLL4* were detected (Fig. 2B). In HCC131, two transcripts were observed; one transcript, a major form, showed in-frame fusion containing intron 3 and the other transcript retained intron 4 that led to the creation of the termination codon in exon 6. In HCC002, three transcripts were observed; one transcript



	HBx (AB033550.1)	MLL4 (AN000671.1)	Frame (stop codon)
1	1374_1827	17754_(intron 3)	out (exon 6)
2	1374_1827	17754_(intron 3)	in
3	1374_1762	17543_(intron 3)	out (intron 3)
4	1374_1716	17870_(exon 4)	out (exon 4)
5	1374_1634	17870_(exon 4)	in
6	1374_1794	17818_(intron 3)	out (intron 3)
7	1374_1794	17818_(intron 3)	out (intron 3)
8	1374_1716	17843_(intron 3)	out (exon 4)
9	1374_1716	17870_(exon 4)	out (exon 4)
10	1374_1630	17839_(intron 3)	in
11	1374_1807	17517_(intron 3)	out (intron 3)
12	1374_1646	17870_(exon 4)	in
13	1374_1634	17870_(exon 4)	in

FIGURE 2. RT-PCR analysis of HBx/*MLL4* fusion transcripts. **A:** Various transcripts were observed for each of the HCC tissues by RT-PCR. **B:** Schematic representation of the fusion transcripts from four HCC tissues (HCC131, HCC002, HCC143, and HCC146), and adjacent sequences between HBx (3′ end) and *MLL4* (5′ end) are summarized. HBx cDNA (black boxes) and *MLL4* gene (exon 4 and 5 as white boxes and intron 3 and 4 as gray boxes) are shown. Spliced out sequences are indicated by bars. Location of 5′ end of *MLL4* in intron 3 or exon 4 is also shown. Reading frame based on HBx cDNA followed by *MLL4* is indicated for individual chimeric transcripts. Location of the aberrant stop codon is also shown except for in-frame transcripts. See the Supplementary Appendix for more information.

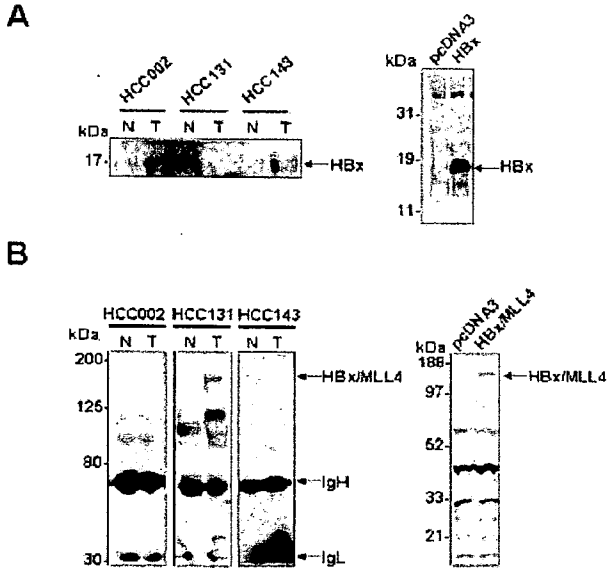


FIGURE 3. Immunodetection of HBx fusion proteins in HCC samples. **A:** Western blot analysis of tumor tissues (T) from HCC002, HCC131, and HCC143 and the adjacent nontumor tissues (N) by using the monoclonal anti-HBx antibody. Recombinant full-length HBx protein expressed in HepG2 cells are also shown. **B:** Immunoprecipitation followed by immunodetection with HBx antibody detected HBx/MLL4 putative fusion protein specifically in HCC131 T (left panel). Western blot using Flag antibody detected an approximately 170-kDa HBx/MLL4 fusion protein transiently expressed in HepG2 cells (right panel). See the Supplementary Appendix for more information.

retained intron 3 causing premature termination and the other two transcripts spliced out intron 3 using distinct 5′ splice sites, resulting in one transcript (nucleotide position 261) showing the in-frame transcript and the other (nucleotide position 343) the premature termination. Similar patterns were observed for HCC146. In HCC143, five species were observed, and the splicing junction of the in-frame transcript was CC-GT, and does not conform to the GT-AG rule.

HBx-MLL4 Fusion Proteins Expressed in HCC

To confirm that the fusion transcripts were translated, the expression of HBx-related proteins in the tumor and adjacent liver tissues were tested by immunodetection with an antibody against HBx protein. Western blot analysis showed that an approximately 17-kDa protein, which represents a short HBx fusion protein compared to recombinant full-length HBx protein expressed in HepG2 cells, is selectively expressed in the HCC002 and HCC143 tumor tissues (Fig. 3A). Immunoprecipitation followed by Western blot analysis detected an approximately 170-kDa protein in HCC131 (Fig. 3B). We constructed an expression vector that can express fusion protein consisting of N-terminally Flag-tagged HBx ORF (amino acids 1_154) and *MLL4* coding region beginning from exon 4 (corresponding to amino acids 820_2715, accession number NM_014727.1), and transiently expressed into HepG2 cells. Western blot using Flag antibody clearly detected an approximately 170-kDa protein (Fig. 3B, right panel). MLL is known to be cleaved at a conserved site and this cleavage generates N- and C-terminal fragments [Hsieh et al., 2003]. MLL4 also possesses a conserved site D/GVDD (amino acids

TABLE 2. cDNA Microarray Results Showing Upregulation and Downregulation by HBx, HBx/*MLL4* Fusion, and Truncated HBx (1_87aa) Proteins*

	Gene description	Category	HBx		HBx/MLL4		HBx 1_87aa	
			Mean	SD	Mean	SD	Mean	SD
Upregulated gene name								
OR11A1	Olfactory receptor, family 11, subfamily A, member 1	G protein-coupled receptor	3.99	0.02	—		—	
OPN4	Opsin 4 (melanopsin)	G protein-coupled receptor	3.52	0.1	—		—	
UPB1	Ureidopropionase, beta	Hydrolase	3.09	0.52	—		3.89	2.2
HIST1H4L	H4 histone family, member K	Nucleosome structure	3.08	0.3	—		2.72	0.3
HIST1H4I	H4 histone family, member M	Nucleosome structure	2.8	0.02	—		—	
ELL3	Elongation factor RNA polymerase II-like 3	Transcriptional regulation	2.59	0.06	—		—	
BAI1	Brain-specific angiogenesis inhibitor 1	Cell adhesion	2.57	0.62	—		—	
CEP290	Centrosomal protein 290kDa	Centrosomal protein	2.51	0.66	—		—	
HIST1H4B	H4 histone family, member I	Nucleosome structure	2.38	0.01	—		—	
CDC2L1	Cell division cycle 2-like 1	Cell cycle	2.36	0.4	—		2.36	0.04
OR2C1	Olfactory receptor, family 2, subfamily C, member 1	G protein-coupled receptor	2.15	0.15	—		—	
DNCL2B	Dynein, light chain 2B	Motor protein	2.11	0.04	—		—	
ZNF354B	Zinc finger protein 354B	Transcriptional regulation	2.09	0.06	—		—	
MLL4	Mixed-lineage leukemia 4	Transcriptional regulation	—		31	3.25	—	
Downregulated								
AVIL	Advillin	Actin-binding protein	3	0.42	5.2	1.46	3.51	0.68
ENO2	Enolase 2	Hydratase	2.19	0.21	—		—	
KERA	Keratocan	Extracellular matrix	—		4.91	1.37	—	
UBXD1	UBX domain containing 1	Unknown	—		4.76	0.34	—	
PIAS3	Protein inhibitor of activated STAT3	Signal transduction	—		4.68	0.45	—	
MYBPC2	Fast-type myosin binding protein C	Unknown	—		4	0.48	2.48	0.12
PITPNM	Phosphatidylinositol-transfer protein membrane-associated	Cytokinesis	—		3.58	0.83	—	
EHD2	EH-domain containing 2	Endocytosis	—		3.4	0.43	—	
GJB1	Connexin 32	Gap junction	—		3.01	0.62	—	
WASL	Wiskott-Aldrich syndrome-like	Actin polymerization	—		2.48	0.22	—	
TNRC6C	Trinucleotide repeat containing 6c	Unknown	—		2.31	0.28	—	
TBC1D10B	TBC1 domain family, member 10B	Unknown	—		2.19	0.08	—	

*The experiments were performed twice, and the mean and standard deviation (SD) values were determined for each gene.

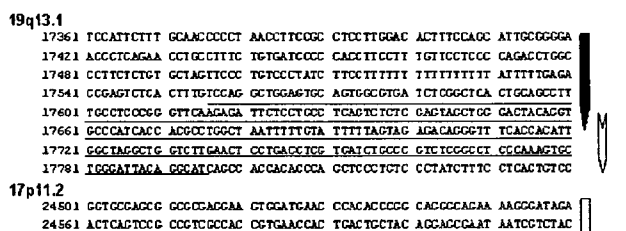


FIGURE 4. **Reciprocal translocation found in intron 3 of the *MLL4* locus. Chromosomal rearrangement between chromosome 19q13.1 (accession number AD000671.1) and chromosome 17p11.2 (accession number AC087294.18) are shown. The sequences of Alu elements are underlined. The downward pointing arrows on the respective chromosomes indicate the newly synthesized chromosomes (black-to-black and white-to-white) after the recombination events occurred via Alu elements.**

2062_2066), indicating a 170-kDa protein could be a posttranslationally modified product.

Functional Elucidation of HBx/MLL4 Fusion Protein by DNA Microarray

To provide mechanistic insights into molecular etiology such as altered target gene expression regulated by HBx/MLL4 fusion protein, we transiently overexpressed full-length HBx and HBx/MLL4 fusion proteins in HepG2 cells. We employed cDNA microarray technology and identified 13 genes that were upregulated and two genes that were downregulated by HBx protein (Table 2). In contrast, no gene (except for *MLL4* itself) was upregulated and 11 genes were downregulated by HBx/MLL4 fusion protein (Table 2). Uniquely, only one gene, Advillin (*AVIL*) was identified as a common target between HBx and HBx/MLL4 fusion proteins. We checked whether C-terminally truncated HBx protein (amino acids 1_87) could regulate the expression of genes identified by above experiments, because HBx protein in HCC002 and HCC143 only had N-terminal 87 and 86 amino acid residues. Three genes were upregulated and two genes, including *AVIL*, were downregulated by truncated HBx protein (Table 2). Taken together, these data predict that HBx/MLL4 fusion protein would suppress the expression pattern of specific genes.

Alu-Mediated Chromosomal Translocation of *MLL4* to 17p11.2 in HCC

We extended the search for HBV DNA integration into intron 3 of the *MLL4* gene in other HCC samples positive for anti-HBc antibody (Supplementary Table S2). The sequencing analyses failed to detect HBV/*MLL4* DNA sequences, instead demonstrated chimeric sequences between the *MLL4* gene and a particular region of chromosome 17p11.2 (Fig. 4). We detected 22 translocations from 32 HCC samples (Table 1). The sequencing analyses of the translocation products revealed an about 240-bp region at the junction that is highly shared by two chromosomes (approximately 85%) containing Alu elements, suggesting that Alu-mediated homologous recombination facilitated translocation (Fig. 4).

DISCUSSION

The classical mechanism by which tumor-associated viruses contribute to oncogenesis is activation of cellular genes with oncogenic potential through viral genome integration into the cellular genome. HBV genome integration into *SERCA1* (sarco/endoplasmic reticulum calcium ATPase) have been demonstrated [Chami et al., 2000]. The resultant chimeric HBx/SERCA1 protein proposed to be implicated in oncogenesis via an apoptotic

mechanism [Chami et al., 2000]. Reports from two groups, including our observation, demonstrate that the promoter region of the *TERT* gene is targeted by HBV in several HCC tissues [Ferber et al., 2003; Paterlini-Brechot et al., 2003]. Therefore, the *TERT* gene most likely serves as a nonrandom integration site of the viral genome in a subset of HBV-positive HCCs, and the oncogenic HBV DNA integrations may possess the preferential sites. In this study, we further demonstrated four cases of integrations into the *MLL4* gene in HBsAg-positive HCC samples. Sequencing analyses revealed that all of the host sites were within 300 bp of intron 3, flanked with the Alu element of the *MLL4* gene. Recently, HBV DNA integration into *MLL4* gene in three Japanese HCC patients, two cases into exon 3 integration and one into intron 3, were reported [Tamori et al., 2005]. These results support the hypothesis that the oncogenic viral integrations into hepatocytes are not entirely random.

HBV integration into intron 3 of *MLL4* resulted in several fusion transcripts between HBx and *MLL4* that could be directly implicated in liver oncogenesis, albeit the C-terminally truncated HBx protein, as observed in HCC002 and HCC143, might be more closely related to oncogenesis. Our cDNA microarray experiments indicate that HBx/MLL4 fusion protein suppressed the unique genes. It might be speculated that the fusion gene product lacking an AT hook, which is encoded in exons 1–3 of *MLL4*, is directly related to oncogenesis. Further investigation of HBx/MLL4-dependent or N-terminal MLL4-dependent transcriptional regulation may provide a novel insight into the elucidation of etiology of hepatic oncogenesis.

The *MLL4* gene, originally reported as a second human homolog of the *MLL* gene, is mapped to chromosome 19q13.1 [FitzGerald and Diaz, 1999], where gene amplification was reported in HBV-related HCCs [Marchio et al., 1997; Huntsman et al., 1999] and frequent genome rearrangements in solid tumor were reported [Curtis et al., 1998]. We detected the chromosomal translocation of the *MLL4* locus to chromosome 17 in 22 tumors out of 32 samples. The chromosomal rearrangement occurred between intron 3 of the *MLL4* gene of chromosome 19q13.1 and chromosome 17p11.2. The two chromosomal regions share nearly identical Alu elements, indicating that Alu-mediated recombination most likely explains the genome rearrangement. HBV infection and subsequent hepatitis induced DNA damage such as double-strand breaks [Dandri et al., 2002; Bill and Summers, 2004]; therefore, the genome repair mechanism is essential for maintaining the genome integrity and cellular viability.

In conclusion, we detected the translocation breakpoint point in the intron 3 of *MLL4* gene that provides one of the preferential targets for HBV integrations. Indeed we also found recurrent integrations of HBV DNA into intron 3 of *MLL4* gene in four HCC cases, and chimeric HBx/*MLL4* transcripts and HBx/MLL4 proteins, suggesting that the insertional mutagenesis could be functionally relevant to liver oncogenesis.

ACKNOWLEDGMENTS

We are grateful to Dr. Murakami (National Cancer Center Research Institute, Japan) and Dr. Miyamura, Dr. Suzuki, and Dr. Shoji (National Institute of Infectious Disease, Japan) for their technical support and helpful comments. This work was partly supported by grant from the CREST of Japan Science and Technology (II) to I.I.

REFERENCES

Bill CA, Summers J. 2004. Genomic DNA double-strand breaks are targets for hepadnaviral DNA integration. Proc Natl Acad Sci USA 101:11135–11140.

Block TM, Mehta AS, Fimmel CJ, Jordan R. 2003. Molecular viral oncology of hepatocellular carcinoma. Oncogene 22:5093–5107.

Brechot C, Gozuacik D, Murakami Y, Paterlini-Brechot P. 2000. Molecular bases for the development of hepatitis B virus (HBV)-related hepatocellular carcinoma (HCC). Semin Cancer Biol 10:211–231.

Chami M, Gozuacik D, Saigo K, Capiod T, Falson P, Lecoeur H, Urashima T, Beckmann J, Gougeon ML, Claret M, le Maire M, Brechot C, Paterlini-Brechot P. 2000. Hepatitis B virus-related insertional mutagenesis implicates SERCA1 gene in the control of apoptosis. Oncogene 19:2877–2886.

Curtis LJ, Li Y, Gerbault-Seureau M, Kuick R, Dutrillaux AM, Goubin G, Fawcett J, Cram S, Dutrillaux B, Hanash S, Muleris M. 1998. Amplification of DNA sequences from chromosome 19q13.1 in human pancreatic cell lines. Genomics 53:42–55.

Dandri M, Burda MR, Burkle A, Zuckerman DW, Will H, Rogler CE, Greten H, Petersen J. 2002. Increase in de novo HBV DNA integrations in response to oxidative DNA damage or inhibition of poly(ADP-ribosyl)ation. Hepatology 35:217–223.

Ferber MJ, Montoya DP, Yu C, Aderca I, McGee A, Thorland EC, Nagorney DM, Gostout BS, Burgart LJ, Boix L, Bruix J, McMahon BJ, Cheung TH, Chung TK, Wong YF, Smith DI, Roberts LR. 2003. Integrations of the hepatitis B virus (HBV) and human papillomavirus (HPV) into the human telomerase reverse transcriptase (hTERT) gene in liver and cervical cancers. Oncogene 22:3813–3820.

FitzGerald KT, Diaz MO. 1999. MLL2: a new mammalian member of the trx/MLL family of genes. Genomics 59:187–192.

Gozuacik D, Murakami Y, Saigo K, Chami M, Mugnier C, Lagorce D, Okanoue T, Urashima T, Brechot C, Paterlini-Brechot P. 2001. Identification of human cancer-related genes by naturally occurring hepatitis B virus DNA tagging. Oncogene 20:6233–6240.

Hsieh JJ, Ernst P, Erdjument-Bromage H, Tempst P, Korsmeyer SJ. 2003. Proteolytic cleavage of MLL generates a complex of N- and C-terminal fragments that confers protein stability and subnuclear localization. Mol Cell Biol 23:186–194.

Huntsman DG, Chin SF, Muleris M, Batley SJ, Collins VP, Wiedemann LM, Aparicio S, Caldas C. 1999. MLL2, the second human homolog of the Drosophila trithorax gene, maps to 19q13.1 and is amplified in solid tumor cell lines. Oncogene 18:7975–7984.

Marchio A, Meddeb M, Pineau P, Danglot G, Tiollais P, Bernheim A, Dejean A. 1997. Recurrent chromosomal abnormalities in hepatocellular carcinoma detected by comparative genomic hybridization. Genes Chromosomes Cancer 18:59–65.

Mitelman F, Mertens F, Johansson B. 1997. A breakpoint map of recurrent chromosomal rearrangements in human neoplasia. Nat Genet 15:417–474.

Paterlini-Brechot P, Saigo K, Murakami Y, Chami M, Gozuacik D, Mugnier C, Lagorce D, Brechot C. 2003. Hepatitis B virus-related insertional mutagenesis occurs frequently in human liver cancers and recurrently targets human telomerase gene. Oncogene 22:3911–3916.

Siebert PD, Chenchik A, Kellogg DE, Lukyanov KA, Lukyanov SA. 1995. An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Res 23:1087–1088.

Tamori A, Yamanishi Y, Kawashima S, Enomoto M, Tanaka H, Kudo S, Shiomi S, Nishiguchi S. 2005. Alteration of gene expression in human hepatocellular carcinoma with integrated hepatitis B virus DNA. Clin Cancer Res 11:5821–5826.

FEBS Letters 582 (2008) 351–358

Expressed sequence tags from cynomolgus monkey (*Macaca fascicularis*) liver: A systematic identification of drug-metabolizing enzymes

Yasuhiro Uno[a,*], Yutaka Suzuki[b,*], Hiroyuki Wakaguri[b], Yoshiko Sakamoto[a], Hitomi Sano[a], Naoki Osada[c], Katsuyuki Hashimoto[c], Sumio Sugano[b], Ituro Inoue[a,d]

[a] *Division of Genetic Diagnosis, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan*
[b] *Department of Medical Genome Sciences, Graduate School of Frontier Sciences, University of Tokyo, 4-6-1 Shirokanedai, Minatoku, Tokyo 108-8639, Japan*
[c] *Department of Biomedical Resources, National Institute of Biomedical Innovation, Ibaraki, Osaka, Japan*
[d] *Division of Molecular Life Science, School of Medicine, Tokai University, Shimokasuya 134, Isehara, Kanagawa 259-1193, Japan*

Received 1 October 2007; revised 14 December 2007; accepted 18 December 2007

Available online 31 December 2007

Edited by Takashi Gojobori

Abstract The liver, a major organ for drug metabolism, is physiologically similar between monkeys and humans. However, the paucity of identified genes has hampered a deep understanding of drug metabolism in monkeys. To provide such a genetic resource, 28655 expressed sequence tags (ESTs) were generated from a cynomolgus monkey liver full-length enriched cDNA library, which contained 23 unique ESTs homologous to human drug-metabolizing enzymes. Our comparative genomics approach identified nine lineage-specific candidate ESTs, including three drug-metabolizing enzymes, which could be important for understanding the physiological differences between monkeys and humans.

Keywords: Cynomolgus monkey; Drug metabolism; Drug-metabolizing enzyme; Expressed sequence tags; Lineage-specific gene; Liver

1. Introduction

Cynomolgus monkeys have been used as an animal model for the investigation of human physiology and disease because of their close genetic and physiological similarities to humans. Application of this animal model includes predicting metabolic fate of newly developed drugs due to pharmacokinetics similar to humans. However, we now know that differences in metabolic properties are occasionally seen for some drugs between monkeys and humans [1–7] possibly due to differences in genetic components essential for drug metabolism between the two lineages, such as lineage-specific genes and alternatively spliced transcripts. However, limited numbers of lineage-specific genes identified in monkeys have hampered complete knowledge of lineage differences in drug metabolism.

An expressed sequence tag (EST)-sequencing approach has been a rapid and efficient way to identify novel cDNAs that provide a basis to investigate genetic components essential to various physiological functions. In non-human primates, efforts have been made for the comprehensive identification of ESTs in chimpanzees [8], rhesus monkeys [9,10], and cynomolgus monkeys [11–13]. However, liver tissue has not been extensively sequenced for ESTs, thus only limited genetic information is available on liver physiological function such as drug metabolism. With the completion of a draft of the rhesus monkey genome sequence [14], EST analysis of macaques should be more feasible and accurate.

To provide a monkey genetic resource, 28655 ESTs from cynomolgus monkey liver were generated. These macaque ESTs analyzed against the rhesus genome identified 1064 unique ESTs, most of which (77.0%) matched the human RefSeq database. cDNAs highly homologous to human drug-metabolizing genes were identified, including those of cytochrome P450 (CYP), UDP-glucuronosyltransferase (UGT), glutathione *S*-transferase (GST), sulfotransferase (SULT), and flavin-containing monooxygenase (FMO). Moreover, our method to select lineage-specific ESTs successfully identified novel transcripts related to drug metabolism. This genetic information should help in discerning various physiological characteristics, including drug metabolism in monkeys.

2. Materials and methods

2.1. cDNA library construction and EST sequencing

Liver samples were collected from three adult cynomolgus monkeys (two males and one female) and used to generate a full-length enriched cDNA library using the pME18S-FL3 vector by the oligo-capping method as previously described [15]. Purified DNA was sequenced using the ABI PRISM® BigDye™ Terminator Cycle Sequencing Ready Reaction Kit, Version 2.0 (Applied Biosystems, Foster City, CA), followed by electrophoresis with ABI-3700 DNA Analyzer (Applied Biosystems) according to the manufacturer's instructions. Primers (5′-GGATGTTGCCTTTACTTCTA-3′ and 5′-TTTTTTTTTTTT-TTTTTV-3′) were used for single-pass sequencing of 5′ and 3′-ends for each cDNA, respectively.

*Corresponding authors.
E-mail addresses: unoxx001@pharm.hokudai.ac.jp (Y. Uno), ysuzuki@k.u-tokyo.ac.jp (Y. Suzuki).

Abbreviations: CYP, cytochrome P450; EST, expressed sequence tag; FMO, flavin-containing monooxygenase; GST, glutathione *S*-transferase; ORF, open reading frame; SULT, sulfotransferase; UGT, UDP-glucuronosyltransferase

0014-5793/$32.00
doi:10.1016/j.febslet.2007.12.031

2.2. Sequence data analysis

Vector sequence was trimmed and sequence quality was inspected using Phred (University of Washington). Only EST sequences longer than 200 bases were used. Generated EST sequences were first computationally mapped to the *Macaca mulatta* genomic sequence (rheMac2, UCSC Genome Browser). Computational mapping was carried out as previously described by consequential use of sequence alignment programs, BLAT and SIM4 [16]. Only ESTs over the entire sequence length that mapped perfectly at unique positions on the macaque genome were regarded as "mapped". Further information for each cDNA is presented in our database, DBTSS (http://dbtss.hgc.jp), and a user manual has been published [16].

The macaque genomic sequences to which our cynomolgus ESTs mapped were examined for any corresponding human genomic and RefSeq sequence. If any, the corresponding macaque EST was correlated with the human RefSeq gene. Based on information from the correlated human RefSeq gene, GO (Gene Ontology) classification was carried out for macaque ESTs using GO slim (http://www.geneontology.org/) for "Biological Process", Molecular Function", and "Cellular Component".

2.3. Identification of putative macaque-specific transcripts

To identify macaque ESTs that do not match to human genes, the ESTs were analyzed by either a genome- or cDNA-based approach. In the genome-based approach, we selected the EST sequence located outside human-macaque alignable regions according to the genome–genome alignment in the UCSC Genome Browser. In the cDNA-based approach, ESTs were first searched with the human RefSeq database using BLASTN (cut-off = 1.0e−100). Those ESTs with no hits were clustered with each other (cut-off = 0.0; >98% identity) and clusters containing more than 10 ESTs were selected. Those clustered cDNAs were searched against the human RefSeq database again (1.0e−50), and the generated sequence alignments were further manually inspected. For the macaque-specific transcript candidates, complete sequences were determined by primer-walking.

3. Results and discussion

3.1. Sequencing and clustering of macaque liver ESTs

A full-length cDNA library was constructed from cynomolgus monkey (*Macaca fascicularis*) livers using the oligo-capping method [15]. One-pass sequencing at 5′ and 3′-ends of the liver cDNA clones and sequence processing generated a total of 28 655 high quality ESTs (deposited in GenBank under Accession Nos. BB873801–BB902455). Only 3′ ESTs (27 959 entries) were further analyzed. Of these ESTs, 14 727 (53%) were successfully mapped to 1064 different regions in the rhesus macaque genome. Of the 1064 regions, 819 (77%) reside in genomic regions highly homologous to human RefSeq genes as revealed by a genome–genome comparison, and were annotated with human RefSeq genes (Table 1). Clustering of 27 959 ESTs was carried out by calculating the number of ESTs that mapped to the same region, which should represent the cluster size for the corresponding gene. This analysis for the 1064 mapped regions in the genome indicated that these 1064 unique ESTs consisted of 525 contigs (49.3%) and 539 singletons (50.7%). The number of members in each cluster ranged up to 4354, with a 26.9 average. The gene expression profile based on our EST data reflected liver functional characteristics because the most abundantly expressed genes were hepatocyte-specific markers, such as albumin, fibrinogen gamma and beta polypeptides, haptoglobin, and alcohol dehydrogenase, all of which comprised more than half of the identified ESTs (Table 1). Such high redundancy of hepatic ESTs from the non-normalized cDNA libraries has been also seen for human libraries [17–19].

3.2. Functional classification of ESTs

Provisional functional classification was carried out using GO slim terms based on the human RefSeq genes that correlated with our macaque ESTs (Fig. 1). Out of 819 unique ESTs that matched a human RefSeq entry, 786 (96.0%) were assigned to at least one main category; Biological Process, Molecular Function, and Cellular Component, to which 520, 458, and 373 sequences (48.9%, 43.0%, and 35.1%) were classified, respectively. Sequences from 133 ESTs (16.2%) were annotated into all three categories. The largest EST groups include metabolism, transcription, protein biosynthesis, electron transport, transport, signal transduction, and lipid metabolism (Fig. 1A) for Biological Process, and protein binding, transferase activity, and nucleotide binding for Molecular Function (Fig. 1B).

3.3. ESTs relevant to drug metabolism

Our major objective was identification of cDNAs important for drug metabolism, namely those encoding drug-metabolizing enzymes, which belong to CYP, UGT, GST, SULT, and FMO families. The 446 ESTs in 23 clusters were highly homologous to genes for such drug-metabolizing enzymes in humans (Table 2). For the CYP family, only ESTs for the CYP1 to CYP4 subfamilies are indicated in the list because of their importance in drug metabolism. The CYP family contained 231 entries (51.8%), the largest group among the ESTs for drug-metabolizing enzymes. CYP, a phase I drug-metabolizing enzyme, is involved in hydroxylation of a large number of

Table 1
Genes abundantly expressed in liver (>100 reads)

Contig number	ESTs	Human RefSeq ID	Annotation
15043	4354	NM_000477	Albumin
15223	3763	NM_000509	Fibrinogen gamma chain
15577	1097	NM_005141	Fibrinogen beta chain
10429	277	NM_005143	Haptoglobin
19271	253	NM_000668	Alcohol dehydrogenase 1B (class I), beta polypeptide
17733	245	NM_001085	Serpin peptidase inhibitor, clade A, member 3
5070	227	NM_000035	Aldolase B, fructose-bisphosphate
6405	158	NM_016413	Carboxypeptidase B2 (plasma, carboxypeptidase U)
5438	154	NM_000638	Vitronectin
11141	130	NM_000354	Serpin peptidase inhibitor, clade A, member 7
12017	125	NM_001622	Alpha-2-HS-glycoprotein
15222	109	NM_000508	Fibrinogen alpha chain
17083	106	NM_001756	Serpin peptidase inhibitor, clade A, member 6

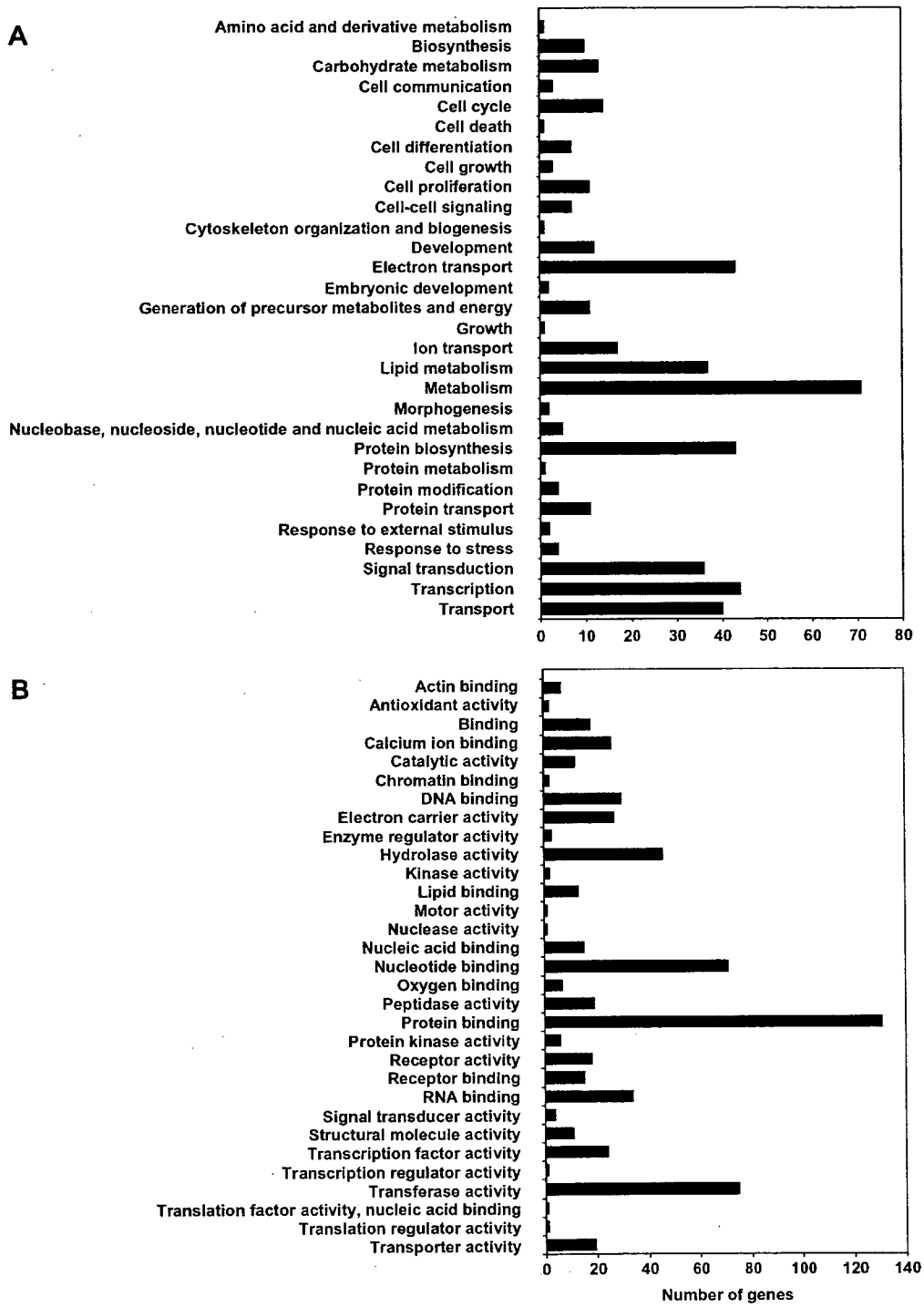


Fig. 1. Functional classification of cynomolgus liver ESTs. All non-redundant ESTs were assigned to each functional category as described in Section 2. Biological process (A) and Molecular function (B) are shown.

drugs [20]. Among the CYP ESTs identified, 124 (53.7%) belonged to the CYP2C subfamily that is important for metabolism of ~20% of all prescribed drugs such as tolbutamide, phenytoin, and warfarin [21]. Fifty-two ESTs belonged to the CYP3A subfamily. In humans, genes in the CYP3A subfamily (especially *CYP3A4*) are essential for drug metabolism, and are involved in the metabolism of more than half the currently prescribed drugs. Moreover, human CYP3A4 and CYP3A5 occupy more than half of the total CYP protein content in liver [20], contributing substantially overall drug

Table 2
Cynomolgus ESTs highly homologous to human drug-metabolizing enzyme families, CYP, UGT, GST, SULT, and FMO

Family	Contig number	Number of ESTs	Accession number	Matched human cDNA
CYP	18729	89	NM_0007700	CYP2C8
	12912	38	NM_017460	CYP3A4
	463	30	NM_000106	CYP2D6
	19109	21	NM_000769	CYP2C19
	19111	14	NM_000771	CYP2C9
	12910	14	NM_000777	CYP3A5
	19262	13	NM_000773	CYP2E1
	7671	8	NM_023944	CYP4F12
	8451	3	NM_000775	CYP2J2
	8410	1	NM_000778	CYP4A11
UGT	15423	75	NM_001074	UGT2B7
	2531	41	NM_019093	UGT1A3
	15424	30	NM_050394	UGT2B28
GST	19270	9	NM_145740	GSTA1
	14165	8	NM_000846	GSTA2
	9619	7	NM_000851	GSTM5
	1616	7	NM_145792	MGST1
	19167	3	NM_004832	GSTO1
	17697	2	NM_145870	GSTZ1
SULT	10631	10	NM_001055	SULT1A1
	19266	2	NM_001054	SULT1A2
	3203	1	NM_006588	SULT1C2
FMO	10051	20	NM_001002294	FMO3

metabolism in humans. Thirty ESTs were highly similar to human CYP2D6. In the human genome, three *CYP2D* genes are present including one functional *CYP2D* gene (*CYP2D6*) and two pseudogenes (*CYP2D7* and *CYP2D8*). CYP2D6 accounts for 5% of the total hepatic CYP content and is responsible for the metabolism of 25% of all drugs oxidized by CYPs [20]. In cynomolgus monkeys, CYP2D17, which is highly homologous to human CYP2D6, has been isolated [22]. Meanwhile, marmoset is known to have two functional CYP2Ds with different metabolic properties, CYP2D19 and CYP2D30 [23]. Further in-depth analysis of our EST clones could reveal whether *CYP2D17* is the only *CYP2D* gene expressed in cynomolgus monkey liver. Characterization of these CYP EST clones is currently in progress, such as full-length sequencing, tissue expression patterns, and metabolic assays, the outcome of which has been partly published [24,25].

Clusters for other drug-metabolizing enzymes of UGT, GST, SULT, and FMO families contained 146, 36, 13, and 20 ESTs, respectively (Table 2). UGT, a phase II drug-metabolizing enzyme, catalyzes the conjugation of various drugs to assist drug excretion and is composed of UGT1A, UGT2A, and UGT2B subfamilies in humans. The 146 ESTs for the UGT family were grouped into three clusters. Forty-one EST sequences were highly homologous to human UGT1A3. The *UGT1A* gene locus contains 13 distinct first exons (promoters) followed by exons 2–5 that are shared among all 13 transcripts, giving rise to nine different proteins (four pseudogenes) in humans [26]. Considering that these ESTs were 3′ cDNAs, the 41 EST clones possibly encode multiple *UGT1A* genes. Because only four *UGT1A* genes have been identified for macaques, further sequence analysis of these UGT1A ESTs could lead to the identification of novel *UGT1A* genes in this lineage. Seventy-five and 30 ESTs matched human UGT2B7 and UGT2B28, respectively. Initial full-length sequencing of the UGT2B EST clones revealed that the clones contained the UGT2B33 cDNA (GenBank Accession No. AB371703) newly identified in cynomolgus monkeys as well as the previously identified cDNAs for cynomolgus UGT2B9, UGT2B18, UGT2B19, UGT2B20, UGT2B23, and UGT2B30 (GenBank Accession Nos. U91582, AF016310, AF112112, AF072223, AF112113, and AF401657, respectively). In contrast to *UGT1A*, *UGT2B* genes have been frequently duplicated in many mammalian lineages [26]; therefore, some of these *UGT2Bs* are possibly lineage-specific genes as discussed below.

Other EST sequences were highly homologous to six human genes, two genes, and one gene in the GST, SULT, and FMO families, respectively (Table 2). GST is another phase II enzyme, catalyzing the conjugation of electrophilic substrates to glutathione and is composed of at least 16 genes for cytosolic, mitochondrial, and microsomal GSTs in humans [27]. SULT is also a gene family comprising at least 10 human genes, catalyzing sulfate conjugation of a wide variety of drugs [28]. FMO is a family of flavoproteins, catalyzing oxygenation of various drugs containing sulfur, nucleophilic nitrogen, and phosphorus heteroatoms [29]. Full-length sequencing and functional characterization of these EST clones are currently under investigation, by which novel genes could be identified because limited numbers of genes have been identified for these enzymes in monkeys. These results suggest that our EST-sequencing approach successfully identified a number of cDNA clones for various drug-metabolizing enzymes in monkeys.

3.4. Identification of lineage-specific genes

In order to better utilize monkeys as an animal model, it is essential to understand similarities or differences in genes expressed between monkeys and humans. The EST data should provide essential information on lineage-specific genes and transcripts. To identify macaque-specific transcripts, 27959 3′ ESTs were analyzed by either a genome- or cDNA-based approach. In the genome-based approach, we found 77 EST clusters, for which at least a part of the sequences were located outside human-macaque alignable regions. In the cDNA-based approach, we identified 12 clusters containing >10 ESTs that were unmatched to any human RefSeq genes according to BLASTN (cut-off = 1.0e−100). Clones available for the 10 remaining candidate clusters after subsequent manual inspection, along with clones for the 29 clusters randomly selected from 77 candidates in the genome-based approach, were subjected to full-length sequencing (excluding 1 overlapping clone). Sequence analysis of these 38 clones confirmed that nine clones contained lineage-specific candidate genes. Of these, six clones matched to human RefSeq sequences (Table 3): two clones lack a portion of human genome sequence and the other four matched to more than one member of a gene family. Thus, these four clones were potentially lineage-specific genes and were further characterized as described below.

One candidate clone (Qlv-U097A-G10) encoded CYP2C76 with a relatively low homology (~80%) to members of the human CYP2C subfamily, CYP2C8, CYP2C9, CYP2C18, and CYP2C19 (Table 3). The extent of homology was much lower than those for other ESTs (~95%). Our characterization

Table 3
Potential lineage-specific ESTs in cynomolgus monkey

Clone ID	GenBank Accession number	Nucleotide (bp)	ORF[a] (Number of amino acids)	Cynomolgus sequence	The most highly homologous human RefSeq cDNAs	Genome- or cDNA-based approach	Aligned location
Novel member of gene family							
Qlv-U042A-F11	AB362497	1637	454	None	CFH, CFHR3/4	Genome/cDNA	Intergenic
Qlv-U097A-G10	AB362507	1986	489	CYP2C76	CYP2C8/9/18/19	cDNA	Intergenic
Qlv-U346A-B11[b]	AB371605	1758	472	CYP2A23	CYP2A6/7/13	cDNA	Intergenic
Qlv-U405A-G11	AB362508	2225	528	UGT2B19	UGT2B4	cDNA	Intergenic
Partially unmatched to human genome							
Qlv-U244A-C6[b]	AB362499	1612	305	None	TSPAN12	Genome	Intergenic
Qlv-U258A-D7[b]	AB362500	1984	89	None	SS18L1	Genome	Intergenic
Unmatched to human genome							
Qlv-U050A-D10	AB362503	1700	34	None	None	Genome	Intron
Qlv-U295A-A3	AB362504	2278	118	None	None	Genome	Intergenic
Qlv-U389A-C1	AB362506	2043	90	None	None	Genome	Intergenic

[a]The longest ORF was selected.
[b]Transcript variants with different exon-intron structure from human homologs.

of CYP2C76 (GenBank Accession No. DQ074807) at the RNA, protein, and genomic level revealed that this CYP2C did not have any human ortholog because the corresponding genes were not found in the human genome [24]. Moreover, this CYP2C76 was at least partly responsible for lineage differences in drug metabolism [30]. These results confirmed that our comparative genomic approach succeeded in identifying macaque-specific transcripts that are absent in humans.

One clone (Qlv-U405A-G11) identified as a lineage-specific candidate contained the cDNA for UGT2B19 previously reported [31]. Cynomolgus UGT2B19 as well as UGT2B30 cDNAs were both highly homologous (92%) to human UGT2B4 cDNA [32]. A phylogenetic comparison (Fig. 2) indicated that the 1-to-1 orthologous relationship to the human UGT2Bs could not be determined for these cynomolgus UGT2Bs, raising the possibility that *UGT2B19* might be a lineage-specific gene. *UGT2B19* is expressed in cynomolgus monkey liver and prostate and has enzymatic activity to xenobiotics (1-naphthol) and steroids (testosterone) [31]. The UGT2B subfamily consisted of a number of member genes including a lineage-specific candidate [26], suggesting that UGT2B19 and other functional UGT2B enzymes in cynomolgus monkeys contribute not only to overall drug metabolism in monkey liver but also possibly to differences in drug metabolism.

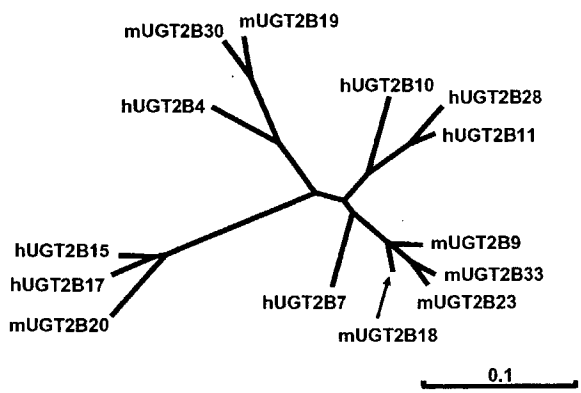


Fig. 2. A phylogenetic comparison of UGTs between macaque and human. The phylogenetic tree was based on amino acid sequence using the Clustal W program. Deduced amino acid sequences were used for cynomolgus monkeys (m) and human (h).

Another lineage-specific candidate clone (Qlv-U346A-B11) was cynomolgus CYP2A23 variant (tentatively named CYP2A23v), containing exons 1–8 with a partial intron 8 sequence and thus, lacking the entire exon 9 as compared to a complete CYP2A23 transcript. CYP2A23 and another cynomolgus CYP2A, CYP2A24, were both highly homologous (~95%) to the three human CYP2As, specifically CYP2A6, CYP2A7, and CYP2A13, indicating the difficulty in determining the orthologous relationship of CYP2A23 and CYP2A24 to human CYP2As [25]. This novel CYP2A23 transcript variant encodes a protein of 472 amino acids and lacks a part of a heme-binding region essential for CYP proteins (Fig. 3). The protein generated from this transcript, therefore, might not function as a drug-metabolizing enzyme. A similar transcript variant was also identified for CYP2C76 and UGT2B19 (data not shown). It remains to be determined whether the presence of these transcripts lacking a functional domain is limited to the animals that provided liver samples for the cDNA library construction and what roles these transcript variants play in drug metabolism.

Other than those for drug metabolism, one lineage-specific candidate (Qlv-U042A-F11) had high sequence homology to complement factor H (CFH) family genes in humans (Table 3). CFH (also called Factor H), an important complement regulator, forms a gene family along with CFH-related proteins (CFHL1-5) in humans [33]. This macaque transcript contained an open reading frame (ORF) of 454 amino acids. CFH and other genes important for immune response and T cell-mediated immunity such as *immunoglobulin-like* genes and MHC-related genes have been identified in macaques as the genes that went under positive selection [13,14,34], and thus, our finding of lineage-specific CFH-like sequence in macaques is not surprising. Further analysis of this CFH-like sequence indicated that the first 19 amino acids and the remaining amino acids were highly similar to CFH-related proteins (CFHR3 and CFHR4) and CFH in humans, respectively (data not shown), raising the possibility that this novel transcript might be a hybrid of CFH and CFH-related genes. In humans, a hybrid transcript of CFH and CFHR1 has been identified and implicated in atypical haemolytic uraemic syndrome [35].

```
mCYP2A23v   1:MLASGLLLVALLACLTVMVLMSVWQQRNSKGKLPPGPTPLPFIGNYLQLNTEQMYNSLMKISERYGPVFTIHLGPRRVVVLCGYDAVKEALVDQAEEFSG 100
mCYP2A23    1:MLASGLLLVALLACLTVMVLMSVWQQRNSKGKLPPGPTPLPFIGNYLQLNTEQMYNSLMKISERYGPVFTIHLGPRRVVVLCGYDAVKEALVDQAEEFSG 100
mCYP2A24    1:MLASGLLLVALLACLTVMVLMSVWQQRNSKGKLPPGPTPLPFIGNYLQLNTEQMCNSIMKISERYGPVFTIHLGPRRVVVLCGYDAVKEALVDQAEEFSG 100
hCYP2A6     1:MLASGMLLVALLVCLTVMVLMSVWQQRKSKGKLPPGPTPLPFIGNYLQLNTEQMYNSLMKISERYGPVFTIHLGPRRVVVLCGHDAVREALVDQAEEFSG 100
hCYP2A7     1:MLASGLLLVALLACLTVMVLMSVWQQRKSRGKLPPGPTPLPFIGNYLQLNTEHICDSIMKFSECYGPVFTIHLGPRRVVVLCGHDAVREALVDQAEEFSG 100
hCYP2A13    1:MLASGLLLVTLLACLTVMVLMSVWRQRKSRGKLPPGPTPLPFIGNYLQLNTEQMYNSLMKISERYGPVFTIHLGPRRVVVLCGHDAVKEALVDQAEEFSG 100
              *****.***.**.**********.** *.**********************....*.**.**.****************** ***.**********

mCYP2A23v 101:RGEQATFDWLFKGYGVVFSNGERAKQLRRFSIATLRDFGVGKRGIEERIQEEAGFLIEALRDTQGANIDPTFFLSRTVSNVISSIVFGDRFDYEDKEFLS 200
mCYP2A23  101:RGEQATFDWLFKGYGVVFSNGERAKQLRRFSIATLRDFGVGKRGIEERIQEEAGFLIEALRDTQGANIDPTFFLSRTVSNVISSIVFGDRFDYEDKEFLS 200
mCYP2A24  101:RGEQATFDWVFKGYGVVFSNGERAKQLRRFSIATLRDFGVGKRGIEERIQEEAGFLIEALRDTHGANIDPTFFLSRTVSNVISSIVFGDRFDYKDKEFLS 200
hCYP2A6   101:RGEQATFDWVFKGYGVVFSNGERAKQLRRFSIATLRDFGVGKRGIEERIQEEAGFLIDALRGTGGANIDPTFFLSRTVSNVISSIVFGDRFDYKDKEFLS 200
hCYP2A7   101:RGEQATFDWVFKGYGVAFSNGERAKQLLRFAIATLRDFGVGKRGIEERIQEESGFLIEAIRSTHGANIDPTFFLSRTVSNVISSIVFGDRFDYEDKEFLS 200
hCYP2A13  101:RGEQATFDWLFKGYGVAFSNGERAKQLRRFSIATLRGFGVGKRGIEERIQEEAGFLIDALRGTHGANIDPTFFLSRTVSNVISSIVFGDRFDYEDKEFLS 200
              ********* ******.**********.**.******.***************.****.*.* * ******************************.******

mCYP2A23v 201:LLRMMLGSFQFTATSAGQLYEMFSSVMKHLPGPQQQAFKELQGLEDFIAKKVEHNRRTLDPNSPRDFIDSFLIRMQEEEKNPNTEFHLKNLVLTSLNLFF 300
mCYP2A23  201:LLRMMLGSFQFTATSAGQLYEMFSSVMKHLPGPQQQAFKELQGLEDFIAKKVEHNRRTLDPNSPRDFIDSFLIRMQEEEKNPNTEFHLKNLVLTSLNLFF 300
mCYP2A24  201:LLGMMLAIFQFTSTSTGQLYEMFSSVMKHLPGPQQQAFKELQGLEDFIAKKVEHNQRTLDPNSPRDFIDSFLIRMQEEEKNPNTEFYLKNLMMTTLNLFI 300
hCYP2A6   201:LLRMMLGIFQFTSTSTGQLYEMFSSVMKHLPGPQQQAFQLLQGLEDFIAKKVEHNQRTLDPNSPRDFIDSFLIRMQEEEKNPNTEFYLKNLVMTTLNLFI 300
hCYP2A7   201:LLSMMLGIFQFTSTSTGQLYEMFSSVMKHLPGPQQQAFKLLQGLEDFIAKKVEHNQRTLDPNSPQDFIDSFLIHMQEEEKNPNTEFYLKNLMMSTLNLFI 300
hCYP2A13  201:LLRMMLGSFQFTATSTGQLYEMFSSVMKHLPGPQQQAFKELQGLEDFIAKKVEHNQRTLDPNSPRDFIDSFLIRMQEEEKNPNTEFYLKNLVMTTLNLFF 300
              **.***. **** **.**********************..***************.********.********.************.****....****

mCYP2A23v 301:GGTETVSTTLRYGFLLLMKHPEVEAKVHEEIDRVIGKNRQPKFEDWAKMPYTEAVIHEIQRFGDMLPFGVAHRVIKDTKFRDFFLPKGTEVFPMLGSVLK 400
mCYP2A23  301:GGTETVSTTLRYGFLLLMKHPEVEAKVHEEIDRVIGKNRQPKFEDWAKMPYTEAVIHEIQRFGDMLPFGVAHRVIKDTKFRDFFLPKGTEVFPMLGSVLK 400
mCYP2A24  301:AGTETVSTTLRYGFLLLMKYPEVEAKVHEEIDRVIGKNRQPKFEDRVKMPYMEAVIHEIQRFGDVIPMSLARRVNKDTKFRDFFLPKGTEVFPMLGSVLR 400
hCYP2A6   301:GGTETVSTTLRYGFLLLMKHPEVEAKVHEEIDRVIGKNRQPKFEDRAKMPYMEAVIHEIQRFGDVIPMSLARRVKKDTKFRDFFLPKGTEVFPMLGSVLR 400
hCYP2A7   301:AGTETVSTTLRYGFLLLMKHPEVEAKVHEEIDRVIGKNRQPKFEDRTKMPYMEAVIHEIQRFGDVIPMSLARRVKKDTKFRDFFLPKGTEVFPMLGSVLR 400
hCYP2A13  301:AGTETVSTTLRYGFLLLMKHPEVEAKVHEEIDRVIGKNRQPKFEDRAKMPYTEAVIHEIQRFGDMLPMGLAHRVNKDTKFRDFFLPKGTEVFPMLGSVLR 400
              ******************.*************************..**** ************ *. .* ** ************************.

mCYP2A23v 401:DPKFFSNPQDFNPQHFLDEKGQFKKSDAFVPFSIGKRPLFAARPLLTPAGASLTQLPSVLCSLVFLQLGSSC                          472
mCYP2A23  401:DPKFFSNPQDFNPQHFLDEKGQFKKSDAFVPFSIGKRNCFGEGLARMELFLFLTTIMQNFRFKSPQSPKDIDVSPKHMGFATIPPNYTMSFLPR  494
mCYP2A24  401:DPRFFSNPQDFNPQHFLDEKGQFKKSDAFVPFSIGKRNCFGEGLARMELFLFFTTIMQNFRFKSPQSPKDIDVSPKHAGFATIPRNYTMSFLPR  494
hCYP2A6   401:DPSFFSNPQDFNPQHFLNEKGQFKKSDAFVPFSIGKRNCFGEGLARMELFLFFTTVMQNFRLKSSQSPKDIDVSPKHVGFATIPRNYTMSFLPR  494
hCYP2A7   401:DPSFFSNPQDFNPQHFLDDKGQFKKSDAFVPFSIGKRNCFGEGLARMELFLFFTTVMQNFRLKSSQSPKDIDVSPKHVVFATIPRNYTMSFLPR  494
hCYP2A13  401:DPRFFSNPRDFNPQHFLDKKGQFKKSDAFVPFSIGKRYCFGEGLARMELFLFFTTIMQNFRFKSPQSPKDIDVSPKHVGFATIPRNYTMSFLPR  494
              ** *****.********..******************..*.............*. ..... .. .......
```

Fig. 3. Alignment of the amino acid sequences deduced from cynomolgus monkey (m) and human (h) CYP2A cDNAs. The sequences were aligned using the Clustal W program. Asterisks and dots under the sequences indicate identical amino acids and conservatively unchanged amino acids, respectively. A black line under the amino acid sequences indicates the putative heme-binding region. The CYP2A23 variant (mCYP2A23v) newly identified lacks half of the putative heme-binding region.

These results suggest that our approach of lineage-specific gene identification successfully identified potential lineage-specific genes or transcripts, possibly relevant to the immune system. Further investigation of other ESTs should help make better use of the macaque for immunological studies.

Three candidate genes were unmatched to any human RefSeq sequence, and thus could be apparent lineage-specific genes (Table 3). The two candidate genes (Qlv-U295A-A3 and Qlv-U389A-C1) reside in intergenic regions of the macaque genome, which might be novel genes in monkeys. This was supported by the fact that these two sequences did not match any human ESTs by BLAST (data not shown). The two transcripts contained relatively small ORFs (<100 amino acids). Transcripts with small ORFs have been identified in mice and humans, some of which could be actually translated in vitro [36,37]. Alternatively, these mRNAs might be functioning as non-coding RNAs. A large proportion of transcripts are non-coding RNAs, including those having essential functions in transcriptional and translational control [38,39].

4. Conclusion

The data presented here provide an overview of genes expressed in cynomolgus liver to investigate liver physiology for macaques. ESTs for genes encoding a variety of drug-metabolizing enzymes hold great promise in deepening our understanding of drug metabolism in monkeys, which in turn helps to elucidate lineage differences between monkeys and humans. Indeed, our characterization of CYP2C ESTs has identified lineage-specific CYP2C76, which is responsible for lineage differences in drug metabolism [24,30]. Furthermore, the ESTs generated in this study can be a resource for the production of microarrays. Given that our cDNA library was generated with RNAs from only three animals, the EST sequencing using the library originated from the RNA samples of more animals would be useful for the identification of the allelic variants expressed in liver.

Many drug-metabolizing enzyme genes are confined to gene families, many of which have been subjected to gene duplication or gene loss during evolution, resulting in family size differences [40]. This indicates that lineage-specific genes could be identified for gene families even between evolutionarily close lineages such as monkeys and humans. Moreover, physiological differences should partly result from differences at the transcriptional level, for example, by alternative splicing and non-coding RNAs [41]. Further investigation of our EST data will lead to the identification of lineage-specific transcripts generated by alternative splicing and lineage-specific gene gain/loss, as the efforts for identifying such transcripts have succeeded partly in macaques [9,13]. The identified lineage-specific transcripts and genes will help lead to a better understanding

of the physiological differences between monkeys and humans, leading to more efficient utilization of monkeys as an animal model.

References

[1] Stevens, J.C., Shipley, L.A., Cashman, J.R., Vandenbranden, M. and Wrighton, S.A. (1993) Comparison of human and rhesus monkey *in vitro* phase I and phase II hepatic drug metabolism activities. Drug Metab. Dispos. 21, 753–760.

[2] Sharer, J.E., Shipley, L.A., Vandenbranden, M.R., Binkley, S.N. and Wrighton, S.A. (1995) Comparisons of phase I and phase II *in vitro* hepatic enzyme activities of human, dog, rhesus monkey, and cynomolgus monkey. Drug Metab. Dispos. 23, 1231–1241.

[3] Guengerich, F.P. (1997) Comparisons of catalytic selectivity of cytochrome *P*450 subfamily enzymes from different species. Chem. Biol. Interact. 106, 161–182.

[4] Shimada, T., Mimura, M., Inoue, K., Nakamura, S., Oda, H., Ohmori, S. and Yamazaki, H. (1997) Cytochrome P450-dependent drug oxidation activities in liver microsomes of various animal species including rats, guinea pigs, dogs, monkeys, and humans. Arch. Toxicol. 71, 401–408.

[5] Weaver, R.J., Dickins, M. and Burke, M.D. (1999) A comparison of basal and induced hepatic microsomal cytochrome P450 monooxygenase activities in the cynomolgus monkey (*Macaca fascicularis*) and man. Xenobiotica 29, 467–482.

[6] Bogaards, J.J., Bertrand, M., Jackson, P., Oudshoorn, M.J., Weaver, R.J., van Bladeren, P.J. and Walther, B. (2000) Determining the best animal model for human cytochrome P450 activities: a comparison of mouse, rat, rabbit, dog, micropig, monkey and man. Xenobiotica 30, 1131–1152.

[7] Narimatsu, S., Kobayashi, N., Masubuchi, Y., Horie, T., Kakegawa, T., Kobayashi, H., Hardwick, J.P., Gonzalez, F.J., Shimada, N., Ohmori, S., Kitada, M., Asaoka, K., Kataoka, H., Yamamoto, S. and Satoh, T. (2000) Species difference in enantioselectivity for the oxidation of propranolol by cytochrome P450 2D enzymes. Chem. Biol. Interact. 127, 73–90.

[8] Sakate, R., Osada, N., Hida, M., Sugano, S., Hayasaka, I., Shimohira, N., Yanagi, S., Suto, Y., Hashimoto, K. and Hirai, M. (2003) Analysis of 5′-end sequences of chimpanzee cDNAs. Genome Res. 13, 1022–1026.

[9] Magness, C.L., Fellin, P.C., Thomas, M.J., Korth, M.J., Agy, M.B., Proll, S.C., Fitzgibbon, M., Scherer, C.A., Miner, D.G., Katze, M.G. and Iadonato, S.P. (2005) Analysis of the *Macaca mulatta* transcriptome and the sequence divergence between *Macaca* and human. Genome Biol. 6, R60.

[10] Li, Y. and Su, B. (2006) No accelerated evolution of 3′UTR region in human for brain-expressed genes. Gene 383C, 38–42.

[11] Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirata, M., Suto, Y., Hirai, M., Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K. (2001) Assignment of 118 novel cDNAs of cynomolgus monkey brain to human chromosomes. Gene 275, 31–37.

[12] Osada, N., Hida, M., Kusuda, J., Tanuma, R., Hirata, M., Suto, Y., Hirai, M., Terao, K., Sugano, S. and Hashimoto, K. (2002) Cynomolgus monkey testicular cDNAs for discovery of novel human genes in the human genome sequence. BMC Genomics 3, 36.

[13] Chen, W.H., Wang, X.X., Lin, W., He, X.W., Wu, Z.Q., Lin, Y., Hu, S.N. and Wang, X.N. (2006) Analysis of 10,000 ESTs from lymphocytes of the cynomolgus monkey to improve our understanding of its immune system. BMC Genomics 7, 82.

[14] The Rhesus Macaque Genome Sequencing and Analysis Consortium (2007) Evolutionary and biomedical insights from the rhesus macaque genome. Science 316, 222–234.

[15] Suzuki, Y. and Sugano, S. (2003) Construction of a full-length enriched and a 5′-end enriched cDNA library using the oligo-capping method. Method Mol. Biol. 221, 73–91.

[16] Yamashita, R., Suzuki, Y., Wakaguri, H., Tsuritani, K., Nakai, K. and Sugano, S. (2006) DBTSS: database of human transcription start sites, progress report 2006. Nucl. Acid Res. 34, D86–89.

[17] Xu, X.R., Huang, J., Xu, Z.G., Qian, B.Z., Zhu, Z.D., Yan, Q., Cai, T., Zhang, X., Xiao, H.S., Qu, J., Liu, F., Huang, Q.H., Cheng, M., Li, N.G., Du, J.J., Hu, W., Shen, K.T., Lu, G., Fu, G., Zhong, M., Xu, S.H., Gu, W.Y., Huang, W., Zhao, X.T., Hu, G.X., Gu, J.R., Chen, Z. and Han, Z.G. (2001) Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver. Proc. Natl. Acad. Sci. USA 98, 15089–15094.

[18] Yu, Y., Zhang, C., Zhou, G., Wu, S., Qu, X., Wei, H., Xing, G., Dong, C., Zhai, Y., Wan, J., Ouyang, S., Li, L., Zhang, S., Zhou, K., Zhang, Y., Wu, C. and He, F. (2001) Gene expression profiling in human fetal liver and identification of tissue- and developmental-stage-specific genes through compiled expression profiles and efficient cloning of full-length cDNAs. Genome Res. 11, 1392–1403.

[19] Otsuka, M., Arai, M., Mori, M., Kato, M., Kato, N., Yokosuka, O., Ochiai, T., Takiguchi, M., Omata, M. and Seki, N. (2003) Comparing gene expression profiles in human liver, gastric, and pancreatic tissues using full-length-enriched cDNA libraries. Hepatol. Res. 27, 76–82.

[20] Guengerich, F.P. (2005) Human cytochrome P450 enzymes in: (Ortiz de Montellano, P., Ed.), third ed, Cytochrome P450: Structure, Mechanism, and Biochemistry, pp. 377–530, Kluwer Academic/Plenum Publishers, New York.

[21] Goldstein, J.A. (2001) Clinical relevance of genetic polymorphisms in the human CYP2C subfamily. Brit. J. Clin. Pharmacol. 52, 349–355.

[22] Mankowski, D.C., Laddison, K.J., Christopherson, P.A., Ekins, S., Tweedie, D.J. and Lawton, M.P. (1999) Molecular cloning, expression, and characterization of CYP2D17 from cynomolgus monkey liver. Arch. Biochem. Biophys. 372, 189–196.

[23] Hichiya, H., Kuramoto, S., Yamamoto, S., Shinoda, S., Hanioka, N., Narimatsu, S., Asaoka, K., Miyata, A., Iwata, S., Nomoto, M., Satoh, T., Kiryu, K., Ueda, N., Naito, S., Tucker, G.T. and Ellis, S.W. (2004) Cloning and functional expression of a novel marmoset cytochrome P450 2D enzyme, CYP2D30: comparison with the known marmoset CYP2D19. Biochem. Pharmacol. 68, 165–175.

[24] Uno, Y., Fujino, H., Kito, G., Kamataki, T. and Nagata, R. (2006) CYP2C76, a novel CYP in cynomolgus monkey, is a major CYP2C in liver, metabolizing tolbutamide and testosterone. Mol. Pharmacol. 70, 477–486.

[25] Uno, Y., Hosaka, S., Matsuno, K., Nakamura, C., Kito, G., Kamataki, T. and Nagata, R. (2007) Characterization of cynomolgus monkey cytochrome P450 (CYP) cDNAs: Is *CYP2C76* the only monkey-specific *CYP* gene responsible for species differences in drug metabolism? Arch. Biochem. Biophys. 466, 98–105.

[26] Mackenzie, P.I., Bock, K.W., Burchell, B., Guillemette, C., Ikushiro, S., Iyanagi, T., Miners, J.O., Owens, I.S. and Nebert, D.W. (2005) Nomenclature update for the mammalian UDP glycosyltransferase (*UGT*) gene superfamily. Pharmacogenet. Genomics 15, 677–685.

[27] Hayes, J.D., Flanagan, J.U. and Jowsey, I.R. (2005) Glutathione transferases. Annu. Rev. Pharmacol. Toxicol. 45, 51–88.

[28] Blanchard, R.L., Freimuth, R.R., Buck, J., Weinshilboum, R.M. and Coughtrie, M.W. (2004) A proposed nomenclature system for the cytosolic sulfotransferase (SULT) superfamily. Pharmacogenetics 14, 199–211.

[29] Cashman, J.R. and Zhang, J. (2006) Human flavin-containing monooxygenases. Annu. Rev. Pharmacol. Toxicol. 46, 65–100.

[30] Uno, Y., Kumano, T., Kito, G., Nagata, R., Kamataki, T. and Fujino, H. (2007) CYP2C76-mediated species difference in drug metabolism: A comparison of pitavastatin metabolism between monkeys and humans. Xenobiotica 37, 30–43.

[31] Belanger, G., Barbier, O., Hum, D.W. and Belanger, A. (1999) Molecular cloning, expression and characterization of a monkey steroid UDP-glucuronosyltransferase, UGT2B19, that conjugates testosterone. Eur. J. Biochem. 260, 701–708.

[32] Girard, C., Barbier, O., Turgeon, D. and Belanger, A. (2002) Isolation and characterization of the monkey *UGT2B30* gene that encodes a uridine diphosphate-glucuronosyltransferase enzyme active on mineralocorticoid, glucocorticoid, androgen and oestrogen hormones. Biochem. J. 365, 213–222.

[33] Male, D.A., Ormsby, R.J., Ranganathan, S., Giannakis, E. and Gordon, D.L. (2000) Complement factor H: sequence analysis of