If the secretion increases 100-fold, the blood concentration of the protein becomes approximately 2-fold; however, the difference is still less than that of individual variability. Alpha-fetoprotein (AFP), a clinical marker for HCC, is elevated from 3 ng/mL (i.e., average level in healthy individuals) to ≥20 ng/mL in a cancer patient,[11] which indicates HCC cells secrete >500-fold AFP by comparison to noncancerous cells. In fact, AFP was found to be a major component of the culture medium of a human hepatoma cell line (HepG2; data not shown). Thus, it is relatively straightforward to identify proteins that are significantly overexpressed by cancer cells (e.g., AFP). However, such proteins are not always found. These considerations reveal the difficulty of finding a "cancer-specific biomarker" based on the quantification of proteins.

We have established an integrated strategy for high-throughput discovery of serum biomarker candidates by utilizing a number of advanced technologies. Specifically, we developed techniques to analyze glycan biosynthesis-related genes (glycogenes), glycan structures, and glycoproteins as described previously.[12,13] We designed the strategy based on a well-known fact that the glycan structure on the cell surface and secretory proteins derived from cancerous cells is different from that derived from the same kind of noncancerous cells. Thus, we believe a qualitative alteration in the glycan structure, rather than a quantitative change in the amount of protein, will be a good index for developing serum biomarkers. The strategy comprises a 4-step analyses: (1) glycogene expression profiling by quantitative real-time PCR array for cancer cell lines, to predict possible and significant glycan structural motifs such as α1,6-fucose, bisecting GlcNAc (GlcNAcβ1-4Man), β1,6-branch-(GlcNAcβ1-6Man), poly-N-acetyllactosamine extension, Lewis antigen, and LacdiNAc (GalNAcβ1-4GlcNAc); (2) comparative glycan profiling by lectin array analysis of proteins secreted from cancer cells into the culture media, to determine cancer-associated glycan alteration and to select lectins to capture glycoproteins or glycopeptides displaying the altered glycans; (3) identification of core proteins carrying the cancer-associated glycans by LC−MS analysis; and (4) selection of promising glycoprotein candidates by comparative profiling of the identified glycoproteins in cancerous samples and healthy controls, followed by bioinformatics to establish tissue specificity and the expression level in the liver.

In accordance with this strategy, we digested glycoprotein mixtures prepared from the culture media of hepatoma cell lines, HepG2 and HuH-7, and also from the sera of HCC patients having a background of LC and those of healthy volunteers (HV). Glycopeptides were separately captured using AAL, DSA, and RCA120 lectins. Many glycoprotein biomarker candidates were identified (744 for AAL and 234 for DSA). In all, 21 AAL-bound glycoproteins were selected to determine the differences in their glycan structure between HCC patients and HV. All selected glycoproteins showed increases of AAL-reactivity in HCC sera, suggesting these would be valuable biomarkers of HCC, LC, or fibrosis progression. Our results confirm the effectiveness of this strategy in identifying cancer glycobiomarker candidates in a high-throughput manner.

METHODS

Materials

We collected sera from 8 HCC patients at Nagoya City University Hospital (NCUH) and control sera from 14 randomly selected healthy volunteers (HV) of 28−48 years of age at AIST, after obtaining the approval from the respective institutional ethics committees and the informed consent from all individuals for the use of clinical specimens. Clinical information of the HCC patients and information on sex and age of HV are presented in Supplementary Table 1. HV did not have hepatitis, other hepatic diseases, renal failure, clinical cardiovascular disease, or malignancies. Because we aimed at discovering not only biomarker candidates of HCC development but also specific markers for various stages between early phase fibrosis and severe liver cirrhosis, we used sera of HV, who were relatively younger than HCC patients, as control samples. To inactivate hepatitis viruses, each serum sample (both HV and HCC patients) was diluted 10-fold with Dulbecco's phosphate buffer containing 0.15 M NaCl (D-PBS) and 0.2% SDS and heated at 98 °C for 15 min before use. Human hepatoma cell lines (HepG2 and HuH-7) were obtained from RIKEN Cell Bank (Ibaraki, Japan). Details of all antibodies used in this study are presented in Supplementary Table 2.

Other reagents used in this study were as follows. Guanidine-HCl and neuraminidase (*Arthrobacter ureafaciens*, EC 3.2.1.18) were purchased from Nacalai Tesque (Kyoto, Japan). Dithiothreitol (DTT), iodoacetamide (IAA), trichloroacetic acid (TCA), methyl α-D-mannopyranoside, lactose, ethanol, 1-butanol, and trifluoroacetic acid (TFA) were from Wako Pure Chemicals (Osaka, Japan). Tosyl phenylalanyl chloromethyl ketone (TPCK)-treated trypsin was obtained from Thermo Fisher Scientific (Waltham, MA). Stable isotope-labeled water ($H_2{}^{18}O$, 99% atom% ^{18}O) was a product of Taiyo Nippon Sanso Corp. (Tokyo, Japan). L-Fucose and phenylmethylsulfonyl fluoride (PMSF) were obtained from Sigma-Aldrich (St Louis, MO). Glycopeptidase F (peptide-*N*-glycanase; PNGase) was from TakaraBio (Kyoto, Japan). The HPLC lectin column (LA-AAL, *Aleuria aurantia* lectin) and chitooligosaccharide were from Seikagaku Biobusiness (Tokyo, Japan). DSA (*Datura stramonium* agglutinin)-agarose and RCA (*Ricinus communis* agglutinin) 120-agarose were from Vector Laboratories, Inc. (Burlingame, CA). All solutions were prepared using ultrapure water. Analytical grade reagents were used unless stated otherwise.

HepG2 and HuH-7 were cultured in a Dulbecco's modified Eagle's medium (D-MEM; Life Technologies, Carlsbad, CA) supplemented with 10% heat-inactivated fetal bovine serum (FBS; JRH Biosciences, Lenexa, KS), 100 units/mL penicillin (Life Technologies), and 100 μg/mL streptomycin (Life Technologies) and maintained at 37 °C in an incubator with 5% CO_2. The cells at 60−80% confluence were washed 3 times with D-PBS and further cultured in FBS- and antibiotics-free D-MEM for 48 h. The final media were harvested, and the supernatants were filtered using a 0.45-μm disc filter (Millipore, Billerica, MA).

Preparation of Tryptic Digests of Culture Medium and Serum Proteins

The pooled HV serum was prepared by mixing equal volumes of heated sera of 14 HV. Proteins in the serum and culture media were collected by TCA precipitation (final concentration of 10%, w/w), S-reduced, and then alkylated as described previously.[14] For discovery experiments, we used 5 HCC sera (S-1 to S-5) listed in Supplementary Table 2. Proteins were digested with TPCK-treated trypsin, using an E/S ratio of 1/100 (w/w), at 37 °C for 16 h. Progression of the digestion was monitored by SDS-polyacrylamide gel electrophoresis followed by CBB staining. Protein concentrations were assayed by the BCA method using bovine serum albumin as a standard.

dx.doi.org/10.1021/pr301217b | J. Proteome Res. XXXX, XXX, XXX−XXX

Lectin Affinity Capture of Glycopeptides from Tryptic Digests

Lectin affinity chromatography was performed as described previously.[14] Briefly, an aliquot of the tryptic digest was applied to an AAL column (4.6 mm i.d. × 150 mm) equilibrated with 10 mM HEPES-NaOH, pH 7.5. Glycopeptides were eluted with the same buffer containing 5 mM fucose. The pass-through fraction was applied again onto the same column, and residual glycopeptides were recovered using an identical procedure. Similarly, the trypsin digest was applied onto a DSA column (10 mm i.d. × 20 mm) after desialylation treatment with neuraminidase from *Arthrobacter ureafaciens*. Glycopeptides were then eluted with 1% chito-oligosaccharide (GlcNAc oligomers), and the eluate was directly applied onto an online RCA120 column (10 mm i.d. × 20 mm) to remove GlcNAc oligomers. Glycopeptides trapped on the RCA120 column were eluted with 0.1 M lactose. In addition, glycopeptides in the pass-through fractions of the AAL column chromatography were applied onto a RCA120 column (10 mm i.d. × 20 mm). Hereafter a fraction (protein/peptide) bound to a lectin column is denoted as "lectin(+)", e.g., AAL(+). Similarly, a passed (unbound) fraction is referred to as, e.g., DSA(−). Subsequently, all glycopeptide samples prepared (AAL(+), DSA(+), and AAL(−)/RCA120(+)) were further purified by hydrophilic interaction chromatography on a Sepharose CL-4B column as described previously.[14]

Enzymatic Stable Isotope Labeling of Glycopeptides

Purified glycopeptides were treated with PNGase (2.5 mU, 37 °C overnight) in stable isotope ^{18}O-labeled water to remove the glycan moiety and concomitantly label the glycosylated Asn of glycopeptides with the isotope, as described previously.[15] This step is referred to as isotope-coded glycosylation site-specific tagging (IGOT).

Identification of the Labeled Peptides by Nanoflow LC–MS Analysis

Stable isotope-labeled peptides were analyzed by the LC–MS method, as described previously.[16] Briefly, the peptide mixture was injected onto a C18 trap column (0.5 mm i.d. × 1 mm). After washing, the column was connected to a nanoflow LC system (flow rate: 100 nL/min), and the peptides were separated on a reverse phase (C18) tip column (150 μm i.d. × 70 mm) using a linear gradient of MeCN (0–35% in 0.1% formic acid) for 70 min. The eluted peptides were sprayed directly into a quadrupole time-of-flight hybrid mass spectrometer (Q-TOF Ultima; Waters-Micromass, Beverly, MA). The spectrometer was operated in data-dependent MS/MS mode.

MS/MS spectra were processed using MassLynx software (version 4.0, Micromass) to create peak list files with smoothing by the Savitzky-Golay method (window channels, ±3). The files were processed by the MASCOT algorithm (Version 2.1, Matrix Science) to assign peptides using the refseq protein sequence database (34,276 entries, downloaded on July 2007). The database search was performed by the MS/MS ion search method with the following parameters. Enzyme: Trypsin + Lys-C; Fixed modification: Carbamidomethyl (Cys); Variable modifications: Gln > pyro-Glu (N-term Gln), Oxidation (Met), Pyro-carbamidomethyl (N-term Cys), IGOT (Asn > Asp+^{18}O = +2.988261 Da, custom-made) (Asn); Peptide mass tolerance: 200 ppm; Fragment mass tolerance: 0.5 Da; Max missed cleavage: 2. All results from the peptide search were exported as CSV files and processed by Microsoft Excel. First, we selected the peptides with rank 1 and an expectation value <0.05. Then, we selected the peptides that contained one or more aspartic acids labeled with ^{18}O atoms (IGOT) at the position of Asn in the consensus sequence for *N*-glycosylation, Asn-Xaa-(Ser/Thr), where Xaa is any residue except Pro.

Immunoprecipitation and Lectin Array Analysis of the Glycoprotein Marker Candidates

Selected glycoproteins were enriched with their specific antibodies. Aliquots (10 μL each) of the pooled sera of HCC patients and HV were incubated overnight with 1 μg of biotinylated antibody at room temperature. The resultant complexes were trapped by 20 μL of streptavidin-conjugated magnetic beads (Dynabeads MyOne Streptavidin T1; Invitrogen) and then released by heating with 20 μL of 0.2% SDS in TBS. Glycan profiles of the released proteins were analyzed by the antibody-assisted lectin profiling (ALP) method.[17] Briefly, we analyzed array signals by using LecChip and GlycoStation Reader (GP Biosciences Ltd., Yokohama, Japan) according to a prescribed method.[17] The data were processed with the program software, Array-Pro Analyzer (Ver. 4.5, Media Cybernetics, Inc., Silver Spring, MD).

AAL Fractionation of Serum Proteins

The denatured serum (10 μL each) was incubated with 100 μL of AAL-agarose at 4 °C for 16 h with gentle shaking. After recovery of the supernatant (AAL(−) fraction), the gel was washed three times with PBSTx. Then, the gel was suspended with an AAL elution buffer (PBSTx containing 500 mM L-fucose) and shaken at 4 °C for 16 h. The supernatant (AAL(+) fraction) was then collected. Both AAL(+) and AAL(−) fractions were concentrated by ultrafiltration using an Amicon Ultra unit (Ultracel 3 kDa, Millipore Amicon) to adjust the volume of each fraction.

LCA-AAL Serial Lectin Fractionation of Serum Glycoproteins

The glycoproteins in the pooled sera of HCC patients and HV were separated by serial lectin fractionation using LCA- and AAL-conjugated gels (Kuno, A., et al. manuscript in preparation). Briefly, the 100-fold diluted serum (100 μL) was incubated with 100 μL of LCA-agarose at 4 °C for 5 h with gentle shaking, and the supernatant (LCA(−) fraction) was recovered. After washing, LCA-bound proteins were eluted with 100 μL of an LCA elution buffer (PBSTx containing 500 mM α-methyl-D-mannopyranoside). Then, the LCA(−) fraction was incubated with 50 μL of AAL-agarose at 4 °C for 16 h, and the supernatant (LCA(−)/AAL(−) fraction) was recovered. The LCA(−)/AAL(+) fraction was recovered with 500 mM L-fucose.

Western Blot Analysis

Protein samples including an equivalent volume of serum were subjected to 10% SDS-PAGE and transferred to a PVDF membrane (BioRad Laboratories, Hercules, CA). The membranes were blocked and incubated with each antibody. Finally, the treated membrane was reacted with the Western Lightning chemiluminescence reagent (Perkin-Elmer, Boston, MA) at room temperature and exposed to Hyperfilm ECL (GE Healthcare) for 60 or 300 s.

■ RESULTS

Selection of Lectins for Capturing Glycopeptides Carrying HCC-Related Glycans

A tissue-specific protein carrying a cancer-associated glycan is suitable as a biomarker for the cancer. To develop such glycoprotein biomarkers, we designed an integrated strategy as described previously.[12] In accordance with this strategy (Figure

dx.doi.org/10.1021/pr301217b | J. Proteome Res. XXXX, XXX, XXX–XXX

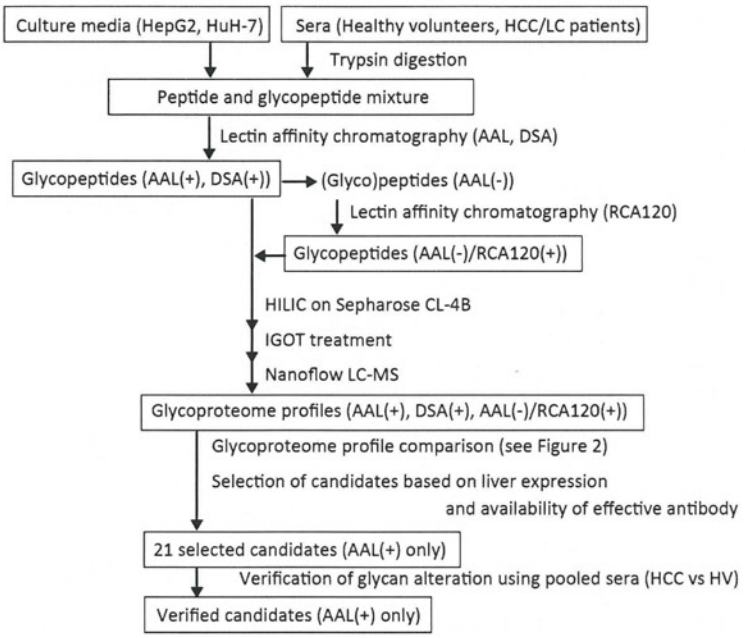

Figure 1. Experimental workflow of the discovery of serum glyco-biomarker candidates for liver disease. Glycopeptide fractions were captured with AAL- or DSA-affinity columns from tryptic digests of culture media (HepG2 and HuH-7) and serum taken from healthy volunteers and HCC patients. RCA-bound glycopeptides were recovered from pass fractions of the AAL-column for serum samples. After purification by HILIC, glycopeptides were identified by the IGOT-LC–MS method. From the AAL(+) glycoproteins, 21 glycoproteins were selected for further verification of HCC-associated glycan alteration (enhanced fucosylation). See Methods for further details.

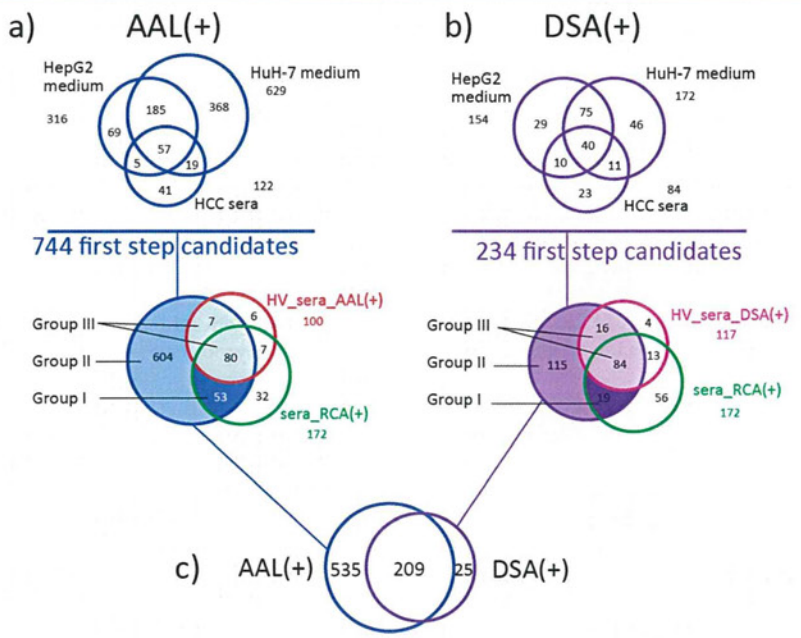

Figure 2. Venn diagram presentation for glycoprotein profile comparison. Lectin-bound glycopeptides were prepared from culture media (HepG2 and HuH-7) and sera of HCC patients and then identified by the IGOT-LC–MS method. In all, (a) 744 AAL(+) and (b) 234 DSA(+) glycoproteins were identified. AAL(+) and DSA(+) glycoproteins were identified from the sera of healthy volunteers (HV; red circle), and RCA120(+) glycoproteins were identified from the sera of HV and HCC patients (green circle). By comparing the glycoprotein profiles, HCC_AAL(+) proteins were classified into 3 groups. (c) Glycoproteins of HCC_AAL(+) and DSA(+) glycoproteins were compared. Most DSA(+) glycoproteins also possess fucosylated glycans.

1), we analyzed the glycan profiles of culture media of HCC cell lines (HepG2 and HuH-7)[13] and found that the lectin array signals of AAL and DSA as well as RCA120 were significant.

AAL has specificity for fucosylated glycans[18] and DSA for branched asialoglycans (tri- or tetra-antennary *N*-glycans).[19] It is known that the levels of these glycan motifs are increased by

dx.doi.org/10.1021/pr301217b | *J. Proteome Res.* XXXX, XXX, XXX–XXX

Table 1. Biomarker Candidates Selected for Further Verification of HCC-Related Glycan Alteration (Fucosylation)[a]

				applicability of antibody		results of verification of the candidatens			
								serial lectin fractionation and WB	
candidate no. at verification step	group in Venn diagram (Figure 2)	gene symbols	description	WB (for sera)	IP (from sera)	IP and Lectin array (AAL)	AAL fractionation and WB	LCA(+)	LCA(−) AAL(+)
1	I	CPB2	plasma carboxypeptidase B2	√	√	ND	↑	→	→
2	I	CPN2	carboxypeptidase N, polypeptide 2, 83 kD	√	√	↑	NT	NT	NT
3	I	CSF1R	colony stimulating factor 1 receptor	√	√	↑	↑	↑	↑
4	I	FN1	fibronectin 1 isoform 7	√	−	NT	NT	→	↑
5	I	LAMP2	lysosomal-associated membrane protein 2	√	√	ND	NT	↑	→
6	I	SERPINA7	serine (or cysteine) proteinase inhibitor, clade A, member 7	√	−	NT	NT	→	↑
7	I	SHBG	sex hormone-binding globulin	√	−	NT	↑	↑	↑
8	II	ICAM2	intercellular adhesion molecule 2	√	−	NT	↑	→	→
9	II	SPARCL1	SPARC-like 1	√	√	→	↑	→	↑
10	III	AFM	afamin	√	−	NT	NT	↑	↑
11	III	AHSG	alpha-2-HS-glycoprotein	√	√	NT	NT	↑	→
12	III	APOD	apolipoprotein D	√	−	NT	NT	↑	→
13	III	AZGP1	alpha-2-glycoprotein 1, zinc	√	−	NT	NT	↑	↑
14	III	ICOSLG	inducible T-cell costimulator ligand	√	√	↑	↑	NT	NT
15	III	ORM1/2	orosomucoid 1/2 (= AGP)	√	√	↑[b]	NT	→	↑
16	III	PIGR	polymeric immunoglobulin receptor	√	√	↑	NT	↑	↑
17	III	PTGDS	prostaglandin H2 D-isomerase	√	−	NT	NT	↑	→
18	III	SEPP1	selenoprotein P isoform 1 precursor	√	−	NT	↑	↑	↑
19	III	SERPINA1	serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1	√	√	NT	NT	→	↑
20	III	SERPINA3	serpin peptidase inhibitor, clade A, member 3	√	√	NT	NT	→	↑
21	III	TF	transferrin	√	√	NT	NT	↑	→

[a]Symbols: √, applicable; −, not applicable; NT, not tested; ND, not detected; ↑, enhanced in HCC; →, not enhanced in HCC. [b]See ref 32.

cancer-associated glycan alterations.[20−22] In addition, we showed that *N*-glycan fucosylation on AFP in the culture media was enhanced by comparison with that of placenta.[13] Therefore, we selected these lectins, AAL and DSA, to enrich HCC marker candidates. RCA120 recognizes Galβ1−4-GlcNAc-terminated glycans, which are present in the majority of serum glycoproteins after neuraminidase treatment.[23,24] Thus, RCA120 was used to comprehensively obtain glycopeptides from the samples in order to estimate their abundance.

Identification of Lectin-Captured Glycoproteins in the Culture Media and Serum Samples Using the IGOT-LC−MS Method

First, we collected glycopeptide subsets by AAL affinity chromatography from tryptic digests of the culture media of HepG2 and HuH-7. In order to evaluate the capturing specificity of the AAL column under the conditions employed, glycopeptides derived from the serum of an HCC patient were separated using the AAL column, and *N*-glycans released from AAL(+) and AAL(−) glycopeptides were analyzed by MALDI-TOF MS (Supplementary Figure 1 and Supplementary Method). As a result, almost all signals of AAL(+) glycans were assigned to have one or more deoxyhexose (dHex) residue(s) and only one slight signal having no dHex was observed, suggesting that the glycopeptides having fucosylated glycans were highly enriched in the AAL(+) fraction and that AAL has enough selectivity to capture glycopeptides having fucosylated glycans. However, because this result could not ensure that all individual glycopeptides were fucosylated, we dealt the AAL(+) glycopeptides as putative fucosylated glycopeptides. After purification by HILIC, the AAL(+) glycopeptides were treated with PNGase F in $H_2{}^{18}O$ and identified by LC−MS analysis followed by Mascot search. In total, 316 and 629 glycoproteins were identified from HepG2 and HuH-7, respectively (Figure 2a, Supplementary Tables 3 and 4). Similarly, we identified 122 AAL(+) glycoproteins from serum samples of 5 HCC patients. In total, 744 AAL(+) glycoproteins were identified as first-step biomarker candidates. Exemplary MS/MS spectra are shown in Supplementary Figure 2 to support the reliability of glycopeptide identification by the IGOT-LC−MS method. Next, we analyzed AAL(+) glycopeptides from HV sera and identified 100 glycoproteins. Finally, we used RCA120 to obtain glycopeptides from HV and HCC sera in order to estimate their abundance in serum and the number of *N*-glycosylated sites on each glycoprotein. Using this procedure, we identified 172 glycoproteins as abundant glycoproteins (Figure 2a).

The same experimental protocol allowed us to identify 154 and 172 DSA(+) glycoproteins from the culture media of HepG2 and HuH-7, respectively (Figure 2b). In addition, 84 DSA(+) glycoproteins were identified from HCC sera; in total, 234 glycoproteins were identified in HCC samples. Furthermore, 117 DSA(+) glycoproteins were identified from HV sera (Figure 2b). Details of the identified glycopeptides are given in Supplementary Table 3. Identified proteins in each sample are listed in Supplementary Tables 4 (AAL) and 5 (DSA). These results were registered to the glycoprotein database, GlycoProtDB, which was constructed using our experimental-based information and is

dx.doi.org/10.1021/pr301217b | J. Proteome Res. XXXX, XXX, XXX−XXX

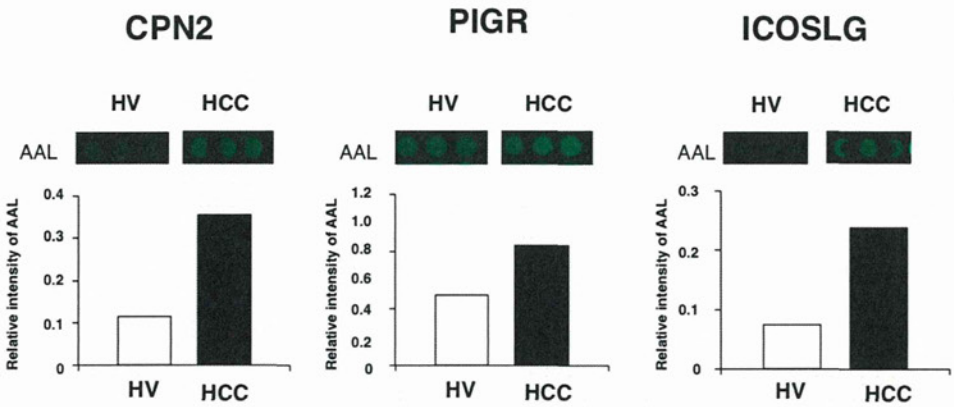

Figure 3. Detection of HCC-associated enhancement of fucosylation on the candidate glycoproteins by the immunoprecipitation-ALP method. Candidate glycoproteins were enriched from pooled sera of HV and HCC patients by immunoprecipitation with their antibodies and analyzed by lectin array (ALP method). Signal images for AAL spots (above, $n = 3$). Histogram indicating relative intensity of AAL to DSA (relative intensity of AAL signal against DSA signal).

currently available for mouse and *C. elegans* glycoproteins (http://jcggdb.jp/rcmg/gpdb/).[16] The overlap of AAL(+) and DSA(+) proteins was 209 (Figure 2c, Supplementary Table 6).

Selection of Biomarker Candidates for Further Verification by Glycoproteome Profiles and Availability of Their Antibodies

As described above, we identified 744 AAL(+) proteins in the media of HepG2 and HuH-7 and the serum samples of HCC patients. These proteins correspond to 398 gene symbols. Among them, 80 glycoproteins (gene symbols) are previously reported to have core-fucose in part or to bind with fucose-recognizing lectins.[25−28] The remaining 318 proteins are newly identified in this study as AAL-reactive proteins. Considering the selectivity of the AAL column under the employed conditions, as shown in Supplementary Figure 1, these are putative fucosylated glycoproteins that had not been formerly identified as fucosylated. Next, as a reference to select candidates to be verified, we classified the first-step AAL(+) candidates obtained from HCC samples into 3 groups by comparing the glycoproteome profiles as follows (Figure 2a). First, we compared these AAL(+) proteins with those in pooled HV serum (n = 100), and found 87 proteins in common. These common proteins were classified into Group III, whose *N*-glycans are also fucosylated in the healthy state. The other proteins were then compared with RCA120(+) serum proteins. This comparison identified 53 common proteins, suggesting that they are abundant in the serum. These proteins were classified into Group I. The remaining 604 proteins were designated as Group II. These proteins were detected only in the culture media, suggesting that they were less abundant in serum, but concentrated in the media. At this point, we minimized the number of proteins by using the gene symbols to avoid redundant identifications, such as splice variants and family proteins.

Next, we further selected the candidates on the basis of the tissue expression level and specificity for liver, as well as the availability of antibodies, which is essential to enrich or detect each candidate protein for verification. Moreover, glycoproteins supposedly produced by cells other than hepatocytes (e.g., immunoglobulins) were deleted from the list. We also eliminated some types of glycoproteins, such as complement factors and coagulation factors, as these were considered to be inappropriate for serological tests, because it is presumed easily that the level of complement components is influenced by inflammation in subject patients and the major parts of coagulation factors and fibrinogen are removed from serum samples during blood clotting (in preparation of serum from blood). Consequently, a total of 21 glycoproteins (gene symbols) were selected for further verification (Table 1). These candidates included glycoproteins known to have fucosylated glycans.[29,30] Nonetheless, we verified the glycan alteration between HCC and HV to confirm the effectiveness of our strategy.

The DSA(+) glycoproteins were also classified into 3 groups (Figure 2b) in the same manner as for the AAL(+) proteins. However, we did not select these candidates for further verification for the following reasons. First, although Group I (19) included glypican 3 (GPC3), which was reported to be a useful tumor marker of HCC and some other cancers,[31] other proteins were mostly coagulation factors, complement components, and fibrinogens, which are deemed unsuitable as described above. Thus, we could not find any suitable candidates in Group I of DSA(+) except GPC3. The DSA(+) candidates belonging to Group II (115) were identified only in the culture media, which means low abundance in serum. Furthermore, many of the DSA(+) proteins in Group III (100) were the same as those selected from the AAL(+) proteins. A further technical difficulty was that DSA(+) proteins were barely released from DSA-agarose in the preliminary experiments. Thus, subsequent verification of HCC-associated glycan alteration was performed only for the AAL(+) candidates.

Verification of HCC/LC-Associated Glycan Alteration (Enhanced Fucosylation) on the Candidate Glycoproteins

We verified HCC/LC-associated glycan alteration of the selected candidates using pooled sera of HV and HCC patients (with LC as background) by one or more of the following methods: (1) immunoprecipitation followed by lectin array analysis, (2) AAL-protein fractionation followed by Western blotting, and (3) LCA/AAL-serial lectin fractionation followed by Western blotting. Because these analyses require appropriate antibody for each protein, we tested commercially available antibodies using a pooled serum of HV (data not shown). Antibodies applicable for our verification analyses are listed in Supplementary Table 2.

dx.doi.org/10.1021/pr301217b | J. Proteome Res. XXXX, XXX, XXX−XXX

1. Immunoprecipitation Followed by Lectin Array Analysis. Enhancement of fucosylation associated with HCC was confirmed by purification of the candidate proteins using immunoprecipitation followed by antibody-assisted lectin profiling (ALP) analysis[17] Among the 8 candidate proteins analyzed by this method, carboxypeptidase N polypeptide 2 (CPN2), polymeric immunoglobulin receptor (PIGR) and inducible T-cell costimulator ligand (ICOSLG) showed an apparent increase of AAL signal in the HCC sera compared to HV. As shown in Figure 3, the AAL signal of HCC sera was increased compared with the respective HV sera by the following factors: 3.1-fold for CPN2, 1.7-fold for PIGR, and 4.1-fold for ICOSLG. Results of the remaining candidates are shown in Supplementary Figure 3a and summarized in Table 1. The number of candidate proteins applicable to the ALP analysis is limited due to the lack of suitable antibodies. However, the ALP analysis also provided the reactivity information against other lectins, which is useful for reselecting the most appropriate lectin(s) for detection of glycan alteration associated with HCC/LC (fibrosis) as exemplified by α1-acid glycoprotein (AGP or ORM1/2).[32]

2. AAL-Fractionation Followed by Western Blotting Analysis. We fractionated pooled sera of HV and HCC patients into AAL(+) and AAL(−) fractions using an AAL column. To estimate the amount of candidates in each fraction, the protein fractions and untreated sera were subjected to SDS-PAGE, and then each candidate was visualized by Western blotting. The levels of all 7 candidates tested by this analysis increased in the AAL(+) fraction of HCC sera compared with HV, suggesting enhanced fucosylation associated with HCC/LC. Specifically, as shown in Figure 4, sex hormone binding globulin (SHBG: Group I), colony stimulating factor 1 receptor (CSF1R: Group I), intracellular adhesion molecule 2 (ICAM2: Group II), and selenoprotein P (SEPP1: Group III) indicated significantly increased signals in AAL(+) of HCC sera compared with that of HV sera. Results of other 3 candidates are shown in Supplementary Figure 3b.

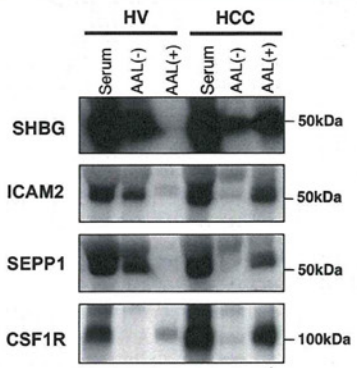

Figure 4. Detection of HCC-associated enhancement of fucosylation on the candidate glycoproteins by AAL-fractionation followed by Western blotting. Pooled sera of HV and HCC patients were fractionated with AAL-agarose. AAL(+), AAL(−), and untreated sera were separated by SDS-PAGE. The candidate glycoproteins were detected by Western blotting. Blot images of each candidate are shown.

3. LCA-AAL Serial Lectin Fractionation Followed by Western Blotting Analysis. As described above, branching and fucosylation of glycans are known to be enhanced in association with HCC.[20,22] Thus, measuring the glycoproteins that underwent both fucosylation and branching might be a clearer indication of glycan alteration by HCC compared to estimating the overall fucosylation level. In many cases, fucosylated glycoproteins are enriched by LCA or AAL; however, their specificities are different. LCA binds specifically to α1,6-fucosylated biantennary glycans, and AAL binds to fucosylated glycans regardless of branched structures.[33] Therefore, we separated the fucosylated candidate glycoproteins into 2 fractions by LCA-AAL serial lectin fractionation (refer to schematic shown in Figure 5a). First, a serum sample was applied to the LCA column to capture glycoproteins having α1,6-fucosylated biantennary glycans, and then the pass fraction (LCA(−)) was applied to the AAL column to recover the remaining glycoproteins having fucosylated glycans. Thus, by serial fractionation glycoproteins having both fucosylated and branched glycans are enriched in the LCA(−)/AAL(+) fraction. These fractions were separated by SDS-PAGE, and then each candidate protein was detected by Western blotting (Figure 5b). As a result, ORM1/2 (AGP), which is known to have highly branched glycans, was detected only in the LCA(−)/AAL(+) fraction (data not shown), indicating that the serial fractionation worked as intended. As summarized in Table 1, the ratio of the LCA(−)/AAL(+) fraction was higher in the HCC sera than HV sera for 12 of the 19 candidates tested (Figure 5b for SHBG, PIGR, CSF1R, and SEPP1; Supplementary Figure 3c for the remaining). These observations suggest that HCC/LC-associated increases of fucosylation and branching occur on the common glycans. In total, 16 of the 21 candidates were identified in both AAL(+) and DSA(+) fractions. Hence, glycans on a wide range of glycoproteins display both fucosylation and branching. However, the bands of proteins in the LCA(+) fraction also increased in HCC sera, suggesting that fucosylation associated with HCC/LC is independent of branching. Conversely, there were also proteins whose associated glycan was less highly branched or not branched at all. For 6 candidates, only the rate of the LCA(+) fraction was higher in the HCC sera (e.g., LAMP2 and SPARCL1). Thus, serial lectin fractionation may be useful for confirming the onset phase of fucosylation and branching individually in the course of chronic fibrosis, LC to HCC.

■ DISCUSSION

Currently, AFP, PIVKA-II (protein induced by vitamin K absence or antagonists-II, also known as des-γ-carboxy prothrombin [DCP]) and AFP-L3% (ratio of LCA-reactive AFP to total AFP) have been used for the diagnosis of HCC. However, the sensitivity and specificity of these markers are insufficient for diagnosis of early stage HCC.[34] Although many candidate serum biomarkers for HCC have been reported by proteomics- and/or glycoproteomics-based research, there is still a clinical need to develop more reliable markers for HCC.

However, it is fundamentally difficult to discover a serum biomarker for early stage cancer by comparing serum protein profiles between patients and healthy volunteers or disease controls. Empirically, we know that glycan alteration of particular cells is a good index for detecting a change of cell status such as differentiation and carcinogenesis. However, it is also difficult to detect early stage cancer by comparative profiling of the entire glycome of serum glycoproteins, i.e., by analyzing glycans released from serum glycoproteins *in vitro*. This is because the altered glycans emerge "ectopically" in association with carcinogenesis and thus do not have structures "specific to the cancerous cells". For example, sialyl Lewis X is known as a cancer-related antigen in colon,[35] non-small cell lung,[36] breast,[21] and pancreatic

dx.doi.org/10.1021/pr301217b | J. Proteome Res. XXXX, XXX, XXX−XXX

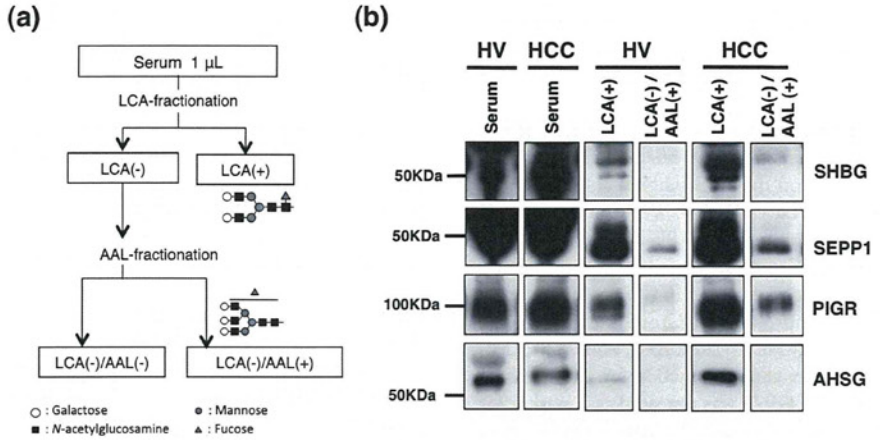

Figure 5. Detection of HCC-associated enhancement of fucosylation on the candidate glycoproteins by LCA/AAL-serial lectin chromatography followed by Western blotting. (a) Outline of the serial fractionation. Serum was fractionated with LCA-agarose. Glycoproteins in the LCA(+) fraction were presumed to have core-fucosylated biantennary glycans. Then, LCA(−) proteins were fractionated with AAL-agarose. Glycans of AAL(+) glycoproteins were presumed to be both fucosylated and branched. (b) Lectin-bound fractions were separated by SDS-PAGE, and then the candidate proteins in each fraction were detected by Western blotting.

cancers.[37] However, the glycan structure is also expressed by noncancerous cells such as endothelial cells of high endothelial venules (HEV) and lymphocytes associated with inflammation.[38] Similarly, although enhanced fucosylation is recognized as a glycan alteration related to HCC, fucosylation does not necessarily occur only in HCC, and fucosylated glycans are expressed ubiquitously. Therefore, it is important to detect glycan alteration on hepatic cell-specific proteins, especially when seeking a serum marker. In order to clarify glycan alteration associated with HCC, we first analyzed glycomes of the "culture media" of HCC cell lines (HepG2 and HuH-7) by lectin array as described previously.[13] Proteins in the culture media are undoubtedly secreted by HCC cells and carry glycans synthesized in the cancer cells. Thus, we used the culture media for the glycome analysis. In this sense, it is important to use a cell line that maintains *in vivo* characteristics of the cancer for the discovery procedure. In HCC, enhanced fucosylation and branching were reported in both cell lines and cancer tissues.[20,22] As expected, we found that the array signals of AAL and DSA in the media were very strong. Therefore, we selected the two lectins, AAL and DSA, to enrich glycoproteins carrying their cognate glycans, i.e., fucosylated and branched glycans, respectively. As exemplified by AFP-L3, which is AFP bearing α1,6-(core) fucosylated *N*-glycan, enhancement of fucosylation in HCC is a promising marker for cancer-associated glycan alterations.[39,40] Indeed, AFP-L3 is used as a U.S. Food and Drug Administration-approved HCC serum marker.

Neutral loss-triggered MS/MS (Electron Transfer Dissociation or Collision-Induced Dissociation) identification after partial deglycosylation with Endo F3 is a novel technology for assigning core fucose-bearing glycopeptides without employing lectin affinity. This method has been developed and used for marker discovery of HCC.[28,41] In general, however, glycoproteins having fucosylated *N*-glycans were enriched using fucose-recognizing lectins such as AAL, LCA, PSA, and VFA to capture not only the core fucose-bearing glycoproteins/peptides but also terminal α1,2/1,3-fucose-bearing molecules.[42]

Next, we identified AAL(+) and DSA(+) glycopeptides from the culture media of HCC cell lines to ensure they were actually produced by liver cancer cells. In total, 703 AAL(+) and 211 DSA(+) glycoproteins were identified from the media (Figure 2). These represent currently the largest data sets of human glycoproteins having a particular glycan motif (fucosylation or branch). It is well recognized that proteins obtained by affinity chromatography include those bound to the column nonspecifically or indirectly. For this reason, we captured glycopeptides, rather than glycoproteins, to minimize nonspecific identification. The IGOT method employed in this study also served to distinguish glycopeptides from non-*N*-glycosylated peptides with high precision.[15]

In addition, we also identified AAL(+) and DSA(+) glycopeptides from HCC sera. As the majority of serum glycoproteins are known to be derived from liver, serum samples are appropriate only for the marker discovery of hepatic diseases. We identified 703 and 122 AAL(+) glycoproteins from the culture media and the patients' sera, respectively. Roughly two-thirds of the AAL(+) glycoproteins from HCC sera (81/122) were also detected in the media, suggesting these glycoproteins were derived from liver cells. The remaining AAL(+) glycoproteins (41) were identified solely from the HCC sera, some of which are known to be primarily expressed in the liver. Thus, we selected 5 glycoproteins as candidates for further verification from the serum AAL(+) proteins.

Next, we compared the glycoproteome profiles as presented in Figure 2 to classify the candidates into three groups based on the following two factors. Due to the reasons outlined above, we confine our discussion to AAL(+) proteins only. The first factor is the degree of fucosylation under a physiological state. A preferable cancer marker is a glycoprotein having little or no fucosylated glycans in the healthy state. Thus, we divided the AAL(+) candidate proteins according to whether they were identified only in the HCC samples or in common with HV. Glycoproteins identified in common were classified into Group III (Figure 2). Another factor is the abundance of the core protein in serum. The RCA120(+) proteins identified from serum samples were thought to be highly abundant in serum. In addition, the number of RCA(+) peptides identified from each protein indicates the number of glycosylation sites on the target glycan. The more glycans the protein has, the easier the protein can be detected, for example, by sandwich ELISA using

dx.doi.org/10.1021/pr301217b | J. Proteome Res. XXXX, XXX, XXX–XXX

antibodies and lectins, because multivalent binding strengthens the affinity between glycans on the glycoprotein and lectin molecules. The AAL(+) proteins identified in HCC samples that were in common with RCA120(+) proteins were classified into Group I. The remaining proteins were designated into Group II as shown in Figure 2. From each group, we selected 21 candidates for further verification based on the availability of antibodies and their abundance in serum. These candidates were then verified by one or more of the subsequent methods (Table 1).

Among 7 candidates in Group I, 4 proteins were already known to have enhanced fucosylation in HCC by comparison to the healthy state.[25−28] Glycoproteins of Group I include AFP, GP73 (GOLM1), and GPC3. Fucosylated AFP, known as AFP-L3, is actually being used in clinical practice for diagnosis of HCC. Moreover, fucosylated GP73 was reported as a marker for HCC with higher sensitivity than AFP.[26,43] These results suggest our approach for identifying candidate markers has been successful. In this study, all selected candidates in Group I (including CPB2, CSF1R, LAMP2, and SHBG) were newly identified AAL(+) glycoproteins displaying enhanced levels of fucosylation (Table 1). Among them, we expected CSF1R and CPN2 to be promising candidate HCC markers. Specifically, in comparison with other candidates, CSF1R and CPN2 show lower fucosylation in HV and increased expression in the liver and have many potential *N*-glycosylation sites (7 in CPN2;[44] 9 in CSF1R[45]), which would facilitate detection by sandwich ELISA using their respective antibodies and AAL lectin.

Glycoproteins in Group III included many classical plasma proteins as represented by transferrin, hemopexin, alpha 2-macroglobulin, kininogen, ceruloplasmin, complement factors, and so on.[46] Among 11 selected candidates, all but ICOSLG were reported to show elevated fucosylation in HCC.[25] However, these proteins in Group III were suggested to have fucosylated glycans in healthy individuals. This also suggests that these Group III proteins might be biomarkers that reflect the status of the liver as a whole rather than a local change, i.e., carcinogenesis. Namely, if the ratio of fucosylation on a glycan of a candidate biomarker is 1% in the healthy state and enhanced to 100% in HCC, the amount of this protein having fucosylated glycan would increase only 2-fold at most in a patient with a small cancer, i.e., 1% by weight (1*100:1*99 + 100 = 1:2). Thus, as discussed above, these proteins may become surrogate fibrosis markers. These markers are also clinically desirable, because a quantitative index indicating the degree of fibrosis is important in identifying high-risk individuals, who should be referred for imaging and cytological diagnosis (biopsy), e.g., hepatitis patients infected by HBV or HCV. Moreover, such an approach will be useful for evaluating the benefit of treatment or medication in patients and for the development of new medicines for liver fibrosis.[47] In this sense, ORM1/2 (AGP), PIGR, SERPINA3, AZGP1, ICOSLG, PTGDS, and SEPP1 might be good markers for fibrosis. In fact, we found that glycan alteration on ORM1/2 correlated with progression of liver fibrosis, which can diagnose LC with an accuracy of over 90%.[5,48,49]

In Group II, the ratio of AAL(+) of 2 candidates, SPARCL1 and ICAM2, increased in HCC. Elevated expression of SPARCL1 is known to be associated with the progression of hepatic disease along with sinusoidal capillary transformation.[50] As the serum level of these candidates is low, improvement of detection sensitivity will be needed for practical use as biomarkers.

As demonstrated in this study, we could identify many glycoproteins carrying glycans with enhanced reactivity to AAL in HCC sera and hepatoma cell lines using our high-throughput strategy. Furthermore, this approach also enabled us to provide a large set of data on human glycoproteins having particular glycan motifs through our glycoprotein database, GlycoProtDB, http://jcggdb.jp/rcmg/gpdb/.

In this study, we also carried out a verification of glycan alteration between HCC patients and HV for many candidate proteins. The verification procedure involved three methods including the newly developed technique of serial lectin fractionation. Because we used sera of HCC patients having LC as samples for the discovery and verification, the glycan alteration is thought to be associated with HCC or LC. Therefore, for each marker candidate, it should be verified at which stage on the long-term pathological course toward HCC the enhancement of glycan alteration (fucosylation) occurs. Detailed comparison using sera of patients at various stages of the disease may allow us to identify stage-specific biomarkers. We have already verified that glycan alteration on ORM1/2 (AGP in Group III) can distinguish the fibrosis stage between F3 and F4 (LC) by immunopurification followed by a lectin array analysis using sera from over 100 patients.[32] Furthermore, subsequent validation studies of this marker using an automated bedside clinical chemistry analyzer in more than 1000 HBV/HCV-infected patients have demonstrated that the diagnostic accuracy for detecting LC is superior to that of other fibrosis indices or instruments.[5,48,49] Thus, we expect that a novel and clinically applicable glycobiomarker for early stage HCC can be selected using the stage-established samples in a future study. Furthermore, we believe that this approach is applicable to discover biomarkers not only for other cancers but also for fibrogenic diseases such as Alzheimer's disease and fibrosis of lung and kidney.

■ ASSOCIATED CONTENT

Supporting Information

This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel: +81-29-861-3200. Fax: +81-29-861-3201. E-mail: h.narimatsu@aist.go.jp.

Present Addresses

¶Institute of Biological Chemistry, Academia Sinica, 128 Academia Road, Sec. 2, Nankang, Taipei 11529, Taiwan.

§The Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Chiba 272−8516, Japan.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported in part by a grant of "Medical Glycomics Project" from New Energy and Industrial Technology Development Organization (NEDO). We thank Dr. Hiroaki Tateno, Ms. Jinko Murakami and Yoshiko Kubo for preparation of the lectin array; Ms. Sachiko Unno and Kozue Saitou for lectin array analyses; Ms. Nana Tanaka and Yoshie Takahashi and Mr. Norihito Hayatsu for technical assistance; Ms. Naoko Tetsura for

data processing of mass spectrometry, and Dr. Mika Fujita and Ms. Azusa Tomioka for MS analyses of glycans. We appreciate Drs. Takashi Sato, Hiromichi Sawaki, Masanori Gotoh and Hiromi Ito and Mr. Toshihide Shikanai for valuable discussion. We also thank Ms. Azumi Takahashi for assistance in the preparation of this manuscript. All persons are members of our research center.

REFERENCES

(1) Umemura, T.; Ichijo, T.; Yoshizawa, K.; Tanaka, E.; Kiyosawa, K. Epidemiology of hepatocellular carcinoma in Japan. *J. Gastroenterol.* **2009**, *44* (Suppl 19), 102–7.

(2) Yoshida, H.; Shiratori, Y.; Moriyama, M.; Arakawa, Y.; Ide, T.; Sata, M.; Inoue, O.; Yano, M.; Tanaka, M.; Fujiyama, S.; Nishiguchi, S.; Kuroki, T.; Imazeki, F.; Yokosuka, O.; Kinoyama, S.; Yamada, G.; Omata, M. Interferon therapy reduces the risk for hepatocellular carcinoma: national surveillance program of cirrhotic and noncirrhotic patients with chronic hepatitis C in Japan. IHIT Study Group. Inhibition of Hepatocarcinogenesis by Interferon Therapy. *Ann. Intern. Med.* **1999**, *131* (3), 174–81.

(3) Cadranel, J. F.; Rufat, P.; Degos, F. Practices of liver biopsy in France: results of a prospective nationwide survey. For the Group of Epidemiology of the French Association for the Study of the Liver (AFEF). *Hepatology* **2000**, *32* (3), 477–81.

(4) Castera, L.; Negre, I.; Samii, K.; Buffet, C. Pain experienced during percutaneous liver biopsy. *Hepatology* **1999**, *30* (6), 1529–30.

(5) Ito, K.; Kuno, A.; Ikehara, Y.; Sugiyama, M.; Saito, H.; Aoki, Y.; Matsui, T.; Imamura, M.; Korenaga, M.; Murata, K.; Masaki, N.; Tanaka, Y.; Hige, S.; Izumi, N.; Kurosaki, M.; Nishiguchi, S.; Sakamoto, M.; Kage, M.; Narimatsu, H.; Mizokami, M. LecT-hepa, a glyco-marker derived from multiple lectins, as a predictor of liver fibrosis in chronic hepatitis C patients. *Hepatology* **2012**, *56* (4), 1448–56.

(6) Wai, C. T.; Greenson, J. K.; Fontana, R. J.; Kalbfleisch, J. D.; Marrero, J. A.; Conjeevaram, H. S.; Lok, A. S. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology* **2003**, *38* (2), 518–26.

(7) Sato, M.; Tateishi, R.; Yasunaga, H.; Horiguchi, H.; Yoshida, H.; Matsuda, S.; Koike, K., Mortality and morbidity of hepatectomy, radiofrequency ablation, and embolization for hepatocellular carcinoma: a national survey of 54,145 patients. *J. Gastroenterol.* **2012**.

(8) Minami, Y.; Kudo, M. Radiofrequency ablation of hepatocellular carcinoma: a literature review. *Int. J. Hepatol.* **2011**, 104685.

(9) Takayama, T.; Makuuchi, M.; Hirohashi, S.; Sakamoto, M.; Yamamoto, J.; Shimada, K.; Kosuge, T.; Okada, S.; Takayasu, K.; Yamasaki, S. Early hepatocellular carcinoma as an entity with a high rate of surgical cure. *Hepatology* **1998**, *28* (5), 1241–6.

(10) Rifai, N.; Gillette, M. A.; Carr, S. A. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat. Biotechnol.* **2006**, *24* (8), 971–83.

(11) Tamura, Y.; Igarashi, M.; Kawai, H.; Suda, T.; Satomura, S.; Aoyagi, Y. Clinical advantage of highly sensitive on-chip immunoassay for fucosylated fraction of alpha-fetoprotein in patients with hepatocellular carcinoma. *Dig. Dis. Sci.* **2010**, *55* (12), 3576–83.

(12) Narimatsu, H.; Sawaki, H.; Kuno, A.; Kaji, H.; Ito, H.; Ikehara, Y. A strategy for discovery of cancer glyco-biomarkers in serum using newly developed technologies for glycoproteomics. *FEBS J.* **2010**, *277* (1), 95–105.

(13) Ito, H.; Kuno, A.; Sawaki, H.; Sogabe, M.; Ozaki, H.; Tanaka, Y.; Mizokami, M.; Shoda, J.; Angata, T.; Sato, T.; Hirabayashi, J.; Ikehara, Y.; Narimatsu, H. Strategy for glycoproteomics: identification of glyco-alteration using multiple glycan profiling tools. *J. Proteome Res.* **2009**, *8* (3), 1358–67.

(14) Kaji, H.; Yamauchi, Y.; Takahashi, N.; Isobe, T. Mass spectrometric identification of N-linked glycopeptides using lectin-mediated affinity capture and glycosylation site-specific stable isotope tagging. *Nat. Protoc.* **2006**, *1* (6), 3019–27.

(15) Kaji, H.; Saito, H.; Yamauchi, Y.; Shinkawa, T.; Taoka, M.; Hirabayashi, J.; Kasai, K.; Takahashi, N.; Isobe, T. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat. Biotechnol.* **2003**, *21* (6), 667–72.

(16) Kaji, H.; Shikanai, T.; Sasaki-Sawa, A.; Wen, H.; Fujita, M.; Suzuki, Y.; Sugahara, D.; Sawaki, H.; Yamauchi, Y.; Shinkawa, T.; Taoka, M.; Takahashi, N.; Isobe, T.; Narimatsu, H. Large-scale identification of N-glycosylated proteins of mouse tissues and construction of a glycoprotein database, GlycoProtDB. *J. Proteome Res.* **2012**, *11* (9), 4553–66.

(17) Kuno, A.; Kato, Y.; Matsuda, A.; Kaneko, M. K.; Ito, H.; Amano, K.; Chiba, Y.; Narimatsu, H.; Hirabayashi, J. Focused differential glycan analysis with the platform antibody-assisted lectin profiling for glycan-related biomarker verification. *Mol. Cell. Proteomics* **2009**, *8* (1), 99–108.

(18) Matsumura, K.; Higashida, K.; Ishida, H.; Hata, Y.; Yamamoto, K.; Shigeta, M.; Mizuno-Horikawa, Y.; Wang, X.; Miyoshi, E.; Gu, J.; Taniguchi, N. Carbohydrate binding specificity of a fucose-specific lectin from Aspergillus oryzae: a novel probe for core fucose. *J. Biol. Chem.* **2007**, *282* (21), 15700–8.

(19) Green, E. D.; Brodbeck, R. M.; Baenziger, J. U. Lectin affinity high-performance liquid chromatography: interactions of N-glycanase-released oligosaccharides with leukoagglutinating phytohemagglutinin, concanavalin A, Datura stramonium agglutinin, and Vicia villosa agglutinin. *Anal. Biochem.* **1987**, *167* (1), 62–75.

(20) Moriwaki, K.; Miyoshi, E. Fucosylation and gastrointestinal cancer. *World J. Hepatol.* **2010**, *2* (4), 151–61.

(21) Listinsky, J. J.; Siegal, G. P.; Listinsky, C. M. The emerging importance of alpha-L-fucose in human breast cancer: a review. *Am. J. Transl. Res.* **2011**, *3* (4), 292–322.

(22) Guo, H. B.; Zhang, Y.; Chen, H. L. Relationship between metastasis-associated phenotypes and N-glycan structure of surface glycoproteins in human hepatocarcinoma cells. *J. Cancer Res. Clin. Oncol.* **2001**, *127* (4), 231–6.

(23) Shinohara, Y.; Sota, H.; Kim, F.; Shimizu, M.; Gotoh, M.; Tosu, M.; Hasegawa, Y. Use of a biosensor based on surface plasmon resonance and biotinyl glycans for analysis of sugar binding specificities of lectins. *J. Biochem.* **1995**, *117* (5), 1076–82.

(24) Itakura, Y.; Nakamura-Tsuruta, S.; Kominami, J.; Sharon, N.; Kasai, K.; Hirabayashi, J. Systematic comparison of oligosaccharide specificity of Ricinus communis agglutinin I and Erythrina lectins: a search by frontal affinity chromatography. *J. Biochem.* **2007**, *142* (4), 459–69.

(25) Comunale, M. A.; Wang, M.; Hafner, J.; Krakover, J.; Rodemich, L.; Kopenhaver, B.; Long, R. E.; Junaidi, O.; Bisceglie, A. M.; Block, T. M.; Mehta, A. S. Identification and development of fucosylated glycoproteins as biomarkers of primary hepatocellular carcinoma. *J. Proteome Res.* **2009**, *8* (2), 595–602.

(26) Drake, R. R.; Schwegler, E. E.; Malik, G.; Diaz, J.; Block, T.; Mehta, A.; Semmes, O. J. Lectin capture strategies combined with mass spectrometry for the discovery of serum glycoprotein biomarkers. *Mol. Cell. Proteomics* **2006**, *5* (10), 1957–67.

(27) Jia, W.; Lu, Z.; Fu, Y.; Wang, H. P.; Wang, L. H.; Chi, H.; Yuan, Z. F.; Zheng, Z. B.; Song, L. N.; Han, H. H.; Liang, Y. M.; Wang, J. L.; Cai, Y.; Zhang, Y. K.; Deng, Y. L.; Ying, W. T.; He, S. M.; Qian, X. H. A strategy for precise and large scale identification of core fucosylated glycoproteins. *Mol. Cell. Proteomics* **2009**, *8* (5), 913–23.

(28) Chen, R.; Wang, F.; Tan, Y.; Sun, Z.; Song, C.; Ye, M.; Wang, H.; Zou, H. Development of a combined chemical and enzymatic approach for the mass spectrometric identification and quantification of aberrant N-glycosylation. *J. Proteomics* **2012**, *75* (5), 1666–74.

(29) Hadziyannis, E.; Sialevris, K.; Georgiou, A.; Koskinas, J. Analysis of serum alpha-fetoprotein-L3% and des-gamma carboxyprothrombin markers in cases with misleading hepatocellular carcinoma total alpha-fetoprotein levels. *Oncol. Rep.* **2012**, No. 23174906.

(30) Biondi, A.; Malaguarnera, G.; Vacante, M.; Berretta, M.; D'Agata, V.; Malaguarnera, M.; Basile, F.; Drago, F.; Bertino, G. Elevated serum levels of Chromogranin A in hepatocellular carcinoma. *BMC Surg.* **2012**, *12* (Suppl 1), S7.

(31) Mounajjed, T.; Zhang, L.; Wu, T. T. Glypican-3 expression in gastrointestinal and pancreatic epithelial neoplasms. *Hum. Pathol.* **2012**, *44* (4), 542–50.

(32) Kuno, A.; Ikehara, Y.; Tanaka, Y.; Angata, T.; Unno, S.; Sogabe, M.; Ozaki, H.; Ito, K.; Hirabayashi, J.; Mizokami, M.; Narimatsu, H. Multilectin assay for detecting fibrosis-specific glyco-alteration by means of lectin microarray. *Clin. Chem.* **2011**, *57* (1), 48–56.
(33) Tateno, H.; Nakamura-Tsuruta, S.; Hirabayashi, J. Comparative analysis of core-fucose-binding lectins from Lens culinaris and Pisum sativum using frontal affinity chromatography. *Glycobiology* **2009**, *19* (5), 527–36.
(34) Masuzaki, R.; Karp, S. J.; Omata, M. New serum markers of hepatocellular carcinoma. *Semin. Oncol.* **2012**, *39* (4), 434–9.
(35) Yusa, A.; Miyazaki, K.; Kimura, N.; Izawa, M.; Kannagi, R. Epigenetic silencing of the sulfate transporter gene DTDST induces sialyl Lewisx expression and accelerates proliferation of colon cancer cells. *Cancer Res.* **2010**, *70* (10), 4064–73.
(36) Mizuguchi, S.; Inoue, K.; Iwata, T.; Nishida, T.; Izumi, N.; Tsukioka, T.; Nishiyama, N.; Uenishi, T.; Suehiro, S. High serum concentrations of Sialyl Lewisx predict multilevel N2 disease in non-small-cell lung cancer. *Ann. Surg. Oncol.* **2006**, *13* (7), 1010–8.
(37) Peracaula, R.; Tabares, G.; Lopez-Ferrer, A.; Brossmer, R.; de Bolos, C.; de Llorens, R. Role of sialyltransferases involved in the biosynthesis of Lewis antigens in human pancreatic tumour cells. *Glycoconjugate J.* **2005**, *22* (3), 135–44.
(38) Lee, J. K.; Bistrup, A.; van Zante, A.; Rosen, S. D. Activities and expression pattern of the carbohydrate sulfotransferase GlcNAc6ST-3 (I-GlcNAc6ST): functional implications. *Glycobiology* **2003**, *13* (4), 245–54.
(39) Taketa, K.; Ichikawa, E.; Sakuda, H.; Iwamasa, T.; Hayakawa, M.; Taga, H.; Hirai, H. Lectin reactivity of alpha-fetoprotein in a case of renal cell carcinoma. *Tumour Biol.* **1989**, *10* (5), 275–80.
(40) Medical devices; immunology and microbiology devices; classification of AFP-L3% immunological test systems. Final rule. *Fed. Regist.* **2005**, *70* (191), 57748–50.
(41) Zhao, Y.; Jia, W.; Wang, J.; Ying, W.; Zhang, Y.; Qian, X. Fragmentation and site-specific quantification of core fucosylated glycoprotein by multiple reaction monitoring-mass spectrometry. *Anal. Chem.* **2011**, *83* (22), 8802–9.
(42) Comunale, M. A.; Lowman, M.; Long, R. E.; Krakover, J.; Philip, R.; Seeholzer, S.; Evans, A. A.; Hann, H. W.; Block, T. M.; Mehta, A. S. Proteomic analysis of serum associated fucosylated glycoproteins in the development of primary hepatocellular carcinoma. *J. Proteome Res.* **2006**, *5* (2), 308–15.
(43) Marrero, J. A.; Feng, Z.; Wang, Y.; Nguyen, M. H.; Befeler, A. S.; Roberts, L. R.; Reddy, K. R.; Harnois, D.; Llovet, J. M.; Normolle, D.; Dalhgren, J.; Chia, D.; Lok, A. S.; Wagner, P. D.; Srivastava, S.; Schwartz, M. Alpha-fetoprotein, des-gamma carboxyprothrombin, and lectin-bound alpha-fetoprotein in early hepatocellular carcinoma. *Gastroenterology* **2009**, *137* (1), 110–8.
(44) Tan, F.; Weerasinghe, D. K.; Skidgel, R. A.; Tamei, H.; Kaul, R. K.; Roninson, I. B.; Schilling, J. W.; Erdos, E. G. The deduced protein sequence of the human carboxypeptidase N high molecular weight subunit reveals the presence of leucine-rich tandem repeats. *J. Biol. Chem.* **1990**, *265* (1), 13–9.
(45) Wang, Z. E.; Myles, G. M.; Brandt, C. S.; Lioubin, M. N.; Rohrschneider, L. Identification of the ligand-binding regions in the macrophage colony-stimulating factor receptor extracellular domain. *Mol. Cell. Biol.* **1993**, *13* (9), 5348–59.
(46) Mitchell, P. Proteomics retrenches. *Nat. Biotechnol.* **2010**, *28* (7), 665–70.
(47) Rosenberg, W. M.; Voelker, M.; Thiel, R.; Becka, M.; Burt, A.; Schuppan, D.; Hubscher, S.; Roskams, T.; Pinzani, M.; Arthur, M. J. Serum markers detect the presence of liver fibrosis: a cohort study. *Gastroenterology* **2004**, *127* (6), 1704–13.
(48) Kuno, A.; Ikehara, Y.; Tanaka, Y.; Saito, K.; Ito, K.; Tsuruno, C.; Nagai, S.; Takahama, Y.; Mizokami, M.; Hirabayashi, J.; Narimatsu, H. LecT-Hepa: A triplex lectin-antibody sandwich immunoassay for estimating the progression dynamics of liver fibrosis assisted by a bedside clinical chemistry analyzer and an automated pretreatment machine. *Clin. Chim. Acta* **2011**, *412* (19–20), 1767–72.
(49) Du, D.; Zhu, X.; Kuno, A.; Matsuda, A.; Tsuruno, C.; Yu, D.; Zhang, Y.; Ikehara, Y.; Tanaka, Y.; Zhang, X.; Narimatsu, H. Comparison of LecT-Hepa and FibroScan for assessment of liver fibrosis in hepatitis B virus infected patients with different ALT levels. *Clin. Chim. Acta* **2012**, *413* (21–22), 1796–9.
(50) Lau, C. P.; Poon, R. T.; Cheung, S. T.; Yu, W. C.; Fan, S. T. SPARC and Hevin expression correlate with tumour angiogenesis in hepatocellular carcinoma. *J. Pathol.* **2006**, *210* (4), 459–68.

Research Article

EASL EUROPEAN ASSOCIATION FOR THE STUDY OF THE LIVER | JOURNAL OF HEPATOLOGY

Effect of nucleos(t)ide analogue therapy on hepatocarcinogenesis in chronic hepatitis B patients: A propensity score analysis

Takashi Kumada[1,*], Hidenori Toyoda[1], Toshifumi Tada[1], Seiki Kiriyama[1], Makoto Tanikawa[1], Yasuhiro Hisanaga[1], Akira Kanamori[1], Takurou Niinomi[1], Satoshi Yasuda[1], Yusuke Andou[1], Kenta Yamamoto[1], Junko Tanaka[2]

[1]*Department of Gastroenterology and Hepatology, Ogaki Municipal Hospital, Ogaki, Japan;* [2]*Department of Epidemiology, Infectious Disease Control and Prevention, Graduate School of Biomedical Sciences, Hiroshima University, Hiroshima, Japan*

Background & Aims: Some patients with chronic hepatitis B virus (HBV) infection progress to hepatocellular carcinoma (HCC). However, the long-term effect of nucleos(t)ide analogue (NA) therapy on progression to HCC is unclear.

Methods: Therefore, we compared chronic hepatitis B patients who received NA therapy to those who did not, using a propensity analysis.

Results: Of 785 consecutive HBV carriers between 1998 and 2008, 117 patients who received NA therapy and 117 patients who did not, were selected by eligibility criteria and propensity score matching. Factors associated with the development of HCC were analyzed. In the follow-up period, HCC developed in 57 of 234 patients (24.4%). Factors significantly associated with the incidence of HCC, as determined by Cox proportional hazards models, include higher age (hazard ratio, 4.36 [95% confidence interval, 1.33–14.29], $p = 0.015$), NA treatment (0.28 [0.13–0.62], $p = 0.002$), basal core promoter (BCP) mutations (12.74 [1.74–93.11], $p = 0.012$), high HBV core-related antigen (HBcrAg) (2.77 [1.07–7.17], $p = 0.036$), and high gamma glutamyl transpeptidase levels (2.76 [1.49–5.12], $p = 0.001$).

Conclusions: NA therapy reduced the risk of HCC compared with untreated controls. Higher serum levels of HBcrAg and BCP mutations are associated with progression to HCC, independent of NA therapy.

Introduction

An estimated 350 million individuals worldwide are chronically infected with hepatitis B virus (HBV), of whom 1 million die annually from HBV-related liver disease [1]. Chronic HBV infection is recognized as a major risk factor for the development of hepatocellular carcinoma (HCC) [1,2]. Hepatitis B surface antigen (HBsAg)-positive patients have a 70-fold increased risk of developing HCC compared to HBsAg seronegative counterparts [3,4]. HBV infection is endemic in Southeast Asia, China, Taiwan, Korea, and sub-Saharan Africa, where up to 85–95% of patients with HCC are HBsAg positive [5]. HCC is the third and fifth leading cause of cancer death in men and women, respectively, and the number of deaths and the mortality rate from HCC have greatly increased in Japan since 1975 [6]. Hepatitis C virus (HCV)-related HCC accounts for 75% of all HCCs in Japan and HBV-related HCC accounts for 15% [6].

In 2004, Liaw *et al.* reported a significant reduction in HCC in 651 adults receiving lamivudine after adjustment for baseline variables (hazard ratio, 0.49 [95% confidence interval (95% CI), 0.25–0.99], $p = 0.047$) [7]. However, the results were not significant after exclusion of 5 patients who developed HCC within 1 year of randomization (0.47 [0.22–1.00], $p = 0.052$). Therefore, in 2009, the National Institutes of Health Consensus Development Conference concluded that there was insufficient evidence to assess whether nucleos(t)ide analogue (NA) therapy can prevent the development of HCC [8].

The long-term use of lamivudine has not been recommended because of tyrosine–methionine–aspartate–aspartate (YMDD) mutations, which have occasionally been associated with severe and even fatal flares of hepatitis [9,10]. Therefore, adefovir dipivoxil should be added immediately in patients with virological or biochemical breakthroughs or no response. Currently, there are 2 nucleoside agents (lamivudine, entecavir) and 1 nucleotide agent (adefovir dipivoxil) available for treatment of HBV infection in Japan. The agent with the higher genetic barrier to resistance, entecavir, is considered the initial drug of choice [11]. Recently, 3 studies on lamivudine suggested that long-term sustained viral suppression was associated with a reduced likelihood of developing HCC [12–14].

In this study, we sought to determine if NA therapy was associated with a reduction in the development of HCC. Since the validity of treatment effects in observational studies may be limited by selection bias and confounding factors, we performed a propensity analysis [15].

Keywords: HBcrAg; BCP; Gamma-GTP; Average integration value; HBV DNA.
Received 25 July 2012; received in revised form 17 October 2012; accepted 22 October 2012; available online 30 October 2012

* Corresponding author. Address: Department of Gastroenterology and Hepatology, Ogaki Municipal Hospital, 4-86, Minaminokawa-cho, Ogaki, Gifu 503-8052, Japan. Tel.: +81 584 81 3341; fax: +81 584 75 5715.
E-mail address: hosp3@omh.ogaki.gifu.jp (T. Kumada).

Abbreviations: HCC, hepatocellular carcinoma; HBV, hepatitis B virus; NA, nucleos(t)ide analogue; HBcrAg, HBV core-related antigen; BCP, basal core promoter; gamma-GTP, gamma glutamyl transpeptidase.

Research Article

Materials and methods

Patient selection

The study protocol was approved by the Institutional Ethics Committee of Ogaki Municipal Hospital in January 2011, and was in compliance with the Declaration of Helsinki. Written informed consent for the use of stored serum samples for the study was obtained from all patients.

Between 1998 and 2008, 1220 consecutive HBsAg-positive patients, who visited the Department of Gastroenterology and Hepatology at Ogaki Municipal Hospital, were prospectively enrolled in our HCC surveillance program. Of these, 785 patients met the following inclusion criteria: HBsAg positive for more than 6 months, no evidence of HCV co-infection, exclusion of other causes of chronic liver disease (alcohol consumption >80 g/day, hepatotoxic drugs, autoimmune hepatitis, primary biliary cirrhosis, hemochromatosis, or Wilson's disease), follow-up duration of greater than 3 years, no evidence of HCC for at least 1 year from the start of the follow-up period, receiving no interferon treatment, and receiving NA therapy for more than 1 year before the detection of HCC (**Fig. 1**). In patients on NA therapy, the date of NA therapy initiation was considered the starting point of the follow-up period.

Of these 785 patients, 148 received NA therapy (NA group) and 637 patients did not receive NA therapy (non-NA group) during the follow-up period. To reduce the confounding effects of covariates, we used propensity scores to match NA patients to unique non-NA patients. Six covariates including age, sex, HBV DNA concentration, hepatitis B e antigen (HBeAg), platelet count, and alanine aminotransferase (ALT) activity were taken into account at the start of follow-up. We computed the propensity score by using logistic regression with the independent variable including age (⩽40 years or >40 years), sex (female or male), HBV DNA concentration (⩽5.0 log copies/ml or >5.0 log copies/ml), HBeAg (negative or positive), platelet count (>150 × 10^3/m^3 or ⩽150 × 10^3/m^3), and ALT activity (⩽40 IU/ml or >40 IU/ml), as shown in previous reported cut-off values according to the indication for NA therapy [16–19]. This model yielded a *c* statistic of 0.85 (95% confidence interval [CI], 0.82–0.88), indicating very good ability of the propensity score model to predict treatment status. We sought to match each patient who received NA therapy to a patient who did not receive NA therapy, having a propensity by using greedy 5–1 digit matching [20]. Once this threshold was exceeded, a patient with NA therapy was excluded. This score ranged from 0.09198 to 0.98967 and, in effect, represented the probability that a patient would be receiving NA. We were able to match 117 patients with NA therapy to 117 unique patients without NA therapy. The follow-up period ended on 31 December, 2011 or the date when HCC occurrence was identified.

Surveillance and diagnosis

All patients were followed up at our hospital at least every 6 months. During each follow-up examination, platelet count, ALT, gamma glutamyl transpeptidase (gamma-GTP), total bilirubin, alkaline phosphatase (ALP), albumin, and alpha-fetoprotein (AFP) levels were measured. We used commercially available kits to test blood samples for HBsAg, HBeAg, and anti-HBe (Abbott Japan Co., Ltd., Tokyo, Japan). Before November 2007, the serum HBV DNA concentration was monitored by a polymerase chain reaction assay (COBAS Amplicor HBV monitor test, Roche Diagnostics K. K., Tokyo, Japan) with a lower detection limit of approximately 2.6 log copies/ml, and after December 2007, it was monitored with another polymerase chain reaction assay (COBAS AmpliPrep-COBAS TaqMan HBV Test, Roche Diagnostics K. K.), with a lower detection limit of approximately 2.1 log copies/ml. HBV genotyping was performed as described previously [21]. Serum levels of HBV core-related antigen (HBcrAg) were measured using a chemiluminescence enzyme immunoassay (CLEIA) as described previously [22,23]. Precore nucleotide 1896 and basal core promoter (BCP) dinucleotide 1762/1764 were determined using the line probe assay (INNO-LiPA HBV PreCore assay; Innogenetics NV) [24,25]. The probes were designed to determine the nucleotides at position 1896 (G *vs.* A) in the precore region and positions 1762 (A *vs.* T) and 1764 (G *vs.* A and G *vs.* T) in the BCP region. A line probe assay was used to identify any emergence of YMDD mutations (INNO-LiPA HBV DR assay; Innogenetics NV).

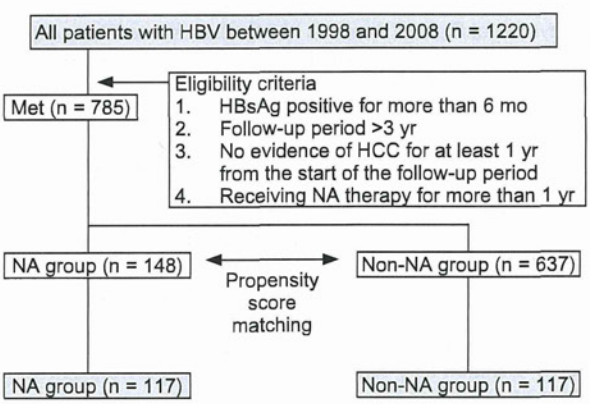

Fig. 1. Flowchart of the patient selection process.

Platelet count, ALT, gamma-GTP, total bilirubin, ALP, albumin, AFP, and HBV DNA values were expressed as average integration values [26,27] after the start of follow-up.

According to the Clinical Practice Guidelines for Hepatocellular Carcinoma in Japan [28], we performed ultrasound (US) and monitoring of 3 biomarkers (AFP, *Lens culinaris* agglutinin-reactive fraction of alpha-fetoprotein [AFP-L3], and des-gamma-carboxy prothrombin [DCP]) every 3–4 months, and dynamic magnetic resonance imaging (MRI) every 12 months, for patients with cirrhosis under surveillance. For patients with chronic hepatitis, we performed US and monitoring of the 3 biomarkers every 6 months. Histological examinations were performed in 91 out of 234 patients. Among them, cirrhosis was diagnosed in 32 patients. In the remaining 143 patients, the diagnosis of cirrhosis was made according to typical US findings, e.g., superficial nodularity, a coarse parenchymal echo pattern, and signs of portal hypertension (splenomegaly >120 mm, dilated portal vein diameter >12 mm, patent collateral veins, or ascites) [29–31]. Patients who did not satisfy these criteria were classified as having chronic hepatitis. One hundred and forty-two patients were diagnosed with chronic hepatitis and 92 patients with cirrhosis. For diagnostic confirmation of HCC, patients underwent dynamic MRI. A histological diagnosis of HCC was made in 28 patients (surgical specimen, 23 patients; US-guided needle biopsy specimen, 5 patients). The remaining 29 patients were diagnosed with HCC based on typical dynamic MRI findings, including hypervascularity in the arterial phase with washout in the portal venous or delayed phase [32].

Treatments

In the NA group, 117 patients received NA therapy including 18 patients with lamivudine, 28 patients with lamivudine and adefovir dipivoxil, and 71 patients with entecavir. The indications for NA therapy followed the guidelines of the American Association for the Study of Liver Diseases (AASLD), the European Association for the Study of the Liver (EASL), or the Asian Pacific Association for the Study of the Liver (APASL) [33–35]. In contrast, of the 117 patients not on NA therapy, 104 did not receive treatment before NA was not yet approved in Japan and the remaining 13 patients declined NA therapy.

Statistical analysis

Continuous variables are expressed as medians (range). The Mann–Whitney *U* test was used for continuous variables, and the Chi-square test with Yates' correction or Fisher's exact test was used for categorical variables. Actuarial analysis of the cumulative incidence of hepatocarcinogenesis was performed using the Kaplan–Meier method, and differences were tested with the log-rank test. The Cox proportional hazards model and the forward selection method were used to estimate the relative risk of HCC associated with age (⩽40 years or >40 years), sex (female or male), treatment (NA or no NA), HBsAg (⩽3.0 log IU/ml or >3.0 log IU/ml), HBV DNA level (⩽5.0 log copies/ml or >5.0 log copies/ml), HBeAg (negative or positive), precore region (wild type or mutant), BCP (wild type or mutant type), HBcrAg (⩽3.0 logU/ml or >3.0 logU/ml), platelet count (>150 × 10^3/m^3 or ⩽150 × 10^3/m^3), ALT (⩽40 IU/ml or >40 IU/ml), total bilirubin, gamma-GTP, ALP, albumin, and AFP (⩽10 ng/ml or >10 ng/ml) for univariate and multivariate analyses. We used the minimum or maximum of the reference values at our institution as cut-off values for total bilirubin, gamma-GTP, ALP, and albumin. We conducted a sensitivity analysis to determine the magnitude of an unmeasured confounder [36].

We considered *p* values of 0.05 or less to be significant. Statistical analysis was performed with SPSS, version 18.0 for Windows (International Business Machines Corporation, Tokyo, Japan).

JOURNAL OF **HEPATOLOGY**

Table 1. Baseline characteristics of all patients.

	NA group (n = 148)	Non-NA group (n = 637)	*p* value	Standardized difference in %
Age (yr)	53 (26-81)	48 (4-85)	<0.0001	40.6
Sex (female/male)	60/88	285/352	0.5378	6.1
Genotype (A/B/C/D/F/n.d.)	2/5/137/0/1/2	24/60/389/2/0/162	<0.0001	37.6
HBsAg (log_{10} IU/ml)	3.5 (-0.1-5.5)	3.3 (-1.3-7.9)	<0.0001	53.8
HBV DNA (log_{10} copies/ml)	7.0 (2.6-9.6)	3.8 (2.3-9.9)	<0.0001	99.9
HBeAg (±)	76/72	151/486	<0.0001	62.8
Precore region (W/M/n.d.)	30/109/9	88/381/168	0.4652	0.0
BCP (W/M/n.d.)	33/123/10	135/279/205	0.0074	27.3
HBcrAg (log_{10} U/ml)	5.9 (2.9-7.0)	3.0 (2.9-7.0)	<0.0001	96.7
Platelet count (x10³/m³)	150 (32-388)	188 (37-503)	<0.0001	-59.7
ALT (IU/ml)	65 (7-1088)	26 (5-3410)	<0.0001	44.1
AFP (ng/ml)	3.9 (0.8-3363)	2.9 (0.8-3686)	0.0062	-6.2
Cirrhosis (presence/absence)	62/86	91/546	<0.0001	59.1
Child-Pugh classification (A/B)	132/16	618/19	0.0002	32.7
Follow-up duration (yr)	12.8 (3.1-19.6)	13.7 (3.1-20.0)	0.1565	-16.9
Administration period (yr)	6.5 (1.5-11.0)	-	-	-
Propensity score	0.58093 (0.09198-0.98686)	0.95253 (0.12913-0.98967)	<0.0001	-132.3

NA, nucleos(t)ide analogue; n.d., not done; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; HBeAg, hepatitis B e antigen; W, wild type; M, mutant type; BCP, basal core promoter; HBcrAg, hepatitis B core-related antigen; ALT, alanine aminotransferase; AFP, alpha-fetoprotein; Child-Pugh classification, reference no [50]. Standardized difference in%; $100(X_{NA} - X_{non\text{-}NA})/([S^2_{NA} + S^2_{non\text{-}NA}]/2)^{1/2}$, where for each covariate X_{NA} and $X_{non\text{-}NA}$ are the sample means in NA and non-NA groups, respectively, and S^2_{NA} and $S^2_{non\text{-}NA}$ are the corresponding sample variances.

Results

Patient characteristics

Table 1 shows baseline characteristics of all 785 patients before propensity matching. There were significant differences in age, HBV genotype, HBsAg, HBV DNA concentration, presence of HBeAg, BCP mutations, HBcrAg, platelet counts, ALT level, AFP level, presence of cirrhosis, and Child-Pugh classification. The baseline characteristics of the 234 study patients after propensity matching are summarized in Table 2. There are no significant differences in age, sex, HBV genotype, HBsAg, HBV DNA concentration, presence of HBeAg, precore region mutations, BCP mutations, platelet counts, ALT concentration, Child-Pugh classification, and follow-up duration. HBcrAg concentration was significantly higher in the NA group than in the non-NA group. NA was administered a median of 6.1 years (range: 1.5–10.7 years).

Factors associated with the incidence of hepatocarcinogenesis

Factors associated with the incidence of HCC as determined by the Cox proportional hazard models and the forward selection method were analyzed in all 785 patients. High age (hazard ratio, 6.43 [95% CI, 2.71–15.26], p <0.001), male sex (3.43 [1.67–7.02], p = 0.002), NA treatment (0.28 [0.21–0.85], p = 0.017), BCP mutation (19.96 [2.27–141.90], p = 0.03), high HBcrAg levels (8.21 [3.40–19.85], p <0.001), and high AFP levels (2.49 [1.43–4.34], p = 0.001) were significantly associated with the incidence of HCC.

HCC developed in 57 of 234 patients (24.4%) during follow-up after propensity matching. The 5-year, 7-year, and 10-year cumulative incidences of HCC were 9.6%, 20.4%, and 33.4%, respectively. The 5-year, 7-year, and 10-year cumulative incidences of HCC were 2.7%, 3.3%, and 3.3%, respectively, in patients on NA therapy (n = 117) and 11.3%, 26.0%, and 40.0% in patients not on NA therapy (n = 117). Hepatocarcinogenesis occurred at significantly higher rates in the non-NA group (p = 0.0094, Fig. 2). The 5-year, 7-year, and 10-year cumulative incidences of HCC were 0.0%, 0.0%, and 0.0%, respectively, in patients with wild type BCP (n = 38) and 11.0%, 25.2%, and 41.9% in patients with mutant BCP (n = 112; p = 0.0006, Fig. 3). Factors associated with the incidence of HCC as determined by the Cox proportional hazard models and the forward selection method are listed in Table 3. Higher age (hazard ratio, 4.36 [95% CI, 1.33–14.29], p = 0.015), NA treatment (0.28 [0.13–0.62], p = 0.002), BCP mutation (12.74 [1.74–93.11], p = 0.012), high HBcrAg levels (2.77 [1.07–7.17], p = 0.036), and high gamma-GTP levels (2.76 [1.49–5.12], p = 0.001) were significantly associated with the incidence of HCC. In addition, 2 patients died due to hepatic failure during the follow-up period in the non-NA group.

The sensitivity analysis found that the observed relationship between NA treatment and HCC incidence could be diminished by the unmeasured confounder that the high prevalence of the unmeasured confounder is greater in the non-NA group than in the NA group. For example, suppose a binary unmeasured confounder that increased the hazard of HCC incidence (hazard ratio, 1.50) was present in 40% of those who were treated with NA and 80% of those who were not treated with NA. Then, the study's result would become less extreme and would no longer be statistically significant (hazard ratio under sensitivity analysis, 0.48 [95% CI, 0.22–1.05]).

Follow-up data of various parameters in patients on or not on NA therapy

For this analysis, we used the average integration value during the follow-up period (Table 4). ALT, gamma-GTP, ALP, AFP, and

Research Article

Table 2. Baseline characteristics of patients on NA therapy and propensity-matched controls.

	NA group (n = 117)	Non-NA group (n = 117)	*p* value	Standardized difference in %
Age (yr)	52 (27-77)	52 (21-77)	0.9223	1.7
Sex (female/male)	44/73	45/72	0.8929	6.1
Genotype (A/B/C/n.d.)	1/4/109/3	4/7/85/21	0.1232	26.8
HBsAg (log_{10} IU/ml)	3.6 (0.9-5.5)	3.6 (0.9-7.9)	0.1440	29.9
HBV DNA (log_{10} copies/ml)	6.7 (2.6-9.6)	6.5 (2.3-9.6)	0.1273	20.5
HBeAg (±)	57/60	58/59	0.8960	2.0
Precore region (W/M/n.d.)	22/87/8	16/75/26	0.6399	5.1
BCP (W/M/n.d.)	22/88/7	17/70/30	0.9359	0.0
HBcrAg (log_{10} U/ml)	5.9 (2.9-7.0)	4.9 (2.9-7.0)	0.0022	41.2
Platelet count (x10^3/m^3)	143 (32-262)	146 (37-396)	0.6340	-12.1
ALT (IU/ml)	68 (7-1088)	55 (9-3410)	0.0977	1.9
AFP (ng/ml)	2.8 (0.8-402)	3.9 (0.8-1010)	0.3118	-13.5
Cirrhosis (presence/absence)	48/69	44/73	0.6882	6.1
Child-Pugh classification (A/B)	108/9	104/13	0.5024	3.1
Follow-up duration (yr)	12.3 (3.1-19.4)	11.6 (3.1-18.3)	0.7346	-4.5
Administration period (yr)	6.1 (1.5-10.7)	-	-	-
Propensity score	0.65895 (0.11449-0.96977)	0.65895 (0.12913-0.96989)	0.9931	0.0

NA, nucleos(t)ide analogue; n.d., not done; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; HBeAg, hepatitis B e antigen; W, wild type; M, mutant type; BCP, basal core promoter; HBcrAg, hepatitis B core-related antigen; ALT, alanine aminotransferase; AFP, alpha-fetoprotein; Child-Pugh classification, reference no [50]. Standardized difference in%; $100(X_{NA} - X_{non\text{-}NA})/([S^2_{NA} + S^2_{non\text{-}NA}]/2)^{1/2}$, where for each covariate X_{NA} and $X_{non\text{-}NA}$ are the sample means in NA and non-NA groups, respectively, and S^2_{NA} and $S^2_{non\text{-}NA}$ are the corresponding sample variances.

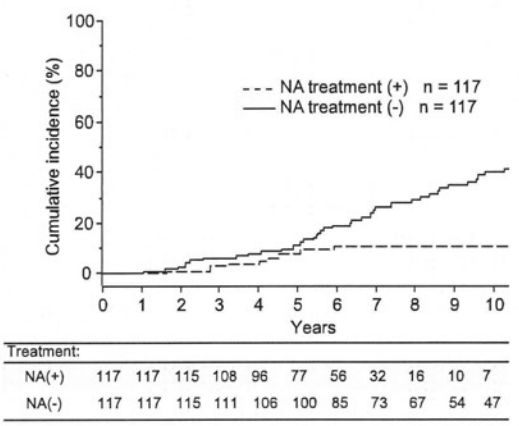

Fig. 2. Incidence of hepatocellular carcinoma (HCC) according to nucleos(t)ide analogue (NA) treatment status. The NA group had a significantly higher rate of progression to HCC than the non-NA group ($p = 0.0094$).

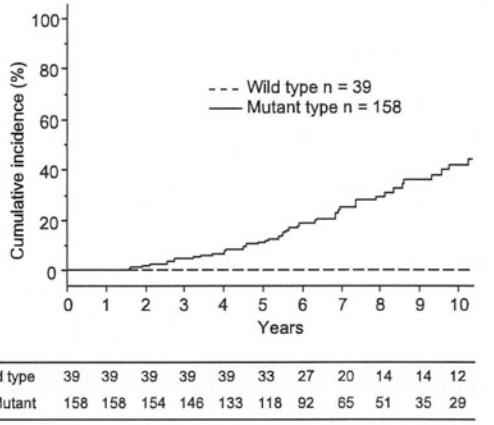

Fig. 3. Incidence of hepatocellular carcinoma (HCC) according to basal core promoter (BCP) mutations. Patients with mutant-type BCP had a significantly higher rate of progression to HCC than those with wild type BCP ($p = 0.0006$).

HBV DNA levels were significantly lower in patients on NA therapy than in patients not on NA therapy. In contrast, platelet counts and albumin levels were significantly higher in patients on NA therapy than in patients not on NA therapy.

Discussion

Our study shows that long-term NA maintenance therapy is associated with the suppression of progression to HCC. Liaw *et al.* reported that lamivudine decreased the risk of HCC in cirrhotic patients [7]. However, it is unclear whether the observed decreased risk of HCC with NA therapy was due to the short observation period in their study. It is very difficult to prove the preventive effect of NA on the development of HCC, because randomized control studies are not ethically possible. In this study, patients on NA therapy were compared to propensity score-matched untreated controls. In these control patients, NA therapy had not yet been approved or was not routinely used for chronic hepatitis B at the time, or was declined by the patient. As opposed to the entire population, these propensity-matched patients were well matched to patients on NA; significant differences included higher HBcrAg levels in the NA group.

Large community-based studies have confirmed that advanced age, male sex, HBeAg positivity, low platelet count,

Table 3. Factors associated with progression to hepatocellular carcinoma among propensity-matched patients (Cox proportional hazard model).

		Adjusted hazard ratio (95% CI)	*p* value
Age (yr)	≤40	1	0.015
	>40	4.36 (1.33-14.29)	
Treatment	no NA	1	0.002
	NA	0.28 (0.13-0.62)	
BCP	wild-type	1	0.012
	mutant-type	12.74 (1.74-93.11)	
HBcrAg (log_{10} U/ml)	≤3.0	1	0.036
	>3.0	2.77 (1.07-7.17)	
γ-GTP (IU/L)	≤56	1	0.001
	>56	2.76 (1.49-5.12)	

NA, nucleos(t)ide analogue; BCP, basal core promoter; HBcrAg, hepatitis B core-related antigen; γ-GTP, gamma glutamyl transpeptidase.

higher ALT levels, elevated AFP levels, and presence of cirrhosis are factors associated with the development of cirrhosis and HCC [17,18]. Platelet count is a useful surrogate marker for the diagnosis of cirrhosis [37]. All subjects were not histologically diagnosed in this study. Therefore, we selected platelet count as a marker of hepatic fibrosis instead of cirrhosis. An elevated ALT level indicates the presence of active disease, and persistently elevated AFP levels are a reflection of an enhanced regenerative state in the liver [16]. In the REVEAL study, a high HBV DNA load was associated with an increased rate of HCC development [17]. A direct correlation was observed between baseline HBV DNA levels and the incidence of HCC, independent of serum ALT concentration. In a model that integrated baseline and follow-up HBV DNA levels, the cumulative incidence of HCC ranged from 1.3% in patients with undetectable levels of HBV DNA to 14.9% in patients with HBV DNA levels greater than or equal to 10^6 copies/ml. Therefore, we have selected factors, such as age, sex, HBeAg serostatus, HBV DNA concentration, platelet count, and ALT for propensity matching.

Although the exact mechanisms of hepatocarcinogenesis by HBV remain unclear, two mechanisms have been proposed [38,39]. One mechanism involves chronic necroinflammation of hepatocytes, cellular injury, and hepatocyte regeneration [40]. The other mechanism involves the direct carcinogenicity of HBV through chromosomal integration [41]. Complete and sustained viral suppression by NA might block both pathways and prevent the development of HCC. It is well known that the rate of HCC is significantly higher in patients with virological breakthrough or no response. In our study, when virological or biochemical breakthrough was observed and the YMDD mutation was detected in patients on lamivudine, adefovir dipivoxil was immediately added. In patients with cirrhosis, especially in the decompensated stage, sustained viral response on NA therapy was not necessarily associated with a preventative effect against the development of HCC, even though the incidence was lower than in a group not on NA [14]. It is not surprising that viral suppression decreased but did not eliminate the risk of HCC, because HBV DNA may have already integrated into the host genome before the initiation of therapy and may have resulted in genomic alternations, chromosomal instability, or both [42,43].

It is reported that patients with HBV genotype C infection have higher HBV DNA levels, higher frequency of pre-S deletions, higher prevalence of BCP T1762/A1764 mutations, and significantly higher chances of developing HCC [16,44–46]. In our study, T1762/A1764 mutations were observed in 158 (80.2%) out of 197 patients and were associated with a higher risk of developing HCC (adjusted hazard ratio, 12.740 [95% CI 1.743–93.108]), independent of NA therapy. However, the BCP T1762/A1764 mutations were detected in HCC patients from Asia and Africa, where HBV genotype C infection is predominant [16].

HBcrAg is a new HBV marker that reflects HBV load and corresponds to HBV DNA levels [21]. HBcrAg is comprised of HBV core antigen (HBcAg) and HBeAg; both are products of the precore/core gene and share the first 149 amino acids of HBcAg. The HBcrAg assay measures HBcAg and HBeAg simultaneously by using monoclonal antibodies that recognize both denatured HBcAg and HBeAg [47]. Serum HBcrAg concentration is well correlated with intrahepatic levels of covalently closed circular DNA (cccDNA) [48]. It is reported that HBcrAg is a useful marker for guiding cessation of NA therapy and evaluation of disease activity [21,49]. In our study, elevated serum HBcrAg concentration was associated with a higher risk of developing HCC (adjusted hazard ratio, 2.767 [95% CI 1.067–7.172]). This is the first report demonstrating a relationship between HBcrAg and HCC.

The present study has several limitations. The retrospective design might have introduced an unintended bias. The propensity matching method was adopted to reduce the confounding effects of covariates. Characteristics of patients who did or did not receive NA therapy were similar except for HBcrAg concentration.

Table 4. Average integration values of various parameters in patients who did or did not receive NA therapy.

	NA group (n = 117)	Non-NA group (n = 117)	*p* value
Platelet count (x10^3/m^3)	17.0 (3.3-37.2)	14.8 (3.3-296)	0.0060
ALT (IU/ml)	28.2 (8.5-88.9)	39.1 (12.2-737.5)	<0.0001
γ-GTP (IU/L)	27.0 (10.9-267.6)	36.2 (9.5-269.7)	0.0427
Total bilirubin (mg/dl)	0.7 (0.3-2.0)	0.7 (0.3-2.6)	0.1554
ALP (IU/L)	242.7 (113.5-1028.8)	265.2 (140.5-1247.6)	0.0127
Albumin (g/dl)	4.4 (3.0-5.0)	4.0 (2.4-4.8)	<0.0001
Alpha-fetoprotein (ng/ml)	2.2 (0.8-106.0)	4.5 (0.9-723.8)	<0.0001
HBV DNA (log_{10} copies/ml)	2.5 (2.1-8.9)	4.6 (2.1-9.3)	<0.0001

NA, nucleos(t)ide analogue; ALT, alanine aminotransferase; γ-GTP, gamma glutamyl transpeptidase; ALP, alkaline phosphatase; HBV, hepatitis B virus.

Research Article

However, the non-NA group included many historical cases when NA therapy was not yet available. In addition, the HBV DNA assay used between 1998 and 2007 was not the most sensitive one.

In conclusion, NA therapy reduces the risk of HCC compared with untreated controls. Higher serum HBcrAg levels and BCP mutations are associated with development of HCC, independent of NA therapy.

Financial support

This work was supported by Health and Labour Sciences Research Grants (Research on Hepatitis) from the Ministry of Health, Labour and Welfare of Japan.

Conflict of interest

The authors who have taken part in this study declared that they do not have anything to disclose regarding funding or conflict of interest with respect to this manuscript.

References

[1] EASL Jury. EASL international consensus conference on hepatitis B, 13–14 September, 2002, Geneva, Switzerland, Consensus statement (short version). J Hepatol 2003;38:533–540.

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

[3] Beasley RP, Hwang LY, Lin CC, Chien CS. Hepatocellular carcinoma and hepatitis B virus. A prospective study of 22 707 men in Taiwan. Lancet 1981;2:1129–1133.

[4] Szmuness W. Hepatocellular carcinoma and the hepatitis B virus: evidence for a causal association. Prog Med Virol 1978;24:40–69.

[5] Rustgi VK. Epidemiology of hepatocellular carcinoma. Gastroenterol Clin North Am 1987;16:545–551.

[6] Kiyosawa K, Umemura T, Ichijo T, Matsumoto A, Yoshizawa K, Gad A, et al. Hepatocellular carcinoma: recent trends in Japan. Gastroenterology 2004;127:S17–S26.

[7] Liaw YF, Sung JJ, Chow WC, Farrell G, Lee CZ, Yuen H, et al. Lamivudine for patients with chronic hepatitis B and advanced liver disease. N Engl J Med 2004;351:1521–1531.

[8] Shamliyan TA, MacDonald R, Shaukat A, Taylor BC, Yuan JM, Johnson JR, et al. Antiviral therapy for adults with chronic hepatitis B: a systematic review for a national institutes of health consensus development conference. Ann Intern Med 2009;150:111–124.

[9] Lai CL, Dienstag J, Schiff E, Leung NW, Atkins M, Hunt C, et al. Prevalence and clinical correlates of YMDD variants during lamivudine therapy for patients with chronic hepatitis B. Clin Infect Dis 2003;36:687–696.

[10] Liaw YF, Chien RN, Yeh CT, Tsai SL, Chu CM. Acute exacerbation and hepatitis B virus clearance after emergence of YMDD motif mutation during lamivudine therapy. Hepatology 1999;30:567–572.

[11] Nguyen MH, Keeffe EB. Chronic hepatitis B: early viral suppression and long-term outcomes of therapy with oral nucleos(t)ides. J Viral Hepat 2009;16:149–155.

[12] Papatheodoridis GV, Dimou E, Dimakopoulos K, Manolakopoulos S, Rapti I, Kitis G, et al. Outcome of hepatitis B e antigen-negative chronic hepatitis B on long-term nucleos(t)ide analog therapy starting with lamivudine. Hepatology 2005;42:121–129.

[13] Yuen MF, Seto WK, Chow DH, Tsui K, Wong DK, Ngai VW, et al. Long-term lamivudine therapy reduces the risk of long-term complications of chronic hepatitis B infection even in patients without advanced disease. Antivir Ther 2007;12:1295–1303.

[14] Eun JR, Lee HJ, Kim TN, Lee KS. Risk assessment for the development of hepatocellular carcinoma: according to on-treatment viral response during long-term lamivudine therapy in hepatitis B virus-related liver disease. J Hepatol 2010;53:118–125.

[15] Joffe MM, Rosenbaum PR. Invited commentary: propensity scores. Am J Epidemiol 1999;150:327–333.

[16] Liaw YF. Natural history of chronic hepatitis B virus infection and long-term outcome under treatment. Liver Int 2009;29:100–107.

[17] Chen CJ, Yang HI, Su J, Jen CL, You SL, Lu SN, et al REVEAL-HBV Study Group. Risk of hepatocellular carcinoma across a biological gradient of serum hepatitis B virus DNA level. JAMA 2006;4:65–73.

[18] Yuen MF, Yuan HJ, Wong DK, Yuen JC, Wong WM, Chan AO, et al. Prognostic determinants for chronic hepatitis B in Asians: therapeutic implications. Gut 2005;54:1610–1614.

[19] Kumada T, Toyoda H, Kiriyama S, Sone Y, Tanikawa M, Hisanaga Y, et al. Incidence of hepatocellular carcinoma in patients with chronic hepatitis B virus infection who have normal alanine aminotransferase values. J Med Virol 2010;82:539–545.

[20] Austin PC. Some methods of propensity-score matching had superior performance to others: results of an empirical investigation and Monte Carlo simulations. Biom J 2009;51:171–184.

[21] Kato H, Orito E, Sugauchi F, Ueda R, Gish RG, Usuda S, et al. Determination of hepatitis B virus genotype G by polymerase chain reaction with hemi-nested primers. J Virol Methods 2001;98:153–159.

[22] Kimura T, Rokuhara A, Matsumoto A, Yagi S, Tanaka E, Kiyosawa K, et al. New enzyme immunoassay for detection of hepatitis B virus core antigen (HBcAg) and relation between levels of HBcAg and HBV DNA. J Clin Microbiol 2003;41:1901–1906.

[23] Wong DK, Tanaka Y, Lai CL, Mizokami M, Fung J, Yuen MF. Hepatitis B virus core-related antigens as markers for monitoring chronic hepatitis B infection. J Clin Microbiol 2007;45:3942–3947.

[24] Liu CJ, Kao JH, Lai MY, Chen PJ, Chen DS. Evolution of precore/core promoter mutations in hepatitis B carriers with hepatitis B e antigen seroreversion. J Med Virol 2004;74:237–245.

[25] Kao JH, Wu NH, Chen PJ, Lai MY, Chen DS. Hepatitis B genotypes and the response to interferon therapy. J Hepatol 2000;33:998–1002.

[26] Kumada T, Toyoda H, Kiriyama S, Sone Y, Tanikawa M, Hisanaga Y, et al. Relation between incidence of hepatic carcinogenesis and integration value of alanine aminotransferase in patients with hepatitis C virus infection. Gut 2007;56:738–739.

[27] Kumada T, Toyoda H, Kiriyama S, Sone Y, Tanikawa M, Hisanaga Y, et al. Incidence of hepatocellular carcinoma in hepatitis C carriers with normal alanine aminotransferase levels. J Hepatol 2009;50:729–735.

[28] Makuuchi M, Kokudo N, Arii S, Futagawa S, Kaneko S, Kawasaki S, et al. Development of evidence-based clinical guidelines for the diagnosis and treatment of hepatocellular carcinoma in Japan. Hepatol Res 2008;38:37–51.

[29] Shen L, Li JQ, Zeng MD, Lu LG, Fan ST, Bao H. Correlation between ultrasonographic and pathologic diagnosis of liver fibrosis due to chronic virus hepatitis. World J Gastroenterol 2006;28:1292–1295.

[30] Iacobellis A, Fusilli S, Mangia A, Clemente R, Festa V, Giacobbe A, et al. Ultrasonographic and biochemical parameters in the non-invasive evaluation of liver fibrosis in hepatitis C virus chronic hepatitis. Aliment Pharmacol Ther 2005;22:769–774.

[31] Caturelli E, Castellano L, Fusilli S, Palmentieri B, Niro GA, del Vecchio-Blanco C, et al. Coarse nodular US pattern in hepatic cirrhosis: risk for hepatocellular carcinoma. Radiology 2003;226:691–697.

[32] Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. Hepatology 2011;53:1020–1022.

[33] Lok AS, McMahon BJ. Chronic hepatitis B: update 2009. Hepatology 2009;50:661–662.

[34] European Association for the Study of the L. EASL clinical practice guidelines: management of chronic hepatitis B. J Hepatol 2009;50:227–242.

[35] Liaw YF, Leung N, Kao JH, Piratvisuth T, Gane E, Han KH, et al. Asian-Pacific consensus statement on the management of chronic hepatitis B: a 2008 update. Hepatol Int 2008;2:263–283.

[36] Greenland S. Basic methods for sensitivity analysis of biases. Int J Epidemiol 1996;25:1107–1116.

[37] Lu SN, Wang JH, Liu SL, Hung CH, Chen CH, Tung HD, et al. Thrombocytopenia as a surrogate for cirrhosis and a marker for the identification of patients at high-risk for hepatocellular carcinoma. Cancer 2006;107:2212–2222.

[38] 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.

[39] Koike K, Tsutsumi T, Fujie H, Shintani Y, Kyoji M. Molecular mechanism of viral hepatocarcinogenesis. Oncology 2002;621:S29–S37.

[40] Chisari FV, Klopchin K, Moriyama T, Pasquinelli C, Dunsford HA, Sell S, et al. Molecular pathogenesis of hepatocellular carcinoma in hepatitis B virus transgenic mice. Cell 1989;59:1145–1156.

[41] Moradpour D, Wands JR. The molecular pathogenesis of hepatocellular carcinoma. J Viral Hepat 1994;1:17–31.

[42] Shafritz DA, Shouval D, Sherman HI, Hadziyannis SJ, Kew MC. Integration of hepatitis B virus DNA into the genome of liver cells in chronic liver disease and hepatocellular carcinoma. Studies in percutaneous liver biopsies and post-mortem tissue specimens. N Engl J Med 1981;305:1067–1073.
[43] Laurent-Puig P, Legoix P, Bluteau O, Belghiti J, Franco D, Binot F, et al. Genetic alterations associated with hepatocellular carcinomas define distinct pathways of hepatocarcinogenesis. Gastroenterology 2001;120: 1763–1773.
[44] Kao JH, Chen PJ, Lai MY, Chen DS. Basal core promoter mutations of hepatitis B virus increase the risk of hepatocellular carcinoma in hepatitis B carriers. Gastroenterology 2003;124:327–334.
[45] Tong MJ, Blatt LM, Kao JH, Cheng JT, Corey WG. Basal core promoter T1762/ A1764 and precore A1896 gene mutations in hepatitis B surface antigen-positive hepatocellular carcinoma: a comparison with chronic carriers. Liver Int 2007;27:1356–1363.
[46] Kusakabe A, Tanaka Y, Inoue M, Kurbanov F, Tatematsu K, Nojiri S, et al. A population-based cohort study for the risk factors of HCC among hepatitis B virus mono-infected subjects in Japan. J Gastroenterol 2011;46:117–124.
[47] Kimura T, Rokuhara A, Sakamoto Y, Yagi S, Tanaka E, Kiyosawa K, et al. Sensitive enzyme immunoassay for hepatitis B virus core-related antigens and their correlation to virus load. J Clin Microbiol 2002;40:439–445.
[48] Suzuki F, Miyakoshi H, Kobayashi M, Kumada H. Correlation between serum hepatitis B virus core-related antigen and intrahepatic covalently closed circular DNA in chronic hepatitis B patients. J Med Virol 2009;81:27–33.
[49] Matsumoto A, Tanaka E, Minami M, Okanoue T, Yatsuhashi H, Nagaoka S, et al. Low serum level of hepatitis B core-related antigen indicates unlikely reactivation of hepatitis after cessation of lamivudine therapy. Hepatol Res 2007;37:661–666.
[50] Pugh RNH, Murray-Lyon IM, Dawson JL, et al. Transection of the oesophagus for bleeding oesophageal varices. Br J Surg 1973;60:646–649.

Journal of Medical Virology 85:449–458 (2013)

Model Incorporating the *ITPA* Genotype Identifies Patients at High Risk of Anemia and Treatment Failure With Pegylated-Interferon Plus Ribavirin Therapy for Chronic Hepatitis C

Masayuki Kurosaki,[1] Yasuhito Tanaka,[2] Nao Nishida,[3] Naoya Sakamoto,[4] Nobuyuki Enomoto,[5] Kentaro Matsuura,[2] Yasuhiro Asahina,[6] Mina Nakagawa,[6] Mamoru Watanabe,[6] Minoru Sakamoto,[5] Shinya Maekawa,[5] Katsushi Tokunaga,[3] Masashi Mizokami,[7] and Namiki Izumi[1]*

[1]*Department of Gastroenterology and Hepatology, Musashino Red Cross Hospital, Tokyo, Japan*
[2]*Department of Virology, Liver Unit, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan*
[3]*Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo, Japan*
[4]*Department of Gastroenterology and Hematology, Hokkaido University, Sapporo, Japan*
[5]*First Department of Internal Medicine, University of Yamanashi, Yamanashi, Japan*
[6]*Department of Gastroenterology and Hepatology, Tokyo Medical and Dental University, Tokyo, Japan*
[7]*Research Center for Hepatitis and Immunology, International Medical Center of Japan Konodai Hospital, Ichikawa, Japan*

This study aimed to develop a model for predicting anemia using the inosine triphosphatase (*ITPA*) genotype and to evaluate its relationship with treatment outcome. Patients with genotype 1b chronic hepatitis C (n = 446) treated with peg-interferon alpha and ribavirin (RBV) for 48 weeks were genotyped for the *ITPA* (rs1127354) and *IL28B* (rs8099917) genes. Data mining analysis generated a predictive model for anemia (hemoglobin (Hb) concentration <10 g/dl); the CC genotype of *ITPA*, baseline Hb <14.0 g/dl, and low creatinine clearance (CLcr) were predictors of anemia. The incidence of anemia was highest in patients with Hb <14.0 g/dl and CLcr <90 ml/min (76%), followed by Hb <14.0 g/dl and *ITPA* CC (57%). Patients with Hb ≥14.0 g/dl and *ITPA* AA/CA had the lowest incidence of anemia (17%). Patients with two predictors (high-risk) had a higher incidence of anemia than the others (64% vs. 28%, $P < 0.0001$). At baseline, the *IL28B* genotype was a predictor of a sustained virological response [adjusted odds ratio 9.88 (95% confidence interval 5.01–19.48), $P < 0.0001$]. In patients who achieved an early virological response, the *IL28B* genotype was not associated with a sustained virological response, while a high risk of anemia was a significant negative predictor of a sustained virological response [0.47 (0.24–0.91), $P = 0.026$]. For high-risk patients with an early virological response, giving >80% of the planned RBV dose increased sustained virological responses by 24%. In conclusion, a predictive model incorporating the *ITPA* genotype could identify patients with a high risk of anemia and reduced probability of sustained virological response. ***J. Med. Virol. 85:449–458, 2013.***

KEY WORDS: hemolytic anemia; ribavirin; creatinine clearance; antiviral therapy

INTRODUCTION

Hepatitis C virus (HCV) infection is a leading cause of cirrhosis and hepatocellular carcinoma worldwide [Kim, 2002]. The rate of eradication of HCV by pegylated interferon (PEG-IFN) plus ribavirin (RBV), defined as a sustained virological response, is around 50% in patients with HCV genotype 1 [Manns et al., 2001; Fried et al., 2002]. Failure of treatment is attributable to the lack of a virological response or relapse after completion of therapy. Genome-wide association studies and subsequent cohort studies

Grant sponsor: Ministry of Health, Labor and Welfare, Japan.
Conflicts of interest and financial disclosures: None reported.

*Correspondence to: Namiki Izumi, MD, PhD, Department of Gastroenterology and Hepatology, Musashino Red Cross Hospital, 1-26-1 Kyonan-cho, Musashino-shi, Tokyo 180-8610, Japan. E-mail: nizumi@musashino.jrc.or.jp

Accepted 19 November 2012
DOI 10.1002/jmv.23497
Published online 7 January 2013 in Wiley Online Library (wileyonlinelibrary.com).

have shown that single nucleotide polymorphisms (SNPs) located near the *IL28B* gene are the most important determinant of virological response to PEG-IFN/RBV therapy [Ge et al., 2009; Suppiah et al., 2009; Tanaka et al., 2009; Rauch et al., 2010]. On the other hand, among patients with a virological response, the probability of a sustained virological response decreases when the patients become intolerant to therapy because of RBV-induced hemolytic anemia and receive a reduced dose of RBV [McHutchison et al., 2002; Kurosaki et al., 2012]. Genome-wide association studies have shown that variants of the inosine triphosphatase (*ITPA*) gene protect against hemolytic anemia [Fellay et al., 2010; Tanaka et al., 2011]. These variants are associated with a reduced requirement for an anemia-related dose reduction of RBV [Sakamoto et al., 2010; Thompson et al., 2010a; Kurosaki et al., 2011d; Seto et al., 2011]. However, factors other than the *ITPA* gene also contribute to the risk of severe anemia or RBV dose reduction [Ochi et al., 2010; Kurosaki et al., 2011d] and the results of studies on the impact of the *ITPA* genotype on treatment outcome are inconsistent [Ochi et al., 2010; Sakamoto et al., 2010; Thompson et al., 2010a, 2011; Kurosaki et al., 2011d].

Data mining is a novel statistical method used to extract relevant factors from a plethora of factors and combine them to predict the incidence of the outcome of interest [Breiman et al., 1980]. Decision tree analysis, a primary component of data mining analysis, has found medical applications recently [Averbook et al., 2002; Miyaki et al., 2002; Baquerizo et al., 2003; Leiter et al., 2004; Garzotto et al., 2005; Zlobec et al., 2005; Valera et al., 2007] and has proven to be a useful tool for predicting therapeutic efficacy [Kurosaki et al., 2010, 2011a,b,c, 2012] and adverse events [Hiramatsu et al., 2011] in patients with chronic hepatitis C treated with PEG-IFN/RBV therapy. Because the results of data mining analysis are presented as a flowchart [LeBlanc and Crowley, 1995], they are easily understandable and usable by clinicians lacking a detailed knowledge of statistics.

For the general application of this genetic information in clinical practice, this study aimed to construct a predictive model of severe anemia using the *ITPA* genotype, together with other relevant factors. This study also aimed to analyze the impact of the risk of anemia on treatment outcome, after adjustment for the *IL28B* genotype. These analyses were carried out at baseline and during therapy, when the early virological response became evident.

MATERIALS AND METHODS

Patients

Data were collected from a total of 446 genotype 1b chronic hepatitis C patients who were treated with PEG-IFN alpha and RBV at five hospitals and universities throughout Japan. The inclusion criteria were: (1) infection by hepatitis C genotype 1b; (2) no co-infection with hepatitis B virus or human immunodeficiency virus; (3) no other causes of liver disease such as autoimmune hepatitis and primary biliary cirrhosis; and (4) availability of DNA for the analysis of the genetic polymorphisms of *IL28B* and *ITPA*. Patients received PEG-IFN alpha-2a (180 μg) and 2b (1.5 μg/kg) subcutaneously every week and a daily weight-adjusted dose of RBV (600 mg for patients weighing <60 kg, 800 mg for patients weighing 60–80 kg, and 1,000 mg for patients weighing >80 kg) for 48 weeks. Dose reduction or discontinuation of PEG-IFN and RBV was primarily based on the recommendations on the package inserts and the discretion of the physicians at each university and hospital. The standard duration of therapy was set at 48 weeks. No patient received erythropoietin or other growth factors for the treatment of anemia. Written informed consent was obtained from each patient, and the study protocol conformed to the ethical guidelines of the Declaration of Helsinki and was approved by the institutional ethics review committees.

Laboratory Tests

Blood samples obtained before therapy were analyzed for hematologic data, blood chemistry, and HCV RNA. Genetic polymorphisms in SNPs of the *ITPA* gene (rs1127354) and the *IL28B* gene (rs8099917) were determined using ABI TaqMan Probes (Applied Biosystems, Carlsbad, CA) and the DigiTag2 assay, respectively. Baseline creatinine clearance (CLcr) levels were calculated using the formula of Cockcroft and Gault [1976]: for males, CLcr = [(140 − age in years) × body weight in kg] ÷ (72 × serum creatinine in mg/dl) and for females, CLcr = 0.85 × [(140 − age in years) × body weight in kg] ÷ (72 × serum creatinine in mg/dl). The stage of liver fibrosis was scored according to the METAVIR scoring system: F0 (no fibrosis), F1 (mild fibrosis: portal fibrosis without septa), F2 (moderate fibrosis: few septa), F3 (severe fibrosis: numerous septa without cirrhosis), and F4 (cirrhosis). A rapid virological response was defined as undetectable HCV RNA by qualitative PCR with a lower detection limit of 50 IU/ml (Amplicor, Roche Diagnostic Systems, Pleasanton, CA) at week 4 of therapy and a complete early virological response was defined as undetectable HCV RNA at week 12. A sustained virological response was defined as undetectable HCV RNA at 24 weeks after completion of therapy. Severe anemia was defined as hemoglobin (Hb) <10 g/dl.

Statistical Analysis

Database for analysis included the following variables: age, sex, body mass index, serum aspartate aminotransferase (AST) levels, alanine aminotransferase (ALT) levels, gamma-glutamyltransferase (GGT) levels, creatinine levels, CLcr, Hb, platelet count, serum levels of HCV RNA, and the stage of liver fibrosis

TABLE I. Patients' Baseline Characteristics

Age (years)	58.6	(9.6)
Gender: male (n, %)	185	(42%)
Body mass index (kg/m^2)	23.1	(3.7)
AST (IU/L)	59.9	(53.8)
ALT (IU/L)	69.8	(53.8)
GGT (IU/L)	48.5	(41.6)
Creatinine (mg/dl)	0.7	(0.2)
Creatinine clearance (ml/min)	89.5	(23.0)
Hemoglobin (g/dl)	14	(1.4)
Platelet count (10^9/L)	154.5	(52.1)
HCV RNA > 600,000 IU/ml (n, %)	354	(79%)
Liver fibrosis: F3-4 (n, %)	108	(24%)
Initial ribavirin dose (n, %)		
600 mg/day	300	(67%)
800 mg/day	138	(31%)
1,000 mg/day	9	(2%)
Pegylated interferon (n, %)		
alpha2a 180 mcg	58	(13%)
alpha2b 1.5 mcg/kg	388	(87%)
ITPA rs1127354: CC (n, %)	317	(71%)
IL28B rs809917: TT (n, %)	311	(70%)

AST, aspartate aminotransferase; ALT, alanine aminotransferase; GGT, gamma-glutamyltransferase.
Data expressed as mean (standard deviation) unless otherwise mentioned.

(Table I). Based on these data set, a model for predicting the risk of developing severe anemia was constructed by data mining analysis using the IBM-SPSS Modeler 13 as described previously [Kurosaki et al., 2010, 2011a,b,c; Hiramatsu et al., 2011]. Briefly, the software was used to explore the database automatically to search for optimal predictors that discriminated most efficiently patients with severe anemia from those without. The software also determined the optimal cutoff values of each predictor. Patients were divided into two groups according to the predictor and each of the two groups was repeatedly divided in the same way until no significant factor remained or 20 or fewer patients were in a group.

The incidence of severe anemia, the total dose of RBV, and treatment outcome were compared between groups with high and low risks of anemia. On univariate analysis, Student's *t*-test was used for continuous variables, and Fisher's exact test was used for categorical data. Logistic regression was used for multivariate analysis. *P* values of <0.05 were considered significant. SPSS Statistics 18 was used for these analyses.

RESULTS

Predictive Model of Severe Anemia

The incidence of severe anemia in the whole cohort was 49% (Fig. 1). The best predictor of severe anemia was the baseline Hb concentration. Patients with a low baseline Hb concentration (<14 g/dl) were more likely to develop severe anemia (67%) than those with a higher Hb (>14 g/dl) (34%). The second best predictor for those patients with a baseline Hb <14.0 g/dl was CLcr. Patients with a CLcr below 90 ml/min had the highest incidence of severe anemia (76%). In those with a CLcr above >90 ml/min the incidence of severe anemia was 57% in patients with the CC allele of the *ITPA* gene while it was 37% in patients with the CA or AA allele. On the other hand, the second best predictor for those patients with a baseline Hb concentration above 14 g/dl was the *ITPA* genotype. Patients with the AA or AC allele had the lowest incidence of anemia (17%). For those with the *ITPA* CC allele, CLcr was the third best predictor; the optimal cutoff value was 85 ml/min for this group. The incidence of severe anemia was 49% in patients with a CLcr below 85 ml/min while it was 32% in those with a CLcr above 85 ml/min.

Following this analysis, the patients were divided into six groups, with the incidence of severe anemia ranging from 17% to 76%. Three groups with two predictors, having an incidence of anemia >40%, were defined as the high-risk group and the remainder were defined as the low-risk group. The incidence of severe anemia was higher in the high-risk group than the low-risk group (65% vs. 28%, $P = 0.029$) (Fig. 2). Comparison of the *ITPA* genotype and the predictive model showed that the sensitivity for the prediction of severe anemia was similar (75.9% vs. 76.4%) but the specificity of the predictive model was greater (33.6% vs. 59.3%).

The Risk of Anemia Impacts on Sustained Virological Responses by Patients Who Achieved an Early Virological Response

The impact of *IL28B* genotype, *ITPA* genotype, and risk group of anemia on the rate of sustained virological response was studied at baseline and week 12. At baseline, patients with the TT allele of the *IL28B* gene had a significantly higher rate of sustained virological response than those with the TG or GG allele (43% vs. 10%, $P < 0.0001$), the high-risk group for anemia had a significantly lower rate of sustained virological response than the low-risk group (28% vs. 40%, $P = 0.011$), and the *ITPA* genotype was not associated with a sustained virological response (Fig. 3A–C). At week 4, patients with rapid virological response had a high rate of sustained virological response, irrespective of the *IL28B* genotype (TT vs. TG/GG; 97% vs. 100%, $P = 1.000$), the *ITPA* genotype (CC vs. CA/AA; 95% vs. 100%, $P = 1.000$), and the risk of anemia (high vs. low; 95% vs. 100%, $P = 1.000$). Among the patients who did not achieve a rapid virological response, those with the *IL28B* TT allele had a significantly higher rate of sustained virological response than those with the TG or GG allele (38% vs. 8%, $P < 0.0001$), and the high-risk group for anemia had a significantly lower rate of sustained virological response than the low-risk group (24% vs. 35%, $P = 0.015$). At week 12, in patients who achieved a complete early virological response, the *IL28B* genotype was not associated with a sustained virological response, while the high-risk group for anemia had a