



Published in final edited form as:

Hepatology. 2010 December ; 52(6): 2034–2043. doi:10.1002/hep.23943.

Genetic Variations at Loci Involved in the Immune Response Are Risk Factors for Hepatocellular Carcinoma

Robert J. Clifford¹, Jinghui Zhang², Daoud M. Meerzaman¹, Myung-Soo Lyu¹, Ying Hu², Constance M. Cultraro¹, Richard P. Finney², Jenny M. Kelley¹, Sol Efroni³, Sharon I. Greenblum², Cu V. Nguyen¹, William L. Rowe², Sweta Sharma¹, Gang Wu², Chunhua Yan², Hongen Zhang¹, Young-Hwa Chung⁴, Jeong A. Kim⁴, Neung Hwa Park⁵, Il Han Song⁶, Kenneth H. Buetow^{1,2}

¹Laboratory of Population Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD;

²Center for Bioinformatics and Information Technology, National Cancer Institute, Rockville, MD;

³Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan, Israel;

⁴Department of Internal Medicine, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea;

⁵Department of Internal Medicine, University of Ulsan College of Medicine, Ulsan University Hospital, Ulsan, Korea;

⁶Department of Internal Medicine, College of Medicine, Dankook University, Cheon -An, Korea.

Abstract

Primary liver cancer is the third most common cause of cancer-related death worldwide, with a rising incidence in Western countries. Little is known about the genetic etiology of this disease. To identify genetic factors associated with hepatocellular carcinoma (HCC) and liver cirrhosis (LC), we conducted a comprehensive, genome-wide variation analysis in a population of unrelated Asian individuals. Copy number variation (CNV) and single nucleotide polymorphisms (SNPs) were assayed in peripheral blood with the high-density Affymetrix SNP6.0 microarray platform. We used a two-stage discovery and replication design to control for overfitting and to validate observed results. We identified a strong association with CNV at the T-cell receptor gamma and alpha loci ($P < 1 \times 10^{-15}$) in HCC cases when contrasted with controls. This variation appears to be somatic in origin, reflecting differences between T-cell receptor processing in lymphocytes from individuals with liver disease and healthy individuals that is not attributable to chronic hepatitis virus infection. Analysis of constitutional variation identified three susceptibility loci including the class II MHC complex, whose protein products present antigen to T-cell receptors and mediate immune surveillance. Statistical analysis of biologic networks identified variation in the “antigen presentation and processing” pathway as being highly significantly associated with HCC ($P = 1 \times 10^{-11}$). SNP analysis identified two variants whose allele frequencies differ

Address reprint requests to: Robert J. Clifford, Laboratory of Population Genetics, NCI, NIH, Bethesda, MD 20892. clifforr@mail.nih.gov; fax: (301) 480-6641.

Potential conflict of interest: Nothing to report.

significantly between HCC and LC. One of these ($P= 1.74 \times 10^{-12}$) lies in the PTEN homolog TPTE2.

Conclusion: Combined analysis of CNV, individual SNPs, and pathways suggest that HCC susceptibility is mediated by germline factors affecting the immune response and differences in T-cell receptor processing.

Primary liver cancer is the third most common worldwide cause of cancer-related deaths, with a rising incidence in Western countries. The highest incidence in the world occurs in Korea, where the rate among males is 44.9/100,000.^{1,2} Hepatocellular carcinoma (HCC) is responsible for 85%–90% of primary liver cancers, with a high incidence rate (35–50/100,000 in males) in Asian countries like China and South Korea. HCC is associated with several major risk factors including chronic hepatitis B and C infection, consumption of aflatoxin-contaminated foods, excessive consumption of alcohol, and liver cirrhosis (LC).^{3–5} Both the variability in outcome following the same environmental exposure and the clustering of HCC within families suggest genetic susceptibility.^{6–8}

Genetic analysis of HCC susceptibility, to date, has centered on examination of individual candidate genes whose variation may plausibly influence the response to known environmental risk factors.^{6,9,10} Recent technological advances have made it feasible to perform comprehensive, genome-wide searches for genetic factors associated with disease susceptibility and progression. These factors include both single nucleotide and copy number polymorphisms. To date, genome-wide analysis of liver cancer has been limited to the examination of HCC tumor tissue and adjacent uninvolved liver tissue which identify somatic changes associated with the tumor.¹¹ Moreover, these studies have largely focused on changes in gene expression measured at the RNA level. To identify susceptibility loci for liver disease, we conducted an association study analyzing single nucleotide polymorphisms (SNPs) and copy number variations (CNVs) in DNA isolated from peripheral blood; for this work we used the Affymetrix SNP 6.0 microarray, which contains 934,968 SNPs and 945,826 structural variation markers.

Our genome-wide association study (GWAS), the first to focus on HCC, revealed that both constitutional genetic variations and somatic genomic events are risk factors for HCC. We observed an association between germline variants in the MHC class II loci and somatic CNV at T-cell receptor loci and liver disease. Our findings provide genomic evidence that genes involved in the immune response play a critical role in the development of liver cancer.

Materials and Methods

Study Population and Design.

This study involved unrelated HCC and LC patients of Korean ethnicity treated at the Asan Medical Center, Seoul, Korea. Disease diagnosis was confirmed by histopathology. Previous clinical history, enzyme-linked immunosorbent assay (ELISA)-based serum test results for hepatitis B virus (HBV) and hepatitis C virus (HCV), and clinical laboratory data were collected for these individuals; 89% of the HCC cases and 76% of the LC cases were chronically infected with either HBV or HCV.

Two sources of controls were used. The first set of controls for our study was unrelated individuals from the Asan Medical Center. The viral infection status of controls was not ascertained. A second set of controls was HBV+ individuals of Chinese origin (described previously).⁶ The local ethics committees and all subjects gave informed consent before inclusion in the study.

A total of 386 Korean HCC cases, 86 Korean LC cases, 587 Korean controls (Supporting Table S1), and 100 Chinese controls passed the quality control evaluations (DNA integrity measurement, STRP genotyping for assessment of identity, and high SNP call rate from the Affymetrix 6.0 platform) described below. We confirmed through molecular assays that there is no population stratification among the Korean samples (see Supporting Methods). Individuals from the Korean population set were assigned to the discovery (Stage 1) or validation (Stage 2) group based on their order of enrollment in the study. Stage 1 included 271 controls, 180 HCC cases, and 66 LC cases; Stage 2 had 316 controls, 206 HCC patients, and 20 individuals with LC. Key findings from the two-stage analysis were further validated using the Chinese control samples.

DNA and RNA Preparation.

Peripheral blood DNA was extracted using the Blood DNA Kit (Qiagen, Valencia, CA). DNA integrity and quantity were assessed using the Quantifiler Human DNA Quantification kit (Applied Biosystems, Foster City, CA). Polymerase chain reaction (PCR) products were analyzed with an ABI 3130XL Automated DNA Sequencer and the GeneMapper ID v3.2 software (Applied Biosystems, Foster City, CA).

Affymetrix SNP6.0 Assay.

The Affymetrix SNP6.0 assay was performed according to the manufacturer's instructions (Affymetrix, Santa Clara, CA). Assay runs were performed in 96-well plates containing equal numbers of case and control samples, two Asian HapMap samples (chosen from NA18954, NA18971, NA18603, and NA18995) for external genotype validation, the Affymetrix Affy103 control DNA, and a water blank. Cases were randomly selected for each plate one-by-one using a random-number generator. Each case in the discovery phase was paired with its best match in sex and age among the control samples. Processing each Stage 1 case along with a matched control was aimed at minimizing technical variation in experimental results. Controls in the validation phase have limited clinical information and therefore were selected randomly. Each batch of 47–93 SNP6.0 assays was analyzed with the Affymetrix Genotyping Console v. 3.0 *birdseed* program. Samples with a global allele call rate below 98.5% were excluded from further analysis. In all, 90.5% of samples had an SNP call rate 99%. Genotype and CNV data are deposited in caArray (<https://array.nci.nih.gov/caarray/project/bueto-00429>).

Copy Number Analysis.

Given the large number of markers examined in a GWAS, it is critical to control for false discovery by validating observations in an independent population. We employed a two-stage discovery-replication study design for our comparison of HCC patients and healthy controls (Supporting Fig. S1). The study population was divided into independent discovery

(Stage 1) and validation (Stage 2) sets as described above. Stage 1 and Stage 2 samples were analyzed separately for CNV using the Affymetrix Genotyping Console program with default parameters and the HapMap270 reference model. The resulting copy number log₂ratio data served as input for the R *DNA-copy* package, which implements the circular binary segmentation (CBS) algorithm.¹² We converted CBS copy number values to discrete copy number states (high, normal, low) using thresholds two standard deviations from the mean CNV of all autosomal markers in the dataset (described in Supporting Methods). In all, 422,062 nonoverlapping genomic segments were identified in the analysis of the Stage 1 samples. CNV segments associated with HCC were identified using a 2×3 Fisher's exact test. The 2,318 segments with P below 1×10^{-4} in the Stage 1 samples were retested in the Stage 2 samples. For validation, segments had to show an association with disease in the Stage 2 population with a $P < 2.157 \times 10^{-5}$, corresponding to $P = 0.05$ after Bonferroni adjustment for 2,318 tests. We confirmed that age and gender were not confounding variables in our analysis (Supporting Methods).

Because our study population contains only 86 LC patients, we performed a Fisher's exact test on combined Stage 1 and Stage 2 CNV data from LC patients and healthy Korean individuals to identify copy number variants acting as risk factors for cirrhosis. To be considered significant, the resulting P had to be < 0.05 after Bonferroni adjustment for 422,062 comparisons. Analysis aimed at identifying CNV that distinguishes HCC from LC was likewise performed on combined Stage 1 and Stage 2 data. The distribution of high, normal, and low copy number was examined at 208,761 nonoverlapping segments identified through CBS analysis of the 386 HCC and 86 LC individuals.

Genotype Analysis.

Genotype calls were generated with the Affymetrix Power Tools *apt-probeset-genotype* program using default parameters. Files were analyzed in two batches (Stages 1 and 2) to ensure accurate normalization. Noninformative markers, markers with a minor allele frequency below 5% in controls, SNPs for which $< 95\%$ of samples have a quality score ≥ 90 , and SNPs not in Hardy-Weinberg equilibrium in controls ($P < 0.001$) were excluded from further analysis. To test the association of individual SNPs with HCC, cases and controls were divided into training (Stage 1) and testing (Stage 2) sets as described above (Supporting Fig. S2). Single SNP association analysis was performed with *PLINK*,¹³ using a logistical model. The 5,622 SNPs that met a significance threshold of $P < 0.01$ in the Stage 1 discovery set were subjected to a Cochran-Armitage trend test using data from the Stage 2 population. The significance threshold for the trend test (8.89×10^{-6}) was based on a correction for 5,622 comparisons. For cirrhosis, SNP analysis was performed using all LC cases and all controls. Similarly, all HCC and LC cases were used in single SNP analysis aimed at identifying variants that distinguish the two disease states. Linkage disequilibrium (LD) among individual markers was calculated for each chromosome using a C program that implements the *LDSelect* algorithm.¹⁴ SNPs with an r^2 correlation ≥ 0.8 were considered to be in linkage disequilibrium.

Pathway Analysis.

The 1,000 SNPs most strongly associated with disease in the single marker association analysis were selected from Stage 1 and Stage 2. Regions of significance were defined by identifying additional SNPs in LD with these markers. The 1,000 SNPs of interest were then assigned to National Cancer Institute (NCI)-curated pathways (<http://pid.nci.nih.gov>) on the basis of their LD to genes in these pathways. The 1,000 SNPs were then evaluated for statistically significant overrepresentation in pathways using Fisher's hypergeometric density function.¹⁵ This test determines the likelihood of the observed number of associations (e.g., seven SNPs observed within the antigen processing pathway) from a finite population (18,504 total SNPs assigned to pathways, among which there are 16 total SNPs within the antigen processing pathway) in a defined number of draws without replacement (1,000 SNPs of interest).

Real-Time Quantitative PCR Assay.

TaqMan real-time PCR assays (Applied Biosystems) were used to confirm the SNP6.0 CNV results for T-cell receptor alpha complex (TRA@) and T-cell receptor gamma complex (TRG@). Details of the assay are in Supporting Table S2. Copy number determination was performed using the standard curve method of absolute quantitation with normalization to albumin (ALB)¹⁶ as an internal reference. Standard curves were generated from CEPH controls, B-cell-derived lymphoblastoid cell lines that do not undergo rearrangement at the TCR loci, and thus are diploid for ALB, TRA@, and TRG@.

Validation of SNP Genotypes.

The MHC class II region contains clusters of homologous genes. To verify that the SNP6.0 genotype calls for rs2647073 and rs3997872, SNPs showing the highest association to HCC, were not experimental artifacts, we genotyped these markers using an independent genotyping methodology, the TaqMan assay. TaqMan results were in complete agreement with the SNP6.0 genotypes.

Results

CNV Analysis.

Comparison of CNV in HCC patients and Korean controls reveals that a number of genomic regions showed strong association with disease outcome. We identified eight loci where CNV is significantly associated with HCC. Six of these appear to be germline CNVs. The other two, however, involve T-cell receptor loci, which are known to undergo recombination in peripheral blood lymphocytes, the source of DNA for our study. Of the six loci showing germline CNV, the one exhibiting the strongest association with HCC is a small region of chromosome 1p36.33 that contains no known or predicted genes. In this case, low copy number correlates with increased risk for both HCC (unadjusted $P = 5.94 \times 10^{-16}$ for Stage 1, $P = 1.11 \times 10^{-10}$ for Stage 2; Table 1) and LC (unadjusted $P = 6.03 \times 10^{-9}$ for combined Stage 1 and Stage 2; Table 2). The five other regions for which CNV is associated with HCC contain the genes KNG1 (3q27.3); C4orf29 and LARP2 (4q28.2); ALDH7A1, PHAX, C5orf48, and LMNB1 (5q23.2); SRPK2 and PUS7 (7q22.2); and TMPO (12q23.1). Low

copy number at all five of these loci is more frequent in controls than HCC patients (Table 1). We observed no statistically significant association between CNV at these five loci and LC (Table 2). Additionally, none of these loci show significant differences between LC and HCC.

Among the loci showing association of CNV with HCC, the strongest association is seen at the TRG@ and TRA@. In both cases low copy number is more frequent in controls than cases. In HCC versus controls, TRG@ shows an unadjusted P of 3.16×10^{-21} in the Stage 1 training set and $P = 1.85 \times 10^{-28}$ in the Stage 2 testing set; TRA@ has an unadjusted $P = 1.94 \times 10^{-16}$ in Stage 1 and $P = 6.24 \times 10^{-28}$ in Stage 2 (Table 1). We validated these findings using an independent platform by performing a TaqMan assay (t test $P = 2.86 \times 10^{-18}$ for TRA@; $P = 3.56 \times 10^{-26}$ for TRG@ for combined Stage 1 and Stage 2 samples; Supporting Table S9). CNV at the TRG@ and TRA@ loci also differs significantly between control and LC individuals (unadjusted P of 5.66×10^{-12} and 3.17×10^{-13} , respectively, in combined Stage 1 and Stage 2 samples; Table 2). As is seen in HCC, low copy number is more frequent in control than LC individuals.

To confirm our proposal that the observed CNV at TRA@ and TRG@ reflects somatic genomic rearrangement at these loci that occurs in normal T lymphocytes, we inspected publicly accessible CNV data at these T-cell receptor loci in B cells. Because B cells do not exhibit TCR rearrangement, they should be diploid at the TRA@ and TRG@ loci. As expected, neither locus shows CNV in publicly accessible HapMap genotype data, which were generated using DNA isolated from B-cell lymphoblastoid cell lines established at the Centre d'Etude du Polymorphisme Humain (CEPH).¹⁷

We observe no significant association between CNV at the T-cell receptor loci and hepatitis virus status in the cases where viral status is known in the current study population (Supporting Table S4). Likewise, the HBV-positive Chinese control individuals were genotyped using the TaqMan assay for the TRA@ locus. TRA@ copy number was observed to be similar in HBV-positive Chinese individuals and our Korean controls (t test $P = 0.477$), but differed significantly between the HBV-positive Chinese individuals and our HBV-positive Korean cases ($P = 6.572 \times 10^{-13}$) (Supporting Table S5). Hence, we conclude that hepatitis virus infection status, *per se*, does not account for observed CNV differences between our cases and controls.

Single SNP Analysis.

In our investigation of the association of genomic variation with disease, we also examined individual SNPs in HCC patients and healthy Korean controls. The set of SNPs most strongly correlated with HCC by a trend test was enriched for polymorphisms in genes involved in antigen presentation. Three of the eight variants with the highest association to liver cancer (rs9267673, rs2647073, and rs3997872) lie in the MHC class II locus (Table 3). None of the three variants is in LD with either of the others. The variant rs9267673 is located adjacent to the gene C2. rs2647073 is in LD with SNPs in a set of genes that includes HLA-DRB1, HLA-DRB6, HLA-DRB5, and HLA-DRA. The SNP rs3997872, on the other hand, is in LD with SNPs in the HLA-DQA1, HLA-DQB1, HLA-DQA2, and

HLA-DQB2 loci. All three SNPs are independently associated with HCC, showing neither an additive nor multiplicative effect.

Interestingly, in addition to their association with HCC, two of the three SNPs (rs9267673 and rs2647073) show association to LC ($P=0.0052$ and 0.0007 , respectively). In contrast, rs3997872 is only weakly associated with LC ($P=0.0408$) (Supporting Table S3).

Comparison of SNP allele frequencies in HCC and LC patients, identified two variants that distinguish liver cancer from cirrhosis (Table 4). An SNP, rs2880301, is located within the TPTE2 gene; the second, rs2551677, lies in a gene-poor region of 2q14.1. Both polymorphisms are distinct from those identified in the comparison of HCC patients and Korean controls.

We also examined HCC individuals to determine whether risk alleles at SNPs associated with cancer (Table 3) correlate with hepatitis virus infection status. All eight variants show an adjusted $P > 0.11$ for association with HBV and an adjusted $P > 0.48$ for association with HCV (Supporting Table S6). Thus, viral infection status does not account for the observed association between SNPs in LD to immune response genes and liver cancer. Finally, we observe no significant association between SNPs in HLA-DP, which has been implicated in HBV susceptibility in Asian populations¹⁸ and HCC.

Pathway Analysis.

We next evaluated whether multiple SNPs in a common biological network, each with a modest individual effect, were associated with HCC. In order to reduce complexity and statistical noise, we first selected the 1,000 most significant SNPs from Stage 1 and Stage 2 and assigned them to biological pathways based on their linkage disequilibrium to genes in the NCI Protein Interaction Database. We then tested whether any biological pathways were over-represented in this set of 1,000 SNPs. The results, summarized in Table 5, show that “antigen processing and presentation” is the pathway most strongly associated with HCC in combined Stage 1 and Stage 2 data, with an unadjusted P of 1×10^{-11} . We next examined the relationship between SNPs in antigen processing loci and CNV at the T-cell receptor loci. We found that multiple SNPs at the HLA-DQB2 locus were associated with CNVs at the TCR loci. Allelic variants of rs9276427, rs28420297, rs9276429, and rs9276490 are correlated with CNVs at TCR-gamma, all with P below 5×10^{-4} .

Discussion

We performed a multidimensional genomic analysis of HCC and LC, examining the association of CNVs, individual SNPs, and genetic pathways to liver disease. Our GWAS, the first to focus on HCC, reveals that both constitutional genetic variations and somatic genomic events behave as risk factors for HCC.

HCC is frequently preceded by cirrhosis. Because only a subset of LC patients develop HCC, it is of great interest to identify factors that affect the transition from LC to cancer. We identified two SNPs whose allele frequencies differ significantly between HCC and LC (Table 4). The first is located in 2q14.1, ≈ 175 kb from the nearest gene. The second variant

lies within an intron of TPTE2, which encodes a homolog of the PTEN tumor suppressor protein.¹⁹ Our study is the first to suggest TPTE2 is involved in carcinogenesis.

Our analysis of HCC patients and healthy individuals identified six loci where inherited CNV is strongly associated with HCC (Table 1). Several of these have functions plausibly related to the etiology of HCC. SPRK2 encodes a product reported to phosphorylate the HBV core protein.²⁰ Work by Zheng et al.²¹ suggests that SRPK2 can inhibit HBV replication.

Two loci where CNV is associated with liver cancer may play roles in tumorigenesis. TMPO encodes a protein that regulates the subnuclear localization of Rb. Knockdown of TMPO in fibroblasts disrupts cell cycle progression^{22,23}; elevated expression of the gene product has been observed in a variety of primary tumors.²⁴ Consistent with its apparent role in promoting tumor formation, low TMPO copy number is associated with reduced HCC risk. In contrast to TMPO, increased copy number in a small region of 1p36.33 is associated with reduced HCC risk. Deletions at 1p36 have been reported in a wide variety of cancers.^{25–28} Although the 1p36.33 CNV region contains no known or predicted genes, the region does show homology to the mitochondrial genome.²⁹ We are undertaking further analyses to determine whether the observed 1p36.33 CNV reflects variation in mitochondrial or chromosomal DNA.

The most striking outcome of our analysis of SNPs and CNVs is that germline variation may modulate somatic immune events that drive HCC susceptibility. First, we found that reduced copy number at the KNG1 locus, which encodes a protein that promotes T-cell senescence,³⁰ is more frequent in healthy individuals than in liver cancer patients.

Further reinforcing the role of the immune system, individual SNP analyses reveal that the MHC class II locus contains three variants (rs9267673, rs2647073, and rs3997872) strongly associated with HCC. MHC class II molecules present antigen to CD4⁺ (helper) T cells.³¹ The three SNPs may be associated with altered MHC class II proteins that result in an ineffective T-cell response. Interestingly, rs2647073 lies 3.4 kb from rs660895, an SNP recently identified as a risk factor for the autoimmune liver disease biliary cirrhosis.³² Analysis of SNP allele distributions in pathways further reinforces this observation. In multiple SNP analysis, “antigen processing and presentation” emerged as the pathway with the strongest association with HCC. Among the SNPs in this pathway, multiple variants at the HLA-DQB2 locus were observed to be associated with CNVs at the TCR loci.

Analysis of copy number variation at TCR gene complexes supports the findings from the SNP analyses. Healthy individuals, on average, have lower copy number at the T-cell receptor loci TRA@ and TRG@ than do persons with HCC (Fig. 1). T-cell maturation involves TCR gene rearrangements that eliminate large portions of the T-cell receptor loci. Thus, successful T-cell receptor rearrangements appear to occur less frequently in cancer patients. Because TCR CNV is absent in DNA samples derived from liver tissue or immortalized B cells, the observed findings are attributable to somatic events occurring in T lymphocytes. CNV patterns at TRA@ suggest that rearrangement events generate functional alpha chain more frequently than delta chain. Low copy number segments observed in

individual samples frequently encompass the TCR delta constant region, but rarely include the TCR alpha constant region (Fig. 2).

Support for the idea that altered T-cell activation contributes directly to carcinogenesis in the liver, rather than simply being a systemic reaction to cancer, comes from the strong association we see between CNV at the T-cell receptor loci and liver cirrhosis, a risk factor for and precursor to HCC (Table 2). Two of the three MHC class II locus SNPs whose genotypes correlate with HCC, rs9267673 and rs2647073, also exhibited strong association with LC (Table 3; Supporting Table S4).

Although the role of the immune system in constitutional susceptibility to HCC is new, the involvement of the immune system in HCC carcinogenesis has been previously suggested in clinical studies and research involving model organisms. Increased activity of helper T cells, which promote inflammation, is associated with HCC.³³ Conversely, activation and proliferation of cytotoxic T lymphocytes is suppressed in individuals with HCC.^{34,35} Further, chronic inflammation has been implicated in the development of liver cancer in both animal models and in humans.^{36–38}

Our work complements and extends these findings by providing a genetic basis for the clinical observations and extending the findings to HCC susceptibility. Our results indicate that germline polymorphisms at the MHC class II locus may affect the generation and proliferation of T cells, with particular rearrangement patterns at TCR loci. From this observation, we propose that the T-cell repertoire of each individual plays a critical role in liver cancer susceptibility and that biological processes affecting T-cell maturation or immune surveillance may represent important etiologic mechanisms for the development of HCC in humans.

With additional validation, the findings of this study may have additional practical clinical benefit. Using DNA obtained from peripheral blood it is possible to assess the status of the germline polymorphisms at the MHC class II loci. Such an assay may allow identification of individuals at increased risk of HCC for more intensive follow-up and monitoring. Similarly, TCR copy number status can be assessed using peripheral blood and an inexpensive TaqMan assay. With validation, this simple test could serve as a noninvasive screen for HCC. Ongoing work will focus on the development of a sensitive and accurate HCC classifier based on CNV loci identified in our study.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment:

We thank Drs. Dinah Singer, Jung-Hyun Park, Katherine McGlynn, and Lalage Wakefield for critical reading of the article. We thank Dr. Barbara Dunn for insightful and valuable comments on the article. We thank Gretchen Carpintero, Grace Yanagawa, and Dhivya Jayaraman for technical assistance. Peripheral blood samples from HBV(+) Chinese individuals enrolled the Haimen City Anti-epidemic Study were kindly provided by Drs. W. Thomas London, Alison Evans, Gang Chen, Wen-Yao Lin, and Fu-Min Shen.

Supported by National Institutes of Health (NIH) intramural grant Z01 CP 010139.

Abbreviations:

CNV	copy number variation
GWAS	genome-wide association study
HBV	hepatitis B virus
HCC	hepatocellular carcinoma
HCV	hepatitis C virus
LC	liver cirrhosis
SNP	single nucleotide polymorphism
TCR	T-cell receptor
TRA@	T-cell receptor alpha complex
TRG@	T-cell receptor gamma complex

References

1. Won YJ, Sung J, Jung KW, Kong HJ, Park S, Shin HR, et al. Nation-wide cancer incidence in Korea, 2003–2005. *Cancer Res Treat* 2009;41: 122–131. [PubMed: 19809561]
2. Curado MP, Edwards B, Shin HR, Storm H, Ferlay J, Heanue M, et al. *Cancer Incidence in Five Continents, Vol. IX*. Lyon, France: IARC Scientific Publications No. 160; 2007.
3. Parkin DM. Global cancer statistics in the year 2000. *Lancet Oncol* 2001;2:533–543. [PubMed: 11905707]
4. Parkin DM, Bray FI, Devesa SS. Cancer burden in the year 2000. The global picture. *Eur J Cancer* 2001;37(Suppl 8):S4–S66. [PubMed: 11602373]
5. El-Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology* 2007;132:2557–2576. [PubMed: 17570226]
6. McGlynn KA, Rosvold EA, Lustbader ED, Hu Y, Clapper ML, Zhou T, et al. Susceptibility to hepatocellular carcinoma is associated with genetic variation in the enzymatic detoxification of aflatoxin B1. *Proc Natl Acad Sci U S A* 1995;92:2384–2387. [PubMed: 7892276]
7. Donato F, Gelatti U, Chiesa R, Albertini A, Bucella E, Boffetta P, et al. A case-control study on family history of liver cancer as a risk factor for hepatocellular carcinoma in North Italy. *Brescia HCC Study. Cancer Causes Control* 1999;10:417–421. [PubMed: 10530612]
8. Shih WL, Yu MW, Chen PJ, Yeh SH, Lo MT, Chang HC, et al. Localization of a susceptibility locus for hepatocellular carcinoma to chromosome 4q in a hepatitis B hyperendemic area. *Oncogene* 2006;25: 3219–3224. [PubMed: 16407824]
9. Qi P, Chen YM, Wang H, Fang M, Ji Q, Zhao YP, et al. –509C>T polymorphism in the TGF-beta1 gene promoter, impact on the hepatocellular carcinoma risk in Chinese patients with chronic hepatitis B virus infection. *Cancer Immunol Immunother* 2009;58:1433–1440. [PubMed: 19169878]
10. Tanabe KK, Lemoine A, Finkelstein DM, Kawasaki H, Fujii T, Chung RT, et al. Epidermal growth factor gene functional polymorphism and the risk of hepatocellular carcinoma in patients with cirrhosis. *JAMA* 2008;299:53–60. [PubMed: 18167406]
11. Aravalli RN, Steer CJ, Cressman EN. Molecular mechanisms of hepatocellular carcinoma. *HEPATOLOGY* 2008;48:2047–2063. [PubMed: 19003900]
12. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 2004;5:557–572. [PubMed: 15475419]

13. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–575. [PubMed: 17701901]
14. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004;74:106–120. [PubMed: 14681826]
15. Beal S Fisher's hypergeometric test for a comparison in a finite population. *Am Stat* 1976;30:165–168.
16. Schaeffeler E, Schwab M, Eichelbaum M, Zanger UM. CYP2D6 genotyping strategy based on gene copy number determination by TaqMan real-time PCR. *Hum Mutat* 2003;22:476–485. [PubMed: 14635107]
17. Dausset J, Cann H, Cohen D, Lathrop M, Lalouel JM, White R. Centre d'étude du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome. *Genomics* 1990;6:575–577. [PubMed: 2184120]
18. Kamatani Y, Wattanapokayakit S, Ochi H, Kawaguchi T, Takahashi A, Hosono N, et al. A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. *Nat Genet* 2009;41:591–595. [PubMed: 19349983]
19. Walker SM, Downes CP, Leslie NR. TPIP: a novel phosphoinositide 3-phosphatase. *Biochem J* 2001;360:277–283. [PubMed: 11716755]
20. Daub H, Blencke S, Habenberger P, Kurtenbach A, Dennenmoser J, Wissing J, et al. Identification of SRPK1 and SRPK2 as the major cellular protein kinases phosphorylating hepatitis B virus core protein. *J Virol* 2002;76:8124–8137. [PubMed: 12134018]
21. Zheng Y, Fu XD, Ou JH. Suppression of hepatitis B virus replication by SRPK1 and SRPK2 via a pathway independent of the phosphorylation of the viral core protein. *Virology* 2005;342:150–158. [PubMed: 16122776]
22. Dorner D, Vlcek S, Foeger N, Gajewski A, Makolm C, Gotzmann J, et al. Lamina-associated polypeptide 2alpha regulates cell cycle progression and differentiation via the retinoblastoma-E2F pathway. *J Cell Biol* 2006;173:83–93. [PubMed: 16606692]
23. Pekovic V, Harborth J, Broers JL, Ramaekers FC, van Engelen B, Lammens M, et al. Nucleoplasmic LAP2alpha-lamin A complexes are required to maintain a proliferative state in human fibroblasts. *J Cell Biol* 2007;176:163–172. [PubMed: 17227891]
24. Parise P, Finocchiaro G, Masciadri B, Quarto M, Francois S, Mancuso F, et al. Lap2alpha expression is controlled by E2F and deregulated in various human tumors. *Cell Cycle* 2006;5:1331–1341. [PubMed: 16760672]
25. Aarts M, Dannenberg H, deLeeuw RJ, van Nederveen FH, Verhofstad AA, Lenders JW, et al. Microarray-based CGH of sporadic and syndrome-related pheochromocytomas using a 0.1–0.2 Mb bacterial artificial chromosome array spanning chromosome arm 1p. *Genes Chromosomes Cancer* 2006;45:83–93. [PubMed: 16215979]
26. Han W, Han MR, Kang JJ, Bae JY, Lee JH, Bae YJ, et al. Genomic alterations identified by array comparative genomic hybridization as prognostic markers in tamoxifen-treated estrogen receptor-positive breast cancer. *BMC Cancer* 2006;6:92. [PubMed: 16608533]
27. Poetsch M, Dittberner T, Woenckhaus C. Microsatellite analysis at 1p36.3 in malignant melanoma of the skin: fine mapping in search of a possible tumour suppressor gene region. *Melanoma Res* 2003;13:29–33. [PubMed: 12569282]
28. Rajgopal A, Carr IM, Leek JP, Hodge D, Bell SM, Roberts P, et al. Detection by fluorescence in situ hybridization of microdeletions at 1p36 in lymphomas, unidentified on cytogenetic analysis. *Cancer Genet Cytogenet* 2003;142:46–50. [PubMed: 12660032]
29. Jeon JP, Shim SM, Nam HY, Baik SY, Kim JW, Han BG. Copy number increase of 1p36.33 and mitochondrial genome amplification in Epstein-Barr virus-transformed lymphoblastoid cell lines. *Cancer Genet Cytogenet* 2007;173:122–130. [PubMed: 17321327]
30. Acuna-Castillo C, Aravena M, Leiva-Salcedo E, Perez V, Gomez C, Sabaj V, et al. T-kininogen, a cystatin-like molecule, inhibits ERK-dependent lymphocyte proliferation. *Mech Ageing Dev* 2005;126: 1284–1291. [PubMed: 16140359]

31. Rudolph MG, Stanfield RL, Wilson IA. How TCRs bind MHCs, peptides, and coreceptors. *Annu Rev Immunol* 2006;24:419–466. [PubMed: 16551255]
32. Hirschfield GM, Liu X, Xu C, Lu Y, Xie G, Gu X, et al. Primary biliary cirrhosis associated with HLA, IL12A, and IL12RB2 variants. *N Engl J Med* 2009;360:2544–2555. [PubMed: 19458352]
33. Pentcheva-Hoang T, Corse E, Allison JP. Negative regulators of T-cell activation: potential targets for therapeutic intervention in cancer, autoimmune disease, and persistent infections. *Immunol Rev* 2009;229:67–87. [PubMed: 19426215]
34. Ormandy LA, Hillemann T, Wedemeyer H, Manns MP, Greten TF, Korangy F. Increased populations of regulatory T cells in peripheral blood of patients with hepatocellular carcinoma. *Cancer Res* 2005;65: 2457–2464. [PubMed: 15781662]
35. Unitt E, Rushbrook SM, Marshall A, Davies S, Gibbs P, Morris LS, et al. Compromised lymphocytes infiltrate hepatocellular carcinoma: the role of T-regulatory cells. *HEPATOLOGY* 2005;41:722–730. [PubMed: 15791620]
36. Mantovani A, Allavena P, Sica A, Balkwill F. Cancer-related inflammation. *Nature* 2008;454:436–444. [PubMed: 18650914]
37. Park EJ, Lee JH, Yu GY, He G, Ali SR, Holzer RG, et al. Dietary and genetic obesity promote liver inflammation and tumorigenesis by enhancing IL-6 and TNF expression. *Cell*;140:197–208. [PubMed: 20141834]
38. Mauad TH, van Nieuwkerk CM, Dingemans KP, Smit JJ, Schinkel AH, Notenboom RG, et al. Mice with homozygous disruption of the *mdr2* P-glycoprotein gene. A novel animal model for studies of non-suppurative inflammatory cholangitis and hepatocarcinogenesis. *Am J Pathol* 1994;145:1237–1245. [PubMed: 7977654]

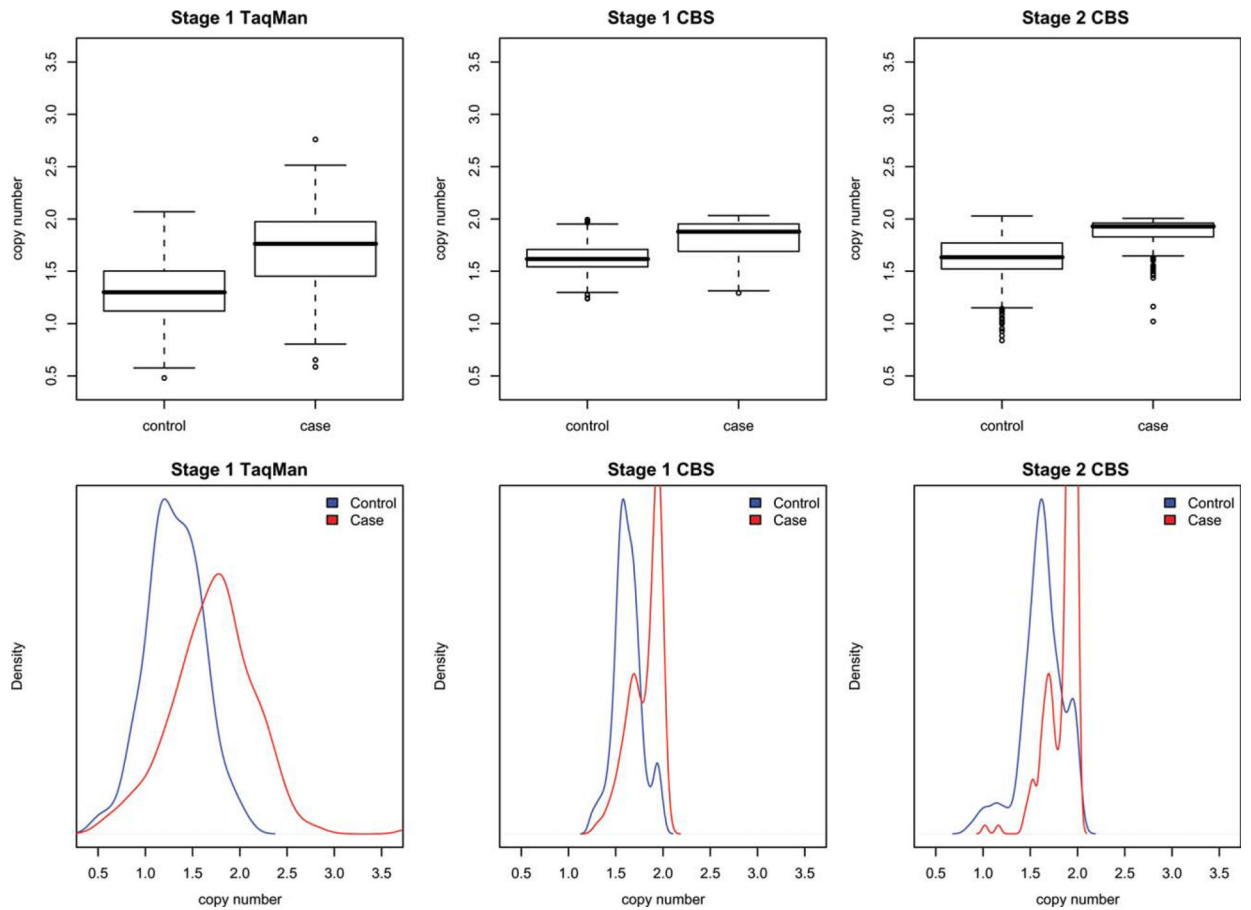


Fig. 1. Copy number distributions for the TCR-gamma locus. Top panels show boxplots of copy number at the TRG@ locus for cases and controls as determined by quantitative PCR assay (TaqMan) or the Affymetrix SNP6.0 assay (CBS). Lower panels show the corresponding density distributions. Blue lines indicate copy number for controls; red lines indicate data for cases.

TRA@ CNV Patterns in Stage 1 Samples

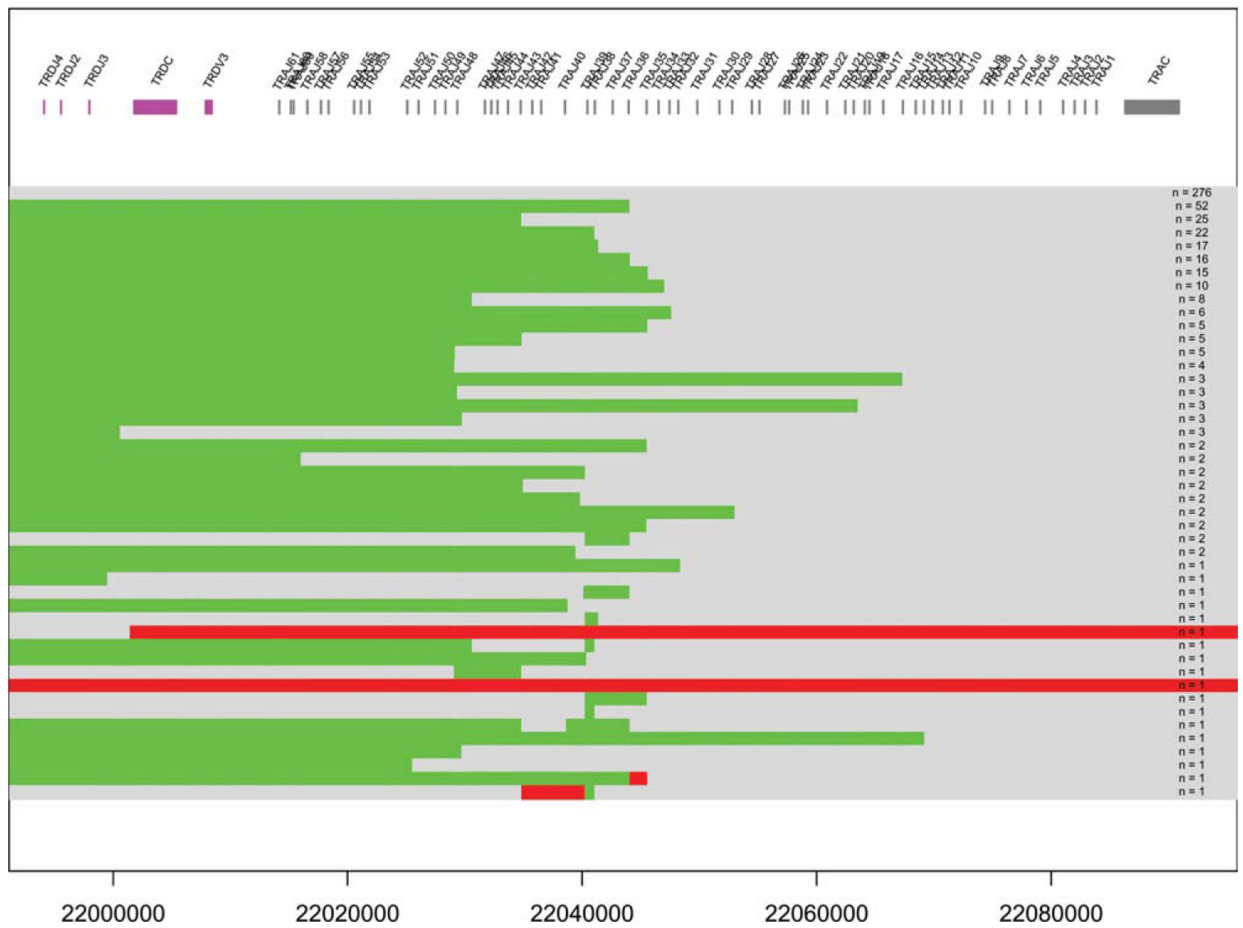


Fig. 2. Patterns of copy number alterations in TRA@. A portion of the TCR-alpha complex is shown at the top of the figure. TCR-delta loci are shaded magenta; TCR-alpha loci are shaded gray. Green bars indicate regions of low copy number observed in one or more case or control sample. Red bars indicate regions of high copy number observed in one or more sample. Chromosome 14 nucleotide positions are at the bottom of the figure.

Table 1.

Copy Number Variation Associated with HCC

Loc	Location	Chr	Start	End	Stage	FET P-value	Copy Number	Control Frequency	Case Frequency	OR-Fisher (95% CI)
1p36.33	NA	1	554484	557616	I	5.94E-16	low	0.018	0.250	17.63 (6.80 – 58.24)
KNG1	Exon	3	187932059	187932610	II	1.11E-10**	Low	0.009	0.141	17.01 (5.16 – 88.42)
					I	6.62E-05	Low	0.081	0.006	0.06 (0.00 – 0.40)
C4orf29, LARP2	Exon	4	129150816	129254964	II	5.31E-07*	Low	0.161	0.029	0.16 (0.05 – 0.37)
					I	1.05E-05	Low	0.162	0.033	0.18 (0.06 – 0.43)
ALDH7A1, PHAX, C5orf48, LMNB1	Exon	5	125912833	126165122	II	2.06E-05*	Low	0.095	0.010	0.09 (0.01 – 0.38)
					I	1.87E-08	Low	0.199	0.028	0.12 (0.04 – 0.29)
TRG@	Exon	7	38253040	38369021	II	9.53E-07*	Low	0.127	0.015	0.10 (0.02 – 0.33)
					I	3.16E-21	Low	0.727	0.278	0.15 (0.09 – 0.22)
SRPK2, PUS7	Exon	7	104782695	104892037	II	1.85E-28**	Low	0.617	0.141	0.10 (0.06 – 0.16)
					I	2.33E-05	Low	0.122	0.017	0.12 (0.02 – 0.40)
TMPO	Exon	12	97464439	97468567	II	6.71E-06*	Low	0.177	0.049	0.24 (0.11 – 0.48)
					I	8.11E-06	Low	0.085	0.000	0.00 (0.00 – 0.24)
TRA@	Exon	14	21480335	22041105	II	2.06E-05*	Low	0.095	0.010	0.09 (0.01 – 0.38)
					I	1.94E-16	Low	0.631	0.244	0.19 (0.12 – 0.29)
					II	6.24E-28**	Low	0.563	0.107	0.09 (0.05 – 0.15)

* P-value < 0.05 after Bonferroni adjustment for 2,318 comparisons.

** P-value < 2.6×10^{-7} after Bonferroni adjustment for 2,318 comparisons.

Table 2.

Copy Number Variation Associated with LC

Loci	Location	Chr	Start	End	Stage	FET P-value	Copy Number	Control Frequency	Case Frequency	OR Fisher (95% CI)
1p36.33	NA	1	554484	557616	I+II	6.03E-09*	Low	0.014	0.163	13.97 (5.26 – 39.84)
6q12	NA	6	67741207	67744804	I+II	2.46E-10*	Low	0.099	0.384	5.66 (3.27 – 9.73)
TRG@	Exon	7	38260614	38337317	I+II	3.17E-13*	Low	0.685	0.267	0.17 (0.10 – 0.29)
TRA@	Exon	14	21616185	22038824	I+II	5.66E-12*	Low	0.595	0.221	0.19 (0.11 – 0.34)
KNIG1	Exon	3	187932059	187932610	I+II	0.0302	n.d.	n.d.	n.d.	n.d.
C4orf29, LARP2	Exon	4	129150816	129254964	I+II	0.0103	n.d.	n.d.	n.d.	n.d.
ALDH7A1, PHAX, C5orf48, LMNB1	Exon	5	125912833	126165122	I+II	0.0009	n.d.	n.d.	n.d.	n.d.
SRPK2, PUS7	Exon	7	104782695	104892037	I+II	0.0021	n.d.	n.d.	n.d.	n.d.
TMPO	Exon	12	97464439	97468567	I+II	0.0009	n.d.	n.d.	n.d.	n.d.

* P-value 0.0025 after Bonferroni adjustment for 422,062 comparisons.

Table 3. Top 10 SNPs Associated with Hepatocellular Carcinoma (HCC) in the Replication Study

SNP	Chr	Position	Closest Gene	Distance to Gene	Stage	MAF		P-value	
						Control	Case	OR (95% CI)	Logistic Trend
rs4678680 (G>T)	3	32995039	GLB1	18kb upstream	I	0.06	0.15	2.56 (1.63–4.00)	1×10 ⁻⁴
					II	0.07	0.14	2.06 (1.37–3.11)	2×10 ⁻³
					Combined	0.07	0.14	2.27 (1.68–3.08)	6×10 ⁻⁷
rs12663434 (A>C)	6	90790803	BACH2	Intron 5	I	0.25	0.18	0.68 (0.49–0.95)	5×10 ⁻³
					II	0.22	0.13	0.53 (0.37–0.75)	2×10 ⁻⁴
					Combined	0.23	0.15	0.60 (0.48–0.77)	3×10 ⁻⁵
rs7749730 (G>A)	6	90791705	BACH2	Intron 5	I	0.25	0.18	0.65 (0.47–0.91)	2×10 ⁻³
					II	0.22	0.13	0.53 (0.37–0.75)	2×10 ⁻⁴
					Combined	0.23	0.15	0.59 (0.46–0.75)	1×10 ⁻⁵
rs9444730 (G>T)	6	90790058	BACH2	Intron 5	I	0.25	0.18	0.68 (0.49–0.95)	6×10 ⁻³
					II	0.23	0.13	0.50 (0.35–0.71)	8×10 ⁻⁵
					Combined	0.24	0.15	0.59 (0.46–0.75)	1×10 ⁻⁵
rs9267673 (T>C)	6	31991658	C2	12kb upstream	I	0.07	0.14	2.02 (1.31–3.14)	2×10 ⁻³
					II	0.08	0.15	1.94 (1.31–2.86)	4×10 ⁻⁴
					Combined	0.08	0.15	1.97 (1.47–2.64)	3×10 ⁻⁶
rs2647073 (C>A)	6	32681992	HLA-DRB1 *	16kb downstream	I	0.05	0.10	2.03 (1.22–3.36)	4×10 ⁻³
					II	0.07	0.12	1.89 (1.23–2.90)	2×10 ⁻³
					Combined	0.06	0.11	1.94 (1.40–2.69)	7×10 ⁻⁵
rs3997872 (A>T)	6	32688595	HLA-DRB1 †	23kb downstream	I	0.11	0.05	0.44 (0.26–0.76)	3×10 ⁻³
					II	0.11	0.06	0.49 (0.30–0.79)	2×10 ⁻³
					Combined	0.11	0.05	0.47 (0.33–0.67)	4×10 ⁻⁵
rs12100561 (A>G)	14	89370788	C14orf143	Intron 5	I	0.40	0.49	1.45 (1.11–1.90)	4×10 ⁻³
					II	0.38	0.49	1.58 (1.23–2.04)	3×10 ⁻⁴
					Combined	0.39	0.49	1.52 (1.26–1.83)	5×10 ⁻⁶
rs4509055 (G>A)	5	91126225	Intergenic		I	0.39	0.31	0.68 (0.52–0.91)	3×10 ⁻³
					II	0.30	0.40	1.54 (1.19–2.00)	1×10 ⁻³

SNP	Chr	Position	Closest Gene	Distance to Gene	Stage	MAF		OR (95% CI)	P-value	
						Control	Case		Logistic	Trend
rs8130508 (G>A)	21	17971736	Intergenic		Combined	0.35	0.36	1.06 (0.87–1.28)	0.57	0.57
					I	0.37	0.47	1.49 (1.14–1.95)	6×10^{-3}	
					II	0.43	0.34	0.67 (0.52–0.87)	9×10^{-4}	
				Combined	0.41	0.40	0.98 (0.81–1.18)	0.81	0.81	

* The SNP is in LD with the upstream genes including HLA-DRB1.

[†]The SNP is in LD with the genes downstream of HLA-DRB1.

Table 4. SNPs That Distinguish Hepatocellular Carcinoma (HCC) from Liver Cirrhosis (LC)

SNP	Chr	Position	Closest Gene	Distance to Gene	Stage	Minor Allele Frequency		OR (95% CI)	P-value
						LC	HCC		
rs2880301 (T>C)	13	18998534	TPTE2	Intron 1	Combined	0.475	0.198	0.273 (0.191–0.390)	1.74×10 ⁻¹²
rs2551677 (G>A)	2	118114059	DDX18	175kb upstream	Combined	0.118	0.311	3.383 (2.070–5.529)	1.41×10 ⁻¹⁰

Table 5. Biological Pathways Identified by Gene-Enrichment Analysis of Significant SNPs

Pathway	P-value			Significant Genes in the Pathways*
	Stage I	Stage II	Combined	
Antigen processing and presentation (Kegg)	0.035	8×10^{-3}	1×10^{-11}	<i>HLA-A</i> , <i>HLA-B</i> , <i>HLA-C</i> , <i>HLA-DOA</i> , <i>HLA-DOB</i> , <i>HLA-DQA1</i> , <i>HLA-DQA2</i> , <i>HLA-DQB1</i> , <i>HLA-DRA</i> , <i>HLA-DRB1</i> , <i>HLA-DRB5</i> , <i>HLA-E</i> , <i>HLA-F</i> , <i>HLA-G</i> , <i>LTA</i> , <i>TAP1</i> , <i>TAP2</i>
Cell adhesion molecules (CAMs) (Kegg)	0.042	1×10^{-3}	4×10^{-3}	<i>CD58</i> , <i>HLA-A</i> , <i>HLA-B</i> , <i>HLA-C</i> , <i>HLA-DOA</i> , <i>HLA-DOB</i> , <i>HLA-DQA1</i> , <i>HLA-DQA2</i> , <i>HLA-DQB1</i> , <i>HLA-DRA</i> , <i>HLA-DRB1</i> , <i>HLA-DRB5</i> , <i>HLA-E</i> , <i>HLA-F</i> , <i>HLA-G</i> , <i>NRCAM</i> , <i>NRXN1</i> , <i>NRXN2</i>
Antigen processing and presentation (Biocarta)	8×10^{-4}	2×10^{-3}	6×10^{-6}	<i>HLA-A</i> , <i>HLA-DRA</i> , <i>HLA-DRB1</i> , <i>TAP1</i> , <i>TAP2</i>
Classical complement pathway (Biocarta)	0.023	4×10^{-3}	2×10^{-5}	<i>C1QB</i> , <i>C2</i> , <i>C4A</i> , <i>C8A</i> , <i>C8B</i>
Lectin induced complement pathway (Biocarta)	0.018	0.034	2×10^{-4}	<i>C2</i> , <i>C4A</i> , <i>C8A</i> , <i>C8B</i>

* Genes identified in Stage I are italic and those found in Stage 2 are underlined. Comp--retain UND and IT.