國立台灣大學公共衛生學院流行病學研究所

碩士論文

Graduate Institute of Epidemiology

College of Public Health

National Taiwan University

Master Thesis

染色體 1p32.2-36.1 區域肝癌易感受基因連鎖高峰之定位：候選基因策略

Identification of Hepatocellular Carcinoma Susceptibility Gene by Peakwide Mapping on Chromosome 1p32.2-36.1: Application of Positional Candidate Gene Approach

研究生：王雅蕙

Student: Ya-Hui Wang

指導教授：于明暉　博士

Adviser: Ming-Whei Yu, Ph.D.

中華民國九十七年七月

July, 2008

致謝

碩士期間我遇到了一位非常好的指導教授—于明暉老師，于老師十分有耐心的帶領著初入流病領域的我，每次討論 1~2 小時似乎是司空見慣，遇到挫折時老師總會適時的安撫，讓我能夠調適心情後，重新振作，至今回想起來，仍是感動萬分。另外特別感謝范盛娟老師、熊昭老師、簡國龍老師、高嘉宏醫師，在論文上給了許多寶貴的意見。再來要感謝我們實驗室最尊敬的大學長惟量學長，沒有他，我們大概畢不了業，不論大小麻煩，他總能幫我們解決。還要謝謝同在 522 室一起努力奮鬥的伙伴：可愛的助理們—耘喬、曉玲、淑瑜和貞萍，及杞蓉學姐；學弟妹們—世韻、宛瑩、瑋怡、俐婷、哲右；最親愛的同學們--珮文、豐仔、家綺、宣豪、志偉，因為有你們的陪伴，總能讓陷入沮喪的我，恢復愉快的心情，不論是在實驗上、生活上在你們的身上，我真的學習到許多。特別感謝靜芬學姐的協助，她之前連鎖分析研究結果，是能夠完成這份碩士論文的重要基石。

此外，544 碩研室常見的固定班底—宜瑾、季侑、昌駿、韻瑾、琳娜、舒涵及雅堂學長、德天學姐，在這段碩士生活內受到你們不少照顧，讓生活充滿了快樂的回憶。另外要感謝伯璋學長在電腦方面的協助，使得分析得以順利進行。還有最重要的機器手和 ABI--7900，雖然偶爾會鬧脾氣，但還是順利完成最重要的實驗。這一路上不斷陪伴著我—雅茹、佳幸、珮怡、琇、小薰、Karrie 等眾多好友們，在悲傷難過時聽我哭訴、在遭遇挫敗時給我鼓勵、在高興歡愉時替我慶祝，因為有妳們，讓我總能擦乾眼淚後再度努力。

最後要謝謝我的家人，特別是我母親，要不是她的一再鼓勵，我想我應該不可能會堅持到最後。<夢想在左手，朋友是右手，未來才會變得有看頭>，在這兩年內我深深的感受的這歌詞的含意，再多的言語都無法完整表達出我心中的感謝，在此，僅能以此小小的成果獻給你們!

王雅蕙　　謹誌

中華民國九十七年七月

中文摘要

背景：在肝腫瘤組織的研究中已經常發現 1 號染色體短臂(1p)有高頻率染色體缺損現象。先前利用微衛星標記針對肝細胞癌多發病例家族為研究樣本，進行連鎖分析，顯示在 1p32.2-36.1 區域有高連鎖訊息。

材料與方法：在 1p32.2-36.1 連鎖區域中，我們應用候選基因策略挑選出 19 個候選基因，並利用單一核苷酸多型性標記(SNP)進行相關性分析。首先，以 240 家肝細胞癌病例家族為研究樣本，進行家族基礎的相關性研究；接著用一個獨立的病例對照樣本(包括 855 個病例和 875 個對照)，研究對象均為男性且 B 型肝炎病毒表面抗原(HBsAg)陽性者，進一步驗證先前在家族研究相關性發現。為考慮多重比較的問題，SNP 標記的顯著相關性將利用 false discovery rate (FDR)的 q 值進行控制。

結果：家族樣本中利用 PDT 分析，我們觀察到在 Haplotype block 有五個 SNPs 和肝細胞癌有顯著的相關性。其中 SNP13 位於 RBBP4 基因的 3'UTR 區域上，顯示有最強的相關性（nominal P=0.0047; empirical P=0.0025; q=0.0188）。病例對照研究中，在相同的 Haplotype block 上，除了 SNP13 再次被驗證和肝細胞癌有相關性外，尚有兩個 SNPs(SNP14 和 SNP15)也呈現顯著相關。此外，帶有 SNP13 的'C'對偶型（minor allele）者，會顯著的增加肝細胞癌危險性(odds ratio [95%信賴區間]: heterozygotes 為 1.36 [1.11-1.65] ; homozygotes 為 1.29 [0.90-1.84])。半型分析結果顯示：SNP13-SNP14-SNP15 中'C-A-C'半型與肝細胞癌呈現顯著相關（OR=1.32，95%信賴區間=1.06-1.63）。

結論：本研究結果顯示在家族和病例對照研究中，SNP13 和肝細胞癌均呈現一致的相關。

關鍵字：候選基因策略、家族研究、病例對照、關連性分析、肝癌易感受基因

Abstract

Background: In hepatocellular carcinomas (HCCs), frequent allelic loss on chromosome 1p has been reported. Using linkage analysis on multiplex families, a HCC-susceptibility locus has been mapped to a broad region of chromosome 1p32.2-36.1.

Materials and Methods: Here we have used a positional candidate gene strategy, based on association mapping with single nucleotide polymorphisms (SNPs) on 19 candidate genes within the linked region among 240 families with HCC, followed by a case-control analysis involving an independent set of 855 cases and 875 controls. Significance of the association was assessed by the false-discovery rate q value, which accounts for multiple testing.

Results: In the family sample, we observed a significant association between HCC and five single-nucleotide polymorphisms (SNPs) in a haplotype block by using the pedigree disequilibrium test. SNP 13, located in the 3' untranslated region (UTR) of the retinoblastoma binding protein 4 (RBBP4) gene, showed the strongest evidence (nominal P=0.0047; empirical P=0.0025; q=0.0188). Further case-control analysis confirmed the genetic association between SNP13 and HCC, and identified additional two SNPs in the same haplotype block. The C allele (minor allele) of SNP13 conferred an increased risk for HCC (odds ratio [95% confidence interval]: 1.36 [1.11-1.65] for heterozygotes; 1.29 [0.90-1.84] for homozygotes). SNP13 and two neighboring SNPs fell on a common haplotype ('C-A-C' at SNP13-SNP14-SNP15), which was also associated with an increase risk of HCC.

Conclusion:

SNP13 was consistently associated with HCC in both family and case-control sample.

Key words: Positional candidate gene approach、family study、case-control study、association study、HCC susceptibility gene

Context

Introduction

Hepatocellular carcinoma (HCC) is one of the most common malignant tumor worldwide (1). Persistent viral infections with hepatitis B virus (HBV) and/or hepatitis C virus (HCV) have been well documented as the most important etiological factors for HCC (2). In Taiwan where HCC incidence is high, the annual age-standardized incidence rate and mortality rate of HCC are 37.99 and 26.93 per 10^5 people, respectively (3). The population-attributable risk percentage was estimated as approximately 70% for HBV surface antigen (HBsAg) carrier status in Taiwan (4). Other environment risk factors, such as alcohol drinking, cigarette smoking, and dietary exposure to aflatoxins, have also been implicated in the etiology of HCC (5). They occur widely, and may modulate the risk of developing HCC associated with hepatitis virus infection (6, 7).

Besides environmental risk factors, family history has also been associated with an increased risk for HCC. According to a previous large-scale cohort study of 4808 male HBV surface antigen (HBsAg) carriers, a first-degree family history of HCC was statistically significantly associated with the incidence of HCC even after controlling other potential confounders (8). This finding suggests a rationale for studying the genetic component of HCC.

There have been many candidate gene association studies carried out with the use of the case-control study design. These include genes involved in the metabolism of xenobiotic carcinogens (9-12), androgen signaling (13), immunity (14), DNA repair (15), tumor-suppressor function (16), and liver regeneration (17). However, only a limited number of functional genetic variants as far have been identified to be associated with HCC. This highlights the importance of finding additional susceptibility genes by use of explorative approaches.

Genome-wide association (GWA) is increasingly a method of choice for systematic evaluation of genetic variants in complex disease (18). GWA studies are attractive because they do not rely on a prior knowledge of gene function. However, cost is still prohibitive for this approach. Linkage analysis and subsequent identification of causal genetic variants by association mapping provide an alternative approach.

Loss or gain of DNA fragments is critical in the pathogenesis of cancer. Using the Genome Imbalance Map (GIM) algorithm, which simultaneously detects DNA copy number alterations and loss of heterozygosity (LOH) events, it has been reported that gains of 1q, 5p, 5q, 6p, 7q, 8q, 17q and 20q, and LOH of 1p, 4q, 6q, 8p, 10q, 13q, 16p, 16q and 17p were significantly associated with HCC (19-22). Among these common genetic alterations, over 30% of reported HCCs display allelic loss on chromosome 1p. Tumor with this deletion showed an association with the early stage of liver carcinogenesis, suggesting a biological role for this deletion in tumor development (23). The common LOH region was mapped to 1p35-36, which is the consensus LOH locus in neuroblastomas (24, 25). LOH at 1p36 was associated with age at diagnosis and high-risk disease features, such as metastatic disease, unfavorable histopathology, and *MYCN* oncogene amplification in neuroblastoma patients (26).

To pinpoint the HCC-susceptibility locus on chromosome 1p, a linkage study of 71 Taiwanese multiplex families has been performed. The positive linkage findings appeared to be dispersed across 1p32.2-36.1 (27). In this region, there are 652 annotated genes (National Center for Biotechnology Information [NCBI]). The linkage region is large, extending over the 28 Mb between D1S1622 and D1S2742. Several approaches have been undertaken to identify susceptibility genes located in

linkage regions (28). We used the positional candidate gene strategy involving both family-based association analysis and case-control analysis. Candidate genes were chosen on the basis of their proven or suggested functional involvement in the pathogenesis of cancer. To our knowledge, of these, only mutY homolog (E. coli) (MUTYH) (29) and v-myc myelocytomatosis viral oncogene homolog 1, lung carcinoma derived (avian) (MYCL1) (30) have previously been tested for a genetic association with cancer.

Materials and Methods

Study Design

We initially used the family sample to identify SNPs associated with transmission disequilibrium of HCC. Then an independent set of case-control sample was used to seek supportive evidence for association findings discovered by the family sample.

Family data set

During 1997-2005, inpatients of HCC diagnosis at four major medical centers in northern Taiwan (National Taiwan University Hospital, Chang-Gung Memorial Hospital, Mackay Memorial Hospital, and Taipei Veterans General Hospital were recruited as probands. The diagnostic criteria of HCC was confirmed by: [1] either histological or cytological finding or [2] increased serum α-fetoprotein (AFP) ($\geqq$ 400 ng/ml) combined with at least one positive image on sonography, angiography, or/and computed tomography.

Families were collected through family-history questionnaire inquired from probands. Family members of probands were informed about the study, and those who were willing to participate in this study by clinical evaluation and questionnaire interview were recruited. There are 240 families in this data set, which consisted of 1,059 individuals (290 affected and 769 unaffected individuals) from 166 singleton families and 74 multiplex families.

Independent case-control data sets

HBsAg positive, male HCC cases were selected from the database of proband patients recruited from three hospitals: National Taiwan University Hospital, Chang-Gung Memorial Hospital, and Taipei Veterans General Hospital. The source of controls involved: [1] a cohort of male HBsAg carries follow-up study (31), in which subjects were recruited from Government Employee Center Clinics between 1988 and 1992 during regular health examinations and have been followed for incident HCC by

periodical clinical evaluate since 1997; [2] Outpatients who attended a special clinic for routine follow-up of HBsAg carries at Chang Gung Memorial Hospital. An essential inclusion criterion for eligible control subjects is ALT measured at two or more visits with an interval between visits at least one year, and sustained ALT normality. The control subjects were 1:1 frequency matched to case patients by sex, ethnic background (depend on patients' parental ethnicity) and birth year (within 5 years). There are 855 cases and 875 controls in this data set.

Questionnaires and DNA collection

Personal information was collected from medical records and interviews, which were carried out by trained research assistants. All participants signed informed consent form, and were asked to donate a 10 mL of blood sample and provide personal information according to a structured questionnaire, including items of socioeconomic demographic characteristics, lifestyle habits, and personal and family medical histories of major chronic diseases.

Genomic DNA was extracted from peripheral white blood cells or buccal brush samples (1145 cases and 1644 controls) by use of QLAamp DNA Blood Mini kit (Qiagen, Hilden, Germany) according to a standard protocol.

Positional candidate gene approach

The linkage region 1p32.2-36.1 is large, extending over 34.7 cM, which comprises 652 (115 found in OMIM (http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim) record (32)) annotated genes at NCBI website (http://www.ncbi.nlm.nih.gov). To identify the HCC-susceptibility gene at 1p32.2-36.1, positional candidate genes were selected. These genes were chosen by mining various database sources, including OMIM and OncoDB.HCC (http://oncodb.hcc.ibms.sinica.edu.tw) (33), based on their expression in the liver or their roles in carcinogenesis and/or hepatocellular carcinogenesis. A total of

19 candidate genes were thus selected (Table 1).

Identification of SNPs

At the first discovery stage, we selected at least one SNP in the intragenic sequences of each of the 19 candidate gene (Table 4). We identified SNPs with minor allele frequencies (MAF) $\geqq 0.1$, that were both informative (i.e., tagSNPs) and of high functional significance. SNPs with potential functions were selected from exonic/UTR regions from NCBI website and/or from PupaSuite (http://pupasuite.bioinfo.cipf.es) (34).

We selected a second series of SNPs to validate the initial results upon finding significant SNPs. SNPs investigated at this stage were tagSNPs selected through use of QuickSNP (http://bioinformoodics.jhmi.edu/quickSNP.pl) (35), PupaSuite and/or HapMap (http://www.hapmap.org/) (36), based on $r^2 > 0.8$. Only SNPs with a MAF $\geqq$ 0.05 in Asian population were selected (Table 4). Finally, there were 11 tagSNPs selected.

Genotyping

SNP genotyping was used real-time polymerase chain reaction (PCR), which was performed on ABI Prism 7900HT sequence detection system (Applied Biosystems). All probes of SNPs were designed form custom TaqMan® SNP genotyping assays. Each real-time PCR reaction (5 μL in total) comprised 2 μL genomic DNA, 2.5 μL TaqMan® Universal PCR Master Mix Buffer, 0.08 μL TaqMan probe, and 0.42 μL ddH_2O. Amplification was performed using the following conditions: 50 cycles of $50^{\circ}C$ for 2 min, at $95^{\circ}C$ for 10 min, at $95^{\circ}C$ for 15s, and at $60^{\circ}C$ for 1.5mins.

Genotypes were checked for Mendelian inheritance inconsistency using the PedCheck software (37). Families with Mendelian inheritance inconsistencies were dismissed from subsequent analysis.

Statistical analysis

Family-based study

Both the pedigree disequilibrium test (PDT) and family-based association test (FBAT) were used. The PDT test was performed with the UNPHASED software version 2.404 (http://www.mrc-bsu.cam.ac.uk/personal/frank/software/unphased/) (38), while the FBAT was carried out with the FBAT software version 1.7.3 (http://www.biostat.harvard.edu/~fbat/fbat.htm) (39). We used haplotype FBAT (HBAT) (40) and UNPHASED software to examine associations between haplotypes and HCC. For all association analyses in UNPHASED software, we calculated both nominal *P* values and empirical *P* values that derived from 10000 simulating replicates. Significance of the association was assessed by the false-discovery rate q value, which accounts for multiple testing.

LD and R_h^2

The Haploview software version 4.1 (http://www.broad.mit.edu/mpg/haploview/) (41) was used to compute pair-wise linkage disequilibrium (LD) (measured as Pearson correlation coefficient (r^2)) and predict haplotype block. In order to evaluate whether a set of tagSNPs within a block we chosen, the squared correlation coefficient R_h^2 were calculated by use of the program tagSNPs.exe (42).

Case-control analysis

All SNPs were checked for Hardy-Weinberg equilibrium (HWE) in the control group by use of the PowerMarker softwareversion 3.25 (http://statgen.ncsu.edu/powermarker/) (43). The χ^2 test and unconditional logistic regression model were used to analyze the relationships between genotypes or haplotypes and HCC and to calculate the odds ratios (ORs) and their 95% confidence intervals (CIs). All the analyses were conducted with use of the SAS version 9.1 (SAS Institute, Cary, NC) software package.

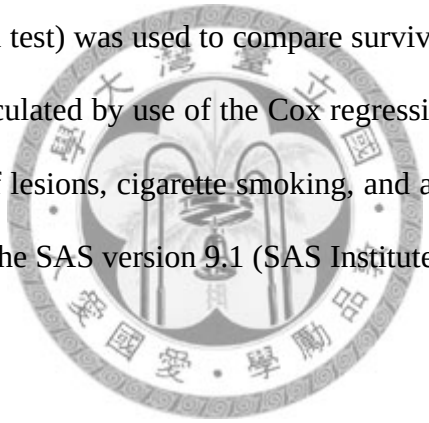
Joint Analysis

To increase statistical power and re-confirm association findings, we further

performed joint linkage and association analyses of a panel of 6 SNPs in the implicated region using the software packages Linkage and Association Modeling in Pedigrees (LAMP) version 0.0.9 (http://www.sph.umich.edu/csg/abecasis/lamp/index.html) (44) and Pseudomarker version 1.0.5 (http://www.helsinki.fi/~tsjuntun/pseudomarker/) (45). In this analysis, the family sample and the case-control sample were combined.

Survival analysis

Survival was calculated from the date of hospital admission to the last reported search for death entries (December 31, 2005) in the national death certification system. Patients whose causes of death were not due to HCC or cirrhosis were censored in the survival analysis. The Kaplan-Meier method was used to generate survival curves, and the log-rank test (or Wilcoxon test) was used to compare survival curves between groups. Hazard ratios (HRs) were calculated by use of the Cox regression model with adjustment for age, AFP level, number of lesions, cigarette smoking, and alcohol consumption. Data analysis was performed with the SAS version 9.1 (SAS Institute, Cary, NC).



Results

Characteristics of study family

There were 166 singleton families and 74 multiplex families. The characteristics of the 240 families are presented in Table 2. A total of 290 affected and 769 unaffected individuals were genotyped. The median age at diagnosis of HCC was 44 (range: 16-73) years. Among affected individuals, 87.9% were HBsAg positive and 11% were anti-HCV positive. Among unaffected individuals, 37.8% were HBsAg positive and 3.6% were anti-HCV positive. In addition to HBsAg positivity (P<0.0001) and anti-HCV positivity (P<0.0001), affected individuals also had a significantly higher prevalence of smoking (P<0.0001) and alcohol abuse (P<0.0001) (Table 2).

Characteristics of an independent set of case and control subjects

All subjects were HBsAg positive. The median age at diagnosis in cases was 51 years (range: 18 – 73 years). Cases (78.1%) were slightly more likely than controls (73.1%) to be Fukien Taiwanese (P=0.0172). Fifteen percent of the cases had a first-degree relative with HCC, while the corresponding figure is only 6.2% in controls (P<0.0001). Like the family sample, the prevalence of anti-HCV positivity (P<0.0001), cigarette smoking (P<0.0001), and alcohol consumption (P<0.0001) was significantly higher in the cases than in the controls (Table 3).

Characteristics of SNPs

At the first discovery stage, a total of 25 SNPs with a MAF>0.05 were selected by using in silico analysis of multiple websites, including PupaSuite, PolyPhen (http://genetics.bwh.harvard.edu/pph/) (46) , and Sorting Intolerant From Tolerant (SIFT) (http://blocks.fhcrc.org/sift/SIFT.html) (47). The selection of SNPs was prioritized on the basis of potential function predicted. After laboratory analysis, Two SNPs (rs7264 and rs3917980) were disregarded because of failing PCR. One (rs3134614) was excluded

from analysis due to no polymorphism in our study population (Table 4). No SNPs investigated were detected as departures from Hardy-Weinberg equilibrium by an exact test in the family sample or case-control sample.

Family-based association analysis

Table 5 presents the results from FBAT and PDT with the use of all 33 SNPs investigated. At the first stage of analysis with 22 SNPs, the PDT demonstrated strong evidence of association with HCC for SNP13 (rs9851), located in the 3' untranslated region (UTR) of the retinoblastoma binding protein 4 (RBBP4) gene (nominal P=0.0047; empirical P=0.0025; q=0.0188). Weaker evidence of association was found for a second SNP located within the 5'UTR of the aldo-keto reductase family 1 (AKR1A1) gene (for SNP27: nominal P=0.0380; empirical P=0.0379). Tests based on FBAT generally yielded lower level of significance (nominal P values were 0.0343 and 0.0233 for SNP13 and SNP27, respectively), although similar results were observed.

Using all HapMap SNPs (http://www.hapmap.org/cgi-perl/gbrowse/hapmap_B36/) with a MAF>0.05 derived from the Asian samples (Figure 1), we found that SNP13 resides within a large haplotype block extending from 32.72 Mb to 32.86 Mb over an interval of approximately 140 kb. To validate the initial association seen in the RBBP4 gene, we genotyped additional eleven tagSNPs in the implicated haplotype block, and found that four other SNPs (SNP11, SNP14, SNP16, and SNP17) continued to show evidence of association with HCC. Except one in RBBP4, three of these four HCC-associated SNPs lie in the neighboring syncoilin, intermediate filament 1 (SYNC1) gene (Table 4).

Case-Control Study

To replicate the findings from the family-based association analysis, we tested 6 SNPs representative of the implicated haplotype block in a second sample involving a

total of 855 cases and 875 controls. Three SNPs (SNP13, SNP14, and SNP15) were significantly associated with HCC. Notably, SNP13 was consistently associated with HCC in both data sets, i.e., the family sample and the case-control sample (Table 6).

Joint analysis

Since analyzing one large data set has more power in detecting genetic association for complex disease than is examining two smaller data sets separately as replication sets, joint analysis with the six SNPs for the initial family sample and the independent validation set of case-control sample combined was then performed. In the joint analysis, SNP13 was highly significant, irrespective of using Pseudomarker (P=0.0021 for LD under linkage; P= 0.0055 for joint linkage and LD) or LAMP (P=0.0086). Weaker association was found for SNP15 (from Pseudomarker: P=0.0158 for LD under linkage, P= 0.0314 for joint linkage and LD; from LAMP: P=0.0570), which was in LD with SNP13 (r^2=0.60). For SNP14, while no association with HCC was seen by Pseudomarker, the LAMP method showed association with a nominal P of 0.0340.

LD structure

Figure 2 illustrates the LD structure based on the 6 SNPs genotyped for both data sets. The 875 unrelated control subjects included in the case-control study were used for this analysis. As can be seen, SNP13 was in LD with SNPs SNP11 (r^2= 0.93), SNP12 (r^2=0.55), and SNP15 (r^2=0.60). This LD block extending from RBBP4 to SYNC1 and overlapping the two genes.

Haplotype analysis

The results of haplotype analysis are presented in Table 8. Haplotypes 2C and 3C were statistically significantly associated with HCC. The magnitude of the OR associated with carrying one copy of Hap1B-1C or Hap3C-3D was indistinguishable from that associated with carrying the allele ‘C” for SNP13 (Table 6). We also obtained a global P

value for transmission disequilibrium of haplotypes within the same haplotype block of 0.7355 by the PDT (global P=0.1324 from HBAT) (data not shown).

Survival analysis

Haplotype ‘C-A-C’ at SNP13-SNP14-SNP15 was marginally significantly associated with poor survival in the unselected cases that were included in the prior case-control analysis and had a tumor size ≦ 3 cm at hospital admission (P=0.0624). After adjusting age (continuous variable), serum AFP level (≦400 vs.＞400 ng/mL), number of lesions(solitary vs. multiple), cigarette smoking, and alcohol consumption, the hazard ratio of disease specific mortality was 1.46 (95% CI =1.00-2.13; P=0.0494) for haplotype ‘C-A-C’ as compared with other haplotypes. In contrast, no such association was observed among unselected cases with a tumor size＞3 cm (Figure 3).

Discussion

The search of HCC-susceptibility genes has been largely limited to a candidate-gene approach with the use of the case-control study design (11, 14, 17). The present study is unique because we aimed to discover HCC-susceptibility gene(s) via a positional candidate gene association study. Our study was motivated by a previous linkage study using 37 microsatellite markers spanning 1p on 71 multiplex families. That study found suggestive evidence of linkage in three regions: 1p35.3 (LOD score=4.14), 1p32.2-34.2 (LOD score=2.32), and 1p32.1 (LOD score=1.33) after adjusting for covariates including age at onset, sex, HBV genotype, and/or HBV viral load (27).

By examining the associations between SNPs in 19 candidate genes which are located within the linked region of 1p32.2-36.1 and HCC, we identified a SNP (rs9851) in the 3'UTR of the RBBP4 gene associated with HCC. This association was initially found by a family-based association study, which can avoid bias due to population stratification (48, 49). Following the initial study, we attempted replication of the genetic association of rs9851 with HCC in a case-control study that comprised subjects with no familial relationship between each other or with any of the families used in the family sample. The case-control analysis yielded a consistent association, and estimated that harboring the C allele of this SNP was associated with a 1.36 fold increase in the risk for HCC.

SNP rs9851 maps to a 140 kb LD block containing four genes and sixty-three SNPs with a MAF>0.05 listed in HapMap. In this LD block, four neighboring SNPs for rs9851 were also observed to be associated with HCC by the PDT. However, we have failed to find an association between these SNPs and HCC in the case-control analysis, though we identified two additional SNPs; one of which resides in the SYNC1 gene.

The discovery of associations with different set of SNPs does not indicate

non-replication. Indeed, the association of disease with a single allele in all data sets is not a universal observation for complex diseases (50). Given that coverage of known haplotypes encompassed 87% of alleles in the LD block, inconsistent results between studies may be only due to the unequal power for different designs. Low-powered attempts at replication conclude no evidence for association can be potentially misleading. Thus, our finding of multiple significant SNPs identified within a same LD block either with case-control study of unrelated subjects or with family-based association study adds support to the hypothesis that there is probable existence of causal variants in this region.

In this study, the most notable association with HCC was found for SNPs lying in or flanking the 3'UTR of RBBP4 that overlaps SYNC1. Based on the gene-wide SNP analyses, the FDR q value for the associations found in the RBBP4 gene ranged from 0.0188 to 0.0430, with 2 below the 0.2 threshold. RBBP4 was originally identified as a binding partner for the retinoblastoma protein (51, 52). Although its role in hepatocellular carcinogenesis has not been stressed in previous investigations, RBBP4 is a strong cancer candidate gene because it has been shown to be involved in several areas of chromatin metabolism, including nucleosome assembly as well as histone modification, and other important cellular processes (53, 54). Additionally, it is suggested that RBBP4 regulates cytoskeletal organization and morphology by increasing K-Ras activity and signaling through mitogen-activated protein kinase (51). Expression of RBBP4 was found to be dramatically reduced in cervical cancer-derived cells and human cervical cancer tissues. Down-regulation of RBBP4 in cervical mucosa epithelial cells is critical for determining the transforming activity of human papillomavirus 16 in inducing cervical cancer (55).

We found that SNP rs9851 in the 3'UTR of RBBP4 gave the strongest evidence of association with HCC. There was also some evidence of transmission disequilibrium with SNPs that were in LD with rs9851 in the PDT results or case-control analysis. Despite the

fact that rs9851 lies within a sequence conserved across different species, bioinformatic prediction revealed that it did not have any known functions.

However, 3'UTR sequences are important for the regulation of transcript cleavage and polyadenylation. Perturbations in 3'UTR-mediated regulation can lead to loss of control over one or more genes (56). On the other hand, 3'UTRs are preferential target sits for micro-RNA. Accumulating evidence suggests that genetic polymorphisms in the 3'UTR may affect gene and protein expression. For example, a 3'UTR SNP in the SLITRK1 gene strengthens an existing micro-RNA target binding site, thereby amplifying the down-regulation of SLITRK1; this polymorphism has been associated with Tourette syndrome (57).

Our study also has several limitations. First, screening the candidate genes in the linked region is substantial but not exhaustive. Second, only one-to-two SNPs were used at the discovery stage. Third, we did not perform resequencing or type all of the polymorphisms in the implicated LD block but, instead, relied on the LD between SNPs at this location. Therefore, our study could not conclusively discount other genes or genetic variants in 1p32.2-36.1, which may have contribution to the etiology of HCC. Fourth, replication study is usually necessary for the confirmation of an association finding. Only one of the 6 SNPs in the critical region was replicated in our data. This finding suggests that a larger sample size would be required for reliable replication. Finally, the case-control study was limited to men. Because there is a striking male-to-female sex ratio of HCC incidence, it is worthwhile to examine the association observed in this study among women.

References

1. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin 2005;55:74-108.
2. Lu SN, Su WW, Yang SS, Chang TT, Cheng KS, Wu JC, et al. Secular trends and geographic variations of hepatitis B virus and hepatitis C virus-associated hepatocellular carcinoma in Taiwan. Int J Cancer 2006;119:1946-52.
3. Taiwan BoHPDoHtEY. Cancer registry annual report, 2005 Taiwan. In. Taiwan; 2007.
4. Yu MW, You SL, Chang AS, Lu SN, Liaw YF, Chen CJ. Association between hepatitis C virus antibodies and hepatocellular carcinoma in Taiwan. Cancer Res 1991;51:5621-5.
5. Yu MC, Yuan JM. Environmental factors and risk for hepatocellular carcinoma. Gastroenterology 2004;127:S72-8.
6. Ross RK, Yuan JM, Yu MC, Wogan GN, Qian GS, Tu JT, et al. Urinary aflatoxin biomarkers and risk of hepatocellular carcinoma. Lancet 1992;339:943-6.
7. Yu MW, Chen CJ. Hepatitis B and C viruses in the development of hepatocellular carcinoma. Crit Rev Oncol Hematol 1994;17:71-91.
8. Yu MW, Chang HC, Liaw YF, Lin SM, Lee SD, Liu CJ, et al. Familial risk of hepatocellular carcinoma among chronic hepatitis B carriers and their relatives. J Natl Cancer Inst 2000;92:1159-64.
9. Yu MW, Gladek-Yarborough A, Chiamprasert S, Santella RM, Liaw YF, Chen CJ. Cytochrome P450 2E1 and glutathione S-transferase M1 polymorphisms and susceptibility to hepatocellular carcinoma. Gastroenterology 1995;109:1266-73.
10. Hsieh LL, Huang RC, Yu MW, Chen CJ, Liaw YF. L-myc, GST M1 genetic polymorphism and hepatocellular carcinoma risk among chronic hepatitis B carriers. Cancer Lett 1996;103:171-6.
11. Chen CJ, Yu MW, Liaw YF, Wang LW, Chiamprasert S, Matin F, et al. Chronic

hepatitis B carriers with null genotypes of glutathione S-transferase M1 and T1 polymorphisms who are exposed to aflatoxin are at increased risk of hepatocellular carcinoma. Am J Hum Genet 1996;59:128-34.

12. Yu MW, Chiu YH, Yang SY, Santella RM, Chern HD, Liaw YF, et al. Cytochrome P450 1A1 genetic polymorphisms and risk of hepatocellular carcinoma among chronic hepatitis B carriers. Br J Cancer 1999;80:598-603.
13. Yu MW, Cheng SW, Lin MW, Yang SY, Liaw YF, Chang HC, et al. Androgen-receptor gene CAG repeats, plasma testosterone levels, and risk of hepatitis B-related hepatocellular carcinoma. J Natl Cancer Inst 2000;92:2023-8.
14. Chen CC, Yang SY, Liu CJ, Lin CL, Liaw YF, Lin SM, et al. Association of cytokine and DNA repair gene polymorphisms with hepatitis B-related hepatocellular carcinoma. Int J Epidemiol 2005;34:1310-8.
15. Yu MW, Yang SY, Pan IJ, Lin CL, Liu CJ, Liaw YF, et al. Polymorphisms in XRCC1 and glutathione S-transferase genes and hepatitis B-related hepatocellular carcinoma. J Natl Cancer Inst 2003;95:1485-8.
16. Yu MW, Yang SY, Chiu YH, Chiang YC, Liaw YF, Chen CJ. A p53 genetic polymorphism as a modulator of hepatocellular carcinoma risk in relation to chronic liver disease, familial tendency, and cigarette smoking in hepatitis B carriers. Hepatology 1999;29:697-702.
17. 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.
18. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 2005;6:95-108.
19. Sheu JC, Lin YW, Chou HC, Huang GT, Lee HS, Lin YH, et al. Loss of heterozygosity and microsatellite instability in hepatocellular carcinoma in Taiwan.

Br J Cancer 1999;80:468-76.

20. Midorikawa Y, Yamamoto S, Ishikawa S, Kamimura N, Igarashi H, Sugimura H, et al. Molecular karyotyping of human hepatocellular carcinoma using single-nucleotide polymorphism arrays. Oncogene 2006;25:5581-90.
21. 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;120:1763-73.
22. Lu T, Hano H, Meng C, Nagatsuma K, Chiba S, Ikegami M. Frequent loss of heterozygosity in two distinct regions, 8p23.1 and 8p22, in hepatocellular carcinoma. World J Gastroenterol 2007;13:1090-7.
23. Sun M, Eshleman JR, Ferrell LD, Jacobs G, Sudilovsky EC, Tuthill R, et al. An early lesion in hepatic carcinogenesis: loss of heterozygosity in human cirrhotic livers and dysplastic nodules at the 1p36-p34 region. Hepatology 2001;33:1415-24.
24. Yeh SH, Chen PJ, Chen HL, Lai MY, Wang CC, Chen DS. Frequent genetic alterations at the distal region of chromosome 1p in human hepatocellular carcinomas. Cancer Res 1994;54:4188-92.
25. Weith A, Martinsson T, Cziepluch C, Bruderlein S, Amler LC, Berthold F, et al. Neuroblastoma consensus deletion maps to 1p36.1-2. Genes Chromosomes Cancer 1989;1:159-66.
26. Cohn SL, London WB, Huang D, Katzenstein HM, Salwen HR, Reinhart T, et al. MYCN expression is not prognostic of adverse outcome in advanced-stage neuroblastoma with nonamplified MYCN. J Clin Oncol 2000;18:3604-13.
27. Chuang CF. Linkage mapping hepatocellular-carcinoma susceptibility loci on chromosome 1p (Master thesis); 2006.
28. Zhu M, Zhao S. Candidate gene identification approach: progress and challenges. Int J Biol Sci 2007;3:420-7.

29. Ali M, Kim H, Cleary S, Cupples C, Gallinger S, Bristow R. Characterization of Mutant MUTYH Proteins Associated With Familial Colorectal Cancer. Gastroenterology 2008.
30. Spinola M, Falvella FS, Galvan A, Pignatiello C, Leoni VP, Pastorino U, et al. Ethnic differences in frequencies of gene polymorphisms in the MYCL1 region and modulation of lung cancer patients' survival. Lung Cancer 2007;55:271-7.
31. Yu MW, Yang YC, Yang SY, Cheng SW, Liaw YF, Lin SM, et al. Hormonal markers and hepatitis B virus-related hepatocellular carcinoma risk: a nested case-control study among men. J Natl Cancer Inst 2001;93:1644-51.
32. Hamosh A, Scott AF, Amberger J, Valle D, McKusick VA. Online Mendelian Inheritance in Man (OMIM). Hum Mutat 2000;15:57-61.
33. Su WH, Chao CC, Yeh SH, Chen DS, Chen PJ, Jou YS. OncoDB.HCC: an integrated oncogenomic database of hepatocellular carcinoma revealed aberrant cancer target genes and loci. Nucleic Acids Res 2007;35:D727-31.
34. Conde L, Vaquerizas JM, Dopazo H, Arbiza L, Reumers J, Rousseau F, et al. PupaSuite: finding functional single nucleotide polymorphisms for large-scale genotyping purposes. Nucleic Acids Res 2006;34:W621-5.
35. Grover D, Woodfield AS, Verma R, Zandi PP, Levinson DF, Potash JB. QuickSNP: an automated web server for selection of tagSNPs. Nucleic Acids Res 2007;35:W115-20.
36. Thorisson GA, Smith AV, Krishnan L, Stein LD. The International HapMap Project Web site. Genome Res 2005;15:1592-3.
37. O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet 1998;63:259-66.
38. Dudbridge F. Pedigree disequilibrium tests for multilocus haplotypes. Genet Epidemiol 2003;25:115-21.

39. Laird NM, Horvath S, Xu X. Implementing a unified approach to family-based tests of association. Genet Epidemiol 2000;19:S36-42.

40. Horvath S, Xu X, Lake SL, Silverman EK, Weiss ST, Laird NM. Family-based tests for associating haplotypes with general phenotype data: application to asthma genetics. Genet Epidemiol 2004;26:61-9.

41. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-5.

42. Stram DO, Haiman CA, Hirschhorn JN, Altshuler D, Kolonel LN, Henderson BE, et al. Choosing haplotype-tagging SNPS based on unphased genotype data using a preliminary sample of unrelated subjects with an example from the Multiethnic Cohort Study. Hum Hered 2003;55:27-36.

43. Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 2005;21:2128-9.

44. Li M, Boehnke M, Abecasis GR. Joint modeling of linkage and association: identifying SNPs responsible for a linkage signal. Am J Hum Genet 2005;76:934-49.

45. Goring HH, Terwilliger JD. Linkage analysis in the presence of errors IV: joint pseudomarker analysis of linkage and/or linkage disequilibrium on a mixture of pedigrees and singletons when the mode of inheritance cannot be accurately specified. Am J Hum Genet 2000;66:1310-27.

46. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic Acids Res 2002;30:3894-900.

47. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res 2001;11:863-74.

48. Bertram L, Hiltunen M, Parkinson M, Ingelsson M, Lange C, Ramasamy K, et al. Family-based association between Alzheimer's disease and variants in UBQLN1. N Engl J Med 2005;352:884-94.

49. Douglas JA, Zuhlke KA, Beebe-Dimmer J, Levin AM, Gruber SB, Wood DP, et al. Identifying susceptibility genes for prostate cancer--a family-based association study of polymorphisms in CYP17, CYP19, CYP11A1, and LH-beta. Cancer Epidemiol Biomarkers Prev 2005;14:2035-9.
50. Rogaeva E, Meng Y, Lee JH, Gu Y, Kawarai T, Zou F, et al. The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. Nat Genet 2007;39:168-77.
51. Scuto A, Zhang H, Zhao H, Rivera M, Yeatman TJ, Jove R, et al. RbAp48 regulates cytoskeletal organization and morphology by increasing K-Ras activity and signaling through mitogen-activated protein kinase. Cancer Res 2007;67:10317-24.
52. Qian YW, Wang YC, Hollingsworth RE, Jr., Jones D, Ling N, Lee EY. A retinoblastoma-binding protein related to a negative regulator of Ras in yeast. Nature 1993;364:648-52.
53. Henikoff S. Versatile assembler. Nature 2003;423:814-5, 817.
54. Muller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, et al. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell 2002;111:197-208.
55. Kong L, Yu XP, Bai XH, Zhang WF, Zhang Y, Zhao WM, et al. RbAp48 is a critical mediator controlling the transforming activity of human papillomavirus type 16 in cervical cancer. J Biol Chem 2007;282:26381-91.
56. Chang H, Fujita T. PicSNP: a browsable catalog of nonsynonymous single nucleotide polymorphisms in the human genome. Biochem Biophys Res Commun 2001;287:288-91.
57. Abelson JF, Kwan KY, O'Roak BJ, Baek DY, Stillman AA, Morgan TM, et al. Sequence variants in SLITRK1 are associated with Tourette's syndrome. Science 2005;310:317-20.

58. Werner SR, Lee PA, DeCamp MW, Crowell DN, Randall SK, Crowell PL. Enhanced cell cycle progression and down regulation of p21(Cip1/Waf1) by PRL tyrosine phosphatases. Cancer Lett 2003;202:201-11.

59. Dumaual CM, Sandusky GE, Crowell PL, Randall SK. Cellular localization of PRL-1 and PRL-2 gene expression in normal adult human tissues. J Histochem Cytochem 2006;54:1401-12.

60. Wang Y, Li ZF, He J, Li YL, Zhu GB, Zhang LH. Expression of the human phosphatases of regenerating liver (PRLs) in colonic adenocarcinoma and its correlation with lymph node metastasis. Int J Colorectal Dis 2007;22:1179-84.

61. Barlat I, Maurier F, Duchesne M, Guitard E, Tocque B, Schweighoffer F. A role for Sam68 in cell cycle progression antagonized by a spliced variant within the KH domain. J Biol Chem 1997;272:3129-32.

62. Duran A, Linares JF, Galvez AS, Wikenheiser K, Flores JM, Diaz-Meco MT, et al. The signaling adaptor p62 is an important NF-kappaB mediator in tumorigenesis. Cancer Cell 2008;13:343-54.

63. Lazer G, Pe'er L, Schapira V, Richard S, Katzav S. The association of Sam68 with Vav1 contributes to tumorigenesis. Cell Signal 2007;19:2479-86.

64. Ahlemann M, Zeidler R, Lang S, Mack B, Munz M, Gires O. Carcinoma-associated eIF3i overexpression facilitates mTOR-dependent growth transformation. Mol Carcinog 2006;45:957-67.

65. Song H, Xia SL, Liao C, Li YL, Wang YF, Li TP, et al. Genes encoding Pir51, Beclin 1, RbAp48 and aldolase b are up or down-regulated in human primary hepatocellular carcinoma. World J Gastroenterol 2004;10:509-13.

66. Pacifico F, Paolillo M, Chiappetta G, Crescenzi E, Arena S, Scaloni A, et al. RbAp48 is a target of nuclear factor-kappaB activity in thyroid cancer. J Clin Endocrinol Metab 2007;92:1458-66.

67. Teufel A, Malik N, Mukhopadhyay M, Westphal H. Frcp1 and Frcp2, two novel fibronectin type III repeat containing genes. Gene 2002;297:79-83.
68. Wang H, Wang L, Erdjument-Bromage H, Vidal M, Tempst P, Jones RS, et al. Role of histone H2A ubiquitination in Polycomb silencing. Nature 2004;431:873-8.
69. Wolfler A, Erkeland SJ, Bodner C, Valkhof M, Renner W, Leitner C, et al. A functional single-nucleotide polymorphism of the G-CSF receptor gene predisposes individuals to high-risk myelodysplastic syndrome. Blood 2005;105:3731-6.
70. Chakraborty A, White SM, Lerner SP. Granulocyte colony-stimulating factor receptor signals for beta1-integrin expression and adhesion in bladder cancer. Urology 2004;63:177-83.
71. Chakraborty A, Guha S. Granulocyte colony-stimulating factor/granulocyte colony-stimulating factor receptor biological axis promotes survival and growth of bladder cancer cells. Urology 2007;69:1210-5.
72. Hasselblatt M, Jeibmann A, Riesmeier B, Maintz D, Schabitz WR. Granulocyte-colony stimulating factor (G-CSF) and G-CSF receptor expression in human ischemic stroke. Acta Neuropathol 2007;113:45-51.
73. Kastury K, Taylor WE, Shen R, Arver S, Gutierrez M, Fisher CE, et al. Complementary deoxyribonucleic acid cloning and characterization of a putative human axonemal dynein light chain gene. J Clin Endocrinol Metab 1997;82:3047-53.
74. Primdahl H, von der Maase H, Christensen M, Wolf H, Orntoft TF. Allelic deletions of Rb and L-myc in urine sediments from patients with bladder tumors or carcinoma in situ. Oncol Rep 2002;9:551-5.
75. Kambara T, Sharp GB, Nagasaka T, Takeda M, Sasamoto H, Nakagawa H, et al. Allelic loss of a common microsatellite marker MYCL1: a useful prognostic factor of poor outcomes in colorectal cancer. Clin Cancer Res 2004;10:1758-63.

76. Suzuki H, Yamashiro K. L-myc restriction fragment length polymorphism and histological pattern of invasion in lung adenocarcinoma. Oncol Rep 2002;9:345-7.
77. Baudino TA, McKay C, Pendeville-Samain H, Nilsson JA, Maclean KH, White EL, et al. c-Myc is essential for vasculogenesis and angiogenesis during development and tumor progression. Genes Dev 2002;16:2530-43.
78. Fujii T, Kawahara A, Basaki Y, Hattori S, Nakashima K, Nakano K, et al. Expression of HER2 and estrogen receptor alpha depends upon nuclear localization of Y-box binding protein-1 in human breast cancers. Cancer Res 2008;68:1504-12.
79. Shiota M, Izumi H, Onitsuka T, Miyamoto N, Kashiwagi E, Kidani A, et al. Twist promotes tumor cell growth through YB-1 expression. Cancer Res 2008;68:98-105.
80. Iizuka N, Oka M, Yamada-Okabe H, Mori N, Tamesa T, Okada T, et al. Differential gene expression in distinct virologic types of hepatocellular carcinoma: association with liver cirrhosis. Oncogene 2003;22:3007-14.
81. Liu H, Cheng EH, Hsieh JJ. Bimodal degradation of MLL by SCFSkp2 and APCCdc20 assures cell cycle execution: a critical regulatory circuit lost in leukemogenic MLL fusions. Genes Dev 2007;21:2385-98.
82. Kidokoro T, Tanikawa C, Furukawa Y, Katagiri T, Nakamura Y, Matsuda K. CDC20, a potential cancer therapeutic target, is negatively regulated by p53. Oncogene 2008;27:1562-71.
83. Tao H, Shinmura K, Suzuki M, Kono S, Mibu R, Tanaka M, et al. Association between genetic polymorphisms of the base excision repair gene MUTYH and increased colorectal cancer risk in a Japanese population. Cancer Sci 2008;99:355-60.
84. Kobayashi K, Shida A, Yamada H, Ishibashi Y, Nakayama R, Toriumi Y, et al. Frequent splicing aberration of the base excision repair gene hMYH in human gastric cancer. Anticancer Res 2008;28:215-21.

85. Barnetson RA, Devlin L, Miller J, Farrington SM, Slater S, Drake AC, et al. Germline mutation prevalence in the base excision repair gene, MYH, in patients with endometrial cancer. Clin Genet 2007;72:551-5.

86. Wang J, Ni H, Chen L, Liu YX, Chen CB, Song WQ. Preparation and analysis of cSNP chip on hepatocellular carcinoma-related genes. Hepatobiliary Pancreat Dis Int 2005;4:398-402.

87. Fromme JC, Banerjee A, Huang SJ, Verdine GL. Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. Nature 2004;427:652-6.

88. Hoshino I, Matsubara H, Hanari N, Mori M, Nishimori T, Yoneyama Y, et al. Histone deacetylase inhibitor FK228 activates tumor suppressor Prdx1 with apoptosis induction in esophageal cancer cells. Clin Cancer Res 2005;11:7945-52.

89. Kim HJ, Chae HZ, Kim YJ, Kim YH, Hwangs TS, Park EM, et al. Preferential elevation of Prx I and Trx expression in lung cancer cells following hypoxia and in human lung cancer tissues. Cell Biol Toxicol 2003;19:285-98.

90. Rabilloud T, Heller M, Gasnier F, Luche S, Rey C, Aebersold R, et al. Proteomics analysis of cellular response to oxidative stress. Evidence for in vivo overoxidation of peroxiredoxins at their active site. J Biol Chem 2002;277:19396-401.

91. Petrash JMA, N. H.; Sadana, I.; Srivastava, S. K. Biochemical and genetic interrelationship between aldose reductase, aldose reductase M and aldehyde reductase in human tissues.(Abstract) Am. J. Hum. Genet. 1981; 33.

92. Wang SS, Davis S, Cerhan JR, Hartge P, Severson RK, Cozen W, et al. Polymorphisms in oxidative stress genes and risk for non-Hodgkin lymphoma. Carcinogenesis 2006;27:1828-34.

93. Giang DK, Cravatt BF. Molecular characterization of human and mouse fatty acid amide hydrolases. Proc Natl Acad Sci U S A 1997;94:2238-42.

94. Di Marzo V, Melck D, Orlando P, Bisogno T, Zagoory O, Bifulco M, et al. Palmitoylethanolamide inhibits the expression of fatty acid amide hydrolase and enhances the anti-proliferative effect of anandamide in human breast cancer cells. Biochem J 2001;358:249-55.

95. Ligresti A, Bisogno T, Matias I, De Petrocellis L, Cascio MG, Cosenza V, et al. Possible endocannabinoid control of colorectal cancer growth. Gastroenterology 2003;125:677-87.

96. Imaoka S, Yoneda Y, Sugimoto T, Hiroi T, Yamamoto K, Nakatani T, et al. CYP4B1 is a possible risk factor for bladder cancer in humans. Biochem Biophys Res Commun 2000;277:776-80.

97. Hiratsuka M, Nozawa H, Konno Y, Saito T, Konno S, Mizugaki M. Human CYP4B1 gene in the japanese population analyzed by denaturing HPLC. Drug Metab Pharmacokinet 2004;19:114-9.

98. Powell PK, Wolf I, Lasker JM. Identification of CYP4A11 as the major lauric acid omega-hydroxylase in human liver microsomes. Arch Biochem Biophys 1996;335:219-26.

99. Yeh CS, Wang JY, Cheng TL, Juan CH, Wu CH, Lin SR. Fatty acid metabolism pathway play an important role in carcinogenesis of human colorectal cancers by Microarray-Bioinformatics analysis. Cancer Lett 2006;233:297-308.

100. Cho BH, Park BL, Kim LH, Chung HS, Shin HD. Highly polymorphic human CYP4A11 gene. J Hum Genet 2005;50:259-63.

101. Guan KL, Jenkins CW, Li Y, Nichols MA, Wu X, O'Keefe CL, et al. Growth suppression by p18, a p16INK4/MTS1- and p14INK4B/MTS2-related CDK6 inhibitor, correlates with wild-type pRb function. Genes Dev 1994;8:2939-52.

102. Luedde T, Tacke F, Manns MP, Trautwein C. p18(INK4c) expression in hepatocellular carcinoma. Hepatology 2005;41:405; author reply 406.

103. Arcellana-Panlilio MY, Egeler RM, Ujack E, Magliocco A, Stuart GC, Robbins SM, et al. Evidence of a role for the INK4 family of cyclin-dependent kinase inhibitors in ovarian granulosa cell tumors. Genes Chromosomes Cancer 2002;35:176-81.

104. Sanchez-Aguilera A, Delgado J, Camacho FI, Sanchez-Beato M, Sanchez L, Montalban C, et al. Silencing of the p18INK4c gene by promoter hypermethylation in Reed-Sternberg cells in Hodgkin lymphomas. Blood 2004;103:2351-7.

105. van Veelen W, van Gasteren CJ, Acton DS, Franklin DS, Berger R, Lips CJ, et al. Synergistic effect of oncogenic RET and loss of p18 on medullary thyroid carcinoma development. Cancer Res 2008;68:1329-37.

106. Huang H, Gallegos AM, Zhou M, Ball JM, Schroeder F. Role of the sterol carrier protein-2 N-terminal membrane binding domain in sterol transfer. Biochemistry 2002;41:12149-62.

107. Yamamoto R, Kallen CB, Babalola GO, Rennert H, Billheimer JT, Strauss JF, 3rd. Cloning and expression of a cDNA encoding human sterol carrier protein 2. Proc Natl Acad Sci U S A 1991;88:463-7.

Table 1. Gene name and function of 19 candidate genes in this study

Gene Symbol	Gene Name	Known function (reference)	Comprehensive searching				
			Gene expression in the liver	Carcinogenesis	Microarray*	OncoDB.HCC	
			Yes (reference) / No	Yes (reference) / No	Yes(reference) / No	Yes / No	Evidence
PTP4A2	protein tyrosine phosphatase type IVA, member 2	a regulator of cellular processes (58)	Yes (59)	Yes (60)	No	No	
KHDRBS1	KH domain containing, RNA binding, signal transduction associated 1	a regulator of cellular processes (61)	Yes (61)	Yes (62, 63)	No	Yes	Experiment
EIF3I	eukaryotic translation initiation factor 3, subunit I	affects mRNA translation process and protein synthesis (64)	No	Yes (64)	No	Yes	Experiment
RBBP4	retinoblastoma binding protein 4	affects chromatin assembly and regulates Ras signal pathway (51)	No	Yes (55, 65, 66)	Yes (65)	No	
FNDC5	fibronectin type III domain containing 5	a kind of cell adhesion molecules (67)	Yes(67)	No	Yes†	Yes	Stanford microarray
PHC2	polyhomeotic homolog 2 (Drosophila)	affects chromatin assembly, specifically monoubiquitinates H2A (68)	No	Yes (68)	Yes†	Yes	Stanford microarray
CSF3R	colony stimulating factor 3 receptor (granulocyte)	function in cell surface adhesion (69), or recognition processes (72)	No	Yes (70, 71)	No	No	
DNALI1	dynein, axonemal, light intermediate chain 1	a candidate gene for immotile cilia syndrome (ICS) (73)	No	No	Yes†	Yes	Stanford microarray
MYCL1	v-myc myelocytomatosis viral oncogene homolog 1, lung carcinoma derived(avian)	essential factor in tumor progression and development (77)	No	Yes (74-76)	No	No	
YBX1	Y box binding protein 1	induces tumor progression and promotes cell growth (79)	No	Yes (78, 79)	Yes (80)	Yes	Experiment

Table 1. (Continued)

Gene Symbol	Gene Name	Known function (reference)	Comprehensive searching				
			Gene expression in the liver	Carcinogenesis	Microarray	OncoDB.HCC	
			Yes (reference) / No	Yes (reference) / No	Yes(reference) / No	Yes / No	Evidence
CDC20	cell division cycle 20 homolog (S. cerevisiae)	a regulator in the cell cycle (81)	No	Yes (82)	Yes†	Yes	Stanford microarray
MUTYH	mutY homolog (E. coli)	an oxidative DNA damage repair enzymes (87)	No	Yes (83-85)	Yes (86)	No	
PRDX1	peroxiredoxin 1	a kind of antioxidant enzymes, may affects cancer progression (88, 90)	No	Yes (88, 89)	Yes†	Yes	at least 3 array data
AKR1A1	aldo-keto reductase family 1, member A1 (aldehyde reductase)	a kind of reduction of biogenic and xenobiotic aldehydes (91)	Yes (91)	Yes (92)	No	No	
FAAH	fatty acid amide hydrolase	to hydrolysis of fatty acid amides (93)	Yes (93)	Yes (94, 95)	No	No	
CYP4B1	cytochrome P450 (CYP), family 4,subfamily B, polypeptide 1	a kind of CYP enzymes, which affects drug metabolism and synthesis of cholesterol & lipids (97)	No	Yes (96)	No	No	
CYP4A11	cytochrome P450, family 4, subfamily A, (CYP4A) polypeptide 11	a kind of CYP4A enzymes, which involved in the metabolism of medium- and long-chain fatty acids (100)	Yes (98)	Yes (99)	No	No	
CDKN2C	cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4)	a regulator of cellular processes (101)	Yes (102)	Yes (103-105)	Yes†	Yes	Stanford microarray Experiment
SCP2	sterol carrier protein 2	an intracellular lipid transfer protein (106)	Yes (107)	?	Yes†	Yes	at least 3 array data

* : Genes significantly up- or down-regulated in HCC microarray/proteomic reports.

† : Refer to OncoDB.HCC website

Table 2. Characteristics of study subjects in the family sample

	Affecteds		Unaffecteds		P value
	No.	(%)	No.	(%)	
No. of subjects	290	(27.4)	769	(72.6)	
Age at diagnosis (years)					
Median	44				
Range	16 - 73				
Age at recruitment (years)					< 0.0001
Median	44.6		48.6		
Range	17 - 77		13 - 82		
Sex					< 0.0001
Male	249	(85.9)	312	(40.6)	
Female	41	(14.1)	457	(59.4)	
Ethnicity					0.3292
Fukien Taiwanese	219	(75.5)	551	(71.7)	
Hakka Taiwanese	23	(7.9)	77	(10.0)	
Mainland	25	(8.6)	58	(7.5)	
Other	23	(7.9)	83	(10.8)	
HBsAg status					< 0.0001
Positive	255	(87.9)	287	(37.8)	
Negative	35	(12.1)	473	(62.2)	
Missing	0		9		
Anti-HCV status					< 0.0001
Positive	30	(11.0)	27	(3.6)	
Negative	242	(89.0)	731	(96.4)	
Missing	18		11		
Smoking					< 0.0001
Yes	147	(50.7)	222	(28.9)	
No	143	(49.3)	545	(71.1)	
Missing	0		2		
Drinking					< 0.0001
Yes	84	(29.0)	129	(16.8)	
No	206	(71.0)	638	(83.2)	
Missing	0		2		
Alcoholism (≧80gm/day)					< 0.0001
Yes	34	(12.0)	17	(2.3)	
No	250	(88.0)	727	(97.7)	
Missing	6		25		

Table 3. Characteristics of study subjects in case-control analysis

	Cases (n=855)		Controls (n=875)		P value
	No.	(%)	No.	(%)	
Age at diagnosis (years)					
Median	51				
Range	18 - 73				
Birth year					0.0578
< 1940	205	(24.0)	170	(19.4)	
1940-1949	266	(31.1)	270	(30.9)	
1950-1959	229	(26.8)	276	(31.5)	
> 1960	155	(18.1)	159	(18.2)	
Ethnicity					0.0172
Fukien Taiwanese	668	(78.1)	640	(73.1)	
Hakka Taiwanese	97	(11.4)	102	(11.7)	
Mainland	56	(6.6)	93	(10.6)	
Other	34	(4.0)	40	(4.6)	
First-degree relatives with HCC					< 0.0001
Yes	130	(15.2)	54	(6.2)	
No	725	(84.8)	821	(93.8)	
Anti-HCV status					< 0.0001
Positive	79	(9.4)	98	(4.4)	
Negative	761	(90.6)	836	(95.6)	
Missing	15		1		
Smoking					< 0.0001
Yes	560	(65.5)	309	(35.3)	
No	295	(34.5)	566	(64.7)	
Missing	0		0		
Drinking					< 0.0001
Yes	340	(39.8)	160	(18.3)	
No	515	(60.2)	715	(81.7)	
Missing	0		0		
Alcoholism (≧80gm/day)					< 0.0001
Yes	107	(12.5)	16	(1.8)	
No	746	(87.5)	857	(98.2)	
Missing	2		2		

Table 4. Characteristics of 36 SNPs

SNP #	dbSNP rs number	Gene	Cytogentic location	Physical map location (bp)	SNP type	Alleles	MAF	Predicted function	Tolerance index			
									SIFT		PolyPhen	
								PupaSuit[§]	Score	Predicted	Score	Predicted
			D1S1622									
			D1S2832									
1	rs646689	PTP4A2	1p35	32144613	3'UTR	A/G	f(A) = 0.40	ESE / Mm cons.				
2	rs12094507	KHDRBS1	1p32	32273802	Intron 4	A/G	f(G) = 0.37					
3	rs3738002	KHDRBS1	1p32	32275017	Intron 4	A/T	f(A) = 0.18	Mm cons.				
4	rs16834931	EIF3I	1p34.1	32462108	Intron 2	A/G	f(G) = 0.18					
5	rs589034	EIF3I	1p34.1	32462444	Intron 3	A/G	f(A) = 0.34					
6	rs6678903*	ZBTB8	1p35.1	32835141	Intron 4	C/T	f(C) = 0.13					
7	rs704887*	ZBTB8	1p35.1	32838071	Intron 4	A/G	f(A) = 0.42					
8	rs704885*	ZBTB8	1p35.1	32839178	3' UTR	A/G	f(A) = 0.12					
9	rs3954228*	ZBTB8OS	1p35.1	32873402	Intron 1	C/T	f(T) = 0.14					
10	rs2934766*	RBBP4	1p35.1	32892180	Intron 2	A/G	f(A) = 0.13					
11	rs359955*	RBBP4	1p35.1	32907361	Intron 6	A/C	f(A) = 0.22					
12	rs359956*	RBBP4	1p35.1	32907576	Intron 7	A/G	f(A) = 0.11					
13	rs9851	RBBP4	1p35.1	32918329	3'UTR	C/T	f(C) = 0.26[†]	Mm cons.				
14	rs16835131*	SYNC1	1p33.4-33	32921522	Intron 3	A/G	f(A) = 0.10					
15	rs697147*	SYNC1	1p33.4-33	32924929	Intron 1	C/T	f(C) = 0.40					
16	rs1482958*			32950652	intergenic	C/T	f(T) = 0.12					
17	rs419145*			32951743	intergenic	A/G	f(A) = 0.19					
18	rs3480	FNDC5	1p35.1	33100752	5'UTR	A/G	f(G) = 0.28	ESE				
	rs7264 (PCR failure)	PHC2	1p34.3	33561969	3'UTR	C/T	f(T) = 0.34	Mm cons.				
19	rs5861	PHC2	1p34.3	33562107	3'UTR	A/C	f(C) = 0.29	ESE / Mm cons.				
	rs3917980 (PCR failure)	CSF3R	1p35-34.3	36709652	Exon 10 (synonymous)	C/T	f(C) = 0.16	ESE / Mm cons.				
			D1S255									
20	rs11540746	DNALI1	1p35.1	37795903	Exon 2 (missense)	C/T	f(T) = 0.35	ESE			0.688	Benign
21	rs6619	DNALI1	1p35.1	37803283	3'UTR	C/T	f(T) = 0.47	ESE / Mm cons.				
	rs3134614 (all homozygous)	MYCL1	1p34.2	40135641	Exon 2 (missense)	G/G	all homozygous	ESE / Mm cons.			0.216	Benign
			D1S2743									
			D1S3721									
22	rs10789424	YBX1	1p34	42931077	Intron 2	A/G	f(A) = 0.38					
23	rs12030724	YBX1	1p34	42932605	Intron 3	A/T	f(T) = 0.30	Mm cons.				
24	rs710251	CDC20	1p34.2-33	43598899	Intron 7	A/C	f(A) = 0.46	Triplex/Mm cons.				
25	rs3219489	MUTYH	1p34.3-32.1	45570092	Exon 12 (missense)	C/G	f(C) = 0.46	ESE	0.56	Tolerated	1.592	Probably damaging
26	rs17522918	PRDX1	1p34.1	45760161	5'UTR	G/T	f(T) = 0.17					
27	rs9147	AKR1A1	1p33-32	45789274	5'UTR	C/T	f(T) = 0.35	Mm cons.				
28	rs324420	FAAH	1p35-34	46643348	Exon 3 (missense)	A/C	f(A) = 0.13	ESE / Mm cons.	0.49	Tolerated	1.319	Benign
29	rs4646487	CYP4B1	1p32	47051762	Exon 5 (missense)	C/T	f(C) = 0.50	ESE / Mm cons.	0.03	Damaging	1.423	Benign
30	rs2297809	CYP4B1	1p32	47055359	Exon 9 (missense)	A/G	f(A) = 0.29	ESE / Mm cons.	0	Damaging	3.031	Probably damaging
31	rs9333049	CYP4A11	1p33	47167783	3'UTR	A/C	f(A) = 0.11[‡]					
			D1S197									
32	rs12855	CDKN2C	1p32	51212681	3'UTR	C/T	f(T) = 0.11	ESE / Mm cons.				
33	rs6657017	SCP2	1p32	53232049	3'UTR	A/G	f(A) = 0.36					
			D1S2742									

*: A second series of 11 tagSNPs; †: MAF information from CEPH; ‡: MAF information from PGA-EUROPEAN-PANEL

§: ESE, exonic splicing enhancer; Mm con., Mus Musculus conserves regions.

Table 5. Results of single locus analysis from FBAT and PDT

SNP #	Gene	Allele	MAF	FBAT						PDT									
										TDT				Sibships					
										T†		NT†		Affected		Unaffected		Nominal P value	Empirical P value‡
				No. of Informative families	S	E(S)	Var(S)	Z	Nominal P value	No.	(%)	No.	(%)	No.	(%)	No.	(%)		
SNP1	PTP4A2	A	0.37	130	131.0	124.9	40.8	1.0	0.3384	54	0.42	45	0.3462	340	0.385	341	0.387	0.7221	0.6968
SNP2	KHDRBS1	G	0.37	143	117.0	122.3	43.8	-0.8	0.4258	63	0.43	60	0.411	310	0.328	339	0.358	0.2770	0.2613
SNP3	KHDRBS1	A	0.27	129	98.0	99.7	36.4	-0.3	0.7798	45	0.32	42	0.2958	238	0.256	257	0.277	0.4829	0.4633
SNP4	EIF3S2	G	0.26	125	96.0	97.1	36.8	-0.2	0.8539	44	0.3	44	0.2973	239	0.258	252	0.272	0.5762	0.6216
SNP5	EIF3S2	A	0.31	130	115.0	111.6	39.8	0.5	0.5881	44	0.31	42	0.2958	315	0.342	286	0.31	0.1874	0.2066
SNP6*	ZBTB8	C	0.10	67	38.0	37.0	17.3	0.2	0.8120	16	0.11	13	0.0915	82	0.09	93	0.102	0.6506	0.6696
SNP7*	ZBTB8	A	0.48	136	145.0	143.9	43.1	0.2	0.8610	65	0.48	71	0.5221	477	0.551	464	0.536	0.7723	0.7419
SNP8*	ZBTB8	A	0.18	101	70.0	65.7	27.6	0.8	0.4089	29	0.22	21	0.1615	135	0.15	140	0.156	0.8613	0.9033
SNP9*	ZBTB8OS	T	0.15	80	54.0	59.1	24.0	-1.0	0.2991	23	0.16	26	0.1757	111	0.121	121	0.132	0.3797	0.4199
SNP10*	RBBP4	A	0.13	81	47.0	51.9	21.8	-1.0	0.2953	24	0.16	23	0.1533	96	0.102	117	0.125	0.1635	0.1849
SNP11*	RBBP4	A	0.27	127	86.0	97.9	37.3	-1.9	0.0521	43	0.31	47	0.3357	192	0.21	234	0.256	**0.0215**	**0.0185**
SNP12*	RBBP4	A	0.19	106	66.0	67.7	28.7	-0.3	0.7468	33	0.24	28	0.2029	128	0.148	146	0.169	0.4347	0.3972
SNP13	RBBP4	C	0.27	119	82.0	94.7	36.0	-2.1	**0.0343**	41	0.31	43	0.3209	188	0.209	244	0.271	**0.0047**	**0.0025**
SNP14*	SYNC1	A	0.09	53	25.0	32.7	14.3	-2.0	**0.0418**	10	0.08	13	0.0985	49	0.055	74	0.082	**0.0390**	**0.0332**
SNP15*	SYNC1	C	0.39	150	136.0	144.5	49.0	-1.2	0.2251	62	0.42	62	0.4247	326	0.352	367	0.396	0.0950	0.1046
SNP16*		T	0.15	87	48.0	57.0	23.0	-1.9	0.0617	23	0.16	27	0.1824	103	0.111	133	0.143	**0.0322**	**0.0283**
SNP17*		G	0.21	103	63.0	72.5	28.4	-1.8	0.0739	32	0.22	37	0.2569	157	0.169	188	0.202	**0.0495**	**0.0414**
SNP18	FNDC5	G	0.25	117	90.0	90.1	35.7	0.0	0.9816	48	0.32	33	0.2171	182	0.194	203	0.216	0.7820	0.7552
SNP19	PHC2	C	0.22	111	71.0	82.5	34.2	-2.0	**0.0486**	28	0.19	37	0.2534	203	0.211	212	0.221	0.3982	0.4317
SNP20	DNALI1	T	0.41	124	121.0	122.3	44.0	-0.2	0.8450	56	0.41	54	0.3913	341	0.386	347	0.393	0.8696	0.9032
SNP21	DNALI1	T	0.46	136	144.0	150.0	46.9	-0.9	0.3828	66	0.45	69	0.4726	411	0.442	433	0.466	0.3279	0.3465
SNP22	YBX1	A	0.40	138	127.0	131.9	46.2	-0.7	0.4677	51	0.35	61	0.4178	367	0.39	371	0.394	0.5670	0.6032
SNP23	YBX1	T	0.31	120	102.0	100.6	38.7	0.2	0.8246	50	0.36	40	0.2899	279	0.306	289	0.317	1.0000	1.0000
SNP24	CDC20	A	0.45	142	156.0	155.2	48.6	0.1	0.9033	76	0.49	77	0.4936	412	0.427	395	0.41	0.5057	0.4748
SNP25	MUTYH	C	0.40	126	122.0	116.2	42.3	0.9	0.3738	57	0.45	46	0.3594	317	0.398	300	0.377	0.2450	0.2813
SNP26	PRDX1	T	0.15	82	62.0	56.2	27.0	1.1	0.2666	23	0.17	18	0.1304	148	0.157	135	0.143	0.2511	0.2868
SNP27	AKR1A1	T	0.37	142	140.0	124.4	47.2	2.3	**0.0233**	64	0.43	47	0.3176	328	0.377	293	0.337	**0.0380**	**0.0379**
SNP28	FAAH	A	0.14	64	49.0	46.1	17.7	0.7	0.4939	7	0.06	13	0.1182	144	0.18	123	0.154	0.3604	0.3410
SNP29	CYP4B1	T	0.20	85	61.0	56.0	22.2	1.1	0.2879	32	0.23	24	0.1714	157	0.179	148	0.169	0.3156	0.2919
SNP30	CYP4B1	A	0.23	95	63.0	65.4	28.5	-0.4	0.6575	34	0.28	26	0.2131	153	0.18	172	0.202	0.5514	0.5839
SNP31	CYP4A11	A	0.10	66	25.0	35.5	16.6	-2.6	**0.0097**	8	0.06	19	0.1418	62	0.067	69	0.075	0.2120	0.1909
SNP32	CDKN2C	T	0.11	61	36.0	39.7	16.6	-0.9	0.3688	12	0.08	16	0.1081	94	0.101	94	0.101	0.8111	0.8636
SNP33	SCP2	A	0.36	142	130.0	129.0	45.4	0.2	0.8764	46	0.36	41	0.3203	338	0.376	330	0.367	0.5811	0.6159

*: A second series of 11 tagSNPs

†: T represents "transmitted"; NT represents "not transmitted"

‡: Empirical P values derived from 10000 simulating replicates

Table 6. Single-locus analysis in 855 unrelated cases and 875 unrelated controls

SNP #	Gene	Genotype-based							Allele-based		
		No. of Cases	No. of Controls	OR	95% CI	$P_{additive}$	$P_{dominant}$	$P_{recessive}$	MAF Cases	MAF Controls	P value
SNP10	RBBP4					0.3867	0.5354	0.4382	0.12	0.13	0.3762
GG		644	642		1.00						
AG		173	186	0.93	(0.73-1.17)						
AA		16	20	0.78	(0.41-1.55)						
SNP11	RBBP4					0.6366	0.8317	0.4742	0.29	0.28	0.6393
CC		397	422		1.00						
AC		344	336	1.09	(0.89-1.33)						
AA		63	67	1.00	(0.69-1.45)						
SNP12	RBBP4					0.9683	0.2949	0.6587	0.19	0.19	0.9685
GG		526	543		1.00						
AG		257	245	1.08	(0.88-1.34)						
AA		23	31	0.77	(0.44-1.33)						
SNP13	RBBP4					**0.0076**	0.5405	**0.0022**	0.32	0.28	**0.0089**
TT		378	453		1.00						
CT		399	353	**1.36**	**(1.11-1.65)**						
CC		73	68	1.29	(0.90-1.84)						
SNP14	SYNC1					0.0793	0.5922	**0.0492**	0.11	0.10	0.0856
GG		638	687		1.00						
AG		178	148	**1.30**	**(1.02-1.65)**						
AA		5	7	0.77	(0.24-2.44)						
SNP15	SYNC1					**0.0284**	0.2297	**0.0264**	0.41	0.38	**0.0310**
TT		280	333		1.00						
CT		431	417	1.23	(0.99-1.78)						
CC		135	121	1.33	(1.00-1.78)						

Table 7. Results from joint linkage and association analysis

Gene and SNP #	SNP type	Minor allele frequency (MAF)		P value for						
				Pseudomarker				LAMP		
		Affecteds	Unaffecteds	Linkage	LD under Linkage	LD under No linkage	LD and Linkage	Linkage	Association	Other linked variants
RBBP4										
SNP10†	Intron 2	0.12	0.14	0.4874	1.0000	1.0000	1.0000	0.97	0.2800	0.93
SNP11†	Intron 6	0.28	0.29	0.3368	0.9503	0.9132	0.7919	0.88	0.2800	0.59
SNP12†	Intron 7	0.19	0.19	0.4708	0.9816	0.9787	0.9679	0.99	0.1010	0.69
SNP13†	3' UTR	0.31	0.28	0.5000	**0.0021**	**0.0023**	**0.0055**	1.00	**0.0086**	0.44
SYNC1										
SNP14†	Intron 3	0.09	0.09	0.5000	0.8665	0.8665	0.9262	-	**0.0340**	-
SNP15†	Intron 1	0.38	0.38	0.3203	**0.0158**	**0.0144**	**0.0314**	0.79	0.0570	0.77

*: The MAF of the affecteds is based on all affecteds from singleton families, one randomly selected affecteds from multiplex families(n=235);
the MAF of the unaffecteds is based on all unaffecteds from singleton families, one randomly selected unaffecteds from multiplex families(n=226).

†: The MAF of the affecteds is based on all affecteds from singleton families, one randomly selected affecteds from multiplex families(n=235) and all cases (n=855);
the MAF of the unaffecteds is based on all unaffecteds from singleton families, one randomly selected unaffecteds from multiplex families(n=226) and all controls (n=875)

Table 8. Haplotype analysis in the critical region extending from 32.783 Mb to 32.831 Mb

	SNP10	SNP11	SNP12	SNP13	SNP14	SNP15	0 copy		1 copy		2 copies	
							case/ control	OR	case/ control	OR (95% CI)	case/ control	OR (95% CI)
RBBP4 gene												
Hap 1A	G	A	G	T			64/70	1.0	388/339	1.25 (0.87 - 1.81)	366/433	0.93 (0.64 - 1.33)
Hap 1B	G	C	G	C			656/700	1.0	157/135	1.24 (0.96 - 1.60)	5/7	0.76 (0.24 - 2.41)
Hap 1C	G	C	A	C			719/753	1.0	98/82	1.25 (0.92 - 1.71)	1/7	0.15 (0.02 - 1.22)
Hap 1D	A	C	A	C			634/651	1.0	169/178	0.98 (0.77 - 1.24)	15/13	1.19 (0.56 - 2.51)
SYNC1 gene												
Hap 2A					G	C	405/447	1.0	349/337	1.14 (0.94 - 1.40)	67/69	1.07 (0.75 - 1.54)
Hap 2B					G	T	125/116	1.0	418/405	0.96 (0.72 - 1.28)	278/332	0.78 (0.58 - 1.05)
Hap 2C					A	C	643/699	1.0	173/147	**1.28 (1.00 - 1.63)**	5/7	0.78 (0.25 - 2.46)
RBBP4 + SYNC1 genes												
Hap 3A	G	A	G	T	G	C	645/685	1.0	144/143	1.07 (0.83 - 1.38)	12/8	1.59 (0.65 - 3.92)
Hap 3B	G	A	G	T	G	T	128/121	1.0	423/402	1.00 (0.75 - 1.32)	250/313	0.76 (0.56 - 1.02)
Hap 3C	G	C	G	C	A	C	631/689	1.0	165/140	**1.29 (1.00 - 1.65)**	5/7	0.78 (0.25 - 2.47)
Hap 3D	G	C	A	C	G	C	700/746	1.0	100/83	1.28 (0.94 - 1.75)	1/7	0.15 (0.02 - 1.24)
Hap 3E	A	C	A	C	G	C	619/644	1.0	168/179	0.98 (0.77 - 1.24)	14/13	1.12 (0.52 - 2.40)

Table 9. Clinical features of case patients with HCC in case-control study

Variables	Cases (n=855)	
	No.	(%)
Vital status		
Alive	197	23.04
Deceased		
HCC	572	66.90
Liver cirrhosis	54	6.32
Other causes	32	3.74
Tumor size (cm)		
≦ 2	181	23.35
3 - 5	244	31.48
＞ 5	350	45.16
Missing	80	
No. of lesions		
1	458	55.11
2 - 3	88	10.59
≧ 4	285	34.30
Missing	24	
AFP levels (ng/mL)		
≦ 20	104	12.29
21 - 200	175	20.69
201 - 400	69	8.16
401 - 1000	109	12.88
＞ 1000	389	45.98
Missing	9	

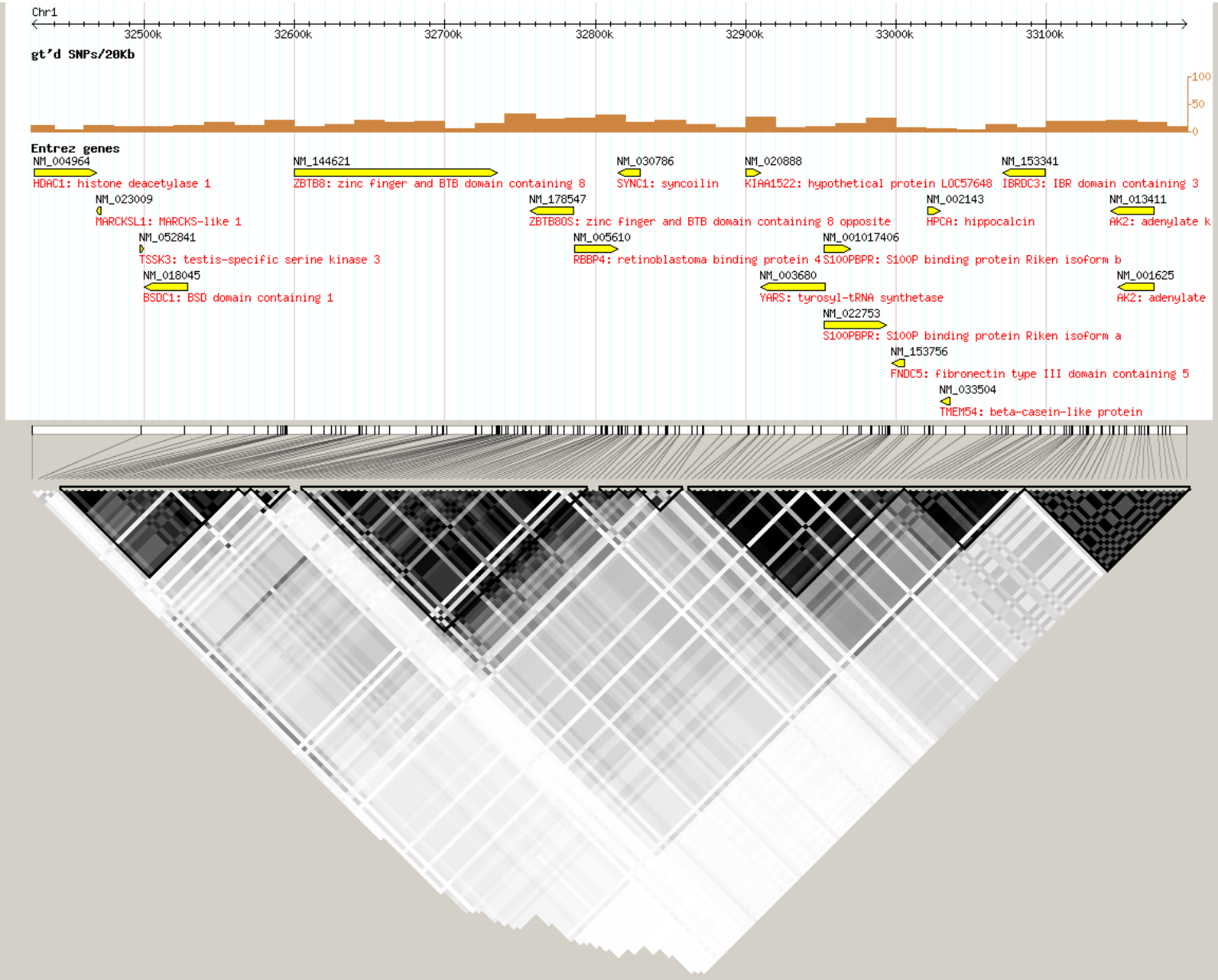
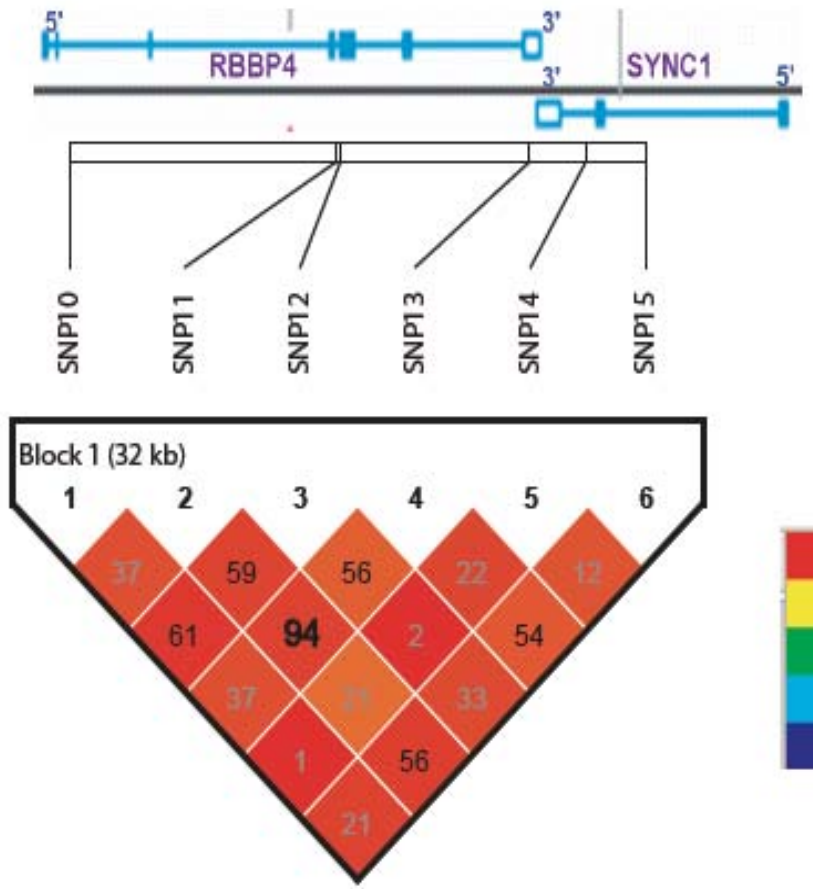


Figure 1. LD structure of all HapMap SNPs derived from the Asian samples. Pairwise r^2 values are color-coded: black, high r^2 values; white, low r^2 values.

A



B

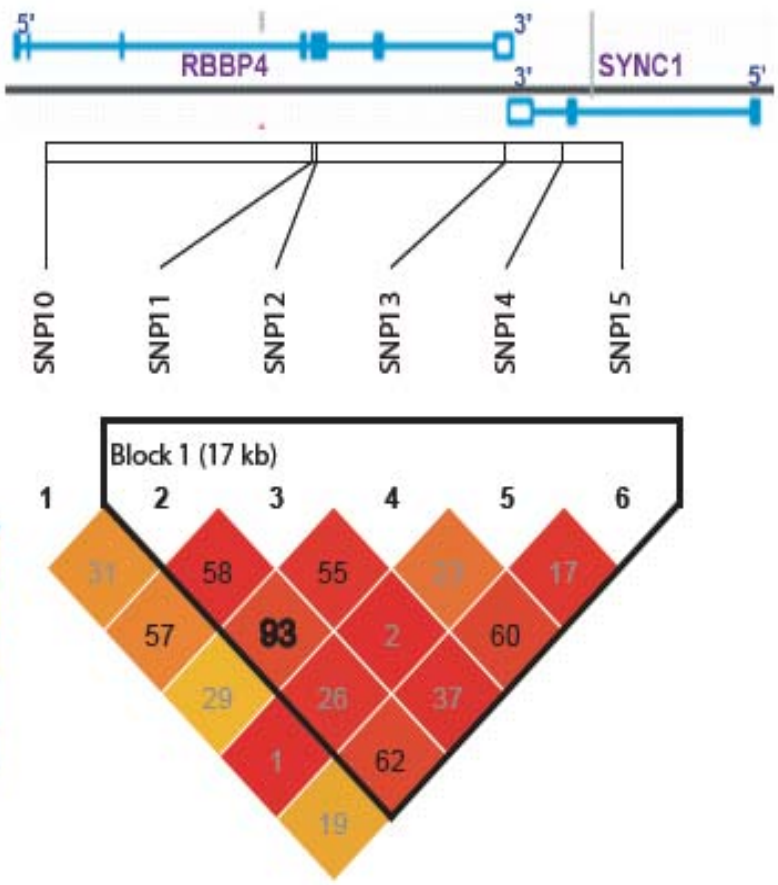


Figure 2. D' and r^2 between 6 SNPs genotyped (A) in the family sample and (B) in 875 unrelated controls subjects

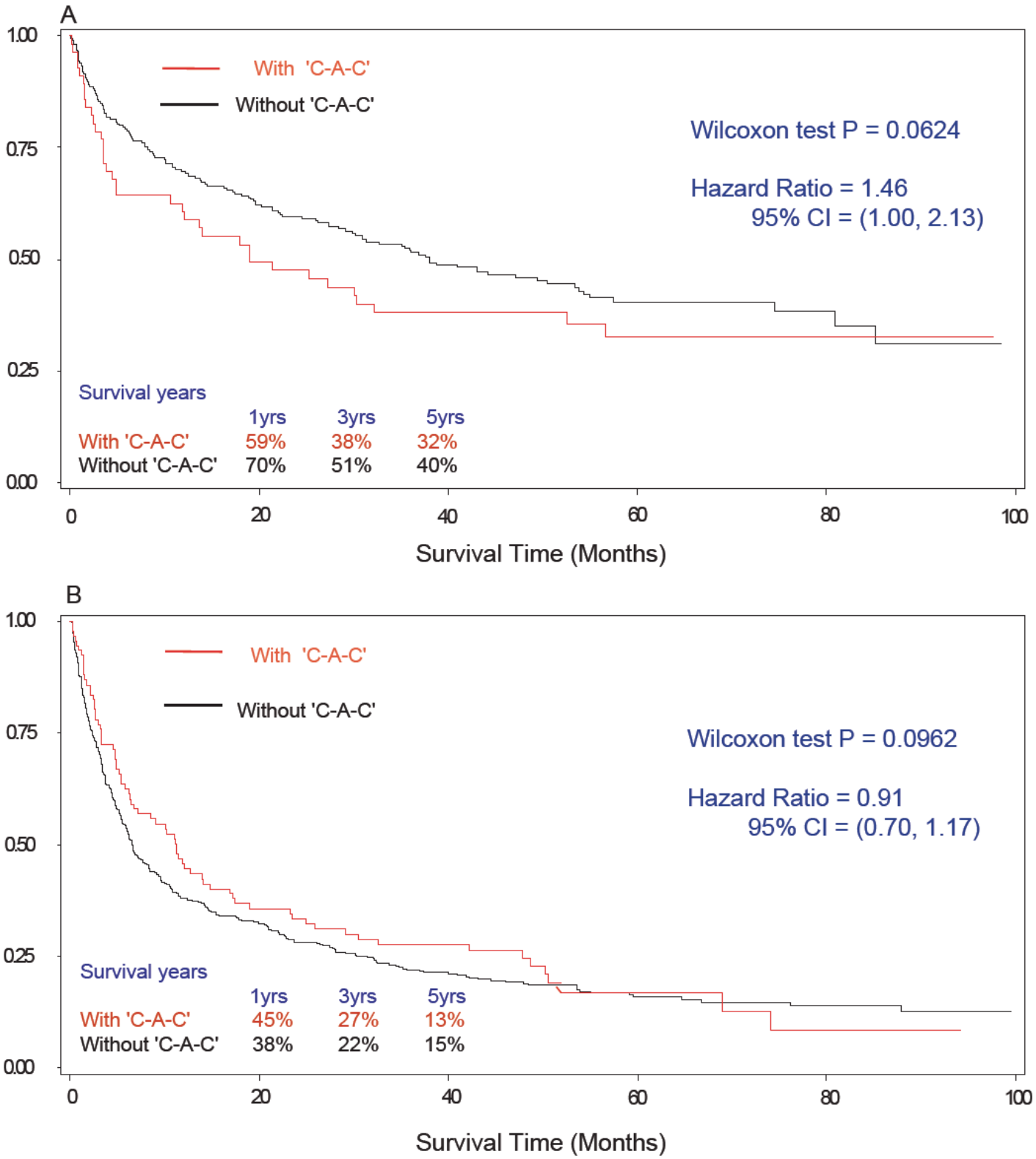


Figure 3. Cumulative survival according to tumor size and the presence or absence of haplotype 'C-A-C' at SNP 13-15 (A) tumor size≦3cm (n=281) (B) tumor size＞3cm (n=449)