Published in final edited form as:

Dig Dis Sci. 2014 May ; 59(5): 986–991. doi:10.1007/s10620-013-2984-3.

Molecular evolution of genetic susceptibility to Hepatocellular Carcinoma

Saowanee Ngamruengphong and Tushar Patel

Departments of Transplantation and Cancer Biology Mayo Clinic, Jacksonville, Florida

Abstract

Background—Hepatocellular carcinoma (HCC) is a leading cancer-related cause of death worldwide. There are widespread global differences in risk of HCC. Although the impact of geographic prevalence of specific causes of chronic liver disease on HCC is recognized, the contribution of the underlying genetic architecture to the risk of HCC remains undefined. Our aims were to characterize evolutionary trends in genetic susceptibility to HCC.

Methods—We examined genetic risk associated with HCC risk alleles identified from genome wide association studies and correlated these with geographic location and with temporal and spatial patterns of human migration.

Results—A moderate increase in differentiation was noted for rs2596542 ($F_{st}=0.106$) and rs17401966 ($F_{st}=0.116$), SNPs associated with an increased risk of HCC in patients with chronic HBV and HCV respectively. Both of these SNPs show a recent increase in allelic frequency with the most recent human migrations into East Asia, Oceania and the Americas. In contrast another SNP associated with an increased risk of HCC, rs9275572, showed a lack of differentiation ($F_{st}=0.09$) with stable allelic expression across populations. The genetic risk score for HCC, based on the allelic frequency and risk odds ratio of five SNPs associated with increased risk of HCC, was greatest in populations from Africa and decreased with subsequent migration into Europe and Asia. However, a major increase was noted with most recent migrations into Oceania and the Americas.

Conclusions—There are differences in directional differentiation of HCC risk alleles across human populations that can contribute to population-based differences in HCC prevalence.

Keywords

Single nucleotide polymorphisms; risk alleles; liver cancer; migration; population genetics

Introduction

Hepatocellular carcinoma (HCC) is the fifth most common cancer and third leading cancer-related cause of death worldwide. The incidence of HCC is increasing in many parts of the world such as North America, Europe and Japan [1]. There are very marked differences in

the incidence of HCC in different geographic regions. The incidence rate is highest in sub-Saharan Africa and in Eastern Asia; whereas North America, South America, Northern Europe, and Oceania have a low incidence of HCC. This geographic variability is likely multifactorial. Risk factors for HCC include cirrhosis, chronic viral hepatitis, alcohol and fatty liver disease [2,3]. Only a fraction of individuals with chronic viral hepatitis or fatty liver disease develop HCC. It is likely that HCC arises as a consequence of complex interactions between genetic factors and tissue responses to environmental exposures or hepatotropic virus infections [3]. Some genetic conditions such as hereditary hemochromatosis are associated with a higher risk of HCC [4]. Although the impact of geographic variations in the prevalence of specific diseases that predispose to HCC is evident, the contribution of the underlying genetic architecture to the risk of HCC remains undefined. In addition to affecting predisposition to chronic liver disease or its progression and response to therapy, host genetic factors could modulate susceptibility to HCC formation by affecting key mediators of tumorigenesis independent of etiology.

The impact of genetic variations on risk of HCC remains largely undefined although hepatocarcinogenesis arises from complex interactions between genetic and environmental factors. Recent genome wide association studies (GWAS) have identified genetic polymorphisms that are associated with increased susceptibility to HCC in defined populations with chronic hepatitis B [5] or chronic hepatitis C [6]. Interaction between allele frequencies and environmental variables may contribute to geographic differences in disease prevalence. Thus, genomic variations across populations may contribute to geographic variations in disease prevalence.

A knowledge of the impact of genetic variations on human disease risk is critical for developing more effective strategies for disease screening, prevention and risk mitigation. There is evidence that negative selection has globally reduced population differentiation at the level of disease-related genes from the International HapMap Project of Human Genome [7]. The impact of these variations on risk of diseases such as HCC remains unknown. Thus, our goals were to characterize selected risk factors for HCC, to determine if susceptibility for HCC varies across geographic populations. We hypothesized that negative selection would reduce susceptibility to single nucleotide polymorphisms (SNPs) associated with an increased risk of HCC. By correlating temporal and spatial changes in expression of these genetic polymorphisms with trends in human migration and with data regarding frequency of risk alleles in these populations, we derived information regarding the potential changes with human migration and selection effects for these specific risk alleles for HCC.

Methods

Identification of HCC risk alleles

SNPs associated with HCC susceptibility were identified from an electronic search of the Pubmed database in November 2011, using the search terms: “hepatocellular carcinoma”, “allele or polymorphism” and “risk or association”. We selected studies that (i) provided genotype data mutation with risk of HCC; (ii) had a case-control or cohort design; and (iii) compared patients with HCC with control subjects who were free of cancer. Control subjects could be healthy individuals, or patients with chronic liver disease. We extracted data on

risk ratio for each SNPs from these published studies. Data on the SNPs and their selection are provided in the respective publications and available in the public domain. When multiple studies reported data on single SNPs, we used pooled odds ratio (OR) from published meta-analysis of these studies. SNPs with an increased risk of HCC associated with defined causes of liver disease were also identified using an online database of SNP-trait associations that are extracted from published GWAS (www.genome.gov/GWASStudies).

Characterization of evolutionary trends

Genetic variations in HCC risk may contribute to differences in geographic variations in prevalence of HCC. Moreover, gene variants associated with an increased risk for HCC may confer unrecognized benefits to their human carriers. In order to examine these population genetic factors, temporal and spatial trends in expression of genetic polymorphisms associated with HCC were examined by correlating individual alleles with patterns of human migration.

Determination of allelic frequency and population differentiation

The frequency of the risk alleles across 53 diverse anthropologically defined human population samples was obtained from the Human Genome Diversity Project database [8,9,10] and using the Allele Frequency Database (ALFRED) [11,12,13]. These populations are derived from all continents, and represent anthropologically distinct populations that were in place prior to modern migrations after the great diasporas of the fifteenth and sixteenth centuries [8]. The relationship between their genetic profiles have been used to understand patterns of human migration [14]. The fixation index (F_{st}) was used to describe the degree of population differentiation for each SNP. $F_{st} = (HT - HS)/HT$, in which HT and HS represent heterozygosity of the total population and of the subpopulation, respectively. F_{st} is used as a measure of gene flow and similarity across subpopulations [7]. In the present study, the F_{st} value across all SNP classes were interpreted as low (below 0.1), moderate (between 0.10-0.15) or high (above 0.15) differentiation.

Genetic risk score

A genetic risk score was determined from the risk odds ratio and the frequency of the risk alleles across 53 human populations using the Human Diversity Genome Project and ALFRED databases [11,12]. The genetic risk associated with each risk allele was then correlated with geographic location and with temporal and spatial patterns of human migration.

Ethics

The analysis was carried out using aggregated data reported in the public domain without any identifying information and did not require any institutional review board approval.

Results

Geographic variation in the disease prevalence can partly be due to differences in risk allele frequencies among different populations. Natural selection results in large allele frequency

differences between populations [15,16]. Previous studies have shown that individual disease-associated alleles that have risen to a high frequency due to selection may be responsible for differences in disease prevalence between populations [17]. We identified 5 gene associated SNPs, (rs1800562 of HFE [18], rs2596542 of MICA [6], rs2267716 of CRHR2 [19], rs9275572 of HLA-DQ/DR [6] and rs17401966 of KIF1B [5] that were associated with an increased risk of HCC (Table 1). Examination of the allele frequency of these HCC-associated SNPs in 53 distinct human populations revealed that these populations substantially differed in their SNP expression and susceptibility to HCC. SNPs with an increased risk of HCC associated with HBV (rs17401966) or HCV (rs9275572, rs2596542) from GWAS data were further analyzed for genetic risk in different geographic regions (Fig 1). Individually, two of the HCC-associated SNPs (rs2596542 and rs17401966) showed heterogeneity in allele frequencies across human populations. The allele frequency of these two alleles between select populations ranged from 0 to 0.95. Thus, these HCC associated risk alleles are present at considerably different frequencies between geographically distinct populations. We applied *Fst* statistics to determine the degree of population differentiation for each HCC associated SNPs. A moderate increase in differentiation was noted for rs2596542 (*Fst*=0.106) and rs17401966 (*Fst*=0.116). Both of these SNPs show a recent increase in allelic frequency with the most recent human migrations into East Asia, Oceania and the Americas (Fig 1). In contrast, rs9275572 showed a lack of differentiation (*Fst*=0.09) with stable allelic expression across populations although allele frequencies showed some variation across populations, ranging from 0 to 0.58.

The *Fst* and risk ratios for each HCC-associated SNPs are shown in Table 1. For the association between rs1800562 and HCC, we used a pooled OR of 5.20 (95%CI: 2.69-10.08) from the meta-analysis of 9 studies reported by Jin et.al[18]. The risk alleles of other SNPs individually had a small effect on HCC risk, with the odds ratios less than 2. However, when multiple SNPs are combined, the risk ratio may be sufficiently large to be useful in risk prediction [20], as has been shown in breast cancer and prostate cancer [21,22]. Thus, we calculated a genetic risk score for HCC from the combined risk of five SNPs associated with increased risk of HCC [rs1800562, rs2596542, rs2267716, rs9275572 and rs17401966]. This HCC genetic risk score was then analyzed in 53 human populations and correlated with proposed patterns of pre-modern human migration. Historically, migrations are believed to have occurred out of Africa and spread to Europe and Asia, and then to the Americas and the Pacific Islands. The genetic risk score for HCC was greatest in populations from Africa namely the Mbuti pygmy and San populations. However this decreased with subsequent migration into Europe and Asia and particularly for the Lahu and Yi populations, although a major increase was noted with most recent migrations into Oceania and the Americas, such as for the Karitiana populations (Fig 2). We next calculated a genetic risk score for HCC based on the combined risk of two SNPs; rs2596542 and rs17401966 associated with a moderately increased risk of HCC in patients with chronic HCV and HBV respectively. The combined risk associated with these SNPs was high in populations from Africa and slightly decreased with subsequent migration into Europe and Central Asia. The risk remains high in several populations of Eastern Asia and a major increase was noted with the most recent migrations into Oceania and the Americas (Fig 3).

Discussion

These data indicate differences in directional differentiation of HCC risk alleles across human populations, and provide insights into the contribution of human migration and hepatocarcinogenesis. The observed differences in directional differentiation of HCC risk alleles across human populations suggest that the genetic susceptibility to HCC could have been modified by factors other than by genetic drift alone [7]. Thus, HCC associated traits could have undergone evolution by selection pressure in turn contributing to differences in the prevalence of HCC among human populations. Future studies are needed to clarify which risk allele and the extent to which the differences in risk allele frequencies between populations can contribute to differences in disease prevalence between these populations.

There are several limitations of our study. First, in a majority of the candidate-gene case-control studies and GWAS, the association of genetic traits and the risk of HCC have been evaluated in European or Asian populations. It is unknown whether the strength of association of genetic polymorphism and risk of HCC are similar in other populations. The impact of specific genetic traits in susceptibility for HCC may vary from one population to another. Therefore, studies conducted in more diverse populations are needed to determine the ability of risk alleles in disease prediction in different populations. Numerous studies have evaluated the association between SNPs and the presence of HCC [23]. In order to characterize the patterns of geographic difference of risk allele, we selected SNPs for which there was data available with respect to their allele frequencies in diverse human populations. This selection may diminish the value of an individually applied genetic risk score because of the possibility for exclusion of genetic polymorphisms that may have been more informative. In addition to changes in SNPs, genetic variation across species can arise from structural variations such as insertions and deletions, and variations in copy number. The widespread occurrence of copy number variants within the genome highlights the potential importance of these variants in genetic diversity, and potentially in disease risk which warrant further evaluation. The risk score concept is based on a limited number of risk alleles across different etiologies of liver cancer to illustrate the concepts. Although the combined 5 SNP genetic risk score was relatively lower in east Asia, there is a high incidence of HCC in that region. This inconsistency reflects the choice of SNPs used. Indeed, there was an increase in allelic frequency of two of the HCC-associated SNPs (rs2596542 and rs17401966), and a risk score based on these reflected an increase risk in East Asia. Therefore, combination of multiple genetic risk variants may not be superior to individual or selected variants, particularly if the latter predispose to geographically prevalent environmental factors associated with disease. The incidence of HCC will reflect the prevalence of specific genetic risk factors as well as environmental risk factors which may vary geographically, such as chronic hepatitis B which is highly prevalent in east Asia. As additional population-based risk alleles become characterized, future studies could incorporate these into more geographically precise disease-specific risk scores. Despite these limitations, the present study provides a first step towards understanding the evolutionary genetics of this disease and novel insights to understanding the global genetic epidemiology of HCC.

In conclusion, population based studies of human genomic variations provide insights into the role of migration and genetic factors in hepatocarcinogenesis. There are large differences in polymorphism that influence risk of HCC among the worldwide populations. Variations in allelic frequency of HCC associated SNPs can contribute to population-based differences in HCC prevalence, and need to be considered in developing regional strategies for screening and surveillance for HCC. It is expected that future risk predictive models incorporating genetic risk and other clinical risk factors will provide sufficient discrimination to stratify individuals according to biologically relevant risk groups. This may in turn have important implications for genetic based screening and preventive strategies for HCC.

Abbreviations

GWAS	genome wide association studies
HCC	hepatocellular carcinoma
SNP	single nucleotide polymorphism

References

1. Bosch FX, Ribes J, Diaz M, Cleries R. Primary liver cancer: worldwide incidence and trends. *Gastroenterology*. 2004; 127:S5–S16. [PubMed: 15508102]
2. Welzel TM, Graubard BI, Zeuzem S, El-Serag HB, Davila JA, et al. Metabolic syndrome increases the risk of primary liver cancer in the United States: a study in the SEER-Medicare database. *Hepatology*. 2011; 54:463–471. [PubMed: 21538440]
3. El-Serag HB. Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology*. 2012; 142:1264–1273. e1261. [PubMed: 22537432]
4. Villanueva A, Newell P, Hoshida Y. Inherited hepatocellular carcinoma. *Best Pract Res Clin Gastroenterol*. 2010; 24:725–734. [PubMed: 20955973]
5. Zhang H, Zhai Y, Hu Z, Wu C, Qian J, et al. Genome-wide association study identifies 1p36.22 as a new susceptibility locus for hepatocellular carcinoma in chronic hepatitis B virus carriers. *Nat Genet*. 2010; 42:755–758. [PubMed: 20676096]
6. Kumar V, Kato N, Urabe Y, Takahashi A, Muroyama R, et al. Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. *Nat Genet*. 2011; 43:455–458. [PubMed: 21499248]
7. Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L. Natural selection has driven population differentiation in modern humans. *Nat Genet*. 2008; 40:340–345. [PubMed: 18246066]
8. Cavalli-Sforza LL. The Human Genome Diversity Project: past, present and future. *Nat Rev Genet*. 2005; 6:333–340. [PubMed: 15803201]
9. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, et al. Worldwide human relationships inferred from genome-wide patterns of variation. *Science*. 2008; 319:1100–1104. [PubMed: 18292342]
10. [1 August 2012] The Human Genome Diversity Project database. Available: ftp://ftp.cephb.fr/hgdp_supp1.
11. Rajeevan H, Soundararajan U, Kidd JR, Pakstis AJ, Kidd KK. ALFRED: an allele frequency resource for research and teaching. *Nucleic Acids Res*. 2012; 40:D1010–1015. [PubMed: 22039151]
12. Rajeevan H, Cheung KH, Gadagkar R, Stein S, Soundararajan U, et al. ALFRED: an allele frequency database for microevolutionary studies. *Evol Bioinform Online*. 2005; 1:1–10. [PubMed: 19325849]
13. [1 August 2012] Allele Frequency Database. Available: <http://alfred.med.yale.edu/>.

14. Zhivotovsky LA, Rosenberg NA, Feldman MW. Features of evolution and expansion of modern humans, inferred from genomewide microsatellite markers. *Am J Hum Genet.* 2003; 72:1171–1186. [PubMed: 12690579]
15. Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, et al. Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet.* 2004; 74:1111–1120. [PubMed: 15114531]
16. Myles S, Somel M, Tang K, Kelso J, Stoneking M. Identifying genes underlying skin pigmentation differences among human populations. *Hum Genet.* 2007; 120:613–621. [PubMed: 16977434]
17. Myles S, Tang K, Somel M, Green RE, Kelso J, et al. Identification and analysis of genomic regions with large between-population differentiation in humans. *Ann Hum Genet.* 2008; 72:99–110. [PubMed: 18184145]
18. Jin F, Qu LS, Shen XZ. Association between C282Y and H63D mutations of the HFE gene with hepatocellular carcinoma in European populations: a meta-analysis. *J Exp Clin Cancer Res.* 2010; 29:18. [PubMed: 20196837]
19. Gu X, Qi P, Zhou F, Ji Q, Wang H, et al. An intronic polymorphism in the corticotropin-releasing hormone receptor 2 gene increases susceptibility to HBV-related hepatocellular carcinoma in Chinese population. *Hum Genet.* 2010; 127:75–81. [PubMed: 19813023]
20. Easton DF, Eeles RA. Genome-wide association studies in cancer. *Hum Mol Genet.* 2008; 17:R109–115. [PubMed: 18852198]
21. Wacholder S, Hartge P, Prentice R, Garcia-Closas M, Feigelson HS, et al. Performance of common genetic variants in breast-cancer risk models. *N Engl J Med.* 2010; 362:986–993. [PubMed: 20237344]
22. Zheng SL, Sun J, Wiklund F, Smith S, Stattin P, et al. Cumulative association of five genetic variants with prostate cancer. *N Engl J Med.* 2008; 358:910–919. [PubMed: 18199855]
23. Nahon P, Zucman-Rossi J. Single nucleotide polymorphisms and risk of hepatocellular carcinoma in cirrhosis. *J Hepatol.* 2012; 57:663–674. [PubMed: 22609306]

Figure 1. Geographic variations of risk allele frequencies for SNPs associated with hepatocellular carcinoma

Variations of risk allele frequencies for rs17401966, rs9275572, and rs2596542 across each reference human population is shown. The Y axis displays risk allele frequencies in each of the human populations with each color representing a different geographic location. Both rs17401966 and rs2596542 show a recent increase in allelic frequencies with the most recent human migrations into East Asia, Oceania and the Americas, whereas, the allelic expression of SNP rs9275572 is stable across populations.

Figure 2. Global genetic risk score and population differentiation for hepatocellular carcinoma

The genetic risk score for HCC was greatest in populations from Africa (Mbuti, San), and decreased with subsequent migration into Europe and Asia (Lahu, Yi). A major increase was noted with most recent migrations into Oceania and the Americas (Karitiana).

Figure 3. Genetic risk prediction based on selected SNPs

rs2596542 and rs17401966 are associated with an increased risk of HCC in patients with chronic HCV and HBV respectively. The genetic risk sfor HCC predicted using these was high in populations from Africa (Mbuti, San), with a slight decrease with subsequent migration into Europe and Central Asia (Adygei, Sindhi). The risk remains high in several populations of Eastern Asia and a major increase was noted with migrations into Oceania and the Americas.

Table 1Risk allele frequencies and *Fst* for HCC-associated SNPs

SNP	Risk alleles	Gene	Diagnosis of HCC	Disease background	OR	Lower	Upper	<i>Fst</i>
rs1800562	A	HFE	H, L	Alcohol, HCV	5.2	2.69	10.08	0.052
rs2596542	A	MICA	H, L	HCV	1.36	1.22	1.51	0.106
rs2267716	C	CRHR2	H	HBV	1.55	1.13	2.15	0.244
rs9275572	A	HLA-DQ/DR	H, L	HCV	1.3	1.19	1.42	0.09
rs17401966	A	KIF1B	H	HBV	1.64	1.26	1.83	0.116

H, diagnosis by histopathological confirmation; L, diagnosis by laboratory/imaging findings; OR, odds ratio;