論文の内容の要旨

論文題目

Genome-wide association studies on hepatocellular carcinoma after eradication of hepatitis C virus in Japanese and on chronic hepatitis B in Thai

(日本人集団における C 型肝炎ウイルス排除後の肝発癌およびタイ人集団における B 型慢性肝炎のゲノムワイド関連解析)

氏名　サイデ　アシュリ

Hepatitis refers to an inflammation of the liver cells and damage to the liver with different types and causes, but similar symptoms. If left untreated, it can develop into chronic (long-term) hepatitis, liver failure, liver cancer and even death. Viral hepatitis is caused by at least 5 different viruses. The most three common viruses are hepatitis A virus (HAV), hepatitis B virus (HBV), and hepatitis C virus (HCV). Based on the World Health Organization, an estimated 1.4 million deaths occur annually from acute infection and hepatitis-related liver cirrhosis and cancer. Approximately, 47% of hepatitis-related deaths are attributable to HBV, 48% to HCV, and the remainder to HAV and hepatitis E virus. Not all of those who have acute infection with HBV and HCV develop hepatitis-related diseases, leading to the speculation that there may be genetic factors involved in the host susceptibility to viral hepatitis-related diseases. The present study focuses on using genome-wide association study (GWAS) in order to search for genetic susceptibility factors affecting host susceptibility to HBV- and HCV-related liver diseases. In the first part of the study, I attempted to identify genetic factors affecting host susceptibility to HCV-associated hepatocellular carcinoma (HCV-HCC) using whole-genome imputation data obtained from Japanese hepatitis C patients. In the second part, I focused on chronic hepatitis B (CHB) in the Thai population.

Hepatitis C starts with an acute phase. Acute hepatitis C (AHC) is usually subclinical and remains undiagnosed as it does not present symptoms. The virus is spontaneously cleared in 15-45% of AHC patients while remaining 55-85% of patients develop chronic hepatitis C (CHC). About 20% of those with CHC develop liver cirrhosis (LC) within 2 or 3 decades and 1-4% of HCV-induced cirrhosis patients have an annual risk of developing HCC. The goal of treatment of hepatitis C is to eradicate the virus (known as sustained virologic response (SVR)), decrease the probability of virus transmission, and reduce the risk of end-stage liver diseases, such as LC and HCC. Interferon (IFN) monotherapy and peg-IFN and ribavirin combination therapy led to the

achievement of SVR in a proportion of patients. Direct-acting antivirals (DAAs) increased the hepatitis C treatment rate (up to 95%). However, despite the availability of highly effective treatments, the incidence of HCV-induced HCC is not disappeared even after SVR.

Numerous studies have been conducted in different populations to try to identify host genetic factors that are involved in host susceptibility to development of HCC induced by HCV, including candidate-gene association studies and more recently, GWASs. GWAS is a hypothesis-free approach used for investigating hundreds of thousands of single nucleotide polymorphisms (SNPs) located all over the genome, in order to detect common genetic variants contributing to the susceptibility to common and complex traits. A number of genes, genic regions and loci have been found to be associated with HCV-HCC. However, the first GWAS on HCV-HCC after achieving SVR by IFN-therapy, was conducted by a group of researchers at our laboratory and led to the identification of a SNP (rs17047200) in the intron of *TLL1* gene as a genetic susceptibility factor for HCV-HCC in the Japanese population. However, in the previous study, only SNPs typed by GWAS array were analyzed, so this study was designed with the aim of conducting GWAS using whole-genome imputation data of the same sample set, in order to detect genetic factors associated with HCV-HCC, not typed by GWAS array. Whole-genome imputation is a technique used to predict untyped variants using linkage disequilibrium (LD) structure of genomic regions and a reference panel such as HapMap.

Case and control samples of the current study were collected in 39 hospitals throughout Japan. The case samples were 123 HCV-induced HCC patients (including 93 males and 30 females, m:f = 3.1:1) who developed HCC after resolving HCV by IFN-therapy ≥ 1 year after the end of treatment and control samples included 336 hepatitis C patients (including 176 males and 160 females, m:f = 1.1:1) who achieved SVR by IFN-therapy and did not develop HCC ≥ 5 years after the end of treatment. Genotyping of DNA samples was performed using Affymetrix Axiom Genome-Wide ASI 1 array, designed for typing of more than 600K SNPs in Asian samples. After performing a pre-imputation quality control, IKJPN reference panel and IMPUTE2 software were used for conducting whole-genome imputation. A pre-GWAS quality control of genotypic data was conducted and SNPs with call rate < 0.97, MAF < 0.01, and those SNPs that failed the Hardy-Weinberg Equilibrium (HWE) test were filtered out. Samples with call rates less than 0.97 and those that showed cryptic relatedness were also excluded. A principal component analysis (PCA) led to the identification and removal of population outliers. Then, an association analysis was conducted on the data, using logistic regression test adjusting for gender. One locus (rs9459344, p-value = 9.42E-09, OR = 1.94, 95% CI = 1.94-3.97) showed genome-wide significance (P <

5.0E-08) and some other loci showed suggestive level of significance (P < 1.0E-04). Ninety-six SNPs (one genotyped and one imputed from each locus) were selected for subsequent validation and replication. Sixty-nine SNPs were genotyped successfully and retained for association analysis after quality control. All SNPs were validated in the discovery set with the consistency rate of 97.2% to 99.5% and the association of these SNPs in an independent replication set (including 130 cases (100 males and 30 females) with HCV-HCC after SVR by IFN-therapy ≥ 1 year after the end of the treatment and 156 controls (175 males and 181 females) that achieved SVR but did not develop HCC ≥ 3 years after the end of the treatment) was statistically examined using a logistic regression test adjusting for gender. None of SNPs passed the significance level after Bonferroni correction (P = 7.2E-04).

As the male to female ratio of cases deviated evidently from 1:1, X chromosome wide association analysis was also conducted to examine if any variants located on X chromosome affect the susceptibility of patients to HCV-HCC. However, none of SNPs of X chromosome could pass the significance threshold.

Finally, I conducted a gene-based and pathway-based association analysis using the results of GWAS. Significant results of gene-based analysis (P < 0.01) was used as input for pathway and gene ontology (GO) enrichment analysis. Some GO processes related to regulation of cell cycle and blood vessels formation and size were found to be significantly associated with HCV-HCC. Results of enrichment analyses can be interesting avenues for further investigations into their effects on the development of HCV-induced HCC.

In the second part of this study, I conducted a GWAS using genotypic data obtained from 329 Thai CHB cases (178 males and 151 females) and 320 Thai healthy controls (162 males and 158 females) using Affymetrix Axiom ASI 1 array. Quality control was conducted on the data and those SNPs with call rate missingness > 0.03, MAF < 0.05, and failure of HWE test were removed. Samples with call rate < 0.97, those with cryptic relatedness, and those that did not lie on the main Thai cluster in PCA plot were excluded from the data. Association analysis was conducted using chi-square test and p-values were obtained using a two-by-two contingency table. One locus in HLA class II region (*HLA-DPA1/DPB1*) passed the genome wide significance level (rs7770370, located in the intron of *HLA-DPB1* gene, p-value = 8.06E-10, OR = 0.49, 95%CI = 0.39-0.61). HLA region, located on the short arm of chromosome 6 (6p21), contains more than 220 genes and 21,000 alleles involved in immune responses. The significant association of HLA class II genes with CHB has been reported in different populations before. HLA allele imputation was conducted on the genotypic data in order to identify primary HLA genes and alleles conferring

susceptibility to or protection against CHB. After conducting a case-control association analysis on HLA alleles using Pearson's chi-square test, *HLA-DPA1*01:03* (Pc = 1.21E-06, OR = 0.54) and *HLA-DPB1*02:01* (Pc = 1.00E-03, OR = 0.5) exhibited significant protective effects in CHB susceptibility. *HLA-DPA1*02:02* (Pc = 6.32E-05, OR = 1.63), *HLA-DPB1*05:01* (Pc = 5.20E-05, OR = 1.72), *HLA-DPB1*13:01* (Pc = 2.16E-02, OR = 1.60), and *HLA-DQB1*03:03* (Pc = 6.47E-04, OR = 1.84) were significantly associated with susceptibility to CHB in the Thai population. The association of all these HLA alleles with CHB has been reported before. Then two-locus and three-locus HLA haplotypes were estimated by Arlequin software and association tests were performed using Pearson's chi-squared test. It was observed that *HLA-DPA1*01:03-DPB1*02:01* (Pc = 3.39E-04, OR = 0.47) was the only protective haplotype, while *HLA-DPA1*02:02-DPB1*05:01* (Pc = 1.33E-04, OR = 1.68)*, HLA-DQB1*03:03-DPB1*05:01* (Pc = 3.30E-03, OR = 2.27), *HLA-DQB1*05:02-DPB1*13:01* (Pc = 9.20E-03, OR = 5.4), *HLA-DQB1*03:03-DPA1*02:02-DPB1*05:01* (Pc = 9.68E-03, OR = 2.7) were found to be significantly associated with susceptibility to HBV persistence.

Association of a locus on chromosome 3 (rs1061307, located on 5' UTR of *PLSCR1* gene on chromosome 3, p-value = 5.92E-06, OR = 0.47, 95%CI = 0.33-0.65) and after performing whole-genome imputation associations of 3 more loci on chromosomes 4, 10, and 12 were also detected at suggestive level of significance (rs62321986, located 28 kb 5' flanking region of *PDLIM5* gene on chromosome 4, p-value = 7.69-E06, OR = 2.5, 95% CI = 1.68-3.96; rs144998273, located in the intron of *SGPL1* gene on chromosome 10, p-value = 4.65E-06, OR = 2.5, 95% CI = 1.67-3.87; and rs1828682, located in the intron of *MGST1* gene on chromosome 12, p-value = 4.82E-06, OR = 0.32, 95% CI = 0.19-0.53, respectively). *PLSCR1*, *PDLIM5, SGPL1,* and *MGST1* genes have not been reported to be associated with HBV-related diseases so far. Further replication and functional studies are required to establish their associations with CHB.

In conclusion, the present study was conducted to look for susceptibility/resistance genetic factors for HCV-HCC in the Japanese and CHB in the Thai. Further studies are required to elucidate the functional effects of SNPs on the genes. Second part of this study was the first GWAS conducted on chronic hepatitis B in the Thai population and could successfully replicate findings of previous studies and detected some novel candidate loci. In order to establish the novel findings in the Thai population more studies with larger sample sizes are required.