



Identification of additional *EHMT2* variant associated with the risk of chronic hepatitis B by GWAS follow-up study

Joong-Gon Shin^{1,2} · Hyun Sub Cheong³ · Jason Yongha Kim¹ · Jeong-Hoon Lee⁴ · Su Jong Yu⁴ · Jung-Hwan Yoon⁴ · Jae Youn Cheong⁵ · Sung Won Cho⁶ · Neung Hwa Park⁷ · Suhg Namgoong^{1,3} · Lyoung Hyo Kim³ · Yoon Jun Kim⁴ · Hyoung Doo Shin^{1,2,3}

Received: 18 May 2017 / Revised: 12 July 2017 / Accepted: 13 July 2017 / Published online: 14 December 2017
© Macmillan Publishers Limited, part of Springer Nature 2017

Abstract

Chronic hepatitis B (CHB) is a precursor to liver cirrhosis and hepatocellular carcinoma, caused by a Hepatitis B viral infection. Genome-wide association studies (GWASs) have been conducted to find genes associated with CHB risk. In previous GWAS, *EHMT2* was identified as one of the susceptibility genes for CHB. To further characterize this association and discover possible causal variants, we conducted an additional association study. A total of 11 *EHMT2* single-nucleotide polymorphisms (SNP) were selected and genotyped in 3902 subjects (1046 CHB patients and 2856 controls). An additional eight imputed SNPs were also included in further analysis. As a result, *rs35875104* showed a strong association with the CHB, along with the previously reported genetic marker for CHB risk, *rs652888* (odds ratio (OR) = 0.53, $P = 2.20 \times 10^{-8}$ at *rs35875104* and OR = 1.58, $P = 9.90 \times 10^{-12}$ at *rs652888*). In addition, linkage disequilibrium and conditional analysis identified one SNP (*rs35875104*) as a novel genetic marker for CHB susceptibility. The GRSs (genetic risk scores) were calculated to visualize the combined genetic effects of all known CHB-associated loci, including *EHMT2* *rs35875104*, which was additionally identified in this study. The findings from the present study may be useful for further understanding of the genetic etiology of CHB.

Introduction

The hepatitis B virus (HBV) is a global public health problem that affects roughly 350 million people worldwide [1]. HBV infection often leads to chronic hepatitis B (CHB),

which may eventually progress to liver failure, liver cirrhosis, or hepatocellular carcinoma (HCC) [2, 3]. The disease is of particular importance to certain Asian regions because the prevalence of HBV infection is higher there compared to other parts of the world [4]. A number of genome-wide association studies (GWASs) of CHB have been conducted recently to identify causal genes of the disease. One Japanese study reported that genes such as *HLA-DPA1* and *HLA-DPBI* are significantly associated with CHB risk [5]. Another study, also conducted in a

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41435-017-0004-x>) contains supplementary material, which is available to authorized users.

✉ Yoon Jun Kim
yoonjun@snu.ac.kr

✉ Hyoung Doo Shin
hdshin@sogang.ac.kr

- ¹ Department of Life Science, Sogang University, 1 Shinsu-dong, Mapo-gu, Seoul 121-742, Republic of Korea
- ² Research Institute for Basic Science, Sogang University, Seoul 121-742, Republic of Korea
- ³ Department of Genetic Epidemiology, SNP Genetics, Sogang University, Inc., Taihard building 1007, 1 Shinsu-dong, Mapo-gu, Seoul 121-742, Republic of Korea

⁴ Department of Internal Medicine and Liver Research Institute, Seoul National University, 28 Yungun-dong, Chongro-Gu, Seoul 110-744, Republic of Korea

⁵ Department of Internal Medicine, Seoul National University Boramae Medical Center, 20 Boramae-ro 5-gil, Dongjak-gu, Seoul 156-707, Republic of Korea

⁶ Department of Gastroenterology, Ajou University School of Medicine, San-5 Wonchon-dong, Youngtong-gu, Suwon 442-721, Republic of Korea

⁷ Department of Internal Medicine, Ulsan University Hospital, 877 Bangeojin Sunhwan-doro, Dong-gu, Ulsan 682-714, Republic of Korea

Table 1 Association of *EHMT2* genetic polymorphisms with the risk of CHB and HCC

Marker	Position	Allele	MAF	Comparing groups													
				All subjects (<i>n</i> = 3902)		CHB (<i>n</i> = 1046)		PC (<i>n</i> = 2856)		CHB vs. PC		HCC (-) vs. PC		HCC (+) vs. PC		HCC (+) vs. HCC (-)	
				All CHB (<i>n</i> = 1046)	HCC (-; <i>n</i> = 669)	HCC (+; <i>n</i> = 377)	OR (95% CI)	<i>P</i> * <i>P</i> _{corr} **	OR (95% CI)	<i>P</i> * <i>P</i> _{corr} **	OR (95% CI)	<i>P</i> * <i>P</i> _{corr} **	OR (95% CI)	<i>P</i> * <i>P</i> _{corr} **	OR (95% CI)	<i>P</i> * <i>P</i> _{corr} **	
<i>rs118097312</i>	Promoter	A > C	0.046	0.046	0.040	0.056	0.045	1.03 (0.81-1.31)	0.80	—	0.88 (0.65-1.19)	0.41	—	1.02 (0.89-1.13)	0.92	1.51 (0.66-3.45)	0.33
<i>rs7887</i>	Exon3	A > C	0.440	0.413	0.414	0.410	0.439	0.90 (0.81-0.99)	0.03	NS	0.90 (0.80-1.02)	0.09	—	0.92 (0.82-1.02)	0.11	1.38 (0.87-2.16)	0.17
<i>rs149384831</i>	Exon3	C > T	0.004	0.004	0.001	0.009	0.004	1.02 (0.46-2.30)	0.95	—	0.18 (0.03-1.37)	0.10	—	0.49 (0.16-1.45)	0.17	—	—
<i>rs114386644</i> †	Intron3	A > G	0.098	0.109	0.102	0.123	0.100	1.12 (0.95-1.32)	0.17	—	1.04 (0.86-1.27)	0.67	—	1.11 (0.80-1.53)	0.55	1.20 (0.62-2.32)	0.60
<i>rs555586</i>	Exon7	G > A	0.096	0.108	0.101	0.121	0.094	1.16 (0.99-1.37)	0.07	—	1.05 (0.86-1.28)	0.63	—	0.89 (0.74-1.06)	0.19	1.22 (0.63-2.36)	0.56
<i>rs611572</i>	Intron7	C > T	0.069	0.068	0.072	0.061	0.069	0.98 (0.81-1.19)	0.86	—	1.04 (0.83-1.31)	0.74	—	0.93 (0.75-1.15)	0.49	0.20 (0.03-1.43)	0.11
<i>rs142338646</i> †	Intron7	A > G	0.068	0.067	0.072	0.059	0.068	1.00 (0.82-1.22)	0.98	—	1.06 (0.84-1.34)	0.61	—	0.95 (0.63-1.44)	0.82	0.21 (0.03-1.50)	0.12
<i>rs35875104</i>	Intron11	T > C	0.073	0.042	0.041	0.042	0.075	0.53 (0.42-0.67)	2.20 × 10⁻⁸	8.36 × 10⁻⁷	0.54 (0.41-0.72)	3.34 × 10⁻⁵	0.0006	0.73 (0.56-0.95)	0.02	0.56 (0.13-0.83)	0.04
<i>rs113973079</i>	Exon12	C > T	0.068	0.068	0.072	0.062	0.068	1.00 (0.82-1.22)	0.98	—	1.07 (0.85-1.35)	0.57	—	0.94 (0.76-1.16)	0.56	0.19 (0.03-1.40)	0.10
<i>rs115485095</i> †	Intron12	A > G	0.093	0.107	0.100	0.120	0.093	1.14 (0.97-1.35)	0.11	—	1.07 (0.88-1.31)	0.48	—	1.14 (0.82-1.58)	0.44	1.23 (0.64-2.39)	0.53
<i>rs116027812</i> †	Intron18	A > G	0.067	0.067	0.072	0.059	0.067	1.01 (0.82-1.23)	0.95	—	1.07 (0.85-1.36)	0.55	—	0.96 (0.64-1.46)	0.86	0.21 (0.03-1.50)	0.12
<i>rs146726232</i>	Exon22	C > T	0.001	0.002	-	0.007	0.001	3.69 (0.99-13.78)	0.05	NS	—	—	—	2.41 (0.41-14.14)	0.36	—	—
<i>rs2844458</i>	Intron22	A > C	0.468	0.479	0.480	0.477	0.468	1.04 (0.94-1.15)	0.40	—	1.05 (0.93-1.18)	0.44	—	1.07 (0.96-1.19)	0.21	1.44 (0.92-2.25)	0.11
<i>rs652888</i>	Intron22	A > G	0.142	0.205	0.186	0.237	0.142	1.58 (1.38-1.79)	9.90 × 10⁻¹²	3.76 × 10⁻¹⁰	1.38 (1.18-1.61)	5.20 × 10⁻⁵	0.001	1.24 (1.07-1.43)	0.004	1.14 (0.68-1.92)	0.06
<i>rs41267090</i>	Intron27	A > T	0.087	0.109	0.105	0.115	0.087	1.29 (1.10-1.52)	0.003	NS	1.25 (1.02-1.52)	0.03	NS	1.18 (0.98-1.41)	0.08	0.48 (0.19-0.23)	0.13
<i>rs589428</i> †	Intron27	G > T	0.093	0.106	0.098	0.120	0.094	1.15 (0.98-1.35)	0.10	—	1.06 (0.87-1.29)	0.57	—	1.16 (0.84-1.61)	0.39	1.23 (0.64-2.39)	0.53
<i>rs570263</i> †	3' Flanking	C > T	0.093	0.106	0.098	0.120	0.094	1.15 (0.98-1.35)	0.10	—	1.06 (0.87-1.29)	0.57	—	1.16 (0.84-1.61)	0.39	1.23 (0.64-2.39)	0.53

Table 1 (continued)

Marker	Position	Allele	MAF	Comparing groups													
				CHB (<i>n</i> = 1046)		PC (<i>n</i> = 2856)		CHB vs. PC		HCC (-) vs. PC		HCC (+) vs. PC		HCC (+) vs. HCC (-)			
				All subjects (<i>n</i> = 3902)	HCC (-; <i>n</i> = 669)	HCC (+; <i>n</i> = 377)	All CHB (<i>n</i> = 1046)	OR (95% CI)	<i>P</i> *	<i>P</i> _{corr} **	OR (95% CI)	<i>P</i> *	<i>P</i> _{corr} **	OR (95% CI)	<i>P</i> *	<i>P</i> _{corr} **	
<i>rs146903072</i> [†]	3'	G >	0.064	0.065	0.070	0.056	0.064	1.03	0.75	—	1.10	0.41	—	0.98	0.92	0.22	0.13
	Flanking	A						(0.84–1.26)			(0.87–1.40)			(0.64–1.49)		(0.03–1.58)	
<i>rs605203</i> [†]	3'	A >	0.093	0.107	0.098	0.120	0.094	1.15	0.10	—	1.06	0.56	—	1.16	0.39	1.23	0.54
	Flanking	C						(0.98–1.35)			(0.87–1.29)			(0.83–1.60)		(0.64–2.38)	

MAF minor allele frequency, CHB chronic hepatitis B, HCC hepatocellular carcinoma, PC population control, OR odds ratio, CI confidence interval, NS not significant.

P* value of logistic analysis under additive model by adjusting for sex and age as covariates. *P* value after Bonferroni correction for multiple testing. ****P* value of logistic analysis under additive model by adjusting for sex, age, and liver cirrhosis as covariates. [†]Imputed SNPs.

Amino-acid information of SNPs in exon region was presented in the Supplementary Table S2

Significant associations are shown in bold face (*P* and *P*_{corr} < 0.05)

Japanese population, found variants of *HLA-DQ* to be associated with CHB [6]. Moreover, genetic variants of *GRIN2A* showed association with the risk of CHB in Chinese populations [7]. Previously, our group conducted a GWAS of CHB in a Korean population [8]. In the study, we reported initial associations of genes such as *euchromatic histone-lysine-methyltransferase 2 (EHMT2)* and *transcription factor 19 (TCF19)* with the risk of CHB.

EHMT2 encodes a histone methyltransferase, which plays a critical role in catalyzing the transfer of methyl groups in the histone proteins. Histone methylation is an important component of epigenetic gene regulation, and methylated histones have been found to either repress or activate transcription of various genes involved in DNA replication and repair, as well as cell proliferation [9, 10]. Consequently, *EHMT2* has been associated with various forms of cancer. A genetic variant of *EHMT2 (rs535586)* has shown a significant association with the risk of breast cancer and colorectal cancer [11, 12], and higher expressions of *EHMT2* were observed in lung and gastric cancers [13, 14]. In addition to liver cancer, epigenetic change has been associated with other liver diseases [15]. On the basis of our previous GWAS, which reported that significant association exists between *rs652888* of *EHMT2* and CHB risk ($P = 2.0 \times 10^{-9}$), we conducted a fine-mapping association follow-up study to identify possible causal variants. Furthermore, GRSs (genetic risk scores) of all known genetic factors for CHB susceptibility were calculated to evaluate the combined genetic effects regarding CHB risk.

Results

For genotyping of *EHMT2* polymorphisms, 11 single-nucleotide polymorphisms (SNPs; one in the promoter region, five in the intron region, and five in the coding region) were selected and genotyped in 3902 study subjects composed of 1046 CHB patients and 2856 healthy controls. Ethnicity of all study subjects was Korean, and study samples did not have any ancestral diversity, which was confirmed by Principal components analysis (PCA; Supplementary Fig. S1). Detailed information about the 11 genotyped SNPs, such as allele, position, minor allele frequency (MAF), heterozygosity, and Hardy–Weinberg equilibrium (HWE), in this study is shown in Supplementary Table S1. Genetic composition of the 11 genotyped SNPs in this study was not significantly different from that of other Asian populations, specifically Chinese and Japanese (Supplementary Table S2). In order to maximize coverage, additional imputation analysis was conducted using the 1000 Genomes database as a reference, and eight *EHMT2* SNPs (*rs114386644*, *rs142338646*, *rs115485095*, *rs116027812*, *rs589428*, *rs570263*, *rs146903072*, and *rs605203*) were included in

Table 2 Results of conditional association analysis of two *EHMT2* SNPs with the risk of CHB

Marker	MAF	CHB (<i>n</i> = 1046)	PC (<i>n</i> = 2856)	OR (95% CI)	<i>P</i> *	Conditional <i>P</i> value by			
						<i>HLA-DP</i>	<i>HLA-DQ</i>	<i>TCF19</i>	<i>EHMT2</i>
<i>rs35875104</i>	0.042	0.075	0.075	0.53 (0.42–0.67)	2.20×10^{-8}	<i>rs9277535</i> [†]	<i>rs2856718</i> [†]	<i>rs1419881</i> [†]	<i>rs652888</i> [†]
<i>rs652888</i>	0.205	0.140	0.140	1.58 (1.38–1.79)	9.90×10^{-12}	<i>rs3077</i> [†]	8.41×10^{-7}	3.73×10^{-6}	7.36×10^{-6}
							1.40×10^{-5}	1.93×10^{-8}	2.26×10^{-9}

MAF minor allele frequency, CHB chronic hepatitis B, PC population control, OR odds ratio, CI confidential interval

**P* value of logistic regression analysis by adjusting for sex and age as covariates. [†]Previously identified CHB-susceptible loci
Significant associations are shown in bold face

further association analysis. Among the 19 investigated polymorphisms, 16 *EHMT2* polymorphisms were used for linkage disequilibrium (LD) block construction. Three genetic variants (*rs118097312*, *rs114386644*, and *rs146726232*) were excluded from LD block construction due to its low frequency (MAF <5%). As a result, one LD block was constructed and it contains eight major haplotypes (frequency > 5%; Supplementary Fig. S2).

Logistic association analyses were performed between 19 *EHMT2* SNPs and the risk of CHB (Table 1). As a result, four SNPs (*rs7887*, *rs35875104*, *rs652888*, and *rs41267090*) showed genetic effects on the CHB with a significance (*P* < 0.05). Among the four SNPs, two in the intron region, including the previously identified CHB-associated genetic marker *rs652888*, showed strong associations with CHB susceptibility (odds ratio (OR) = 0.53, *P* = 2.20×10^{-8} at *rs35875104* and OR = 1.58, *P* = 9.90×10^{-12} at *rs652888*), although no SNPs were associated with HCC after multiple correction. Additional subgroup analyses were conducted to identify the statistical significances of CHB-related HCC progression, and *rs35875104* was found to be marginally associated with the CHB-related HCC progression (OR = 0.56, *P* = 0.04; Table 1). In the haplotype analysis, six haplotypes (*ht2*, *ht3*, *ht4*, *ht5*, *ht7*, and *ht8*) showed genetic effects on the CHB with a significance (*P* < 0.05). Among the six haplotypes, *ht7* and *ht8* were tagged with *rs652888* and *rs35875104*, respectively, and showed genetic effects on the CHB in the same direction as *rs652888* and *rs35875104*, respectively (OR = 2.14, *P* = 7.84×10^{-15} at *ht7* and OR = 0.51, *P* = 7.58×10^{-7} at *ht8*; Supplementary Table S3).

In order to confirm whether the significance of the newly identified genetic marker (*rs35875104*) was independent or influenced by known CHB risk loci, LD calculation and conditional analysis were conducted with *rs35875104* and six known and nearby CHB-susceptible loci (*rs9277535* and *rs3077* of *HLA-DP*; *rs2856716* and *rs7453920* of *HLA-DQ*; *rs1419881* of *TCF19*; *rs652888* of *EHMT2*). The *EHMT2* SNP *rs35875104* did not show tight LDs with any known, nearby CHB-susceptible loci (pairwise $r^2 \leq 0.07$; Supplementary Fig. S3). After adjusting for already known and nearby CHB-associated loci, *rs35875104* still retained its significant association with CHB risk, suggesting an independent genetic effect of *rs35875104* on CHB susceptibility (Table 2). Subsequent in silico analysis using SNP FuncPred (<https://snpinfo.niehs.nih.gov/snpinfo/snpfunc.html>) was conducted to predict the potential role of *rs35875104*. The SNP *rs35875104* has the potential to affect the exon-splicing mechanism by regulating the binding of splicing factors, such as SRp55 and SF2ASFs (Supplementary Table S4).

To investigate the detailed genetic effects of all seven CHB loci (*rs9277535* and *rs3077* of *HLA-DP*; *rs2856716* and *rs7453920* of *HLA-DQ*; *rs1419881* of *TCF19*;

Table 3 Determination of genetic risk score based on allele test of CHB-susceptible loci in a Korean population

Gene/loci	Genotype	Genotype distribution		Allele test			GRS [†]
		CHB (<i>n</i> = 1046)	PC (<i>n</i> = 2856)	Allele	OR (95% CI)	<i>P</i> *	
<i>HLA-DP1</i>	G/G	463 (44.3%)	750 (26.3%)	G	Reference		1
<i>rs9277535</i> [‡]	A/G	462 (44.2%)	1435 (50.2%)	A	0.53 (0.48–0.59)	4.92 × 10⁻³²	0.53
	A/A	121 (11.6%)	671 (23.5%)				0.27
<i>HLA-DP1</i>	G/G	538 (51.4%)	912 (31.9%)	G	Reference		1
<i>rs3077</i> [‡]	A/G	412 (39.4%)	1418 (49.6%)	A	0.53 (0.48–0.59)	1.19 × 10⁻³⁰	0.53
	A/A	96 (9.2%)	526 (18.4%)				0.27
<i>HLA-DQ1</i>	C/C	250 (23.9%)	947 (33.2%)	C	Reference		1
<i>rs2856718</i> [‡]	C/T	458 (43.8%)	1413 (49.5%)	T	1.63 (1.47–1.80)	1.86 × 10⁻²¹	1.63
	T/T	338 (32.3%)	496 (17.4%)				3.27
<i>HLA-DQ1</i>	G/G	817 (78.1%)	1782 (62.4%)	G	Reference		1
<i>rs7453920</i> [‡]	A/G	212 (20.3%)	945 (33.1%)	A	0.50 (0.43–0.58)	7.92 × 10⁻²¹	0.50
	A/A	17 (1.6%)	129 (4.5%)				0.26
<i>TCF191</i>	T/T	460 (44.0%)	976 (34.2%)	T	Reference		1
<i>rs1419881</i> [‡]	C/T	462 (44.2%)	1359 (47.6%)	C	0.71 (0.64–0.79)	9.80 × 10⁻¹¹	0.71
	C/C	124 (11.9%)	521 (18.2%)				0.36
<i>EHMT21</i>	T/T	962 (92.0%)	2452 (85.9%)	T	Reference		1
<i>rs35875104</i>	C/T	81 (7.7%)	391 (13.7%)	C	0.55 (0.43–0.70)	5.68 × 10⁻⁷	0.55
	C/C	3 (0.3%)	13 (0.5%)				0.28
<i>EHMT21</i>	A/A	669 (64.0%)	2096 (73.5%)	A	Reference		1
<i>rs652888</i> [‡]	A/G	326 (31.2%)	705 (24.7%)	G	1.56 (1.37–1.77)	1.74 × 10⁻¹¹	1.56
	G/G	51 (4.9%)	52 (1.8%)				3.12

CHB chronic hepatitis B, PC population control, OR odds ratio, CI confidence interval, GRS genetic risk score.

**P* value of logistic analysis under additive model by adjusting for sex and age as covariates. [†]GRS was calculated by multiplying the number of minor alleles by effect size (OR) of the SNP. [‡]Previously identified CHB-susceptible loci

Significant associations are shown in bold face

rs35875104 and *rs652888* of *EHMT2* in a Korean population, referent analysis based on the allele distribution of each SNP was performed. GRSs of genotypes were calculated using the OR from referent analysis (Table 3). To visualize the combined genetic effects of all seven CHB loci, including *EHMT2 rs35875104*, combined genetic effects in individuals were evaluated using the GRSs from Table 3 (Fig. 1 and Supplementary Table S5). The cumulative GRSs ranged from 3.44 (most protected group) to 11.39 (most susceptible group), and CHB patients showed significantly higher cumulative GRSs than healthy controls (Fig. 1a). Furthermore, individuals with higher cumulative GRSs showed significantly increased ORs. In particular, individuals with GRSs less than 4.8 showed an OR of 0.31

(\log_{10} OR = -0.51), while individuals with GRSs over 8.8 showed an OR of 2.41 (\log_{10} OR = 0.38; Fig. 1b).

Discussion

The role of histone methyltransferase in *EHMT2* influences the transcription of several genes that are related to various types of cancer, such as pancreatic adenocarcinoma, leukemia, breast cancer, and liver cancer [16–19]. Histone methylation plays dynamic and crucial roles in regulating the chromatin structure. Precise coordination and organization of open and closed chromatin regions control normal cellular processes such as DNA replication, repair, and transcription [20].

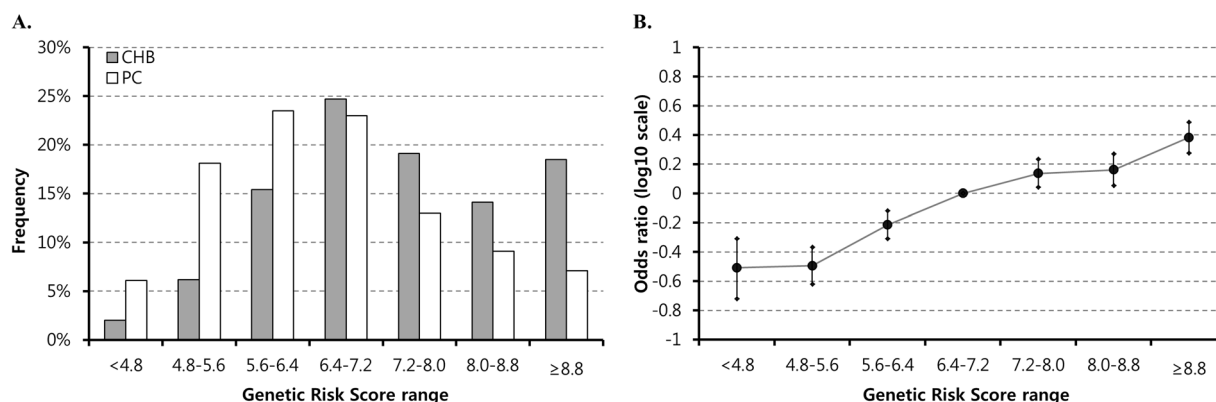


Fig. 1 Combined genetic impact of susceptible alleles from seven CHB genetic markers on the risk of CHB. **a** A comparison of GRS between CHB patients and PC. The ranges of GRS are shown below

the bars. **b** Odds ratios of different GRS range in log10 scale. Median GRS in controls was used as the reference. *GRS* genetic risk score, *CHB* chronic hepatitis B, *PC* population control

A recent study reported that a change in histone methylation may affect mRNA expression of hepatocyte nuclear factor 4 α (HNF4 α) [21]. HNF4 α is a transcription factor found in the liver and is essential for liver development and function. A previous study showed that the overexpression of HNF4 α in a human liver cancer cell line transfected with the HBV genome led to a significant increase in viral DNA synthesis [22]. Considering this evidence, a decreased expression of *EHMT2* may lead to a loose epigenetic modification and a more abundant expression of HNF4 α , which consequently may lead to an increase in HBV synthesis and ultimately, CHB.

In this study, we conducted a follow-up to our previous GWAS by performing a fine-mapping association analysis of *EHMT2* to identify possible causal variants. *EHMT2* *rs652888* was found to have a significant association with CHB risk in previous GWAS studies, and its genetic effect on the CHB was also validated [8, 23, 24]. However, we speculated that it might not be the only causal SNP. Accordingly, 11 *EHMT2* SNPs, including *rs652888*, were selected for genotyping, and additional imputation analysis was conducted. To identify the novel causal variant, LD calculation and conditional analysis were conducted. As a result, one SNP in the intron region, *rs35875104*, showed an independent genetic effect on CHB, and subsequent in silico analysis was conducted to predict the function of *rs35875104*. The analysis showed that the *rs35875104* variant was located in the exonic splicing enhancer region and promotes exon splicing that can induce exon skipping, and the production of a nonfunctional protein. The C allele of *rs35875104* led to binding of splicing factors (SRp55, SF2ASF1, and SF2ASF2) to the motifs, including variant “C” allele of *rs35875104* (Supplementary Table S4). Considering that genetic alterations of the splicing factor-binding site can be a critical pathogen in human diseases [25], *rs35875104* may have a role in the mechanism related to histone methyltransferase during viral infection.

In addition to the association analyses, combined genetic effects in individuals were evaluated using seven already known CHB-susceptible loci (Supplementary Table S5). The results indicated that seven investigated SNPs in this study might have combined genetic effects on disease susceptibility, as anticipated. When the genetic effects of individual SNPs were considered together, a stronger and more consistent effect on disease could be observed. Consistent with a number of studies that suggest GRS as a significant predictor of disease susceptibility in individuals [26–28], the current study may provide prospecting ideas for genetic risk prediction in individuals.

HBV infection is a major factor for HCC development; more than 50% of HCC patients already have CHB [29]. Although the mechanisms of hepatocarcinogenesis in CHB patients are not fully understood, several studies have provided evidence that epigenetic modifications, including histone methylation, might affect HCC development [30, 31]. HBV infection can cause host epigenetic condition changes by affecting histone modification, and several genes, including tumor suppressor gene *CDKN2A*, have been found to be downregulated with high *EHMT2* expression in HCC patients [32]. As a result of analyses conducted to investigate the genetic effects of *EHMT2* SNPs on HBV infection and/or HCC progression, two SNPs, *rs35875104* and *rs652888*, showed genetic effects on the HBV infection (OR = 0.54, $P = 0.0006$ and OR = 1.38, $P = 0.001$, respectively) and HCC (OR = 0.73, $P = 0.02$ and OR = 1.24, $P = 0.004$, respectively). Among them, *rs35875104* was found to be marginally associated with the CHB-related HCC progression (OR = 0.56, $P = 0.04$; Table 1). Considering the importance of epigenetic modification during HCC development, the genetic effects induced by *EHMT2* variants on HCC in this study suggest that *EHMT2* polymorphisms may have a role in the mechanisms related to HCC progression from HBV infection by affecting epigenetic modification.

Next, we compared the genetic composition of *EHMT2* among ethnicities to investigate the ethnic differences in *EHMT2* polymorphisms. Frequency analysis and Fisher's exact test were conducted among various ethnic groups: Korean from the present study and other Asian populations (Chinese and Japanese), as well as African and Caucasian populations from the 1000 Genomes database. Frequencies of minor alleles were found to be more similar among Asian populations (Supplementary Table S3). In the additional PCA analysis, all the study subjects in this study were clustered in the Asian population (Supplementary Fig. S2A). These results suggest that genetic effects of CHB might be influenced by ethnic differences.

There are several limitations in this study, e.g., the use of population controls (PC) whose responses to an HBV infection are unknown, and no information for infection age in case of patients, by which we could not distinguish chronic carriers infected in infancy from those infected in adulthood. Although the population controls would not secure the best statistical power, it would be an alternative way to detect the genetic effects on diseases, when considering the difficulty of recruitment of large number of disease controls. Also, study subjects in this study were mostly the same as in our previous GWAS [8]. Although an independent sample of subjects was not used, the main purpose of this study was to identify the novel causal CHB risk loci that had not been identified in our previous GWAS. Moreover, in the PCA analysis, our study subjects were distinguished from the 1000 Genome project samples of other Asian populations, Chinese and Japanese (Supplementary Fig. S2B). On the basis of the PCA analysis, our study samples did not have any ancestral diversity, which increases the power and validity of the study results.

In conclusion, a follow-up study to a previous GWAS identified a novel genetic marker (*rs35875104*) for CHB susceptibility. In addition, it confirmed the genetic effect of *rs652888* on CHB risk, this study only determined the potential functions for significantly associated variants using in silico analysis and further functional evaluations are required. In addition, the two SNPs were marginally associated with the risk of HCC. Furthermore, combined genetic effects were estimated with known CHB-susceptible loci, and these data might be used to help predict the CHB susceptibility.

Materials and methods

Study subjects

A total of 3902 subjects (1046 cases and 2856 controls) were recruited for the study. CHB patients ($n = 1046$) were obtained from the outpatient clinic of the Liver Unit and the

Center for Health Promotion at Seoul National University Hospital, Ajou University Medical Center, and Ulsan University Hospital (Seoul, Korea). The PC samples ($n = 2856$) were provided by Korea BioBank, the Center for Genome Science, the National Institute of Health, and Korea Centers for Disease Control and Prevention. The seropositivity of the hepatitis B surface antigen (HBsAg; Enzygnost® HBsAg 5.0; Dade Behring, Marburg, Germany) over a 6-month period was used to diagnose chronic HBV infection (Supplementary Table S6). HBsAg detection is determined using the Enzygnost® HBsAg 5.0 Kit assay. A blood sample of 100 μ l is added to a microplate well coated with sheep polyclonal antibodies against HBsAg, and the plate is then loaded into an assay processor, which performs all steps automatically, such as incubation, binding of the antigen, and the washing processes. Diagnosis of HCC was based on imaging findings of nodules that were larger than 1 cm, showing intense arterial uptake, followed by washout of contrast in the venous-delayed phases, in a four-phase multidetector CT scan or dynamic contrast-enhanced MRI and/or biopsy [33]. The ethnicity of all patients and healthy controls was Korean, and there was no ancestral diversity among the study subjects. The study protocol conformed to the Declaration of Helsinki. The study was approved by the institutional review board of Seoul National University Hospital, Ajou University Medical Center, and Ulsan University Hospital. All the subjects participating in the study provided written informed consent.

SNP genotyping

Candidate SNPs of *EHMT2* were selected from genotype data from Japanese and Han Chinese populations in the 1000 Genomes database (<http://browser.1000genomes.org/index.html>) based on the following conditions: (1) MAF > 5%, (2) LD status (LD coefficient (r^2) > 0.98), and (3) low-frequency (MAF \leq 5%) promoter region and non-synonymous SNPs. A total of 11 *EHMT2* SNPs were selected from 105 reported SNPs in the 1000 Genomes Asian population database and were genotyped in 1046 CHB patients and 2856 healthy controls at the multiplex level using the Golden Gate genotyping system (Illumina Inc., San Diego, CA, USA). Approximately 250 ng of genomic materials were used to genotype each sample that had undergone DNA activation, binding to paramagnetic particles, hybridization to oligonucleotides, washing, extension, ligation, amplification by polymerase chain reaction, and hybridization to the Bead plate in an appropriate hybridization buffer. Image intensities were scanned using the BeadXpress® Reader and they were genotyped using the GenomeStudio® software (Illumina Inc.). The genotype quality score for retaining data was set to 0.25. The locations of the genotyped SNPs are shown in Supplementary Fig. S2.

Statistical analysis

LD was obtained using Haploview v4.2 software downloaded from the Broad Institute (<http://www.broadinstitute.org/mpg/haploview>), with examination of Lewontin's D' (D') and the LD coefficient r^2 between all pairs of bi-allelic loci [34]. In order to increase coverage, imputation analysis was performed using IMPUTE2 software [35], with the genotypes from the 1000 Genomes database as a reference. The following criteria were applied to remove improper results from the imputation analysis: (1) call rate > 95%, (2) MAF > 5%, and (3) HWE P value > 0.05. Haplotypes of investigated SNPs were estimated using PHASE software [36]. Logistic regression models were used to compare genotype distributions, including MAF and HWE, among CHB patients and controls, and to calculate ORs, 95% confidence intervals, and corresponding P values adjusted for age (continuous value) and sex (male = 0, female = 1) as covariates using SAS, version 9.4 (SAS Inc., Cary, NC, USA). In corrections for multiple comparisons, Bonferroni correction for multiple testing was applied to the P values based on the number of independent SNPs and the analysis model. Additional subgroup analyses were also performed. Along with the age and sex, liver cirrhosis was also used as a covariate for the association analysis between HCC (+) and HCC (−) cases. In order to investigate whether the association signals of *EHMT2* SNPs are independent or affected by known CHB-susceptible loci, conditional logistic regression analyses were performed. Referent model analysis based on the allele distribution of each SNP was also performed to assess the detailed genetic effects. Six known CHB-susceptible loci from our previous study (*rs9277535* and *rs3077* of *HLA-DP*; *rs2856718* and *rs7453920* of *HLA-DQ*; *rs1419881* of *TCF19*; *rs652888* of *EHMT2*) [8] were used for the conditional analysis and referent analysis. Based on the results from referent analysis, GRSs were calculated by multiplying the number of minor alleles by effect size (OR) of the SNP. Then, the combined genetic effects for each individual were calculated as the sum of the GRSs. We also examined the distribution of cumulative GRSs, ORs, 95% confidence intervals, and corresponding P values between patients and controls. In addition, Fisher's exact test was conducted using SAS, version 9.4 (SAS Inc.). PCA was conducted to check for possible ethnic diversity using the genotype data from our previous GWAS [8] and the 1000 Genomes database.

Acknowledgements We thank the patients and their parents without whom this study would not have been possible. This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (No. HI16C1074, No. A101023); Korea Science and Engineering

Foundation (KOSEF), funded by the Korean government (MEST; No. 2011-0004453); a Sogang University Research Grant for 2011 (SRF-201114006.01); and the SNUH Research Fund (04-2016-0300).

Author contributions Conceived and designed the experiments: Y.J.K. and H.D.S. Performed the experiments: J.G.S., H.S.C., J.Y.K., S.N., and L.H.K. Analyzed the data: J.G.S., H.S.C., and H.D.S. Provided and cared for study patients: J.H.L., S.J.Y., J.H.Y., J.Y.C., S.W.C., N.H.P., and Y.J.K. Contributed to the writing of the manuscript: J.G.S., Y.J.K., and H.D.S.

Compliance with ethical standards

Competing interests The authors declare that they have no competing financial interest.

References

1. Kane MA. [World-wide epidemiology of hepatitis B]. *Soz Preventivmed.* 1998;43(Suppl 1):S98–100 S24-6.
2. Chu CM, Karayiannis P, Fowler MJ, Monjardino J, Liaw YF, Thomas HC. Natural history of chronic hepatitis B virus infection in Taiwan: studies of hepatitis B virus DNA in serum. *Hepatology.* 1985;5:431–4.
3. Hadziyannis SJ, Vassilopoulos D, Hepatitis B e antigen-negative chronic hepatitis B. *Hepatology.* 2001;34:617–24.
4. Liaw YF, Leung N, Kao JH, Piratvisuth T, Gane E, Han KH, et al Asian-Pacific consensus statement on the management of chronic hepatitis B: a 2008 update. *Hepatol Int.* 2008;2:263–83.
5. Kamatani Y, Wattanapokayakit S, Ochi H, Kawaguchi T, Takahashi A, Hosono N, et al A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. *Nat Genet.* 2009;41:591–5.
6. Mbarek H, Ochi H, Urabe Y, Kumar V, Kubo M, Hosono N, et al A genome-wide association study of chronic hepatitis B identified novel risk locus in a Japanese population. *Hum Mol Genet.* 2011;20:3884–92.
7. Liu L, Li J, Yao J, Yu J, Zhang J, Ning Q, et al A genome-wide association study with DNA pooling identifies the variant rs11866328 in the GRIN2A gene that affects disease progression of chronic HBV infection. *Viral Immunol.* 2011;24:397–402.
8. Kim YJ, Kim HY, Lee JH, Yu SJ, Yoon JH, Lee HS, et al. A genome-wide association study identified new variants associated with the risk of chronic hepatitis B. *Hum Mol Genet.* 2013;22:4233–8.
9. Jones PL, Veenstra GJ, Wade PA, Vermaak D, Kass SU, Landsberger N, et al. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet.* 1998;19:187–91.
10. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet.* 2003;33:245–54.
11. Cebrian A, Pharoah PD, Ahmed S, Roper S, Fraga MF, Smith PL, et al. Genetic variants in epigenetic genes and breast cancer risk. *Carcinogenesis.* 2006;27:1661–9.
12. de Vogel S, Wouters KA, Gottschalk RW, van Schooten FJ, de Goeij AF, de Bruine AP, et al. Genetic variants of methyl metabolizing enzymes and epigenetic regulators: associations with promoter CpG island hypermethylation in colorectal cancer. *Cancer Epidemiol Biomarkers Prev.* 2009;18:3086–96.
13. Lee SH, Kim J, Kim WH, Lee YM. Hypoxic silencing of tumor suppressor RUNX3 by histone modification in gastric cancer cells. *Oncogene.* 2009;28:184–94.
14. Chen MW, Hua KT, Kao HJ, Chi CC, Wei LH, Johansson G, et al. H3K9 histone methyltransferase G9a promotes lung cancer

- invasion and metastasis by silencing the cell adhesion molecule Ep-CAM. *Cancer Res.* 2010;70:7830–40.
15. Mann DA. Epigenetics in liver disease. *Hepatology.* 2014;60:1418–25.
 16. Son HJ, Kim JY, Hahn Y, Seo SB. Negative regulation of JAK2 by H3K9 methyltransferase G9a in leukemia. *Mol Cell Biol.* 2012;32:3681–94.
 17. Dong C, Wu Y, Yao J, Wang Y, Yu Y, Rychahou PG, et al. G9a interacts with Snail and is critical for Snail-mediated E-cadherin repression in human breast cancer. *J Clin Invest.* 2012;122:1469–86.
 18. Yuan Y, Tang AJ, Castoreno AB, Kuo SY, Wang Q, Kuballa P, et al. Gossypol and an HMT G9a inhibitor act in synergy to induce cell death in pancreatic cancer cells. *Cell Death Dis.* 2013;4:e690.
 19. Hung SY, Lin HH, Yeh KT, Chang JG. Histone-modifying genes as biomarkers in hepatocellular carcinoma. *Int J Clin Exp Pathol.* 2014;7:2496–507.
 20. Kouzarides T. Histone methylation in transcriptional control. *Curr Opin Genet Dev.* 2002;12:198–209.
 21. Zhang Q, Lei X, Lu H. Alterations of epigenetic signatures in hepatocyte nuclear factor 4alpha deficient mouse liver determined by improved ChIP-qPCR and (h)MeDIP-qPCR assays. *PLoS ONE.* 2014;9:e84925.
 22. Yu X, Mertz JE. Distinct modes of regulation of transcription of hepatitis B virus by the nuclear receptors HNF4alpha and COUP-TF1. *J Virol.* 2003;77:2489–99.
 23. Jiang DK, Ma XP, Yu H, Cao G, Ding DL, Chen H, et al. Genetic variants in five novel loci including CFB and CD40 predispose to chronic hepatitis B. *Hepatology.* 2015;62:118–28.
 24. Posuwan N, Payungporn S, Tangkijvanich P, Ogawa S, Murakami S, Iijima S, et al. Genetic association of human leukocyte antigens with chronicity or resolution of hepatitis B infection in thai population. *PLoS ONE.* 2014;9:e86007.
 25. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature.* 2011;478:64–9.
 26. Cheong HS, Lee JH, Yu SJ, Yoon JH, Lee HS, Cheong JY, et al. Association of VARS2-SFTA2 polymorphisms with the risk of chronic hepatitis B in a Korean population. *Liver Int.* 2015;35:1934–40.
 27. Fava C, Sjogren M, Olsson S, Lovkvist H, Jood K, Engstrom G, et al. A genetic risk score for hypertension associates with the risk of ischemic stroke in a Swedish case-control study. *Eur J Hum Genet.* 2015;23:969–74.
 28. Zhou XJ, Qi YY, Hou P, Lv JC, Shi SF, Liu LJ, et al. Cumulative effects of variants identified by genome-wide association studies in IgA nephropathy. *Sci Rep.* 2014;4:4904.
 29. El-Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology.* 2007;132:2557–76.
 30. Guerrieri F, Belloni L, Pediconi N, Levrero M. Molecular mechanisms of HBV-associated hepatocarcinogenesis. *Semin Liver Dis.* 2013;33:147–56.
 31. Herath NI, Leggett BA, MacDonald GA. Review of genetic and epigenetic alterations in hepatocarcinogenesis. *J Gastroenterol Hepatol.* 2006;21:15–21.
 32. Kondo Y, Shen L, Suzuki S, Kurokawa T, Masuko K, Tanaka Y, et al. Alterations of DNA methylation and histone modifications contribute to gene silencing in hepatocellular carcinomas. *Hepatol Res.* 2007;37:974–83.
 33. Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. *Hepatology.* 2011;53:1020–2.
 34. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005;21:263–5.
 35. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009;5(6):e1000529.
 36. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet.* 2001;68(4):978–89.