

RESEARCH ARTICLE

Open Access



A rare *CHD5* haplotype and its interactions with environmental factors predicting hepatocellular carcinoma risk

Qin Xiao^{1,2†}, Lianzhou Chen^{3†}, Haiqing Luo^{1,4†}, Hongmei Li^{1,5†}, Qingming Kong⁶, Fei Jiao⁷, Shifeng Pang¹, Ming Zhang⁸, Feifei Lan⁹, Wenguo Fan¹⁰, Hui Luo^{1*}, Tao Tao^{8*} and Xiao Zhu^{1*}

Abstract

Background: *CHD5* is a conventional tumour-suppressing gene in many tumours. The aim of this study was to determine whether *CHD5* variants contribute to the risk of hepatocellular carcinoma (HCC).

Methods: Gene variants were identified using next-generation sequencing targeted on referenced mutations followed by TaqMan genotyping in two case-control studies.

Results: We discovered a rare variant (haplotype AG) in *CHD5* (rs12564469-rs9434711) that was markedly associated with the risk of HCC in a Chinese population. A logistical regression model and permutation test confirmed the association. Indeed, the association quality increased in a gene dose-dependent manner as the number of samples increased. In the stratified analysis, this haplotype risk effect was statistically significant in a subgroup of alcohol drinkers. The false-positive report probability and multifactor dimensionality reduction further supported the finding.

Conclusions: Our results suggest that the rare *CHD5* gene haplotype and alcohol intake contribute to the risk of HCC. Our findings can be valuable to researchers of cancer precision medicine looking to improve diagnosis and treatment of HCC.

Keywords: *CHD5*, Gene haplotype, Hepatocellular carcinoma, Alcohol intake, Risk

Background

Hepatocellular carcinoma (HCC) is the most common primary liver cancer and has the worst prognoses of all malignancies. The etiological background of HCC patients differs between patients from different regions. In China, chronic hepatitis B virus (HBV) infection is the most important risk factor for HCC; two-thirds of the worldwide HBV carriers are Chinese, and approximately 20% of them have a chronic HBV infection [1].

Chromodomain helicase DNA-binding protein 5 (*CHD5*) is on the *Homo sapiens* chromosome 1p36.31. It

is one of the nine members of the CHD-binding enzymes and belongs to the snf2 DNA helicase/methylase superfamily [2]. *CHD5* consists of 42 exons coding for a 223 kDa protein. Based on its protein sequence, it contains two PHD zinc fingers, two chromodomains and a helicase/ATPase domain.

Evidence that *CHD5* functions as a tumour suppressor in human cancers has emerged principally from studies of neuroblastoma, wherein loss of the *CHD5* locus on chromosome 1p36.3 is very common. *CHD5* has garnered considerable interest owing to its ability to severely impact clonogenicity and tumorigenicity. Although its expression was thought to be restricted to neural-related tissues, it was subsequently found to be a tumour suppressor in neuroblastoma [3], melanoma [4], lung cancer [5], breast cancer [6], ovarian cancer [7], gastric cancer [8], colorectal cancer [9] and HCC [10]. *CHD5* loss leads to a wide range of cellular consequences, and it, therefore, remains a promising

* Correspondence: luohui@gdmu.edu.cn; tao_tao79@163.com; biozhu@yahoo.com

†Qin Xiao, Lianzhou Chen, Haiqing Luo and Hongmei Li contributed equally to this work.

¹Guangdong Provincial Key Laboratory of Medical Molecular Diagnostics, Dongguan Scientific Research Center, Guangdong Medical University, Dongguan, China

⁸Department of Gastroenterology, Zibo Central Hospital, Zibo, China
Full list of author information is available at the end of the article



candidate for further investigation in HCC. In this study, we tested the hypothesis that single-nucleotide polymorphisms (SNPs) in the 1p36 region of *CHD5* are associated with HCC.

Methods

Study subjects

First, 280 unrelated HCC patients and 255 healthy controls (admitted to the Zibo Central Hospital in North China between 2006 and 2010) were recruited for our study. Then, 549 HCC patients and 510 controls (admitted to the Peking University Shenzhen Hospital between 2007 and 2010, the First Affiliated Hospital at the Sun Yat-Sen University between 2007 and 2015, and the Cancer Hospital of Guangzhou Medical University between 2009 and 2011 in South China) were enrolled in the replication study. The selection criteria for the controls included no individual/family history of cancer or diabetes; no history of HBV, HCV, tuberculosis or HIV infection and frequency of age (± 5 years) and sex matching those of the patients. All patients were newly diagnosed, previously untreated (no radiotherapy or chemotherapy) and were proven to have no other tumours. We used published diagnostic criteria for HCC [11, 12]. The definition of 'Ever or current smokers' is those who had smoked more than 100 cigarettes, which is equal to five packs in their whole life before the date they were diagnosed with cancer or before the date they were interviewed for the controls [13, 14]. The definition of 'Ever or current drinkers' were those who have consumed alcoholic beverages \geq one time per week for 6 months or more previously; otherwise, they were defined as non-drinkers [15]. The purpose of frequency matching was to control confounding factors while evaluating the main effect of *CHD5* polymorphisms. All patients and controls were *Han* Chinese in origin and lived in China. Relevant biographical features of the subjects are summarised in Table 1.

The committee of ethics in Guangdong Medical University authorised the experimental and research protocols of this study. Experiments on humans were performed in accordance with relevant guidelines and regulations. After clearly explaining the purpose of the study to the participants, all controls and patients (or relatives of patients who already died) provided written informed consent. The study also adhered to tenets in the Helsinki declaration. All potential participants who declined to participate or ended up not participating were eligible for treatment, and non-participation did not result in any disadvantages for patients.

Targeted next-generation sequencing (NGS) and identification of genetic variants

Aliquots of buffy coat and plasma separated from blood samples were stored at -80 °C until subsequent

treatment. All samples were included in the combined study. Genomic DNA was extracted from peripheral whole blood cells using the QIAamp system (QIAGEN Co.). Genomic DNA from 255 controls and 280 HCC patients were randomly sheared by sonication to an average size of 250 bp per fragment. Target enrichment technology was used as described by Anna Kiiäläinen et al. [16]. The enriched libraries were loaded onto the HiSeq system 2000 and approximately 90-bp paired-end reads were produced using the NGS technology (Illumina Genome Analyzer). We will use fastq short reads to align the NCBI build 37.1 hg19 [17]. Single-nucleotide variants (SNV) that obey the criteria that a. P for Hardy–Weinberg equilibrium (HWE, $<10^{-4}$), b. duplicated paired-end reads, c. overall depth $\leq 8\times$, d. SNP within 10 bp of a gap, or e. copy number variant ≥ 2 were then filtered [18]. For these concerns, only qualified SNPs were considered for this evaluation, so a 164-SNP set was used for the primary case-control study. Plink was used to calculate single-nucleotide variants [19], and the Haploview was used to perform visualisation [20].

Population risk evaluation, linkage disequilibrium (LD) mapping and gene–gene interactions

We used the chi-square and Mann–Whitney U tests to compare and evaluate the clinical data between the patients and controls in discovery, replication and the combined groups. The risk evaluation was assessed using the Pearson chi-square test. Because 164 SNPs were genotyped, the Bonferroni-corrected P value for association studies is $0.05/164 = 0.0003$ for single SNPs.

A gene–gene interaction in this study is defined as an SNP–SNP interaction and was conducted with LD mapping. To estimate the degree of LD between pairs of loci, the standardised disequilibrium coefficient (D') was calculated and haplotype blocks were defined using the Haploview programme [20]. The haplotypic imputation, reconstruction and frequency estimations were conducted with an expectation–maximisation algorithm [21]. $n_e = 1/\sum p_i^2$ was used to calculate the number of effective haplotypes, and P_i was the estimate of individual haplotype frequency [22]. P_i was calculated because the phase of the genotype was known and it was chosen in compliance with the homologous probabilities of occurrence that had a higher likelihood (>0.95 as cut-point).

Permutation test and quantile–quantile (Q–Q) analysis

We performed permutation tests for 10^5 permutations, in which subjects' phenotypes were randomly realigned. P values (permutation or empirical P values) were specified as permutation values that were at least as extreme as the original statistics divided by the total permutation numbers. For better estimation of empirical P values, SNPs were reconsidered with 10^5 permutations. Permutations

Table 1 Clinical and laboratory features of the subjects included in the study

Characteristics	Discovery study			Replication study			Combined study		
	Cases (%)	Controls (%)	<i>P</i>	Cases (%)	Controls (%)	<i>P</i>	Cases (%)	Controls (%)	<i>P</i>
<i>n</i>	280	255		549	510		829	765	
Age (ys, mean ± SD)	55.1 ± 14.6	41.5 ± 9.1	< 0.001 ^a	56.6 ± 11.3	47.2 ± 10.7	< 0.001 ^a	56.0 ± 13.6	44.8 ± 10.3	< 0.001 ^a
Gender (F/M)	53/227	91/164	< 0.001 ^b	125/424	167/343	< 0.001 ^b	178/651	258/507	< 0.001 ^b
Smoking	99 (35.36)	56 (21.96)	0.001 ^b	231 (42.08)	145 (28.43)	< 0.001 ^b	330 (39.81)	201 (26.27)	< 0.001 ^b
Missing	5 (1.79)	7 (2.75)		22 (4.01)	26 (5.10)		27 (3.26)	33 (3.98)	
Drinking	95 (33.93)	54 (21.18)	0.001 ^b	210 (38.25)	129 (25.29)	< 0.001 ^b	305 (36.79)	183 (23.92)	< 0.001 ^b
Missing	8 (2.86)	7 (2.75)		28 (5.10)	29 (5.69)		36 (4.34)	36 (4.71)	
HBsAg+	224 (80.00)	0 (0.00)		419 (76.32)	0 (0.00)		643 (77.56)		
Anti-HCV	0 (0.00)	0 (0.00)		4 (0.73)	0 (0.00)		4 (0.48)	0 (0.00)	
Anti-HIV	0 (0.00)	0 (0.00)		2 (0.36)	0 (0.00)		2 (0.24)	0 (0.00)	
Serum AFP (>25 µg/L)	233 (83.21)	0 (0.00)		431 (78.51)	0 (0.00)		664 (80.10)	0 (0.00)	
Tumor size (cm)									
≤5	65 (23.21)			139 (25.32)			204 (24.61)		
>5, ≤10	93 (33.21)			273 (49.73)			366 (44.15)		
>10	122 (43.57)			137 (24.95)			259 (31.24)		
Cirrhosis									
No	16 (5.71)			38 (6.92)			54 (6.51)		
Yes	260 (92.86)			504 (91.80)			764 (92.16)		
Missing	4 (1.43)			7 (1.28)			11 (1.33)		
Tumor morphology									
No residual tumor	19 (6.79)			43 (7.83)			62 (7.48)		
Uninodular tumor	55 (19.64)			89 (16.21)			144 (17.37)		
Multinodular tumor	107 (38.21)			228 (41.53)			335 (40.41)		
Massive tumor	92 (32.86)			168 (30.60)			260 (31.36)		
Missing	7 (2.50)			21 (3.83)			28 (3.38)		
Differentiation									
Well	31 (11.07)			77 (14.03)			108 (13.03)		
Moderate	78 (27.86)			195 (35.52)			273 (32.93)		
Poor	171 (61.07)			277 (50.46)			448 (54.04)		
Metastasis									
Absent	81 (28.93)			189 (34.43)			270 (32.57)		
Present	193 (68.93)			347 (63.21)			540 (65.14)		
Missing	6 (2.14)			13 (2.37)			19 (2.29)		
TNM stage									
I	53 (18.93)			148 (26.96)			201 (24.25)		
II	95 (33.93)			230 (41.89)			325 (39.20)		
III	64 (22.86)			110 (20.04)			174 (20.99)		
IV	68 (24.29)			61 (11.11)			129 (15.56)		

F females, M males, SD standard deviation, AFP alpha fetoprotein, TNM tumor, nodes, metastasis-classification

^aKruskal-Wallis test for continuous variables

^bChi square test for categorical variables

were used to redistribute controls and patients. By convention if $P < 0.05$, the difference was considered statistically significant.

A Q–Q plot was then graphed to check the P value distribution. The ‘cumulative distribution function’ of the normal density and q th quantile of a Gauss distribution

was signified by $\Phi(z)$ and ξ_q , respectively, ($\Phi(\xi_q) = q$). Therefore, the probability $<\xi_q$ is actually just q . The theoretical quantile was defined by the inverse of the normal cumulative distribution function. Especially, the theoretical fitting the empirical quantile $z_{(i)}$ should be

$$\xi_q = q \approx \frac{i-0.5}{n},$$

for $i = 1, 2, 3, \dots, n$.

SNP selection and TaqMan genotyping in the following replication study

SNPs in *CHD5* were selected on the basis of 'significant SNPs' found in the discovery-targeted NGS results of 255 controls and 280 HCC samples. Next, genomic DNAs from all other subjects (510 controls and 549 patients) were genotyped using TaqMan probes with the ABI 7500 Fast System (Applied Biosystems, Foster City, CA) for the selected two SNPs in haplotypic block 3 (rs12564469 and rs9434711). PCRs were performed with 50 ng DNA in 25- μ l total volume containing 0.25 μ l Taq polymerase, 2.5 μ l PCR mix, 0.625 μ l of each primer and 2.5 μ l dNTPs for 40 cycles of denaturation (95 °C) for 10 min, annealing (92 °C) for 15 s and extension (60 °C) for 1 min. Associations of the potential risk SNPs or haplotypes with HCC were further evaluated by stratification analysis with subgroups of age, sex, smoking and drinking status. P_i was defined as the division of the two P numbers, which means the larger in absolute terms indicating more meaningful value.

False-positive report probability (FPRP) analysis

To avoid the possibility of false-positives inherent to performing multiple tests, a Bayesian statistical test—the FPRP—was performed for all significance in genetic association studies [23]. According to the method proposed, an FPRP value of ≤ 0.2 was regarded as pointing to a significant association, and a prior probability of 0.1 to check ORs of 1.50/0.67 was applied for risk/protective functions. The statistical power was calculated according to the case/control numbers and OR/ P values in the study.

Gene–environment interactions

The possible gene–environment interactions with high-order in the associations were evaluated using the multiple dimension reduction (MDR) programme [24]. Briefly, we carried out a 100-fold cross-validation and 1000-fold permutations under the assumption of no association. The maximum cross-validation consistency (CVC) and minimum average prediction error were requirements for the best interaction model.

Statistical software

The SPSS 22.0 for Windows (SPSS, Chicago, IL) and R scripts (3.0.2 Suite) software were used for statistical analyses.

Results

Population association risk (PAR) in the discovery study

We detected a total of 164 single-base substitutions analysing the targeted NGS results (Fig. 1a and Additional file 1: Table S1). Of these, eight were in a promoter region, 129 were intronic and 27 were in coding exons. A case-control study was conducted and the results indicated potential associations between the risk of HCC and the *CHD5* polymorphisms rs9434741 (PAR = 0.0051), rs2273032 (PAR = 0.0089) and rs12067480 (PAR = 0.0261) in the Han population (Fig. 1b and Additional file 1: Table S1). But they lost statistical significance after performing a Bonferroni correction. They also lost their significance after 10^5 permutation tests (for example, $P = 0.3156$ for rs9434741, Fig. 1c). Q–Q plots were used to compare with the observed chi-square results with the distribution expected under the null hypothesis, there was deviation from expectation at a higher value of approximately 2.8 (Fig. 1d). After removing rs9434741, there were no significant curve changes compared with the expected distribution (Fig. 1e).

LD and haplotypic analysis in the discovery study

Direct sequencing results revealed a total of 164 SNPs in *CHD5*. We identified three blocks with high LD (Fig. 1a). Block 1 includes SNP3–SNP6 (rs12037962, rs11587, rs41307753 and rs3810989). Block 2 includes SNP35–SNP38 (rs2273041, rs2273040, rs2273038 and rs55930553). Block 3 includes SNP115 and SNP116 (rs12564469 and rs9434711). Blocks were reconstructed according to their frequencies. The results of the haplotype-based case-control study between the HCC and control groups are shown in Table 2. We found that a haplotype AG in block 3 showed a significant association with HCC ($P = 1.94 \times 10^{-5}$). It remained significant according to unconditional logistic regression analysis after adjustment for age, sex, smoking and drinking status ($P_{\text{corrected}} = 5.73 \times 10^{-5}$) and after 10^5 permutation tests ($P = 4.00 \times 10^{-5}$).

Population association and haplotypic analysis based on selected SNPs in the replication and combined studies

We selected SNPs rs12564469 and rs9434711 in block 3 from the first SNP discovery study for the next study. Replicative results showed no associations for rs12564469 (PAR = 0.0800, $P_{\text{adjusted}} = 0.1029$, $P_{\text{permutation}} = 0.1062$) or for rs9434711 (PAR = 0.8718, $P_{\text{adjusted}} = 0.8485$, $P_{\text{permutation}} = 0.9601$). Finally, a combined study including discovery and replicative cohort data was conducted. Combined

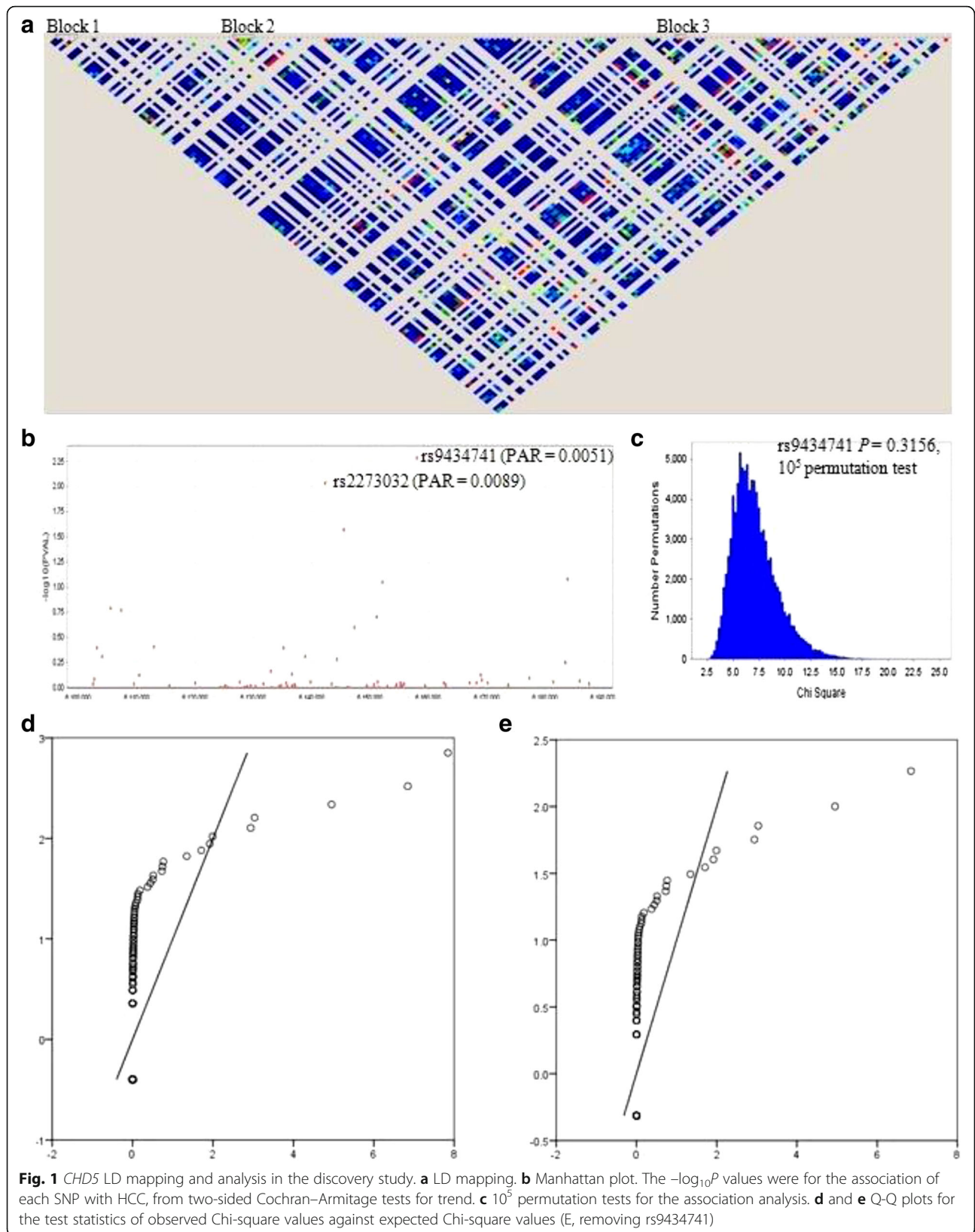


Fig. 1 CHD5 LD mapping and analysis in the discovery study. **a** LD mapping. **b** Manhattan plot. The $-\log_{10}P$ values were for the association of each SNP with HCC, from two-sided Cochran–Armitage tests for trend. **c** 10^5 permutation tests for the association analysis. **d** and **e** Q-Q plots for the test statistics of observed Chi-square values against expected Chi-square values (E, removing rs9434741)

Table 2 Haplotype frequencies in the discovery, replication and combined studies

Haplotypes	Case, Control Ratio Counts ^a	Case, Control Frequencies ^b	Chi Square	PAR	$P_{corrected}$ ^c	$P_{Permutation}$ ^d
Discovery study						
Block 1						
GGCA	176.0: 384.0, 171.1: 338.9	0.314, 0.335	0.54	0.4623	0.2970	0.9976
GACA	176.0: 384.0, 171.1: 338.9	0.314, 0.335	0.54	0.4623	0.2970	0.9976
AGCG	67.5: 492.5, 61.1: 448.9	0.121, 0.120	0.002	0.9651	0.8263	1
AACG	67.5: 492.5, 61.1: 448.9	0.121, 0.120	0.002	0.9651	0.8263	1
AGCA	29.5: 530.5, 20.9: 489.1	0.053, 0.041	0.795	0.3727	0.6037	0.9876
AACA	29.5: 530.5, 20.9: 489.1	0.053, 0.041	0.795	0.3727	0.6037	0.9876
Block 2						
CCCG	417.3: 126.7, 375.9: 118.1	0.767, 0.761	0.055	0.8153	0.7749	1
TTTA	42.0: 502.0, 37.3: 456.7	0.077, 0.076	0.011	0.9171	0.9690	1
CCTG	36.0: 508.0, 35.5: 458.5	0.066, 0.072	0.128	0.7208	0.8452	1
TTCA	25.5: 518.5, 25.2: 468.8	0.047, 0.051	0.098	0.7544	0.7805	1
CCCA	9.5: 534.5, 8.4: 485.6	0.017, 0.017	0.002	0.9622	0.9417	1
TTTG	7.7: 536.3, 5.8: 488.2	0.014, 0.012	0.121	0.7282	0.7548	1
Block 3						
AA	324.0: 170.0, 289.0: 151.0	0.656, 0.657	0.001	0.9757	0.8983	1
GG	143.8: 350.2, 148.8: 291.2	0.291, 0.338	2.399	0.1214	0.1665	0.5747
AG	26.2: 467.8, 2.2: 437.8	0.053, 0.005	18.248	1.94×10^{-5}	5.73×10^{-5}	4.00×10^{-5}
Replication study						
Block 3						
AA	630.9: 341.1, 579.0: 309.0	0.649, 0.652	0.018	0.8945	0.8714 ⁸	0.9893
GG	294.5: 677.5, 303.5: 584.5	0.303, 0.342	3.202	0.0735	0.1069	0.1542
AG	46.6: 925.4, 5.5: 882.5	0.048, 0.006	29.716	5.038×10^{-8}	7.571×10^{-8}	0.00001
Combined study						
Block 3						
AA	954.9: 511.1, 868.0: 460.0	0.651, 0.654	0.015	0.9012	0.9467	0.9909
GG	438.3: 1027.7, 452.3: 875.7	0.299, 0.341	5.556	0.0184	0.0383	0.0410
AG	72.7: 1393.3, 7.7: 1320.3	0.050, 0.006	47.941	4.393×10^{-12}	5.514×10^{-11}	0.00001

Block 1, rs12037962, rs11587, rs41307753 and rs3810989

Block 2, rs2273041, rs2273040, rs2273038 and rs55930553

Block 3, rs12564469 and rs9434711

^aNumber of haplotypes were compared in cases versus controls: Haplotype(1):haplotype(others) cases, Haplotype(1):haplotype(others) controls^bFrequency of the haplotype^cCalculated in logistical regression models with adjustment for age, gender, smoking and drinking status; $p < 0.005$ means significant value by Bonferroni correction based on the total number of markers genotyped^dEmpirical p -value based on 10^5 permutations of case-control status using the max(T) procedure. $p < 0.05$ means significant value

results also showed no association for rs12564469 (PAR = 0.0210, $P_{adjusted}$ = 0.0290, $P_{Permutation}$ = 0.0286) and for rs9434711 (PAR = 0.8829, $P_{adjusted}$ = 0.9137, $P_{Permutation}$ = 0.9704; Table 3).

The results of the haplotype-based replication and combined studies between the HCC and control groups are shown in Table 2. We observed increased frequencies of haplotype AG in HCC patients compared with those seen in healthy controls both in the replication study (PAR = 5.038×10^{-8} , $P_{adjusted}$ = 7.571×10^{-8} , $P_{Permutation}$ = 0.00001)

and in the combined study (PAR = 4.393×10^{-12} , $P_{adjusted}$ = 5.514×10^{-11} , $P_{Permutation}$ = 0.00001).

Stratification analysis of haplotypes

The association of haplotype AG (block 3) with the risk of HCC in subgroups such as age, sex, smokers and drinkers were evaluated further using replication and combined studies (Table 4). We found that those individuals carrying haplotype AG had a significantly increased risk of HCC, and the risk was

Table 3 rs12564469 and rs9434711 in replication and combined studies

	Alleles ^a	Case, Control Ratio Counts ^b	Case, Control Frequencies ^c	Chi square	PAR ^d	P _{adjusted} ^e	P _{permutation} ^f
Replication							
rs12564469	A > G	659:289, 570:298	0.695, 0.657	3.065	0.0800	0.1029	0.1062
rs9434711	A > G	341:629, 309:579	0.352, 0.348	0.026	0.8718	0.8485	0.9601
Combined							
rs12564469	A > G	1003:431, 857:445	0.699, 0.658	5.328	0.0210	0.0290	0.0286
rs9434711	A > G	511:953, 460:868	0.349, 0.346	0.022	0.8829	0.9137	0.9704

^aThe major allele is listed first, then the minor allele

^bNumber of alleles were compared in cases versus controls: allele(1):allele(2) cases, allele(1):allele(2) controls

^cFrequency of the association allele

^dPAR, population attributable risk

^eCalculated in logistical regression models with adjustment for age, gender, smoking and drinking status

^fP for 10⁵ permutation test

increased in patients of >55 years ($P = 6.04 \times 10^{-8}$ and $P_i (P_2/P_1) = 5.12 \times 10^{-4}$) and in drinkers ($P = 9.43 \times 10^{-8}$ and $P_i (P_2/P_1) = 3.25 \times 10^{-6}$).

FPRP

The significant associations of FPRP values for block 3 haplotype AG (vs. AA + GG) at different levels of prior probability are listed in Table 5. FPRP values of haplotype AG for HCC risk in patients >55 years were <0.20 for the assigned prior probability (0.017 for the prior probability of 0.1 in the replication study; 0.004 and 0.010 for the prior probabilities of 0.1 and 0.01, respectively, in the combined study). For the risk of HCC in alcohol drinkers, when the assumptions of prior probability were 0.1 and 0.01, all findings were significant not only in the discovery study but also in the replication and combined studies (FPRP < 0.20). Moreover, when the assumption of prior probability was 0.001, this association was still prominent in the combined study (FPRP = 0.069).

Association of high-order interactions with HCC risk by MDR

The interactions of high-order assessed with MDR were conducted, including the potential risk haplotype AG and four known risk factors (age, sex, smoking and drinking status), in order to check whether possible gene–environmental interactions in association with the risk of HCC exists. In the discovery study, we noticed that the best one-factor model was drinking status, with the highest CVC (99/100, the same model is selected as the best model 99 out of 100 times) and the lowest prediction error (0.385). The best model for two-factors was drinking status plus haplotype AG, with the highest CVC (96/100) and the lowest prediction error (0.403). Interestingly, the model with 5-factors had a maximum CVC (100/100) and a minimum prediction error (0.378). This is a model with better prediction than the model

with one factor. Same results were found in the replication study and the combined study (Table 6).

Discussion

Studies have found that the chromosome aberration of 1p36 deletion is not frequent in HCC. It remains to be determined whether the common SNPs in *CHD5* are associated with the risk of HCC. *CHD5* is a tumour-suppressing gene of the chromodomain gene family, first identified as a tumour-suppressing gene mapping to 1p36.31 [25].

The integration of clinical phenotypes and genomic information may enable precision cancer medicine through NGS approaches [26]. Results of our targeted NGS and TaqMan genotyping revealed no significant associations with the risk of HCC neither in the discovery study nor in the replication and combined studies. For two data sets, it is important to identify whether the hypothesis of a common distribution is proven to be true. The Q–Q plot offers more insight into the discrepancy than any other statistical analysis such as the Kolmogorov Smirnov 2-sample test or the chi-square test. However, we did not find any significant change after removing rs9434741, which suggests that the most likely associated SNP is not a risk locus.

Nonetheless, we inadvertently found a positive association of a rare haplotype AG (block 3: rs-12564469-rs9434711) in *CHD5* and HCC, which has not been reported to date. Importantly, this association quality increased in a gene dose-dependent manner as the number of samples increased (PAR in Table 2). Thus, our results support the idea that the 1p36 region plays a role in HCC. We believe it is possible that hereditary mutations of tumour-suppressing genes in the 1p36 region contribute to the aggressive properties of liver cancer. Hereditary changes in the 1p36 region are extraordinarily common in human tumours, occurring in malignancies of epithelial, neural and haematopoietic origin [25]. Genetic mutations of the tumour-suppressing gene *CHD5* have conducted to the understanding of human oncogenesis.

Table 4 Stratification analysis for associations between block 3 (rs12564469-rs9434711) haplotypes and HCC risk in the discovery, replication and combined studies

Variables	Discovery study			Replication study			Combined study					
	Cases/controls	OR (95% CI)	P	Cases/controls	OR (95% CI)	P	Cases/controls	OR (95% CI)	P			
	AG AA+GG			AG AA+GG			AG AA+GG					
Age (ys)												
≤55	10/06	123.2/128.8	8.57 (1.28–7603)	0.010	132/17	231.5/198.5	5.11 (1.20–23.49)	0.026	23.2/2.3	354.7/327.3	9.62 (2.37–41.84)	1.18 × 10 ⁻⁴
>55	16.2/1.6	344.6/309	6.39 (1.60–28.85)	0.007	33.4/3.8	651/684	7.88 (2.64–20.18)	8.23 × 10 ⁻⁵	49.6/5.4	995.6/993	9.13 (3.36–22.70)	6.04 × 10 ⁻⁸
P _i			0.7					3.17 × 10 ⁻³				5.12 × 10 ⁻⁴
Gender												
Females	6/0.2	70.2/139.6	2.81 (2.22–3.47)	0.008	13.6/1.8	231.4/268	7.35 (1.76–33.38)	0.008	19.6/2	301.6/407.6	12.43 (3.03–54.68)	6.36 × 10 ⁻⁵
Males	20.2/2	397.6/298.2	7.15 (1.55–2907)	0.009	33/3.7	694/614.5	7.09 (2.38–17.14)	7.95 × 10 ⁻⁴	53.2/5.7	1091.6/912.7	7.35 (3.02–15.98)	8.61 × 10 ⁻⁷
P _i			1.125					0.099				0.014
Drinking												
Never	3/2.2	320.2/333.8	1.78 (0.30–11.79)	0.505	9.6/1.5	574.4/652	5.17 (1.19–22.93)	0.025	12.6/3.7	894.6/985.8	3.36 (1.11–9.83)	0.029
Ever+current	23.2/0	147.6/104	1.64 (1.47–1.75)	1.07 × 10 ⁻⁴	37/4	351/230.5	5.88 (2.03–16.01)	8.76 × 10 ⁻⁵	60.2/4	498.6/334.5	9.78 (3.56–24.83)	9.43 × 10 ⁻⁸
P _i			2.12 × 10 ⁻⁴					3.50 × 10 ⁻³				3.25 × 10 ⁻⁶

P, P value for haplotype model, which obtained in logistic regression with adjustment for age, sex, smoking status and drinking status

P_i means P_i/P_t

Table 5 FPRP values for associations between HCC risk and block 3 haplotype frequencies (AG vs. AA+GG)

Variables	Statistical power ^a	Prior probability			
		0.1	0.01	0.001	0.0001
HCC risk in >55 years old group					
Discovery study	0.704	0.216	0.493	0.721	0.885
Replication study	0.689	0.017	0.271	0.525	0.843
Combined study	0.837	0.004	0.010	0.347	0.706
HCC risk in drinking group					
Discovery study	0.792	0.003	0.013	0.298	0.635
Replication study	0.658	0.005	0.017	0.424	0.757
Combined study	1	< 0.001	0.005	0.069	0.236

Block 3, rs12564469 and rs9434711

If the prior probability <0.20, the results in FPRP are in bold

It seems that the risk effect of the haplotype AG was more evident in the drinkers' subgroup (Ref: P_i in Table 4) with the stratified analysis. One of the possible comments is that the sample size is smaller in subgroups. Nevertheless, the results of the FPRP analysis for those findings showed that the drinkers group remained significant at the prior probability level of 0.1. We believe that in drinkers,

Table 6 MDR analysis for the prediction of HCC risk with and without haplotype AG

Best interaction models	Cross-validation	Average prediction error	P^a
Discovery study			
1	99/100	0.385	< 0.0001
1,2	96/100	0.403	< 0.0001
1,2,3	100/100	0.401	< 0.0001
1,2,3,4	87/100	0.380	< 0.0001
1,2,3,4,5 ^b	100/100	0.378	< 0.0001
Replication study			
1	95/100	0.412	< 0.0001
1,2	94/100	0.417	< 0.0001
1,2,3	98/100	0.389	< 0.0001
1,2,3,4	90/100	0.383	< 0.0001
1,2,3,4,5 ^b	100/100	0.368	< 0.0001
Combined study			
1	96/100	0.399	< 0.0001
1,2	94/100	0.410	< 0.0001
1,2,3	99/100	0.396	< 0.0001
1,2,3,4	89/100	0.382	< 0.0001
1,2,3,4,5 ^b	100/100	0.375	< 0.0001

Labels: 1, drinking status; 2, haplotype AG (block 3); 3, age; 4, smoking status; 5, gender

^a P value for 1000-fold permutation test^bThe best model with maximum cross-validation consistency and minimum prediction error rate

alcohol-related carcinogens may cause DNA damage [27] and that accumulated DNA damage caused by the regular carcinogenic exposure to alcoholic drinks [28, 29] might enhance the effect of genetic instability.

Next, we conducted a high-order gene (haplotype)-environment interaction analysis with MDR testing to support the above results. The best interaction model revealed that the *CHD5* haplotype AG interacted with the drinking status with a maximal CVC and minimal prediction error, which was more obvious in the interaction entropy analysis. Our results suggested that the stratification testing reliably identified alcohol drinking as a risk factor.

Our recent study had reported that the *CHD5* rs12564469-rs9434711 region might functionally contribute to HCC prognosis and *CHD5* mRNA expressions [30]. It is possible that *CHD5* plays an essential role in cancer development. The expression of multiple genes that regulate pathways in the tumorigenic process was modulated by *CHD5* [31]. Apoptosis, cellular senescence and neonatal death will occur by excessive activation of these tumour-suppressive pathways, dependent on p53, p19 and p16. *CHD5* expression seems to be restricted to neural-derived tissues, as opposed to *CHD4* which is expressed in all tissues. *CHD5* mRNA cannot be detected in the liver, placenta, spleen, bone marrow, thyroid, stomach, pancreas, small intestine, colon or prostate [8, 30]. Because of this, expression of the candidate tumour-suppressing genes was sequentially disrupted by specific shRNAs. What is more, *CHD5* expression is down-regulated in HCC tissues and HepG2, and the expression level of *CHD5* was inversely correlated with the expression of oncogene miR-454 in HCC tissues [32]. Therefore, *CHD5* as the cause of the observed phenotype was identified.

Alternatively, *CHD5* or a *CHD5*-containing complex could interact with p53 directly. A similar model for a MTA2-containing NuRD complex regulating the p53-mediated transactivation by modulating the p53 acetylation status [33] was suggested. *CHD5* may function in a similar manner since it was shown to be part of a NuRD-like complex [34]. Both the interactions and functions are equally important for the development of HCC. The genetic engineering mice with a heterozygous deficiency of the (human) 1p36 locus were prone to develop non-neural tumours (lymphoma, squamous cell carcinoma and hibernoma). *CHD5* was found to positively regulate p53 presumably via p14/p19ARF [35, 36]. But the exact molecular mechanisms could not be defined.

Conclusions

In short, we identified a rare haplotype in *CHD5* that was significant associated with the risk of HCC. Our results highlight the breadth of precision medicine by providing clues to help the advancement of effective diagnostic, management and prevention tools against cancer. Nonetheless, larger sample size studies are needed to corroborate our findings.

Additional file

Additional file 1: Table S1. A case-control study in the discovery study with targeted NGS. (XLS 37 kb)

Abbreviations

CHD5: Chromodomain helicase DNA-binding protein 5; FPRP: False-positive report probability Population risk evaluation; HCC: Hepatocellular carcinoma; LD: Linkage disequilibrium; MDR: Multiple dimension reduction; NGS: Targeted next-generation sequencing; Q-Q: Permutation test and quantile-quantile; SNPs: Single-nucleotide polymorphisms

Acknowledgments

We would like to thank all the participants that contribute to this work. The authors would like to thank Enago (www.enago.cn) for the English language review.

Funding

Supported by Guangdong Provincial Science and Technology Programs (2014A020212653 and 2016A050503046); The Public Service Platform of South China Sea for R&D Marine Biomedicine Resources (2018008), and Science and Technology Research Project in Dongguan City (2014108101048).

Availability of data and materials

The datasets generate and analyzed in this study are not publicly available for the reason of protecting patients' privacy, but are available from the corresponding author (TT) on reasonable request.

Authors' contributions

TT, XZ and HL designed the study. LC, QX, HaiL, HLi, MZ, FL and XZ analyzed the patient data and carried out the genotyping. QK, FJ and SP performed the statistical analyzes. XZ, WF and TT wrote the manuscript. LC, HaiL, MZ and TT contributed samples and patient information. All authors read and approved the final draft of the manuscript.

Ethics approval and consent to participate

The ethics committee of Guangdong Medical University authorised the experimental and research protocols of this study. All procedures performed in studies were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. All controls and patients (or relatives of patients who already died) provided written informed consent.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Guangdong Provincial Key Laboratory of Medical Molecular Diagnostics, Dongguan Scientific Research Center, Guangdong Medical University, Dongguan, China. ²Department of Blood Transfusion, Peking University Shenzhen Hospital, Shenzhen, China. ³Digestive System Tumor Tissue Bank,

Center of Surgery Laboratory, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China. ⁴The Affiliated Hospital Cancer Center, Guangdong Medical University, Zhanjiang, China. ⁵Department of Pathology, Guangdong Medical University, Dongguan, China. ⁶Immunity and Biochemical Research Lab, Zhejiang Academy of Medical Sciences, Hangzhou, China. ⁷Department of Biochemistry and Molecular Biology, Binzhou Medical University, Yantai, China. ⁸Department of Gastroenterology, Zibo Central Hospital, Zibo, China. ⁹Forensic Identification Institute, Guangdong Women and Children Hospital, Guangzhou, China. ¹⁰Guanghua School of Stomatology, Hospital of Stomatology, Sun Yat-sen University, Guangzhou, China.

Received: 31 December 2017 Accepted: 24 May 2018

Published online: 15 June 2018

References

- Zhang Y, Liu H, Yi R, Yan T, He Y, Zhao Y, Liu J. Hepatitis B virus whole-X and X protein play distinct roles in HBV-related hepatocellular carcinoma progression. *J Exp Clin Cancer Res.* 2016;35(1):87.
- Baykara O, Tansarikaya M, Bulut P, Demirkaya A, Buyru N. *CHD5* is a potential tumor suppressor in non small cell lung cancer (NSCLC). *Gene.* 2017;618:65–8.
- Higashi M, Kolla V, Iyer R, Naraparaju K, Zhuang T, Kolla S, Brodeur GM. Retinoic acid-induced *CHD5* upregulation and neuronal differentiation of neuroblastoma. *Mol Cancer.* 2015;14:150.
- Lang J, Tobias ES, Mackie R. Preliminary evidence for involvement of the tumour suppressor gene *CHD5* in a family with cutaneous melanoma. *Br J Dermatol.* 2011;164(5):1010–6.
- Zhao R, Yan Q, Lv J, Huang H, Zheng W, Zhang B, Ma W. *CHD5*, a tumor suppressor that is epigenetically silenced in lung cancer. *Lung Cancer.* 2012; 76(3):324–31.
- Wu X, Zhu Z, Li W, Fu X, Su D, Fu L, Zhang Z, Luo A, Sun X, Fu L, et al. Chromodomain helicase DNA binding protein 5 plays a tumor suppressor role in human breast cancer. *Breast Cancer Res.* 2012;14(3):R73.
- Wong RR, Chan LK, Tsang TP, Lee CW, Cheung TH, Yim SF, Siu NS, Lee SN, Yu MY, Chim SS, et al. *CHD5* downregulation associated with poor prognosis in epithelial ovarian cancer. *Gynecol Obstet Investig.* 2011; 72(3):203–7.
- Xu G, Zhu H, Zhang M, Xu J. Histone deacetylase 3 is associated with gastric cancer cell growth via the miR-454-mediated targeting of *CHD5*. *Int J Mol Med.* 2018;41(1):155–63.
- Fatemi M, Paul TA, Brodeur GM, Shokrani B, Brim H, Ashktorab H. Epigenetic silencing of *CHD5*, a novel tumor-suppressor gene, occurs in early colorectal cancer stages. *Cancer.* 2014;120(2):172–80.
- Fang QL, Yin YR, Xie CR, Zhang S, Zhao WX, Pan C, Wang XM, Yin ZY. Mechanistic and biological significance of DNA methyltransferase 1 upregulated by growth factors in human hepatocellular carcinoma. *Int J Oncol.* 2015;46(2):782–90.
- Zhu X, Zhang J, Fan W, Wang F, Yao H, Wang Z, Hou S, Tian Y, Fu W, Xie D, et al. The rs391957 variant cis-regulating oncogene *GRP78* expression contributes to the risk of hepatocellular carcinoma. *Carcinogenesis.* 2013; 34(6):1273–80.
- Sangiovanni A, Colombo M. Treatment of hepatocellular carcinoma: beyond international guidelines. *Liver Int.* 2016;36(Suppl 1):124–9.
- Brailion A. Recurrence in early-stage hepatocellular carcinoma: CpG methylation and smoking. *J Clin Oncol.* 2017;35(18):2097–8.
- Tarabichi Y, Kats DJ, Kaelber DC, Thornton JD. The impact of fluctuations in pack-year smoking history in the electronic health record on lung cancer screening practices. *Chest.* 2018;153(2):575–8.
- Offermans NSM, Ketcham SM, van den Brandt PA, Weijnenberg MP, Simons C. Alcohol intake, *ADH1B* and *ADH1C* genotypes, and the risk of colorectal cancer by sex and subsite in the Netherlands Cohort Study. *Carcinogenesis.* 2018;39(3):375–88.
- Kiialainen A, Karlberg O, Ahlford A, Sigurdsson S, Lindblad-Toh K, Syvanen AC. Performance of microarray and liquid based capture methods for target enrichment for massively parallel sequencing and SNP discovery. *PLoS One.* 2011;6(2):e16486.
- Garcia-Seco D, Zhang Y, Gutierrez-Manero FJ, Martin C, Ramos-Solano B. RNA-Seq analysis and transcriptome assembly for blackberry (*Rubus sp.* Var. *Lochness*) fruit. *BMC Genomics.* 2015;16:5.

18. Perkins BA, Caskey CT, Brar P, Dec E, Karow DS, Kahn AM, Hou YC, Shah N, Boeldt D, Coughlin E, et al. Precision medicine screening using whole-genome sequencing and advanced imaging to identify disease risk in adults. *Proc Natl Acad Sci U S A*. 2018;115(14):3686–91.
19. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81(3):559–75.
20. Smith AV. Manipulating HapMap Data Using HaploView. *CSH Protoc*. 2008; 2008.pdb prot5025.
21. Dai JY, Ruczinski I, LeBlanc M, Kooperberg C. Imputation methods to improve inference in SNP association studies. *Genet Epidemiol*. 2006;30(8): 690–702.
22. Long JC, Williams RC, Urbanek M. An E-M algorithm and testing strategy for multiple-locus haplotypes. *Am J Hum Genet*. 1995;56(3):799–810.
23. Obazee O, Capurso G, Tavano F, Archibugi L, De Bonis A, Greenhalf W, Key T, Pasquali C, Milanetto AC, Hackert T, et al. Common genetic variants associated with pancreatic adenocarcinoma may also modify risk of pancreatic neuroendocrine neoplasms. *Carcinogenesis*. 2018;39(3):360–7.
24. Moore JH, Andrews PC. Epistasis analysis using multifactor dimensionality reduction. *Methods Mol Biol*. 2015;1253:301–14.
25. Henrich KO, Schwab M, Westermann F. 1p36 tumor suppression—a matter of dosage? *Cancer Res*. 2012;72(23):6079–88.
26. Gagan J, Van Allen EM. Next-generation sequencing to guide cancer therapy. *Genome Med*. 2015;7(1):80.
27. Cederbaum AI, Lu Y, Wu D. Role of oxidative stress in alcohol-induced liver injury. *Arch Toxicol*. 2009;83(6):519–48.
28. Xie Z, Tan G, Ding M, Dong D, Chen T, Meng X, Huang X, Tan Y. Foxm1 transcription factor is required for maintenance of pluripotency of P19 embryonal carcinoma cells. *Nucleic Acids Res*. 2010;38(22):8027–38.
29. El Idrissi M, Hervieu V, Merle P, Mortreux F, Wattel E. Cause-specific telomere factors deregulation in hepatocellular carcinoma. *J Exp Clin Cancer Res*. 2013;32:64.
30. Zhu X, Kong Q, Xie L, Chen Z, Li H, Zhu Z, Huang Y, Lan F, Luo H, Zhan J, et al. The single-nucleotide polymorphisms in CHD5 affect the prognosis of patients with hepatocellular carcinoma. *Oncotarget*. 2018;9(17):13222–30.
31. Tamura S, Wang Y, Veeneman B, Hovelson D, Bankhead A 3rd, Broses LJ, Lorenzatti Hiles G, Liebert M, Rubin JR, Day KC, et al. Molecular correlates of in vitro responses to dacomitinib and afatinib in bladder cancer. *Bladder Cancer*. 2018;4(1):77–90.
32. Yu L, Gong X, Sun L, Yao H, Lu B, Zhu L. miR-454 functions as an oncogene by inhibiting CHD5 in hepatocellular carcinoma. *Oncotarget*. 2015;6(36): 39225–34.
33. Luo J, Su F, Chen D, Shiloh A, Gu W. Deacetylation of p53 modulates its effect on cell growth and apoptosis. *Nature*. 2000;408(6810):377–81.
34. Kolla V, Naraparaju K, Zhuang T, Higashi M, Kolla S, Blobel GA, Brodeur GM. The tumour suppressor CHD5 forms a NuRD-type chromatin remodelling complex. *Biochem J*. 2015;468(2):345–52.
35. Kolla V, Zhuang T, Higashi M, Naraparaju K, Brodeur GM. Role of CHD5 in human cancers: 10 years later. *Cancer Res*. 2014;74(3):652–8.
36. Bagchi A, Papazoglu C, Wu Y, Capurso D, Brodt M, Francis D, Bredel M, Vogel H, Mills AA. CHD5 is a tumor suppressor at human 1p36. *Cell*. 2007; 128(3):459–75.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

