Associations of the Polymorphisms of the NHEJ Pathway Genes With HIV-1 Infection and Aids Progression Among Men Who Have Sex With Men in Northern China

Xuelong Zhang
Harbin Medical University
Kaili Wang
Infectious Disease Hospital of Heilongjiang Province
Han Mo
Harbin Medical University
Yuanting Hu
Harbin Medical University
Xun Yang
Harbin Medical University
Jiawei Wu
Harbin Medical University
Bangquan Liu
Harbin Medical University
Lidan Xu
Harbin Medical University
Haiming Sun
Harbin Medical University
Xueyuan Jia
Harbin Medical University
Ping Wang
Harbin Medical University
Wenjing Sun
Harbin Medical University
Songbin Fu
Harbin Medical University
Yuandong Qiao (qiaoyuandong@hrbmu.edu.cn)
Harbin Medical University

Research Article

Keywords: NHEJ, HIV-1, AIDS, SNP, MSM

Posted Date: May 20th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-505071/v1

Abstract

Background: Men who have sex with men (MSM) are at high risk of HIV infection. Non-homologous end joining (NHEJ) pathway is the main way of double-stranded DNA break (DSB) repair in the higher eukaryotes, and can repair the DSB timely at any time in cell cycle. The objective of this study was to investigate the association of SNPs of the NHEJ pathway genes with susceptibility to HIV-1 infection and AIDS progression among MSM residing in northern China.

Results: In the present study, a total of 481 HIV-1 seropositive men and 493 HIV-1 seronegative men were included. And genotyping of 22 SNPs in NHEJ pathway genes was performed using the SNPscan^{TM} Kit. Our results disclosed significant associations of XRCC6 rs132770 and XRCC4 rs1056503 genotypes with susceptibility to HIV-1 infection. The generalized multifactor dimensionality reduction (GMDR) analysis found a significant SNP-SNP interaction between the XRCC6 and XRCC4 variants in the risk of HIV-1 infection. In stratified analysis, the positive effects of XRCC5 rs16855458 and LIG4 rs1805388 on the CD4+ T cell count and clinical phase of disease were validated.

Conclusions: Our results confirmed that the NHEJ gene polymorphisms played an important role in HIV-1 infection and AIDS progression in the northern Chinese MSM population.

Background

Acquired immune deficiency syndrome (AIDS) due to the infection of human immunodeficiency virus (HIV) is a chronic infectious disease and continues to be a major global public health issue. There were approximately 37.9 million people living with HIV at the end of 2018 and 1.7 million people becoming newly infected with HIV in 2018 globally (https://www.who.int/en/news-room/fact-sheets/detail/hiv-aids). In China, people living with HIV and AIDS patients were a total of 1,250,000 cases and 262,000 cases of death by the end of September in 2018 (http://www.chinaaids.cn/). The significant increase in the proportion of behavior spread of men who have sex with men (MSM) is the dominant pathway of all kinds of HIV infection routes. The individuals with different susceptibility to HIV infection and clinical disease progression arise from different genetic backgrounds of the host [1]. The finding of AIDS-related genes with single nucleotide polymorphisms (SNPs) is an important breakthrough that can help us to explore the role of host genetic background in HIV infection, reveal the pathogenesis of AIDS, predict the disease process and develop new drugs and vaccines [2].

Double-stranded DNA break (DSB) is one of the main reasons for the gene mutation and chromosome break, and plays an important role in tumorigenesis and progression of tumors [3]. Non-homologous end joining (NHEJ) pathway is the main approach of DSB repair (DSBR) in the higher eukaryotes, and can repair DSBs timely at any time in cell cycle [4, 5]. There are five core genes (*XRCC7, XRCC6, XRCC5, XRCC4* and *LIG4*) in the NHEJ pathway that encode five proteins (DNA-PK, Ku70, Ku80, XRCC4 and LIG4), respectively. Studies have shown that NHEJ gene polymorphisms are associated with susceptibility to a wide variety of cancers and disease progression. For instance, *XRCC7* gene polymorphisms play an important role in prostate cancer [6], bladder cancer [7], liver cancer [8], thyroid cancer [9] and lung cancer [10]. The other gene polymorphisms such as *XRCC4*, *XRCC5*, *XRCC6* and *LIG4* SNPs are also associated with many different types of cancers [11–15].

We believe that NHEJ pathway is associated with HIV-1 infection because the DSB in host genome DNA occurs in the process of HIV-1 integration based on previous functional studies of NHEJ genes. For example, the DNA-PK protein interacts with HIV-1 Tat to regulate HIV-1 replication and transcription [16, 17]. HIV-1 proviral DNA integration triggers cell death during HIV-1 infection because of the activation of DNA-PK, which causes phosphorylation of p53 and histone gamma-H2AX [18]. The Ku70 and Ku80 proteins are closely associated with HIV-1 integrase that is beneficial to virus integration and replication [19, 20] and protect cells against toxicity induced by HIV-1 integrase or integration [21]. However, the functional researches of other NHEJ genes and the association studies of NHEJ gene polymorphisms in the process of HIV-1 infection have not been reported. Up to now, the role of SNPs in NHEJ genes and their importance in HIV-1 infection and AIDS progression remain unclear. Therefore, to assess the involvement of NHEJ gene polymorphisms, we performed an association study of 22 SNPs in *XRCC7, XRCC6, XRCC5, XRCC4* and *LIG4* genes in 974 northern Chinese individuals. Participants were genotyped to investigate whether the polymorphisms in five genes were associated with the susceptibility to HIV-1 infection and the progression of AIDS.

Results

Hardy-Weinberg equilibrium test

In this study, 479 HIV-1-infected and 487 HIV-1-uninfected individuals from northern China were genotyped for 22 SNPs in NHEJ genes. The success rates were > 98 % for all SNPs and a mismatch rate of 0% detected in 50 replicate samples. As shown in Table 1, all 22 SNPs did not deviate from the Hardy-Weinberg equilibrium in the control group ($P > 0.05$).

Table 1
Hardy-Weinberg equilibrium test for the 22 NHEJ SNPs in controls

gene	Chr[a]	SNPs	Major/Minor allele	*P* for HWET[b]
XRCC7	8	rs7830743	A/G	0.248
	8	rs7003908	A/C	0.78
XRCC6	22	rs5751129	T/C	0.677
	22	rs2267437	C/G	0.178
	22	rs132770	G/A	0.468
	22	rs132774	G/C	0.568
XRCC5	2	rs828907	G/T	0.307
	2	rs705649	G/A	0.185
	2	rs16855458	C/A	0.762
	2	rs3770502	C/T	0.501
	2	rs9288516	T/A	0.504
	2	rs3835	G/A	0.529
	2	rs1051677	T/C	0.92
	2	rs2440	A/G	0.055
XRCC4	5	rs6869366	T/G	0.936
	5	rs2075685	G/T	0.476
	5	rs10040363	A/G	0.247
	5	rs963248	C/T	0.127
	5	rs35268	T/C	0.397
	5	rs1056503	G/T	0.051
LIG4	13	rs1805388	G/A	0.81
	13	rs1805389	G/A	0.994

[a] Chr: chromosome.

[b] Hardy-Weinberg equilibrium test.

Associations of NHEJ gene polymorphisms with HIV-1 infection

To explore the possible associations, the genotype distribution of the 22 SNPs was investigated and the differences of genotype frequencies between cases and controls were analyzed under three genetic models (codominant model, dominant model and recessive model). As shown in Fig. 1, a significant association was found for *XRCC6* rs132770 under the codominant ($P = 0.005$, OR = 10.51, 95% CI 2.000-55.251) and recessive ($P = 0.006$, OR = 10.45, 95% CI 1.986–54.933) inheritance models. Consistent with the aforementioned findings, for the *XRCC4* rs1056503, the TT genotype significantly increased the risk of HIV-1 infection compared

with GG genotype under the codominant model (P = 0.035, OR = 1.698, 95% CI 1.037–2.779), and the TT genotype increased the risk of HIV-1 infection compared with TG and GG genotypes under the recessive model (P = 0.028, OR = 1.707, 95% CI 1.060–2.750). However, no association with HIV-1 infection was observed in any genetic models for the rest 20 SNPs (P > 0.05).

Analysis of the SNP-SNP interaction

Then, the GMDR method was used to study the association of 10 SNPs within *XRCC6* and *XRCC4* gene with high-order interactions on HIV-1 infection. Through the 10-fold cross-validation, the best four-locus model involving *XRCC6* (rs2267437) and *XRCC4* (rs10040363, rs963248 and rs1056503) was found. The model had the testing balanced accuracy of 53.42%, the maximum cross-validation consistency of 10/10, and a sign test *P*-value 0.010 (Fig. 2). In order to obtain the ORs for the joint effects of the four SNPs on HIV-1 infection, traditional statistical methods were applied to this four-locus model to aid in interpretation, which identified three significant genotype combinations from all possible high-risk genotype combinations. In this four-locus (rs1056503-rs2267437-rs10040363-rs963248) model, the ORs for the three significant high-risk genotype combinations (TT)-(CC)-(AG/GG)-(TC/CC), (TT)-(CC)-(AA)-(TC/CC), and (TT)-(CC)-(AA)-(TT) were 6.667 (P = 0.035), 7.333 (P = 0.026) and 6.667 (P = 0.035), respectively (Table 2).

Table 2

Combined effects of rs1056503, rs2267437, rs10040363 and rs963248 on HIV-1 infection

rs1056503	rs2267437	rs10040363	rs963248	*P* value	OR (95% CI)
TT	CG + GG	AG + GG	TC + CC	-	1.000
TT	CC	AG + GG	TC + CC	0.035	6.667 (1.145–38.833)
TT	CG + GG	AA	TC + CC	0.848	1.200 (0.185–7.770)
TT	CG + GG	AG + GG	TT	1.000	1.000 (0.125–7.995)
TT	CC	AA	TC + CC	0.026	7.333 (1.272–42.294)
TT	CC	AG + GG	TT	0.756	0.667 (0.051–8.639)
TT	CG + GG	AA	TT	0.642	24.556 (1.991-302.866)
TT	CC	AA	TT	0.035	6.667 (1.145–38.833)

Bold italic indicates statistical significance.

Analysis of haplotype associations

Analysis of LD between the SNPs in NHEJ genes was performed with HaploView software. There was strong LD between the four SNPs in *XRCC6* gene, eight SNPs in *XRCC5* gene, six SNPs in *XRCC4* gene and two SNPs in *LIG4* gene. There were no significant differences in the frequencies of all haplotypes between HIV-1-infected individuals and healthy cohorts and no association with the susceptibility to HIV-1 infection (P > 0.05). Table 3 shows all blocks and haplotypes identified and the frequencies of these haplotypes.

Table 3
The frequency of haplotypes of the NHEJ genes in cases and controls

gene	Haplotype	Frequency	Haplotype frequencies in the cases	Haplotype frequencies in the controls	Chi Square	*P*
XRCC6	TCGG	0.671	0.665	0.677	0.309	0.578
	TGGG	0.237	0.233	0.241	0.175	0.675
	CCAC	0.070	0.081	0.060	3.181	**0.075**
	CCGC	0.021	0.021	0.022	0.012	0.912
XRCC5	Block 1					
	GG	0.771	0.777	0.765	0.335	0.563
	TA	0.223	0.219	0.227	0.169	0.681
	Block 2					
	TG	0.468	0.470	0.466	0.040	0.841
	AG	0.452	0.448	0.456	0.149	0.699
	TA	0.080	0.082	0.078	0.115	0.735
	Block 3					
	TA	0.693	0.705	0.683	1.101	0.294
	CG	0.165	0.164	0.166	0.009	0.923
	TG	0.142	0.132	0.152	1.649	0.199
XRCC4	Block 1					
	TG	0.806	0.806	0.805	0.007	0.931
	TT	0.141	0.135	0.147	0.597	0.440
	GT	0.053	0.059	0.048	1.095	0.295
	Block 2					
	CT	0.555	0.536	0.573	2.552	0.110
	TT	0.308	0.327	0.290	3.095	**0.079**
	CC	0.133	0.132	0.134	0.020	0.888
LIG4	GG	0.814	0.815	0.812	0.035	0.852
	AA	0.102	0.101	0.104	0.030	0.863
	AG	0.084	0.084	0.084	0.005	0.941

The values in bold indicate a trend towards association.

Association analysis for NHEJ gene SNPs with CD4 + T cell count and clinical phase in AIDS patients

In order to discover the relationship between the NHEJ gene polymorphisms and the progression of AIDS, differences in allele frequencies were analyzed between the case subgroups which were divided on CD4 + T-lymphocyte count and clinical stage as an index. The CD4 + T cell counts of the study participants ranged from 3 to 1038 cells/μl (mean ± SD, 335.57 ± 198.79). The associations between SNPs and CD4 + T cell counts were used to assess the influence of these polymorphisms on immunity status. As shown in Table 4, there were significant differences of genotype frequencies for *XRCC5* rs16855458 and *LIG4* rs1805388

between different case groups ($P < 0.05$). The subjects with AA or AC of rs16855458 have the significantly lower CD4 + T-lymphocyte count, compared to subjects with CC genotype ($P = 0.025$, OR = 1.538, 95% CI 1.054–2.243). In addition, the subjects with AA or AG of rs1805388 have the higher progression risk of AIDS, compared to subjects with GG genotype ($P = 0.036$, OR = 1.506, 95% CI 1.027–2.209). However, other SNPs were not associated with the $CD4^+$ T-lymphocyte count and clinical stages ($P > 0.05$). These results suggest that rs16855458 and rs1805388 were associated with the clinical features and that the polymorphisms in *XRCC5* and *LIG4* genes likely play an important role in the progression of AIDS in the northern Chinese population.

Table 4
Association between the 22 SNPs and the clinical features of AIDS

gene	gene polymorphisms	Genotype	$CD4^+$T-lymphocyte count [a]		*P*	OR (95%CI)	clinical stage[b]		*P*	OR (95%CI)
			< 350 cells/μl	> 350 cells/μl			phase III + IV	phase I + II		
XRCC7	rs7830743	GG + AG	37	35	0.550	0.858(0.519–1.418)	29	43	0.438	0.817(0.491–1.361)
		AA	223	181			184	223		
	rs7003908	CC + CA	125	92	0.232	1.248(0.868–1.795)	100	118	0.572	1.110 (0.773–1.594)
		AA	135	124			113	148		
XRCC6	rs5751129	CC + CT	49	39	0.825	1.054(0.662–1.679)	40	48	0.837	1.050(0.660–1.671)
		TT	211	177			173	218		
	rs2267437	GG + CG	118	82	0.103	1.358(0.940–1.961)	95	107	0.335	1.196(0.831–1.723)
		CC	142	134			118	159		
	rs132770	AA + AG	38	31	0.908	1.031(0.617–1.722)	27	43	0.282	0.752(0.448–1.264)
		GG	220	185			187	224		
	rs132774	CC + CG	50	39	0.743	1.081(0.679–1.719)	40	49	0.920	1.024(0.645–1.627)
		GG	210	177			173	217		
XRCC5	rs828907	TT + GT	104	86	0.967	1.008(0.697–1.457)	92	99	0.185	1.283(0.888–1.853)
		GG	156	130			121	167		
	rs705649	AA + GA	105	85	0.819	1.044(0.722–1.509)	92	99	0.185	1.283(0.888–1.853)
		GG	155	131			121	167		
	rs16855458	AA + AC	109	69	0.025	1.538(1.054–2.243)	87	92	0.160	1.306(0.900-1.895)
		CC	151	147			126	174		
	rs3770502	TT + CT	81	61	0.489	1.150(0.774–1.708)	66	76	0.499	1.146(0.772-1.700)
		CC	179	155			144	190		
	rs9288516	AA + TA	183	145	0.445	1.164(0.789–1.717)	144	185	0.649	0.914(0.620–1.347)
		TT	77	71			69	81		

Bold italic indicates statistical significance.

[a] The $CD4^+$ T-lymphocyte counts were divided into two groups: Category 1, < 350 cells/μl; Category 2, > 350 cells/μl.

[b] Clinical stage: Category A, Clinical phase III + IV; Category B, Clinical phase I + II.

gene	gene polymorphisms	Genotype	$CD4^+$T-lymphocyte count [a]		*P*	OR (95%CI)	clinical stage[b]		*P*	OR (95%CI)
			< 350 cells/μl	> 350 cells/μl			phase III + IV	phase I + II		
	rs3835	AA + AG	36	35	0.495	0.839(0.506–1.390)	33	39	0.804	1.066(0.645–1.763)
		GG	222	181			181	228		
	rs1051677	CC + TC	83	61	0.384	1.192(0.803–1.768)	68	77	0.481	1.151(0.778–1.703)
		TT	177	155			145	189		
	rs2440	GG + AG	127	113	0.451	0.870(0.607–1.249)	96	136	0.767	0.947(0.660–1.358)
		AA	133	103			107	130		
XRCC4	rs6869366	GG + GT	36	20	0.124	1.575(0.883–2.811)	23	33	0.587	0.855(0.485–1.505)
		TT	224	196			190	233		
	rs2075685	TT + GT	77	80	0.087	0.715(0.487–1.050)	62	95	0.126	0.739(0.502–1.089)
		GG	183	136			151	171		
	rs10040363	GG + AG	72	66	0.510	0.875(0.588–1.302)	59	79	0.613	0.902(0.605–1.345)
		AA	187	150			154	186		
	rs963248	TT + TC	136	132	0.054	0.698(0.484–1.007)	114	155	0.298	0.825(0.574–1.186)
		CC	124	84			99	111		
	rs35268	CC + TC	70	51	0.409	1.192(0.786–1.808)	60	61	0.191	1.318(0.872–1.992)
		TT	190	165			153	205		
	rs1056503	TT + TG	128	118	0.241	0.805(0.561–1.156)	107	140	0.602	0.909(0.633–1.303)
		GG	132	98			106	126		
LIG4	rs1805388	AA + AG	87	69	0.726	1.071(0.729–1.574)	81	77	0.036	1.506(1.027–2.209)
		GG	173	147			132	189		
	rs1805389	AA + AG	52	39	0.591	1.135(0.716–1.799)	48	45	0.124	1.429(0.907–2.249)
		GG	208	177			165	221		

Bold italic indicates statistical significance.

[a] The $CD4^+$ T-lymphocyte counts were divided into two groups: Category 1, < 350 cells/μl; Category 2, > 350 cells/μl.

[b] Clinical stage: Category A, Clinical phase III + IV; Category B, Clinical phase I + II.

Discussion

According to the molecular mechanism of HIV-1 infection, viral DNA is inserted into the host genomic DNA in the process of HIV-1 integration. The integration process was equivalent to genomic DNA with DSBs in host cells under the action of HIV-1 and then the signal of damage repair would start NHEJ pathway. Thus, we believed that the NHEJ genes were involved in HIV-1 infection and the disease progression. To the best of our knowledge, this comprehensive study is the first to systematically evaluate the association between the polymorphisms in NHEJ genes and the susceptibility to HIV-1 infection and the progression of AIDS.

In our study, the differences of genotype frequencies of *XRCC6* rs132770 and *XRCC4* rs1056503 were seen between the cases and the controls under different genetic models. Our results implied a positive association of the polymorphisms in NEHJ genes with the susceptibility to HIV-1 infection in the northern Chinese MSM population. The *XRCC6* gene codes Ku70 protein, which functions as a single-stranded DNA- and ATP-dependent helicase and may be involved in the repair of non-homologous DNA ends such as that required for DSB repair. The Ku 70 protein also interacts with HIV-1 integrase in the process of the HIV-1 infection [19–21]. The rs1056503 is a synonymous codon in *XRCC6* gene. The association may be due to the fact that the rs1056503 affects the mRNA expression by alternative splicing, and regulates the XRCC6 protein function; or this SNP is closely linked to another SNP which is associated to HIV-1 infection. Similar to our findings, it was also reported that different *XRCC6* genotypes could contribute to the susceptibility of another infectious disease, namely hepatocellular carcinoma [22, 23].

The *XRCC4* gene codes XRCC4 protein which can activate and enhance the activity of LIG4 protein, and plays an important role in NEHJ repair pathway [24]. The *XRCC4* gene mutations can lead to the occurrence of small head dwarfism [25]. Although the SNPs in *XRCC4* gene can influence the susceptibility and progression of infectious disease such as liver cancer [26, 27], their effects on HIV-1 infection have not been reported yet. Our study suggested that *XRCC4* gene polymorphisms were associated with HIV-1 infection, which was in accordance with the above reports. Thus, the association in this study could be explained as following. The rs132770 is located in *XRCC4* gene 5′ regulatory region, which can cause the changes in mRNA expression levels and Ku70 protein function. The functional changes in Ku70 protein will affect the expression level of HIV-1 integrase or the process of HIV-1 infected into the host cell. Moreover, in analysis of SNP-SNP interaction, our results provide evidence for a four-locus interaction between the *XRCC6* and *XRCC4* variants in the risk of HIV-1 infection, and highlight further the importance of multilocus effects in the genetic component of HIV-1 infection.

As an indicator of AIDS clinical characteristics, CD4 + T cell count reflects the count of the patient's body immune cells. The AIDS patients with CD4 + T cell count less than 350 cells/μl should be given antiretroviral therapy or other treatments according to the World Health Organization (WHO) [28–30]. In our study, we found that there was a significant difference of rs16855458 genotype frequency in *XRCC5* gene between two case subgroups, and genotypes of AA and AC were associated with a small number of CD4 + T cells. This result showed that *XRCC5* rs16855458 was associated with progression of AIDS. The *XRCC5* gene codes Ku80 protein which forms Ku heterodimer with Ku70 protein. Functional studies showed that changes in expression levels of Ku80 protein are the main reason of tumor development and can be used as a predictor of patient survival as well as treatment outcome [31, 32]. In the process of HIV-1 infection, the *XRCC5* gene is closely related to HIV-1 integration and translation [33–35]. We propose that the rs16855458 in *XRCC5* gene intron may regulate the transcription and the expression of the *XRCC5* gene by alternative splicing, which interacts with HIV-1 to promote its integration and translation, leading to the decrease in the $CD4^+$ T-lymphocyte count and the AIDS acceleration. Similar to our findings, the polymorphisms of *XRCC5* gene were also reported to be associated with viral disease such as liver cancer [23].

In this study, we divided the cases into two subgroups by clinical stage as an index, which is a clinical feature of AIDS and directly reflects the disease progression. The clinical symptoms of patients with phase I and II are mild and just show HIV-1 antibody positive. On the contrary, patients with phase III and IV have serious clinical symptoms such as nervous system lesions, continuous fever and diarrhea, sepsis and various kinds of tumors caused by the loss of immune functions, and should be timely given the anti-retroviral therapy or other treatments. The result of our studies revealed that there was a significant difference of genotype frequency of *LIG4* rs1805388 between MSM cases with clinical phase I + II and those with clinical phase III + IV, and that genotype AA/AG could significantly promote the disease progression of AIDS. The *LIG4* gene codes LIG4 protein which connects the DSB end and then completes the NHEJ repair. Previous study showed that *LIG4* gene polymorphisms were associated with clinical features of cancer such as treatment outcome, progression-free survival, and overall survival [36]. The LIG4 gene mutation can not only lead to abnormal development of the immune defects but also cause severe combined immunodeficiency disease of normal individuals

[37]. The rs1805388 is located in the exon region of *LIG4* gene, which is a missense mutation of threonine and isoleucine. Here we propose that the reason for association was the functional changes in LIG4 protein which directly affected the AIDS clinical stage.

Conclusions

In conclusion, our results confirmed that the NHEJ gene polymorphisms play an important role in HIV-1 infection and AIDS progression among MSM in northern China. Up to now, the mechanism underlying the interaction between the NHEJ genes and HIV-1/AIDS remains unclear, and our study opens a new field for further studying on the functional significance and the underlying mechanism of the association between the NHEJ gene polymorphisms and HIV-1/AIDS. However, the results and conclusions of this study needs to be tested and verified with more association studies and subsequent function researches in different races.

Methods

Subjects

In the present study, a total of 481 HIV-1 seropositive men were recruited from Heilongjiang Center for Disease Control and Prevention (CDC). The age of the HIV-1 infected individuals ranged from 16 to 75 years old (mean age ± SD, 35.3 ± 11.55) and the average CD4 + T-lymphocyte count at that time point was 335 cells/μl (range, 3-1038 cells/μl). All patients had acquired HIV-1 infection through male-male homosexual transmission. We categorized these patients as Category 1 (T-lymphocytes < 350 cells/μl) or Category 2 (T-lymphocytes > 350 cells/μl) by the CD4 + T-lymphocyte count, and as Category A (Clinical phase III + IV) or Category B (Clinical phase I + II) by the clinical stage.

Correspondingly, 493 HIV-1 seronegative men age-matched to the HIV-1 patients were randomly selected as a control group from the Second Affiliated Hospital of Harbin Medical University, Harbin, China. The age of the uninfected controls ranged from 16 to 75 years (mean age ± SD, 35.3 ± 11.59). All participants provided informed consent approved by local ethics review board.

SNPs selection and genotyping

Based on the published literature, 22 candidate SNPs in NHEJ pathway genes were included in the present study. Of these, two SNPs (rs7830743 and rs7003908) were from *XRCC7*, four SNPs (rs132770, rs5751129, rs2267437 and rs132774) were from *XRCC6*, eight SNPs (rs828907, rs705649, rs16855458, rs3770502, rs9288516, rs3835, rs1051677 and rs2440) were from *XRCC5*, six SNPs (rs1056503, rs6869366, rs2075685, rs10040363, rs963248 and rs35268) were from *XRCC4* and two SNPs (rs1805388 and rs1805389) were from *LIG4*.

Genomic DNA was extracted from 200 μl of peripheral blood of all participants using the QIAamp blood kit (Qiagen, Germany) according to the manufacturer's protocol. All 22 SNPs were genotyped using a custom-designed 48-Plex SNPscan™ Kit (Genesky Bio-technologies Inc., Shanghai, China), based on a method of high-throughput SNP genotyping utilizing double ligation and multiplex fluorescence PCR. For quality control, a 5% random sample of cases and controls was genotyped twice to verify the genotyping accuracy, the reproducibility was 100%.

Statistical analysis

The genotype and allele frequencies were calculated through directly counting the numbers after the genotypes of the cases and controls were determined. Chi-square test was used for examining the deviation from Hardy-Weinberg equilibrium (HWE) for all SNPs of the control group, the association between the genotype frequency and susceptibility to HIV-1 infection and the clinical features of cases (such as the CD4 + T-lymphocyte count and clinical stage). Odds ratios (ORs) and 95% confidence intervals (95% CI) were estimated as the relative risk associated with SNPs. The generalized multifactor dimensionality reduction (GMDR) software (http://www.ssg.uab.edu/gmdr/) was applied to assess SNP-SNP interactions. SPSS 23.0 (IBM-SPSS, Inc., Chicago, IL, USA) was used for all statistical analyses. The analyses of linkage disequilibrium (LD) and the haplotypes frequencies were performed using the HaploView software (ver. 4.2, http://sourceforge.net/projects/haploview/). The differences with a *P* value less than 0.05 were considered statistically significant.

Abbreviations

MSM: men who have sex with men; NHEJ: non-homologous end joining; DSB: double-stranded DNA break; DSBR: DSB repair; SNP: single nucleotide polymorphism; AIDS: Acquired immune deficiency syndrome; HIV: human immunodeficiency virus; GMDR: generalized multifactor dimensionality reduction; WHO: World Health Organization; HWE: Hardy-Weinberg equilibrium; ORs: Odds ratios; 95% CI: 95% confidence intervals; LD: linkage disequilibrium.

Declarations

Ethics approval and consent to participate

The study protocol was approved by the local ethics review board (No.: HMUIRB20180019) and all experimental procedures complied with the Declaration of Helsinki. All participates gave written informed consent to take part in the present study.

Consent for publication

Not applicable.

Availability of data and materials

The data-sets used and/or analysed during the current study available from the corresponding author on reasonable request.

Funding

This work was funded by the National Natural Science Foundation of China (81373220) and the post-doctoral foundation of Heilongjiang Province (LRB 08-340 and LBH-Q11029) of China.

Authors' contributions

Xuelong Zhang, Yuandong Qiao and Songbin Fu conceived and designed the study; Han Mo, Chuntao Wang, Jiawei Wu, Bangquan Liu, Haiming Sun, Ping Wang and Kaili Wang performed the experiments; Xuelong Zhang, Han Mo, Chuntao Wang, Lidan Xu and Xueyuan Jia analyzed the data; and Xuelong Zhang, Wenjing Sun, Kaili Wang and Han Mo prepared the manuscript. All authors revised and approved the final draft.

Acknowledgements

We gratefully acknowledge the numerous sample donors for making this work possible.

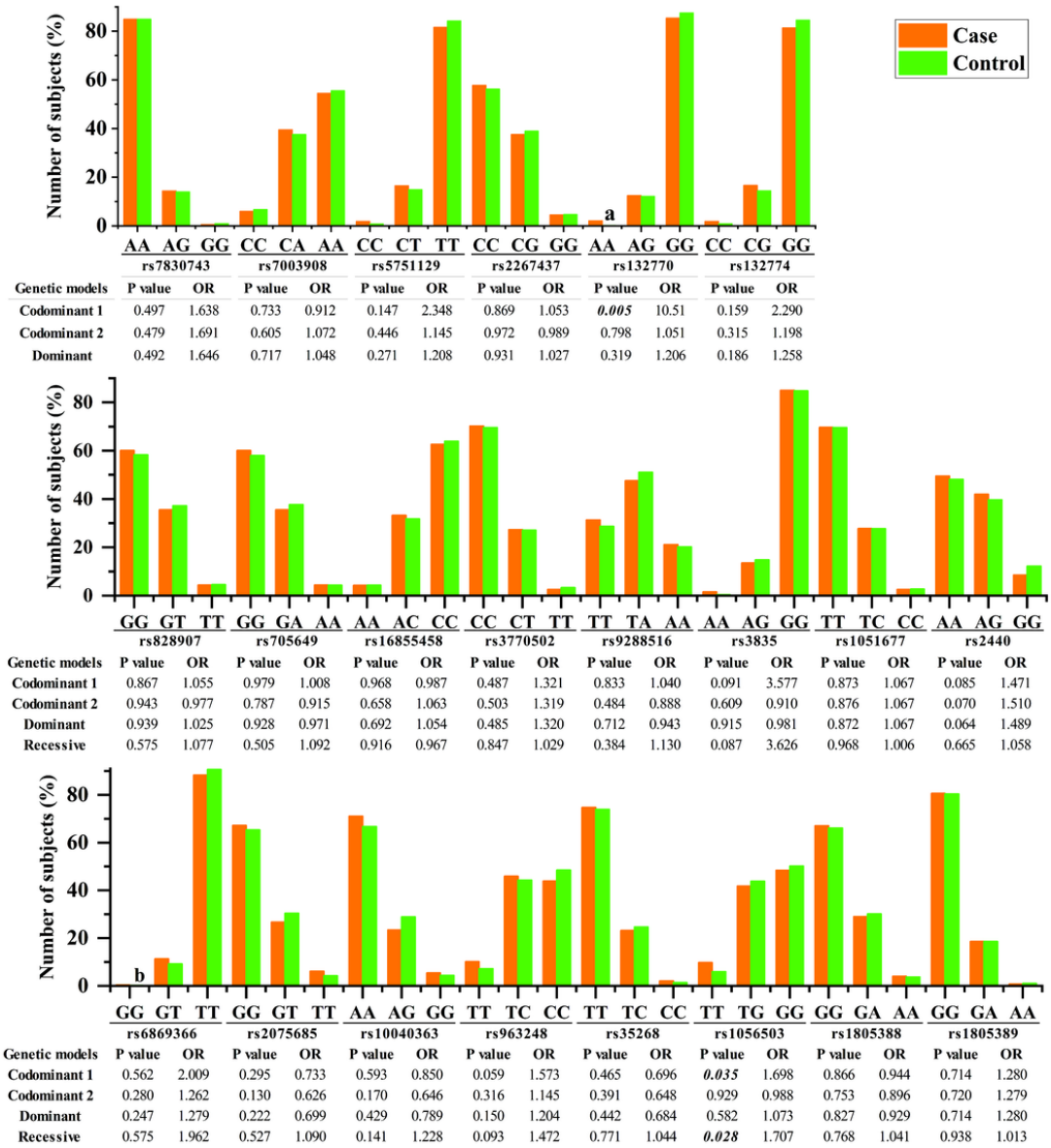
References

1. McLaren PJ, Carrington M: The impact of host genetic variation on infection with HIV-1. Nature immunology 2015, 16(6):577-583.
2. van Manen D, van 't Wout AB, Schuitemaker H: Genome-wide association studies on HIV susceptibility, pathogenesis and pharmacogenomics. Retrovirology 2012, 9:70.
3. Sishc BJ, Davis AJ: The Role of the Core Non-Homologous End Joining Factors in Carcinogenesis and Cancer. Cancers 2017, 9(7).
4. Chang HHY, Pannunzio NR, Adachi N, Lieber MR: Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nature reviews 2017, 18(8):495-506.
5. Zhao X, Wei C, Li J, Xing P, Li J, Zheng S, Chen X: Cell cycle-dependent control of homologous recombination. Acta biochimica et biophysica Sinica 2017, 49(8):655-668.
6. Mandal RK, Kapoor R, Mittal RD: Polymorphic variants of DNA repair gene XRCC3 and XRCC7 and risk of prostate cancer: a study from North Indian population. DNA and cell biology 2010, 29(11):669-674.
7. Zhi Y, Yu J, Liu Y, Wei Q, Yuan F, Zhou X, Song B, Chen Z, Yang J: Interaction between polymorphisms of DNA repair genes significantly modulated bladder cancer risk. International journal of medical sciences 2012, 9(6):498-505.
8. Hsieh YH, Chang WS, Tsai CW, Tsai JP, Hsu CM, Jeng LB, Bau DT: DNA double-strand break repair gene XRCC7 genotypes were associated with hepatocellular carcinoma risk in Taiwanese males and alcohol drinkers. Tumour Biol 2015, 36(6):4101-4106.

9. Rahimi M, Fayaz S, Fard-Esfahani A, Modarressi MH, Akrami SM, Fard-Esfahani P: The role of Ile3434Thr XRCC7 gene polymorphism in differentiated thyroid cancer risk in an Iranian population. Iranian biomedical journal 2012, 16(4):218-222.
10. Hsia TC, Chang WS, Chen WC, Liang SJ, Tu CY, Chen HJ, Liang JA, Tsai CW, Hsu CM, Tsai CH et al: Genotype of DNA double-strand break repair gene XRCC7 is associated with lung cancer risk in Taiwan males and smokers. Anticancer research 2014, 34(12):7001-7005.
11. Corral R, Lewinger JP, Van Den Berg D, Joshi AD, Yuan JM, Gago-Dominguez M, Cortessis VK, Pike MC, Conti DV, Thomas DC et al: Comprehensive analyses of DNA repair pathways, smoking and bladder cancer risk in Los Angeles and Shanghai. International journal of cancer 2014, 135(2):335-347.
12. Hasan SK, Buttari F, Ottone T, Voso MT, Hohaus S, Marasco E, Mantovani V, Garagnani P, Sanz MA, Cicconi L et al: Risk of acute promyelocytic leukemia in multiple sclerosis: coding variants of DNA repair genes. Neurology 2011, 76(12):1059-1065.
13. Makkoch J, Praianantathavorn K, Sopipong W, Chuaypen N, Tangkijvanich P, Payungporn S: Genetic Variations in XRCC4 (rs1805377) and ATF6 (rs2070150) are not Associated with Hepatocellular Carcinoma in Thai Patients with Hepatitis B Virus Infection. Asian Pac J Cancer Prev 2016, 17(2):591-595.
14. Mumbrekar KD, Goutham HV, Vadhiraja BM, Bola Sadashiva SR: Polymorphisms in double strand break repair related genes influence radiosensitivity phenotype in lymphocytes from healthy individuals. DNA repair 2016, 40:27-34.
15. Yin M, Liao Z, Liu Z, Wang LE, O'Reilly M, Gomez D, Li M, Komaki R, Wei Q: Genetic variants of the nonhomologous end joining gene LIG4 and severe radiation pneumonitis in nonsmall cell lung cancer patients treated with definitive radiotherapy. Cancer 2012, 118(2):528-535.
16. Tyagi S, Ochem A, Tyagi M: DNA-dependent protein kinase interacts functionally with the RNA polymerase II complex recruited at the human immunodeficiency virus (HIV) long terminal repeat and plays an important role in HIV gene expression. The Journal of general virology 2011, 92(Pt 7):1710-1720.
17. Zhang SM, Zhang H, Yang TY, Ying TY, Yang PX, Liu XD, Tang SJ, Zhou PK: Interaction between HIV-1 Tat and DNA-PKcs modulates HIV transcription and class switch recombination. International journal of biological sciences 2014, 10(10):1138-1149.
18. Cooper A, Garcia M, Petrovas C, Yamamoto T, Koup RA, Nabel GJ: HIV-1 causes CD4 cell death through DNA-dependent protein kinase during viral integration. Nature 2013, 498(7454):376-379.
19. Zheng Y, Ao Z, Wang B, Jayappa KD, Yao X: Host protein Ku70 binds and protects HIV-1 integrase from proteasomal degradation and is required for HIV replication. The Journal of biological chemistry 2011, 286(20):17722-17735.
20. Zheng Y, Yao X: Posttranslational modifications of HIV-1 integrase by various cellular proteins during viral replication. Viruses 2013, 5(7):1787-1801.
21. Warrilow D, Tachedjian G, Harrich D: Maturation of the HIV reverse transcription complex: putting the jigsaw together. Reviews in medical virology 2009, 19(6):324-337.
22. Hsu CM, Yang MD, Chang WS, Jeng LB, Lee MH, Lu MC, Chang SC, Tsai CW, Tsai Y, Tsai FJ et al: The contribution of XRCC6/Ku70 to hepatocellular carcinoma in Taiwan. Anticancer research 2013, 33(2):529-535.
23. Li R, Yang Y, An Y, Zhou Y, Liu Y, Yu Q, Lu D, Wang H, Jin L, Zhou W et al: Genetic polymorphisms in DNA double-strand break repair genes XRCC5, XRCC6 and susceptibility to hepatocellular carcinoma. Carcinogenesis 2011, 32(4):530-536.
24. Foster RE, Nnakwe C, Woo L, Frank KM: Monoubiquitination of the nonhomologous end joining protein XRCC4. Biochemical and biophysical research communications 2006, 341(1):175-183.
25. Murray JE, van der Burg M, H IJ, Carroll P, Wu Q, Ochi T, Leitch A, Miller ES, Kysela B, Jawad A et al: Mutations in the NHEJ component XRCC4 cause primordial dwarfism. American journal of human genetics 2015, 96(3):412-424.
26. Jung SW, Park NH, Shin JW, Park BR, Kim CJ, Lee JE, Shin ES, Kim JA, Chung YH: Polymorphisms of DNA repair genes in Korean hepatocellular carcinoma patients with chronic hepatitis B: possible implications on survival. Journal of hepatology 2012, 57(3):621-627.
27. Long XD, Zhao D, Wang C, Huang XY, Yao JG, Ma Y, Wei ZH, Liu M, Zeng LX, Mo XQ et al: Genetic polymorphisms in DNA repair genes XRCC4 and XRCC5 and aflatoxin B1-related hepatocellular carcinoma. Epidemiology (Cambridge, Mass 2015, 24(5):671-681.

28. Palella FJ, Jr., Deloria-Knoll M, Chmiel JS, Moorman AC, Wood KC, Greenberg AE, Holmberg SD: Survival benefit of initiating antiretroviral therapy in HIV-infected persons in different CD4+ cell strata. Annals of internal medicine 2003, 138(8):620-626.
29. Siegfried N, Uthman OA, Rutherford GW: Optimal time for initiation of antiretroviral therapy in asymptomatic, HIV-infected, treatment-naive adults. The Cochrane database of systematic reviews 2010(3):CD008272.
30. Sterne JA, May M, Costagliola D, de Wolf F, Phillips AN, Harris R, Funk MJ, Geskus RB, Gill J, Dabis F et al: Timing of initiation of antiretroviral therapy in AIDS-free HIV-1-infected patients: a collaborative analysis of 18 HIV cohort studies. Lancet (London, England) 2009, 373(9672):1352-1363.
31. Li W, Xie C, Yang Z, Chen J, Lu NH: Abnormal DNA-PKcs and Ku 70/80 expression may promote malignant pathological processes in gastric carcinoma. World journal of gastroenterology 2013, 19(40):6894-6901.
32. Tseng SH, Yang CC, Yu EH, Chang C, Lee YS, Liu CJ, Chang KW, Lin SC: K14-EGFP-miR-31 transgenic mice have high susceptibility to chemical-induced squamous cell tumorigenesis that is associating with Ku80 repression. International journal of cancer 2015, 136(6):1263-1275.
33. Jeanson L, Mouscadet JF: Ku represses the HIV-1 transcription: identification of a putative Ku binding site homologous to the mouse mammary tumor virus NRE1 sequence in the HIV-1 long terminal repeat. The Journal of biological chemistry 2002, 277(7):4918-4924.
34. Jeanson L, Subra F, Vaganay S, Hervy M, Marangoni E, Bourhis J, Mouscadet JF: Effect of Ku80 depletion on the preintegrative steps of HIV-1 replication in human cells. Virology 2002, 300(1):100-108.
35. Manic G, Maurin-Marlin A, Laurent F, Vitale I, Thierry S, Delelis O, Dessen P, Vincendeau M, Leib-Mosch C, Hazan U et al: Impact of the Ku complex on HIV-1 expression and latency. PloS one 2013, 8(7):e69691.
36. Jiang YH, Xu XL, Ruan HH, Xu WZ, Li D, Feng JG, Han QB, Mao WM: The impact of functional LIG4 polymorphism on platinum-based chemotherapy response and survival in non-small cell lung cancer. Medical oncology (Northwood, London, England) 2014, 31(5):959.
37. van der Burg M, van Veelen LR, Verkaik NS, Wiegant WW, Hartwig NG, Barendregt BH, Brugmans L, Raams A, Jaspers NG, Zdzienicka MZ et al: A new type of radiosensitive T-B-NK+ severe combined immunodeficiency caused by a LIG4 mutation. The Journal of clinical investigation 2006, 116(1):137-145.

Figures



Genetic models	rs7830743		rs7003908		rs5751129		rs2267437		rs132770		rs132774	
	P value	OR	P value	OR	P value	OR	P value	OR	P value	OR	P value	OR
Codominant 1	0.497	1.638	0.733	0.912	0.147	2.348	0.869	1.053	***0.005***	10.51	0.159	2.290
Codominant 2	0.479	1.691	0.605	1.072	0.446	1.145	0.972	0.989	0.798	1.051	0.315	1.198
Dominant	0.492	1.646	0.717	1.048	0.271	1.208	0.931	1.027	0.319	1.206	0.186	1.258

Genetic models	rs828907		rs705649		rs16855458		rs3770502		rs9288516		rs3835		rs1051677		rs2440	
	P value	OR	P value	OR	P value	OR	P value	OR	P value	OR	P value	OR	P value	OR	P value	OR
Codominant 1	0.867	1.055	0.979	1.008	0.968	0.987	0.487	1.321	0.833	1.040	0.091	3.577	0.873	1.067	0.085	1.471
Codominant 2	0.943	0.977	0.787	0.915	0.658	1.063	0.503	1.319	0.484	0.888	0.609	0.910	0.876	1.067	0.070	1.510
Dominant	0.939	1.025	0.928	0.971	0.692	1.054	0.485	1.320	0.712	0.943	0.915	0.981	0.872	1.067	0.064	1.489
Recessive	0.575	1.077	0.505	1.092	0.916	0.967	0.847	1.029	0.384	1.130	0.087	3.626	0.968	1.006	0.665	1.058

Genetic models	rs6869366		rs2075685		rs10040363		rs963248		rs35268		rs1056503		rs1805388		rs1805389	
	P value	OR	P value	OR	P value	OR	P value	OR	P value	OR	P value	OR	P value	OR	P value	OR
Codominant 1	0.562	2.009	0.295	0.733	0.593	0.850	0.059	1.573	0.465	0.696	***0.035***	1.698	0.866	0.944	0.714	1.280
Codominant 2	0.280	1.262	0.130	0.626	0.170	0.646	0.316	1.145	0.391	0.648	0.929	0.988	0.753	0.896	0.720	1.279
Dominant	0.247	1.279	0.222	0.699	0.429	0.789	0.150	1.204	0.442	0.684	0.582	1.073	0.827	0.929	0.714	1.280
Recessive	0.575	1.962	0.527	1.090	0.141	1.228	0.093	1.472	0.771	1.044	***0.028***	1.707	0.768	1.041	0.938	1.013

Figure 1

Genetic association of NHEJ polymorphisms between cases and controls. The bar marked by the letters a and b corresponds to the ordinate of the minimum value of 0.2%. Codominant 1, the first column homozygote versus the third column homozygote; Codominant 2, heterozygote versus the third column homozygote. Bold italic indicates statistical significance.

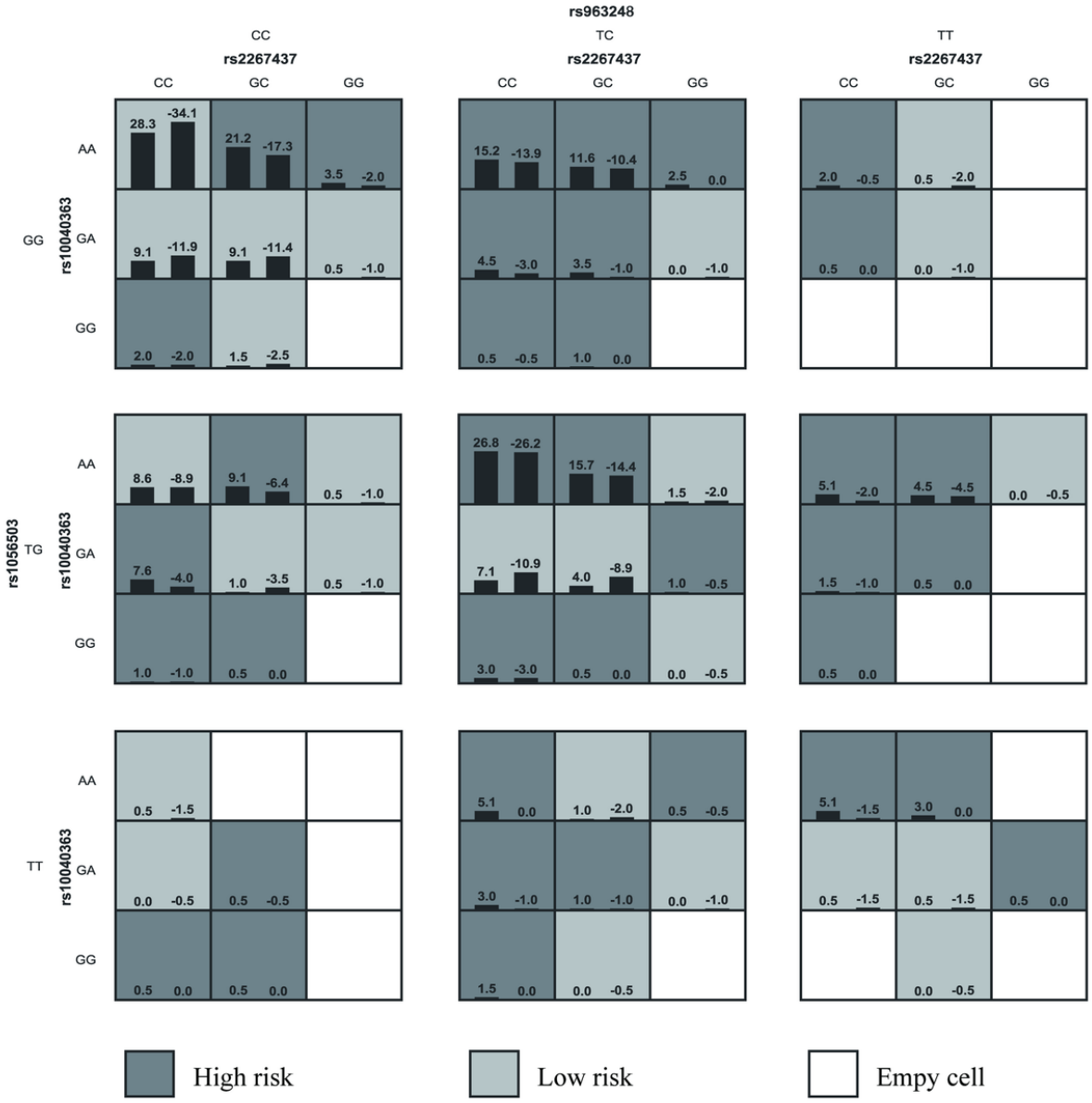


Figure 2

Best four-locus SNP-SNP interaction model identified by the generalized multifactor dimensionality reduction method. High-risk cells are in dark, low-risk cells are in grey, and empty cells are indicated by no shading. In each cell, the left bar represents case while the right bar represents control. The heights of the bars are proportional to the sum of samples in each group. Note that the patterns of high-risk and low-risk cells differ across each of the different multilocus dimensions, presenting evidence of SNP-SNP interaction or epistasis.