

A genetic study and meta-analysis of the genetic predisposition of prostate cancer in a Chinese population

Jacek Marzec^{1,*}, Xueying Mao^{1,*}, Meiling Li^{2,*}, Meilin Wang^{3,*}, Ninghan Feng^{4,5,*}, Xin Gou^{6,*}, Guomin Wang^{7,*}, Zan Sun^{8,*}, Jianfeng Xu^{9,10,*}, Hua Xu^{11,*}, Xiaoping Zhang^{12,*}, Shan-Chao Zhao¹³, Guoping Ren¹⁴, Yongwei Yu¹⁵, Yudong Wu¹⁶, Ji Wu¹⁷, Yao Xue³, Bo Zhou¹⁸, Yanling Zhang¹⁴, Xingxing Xu², Jie Li⁶, Weiyang He⁶, Sara Benlloch¹⁹, Helen Ross-Adams²⁰, Li Chen¹², Jucong Li¹³, Yingqia Hong¹³, Zsofia Kote-Jarai²¹, Xingang Cui²², Jianguo Hou²³, Jianming Guo⁷, Lei Xu⁷, Changjun Yin⁵, Yuanping Zhou²⁴, David E. Neal²⁰, Tim Oliver¹, Guangwen Cao², Zhengdong Zhang³, Douglas F. Easton¹⁹, Claude Chelala¹, The PRACTICAL Consortium, The CHIPGECS Group, Ali Amin Al Olama¹⁹, Rosalind A. Eeles^{21,25}, Hongwei Zhang², Yong-Jie Lu¹

¹Centre for Molecular Oncology, Barts Cancer Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, EC1M 6BQ, UK

²Department of Epidemiology, Second Military Medical University, Shanghai, 200433, China

³Department of Molecular and Genetic Toxicology, The Key Laboratory of Modern Toxicology, School of Public Health, Nanjing Medical University, Nanjing, 210029, China

⁴Department of Urology, Wuxi Second People's Hospital, Nanjing Medical University, Wuxi, 214002, China

⁵Department of Urology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, 210029, China

⁶Department of Urology, The First Affiliated Hospital, Chongqing Medical University, Chongqing, 400016, China

⁷Department of Urology, Zhongshan Hospital, Fudan University Medical College, Shanghai, 200032, China

⁸Liaoning People's Hospital and Center of Experiment and Technology, China Medical University, Shenyang, 110001, China

⁹Program for Personalized Cancer Care, North Shore University Health System, Evanston, IL 60201, U.S.A

¹⁰Fudan Institute of Urology, Huashang Hospital, Fudan University, Shanghai, 200040, China

¹¹Department of Urology, Tongji Hospital, Huazhong Science and Technology University, Wuhan, 430030, China

¹²Department of Urology, Xiehe Hospital, Huazhong Science and Technology University, Wuhan, 430022, China

¹³Department of Urology, Nanfang Hospital, Southern Medical University, Guangzhou, 510515, China

¹⁴Department of Pathology, The First Affiliated Hospital, Zhejiang University Medical College, Hangzhou, 310009, China

¹⁵Department of Pathology, Changhai Hospital, The Second Military Medical University, Shanghai, 200433, China

¹⁶Department of Urology, First Affiliated Hospital, Medical College, Zhengzhou University, Zhengzhou, 450003, China

¹⁷Department of Urology, North Sichuan Medical College, Nanchong, 637000, China

¹⁸Department of Nutrition Science, Shenyang Medical College, Shenyang, 110034, China

¹⁹Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge-Strangeways Research Laboratory, Cambridge, CB1 8RN, UK

²⁰Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge, CB2 0RE, UK

²¹Division of Genetics and Epidemiology, The Institute of Cancer Research, London, SM2 5NG, UK

²²Department of Urology, Changzheng Hospital, The Second Military Medical University, Shanghai, 200003, China

²³Department of Urology, Changhai Hospital, The Second Military Medical University, Shanghai, 200433, China

²⁴Department of Infectious Diseases, Nanfang Hospital, Southern Medical University, Guangzhou, 510515, China

²⁵The Royal Marsden NHS Foundation Trust, London and Surrey, SM2 5NG, UK

*These authors have contributed equally to this work

Correspondence to: Yong-Jie Lu, e-mail: y.j.lu@qmul.ac.uk

Hongwei Zhang, e-mail: hwzhang@smmu.edu.cn

Keywords: Chinese prostate cancer, genetic risk, single-nucleotide polymorphism, predisposition, population difference

Received: September 02, 2015

Accepted: January 23, 2016

Published: February 8, 2016

ABSTRACT

Prostate cancer predisposition has been extensively investigated in European populations, but there have been few studies of other ethnic groups. To investigate prostate cancer susceptibility in the under-investigated Chinese population, we performed single-nucleotide polymorphism (SNP) array analysis on a cohort of Chinese cases and controls and then meta-analysis with data from the existing Chinese prostate cancer genome-wide association study (GWAS). Genotyping 211,155 SNPs in 495 cases and 640 controls of Chinese ancestry identified several new suggestive Chinese prostate cancer predisposition loci. However, none of them reached genome-wide significance level either by meta-analysis or replication study. The meta-analysis with the Chinese GWAS data revealed that four 8q24 loci are the main contributors to Chinese prostate cancer risk and the risk alleles from three of them exist at much higher frequencies in Chinese than European populations. We also found that several predisposition loci reported in Western populations have different effect on Chinese men. Therefore, this first extensive single-nucleotide polymorphism study of Chinese prostate cancer in comparison with European population indicates that four loci on 8q24 contribute to a great risk of prostate cancer in a considerable large proportion of Chinese men. Based on those four loci, the top 10% of the population have six- or two-fold prostate cancer risk compared with men of the bottom 10% or median risk respectively, which may facilitate the design of prostate cancer genetic risk screening and prevention in Chinese men. These findings also provide additional insights into the etiology and pathogenesis of prostate cancer.

INTRODUCTION

The incidence and mortality rates of prostate cancer vary dramatically among countries with 24-fold and 10-fold difference, respectively, in 2008 [1]. While it is the most common male cancer in most of America, Oceania, western and northern European countries, its incidence in China is much lower. Despite a rapid increase of prostate cancer incidence in China in recent years, the variation between China and Western countries is still large. Both genetic and environmental factors may contribute to this finding, potentially through the induction of different genetic changes [2].

Previous prostate cancer predisposition studies, particularly genome-wide association studies (GWAS), have been mainly performed in populations from European ancestors [3–20]. While certain loci may contribute to prostate cancer in multiple ethnic populations, population differences have been previously demonstrated [3, 16, 18, 21]. Only one GWAS has been performed in the Chinese population [18]. In association with the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) consortium for prostate cancer predisposition study [6, 8, 15], we investigated genetic predisposition to prostate cancer in men of Han Chinese ancestry using the custom Illumina single-nucleotide polymorphism (SNP) array (iCOGS) [8]. Meta-analysis with the Chinese GWAS data [18] revealed few suggestive loci. Importantly, we found that four loci

on 8q24 contribute in a large proportion of Chinese men to a considerable risk of prostate cancer, which are different in the European population.

RESULTS

Genotyping and analysis

We genotyped 495 prostate cancer cases and 640 age matched non-cancer male controls (Supplementary Table S1) using the iCOGS SNP array [8]. No population outliers were identified in population structure analysis (Supplementary Figure S1) and 170,264 SNPs in 455 cases and 614 controls were retained for association analysis after systematic quality control. The quantile-quantile plot based on the association results showed minimal evidence of inflation ($\lambda=1.04$) (Supplementary Figure S2). We found eight SNPs independently associated with prostate cancer with $P < 1 \times 10^{-4}$ and $r^2 < 0.5$ for linkage disequilibrium (LD) (Table 1). Five of them have not been reported as associated with prostate cancer risk (Supplementary Figure S3) and the other three, all located at 8q24 regions, were previously reported (Table 1). However, the haplotype pattern observed in our genotype data suggests that SNP rs7463708, which is located within the previously reported 8q24 predisposition loci region 2, is potentially independent from region 2 for prostate cancer predisposition in Chinese population (Figure 1) and demonstrates the strongest association with

Table 1: SNPs independently associated with prostate cancer risk with $P < 1 \times 10^{-4}$ at the array genotyping stage of Chinese samples

| SNP | Location ^a | Alleles ^b | Study | MAF | | OR (95% CI) ^c | P_{GC} ^d | Previously reported |
|------------|------------------------|----------------------|-----------------------|-------|---------|--------------------------|------------------------|---------------------|
| | | | | Case | Control | | | |
| rs12567052 | 1q32.1 (200494416) | G/A | Chinese ^e | 0.241 | 0.321 | 0.67 (0.55-0.81) | 5.89×10^{-5} | No |
| | | G/A | European ^f | 0.404 | 0.399 | 1.02 (1.00-1.05) | 1.69×10^{-2} | |
| rs10235505 | 7p21.3 (7441152) | G/A | Chinese ^e | 0.347 | 0.266 | 1.45 (1.20-1.74) | 9.09×10^{-5} | No |
| | | G/A | European ^f | 0.247 | 0.244 | 1.02 (0.99-1.05) | 2.30×10^{-1} | |
| rs1532980 | 8p23.1 (9848617) | T/C | Chinese ^e | 0.428 | 0.486 | 0.70 (0.59-0.84) | 8.65×10^{-5} | No |
| | | T/C | European ^f | 0.252 | 0.248 | 1.02 (0.99-1.05) | 2.42×10^{-1} | |
| rs7463708 | 8q24.21 (128104055) | T/G | Chinese ^e | 0.232 | 0.331 | 0.60 (0.49-0.74) | 7.00×10^{-7} | ? |
| | | G/T | European ^f | 0.280 | 0.260 | 1.11 (1.07-1.14) | 6.87×10^{-10} | |
| rs7013255 | 8q24.21 (128130487) | T/G | Chinese ^e | 0.340 | 0.257 | 1.46 (1.21-1.76) | 5.90×10^{-5} | Yes |
| | | T/G | European ^f | 0.053 | 0.032 | 1.69 (1.58-1.81) | 4.17×10^{-48} | |
| rs13255059 | 8q24.21 (128530616) | G/A | Chinese ^e | 0.206 | 0.130 | 1.74 (1.37-2.20) | 4.22×10^{-6} | Yes |
| | | G/A | European ^f | 0.150 | 0.109 | 1.45 (1.39-1.50) | 5.24×10^{-69} | |
| rs10746597 | 9q21.31 (82344077) | A/G | Chinese ^e | 0.354 | 0.271 | 1.52 (1.25-1.85) | 2.58×10^{-05} | No |
| | | A/G | European ^f | 0.415 | 0.416 | 0.99 (0.97-1.02) | 7.04×10^{-1} | |
| rs1335214 | 9q33.2 (122847182) | G/C | Chinese ^e | 0.467 | 0.376 | 1.46 (1.22-1.74) | 2.82×10^{-05} | No |
| | | C/G | European ^f | 0.420 | 0.421 | 0.99 (0.97-1.02) | 8.16×10^{-1} | |

Abbreviations: MAF, minor allele frequency.

^aChromosomal and physical (in bracket) locations based on NCBI Human Genome Build 37.

^bMajor/minor allele.

^cAllelic OR with 95% CI for the minor allele in association with prostate cancer risk.

^d P_{GC} values in Chinese study indicate λ -corrected P values.

^eiCOGS study on Chinese population.

^fPRACTICAL iCOGS study on European population.

prostate cancer risk at the array stage. For additionally exploratory purposes, we performed a stepwise logistic regression analysis using the three 8q24 SNPs ($P < 1 \times 10^{-4}$ and $r^2 < 0.5$ for LD) and all previously reported prostate cancer predisposition loci located in 8q24 region. The multivariate analysis performed using the stepwise-selected model showed that rs7463708 exhibits independent association with Chinese prostate cancer (Supplementary Table S2). This adds further evidence that rs7463708 represents loci associated with prostate

cancer in Chinese population independent from other 8q24 predisposition SNPs.

Population genetic risk similarities and differences

To explore the population genetic risk differences we checked our data for previously reported prostate cancer predisposition loci in European and Japanese populations. We found suggestive association ($P < 0.05$) in only nine

of the 91 (10%) prostate cancer risk SNPs previously identified in European descendants and successfully measured by iCOGS array or imputed. Three of them are located at 8q24 (Supplementary Table S3). Whereas three of the seven (43%) predisposition SNPs initially identified in Japanese population showed association with prostate cancer ($P < 0.05$) in our data (Supplementary Table S3 and Supplementary Figure S4). We further estimated the heterogeneity for these predisposition loci between European and Chinese descendants using our Chinese and the PRACTICAL European array data. 24 of 96 showed opposite odds ratio (OR) direction and seven loci demonstrated significant differences ($P_{\text{het}} < 0.05$) in prostate cancer association between Chinese and European data sets (Supplementary Table S3 and Supplementary Figure S5). For the remaining previously reported prostate cancer predisposition loci, no evidence for significant difference between those two populations was found.

Chinese prostate cancer predisposition meta-analysis

To increase the statistical power and reduce potential false-positive findings we performed meta-analysis with the Chinese GWAS data [18] using SNPs associated with prostate cancer at $P < 0.001$ in our iCOGS array data (Supplementary Table S4). After combining the results from both studies using fixed-effect model, nine loci reached an association with prostate cancer at $P < 1 \times 10^{-3}$ (Table 2 and Supplementary Table S4). Two previously reported loci at 8q24 (region 1 and 2) reached the GWAS significance ($P < 5 \times 10^{-8}$) and one (region 3) reached association significance of $P = 2.5 \times 10^{-7}$. In addition, SNP rs1456315 closely linked with rs7463708 ($r^2 = 0.83$), which is absent in Chinese GWAS data, demonstrated the strongest evidence of prostate cancer association in Chinese men (Table 2 and Supplementary Table S4). This is consistent with the Chinese

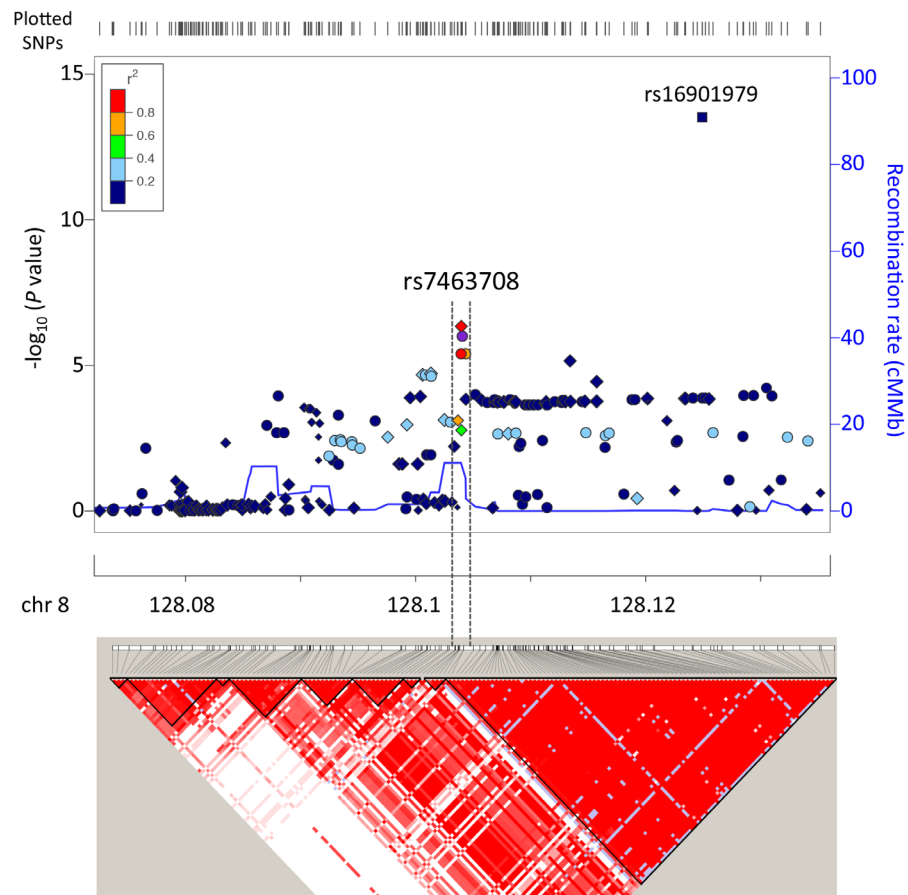


Figure 1: Regional association plot and LD map around SNP rs7463708. Top and bottom panels illustrate regional associations and LD map, respectively, around SNP rs7463708 potentially independently associated with prostate cancer. SNPs genotyped in this study, imputed on the basis of the 1000 Genomes Project data (Phase 1 integrated data version 3, March 2012) and rs16901979 (P value based on published data) associated with prostate cancer are indicated by circles, diamonds and square, respectively. Each point size is proportional to sample size. Symbol colors represent the LD of rs7463708 (purple circle) with nearby SNPs. The legend for LD measure (r^2) is on the top left corner of the plot. The left Y-axis illustrates the $-\log_{10}$ association P values of SNPs and the Y-axis on the right shows the recombination rate estimated from the 1000 Genomes Project using Chinese and Japanese population data. The LD map is based on r^2 values in CHB (Han Chinese in Beijing, China) and JPT (Japanese in Tokyo, Japan) samples from the 1000 Genomes Project.

Table 2: Results summary of prostate cancer susceptibility loci with $P < 1 \times 10^{-3}$ in meta-analysis of Chinese iCOGS and Chinese GWAS data

| SNP ^a | Location | Alleles ^b | Reported loci | Study | OR (95% CI) ^c | P |
|------------------|-----------------|----------------------|---------------|----------|--------------------------|------------------------|
| rs4539815 | 2q12 | T/C | No | iCOGS | 1.42 (1.17-1.74) | 5.50×10^{-4} |
| | | | | GWAS | 1.14 (1.00-1.31) | 6.45×10^{-2} |
| | | | | Combined | 1.22 (1.09-1.37) | 4.43×10^{-4} |
| rs12622816 | 2q31 | G/A | rs12621278 | iCOGS | 0.72 (0.59-0.87) | 8.92×10^{-4} |
| | | | | GWAS | 0.86 (0.76-0.98) | 1.94×10^{-2} |
| | | | | Combined | 0.82 (0.73-0.91) | 1.56×10^{-4} |
| rs13319291 | 3q22 | G/A | No | iCOGS | 0.71 (0.59-0.84) | 1.37×10^{-4} |
| | | | | GWAS | 0.90 (0.80-1.02) | 8.40×10^{-2} |
| | | | | Combined | 0.83 (0.76-0.92) | 3.64×10^{-4} |
| rs1456315 | 8q24 (region 2) | A/G | No | iCOGS | 0.61 (0.50-0.75) | 3.78×10^{-6} |
| | | | | GWAS | 0.61 (0.53-0.70) | 2.76×10^{-13} |
| | | | | Combined | 0.61 (0.55-0.68) | 7.93×10^{-18} |
| rs7013255 | 8q24 (region 2) | A/G | rs16901979 | iCOGS | 1.46 (1.22-1.76) | 5.90×10^{-5} |
| | | | | GWAS | 1.41 (1.24-1.60) | 1.07×10^{-7} |
| | | | | Combined | 1.43 (1.29-1.58) | 2.28×10^{-11} |
| rs12682374 | 8q24 (region 3) | G/C | rs6983267 | iCOGS | 1.36 (1.14-1.61) | 5.56×10^{-4} |
| | | | | GWAS | 1.26 (1.12-1.42) | 1.01×10^{-4} |
| | | | | Combined | 1.29 (1.17-1.42) | 2.49×10^{-7} |
| rs7824868 | 8q24 (region 1) | C/T | rs4242382 | iCOGS | 1.73 (1.37-2.18) | 4.61×10^{-6} |
| | | | | GWAS | 1.54 (1.32-1.80) | 5.14×10^{-8} |
| | | | | Combined | 1.60 (1.40-1.82) | 2.02×10^{-12} |
| rs6598099 | 11q13 | C/T | No | iCOGS | 1.44 (1.18-1.77) | 3.97×10^{-4} |
| | | | | GWAS | 1.12 (0.98-1.28) | 8.53×10^{-2} |
| | | | | Combined | 1.21 (1.08-1.35) | 8.18×10^{-4} |
| rs1893384 | 18q11 | C/A | No | iCOGS | 2.55 (1.50-4.31) | 5.20×10^{-4} |
| | | | | GWAS | 1.37 (0.94-1.99) | 9.95×10^{-2} |
| | | | | Combined | 1.69 (1.24-2.29) | 7.71×10^{-4} |

Abbreviation: GWAS, genome-wide association study.

^aSNP with the strongest evidence of prostate cancer association in Chinese men accordingly to meta-analysis of Chinese iCOGS and Chinese GWAS data sets.

^bMajor/minor allele.

^cOR with 95% CI for the minor allele in association with prostate cancer risk.

GWAS study identifying rs1456315 as the only SNP with a genome-wide significance level at the GWAS stage [18]. In this Chinese and a previous Japanese study [16], rs1456315 was considered to be within the same LD as the previously reported 8q24 region 2 identified in populations of European ancestry. However, the weak linkage ($r^2=0.14$) between rs1456315 and region 2 prostate cancer risk susceptibility locus (rs16901979) (Figure 1 and 2D), together with strong

association of rs1456315 (OR=0.66, 95% confidence intervals (CI)=0.52-0.82, $P=2.81 \times 10^{-4}$) remaining after conditioning on the effects of all the other predisposition loci at 8q24 region using iCOGS array data, suggests that these two loci are independently associated with Chinese prostate cancer. Apart from SNPs at 8q24 region, no other loci were associated with prostate cancer with $P < 1 \times 10^{-4}$ (Table 2 and Supplementary Table S4).

We checked the risk allele frequencies of these 8q24 SNPs in Chinese, Japanese and European populations. Surprisingly, we found that apart from 8q24 region 3, the risk allele frequencies of all SNPs are much higher in Chinese than European population and, for rs1456315 and rs13254738, the risk alleles are the major alleles in Chinese population but minor alleles in European population. The risk allele frequencies for many of these 8q24 SNPs are also higher in Chinese than Japanese population, although the differences are less apparent. Only for some SNPs at 8q24 region 1, the risk allele frequencies are slightly lower in Chinese than Japanese population (Figure 2A). This study also benefits from the direct comparison of the combined Chinese array data with PRACTICAL European population data of 25,074 cases and 24,272 controls. Apart from SNPs between rs1073997 and rs16901984 on 8q24 region 2 (Figure 2D), where the risk allele frequencies are very low (<5%) in European population, all the remaining risk SNPs had higher ORs in Chinese than the European men (Figure 2B). Interestingly, the two SNPs (rs1456315 and rs13254738) demonstrating the highest association with prostate cancer in Chinese population ($P=7.93\times 10^{-18}$ and 2.22×10^{-16} respectively) are not associated with prostate cancer at the GWAS level ($P<5\times 10^{-8}$) in European men (Figure 2C). Due to the high frequency of those risk alleles, the four 8q24 SNPs (rs1456315, rs16901979, rs6983267 and rs4242382) independently associated with Chinese prostate cancer coexist in a considerable large proportion of Chinese men, i.e. 6.5% subjects have six or more of these risk alleles. Based on these four loci, the prostate cancer risk for men in the top 10% of the polygenic risk score (PRS) distribution was 5.91 (95% CI, 2.68-13.04) fold compared with men in the bottom 10%, and around two-fold (1.96, 95% CI, 1.36-2.82) compared with the median risk (Supplementary Table S5). When we looked only at the top 5% of the population, the prostate cancer risk was 8.89 (95% CI, 3.46-22.83) fold compared with men in the bottom 5%, and 2.32 (95% CI, 1.55-3.49) fold compared with the median risk.

Replication of unreported SNPs

As none of the new suggestive loci ($P<0.001$) from our iCOGs analysis were supported by the Chinese GWAS data, we attempted to replicate the five association signals passing nominal significance $P<1\times 10^{-4}$ at the iCOGS stage (rs12567052, rs10235505, rs1532980, rs10746597 and rs1335214) using the TaqMan™ method in additional cohort of 1,940 cases and 2,820 controls of Han Chinese ancestry. Association analysis using the same approach as in the iCOGS genotyping stage confirmed two SNPs, rs12567052 at 1q32.1 ($P=1.9\times 10^{-4}$) and rs10235505 at 7p21.3 ($P=9.9\times 10^{-4}$), to have significant association with prostate cancer, while the other three were either not significantly associated ($P>0.05$) or demonstrated opposite effect to that observed at the array

stage (Supplementary Table S6). After combining the results from both stages using a meta-analysis assuming fixed-effect, the rs12567052[A] allele at 1q32.1 was associated with an OR of 0.79 for prostate cancer risk (95% CI=0.72-0.86, $P=2.41\times 10^{-7}$) and the rs10235505[A] allele at 7p21.3 was associated with an OR of 1.22 (95% CI=1.12-1.33, $P=2.4\times 10^{-6}$) (Supplementary Table S6). For both SNPs, the allele frequency differed considerably between the populations of European and Chinese (Table 1). In the PRACTICAL iCOGS cohort, which mainly contains European descendants [8], the rs12567052[A] putative protective allele in Chinese population showed significantly higher frequency in cases than controls ($P=0.017$). The rs10235505 did not show association with prostate cancer risk ($P=0.23$) (Table 1).

DISCUSSION

While more than 100 prostate cancer predisposition loci have been identified [3-20, 22, 23], the majority are from European patients. Identifying the predisposition loci in Asian populations, where prostate cancer incidence is relatively low, may not only help to prevent an increase of incidence rate in these populations, but also contribute to our understanding of the molecular and genetic mechanisms underlying prostate carcinogenesis by evaluating population differences.

In this study we identified a number of suggestive prostate cancer-associated genetic loci in Chinese Han population using the custom iCOGS array. Two of them, rs12567052 and rs10235505, remained to be associated at the replication stage, but none reached the genome-wide level significance when combined all the data. Further studies with much larger sample size are required to determine if rs12567052 and rs10235505 are associated with Chinese prostate cancer risk. The SNP with the strongest association at the array stage was rs7463708 at 8q24 (OR=0.60, 95% CI=0.49-0.74, $P=7.00\times 10^{-7}$), which is consistent with the Chinese GWAS study identifying rs1456315, a SNP in strong linkage with rs7463708, as the only SNP with a genome-wide significance level at the GWAS stage [18]. While this SNP is located within previously reported prostate cancer predisposition loci 8q24 region 2 identified in populations of European ancestry, the haplotype pattern observed in our genotype data and stepwise logistic regression analysis suggest that these two loci are independently associated with prostate cancer in Chinese population.

By comparing our data with prostate cancer predisposition loci reported in European and Japanese populations [4, 6-17, 20, 24-29], we found that many prostate cancer predisposition loci identified in European population are not associated with prostate cancer in Chinese men. Clearly the sample size of our study is relatively small, hence the statistical power is limited and with increased sample size more reported predisposition

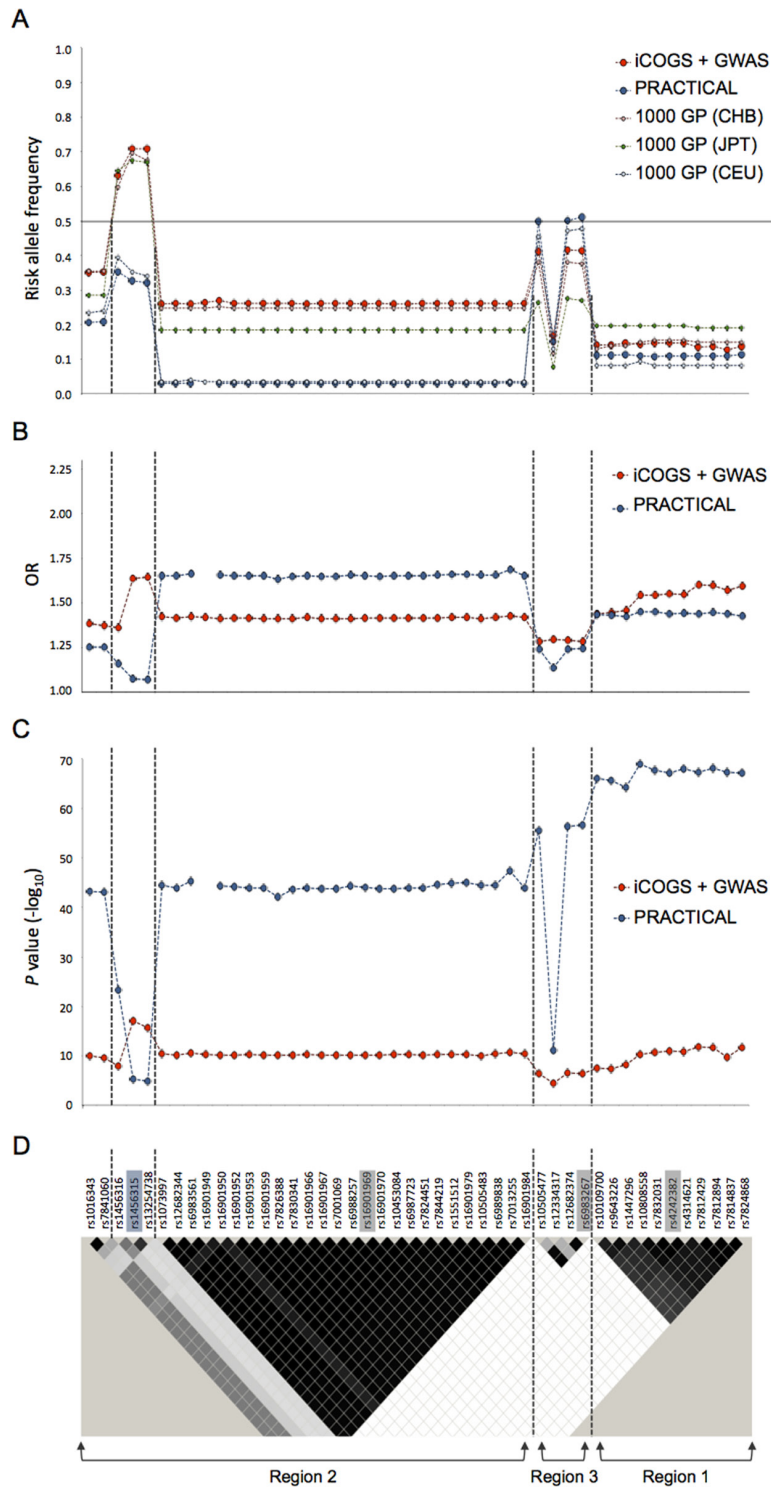


Figure 2: The risk allele frequency, OR and P value plots at the four 8q24 prostate cancer-associated loci in Chinese, Japanese and European descendants. **A.** Risk allele frequency, **B.** OR and **C.** P value plots of SNPs with $P < 1 \times 10^{-4}$ in meta-analysis results of Chinese data. **D.** LD map of 8q24 region based on r^2 values computed using iCOGS Chinese genotyping data. Index SNPs are highlighted in grey on the LD plot. The independently associated SNP rs1456315 in Chinese iCOGS data is highlighted in blue. Previously reported prostate cancer-associated regions 1, 2 and 3 are indicated below the LD map. The vertical dashed lines distinguish the four 8q24 loci independently associated with Chinese prostate cancer accordingly to Chinese iCOGS data. iCOGS: iCOGS array data from Chinese ancestors; PRACTICAL: iCOGS array data from European ancestors; 1000 GP: 1000 Genome Project data; GWAS: Chinese GWAS data.

loci to the European men are likely to be confirmed in the Chinese population. However, the much greater proportion of predisposition loci identified in Japanese population compared to those identified in European populations, which demonstrated an association ($P < 0.05$) in our Chinese cohort, suggests that these two Eastern Asian populations have more similar genetic risk to prostate cancer than European descent. Moreover, we observed prominent difference in allele frequency in the 8q24 loci and several loci reported in Western populations that have significantly different effect on Chinese men, demonstrating that certain population difference in prostate cancer predisposition exists. However, while we identified differences between Chinese and European populations for certain previously reported prostate cancer predisposition loci, the majority of them did not demonstrate significant heterogeneity between data sets derived from Chinese and European descendants. This is convergent with the presumption that many genetic predisposition loci may contribute to prostate cancer in multiple ethnic populations.

The current sample size of microarray data for Chinese prostate cancer predisposition study is still limited. Therefore, susceptibility loci with very low penetrance may not be detected due to insufficient statistical power. Further Chinese studies with larger cohort of samples are required to comprehensively characterize the genetic predisposition of prostate cancer in Chinese men. However, the much greater significance of prostate cancer association at these four 8q24 SNPs compared with other SNPs in the Chinese array data sets indicate that other loci potentially to be confirmed or identified in Chinese population may have much smaller contribution to prostate cancer risk in Chinese population. As a large proportion (around 6.5%) of Chinese men carry six or more of these 8q24 risk alleles and the top 10% of the population have six- and two-fold prostate cancer risk compared with men of the bottom 10% and median risk respectively, these four 8q24 SNPs may be implemented in a simple prostate cancer genetic risk screening kit to detect these high-risk Chinese men for individualized cancer prevention. In addition to individual genetic risk assessment for cancer prevention, SNP-based prediction model for prostate cancer diagnosis has also been developed [30, 31]. In a previously reported 24 SNP-based genetic score for prostate cancer diagnosis prediction in Chinese population, three of the four SNPs in our study, rs16901979, rs6983267 and rs1447295, have been genotyped. We compared the performance of our three-SNP combination in comparison to the 24 SNPs and the area under the curve (AUC) for our three SNPs was 0.646 and the AUC for the 24 SNPs was 0.658. Although the 24 SNPs performed slightly better than the three SNPs, the cancer prediction value is similar. It is possible that including the 4th SNP, rs1456315, which demonstrated the strongest signal in the Chinese iCOGS and GWAS, could

further improve the AUC, suggesting these four 8q24 SNPs may also be useful for diagnosis prediction. This genetic score data also support our observation that the other SNPs have limited contribution to prostate cancer risk in Chinese men.

While those findings suggest that Chinese population has fewer genetic risk loci for prostate cancer than European population, which is consistent with the prostate cancer incidence disparity between these two populations, it is surprising that the risk alleles are more frequent and ORs are greater at these four 8q24 loci in Chinese than European descendants. Our findings suggest that Chinese population either generally has a lower genetic risk or is better protected by healthy diet than European population, resulting in much lower incidence and mortality of prostate cancer. If the former is true, it means that the recent rapid increase in Chinese prostate cancer incidence will peak soon and won't ever achieve European levels. However, if the latter is the case, with dietary change and increased population, prostate cancer will become a common cancer in China.

In summary, while we did not identify new prostate cancer predisposition loci at genome-wide significance level in Chinese population using our SNP array data, meta-analysis and replication study, the meta-analysis with Chinese GWAS showed that four 8q24 loci increase the risk of prostate cancer considerably in a large proportion of Chinese men. These four 8q24 loci bear great potential for designing a simple prostate cancer risk screening kit to identify risk individuals in the large Chinese population. We also identified differences between Chinese and European populations for certain previously reported prostate cancer predisposition loci. These findings will help to understand the predisposition and molecular mechanisms of prostate carcinogenesis and improve cancer prevention.

MATERIALS AND METHODS

Study design

In this study 211,155 SNPs were genotyped in 495 cases and 640 age-matched non-cancer male controls from the Han Chinese population using iCOGS array [8] at Beijing Genome Institute. The array was designed by the international consortium, COGS, to detect genetic variants related to prostate, breast and ovarian cancers. The replication of nominally associated SNPs within unreported loci was performed in additional 1,940 cases and 2,820 controls. Demographic and clinical features of subjects for the prostate cancer predisposition study are summarized in Supplementary Table S1. These male Han Chinese subjects were recruited from several regions of China (south, southeast, southwest, middle and northeast) by members of the Chinese Prostate Cancer Genetic and Environmental Correlation Study (CHIPGECS)

collaborative group. The cases were hospital based with pathological diagnosis of primary prostate cancer. Similar numbers of cancer-free controls and cases were recruited from the community or hospital subjects undergoing routine physical examination in each recruitment region in China. Controls are all males with comparable ages to the cases, particularly for the first stage conducted with iCOGS array. Blood samples were obtained with informed consent and the study was approved by the Institutional Ethical Committee of each participating institution. The Chinese GWAS [18] data used for meta-analysis contained 1,417 cases and 1,008 cancer-free male controls from the Han Chinese population. The GWAS study subjects selection and genotype data quality control methods were previously described [18].

Genotype quality control

Systematic quality control steps were conducted on the raw iCOGS genotyping data resulting in the exclusion of 15 samples with an overall genotyping success rate less than 95%, four samples with high B allele frequency (BAF) variance, ten samples predicted as female based on X chromosome specific inbreeding coefficient (F) less than 0.2, and two samples with low heterozygosity ($F > 0.1$). Additional 35 samples identified as duplicates or second-degree relatives were excluded on the basis of identity-by-descent analysis performed using PLINK (v1.07) [32].

Ancestry and population stratification were determined by principal component analysis. The principal components were calculated for a subset of 66,878 uncorrelated ($r^2 < 0.5$) autosomal SNPs that passed quality control and were present on both the iCOGS array and in HapMap 3 data from four reference populations including CEU (Utah residents with Northern and Western European ancestry), CHB (Han Chinese in Beijing, China), JPT (Japanese in Tokyo, Japan) and YRI (Yoruba in Ibadan, Nigeria) (Supplementary Figure S1A). Subsequently, principal components were calculated for a set of 71,591 uncorrelated autosomal markers present on the iCOGS array alone (Supplementary Figure S1B).

The exclusion criteria for SNPs were genotyping call rate less than 95% (3,174 SNPs), minor allele frequency (MAF) less than 1% (37,920 SNPs) and genotype frequency that deviated from expected Hardy-Weinberg equilibrium among control samples ($P < 1 \times 10^{-4}$) (1,039 SNPs).

Statistical analyses

The association between each SNP and the disease phenotype was estimated by per-allele OR and 95% CI using unconditional logistic regression implemented in PLINK, assuming an additive genetic model. A genomic control (GC) method was applied to adjust the P values for unknown genetic heterogeneity. Ungenotyped SNPs within 100kb flanking regions on each side of associated

SNPs ($P < 1 \times 10^{-4}$) were imputed with IMPUTE2 (version 2.3.0) software [33] using 1000 Genomes Project data (Phase 1 integrated data version 3, March 2012) from Chinese and Japanese populations as reference haplotype maps. A pre-phasing technique was applied to SNP array data to produce best-guess haplotypes and speed up the imputation process [34]. A posterior probability greater than 0.9 was applied to call imputed genotypes. Haploview software (v4.2) [35] was used to determine pair-wise LD structure across interrogated genomic regions. CLUMP analysis with PLINK was performed to identify SNPs independently associated with the disease phenotype. Akaike information criterion (AIC) was used for stepwise logistic regression model selection. Conditional and stepwise logistic regression analyses were performed in R.

Results obtained in independent genotyping analysis were combined with fixed-effect meta-analysis implemented in PLINK. The heterogeneity between studies was tested with Cochran's Q test statistic. Results that obtained P value less than 5.0×10^{-8} were considered statistically significant at the genome-wide level.

The different prostate cancer risk strata were defined by PRS constructed from the summed genotypes weighted by the estimated per-allele log-ORs obtained in meta-analysis for each of the four 8q24 loci (Supplementary Table S7) [36].

ACKNOWLEDGMENTS

This work was conducted on behalf of the CHIPGECs and The PRACTICAL consortia (see Supplementary Consortia). We acknowledge the contribution of doctors, nurses and postgraduate research students at the CHIPGENCS sample collecting centers. We thank Orchid and Rosetrees for funding support. This work was also supported by National Natural Science foundation of China for funding support to H Zhang (Grant No: 30671793 and 81072377), N Feng (Grant No: 81272831), X Zhang (Grant No: 30572139, 30872924 and 81072095), S Zhao (Grant No: 81072092 and 81328017), Y Yu (Grant No: 81172448) and Program for New Century Excellent Talents in University from Department of Education of China (NCET-08-0223) and the National High Technology Research and Development Program of China (863 Program 2012AA021101) to X Zhang.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

1. Center MM, Jemal A, Lortet-Tieulent J, Ward E, Ferlay J, Brawley O, Bray F. International variation in prostate cancer incidence and mortality rates. *European urology*. 2012; 61: 1079-1092.

2. Mao X, Yu Y, Boyd LK, Ren G, Lin D, Chaplin T, Kudahetti SC, Stankiewicz E, Xue L, Beltran L, Gupta M, Oliver RT, Lemoine NR, et al. Distinct genomic alterations in prostate cancers in Chinese and Western populations suggest alternative pathways of prostate carcinogenesis. *Cancer Res.* 2010; 70: 5207-5212.
3. Akamatsu S, Takata R, Haiman CA, Takahashi A, Inoue T, Kubo M, Furihata M, Kamatani N, Inazawa J, Chen GK, Le Marchand L, Kolonel LN, Katoh T, et al. Common variants at 11q12, 10q26 and 3p11.2 are associated with prostate cancer susceptibility in Japanese. *Nat Genet.* 2012; 44: 426-429, S421.
4. Amundadottir LT, Sulem P, Gudmundsson J, Helgason A, Baker A, Agnarsson BA, Sigurdsson A, Benediktsdottir KR, Cazier JB, Sainz J, Jakobsdottir M, Kostic J, Magnusdottir DN, et al. A common variant associated with prostate cancer in European and African populations. *Nat Genet.* 2006; 38: 652-658.
5. Boyd LK, Mao X, Lu YJ. The complexity of prostate cancer: genomic alterations and heterogeneity. *Nat Rev Urol.* 2012; 9: 652-664.
6. Eeles RA, Kote-Jarai Z, Al Olama AA, Giles GG, Guy M, Severi G, Muir K, Hopper JL, Henderson BE, Haiman CA, Schleutker J, Hamdy FC, Neal DE, et al. Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat Genet.* 2009; 41: 1116-1121.
7. Eeles RA, Kote-Jarai Z, Giles GG, Olama AA, Guy M, Jugurnauth SK, Mulholland S, Leongamornlert DA, Edwards SM, Morrison J, Field HI, Southey MC, Severi G, et al. Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet.* 2008; 40: 316-321.
8. Eeles RA, Olama AA, Benlloch S, Saunders EJ, Leongamornlert DA, Tymrakiewicz M, Ghousaini M, Luccarini C, Dennis J, Jugurnauth-Little S, Dadaev T, Neal DE, Hamdy FC, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat Genet.* 2013; 45: 385-391, 391e381-382.
9. Gudmundsson J, Sulem P, Gudbjartsson DF, Blondal T, Gylfason A, Agnarsson BA, Benediktsdottir KR, Magnusdottir DN, Orlygsdottir G, Jakobsdottir M, Stacey SN, Sigurdsson A, Wahlfors T, et al. Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat Genet.* 2009; 41: 1122-1126.
10. Gudmundsson J, Sulem P, Gudbjartsson DF, Masson G, Agnarsson BA, Benediktsdottir KR, Sigurdsson A, Magnusson OT, Gudjonsson SA, Magnusdottir DN, Johannsdottir H, Helgadottir HT, Stacey SN, et al. A study based on whole-genome sequencing yields a rare variant at 8q24 associated with prostate cancer. *Nat Genet.* 2012; 44: 1326-1329.
11. Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, Rafnar T, Bergthorsson JT, Agnarsson BA, Baker A, Sigurdsson A, Benediktsdottir KR, Jakobsdottir M, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet.* 2007; 39: 631-637.
12. Gudmundsson J, Sulem P, Rafnar T, Bergthorsson JT, Manolescu A, Gudbjartsson D, Agnarsson BA, Sigurdsson A, Benediktsdottir KR, Blondal T, Jakobsdottir M, Stacey SN, Kostic J, et al. Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. *Nat Genet.* 2008; 40: 281-283.
13. Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, Manolescu A, Rafnar T, Gudbjartsson D, Agnarsson BA, Baker A, Sigurdsson A, Benediktsdottir KR, Jakobsdottir M, et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet.* 2007; 39: 977-983.
14. Haiman CA, Chen GK, Blot WJ, Strom SS, Berndt SI, Kittles RA, Rybicki BA, Isaacs WB, Ingles SA, Stanford JL, Diver WR, Witte JS, Hsing AW, et al. Genome-wide association study of prostate cancer in men of African ancestry identifies a susceptibility locus at 17q21. *Nat Genet.* 2011; 43: 570-573.
15. Kote-Jarai Z, Olama AA, Giles GG, Severi G, Schleutker J, Weischer M, Campa D, Riboli E, Key T, Gronberg H, Hunter DJ, Kraft P, Thun MJ, et al. Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. *Nat Genet.* 2011; 43: 785-791.
16. Takata R, Akamatsu S, Kubo M, Takahashi A, Hosono N, Kawaguchi T, Tsunoda T, Inazawa J, Kamatani N, Ogawa O, Fujioka T, Nakamura Y, Nakagawa H. Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. *Nat Genet.* 2010; 42: 751-754.
17. Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, Orr N, Yu K, Chatterjee N, Welch R, Hutchinson A, Crenshaw A, Cancel-Tassin G, Staats BJ, et al. Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet.* 2008; 40: 310-315.
18. Xu J, Mo Z, Ye D, Wang M, Liu F, Jin G, Xu C, Wang X, Shao Q, Chen Z, Tao Z, Qi J, Zhou F, et al. Genome-wide association study in Chinese men identifies two new prostate cancer risk loci at 9q31.2 and 19q13.4. *Nat Genet.* 2012; 44: 1231-1235.
19. Yeager M, Chatterjee N, Ciampa J, Jacobs KB, Gonzalez-Bosquet J, Hayes RB, Kraft P, Wacholder S, Orr N, Berndt S, Yu K, Hutchinson A, Wang Z, et al. Identification of a new prostate cancer susceptibility locus on chromosome 8q24. *Nat Genet.* 2009; 41: 1055-1057.
20. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, Minichiello MJ, Fearnhead P, Yu K, Chatterjee N, Wang Z, Welch R, Staats BJ, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet.* 2007; 39: 645-649.
21. Yamada H, Penney KL, Takahashi H, Katoh T, Yamano Y, Yamakado M, Kimura T, Kuruma H, Kamata Y, Egawa S, Freedman ML. Replication of prostate cancer risk loci

- in a Japanese case-control association study. *Journal of the National Cancer Institute*. 2009; 101: 1330-1336.
22. Al Olama AA, Kote-Jarai Z, Berndt SI, Conti DV, Schumacher F, Han Y, Benlloch S, Hazelett DJ, Wang Z, Saunders E, Leongamornlert D, Lindstrom S, Jugurnauth-Little S, et al. A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. *Nat Genet*. 2014; 46: 1103-1109.
 23. Amin Al Olama A, Kote-Jarai Z, Schumacher FR, Wiklund F, Berndt SI, Benlloch S, Giles GG, Severi G, Neal DE, Hamdy FC, Donovan JL, Hunter DJ, Henderson BE, et al. A meta-analysis of genome-wide association studies to identify prostate cancer susceptibility loci associated with aggressive and non-aggressive disease. *Hum Mol Genet*. 2013; 22: 408-415.
 24. Al Olama AA, Kote-Jarai Z, Giles GG, Guy M, Morrison J, Severi G, Leongamornlert DA, Tymrakiewicz M, Jhavar S, Saunders E, Hopper JL, Southey MC, Muir KR, et al. Multiple loci on 8q24 associated with prostate cancer susceptibility. *Nat Genet*. 2009; 41: 1058-1060.
 25. Duggan D, Zheng SL, Knowlton M, Benitez D, Dimitrov L, Wiklund F, Robbins C, Isaacs SD, Cheng Y, Li G, Sun J, Chang BL, Marovich L, et al. Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. *J Natl Cancer Inst*. 2007; 99: 1836-1844.
 26. Sun J, Zheng SL, Wiklund F, Isaacs SD, Purcell LD, Gao Z, Hsu FC, Kim ST, Liu W, Zhu Y, Stattin P, Adami HO, Wiley KE, et al. Evidence for two independent prostate cancer risk-associated loci in the HNF1B gene at 17q12. *Nat Genet*. 2008; 40: 1153-1155.
 27. Schumacher FR, Berndt SI, Siddiq A, Jacobs KB, Wang Z, Lindstrom S, Stevens VL, Chen C, Mondul AM, Travis RC, Stram DO, Eeles RA, Easton DF, et al. Genome-wide association study identifies new prostate cancer susceptibility loci. *Hum Mol Genet*. 2011; 20: 3867-3875.
 28. Hsu FC, Sun J, Wiklund F, Isaacs SD, Wiley KE, Purcell LD, Gao Z, Stattin P, Zhu Y, Kim ST, Zhang Z, Liu W, Chang BL, et al. A novel prostate cancer susceptibility locus at 19q13. *Cancer Res*. 2009; 69: 2720-2723.
 29. Sun J, Zheng SL, Wiklund F, Isaacs SD, Li G, Wiley KE, Kim ST, Zhu Y, Zhang Z, Hsu FC, Turner AR, Stattin P, Liu W, et al. Sequence variants at 22q13 are associated with prostate cancer risk. *Cancer Res*. 2009; 69: 10-15.
 30. Aly M, Wiklund F, Xu J, Isaacs WB, Eklund M, D'Amato M, Adolfsson J, Gronberg H. Polygenic risk score improves prostate cancer risk prediction: results from the Stockholm-1 cohort study. *European urology*. 2011; 60: 21-28.
 31. Jiang H, Liu F, Wang Z, Na R, Zhang L, Wu Y, Zheng J, Lin X, Jiang D, Sun J, Zheng SL, Ding Q, Xu J. Prediction of prostate cancer from prostate biopsy in Chinese men using a genetic score derived from 24 prostate cancer risk-associated SNPs. *The Prostate*. 2013; 73: 1651-1659.
 32. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81: 559-575.
 33. 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: e1000529.
 34. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*. 2012; 44: 955-959.
 35. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005; 21: 263-265.
 36. Amin Al Olama A, Benlloch S, Antoniou AC, Giles GG, Severi G, Neal DE, Hamdy FC, Donovan JL, Muir K, Schleutker J, Henderson BE, Haiman CA, Schumacher FR, et al. Risk Analysis of Prostate Cancer in PRACTICAL, a Multinational Consortium, Using 25 Known Prostate Cancer Susceptibility Loci. *Cancer epidemiology, biomarkers & prevention*. 2015; 24: 1121-1129.