Genotypic variants at 2q33 and risk of esophageal squamous cell carcinoma in China: a meta-analysis of genome-wide association studies

Christian C. Abnet[1,*,†], Zhaoming Wang[1,3,†], Xin Song[4,5,†], Nan Hu[1,†], Fu-You Zhou[7,†], Neal D. Freedman[1,†], Xue-Min Li[8,†], Kai Yu[1], Xiao-Ou Shu[9], Jian-Min Yuan[10], Wei Zheng[9], Sanford M. Dawsey[1], Linda M. Liao[1], Maxwell P. Lee[2], Ti Ding[11], You-Lin Qiao[12], Yu-Tang Gao[13], Woon-Puay Koh[14], Yong-Bing Xiang[13], Ze-Zhong Tang[11], Jin-Hu Fan[12], Charles C. Chung[1,3], Chaoyu Wang[1], William Wheeler[15], Meredith Yeager[1,3], Jeff Yuenger[1,3], Amy Hutchinson[1,3], Kevin B. Jacobs[1,3], Carol A. Giffen[15], Laurie Burdett[1,3], Joseph F. Fraumeni Jr[1], Margaret A. Tucker[1], Wong-Ho Chow[1], Xue-Ke Zhao[4,5], Jiang-Man Li[4,5], Ai-Li Li[4,16], Liang-Dan Sun[17], Wu Wei[18], Ji-Lin Li[19], Peng Zhang[4], Hong-Lei Li[5], Wen-Yan Cui[4], Wei-Peng Wang[5,8], Zhi-Cai Liu[16], Xia Yang[5], Wen-Jing Fu[5], Ji-Li Cui[4], Hong-Li Lin[4], Wen-Liang Zhu[5], Min Liu[5], Xi Chen[5], Jie Chen[5], Li Guo[5], Jing-Jing Han[5], Sheng-Li Zhou[5], Jia Huang[5], Yue Wu[5], Chao Yuan[5], Jing Huang[5], Ai-Fang Ji[5,18], Jian-Wei Kul[5,20], Zhong-Min Fan[5], Jian-Po Wang[5,7], Dong-Yun Zhang[5], Lian-Qun Zhang[5,7], Wei Zhang[21], Yuan-Fang Chen[22], Jing-Li Ren[6], Xiu-Min Li[4], Jin-Cheng Dong[5,23], Guo-Lan Xing[5], Zhi-Gang Guo[4], Jian-Xue Yang[5,24], Yi-Ming Mao[5,24], Yuan Yuan[5], Er-Tao Guo[5], Wei Zhang[5], Zhi-Chao Hou[5], Jing Liu[5], Yan Li[5], Sa Tang[5], Jia Chang[5], Xiu-Qin Peng[5,24], Min Han[25], Wan-Li Yin[5], Ya-Li Liu[4], Yan-Long Hu[4], Yu Liu[5], Liu-Qin Yang[26], Fu-Guo Zhu[25], Xiu-Feng Yang[27], Xiao-Shan Feng[24], Zhou Wang[28], Yin Li[29], She-Gan Gao[24], Hai-Lin Liu[4], Ling Yuan[29], Yan Jin[4], Yan-Rui Zhang[30], Ilyar Sheyhidin[31], Feng Li[32], Bao-Ping Chen[33], Shu-Wei Ren[34], Bin Liu[35], Dan Li[5,36], Gao-Fu Zhang[36], Wen-Bin Yue[5,37], Chang-Wei Feng[6], Qirenwang Qige[38], Jian-Ting Zhao[21], Wen-Jun Yang[39], Guang-Yan Lei[40], Long-Qi Chen[41], En-Min Li[42], Li-Yan Xu[42], Zhi-Yong Wu[42], Zhi-Qin Bao[8], Ji-Li Chen[8], Xian-Chang Li[15], Xiang Zhuang[5], Ying-Fa Zhou[5], Xian-Bo Zuo[17], Zi-Ming Dong[43], Lu-Wen Wang[5], Xue-Pin Fan[5], Jin Wang[5], Qi Zhou[43], Guo-Shun Ma[44], Qin-Xian Zhang[43], Hai Liu[45], Xin-Ying Jian[19], Sin-Yong Lian[16], Jin-Sheng Wang[18], Fu-Bao Chang[46], Chang-Dong Lu[7], Jian-Jun Miao[8], Zhi-Guo Chen[4], Ran Wang[5], Ming Guo[5], Zeng-Lin Fan[8], Ping Tao[47], Tai-Jing Liu[36], Jin-Chang Wei[19], Qing-Peng Kong[48], Lei Fan[8], Xian-Zeng Wang[49], Fu-Sheng Gao[35], Tian-Yun Wang[4], Dong Xie[50], Li Wang[5], Shu-Qing Chen[51], Wan-Cai Yang[4], Jun-Yan Hong[51], Liang Wang[52], Song-Liang Qiu[5,†], Alisa M. Goldstein[1,†], Zhi-Qing Yuan[4,†], Stephen J. Chanock[1,†], Xue-Jun Zhang[17,†], Philip R. Taylor[1,†] and Li-Dong Wang[4,5,*]

[1]Division of Cancer Epidemiology and Genetics and [2]Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA, [3]Core Genotyping Facility, NCI-Frederick, SAIC-Frederick, Frederick, MD, USA, [4]Cancer Research Center, Xinxiang Medical University, Xinxiang, Henan 453003, China, [5]Henan Key

---

*To whom correspondence should be addressed at: Nutritional Epidemiology Branch, DCEG, NCI, EPS 320 Bethesda, MD 20892-7236, USA. Tel: +1 3015941511; Fax: +1 3014966829; Email: abnetc@mail.nih.gov (C.C.A.); Cancer Research Center, Xinxiang Medical University, Xinxiang, Henan, China. Tel: +86 13937183227; Email: ldwang@xxmu.edu.cn (L.-D.W.)
†C.C.A., Z.W., X.S., N.H., F-Y.Z., N.D.F., X.-M.L. are co-first authors; S.-L.Q., A.M.G., Z.-Q.Y., S.J.C., X.-J.Z., P.R.T. and L.-D.W. are co-last authors.

Laboratory for Esophageal Cancer Research of The First Affiliated Hospital and [6]Department of Basic Oncology and Pathology, The Second Affiliated Hospital, Zhengzhou University, Zhengzhou, Henan 450052, China, [7]Department of Thoracic Surgery and Tumor Prevention Treatment, Anyang Tumor Hospital, Anyang, Henan 455000, China, [8]Department of Pathology and Thoracic Surgery, Centre for Health Screening and Endoscopy, Cixian Hospital, Cixian, Hebei 056500, China, [9]Department of Medicine and Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, TN, USA, [10]University of Pittsburgh Cancer Institute, Pittsburg, PA, USA, [11]Shanxi Cancer Hospital, Taiyuan, Shanxi, China, [12]Department of Epidemiology, Cancer Institute (Hospital), Chinese Academy of Medical Sciences, Beijing, China, [13]Shanghai Cancer Institute, Shanghai, China, [14]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore, [15]Information Management Services, Inc., Silver Spring, MD, USA, [16]Department of Thoracic Surgery, Tumor Hospital of Linzhou, Linzhou, Henan 456550, China, [17]Key Laboratory of Dermatology (Anhui Medical University), Ministry of Education, Hefei, Anhui 230032, China, [18]Institute of Hematologic Disease and Department of Pathology, Changzhi Medical University, Changzhi, Shanxi 046000, China, [19]Department of Pathology and Thoracic Surgery, Linzhou Esophageal Cancer Hospital, Linzhou, Henan 456592, China, [20]Department of Oncology, The First Affiliated Hospital of Nanyang Medical College, Nanyang, Henan 473058, China, [21]Department of Oncology, The First People's Hospital of Hebi, Hebi, Henan 458000, China, [22]Department of Biostatistics, Hebi City Public Health Bureau, Hebi, Henan 458000, China, [23]Qixian Commerce Bureau, Hebi, Henan 456750, China, [24]Department of Oncology, The First Affiliated Hospital, Henan University of Science and Technology, Luoyang, Henan 471003, China, [25]Department of Gastroenterology and Thoracic Surgery, The First People's Hospital of Shangqiu, Shangqiu, Henan 476100, China, [26]Department of Radiotherapy, Xinxiang Central Hospital, Xinxiang, Henan 453000, China, [27]Department of Surgery, Hebi Dahejian Hospital, Hebi, Henan 458000, China, [28]Department of Thoracic Surgery, Provincial Hospital Affiliated to Shandong University, Jinan, Shandong 250021, China, [29]Department of Thoracic Surgery and Radiotherapy, Cancer Hospital of Zhengzhou University, Zhengzhou, Henan 450008, China, [30]Department of Gastroenterology, People's Hospital of Zhengzhou University, Zhengzhou, Henan 450003, China, [31]Department of Thoracic Surgery and Medical Research Center, The First Affiliated Hospital, Xinjiang Medical University, Urumqi, Xinjiang 830054, China, [32]Department of Pathology, Shihezi University School of Medicine, Shihezi, Xinjiang 832002, China, [33]Department of Nephrology, Huaihe Hospital, Henan University, Kaifeng, Henan 475000, China, [34]Department of Oncology, Xinyang Central Hospital, Xinyang, Henan 464000, China, [35]Department of Gastroenterology, Beijing Tongren Hospital Affiliated to Capital Medical University, Beijing 100176, China, [36]Department of Surgery, The Red Cross Hospital of Huojia, Xinxiang, Henan 453800, China, [37]Department of Oncology, Puyang City Oil Field Hospital, Puyang, Henan 457000, China, [38]Department of Internal Mongolia Medicine, The Affiliated Hospital, Inner Mongolia Medical College, Huhhot, Inner Mongolia 010050, China, [39]Department of Biotechnology, Ningxia Medical University, Yinchuan, Ningxia 750004, China, [40]Department of Thoracic Surgery, Tumor Hospital of Shaanxi Province, Xi'an, Shaanxi 710061, China, [41]Department of Thoracic and Cardiovascular Surgery, West China Hospital, Sichuan University, Chengdu, Sichuan 610041, China, [42]Department of Oncologic Pathology and Department of Biochemistry and Molecular Biology, Medical College of Shantou University, Shantou, Guangdong 515041, China, [43]Department of Oncology and Pathology, Basic Medical College of Zhengzhou University, Zhengzhou, Henan 450052, China, [44]Department of Biostatistics, Neihuang Public Health Bureau, Anyang, Henan 4563000, China, [45]Department of Gastroenterology, The Third People's Hospital of Yunnan Province, Kunming, Yunnan 650011, China, [46]Department of Thoracic Surgery, Chinese Medicine Hospital of Linzhou, Linzhou, Henan 456550, China, [47]Department of Oncologic Pathology, Langzhongshi People's Hospital, Langzhong, Sichuan 637400, China, [48]State Key Laboratory of Genetic Resource and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China, [49]Department of Thoracic Surgery, Linzhou People's Hospital, Linzhou, Henan 456550, China, [50]Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China, [51]College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China and [52]MCW Cancer Center, Department of Pathology, Medical College of Wisconsin, Milwaukee, WI, USA

Genome-wide association studies have identified susceptibility loci for esophageal squamous cell carcinoma (ESCC). We conducted a meta-analysis of all single-nucleotide polymorphisms (SNPs) that showed nominally significant *P*-values in two previously published genome-wide scans that included a total of 2961 ESCC cases and 3400 controls. The meta-analysis revealed five SNPs at 2q33 with $P < 5 \times 10^{-8}$, and the strongest signal was rs13016963, with a combined odds ratio (95% confidence interval) of 1.29 (1.19–1.40) and $P = 7.63 \times 10^{-10}$. An imputation analysis of 4304 SNPs at 2q33 suggested a single association signal, and the strongest imputed SNP associations were similar to those from the genotyped SNPs. We conducted an ancestral recombination graph analysis with 53 SNPs to identify one or more haplotypes that harbor the variants directly responsible for the detected association signal. This showed that the five SNPs exist in a single haplotype along with 45 imputed SNPs in strong linkage disequilibrium, and the strongest candidate was rs10201587, one of the genotyped SNPs. Our meta-analysis found genome-wide significant SNPs at 2q33 that map to the *CASP8*/*ALS2CR12*/*TRAK2* gene region. Variants in *CASP8* have been extensively studied across a spectrum of cancers with mixed results. The locus we identified appears to be distinct from the widely studied rs3834129 and rs1045485 SNPs in *CASP8*. Future studies of esophageal and other cancers should focus on comprehensive sequencing of this 2q33 locus and functional analysis of rs13016963 and rs10201587 and other strongly correlated variants.

INTRODUCTION

Worldwide, esophageal cancer causes more than 400 000 deaths each year (1) and within the People's Republic of China it ranks fourth as a cause of cancer-related deaths (2). As seen throughout the economically developing world, esophageal squamous cell carcinoma (ESCC) predominates, which differs from recent trends in populations of European descent, where esophageal adenocarcinoma rates now exceed ESCC (3). In contrast to low-incidence populations, tobacco smoking and excessive alcohol consumption do not appear to be major risk factors in the geographically limited areas of China with a heavy burden of ESCC (4). Numerous studies show that a family history of esophageal cancer increases risk in China (5), but compared with other cancer types, few studies have comprehensively assessed the contribution of common genetic variants to ESCC risk.

Recently, two genome-wide association studies (GWASs) examined the contribution of common genetic variants to ESCC risk in Han Chinese (6,7). Both studies reported a strong association with single-nucleotide polymorphisms (SNPs) on chromosome 10q23, which harbors a plausible candidate gene, *PLCE1*. Variants in *PLCE1* have also been linked to an inherited nephrotic syndrome and dengue fever shock syndrome (8,9). To further explore the role of common variants in ESCC risk, we combined risk estimate data from the two studies and completed a meta-analysis.

RESULTS

Meta-analysis of known risk variants at 10q23

The joint data set included 2961 ESCC cases and 3400 controls, including 2024 cases and 2708 controls from one scan (6) and 937 cases and 692 controls from the other (7). In the joint data set, we examined associations at the previously reported susceptibility locus at 10q23 (Supplementary Material, Table S1); rs2274223, a nonsynonymous SNP in *PLCE1* that was independently reported in the previous GWAS,

showed a combined per allele odds ratio (OR) [95% confidence interval (CI)] of 1.39 (1.27–1.51) with $P = 1.44 \times 10^{-13}$. Six other SNPs also showed highly significant associations, with the strongest at rs3765524 with a *P*-value of 3.15×10^{-14}.

Meta-analysis of all hits with $P < 0.05$

Using the combined data set, we discovered an association at 2q33 that achieved genome-wide significance. We found five SNPs at this locus with $P < 5 \times 10^{-8}$ in the combined data set (Table 1). The strongest signal was rs13016963, with a combined OR (95%CI) of 1.29 (1.19–1.40) and $P = 7.63 \times 10^{-10}$. These SNPs are in high linkage disequilibrium (LD) and map to a region including *CASP8*, *ALS2CR12* and *TRAK2* (Fig. 1). In models conditioned on the most notable marker, rs10201587, the associations for the other five SNPs were attenuated, which suggests that our findings point to a single association signal.

Imputation analysis of variants at 2q33

The association between cancer risk and variants at the 2q33 locus, which includes *CASP8*, has been tested for many tumor types in over 50 studies (10,11). Since the SNPs tested and the genomic location associated varied by study, organ and histology, to better define our association signal, we imputed SNPs with the IMPUTE2 program (12) in the National Cancer Institute (NCI) scan using a hybrid reference of the 1000 Genomes Asian set and the Asian component of the Division of Cancer Epidemiology and Genetics (DCEG) Reference Imputation Set (13). We imputed 4304 SNPs in the 2q33 region with a mean certainty of 80.7% based on the information measure of the IMPUTE2 program and the SNPs associated with cancer risk with $P < 1 \times 10^{-5}$ that are listed in Supplementary Material, Table S2. The imputed SNP (rs6745435) with the strongest association signal was only marginally better than the strongest association for a genotyped SNP in the NCI scan (rs10201587) (Fig. 2). These

Table 1. The associations between SNPs at 2q33.1 and risk of ESCC in Chinese subjects

NCBI dbSNP identifier; NCBI genome build 36 (major, minor allele)	Location	NCI scan			China scan			Combined		$P_{\text{Heterogeneity}}$
		MAF: controls, cases	$P_{\text{1df score}}$	OR (95% CI) per allele	MAF: controls, cases	$P_{\text{1df score}}$	OR (95% CI) per allele	$P_{\text{1df score}}$	OR (95% CI) per allele	
rs10931936 (C,T)	201 852 173	0.269, 0.307	5.17E − 06	1.27 (1.17–1.38)	0.243, 0.303	1.44E − 04	1.36 (1.16–1.59)	4.74E − 09	1.27 (1.17–1.38)	0.36
rs13016963 (G,A)	201 871 056	0.272, 0.313	1.08E − 06	1.29 (1.19–1.40)	0.248, 0.306	1.16E − 04	1.36 (1.16–1.59)	7.63E − 10	1.29 (1.19–1.40)	0.43
rs9288318 (A, C)	201 904 308	0.268, 0.311	3.49E − 07	1.28 (1.18–1.39)	0.249, 0.302	9.13E − 04	1.30 (1.11–1.52)	1.35E − 09	1.28 (1.18–1.39)	0.83
rs10201587 (G, A)	201 911 036	0.270, 0.315	1.77E − 07	1.29 (1.19–1.40)	0.250, 0.302	1.21E − 03	1.29 (1.11–1.51)	8.71E − 10	1.29 (1.19–1.40)	0.93
rs7578456 (G, A)	201 943 593	0.240, 0.277	1.18E − 05	1.27 (1.17–1.38)	0.213, 0.270	2.08E − 04	1.36 (1.16–1.61)	1.60E − 08	1.27 (1.17–1.38)	0.34

SNPs were also in high LD and when we tested the 34 imputed SNPs in models conditioned on rs10201587 (Supplementary Material, Table S3), we found that all other SNP associations were attenuated, suggesting that these associations are from a single association signal.

Assessment of recombination hotspots across 2q33

The five genotyped SNPs that exceeded genome-wide significance localize between two recombination hotspots (positions 201819605 and 202211605, medians of 2 kb inferred hotspot intervals), which contain all SNPs with P-value $< 5 \times 10^{-8}$. The top SNP in the combined data (rs13016963) and two others (rs9288318 and rs10201587) are located in introns of *ALS2CR12* (amyotrophic lateral sclerosis 2 chromosomal, candidate 12) (Fig. 1). These are in high LD with rs10931936 ($r^2 = 0.982$) located in an intron of *CASP8* (caspase 8, apoptosis-related cysteine peptidase), and rs9288318 ($r^2 = 0.865$), rs10201587 ($r^2 = 0.874$) and rs7578456 ($r^2 = 0.734$), which are located in the region between *ALS2CR12* and *TRAK2*, adding further support to a single locus.

Ancestral recombination graph analyses at 2q33

We conducted an inferred ancestral recombination graph (ARG) analysis (14) to identify one or more haplotypes that harbor the variants directly responsible for the detected association signal. We used data from the NCI scan (4732 subjects and 53 genotyped SNPs) to identify the location with the most likely functional SNPs by reconstructing the genealogical history based on haplotypes inferred using the PHASE program (15). The Margarita program determines whether a possible mutation resides on the marginal tree by comparing the frequency of the branches between cases and controls. The most probable pair of haplotypes for each subject was selected for analysis in the Margarita program, and during this process subjects whose highest haplotype pair probability was < 50% were excluded from further analysis. We estimated 100 ARG genealogies and calculated permutation P-values (10^6 permutations) for each of the 53 SNPs. In a comparison of the location and strength of association from the ARG analysis with that from the standard GWAS P-values for each SNP (Fig. 3), both methods indicate that the strongest signal is localized to a region near rs10201587 (chr2: 201 911 036, hg18). The background association signal of the ARG analysis points to a haplotype containing the five SNPs and further distinguishes between risk haplotypes and protective haplotypes. In Figure 4, all haplotypes highlighted in green are predicted not to harbor the risk allele, whereas those in red are predicted to carry the risk allele. The ARG analysis showed that among all the marginal trees, rs10201587 had the lowest permutation P-value and has the greatest ability to segregate cases and controls, but four additional SNPs in near-perfect LD with rs10201587 (rs3769823, rs10931936, rs13016963 and rs9288318) provide the same discrimination when we restricted the data to haplotypes with a frequency of at least 1%. The complete separation of haplotypes with a frequency > 1% identified the five genotyped SNPs and 45 imputed SNPs in strong LD with an $r^2 > 0.8$.
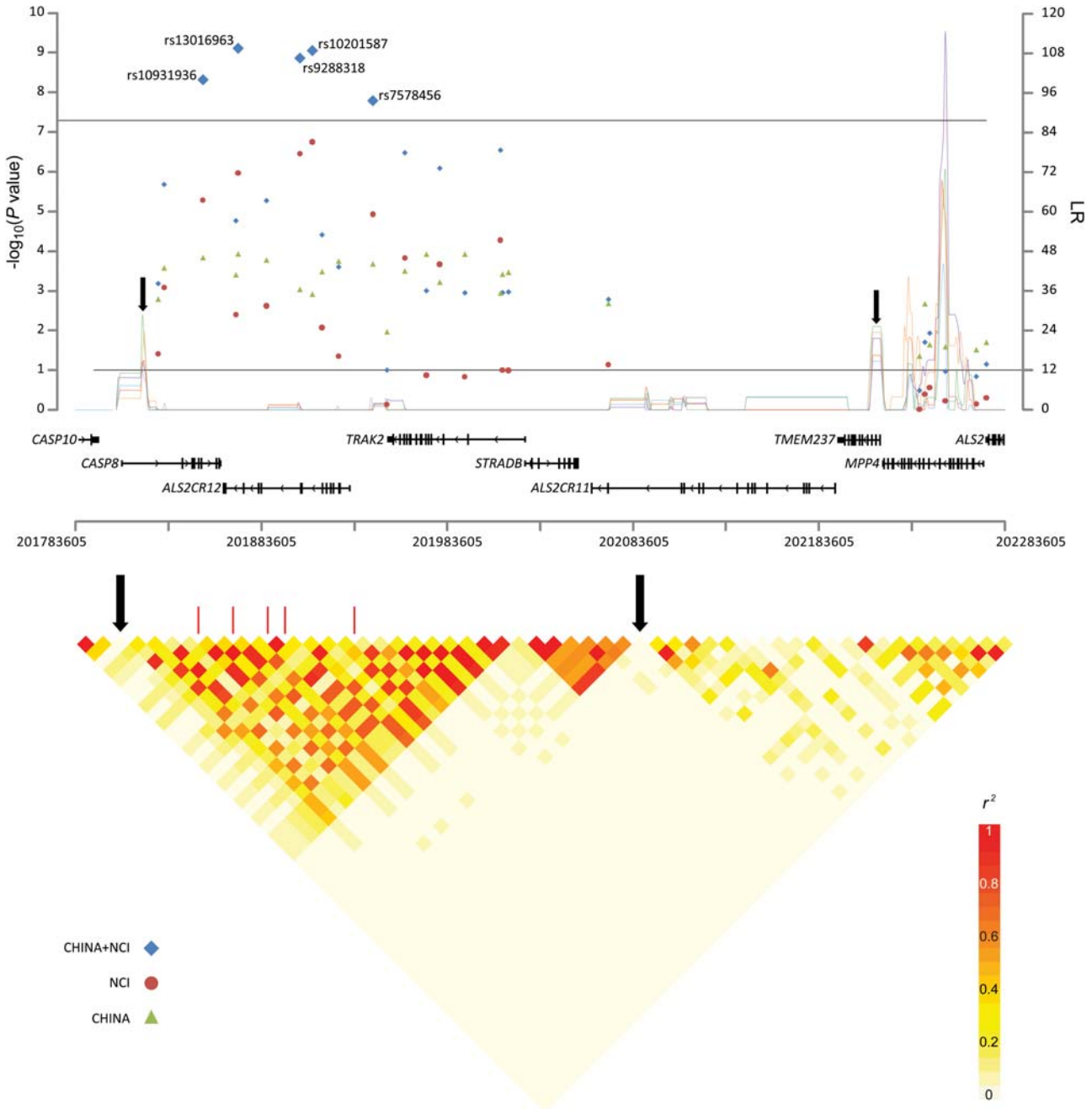
Figure 1. Association results, recombination and LD plots for the region of 2q33 with risk of ESCC. *P*-values derived from 1 df trend tests across a region of 2q33.1 bounded by rs12693932 and rs3731707, a distance of 473 kb, were plotted. The five colored-line graphs in the upper panel show likelihood ratio statistics (*Y*-axis on the right) for recombination hotspots from the SequenceLDhot software. The top horizontal line indicates a *P*-value of 5.0×10^{-8}, and the bottom horizontal line indicates a likelihood ratio statistic cut-off to predict the presence of a hotspot with a false-positive rate of 1 in 3700 independent tests (32). The five different colored lines represent five independent samplings used to estimate the location of the hotspots. The trend *P*-values from the Chinese, NCI and combined samples were plotted in green triangles, red circles and blue diamonds, respectively. The bottom panel depicts the LD pattern of the region in r^2, and solid black arrows indicate two flanking recombination hotspots containing five SNPs that exceed genome-wide significance (*P*-values $<5 \times 10^{-8}$). The short red vertical lines on the LD heat map indicate the locations of the five genome-wide significant SNPs. The sets of black arrows point to the two recombination hotspots determined in the randomly selected subsets of controls. Since the two panels are on different scales, each has a set of arrows to indicate the hotspots.

DISCUSSION

In this study, we combined results from two previous GWASs in Han Chinese and discovered a new association between highly correlated variants at 2q33 and risk of ESCC, which maps to the *CASP8/ALS2CR12/TRAK2* gene region. The strongest association was observed for rs13016963 [OR (95%CI) of 1.29 (1.19–1.40) and $P = 7.63 \times 10^{-10}$]. In an imputation analysis of over 4000 SNPs, no stronger signal was observed with ESCC risk. The notable SNPs are strongly
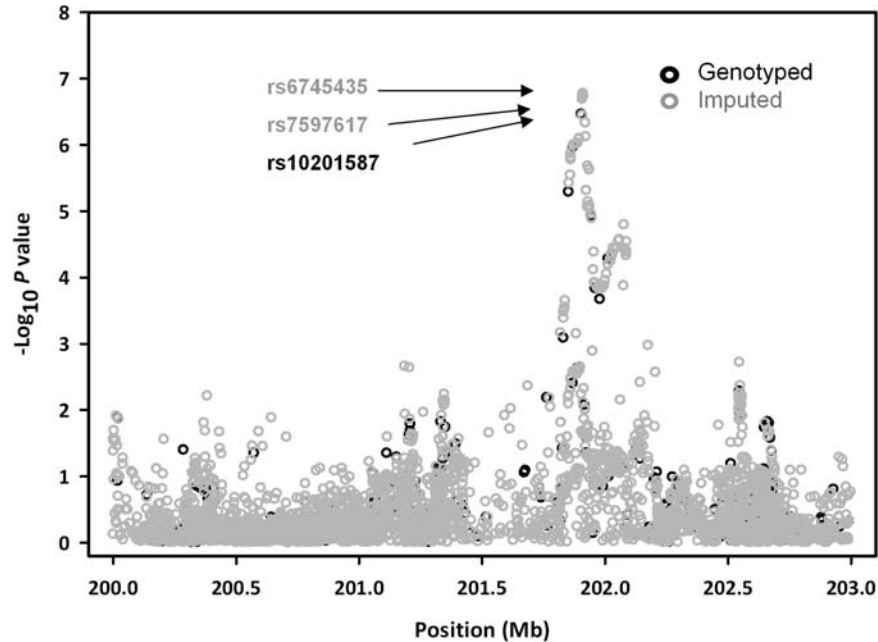
Figure 2. Comparison of results for genotyped (in black) and imputed (in grey) SNPs at 2q33 for their association with risk of ESCC. We plotted *P*-values for 426 genotyped SNPs and 3878 imputed SNPs. For the imputation, we used a hybrid reference of 1000 Genomes Asian set and the Asian component of the DCEG Reference Imputation Set. This figure shows that the genotyped SNPs were as strong as any SNPs in the imputation analysis.

correlated and reside within the boundaries of two recombination hotspots that we determined in the control subjects of the NCI GWAS. Since the conditional analysis did not preserve independent significance and in fact demonstrated substantial diminution of the tested signal, we conclude that there is one common allele on 2q33 associated with risk for ESCC.

We also used ARG analyses to further explore the region by reconstructing the genealogical history in this stable population to compare the frequency of branches of cases and controls to further localize the most likely functional variant(s). In the ARG analysis based on a permutation test, the results showed that rs10201587 had the strongest signal, but rs13016963 and three other SNPs were in near perfect LD with this SNP among the haplotypes tested. Based on the imputation analysis, there were 45 additional SNPs in strong LD, which could also be responsible for the direct association. Confirmation of the variants role will require additional study, including re-sequence analysis and functional studies of the optimal variants.

Numerous studies have investigated whether variation in the *CASP8* gene region alters cancer risk, including cancers of the lung (16), breast (17), pancreas (18), non-Hodgkin lymphoma (19), squamous head and neck cancers (20) and others (10,21). Although a recent GWAS showed that rs13016963 was significantly associated with risk of melanoma (22), most previous studies tested ostensibly functional variants in the *CASP8* promoter (-652 6N del, rs3834129), a variant that leads to an amino acid change in the *CASP8* gene product (D302H, rs1045485), or a base substitution in the 3′UTR (Ex14−271 A>T). Overall, these studies have produced mixed results, primarily due to their small size and inadequate replication efforts (10,21,23), but they did include one finding of a

significant association for rs3834129 with ESCC in Han Chinese (11). However, we found no evidence for an association between rs3834129 and risk of ESCC. rs6747918, which is highly correlated with rs3834129 in Han Chinese ($r^2 = 0.8$), had an OR (95% CI) of 1.06 (0.96−1.17) in our study ($P = 0.256$).

Two previous studies have investigated associations between risk of ESCC and other variants in the 3′ end of *CASP8* or further downstream in the region of *ALS2CR12* and *TRAK2*. A small study from our group suggested an association with rs1406121 in *ALS2CR12* (24), but this finding was not replicated in a sample of 300 ESCC cases and matched controls (25). In our current larger data set, a proxy for rs1406121 [rs7577057, $r^2 = 0.963$ in the Beijing Han Chinese (CHB) population] showed a nominal but not genome-wide significant association with an OR of 0.87 (0.81−0.93), $P = 0.00015$. This SNP has r^2 (0.284) with the strongest signal (rs13016963) seen in our meta-analysis.

In India, Umar *et al.* (23) reported an association between risk of ESCC and the IVS12−19G>A (rs3769818) variant in *CASP8* [per allele OR 3.36 (1.07−10.61), $P = 0.039$]. In our data, we did not genotype rs3769818 directly, but rs13016963 is in high LD with rs3769818 ($r^2 = 0.733$ in the NCI study population) and showed a strong association (Table 1). Furthermore, Umar suggested that the association was limited to men. In our data, we observed no significant difference in the association for men and women [OR_{men} (95% CI) = 1.30 (1.15−1.46), $P = 1.49 \times 10^{-5}$; OR_{women} (95% CI) = 1.20 (1.03−1.40), $P = 0.023$].

CASP8 encodes caspase-8, a cysteine-aspartic acid protease that in its mature form initiates apoptosis and in its immature form, procaspase-8, helps control cell migration and adhesion (26). Genetic variants altering caspase-8 expression or
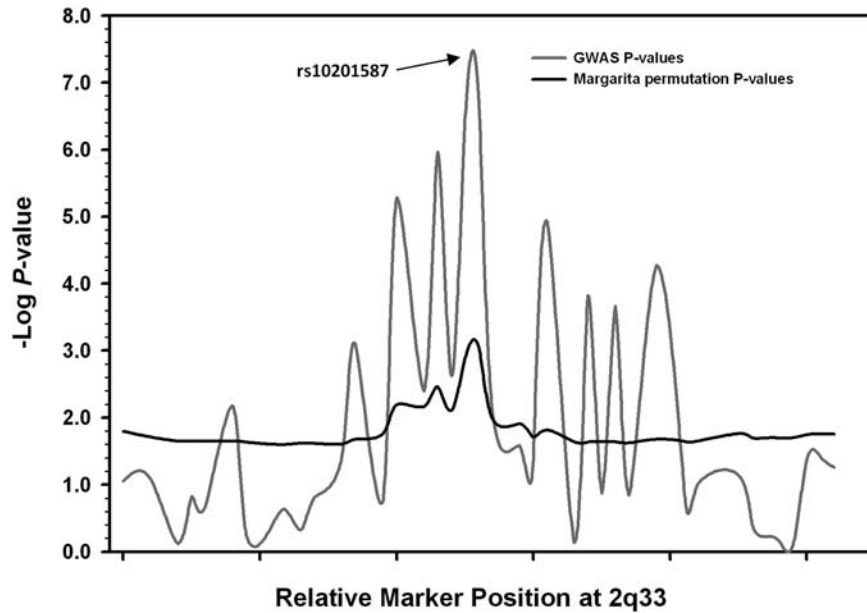
Figure 3. Comparison of GWAS and Margarita *P*-values for 40 SNPs at 2q33. The GWAS *P*-values are uncorrected, whereas the Margarita *P*-values (14) come from permutation tests and do not need correction for multiple comparisons. The figure includes data from 40 SNPs and extends from rs12470378 to rs10931959. Both methods show the strongest signal at rs10201587.

function have the potential to affect tumorigenesis through either of these pathways. Earlier work from our group showed loss of *CASP8* expression from both tumors and esophageal squamous dysplasia (27), the precursor lesion for ESCC. The association signal we see at 2q33 could also be attributed to other genes in the region, *ALS2CR12* or *TRAK2*. The ALS2CR12 protein is strongly expressed in normal squamous esophageal tissue (http://www.proteinatlas.org/ENSG00000155749/normal/

esophagus) and is named as a candidate gene for juvenile amyotrophic lateral sclerosis (28), but the connection to cancer is unclear. The SNPs associated with ESCC in this study could also be tagged to *TRAK2* (29).

Our meta-analysis identified an association between ESCC risk and a single locus at 2q33 that harbors a plausible candidate gene, *CASP8* in ESCC. This locus appears to play a role in the risk of other cancers, but the pattern of association between variants at this locus and different cancer types has varied widely in previous studies, similar to what has been observed for the *TERT/CLPTM1L* locus on 5p15 (30). Future studies of esophageal and other cancers should focus on comprehensive sequencing of this 2q33 locus and functional analysis of rs13016963 and rs10201587and other strongly correlated variants.

MATERIALS AND METHODS

Subjects and genotyping

The two studies contributing data to this meta-analysis have been described in detail in the original publications (6,7). After publication, additional GWAS data were available for 133 ESCC cases and 615 controls in one scan and these were included in this analysis (6). The newly genotyped additional data were processed by the Core Genotyping Facility GWAS pipeline and similar quality control (QC) filtering metrics were applied to ensure good quality data retained in the downstream analysis. We excluded the following: (i) samples with missing rates >6%; (ii) loci with missing rates >5%; (iii) samples with abnormal mean heterozygosity values of either >30 or 25%; (iv) gender discordant samples; (v) unexpected duplicate pairs. In total, this analysis used data on 2961 ESCC cases and 3400 controls, including 2024 cases and 2708 controls from one scan (vi) and 937 cases and 692 controls from the other (vii). The NCI scan used the Illumina 660W Quad array for genotyping, whereas the China scan used the Illumina 610 Quad array. The details of these methods and the quality assurance and QC metrics are available in the prior publications (6,7).

From each study, variants with a nominal *P*-value <0.05 from a two-sided linear trend test were tabulated and beta estimates and standard errors sent to the NCI analytic core. From the China scan, this constituted 29 971 SNPs, and from the NCI scan this constituted 26 177 SNPs.

We designed four TaqMan assays for rs10201587, rs13016963, rs7578456 and rs9288318, respectively, and genotyped a total of 340 samples randomly chosen from the Shanxi or Singapore studies, out of which 303 samples were previously genotyped on the Illumina 660W arrays and passed after genotyping QC filtering. The overall concordance rate was >99.7% between the TaqMan data and GWAS data for these four SNPs in the 303 samples.

Statistical analysis

We performed a fixed effect meta-analysis using the inverse variance method to estimate the combined ORs and 95% CIs. The *P*-value for heterogeneity was calculated by Cochran's

Figure 4. Haplotypes for 53 SNPs at 2q33 and risk of ESCC. Thirteen monomorphic SNPs among all the haplotypes with frequency >1% were excluded from the figure. Haplotype pairs for each subject were inferred using PHASE v2 (15) and used to generate 100 genealogies. ARG result for haplotypes with a frequency >1% containing five SNPs (rs10201587, rs3769823, rs10931936, rs13016963 and rs9288318) showed complete separation between those predicted to be associated with cancer (red) or not (green). All prediction frequencies were 0 or 1.0. rs10201587 was always the best segregator, but in these haplotypes the other four SNPs were in near perfect LD with this SNP. The ARG model prediction suggests that rs10201587 (allele frequency 28.9%) is the most likely SNP, or in strong LD with the functional variant.

Q, which is distributed as a χ^2 statistic with 1 degree of freedom (df).

Imputation

For the imputation reference set, we used a combination of 60 CHB+JPT subjects from 1000 genomes low coverage July 2010 data set (31), 29 additional HapMap CHB+JPT subjects and an internal imputation data set of 74 subjects scanned with the Illumina 2.5M chip (13). The IMPUTE2 program (12) was used with the recommended default settings to impute a 3 Mb window (200–203 Mb, hg18) on chr2, which encompasses our newly discovered locus. The association for the imputed SNPs was analyzed using SNPTEST (31) based on allelic dosage for the genotypic term, including adjustments for study, age, sex and the first principal component from the PCA of the population stratification SNPs.

Determination of recombination hotspots in 2q33

To identify recombination hotspots in the region, we used SequenceLDhot (32), a program that uses the approximate marginal likelihood method (33) and calculates likelihood ratio statistics at a set of possible hotspots. From the controls, 100 individuals were sampled five times without replacement

for five independent recombination hotspot analyses and these five random samples are represented as five different colored lines in Figure 1. Specifically, genotypes of 90 SNPs spanning chr2: 201 432 605–202 408 291 were phased using PHASE v2.1 (15) to calculate background recombination rates. The PHASE outcome was used as direct input for the SequenceLDhot program. For the plot, the likelihood ratio statistic values between 201783605 and 202283605 were plotted. The reason we tested a wider range than the plot was to capture flanking hotspots that contained all highly significant signals. The LD was calculated as r^2 for 55 SNPs that were genotyped in both data sets within an ~473 kb region bounded by rs12693932 and rs3731707 (chr2: 201 801 640–202 275 009, UCSC genome build hg18), and a heat map was drawn using the snp.plotter program (34).

ARG analyses

The genotype data from the NCI scan for 4732 individuals on 53 loci in the 2q33 region were analyzed with the PHASE v2.1.1 program (15) to statistically infer all probable pairs of haplotypes for each individual. The most probable pair of haplotypes for each subject was selected for analysis in the Margarita program (14). A total of 141 (<3%) individuals with ambiguous phasing and a highest haplotype pair probability

of <50% were excluded; 4591 (1974 cases and 2617 controls) remained in the downstream ARG analysis. An ensemble of 100 ARGs was inferred and 1 000 000 permutations were done for the best cut (determined by the allelic test *P*-value) at each marginal tree to estimate the *P*-value at each locus. Haplogroups with frequency >1% were categorized into protective or at-risk groups.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

Conflict of Interest statement. The authors have no conflicts of interest to declare.

FUNDING

REFERENCES

1. Parkin, D.M., Bray, F., Ferlay, J. and Pisani, P. (2005) Global cancer statistics, 2002. *CA Cancer J. Clin.*, **55**, 74–108.
2. Chen, W.-Q., Zhang, S.-W., Zou, X.-N. and Zhao, P. (2011) Cancer incidence and mortality in China, 2006. *Chin. J. Cancer Res.*, **23**, 3–9.
3. Brown, L.M., Devesa, S.S. and Chow, W.H. (2008) Incidence of adenocarcinoma of the esophagus among white Americans by sex, stage, and age. *J. Natl Cancer Inst.*, **100**, 1184–1187.
4. Tran, G.D., Sun, X.D., Abnet, C.C., Fan, J.H., Dawsey, S.M., Dong, Z.W., Mark, S.D., Qiao, Y.L. and Taylor, P.R. (2004) Prospective study of risk factors for esophageal and gastric cancers in the Linxian general population trial cohort in China. *Int. J. Cancer*, **113**, 176–181.
5. Gao, Y., Hu, N., Han, X., Giffen, C., Ding, T., Goldstein, A. and Taylor, P. (2009) Family history of cancer and risk for esophageal and gastric cancer in Shanxi, China. *BMC Cancer*, **9**, 269.
6. Abnet, C.C., Freedman, N.D., Hu, N., Wang, Z., Yu, K., Shu, X.O., Yuan, J.M., Zheng, W., Dawsey, S.M., Dong, L.M. *et al.* (2010) A shared susceptibility locus in PLCE1 at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. *Nat. Genet.*, **42**, 764–767.
7. Wang, L.D., Zhou, F.Y., Li, X.M., Sun, L.D., Song, X., Jin, Y., Li, J.M., Kong, G.Q., Qi, H., Cui, J. *et al.* (2010) Genome-wide association study of esophageal squamous cell carcinoma in Chinese subjects identifies susceptibility loci at PLCE1 and C20orf54. *Nat. Genet.*, **42**, 759–763.
8. Hinkes, B., Wiggins, R.C., Gbadegesin, R., Vlangos, C.N., Seelow, D., Nurnberg, G., Garg, P., Verma, R., Chaib, H., Hoskins, B.E. *et al.* (2006) Positional cloning uncovers mutations in PLCE1 responsible for a nephrotic syndrome variant that may be reversible. *Nat. Genet.*, **38**, 1397–1405.
9. Khor, C.C., Chau, T.N., Pang, J., Davila, S., Long, H.T., Ong, R.T., Dunstan, S.J., Wills, B., Farrar, J., Van Tram, T. *et al.* (2011) Genome-wide association study identifies susceptibility loci for dengue shock syndrome at MICB and PLCE1. *Nat. Genet.*, **436**, 1139–1141.
10. Yin, M., Yan, J., Wei, S. and Wei, Q. (2010) CASP8 polymorphisms contribute to cancer susceptibility: evidence from a meta-analysis of 23 publications with 55 individual studies. *Carcinogenesis*, **31**, 850–857.
11. Sun, T., Gao, Y., Tan, W., Ma, S., Shi, Y., Yao, J., Guo, Y., Yang, M., Zhang, X., Zhang, Q. *et al.* (2007) A six-nucleotide insertion-deletion polymorphism in the CASP8 promoter is associated with susceptibility to multiple cancers. *Nat. Genet.*, **39**, 605–613.
12. Howie, B.N., Donnelly, P. and Marchini, J. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.*, **5**, e1000529.
13. Wang, Z., Jacobs, K.B., Yeager, M., Hutchinson, A., Sampson, J., Chatterjee, N., Berndt, S.I., Chung, C.C., Diver, W.R., Gapstur, S.M. *et al.* (2012) Improved imputation of common and uncommon single nucleotide polymorphisms (SNPs) with a new reference set. *Nat. Genet.*, **44**, 6–7.
14. Minichiello, M.J. and Durbin, R. (2006) Mapping trait loci by use of inferred ancestral recombination graphs. *Am. J. Hum. Genet.*, **79**, 910–922.
15. Stephens, M., Smith, N.J. and Donnelly, P. (2001) A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.*, **68**, 978–989.
16. Son, J.W., Kang, H.K., Chae, M.H., Choi, J.E., Park, J.M., Lee, W.K., Kim, C.H., Kim, D.S., Kam, S., Kang, Y.M. and Park, J.Y. (2006) Polymorphisms in the caspase-8 gene and the risk of lung cancer. *Cancer Genet. Cytogenet.*, **169**, 121–127.
17. Cox, A., Dunning, A.M., Garcia-Closas, M., Balasubramanian, S., Reed, M.W., Pooley, K.A., Scollen, S., Baynes, C., Ponder, B.A., Chanock, S. *et al.* (2007) A common coding variant in CASP8 is associated with breast cancer risk. *Nat. Genet.*, **39**, 352–358.
18. Couch, F.J., Wang, X., McWilliams, R.R., Bamlet, W.R., de Andrade, M. and Petersen, G.M. (2009) Association of breast cancer susceptibility variants with risk of pancreatic cancer. *Cancer Epidemiol. Biomarkers Prev.*, **18**, 3044–3048.
19. Lan, Q., Morton, L.M., Armstrong, B., Hartge, P., Menashe, I., Zheng, T., Purdue, M.P., Cerhan, J.R., Zhang, Y., Grulich, A. *et al.* (2009) Genetic variation in caspase genes and risk of non-Hodgkin lymphoma: a pooled analysis of 3 population-based case-control studies. *Blood*, **114**, 264–267.
20. Li, C., Lu, J., Liu, Z., Wang, L.E., Zhao, H., El-Naggar, A.K., Sturgis, E.M. and Wei, Q. (2010) The six-nucleotide deletion/insertion variant in the CASP8 promoter region is inversely associated with risk of squamous cell carcinoma of the head and neck. *Cancer Prev. Res. (Phila.)*, **3**, 246–253.
21. Haiman, C.A., Garcia, R.R., Kolonel, L.N., Henderson, B.E., Wu, A.H. and Le, M.L. (2008) A promoter polymorphism in the CASP8 gene is not associated with cancer risk. *Nat. Genet.*, **40**, 259–260.
22. Barrett, J.H., Iles, M.M., Harland, M., Taylor, J.C., Aitken, J.F., Andresen, P.A., Akslen, L.A., Armstrong, B.K., Avril, M.F., Azizi, E. *et al.* (2011) Genome-wide association study identifies three new melanoma susceptibility loci. *Nat. Genet.*, **43**, 1108–1113.
23. Umar, M., Upadhyay, R., Kumar, S., Ghoshal, U.C. and Mittal, B. (2011) CASP8 -652 6N del and CASP8 IVS12−19G>A gene polymorphisms and susceptibility/prognosis of ESCC: a case control study in northern Indian population. *J. Surg. Oncol.*, **103**, 716–723.

24. Hu, N., Wang, C., Hu, Y., Yang, H.H., Giffen, C., Tang, Z.Z., Han, X.Y., Goldstein, A.M., Emmert-Buck, M.R., Buetow, K.H. *et al.* (2005) Genome-wide association study in esophageal cancer using GeneChip mapping 10K array. *Cancer Res.*, **65**, 2542–2546.

25. Ng, D., Hu, N., Hu, Y., Wang, C., Giffen, C., Tang, Z.Z., Han, X.Y., Yang, H.H., Lee, M.P., Goldstein, A.M. and Taylor, P.R. (2008) Replication of a genome-wide case-control study of esophageal squamous cell carcinoma. *Int. J. Cancer.*, **123**, 1610–1615.

26. Zhao, Y., Sui, X. and Ren, H. (2010) From procaspase-8 to caspase-8: revisiting structural functions of caspase-8. *J. Cell. Physiol.*, **225**, 316–320.

27. Xue, L.Y., Hu, N., Song, Y.M., Zou, S.M., Shou, J.Z., Qian, L.X., Ren, L.Q., Lin, D.M., Tong, T., He, Z.G. *et al.* (2006) Tissue microarray analysis reveals a tight correlation between protein expression pattern and progression of esophageal squamous cell carcinoma. *BMC Cancer*, **6**, 296.

28. Hadano, S., Hand, C.K., Osuga, H., Yanagisawa, Y., Otomo, A., Devon, R.S., Miyamoto, N., Showguchi-Miyata, J., Okada, Y., Singaraja, R. *et al.* (2001) A gene encoding a putative GTPase regulator is mutated in familial amyotrophic lateral sclerosis 2. *Nat. Genet.*, **29**, 166–173.

29. Grishin, A., Li, H., Levitan, E.S. and Zaks-Makhina, E. (2006) Identification of gamma-aminobutyric acid receptor-interacting factor 1 (TRAK2) as a trafficking factor for the K+ channel Kir2.1. *J. Biol. Chem.*, **281**, 30104–30111.

30. Chung, C.C. and Chanock, S.J. (2011) Current status of genome-wide association studies in cancer. *Hum. Genet.*, **130**, 59–78.

31. 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**, 1061–1073.

32. Fearnhead, P. (2006) SequenceLDhot: detecting recombination hotspots. *Bioinformatics*, **22**, 3061–3066.

33. Fearnhead, P. and Donnelly, P. (2002) Approximate likelihood methods for estimating local recombination rates. *J. Royal Stat. Soc. B (Stat. Methodol.)*, **64**, 657–680.

34. Luna, A. and Nicodemus, K.K. (2007) snp.plotter: an R-based SNP/haplotype association and linkage disequilibrium plotting package. *Bioinformatics*, **23**, 774–776.