Joint analysis of three genome-wide association studies of esophageal squamous cell carcinoma in Chinese populations

A full list of authors and affiliations appears at the end of the article.

[#] These authors contributed equally to this work.

Abstract

We conducted a joint (pooled) analysis of three genome-wide association studies (GWAS) [1-3] of esophageal squamous cell carcinoma (ESCC) in ethnic Chinese (5,337 ESCC cases and 5,787 controls) with 9,654 ESCC cases and 10,058 controls for follow-up. In a logistic regression model adjusted for age, sex, study, and two eigenvectors, two new loci achieved genome-wide significance, marked by rs7447927 at 5q31.2 (per-allele odds ratio (OR) = 0.85, 95% CI 0.82-0.88; *P*=7.72x10[−20]) and rs1642764 at 17p13.1 (per-allele OR= 0.88, 95% CI 0.85-0.91; *P*=3.10x10[−13]). rs7447927 is a synonymous single nucleotide polymorphism (SNP) in *TMEM173* and rs1642764 is an intronic SNP in *ATP1B2,* near *TP53*. Furthermore, a locus in the HLA class II region at 6p21.32 (rs35597309) achieved genome-wide significance in the two populations at highest risk for ESSC (OR=1.33, 95% CI 1.22-1.46; *P*=1.99x10[−10]). Our joint analysis identified new ESCC susceptibility loci overall as well as a new locus unique to the ESCC high risk Taihang Mountain region.

[**]Correspondence should be addressed to: Dr. Li-Dong Wang (ldwang@zzu.edu.cn); Dr. Philip Taylor (ptaylor@mail.nih.gov); Dr. Dongxin Lin (lindx72@cicams.ac.cn)..

Esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma are distinct diseases with different etiologies. ESCC remains the more common type of esophageal carcinoma in economically-developing countries as well as globally. Approximately half of the world's 500,000 new ESCC cases annually occur in China where the disease is a major public health problem. Three genome wide association studies (GWAS) have examined ESCC[1-3] and two subsequent analyses used combinations of the three studies[4, 5] to report as many as 12 loci associated with ESCC risk. Four additional loci have been reported based on an interaction with alcohol consumption, a known risk factor for ESCC[4]. Two of the GWAS[1, 2] drew subjects primarily from the Taihang Mountain region of Henan and Shanxi provinces where ESCC occurs at very high rates. Total mortality due to ESCC and gastric cardia cancer can exceed 20% in the highest risk communities in this region[6]. The third GWAS[3] drew subjects from a range of locations in China including a substantial collection from Beijing, where the population is comprised of people who originate from all provinces. The contribution of lifestyle risk factors for ESCC differs widely between locations in China, with alcohol consumption being a notable example. Heavy consumption of alcoholic beverages is a known cause of ESCC[7], but historically was uncommon in the high incidence regions of China[8]. Therefore, alcoholic beverages played little role in the extraordinarily high rates, leaving the high incidence largely unexplained. To discover additional novel ESCC susceptibility alleles, we conducted a joint analysis of the three original GWAS and followed promising signals in an independent set of new cases and controls.

Supplementary Table 1 describes the subjects from the three underlying GWAS, which include additional subjects that were scanned after the first round publications plus the additional subjects used for replication of the top hits. In our joint analysis we investigated for the first time the individual GWAS data drawn from three studies in China, which consisted of 5,337 ESCC cases and 5,787 controls of Han Chinese ethnicity. The NCI GWAS used the Illumina 660W-Quad SNP microarray, the Henan GWAS used the Illumina 610-Quad SNP microarray, while the Beijing GWAS used the Affymetrix GeneChip Human Mapping 6.0 set. **Supplementary Figure 1** shows eigenvector plots from a principal components analysis (PCA) of 25,676 uncorrelated genotyped SNPs ($r^2 < 0.01$ in our combined control set). The results show distinct clusters corresponding to different source populations. However, only the first two eigenvectors were significant in the base cancer risk model and were thus used to adjust for population stratification in the final estimates for our joint analysis. To explore untyped common variants, we assessed variants based on the imputation of 40.5 million variants using version 3 of the 1000 Genomes data as the reference for each of the three data sets before combining them as described in the **ONLINE METHODS and Supplementary Table 2**. After filtering out SNPs with MAF < 1% or imputation information criteria (INFO) < 0.3, we advanced 7,556,215 SNPs to the association analysis. The inflation factor λ for the joint analysis is 1.01 for all SNPs, which indicates that population stratification should not be a concern. **Supplementary Figure 2** provides a quantile-quantile plot for the joint case-control comparison with all SNPs and after exclusion of SNPs within 500 Kb of loci previously reported to be associated with ESCC. **Supplementary Figures 3-5** provide individual quantile-quantile plots for each of the three underlying GWAS data sets. Fourteen promising SNPs based on the joint analysis

(**see ONLINE METHODS**) were genotyped in an additional 9,654 ESCC cases and 10,058 controls divided between subjects from Henan Province and Beijing.

The joint analysis identified two new genome-wide significant loci at 5q31.2 (**Figure 1a**) and 17p13.1 (**Figure 1b**) associated with risk of ESCC in the pooled data of individual genotypes of all three GWAS and two replication studies (**Table 1 and Supplementary Table 3**). At 5q31.2, rs7447927, a synonymous SNP located in transmembrane protein 173 (*TMEM173),* had a combined per allele odds ratio (OR) and 95% confidence interval (95% CI) of 0.85 (0.82-0.88), P=7.72x10^{-20}. At 17p13.1, rs1642764, an intronic SNP located in ATPase, Na+/K+ transporting, beta 2 polypeptide (*ATP1B2)*, which is just telomeric to the tumor suppressor gene, *TP53,* had a combined per allele OR (95%CI) of 0.88 (0.85-0.91), P=3.10x10^{-13}. No statistically significant heterogeneity was observed across the three pooled GWAS and two replication studies for either locus (**Table 1**). Furthermore, the associations were confirmed independently in the follow-up sets collected from both high- and low-risk populations.

On further analysis, we observed an additional susceptibility locus that showed geographic differences such that the significant association was observed only in the two GWAS[1, 2] which included subjects from populations at the highest risk for ESCC. In joint analyses using all three GWAS, a SNP at 6p21.32 showed a nearly genome-wide significant association, however, statistically significant heterogeneity among studies (*P*=0.015) was evident when subjects from the Beijing GWAS and replication subjects were included (i.e., among the three pooled GWAS and two replication studies as shown in **Table 1**). The test for heterogeneity became non-significant when the Beijing GWAS and Beijing replication subjects were excluded (**Table 1**). **Table 1** also shows that when the analyses were restricted to subjects from the high incidence regions, rs35597309, located in the HLA Class II gene region between *HLA-DRB1* and *HLA-DQA1*, had a per allele OR (95% CI) of 1.33 (1.22-1.46), *P*=1.99 x 10^{-10} (**Supplementary Figure 6**). This heterogeneity between high- and low-risk regions was also evident when data from the three separate GWAS were examined (**Supplementary Table 4**). Further genotyping and possibly sequencing are necessary to map the susceptibility alleles across the HLA Class II region due to its complex structure defined by long-range haplotypes.

Finally, our joint analysis observed a promising association with rs61271866 (*P*=5.18 x 10^{-8}), an intergenic SNP at 9p21.3 that includes the cyclin-dependent kinase inhibitor 2B (*CDKN2B*)-*CDKN2A* gene cluster (**Supplementary Table 3**). Variants in this region have been associated with risk of melanoma[9], childhood acute lymphoblastic leukemia[10], chronic lymphocytic leukemia[11], and glioma[12], as well as ESCC[13] in a prior study using a subset of the samples examined here. This SNP showed heterogeneity among individual studies in the GWAS (**Supplementary Table 4**) and between the joint Stage 1 and replication phase results (**Supplementary Table 3**). The heterogeneity picture for the 9p21.3 locus is more varied as compared to the 6p21.32 HLA Class II locus, indicating that validation of this finding will require additional work.

For these four SNPs (rs7447927, rs1642764, rs35597309, and rs61271866), we tested for interactions by use of alcohol (**Supplementary Table 5**) or tobacco (**Supplementary Table**

6). Because of the substantial differences in the degree of alcohol and tobacco use by population, we did these tests separately for each of the three underlying studies. In total this constituted 24 tests and we found one that was nominally significant (rs35597309 and tobacco in the NCI GWAS, **Supplementary Table 6**), but the difference did not replicate in the two other studies and thus, is most likely due to chance. We further note that the number of GWAS subjects with these covariate data from Henan was limited.

Here we report a new finding of an association between rs7447927, a synonymous SNP in *TMEM173*, and ESCC risk. *TMEM173* (also known as *STING*) facilitates innate immune reactions to viruses and bacteria through the production of type 1 interferon[14]. The only previous GWAS hit at this locus (rs13181561) was demonstrated to be associated with the modulation of interferon-α responses to the smallpox vaccine[15] and is highly correlated (r^2=0.956 in 1000 Genomes data for CHB population) with rs7447927. rs13181561 is also tagged as an expression quantitative trait locus in lymphoblastoid cells (**Supplementary Table 7**) that alters expression of genes in segment AC135457.2. This genomic region includes the sodium-dependent vitamin C transporter (*SLC23A1*), which is critical for vitamin C transport[16]. Interestingly, low vitamin C has been implicated in risk of ESCC[17].

The new susceptibility locus at 17p13.1 is marked by rs1642764, an intronic SNP located within the *ATP1B2* gene; this variant resides in an LD block that includes the 3' region of *TP53* (**Figure 1b**). Alteration of *TP53* regulation or function could be a plausible explanation for the observed association between rs1642764 and risk of ESCC. It is noteworthy that no prior cancer GWAS has reported a conclusive association with a variant in or around *TP53*. Recently, a candidate gene study of genotyped and imputed SNPs across *TP53* reported a strong association of a SNP with a low minor allele frequency, rs78378222, with glioma risk (P=6.86 x 10^{-24} with a MAF=0.013 in a population of European ancestry)[18]. We also observe that *TP53* is frequently inactivated in ESCC[19]. Our target SNP in this region is in LD with several other SNPs, notably rs1050541 (r^2=0.575), which alters a binding site for *RAD21* (**Supplementary Table 7**). *RAD21* is a key component of the cohesion complex, which binds to DNA and is essential for mitosis, homologous DNA repair, and enhancer activities and may be relevant to cancer[20]. Alternatively, this SNP falls between recombination hotspots that includes the *SHBG* gene, and variants bracketing this SNP are known to be associated with sex hormone binding globulin regulation and serum testosterone concentration[21]. Esophageal cancer is male predominant in low incidence populations, but this has typically been attributed to greater tobacco smoking and alcohol consumption by men compared to women. Recent studies have suggested that hormonal factors may play a role in the development of ESCC[22, 23].

In published GWAS, HLA Class II genetic variants in close proximity (<20 kb) of rs35597309 on 6p21.32 have been associated with multiple cancers including nasopharyngeal carcinoma[24], hepatocellular carcinoma[25], lung cancer in never smokers[26], and familial chronic lymphocytic leukemia[27] as well as autoimmune diseases, including Crohn's disease[28] and lupus [29]. This SNP resides between *HLA-DRB1* and *HLADQA1*, both MHC class 2 genes that function in antigen presentation. The HLA region is large, complex, and shows unusually long-range LD that makes the interpretation of GWAS hits in this region difficult. rs35597309 is in perfect LD (r^2=1.0) with 34 other SNPs within 2 MB

(**Supplementary Table 7**) including two missense mutations in *HLADQA1* and a host of putative protein binding sites, enhancers, and regulatory motifs. But the LD with top hits in this region from previously reported studies of cancer was low; $r^2 < 0.1$ for rs2860580 (nasopharyngeal carcinoma), rs9272105 (hepatocellular carcinoma), rs2395185 (lung cancer), and rs674313 (chronic lymphocytic leukemia) in 1000 Genomes data for CHB+JPT populations.

Our results provide evidence for an association between variants at 6p21.32 and risk of ESCC, although the association was restricted to the studies which examined subjects from the high incidence Taihang Mountain region. While it is plausible that our results could be due to a gene-environment interaction between HLA genotypes and an uncharacterized risk factor specific to the Taihang Mountains, it is also possible that chance could account for this finding. The result was, however, confirmed in the Henan replication set (*P*=0.001). We know that populations in this region suffer from very high rates of ESCC, which is likely multi-factorial and could involve immune challenges. It is also important to note that the Han Chinese population is genetically diverse and this difference in association may be due to true differences in the genomic structure of the HLA regions between subjects from the Taihang Mountains and other parts of China.

But we note that the allele frequency is similar between the Henan and Beijing replication sets (**Table 1**). Future work should use genomic methods specifically designed to investigate the HLA region to explore these results. We also note that a region at 6p22.1 that is linked to other HLA alleles has been associated with risk of Barrett's esophagus, the precursor lesion for esophageal adenocarcinoma[30].

Supplementary Table 8 shows the individual GWAS and joint analysis results for the 12 main effect GWAS loci reported in five previous publications[1-5]. Joint analysis of the pooled data from the three GWAS did not strengthen all previously reported loci. We observed strong associations for SNPs in loci harboring *PLCE1, CASP8, RUNX1,* and *CHEK2,* but no signal for four loci. The Beijing GWAS showed geographic differences in associations for some SNPs[3] and this variation may be attributed, in part, to differences in environmental exposures and habits, such as alcohol consumption. A previous analysis by Wu *et al.* [4] using partial data from two of the studies also showed substantial heterogeneity between GWAS. Of 18 hits identified in analyses of both main effects and alcohol interactions in the Beijing data, only four replicated in subjects from the Shanxi Upper Gastrointestinal Cancer Genetics Study[1].

In conclusion, we present the joint analysis of the individual genotype data from three previously published GWAS in China and have established two new loci associated with risk of ESCC across all three studies, and two promising signals, the most notable, in the HLA class II region. The latter appears to be present only in subjects from the high incidence Taihang Mountain region of China. Environmental factors have previously been shown to be of varied relevance for ESCC in different Chinese populations and this may have led to differential GWAS findings. Here we find additional evidence for distinct results among these populations. Etiologic heterogeneity may play an important role in interpreting GWAS results and should be considered as GWAS are extended to understudied populations

with distinct lifestyles. Lastly, further work is needed to fine-map the regions to identify the optimal alleles for laboratory studies that will provide an understanding of the basic biology underlying the ESCC susceptibility alleles and their interaction with environmental factors.

ONLINE METHODS

Subject selection and genotyping

This study pooled the individual genotype data of subjects from three independent GWAS of esophageal squamous cell carcinoma in Han Chinese populations, which were part of three earlier reports from the NCI[1], Henan[2], and Beijing study groups[3]. The numbers of subjects differs in some cases from those listed in the original publications because additional subjects were genotyped using the same platform subsequent to the original publications; these subjects were included in the joint analysis of the individual genotype data. For the NCI GWAS, subjects came from four prospective cohort studies and one large case-control study as reported in Abnet et al.[1] and all subjects used in replication in the original paper were subsequently genotyped using the Illumina 660W-Quad microarray. For the Henan scan, subjects were collected from many hospitals in Henan Province and a smaller 'genetically-matched' subset[2] (1,076 cases, 713 controls) was selected for this joint analysis. Furthermore, 299 cases and 370 controls were added who had subsequently been genotyped using the Illumina 610-Quad SNP microarray *after* the first publication. For the Beijing scan, subjects were collected from four different localities as previously reported in Wu et al.[4] and all subjects were included in the current study. These subjects had been genotyped using the Affymetrix GeneChip Human Mapping 6.0 set. Subjects included in the replications from both Henan[2] and Beijing[3] were identified and recruited using the same approaches as for subjects included in the GWAS from each of these respective sites and as described in the initial publications from each of their GWAS. A description of the included subjects is given in **Supplementary Table 1**.

GWAS data

The details of the analytic pre-processing for the three GWAS were included in each of the primary papers. In addition to the quality control procedures performed in the previous primary publications for all three studies, SNPs with a call rate < 95% or Hardy-Weinberg proportion test *P*-value < 0.000001 or minor allele frequency < 1% were further removed prior to imputation for the current analysis (**Supplementary Table 2**). We also searched for potential duplicates or first degree relatives across all three GWAS with glu ibds module (http://code.google.com/p/glu-genetics/) using the set of 25,676 independent SNPs with pair-wise $r^2 < 0.01$ estimated from the GWAS control population. A total of nine pairs of duplicates were found; all were between Henan[2] and the NIT component of the NCI scan[1], which enrolled subjects in the same high incidence area of Henan Province. As a result, we excluded nine individuals (one from each pair) from the Henan study for the joint analysis. Application of standardized QC procedures for subjects and for SNPs resulted in the exclusion of an additional small proportion of subjects such that the final numbers of subjects in the current analysis are slightly different from those reported previously.

Imputation analysis

Imputation was conducted separately for each scan using IMPUTE2 software version 2.2.2 (http://mathgen.stats.ox.ac.uk/impute/impute_v2.html) and version 3 of the 1000 Genomes Project data as the reference set. First, the genomic coordinates were lifted over from NCBI human genome build 36 to build 37 using the UCSC lift over tool (http://hgdownload.cse.ucsc.edu/downloads.html). The few loci that failed to be lifted over were also excluded from the imputation. Second, the strand of the inference data was aligned with the 1000 Genomes data by simple allele state comparison or allele frequency matching for A/T and G/C SNPs. We implemented a 4-Mb sliding window to impute across the genome, resulting in 744 jobs running in parallel on the NIH BIOWULF cluster (http://biowulf.nih.gov/). A pre-phasing strategy with SHAPEIT software version 1 (http://www.shapeit.fr/) was adopted to improve the imputation performance. The phased haplotypes from SHAPEIT were fed directly into IMPUTE2. Imputed loci with INFO < 0.3 or MAF < 0.01 were excluded from further association analysis. To technically validate our imputation findings, we optimized TaqMan assays for rs7447927, rs1642764, and rs35597309. We genotyped a set of 892 samples from NCI and another set of 752 samples from Beijing. For the NCI set, the concordance rates between the imputed genotypes (using a posterior probability threshold of 0.95) and TaqMan genotypes were 99.3%, 96.5%, and 99.1%, respectively and for the Beijing set, the concordance rates between the imputed genotypes (using a posterior probability threshold of 0.95) and TaqMan genotypes were 96.7%, 90.4%, and 99.6%, respectively.

Association analysis

The imputed genotypes were merged using GTOOL software version 0.7.5 (http://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html) and the association testing was performed using SNPTEST software version 2.2 (https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html), with adjustment for age, sex, study, and the top two eigenvectors, which controls for population stratification. In the joint analysis baseline model (not including SNP effects) adjusting for age, sex, study, and all top ten eigenvectors (EVs), the top two eigenvectors were significantly associated with case status (P-value < 0.05), and were, therefore, included in the final joint analysis association models used to test SNP effects across all three studies. In a sensitivity analysis, we also conducted the association analyses in each of the three GWAS separately, followed by meta-analyses that used the fixed effect inverse variance method to combine the beta estimates and standard errors from each scan. In this second approach, we generated the set of eigenvectors based on each GWAS and identified significant eigenvectors to control for population stratification in each individual GWAS (EV1 for NCI; EV1, EV5 and EV7 for Beijing; no EV was needed for Henan). The meta-analysis (**Supplementary Table 4**) produced very similar results to our current joint analysis (**Table 1**), so we presented results for Stage 1 from the joint analysis as our primary analysis. The P for heterogeneity was calculated using Cochran's Q, which is distributed as a chi-square statistic with (n-1) degrees of freedom where n is the number of sets included in the meta-analysis. For exploring the gene and environment interaction with use of alcohol or tobacco, we performed stratified analyses for the four novel SNP associations and assessed the risk heterogeneity between drinkers and

nondrinkers (**Supplementary Table 5**), and between smokers and nonsmokers (**Supplementary Table 6**). To evaluate population stratification, we examined QQ plots before and after eigenvector adjustment for the joint analysis (**Supplementary Figure 7**), for Beijing (**Supplementary Figure 8**), and for NCI (**Supplementary Figure 9**). No figure is shown for Henan as no adjustment was required. Further, we examined the association with risk for the four novel SNP associations we report here before and after eigenvector adjustment (**Supplementary Table 9**).

Recombination hotspot inference

Likelihood ratio statistics for recombination hotspots were estimated by SequenceLDhot software based on background recombination rates inferred by PHASE v2.1 using the 1000 Genomes CHB data.

Replication genotyping and analysis

After SNPs from previously reported ESCC risk loci were excluded, the top SNPs with *p* values less than 1.0×10^{-5} (n=14) from our Stage 1 analysis were selected for replication testing in both Beijing and Henan (**Supplementary Table 3**). However, when the imputation was updated with the addition of more covariate data on subjects, the new Stage 1 *p*-value for rs4252725 was only 1.0×10^{-4}. At that point, primers for those top 14 SNPs from the initial analysis had already been designed and validated, so we proceeded to test all 14 of these SNPs. Therefore, we reported replication results for all the SNPs that we advanced to replication despite the updated analysis of our initial results and the appearance of a shift in our criterion. All SNPs genotyped in samples from the additional Beijing subjects used optimized TaqMan assays, whereas the Henan replication subjects were genotyped using Sequenom (11 SNPs) and TaqMan (three SNPs) assays. Three Sequenom assays failed genotyping and one (rs7822239) was repeated using TaqMan because it was nominally significant ($P = 0.02$) in the Beijing replication set. Samples with completion rates less than 80% in either replication were excluded from association analysis. Association analyses used log additive models with a trend effect and were adjusted for sex and age. Replication and Stage 1 results were combined using a fixed effect meta-analysis.

In silico bioinformatics analysis

Using 1000 Genomes CHB data, we identified all SNPs with $r^2 > 0.8$, 0.8, 0.5 (because no SNPs passed the 0.8 threshold), respectively, for the lead SNP in each of the three novel regions we identified. We then used HaploReg[31] and RegulomeDB[32] to explore potential functional annotations within the ENCODE data in the genome surrounding our lead SNPs (**Supplementary Table 7**).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Chen Wu[#1], Zhaoming Wang[#2,3], Xin Song[#4,5], Xiao-Shan Feng[#6], Christian C. Abnet[#2], Jie He[#7], Nan Hu[#2], Xian-Bo Zuo[#8], Wen Tan[#1], Qimin Zhan[1], Zhibin Hu[9], Zhonghu He[10], Weihua Jia[11,12], Yifeng Zhou[13], Kai Yu[2], Xiao-Ou Shu[14,15], Jian-Min Yuan[16], Wei Zheng[14,15], Xue-Ke Zhao[5], She-Gan Gao[6], Zhi-Qing Yuan[4], Fu-You Zhou[17], Zong-Min Fan[5], Ji-Li Cui[4,5], Hong-Li Lin[4,5], Xue-Na Han[5], Bei Li[5], Xi Chen[5,18], Sanford M. Dawsey[2], Linda Liao[2], Maxwell P. Lee[19], Ti Ding[20], You-Lin Qiao[21], Zhihua Liu[1], Yu Liu[1], Dianke Yu[1], Jiang Chang[1], Lixuan Wei[1], Yu-Tang Gao[22], Woon-Puay Koh[23], Yong-Bing Xiang[22], Ze-Zhong Tang[20], Jin-Hu Fan[21], Jing-Jing Han[5,18], Sheng-Li Zhou[5], Peng Zhang[5], Dong-Yun Zhang[5], Yuan Yuan[5], Ying Huang[1], Chunling Liu[1], Kan Zhai[1], Yan Qiao[1], Guangfu Jin[9], Chuanhai Guo[10], Jianhua Fu[11,12], Xiaoping Miao[24], Changdong Lu[17], Haijun Yang[17], Chaoyu Wang[2], William A. Wheeler[25], Mitchell Gail[2], Meredith Yeager[2,3], Jeff Yuenger[2,3], Er-Tao Guo[5], Ai-Li Li[5,26], Wei Zhang[5], Xue-Min Li[27], Liang-Dan Sun[8], Bao-Gen Ma[28,29,30], Yan Li[5], Sa Tang[5], Xiu-Qing Peng[5,6], Jing Liu[5], Amy Hutchinson[2,3], Kevin Jacobs[2,3], Carol Giffen[25], Laurie Burdette[2,3], Joseph F. Fraumeni Jr[2], Hongbing Shen[9], Yang Ke[10], Yixin Zeng[11,12], Tangchun Wu[24], Peter Kraft[31], Charles C. Chung[2,3], Margaret A. Tucker[2], Zhi-Chao Hou[5], Ya-Li Liu[4,5], Yan-Long Hu[4,5], Yu Liu[5], Li Wang[4,32], Guo Yuan[4,5], Li-Sha Chen[4,5], Xiao Liu[5], Teng Ma[5], Hui Meng[5], Li Sun[5], Xin-Min Li[5], Xiu-Min Li[4], Jian-Wei Ku[5,33], Ying-Fa Zhou[5,31], Liu-Qin Yang[34], Zhou Wang[35], Yin Li[36], Qirenwang Qige[37], Wen-Jun Yang[38], Guang-Yan Lei[39], Long-Qi Chen[40], En-Min Li[41,42], Ling Yuan[36], Wen-Bin Yue[5,43], Ran Wang[5], Lu-Wen Wang[5], Xue-Ping Fan[5], Fang-Heng Zhu[34], Wei-Xing Zhao[4], Yi-Min Mao[6], Mei Zhang[5], Guo-Lan Xing[5], Ji-Lin Li[44], Min Han[45], Jing-Li Ren[32], Bin Liu[46], Shu-Wei Ren[47], Qing-Peng Kong[48], Feng Li[49], Ilyar Sheyhidin[50,51], Wu Wei[52,53], Yan-Rui Zhang[28,29,30], Chang-Wei Feng[32], Jin Wang[5], Yu-Hua Yang[54], Hong-Zhang Hao[55], Qi-De Bao[55], Bao-Chi Liu[56], Ai-Qun Wu[57], Dong Xie[58], Wan-Cai Yang[4], Liang Wang[4,59], Xiao-Hang Zhao[60], Shu-Qing Chen[61], Jun-Yan Hong[61,62], Xue-Jun Zhang[8], Neal D Freedman[2], Alisa M. Goldstein[2], Dongxin Lin[**,1], Philip R. Taylor[**, 2], Li-Dong Wang[**,4,5], and Stephen J. Chanock[2]

Affiliations

[1]State Key Laboratory of Molecular Oncology, Cancer Institute & Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.

[2]Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI), National Institutes of Health, Bethesda, Maryland, USA.

[3]Cancer Genome Research Laboratory, Division of Cancer Epidemiology and Genetics, Advanced Technology Program, SAIC-Frederick Inc., NCI-Frederick, Frederick, MD, USA.

[4]Department of Pathology, Cancer Research Center, Basic Medical College, Xinxiang Medical University, Xinxiang, Henan, China.

[5]Henan Key Laboratory for Esophageal Cancer Research of The First Affiliated Hospital, Zhengzhou University, Zhengzhou, Henan, China.

[6]Department of Oncology, Pathology, Clinical Laboratory and Respiratory Medicine, The First Affiliated Hospital, Henan University of Science and Technology, Luoyang, Henan, China.

[7]Department of Thoracic Surgery, Cancer Institute & Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.

[8]Key Laboratory of Dermatology, Anhui Medical University, Anhui, Hefei, China.

[9]Department of Epidemiology and Biostatistics, Cancer Center, Nanjing Medical University, Nanjing, China.

[10]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University School of Oncology, Beijing Cancer Hospital & Institute, Beijing, China.

[11]State Key Laboratory of Oncology in Southern China

[12]Department of Experimental Research, Sun Yat-Sen University Cancer Center, Guangzhou, China.

[13]Laboratory of Cancer Molecular Genetics, Medical College of Soochow University, Suzhou, China.

[14]Department of Medicine, Vanderbilt University, Nashville, Tennessee, USA.

[15]Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, Tennessee, USA.

[16]University of Pittsburgh Cancer Institute, Pittsburg, PA, USA.

[17]Anyang Tumor Hospital, Anyang, Henan, China.

[18]Department of Basic Oncology and Pathology at College of Medicine, Zhengzhou University, Zhengzhou, Henan, China.

[19]Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA.

[20]Shanxi Cancer Hospital, Taiyuan, Shanxi, China.

[21]Department of Epidemiology, Cancer Institute & Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.

[22]Shanghai Cancer Institute, Shanghai, People's Republic of China.

[23]Duke-National University of Singapore, Graduate Medical School, Singapore.

[24]Key Laboratory for Environment and Health (Ministry of Education), School of Public Health, Huazhong University of Sciences and Technology, Wuhan, China.

[25]Information Management Services Inc., Silver Spring, Maryland, USA.

[26]Department of Medical Oncology, Tumor Hospital of Linzhou, Linzhou, Henan, China.

[27]Department of Pathology and Thoracic Surgery, Centre for Health Screening and Endoscopy, Cixian Hospital, Cixian, Hebei, China.

[28]Department of Gastroenterology, Henan Provincial People's Hospital, Zhengzhou, Henan, China.

[29]Department of Hematology, Henan Provincial People's Hospital, Zhengzhou, Henan, China.

[30]Central Laboratory, Henan Provincial People's Hospital, Zhengzhou, Henan, China.

[31]Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA.

[32]Department of Basic Oncology, Pathology and Gastroenterology, The Second Affiliated Hospital, Zhengzhou University, Zhengzhou, Henan, China.

[33]Department of Oncology, The Second Affiliated Hospital of Nanyang Medical College, Nanyang, Henan, China.

[34]Department of Radiotherapy, Pathology and Pediatrics, Central Hospital of Xinxiang, Xinxiang, Henan, China.

[35]Department of Thoracic Surgery, Provincial Hospital Affiliated to Shandong University, Jinan, Shandong, China.

[36]Department of Thoracic Surgery and Radiotherapy, Cancer Hospital of Henan Province, Zhengzhou, Henan, China.

[37]Department of Internal Mongolia Medicine, The Affiliated Hospital, Inner Mongolia Medical College, Huhhot, Inner Mongolia, China.

[38]Department of Biotechnology, Ningxia Medical University, Yinchuan, Ningxia, China.

[39]Department of Thoracic Surgery, Tumor Hospital of Shaanxi Province, Xi'an, Shaanxi, China.

[40]Department of Thoracic and Cardiovascular Surgery, West China Hospital, Sichuan University, Chengdu, Sichuan, China.

[41]Department of Oncologic Pathology, Medical College of Shantou University, Shantou, Guangdong, China.

[42]Department of Biochemistry and Molecular Biology, Medical College of Shantou University, Shantou, Guangdong, China.

[43]Department of Oncology, Puyang City Oil Field General Hospital, Puyang, Henan, China.

[44]Department of Pathology and Thoracic Surgery, Linzhou Esophageal Cancer Hospital, Linzhou, Henan, China.

[45]Department of Gastroenterology and Thoracic Surgery, The First People's Hospital of Shangqiu, Shangqiu, Henan, China.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

[46]Department of Gastroenterology, Beijing Tongren Hospital Affiliated to Capital Medical University, Beijing, China.

[47]Department of Oncology, Xinyang Central Hospital, Xinyang, Henan, China.

[48]State Key Laboratory of Genetic Resource and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China.

[49]Department of Pathology, Shihezi University School of Medicine, Shihezi, Xinjiang, China.

[50]Department of Thoracic Surgery, Xinjiang Medical University, Urumqi, Xinjiang, China.

[51]Medical Research Center, The First Affiliated Hospital, Xinjiang Medical University, Urumqi, Xinjiang, China.

[52]Institute of Hematologic Disease, Changzhi Medical University, Changzhi, Shanxi, China.

[53]Department of Pathology, Changzhi Medical University, Changzhi, Shanxi, China.

[54]Department of Surgery, Hebi Dahejian Hospital, Hebi, Henan, China.

[55]Anyang District Hospital, Anyang, Henan, China.

[56]Surgical Department of Shanghai Public Health Clinical Center, Fudan University, Shanghai, China.

[57]Department of Anatomy, the Second Military Medical University of Chinese PLA, Shanghai, China.

[58]Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China.

[59]Medical College of Wisconsin, Cancer Research Center, Milwaukee, WI, USA.

[60]National Laboratory of Molecular Oncology, Cancer Institute & Hospital, CAMS, Beijing, China.

[61]College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, Zhejiang, China.

[62]University of Medicine and Dentistry of New Jersey, Newark, NJ, USA.

ACKNOWLEDGEMENTS

Reference List

1. Abnet CC, et al. A shared susceptibility locus in PLCE1 at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. Nat. Genet. 2010; 42:764–767. [PubMed: 20729852]

2. Wang LD, et al. Genome-wide association study of esophageal squamous cell carcinoma in Chinese subjects identifies susceptibility loci at PLCE1 and C20orf54. Nat. Genet. 2010; 42:759–763. [PubMed: 20729853]

3. Wu C, et al. Genome-wide association study identifies three new susceptibility loci for esophageal squamous-cell carcinoma in Chinese populations. Nat. Genet. 2011; 43:679–684. [PubMed: 21642993]

4. Wu C, et al. Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene-environment interactions. Nat. Genet. 2012; 44:1090–1097. [PubMed: 22960999]

5. Abnet CC, et al. Genotypic variants at 2q33 and risk of esophageal squamous cell carcinoma in China: a meta-analysis of genome-wide association studies. Hum. Mol. Genet. 2012; 21:2132–2141. [PubMed: 22323360]

6. Yang CS. Research on esophageal cancer in China: a review. Cancer Res. 1980; 40:2633–2644. [PubMed: 6992989]

7. Kamangar F, Chow WH, Abnet CC, Dawsey SM. Environmental causes of esophageal cancer. Gastroenterol Clin North Am. 2009; 38:27–57. [PubMed: 19327566]

8. Tran GD, et al. Prospective study of risk factors for esophageal and gastric cancers in the Linxian general population trial cohort in China. Int. J Cancer. 2004; 113:176–181.

9. Bishop DT, et al. Genome-wide association study identifies three loci associated with melanoma risk. Nat. Genet. 2009; 41:920–925. [PubMed: 19578364]

10. Sherborne AL, et al. Variation in CDKN2A at 9p21.3 influences childhood acute lymphoblastic leukemia risk. Nat. Genet. 2010; 42:492–494. [PubMed: 20453839]

11. Berndt SI, et al. Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia. Nat. Genet. 2013

12. Rajaraman P, et al. Genome-wide association study of glioma and meta-analysis. Hum. Genet. 2012; 131:1877–1888. [PubMed: 22886559]

13. Gu F, et al. Common genetic variants in the 9p21 region and their associations with multiple tumours. Br. J. Cancer.10. 2013

14. Ishikawa H, Barber GN. STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. Nature. 2008; 455:674–678. [PubMed: 18724357]

15. Kennedy RB, et al. Genome-wide analysis of polymorphisms associated with cytokine responses in smallpox vaccine recipients. Hum. Genet. 2012; 131:1403–1421. [PubMed: 22610502]

16. Eck P, et al. Genomic and functional analysis of the sodium-dependent vitamin C transporter SLC23A1-SVCT1. Genes Nutr. 2007; 2:143–145. [PubMed: 18850166]

17. Mirvish SS. Role of N-nitroso compounds (NOC) and N-nitrosation in etiology of gastric, esophageal, nasopharyngeal and bladder cancer and contribution to cancer of known exposures to NOC. Cancer Lett. 1995; 93:17–48. [PubMed: 7600541]

18. Enciso-Mora V, et al. Low penetrance susceptibility to glioma is caused by the TP53 variant rs78378222. Br. J. Cancer. 2013; 108:2178–2185. [PubMed: 23571737]

19. Hu N, et al. Frequent inactivation of the TP53 gene in esophageal squamous cell carcinoma from a high-risk population in China. Clin. Cancer Res. 2001; 7:883–891. [PubMed: 11309337]

20. Remeseiro S, Losada A. Cohesin, a chromatin engagement ring. Curr. Opin. Cell Biol. 2013; 25:63–71. [PubMed: 23219370]

21. Coviello AD, et al. A genome-wide association meta-analysis of circulating sex hormone-binding globulin reveals multiple Loci implicated in sex steroid hormone regulation. PLoS. Genet. 2012; 8:e1002805. [PubMed: 22829776]

22. Freedman ND, et al. The association of menstrual and reproductive factors with upper gastrointestinal tract cancers in the NIH-AARP cohort. Cancer. 2010; 116:1572–1581. [PubMed: 20186831]

23. Wang QM, Qi YJ, Jiang Q, Ma YF, Wang LD. Relevance of serum estradiol and estrogen receptor beta expression from a high-incidence area for esophageal squamous cell carcinoma in China. Med. Oncol. 2011; 28:188–193. [PubMed: 20195802]

24. Bei JX, et al. A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci. Nat. Genet. 2010; 42:599–603. [PubMed: 20512145]

25. Li S, et al. GWAS identifies novel susceptibility loci on 6p21.32 and 21q21.3 for hepatocellular carcinoma in chronic hepatitis B virus carriers. PLoS. Genet. 2012; 8:e1002791. [PubMed: 22807686]

26. Lan Q, et al. Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. Nat. Genet. 2012; 44:1330–1335. [PubMed: 23143601]

27. Slager SL, et al. Genome-wide association study identifies a novel susceptibility locus at 6p21.3 among familial CLL. Blood. 2011; 117:1911–1916. [PubMed: 21131588]

28. Okada Y, et al. HLA-Cw*1202-B*5201-DRB1*1502 haplotype increases risk for ulcerative colitis but reduces risk for Crohn's disease. Gastroenterology. 2011; 141:864–871. [PubMed: 21699788]

29. Han JW, et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nat. Genet. 2009; 41:1234–1237. [PubMed: 19838193]

30. Su Z, et al. Common variants at the MHC locus and at chromosome 16q24.1 predispose to Barrett's esophagus. Nat. Genet. 2012; 44:1131–1136. [PubMed: 22961001]

31. Boyle AP, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012; 22:1790–1797. [PubMed: 22955989]

32. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012; 40:D930–D934. [PubMed: 22064851]
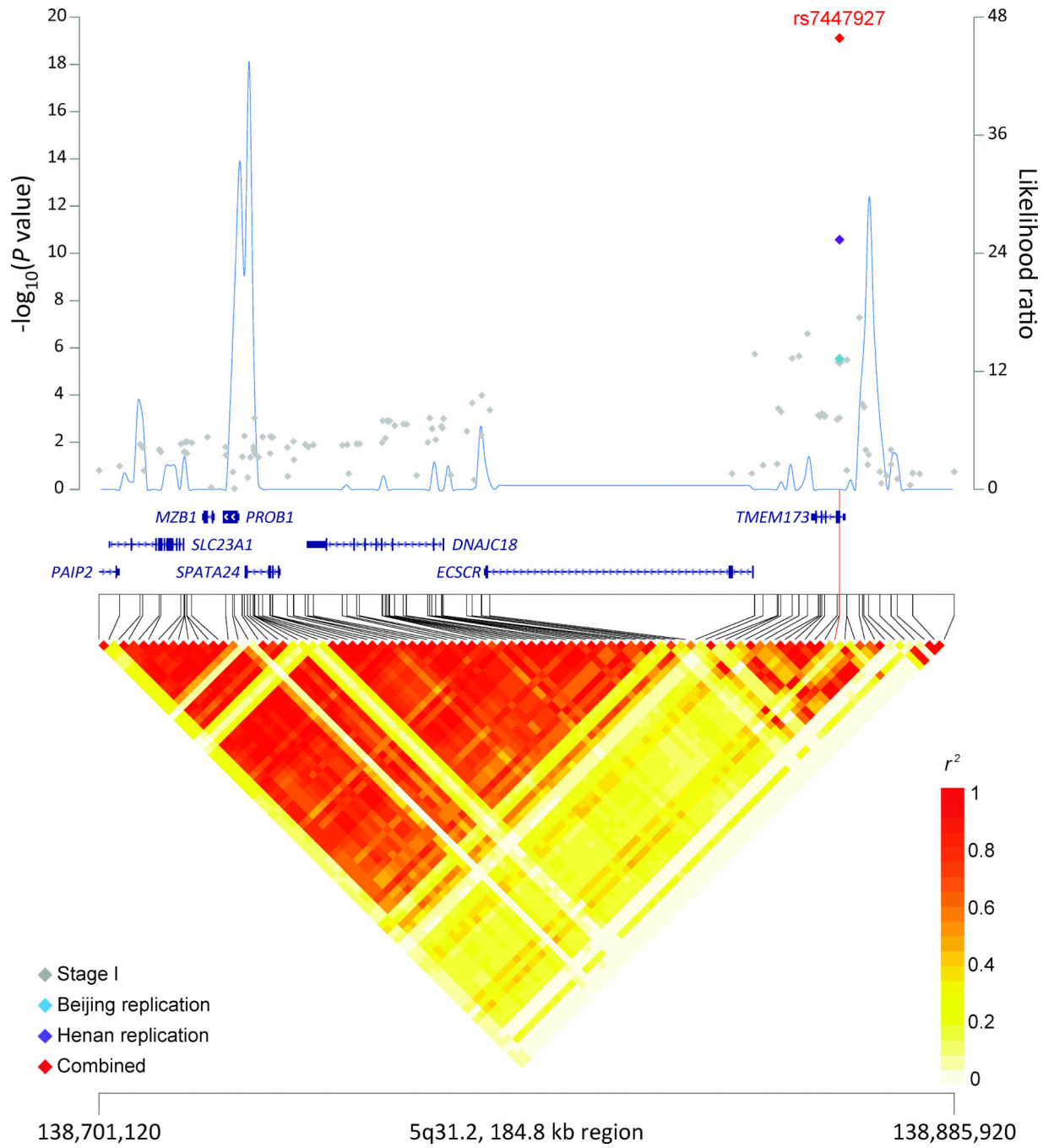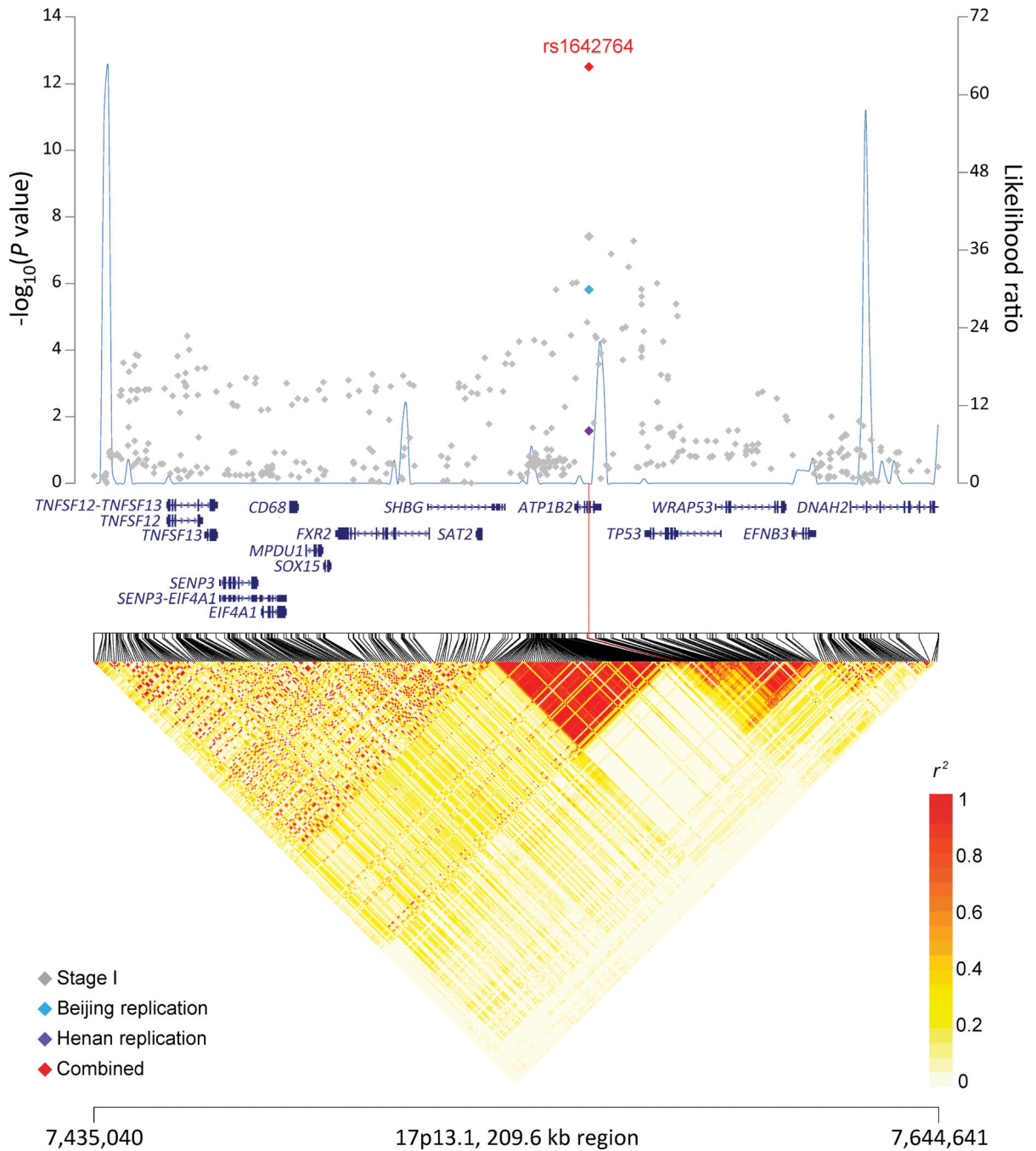
Figure 1. Association results, recombination, and linkage disequilibrium plots for the 5q31.2 and 17p13.1 regions identified in stage 1 (joint analysis of three independent GWAS), for the two replications sets (Beijing and Henan), and the joint estimate from all data

(a) 5q31.2:138,701,120–138,885,920 and (b) 17p13.1:7,435,040–7,644,641. Association results from a trend test in –log10 *P* values (y axis, left; gray diamonds, stage I association result; sky blue diamonds, Beijing replication result; purple diamonds, Henan replication results; red diamonds, joint result) of the SNPs are shown according to their chromosomal positions (x axis). Linkage disequilibrium structure based on the 1000 Genomes CHB data (n=91) was visualized by snp.plotter software. The line graph shows likelihood ratio

statistics (y axis, right) for recombination hotspot by SequenceLDhot software based on the background recombination rates inferred by PHASE v2.1 using the 1000 Genomes CHB data. Physical locations are based on NCBI human genome build 37. Gene annotation was based on the NCBI RefSeq genes from the UCSC Genome Browser.

Table 1

Association between SNPs at three loci and risk for esophageal squamous cell carcinoma in a joint analysis of three independent genome-wide association studies of Han Chinese subjects.

NCBI dbSNP 137 identifier (Reference Allele, Effect Allele)	Cytoband	Nearest Gene	Study	Controls	Cases	Effect Allele Frequency in Controls	Effect Allele Frequency in Cases	OR	(95% CI)	P-value	Pheterogeneity
			All China								
rs7447927 (C,G)	5q31.2	TMEM173	3 scan analysis (Stage 1)	5786	5336	0.460	0.451	0.87	(0.82-0.92)	4.07E-06	
			Henan Replication	4802	4486	0.489	0.438	0.80	(0.75-0.86)	2.70E-11	
			Beijing Replication	5079	4797	0.431	0.397	0.87	(0.82-0.92)	2.92E-06	
			Combined	15667	14619			0.85	(0.82-0.88)	7.72E-20	1.31E-01
rs1642764(C,T)	17p13.1	ATP1B2,TP53,p53	3 scan analysis (Stage 1)	5786	5336	0.499	0.456	0.84	(0.79-0.89)	3.91E-08	
			Henan Replication	4634	4694	0.484	0.470	0.93	(0.87-0.99)	2.67E-02	
			Beijing Replication	5054	4777	0.520	0.483	0.86	(0.81-0.92)	1.55E-06	
			Combined	15474	14807			0.88	(0.85-0.91)	3.10E-13	7.70E-02
rs35597309(G,A)	6p21.32	HLA class II genes	3 scan analysis (Stage 1)	5787	5336	0.072	0.094	1.32	(1.18-1.47)	6.17E-07	
			Henan Replication	4659	4597	0.067	0.085	1.23	(1.09-1.38)	1.00E-03	
			Beijing Replication	5079	4786	0.075	0.080	1.05	(0.94-1.17)	3.59E-01	
			Combined	15525	14719			1.19	(1.12-1.27)	1.18E-07	1.50E-02
			Taihang mountains only								
rs35597309(G,A)	6p21.32	HLA class II genes	NCI Scan	2707	2021	0.077	0.109	1.43	(1.23-1.67)	4.27E-06	
			Henan Scan	1082	1375	0.063	0.093	1.55	(1.22-1.97)	3.50E-04	
			Henan Replication	4659	4597	0.067	0.085	1.23	(1.09-1.38)	1.00E-03	
			Combined	8448	7993			1.33	(1.22-1.46)	1.99E-10	1.22E-01

One of these new loci showed significant heterogeneity (p=0.015) among the three studies and the associations are also reported using only the two GWAS and one replication set that primarily used subjects from populations with high incidence in the Taihang Mountains of north central China.