HHS Public Access

Author manuscript

Gut. Author manuscript; available in PMC 2018 April 01.

Published in final edited form as:

Gut. 2017 April ; 66(4): 581–587. doi:10.1136/gutjnl-2015-310612.

Identification of new susceptibility loci for gastric noncardia adenocarcinoma: Pooled results from two Chinese genome-wide association studies

Zhaoming Wang[#1,2,3], **Juncheng Dai**[#3], **Nan Hu**[#1], **Xiaoping Miao**[#4], **Christian C. Abnet**[#1], **Ming Yang**[#5], **Neal D. Freedman**[#1], **Jinfei Chen**[#6], **Laurie Burdette**[1,2], **Xun Zhu**[3], **Charles C. Chung**[1,2], **Chuanli Ren**[3,7], **Sanford M Dawsey**[1], **Meilin Wang**[8], **Ti Ding**[9], **Jiangbo Du**[3], **Yu-Tang Gao**[10], **Rong Zhong**[4], **Carol Giffen**[11], **Wenting Pan**[5], **Woon-Puay Koh**[12], **Ningbing Dai**[3], **Linda M Liao**[1], **Caiwang Yan**[3], **You-Lin Qiao**[13], **Yue Jiang**[3], **Xiao-Ou Shu**[14], **Jiaping Chen**[3], **Chaoyu Wang**[1], **Hongxia Ma**[3], **Hua Su**[1], **Zhendong Zhang**[8], **Lemin Wang**[1], **Chen Wu**[15], **Yong-Bing Xiang**[10], **Zhibin Hu**[3], **Jian-Min Yuan**[16], **Lu Xie**[3], **Wei Zheng**[14], **Dongxin Lin**[15], **Stephen J. Chanock**[1,#], **Yongyong Shi**[17,#], **Alisa M. Goldstein**[1,#], **Guangfu Jin**[3,#], **Philip R. Taylor**[1,#], and **Hongbing Shen**[3,#]

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA

[2]Cancer Genomics Research Laboratory, Leidos Biomedical Research Inc., Gaithersburg, MD, USA

[3]Department of Epidemiology and Biostatistics, Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Personalized Medicine, School of Public Health, Nanjing Medical University, Nanjing, P.R. China

Correspondence should be addressed to: Hongbing Shen, Department of Epidemiology and Biostatistics, Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Personalized Medicine, School of Public Health, Nanjing Medical University, Nanjing, P.R. China; hbshen@njmu.edu.cn.

[#]These authors co-supervised this work.

AUTHOR CONTRIBUTIONS

H.B.S, P.R.T., G.J., A.M.G., Y.S., S.J.C., Z.W., J.C.D., N.H., X.M., C.C.A., M.Y., N.D.F., and J.F.C. organized and designed the study; Z.W., J.C.D., H.B.S., P.R.T., A.M.G., and S.J.C. wrote the first draft of the manuscript; Z.W. and J.C.D. contributed to the design and execution of statistical analysis; H.B.S, P.R.T., G.J., A.M.G., Y.S., S.J.C., Z.W., J.C.D., N.H., X.M., C.C.A., M.Y., N.D.F., and J.F.C. contributed to the writing of the manuscript; Z.W., J.C.D., H.B.S., G.J., Z.H., and L.B. conducted and supervised the genotyping of samples; H.S., L.W., and N.H. conducted the qRT-PCR experiments; M.Y., X.Z., C.C.C., C.R., S.M.D., M.W., T.D., J.B.D., Y.T.G., R.Z., C.G., W.P., W.P.K., N.D., L.M.L., C.Y., Y.L.Q., Y.J., X.O.S., J.P.C., C.W., H.M.,Z.Z., C.W., Y.B.X., Z.H., J.M.Y., L. X., W.Z., and D.L. contributed to the conduct of the epidemiological studies or contributed samples to the GWAS or follow-up genotyping.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

URLs

GTEx, http://www.gtexportal.org/;
GTEx *MUC1* expression, http://www.gtexportal.org/home/gene/MUC1
GLU, https://code.google.com/p/glu-genetics/;
IMPUTE2 software version 2.2.2, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html;
LocusZoom, http://csg.sph.umich.edu/locuszoom/;
PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/;
SHAPEIT software version 2, https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html;
SNPTEST software version 2.2, https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html.
UCSC liftOver tool, http://hgdownload.cse.ucsc.edu/downloads.html

[4]Department of Epidemiology and Biostatistics, and the Ministry of Education Key Lab of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, P.R. China

[5]State Key Laboratory of Chemical Resource Engineering, Beijing Laboratory of Biomedical Materials, College of Life Science and Technology, Beijing University of Chemical Technology, Beijing, China

[6]Department of Oncology, Nanjing First Hospital, Nanjing Medical University, Nanjing, P.R. China

[7]Clinical Medical Testing Laboratory, Northern Jiangsu People's Hospital and Clinical Medical College of Yangzhou University, Yangzhou, P.R. China

[8]Department of Genetic Toxicology, the Key Laboratory of Modern Toxicology of Ministry of Education, School of Public Health, Nanjing Medical University, Nanjing, P.R. China

[9]Shanxi Cancer Hospital, Taiyuan, Shanxi, P.R. China

[10]Shanghai Cancer Institute, Shanghai, P.R. China

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

[12]Duke-NUS Graduate Medical School Singapore, and Saw Swee Hock School of Public Health, National University of Singapore, Singapore

[13]Department of Epidemiology, Cancer Institute and Hospital Chinese Academy of Medical Sciences, Beijing, P.R. China

[14]Department of Medicine and Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, Tennessee, USA

[15]Department of Etiology & Carcinogenesis, Cancer Institute and Hospital Chinese Academy of Medical Sciences, Beijing, P.R. China

[16]Division of Cancer Control and Population Sciences, University of Pittsburgh Cancer Institute; and Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburg, PA, USA

[17]Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders, Ministry of Education, Shanghai Jiao Tong University, Shanghai, P.R. China

[#] These authors contributed equally to this work.

Abstract

Objective—Although several genome-wide association studies (GWAS) of noncardia gastric cancer have been published, more novel association signals could be exploited by combining individual studies together, which will further elucidate the genetic susceptibility of noncardia gastric cancer.

Design—We conducted a meta-analysis of two published Chinese GWAS studies (2,031 noncardia gastric cancer cases and 4,970 cancer-free controls) and followed by genotyping of additional 3,564 cases and 4,637 controls in two stages.

Results—The overall meta-analysis revealed two new association signals. The first was a novel locus at 5q14.3 and marked by rs7712641 (per-allele odds ratio (OR) = 0.84, 95% CI 0.80-0.88; $P = 1.21 \times 10^{-11}$). This SNP marker maps to the intron of the long-noncoding RNA, lnc-POLR3G-4 (*XLOC_004464*), which we observed has lower expression in noncardia gastric tumor compared to matched normal tissue ($P_{wilcoxon\ signed\text{-}rank} = 7.20 \times 10^{-4}$). We also identified a new signal at the 1q22 locus, rs80142782 (per-allele OR=0.62; 95% CI 0.56-0.69; $P = 1.71 \times 10^{-19}$), which was independent of the previously reported SNP at the same locus, rs4072037 (per-allele OR=0.74; 95% CI 0.69-0.79; $P = 6.28 \times 10^{-17}$). Analysis of the new SNP conditioned on the known SNP showed that the new SNP remained genome-wide significant ($P_{conditional} = 3.47 \times 10^{-8}$). Interestingly, rs80142782 has a minor allele frequency (MAF) of 0.05 in East Asians but is monomorphic in both European and African populations.

Conclusions—These findings add new evidence for inherited genetic susceptibility to noncardia gastric cancer and provide further clues to its etiology in the Han Chinese population.

Keywords

Gastric cancer; GWAS; Genetic epidemiology

INTRODUCTION

Globally, gastric cancer remains the third leading cause of cancer death in both sexes[1], with more than half of gastric cancer cases worldwide occurring in East Asia, predominantly in China. Most cases of gastric cancer are sporadic[2], and its etiology is related to both genetic susceptibility and epidemiological risk factors[3] such as age, sex, *Helicobacter pylori* infection[4, 5], family history, excessive salt intake, and tobacco smoking. Anatomically, gastric cancer is classified into cardia and noncardia gastric cancer, which are characterized by distinct risk factors and clinical features[4, 6-8].

Recently, several genome-wide association studies (GWAS) of gastric adenocarcinoma were conducted in East Asians[9-12]. Notable findings include single nucleotide polymorphism (SNP) markers mapping to 8q24.3, for both an intronic SNP (rs2976392) and an exonic SNP (rs2294008) in the Prostate Stem Cell Antigen gene (*PSCA)*; and two markers (rs2075570 and rs2070803) near the Mucin 1 gene (*MUC1)* on 1q22[9]. The findings on 1q22 locus and gastric cancer risk were further replicated in several follow-up studies and additional evidence pointed to the nonsynonymous SNP, rs4072037, as the functional variant underlying the observed association[13-18]. In addition, 3q13.31 marked by rs9841504, 5p13.1 marked by rs13361707 or rs10074991 and 6p21.1 marked by rs2294693 were reported to be associated with noncardia gastric adenocarcinoma in China[11,12], whereas 10q23 marked by rs2274223, a nonsynonymous SNP located in *PLCE1*, was associated with cardia but not noncardia gastric cancer[10]. Together, these data indicate that five chromosomal regions, 1q22, 3q13.31, 5p13.1, 6p21.1 and 8q24.3, have strong evidence for harboring one or more susceptibility alleles for noncardia gastric cancer. Based on the experience of other cancer sites, additional loci will likely be found by interrogation of increasingly larger studies.

MATERIALS AND METHODS

Primary GWAS scan data

For the NCI GWAS, subjects were drawn from four prospective cohort studies and one large case-control study as reported in Abnet *et al.*[10]. In addition, all subjects used in replication in the original paper were subsequently genotyped using the Illumina 660W-Quad microarray; this included scanning of 725 additional gastric cancer cases and 608 additional controls. The current analysis included all 1,025 noncardia gastric cancer cases and 2,697 controls from the NCI Upper Gastrointestinal Cancer GWAS.

For the Nanjing and Beijing GWAS, individuals were derived from separate case-control studies conducted in Nanjing (565 cases and 1,162 controls) and Beijing (468 cases and 1,123 controls), as previously reported in Shi Y *et al.*[11], where individuals were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0.

Replication samples

The first-stage replication included 1,145 cases and 2,253 controls were derived from Jiangsu Province. The second-stage replication included 2,419 cases and 2,384 controls, which were derived from Beijing, Hubei and Shangdong Province.

Gastric cancer cases tested for expression of the lncRNA associated with the GWAS SNPs came from our UGI Cancer Genetic Studies (URL: http://dceg.cancer.gov/about/staff-directory/biographies/O-Z/taylor-philip). Genotypes for all these cases are known because they were also all participants in our previous GWAS [10].

All study individuals provided informed consent and both the institutional review boards of NCI and Nanjing Medical University approved all procedures and all experiments, which were conducted in accordance with the approved guidelines.

Genotype imputation

In addition to the quality control procedures performed in the previous primary publications for both previous GWAS, SNPs with call rate of <95%, *P* value for Hardy-Weinberg Equilibrium (HWE) in controls≤1.0×10^{-6} or MAF <1% in controls were further removed before imputation. Imputation was conducted separately for the NCI (Illumina 660W) and the Nanjing+Beijing (Affymetrix 6.0) scan data using IMPUTE2 software version 2.2.2 and taking all populations in the 1000 Genomes Project Release 1 Version 3 as the reference set, which automatically finds haplotypes from the best matching population from the entire reference set to do the imputation. First, genomic coordinates for NCBI human genome Build 36 were converted to those for NCBI human genome Build 37 using the UCSC liftOver tool. The few loci for which coordinates could not be converted were also excluded from imputation. Second, the strand of the inference data was aligned with the 1000 Genomes Project 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 windows. A pre-phasing strategy with SHAPEIT software version 2 was

adopted to improve the imputation performance. The phased haplotypes from SHAPEIT were fed directly into IMPUTE2.

SNP selection and replication genotyping

The meta-analysis included 6,223,896 SNPs based on the intersection of the three imputed datasets. Individual SNPs for the stage 2A replication were selected based on the following criteria: (i) SNPs with INFO score ≥ 0.5; (ii) MAF in control set ≥ 0.01; (iii) *P* value for HWE in control set > 1.0×10^{-4} in each set; (iv) P_{het}> 1.0×10^{-4} and I^2<75% in meta analysis; (v) LD pruning: included only one SNP with the lowest *P* value when the pair-wise $r^2\geq0.3$ within a distance of 200kb; (vi) Exclusion of previously identified loci associated with risk for noncardia gastric cancer. After applying the above criteria, we then picked the top 48 SNPs ($P_{meta} \leq 2.58 \times 10^{-5}$). For 1q22 and 8q24, there were two SNPs (rs80142782; rs76845414) retained in our LD filtered list, so we included back two more SNPs (rs4072037; rs2294008) previously reported for each of these regions in order to search for potential secondary signals for these two known loci to derive an initial list of 50 SNPs. Subsequently, 13 SNPs failed either Sequenom assay design or genotyping studies. As a result, a total of 37 SNPs (Supplementary Table 2) were successfully genotyped using iPLEX Sequenom MassARRAY platform (Sequenom, CA, USA) in stage 2A replication (1,145 cases and 2,253 controls).

Five SNPs with *P*< 0.05 in stage 2A without significant heterogeneity (P_{het}> 1.0×10^{-4} and I^2<75%), were advanced for TaqMan assays (Applied Biosystems) in stage 2B replication (2,419 cases and 2,384 controls) (Supplementary Table 3). Further information on primers and probes are available upon request. For quality control purposes: (i) case and control samples were mixed on each plate; (ii) genotyping was performed blind to case/control status; (iii) two water controls were used in each plate as blank controls.

Quantitative Reverse Transcription-PCR

Total RNA was extracted from each patient's matched frozen tumor and normal surgical resection tissues using All Prep DNA/RNA/Protein kit (QIAGEN) in accordance with the manufacturer's instructions. RNA quality and quantity were determined using the RNA Nano Chip/Agilent 2100 Bioanalyzer (Agilent Technologies). Reverse transcription of RNA was done by adding 0.2-2ug total RNA, 1 uL of oligo(dT)12-18 (500 ug/mL), 1 uL (200 units) of SuperScript II reverse transcriptase, 1 uL (2 units) of E-coli RNase, and 1 uL of 10 mmol/L deoxynucleotide triphosphate (Invitrogen) in total volume of 20 ul. All real-time PCRs were done using an ABI 7300 Sequence Detection System. Primer and probe for the target gene and the internal control gene (GAPDH) were designed and ordered from ABI (Assay ID: AJBJXUX; Part Number: 4441114). A singleplex reaction mix was prepared according to the manufacturer's protocol of "Assays-on-Demand Gene Expression Products", including 10 uL Taqman Universal PCR Master Mix, No AmpErase UNG (2×), 1 uL of 20× Assays on-Demand Gene Expression Assay Mix (all Gene Expression assays have a FAM reporter dye at the 5' end of the TaqMan MGB probe and a nonfluorescent quencher at the 3' end of the probe), and 9 uL of cDNA (1000ng) diluted in RNase-free water to a total volume of 20 uL. Each sample for the gene was run in triplicate and the

expression level was averaged over all runs. The thermal cycling conditions included an initial denaturation step at 95°C for 10 min, 40 cycles at 95°C for 15 s, and 60°C for 1 min.

Statistical analysis

Association testing was performed using SNPTEST software version 2.2, with adjustment for age, sex, and study variables for NCI. Two eigenvectors (ev4 and ev8) were significantly associated with case status ($P < 0.05$) in the baseline model (not including SNP effects) which was adjusted for age, sex, study, and all top ten eigenvectors; and therefore these two significant eigenvectors were also included to adjust for population stratification in final association models. For Nanjing and Beijing, age, sex, smoking, and alcohol consumption were adjusted in baseline models. Three eigenvectors (ev1, ev4, and ev9) for Nanjing and five eigenvectors (ev1, ev3, ev7, ev9, and ev10) for Beijing were adjusted for population stratification separately, which were also significantly associated with case status ($P < 0.05$). In each replication study, we adjusted for gender, age, smoking, and alcohol consumption only.

For the meta-analysis we used the meta module implemented in GLU (see URLs). Strand flipping was handled by comparing alleles either with direct matching or with reverse complement matching). For A/T or G/C SNPs, strand matching was based on allele frequency checking. The fixed-effects inverse variance method was used to combine the β estimates and standard errors from each GWAS scan as well as the replication stages. The P value for heterogeneity was calculated using Cochran's Q, which is distributed as a χ^2 statistic with $(n - 1)$ degrees of freedom, where n is the number of sets included in the meta-analysis. Data analysis and management was performed with GLU or PLINK (see URLs).

PLINK was also used for the conditional haplotype analysis. Wilcoxon signed-rank test (R package) was applied to the tumor/normal paired quantitative RT-PCR data to assess the RNA expression level.

In silico bioinformatics analysis

We utilized GTEx (see URLs) for eQTL information for associated SNPs (Supplementary Table 5). We also searched HaploReg v3[19] to explore potential functional annotations within Encyclopedia of DNA Elements (ENCODE) data for the genomic regions surrounding our lead SNPs (Supplementary Table 6).

RESULTS

To discover additional susceptibility alleles for noncardia gastric cancer in the Han Chinese population, we conducted a combined analysis of two previously published GWAS[10, 11] after imputing the genetic data with the 1000 Genome Project data phase 1 release version 3[20]. The combined data set included a total of 6,223,896 SNPs for a fixed-effects meta-analysis of 2,031 cases and 4,970 cancer-free controls (Supplementary Table 1). Quantile-quantile (QQ) and Manhattan plots based on stage 1 meta-analysis P values are shown in Supplementary Figures 1 and 2, respectively. We further followed up 37 promising loci (see Methods) and genotyped them in an independent set of 1,145 cases and 2,253 controls in stage 2A (Supplementary Table 2). Finally, for stage 2B, we advanced five loci that were

nominally significant in stage2A to a second independent set of 2,419 cases and 2,384 controls (Supplementary Table 3).

Based on the overall meta-analysis including two discovery GWAS scans and two replication studies, we identified two novel risk loci for noncardia gastric cancer, the first one is rs7712641 at 5q14.3 (per-allele odds ratio (OR) = 0.84, 95% CI 0.80-0.88; $P = 1.21\times10^{-11}$). No heterogeneity was observed across the two GWAS scans and two replication studies (P_{het}=0.56) (Table 1). There are no protein-coding genes within the 1Mb of the associated SNP (chr5: 88,346,298-89,459,630; hg19) (Figure 1a). However, rs7712641 is located in the intron of lnc-POLR3G-4 (*XLOC_004464*), a long noncoding RNA (lncRNA) which is poorly characterized. To explore the possible effect of the SNP marker on the lncRNA, lnc-POLR3G-4, we extracted total RNA from 75 matched gastric noncardia adenocarcinoma and adjacent normal tissue pairs and performed a qRT-PCR analysis to measure its expression abundance. We found that expression differed between tumor and normal tissues ($P_{wilcoxon\ signed\text{-}rank}=7.2\times10^{-4}$); with the majority of pairs (50 of 75) showing lower expression in tumor compared to normal tissue (Figure 2). These data provided preliminary evidence that this lncRNA could function in a manner resembling a tumor suppressor gene. However, the association between rs7712641 and lnc-POLR3G-4 expression in normal tissue was negative (P=0.99), which does not support the the notion of a functional role for this SNP. More functional studies are warranted to clarify the complicated phenomena.

As anticipated, the current study, which included samples from these previous GWAS reports[10, 11], also replicated the association with rs4072037 (Table 1; per-allele OR=0.74; 95% CI 0.69-0.79; $P = 6.28\times10^{-17}$) at 1q22. However, we also identified a second strong signal in this region and by doing so established an independent, new genome-wide significant SNP rs80142782 (Table 1 **and** Figure 1b; per-allele OR=0.62; 95% CI 0.56-0.69; $P = 1.71\times10^{-19}$). Based on the evidence at hand, it seems that rs80142782 is likely an independent primary signal (rs4072037 as a secondary signal) at this locus. This evidence includes: (i) both SNPs are about 323kb apart and have moderately low ($r^2 = 0.3$) pair-wise linkage disequilibrium (LD) in 1,000 Genome s Project data for Asians; (ii) rs80142782 conditioned on rs4072037 remained genome-wide significant ($P_{conditional} = 3.47\times10^{-8}$, Supplementary Table 5), although rs4072037 conditioned on rs80142782 did not ($P_{conditional} = 2.95\times10^{-6}$, Supplementary Table 4**)**; (iii) We used the haplotype inference method implemented in the plink haplotype test “--chap” option. For rs4072037 and rs80142782, there are three inferred haplotypes with frequencies greater than 1% from all four possible ones. The two models compared in the conditional haplotype likelihood ratio test are: (1) the null model: {CC} {CT, TT}; and (2) the alternative model: {CC} {CT} {TT}, where each {set} allows a unique effect. The conditional haplotype analysis demonstrated that the effect size of haplotype CC differed from that of CT among the possible haplotypes formed by these two SNPs ($P_{likelihood\ ratio} = 9.11\times10^{-9}$); (iv) Finally, it is notable that rs80142782 is Asian specific with a minor allele frequency (MAF) of 0.05 in Asians but monomorphic in both European and African populations. Thus, rs80142782 appears to be a better signal specifically marked at 1q22 in Asian populations. Further validation in additional studies

with even larger sample sizes will be required to determine if these two SNPs are truly independent signals tagging two different causal variants.

The previously reported SNP marker, rs4072037 at 1q22, is a synonymous SNP in *MUC1* which is a member of the mucin family that collectively forms the protective mucous barrier on epithelial surfaces. Its expression was highest in stomach among all normal tissues examined by the Genotype-Tissue Expression (GTEx) project (see URLs). Evidence suggests that rs4072037 is the functional variant for this locus because it alters transcriptional regulation and determines splice variants in *MUC1*[15]. Although *MUC1* is a putative candidate gene for gastric cancer risk, it is also interesting to note that GTEx data show that rs4072037 is an eQTL for several neighboring genes (including *THBS3*, *GBAP1*, *GBA*, and *RP11-263K19.4*) in other tissues (Supplementary Table 5**)**. Rs80142782 may act on the *ASH1L* gene based on its close proximity. *ASH1L* encodes a member of the trithorax group of transcriptional activators and functions as an epigenetic regulator by histone methylation (H3K4 methyltransferase), and is frequently altered in lung cancer tumors and cell lines[21, 22], esophageal squamous cell carcinoma tumor tissue[23], and colorectal cancer cell lines[24]. It was also implicated in inflammatory autoimmune disease[25].

DISCUSSION

Our study identified a new risk locus at 5q14.3 marked by rs7712641 which lies in the intron of a lncRNA with little known prior functional characterization. Other lncRNAs implicated in cancers include *PCA3* and *PCGEM1* in prostate tumor[26], and *MALAT1* in tumors of the colorectum, liver, pancreas, lung, breast, and prostate[27, 28]. It is remarkable that a recent comprehensive transcriptome analysis nominated a total of 7942 lineage- or cancer-associated lncRNA genes[29], for which further functional investigations are warranted. Overexpression or knockdown of this lncRNA may be informative in identifying target genes through analysis of differential gene expression profiles in noncardia gastric tumor cell lines.

Our analysis also revealed an apparently stronger associated SNP (rs80142782) at 1q22 than the previously identified rs4072037. Our data indicates that the association with rs80142782 is independent of rs4072037. Both the new locus at 5q14.3 marked by rs7712641 and the new independent signal at 1q22 marked by rs80142782 could contribute to epigenome regulation. Haploreg data show that both of these SNPs (or SNPs in high LD with them) locate to sites of multiple regulatory elements, including promoter histone marks, enhancer histone marks, and DNAse hypersensitivity (Supplementary Table 6). Further functional validation studies are warranted to understand the contribution of these susceptibility alleles to gastric carcinogenesis.

Supplementary Table 7 shows the results from our meta-analysis of two GWAS scans for previously reported variants from the literature. Notably we confirmed a prior independent GWAS report[9] of an association between multiple SNPs in *PSCA* at 8q24.3 and risk of noncardia gastric cancer. We also confirmed the association for rs13361707 in *PRKAA1*[11], but there was no additional evidence to support an association for rs9841504 in *ZBTB20* with noncardia gastric cancer in NCI data (P=0.27). Recently, Mocellin *et al.*[30] collected

published data and nominated a list of 11 SNPs at eight loci with a high level of cumulative evidence for susceptibility to gastric cancer. Among these 11 SNPs, four (at 2q33.1, 3p24.1, 6p21.33 and 11q13.2 respectively) were identified beyond those previously established in GWAS findings, but none of these SNPs was associated ($P < 0.05$) with gastric noncardia cancer risk in the current meta-analysis.

In summary, by combining two preexisting GWAS scans of noncardia gastric cancer to increase the sample size for the discovery stage and adding over 8,000 individuals for further replication, we identified two novel loci. In the future, additional studies are warranted with larger sample size and/or with a design that considers heterogeneity of the gastric cancer where, for example, the Cancer Genome Atlas (TCGA) recently reported four molecular subtypes for gastric cancer based on multi-omics profiling analyses[31].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

FUNDING

This work was supported by the Intramural Research Program of the National Cancer Institute, the National Institutes of Health, the Division of Cancer Epidemiology and Genetics, and the Center for Cancer Research.

This work was also supported in part by the National Basic Research Program (973) (2013CB910304), Program of National Natural Science Foundation of China (81230067), National Natural Science Foundation of China (81573228, 81521004 81422042, 81373090); Science Foundation for Distinguished Young Scholars in Jiangsu (BK20130042); Jiangsu Natural Science Foundation (BK2012443, BK2012841); Jiangsu Province Clinical Science and Technology Projects (Clinical Research Center, BL2012008); National Program for Support of Top-notch Young Professionals, National Natural Science Foundation of China (NSFC-81222038); National High-Tech Research and Development Program of China (2015AA020950); and Priority Academic Program for the Development of Jiangsu Higher Education Institutions (Public Health and Preventive Medicine).

REFERENCES

1. Ferlay, Jea. GLOBOCAN 2012 v1.0, Cancer incidence and Mortality Worldwide: IARC CancerBase No. 11. International Agency for Research on Cancer; 2013.
2. McLean MH, El-Omar EM. Genetics of gastric cancer. Nat Rev Gastroenterol Hepatol. 2014; 11:664–74. [PubMed: 25134511]
3. Karimi P, Islami F, Anandasabapathy S, et al. Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention. Cancer Epidemiol Biomarkers Prev. 2014; 23:700–13. [PubMed: 24618998]
4. Kamangar F, Qiao YL, Blaser MJ, et al. Helicobacter pylori and oesophageal and gastric cancers in a prospective study in China. Br J Cancer. 2007; 96:172–6. [PubMed: 17179990]
5. Dawsey SM, Mark SD, Taylor PR, et al. Gastric cancer and H pylori. Gut. 2002; 51:457–8. [PubMed: 12171977]
6. Gastric cancer and Helicobacter pylori: a combined analysis of 12 case control studies nested within prospective cohorts. Gut. 2001; 49:347–53. [PubMed: 11511555]
7. Mark SD, Qiao YL, Dawsey SM, et al. Prospective study of serum selenium levels and incident esophageal and gastric cancers. J Natl Cancer Inst. 2000; 92:1753–63. [PubMed: 11058618]
8. Taylor PR, Qiao YL, Abnet CC, et al. Prospective study of serum vitamin E levels and esophageal and gastric cancers. J Natl Cancer Inst. 2003; 95:1414–6. [PubMed: 13130117]

9. Sakamoto H, Yoshimura K, Saeki N, et al. Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer. Nat Genet. 2008; 40:730–40. [PubMed: 18488030]
10. Abnet CC, Freedman ND, Hu N, et al. A shared susceptibility locus in PLCE1 at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. Nat Genet. 2010; 42:764–7. [PubMed: 20729852]
11. Shi Y, Hu Z, Wu C, et al. A genome-wide association study identifies new susceptibility loci for non-cardia gastric cancer at 3q13.31 and 5p13.1. Nat Genet. 2011; 43:1215–8. [PubMed: 22037551]
12. Hu N, Wang Z, Song X, et al. Genome-wide association study of gastric adenocarcinoma in Asia: a comparison of associations between cardia and non-cardia tumours. Gut. 2015
13. Xu Q, Yuan Y, Sun LP, et al. Risk of gastric cancer is associated with the MUC1 568 A/G polymorphism. Int J Oncol. 2009; 35:1313–20. [PubMed: 19885554]
14. Jia Y, Persson C, Hou L, et al. A comprehensive analysis of common genetic variation in MUC1, MUC5AC, MUC6 genes and risk of stomach cancer. Cancer Causes Control. 2010; 21:313–21. [PubMed: 19924550]
15. Saeki N, Saito A, Choi IJ, et al. A functional single nucleotide polymorphism in mucin 1, at chromosome 1q22, determines susceptibility to diffuse-type gastric cancer. Gastroenterology. 2011; 140:892–902. [PubMed: 21070779]
16. Zhang H, Jin G, Li H, et al. Genetic variants at 1q22 and 10q23 reproducibly associated with gastric cancer susceptibility in a Chinese population. Carcinogenesis. 2011; 32:848–52. [PubMed: 21427165]
17. Palmer AJ, Lochhead P, Hold GL, et al. Genetic variation in C20orf54, PLCE1 and MUC1 and the risk of upper gastrointestinal cancers in Caucasian populations. Eur J Cancer Prev. 2012; 21:541–4. [PubMed: 22805490]
18. Song HR, Kim HN, Kweon SS, et al. Common genetic variants at 1q22 and 10q23 and gastric cancer susceptibility in a Korean population. Tumour Biol. 2014; 35:3133–7. [PubMed: 24254309]
19. 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:D930–4. [PubMed: 22064851]
20. Genomes Project C. Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491:56–65. [PubMed: 23128226]
21. Osada H, Tomida S, Yatabe Y, et al. Roles of achaete-scute homologue 1 in DKK1 and E-cadherin repression and neuroendocrine differentiation in lung cancer. Cancer Res. 2008; 68:1647–55. [PubMed: 18339843]
22. Liu J, Lee W, Jiang Z, et al. Genome and transcriptome sequencing of lung cancers reveal diverse mutational and splicing events. Genome Res. 2012; 22:2315–27. [PubMed: 23033341]
23. Song Y, Li L, Ou Y, et al. Identification of genomic alterations in oesophageal squamous cell cancer. Nature. 2014; 509:91–5. [PubMed: 24670651]
24. Mouradov D, Sloggett C, Jorissen RN, et al. Colorectal cancer cell lines are representative models of the main molecular subtypes of primary cancer. Cancer Res. 2014; 74:3238–47. [PubMed: 24755471]
25. Xia M, Liu J, Wu X, et al. Histone methyltransferase Ash1l suppresses interleukin-6 production and inflammatory autoimmune diseases by inducing the ubiquitin-editing enzyme A20. Immunity. 2013; 39:470–81. [PubMed: 24012418]
26. Fu X, Ravindranath L, Tran N, et al. Regulation of apoptosis by a prostate-specific and prostate cancer-associated noncoding gene, PCGEM1. DNA Cell Biol. 2006; 25:135–41. [PubMed: 16569192]
27. Lin R, Maeda S, Liu C, et al. A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas. Oncogene. 2007; 26:851–8. [PubMed: 16878148]
28. Xu C, Yang M, Tian J, et al. MALAT-1: a long non-coding RNA and its important 3' end functional motif in colorectal cancer metastasis. Int J Oncol. 2011; 39:169–75. [PubMed: 21503572]

29. Iyer MK, Niknafs YS, Malik R, et al. The landscape of long noncoding RNAs in the human transcriptome. Nat Genet. 2015; 47:199–208. [PubMed: 25599403]

30. Mocellin S, Verdi D, Pooley KA, et al. Genetic variation and gastric cancer risk: a field synopsis and meta-analysis. Gut. 2015

31. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014; 513:202–9. [PubMed: 25079317]

Significance of this study

What is already known about this subject?

Approximately 40% of all cases of gastric cancer worldwide occur in China, and this form of cancer remains one of the key public health issues in cancer prevention and control.

Several previous genome-wide association studies (GWAS) studies of gastric cancer have reported several associations for common single-nucleotide polymorphisms (SNP).

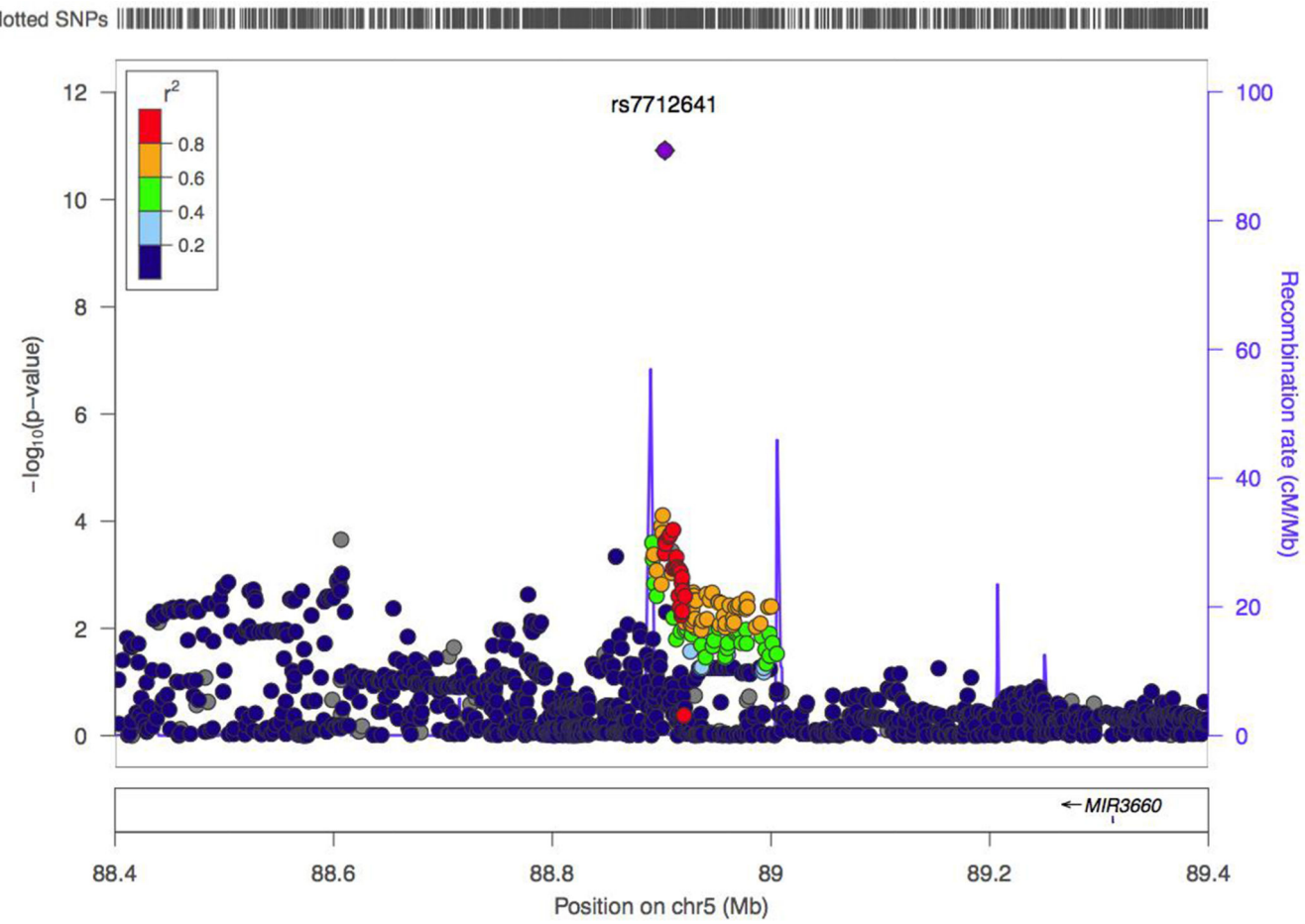
Combining together studies of moderate sample sizes will increase statistical power, so more novel signals can be exploited.

What are the new findings?

Based on a GWAS meta-analysis approach to pool two published Chinese GWAS studies, and followed by two-stage replications (more than 10,000 samples), we identified two novel signals associated with the risk of noncardia gastric cancer. The first one rs7712641 maps to the intron of the long-noncoding RNA, lnc-POLR3G-4 (*XLOC_004464*). Further analysis showed that rs7712641 had significantly lower expression in noncardia gastric tumor compared to matched normal tissue. In addition, we observed a new signal marked by rs80142782 at the 1q22 locus. It was independent of the previously reported SNP rs4072037 and is a common SNP in East Asians but is monomorphic in both European and African populations.

How might it impact on clinical practice in the foreseeable future?

The results from this GWAS meta-analysis will improve our understanding of the etiology of noncardia gastric cancer, which will further shed light on risk prediction and early detection of this malignancy.



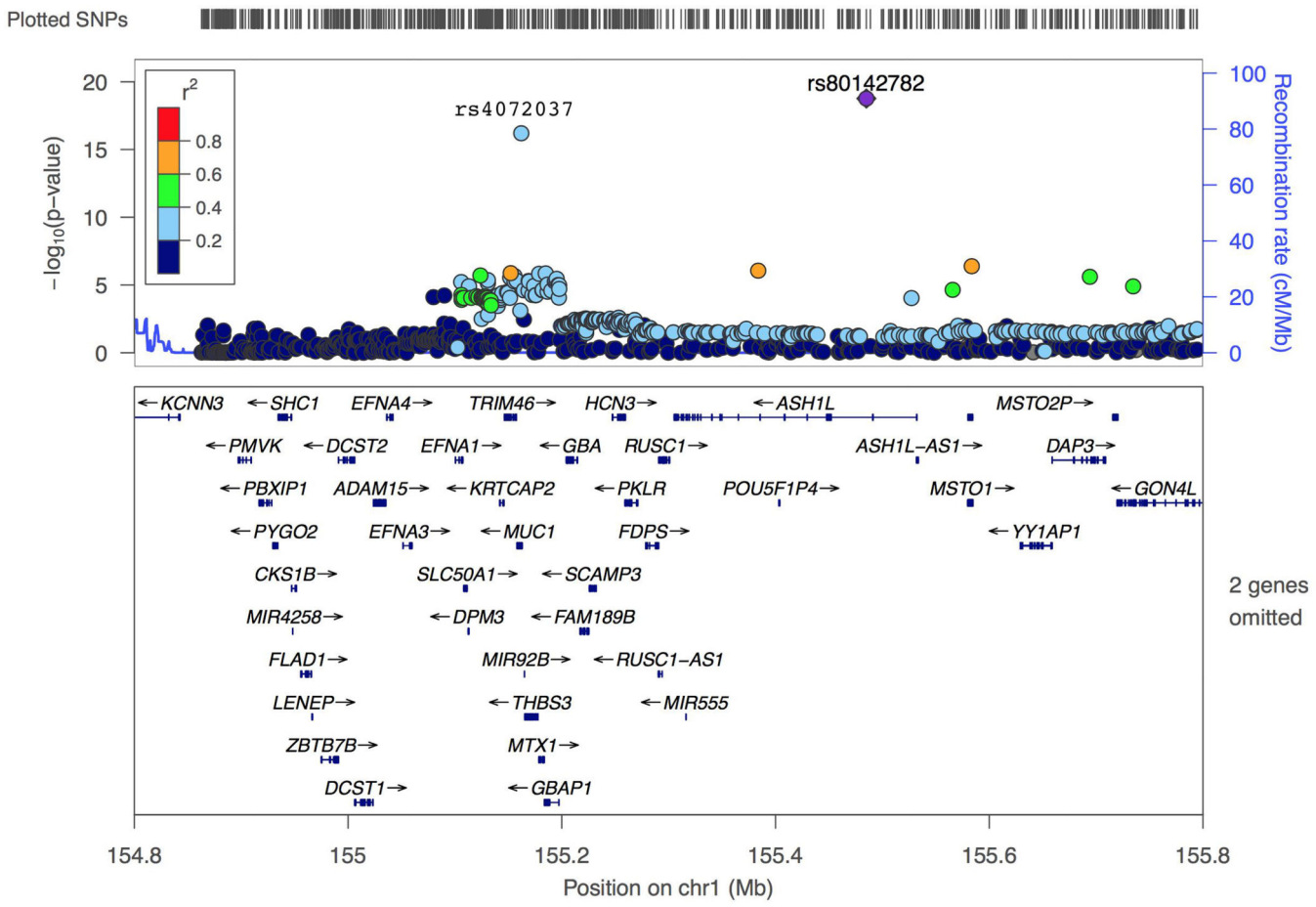


Figure 1.

Regional plots of association results (a) 5q14.3:88,400,000--89,400,000; (b) 1q22:152,940,379–153,897,376. Association results are based on a trend test in which −log10 *P*-values (y axis, left) were plotted against the chromosomal positions based on hg19 (x axis). All *P* values were based on the discovery meta-analysis Stage 1 data except for three index SNPs (rs7712641 in 1a; rs80142782 and rs4072037 in 1b) which were based on all data (Stage 1, Stage2A, and Stage 2B). The line graph shows recombination rate (y axis, right). The LD (r^2) is color coded (figure legend) based on estimates from the 1000 Genomes Mar 2012 release ASN population. The plots were generated using LocusZoom online version (URLs).

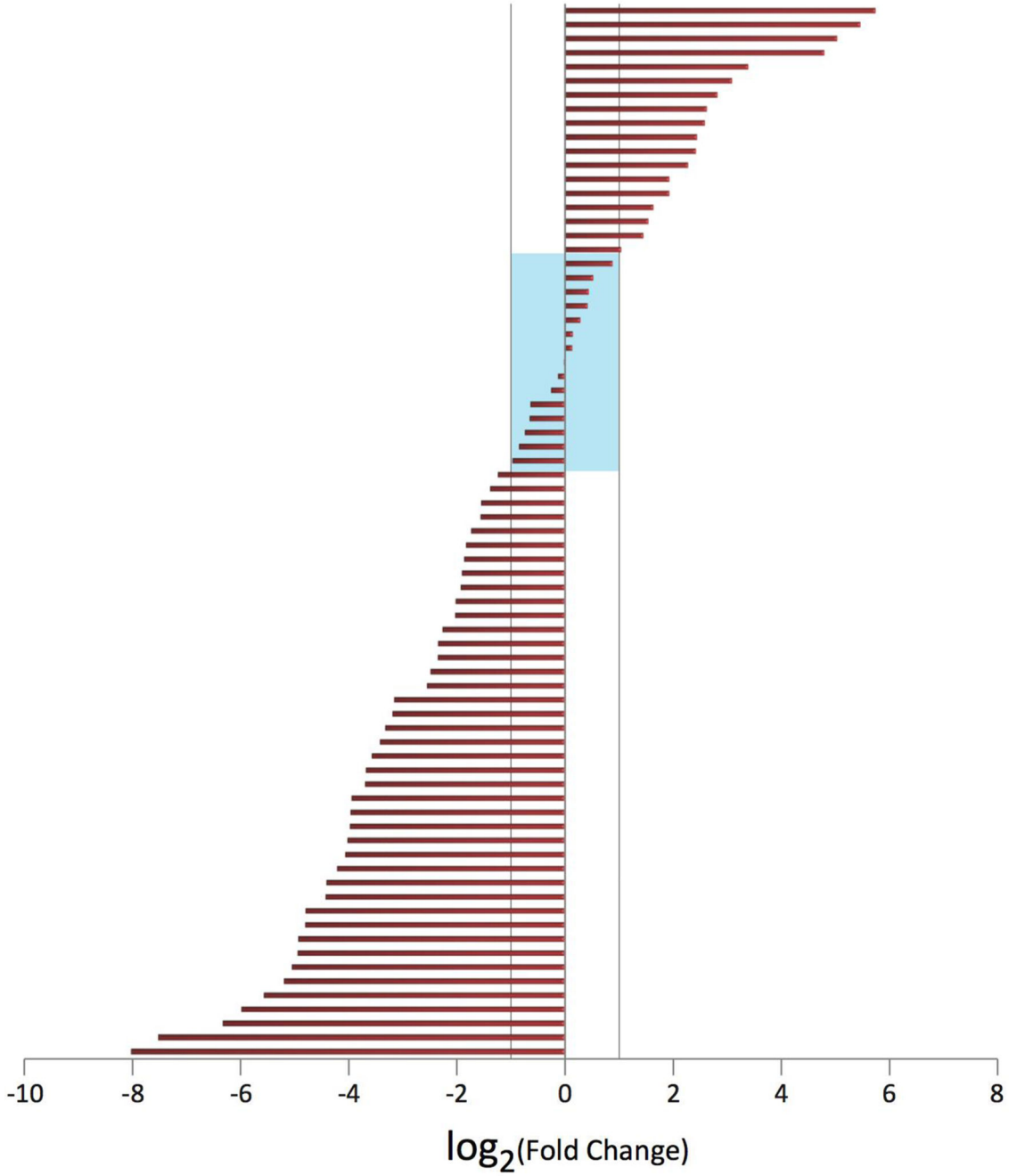


Figure 2.

Differential expression of lnc-POLR3G-4 between noncardia gastric tumor and normal tissue based on a qRT-PCR analysis. X axis is the log2 of tumor:normal fold change. All 75 pairs were sorted in increasing order of fold change and plotted along the Y axis. Two thirds of pairs show lower expression in tumor compared to normal, while expression is higher in normal compared to tumor in one third of pairs. Highlighted in the middle are 15 pairs with fold changes between 0.5 and 2 (or −1 to 1 in log2 scale). Based on all 75 pairs, the

Wilcoxon signed rank test was $P=7.2 \times 10^{-4}$, and the median tumor:normal fold change is 0.3.

Table 1

Four SNPs associated with noncardia gastric cancer risk above genome-wide significance level

Cytoband (BP)	NCBI dbSNP identifier, build 138 (referent, effect allele)	Related Gene	Group	Effect allele frequency (case,control)	OR(95% CI)	*P*	P_{het}	I^2
1q22 155485027	rs80142782 (T,C)	*ASH1L*	NCI	0.06,0.07	0.68 (0.55-0.86)	7.44E-04		
			Beijing	0.03,0.07	0.47 (0.27-0.80)	5.39E-03		
			Nanjing	0.04,0.08	0.56 (0.38-0.82)	2.55E-03		
			Stage 2a	0.05,0.09	0.61(0.50-0.73)	1.46E-07		
			Stage 2b	0.05,0.07	0.64(0.54-0.75)	1.16E-07		
			Combined		**0.62(0.56-0.69)**	**1.71E-19**	**0.66**	**0.00**
1q22 155162067	rs4072037 (T,C)	*MUC1*	NCI	0.14,0.16	0.76(0.65-0.87)	1.31E-04		
			Beijing	0.10,0.14	0.87(0.61-1.22)	4.17E-01		
			Nanjing	0.13,0.16	0.84(0.66-1.08)	1.70E-01		
			Stage 2a	0.13,0.19	0.68(0.59-0.77)	9.63E-09		
			Stage 2b	0.12,0.16	0.75(0.67-0.84)	1.22E-06		
			Combined		**0.74(0.69-0.79)**	**6.28E-17**	**0.44**	**0.00**
5q14.3 88902964	rs7712641 (C,T)	*NA*	NCI	0.44,0.46	0.88(0.79-0.97)	1.14E-02		
			Beijing	0.41,0.47	0.75(0.60-0.93)	9.41E-03		
			Nanjing	0.45,0.48	0.79(0.67-0.94)	7.54E-03		
			Stage 2a	0.43,0.48	0.81(0.73-0.90)	4.91E-05		
			Stage 2b	0.42,0.46	0.86(0.79-0.94)	4.94E-04		
			Combined		**0.84(0.80-0.88)**	**1.21E-11**	**0.56**	**0.00**
8q24.3 143761931	rs2294008 (C,T)	*PSCA*	NCI	0.31,0.29	1.16(1.03-1.31)	1.18E-02		
			Beijing	0.39,0.34	1.29(1.01-1.64)	4.30E-02		
			Nanjing	0.36,0.32	1.22(1.00-1.48)	4.78E-02		
			Stage 2a	0.29,0.25	1.16(1.04-1.30)	9.04E-03		
			Stage 2b	0.31,0.26	1.26(1.15-1.39)	1.24E-06		
			Combined		**1.20(1.15-1.28)**	**5.95E-11**	**0.76**	**0.00**

*NA: not available