nature genetics

Genome-wide association study of esophageal squamous cell carcinoma in Chinese subjects identifies susceptibility loci at *PLCE1* and *C20orf54*

We performed a genome-wide association study of esophageal squamous cell carcinoma (ESCC) by genotyping 1,077 individuals with ESCC and 1,733 control subjects of Chinese Han descent. We selected 18 promising SNPs for replication in an additional 7,673 cases of ESCC and 11,013 control subjects of Chinese Han descent and 303 cases of ESCC and 537 control subjects of Chinese Uygur-Kazakh descent. We identified two previously unknown susceptibility loci for ESCC: *PLCE1* at 10q23 ($P_{\text{Han combined for ESCC}} = 7.46 \times 10^{-56}$, odds ratio (OR) = 1.43; $P_{\text{Uygur-Kazakh for ESCC}} = 5.70 \times 10^{-4}$, OR = 1.53) and *C20orf54* at 20p13 ($P_{\text{Han combined for ESCC}} = 1.21 \times 10^{-11}$, OR = 0.86; $P_{\text{Uygur-Kazakh for ESCC}} = 7.88 \times 10^{-3}$, OR = 0.66). We also confirmed association in 2,766 cases of gastric cardia adenocarcinoma cases and the same 11,013 control subjects (*PLCE1*, $P_{\text{Han for GCA}} = 1.74 \times 10^{-39}$, OR = 1.55 and *C20orf54*, $P_{\text{Han for GCA}} = 3.02 \times 10^{-3}$, OR = 0.91). *PLCE1* and *C20orf54* have important biological implications for both ESCC and GCA. *PLCE1* might regulate cell growth, differentiation, apoptosis and angiogenesis. *C20orf54* is responsible for transporting riboflavin, and deficiency of riboflavin has been documented as a risk factor for ESCC and GCA.

Esophageal squamous cell carcinoma (ESCC) is one of the most common cancers worldwide, as well as the fourth most frequent cause of cancer-related deaths in China[1,2]. Epidemiological and etiological studies have shown that environmental and genetic factors have crucial roles in esophageal carcinogenesis[3,4]. ESCC has a striking geographic distribution worldwide, with higher prevalence in some areas of China[2], central Asia and southern Africa, where nutritional deficiencies, intake of pickled vegetables, nitrosamine-rich or mycotoxin-contaminated foods and low socioeconomic status are likely to contribute to ESCC[5–8]. In Europe and North America, heavy smoking and alcohol consumption are the main environmental risk factors for ESCC[9–11]. In high-risk areas there is a strong tendency toward familial aggregation of ESCC[12,13], suggesting that genetic susceptibility, in combination with exposure to environmental risk factors, contributes to the high rates of ESCC in these areas. The potential for genetic susceptibility is further supported by recent genome-wide association studies (GWASs) on ESCC in Japanese[14] and Chinese[15] populations, in which several susceptibility loci for ESCC have been identified. Gastric cardia adenocarcinoma (GCA) is another common type of cancer in China, which bears many similarities to ESCC in terms of concurrent geographic distribution and environmental risk factors including nutritional deficiencies, low intake of vegetables and fruit, and low socioeconomic status[16]. Moreover, smoking and alcohol seem to be minor risk factors for both ESCC and GCA in the high-incidence areas in China[16,17]. High body mass index (BMI) has been suggested as a risk factor for GCA and esophageal adenocarcinoma in western countries[3,18], but it seems to have weak effect on both ESCC and GCA in China[16]. Currently, the genetic and environmental risk factors that underlie ESCC and GCA remain unclear. This study aims to identify the genetic factors that predispose individuals to ESCC and GCA, particularly in Chinese populations, by using a large two-staged GWAS on ESCC.

In the first stage, we conducted a GWAS analysis using Illumina Human 610-Quad BeadChips in 1,089 people with ESCC and 1,763 control subjects of Chinese Han descent. After SNP- and sample-based quality control filtering, 506,666 SNPs as well as 1,077 cases and 1,733 controls were left for further analyses (**Table 1**). Principal-components analysis[19,20] confirmed that all samples came from individuals of Chinese ancestry (**Supplementary Fig. 1a–d**), although some degrees of population stratification were detected between cases and controls (genome-wide χ^2 inflation factor $\lambda_{gc} = 1.505$) (**Supplementary Fig. 1e–h**). This might be because subjects were recruited from many parts of China[21,22] (see Online Methods). We used two approaches to minimize the effect of population stratification. First, we performed GWAS analysis on the 2,810 samples (1,077 cases and 1,733 controls) using the Cochran-Armitage trend test with genomic control (GC) correction for population stratification. Second, we tested for association in 1,629 samples (937 cases and 692 controls) that were genetically matched after removing 1,181 genetically unmatched subjects by excluding samples with deviated values of principal components as population outliers from GWAS analysis (**Supplementary Fig. 2a–h**). We achieved a minor genome-wide χ^2 inflation factor ($\lambda_{gc} = 1.075$), indicating minimal overall inflation of the genome-wide statistical results (937 cases and 692 controls) due to population stratification. The results from these two approaches identified association at two loci: rs2274223 on 10q23 ($P_{\text{GWAS 1,629 samples}} < 5 \times 10^{-6}$, $P_{\text{GWAS 2,810 samples}} < 5 \times 10^{-8}$) and rs13042395 on 20p13 ($P_{\text{GWAS 1,629 samples}} < 3 \times 10^{-4}$, $P_{\text{GWAS 2,810 samples}} < 5 \times 10^{-8}$; **Supplementary Fig. 3a**,**b**). Furthermore, a quantile-quantile plot of the observed *P* values showed a clear deviation at the tail of the

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

Received 27 January; accepted 29 June; published online 22 August 2010; doi:10.1038/ng.648

Table 1 Summary of samples used in GWAS and replication studies

	Cases			Controls		
	Sample size	Mean age (s.d.)	Male/female	Sample size	Mean age (s.d.)	Male/female
ESCC GWAS[a]	1,077	61.12 (8.76)	626/451	1,733	35.55 (14.83)	950/783
ESCC replication 1[b]	7,673	60.66 (9.08)	4,802/2,871	11,013	47.70 (13.56)	5,638/5,375
ESCC replication 2[c]	303	58.00 (9.00)	214/89	537	43.29 (13.61)	243/294
ESCC combined[d]	9,053	60.61 (9.05)	5,642/3,411	13,283	46.58 (14.02)	6,831/6,452
GCA[e]	2,766	61.08 (9.02)	2,144/622	11,013	47.70 (13.56)	5,638/5,375

[a]ESCC GWAS samples are from Chinese Han ESCC cases and controls. [b]ESCC replication 1 samples are from Chinese Han ESCC cases and controls. [c]ESCC replication 2 samples are from Chinese Uygur-Kazakh ESCC cases and controls. [d]Combined data from the three ESCC groups. [e]Samples for GCA are from Chinese Han GCA cases and shared controls with replication 1.

distribution from the null distribution (the distribution expected if there were no associations). This suggested that the observed *P* values, particularly the ones in the tail of the distribution, are smaller than those expected by chance and probably reflect true genetic association (**Supplementary Fig. 3c**).

In the second stage, we performed a fast-track replication study by genotyping 18 promising SNPs (from 13 loci) in 2 independent cohorts (replication 1: 7,673 individuals with ESCC and 11,013 control subjects of Chinese Han descent; replication 2: 303 individuals with ESCC and 537 control subjects of Chinese Uygur-Kazakh descent). Considering the high geographic concurrence for ESCC and GCA in China, we also tested the 18 SNPs in 2,766 subjects with GCA and the same 11,013 control subjects from replication 1 (**Table 1** and **Supplementary Table 1**) to identify potential risk loci shared by ESCC and GCA.

Of the 18 SNPs, 3 SNPs (rs2274223, rs12263737 and rs13042395) within 2 loci were validated as ESCC susceptibility loci, showing independent association evidence ($P < 0.01$) in the 2 replication samples, and highly significant evidence ($P_{combined} < 5 \times 10^{-8}$) in the combined samples (GWAS: 1,077 subjects with ESCC and 1,733 control subjects, and replication 1). As SNP rs12263737 was highly correlated with SNP rs2274223 and could be completely represented by the latter ($D' = 0.989/r^2 = 0.978$ according to our GWAS data, and $D' = 1/r^2 = 1$ in HapMap CHB data), the 3 significant SNPs represented 2 previously unknown susceptibility loci for ESCC: rs2274223 on 10q23 ($P_{Han\ combined} = 7.46 \times 10^{-56}$, OR = 1.43, 95% CI = 1.37–1.49) and rs13042395 on 20p13 ($P_{Han\ combined} = 1.21 \times 10^{-11}$, OR = 0.86, 95% CI = 0.82–0.90; **Table 2**). We also validated these two loci in a Chinese Uygur-Kazakh population (rs2274223, $P = 5.70 \times 10^{-4}$, OR = 1.53, 95% CI = 1.21–1.95; rs13042395, $P = 7.88 \times 10^{-3}$, OR = 0.66, 95% CI = 0.49–0.88; **Table 2**). Further analyses showed that both loci were also highly associated with GCA in a Chinese Han population (rs2274223, $P = 1.74 \times 10^{-39}$, OR = 1.55, 95% CI = 1.45–1.66; rs13042395, $P = 3.02 \times 10^{-3}$, OR = 0.91, 95% CI = 0.85–0.97; **Table 2**). Logistic regression analysis adjusted for gender and age did not show any significant difference in association at 10q23 and 20p13 (data not shown). Therefore, we report the association results without gender and age adjustment in this study. These results show that genetic variations at 10q23 and 20p13 contribute significantly to risk for ESCC and GCA in Chinese Han and Uygur-Kazakh populations.

To investigate the contribution of confounding factors, such as gender, age, alcohol consumption, smoking and BMI, to the risk for ESCC, we further evaluated the effects of rs2274223 and rs13042395 by conducting stratified analyses in samples from the ESCC replication 1 group. These factors did not significantly alter the effects of the two SNPs on the risk of ESCC (**Supplementary Fig. 4a**,**b**).

To identify the potential susceptibility genes that underlie the two previously unknown associations, we investigated the patterns of recombination and linkage disequilibrium (LD) around the risk-associated SNPs and the gene(s) located within each region harboring the association (**Figs. 1** and **2**). One gene at each locus was implicated: phospholipase C epsilon 1 (*PLCE1*) at 10q23 and chromosome 20 open reading frame 54 (*C20orf54*) at 20p13.

PLCE1 encodes a phospholipase that hydrolyzes phosphatidylinositol 4,5-bisphosphate to 1,2-diacylglycerol and inositol 1,4,5-trisphosphate, and thereby contributes to intracellular signaling. 1,2-Diacylglycerol regulates a variety of target proteins, including protein kinase C (PKC) isozymes[23]. The SNP rs2274223 is a nonsynonymous SNP (causing an amino acid change from His to Arg) in the 26th exon of *PLCE1*. Moreover, PLCE1 contains several Ras binding domains for small G-proteins of the Ras family and acts as an effector of GTPases (Ras, Rap1 and Rap2)[24,25] involved in regulating cell growth, differentiation, apoptosis and angiogenesis[26]. Recent studies have reported that *PLCE1* has an oncogenic role in skin and intestinal carcinogenesis through inflammation signaling pathways[27,28], and in head and neck squamous cell carcinoma progression by binding Ras family small GTPases[24]. Our preliminary immunohistochemical analysis showed that more cells contained PLCE1 in ESCC and GCA

Table 2 Association evidence for two SNPs at 10q23 and 20p13 with ESCC and GCA

Locus	Samples	Allele frequency		OR (95% CI) (allele)	OR (95% CI) (homozygous)	OR (95% CI) (heterozygous)	P_{HWE}[a]	*P* (Armitage)
		Cases	Controls					
10q23 rs2274223 [G]	ESCC GWAS (Chinese Han)	0.27	0.20	1.47 (1.30–1.67)	2.02 (1.80–2.26)	1.44 (1.36–1.53)	–	2.81×10^{-9}
	ESCC replication 1 (Chinese Han)	0.29	0.22	1.42 (1.36–1.49)	2.02 (1.78–2.28)	1.43 (1.34–1.52)	0.68	7.35×10^{-48}
	ESCC replication 2 (Chinese Uygur-Kazakh)	0.26	0.18	1.53 (1.21–1.95)	2.10 (1.10–4.02)	1.59 (1.11–2.16)	0.47	5.70×10^{-4}
	ESCC (All Chinese Han)	–	–	1.43 (1.37–1.49)	1.97 (1.42–2.73)	1.52 (1.29–1.79)	–	7.46×10^{-56}
	GCA (Chinese Han)	0.30	0.22	1.55 (1.45–1.66)	2.34 (1.99–2.75)	1.59 (1.46–1.74)	0.68	1.74×10^{-39}
20p13 rs13042395 [T]	ESCC GWAS (Chinese Han)	0.24	0.32	0.66 (0.58–0.74)	0.75 (0.67–0.84)	0.84 (0.79–0.89)	–	1.51×10^{-11}
	ESCC replication 1 (Chinese Han)	0.28	0.30	0.89 (0.85–0.94)	0.81 (0.72–0.91)	0.89 (0.83–0.94)	0.42	5.30×10^{-6}
	ESCC replication 2 (Chinese Uygur-Kazakh)	0.12	0.17	0.66 (0.49–0.88)	0.12 (0.03–0.51)	0.87 (0.62–1.23)	0.33	7.88×10^{-3}
	ESCC (All Chinese Han)	–	–	0.86 (0.82–0.90)	0.47 (0.35–0.64)	0.63 (0.53–0.74)	–	1.21×10^{-11}
	GCA (Chinese Han)	0.28	0.30	0.91 (0.85–0.97)	0.81 (0.69–0.95)	0.91 (0.84–1.00)	0.42	3.02×10^{-3}

[a]*P* value of HWE in controls.

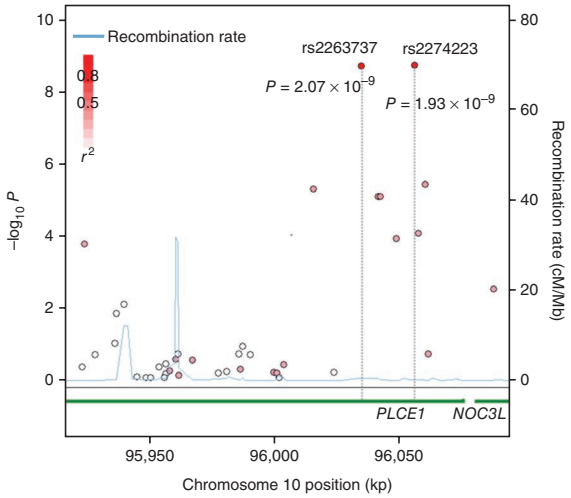

Figure 1 Scatter plot of the association evidence in 10q23 (*PLCE1*) for ESCC. The *P* values of SNPs (shown as –log10 values on *y* axis, from the genome-wide single-marker association analysis) are plotted against their map positions (*x* axis). The color of each SNP spot reflects its r^2 with the top SNP (large red circle) within each association locus, changing from red to white. Estimated recombination rates (based on the combined CHB and JPT samples from the HapMap project) are plotted in light blue. Gene annotations were adapted from the University of California at Santa Cruz Genome Browser (http://genome.ucsc.edu/).

tissues than in normal tissues (80% versus 36%, $\chi^2 = 55.72$, $P < 0.01$ in ESCC and 72% versus 23%, $P < 0.01$, $\chi^2 = 10.47$ in GCA), which further supports the idea that *PLCE1* contributes to esophageal and gastric cardia carcinogenesis (**Supplementary Fig. 5**).

C20orf54, also known as human riboflavin transporter 2 (*RFT2*), encodes an open reading frame protein. The biological function of human RFT2 has not been characterized, but it has an 83% similarity to rat riboflavin transporter 2 (rat RFT2), which is highly expressed in rat small intestine and colon[29]. Rat RFT2 is a riboflavin transporter and has an important role in the intestinal absorption of riboflavin[29]. In addition, rat RFT2 is homologous to rat RFT1. Both rat and human *RFT1* cDNAs have an opening reading frame, and they encode 448- and 450-amino acid proteins, respectively, that show 81.1% identity and 96.4% similarity to one another[30]. The role of riboflavin in cellular homeostasis has been well documented. In particular, riboflavin deficiency is associated with an increased risk of ESCC and GCA[6] and riboflavin supplementation has been reported to reduce the risk of ESCC and GCA[31]. Therefore, human RFT2 is likely to have an important role in esophageal and gastric cardia carcinogenesis that involves modulating riboflavin absorption.

A recent GWAS in the Japanese population showed that rs671 in *ALDH2* and rs1229984 in *ADH1B* were highly associated with ESCC[14]. *ALDH2* and *ADH1B* encode dehydrogenases involved in alcohol metabolism. We typed rs671 in 4,068 subjects with ESCC and 4,090 control subjects shared with replication 1 to determine its association in the Chinese population. In our GWAS data, rs671 showed marginal evidence for association with ESCC ($P = 1.97 \times 10^{-2}$, OR = 0.79, 95% CI = 0.65–0.96), but we did not find any evidence of association in replication 1 for rs671 ($P = 0.601$; OR = 1.02, 95% CI = 0.93–1.12). As our study had more than 90% power to detect the association of rs671 with ESCC, the differences between these two populations might be due to differences in allele and genotype frequency (**Supplementary Table 2a**). Notably, stratification analyses revealed that rs671 showed significant association with ESCC in smoker and alcohol drinker subgroups ($P = 0.017$ and $P = 0.009$, respectively; **Supplementary Table 2b**). The heterogeneity analyses revealed that rs671 increased the risk of ESCC in the smoker ($P_{het} = 0.01$) and drinker populations ($P_{het} = 0.008$; **Supplementary Fig. 6**). This is partially consistent with the study conducted in the Japanese population[14] and suggests that *ALDH2* might modify genetic risk to ESCC differently in the Chinese and Japanese populations.

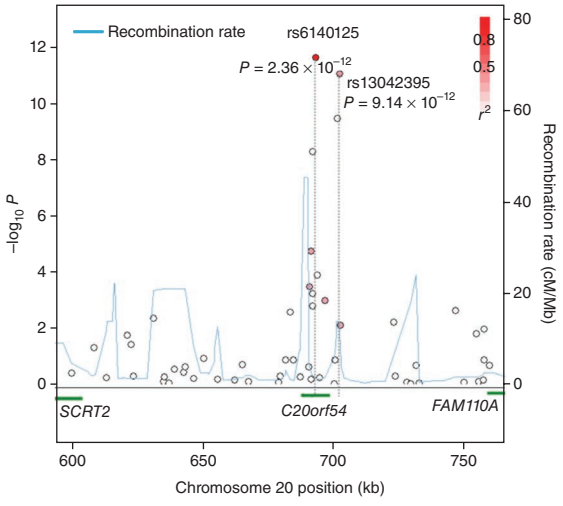

Figure 2 Scatter plot of the association evidence within 20p13 (*C20orf54*) for ESCC. The *P* values of SNPs (shown as –log10 values in *y* axis, from the genome-wide single-marker association analysis) are plotted against their map positions (*x* axis). The color of each SNP spot reflects its r^2 with the top SNP (large red circle) within each association locus, changing from red to white. Estimated recombination rates (based on the combined CHB and JPT samples from the HapMap project) are plotted in light blue. Gene annotations were adapted from the University of California at Santa Cruz Genome Browser (http://genome.ucsc.edu/).

In summary, we performed a large GWAS of ESCC in Chinese populations and identified two previously unknown susceptibility genes, *PLCE1* and *C20orf54*, for both ESCC and GCA. Our study not only adds to the known genetic factors that predispose individuals to ESCC and GCA, but also highlights the importance of genetic factors and genetic heterogeneity in the development of these diseases, which should advance our understanding of the pathogenesis of ESCC and GCA.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank all the patients and their family members whose contributions made this work possible; medical students from Xinxiang Medical University for sample and data collections; Y. Xing (Xinxiang Medical University) for organizing field work for sample collections and finding financial support for this study; the genotyping at Key Laboratory of Dermatology (Anhui Medical University), Ministry of Education, China, Hefei, Anhui for genotyping and statistical and bioinformatic analyses; B.G. Ma (Henan Provincial People's Hospital), W. Huang (Health Department of Henan Province), S.Y. Lian (Tumor Hospital of Linzhou), Z. Wang (Anyang Tumor Hospital), W. Zhang and J.T. Zhao (both Peoples Hospital of Heibi) for field work organization; J.H. Bian (Shantou University), J.J. Liu (Human Genetics, Genome Institute of Singapore) and I. Ding (National Cancer Institute, NIH) for discussions about the

manuscript. Supported by Xinxiang Medical University Key Scientific Program (2009-5), the National Natural Science Foundations (30670956, 30971133), 863 HighTech Key Projects (2006AA02A403, 2007AA02Z161), China Key Program on Basic Research (2007CB516812), Special Scientific Programs from Science and Technology Department (2009-8), Health Department (2009-10) and Education Department (2008-7) of Henan Province and Anhui Provincial Special Scientific Program (2007-7).

AUTHOR CONTRIBUTIONS

L.-D.W. and X.-J.Z. conceived of this study and obtained financial support. L.-D.W. and X.-J.Z. designed the study. Subject recruitment, biological material collection and handling along with DNA extraction in Henan, Hebei and Shanxi Provinces were supervised by L.-D.W., F.-Y.Z., Xue-Min Li and carried out by J.-L.L., X.S., Y.J., H.-L. Liu, J.-M.L., G.-Q.K., H.Q., J. Cui, L.-Q.Z., J.-Z.Y., X.-C.L., J.-L.R., Z.-C.L., W.-J.G., W.W., Y.-R.Z., W.-P.W., B.-P.C., S.-W.R., D.L., J.-W.K., Z.-M.F., S.-L.Z., Z.-G.G., X.-K.Z., N.L., Y.-H.A., F.-F.S., W.-Y.C., S.S., T.G., Jing Huang, C.Y., Jia Huang, Y. Wu, W.-B.Y., C.-W.F., H.-L. Li, L.Y., Y. Wang, J.-Y.T., Y. Lu, Y.Y., W.-L.Z., M.L., W.-J.F., X.Y., H.-J.W., S.-L.H., J. Chen, M.H., H.-Y.W., P.Z., Xiu-Min Li, J.-C.D., G.-L.X., R.W., M.G., Z.-W.C., Z.-Q.Y., Q.L., L.-Q.Y., F.-G.Z., X.-F.Y., X.-S.F., Yin Li, S.-G.G., J.-P.W., J.W., Z.-Q.B., L.G., J.-L.C., G.-C.D., X.Z., Y.-F.Z., Z.-M.D., X.-Y.J., D.-M.F., A.-F.J., J.-S.W., Q.Z., Q.-X.Z., F.-B.C., C.-D.L., J.-J.M., Z.-L.F., T.-J.L., S.-L.Q., Z.-G.C., J.-C.W., T.-Y.W. and X.H. The following authors from the various collaborating groups undertook assembly of case-control series in their respective regions and collected samples and data: I.S., F.L. and X.-M.L. in Xinjiang Uygur Autonomous Region; Q.Q. and L.-T.B. in Inner Mongolia Autonomous Region; Z.-Y.S., E.-M.L., L.-Y.X. and Z.-Y.W. in Guangdong Province; W.-K.C. in Jiangsu Province; G.-Y.L. in Shaanxi Province; Z.W. in Shandong Province; W.-J.Y. in Ningxia Hui Tribe Autonomous Region; L.-Q.C. in Sichuan Province; R.-B.L. in Fujian Province; B.L., F.-S.G., Yu Lan and Y.-J.F. in Beijing; Z.-M.C. in Guangdong Province; H.L. in Yunnan Province; S.-Q.C., W.C.Y., J.-Y.H., D.X., L.W. and Q.-P.K. participated in study design, results discussion and manuscript preparation; L.-D.S., H.-F.Z., Z.Z., X.-B.Z., X.S., G.-Q.K. and L.-Q.Z. performed the two-staged genotyping and sequencing, data manipulation, statistical and bioinformatic analyses. The manuscript was drafted by L.-D.W., and all authors contributed to the final draft.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Parkin, D.M., Bray, F., Ferlay, J. & Pisani, P. Global cancer statistics, 2002. *CA Cancer J. Clin.* **55**, 74–108 (2005).
2. Ke, L. Mortality and incidence trends from esophagus cancer in selected geographic areas of China circa 1970–90. *Int. J. Cancer* **102**, 271–274 (2002).
3. Hongo, M., Nagasaki, Y. & Shoji, T. Epidemiology of esophageal cancer: Orient to Occident. Effects of chronology, geography and ethnicity. *J. Gastroenterol. Hepatol.* **24**, 729–735 (2009).
4. Cheung, W.Y. & Liu, G. Genetic variations in esophageal cancer risk and prognosis. *Gastroenterol. Clin. North Am.* **38**, 75–91 (2009).
5. Munoz, N. & Day, N.E. in *Cancer Epidemiology and Prevention* (eds. Schottenfeld D. & Fraumeni J.F.) 681–706 (Oxford Univ. Press, New York, 1996).
6. Siassi, F. & Ghadirian, P. Riboflavin deficiency and esophageal cancer: a case control-household study in the Caspian Littoral of Iran. *Cancer Detect. Prev.* **29**, 464–469 (2005).
7. Yang, C.X. *et al.* Risk factors for esophageal cancer: a case-control study in Southwestern China. *Asian Pac. J. Cancer Prev.* **6**, 48–53 (2005).
8. Sun, G. *et al.* Fumonisin B1 contamination of home-grown corn in high-risk areas for esophageal and liver cancer in China. *Food Addit. Contam.* **24**, 181–185 (2007).
9. Morita, M. *et al.* Alcohol drinking, cigarette smoking, and the development of squamous cell carcinoma of the esophagus: epidemiology, clinical findings, and prevention. *Int. J. Clin. Oncol.* **15**, 126–134 (2010).
10. Messmann, H. Squamous cell cancer of the oesophagus. *Best Pract. Res. Clin. Gastroenterol.* **15**, 249–265 (2001).
11. Brown, L.M. *et al.* Excess incidence of squamous cell esophageal cancer among US Black men: role of social class and other risk factors. *Am. J. Epidemiol.* **153**, 114–122 (2001).
12. Zhang, G.H. *et al.* Genetic heterogeneity of oesophageal cancer in high-incidence areas of southern and northern china. *PLoS ONE* **5**, e9668 (2010).
13. Zhang, W. *et al.* Segregation analysis of esophageal cancer in a moderately high-incidence area of northern China. *Am. J. Hum. Genet.* **67**, 110–119 (2000).
14. Cui, R. *et al.* Functional variants in ADH1B and ALDH2 coupled with alcohol and smoking synergistically enhance esophageal cancer risk. *Gastroenterology* **137**, 1768–1775 (2009).
15. Hu, N. *et al.* Genome-wide association study in esophageal cancer using GeneChip mapping 10K array. *Cancer Res.* **65**, 2542–2546 (2005).
16. Tran, G.D. *et al.* Prospective study of risk factors for esophageal and gastric cancers in the Linxian general population trial cohort in China. *Int. J. Cancer* **113**, 456–463 (2005).
17. Kamangar, F., Chow, W.H., Abnet, C.C. & Dawsey, S.M. Environmental causes of esophageal cancer. *Gastroenterol. Clin. North Am.* **38**, 27–57 (2009).
18. Chow, W.H. *et al.* Body mass index and risk of adenocarcinomas of the esophagus and gastric cardia. *J. Natl. Cancer Inst.* **90**, 150–155 (1998).
19. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
20. Zhang, F.R. *et al.* Genomewide association study of leprosy. *N. Engl. J. Med.* **361**, 2609–2618 (2009).
21. Chen, J. *et al.* Genetic structure of the Han Chinese population revealed by genome-wide SNP variation. *Am. J. Hum. Genet.* **85**, 775–785 (2009).
22. Xu, S. *et al.* Genomic dissection of population substructure of Han Chinese and its implication in association studies. *Am. J. Hum. Genet.* **85**, 762–774 (2009).
23. Rhee, S.G. Regulation of phosphoinositide-specific phospholipase C. *Annu. Rev. Biochem.* **70**, 281–312 (2001).
24. Bunney, T.D., Baxendale, R.W. & Katan, M. Regulatory links between PLC enzymes and Ras superfamily GTPases: signalling via PLCepsilon. *Adv. Enzyme Regul.* **49**, 54–58 (2009).
25. Bunney, T.D. *et al.* Structural and mechanistic insights into ras association domains of phospholipase C epsilon. *Mol. Cell* **21**, 495–507 (2006).
26. Bourguignon, L.Y., Gilad, E., Brightman, A., Diedrich, F. & Singleton, P. Hyaluronan-CD44 interaction with leukemia-associated RhoGEF and epidermal growth factor receptor promotes Rho/Ras co-activation, phospholipase C epsilon-Ca2+ signaling, and cytoskeleton modification in head and neck squamous cell carcinoma cells. *J. Biol. Chem.* **281**, 14026–14040 (2006).
27. Bai, Y. *et al.* Crucial role of phospholipase Cepsilon in chemical carcinogen-induced skin tumor development. *Cancer Res.* **64**, 8808–8810 (2004).
28. Li, M., Edamatsu, H., Kitazawa, R., Kitazawa, S. & Kataoka, T. Phospholipase Cepsilon promotes intestinal tumorigenesis of Apc(Min/+) mice through augmentation of inflammation and angiogenesis. *Carcinogenesis* **30**, 1424–1432 (2009).
29. Yamamoto, S. *et al.* Identification and functional characterization of rat riboflavin transporter 2. *J. Biochem.* **145**, 437–443 (2009).
30. Yonezawa, A., Masuda, S., Katsura, T. & Inui, K. Identification and functional characterization of a novel human and rat riboflavin transporter, RFT1. *Am. J. Physiol. Cell Physiol.* **295**, C632–C641 (2008).
31. He, Y. *et al.* Effect of riboflavin-fortified salt nutrition intervention on esophageal squamous cell carcinoma in a high incidence area, China. *Asian Pac. J. Cancer Prev.* **10**, 619–622 (2009).

Li-Dong Wang[1,2,*], Fu-You Zhou[2,3], Xue-Min Li[2,4], Liang-Dan Sun[5], Xin Song[1,2], Yan Jin[1], Jiang-Man Li[1,2], Guo-Qiang Kong[2], Hong Qi[2], Juan Cui[2], Lian-Qun Zhang[1], Jie-Zhi Yang[2], Ji-Lin Li[1,6], Xing-Chuan Li[2], Jing-Li Ren[2], Zhi-Cai Liu[7], Wen-Jun Gao[8], Ling Yuan[2,9], Wu Wei[10], Yan-Rui Zhang[11], Wei-Peng Wang[2,4], Ilyar Sheyhidin[12], Feng Li[13], Bao-Ping Chen[14], Shu-Wei Ren[15], Bin Liu[16], Dan Li[1,2], Jian-Wei Ku[17], Zong-Min Fan[2], Sheng-Li Zhou[1], Zhi-Gang Guo[1], Xue-Ke Zhao[1], Na Liu[1], Yong-Hong Ai[1], Fang-Fang Shen[1], Wen-Yan Cui[1], Shuang Song[2], Tao Guo[2], Jing Huang[2], Chao Yuan[2], Jia Huang[2], Yue Wu[2], Wen-Bin Yue[18], Chang-Wei Feng[2], Hong-Lei Li[2], Yan Wang[2], Jin-Ya Tian[2], Yue Lu[2], Yi Yuan[1], Wen-Liang Zhu[2], Min Liu[2], Wen-Jing Fu[2], Xia Yang[2], Han-Jing Wang[2], Suo-Li Han[2], Jie Chen[2], Min Han[19], Hai-Yan Wang[1], Peng Zhang[1], Xiu-Min Li[1], Jin-Cheng Dong[2], Guo-Lan Xing[2], Ran Wang[2], Ming Guo[2], Zhi-Wei Chang[2], Hai-Lin Liu[1], Li Guo[2], Zhi-Qing Yuan[1], Hai Liu[20], Qin Lu[2], Liu-Qin Yang[21], Fu-Guo Zhu[19], Xiu-Feng Yang[22], Xiao-Shan Feng[23], Zhou Wang[24], Yin Li[9], She-Gan Gao[23], Qirenwang Qige[25], Long-Tang Bai[25], Wen-Jun Yang[26], Guang-Yan Lei[27], Zhong-Ying Shen[28], Long-Qi Chen[29], En-Min Li[28], Li-Yan Xu[28], Zhi-Yong Wu[28], Wei-Ke Cao[30], Jian-Po Wang[2,3],

Zhi-Qin Bao[4], Ji-Li Chen[4], Guang-Cheng Ding[2], Xiang Zhuang[2], Ying-Fa Zhou[2], Hou-Feng Zheng[5], Zheng Zhang[5], Xian-Bo Zuo[5], Zi-Ming Dong[2], Dong-Mei Fan[2], Xin He[2], Jin Wang[2], Qi Zhou[2], Qin-Xian Zhang[2], Xin-Ying Jiao[6], Shi-Yong Lian[7], Ai-Fang Ji[10], Xiao-Mei Lu[12], Jin-Sheng Wang[10], Fu-Bao Chang[8], Chang-Dong Lu[3], Zhi-Guo Chen[1], Jian-Jun Miao[4], Zeng-Lin Fan[4], Ruo-Bai Lin[31], Tai-Jiang Liu[32], Jin-Chang Wei[6], Qing-Peng Kong[33], Yu Lan[34], Yu-Jing Fan[34], Fu-Sheng Gao[16], Tian-Yun Wang[1], Dong Xie[35], Shu-Qing Chen[36], Wan-Cai Yang[1,38], Jun-Yan Hong[36,38], Liang Wang[1], Song-Liang Qiu[2], Zhi-Ming Cai[37] & Xue-Jun Zhang[5]

[1]Cancer Research Center, Xinxiang Medical University, Xinxiang, Henan, China. [2]Henan Key Laboratory for Esophageal Cancer Research, Department of Basic Oncology and Pathology at College of Medicine, The First and The Second Affiliated Hospital, Zhengzhou University, Zhengzhou, Henan, China. [3]Department of Thoracic Surgery and Tumor Prevention Treatment, Anyang Tumor Hospital, Anyang, Henan, China. [4]Department of Pathology and Thoracic Surgery, Centre for Health Screening and Endoscopy, Cixian Hospital, Cixian, Hebei, China. [5]Key Laboratory of Dermatology (Anhui Medical University), Ministry of Education, China, Hefei, Anhui, China. [6]Department of Pathology and Thoracic Surgery, Linzhou Esophageal Cancer Hospital, Linzhou, Henan, China. [7]Department of Thoracic Surgery, Tumor Hospital of Linzhou, Linzhou, Henan, China. [8]Department of Thoracic Surgery, People's Hospital and Chinese Medicine Hospital of Linzhou, Linzhou, Henan, China. [9]Department of Thoracic Surgery and Radiotherapy, Cancer Hospital of Henan Province, Zhengzhou, Henan, China. [10]Institute of Hematologic Disease and Department of Pathology, Changzhi Medical University, Changzhi, Shanxi, China. [11]Department of Gastroenterology, Henan Provincial People's Hospital, Zhengzhou, Henan, China. [12]Department of Thoracic Surgery and Medical Research Center, The First Affiliated Hospital, Xinjiang Medical University, Urumqi, Xinjiang, China. [13]Department of Pathology, Shihezi University School of Medicine, Shihezi, Xinjiang, China. [14]Department of Nephrology, Huaihe Hospital, Henan University, Kaifeng, Henan, China. [15]Department of Oncology, Xinyang Central Hospital, Xinyang, Henan, China. [16]Department of Gastroenterology, Beijing Tongren Hospital Affiliated to Capital Medical University, Beijing, China. [17]Department of Oncology, The First Affiliated Hospital of Nanyang Medical College, Nanyang, Henan, China. [18]Department of Oncology, People's Hospital of Puyang, Puyang, Henan, China. [19]Department of Gastroenterology and Thoracic Surgery, The First People's Hospital of Shangqiu, Shangqiu, Henan, China. [20]Department of Gastroenterology, The Third People's Hospital of Yunnan Province, Kunming, Yunnan, China. [21]Department of Radiotherapy, Central Hospital of Xinxiang, Xinxiang, Henan, China. [22]Department of Surgery, Hebi Dahejian Hospital, Hebi, Henan, China. [23]Department of Oncology, The First Affiliated Hospital, Henan University of Science and Technology, Luoyang, Henan, China. [24]Department of Thoracic Surgery, Provincial Hospital Affiliated to Shandong University, Jinan, Shandong, China. [25]Department of Internal Mongolia Medicine, The Affiliated Hospital, Inner Mongolia Medical College, Huhhot, Inner Mongolia, China. [26]Department of Biotechnology, Ningxia Medical University, Yinchuan, Ningxia, China. [27]Department of Thoracic Surgery, Tumor Hospital of Shaanxi Province, Xi'an, Shaanxi, China. [28]Department of Oncologic Pathology, and Department of Biochemistry and Molecular Biology, Medical College of Shantou University, Shantou, Guangdong, China. [29]Department of Thoracic and Cardiovascular Surgery, West China Hospital, Sichuan University, Chengdu, Sichuan, China. [30]Central Laboratory, Huaian No.1 Hospital, Nanjing Medical University, Huaian, Jiangsu, China. [31]Department of Thoracic Surgery, Fujian Medical University Union Hospital, Fuzhou, Fujian, China. [32]Department of Surgery, The Red Cross Hospital of Huojia, Xinxiang, Henan, China. [33]State Key Laboratory of Genetic Resource and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China. [34]Department of Gastroenterology, Beijing Jishuitan Hospital, The Fourth Medical College of Peking University, Beijing, China. [35]Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China. [36]College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, Zhejiang, China. [37]Guangdong Key Laboratory of Male Reprodutive Medicine and Genetics, Peking University Shenzhen Hospital, Shenzhen, Guangdong, China. [38]Present addresses: Department of Pathology, University of Illinois at Chicago, Chicago, Illinois, USA (W.-C.Y.); Department of Environment and Occupational Health, School of Public Health, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey, USA (J.-Y.H.). Correspondence should be addressed to L.-D.W. (ldwang@xxmu.edu.cn).

ONLINE METHODS

Study populations. Cases and control subjects for the GWAS scan and replications were obtained through the ongoing hospital-based ESCC and GCA case-control study, a collaboration among multiple hospitals throughout the high- and low-incidence areas for ESCC and GCA in China that was established in 2007. Most of the Chinese Han samples were obtained from the high-incidence areas for ESCC and GCA in northern China. The Chinese Uygur-Kazakh samples were obtained from western China. The study was approved by each institutional and hospital ethical committee and conducted according to the principles of the Declaration of Helsinki. On the basis of our previous findings that geographic match could be a good proxy for genetic match in the Chinese population[21,22], we matched the cases and controls by geography and ethnicity to minimize the potential population stratification in our replication studies (see **Supplementary Note**).

Specifically, in the GWAS: 1,077 ESCC cases in northern China and 1,733 controls in northern and central China were identified from the ongoing hospital-based ESCC and GCA case-control study. In replication 1: 7,673 ESCC cases and 11,013 controls were identified in northern China from the ongoing hospital-based ESCC and GCA case-control study. In replication 2: 303 ESCC cases and 537 controls of Chinese Uygur-Kazakh were identified from Xinjiang Uygur Autonomous Region in western China. In the GCA study: an additional 2,766 GCA cases and 11,013 controls shared with replication 1 were identified from northern China (**Table 1**).

Genotyping and quality control in GWAS stage. The genome-wide genotyping analysis was conducted using Illumina Human 610-Quad BeadChips at the Key Laboratory of Dermatology (Anhui Medical University). Genotyping was performed as described according to manufacturer's instructions. Briefly, genomic DNA was extracted from peripheral blood leukocytes by standard procedures using Flexi Gene DNA kits (Qiagen). Approximately 200 ng of genomic DNA was used for genotyping analysis. Each sample was whole-genome amplified, fragmented, precipitated and resuspended in appropriate hybridization buffer. Denatured samples were hybridized on prepared Illumina Human 610-Quad BeadChips. After hybridization, the BeadChips oligonucleotides were extended by a single labeled base, which was detected by fluorescence imaging with an Illumina Bead Array Reader. Normalized bead intensity data obtained for each sample were loaded into the Illumina BeadStudio 3.2 software, which converted fluorescence intensities into SNP genotypes. We estimated gender for study samples using the genotypes of sex-specific SNPs in BeadStudio 3.2 software and the results fit well with the recruited gender information. The clustering of genotypes was carried out with Gen-Call software version 6.2.0.4, which assigns a quality score to each locus and an individual genotype confidence score that is based on the distance of a genotype from the center of the nearest cluster. The SNPs were excluded if they showed a call rate lower than 95% in cases or controls, a minor allele frequency <1% in the population or significant deviation from Hardy-Weinberg equilibrium in the controls ($P < 10^{-7}$). We further removed 42 samples owing to duplicated samples, low call rate (<98%) and genetic relatedness and the SNPs on the X, Y and mitochondrial chromosomes as well as the CNV probes. Finally, 506,666 SNPs in 1,077 ESCC cases and 1,733 controls were left for downstream analysis.

SNP selection for replications. The 18 targeted SNPs from our GWAS were further selected for replication on the basis of being shared among 1,629 and 2,810 samples. The detailed selection criteria were as follows: (i) showing high MAF (>0.05 both in cases and controls); (ii) having high P value in Hardy-Weinberg equilibrium test ($P \geq 0.001$ in controls); (iii) having $P < 10^{-4}$ in both the tests (Cochran–Armitage trend test in the cohort of genetically matched 1,629 samples, and GC correction-based Cochran–Armitage trend test in the cohort of 2,810 samples) of the GWAS analysis; and (iv) we recruited one additional SNP rs7230870 on 18q22.3 that was highly correlated with SNP rs4891058. Finally, 18 SNPs from 13 loci were selected for genotyping.

Furthermore, to explore whether *ALDH2* identified in the Japanese population is associated with ESCC in the Chinese Han population, rs671 was also included in the current study.

Principal components analysis. We performed principal components analysis (PCA) following the method used in the software package EIGENSTRAT[19]. The original script from EIGENSTRAT was modified so that we could extract the principal components and plot them. PCA was used to identify genetic outliers. To do this, we combined 1,077 cases and 1,733 controls (2,810 samples, after removal of samples with low call rates and familial relationships) with HapMap samples and performed PCA analysis. The HapMap samples were drawn from: Yoruba in Ibadan, Nigeria (YRI) (n = 57), Japanese in Tokyo, Japan (JPT) (n = 44), Han Chinese in Beijing, China (CHB) (n = 45) and CEPH (Utah residents with ancestry from northern and western Europe) (CEU) (n = 60). We employed a stringent criterion for the removal of genetic outliers and the PCA analysis showed that all the 2,810 samples were of Chinese origin (**Supplementary Fig. 1a–d**), but the cases and controls showed some genetic mismatch.

Genotyping and quality control in follow-up study stage. Genomic DNA was extracted as in the initial study. For the follow-up study, DNA concentration was normalized to 15–20 ng/μl (diluted in 10 mM Tris and 1 mM EDTA) with a Nanodrop Spectrophotometer (ND-1000). Genotypes for SNPs selected to identify susceptibility loci for ESCC were obtained using the Sequenom MassArray System (Sequenom iPLEX assay) at Key Laboratory of Dermatology (Anhui Medical University) according to the manufacturer's instructions. Approximately 15 ng of genomic DNA was used to genotype each sample. Locus-specific PCR and detection primers were designed using the MassARRAY Assay Design 3.0 software (Sequenom). The sample DNAs were amplified by multiplex PCR reactions, and the PCR products were then used for locus-specific single-base extension reaction. The resulting products were desalted and transferred to a 384-element SpectroCHIP array. Allele detection was performed using MALDI-TOF MS. The mass spectrograms were analyzed by the MassARRAY TYPER software (Sequenom). In addition, we genotyped SNPs selected to replicate susceptibility genes previously reported[14] using TaqMan (Applied Biosystems). To evaluate the quality of the genotype data for the replication analysis, 100 randomly selected samples from the GWAS stage were re-genotyped in replication samples by using the Sequenom and TaqMan. The concordance rate among the genotypes from the Illumina, Sequenom and TaqMan analyses was >99%. All the 19 SNPs satisfied the criteria with call rate larger than 95% both in cases and controls and with Hardy-Weinberg equilibrium in the controls ($P \geq 0.01$). Furthermore, we checked the cluster patterns of the 19 SNPs from the genotyping data from the Illumina, Sequenom and TaqMan analyses to confirm their good quality. After quality control, all the 19 SNPs were left for final analysis.

Immunohistochemical staining for PLCE1 expression in ESCC and GCA. We collected 188 samples of surgically resected ESCC with 94 matched normal esophageal tissues and 50 samples of GCA with 13 matched normal gastric cardia from northern Chinese Han populations. All the patients were untreated apart from surgery. The avidin-biotin-peroxidase complex (ABC) method[32] was used for PLCE1 protein immunostaining using anti-human PLCE1 polyclonal antibody (Sigma-Aldrich). Stratification analysis by age and gender did not significantly alter the immunostaining results of PLCE1 (data not shown).

Statistical analysis. We performed the GWAS analysis using the genotypes of 506,666 SNPs in 1,077 ESCC cases and 1,733 controls of Chinese Han origin (**Table 1**). To minimize the adverse impact of population stratifications, the analyses of genotype-phenotype in the GWAS stage were performed in two ways[20]. First, we did the GWAS analysis in the 1,077 cases and 1,733 controls using the Cochran-Armitage trend test with GC correction for population stratification. Second, the Cochran-Armitage trend test was performed in 937 cases and 692 genetically matched controls (1,629 samples created by removing 1,181 unmatched subjects, **Supplementary Fig. 2e–h**), without correction for population stratification. For both the analyses, there were no indications that population stratification caused overall inflation of genome-wide association results and the overall results of the two analyses were consistent at 10q23 and 20p13. For the replication analyses, 18 promising SNPs were selected and genotyped in two independent replication samples (replication 1: 7,673 ESCC cases and 11,013 controls of Chinese Han origin; replication 2: 303 cases and 537 controls of Chinese Uygur-Kazakh origins), with an additional 2,766 GCA

doi:10.1038/ng.648

cases of Chinese Han origins and the same 11,013 controls as in replication 1. The association analyses were performed in the two replication samples and GCA samples individually by using the Cochran-Armitage trend test, and in the combined samples of GWAS (1,077 ESCC cases and 1,733 controls) and replication 1 (7,673 ESCC cases and 11,013 controls) by using Cochran-Mantel-Hanezel stratification analysis. For rs671, which was selected to replicate susceptibility genes previously reported in a Japanese population[14], the same Cochran-Armitage trend test was performed. Heterogeneity analyses for the different subgroups of ESCC were performed using the Breslow-Day test. Recombination plots for each newly discovered susceptibility locus were generated using information from the HapMap project (CHB and JPT samples).

32. Wang, L.D., Hong, J.Y., Qiu, S.L., Gao, H. & Yang, C.S. Accumulation of p53 protein in human esophageal precancerous lesions: a possible early biomarker for carcinogenesis. *Cancer Res.* **53**, 1783–1787 (1993).