Am J Cancer Res 2015;5(10):3056-3071
www.ajcr.us /ISSN:2156-6976/ajcr0014322

Original Article

Germline copy number loss of UGT2B28 and gain of PLEC contribute to increased human esophageal squamous cell carcinoma risk in Southwest China

Liwen Hu[1,4*], Yuanyuan Wu[2*], Xingying Guan[2], Yan Liang[2], Xinyue Yao[3], Deli Tan[1], Yun Bai[2], Gang Xiong[1], Kang Yang[1]

[1]Department of Cardiothoracic Surgery, Southwest Hospital, Third Military Medical University, Chongqing, PR China; Departments of [2]Medical Genetics, [3]Microbiology, College of Basic Medical Science, Third Military Medical University, Chongqing, PR China; [4]Department of Cardiothoracic Surgery, Jinling Hospital, School of Clinical Medicine, Nanjing University, Nanjing, Jiangsu Province, PR China. [*]Equal contributors.

Received August 11, 2015; Accepted September 2, 2015; Epub September 15, 2015; Published October 1, 2015

Abstract: Esophageal squamous cell carcinoma (ESCC) is one of the most common cancers worldwide with poor prognosis. Thus, identification of predictive biomarkers for early diagnosis and intervention is needed to improve patients' survival. Research shows that heritable mutations, such as single nucleotide polymorphisms (SNPs), contribute to human cancer susceptibility significantly. However, the association of copy number variations (CNVs), another major source of genetic variation, with ESCC risk remains poorly clarified. In this study, we aimed to identify ESCC risk-related CNVs based on candidate-gene strategy in a case-control study. A meta-analysis was first performed to identify the most variable chromosome regions of ESCC tissues. Bioinformatic analysis and dual-luciferase reporter assays were carried out to evaluate the properties of all recorded CNVs located on these regions. Six candidate CNVs located within well-known oncogenes and detoxification-associated enzymes were enrolled in the final analysis. A newly developed multiplex gene copy number quantitation method AccuCopy™ was employed to simultaneously genotype all six candidate sites in 404 ESCC patients and 402 cancer-free controls from Southwest China, and in 42 ESCC tissues. qRT-PCR was performed to measure UGT2B28 mRNA in cancerous and corresponding normal tissues. Unconditional logistic regression was applied to test association between germline CNV genotypes and ESCC risk. Relationship between germline copy number variation and somatic copy number alterations was further analyzed. Finally we found that copy number loss of UDP-glucuronosyltransferase family 2, polypeptide B28 (UGT2B28) and gain of plectin (PLEC) conferred increased ESCC risk (Adjusted OR = 2.085, 95% CI = 1.493-2.912, P < 0.001 for UGT2B28. Adjusted OR = 3.725, 95% CI = 1.026-13.533, P = 0.046 for PLEC). mRNA level was lower in UGT2B28 loss genotyped esophageal tissues than in two-copy tissues, indicating that UGT2B28 loss genotypes modify ESCC susceptibility perhaps by decreasing UGT2B28 expression level and enzyme activity. In addition, an association was drawn between germline copy number variations and somatic alterations for PLEC, UGT2B17 and UGT2B28, but not for other candidate loci.

Keywords: Copy number variation, esophageal squamous cell carcinoma, susceptibility, UDP-glucuronosyltransferase, plectin

Introduction

Esophageal cancer is one of the most malignant cancers worldwide and more new cases are identified annually. In China, esophageal cancer is within the top five causes of cancer mortality according to GLOBOCAN 2012 estimates (http://globocan.iarc.fr/), whereas squamous cell carcinoma represents 90% of Chinese cases [1]. Due to the lack of effective measures for early inspection and diagnosis, most patients miss opportunities for surgical resection, and the five-year overall survival rate remains extremely poor (< 20%) [2]. Accordingly identification of ESCC susceptibility related factors may assist with risk assessment and early intervention. Genetic variants have been suggested to significantly contribute to the susceptibility of most common sporadic cancers, including esophageal squamous cell carcinomas (ESCC). Although genome-wide association studies (GWAS) and site-specific genotyp-

ing have identified many genetic polymorphisms associated with ESCC risk [2-5], only a small fraction of the heritability for ESCC can be explained and certified.

As a prevalent form of genetic variation, copy number variation (CNV) is defined as a particular DNA segment ranging from kb to mega bases with variable copies compared to a reference genome, and it contributes to phenotypic diversity by altering gene dosage directly or by interfering long-range gene regulation via position effects [6, 7]. Accumulating evidence suggests that germline CNVs may explain the missing inherited basis of complex human diseases left by SNPs, especially for human cancers [8-10]. But few studies have associated CNVs with ESCC susceptibility, and this may be attributed to the difficulties in identification of promising CNVs, such as low frequency of novel CNVs and limitations of current technologies. For example, although GWAS method has been widely used in identifying disease-associated genetic variations, this method is limited in identifying rare genetic variations, due to the need for large samples and high cost. Therefore new strategies to identify cancer risk-related variations are urgently needed.

In current study, we applied a hypothesis-driven strategy to explore the ESCC risk associated genetic variation, which may make the results more promising. Our assumption was based mainly on previous evidence indicating tight association between germline CNVs and somatic copy numbers aberrations [11-14]. Therefore we speculated that the somatic chromosome abnormalities may be used as indicators for exploring possible germline variations. According to our hypothesis, we first performed a meta-analysis to obtain the most viable regions of ESCC, and then evaluated all recorded CNVs within these regions by bioinformatics and dual-luciferase reporter assays, to select the most functional CNVs sites. Finally we chose CNVs located on PRKCI, MYC, PLEC and CCND1 as our research targets. These four host genes are well-known oncogenes, which were repeatedly reported to be amplified in ESCC tissues, indicating that copy number abnormality of these genes plays crucial role in ESCC progression, but no studies have documented whether germline copy number variations of these genes are associated with ESCC risk. Considering their roles in tumorigenesis of ESCC, it seems reasonable that variations overlapping with these genes are associated with cancer risk.

On the other hand, we took genes encoding enzyme on detoxification of carcinogens into account, for ESCC is believed to be a consequence of gene-environment interactions [4, 15]. Experimental data suggested enzyme activity is directly proportional to the number of gene copies present [16, 17]. Individuals with unusual gene copies may have modified risk for tumor formation. Previous studies showed that xenobiotic metabolizing involved enzymes in two phases, which are represented by Cytochrome P450 (CYP450; Phase I), glutathione S-transferase (GST; phase II) and UDP-glucuronosyltransferase (UGT; phase II) family members respectively. Variations on CYP450 and GST families have been suggested to be associated with ESCC carcinogenesis [18-21], however, little was known about UGT family variations and ESCC risk, and no study has been reported about the association of UGT family copy number variation and ESCC risk. Here we included UDP glucuronosyltransferase family members UGT2B17 and UGT2B28 in our analysis, because these members have been reported to detoxify diverse carcinogenic metabolites and dietary chemopreventives by glucuronidation [22-25]. Previous studies indicated that copy number changes of the UGT family modified the risk of lung [26, 27], and colorectal cancers [28]. Considering the shared etiology among different cancer types, such as exposure to steroid compounds, tobacco and drinking, we hypothesized that ESCC may share some common susceptibility loci with other cancers, such as the UGT family members.

A multiplex gene copy number quantitation method AccuCopy™ was applied to simultaneously genotype all the six CNV sites in 404 cases and 402 controls for its accuracy and high efficiency [11, 29, 30]. Our data demonstrated that in comparison with the common 2-copy genotype, carriers of loss variant genotypes of UGT2B28 had significantly increased risk of ESCC (Adjusted OR = 2.085, 95% CI = 1.493-2.912, $P < 0.001$) in a dose-response manner ($P_{trend} < 0.001$), and additional stratified analyses did not change the results materially. Gain of PLEC was also found to be associated with ESCC risk with borderline statistical significance (Adjusted OR = 3.725, 95% CI = 1.026-13.533, $P = 0.046$). Moreover, mRNA level in UGT2B28 loss genotyped tissues was

Table 1. Distribution of demographic characteristics for cases and controls included in the study

		No. of cases	No. of controls	*P*-value[a]
Sex	Male	281	286	0.621
	Female	123	116	
Age	< 40	10	14	0.335
	40-50	64	60	
	50-60	206	216	
	60-70	106	86	
	> 70	18	26	
Smoking	Never/Occasionally	244	294	< 0.001[b]
	Ever/Frequently	160	108	
Drinking	Never/Occasionally	253	284	0.016[b]
	Ever/Frequently	151	118	
Family history of Cancer	Yes	57	39	0.053
	No	347	363	
Stage	I	32 (7.9%)		
	II	126 (31.2%)		
	III	148 (36.6%)		
	IV	98 (24.3%)		

[a]Chi-squared test results; [b]P < 0.05.

lower than those of 2-copy genotyped tissues, suggesting that loss of UGT2B28 might increase ESCC risk by reducing its mRNA level, and finally decreasing its detoxification capacity.

Material and methods

Study population and tissue samples

The study was approved by the ethics committee of the Third Military Medical University. All subjects were geographically homogenous and genetically unrelated Chinese Han from Southwestern China. Patients with ESCC (n = 404) were pathologically diagnosed through gastroscopic biopsy at Southwest Hospital, Third Military Medical University, from December 2006 to January 2012. TNM staging of ESCC patients was according to the American Joint Committee on Cancer (AJCC). Cancer-free controls (n = 402) were randomly selected from the individuals who had physical examinations at Southwest Hospital during the same period. Patient demographics characteristics are in **Table 1**. Each subject provided 3-5 ml of peripheral blood before any treatment.

Tumor and corresponding normal esophageal epithelial tissues were obtained from patients who underwent esophagectomy at Southwest Hospital. All resected tissue specimens were confirmed by pathologists. Written informed consent was obtained from all participants prior to study initiation or sample collection.

Esophageal cancer cell lines and plasmids

Human esophageal cancer cell lines EC109 and EC9706 were purchased from Cell Bank of Chinese Academy of Sciences, Shanghai, China. All cells were cultured in RPMI-1640 medium supplemented with 10% fetal bovine serum (FBS), at 37°C in a humidified incubator containing 5% CO_2. pGL3-promoter and pRL-TK vector were purchased from Promega (Madison, WI).

Systematic review and meta-analysis

To identify candidate regions for further genotyping, we first selected the most variable chromosome regions in somatic ESCC tissues. A PubMed search was performed on January 1, 2014 to identify articles with key words 'comparative genomic hybridization' and 'esophageal cancer'. Returned reports were scanned manually to identify the relevant publications.

Studies including whole-genome arrays and individual or summarized data on chromosome abnormality frequency in ESCC tissues were eligible for inclusion. Literature regarding cell lines, adenocarcinoma and sequencing methods were excluded from the meta-analysis to minimize study heterogeneity.

For eligible publications, the most frequent chromosomal alterations were listed, and chromosome aberration prevalence was analyzed by meta-analysis. A fixed-effect model was used to combine studies. Study heterogeneity was assessed using the I^2 statistic. A funnel plot graphing study precision against the logit event rate was created and Egger's regression asymmetry test was performed to assess reporting bias. All statistics for the meta-analysis were calculated in Comprehensive Meta-

Analysis, version 2 (Biostat Inc., Englewood, NJ).

In silico analysis and selection of candidate CNVs according to a region-based strategy

In the Database of Genomic Variants (http://dgvbeta.tcag.ca/dgv/app/home, version: GRCh-37/hg19) and UCSC Genome Bioinformatics (http://genome.ucsc.edu/), we evaluated all recorded CNV sites within these variable regions with respect to frequency, recurrence, length, position and predicted regulatory potentiality. Inclusion criteria for candidate CNVs was as follows: i) CNVs were reported by at least three groups, with the greatest frequency > 5/100; ii) the CNV length should be 1-10kb; iii) CNV sites should be located within or near well-known tumor-associated genes, and iv) abundant regulatory signals were present as reflected by regulation tracks of the UCSC genome browser.

Dual-Luciferase reporter gene assay

A dual-luciferase reporter gene assay was used to evaluate the regulatory potentiality of selected candidate sites from introns and non-coding regions, and only those with obvious regulatory ability were enrolled in the final genotyping system. Fragments containing core sequences of the selected CNV sites were amplified with primers containing artificial XhoI/MluI or KpnI/MluI enzymes restriction sites, and cloned into the pGL3-promoter vectors (firefly luciferase, Promega). Cloning primers are shown in Supplementary Table 1. Recombinant plasmids were confirmed by sequencing.

For luciferase activity analysis, EC109 and EC9706 cells (5×10^3/well) were plated in 96-wells overnight, then 100 ng of recombinant plasmids or empty pGL3-promoter vectors were co-transfected with 0.5 ng of pRL-TK vector (*Renilla* luciferase, Promega) into cells using lipofectamine2000 (Invitrogen, Carlsbad, CA). The pRL-TK plasmid was used as an internal reference. Luciferase assay was carried out 48 h post-transfection according to the manufacturer's instructions (Promega). Activity was defined as firefly/*Renilla* ratio and normalized to negative control. Assays were performed in triplicate.

DNA extraction and genotyping

Genomic DNA (gDNA) was extracted from peripheral blood and ESCC tumor tissues using a Wizard Genomic DNA Purification Kit (Promega) according to the manufacturer's instructions. We used a custom-by-design patented Multiplex AccuCopy™ Kit (Genesky Biotechnologies Inc., Shanghai, China) to measure copy number of selected CNV sites in one reaction. To ensure accuracy of genotyping results, two probes (two pairs of primers) were designed for each CNV site, and four reference DNA segments (Reference 1, Reference 2, Reference 3, and Reference 4) were utilized for normalization. Details of this method have been described previously [29]. Probe primers were designed to avoid the frequently mutated regions. One-third of the samples were randomly selected for repeat analysis three times. Probe sequences and position information are shown in Supplementary Table 2 and Supplementary Figure 1.

mRNA extraction and qRT-PCR

Total RNA was isolated from 42 pairs of esophageal tumor and adjacent normal tissues using RNAiso Plus Kit (Takara, Japan). RNA was reverse-transcribed into complementary DNA (cDNA) and subjected to real time SYBR-Green quantitative PCR (Takara) on Illumina Eco™ (Illumina, Santiago, CA), The qRT-PCR primers sequences were as follows: UGT2B28-forward: AGCAGAAAGGGCCAACGTAA, UGT2B28-reverse: CCAGTCTAACAGCTGCTCCC; GAPDH-forward: TGTTGCCATCAATGACCCCTT, GAPDH-reverse: CTCCACGACGTACTCAGCG. UGT2B28 expression was normalized to GAPDH mRNA expression by method of $2^{-\Delta t}$.

Statistical analysis

A Chi-squared test was used to evaluate differences in distribution of demographic characteristics for study subjects. Luciferase-reporter assays were repeated three times in triplicate. Recombinant plasmid regulatory activity was normalized to *Renilla* luciferase, and compared with activity of empty plasmids using an unpaired Student's *t*-test. Copy numbers of each sample were calculated by standard Accucopy™, and then unconditional logistic regression was applied to estimate the association between CNVs and ESCC risk, with or without adjustments for age, sex, smoking/drinking status and family cancer history. The association was reflected by the odds ratios (OR) and 95% Confidential Interval (CI). Analyses stratified by risk factors were also carried out

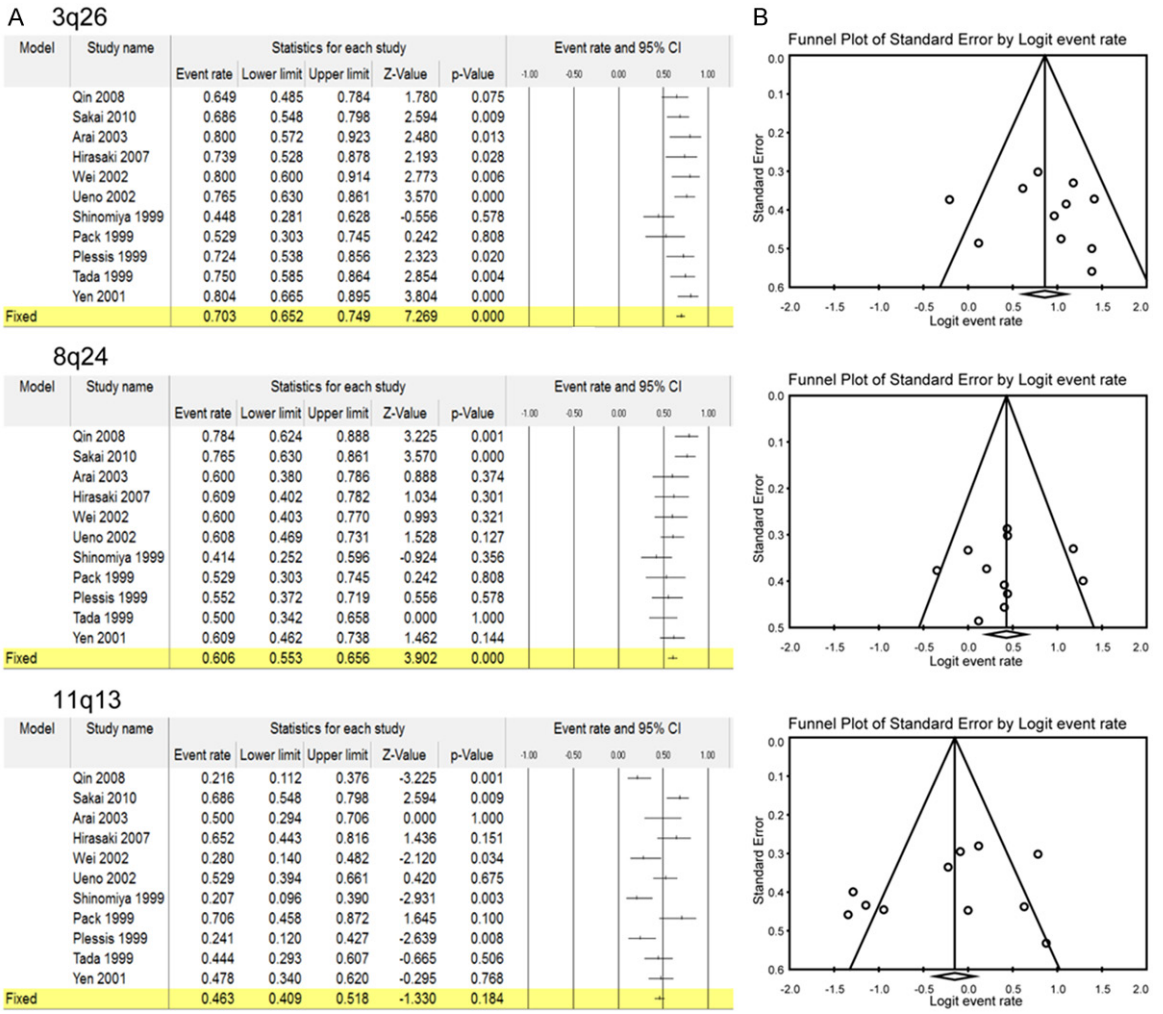


Figure 1. Meta-analysis of the frequently amplified chromosome regions in ESCC tissues. (A) Forest plot of the prevalence of 3q26, 8q24 and 11q13 amplification in ESCC tissues. (B) Funnel plot againsts the logit event rate.

under the same unconditional logistic regression model, and the Breslow-Day test method was used to test homogeneity among strata ORs. Linear regression analysis was applied to test the association between germline CNV and somatic copy number alterations returned by Accucopy™ calculations. UGT2B28 expression in cancerous and normal tissues were normalized to GAPDH, and a paired Student's *t* test was used to compare UGT2B28 expression between cancerous and normal tissues, whereas an unpaired Student's *t* test used for analyzing associations between CNV genotypes and UGT2B28 expression. All statistical analyses were performed using SPSS v16.0 and $P < 0.05$ was considered statistically significant.

Results

Combined analysis of ESCC tissue variable regions suggested variations on four prominent oncogenes as genotyping targets

To guide our selection of promising CNVs for study, a systematic review and meta-analysis were performed to identify the most common chromosomal aberrations in ESCC tissues. A PubMed search returned 142 articles, of which 11 were included in the meta-analysis based on our inclusion criteria [31-41]. According to the literature, amplification of chromosome 3q26, 8q24, and 11q13 were the most commonly identified chromosomal aberration by comparative genomic hybridization (CGH) in

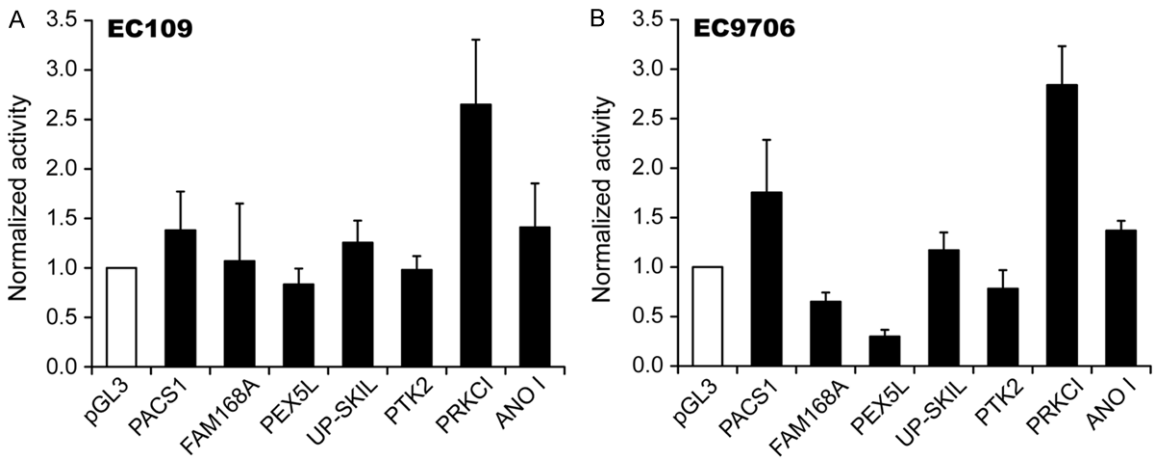


Figure 2. Regulatory activity of the candidate CNVs from introns and non-coding regions. Recombinant plasmids encompassing the core regions of the candidate sites were constructed. The recombinant or empty vectors were co-transfected with pRL-TK. Regulatory activity was normalized to the negative control (empty pGL3-promotor transfected cells). Data are expressed as means ± SDs, and each experiment was performed in triplicate for three times. *P<0.05, **P<0.01.

ESCC tissues, and these data are summarized in our meta-analysis.

Pooling results from the 11 studies of ESCC tissues revealed that 70.3% of these tumors had amplified 3q26 (95% CI: 65.2-74.9%; **Figure 1A**), 60.6% had amplified 8q24 (95% CI: 55.3-65.6%; **Figure 1A**), and 46.3% had amplified 11q13 (95% CI: 40.9-51.8%; **Figure 1A**). Reporting bias was assessed by funnel plot, which graphed study precision against the logit event rate. We observed no evidence of reporting bias for all three regions, as there were no studies in the lower right corner (**Figure 1B**). Egger's regression asymmetry test did not suggest that publication bias was present (Intercept = 0.95719, P = 0.68398 for 3q26; Intercept = -0.71642, P = 0.78468 for 8q24; and Intercept = -3.14956 P = 0.31239 for 11q13). The meta-analysis confirmed that amplification of 3q26, 8q24, and 11q13 are common phenomena in most ESCC tissues, which may guide our selection of candidate CNV sites.

Next, we evaluated all recorded germline CNV sites within 3q26, 8q24, and 11q13 with DGV and UCSC databases, and we preliminarily selected 10 candidate CNV sites according to our inclusion criteria. Among these candidate sites, 7 are from introns or non-coding regions, while the others overlapped with coding regions of target genes. Data about the 10 ten primary candidate sites are in Supplementary Table 3.

To investigate the functional potentiality of the primary candidates from introns or non-coding regions, we constructed recombinant reporter plasmids which encompass the core regions of these CNVs and transiently transfected them into EC109 and EC9706 cells. Recombinant plasmid of variation_59109 (PRKCI) drove significantly higher luciferase expression than the pGL3-promoter plasmid in both cell lines (**Figure 2**), indicating that variation_59109 might increase PRKCI transcription because of its enhancing capacity. Thus, the variation_59109 on PRKCI was selected for genotyping, accompanied by candidates located in exons of oncogenes MYC, PLEC, and CCND1. Position information about candidate CNVs and their host genes are presented in Supplementary Figure 1.

Distribution of demographic characteristics

Selected characteristics of cases and controls are summarized in **Table 1**. There were no significant differences in distributions of sex and age between cases and controls, suggesting that the frequency matching of cases and controls was adequate. However, individuals who usually smoked or drank were over-represented in cases compared with controls, which is consistent with the concept that smoking and drinking are highly risk factors for the development of esophageal cancer. Of the 404 patients with esophageal cancer, 67.8% were stage III or IV.

Table 2. Association between target copy number variations and ESCC risk

	Genotypes	No. of cases	No. of controls	Crude OR (95% CI)[a]	*P*-value	Adjusted OR (95% CI)[b]	*P*-value
PRKCI	2 copies	401	402	1.00 (reference)		1.00 (reference)	
	1 copy	1	0	1.62E+009 (0.000-)	1	1.41E+009 (0.000-)	0.999
	3 copies	2	0	2E+009 (0.000-)	0.999	2E+009 (0.000-)	0.999
CCND1	2 copies	386	377	1.00 (reference)		1.00 (reference)	
	3 copies	18	25	0.703 (0.377-1.310)	0.267	0.659 (0.42-1.268)	0.212
MYC	2 copies	401	401	1.00 (reference)		1.00 (reference)	
	3 copies	3	1	3.000 (0.311-28.963)	0.342	4.086(0.418-39.907)	0.226
PLEC	2 copies	390	399	1.00 (reference)		1.00 (reference)	
	1 copy	1	0	2E+009 (0.000-)	1	2E+009 (0.000-)	1
	3 copies	13	3	4.433 (1.254-15.678)	0.021	3.725 (1.026-13.533)	0.046
UGT2B17	0 copy	283	277	1.00 (reference)		1.00 (reference)	
	1 copy	112	119	0.921 (0.678-1.252)	0.600	0.965 (0.705-1.322)	0.826
	2 copies	9	6	1.594 (0.550-4.622)	0.391	1.485 (0.494-4.465)	0.482
	P_{trend}		0.909				
	Null	283	277	1.00 (reference)		1.00 (reference)	
	carrier	121	125	0.947 (0.702-1.279)	0.724	0.982 (0.722-1.336)	0.908
UGT2B28	2 copies	273	327	1.00 (reference)		1.00 (reference)	
	1 copy	123	73	2.018 (1.449-2.811)	0.000	1.996 (1.422-2.802)	0.000
	0 copy	8	2	4.79 (1.009-22.750)	0.049	5.39 (1.104-26.360)	0.037
	P_{trend}		< 0.0001				
	Wild type	273	327	1.00		1.00	
	Deletion	131	75	2.092 (1.510-2.899)	0.000	2.085 (1.493-2.912)	0.000

[a,b]The unconditional logistic regression model was applied to estimate the association between target CNVs and ESCC risk, with (b) or without (a) adjustments of age, sex, smoking/drinking status and family cancer history.

Gain of PLEC and loss of UGT2B28 were associated with ESCC risk

Data show that most subjects had two copies consistently (**Table 2**) for CNVs within PRKCI and MYC, suggesting a neutral CNV status in a diploid genome. Gain of variation_59109 (PRKCI) was found in two patients but not in controls, and gain of MYC occurred in three patients and one control. It seems that gains of these two loci were over-represented in ESCC patients, but no statistically significant difference between the two groups existed for either (P = 0.999 for PRKCI and P = 0.226 for MYC). This might be explained by low frequency of these CNVs, so larger sample sizes are required to calculate the significance of these two loci. For CCND1, no obvious difference was identified between the two groups either.

Except for the three loci with no significant association with ESCC risk, gain of PLEC was found in 13 cases and 3 controls. A variant 3-copy was significantly associated with increased risk of ESCC [OR = 4.433, 95% CI = 1.254-15.678, P = 0.021]. Additional adjustment for gender, age, smoking/drinking status, and family history of cancer did not change the results virtually (Adjusted OR= 3.725, 95% CI = 1.026-13.533, P = 0.046; **Table 2**).

A high frequency of UGT2B17 null genotype was found among subjects, just 9 two-copy genotyped subjects were identified in cases and 6 in controls, which differed from the copy number status in other populations, suggesting that the copy number status of UGT2B17 may differ by ethnic groups. UGT2B17 carriers did not have modified risk of ESCC compared with those with the UGT2B17 null genotype (OR = 0.947, 95% CI = 0.702-1.279; adjusted OR = 0.982, 95% CI = 0.722-1.336; **Table 2**).

A highly significant association was observed between the UGT2B28 deletion and ESCC risk, with OR = 2.018 (95% CIs = 1.449-2.811) asso-

Table 3. Association of UGT2B28 gene deletion and ESCC risk stratified by esophageal cancer risk factors

		Cases		Controls		Crude OR	Adjusted OR	$P_{homogeneous}$ [a]
		2-copies	Deletion	2-copies	Deletion	Deletion/2-copy	Deletion/2-copy	
Sex	Male	192	89	230	56	1.904 (1.295-2.799)	1.832 (1.218-2.755)	0.529
	Female	82	41	96	20	2.400 (1.303-4.419)	2.653 (1.388-5.072)	
Age	> 50	228	102	262	66	1.776 (1.243-2.538)	1.742 (1.205-2.520)	0.081
	≤ 50	46	28	64	10	3.896 (1.724-8.804)	4.364 (1.841-10.345)	
Smoking	Yes	116	54	78	20	1.816 (1.009-3.268)	1.757 (0.934-3.306)	0.659
	No	158	76	248	56	2.130 (1.430-3.174)	2.123 (1.415-3.186)	
Drinking	Yes	100	51	96	22	2.225 (1.255-3.947)	2.373 (1.244-4.526)	0.693
	No	174	79	230	54	1.934 (1.299-2.880)	1.915 (1.281-2.864)	
Family	Yes	43	14	31	8	1.262 (0.472-3.374)	1.813 (0.370-8.891)	0.302
	No	231	116	295	68	2.179 (1.542-3.077)	2.141 (1.508-3.040)	
Stages	I	22	10	324	78	1.888 (0.859-4.149)	1.586 (0.642-3.921)	
	II	85	41			2.004 (1.281-3.134)	1. 987 (1.257-3.142)	
	III	105	43			1.701 (1.104-2.621)	1.718 (1.106-2.667)	
	IV	67	31			1.922 (1.175-3.144)	1.855 (1.114-3.088)	

[a]The Breslow-Day test method was used to test the homogeneity among strata odds ratios (ORs).

ciated with one-copy deletion and OR = 4.791 (95% CI = 1.009-22.750) associated with two-copy deletion. After adjusting for additional potentially confounding factors, such as age, sex, smoking/drinking status and family history of cancer, the results did not change materially. ORs (95% CIs) were 1.996 (1.422-2.802) associated with one-copy deletion and 5.394 (1.104-26.360) associated with two-copy deletion compared with people with no deletion. The effect of the loss allele occurred in a dose-dependent manner (P_{trend} < 0.001; **Table 2**).

We further assessed possible modifications of surrounding factors in stratification analysis for UGT2B28 loss (**Table 3**). A positive association of UGT2B28 deletion was found for ESCC risk and this was similar when stratified by age, gender, smoking/drinking status and family history of cancer or AJCC grade, with $P_{homogeneous}$ > 0.05 for all surrounding factors.

Associations between germline copy number variation and somatic copy number alteration occurred in UGT2B and PLEC loci

We measured copy number alterations of target loci in 42 ESCC tissues, and found that PRKCI, MYC, PLEC, and CCND1 were frequently amplified in ESCC tissues: frequency of 19/42 (CCND1), 15/42 (MYC), 10/42 (PLEC), and 25/42 (PRKCI) respectively (**Figure 3A-D**), which was consistent with previous reports. Moreover, amplification of PLEC in ESCC tissues was significantly associated with germline copy number gain (P = 0.0296, **Figure 4C**), but the association was not obvious for CCND1 (P = 0.0582, **Figure 4D**). In addition, this association was not present in PRKCI and MYC gene loci due to low frequency of germline variations (**Figure 4A** and **4B**), suggesting a complicated progression for the occurrence of somatic alterations, at least for these two loci.

Of note, somatic deletion of UGT2B17 and UGT2B28 correlated with germline loss significantly (**Figure 4E** and **4F**), and most tissues had equal copy number status as their germline partners (**Figure 3E** and **3F**), suggesting that deletion of the UGT family members in esophageal tissues may be derived from the germline source.

Effects of germline genotypes of UGT2B28 on mRNA expression

Figure 5A shows that UGT2B28 mRNA was down-regulated in ESCC tissues compared with normal ones (**Figure 5A**), suggesting a tumor-suppressor role for UGT2B28 in ESCC. UGT-2B28 mRNA was significantly increased along with increased copy numbers (**Figure 5B**). Subjects with 1-copy genotype showed lower mRNA level than those with the 2-copy geno-

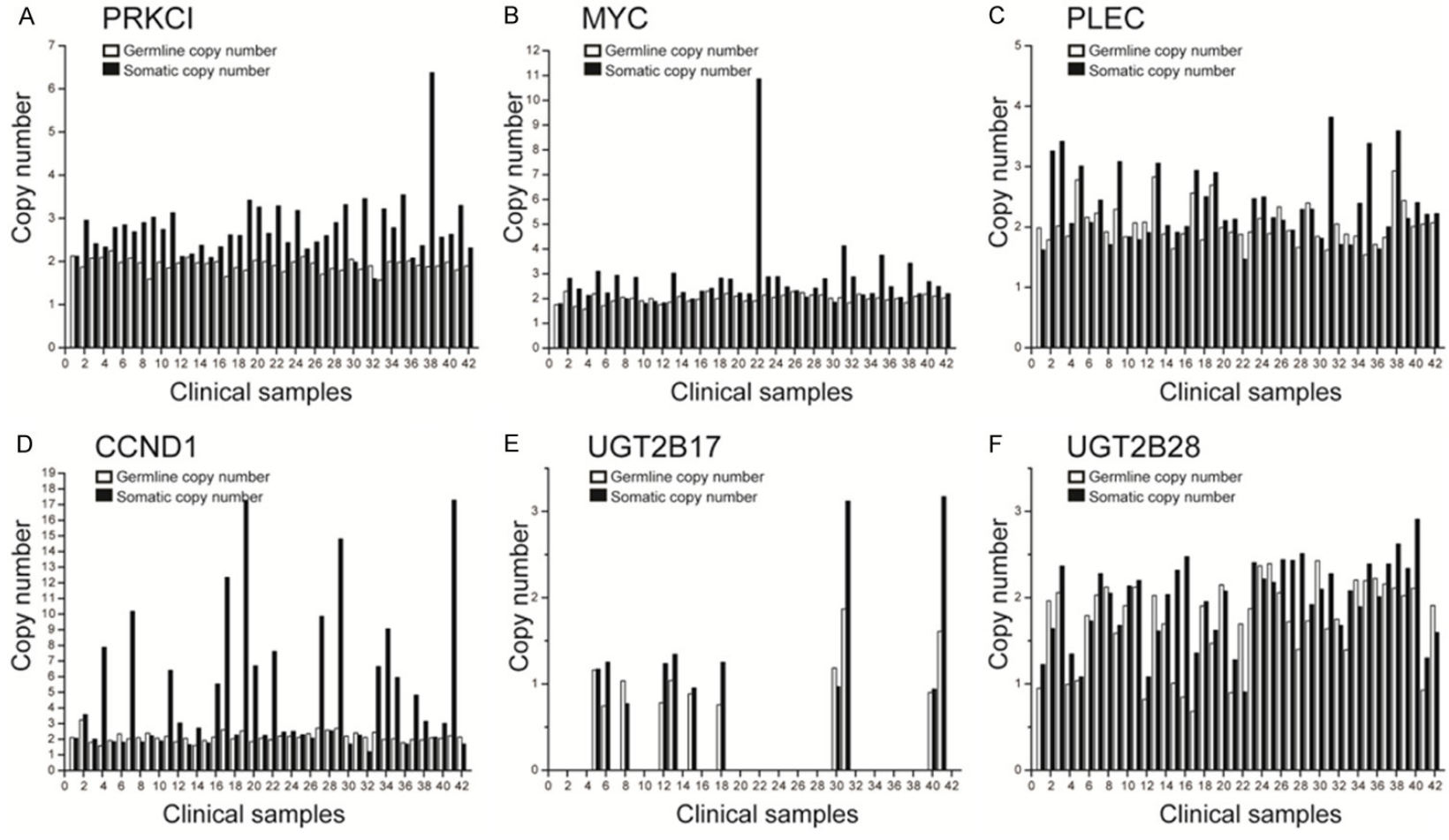


Figure 3. Comparison of the copy numbers alterations in ESCC tissues and corresponding germline copy number variations. The copy numbers of the selected genes in ESCC tissues were calculated by Accucopy™ method as well. Frequent amplification of PRKCI, MYC, PLEC and CCND1 was found in ESCC tissues. Moreover, most of the detected tissues displayed equal copy number status as their germline partners for UGT2B17 and UGT2b28 loci.

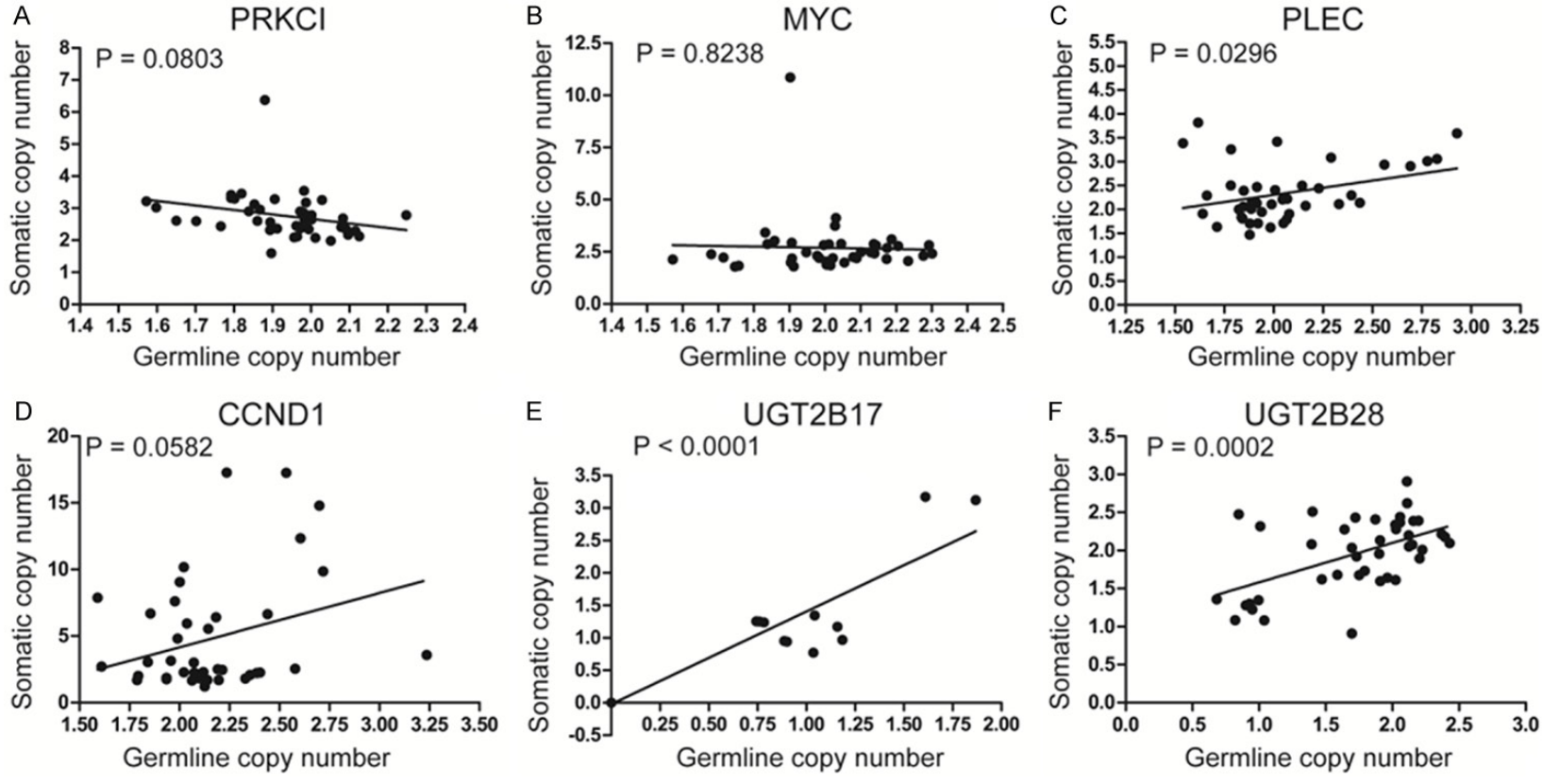


Figure 4. Analysis of the association between germline copy number variation and somatic copy number alteration. Linear regression was applied to test the association between germline copy number variation and somatic copy number alteration. No significant association was found for PRKCI, MYC, and CCND1 loci, while obvious association was found for PLEC, UGT2B17 and UGT2B28 loci.

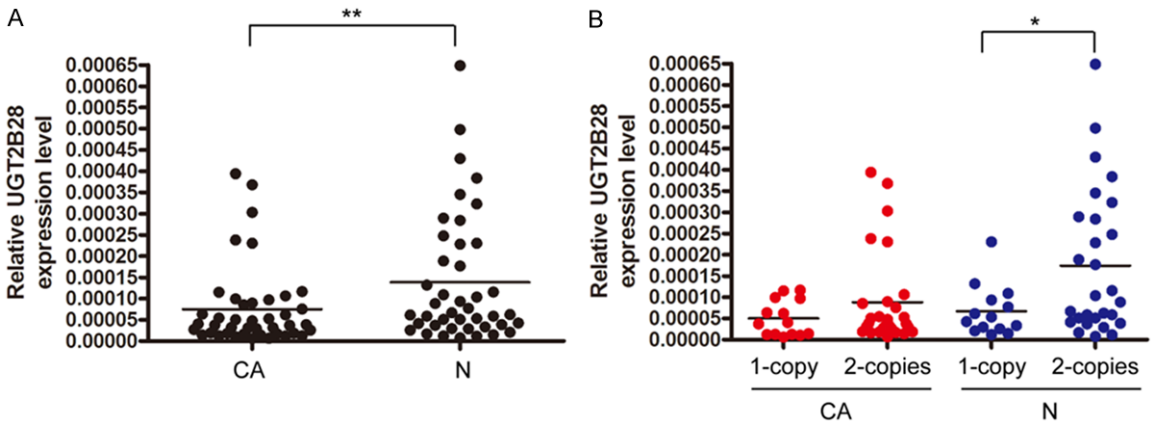


Figure 5. Relative mRNA level of UGT2B28 detected by qRT-PCR. A. Comparison of the expression level of UGT2B28 between tumor tissues and normal counterparts using paired student's *t*-test. B. Association between germline copy numbers and UGT2B28 expression. Loss genotyped tissues expressed significantly lower UGT2B28 than two-copy ones.

type in both tumor and normal tissues, suggesting that loss of UGT2B28 increases ESCC risk by reducing its expression through gene-dosage effects.

Discussion

The association between germline copy number variation and human disease susceptibility has gained increasing attention, but little is known about their effects on ESCC risk. Recently, Sun identified ESCC risk-related ABCC4 loss using a GWAS method, supporting an important role for CNV in the etiology of ESCC [42]. Lack of evidence about a relationship between CNVs and cancer risk may be partly due to the complexity of studying CNVs, which are rare and of low frequency among diverse population. Here, we developed a new hypothesis-driven strategy based on candidate-gene tactics to make choice of the most promising CNV sites.

With this strategy, we took the association between germline genetic variation and somatic genomic alteration into account. Meta-analysis data returned the most variable regions in ESCC tissues, and bioinformatic analysis combined with dual-luciferase reporter assays suggested variation_59109 on PRKCI (3q26), along with those CNVs located on exons of MYC (8q24), PLEC (8q24), and CCND1 (11q13) as our final targets. Roles of the four host genes in tumorigenesis have been well characterized, especially in ESCC [43-46].

Genotyping data confirmed that gain of PLEC contributes to ESCC risk with marginal significance, emphasizing that our strategy and project design was rational. To our knowledge, PLEC (or plectin) is an important structural protein (plakin or cytolinker family member) [47-49]. It can interlink different elements of the cytoskeleton, helping to maintain cell integrity and shape, as well as serve as scaffolding platforms for assembly, positioning, and regulation of signaling complexes. Its functional characteristics determine its important involvement in carcinogenesis, especially in tumor migration and invasion. The oncogenic role of PLEC has been reported in human head and neck squamous cell carcinoma [50], colon carcinoma [51], and in pancreatic cancer [52-54]. Pawar discovered that PLEC was over-expressed in ESCC tissues, indicating a possible participation in ESCC tumorigenesis [45]. Thus, we propose that gain of PLEC may up-regulate its expression and facilitate ESCC occurrence. Its functional role in ESCC needs our further verification.

On the other hand, UGT family members were considered for their indispensable roles in detoxification of carcinogens. The UGT family participates in phase II detoxification metabolism of endobiotics and xenobiotics by catalyzing glucuronidation, such as of genotoxic intermediates NNAL [55] and steroid compounds [56], and as such may decrease cancer risk.

Among this superfamily, UGT2B17 and UGT2B28 are the only UGT genes that exhibit com-

mon whole-gene deletion polymorphisms. Loss of UGT2B28 was found in 131 cases and 75 controls, indicating a significant association with ESCC risk, which support the hypothesis that functional UGT alleles loss may result in reduced genotoxic metabolites. Furthermore, germline copy number variation of UGT2B28 was significantly associated with somatic copy number abbreviations and tissue mRNA level, suggesting that loss genotypes of UGT2B28 led to gene deletion in somatic tissues and decreased UGT2B28 expression, which may reduce detoxification and increase the likelihood of carcinogenesis. In agreement with our data, Yoshida reported that UGT2B28 was deleted and down-regulated in colorectal cancer tissues [57], indicating that UGT2B28 functions as a tumor suppressor in cancers, and its function loss increased cancer risk. However, detailed information about UGT2B28 function is largely undocumented and existing data indicate it mainly participates in steroid inactivation [24, 58]. Steroid compounds, such as bile acids and sex hormones, are known risk factors for ESCC [59-61]. It worth noting that esophageal cancer is two to four times more common among men than women, and both of epidemiological and experimental evidences suggested that sex hormones may play an important role in the development of ESCC [62]. So it's plausible to explain the high ESCC risk induced by UGT2B28 loss by its functional relevance.

In summary, we found that germline copy number gain of PLEC and loss of UGT2B28 contributed to increased ESCC risk and this may be explained by their functional significance and gene-dosage effects. However, there were still some limitations in this study. First, because it's carried out in Chinese southwest Han population, it is uncertain whether our findings can be generalized to other populations; Secondly, this study was restricted in a hospital-based sample, so selection bias is unavoidable. Meanwhile, the sample size for this study is relatively small, which may hinder our analysis for rare CNV sites, such as CNVs on PRKCI and MYC. Larger population-based studies in different ethnic groups and characterization of detailed functions of UGT2B28 and PLEC will be required in further studies.

Acknowledgements

This work was supported by grants from National Natural Science Foundation of China [No. 81172379].

Disclosure of conflict of interest

None.

Address correspondence to: Drs. Gang Xiong and Kang Yang, Department of Cardiothoracic surgery, Southwest Hospital, Third Military Medical Uni0 versity, Gaotanyan St., Shapingba District, Chongqing, People's Republic of China. Tel: +86-23-68754183; Fax: +86-23-68754183; E-mail: xionggang_xw@sina.com.cn (GX); yangkangxwk@163.com (KY)

References

[1] Ren Z, Zhu J, Gu H, Liu R, Chen S, Rong G and Sun B. Decoy receptor 3 polymorphisms are not associated with the risk of esophageal cancer in a Chinese population. Biomarkers 2014; 19: 340-344.

[2] Wu C, Wang Z, Song X, Feng XS, Abnet CC, He J, Hu N, Zuo XB, Tan W, Zhan Q, Hu Z, He Z, Jia W, Zhou Y, Yu K, Shu XO, Yuan JM, Zheng W, Zhao XK, Gao SG, Yuan ZQ, Zhou FY, Fan ZM, Cui JL, Lin HL, Han XN, Li B, Chen X, Dawsey SM, Liao L, Lee MP, Ding T, Qiao YL, Liu Z, Liu Y, Yu D, Chang J, Wei L, Gao YT, Koh WP, Xiang YB, Tang ZZ, Fan JH, Han JJ, Zhou SL, Zhang P, Zhang DY, Yuan Y, Huang Y, Liu C, Zhai K, Qiao Y, Jin G, Guo C, Fu J, Miao X, Lu C, Yang H, Wang C, Wheeler WA, Gail M, Yeager M, Yuenger J, Guo ET, Li AL, Zhang W, Li XM, Sun LD, Ma BG, Li Y, Tang S, Peng XQ, Liu J, Hutchinson A, Jacobs K, Giffen C, Burdette L, Fraumeni JF, Jr., Shen H, Ke Y, Zeng Y, Wu T, Kraft P, Chung CC, Tucker MA, Hou ZC, Liu YL, Hu YL, Wang L, Yuan G, Chen LS, Liu X, Ma T, Meng H, Sun L, Ku JW, Zhou YF, Yang LQ, Qige Q, Yang WJ, Lei GY, Chen LQ, Li EM, Yuan L, Yue WB, Wang R, Wang LW, Fan XP, Zhu FH, Zhao WX, Mao YM, Zhang M, Xing GL, Li JL, Han M, Ren JL, Liu B, Ren SW, Kong QP, Li F, Sheyhidin I, Wei W, Zhang YR, Feng CW, Wang J, Yang YH, Hao HZ, Bao QD, Liu BC, Wu AQ, Xie D, Yang WC, Zhao XH, Chen SQ, Hong JY, Zhang XJ, Freedman ND, Goldstein AM, Lin D, Taylor PR, Wang LD and Chanock SJ. Joint analysis of three genome-wide association studies of esophageal squamous cell carcinoma in Chinese populations. Nat Genet 2014; 46: 1001-1006.

[3] Li WQ, Pfeiffer RM, Hyland PL, Shi J, Gu F, Wang Z, Bhattacharjee S, Luo J, Xiong X, Yeager M, Deng X, Hu N, Taylor PR, Albanes D, Caporaso NE, Gapstur SM, Amundadottir L, Chanock SJ, Chatterjee N, Landi MT, Tucker MA, Goldstein AM and Yang XR. Genetic polymorphisms in the 9p21 region associated with risk of multiple cancers. Carcinogenesis 2014; 35: 2698-2705.

[4] Chang J, Huang Y, Wei L, Ma B, Miao X, Li Y, Hu Z, Yu D, Jia W, Liu Y, Tan W, He Z, Ke Y, Wu T, Shen H, Zeng Y, Wu C and Lin D. Risk prediction of esophageal squamous-cell carcinoma with common genetic variants and lifestyle factors in Chinese population. Carcinogenesis 2013; 34: 1782-1786.

[5] Wu C, Kraft P, Zhai K, Chang J, Wang Z, Li Y, Hu Z, He Z, Jia W, Abnet CC, Liang L, Hu N, Miao X, Zhou Y, Liu Z, Zhan Q, Liu Y, Qiao Y, Jin G, Guo C, Lu C, Yang H, Fu J, Yu D, Freedman ND, Ding T, Tan W, Goldstein AM, Wu T, Shen H, Ke Y, Zeng Y, Chanock SJ, Taylor PR and Lin D. 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.

[6] Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurles ME, Carter NP, Scherer SW and Lee C. Copy number variation: new insights in genome diversity. Genome Res 2006; 16: 949-961.

[7] Haraksingh RR and Snyder MP. Impacts of variation in the human genome on gene regulation. J Mol Biol 2013; 425: 3970-3977.

[8] Melkonian SC, Wang X, Gu J, Matin SF, Tannir NM, Wood CG and Wu X. Mitochondrial DNA copy number in peripheral blood leukocytes and the risk of clear cell renal cell carcinoma. Carcinogenesis 2015; 36: 249-255.

[9] Yu K, Fan J, Ding X, Li C, Wang J, Xiang Y and Wang QS. Association study of a functional copy number variation in the WWOX gene with risk of gliomas among Chinese people. Int J Cancer 2014; 135: 1687-1691.

[10] Yang L, Liu B, Qiu F, Huang B, Li Y, Huang D, Yang R, Yang X, Deng J, Jiang Q, Zhou Y and Lu J. The effect of functional MAPKAPK2 copy number variation CNV-30450 on elevating nasopharyngeal carcinoma risk is modulated by EBV infection. Carcinogenesis 2014; 35: 46-52.

[11] Liu B, Yang L, Huang B, Cheng M, Wang H, Li Y, Huang D, Zheng J, Li Q, Zhang X, Ji W, Zhou Y and Lu J. A functional copy-number variation in MAPKAPK2 predicts risk and prognosis of lung cancer. Am J Hum Genet 2012; 91: 384-390.

[12] Huang L, Yu D, Wu C, Zhai K, Jiang G, Cao G, Wang C, Liu Y, Sun M, Li Z, Tan W and Lin D. Copy number variation at 6q13 functions as a long-range regulator and is associated with pancreatic cancer risk. Carcinogenesis 2012; 33: 94-100.

[13] Walsh KM, Choi M, Oberg K, Kulke MH, Yao JC, Wu C, Jurkiewicz M, Hsu LI, Hooshmand SM, Hassan M, Janson ET, Cunningham JL, Vosburgh E, Sackler RS, Lifton RP, Dewan AT and Hoh J. A pilot genome-wide association study shows genomic variants enriched in the non-tumor cells of patients with well-differentiated neuroendocrine tumors of the ileum. Endocr Relat Cancer 2011; 18: 171-180.

[14] Dworkin AM, Ridd K, Bautista D, Allain DC, Iwenofu OH, Roy R, Bastian BC and Toland AE. Germline variation controls the architecture of somatic alterations in tumors. PLoS Genet 2010; 6: e1001136.

[15] Zhu J, Wang M, Zhu M, He J, Wang JC, Jin L, Wang XF, Xiang JQ and Wei Q. Associations of PI3KR1 and mTOR polymorphisms with esophageal squamous cell carcinoma risk and gene-environment interactions in Eastern Chinese populations. Sci Rep 2015; 5: 8250.

[16] Gruber M, Bellemare J, Hoermann G, Gleiss A, Porpaczy E, Bilban M, Le T, Zehetmayer S, Mannhalter C, Gaiger A, Shehata M, Fleiss K, Skrabs C, Levesque E, Vanura K, Guillemette C and Jaeger U. Overexpression of uridine diphospho glucuronosyltransferase 2B17 in high-risk chronic lymphocytic leukemia. Blood 2013; 121: 1175-1183.

[17] Zhu AZ, Cox LS, Ahluwalia JS, Renner CC, Hatsukami DK, Benowitz NL and Tyndale RF. Genetic and phenotypic variation in UGT2B17, a testosterone-metabolizing enzyme, is associated with BMI in males. Pharmacogenet Genomics 2015; 25: 263-269.

[18] Tan W, Song N, Wang GQ, Liu Q, Tang HJ, Kadlubar FF and Lin DX. Impact of genetic polymorphisms in cytochrome P450 2E1 and glutathione S-transferases M1, T1, and P1 on susceptibility to esophageal cancer among high-risk individuals in China. Cancer Epidemiol Biomarkers Prev 2000; 9: 551-556.

[19] Jain M, Kumar S, Lal P, Tiwari A, Ghoshal UC and Mittal B. Role of GSTM3 polymorphism in the risk of developing esophageal cancer. Cancer Epidemiol Biomarkers Prev 2007; 16: 178-181.

[20] Matejcic M, Li D, Prescott NJ, Lewis CM, Mathew CG and Parker MI. Association of a deletion of GSTT2B with an altered risk of oesophageal squamous cell carcinoma in a South African population: a case-control study. PLoS One 2011; 6: e29366.

[21] Gong FF, Lu SS, Hu CY, Qian ZZ, Feng F, Wu YL, Yang HY and Sun YH. Cytochrome P450 1A1 (CYP1A1) polymorphism and susceptibility to esophageal cancer: an updated meta-analysis of 27 studies. Tumour Biol 2014; 35: 10351-10361.

[22] Grosse L, Paquet S, Caron P, Fazli L, Rennie PS, Belanger A and Barbier O. Androgen glucuronidation: an unexpected target for androgen deprivation therapy, with prognosis and diagnostic implications. Cancer Res 2013; 73: 6963-6971.

[23] Beaulieu M, Levesque E, Tchernof A, Beatty BG, Belanger A and Hum DW. Chromosomal localization, structure, and regulation of the

UGT2B17 gene, encoding a C19 steroid metabolizing enzyme. DNA Cell Biol 1997; 16: 1143-1154.

[24] Levesque E, Turgeon D, Carrier JS, Montminy V, Beaulieu M and Belanger A. Isolation and characterization of the UGT2B28 cDNA encoding a novel human steroid conjugating UDP-glucuronosyltransferase. Biochemistry 2001; 40: 3869-3881.

[25] Zheng Z, Fang JL and Lazarus P. Glucuronidation: an important mechanism for detoxification of benzo[a]pyrene metabolites in aerodigestive tract tissues. Drug Metab Dispos 2002; 30: 397-403.

[26] Gallagher CJ, Muscat JE, Hicks AN, Zheng Y, Dyer AM, Chase GA, Richie J and Lazarus P. The UDP-glucuronosyltransferase 2B17 gene deletion polymorphism: sex-specific association with urinary 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol glucuronidation phenotype and risk for lung cancer. Cancer Epidemiol Biomarkers Prev 2007; 16: 823-828.

[27] Gruber M, Le T, Filipits M, Gsur A, Mannhalter C, Jager U and Vanura K. UDP-glucuronosyltransferase 2B17 genotype and the risk of lung cancer among Austrian Caucasians. Cancer Epidemiol 2013; 37: 625-628.

[28] Angstadt AY, Berg A, Zhu J, Miller P, Hartman TJ, Lesko SM, Muscat JE, Lazarus P and Gallagher CJ. The effect of copy number variation in the phase II detoxification genes UGT2B17 and UGT2B28 on colorectal cancer risk. Cancer 2013; 119: 2477-2485.

[29] Du R, Lu C, Jiang Z, Li S, Ma R, An H, Xu M, An Y, Xia Y, Jin L, Wang X and Zhang F. Efficient typing of copy number variations in a segmental duplication-mediated rearrangement hotspot using multiplex competitive amplification. J Hum Genet 2012; 57: 545-551.

[30] Hu L, Wu Y, Tan D, Meng H, Wang K, Bai Y and Yang K. Up-regulation of long noncoding RNA MALAT1 contributes to proliferation and metastasis in esophageal squamous cell carcinoma. J Exp Clin Cancer Res 2015; 34: 7.

[31] Shinomiya T, Mori T, Ariyama Y, Sakabe T, Fukuda Y, Murakami Y, Nakamura Y and Inazawa J. Comparative genomic hybridization of squamous cell carcinoma of the esophagus: the possible involvement of the DPI gene in the 13q34 amplicon. Genes Chromosomes Cancer 1999; 24: 337-344.

[32] Pack SD, Karkera JD, Zhuang Z, Pak ED, Balan KV, Hwu P, Park WS, Pham T, Ault DO, Glaser M, Liotta L, Detera-Wadleigh SD and Wadleigh RG. Molecular cytogenetic fingerprinting of esophageal squamous cell carcinoma by comparative genomic hybridization reveals a consistent pattern of chromosomal alterations. Genes Chromosomes Cancer 1999; 25: 160-168.

[33] Du Plessis L, Dietzsch E, Van Gele M, Van Roy N, Van Helden P, Parker MI, Mugwanya DK, De Groot M, Marx MP, Kotze MJ and Speleman F. Mapping of novel regions of DNA gain and loss by comparative genomic hybridization in esophageal carcinoma in the Black and Colored populations of South Africa. Cancer Res 1999; 59: 1877-1883.

[34] Tada K, Oka M, Tangoku A, Hayashi H, Oga A and Sasaki K. Gains of 8q23-qter and 20q and loss of 11q22-qter in esophageal squamous cell carcinoma associated with lymph node metastasis. Cancer 2000; 88: 268-273.

[35] Yen CC, Chen YJ, Chen JT, Hsia JY, Chen PM, Liu JH, Fan FS, Chiou TJ, Wang WS and Lin CH. Comparative genomic hybridization of esophageal squamous cell carcinoma: correlations between chromosomal aberrations and disease progression/prognosis. Cancer 2001; 92: 2769-2777.

[36] Wei F, Ni J, Wu SS, Liu H, Xu X, Han YL, Cai Y, Zhang JW, Chen XJ, Pang H, Lu N, Ji L, Wu M and Wang MR. Cytogenetic studies of esophageal squamous cell carcinomas in the northern Chinese population by comparative genomic hybridization. Cancer Genet Cytogenet 2002; 138: 38-43.

[37] Arai H, Ueno T, Tangoku A, Yoshino S, Abe T, Kawauchi S, Oga A, Furuya T, Oka M and Sasaki K. Detection of amplified oncogenes by genome DNA microarrays in human primary esophageal squamous cell carcinoma: comparison with conventional comparative genomic hybridization analysis. Cancer Genet Cytogenet 2003; 146: 16-21.

[38] Ueno T, Tangoku A, Yoshino S, Abe T, Toshimitsu H, Furuya T, Kawauchi S, Oga A, Oka M and Sasaki K. Gain of 5p15 detected by comparative genomic hybridization as an independent marker of poor prognosis in patients with esophageal squamous cell carcinoma. Clin Cancer Res 2002; 8: 526-533.

[39] Hirasaki S, Noguchi T, Mimori K, Onuki J, Morita K, Inoue H, Sugihara K, Mori M and Hirano T. BAC clones related to prognosis in patients with esophageal squamous carcinoma: an array comparative genomic hybridization study. Oncologist 2007; 12: 406-417.

[40] Qin YR, Wang LD, Fan ZM, Kwong D and Guan XY. Comparative genomic hybridization analysis of genetic aberrations associated with development of esophageal squamous cell carcinoma in Henan, China. World J Gastroenterol 2008; 14: 1828-1835.

[41] Sakai N, Kajiyama Y, Iwanuma Y, Tomita N, Amano T, Isayama F, Ouchi K and Tsurumaru M. Study of abnormal chromosome regions in esophageal squamous cell carcinoma by comparative genomic hybridization: relationship of lymph node metastasis and distant metastasis

to selected abnormal regions. Dis Esophagus 2010; 23: 415-421.
[42] Sun Y, Shi N, Lu H, Zhang J, Ma Y, Qiao Y, Mao Y, Jia K, Han L, Liu F, Li H, Lin Z, Li X and Zhao X. ABCC4 copy number variation is associated with susceptibility to esophageal squamous cell carcinoma. Carcinogenesis 2014; 35: 1941-1950.
[43] Yang YL, Chu JY, Luo ML, Wu YP, Zhang Y, Feng YB, Shi ZZ, Xu X, Han YL, Cai Y, Dong JT, Zhan QM, Wu M and Wang MR. Amplification of PRKCI, located in 3q26, is associated with lymph node metastasis in esophageal squamous cell carcinoma. Genes Chromosomes Cancer 2008; 47: 127-136.
[44] Miyawaki Y, Kawachi H, Ooi A, Eishi Y, Kawano T, Inazawa J and Imoto I. Genomic copy-number alterations of MYC and FHIT genes are associated with survival in esophageal squamous-cell carcinoma. Cancer Sci 2012; 103: 1558-1566.
[45] Pawar H, Kashyap MK, Sahasrabuddhe NA, Renuse S, Harsha HC, Kumar P, Sharma J, Kandasamy K, Marimuthu A, Nair B, Rajagopalan S, Maharudraiah J, Premalatha CS, Kumar KV, Vijayakumar M, Chaerkady R, Prasad TS, Kumar RV and Pandey A. Quantitative tissue proteomics of esophageal squamous cell carcinoma for novel biomarker discovery. Cancer Biol Ther 2011; 12: 510-522.
[46] Takeshita H, Ichikawa D, Komatsu S, Tsujiura M, Kosuga T, Deguchi K, Konishi H, Morimura R, Shiozaki A, Fujiwara H, Okamoto K and Otsuji E. Prediction of CCND1 amplification using plasma DNA as a prognostic marker in oesophageal squamous cell carcinoma. Br J Cancer 2010; 102: 1378-1383.
[47] Wiche G. Role of plectin in cytoskeleton organization and dynamics. J Cell Sci 1998; 111: 2477-2486.
[48] Leung CL, Green KJ and Liem RK. Plakins: a family of versatile cytolinker proteins. Trends Cell Biol 2002; 12: 37-45.
[49] Bouameur JE, Favre B and Borradori L. Plakins, a versatile family of cytolinkers: roles in skin integrity and in human diseases. J Invest Dermatol 2014; 134: 885-894.
[50] Katada K, Tomonaga T, Satoh M, Matsushita K, Tonoike Y, Kodera Y, Hanazawa T, Nomura F and Okamoto Y. Plectin promotes migration and invasion of cancer cells and is a novel prognostic marker for head and neck squamous cell carcinoma. J Proteomics 2012; 75: 1803-1815.
[51] McInroy L and Maatta A. Plectin regulates invasiveness of SW480 colon carcinoma cells and is targeted to podosome-like adhesions in an isoform-specific manner. Exp Cell Res 2011; 317: 2468-2478.
[52] Shin SJ, Smith JA, Rezniczek GA, Pan S, Chen R, Brentnall TA, Wiche G and Kelly KA. Unexpected gain of function for the scaffolding protein plectin due to mislocalization in pancreatic cancer. Proc Natl Acad Sci U S A 2013; 110: 19414-19419.
[53] Burch TC, Watson MT and Nyalwidhe JO. Variable metastatic potentials correlate with differential plectin and vimentin expression in syngeneic androgen independent prostate cancer cells. PLoS One 2013; 8: e65005.
[54] Bausch D, Thomas S, Mino-Kenudson M, Fernandez-del CC, Bauer TW, Williams M, Warshaw AL, Thayer SP and Kelly KA. Plectin-1 as a novel biomarker for pancreatic cancer. Clin Cancer Res 2011; 17: 302-309.
[55] Lazarus P, Zheng Y, Runkle EA, Muscat JE and Wiener D. Genotype-phenotype correlation between the polymorphic UGT2B17 gene deletion and NNAL glucuronidation activities in human liver microsomes. Pharmacogenet Genomics 2005; 15: 769-778.
[56] Nadeau G, Bellemare J, Audet-Walsh E, Flageole C, Huang SP, Bao BY, Douville P, Caron P, Fradet Y, Lacombe L, Guillemette C and Levesque E. Deletions of the androgen-metabolizing UGT2B genes have an effect on circulating steroid levels and biochemical recurrence after radical prostatectomy in localized prostate cancer. J Clin Endocrinol Metab 2011; 96: E1550-1557.
[57] Yoshida T, Kobayashi T, Itoda M, Muto T, Miyaguchi K, Mogushi K, Shoji S, Shimokawa K, Iida S, Uetake H, Ishikawa T, Sugihara K, Mizushima H and Tanaka H. Clinical omics analysis of colorectal cancer incorporating copy number aberrations and gene expression data. Cancer Inform 2010; 9: 147-161.
[58] Setlur SR, Chen CX, Hossain RR, Ha JS, Van Doren VE, Stenzel B, Steiner E, Oldridge D, Kitabayashi N, Banerjee S, Chen JY, Schafer G, Horninger W, Lee C, Rubin MA, Klocker H and Demichelis F. Genetic variation of genes involved in dihydrotestosterone metabolism and the risk of prostate cancer. Cancer Epidemiol Biomarkers Prev 2010; 19: 229-239.
[59] McQuaid KR, Laine L, Fennerty MB, Souza R and Spechler SJ. Systematic review: the role of bile acids in the pathogenesis of gastro-oesophageal reflux disease and related neoplasia. Aliment Pharmacol Ther 2011; 34: 146-165.
[60] Quante M, Bhagat G, Abrams JA, Marache F, Good P, Lee MD, Lee Y, Friedman R, Asfaha S, Dubeykovskaya Z, Mahmood U, Figueiredo JL, Kitajewski J, Shawber C, Lightdale CJ, Rustgi AK and Wang TC. Bile acid and inflammation activate gastric cardia stem cells in a mouse model of Barrett-like metaplasia. Cancer Cell 2012; 21: 36-51.
[61] Wang QM, Yuan L, Qi YJ, Ma ZY and Wang LD. Estrogen analogues: promising target for prevention and treatment of esophageal squa-

mous cell carcinoma in high risk areas. Med Sci Monit 2010; 16: HY19-22.

[62] Hyland PL, Freedman ND, Hu N, Tang ZZ, Wang L, Wang C, Ding T, Fan JH, Qiao YL, Golozar A, Wheeler W, Yu K, Yuenger J, Burdett L, Channock SJ, Dawsey SM, Tucker MA, Goldstein AM, Abnet CC and Taylor PR. Genetic variants in sex hormone metabolic pathway genes and risk of esophageal squamous cell carcinoma. Carcinogenesis 2013; 34: 1062-1068.

Supplementary Table 1. Primers used in construction of recombinant plasmids

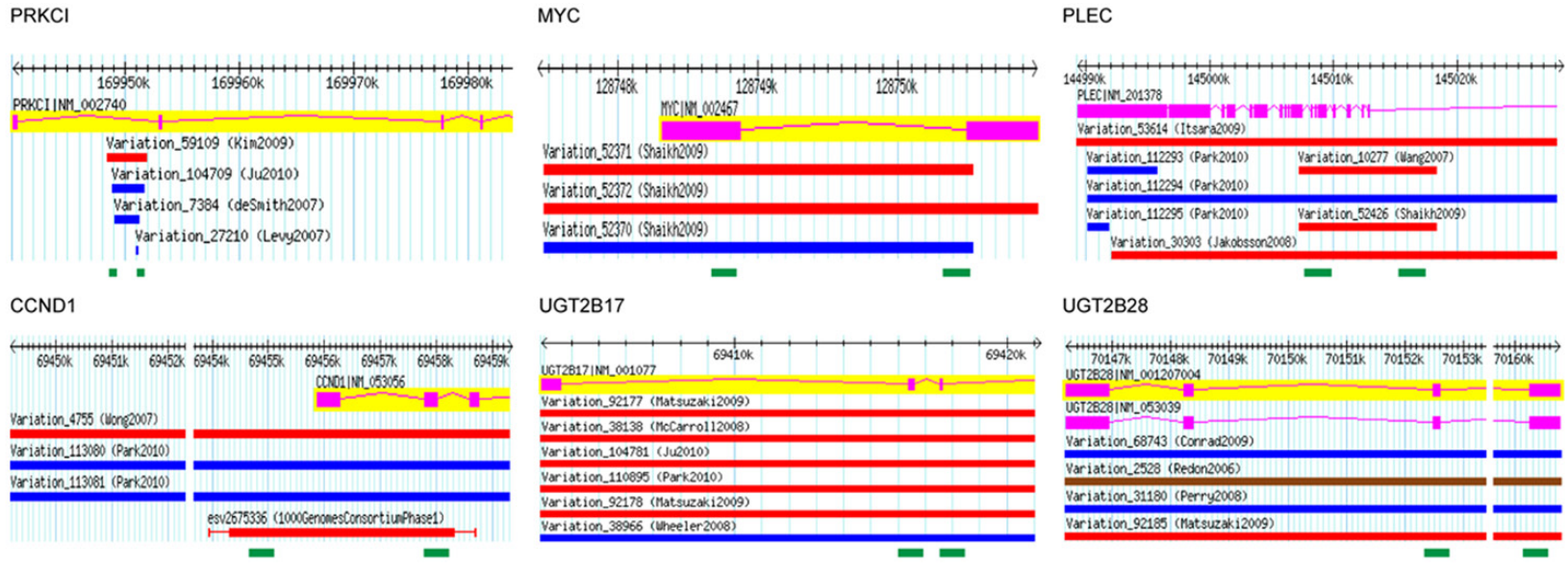
Primers	Sequences
PRKCI-Forward	CGACGCGTTGCCATTTTGGGAATCTTTGTGA
PRKCI-Reverse	CCGCTCGAGCCATCAGAGACTCAATGAAGGAG
PEX5L-Forward	CGACGCGTGACAGTAGCTGTCATCGCAGT
PEX5L-Reverse	CCGCTCGAGTTGCAGGCAAAAATGTCCAGT
FAM168A-Forward	CGACGCGTTCTGCCCCATCCATACCAAC
FAM168A-Reverse	CCGCTCGAGTCAACGGAGAGTGAGTGGCA
PTK2-Forward	CGACGCGTACAAGGTCAAGGATGAGCGG
PTK2-Reverse	CCGCTCGAGTGAAATCCTACCCGAATCTTACTGA
PACS1-Forward	CGACGCGTAGTCTTCTCTGGGTTCATGGC
PACS1-Reverse	CCGCTCGAGGGTTCCGAAGCCTTTGCTCT
Up-SKIL-Forward	CGACGCGTGAAGCAAAGCAGTTAAGGCCC
Up-SKIL-Reverse	CCGCTCGAGTCCTGAATGGACAAAGGCCAG
ANO1-Forward	CGGGTTACCTCGTCCTTTGGCCGATTTCA
ANO1-Reverse	CGACGCGTGCACTGTGAACCCCTTGACT

The red characters indicated the enzymes restriction sites. Mlu I: CGACGCGT; Xho I: CCGCTCGAG; Kpn I: CGGGTTACC.

Supplementary Table 2. Probes used in Accucopy™ system

DNA segments	Chromosome	Location (Ref37 database)	Product length (Sample, reference)	Notes	Primer-forward	Primer-reverse
2p	Chr2	84500611-84500685	103 (+0, -2)	Reference 1	TGAGCCAAAAATTCAGAATACAAGGA	GTTTGCCTGCCTTCCAAGCAA
CCND1-2	Chr11	69457077-69457158	113 (+0, -2)		CCTGGGAGCACATTTTCAGACC	TAATTTTTGAAAGTGCGGCGTG
MALAT1-2	Chr11	65269302-65269383	121 (+0, -2)		TTTTATTAAAGGGGAGGGGCAAAT	GATTGGTGGGGTGGGTTTAGGT
MYC-2	Chr8	128750292-128750405	142 (+0, -2)		CTCCGAGATAGCAGGGGACTGT	CAAGGGGAGAGGTTCGGGAC
UGT2B28_2	Chr4	70152227-70152345	150 (+0, -2)		TTACACACATGCAGACACATACCTACATAA	TAAAGAACCTGAGTGATTGAGTCAGTTAAAA
MYC-1	Chr8	128748849-128748977	161 (+0, -2)		TTTCGGGTAGTGGAAAACCAGGTA	TTCTCCCATTCCTGCGCTATTG
10pL	Chr10	31120531-31120675	173 (+0, -2)	Reference 2	CACTGAGCCCCAGAGACCTGAC	GTTTTCCCTGGAGGTGTGCATT
UGT2B17_2	Chr4	69416236-69416396	193 (+0, -2)		ATGTGAGTATTACAATTTTGTGACCAGGTG	CTCTTATTTTCCACCAGACAATTATTTCAA
PRKCI-1	Chr3	169949038-169949219	210 (+0, -2)		CGTGCCCGGCCTATTCCTAT	TCATCTGCAAGTAGCAATTTTGCC
PLEC-1	Chr8	145007360-145007567	236 (+0, -2)		CCCACCTGCGGTACATCTGC	CACAGCTGATCCCCCTCTGAC
20q	Chr20	35865921-35866142	250 (+0, -2)	Reference 3	AGGGTGCTGGGATCAGAGAGAG	CATTTTGCCACCCTCCAGTAGC
PLEC-2	Chr8	145017625-145017856	260 (+0, -2)		ACAGGCACCTGCCCACTTGT	TTCACTGACATTCAGCTTGGAAACA
PRKCI-2	Chr3	169951166-169951407	270 (+0, -2)		TTGCCATTTTGGGAATCTTTGTG	TGGTTTCGCAATTACTGAGAGCTAAG
UGT2B17_1	Chr4	69417527-69417804	306 (+0, -2, +4)		AATGTGGGTATTCAGCTTACCTCAGAC	TGACCTTCTTGGTAAGACTCTGGAGA
MALAT1-1	Chr11	65266732-65267016	317 (+0, -2)		AGCATGAGGAAGGAAAAGATAAAAGG	ACCTTGAAATCCATGACGCAGG
16p	Chr16	25258129-25258426	326 (+0,-2)	Reference 4	TCCTCCACCAAGCTGATGTGTT	CTATTTCGGGGACAGGCCTGAA
UGT2B28_1	Chr4	70160175-70160485	339 (+0, -2)		CCTTTCATAGACTTGATACATACAGGCCAG	GTCGTCACAAAGTTTTGTCTGTTTTGTTTC
CCND1-1	Chr11	69454551-69454877	355 (+0, -2)		CAGTTTTCTTAATTTGGGGCAGGT	AAGGTGAAGGGACGTCTACACCC

The red lines indicate the four references DNA segments used for normalization.



Supplementary Figure 1. Position and probes information of the target CNV sites. The recorded copy number gains (blue) and losses (red) on Database of Genomic Variants were presented. The green boxes indicate the probes for amplification of the selected CNVs (Not proportional to the actual length of amplification products).

Supplementary Table 3. Comprehensive information of the ten primary candidate CNV sites returned by bioinformatic analysis

Location (GRCh37, hg19)	CNVs across this region	CNV genotype	Reporter	Frequency	Location in genes	Regulatory Signal from ENCODE			
						H3K27Ac Mark 7 cell lines (Intensity)	H3K4Me3 Mark 7 cellines (Intensity)	TF binding sites by chip-seq	Dnasel Hypersensitivity Mark in 125 cell lines (Positive/Total)
chr3: 169948852.169951652	Variation_59109	Loss	Kim	1/1	PRKCI	Moderate	Moderate	CEBPB	67/125
	Variation_104709	Gain	JU	1/1	1st Intron			STAT3	
	Variation_7384	Gain	deSmith	3/50					
chr3: 170062604.170063770	esv_2677963	Loss	1000_Genomes_Consortium	40/1151	Upstream of SKIL	Moderate	Moderate	-	34/125
			Matsuzaki						
			Conrad						
	Variation_91699	Loss	Matsuzaki	27/90					
	Variation_63659	Loss	ju	54/450					
	Variation_79997	Gain		2/90					
	Variation_104710	Gain		1/1					
chr3: 181136547.181138049	Variation_80015	Gain	Matsuzaki	16/90	PEX5L	Moderate	Moderate	-	-
	Variation_98546	Loss	McKernan	1/1	Intron				
	Variation_91727	Loss	Matsuzaki	27/90					
chr8: 128747471.128750540	Variation_52370	Gain	Shaikh	2/2026	MYC	Strong	Strong	E2F1	125/125
	Variation_52371	Loss	Shaikh	3/2026	Cross exons			HEY1	
	Variation_4609	Loss	Wong	4/95				TAF1 et al	
	Variation_8615	Gain	Pinto	1/776					
chr8: 145007187.145018354	Variation_112294	Gain	Park	1/30	PLEC	Strong	Strong	NFKB	124/125
	Variation_52426 et al.	Loss	Shaikh	12/2026	Cross Exons			c-myc	
	Variation_10277								
	Variation_30303	Loss	Wang	1/112					
	Variation_53614	Loss	Jakobsson	19/485					
		Loss	Itsara	3/1854					
chr8: 142002613.142005570	Variation_65286	Gain	Conrad	18/450	PTK2	Moderate	Moderate	-	-
	Variation_31454	Gain	Perry	2/30	1st intron				
	Variation_103828	Gain	Altshuler	32/1184					
	Variation_82807	Gain	Matsuzaki	6/90					
chr11: 65933958.65939397	Variation_85825	Loss	Matsuzaki	19/90	PACS1	Moderate	Moderate	-	4/125
	Variation_65975	Loss	Conrad	49/450	Intron				
	Variation_37486	Loss	Cooper	12/126					
chr11: 69454316.69458322	esv2675336	Loss	1000_Genomes	248/1151	CCND1	Strong	Strong	P300	124/125
			Park					TCF4	
	Variation_113081	Gain	Park	1/30				SP1	
	Variation_113080	Gain	Wong	1/30				CTCF et al	
	Variation_4755	Loss		6/95					
chr11: 69982591.69984115	Variation_65986	Loss	Conrad	82/450	ANO1	Moderate	Moderate	CTCF RAD21	10/125
chr11: 69659795.69662302	Variation_113083	Loss	Park	26+3/30	11th intron				
chr11: 73307810.73308594	Variation_85850	Loss	Matsuzaki	8/90	FAM168A	Strong	Strong	PAX4	16/125
	Variation_75699	Gain	Matsuzaki	4/90	Intron				
	Variation_65991	Gain	Conrad	8/450					