Biostatistics statement

The statistical methods of this study were reviewed by Tian Feng from National Engineering Research Center for Miniaturized Detection Systems. (1) Statistical methods are adequately and appropriately described when they are used to verify the results; (2) The statistical techniques are suitable; (3) Only homogeneous data can be averaged. Standard deviations are preferred to standard errors. The number of observations and subjects (n) is given. Losses in observations, are reported; (4) Values have the 95% confidence limits calculated and have been compared by weighted probit modeling; and (5) The word "significantly" is replaced by its synonyms or the P value (if it indicates statistical significance).

The SPSS18.0 statistical software and Microsoft Excel were used for statistical analysis. In all analyses, the lower frequency allele was coded as the 'risk' allele. All p values presented in this study were two sided, and we used $p \leq 0.05$ as the cutoff value for statistical significance. An exact test was used to assess the variation in each SNP frequency from Hardy-Weinberg equilibrium (HWE) in the control subjects. Differences in SNP genotype distribution between cases and controls were compared by χ^2-test Odds ratios and 95% confidence intervals (CIs) were determined using unconditional logistic regression analysis with adjustments for age and gender.

Associations between SNPs and risk of esophageal cancer were tested in genetic models by analysis with SNP Stats software, obtained from http://bioinfo.iconcologia.net. For the additive model, individuals were assigned a 0, 1 or 2 representing the number of risk alleles they possessed for that SNP. For the dominant model, individuals were coded as 1 if they carried at least 1 risk allele and 0 otherwise; for the recessive model, Individuals were coded as 1 if they were homozygous for the risk allele (two copies) and 0 otherwise. Values of OR and 95% CI were calculated as above. Akaike's Information Criterion and Bayesian Information Criterion were applied to estimate the best-fit model for each SNP.

Name: Tian Feng

Email: fengtianalice@gmail.com

Signature: Tian Feng

Date: Jan 29, 2015