

Association of *DFNA5*, *SYK* & *NELL1* variants along with HPV infection in oral cancer among the prolonged tobacco-chewers

Sharbadeb Kundu¹, Vijayalakshmi Ramshankar⁴, Akalesh Kumar Verma³, Soundara Viveka Thangaraj⁴, Arvind Krishnamurthy⁵, Rajeev Kumar⁶, Ravi Kannan⁶, Sankar Kumar Ghosh^{1,2*}

Supplementary Information

The **Additional File 1.pdf** file contains the *Supplementary Tables* viz. Supplementary Table S1 to Supplementary Table S5. The **Additional File 2.pdf** file contains the *Supplementary Figures* viz. Supplementary Figure S1 to Supplementary Figure S4. The table titles as well as the figure captions are mentioned below.

Supplementary Tables

Additional File 1

Supplementary Table S1. List of populations taken under this study

Supplementary Table S2. List of primers used in the targeted re-sequencing experiment

Supplementary Table S3 List of 115 significant SNPs associated with oral cancer obtained from "Discovery Set" samples

Supplementary Table S4. List of SNPs distributed in four pools for Sequenom experiment

Supplementary Table S5. Overall summary of the strongly associated SNPs obtained for **Case1** and **Case2** after stratified analysis

Supplementary Figures

Additional file 2

Supplementary Figure S1 Immunohistochemistry study of the case1 and case2 groups. Here, BSA was used as negative control and cervical cancer tissue specimen was considered as the positive control for the Immunohistochemistry study.

Supplementary Figure S2. Representative LD plot from Haploview software. (A) LD plot for chromosome 11 [$r^2 = 0.975$ for LD Block 1], (B) LD plot for chromosome 9 [$r^2 > 0.8$ for LD Block 1] and (C) LD plot for chromosome 7 [$r^2 < 0.8$ for LD Block 1 and $r^2 \geq 0.8$ for LD Block 2]. The figure is not to scale.

Supplementary Figure S3. Estimated population structure for 12 Northeast Indian heterogeneous ethnic groups taken under this study. (A) Estimation of the number of clusters (K) by Delta-K method (Ref. 31). (B) The clustering (for K=3) was done based on major language families, which are labeled above the figure. The individual populations are numbered below the figure.

Supplementary Figure S4. Top 10 interacting chemicals with *DFNA5* gene. The red arrow signifies the chemical results in *increased expression* and the green arrow signifies the chemical results in *decreased expression* of the gene.