



Article

Methylation-Based Classification of Cervical Squamous Cell Carcinoma into Two New Subclasses Differing in Immune-Related Gene Expression

Xia Li ^{1,*} and Yunpeng Cai ^{1,*}

¹ Research Center for Biomedical Information Technology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, 1068 Xueyuan Avenue, Shenzhen University Town, Shenzhen 518055, China

* Correspondence: xia.li@siat.ac.cn / bioinflux@126.com (X.L.); Tel.: (86)755-86392203; yp.cai@siat.ac.cn (Y.C.); Tel.: (86)755-86392202

Supplementary Figures

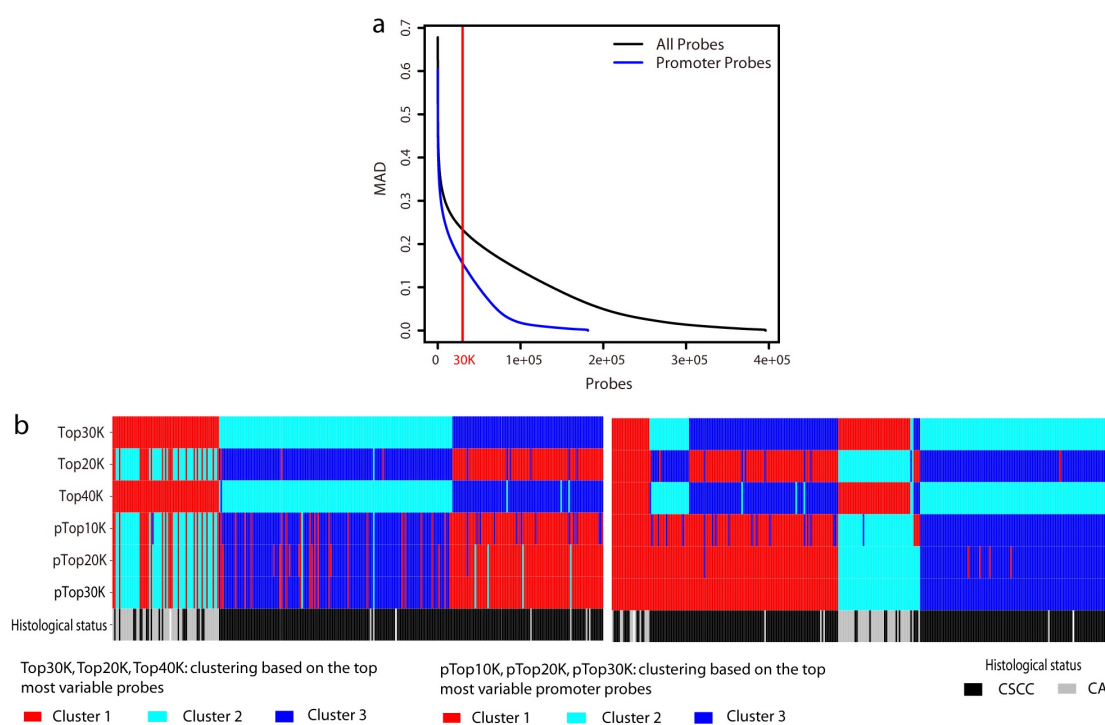


Figure S1. Unsupervised clustering based on the top most variable probes and promoter probes. (a) Distribution of MAD values for all probes and promoter probes. Vertical line in red represents the top 30,000 (30K) most variable probes that showed the highest MAD across beta values. We selected these probes for clustering in this study; (b) The top 20,000 (Top20K) and 40,000 (Top40K) probes, the top 10,000 (pTop10K), 20,000 (pTop20K), and 30,000 (pTop30K) most variable promoter probes were also used to perform unsupervised hierarchical clustering, respectively. Heatmap shows patients separation in each clustering and their histological status. For clarity, patients were ordered according to their cluster in clustering using the top 30,000 probes (left heatmap), and top 30,000 promoter probes (right heatmap), respectively.

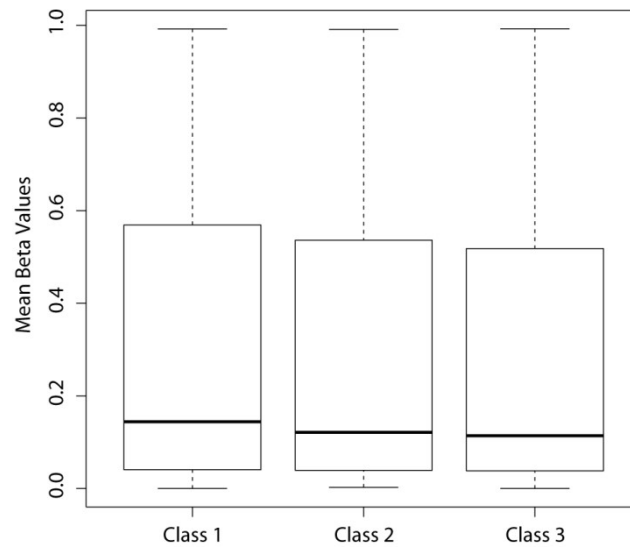


Figure S2. Distribution of mean beta values of probes in each subclass.

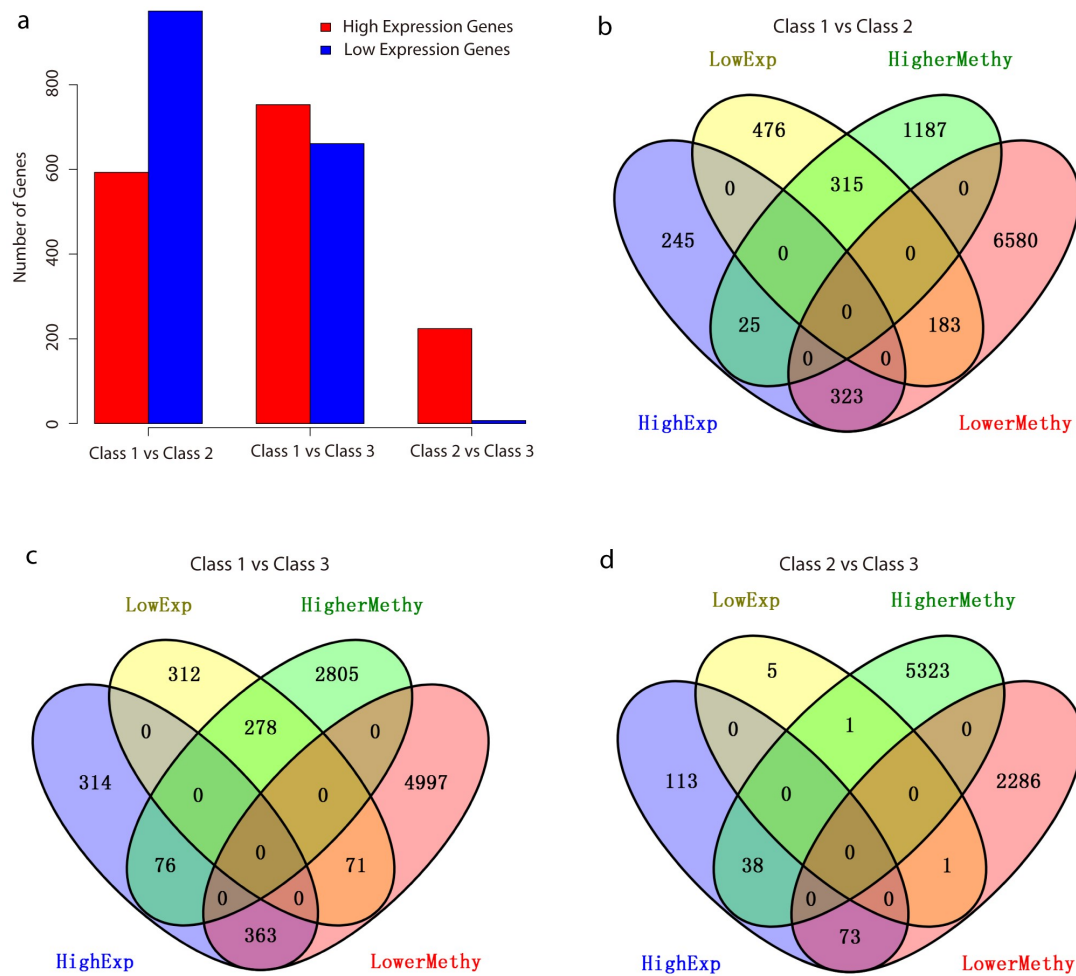


Figure S3. Number of differentially expressed genes and overlap with differentially methylated genes.

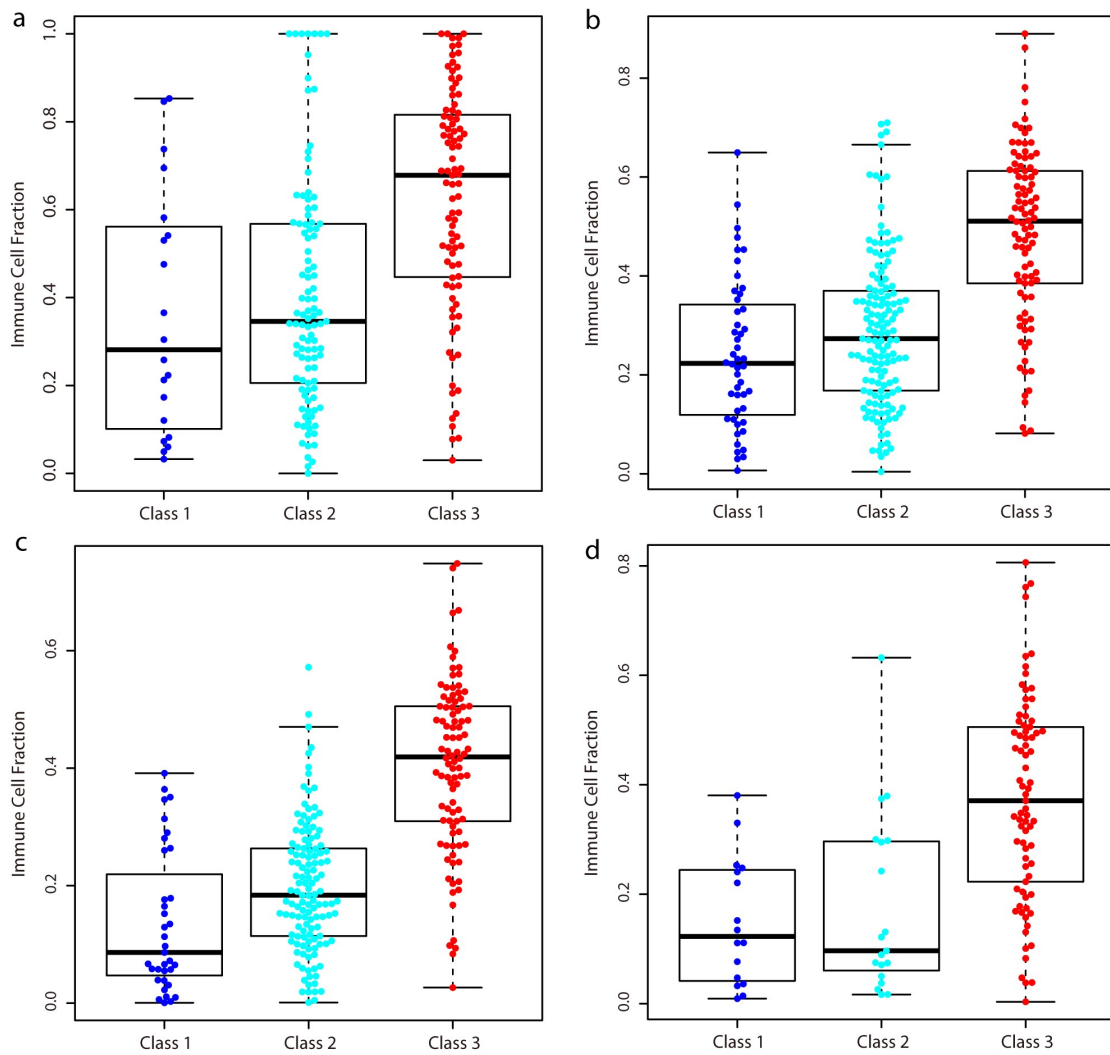


Figure S4. Distribution of immune cell fraction in three groups. Boxplots show the immune cell fraction which was calculated based on top 30,000 most variable probes (a), the differentially methylated probes between Class 1 and Class 2 (b), Class 1 and Class 3 (c), and Class 2 and Class 3 (d), respectively.

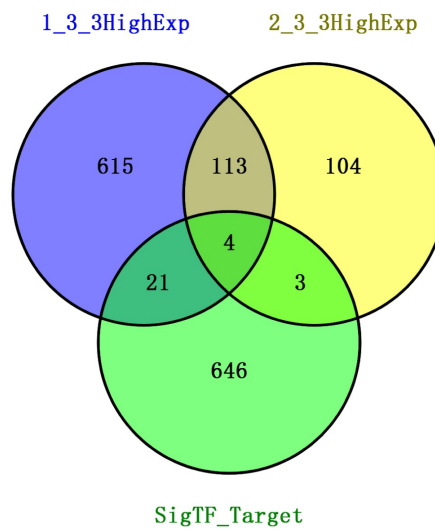


Figure S5. 28 target genes were highly expressed in Class 3. Venn representation of overlaps among highly expressed genes in Class 3 compared to Class 1 ("1_3_3HighExp"), highly expressed genes in Class 3 compared to Class 2 ("2_3_3HighExp"), and target genes of those 5 significantly enriched TFs ("SigTF_targets").

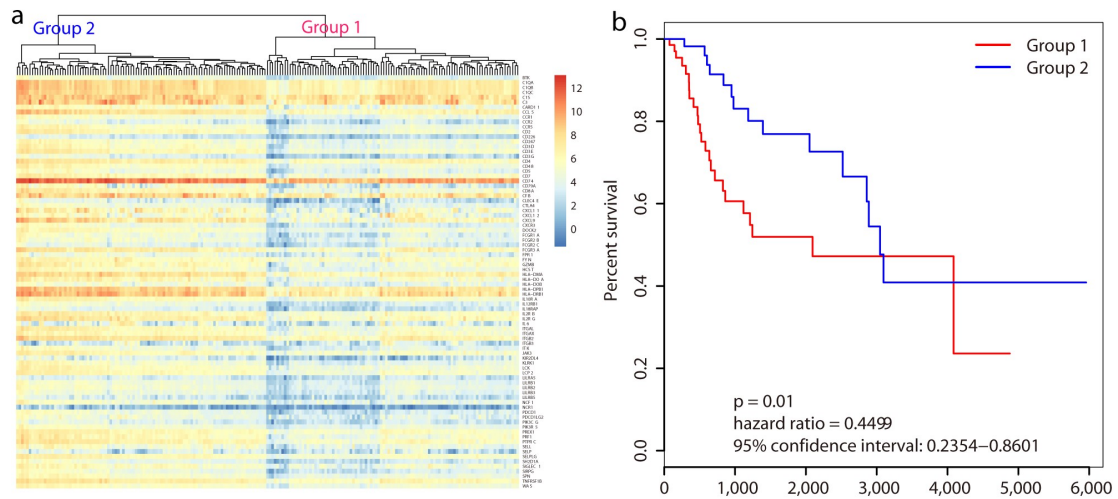


Figure S6. Clinical comparison of the CSCC patients based on the immune genes' expression. We observed a total of 84 immune-related genes were highly expressed in Class 3 as compared with Class 2. The gene expression of these 84 genes could divide patients in CSCC into two groups (a). Survival analysis showed the group with high expression displayed better overall survival (b).

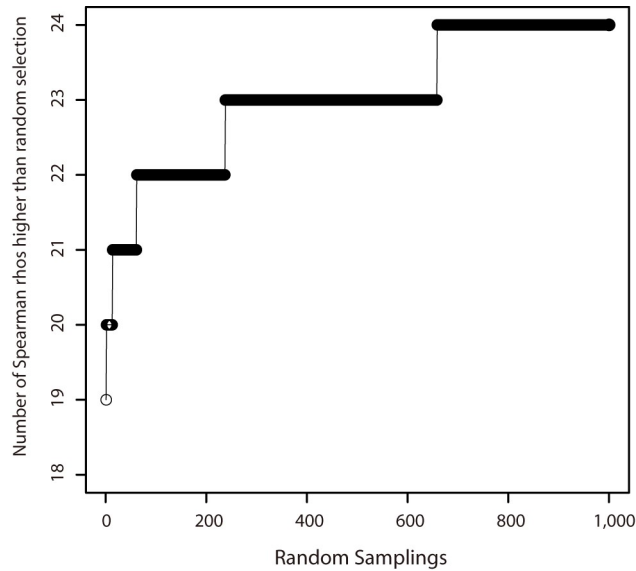


Figure S7. Comparison of Spearman's Rho across 1,000 samplings. Each point represents the number of Spearman's rho among the 24 genes higher than the 24 randomly selected genes' in each sampling.

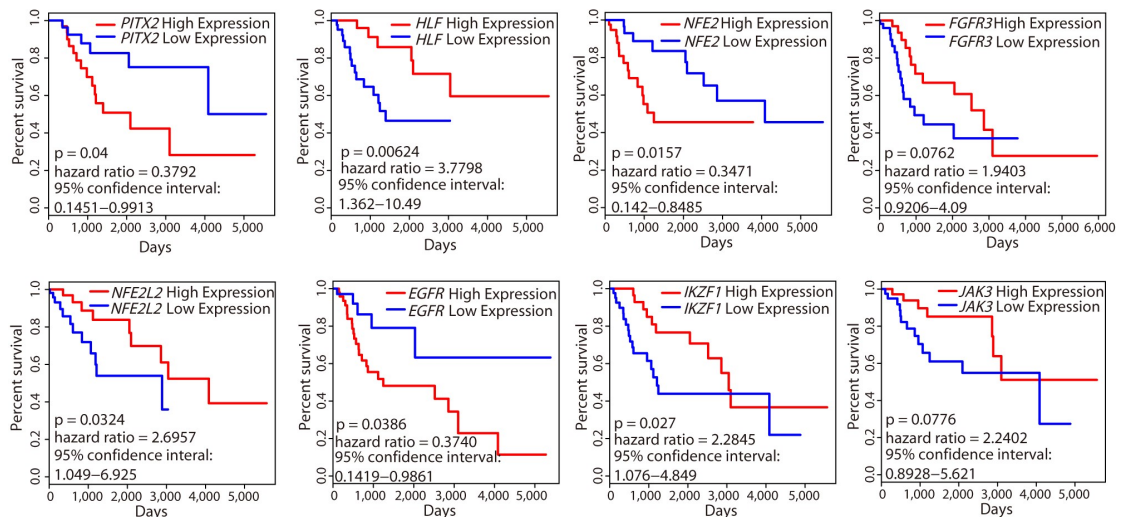


Figure S8. Association between TFs and driver genes' expression and patient survival. Kaplan-Meier curves were constructed to look at the difference in the survival of patients with high and low expression of TFs and driver genes. Patient of high expression was chosen if the expression value was above the fourth quantile, and low expression if below the second quantile. Differences were evaluated by the Log-rank (Mantel-Cox) test.

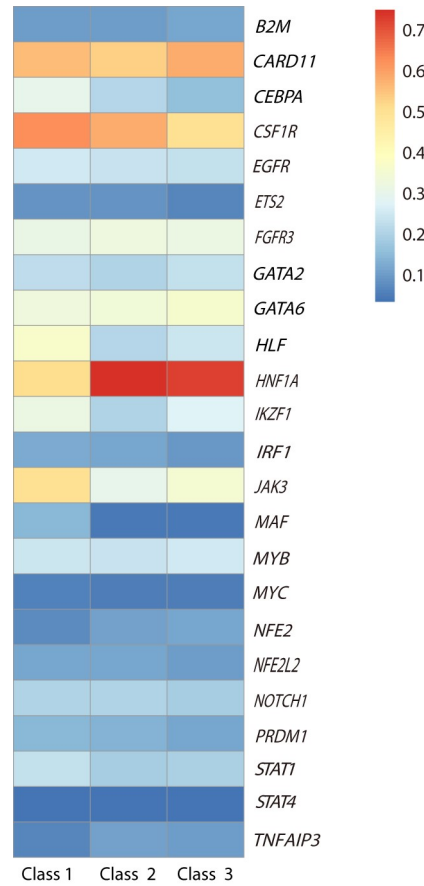


Figure S9. Promoter methylation status of the differentially expressed TFs and driver genes. Heatmap shows the mean beta values of all promoter probes belonging to each gene in each class.

Table S1. Expression of 28 highly expressed TF target genes in cervix and cancer.

Gene	Cervical Cancer Average RPKM ¹	Cervix GTEx RPKM ¹
<i>ARHGAP30</i>	6.1	3
<i>BCL11B</i>	1.9	0.6
<i>CDH13</i>	2.3	12.7
<i>CDK5R1</i>	2.9	0.9
<i>CDKN2B</i>	26.5	6.2
<i>CLIP1</i>	8.1	12.8
<i>FGD2</i>	0.9	3.4
<i>GNAI1</i>	10.8	8.5
<i>HOXA7</i>	1.1	1.3
<i>HOXD10</i>	6.8	12.5
<i>HOXD11</i>	3.7	1.5
<i>HOXD4</i>	0.4	2.3
<i>IRX4</i>	5.4	1.7
<i>LCP1</i>	20.1	4.2
<i>MAFB</i>	18.4	13.7
<i>MARCH1</i>	0.7	1.5
<i>NHLH2</i>	0.3	0.1

<i>PAK6</i>	0.2	1.9
<i>PPP2R3A</i>	3.8	2.8
<i>PVRL1</i>	43.8	7.5
<i>TMEM22</i>	1.1	2.9
<i>C1S</i>	35.2	210.1
<i>CTLA4</i>	2	0.2
<i>SCML4</i>	0.2	0.2
<i>WAS</i>	4.8	2
<i>CARD11</i>	3.3	1.2
<i>CD226</i>	0.2	0.1
<i>FYN</i>	3.5	16.5

¹Gene expression value was obtained from the Human Protein Atlas.