**Supplementary Table 1 The methylation status of the CpG sites located at the four genes in the TCGA dataset**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Gene** | **Chr** | **Pos** | **McaM** | **McoM** | **Pvalue** | **FDR** | **Sens** | **Spec** | **AUC** |
| EOMES-cg06748434 | 3 | 27761408 | 0.629 | 0.448 | 2.10E-05 | 0.000192 | 0.73 | 0.88 | 0.84 |
| EOMES-cg20235075 | 3 | 27761503 | 0.821 | 0.724 | 2.20E-04 | 0.001414 | 0.76 | 0.75 | 0.79 |
| EOMES-cg20739013 | 3 | 27761677 | 0.722 | 0.537 | 3.00E-05 | 0.000257 | 0.94 | 0.69 | 0.83 |
| EOMES-cg21473142 | 3 | 27762095 | 0.735 | 0.389 | 2.60E-06 | 0.000047 | 0.96 | 0.81 | 0.87 |
| EOMES-cg15270892 | 3 | 27762352 | 0.274 | 0.329 | 2.30E-01 | 0.358197 | 0.42 | 0.88 | 0.6 |
| EOMES-cg09665320 | 3 | 27762636 | 0.111 | 0.139 | 8.40E-03 | 0.024102 | 0.6 | 0.81 | 0.71 |
| EOMES-cg16943529 | 3 | 27762805 | 0.035 | 0.109 | 7.70E-03 | 0.022834 | 0.62 | 0.81 | 0.71 |
| EOMES-cg13830203 | 3 | 27762916 | 0.108 | 0.195 | 2.20E-07 | 0.000011 | 0.79 | 1 | 0.91 |
| EOMES-cg24434959 | 3 | 27763102 | 0.052 | 0.137 | 2.60E-01 | 0.387955 | 0.5 | 0.69 | 0.59 |
| EOMES-cg02797548 | 3 | 27763266 | 0.031 | 0.097 | 7.40E-01 | 0.785192 | 0.98 | 0.19 | 0.47 |
| EOMES-cg14557534 | 3 | 27763581 | 0.037 | 0.118 | 1.60E-04 | 0.001077 | 0.77 | 0.81 | 0.8 |
| EOMES-cg12858460 | 3 | 27763737 | 0.051 | 0.136 | 3.80E-06 | 0.00005 | 0.83 | 0.88 | 0.87 |
| EOMES-cg15308664 | 3 | 27763797 | 0.047 | 0.121 | 1.40E-07 | 0.000011 | 0.83 | 0.94 | 0.92 |
| EOMES-cg01176458 | 3 | 27763828 | 0.154 | 0.202 | 6.00E-01 | 0.675346 | 0.29 | 0.94 | 0.54 |
| EOMES-cg07055504 | 3 | 27763865 | 0.029 | 0.069 | 1.80E-01 | 0.298538 | 0.75 | 0.5 | 0.61 |
| EOMES-cg01714811 | 3 | 27764473 | 0.538 | 0.252 | 1.50E-05 | 0.000151 | 0.83 | 0.81 | 0.84 |
| EOMES-cg02694810 | 3 | 27764660 | 0.363 | 0.124 | 1.20E-04 | 0.000885 | 0.76 | 0.88 | 0.8 |
| EOMES-cg16971668 | 3 | 27764810 | 0.592 | 0.25 | 4.00E-06 | 0.00005 | 0.89 | 0.81 | 0.87 |
| EOMES-cg22383888 | 3 | 27764816 | 0.539 | 0.185 | 2.80E-06 | 0.000047 | 0.89 | 0.81 | 0.87 |
| EOMES-cg06014401 | 3 | 27765232 | 0.455 | 0.172 | 1.70E-06 | 0.000044 | 0.89 | 0.88 | 0.88 |
| EOMES-cg15540820 | 3 | 27765283 | 0.599 | 0.292 | 4.80E-06 | 0.000054 | 0.95 | 0.81 | 0.86 |
| SALL1-cg27022663 | 16 | 51170280 | 0.714 | 0.864 | 1.30E-03 | 0.005624 | 0.76 | 0.75 | 0.76 |
| SALL1-cg04962978 | 16 | 51170800 | 0.796 | 0.865 | 2.70E-01 | 0.394419 | 0.45 | 0.88 | 0.59 |
| SALL1-cg08304084 | 16 | 51171082 | 0.737 | 0.879 | 4.70E-01 | 0.567723 | 0.38 | 0.94 | 0.56 |
| SALL1-cg06746670 | 16 | 51171260 | 0.747 | 0.839 | 4.80E-01 | 0.567723 | 0.5 | 0.94 | 0.56 |
| SALL1-cg01679108 | 16 | 51174069 | 0.323 | 0.369 | 6.70E-02 | 0.126138 | 0.62 | 0.75 | 0.65 |
| SALL1-cg02498177 | 16 | 51174131 | 0.523 | 0.633 | 4.50E-02 | 0.094077 | 0.71 | 0.69 | 0.66 |
| SALL1-cg07726904 | 16 | 51174177 | 0.902 | 0.958 | 8.60E-03 | 0.024102 | 0.74 | 0.69 | 0.71 |
| SALL1-cg06099646 | 16 | 51174768 | 0.913 | 0.926 | 9.30E-01 | 0.938115 | 0.18 | 1 | 0.51 |
| SALL1-cg05044764 | 16 | 51175026 | 0.57 | 0.669 | 1.90E-02 | 0.044512 | 0.61 | 0.81 | 0.69 |
| SALL1-cg02018076 | 16 | 51175381 | 0.656 | 0.722 | 4.80E-01 | 0.567723 | 0.39 | 0.94 | 0.56 |
| SALL1-cg16528497 | 16 | 51175541 | 0.719 | 0.881 | 8.00E-04 | 0.004053 | 0.7 | 0.88 | 0.77 |
| SALL1-cg06497955 | 16 | 51175727 | 0.881 | 0.941 | 3.00E-01 | 0.41953 | 0.26 | 1 | 0.58 |
| SALL1-cg03961481 | 16 | 51180521 | 0.395 | 0.21 | 3.70E-03 | 0.01254 | 0.67 | 0.81 | 0.73 |
| SALL1-cg08878368 | 16 | 51183363 | 0.724 | 0.62 | 3.40E-03 | 0.01254 | 0.6 | 0.81 | 0.73 |
| SALL1-cg08199953 | 16 | 51183473 | 0.669 | 0.402 | 9.40E-05 | 0.000727 | 0.88 | 0.69 | 0.81 |
| SALL1-cg02417084 | 16 | 51183533 | 0.485 | 0.361 | 6.60E-02 | 0.125855 | 0.65 | 0.69 | 0.65 |
| SALL1-cg02288754 | 16 | 51183988 | 0.355 | 0.253 | 3.60E-03 | 0.01254 | 0.74 | 0.81 | 0.73 |
| SALL1-cg10297473 | 16 | 51184152 | 0.49 | 0.308 | 5.50E-04 | 0.003243 | 0.87 | 0.75 | 0.77 |
| SALL1-cg04844564 | 16 | 51184205 | 0.439 | 0.305 | 6.30E-03 | 0.019305 | 0.81 | 0.75 | 0.72 |
| SALL1-cg04550052 | 16 | 51184355 | 0.444 | 0.209 | 2.40E-03 | 0.010216 | 0.8 | 0.81 | 0.74 |
| SALL1-cg04698114 | 16 | 51184379 | 0.455 | 0.213 | 4.70E-03 | 0.014906 | 0.81 | 0.75 | 0.72 |
| SALL1-cg00582524 | 16 | 51184392 | 0.546 | 0.296 | 3.00E-03 | 0.011705 | 0.74 | 0.81 | 0.74 |
| SALL1-cg08526074 | 16 | 51184562 | 0.258 | 0.147 | 6.10E-02 | 0.120191 | 0.62 | 0.75 | 0.65 |
| SALL1-cg02864757 | 16 | 51184583 | 0.257 | 0.17 | 2.60E-02 | 0.058745 | 0.6 | 0.81 | 0.68 |
| SALL1-cg05151154 | 16 | 51184786 | 0.19 | 0.171 | 4.80E-01 | 0.567723 | 0.43 | 0.81 | 0.56 |
| SALL1-cg07498275 | 16 | 51184886 | 0.25 | 0.242 | 4.50E-01 | 0.567066 | 0.42 | 0.81 | 0.56 |
| SALL1-cg08806408 | 16 | 51185001 | 0.195 | 0.129 | 1.80E-01 | 0.298538 | 0.45 | 0.81 | 0.61 |
| SALL1-cg06653699 | 16 | 51185082 | 0.181 | 0.122 | 1.10E-01 | 0.192725 | 0.68 | 0.62 | 0.63 |
| SALL1-cg09016242 | 16 | 51185110 | 0.345 | 0.251 | 2.40E-01 | 0.375672 | 0.48 | 0.81 | 0.59 |
| SALL1-cg06232807 | 16 | 51185132 | 0.469 | 0.442 | 5.90E-01 | 0.668148 | 0.43 | 0.81 | 0.54 |
| SALL1-cg01146232 | 16 | 51185147 | 0.381 | 0.353 | 6.70E-01 | 0.7186 | 0.43 | 0.81 | 0.53 |
| SALL1-cg06274671 | 16 | 51185236 | 0.107 | 0.084 | 1.20E-02 | 0.030796 | 0.74 | 0.75 | 0.7 |
| SALL1-cg00310215 | 16 | 51185346 | 0.279 | 0.231 | 6.20E-01 | 0.682384 | 0.56 | 0.69 | 0.54 |
| SALL1-cg22674717 | 16 | 51185394 | 0.118 | 0.13 | 6.10E-01 | 0.682384 | 0.74 | 0.44 | 0.54 |
| SALL1-cg05213609 | 16 | 51185407 | 0.214 | 0.183 | 3.20E-01 | 0.432204 | 0.43 | 0.81 | 0.58 |
| SALL1-cg06724588 | 16 | 51185459 | 0.256 | 0.195 | 4.50E-01 | 0.567066 | 0.43 | 0.81 | 0.56 |
| SALL1-cg07502439 | 16 | 51185461 | 0.173 | 0.12 | 2.90E-01 | 0.401059 | 0.51 | 0.75 | 0.58 |
| SALL1-cg01500945 | 16 | 51185672 | 0.195 | 0.23 | 3.50E-01 | 0.477238 | 0.63 | 0.56 | 0.57 |
| SALL1-cg13755795 | 16 | 51185772 | 0.177 | 0.167 | 4.00E-01 | 0.51148 | 0.5 | 0.69 | 0.57 |
| SALL1-cg05404010 | 16 | 51186211 | 0.131 | 0.167 | 9.70E-01 | 0.966252 | 0.95 | 0.19 | 0.5 |
| SALL1-cg08776356 | 16 | 51186260 | 0.132 | 0.156 | 3.90E-01 | 0.504511 | 0.27 | 1 | 0.57 |
| SALL1-cg00177451 | 16 | 51186266 | 0.14 | 0.164 | 1.30E-01 | 0.228883 | 0.62 | 0.69 | 0.62 |
| SALL1-cg27423760 | 16 | 51186463 | 0.208 | 0.187 | 7.70E-01 | 0.814618 | 0.33 | 0.81 | 0.52 |
| ADHFE1-cg16642695 | 8 | 67343633 | 0.472 | 0.371 | 1.60E-02 | 0.041142 | 0.76 | 0.62 | 0.69 |
| ADHFE1-cg01588438 | 8 | 67344553 | 0.275 | 0.196 | 1.00E-01 | 0.18522 | 0.62 | 0.75 | 0.63 |
| ADHFE1-cg09383816 | 8 | 67344556 | 0.237 | 0.166 | 4.20E-02 | 0.089809 | 0.83 | 0.5 | 0.66 |
| ADHFE1-cg18065361 | 8 | 67344588 | 0.186 | 0.139 | 2.10E-01 | 0.346723 | 0.73 | 0.5 | 0.6 |
| ADHFE1-cg08090772 | 8 | 67344640 | 0.257 | 0.232 | 5.00E-01 | 0.575182 | 0.58 | 0.62 | 0.55 |
| ADHFE1-cg19283840 | 8 | 67344642 | 0.212 | 0.156 | 2.20E-01 | 0.346788 | 0.54 | 0.75 | 0.6 |
| ADHFE1-cg20295442 | 8 | 67344665 | 0.25 | 0.165 | 1.30E-01 | 0.228883 | 0.85 | 0.44 | 0.62 |
| ADHFE1-cg20912169 | 8 | 67344720 | 0.25 | 0.16 | 2.60E-01 | 0.387955 | 0.5 | 0.75 | 0.59 |
| ADHFE1-cg01988129 | 8 | 67344936 | 0.323 | 0.221 | 6.50E-02 | 0.125608 | 0.58 | 0.75 | 0.65 |
| ADHFE1-cg25046651 | 8 | 67345006 | 0.281 | 0.185 | 1.10E-02 | 0.030792 | 0.89 | 0.5 | 0.7 |
| ADHFE1-cg11530289 | 8 | 67350852 | 0.474 | 0.438 | 6.40E-01 | 0.696833 | 0.46 | 0.88 | 0.54 |
| ADHFE1-cg19366217 | 8 | 67351311 | 0.014 | 0.014 | 8.00E-01 | 0.836361 | 0.6 | 0.56 | 0.52 |
| ADHFE1-cg21779231 | 8 | 67351597 | 0.1 | 0.088 | 1.30E-01 | 0.228883 | 0.54 | 0.81 | 0.62 |
| ADHFE1-cg22628608 | 8 | 67352379 | 0.767 | 0.799 | 4.80E-02 | 0.09655 | 0.8 | 0.62 | 0.66 |
| ADHFE1-cg20040765 | 8 | 67358546 | 0.147 | 0.414 | 9.00E-07 | 0.00003 | 0.85 | 0.81 | 0.89 |
| ADHFE1-cg20794855 | 8 | 67365340 | 0.756 | 0.793 | 2.50E-02 | 0.057234 | 0.68 | 0.69 | 0.68 |
| ADHFE1-cg20180050 | 8 | 67371088 | 0.6 | 0.777 | 5.90E-04 | 0.003243 | 0.9 | 0.62 | 0.77 |
| ADHFE1-cg10895168 | 8 | 67373171 | 0.884 | 0.915 | 4.60E-02 | 0.094239 | 0.67 | 0.75 | 0.66 |
| ADHFE1-cg05042439 | 8 | 67380613 | 0.846 | 0.879 | 1.20E-02 | 0.030818 | 0.69 | 0.75 | 0.7 |
| TFPI2-cg18024479 | 7 | 93519220 | 0.672 | 0.507 | 6.10E-04 | 0.003243 | 0.86 | 0.75 | 0.77 |
| TFPI2-cg12973591 | 7 | 93519473 | 0.293 | 0.165 | 2.90E-01 | 0.401059 | 0.6 | 0.81 | 0.58 |
| TFPI2-cg22799321 | 7 | 93519621 | 0.28 | 0.212 | 2.70E-01 | 0.394139 | 0.57 | 0.75 | 0.59 |
| TFPI2-cg20230721 | 7 | 93519855 | 0.257 | 0.162 | 5.00E-01 | 0.575182 | 0.45 | 0.81 | 0.55 |
| TFPI2-cg23141855 | 7 | 93519892 | 0.222 | 0.198 | 5.30E-01 | 0.603311 | 0.5 | 0.81 | 0.55 |
| TFPI2-cg23686014 | 7 | 93519924 | 0.199 | 0.177 | 8.80E-01 | 0.911196 | 0.44 | 0.81 | 0.49 |
| TFPI2-cg24531255 | 7 | 93520012 | 0.148 | 0.142 | 3.80E-01 | 0.504511 | 0.38 | 0.88 | 0.57 |
| TFPI2-cg17338208 | 7 | 93520024 | 0.145 | 0.142 | 2.60E-01 | 0.387955 | 0.33 | 0.94 | 0.59 |
| TFPI2-cg26739865 | 7 | 93520172 | 0.189 | 0.196 | 3.90E-01 | 0.504511 | 0.39 | 0.88 | 0.57 |
| TFPI2-cg22441533 | 7 | 93520175 | 0.086 | 0.098 | 9.30E-01 | 0.938115 | 0.56 | 0.69 | 0.51 |
| TFPI2-cg14377593 | 7 | 93520183 | 0.108 | 0.139 | 3.50E-02 | 0.076386 | 0.63 | 0.75 | 0.67 |
| TFPI2-cg15649801 | 7 | 93520269 | 0.117 | 0.18 | 1.70E-02 | 0.042249 | 0.67 | 0.75 | 0.69 |
| TFPI2-cg10539069 | 7 | 93520275 | 0.124 | 0.182 | 4.70E-03 | 0.014906 | 0.65 | 0.88 | 0.72 |
| TFPI2-cg24084681 | 7 | 93520288 | 0.126 | 0.167 | 2.30E-02 | 0.05442 | 0.46 | 0.94 | 0.68 |
| TFPI2-cg07380959 | 7 | 93520323 | 0.141 | 0.201 | 3.50E-03 | 0.01254 | 0.67 | 0.88 | 0.73 |
| TFPI2-cg03333330 | 7 | 93520346 | 0.191 | 0.268 | 2.80E-03 | 0.011449 | 0.71 | 0.81 | 0.74 |
| TFPI2-cg09558850 | 7 | 93520445 | 0.571 | 0.347 | 1.10E-03 | 0.004822 | 0.58 | 0.88 | 0.76 |
| TFPI2-cg19854521 | 7 | 93520452 | 0.544 | 0.336 | 9.80E-04 | 0.004726 | 0.62 | 0.88 | 0.76 |

**Supplementary Table 2 The methylation status of the 8 selected CpG sites in the GEO dataset and normal PBMC and PBL dataset**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **GSE52826 (N =12)** | | | | | | | **PBMC (N=111)** | | **PBL**  **(N=527)** |
| **CpGsites** | **McaMa** | **McoMb** | **P valuec** | **Sensd** | **Specd** | **AUCd** | | **MnoMe** | **MnoMe** | |
| cg08090772 | 0.1785 | 0.152 | 0.85 | 0.67 | 0.88 | 0.58 | | 0.12 | 0.15 | |
| cg20295442 | 0.178 | 0.0642 | 0.215 | 1 | 0.5 | 0.78 | | 0.05 | 0.08 | |
| cg20912169 | 0.1781 | 0.0464 | 0.215 | 0.75 | 0.88 | 0.78 | | 0.14 | 0.14 | |
| cg16971668 | 0.4532 | 0.2237 | 0.178 | 0.75 | 1 | 0.81 | | 0.06 | 0.06 | |
| cg22383888 | 0.4575 | 0.1415 | 0.024 | 1 | 1 | 1 | | 0.13 | 0.172 | |
| cg04550052 | 0.564 | 0.1679 | 0.024 | 1 | 1 | 1 | | 0.05 | 0.06 | |
| cg04698114 | 0.5364 | 0.1321 | 0.024 | 1 | 1 | 1 | | 0.03 | 0.05 | |
| cg12973591 | 0.5591 | 0.0581 | 0.024 | 1 | 1 | 1 | | 0.06 | 0.07 | |

McaMa represent the mean methylation percentage of the ESCC samples in the GSE52826 dataset, and the McoMb represent the mean methylation percentage of the control samples. The Pvaluec is calculated through the Wilcoxon rank-sum test after multiple test correction with FDR. Sensd is short for sensitivity, and Specd is short for specificity, AUCd is short for the area under curve, all of which were derived with the logistic regression prediction model without adjustment for gender, age and smoking status and alcohol status. MnoMe represent the mean methylation percentage of the normal samples. The methylation data of the PBMC and PBL of the healthy normal samples were obtained from the public dataset from the GEO database.

**Supplementary Table 3 The methylation status of the 4 genomic regions in the Young/Old subgroups**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Subgroup** | **Gene** | **Genomic** | **No.** | **McaMd** | **McoMd** | **Pvaluee** | **Sensf** | **Specf** | **AUCf** |
| **Regionb** | **CpGsitesc** |
| Younga  (N =45) | ADHFE1 | chr8:67344610-67344805 | 24 | 0.24 | 0.15 | 7.51×10-3 | 0.41 | 0.94 | 0.68 |
| EOMES | chr3:27764697-27764940 | 8 | 0.37 | 0.24 | 2.70×10-4 | 0.51 | 0.97 | 0.75 |
| SALL1 | chr16:51184268-51184468 | 18 | 0.35 | 0.17 | 1.00×10-4 | 0.51 | 0.94 | 0.77 |
| TFPI2 | chr7:93519367-93519503 | 13 | 0.24 | 0.13 | 6.79×10-3 | 0.54 | 0.94 | 0.69 |
|  | Combined | - | - | - | - | - | **0.74** | **0.80** | **0.82** |
| Olda  (N = 49) | ADHFE1 | chr8:67344610-67344805 | 24 | 0.23 | 0.15 | 6.54×10-2 | 0.26 | 0.95 | 0.61 |
| EOMES | chr3:27764697-27764940 | 8 | 0.38 | 0.23 | 6.00×10-7 | 0.81 | 0.70 | 0.80 |
| SALL1 | chr16:51184268-51184468 | 18 | 0.39 | 0.20 | 3.02×10-4 | 0.49 | 0.93 | 0.72 |
| TFPI2 | chr7:93519367-93519503 | 13 | 0.32 | 0.14 | 8.87×10-5 | 0.43 | 1.00 | 0.74 |
|  | Combined | - | - | - | - | - | **0.68** | **0.88** | **0.80** |

aThe median ages of the patients were utilized as the criteria for dividing samples into the young and old groups. bGenomic region represents the genomic coverage of the reads with targeted bisulfite sequencing, and the genomic coordinates shown here were based on the hg19 version of the genome. cNo.CpGsites represents the number of the CpGsites in each region. dMcaM represents the mean methylation percentage of the cases in each region, which consists of several CpGsites, while McoM represents the mean methylation percentage of the controls in each region. eP value is calculated through the Wilcoxon rank-sum test following with FDR (false discovery rate) adjustment for multiple correction. fSens is short for sensitivty, while Spec is short for specificity, and the AUC is short for area under curve. The sensitivity, specificity as well as the AUC were both with a logistic regression prediction model without adjustment for gender, age and smoking status and alcohol status. The mean methylation percentage of each in each genomic region was considered as a variable.

**Supplementary Table 4 The methylation status of the 4 genomic regions in the Male/Female subgroups**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Subgroup** | **Gene** | **Genomic** | **No.** | **McaMc** | **McoMc** | **Pvalued** | **Sense** | **Spece** | **AUCe** |
| **Regiona** | **CpGsitesb** |
| Male  (N =69) | ADHFE1 | chr8:67344610-67344805 | 24 | 0.24 | 0.16 | 4.54×10-3 | 0.56 | 0.69 | 0.65 |
| EOMES | chr3:27764697-27764940 | 8 | 0.37 | 0.24 | 2.40×10-6 | 0.65 | 0.76 | 0.76 |
| SALL1 | chr16:51184268-51184468 | 18 | 0.36 | 0.18 | 3.13×10-5 | 0.48 | 0.95 | 0.73 |
| TFPI2 | chr7:93519367-93519503 | 13 | 0.28 | 0.13 | 5.72×10-5 | 0.54 | 0.95 | 0.72 |
|  | Combined | - | - | - | - | - | **0.68** | **0.86** | **0.79** |
| Female  (N = 25) | ADHFE1 | chr8:67344610-67344805 | 24 | 0.24 | 0.13 | 1.88×10-1 | 0.30 | 1.00 | 0.62 |
| EOMES | chr3:27764697-27764940 | 8 | 0.41 | 0.23 | 1.20×10-3 | 0.65 | 0.95 | 0.83 |
| SALL1 | chr16:51184268-51184468 | 18 | 0.39 | 0.18 | 7.10×10-3 | 0.61 | 0.95 | 0.76 |
| TFPI2 | chr7:93519367-93519503 | 13 | 0.28 | 0.13 | 1.43×10-2 | 0.78 | 0.58 | 0.73 |
|  | Combined | - | - | - | - | - | **0.70** | **0.95** | **0.82** |

aGenomic region represents the genomic coverage of the reads with targeted bisulfite sequencing, and the genomic coordinates shown here were based on the hg19 version of the genome. bNo.CpGsites represents the number of the CpGsites in each region. cMcaM represents the mean methylation percentage of the cases in each region, which consists of several CpGsites, while McoM represents the mean methylation percentage of the controls in each region. dP value is calculated through the Wilcoxon rank-sum test following with FDR (false discovery rate) adjustment for multiple correction. eSens is short for sensitivty, while Spec is short for specificity, and the AUC is short for area under curve. The sensitivity, specificity as well as the AUC were both with a logistic regression prediction model without adjustment for gender, age and smoking status and alcohol status. The mean methylation percentage of each in each genomic region was considered as a variable.

**Supplementary Table 5 The methylation status of the 4 genomic regions in the Smokers/Non-smokers subgroups**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Subgroup** | **Gene** | **Genomic** | **No.** | **McaMd** | **McoMd** | **P valuee** | **Sensf** | **Specf** | **AUCf** |
| **Regionb** | **CpGsitesc** |
| Smokeda  (N =58) | ADHFE1 | chr8:67344610-67344805 | 24 | 0.23 | 0.16 | 1.61×10-2 | 0.60 | 0.64 | 0.64 |
| EOMES | chr3:27764697-27764940 | 8 | 0.38 | 0.23 | 7.80×10-6 | 0.68 | 0.79 | 0.78 |
| SALL1 | chr16:51184268-51184468 | 18 | 0.37 | 0.18 | 3.44×10-4 | 0.51 | 0.96 | 0.72 |
| TFPI2 | chr7:93519367-93519503 | 13 | 0.30 | 0.14 | 8.39×10-4 | 0.57 | 0.94 | 0.70 |
| Combined | - | - | - | - | - | **0.64** | **0.94** | **0.80** |
| Non-smokeda  (N = 36) | ADHFE1 | chr8:67344610-67344805 | 24 | 0.26 | 0.13 | 6.71×10-2 | 0.30 | 1.00 | 0.63 |
| EOMES | chr3:27764697-27764940 | 8 | 0.38 | 0.24 | 3.50×10-4 | 0.61 | 0.87 | 0.79 |
| SALL1 | chr16:51184268-51184468 | 18 | 0.36 | 0.18 | 7.30×10-4 | 0.55 | 0.87 | 0.75 |
| TFPI2 | chr7:93519367-93519503 | 13 | 0.25 | 0.13 | 7.30×10-4 | 0.85 | 0.57 | 0.75 |
| Combined | - | - | - | - | - | **0.70** | **0.80** | **0.81** |

aThe smoker subgroup included the former smokers as well as current smokers. bGenomic region represents the genomic coverage of the reads with targeted bisulfite sequencing, and the genomic coordinates shown here were based on the hg19 version of the genome. cNo.CpGsites represents the number of the CpGsites in each region. dMcaM represents the mean methylation percentage of the cases in each region, which consists of several CpGsites, while McoM represents the mean methylation percentage of the controls in each region. eP value is calculated through the Wilcoxon rank-sum test following with FDR (false discovery rate) adjustment for multiple correction. fSens is short for sensitivty, while Spec is short for specificity, and the AUC is short for area under curve. The sensitivity, specificity as well as the AUC were both with a logistic regression prediction model without adjustment for gender, age and smoking status and alcohol status. The mean methylation percentage of each in each genomic region was considered as a variable.

**Supplementary Table 6 The methylation status of the 4 genomic regions in the Alcohol/ Non-alcohol subgroups**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Subgroup** | **Gene** | **Genomic** | **No.** | **McaMd** | **McoMd** | **P valuee** | **Sensf** | **Specf** | **AUCf** |
| **Regionb** | **CpGsitesc** |
| Alcohola  (N =34) | ADHFE1 | chr8:67344610-67344805 | 19 | 0.23 | 0.17 | 2.24×10-1 | 0.52 | 0.70 | 0.59 |
| EOMES | chr3:27764697-27764940 | 6 | 0.36 | 0.24 | 6.70×10-3 | 0.45 | 0.93 | 0.74 |
| SALL1 | chr16:51184268-51184468 | 17 | 0.35 | 0.20 | 2.75×10-2 | 0.45 | 0.96 | 0.68 |
| TFPI2 | chr7:93519367-93519503 | 19 | 0.30 | 0.14 | 1.30×10-2 | 0.55 | 0.93 | 0.71 |
| Combined | - | - | - | - | - | **0.74** | **0.78** | **0.77** |
| Non-alcohola  (N = 58) | ADHFE1 | chr8:67344610-67344805 | 19 | 0.24 | 0.13 | 1.70×10-3 | 0.30 | 1.00 | 0.68 |
| EOMES | chr3:27764697-27764940 | 6 | 0.38 | 0.22 | 2.10×10-7 | 0.66 | 0.88 | 0.81 |
| SALL1 | chr16:51184268-51184468 | 17 | 0.37 | 0.17 | 3.00×10-6 | 0.57 | 0.94 | 0.78 |
| TFPI2 | chr7:93519367-93519503 | 19 | 0.27 | 0.13 | 2.90×10-4 | 0.55 | 0.86 | 0.71 |
| Combined | - | - | - | - | - | **0.66** | **0.96** | **0.84** |

aThe Alcohol subgroup included the samples which had alcohol intake currently and formerly. bGenomic region represents the genomic coverage of the reads with targeted bisulfite sequencing, and the genomic coordinates shown here were based on the hg19 version of the genome. cNo.CpGsites represents the number of the CpGsites in each region. dMcaM represents the mean methylation percentage of the cases in each region, which consists of several CpGsites, while McoM represents the mean methylation percentage of the controls in each region. eP value is calculated through the Wilcoxon rank-sum test following with FDR (false discovery rate) adjustment for multiple correction. fSens is short for sensitivty, while Spec is short for specificity, and the AUC is short for area under curve. The sensitivity, specificity as well as the AUC were both with a logistic regression prediction model without adjustment for gender, age and smoking status and alcohol status. The mean methylation percentage of each in each genomic region was considered as a variable.

**Supplementary Table 7 The designed primers of the five genomic regions for targeted bisulfite sequencing**

|  |  |
| --- | --- |
| **Primer Name** | **Sequence** |
| ADHFE1\_F | GTTTTGAGTTYGATTGGTTTGAGG |
| ADHFE1\_R | CTACRCRTTACAATTACCTCAACAAATAC |
| EOMES\_F | AGTTGTGTTGGYGTGAGTATGAAG |
| EOMES\_R | TTCAAAACACACCTTCCTCTTATCRAAAACA |
| SALL1\_F | GTAAGAAGATGGGGATTGGTGT |
| SALL1\_R | CTTCCCTAACCCCCCTAAAA |
| TFPI2\_F | GAGGTTTGTTTAATATTTGAGAAAATTTAGG |
| TFPI2\_R | AACAACRCCAACAATTTCTACACCTAA |

**Supplementary Figure 1 PCA analysis of the ESCC and EAC adjacent normal tissues.**

**Supplementary Figure 2 PCA analysis for the ESCC and adjacent normal tissues in the validation dataset**

**Supplementary Figure 3 The ROC (Receiver Operating characteristics) curve for the subgroup analyzes**

A-H represent the ROC curve for the prediction model in different subgroups, respectively. A-H each represent the overall ROC curve for the subgroup, which was calculated through a logistic regression model, incorporating the mean methylation percentage of the five genomic regions as the variables, and without the adjustment for gender, age and smoking status and alcohol status.

**Supplementary Figure 4 The expression profiles for the three genes using RNA-seq data from TCGA**