# DNA Methylation

Shicheng Guo

Abstract

DNA methylation is the most important and stable epigenetic biomarkers. However, in order to develop different methylation biomarkers, we need to have deep understanding to the methylation profiles for different tissues and disease status. In this book, I introduced all the knowledge of the DNA methylation biomarkers.

How to understand the genome-wide hypo-methylation in human cancers while all of 3 DNMT genes are over-expressed in majority cancer types? What’s the expression status who interact with DNMTs to silent human repeat regions? large number of methylation abnormal didn’t transfer to gene expression (such as FSTL1).

# DNA Methylation Research for Tumor Suppressor Genes

It will be interesting to know how many TSGs are hyper-methylation and low-expressed in cancer samples. I collected TSGs with two methods including COSMIC.

# Blood Un-methylation Regions (BUR) and Tissue Methylation Regions (TMR)

In order to obtain some interesting tissue hyper-methylation genes/regions, I tried to collected all the tissue-specific hyper-methylated and blood high expression genes. I collected all the TCGA non-blood samples include tumor and normal tissues and identified all the hyper-methylated CpGs (Q25>0.6) and blood high expressing genes (TPM>1.0 in whole-blood samples from GTEx project). Finally, I identified 21 genes (Table xx).

GSC is hypo-methylated in normal blood cells (PBMC, B-cell, Neu, CD4+) while it is hyper-methylated in solid cancer and AML. It looks GSC can be a pan-cancer methylation biomarker, however, it isn’t hyper-methylated in B-ALL cancers. Most interesting genomic region for GSC is chr14:95234459-95234854 (hg19).

# DNA methylation Biomarker Research in ESCC

Table 1. Significant differential expressing genes in ESCA and CHOL DNA methylation project

|  |  |  |  |
| --- | --- | --- | --- |
| ID |  | beta | P |
| 1 | TMEM132C | -4.721181769 | 1.61E-27 |
| 2 | PEG3 | -2.579983213 | 3.92E-20 |
| 3 | USP44 | -1.72167968 | 9.39E-15 |
| 4 | ADHFE1 | -1.666274027 | 9.91E-15 |
| 5 | GFRA1 | -2.707357257 | 9.31E-12 |
| 6 | CDO1 | -2.952271252 | 3.57E-11 |
| 7 | RNLS | -0.676699455 | 1.72E-10 |
| 8 | ITGA8 | -1.769931775 | 3.46E-10 |
| 9 | RUNDC3B | -1.710059176 | 5.56E-10 |
| 10 | AKAP12 | -1.57760242 | 1.61E-09 |
| 11 | SOX11 | 3.343099328 | 3.20E-09 |

The beta and P-values are derived from meta-analysis to Pan-cancer RNA-Seq data. These genes will be also considered in lung and breast cancer project so that finally I can prepare a Pan-cancer analysis with review or mini-review.

chr1:50889198-50889740

# Pan-Cancer Methylation Biomarker (PCMB)

Table xx. Genes hyper-methylated in solid tissue and high expressed in whole-blood

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | MethQ0 | MethQ25 | TPM­\_Blood | B/T Ratio | DMRs in GRCH37 |
| SPI1 | 0.093967 | 0.732906 | 908 | 52.78974 | chr11:47,399,777-47,400,223 |
| S1PR4 | 0.052019 | 0.643889 | 260 | 52.48239 | chr19:3,178,735-3,179,184 |
| NFAM1 | 0.012068 | 0.603244 | 109.2 | 47.21494 | chr22:42,828,069-42,828,533 |
| OSM | 0.059035 | 0.602995 | 49.48 | 46.86289 | chr22:30,662,629-30,662,898 |
| C16orf54 | 0.03849 | 0.634561 | 63.03 | 38.90642 | chr16:29,757,144-29,757,610 |
| BIN2 | 0.038053 | 0.692999 | 235.2 | 37.46268 | chr12:51,717,574-51,718,265 |
| C19orf35 | 0.094108 | 0.627955 | 20.44 | 33.34713 | chr19:2,282,044-2,282,537 |
| ITGB2 | 0.053084 | 0.630435 | 572.7 | 30.19166 | chr21:46,340,250-46,340,802 |
| ICAM3 | 0.121451 | 0.642406 | 403.5 | 28.40834 | chr19:10,449,867-10,450,324 |
| PSTPIP1 | 0.090827 | 0.636502 | 103.6 | 27.15204 | chr15:77,287,327-77,287,771 |
| CD7 | 0.07259 | 0.672038 | 67.17 | 20.91915 | chr17:80,273,516-80,273,968 |
| TBC1D10C | 0.079223 | 0.63471 | 185 | 20.19358 | chr11:67,171,580-67,172,036 |
| GPSM3 | 0.080427 | 0.7026 | 831.1 | 32.28412 | Not Good Marker |
| WAS | 0.066367 | 0.66933 | 329.8 | 31.08416 | Not Good Marker |
| PRF1 | 0.146939 | 0.621175 | 138.9 | 30.98476 | Not Good Marker |
| CFP | 0.022835 | 0.60196 | 244.4 | 44.50527 | Not Good Marker |
| FGR | 0.069841 | 0.745352 | 597.3 | 37.78979 | Not Good Marker |
| IL2RG | 0.045546 | 0.682142 | 422.4 | 24.98398 | Not Good Marker |
| CYTIP | 0.088117 | 0.717905 | 192.4 | 22.86334 | Not Good Marker |
| TAGAP | 0.078277 | 0.780386 | 49.66 | 22.53926 | Not Good Marker |
| PTPRC | 0.043059 | 0.66827 | 229.8 | 22.04525 | Not Good Marker |
| SFRP1 |  |  |  |  |  |
| CDO1 |  |  |  |  |  |
| OPCML |  |  |  |  |  |
| DCLK1 |  |  |  |  |  |
| NEUROG1 |  |  |  |  |  |
| HOXA1 |  |  |  |  |  |

These genes will be also considered in lung and breast cancer project so that finally, I can prepare a paper to introduce solid-tissue or solid tumor biomarker and which are hypo-methylated in blood cells (UMR)

# DNA methylation and Human Cholangiocarcinoma

# DNA Methylation and Human Breast Cancer

In order to identify some novel methylation biomarker for breast cancer. We initialed this project in 2019 with a two stage biomarker validation study design. I selected xx DMR regions in April 1, 2019

1. Frequent mutated genes (50) -> Low-expression in cancer(20) -> hyper-methylation (10) -> BUR (5)
2. Tumor-suppressor genes () -> Low-expression in cancer(20) -> hyper-methylation (10) -> BUR (5)
3. GWAS in BRCA () -> Low-expression in cancer(20) -> hyper-methylation (10) -> BUR (5)
4. Previous published methylation biomarkers to increase totally citation and meta-analysis probability

# DNA Methylation and Stomach Cancer

Compared with other human cancers, patients with gastrointestinal usually have lower proportion who are benefit to immunotherapy, except microsatellite instable (MSI) patients. Approximately 15% of gastric and stomach adenocarcinoma belong to MSI subtype[1].

# DNA methylation Biomarker in Lung Adenocarcinoma

In order to identify some novel methylation biomarker for breast cancer. We initialed this project in 2019 with a two stage biomarker validation study design. I selected xx DMR regions in April 1, 2019

# DNA methylation Biomarker in Lung Squamous Cell Carcinoma

In order to identify some novel methylation biomarker for breast cancer. We initialed this project in 2019 with a two stage biomarker validation study design. I selected xx DMR regions in April 1, 2019

# DNA methylation Biomarker in Lung Squamous Cell Carcinoma

In order to identify some novel methylation biomarker for breast cancer. We initialed this project in 2019 with a two stage biomarker validation study design. I selected xx DMR regions in April 1, 2019

# DNA methylation Biomarker in Hepatocellular carcinoma (HCC)

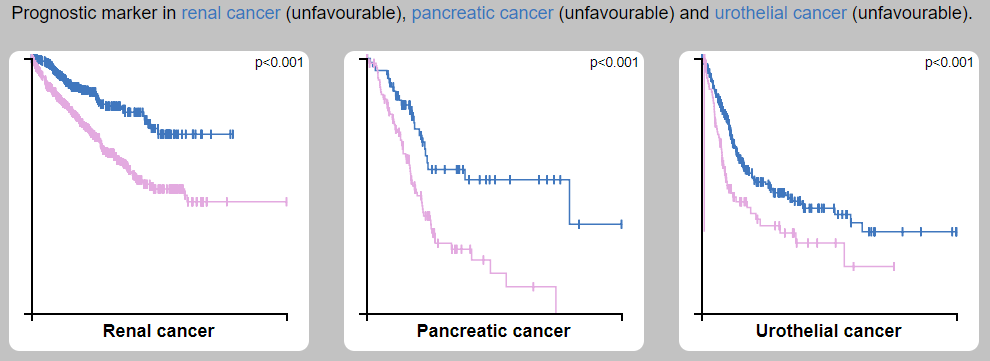
In order to identify some novel methylation biomarker for breast cancer. We initialed this project in 2019 with a two stage biomarker validation study design. I selected xx DMR regions in April 1, 2019

# Significant high or low expression ncRNAs in pan-cancer samples

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | idx | beta | pval | cilb | ciub | i2 | tau2 |
| ENSG00000243384 | 4446 | -5.59032 | 3.76E-26 | -6.62612 | -4.55452 | 94.30997 | 5.487714 |
| ENSG00000267328 | 11433 | -5.12158 | 1.49E-13 | -6.48033 | -3.76283 | 96.97247 | 9.796229 |
| ENSG00000241720 | 11527 | -4.59129 | 2.78E-14 | -5.77407 | -3.40851 | 97.68349 | 7.762307 |
| ENSG00000231246 | 8656 | -4.46878 | 9.35E-20 | -5.43167 | -3.50589 | 97.09272 | 4.783646 |
| ENSG00000258604 | 14177 | -4.24114 | 2.31E-15 | -5.29022 | -3.19206 | 97.17715 | 6.001081 |
| ENSG00000224239 | 12083 | -3.43468 | 1.54E-15 | -4.27896 | -2.59041 | 88.6219 | 3.19598 |
| ENSG00000251226 | 18324 | -3.41835 | 4.40E-15 | -4.27261 | -2.5641 | 93.24342 | 3.684191 |
| ENSG00000275088 | 26927 | -2.69524 | 2.23E-16 | -3.33873 | -2.05174 | 95.65713 | 2.220855 |
| ENSG00000254319 | 14417 | -2.49087 | 7.08E-19 | -3.04104 | -1.9407 | 91.46546 | 1.465469 |
| ENSG00000273682 | 23803 | -2.29412 | 1.26E-19 | -2.79021 | -1.79803 | 86.91031 | 1.081182 |
| ENSG00000262061 | 8524 | -2.17824 | 4.78E-17 | -2.68698 | -1.66951 | 97.20843 | 1.424961 |
| ENSG00000257283 | 12824 | -1.78054 | 1.52E-14 | -2.23459 | -1.32649 | 92.04315 | 0.995086 |
| ENSG00000256139 | 17027 | -1.76796 | 1.34E-14 | -2.21789 | -1.31803 | 95.7985 | 1.07907 |
| ENSG00000274370 | 22769 | -1.66868 | 7.43E-20 | -2.02725 | -1.31012 | 95.78092 | 0.669986 |
| ENSG00000213754 | 21961 | -1.63532 | 1.95E-19 | -1.99081 | -1.27983 | 94.94271 | 0.673716 |
| ENSG00000249456 | 12095 | -1.27253 | 2.59E-14 | -1.59996 | -0.94511 | 92.71392 | 0.494312 |
| ENSG00000266601 | 4020 | -1.19501 | 3.79E-23 | -1.43137 | -0.95865 | 92.59149 | 0.269393 |
| ENSG00000280054 | 18392 | -1.00387 | 9.00E-24 | -1.19961 | -0.80814 | 89.51505 | 0.175499 |
| ENSG00000279982 | 9753 | -0.89974 | 1.59E-14 | -1.12936 | -0.67012 | 87.83427 | 0.23865 |
| ENSG00000230555 | 7357 | -0.82476 | 7.66E-21 | -0.99739 | -0.65214 | 88.64084 | 0.13327 |
| ENSG00000180015 | 7344 | -0.8036 | 2.95E-16 | -0.99625 | -0.61094 | 47.80452 | 0.081522 |
| ENSG00000274220 | 16917 | -0.75817 | 1.92E-14 | -0.95227 | -0.56408 | 90.01046 | 0.170717 |
| ENSG00000228452 | 6043 | -0.60381 | 7.52E-18 | -0.74132 | -0.46631 | 92.40947 | 0.094891 |
| ENSG00000262712 | 27294 | 0.730751 | 1.44E-14 | 0.544568 | 0.916935 | 88.54463 | 0.162474 |
| ENSG00000270804 | 9761 | 0.778765 | 1.48E-16 | 0.593929 | 0.9636 | 93.01978 | 0.173717 |
| ENSG00000269399 | 1428 | 0.830658 | 2.87E-22 | 0.662903 | 0.998414 | 86.85776 | 0.123938 |
| ENSG00000257086 | 30139 | 0.830893 | 1.54E-20 | 0.655594 | 1.006191 | 91.1971 | 0.152908 |
| ENSG00000272455 | 26350 | 0.959423 | 9.76E-18 | 0.740175 | 1.178671 | 93.6502 | 0.248541 |
| ENSG00000270195 | 10894 | 1.485524 | 5.19E-18 | 1.148888 | 1.822159 | 87.3414 | 0.497432 |
| ENSG00000255921 | 16543 | 2.991146 | 1.94E-15 | 2.253243 | 3.729049 | 84.66538 | 2.247614 |
| ENSG00000249001 | 28435 | 4.682048 | 1.03E-14 | 3.495808 | 5.868289 | 94.00953 | 7.225257 |

When I collect Pan-cancer RNA-seq dataset, I found some of ncRNA were significantly over-expressed or low-expressed in cancer samples. I am quite interested in these ncRNAs. Here, I summarized the top 31 ncRNAs and hope later I can make some functional study to these ncRNAs in multiple cancers. What’s the status for tumor suppressor genes in meta-analysis will be quite interesting.

IL1RAP over-expression as risk of overall survival time.



1. Bonneville, R., et al., *Landscape of Microsatellite Instability Across 39 Cancer Types.* JCO Precis Oncol, 2017. **2017**.