# DNA Methylation Research for Human Diseases

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 Gender Specific DNA methylation Regions

# 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).

PCDH10 chr2:177,012,372-177,012,675

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 |

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 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

Baarrrerra et 2008 and Schug et al., 2005 believe tissue-specific genes usually contains less CpGs or low CpGs in the promoter region while house-keeping genes contains CpG island.

**Variable methylated regions**

**Partially methylated domain**

# DNA methylation Biomarker Research in CHOL

# DNA methylation Biomarker Research in BRCA

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

|  |  |  |  |
| --- | --- | --- | --- |
| ID |  | beta | P |
|  | IRF4 | chr6:391,178-391,637 | Hyper5me+low mRNA+ better outcome |
|  | C2orf40 | chr2:106,681,823-106,682,276 | Hyper5me+low mRNA+ better outcome |
|  |  |  |  |

# DNA methylation Biomarker Research in Human Non-Small Cell Lung Cancer (NSCLC)

Adenocarcinoma and 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

1. TSG1(PCDH8,ZIC1,PCDH17,TBX5,IRX1)

2. GWAS(TBX15,PITX2,SIX2,LMX1A,FOXP4)

3. VIP(PCDH8,SATB2,TBX15,HOXD9,HOXD12,SATB2)

# DNA methylation Biomarker Research in Hepatocellular carcinoma (HCC)

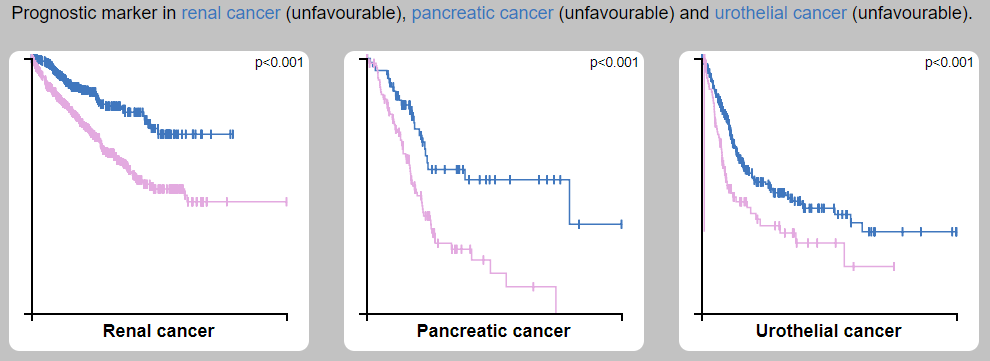
In order to identify some novel methylation biomarker for breast cancer. We initialed this project in 2019 with a two stage biomarker validation study design. I selected xx DMR regions in April 1, 2019

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

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | idx | beta | pval | cilb | ciub | i2 | tau2 | Len |
| ENSG00000243384 | 4446 | -5.59032 | 3.76E-26 | -6.62612 | -4.55452 | 94.30997 | 5.487714 | 3109 |
| ENSG00000267328 | 11433 | -5.12158 | 1.49E-13 | -6.48033 | -3.76283 | 96.97247 | 9.796229 | 660 |
| ENSG00000241720 | 11527 | -4.59129 | 2.78E-14 | -5.77407 | -3.40851 | 97.68349 | 7.762307 | 808 |
| ENSG00000231246 | 8656 | -4.46878 | 9.35E-20 | -5.43167 | -3.50589 | 97.09272 | 4.783646 | 547 |
| ENSG00000258604 | 14177 | -4.24114 | 2.31E-15 | -5.29022 | -3.19206 | 97.17715 | 6.001081 | 478 |
| ENSG00000224239 | 12083 | -3.43468 | 1.54E-15 | -4.27896 | -2.59041 | 88.6219 | 3.19598 | 497 |
| ENSG00000251226 | 18324 | -3.41835 | 4.40E-15 | -4.27261 | -2.5641 | 93.24342 | 3.684191 | 3684 |
| ENSG00000275088 | 26927 | -2.69524 | 2.23E-16 | -3.33873 | -2.05174 | 95.65713 | 2.220855 | 3306 |
| ENSG00000254319 | 14417 | -2.49087 | 7.08E-19 | -3.04104 | -1.9407 | 91.46546 | 1.465469 | 582 |
| ENSG00000273682 | 23803 | -2.29412 | 1.26E-19 | -2.79021 | -1.79803 | 86.91031 | 1.081182 | 129 |
| ENSG00000262061 | 8524 | -2.17824 | 4.78E-17 | -2.68698 | -1.66951 | 97.20843 | 1.424961 | 1013 |
| ENSG00000257283 | 12824 | -1.78054 | 1.52E-14 | -2.23459 | -1.32649 | 92.04315 | 0.995086 | 540 |
| ENSG00000256139 | 17027 | -1.76796 | 1.34E-14 | -2.21789 | -1.31803 | 95.7985 | 1.07907 | 910 |
| ENSG00000274370 | 22769 | -1.66868 | 7.43E-20 | -2.02725 | -1.31012 | 95.78092 | 0.669986 | 611 |
| ENSG00000213754 | 21961 | -1.63532 | 1.95E-19 | -1.99081 | -1.27983 | 94.94271 | 0.673716 | 1436 |
| ENSG00000249456 | 12095 | -1.27253 | 2.59E-14 | -1.59996 | -0.94511 | 92.71392 | 0.494312 | 555 |
| ENSG00000266601 | 4020 | -1.19501 | 3.79E-23 | -1.43137 | -0.95865 | 92.59149 | 0.269393 | 335 |
| ENSG00000280054 | 18392 | -1.00387 | 9.00E-24 | -1.19961 | -0.80814 | 89.51505 | 0.175499 | 2248 |
| ENSG00000279982 | 9753 | -0.89974 | 1.59E-14 | -1.12936 | -0.67012 | 87.83427 | 0.23865 | 3109 |
| ENSG00000230555 | 7357 | -0.82476 | 7.66E-21 | -0.99739 | -0.65214 | 88.64084 | 0.13327 | 760 |
| ENSG00000180015 | 7344 | -0.8036 | 2.95E-16 | -0.99625 | -0.61094 | 47.80452 | 0.081522 | 1122 |
| ENSG00000274220 | 16917 | -0.75817 | 1.92E-14 | -0.95227 | -0.56408 | 90.01046 | 0.170717 | 2557 |
| ENSG00000228452 | 6043 | -0.60381 | 7.52E-18 | -0.74132 | -0.46631 | 92.40947 | 0.094891 | 484 |
| ENSG00000262712 | 27294 | 0.730751 | 1.44E-14 | 0.544568 | 0.916935 | 88.54463 | 0.162474 | 1949 |
| ENSG00000270804 | 9761 | 0.778765 | 1.48E-16 | 0.593929 | 0.9636 | 93.01978 | 0.173717 | 891 |
| ENSG00000269399 | 1428 | 0.830658 | 2.87E-22 | 0.662903 | 0.998414 | 86.85776 | 0.123938 | 2069 |
| ENSG00000257086 | 30139 | 0.830893 | 1.54E-20 | 0.655594 | 1.006191 | 91.1971 | 0.152908 | 853 |
| ENSG00000272455 | 26350 | 0.959423 | 9.76E-18 | 0.740175 | 1.178671 | 93.6502 | 0.248541 | 1523 |
| ENSG00000270195 | 10894 | 1.485524 | 5.19E-18 | 1.148888 | 1.822159 | 87.3414 | 0.497432 | 802 |
| ENSG00000255921 | 16543 | 2.991146 | 1.94E-15 | 2.253243 | 3.729049 | 84.66538 | 2.247614 | 728 |
| ENSG00000249001 | 28435 | 4.682048 | 1.03E-14 | 3.495808 | 5.868289 | 94.00953 | 7.225257 | 821 |

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.



DNA methylation and Hypoxia Signaling. I found 3 hypoxia related genes are hyper-methylated in cancers.

|  |  |  |  |
| --- | --- | --- | --- |
| EGLN3 | chr14:34,419,380-34,419,934 |  |  |
| [P4HTM](https://www.genecards.org/cgi-bin/carddisp.pl?gene=P4HTM) | chr3:49,027,655-49,028,040 |  |  |
| HIF3A | chr19:46,799,993-46,800,409 |  |  |
| BNC1 | chr15:83,952,926-83,953,338 | Joo, CCR, 2013 |  |
| ADAMTS1 | chr21:28,216,680-28,217,166 | Joo, CCR, 2013 |  |

# Hyper-methylation and low expression of transcript factors (TF) in human Pan-cancer dataset

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Symbol | CHR | START | END | CpG | C | N | 5me\_Pval | Q25\_PBMC | Diff\_Exp\_beta | Pval\_Exp |
| KLF4 | chr9 | 1.1E+08 | 1.1E+08 | cg14367995 | 0.283241 | 0.124233 | 1.49E-09 | 0.090364417 | -1.41641 | 2.04E-10 |
| MEF2C | chr5 | 88185768 | 88185769 | cg12621171 | 0.274873 | 0.102616 | 1.80E-09 | 0.076388643 | -1.10716 | 2.51E-06 |
| MEF2C | chr5 | 88185051 | 88185052 | cg04694437 | 0.330285 | 0.174203 | 2.13E-11 | 0.091856577 | -1.10716 | 2.51E-06 |
| MEIS2 | chr15 | 37387525 | 37387526 | cg23677243 | 0.52135 | 0.365146 | 1.04E-08 | 0.098629339 | -1.15831 | 1.66E-06 |
| TAL1 | chr1 | 47695138 | 47695139 | cg06087421 | 0.285292 | 0.080392 | 9.33E-17 | 0.064385391 | -1.34369 | 4.70E-08 |
| TAL1 | chr1 | 47695475 | 47695476 | cg19797376 | 0.50543 | 0.293174 | 8.32E-16 | 0.1582581 | -1.34369 | 4.70E-08 |
| TAL1 | chr1 | 47697429 | 47697430 | cg10961841 | 0.308821 | 0.155139 | 3.86E-12 | 0.069921369 | -1.34369 | 4.70E-08 |
| TAL1 | chr1 | 47694919 | 47694920 | cg10026317 | 0.641 | 0.407413 | 1.63E-13 | 0.174951645 | -1.34369 | 4.70E-08 |
| TAL1 | chr1 | 47695408 | 47695409 | cg06955484 | 0.525787 | 0.371714 | 2.47E-09 | 0.142368231 | -1.34369 | 4.70E-08 |
| TAL1 | chr1 | 47691550 | 47691551 | cg03885399 | 0.380572 | 0.198328 | 6.02E-11 | 0.07954334 | -1.34369 | 4.70E-08 |
| TAL1 | chr1 | 47697948 | 47697949 | cg25764899 | 0.288831 | 0.063993 | 1.34E-13 | 0.022343867 | -1.34369 | 4.70E-08 |
| TAL1 | chr1 | 47697965 | 47697966 | cg04363624 | 0.262785 | 0.082792 | 1.72E-15 | 0.030436096 | -1.34369 | 4.70E-08 |
| ZEB2 | chr2 | 1.45E+08 | 1.45E+08 | cg16405026 | 0.232626 | 0.044749 | 8.62E-08 | 0.030652337 | -1.17618 | 4.31E-08 |
| ZNF132 | chr19 | 58951885 | 58951886 | cg07878486 | 0.24323 | 0.074635 | 4.20E-08 | 0.04744 | -0.58494 | 9.58E-06 |
| ZNF132 | chr19 | 58952108 | 58952109 | cg11618529 | 0.415132 | 0.250965 | 2.74E-14 | 0.202779705 | -0.58494 | 9.58E-06 |
| ZNF662 | chr3 | 42947565 | 42947566 | cg17619311 | 0.304075 | 0.154943 | 7.54E-10 | 0.087906145 | -1.07764 | 6.42E-09 |
| ZNF662 | chr3 | 42947690 | 42947691 | cg24384244 | 0.220833 | 0.050645 | 2.06E-08 | 0.02985 | -1.07764 | 6.42E-09 |
| ZNF781 | chr19 | 38183418 | 38183419 | cg22685966 | 0.280357 | 0.10777 | 4.02E-09 | 0.100566438 | -1.32941 | 1.98E-08 |
| ZNF781 | chr19 | 38183055 | 38183056 | cg25875213 | 0.218791 | 0.035317 | 3.20E-08 | 0.029505886 | -1.32941 | 1.98E-08 |
| ZNF781 | chr19 | 38183259 | 38183260 | cg25324105 | 0.265343 | 0.060187 | 2.45E-10 | 0.042273325 | -1.32941 | 1.98E-08 |
| ZNF781 | chr19 | 38183262 | 38183263 | cg14587524 | 0.277692 | 0.058945 | 9.49E-10 | 0.041567302 | -1.32941 | 1.98E-08 |
| ZNF781 | chr19 | 38183253 | 38183254 | cg03611452 | 0.234615 | 0.062263 | 5.45E-10 | 0.036576216 | -1.32941 | 1.98E-08 |
| ZNF781 | chr19 | 38182773 | 38182774 | cg06940614 | 0.273612 | 0.119931 | 5.12E-09 | 0.11486 | -1.32941 | 1.98E-08 |

**EZH2** is the functional enzymatic component of the Polycomb Repressive Complex 2 (PRC2), which is responsible for healthy embryonic development through the epigenetic maintenance of genes responsible for regulating development and differentiation. EZH2 is highly over-expressed in cancer tissues and cells while no any methylation changes in its promoter region. EZH2 over-expression in cancer cell is regulated by over-expressed TFs which are binding to its promoter and enhancer regions (which TFs are quite interesting). We can find lots of TFs are over-expressing or low-expressing in cancer cells without methylation change, such as EZH2 and EZH2 is risk factors for cancer overall-survival time (OS) in CRC, liver cancer and melanoma while over-expressing is favorable factor for stomach cancer. Interesting thing is EZH2 have distinct performance in two types of lung cancer: LUAD (risk factor) and LUSC (favorable factor).

Methylation Micro-Array Platform

|  |  |
| --- | --- |
| A-GEOD-13534 | Illumina HumanMethylation450 BeadChip (HumanMethylation450\_15017482\_v.1.1) |
| A-GEOD-18809 | Illumina HumanMethylation450 BeadChip (v1.2, extended annotation) |
| A-MEXP-2255 | Infinium HumanMethylation450 BeadChip |
| A-GEOD-21145 | Infinium MethylationEPIC |
| [GPL13534](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL13534) | HumanMethylation450\_15017482 |
| [GPL16304](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL16304) | Illumina HumanMethylation450 BeadChip [UBC enhanced annotation v1.0] |
| GPL18809 | Illumina HumanMethylation450 BeadChip (v1.2, extended annotation) |
| GPL8490 | Illumina HumanMethylation27 BeadChip (HumanMethylation27\_270596\_v.1.2) |
| GPL21145 | Infinium MethylationEPIC |

Other microarray coocurrent with MH450K beadchip

CIMP, MSI, MUTATION and RNA-SEQ in colon cancer

|  |  |
| --- | --- |
| GPL16131 | [CytoScanHD\_Array] Affymetrix CytoScan HD Array |
| GPL570 | [HG-U133\_Plus\_2] Affymetrix Human Genome U133 Plus 2.0 Array |
| GPL22321 | [HG-U133\_Plus\_2] Affymetrix Human Genome U133 Plus 2.0 Array |
| GPL571 | [HG-U133A\_2] Affymetrix Human Genome U133A 2.0 Array |
| GPL17586 | [HTA-2\_0] Affymetrix Human Transcriptome Array 2.0 [transcript (gene) version] |
| GPL5175 | [HuEx-1\_0-st] Affymetrix Human Exon 1.0 ST Array [transcript (gene) version] |
| GPL10739 | [HuGene-1\_0-st] Affymetrix Human Gene 1.0 ST Array [probe set (exon) version] |
| GPL6244 | [HuGene-1\_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version] |
| GPL22286 | [HuGene-1\_1-st] Affymetrix Human Gene 1.1 ST Array [HuGene11stv1\_Hs\_ENSG version 19.0.0] |
| GPL11532 | [HuGene-1\_1-st] Affymetrix Human Gene 1.1 ST Array [transcript (gene) version] |
| GPL16876 | Agilent-020382 Human Custom Microarray 44k (Feature Number version) |
| GPL10332 | [Agilent-026652 Whole Human Genome Microarray 4x44K v2 (Feature Number version)](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL10332) |
| GPL15159 | Agilent-031181 Unrestricted\_Human\_miRNA\_V16.0\_Microarray 030840 (Probe Name version) |
| GPL17077 | Agilent-039494 SurePrint G3 Human GE v2 8x60K Microarray 039381 (Probe Name version) |
| GPL20157 | Agilent-041686 Unrestricted Human miRNA Microarray (miRNA\_ID version) |
| GPL10999 | Illumina Genome Analyzer IIx (Homo sapiens) |
| GPL9183 | Illumina GoldenGate Methylation Cancer Panel I |
| GPL15456 | Illumina HiScanSQ (Homo sapiens) |
| GPL11154 | Illumina HiSeq 2000 (Homo sapiens) |
| GPL16791 | Illumina HiSeq 2500 (Homo sapiens) |
| GPL6947 | Illumina HumanHT-12 V3.0 expression beadchip |
| GPL10558 | [Illumina HumanHT-12 V4.0 expression beadchip](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL10558) |
| GPL10904 | Illumina HumanHT-12 V4.0 expression beadchip (gene symbol) |
| GPL8490 | Illumina HumanMethylation27 BeadChip (HumanMethylation27\_270596\_v.1.2) |
| GPL13534 | Illumina HumanMethylation450 BeadChip (HumanMethylation450\_15017482) |
| GPL18809 | Illumina HumanMethylation450 BeadChip (v1.2, extended annotation) |
| GPL16304 | Illumina HumanMethylation450 BeadChip [UBC enhanced annotation v1.0] |
| GPL8882 | Illumina HumanOmni1-Quad BeadChip |
| GPL21145 | Infinium MethylationEPIC |
| GPL16025 | NimbleGen Homo sapiens Expression Array [100718\_HG18\_opt\_expr] |