**甲基化芯片技术分析**

**gc5k**

甲基化定义 3

基本术语 3

芯片设计 3

27K 4

450K 4

芯片内部质量控制 6

统计分析工具 6

生活习惯对甲基化的影响 8

epiclock 9

返老还童 9

# 甲基化定义

DNA甲基化是指C位点经过化学修饰转变为T的过程，经常发生在CpG位点。当DNA甲基化发生在promoter区段，经常抑制基因表达，但在基因区内的甲基化又常常带来基因的高表达，在植物和哺乳动物中都是如此(Laird 2010)。

## 基本术语

甲基化分析中经常使用的几个术语

Beta value: ，or [minif]其中M和U是甲基化与非甲基化的信号

MValue:

**DMP**: Differentially methylated position: single genomic position that has a different methylated level in two different groups of samples (or conditions).

**DMR**: Differentially methylated region: when consecutive genomic locations are differentially methylated in the same direction.

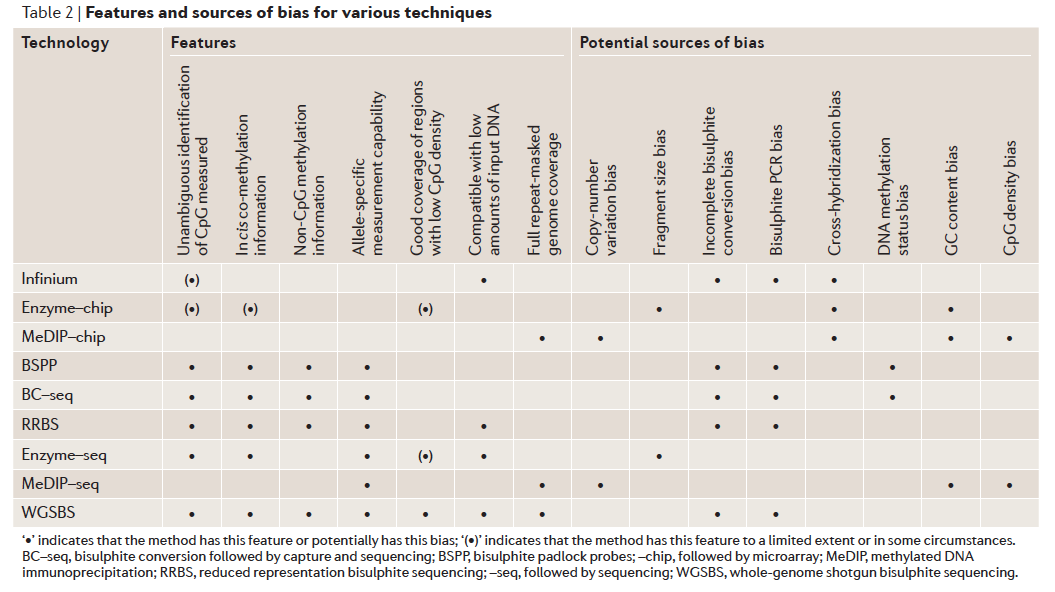
**Array**: One sample.

**Slide**: Physical slide containing 12 arrays (6×2 grid).

**Plate/batch:** Physical plate containing at most 8 slides (96 arrays). For this tutorial, we use **batch** and plate interchangeably.

# 芯片设计

甲基化主要使用测序或则芯片的方式测定(Bock 2012)。

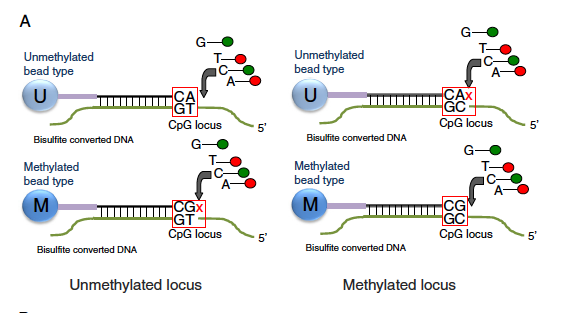


图片来自(Laird 2010)。

目前人类甲基化研究比较成熟的产品是Illumina设计的三款甲基化芯片产品。27K, 450K, 850K (EPIC)

## 27K

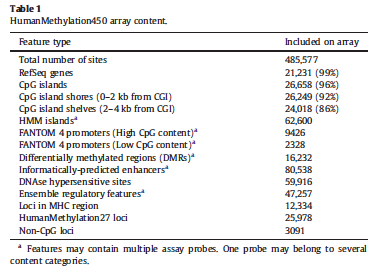
第一款是基于27000个甲基化位点开发的27K芯片[Infinium HumanMethylation27; Illumina, Inc, CA, USA](Bibikova *et al.* 2009)，使用双探针设计的Infinium I probe设计。  
Infinium I methylation-specific assay design consisting of two probes per CpG locus[图片来自(Bibikova *et al.* 2011)]

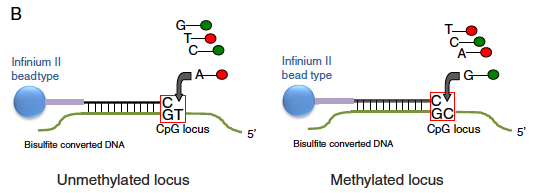


每个位点都采用双探针设计，分别用红、绿两个荧光色道[channel]进行测量。

## 450K

第二款在征集了19个研究所22名研究员[反映产业与实验室的关系]的意见开发了甲基化450K芯片[Infinium HumanMethylation 450K; Illumina, Inc, CA, USA]包含了485,577个位点(Bibikova *et al.* 2011)，94%的27K芯片位点被450K芯片所包含。450K芯片包含96%了当时已知的UCSC所标注的CpG岛，与测序数据的甲基化相比相关性的Rsq大约在95%以上。表格来自(Bibikova *et al.* 2011)。



450K芯片因为保留了27K的很多位点，所以在探针设计上有两类探针135,501个是Infinium I probe（如上图所示），350,076个Infinium II probe[图片来自(Bibikova *et al.* 2011)]，II型的probe采用的是单探针设计。  


使用Illumina自带的GenomeStudio，450K甲基化芯片可以探测位点的甲基化(m)或者非甲基化(u)，甲基化信号可以计算为 (Bibikova *et al.* 2011)；

, 默认, 。

但Lehne[GSE55763]等人建议使用不同的转化(Lehne *et al.* 2015)，甲基化位点的强度信号，经过转化可以得到甲基化的值。

450K芯片两种probe的设计不同，在第三方评估中，450K的两种probe都有系统误差(Dedeurwaerder *et al.* 2011) [GSE:42409; GSE:40279]。

1) 表现为CpG位点的probe杂交并不专一。CpG的probe可以杂交到多个位点，从而测量不准确；从而导致CpG的enrichment分析出现误差。这个问题在27K芯片中也有，当时是常染色体与性染色体的probe出现问题(Chen *et al.* 2011)。Homologous基因家族, duplicated基因和重复的原件都有可能是诱因。可以通过BLAT(Kent 2002)查找多重匹配的probe，而且要兼顾“Therefore, in order to identify cross-reactive sequences for both types of probe sequences, we need to perform BLAT on both strands of unmethylated and methylated genomes (total of4 genomes)”(Chen *et al.* 2011)。

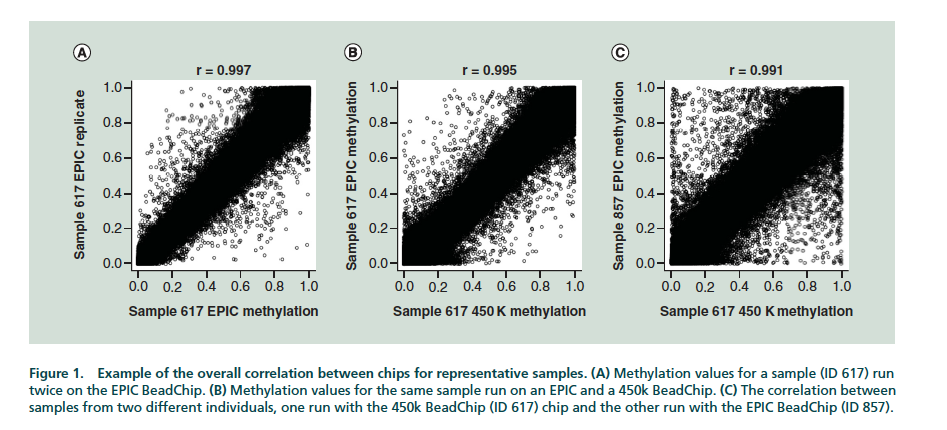
2) CpG的probe包含多肽位点。那么测到的多态性就可能仅仅是多态性而不是甲基化水平

需要额外工作来剔除这些非专一位点(Price *et al.* 2013)；从而导致信号是一个混杂信号。[其实同样的问题也曾经困扰基因芯片表达数据的精度(Benovoy *et al.* 2008)]。也可能就是基于芯片数据的一个死结；另外CNV等结构变异也是可以影响结果的！

3) 因为1-2的原因，后导致后期做GO分析出现偏差，因为基因本不在那些所谓的位点中。

TCGA项目曾经大量使用27K和450K芯片(Weisenberger 2014)，多达8千多个样本。目前450K芯片已经渐渐退出市场。但相关资料依旧可以找到https://support.illumina.com.cn/array/array\_kits/infinium\_humanmethylation450\_beadchip\_kit/downloads.html

第三款是在450芯片基础上开发的850K芯片[2015年12月发布]，也叫EPIC芯片(Moran *et al.* 2016)[GSE75073]。EPIC芯片包含90%以上的450芯片位点[853,307 CpG probe. 450K(482,421)的439,562 CpGs 包含其中]，并且比较有一致的表达(Logue *et al.* 2017)。所使用数据在Mol Psychiatry 2016, 21: 357–363。[图片来自Sadel， Epigenomics, 2017, 11:1363-71]



## 芯片内部质量控制

芯片中加入了各种类型的control probes (Illumina Inc 2010) 或者Fortin的补充材料(Fortin *et al.* 2014)

|  |  |  |
| --- | --- | --- |
| Probe | Number |  |
| Bisulfite Conversion I | 3 red channel + 3 green channel |  |
| Bisulfite Conversion II | 4 probes red channel |  |
| Extension | 2 (A, T) red + 2 (G, C) green |  |
| Hybridization | 3 green channel, corresponding to low, medium and high hybridization signals |  |
| Staining | 1 Green + 1 red |  |
| Non-polymorphic | 2 (C, G) green + 2 (A, T) red |  |
| Target removal | 2 green |  |
| Specificity II | 3 probe in both green and red |  |
| Specificity I | 3 in green + 3 in red |  |
| Normalization |  |  |
| Out-of-band probes (Oob) |  |  |

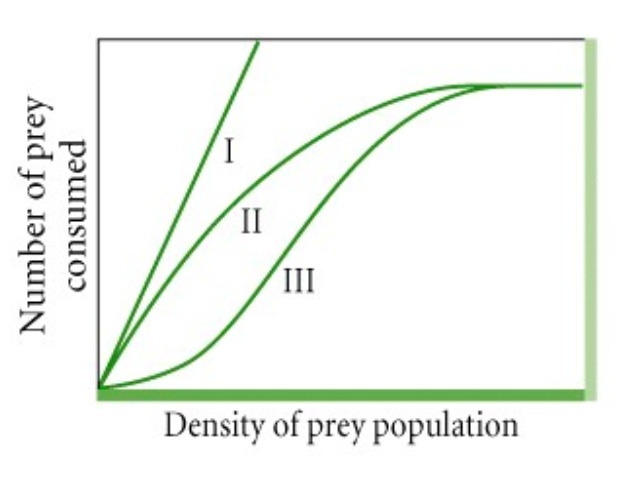
这些probe的存在，为检测质量好坏提供了很大的QC分析基础。

# 统计分析工具

作者通过LILOPOP群体的奖金2600个甲基化群体用450K测量，甲基化芯片数据并不存在所谓绝对最优的分析策略(Lehne *et al.* 2015)。但minif[[1]](#footnote-1)是目前最常用的甲基化分析工具(Aryee *et al.* 2014)以及其更新的版本(Tian *et al.* 2017)，以及针对更大数据集而构建的分析工具(Min *et al.* 2018)。

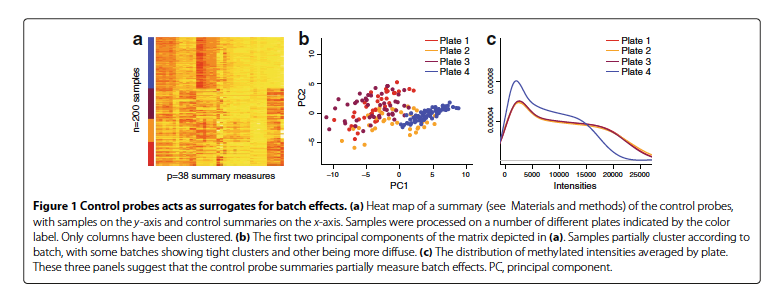
数据标准化是比较重要的一个指标，可以采用的标准化包括  
i) normalization for out-of-band probes [noob] (Triche *et al.* 2013)，对不同染色的信号进行标准化。这是相对比较接近芯片设计的物理层面的标准化方式。

ii) functional normalization (Fortin *et al.* 2014)，适用于存在case-control分组的情况。450K芯片上设计了848个control的probes，其中613个negative control，186个是array之间的标准化控制，另外还有49个留作他用。  
相比Supervised normalization (SNM; Bioinformatics 2010, 26:1308–1315), surrogate variable analysis (SVA; PLoS Genet 2007, 3:1724–1735; Proc Nat Acad Sci 2008, 105:18718–18723), ComBat (Biostatistics 2007, 8:118–127), Remove unwanted variation (RUV; Biostatistics 2012, 13:539–552)。  
Functional response大概可以分为三种[https://en.wikipedia.org/wiki/Functional\_response]



在甲基化中，将甲基化水平在[0-1]的区间分成k等分，然后进行如下的回归模型

[缺乏更细腻的描述]

其中Z是协变量，这里使用以contro probe构建的PC1 和 PC 2作为协变量，矫正后的残差就是再进行矫正  
使用control probe进行主成分分析不需要分析者提供实验设计的变量信息，直接通过主成分展示所得到的样品分布[Fortin et al Fig 1]。  


iii) subset within array normalization（SWAN）(Maksimovic *et al.* 2012)

iv) quantile normalization (Touleimat and Tost 2012)，单一来源材料。  
QN的目是让每个个体最终达到类似的表达方差，去除由于batch等技术原因而可能导致的问题，早在基因表达研究时期就已经引入(Irizarry *et al.* 2003)。quantile normalization算法可以标示如下，有如下四个基因A、B、C、D[行]和三个个体[列]。

|  |  |  |  |
| --- | --- | --- | --- |
| A | 5 | 4 | 3 |
| B | 2 | 1 | 4 |
| C | 3 | 4 | 6 |
| D | 4 | 2 | 8 |

原始状态下，每个个体的基因表达排序后得到

|  |  |  |  |
| --- | --- | --- | --- |
| A | iv | iii | i |
| B | i | i | ii |
| C | ii | iii | iii |
| D | iii | ii | iv |

则对原始矩阵数值进行重新排列

|  |  |  |  |
| --- | --- | --- | --- |
| A | 2 | 1 | 3 |
| B | 3 | 2 | 4 |
| C | 3 | 4 | 6 |
| D | 5 | 4 | 8 |

对单个基因数据进行矫正

|  |  |  |
| --- | --- | --- |
| A | [2+1+3]/3=2.00 | rank i |
| B | [3+2+4]/3=3.00 | rank ii |
| C | [4+4+6]/3=4.67 | rank iii |
| D | [5+4+8]/3=5.67 | rank iv |

QN后的基因表达数据为

|  |  |  |  |
| --- | --- | --- | --- |
| A | 5.67 | 4.67 | 2.00 |
| B | 2.00 | 2.00 | 3.00 |
| C | 3.00 | 4.67 | 4.67 |
| D | 4.67 | 3.00 | 5.67 |

v) beta-mixture normalization (Teschendorff *et al.* 2013)，对于不同设计的probe，根据其probe来源进行标准化。

各种标准化方式各有缺陷，因为对于这类数据驱动的分析，并不存在必然的标准化方式，而是随着目标和对数据的假设而变化(Dedeurwaerder *et al.* 2014)。

因为细胞的杂合会对甲基化产生非常大的影响，分析时候是需要矫正甲基化细胞杂合问题的，甲基化的杂合度是可以估算的(Houseman *et al.* 2012; Jaffe and Irizarry 2014)。

甲基化进行关联分析是现实中会采用的分析策略(Laird 2010)，分析一般集中在两种EWAS分析，主要是找关联的CpG位点或者夹杂GO分析(Zhang *et al.* 2017)。基于甲基化，然后再进行遗传学分析的也有，但不是很多(McRae *et al.* 2014)。另外可以将甲基化作为risk profile使用(Shah *et al.* 2015)。

Kolmogorov-Smirnov[KS]测验用于检验表达差异(Price *et al.* 2013)。也可以使用混合线性模型的方式(Zou *et al.* 2014)，或者矫正细胞类型(Houseman *et al.* 2014)，其中Houseman的算法已经申请专利[http://www.patentsencyclopedia.com/app/20140178348]。

对于差异表达的CpG位点，可以使用注释软件进行分析，目前使用较多的是MAGENTA(Ayellet *et al.* 2010)。

# 生活习惯对甲基化的影响

1 吸烟(Gao *et al.* 2017)、喝酒(Liu *et al.* 2018)

2 群体结构对甲基化的影响，这是一个很困难的问题，因为甲基化测量值受到probe是否包含多态性印象，其实是存在影响的。但这既可以解释为 i)群体结构高的区段甲基化水平差异大(Heyn *et al.* 2013)，ii)也可以说因为snp多态性密度高，导致群体了假阳性的甲基化(Daca-Roszak *et al.* 2015)。

但是也可能有证据表明snp多态性跟甲基化之间存在互作效应(Bell *et al.* 2011; Fraser *et al.* 2012)

甲基化基于111个参考群体的的综合分析可以在这里找到(Roadmap Epigenomics Consortium *et al.* 2015)，甲基化的群体遗传学资料汇编可以在这里找到(Taudt *et al.* 2016)。另外拟南芥也已经完成了参考群体的甲基化项目(Kawakatsu *et al.* 2016)。

# epiclock

关于甲基化最全的记录可以在wiki找到[<https://en.wikipedia.org/wiki/Epigenetic_clock>]

甲基化目前最重要的应用应该是甲基化年龄(Horvath and Raj 2018)。主要有几个方面的成果

1) 71个标记(Hannum *et al.* 2013)

2) Horvath的353个标记(Horvath 2013)，这个大的研究之前是基于一个很小的研究(Bocklandt *et al.* 2011)。Horvath已经讲这套标记转让给了Zymo，并且已经开发出产品[https://www.zymoresearch.eu/dnage]。

3) Horvath最近发布的可以预测疾病的标记系统(Levine *et al.* 2018)

亚洲人中这方面研究不是样本比较小就是数据不公开(Li *et al.* 2018)

最近Horvath也对甲基化年龄进行了GWAS分析(Lu *et al.* 2018)

张学军(Zhou *et al.* 2016)

## 返老还童

Gregory Fahy在Aging Cell上发表了一个非常小型的临床试验[<https://www.nature.com/articles/d41586-019-02638-w>

]，让9名51-65岁的男性白人身上测试了一种鸡尾酒疗法的组合，生长激素+两种糖尿病药物[growth hormone and two diabetes medications]。经过一年试验，甲基化年龄降低2.5岁。

**参考文献**

Aryee M. J., Jaffe A. E., Corrada-Bravo H., Ladd-Acosta C., Feinberg A. P., Hansen K. D., Irizarry R. A., 2014 Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics **30**: 1363–1369.

Ayellet V. S., Groop L., Mootha V. K., Daly M. J., Altshuler D., 2010 Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. PLoS Genet. **6**.

Bell J. T., Pai A. A., Pickrell J. K., Gaffney D. J., Degner J., Gilad Y., Pritchard J. K., 2011 DNA methylation associates with genetic and gene-expression variation in HapMap cell lines. Genome Biol. **12**: R10.

Benovoy D., Kwan T., Majewski J., 2008 Effect of polymorphisms within probe-target sequences on olignonucleotide microarray experiments. Nucleic Acids Res. **36**: 4417–4423.

Bibikova M., Le J., Barnes B., Saedinia-Melnyk S., Zhou L., Shen R., Gunderson K. L., 2009 Genome-wide DNA methylation profiling using Infinium@ assay. Epigenomics **1**: 177–200.

Bibikova M., Barnes B., Tsan C., Ho V., Klotzle B., Le J. M., Delano D., Zhang L., Schroth G. P., Gunderson K. L., Fan J. B., Shen R., 2011 High density DNA methylation array with single CpG site resolution. Genomics **98**: 288–295.

Bock C., 2012 Analysing and interpreting DNA methylation data. Nat. Rev. Genet. **13**: 705–719.

Bocklandt S., Lin W., Sehl M. E., Sanchez F. J., Sinsheimer J. S., Horvath S., Vilain E., 2011 Epigenetic Predictor of Age. PLoS One **101**: 1435–1439.

Chen Y., Choufani S., Ferreira J. C., Grafodatskaya D., Butcher D. T., Weksberg R., 2011 Sequence overlap between autosomal and sex-linked probes on the Illumina HumanMethylation27 microarray. Genomics **97**: 214–222.

Daca-Roszak P., Pfeifer A., Zebracka-Gala J., Rusinek D., Szybińska A., Jarzab B., Witt M., Zietkiewicz E., 2015 Impact of SNPs on methylation readouts by Illumina Infinium HumanMethylation450 BeadChip Array: Implications for comparative population studies. BMC Genomics **16**: 1003.

Dedeurwaerder S., Defrance M., Calonne E., Denis H., Sotiriou C., Fuks F., 2011 Evaluation of the Infinium Methylation 450K technology. Epigenomics **3**: 771–784.

Dedeurwaerder S., Defrance M., Bizet M., Calonne E., Bontempi G., Fuks F., 2014 A comprehensive overview of Infinium Human Methylation450 data processing. Brief. Bioinform. **15**: 929–941.

Fortin J. P., Labbe A., Lemire M., Zanke B. W., Hudson T. J., Fertig E. J., Greenwood C. M. T., Hansen K. D., 2014 Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol. **15**: 1–17.

Fraser H., Lam L., Neumann S., Kobor M., 2012 Population-Specificity of Human DNA Methylation. Genome Biol. **13**: R8.

Gao X., Thomsen H., Zhang Y., Breitling L. P., Brenner H., 2017 The impact of methylation quantitative trait loci (mQTLs) on active smoking-related DNA methylation changes. Clin. Epigenetics **9**: 1–13.

Hannum G., Guinney J., Zhao L., Zhang L., Hughes G., Sadda S., Klotzle B., Bibikova M., Fan J.-B., Gao Y., Deconde R., Chen M., Rajapakse I., Friend S., Ideker T., Zhang K., 2013 Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol. Cell **49**: 359–67.

Heyn H., Moran S., Hernando-herraez I., Sayols S., Gomez A., Sandoval J., Monk D., Hata K., Marques-bonet T., Wang L., Esteller M., 2013 DNA methylation contributes to natural human variation. Genome Res. **23**: 1363–1372.

Horvath S., 2013 DNA methylation age of human tissues and cell types DNA methylation age of human tissues and cell types. Genome Biol. **14**: R115.

Horvath S., Raj K., 2018 DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat. Rev. Genet. **19**: 371–384.

Houseman E. A., Accomando W. P., Koestler D. C., Christensen B. C., Marsit C. J., Nelson H. H., Wiencke J. K., Kelsey K. T., 2012 DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics **13**: 86.

Houseman E. A., Molitor J., Marsit C. J., 2014 Reference-free cell mixture adjustments in analysis of DNA methylation data. Bioinformatics **30**: 1431–1439.

Illumina Inc, 2010 *GenomeStudio Methylation Module v1.8 User Guide (11319130B)*.

Irizarry R. A., Hobbs B., Beazer-barclay Y. D., Antonellis K. J., Scherf U. W. E., Speed T. P., 2003 Exploration , normalization , and summaries of high density oligonucleotide array probe level data. Biostatistics **4**: 249–264.

Jaffe A. E., Irizarry R. A., 2014 Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol. **15**: R31.

Kawakatsu T., Huang S. C., Jupe F., Sasaki E., Schmitz R. J., Urich M. A., Castanon R., Nery J. R., Barragan C., He Y., Chen H., Dubin M., Lee C. R., Wang C., Bemm F., Becker C., O’Neil R., O’Malley R. C., Quarless D. X., Alonso-Blanco C., Andrade J., Becker C., Bemm F., Bergelson J., Borgwardt K., Chae E., Dezwaan T., Ding W., Ecker J. R., 2016 Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions. Cell **166**: 492–506.

Kent W. J., 2002 BLAT — The BLAST-Like Alignment Tool. Genome Res. **12**: 656–664.

Laird P. W., 2010 Principles and challenges of genome-wide DNA methylation analysis. Nat. Rev. Genet. **11**: 191–203.

Lehne B., Drong A. W., Loh M., Zhang W., Scott W. R., Tan S. T., Afzal U., Scott J., Jarvelin M. R., Elliott P., McCarthy M. I., Kooner J. S., Chambers J. C., 2015 A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. Genome Biol. **16**: 37.

Levine M. E., Lu A. T., Quach A., Chen B. H., Assimes T. L., Bandinelli S., Hou L., Baccarelli A. A., Stewart J. D., Li Y., Whitsel E. A., Wilson J. G., Reiner1 A. P., Aviv1 A., Lohman K., Liu Y., Ferrucci L., Horvath S., 2018 An epigenetic biomarker of aging for lifespan and healthspan. Aging (Albany. NY). **10**: 573–591.

Li C., Gao W., Gao Y., Yu C., Lv J., Lv R., Duan J., Sun Y., Guo X., Cao W., Li L., 2018 Age prediction of children and adolescents aged 6 ‐ 17 years : an epigenome ‐ wide analysis of DNA methylation. Aging (Albany. NY). **10**: 1015–1026.

Liu C., Marioni R. E., Hedman A. K., Pfeiffer L., Tsai P. C., Reynolds L. M., Just A. C., Duan Q., Boer C. G., Tanaka T., Elks C. E., Aslibekyan S., Brody J. A., Kühnel B., Herder C., Almli L. M., Zhi D., Wang Y., Huan T., Yao C., Mendelson M. M., Joehanes R., Liang L., Love S. A., Guan W., Shah S., McRae A. F., Kretschmer A., Prokisch H., Strauch K., Peters A., Visscher P. M., Wray N. R., Guo X., Wiggins K. L., Smith A. K., Binder E. B., Ressler K. J., Irvin M. R., Absher D. M., Hernandez D., Ferrucci L., Bandinelli S., Lohman K., Ding J., Trevisi L., Gustafsson S., Sandling J. H., Stolk L., Uitterlinden A. G., Yet I., Castillo-Fernandez J. E., Spector T. D., Schwartz J. D., Vokonas P., Lind L., Li Y., Fornage M., Arnett D. K., Wareham N. J., Sotoodehnia N., Ong K. K., Meurs J. B. J. Van, Conneely K. N., Baccarelli A. A., Deary I. J., Bell J. T., North K. E., Liu Y., Waldenberger M., London S. J., Ingelsson E., Levy D., 2018 A DNA methylation biomarker of alcohol consumption. Mol. Psychiatry **23**: 422–433.

Logue M. W., Smith A. K., Wolf E. J., Maniates H., Stone A., Schichman S. A., McGlinchey R. E., Milberg W., Miller M. W., 2017 The correlation of methylation levels measured using Illumina 450K and EPIC BeadChips in blood samples. Epigenomics **9**: 1363–1371.

Lu A. T., Xue L., Salfati E. L., Chen B. H., Ferrucci L., Levy D., Joehanes R., Murabito J. M., Kiel D. P., Tsai P. C., Yet I., Bell J. T., Mangino M., Tanaka T., McRae A. F., Marioni R. E., Visscher P. M., Wray N. R., Deary I. J., Levine M. E., Quach A., Assimes T., Tsao P. S., Absher D., Stewart J. D., Li Y., Reiner A. P., Hou L., Baccarelli A. A., Whitsel E. A., Aviv A., Cardona A., Day F. R., Wareham N. J., Perry J. R. B., Ong K. K., Raj K., Lunetta K. L., Horvath S., 2018 GWAS of epigenetic aging rates in blood reveals a critical role for TERT. Nat. Commun. **9**: 387.

Maksimovic J., Gordon L., Oshlack A., 2012 SWAN: Subset-quantile Within Array Normalization for Illumina Infinium HumanMethylation450 BeadChips. Genome Biol. **13**: R44.

McRae A. F., Powell J. E., Henders A. K., Bowdler L., Hemani G., Shah S., Painter J. N., Martin N. G., Visscher P. M., Montgomery G. W., 2014 Contribution of genetic variation to transgenerational inheritance of DNA methylation. Genome Biol. **15**: R73.

Min J. L., Hemani G., Davey Smith G., Relton C., Suderman M., 2018 Meffil: efficient normalization and analysis of very large DNA methylation datasets. Bioinformatics: 1–7.

Moran S., Arribas C., Esteller M., 2016 Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics **8**: 389–399.

Price M. E., Cotton A. M., Lam L. L., Farré P., Emberly E., Brown C. J., Robinson W. P., Kobor M. S., 2013 Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. Epigenetics and Chromatin **6**: 1–15.

Roadmap Epigenomics Consortium, Kundaje A., Meuleman W., Ernst J., Bilenky M., Yen A., Heravi-Moussavi A., Kheradpour P., Zhang Z., Wang J., Ziller M. J., Amin V., Whitaker J. W., Schultz M. D., Ward L. D., Sarkar A., Quon G., Sandstrom R. S., Eaton M. L., Wu Y.-C., Pfenning A. R., Wang X., Claussnitzer M., Liu Y., Coarfa C., Harris R. A., Shoresh N., Epstein C. B., Gjoneska E., Leung D., Xie W., Hawkins R. D., Lister R., Hong C., Gascard P., Mungall A. J., Moore R., Chuah E., Tam A., Canfield T. K., Hansen R. S., Kaul R., Sabo P. J., Bansal M. S., Carles A., Dixon J. R., Farh K.-H., Feizi S., Karlic R., Kim A.-R., Kulkarni A., Li D., Lowdon R., Elliott G., Mercer T. R., Neph S. J., Onuchic V., Polak P., Rajagopal N., Ray P., Sallari R. C., Siebenthall K. T., Sinnott-Armstrong N. a., Stevens M., Thurman R. E., Wu J., Zhang B., Zhou X., Beaudet A. E., Boyer L. a., Jager P. L. De, Farnham P. J., Fisher S. J., Haussler D., Jones S. J. M., Li W., Marra M. a., McManus M. T., Sunyaev S., Thomson J. a., Tlsty T. D., Tsai L.-H., Wang W., Waterland R. a., Zhang M. Q., Chadwick L. H., Bernstein B. E., Costello J. F., Ecker J. R., Hirst M., Meissner A., Milosavljevic A., Ren B., Stamatoyannopoulos J. a., Wang T., Kellis M., Kundaje A., Meuleman W., Ernst J., Bilenky M., Yen A., Heravi-Moussavi A., Kheradpour P., Zhang Z., Wang J., Ziller M. J., Amin V., Whitaker J. W., Schultz M. D., Ward L. D., Sarkar A., Quon G., Sandstrom R. S., Eaton M. L., Wu Y.-C., Pfenning A., Wang X., ClaussnitzerYaping Liu M., Coarfa C., Alan Harris R., Shoresh N., Epstein C. B., Gjoneska E., Leung D., Xie W., David Hawkins R., Lister R., Hong C., Gascard P., Mungall A. J., Moore R., Chuah E., Tam A., Canfield T. K., Scott Hansen R., Kaul R., Sabo P. J., Bansal M. S., Carles A., Dixon J. R., Farh K.-H., Feizi S., Karlic R., Kim A.-R., Kulkarni A., Li D., Lowdon R., Elliott G., Mercer T. R., Neph S. J., Onuchic V., Polak P., Rajagopal N., Ray P., Sallari R. C., Siebenthall K. T., Sinnott-Armstrong N. a., Stevens M., Thurman R. E., Wu J., Zhang B., Zhou X., Abdennur N., Adli M., Akerman M., Barrera L., Antosiewicz-Bourget J., Ballinger T., Barnes M. J., Bates D., Bell R. J. a., Bennett D. a., Bianco K., Bock C., Boyle P., Brinchmann J., Caballero-Campo P., Camahort R., Carrasco-Alfonso M. J., Charnecki T., Chen H., Chen Z., Cheng J. B., Cho S., Chu A., Chung W.-Y., Cowan C., Athena Deng Q., Deshpande V., Diegel M., Ding B., Durham T., Echipare L., Edsall L., Flowers D., Genbacev-Krtolica O., Gifford C., Gillespie S., Giste E., Glass I. a., Gnirke A., Gormley M., Gu H., Gu J., Hafler D. a., Hangauer M. J., Hariharan M., Hatan M., Haugen E., He Y., Heimfeld S., Herlofsen S., Hou Z., Humbert R., Issner R., Jackson A. R., Jia H., Jiang P., Johnson A. K., Kadlecek T., Kamoh B., Kapidzic M., Kent J., Kim A., Kleinewietfeld M., Klugman S., Krishnan J., Kuan S., Kutyavin T., Lee A.-Y., Lee K., Li J., Li N., Li Y., Ligon K. L., Lin S., Lin Y., Liu J., Liu Y., Luckey C. J., Ma Y. P., Maire C., Marson A., Mattick J. S., Mayo M., McMaster M., Metsky H., Mikkelsen T., Miller D., Miri M., Mukame E., Nagarajan R. P., Neri F., Nery J., Nguyen T., O’Geen H., Paithankar S., Papayannopoulou T., Pelizzola M., Plettner P., Propson N. E., Raghuraman S., Raney B. J., Raubitschek A., Reynolds A. P., Richards H., Riehle K., Rinaudo P., Robinson J. F., Rockweiler N. B., Rosen E., Rynes E., Schein J., Sears R., Sejnowski T., Shafer A., Shen L., Shoemaker R., Sigaroudinia M., Slukvin I., Stehling-Sun S., Stewart R., Subramanian S. L., Suknuntha K., Swanson S., Tian S., Tilden H., Tsai L., Urich M., Vaughn I., Vierstra J., Vong S., Wagner U., Wang H., Wang T., Wang Y., Weiss A., Whitton H., Wildberg A., Witt H., Won K.-J., Xie M., Xing X., Xu I., Xuan Z., Ye Z., Yen C., Yu P., Zhang X., Zhang X., Zhao J., Zhou Y., Zhu J., Zhu Y., Ziegler S., Beaudet A. E., Boyer L. a., Jager P. L. De, Farnham P. J., Fisher S. J., Haussler D., Jones S. J. M., Li W., Marra M. a., McManus M. T., Sunyaev S., Thomson J. a., Tlsty T. D., Tsai L.-H., Wang W., Waterland R. a., Zhang M. Q., Chadwick L. H., Bernstein B. E., Costello J. F., Ecker J. R., Hirst M., Meissner A., Milosavljevic A., Ren B., Stamatoyannopoulos J. a., Wang T., Kellis M., 2015 Integrative analysis of 111 reference human epigenomes. Nature **518**: 317–330.

Shah S., Bonder M. J., Marioni R. E., Zhu Z., McRae A. F., Zhernakova A., Harris S. E., Liewald D., Henders A. K., Mendelson M. M., Liu C., Joehanes R., Liang L., Levy D., Martin N. G., Starr J. M., Wijmenga C., Wray N. R., Yang J., Montgomery G. W., Franke L., Deary I. J., Visscher P. M., 2015 Improving Phenotypic Prediction by Combining Genetic and Epigenetic Associations. Am. J. Hum. Genet. **97**: 75–85.

Taudt A., Colomé-Tatché M., Johannes F., 2016 Genetic sources of population epigenomic variation. Nat. Rev. Genet. **17**: 319–332.

Teschendorff A. E., Marabita F., Lechner M., Bartlett T., Tegner J., Gomez-Cabrero D., Beck S., 2013 A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics **29**: 189–196.

Tian Y., Morris T. J., Webster A. P., Yang Z., Beck S., Feber A., Teschendorff A. E., 2017 ChAMP: Updated methylation analysis pipeline for Illumina BeadChips. Bioinformatics **33**: 3982–3984.

Touleimat N., Tost J., 2012 Complete pipeline for Infinium® Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. Epigenomics **4**: 325–341.

Triche T. J., Weisenberger D. J., Berg D. Van Den, Laird P. W., Siegmund K. D., 2013 Low-level processing of Illumina Infinium DNA Methylation BeadArrays. Nucleic Acids Res. **41**: e90.

Weisenberger D. J., 2014 Characterizing DNA methylation alterations from The Cancer Genome Atlas. J. Clin. Invest. **124**: 17–24.

Zhang X., Hu Y., Justice A. C., Li B., Wang Z., Zhao H., Krystal J. H., Xu K., 2017 DNA methylation signatures of illicit drug injection and hepatitis C are associated with HIV frailty. Nat. Commun. **8**: 2243.

Zhou F., Shen C., Xu J., Gao J., Zheng X., Ko R., Dou J., Cheng Y., Zhu C., Xu S., Tang X., Zuo X., Yin X., Cui Y., Sun L., Tsoi L. C., Hsu Y. H., Yang S., Zhang X., 2016 Epigenome-wide association data implicates DNA methylation-mediated genetic risk in psoriasis. Clin. Epigenetics **8**: 1–9.

Zou J., Lippert C., Heckerman D., Aryee M., Listgarten J., 2014 Epigenome-wide association studies without the need for cell-type composition. Nat. Methods **11**: 309–311.

1. http://www.bioconductor.org/help/course-materials/2015/BioC2015/methylation450k.html [↑](#footnote-ref-1)