咨询的主要内容（2019年9月4日）

**上午9:00 – 12:00**

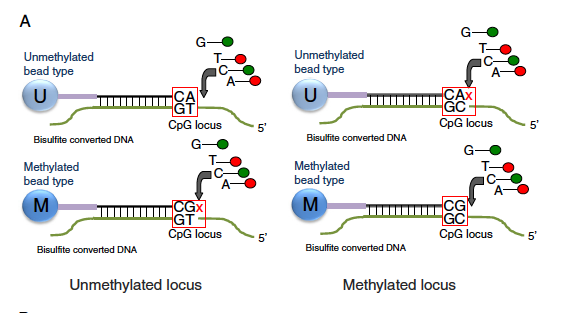
问：DNA甲基化是什么？

答：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)。

问：甲基化芯片

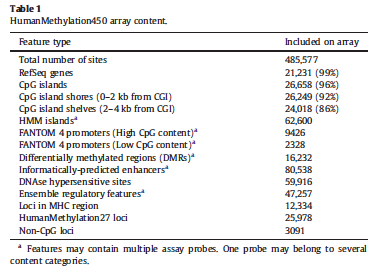
答：甲基化目前普遍使用芯片测序，都是有Illumina公司主导，第一代芯片是

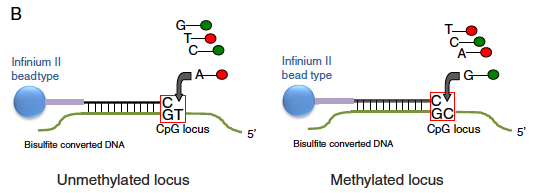
目前人类甲基化研究比较成熟的产品是Illumina设计的三款甲基化芯片产品。27K, 450K, 850K (EPIC)  
第一款是基于27000个甲基化位点开发的27K芯片[Infinium HumanMethylation27; Illumina, Inc, CA, USA](Bibikova *et al.* 2009)，使用双探针设计的Infinium I probe设计。  
Infinium I methylation-specific assay design consisting of two probes per CpG locus[图片来自(Bibikova *et al.* 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。

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。

目前850K芯片已经上市，逐渐会成为主流应用。

**问：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).

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)。

问: 甲基化与疾病的研究进展

答：甲基化最突出的研究成果，是Steve Horvath发表的基于353个CpG甲基化可以广泛预测人体的“内生”年龄(Horvath 2013)，而且与个体的实际年龄相关度极高，Pearson’s correlation达到95%以上。而且，这组预测标记是在跨器官、跨种族的研究中都得到了比较好的验证，就很强的广谱性(Horvath *et al.* 2016)。而且对于病变的器官，其预测的甲基化年龄往往高于物理年龄。

研究中发现甲基化年龄在一系列疾病中会加速

Down Syndrome(Horvath *et al.* 2015)，女性乳腺组织(Horvath 2013)。更加详尽的列表可以在wiki上找到[<https://en.wikipedia.org/wiki/Epigenetic_clock>]。

另外生活历也会导致甲基化年龄加速，比如饮食，教育，锻炼等等因素，都会导致甲基化年龄变化(Quach *et al.* 2017)；而肥胖则会显著加速肝组织的甲基化年龄(Horvath *et al.* 2014)。最近。Steve Horvath与合作者又根据实际数据分析，进一步提出了更广泛可以使用的甲基化标记(Levine *et al.* 2018)，以及跟个人生活历也有关系。

最近Steve Horvath的Nature Reviews Genetics更进一步总结了甲基化在研究进展(Horvath and Raj 2018)。

问：甲基化主要使用的分析包

答：甲基化目前只用最多的分析包是minfi(Aryee *et al.* 2014)和meffil(Min *et al.* 2018)。

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.

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.

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.

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

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

Horvath S., Erhart W., Brosch M., Ammerpohl O., Schönfels W. Von, Ahrens M., Heits N., Bell J. T., Tsai P. C., Spector T. D., Deloukas P., Siebert R., Sipos B., Becker T., Röcken C., Schafmayer C., Hampe J., 2014 Obesity accelerates epigenetic aging of human liver. Proc. Natl. Acad. Sci. U. S. A. **111**: 15538–15543.

Horvath S., Garagnani P., Bacalini M. G., Pirazzini C., Salvioli S., Gentilini D., Blasio A. M. Di, Giuliani C., Tung S., Vinters H. V., Franceschi C., 2015 Accelerated epigenetic aging in Down syndrome. Aging Cell **14**: 491–495.

Horvath S., Gurven M., Levine M. E., Trumble B. C., Kaplan H., Allayee H., Ritz B. R., Chen B., Lu A. T., Rickabaugh T. M., Jamieson B. D., Sun D., Li S., Chen W., Quintana-Murci L., Fagny M., Kobor M. S., Tsao P. S., Reiner A. P., Edlefsen K. L., Absher D., Assimes T. L., 2016 An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. Genome Biol. **17**: 0–22.

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

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 **10**: 573–591.

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.

Quach A., Levine M. E., Tanaka T., Lu A. T., Chen B. H., Ferrucci L., Ritz B., Bandinelli S., Neuhouser M. L., Beasley J. M., Snetselaar L., Wallace R. B., Tsao P. S., Absher D., Assimes T. L., Stewart J. D., Li Y., Hou L., Baccarelli A. A., Whitsel E. A., Horvath S., 2017 Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. Aging **9**: 419–446.