待讨论问题：

1. 导出的表格数据都能查到，是怎么样得到这样的表格，具体流程。
2. 这里的数据源全都来源于encodeproject,还需要结合roadmap的数据吗？还有其他数据来源吗？
3. Cpgpt的数据来源主要来源于array的数据，这种数据和wgbs数据得到的结果差异大吗。
4. 有感叹号的样本是否还需要考虑
5. 需要考虑不同测序平台的差距吗？
6. Cell line那里的类别需要统一吗
7. 有些说明是某些药物治疗情况下，这种要考虑差异吗？

逐条解释如下：

Assay

Whole-genome shotgun bisulfite sequencing (WGBS)

仍然是用 亚硫酸氢盐 处理 DNA 来检测全基因组甲基化状态的测序方法。

Biosample summary

Homo sapiens GM12878

物种：人类（Homo sapiens）。

GM12878：这是一个非常著名的人类B淋巴母细胞系（lymphoblastoid cell line），来自国际HapMap计划，用于很多基因组学和表观遗传学研究（ENCODE计划的标准样本之一）。

Biosample Type

Cell line

样本来自培养的细胞系，而不是组织、血液或单个个体的原代细胞。

Replication type

Isogenic

这里指的是样本来自相同遗传背景的细胞（同一细胞系），避免了不同个体遗传差异带来的影响。

Description

HAIB Whole Genome Bisulfite Sequencing GM12878

HAIB：HudsonAlpha Institute for Biotechnology，美国阿拉巴马州的一个生物技术研究所，很活跃于基因组学领域。

表示这是由 HAIB 机构进行的 GM12878 细胞系 WGBS 实验。

Nucleic acid type

DNA

测的是 DNA（而不是 RNA、蛋白质等）。

Platform

Illumina HiSeq 2500

Illumina 的高通量测序仪，比 HiSeq X Ten 旧一代，但依然能做全基因组测序。

通常可选 2×125bp 或 2×150bp 双端测序模式。