



# 中华人民共和国国家标准

GB/T XXXXX—XXXX

## 人全基因组甲基化高通量测序数据质量评价通用要求

General requirements for the data quality assessment of human whole-genome methylation high-throughput sequencing

(点击此处添加与国际标准一致性程度的标识)

草案版次选择

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX - XX - XX 发布

XXXX - XX - XX 实施

国家市场监督管理总局  
国家标准化管理委员会 发布

## 目 次

前 言 .....	II
1 范围 .....	3
2 规范性引用文件 .....	3
3 术语和定义 .....	3
4 缩略词 .....	
5 质量要求 .....	7
6 评价方法 .....	10
参 考 文 献 .....	11

## 前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由国家药品监督管理局提出。

本文件由全国医用临床检验实验室和体外诊断系统标准化技术委员会（SAC/TC 136）归口。

本文件起草单位：

本文件主要起草人：

# 人全基因组甲基化高通量测序数据质量评价通用要求

## 1 范围

本文件规定了人全基因组甲基化高通量测序数据的质量要求，描述了相应的评价方法。  
本文件适用于基于高通量测序对人基因组DNA进行全基因组甲基化测序的数据质量评价。  
本文件不适用于双脱氧链终止法测序（桑格测序）和单分子测序的数据。

## 2 规范性引用文件

本文件没有规范性引用文件。

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

**人全基因组甲基化高通量测序** human whole genome methylation sequencing

将甲基化转化技术与高通量测序技术相结合，对人类不同个体或群体进行全基因组单碱基分辨率的5-甲基胞嘧啶（5mC）和5-羟甲基胞嘧啶（5hmC）测序。

注：1. 包括人的23对染色体核酸序列、线粒体核酸序列。

2. 甲基化转化技术主要分为将非甲基化胞嘧啶（C）转化成胸腺嘧啶（T）技术和将甲基化胞嘧啶（mC）碱基转化成胸腺嘧啶（T）技术2种。

### 3.2

**重亚硫酸盐处理** bisulfite conversion

采用重亚硫酸盐将DNA中的非甲基化胞嘧啶（C）转化为尿嘧啶（U），而甲基化的胞嘧啶（5-甲基胞嘧啶，5mC）保持不变。U在后续PCR过程中转化为胸腺嘧啶（T）。

注：在重亚硫酸盐处理过程中，5-羟甲基胞嘧啶（5hmC）会形成5-磺酰基甲基胞嘧啶（smC），在测序中仍然被读取为胞嘧啶（C）。

### 3.3

**酶促转化** enzymatic conversion

采用两个关键酶促转化步骤将DNA中的非甲基化胞嘧啶（C）转化为尿嘧啶（U），而甲基化的胞嘧啶（5-甲基胞嘧啶，5mC和5-羟甲基胞嘧啶，5hmC）保持不变。U在后续PCR过程中转化为胸腺嘧啶（T）。

注：首先使用TET2酶和氧化增强剂对5mC和5hmC进行氧化，生成5-羧基胞嘧啶（5caC）和5-羧甲基胞嘧啶（5ghmC），

再使用APOBEC酶对非甲基化胞嘧啶（C）进行脱氨，转化为尿嘧啶（U），而5caC和5ghmC保持不变。

### 3.4

#### TET辅助吡啶硼烷转化 TET-assisted pyridine borane conversion

采用TET酶及吡啶硼烷等还原试剂将DNA中的甲基化胞嘧啶（mC）还原为二氢尿嘧啶（DHU），而非甲基化的胞嘧啶（C）保持不变。DHU在后续PCR过程中转化为胸腺嘧啶（T）。

注：TET酶将5-甲基胞嘧啶（5mC）和5-羟甲基胞嘧啶（5hmC）均氧化成5-羧基胞嘧啶（5caC），然后通过还原试剂（如：吡啶硼烷等）将5caC还原为DHU。

### 3.5

#### 碱基平衡文库 base-balanced library

在高通量测序时，为确保4种碱基（A、T、C、G）测序信号相对均衡，4种碱基（A、T、C、G）的比例相对均衡的高通量测序文库或额外添加后使4种碱基（A、T、C、G）的比例相对均衡的高通量测序文库。

### 3.6

#### 碱基含量 base content

测序片段碱基中不同碱基（A、C、G、T、N）占碱基总数量的百分比。

### 3.7

#### 基因组唯一比对率 unique reads rate

数据过滤后，单末端测序中的一条测序片段（read）或双末端测序中的一对测序片段（reads）与参考基因组比对时，比对到参考基因组唯一位置的reads数与所有比对上的reads数的百分比。

### 3.8

#### C to T转化率 C to T conversion rate

非甲基化胞嘧啶（C）经重亚硫酸盐处理或酶促转化后被成功转化为胸腺嘧啶（T）的碱基数占比。

### 3.9

#### mC to T转化率 mC to T conversion rate

甲基化胞嘧啶（mC）经TET辅助吡啶硼烷转化后被成功转化胸腺嘧啶（T）的碱基数占比。

### 3.10

#### 胞嘧啶平均测序深度 C mean sequencing depth

经过数据过滤、序列比对、去重后获得的一个全基因组甲基化测序样本的基因组中C碱基区域的平均测序深度。

注：1. 计算方法为该样本测序得到能够比对上基因组C碱基区域的碱基总量（bp）与基因组C碱基区域的大小的比值，计量单位为×，即乘数。

2. 根据C碱基在参考基因组上的分布，C碱基区域通常分CG、CHG、CHH（H代表A/T/C碱基）三种情况。

### 3.11

#### 胞嘧啶测序覆盖度 C sequencing coverage

测序片段与参考基因组比对并去重后，参考基因组上至少被测序片段覆盖1次的C碱基数目占C碱基总数的百分比。

注：根据C碱基在参考基因组上的分布，通常分CG、CHG、CHH（H代表A/T/C碱基）三种情况计算覆盖度。

### 3.12

#### 甲基化率 methylation rate

测序片段与参考基因组比对并去重后，特定区域或者整个基因组上甲基化胞嘧啶位点数量与总胞嘧啶位点数量的比值。

甲基化率的计算方法可用式（1）表示：

$$\text{甲基化率} = \frac{\text{甲基化胞嘧啶位点的数目}}{\text{甲基化胞嘧啶位点的数目} + \text{未甲基化的胞嘧啶的数目}} \times 100\% \dots\dots\dots (1)$$

注：1. 根据C碱基在参考基因组上的分布，通常分CG、CHG、CHH（H代表A/T/C碱基）三种情况计算甲基化率。

2. 根据甲基化率划定阈值分别判断低甲基化、中甲基化、高甲基化。

3. 特定区域或者整个基因组上的甲基化率可以通过计算单碱基位点的平均甲基化率来表示。

### 3.13

#### 标准甲基化数据集 standard methylation dataset

基于人基因组标准物质构建的高置信甲基化数据集。

### 3.14

#### 准确度 precision

检测到的真正甲基化位点个数占检测到的全部甲基化位点个数的百分比。

准确度的计算方法可用式（2）表示：

$$PC = \frac{TP}{TP+FP} \times 100\% \dots\dots\dots (2)$$

式中：

PC——甲基化位点检测的准确度。

TP——与标准甲基化数据集甲基化位点一致的检测位点个数。

FP——与标准甲基化数据集甲基化位点不一致的检测位点个数。

注：区别于其他领域precision的含义，precision在本文件专指甲基化位点检测统计结果的准确度。

### 3.15

**灵敏度 sensitivity**

检测到的真正甲基化位点个数占待检测的所有甲基化位点个数的百分比

灵敏度的计算方法可用式（3）表示：

$$SE = \frac{TP}{TP+FN} \times 100\% \dots\dots\dots (3)$$

式中：

SE——甲基化位点检测的灵敏度。

TP——与标准甲基化数据集甲基化位点一致的检测位点个数。

FN——标准甲基化数据集中包含、但未被检出的甲基化位点个数。

注：区别于其他领域sensitivity的含义，sensitivity在本文件专指甲基化位点检测统计结果的灵敏度。

**4 缩略词**

下列缩略语适用于本文件。

DNA：脱氧核糖核酸（Deoxyribonucleic Acid）

TET2：十-十一易位甲基胞嘧啶双加氧酶2（Ten-Eleven Translocation Methylcytosine Dioxygenase 2）

APOBEC：载脂蛋白B信使核糖核酸编辑催化多肽（Apolipoprotein B mRNA Editing Catalyzed Polypeptide）

TET：十-十一易位酶（Ten-Eleven Translocation）

DHU：二氢尿嘧啶（Dihydrouracil）

PCR：聚合酶链式反应（Polymerase Chain Reaction）

bp：碱基对（base pair）

CG：鸟嘌呤和胞嘧啶（Guanine and Cytosine）

PC：准确度（Precision）

TP：真阳（True Positive）

FP：假阳（False Positive）

SE：灵敏度（Sensitivity）

FN：假阴（False Negative）

kb：千碱基对（kilo base pair）

RNA：核糖核酸（Ribonucleic Acid）

OD：光密度（Optical Density）

A260/280：260纳米与280纳米的吸光度比值（Absorbance Ratio at 260 nm to 280 nm）

A260/230：260纳米与230纳米的吸光度比值（Absorbance Ratio at 260 nm to 280 nm）

pH：氢离子浓度指数（Potential of Hydrogen）

PE：双端（Paired-End）

## 5 质量要求

### 5.1 样本质量要求

#### 5.1.1 样本类型

测序样本类型为人基因组DNA。DNA的来源主要为人全血、血液白膜层、正常组织、正常细胞系、肿瘤组织、肿瘤细胞系，样本来源信息清晰，且保存、运输得当，无明显污染。

本文件采用经国家权威机构批准发布的正常人细胞系提取的人基因组DNA作为质量评价指标的实物标准。

注：本文件所涉及的人基因组DNA实物标准信息见附录A。

#### 5.1.2 DNA 样本质量

##### 5.1.2.1 DNA 样本的完整性

基因组DNA完整，没有明显的降解现象。通常使用1%-2%琼脂糖电泳，基因组条带集中在23kb及以上，无明显拖尾或弥散。

##### 5.1.2.2 DNA 样本的纯度

DNA样本透明澄清，无明显黏稠状，无明显色素或杂质。DNA样本中无RNA、蛋白残留。通常使用紫外分光光度计测定，以OD值（A260/280、A260/230）来表示。DNA样本的纯度应符合建库试剂的要求。

##### 5.1.2.3 DNA 样本的体积、浓度和总量

DNA样本的体积、浓度和总量应符合建库试剂盒的要求。

##### 5.1.2.4 DNA 样本的溶解缓冲液组分

DNA样本的溶解缓冲液组分（包括溶解缓冲液的化学试剂配方、pH值等）应符合建库试剂的要求。

### 5.2 建库方法要求

#### 5.2.1 加接头的建库方法

主要包括先进行变性和转化处理、再进行单链模板加接头的单链加接头建库方法，以及先进行双链模板加接头、再进行变性和转化处理的双链加接头建库方法。

#### 5.2.2 转化方法

主要包括C to T转化和mC to T转化两种方式。其中C to T转化方式主要包括重亚硫酸盐处理和酶促转化两种方法，mC to T转化方式主要指TET辅助吡啶硼烷转化的方法。

### 5.3 文库质量要求

#### 5.3.1 文库质量

##### 5.3.1.1 文库的片段大小与分布

应符合建库试剂及测序仪说明书的要求，同时不应存在接头或引物二聚体污染。

#### 5.3.1.2 文库的浓度、体积和总量

应符合建库试剂及测序仪说明书的要求。

### 5.4 测序质量要求

#### 5.4.1 测序读长和测序类型、测序策略

根据人全基因组甲基化高通量测序应用要求及文库主带大小，选择测序读长和测序类型。通常为PE100和PE150。

测序策略应符合建库试剂及测序仪说明书的要求。通常采用重亚硫酸盐和/或酶促转化进行全基因组C to T转化的甲基化测序文库需要添加一定比例的碱基平衡文库，采用TET辅助吡啶硼烷转化进行全基因组mC to T转化的甲基化测序文库无需添加碱基平衡文库。此外，兼容碱基不平衡文库测序的测序仪也无需添加碱基平衡文库。

#### 5.4.2 测序原始碱基数

单个文库的测序原始碱基数应符合特定建库试剂及测序仪人全基因组甲基化高通量测序应用的要求。

#### 5.4.3 碱基识别质量百分比

单张测序芯片或单个测序管道的碱基识别质量百分比和单个样本的碱基识别质量百分比应符合测序仪说明书的要求。通常Q30不低于85%。

#### 5.4.4 标签拆分率

当采用无标签文库进行测序时，无需衡量标签拆分率。

当采用单标签文库或双标签文库进行测序时，标签拆分率应符合测序仪说明书的要求。

### 5.5 单样本测序数据质量要求

#### 5.5.1 碱基含量

应建立数据过滤及比对后的单样本人全基因组甲基化高通量测序数据碱基含量要求。

如：采用重亚硫酸盐和/或酶促转化进行全基因组C to T转化的人全基因组甲基化测序单样本文库的GC含量在20%~23%的范围，采用TET辅助吡啶硼烷转化进行全基因组mC to T转化的人全基因组甲基化测序单样本文库的GC含量在39%~43%的范围。此外，Read1的GC含量应与Read2的GC含量保持一致水平，通常Read1与Read2的GC含量差值的绝对值不大于1%。

#### 5.5.2 基因组唯一比对率

应建立数据过滤及比对后的单样本人全基因组甲基化高通量测序数据基因组唯一比对率要求。通常，常规样本类型（如来源于细胞系、全血、白膜层、组织等）PE测序的基因组唯一比对率不小于85%。

### 5.5.3 有效测序深度

有效测序深度应符合人全基因组甲基化高通量测序的要求，通常不小于 30×。

### 5.5.4 10×测序覆盖率

单样本的10×测序覆盖率应符合人全基因组甲基化高通量测序的要求，通常不小于90%。

目标检测区域的10×测序覆盖率不小于90%。

### 5.5.5 重复测序片段比率

单个文库的重复测序片段比率应不高于制造商规定的要求。通常不高于30%。

### 5.5.6 转化率

通常会将人工合成的全非甲基化标准品作为阴性参照、将人工合成的全甲基化标准品作为阳性对照掺入到待检样本中做内参。内参的转化率应符合人全基因组甲基化高通量测序的要求。

如：采用重亚硫酸盐和/或酶促转化进行全基因组C to T转化的人全基因组甲基化测序单样本文库的阴性参照的C to T转化率应不小于99%、阳性参照的C to T转化率应不大于1%；采用TET辅助吡啶硼烷转化进行全基因组mC to T转化的人全基因组甲基化测序单样本文库的阴性参照的mC to T转化率应不大于1%、阳性参照的mC to T转化率应不小于94%。

### 5.5.7 胞嘧啶平均测序深度

胞嘧啶平均测序深度应不低于人全基因组甲基化高通量测序的要求。通常不小于10×。

### 5.5.8 胞嘧啶测序覆盖度

胞嘧啶测序覆盖度应不低于人全基因组甲基化高通量测序要求。通常不小于90%。

## 5.6 单样本甲基化检测质量要求

### 5.6.1 位点检测质量要求

#### 5.6.1.1 甲基化位点数

当使用人全基因组甲基化标准物质评估时，单样本甲基化位点个数应符合标准物质甲基化位点数的阈值范围。

#### 5.6.1.2 甲基化率

当使用人全基因组甲基化标准物质评估时，单样本甲基化率应符合标准物质甲基化率的阈值范围。

#### 5.6.1.3 甲基化位点准确度和灵敏度

当使用人全基因组甲基化标准物质评估时，甲基化位点的准确度应不小于95%，灵敏度应不小于90%。

#### 5.6.1.4 2个技术重复的位点甲基化率一致性

同一管标准物质或同一管DNA样本的2个技术（从建库到测序）重复的位点甲基化率的皮尔森相关系数应不小于0.95。

## 5.6.2 特定区域检测质量要求

### 5.6.2.1 特定区域甲基化率一致性

当使用人全基因组甲基化标准物质评估时，单个人全基因组甲基化高通量测序数据在特定区域内位点甲基化率与标准物质的位点甲基化率的皮尔森相关系数应不小于0.95。

### 5.6.2.2 特定位点甲基化率一致性（如适用）

当使用特定位点不同甲基化率梯度的人全基因组甲基化标准物质评估时，单样本特定位点的平均测序深度不小于10×时，其甲基化率应符合标准物质特定位点甲基化率的阈值范围。

## 6 评价方法

### 6.1 样本准备

采用人基因组DNA，判断样本质量是否符合5.1的要求

### 6.2 文库制备

采用5.2的建库方法，按照建库试剂说明书进行人全基因组高通量甲基化测序文库构建，判断文库质量是否符合5.3的要求。

### 6.3 高通量测序

按照建库试剂及测序仪说明书进行文库上机测序，判断下机原始数据的质量是否符合5.4的要求。

### 6.4 单样本测序数据

单样本原始测序数据进行过滤、比对后，判断数据质量是否符合5.5的要求。

### 6.5 单样本位点甲基化检测

单样本过滤、比对、甲基化分析处理后，判断甲基化检测质量是否符合5.6的要求。

### A.1 概述

本附录提供了第5章中适用的人基因组DNA实物标准信息：人基因组标准品、人源B淋巴细胞全基因组DNA序列标准物质。

### A.2 用途

本文件中使用的人基因组标准品原料为健康人外周血白细胞构建永生细胞系提取的基因组DNA样本。

### A.3 规格和组成

人基因组标准品样本规格和组成见表A.1。

表A.1 人基因组标准品样本规格和组成

编号	名称	DNA 来源	细胞来源	人参考基因组序列来源
1	人基因组标准品 DNA 样本	NIFDC-HJ 细胞系	正常男性外周血白细胞构建永生细胞系	人参考基因组 hs37d5

### A.4 注意事项

现行国家权威机构发布的标准品、标准物质说明书可在分发单位的网站进行查询下载。说明书的部分内容会根据批次进行变更，以现行有效版本为准。

## 参 考 文 献

[1]GB/T 29859-2013 生物信息学术语

[2]GB/T 35890-2018 高通量测序数据序列格式规范

- [3] 陈铭. 生物信息学[M]. 3版. 北京: 科学出版社, 2018.
- [4] GA/T 1693-2020 法庭科学 DNA二代测序检验规范
- [5] YY/T 1723-2020 高通量基因测序仪
- [6] GB/T 40183-2021 DNA甲基化的测定 焦磷酸测序法
- [7] GB/T XXXXX-XXXX 人全基因组高通量测序数据质量评价方法