



# 中华人民共和国国家标准

GB/T 37872—2019

---

## 目标基因区域捕获质量评价通则

Guidelines for validation of next-generation target region sequencing

2019-08-30 发布

2019-08-30 实施

国家市场监督管理总局  
中国国家标准化管理委员会 发布



## 前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

本标准由国家标准物质研究中心提出并归口。

本标准起草单位：深圳华大生命科学研究院(原深圳华大基因研究院)、中国计量科学研究院、深圳华大智造科技有限公司、深圳华大基因科技有限公司、深圳华大临床检验中心有限公司、艾吉泰康生物科技(北京)有限公司。

本标准主要起草人：耿春雨、王晶、傅书锦、郝世杰、刘心、蒋慧、牛春艳、蔡万世、李雅乔、杜佳婷、李倩一、李岱怡、谢强、唐美芳、刘继龙、王瑞超。



# 目标基因区域捕获质量评价通则

## 1 范围

本标准规定了基于液相捕获技术的目标基因区域捕获质量评价的术语和定义、质量要求和评价方法。

本标准适用于应用高通量基因测序对人类基因组 DNA 样本进行目标基因区域捕获的质量评价。  
本标准不适用于单分子测序。

## 2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

GB/T 29859—2013 生物信息学术语

## 3 术语和定义、缩略语

### 3.1 术语和定义

GB/T 29859—2013 界定的以及下列术语和定义适用于本文件。为了便于使用,以下重复列出了 GB/T 29859—2013 中的某些术语和定义。

#### 3.1.1

##### **测序 sequencing**

测定氨基酸或者核苷酸序列的过程。

[GB/T 29859—2013,定义 2.4.13]

#### 3.1.2

##### **外显子 exon**

真核生物基因的一部分,在剪接后会被保留在成熟核糖核酸分子中的序列。

[GB/T 29859—2013,定义 2.2.8]

#### 3.1.3

##### **内含子 intron**

真核生物基因的一部分,在剪接后未被保留在成熟核糖核酸分子中的序列。

[GB/T 29859—2013,定义 2.2.20]

#### 3.1.4

##### **胚系突变 germline mutation**

遗传自父本、母本或者在胚胎形成时期产生的基因突变。

#### 3.1.5

##### **体细胞突变 somatic mutation**

发生于胚胎形成时期之后,只存在于特定组织部分细胞中的细胞特异性突变。

### 3.1.6

#### 目标基因区域捕获 target region capture

对一个或多个基因的核苷酸序列定制目标基因区域特异性探针,与基因组 DNA 进行杂交,并富集目标基因 DNA 片段的过程。

### 3.2 缩略语

下列缩略语适用于本文件。

DNA:脱氧核糖核酸(deoxyribonucleic acid)

## 4 质量要求

### 4.1 总则

目标基因区域捕获质量要求由测序统计后的平均测序深度、测序覆盖度、捕获特异性确定。

### 4.2 平均测序深度

胚系突变检测的目标基因区域平均测序深度应大于或等于 60 倍,体细胞突变检测(等位基因频率  $\geq 5\%$ )的目标基因区域平均测序深度应大于或等于 200 倍,体细胞突变检测( $1\% \leq$  等位基因频率  $< 5\%$ )的目标基因区域平均测序深度应大于或等于 500 倍。

### 4.3 测序覆盖度

胚系突变检测的测序覆盖度在 60 倍平均测序深度条件下应满足表 1 的要求,体细胞突变检测(等位基因频率  $\geq 5\%$ )的测序覆盖度在 200 倍平均测序深度条件下应满足表 2 的要求,体细胞突变检测( $1\% \leq$  等位基因频率  $< 5\%$ )的测序覆盖度在 500 倍平均测序深度条件下应满足表 3 的要求。

表 1 胚系突变检测的测序覆盖度要求

目标基因区域覆盖深度	测序覆盖度
$\geq 1$ 倍	$\geq 99\%$
$\geq 4$ 倍	$\geq 97\%$
$\geq 10$ 倍	$\geq 92\%$
$\geq 20$ 倍	$\geq 80\%$
$\geq 30$ 倍	$\geq 65\%$

表 2 体细胞突变检测的测序覆盖度要求(等位基因频率 $\geq 5\%$ )

目标基因区域覆盖深度	测序覆盖度
$\geq 1$ 倍	$\geq 99.1\%$
$\geq 4$ 倍	$\geq 98\%$
$\geq 10$ 倍	$\geq 97.5\%$
$\geq 20$ 倍	$\geq 97\%$
$\geq 100$ 倍	$\geq 75\%$

表 3 体细胞突变检测的测序覆盖度要求( $1\% \leq$  等位基因频率  $< 5\%$ )

目标基因区域覆盖深度	测序覆盖度
$\geq 1$ 倍	$\geq 99.3\%$
$\geq 4$ 倍	$\geq 99\%$
$\geq 10$ 倍	$\geq 98\%$
$\geq 20$ 倍	$\geq 97\%$
$\geq 100$ 倍	$\geq 90\%$

#### 4.4 捕获特异性

大于 10 Mb 的目标基因区域探针捕获特异性应不低于 45%，小于或等于 10 Mb 的目标基因区域探针捕获特异性应不低于 12%。

注 1：目标基因区域不包含侧翼序列区域。

注 2：按照比对的碱基数据量进行统计。

注 3：针对目标基因区域较为特殊的(如目标基因区域较小或包含基因组重复序列)可视重复区域长度对该指标进行调整。

## 5 评价方法

### 5.1 试验材料

采用人源基因组 DNA。

### 5.2 文库制备

根据高通量基因测序平台的文库长度、产量、浓度等要求,按照对应的目标基因区域捕获建库流程进行操作。

### 5.3 数据产出与分析

#### 5.3.1 高通量测序

制备完成的文库在高通量基因测序平台进行上机前处理和测序,得到高通量测序原始数据。

#### 5.3.2 数据过滤

根据不同的数据过滤参数,对原始数据进行过滤,去除未达到过滤参数要求的序列或者 N 碱基含量较多的序列。

#### 5.3.3 序列比对

使用局部比对算法按照匹配打分规则,将过滤后得到的序列与参考基因组进行比对,从而确定最佳比对位置。

注:本标准人参考基因组使用 hg19。

### 5.4 结果计算

#### 5.4.1 平均测序深度

基于序列比对后的数据进行重复序列去除,统计比对到目标基因区域的碱基数据量与目标基因区域总长度的比值,按照式(1)进行计算。

$$D = \frac{r}{l} \dots\dots\dots(1)$$

式中:

- $D$  —— 目标基因区域平均测序深度;
- $r$  —— 比对到目标基因区域的碱基数据量;
- $l$  —— 目标基因区域总长度。

#### 5.4.2 测序覆盖度

基于序列比对后的数据进行重复序列去除。

对胚系突变,分别统计目标基因区域内覆盖深度不低于 1 倍、4 倍、10 倍、20 倍和 30 倍的位点数量与目标基因区域总长度的百分比,按照式(2)进行计算。

对体细胞突变,分别统计目标基因区域内覆盖深度不低于 1 倍、4 倍、10 倍、20 倍和 100 倍的位点数量与目标基因区域总长度的百分比,按照式(2)进行计算。

$$c(X) = \frac{d(X)}{l} \times 100\% \dots\dots\dots(2)$$

式中:

- $c(X)$  —— 目标基因区域覆盖度( $\geq X$  倍);
- $d(X)$  —— 目标基因区域内覆盖深度不低于  $X$  倍的位点数量;
- $X$  —— 1 倍、4 倍、10 倍、20 倍、30 倍、100 倍的目标基因区域覆盖深度;
- $l$  —— 目标基因区域总长度。

#### 5.4.3 捕获特异性

基于序列比对后的数据进行重复序列去除,比对到目标基因区域的碱基数据量与比对到全基因组



区域的碱基数据量的比值,按照式(3)进行计算。

$$S = \frac{r}{R} \times 100\% \quad \dots\dots\dots(3)$$

式中:

$S$  —— 目标基因区域捕获特异性;

$r$  —— 比对到目标基因区域的碱基数据量;

$R$  —— 比对到全基因组区域的碱基数据量。

注: 目标基因区域不包括侧翼序列区域。





中 华 人 民 共 和 国  
国 家 标 准  
目标基因区域捕获质量评价通则

GB/T 37872—2019

\*

中国标准出版社出版发行  
北京市朝阳区和平里西街甲2号(100029)  
北京市西城区三里河北街16号(100045)

网址: [www.spc.org.cn](http://www.spc.org.cn)

服务热线: 400-168-0010

2019年7月第一版

\*

书号: 155066 · 1-62713

版权专有 侵权必究



GB/T 37872-2019