

# GSA词条注释

Version 2.1, 2018

---

<b>数据递交</b> .....	2
<b>实验</b> .....	2
实验基本信息 .....	2
文库信息.....	4
<b>测序反应</b> .....	6
测序反应基本信息 .....	6
测序反应数据信息 .....	7

## GSA Submission 数据递交

### \* Alias 别名

最终系统搜索或显示的 GSA 名称

### \* Data Released

选择在指定日期发布数据，日期格式为 (yyyy-mm-dd)

---

## Experiments 实验

### Meta Information 实验基本信息

#### \* Platform 测序平台

测序平台以及测序仪型号

测序平台	仪器型号
LS454	454 GS
	454 GS 20
	454 GS FLX
	454 GS FLX Titanium
	454 GS FLX+
	454 GS Junior
Capillary Technologies	AB 310 Genetic Analyzer
	AB 3130 Genetic Analyzer
	AB 3130xL Genetic Analyzer
	AB 3500 Genetic Analyzer
	AB 3500xL Genetic Analyzer
	AB 3730 Genetic Analyzer
	AB 3730xL Genetic Analyzer
ABI Solid	AB 5500 Genetic Analyzer
	AB 5500xl Genetic Analyzer
	AB 5500x-WI Genetic Analyzer
	AB SOLiD 3 Plus System
	AB SOLiD 4 System
	AB SOLiD 4hq System
	AB SOLiD PI System
	AB SOLiD System 1.0
	AB SOLiD System 2.0
	AB SOLiD System 3.0
BGISEq	BGISEQ-100
	BGISEQ-1000
	BGISEQ-500
CapitalBio Company	BioelectronSeq 4000

Bionano Genomics	BioNano IRYS
	BioNano SAPHYR
Complete Genomics	Complete Genomics
DAAN GENE	DA8600
Helicos BioSciences Corporation	Helicos HeliScope
HYK Genetic	HYK-PSTAR-IIA
Illumina	Illumina Genome Analyzer
	Illumina Genome Analyzer II
	Illumina Genome Analyzer IIx
	Illumina HiScanSQ
	Illumina HiSeq 1000
	Illumina HiSeq 1500
	Illumina HiSeq 2000
	Illumina HiSeq 2500
	Illumina HiSeq 3000
	Illumina HiSeq 4000
	Illumina HiSeq X Ten
	Illumina MiSeq
	Illumina MiniSeq
	Illumina Nextseq 500
	Illumina Nextseq 550
	Illumina iSeq 100
	Illumina NovaSeq 5000
Illumina NovaSeq 6000	
IonTorrent	Ion Torrent PGM
	Ion Torrent Proton
Oxford Nanopore	MinION
	GridION
Berry Genomics	NextSeq CN500
PacBio SMRT	PacBio RS
	PacBio RS II
	PacBio Sequel

**\*Alias 实验别名**

最终系统搜索或显示的 Experiment 名称

**\*Title 标题名称**

可以用来在搜索或显示中调用实验记录的短文本

**\*Project accession 所属项目序列号**

用于建立与 BioProject 间的关联关系

**\* Sample accession 所属样本序列号**

用于建立与 Bio Sample 间的关联关系

**\* Library Construction/Experiment design 实验设计和策略**

用于填写您的实验设计和策略的详细信息，可包括杂交和亲和捕获试剂选择等细节

**Library 文库信息**

用于描述文库正在进行测序的材料的来源和材料可能对测序结果产生影响的任何处理

**Library name 文库名**

文库的名称

**\* Strategy 策略**

文库的测序技术和策略

技术策略	描述信息
WGS	非 PCR 扩增全基因组的随机测序技术
WGA	全基因组随机测序技术
WES	外显子随机测序技术
WXS	外显子特定区域基因组随机测序技术
RNA-Seq	全转录组的随机测序技术
miRNA-Seq	小 RNA 的随机测序技术
WCS	全染色体随机测序技术
CLONE	基于基因组克隆的测序技术
POOLCLONE	混合克隆鸟枪法测序技术
AMPLICON	Overlapping PCR 或 RT-PCR 产物的测序技术
CLONEEND	克隆末端 (5'/3'或两者) 测序技术
FINISHING	终止法测序技术
ChIP-Seq	染色质免疫沉淀测序技术
MNase-Seq	MNase 消化后直接测序技术.
DNase-Hypersensitivity	超敏感位点测序技术
Bisulfite-Seq	亚硫酸氢钠测序技术
Tn-Seq	转座子插入位点测序技术
EST	cDNA 模板的单次测序技术
FL-cDNA	cDNA 模板的全长测序技术
CTS	级联标签测序技术
MRE-Seq	甲基化敏感限制性内切酶测序技术
MeDIP-Seq	甲基化 DNA 免疫沉淀测序技术
MBD-Seq	甲基化组分测序的直接测序技术
Synthetic-Long-Read	长读长测序技术
ATAC-seq	染色质开放性测序技术
ChIA-PET	ChIA-PET 配对末端标签测序技术
FAIRE-seq	借助有机溶剂甲醛对染色体中裸露的 DNA 进行固定，之后通过酚氯仿抽提获取裸露 DNA 的测序技术

Hi-C	染色体构象捕获测序技术，其中生物素标记的核苷酸结合在连接结点，使得能够选择性纯化嵌合 DNA 连接结点，然后进行深度测序
ncRNA-Seq	非编码 RNA 测序技术
RAD-Seq	限制性位点相关 DNA 序列测序技术
RIP-Seq	RNA 免疫沉淀物的直接测序技术(包括 CLIP-Seq、HITS-CLIP 和 PAR-CLIP)
SELEX	指数富集的配体系统进化技术
ssRNA-seq	链特异性测序
Targeted-Capture	靶向捕获测序技术
Tethered Chromatin Conformation Capture	染色质构象捕获测序技术
Other	其他未收录测序技术

**\* Source 样本来源**

文库的实验材料来源类型

样本来源	
GENOMIC	基因组
TRANSCRIPTOMIC	转录组
METATRANSCRIPTOMIC	代谢组
METAGENOMIC	宏基因组
SYNTHETIC	合成 DNA
VIRAL RNA	病毒 RNA
OTHER	其他，未指定的或未知的来源类型

**\*Selection 实验方式选择**

是否使用了任何方法来选择和/或富集被测序的材料

实验方式选择	描述信息
unspecified	非特异性筛选
RANDOM	随机
PCR	利用特定引物对材料进行筛选
RANDOM PCR	利用随机产生的引物对材料进行筛选
RT-PCR	用逆转录 PCR 方法筛选
HMPR	亚甲基化部分限制性消化筛选 (Hypo-methylated partial restriction digest)
MF	甲基筛选 (Methyl Filtrated)
CF-S	Cot-filtered 单拷贝/低拷贝基因组 DNA 方式筛选 (Cot-filtered single/low-copy genomic DNA)
CF-M	Cot-filtered 中度重复基因组 DNA 方式筛选 (Cot-filtered moderately repetitive genomic DNA)
CF-H	Cot-filtered 高重复性基因组 DNA 方式筛选 (Cot-filtered highly repetitive genomic DNA)

CF-T	Cot-filtered 单拷贝基因组 DNA 方式筛选 (Cot-filtered theoretical single-copy genomic DNA)
MDA	多重置换扩增 (multiple displacement amplification, MDA) 利用随机引物和等温扩增可以获得高保真的 DNA 大片段
MSLL	Methylation Spanning Linking Library
cDNA	cDNA 筛选
ChIP	染色质免疫沉淀技术 (Chromatin immunoprecipitation)
MNase	微球菌核酸酶 (MNase) 消化
DNase	脱氧核糖核酸酶 (MNase) 消化
Hybrid Selection	杂交体选择 (法)
Reduced Representation	通常由限制性片段大小选择产生, 包含可控数量的基因座以便于重新取样
Restriction Digest	限制性内切酶用于 DNA 分离
5-methylcytidine antibody	用抗 5-met 抗体选择甲基化 DNA 片段
MBD2 protein methyl-CpG binding domain	MBD2 蛋白甲基-CpG 结合域筛选
CAGE	帽分析基因表达筛选 (Cap-analysis gene expression)
RACE	末端的快速扩增 (Rapid Amplification of cDNA Ends)
size fractionation	Physical selection of size appropriate targets
Padlock probes capture method	环化寡核苷酸探针
Poly-A	polyA 富集筛选
other	其他, 未指定的或未知的筛选方式

### \*Layout 文库布局

测序片段读取配置相关信息

#### \*Insert size (bp)

插入片段长度

#### Nominal size (bp)

接头长度

#### Nominal standard deviation (bp)

接头标准偏差

## Run 测序反应

### General Information 测序反应基本信息

#### \* Alias 测序反应别名

最终系统搜索或显示的 Run 名称

#### \* Run data file type 数据类型

关于提交序列数据的支持格式，我们建议提交格式列表如下：

数据类型	文件扩展名	是否推荐
Fastq format	.fastq.gz .fq.gz .fastq.bz2 .fq.bz2	Yes
BAM format	.bam	Yes
HDF5 format	.bax.h5 .bas.h5	Yes
Reference_FASTA	.fasta.gz, .fa.gz	Yes

### Data Blocks 测序反应数据信息

#### Fastq format (as an example)

FASTQ 是一种存储了生物序列（通常是核酸序列）以及相应的质量评价的文本格式。序列和质量分数都用一个 ASCII 字符编码。

#### \* File Name for Forward 正向端文件名

数据文件名（全名，包括后缀名）必须是唯一的，不可包含空格、连字符、括号、句点、或正反斜杠，仅接受 gzip 和 bzip2 压缩格式，PacBio 系列测序仪可上传 tar 格式，不接受 7\_Zip 或 tar 压缩文件。

#### \* MD5 for Forward file 正向端文件 MD5 码

MD5 校验是一个 32 个字符的字母数字字符串。Mac 和 Linux 系统的用户，只需在本地运行命令行 "md5sum"(Linux) 或 "md5"(Mac OS)便可生成 MD5 码。Windows 用户必须下载第三程序才能生产 MD5 码。

#### \* File Name for Reverse 反向端文件名

数据文件名（全名，包括后缀名）必须是唯一的，不可包含空格、连字符、括号、句点、或正反斜杠，仅接受 gzip 和 bzip2 压缩格式，PacBio 系列测序仪可上传 tar 格式，不接受 7\_Zip 或 tar 压缩文件。

#### \* MD5 for Reverse file 反向端文件 MD5 码

MD5 校验是一个 32 个字符的字母数字字符串。Mac 和 Linux 系统的用户，只需在本地运行命令行 "md5sum"(Linux) 或 "md5"(Mac OS)便可生成 MD5 码。Windows 用户必须下载第三程序才能生产 MD5 码。

#### BAM format (as an example)

BAM 格式其实就是 SAM 格式的二进制格式，是一种有效方法存储和共享数据格式。因为占用存储

空间更小，主要用于存储校准信息，基本可以（经常做）存储读取。

## 1) \* Reference Assembly Name or Accession

### \* Assembly Name or Accession

参考序列名或序列号

### \* Web URL of the Reference Assembly

参考序列对应的网址链接

### \* File Name for bam

BAM 的文件必须可用读 SAMtools 读取。BAM 的文件名格式为 .bam

### \* MD5 for bam file

MD5 校验是一个 32 个字符的字母数字字符串。Mac 和 Linux 系统的用户，只需在本地运行命令"md5sum"(Linux) 或 "md5"(Mac OS)便可生成 MD5 码。Windows 用户必须下载第三方程序才能生产 MD5 码。

## 2) \* Local Assembly file

### \* select one reference file you have uploaded/submit new reference file

选择已提交过的参考序列/提交新的参考序列

### \* Reference file name 参考序列文件名

参考序列文件名称

### \* MD5 for reference file 参考序列文件 MD5 码

MD5 校验是一个 32 个字符的字母数字字符串。Mac 和 Linux 系统的用户，只需在本地运行命令"md5sum"(Linux) 或 "md5"(Mac OS)便可生成 MD5 码。Windows 用户必须下载第三方程序才能生产 MD5 码。

### \* File Name for bam BAM 文件名

提供 bam 文件名称

### \* MD5 for bam file bam 文件名 MD5 码

MD5 校验是一个 32 个字符的字母数字字符串。Mac 和 Linux 系统的用户，只需在本地运行命令"md5sum"(Linux) 或 "md5"(Mac OS)便可生成 MD5 码。Windows 用户必须下载第三方程序才能生产 MD5 码。