PLoB

首页 › Bioinformatics › 正文

samtools常用命令详解

samtools的说明文档：http://samtools.sourceforge.net/samtools.shtml

samtools是一个用于操作sam和bam文件的工具合集。包含有许多命令。以下是常用命令的介绍

1. view

view命令的主要功能是：将sam文件转换成bam文件；然后对bam文件进行各种操作，比如数据的排序(不属于本命令的功能)和提取(这些操作是对bam文件进行的，因而当输入为sam文件的时候，不能进行该操作)；最后将排序或提取得到的数据输出为bam或sam（默认的）格式。

bam文件优点：bam文件为二进制文件，占用的磁盘空间比sam文本文件小；利用bam二进制文件的运算速度快。

view命令中，对sam文件头部的输入(-t或-T）和输出(-h)是单独的一些参数来控制的。

```
01  Usage: samtools view [options] <in.bam>|<in.sam> [region1 [...]]
02  默认情况下不加 region，则是输出所有的 region.
03
04  Options: -b        output BAM
05                     默认下输出是 SAM 格式文件，该参数设置输出 BAM 格式
06           -h        print header for the SAM output
07                     默认下输出的 sam 格式文件不带 header，该参数设定输出sam文件时带
    header 信息
08           -H        print header only (no alignments)
09           -S        input is SAM
10                     默认下输入是 BAM 文件，若是输入是 SAM 文件，则最好加该参数，否则有
    时候会报错。
11           -u        uncompressed BAM output (force -b)
12                     该参数的使用需要有-b参数，能节约时间，但是需要更多磁盘空间。
13           -c        Instead of printing the alignments, only count them and print
    the
14                     total number. All filter options, such as '-f', '-F' and '-q'
    ,
15                     are taken into account.
16           -1        fast compression (force -b)
17           -x        output FLAG in HEX (samtools-C specific)
18           -X        output FLAG in string (samtools-C specific)
19           -c        print only the count of matching records
20           -L FILE   output alignments overlapping the input BED FILE [null]
21           -t FILE   list of reference names and lengths (force -S) [null]
22                     使用一个list文件来作为header的输入
23           -T FILE   reference sequence file (force -S) [null]
24                     使用序列fasta文件作为header的输入
25           -o FILE   output file name [stdout]
26           -R FILE   list of read groups to be outputted [null]
27           -f INT    required flag, 0 for unset [0]
28           -F INT    filtering flag, 0 for unset [0]
29                     Skip alignments with bits present in INT [0]
30                     数字4代表该序列没有比对到参考序列上
31                     数字8代表该序列的mate序列没有比对到参考序列上
32           -q INT    minimum mapping quality [0]
33           -l STR    only output reads in library STR [null]
34           -r STR    only output reads in read group STR [null]
35           -s FLOAT  fraction of templates to subsample; integer part as seed [-1]
36           -?        longer help
```

例子：

```
#将sam文件转换成bam文件
$ samtools view -bS abc.sam > abc.bam
$ samtools view -b -S abc.sam -o abc.bam
```

```
01   #提取比对到参考序列上的比对结果
02   $ samtools view -bF 4 abc.bam > abc.F.bam
03
04   #提取paired reads中两条reads都比对到参考序列上的比对结果，只需要把两个4+8的值12作为过
     滤参数即可
05   $ samtools view -bF 12 abc.bam > abc.F12.bam
06
07   #提取没有比对到参考序列上的比对结果
08   $ samtools view -bf 4 abc.bam > abc.f.bam
09
10   #提取bam文件中比对到caffold1上的比对结果，并保存到sam文件格式
11   $ samtools view abc.bam scaffold1 > scaffold1.sam
12
13   #提取scaffold1上能比对到30k到100k区域的比对结果
14   $ samtools view abc.bam scaffold1:30000-100000 > scaffold1_30k-100k.sam
15
16   #根据fasta文件，将 header 加入到 sam 或 bam 文件中
17   $ samtools view -T genome.fasta -h scaffold1.sam > scaffold1.h.sam
```

2. sort

sort对bam文件进行排序。

```
1   Usage: samtools sort [-n] [-m <maxMem>] <in.bam> <out.prefix>
2   -m 参数默认下是 500,000,000 即500M（不支持K, M, G等缩写）。对于处理大数据时，如果内存
    够用，则设置大点的值，以节约时间。
3   -n 设定排序方式按short reads的ID排序。默认下是按序列在fasta文件中的顺序（即header）
    和序列从左往右的位点排序。
```

例子：

$ samtools sort abc.bam abc.sort

$ samtools view abc.sort.bam | less -S

3.merge

将2个或2个以上的已经sort了的bam文件融合成一个bam文件。融合后的文件不需要则是已经sort过了的。

```
01   Usage:   samtools merge [-nr] [-h inh.sam] <out.bam> <in1.bam> <in2.bam>[...]
02
03   Options: -n         sort by read names
04            -r         attach RG tag (inferred from file names)
05            -u         uncompressed BAM output
06            -f         overwrite the output BAM if exist
07            -1         compress level 1
08            -R STR     merge file in the specified region STR [all]
09            -h FILE    copy the header in FILE to <out.bam> [in1.bam]
10
11   Note: Samtools' merge does not reconstruct the @RG dictionary in the header. Users
12       must provide the correct header with -h, or uses Picard which properly maintains
13       the header dictionary in merging.
```

4.index

必须对bam文件进行默认情况下的排序后，才能进行index。否则会报错。

建立索引后将产生后缀为.bai的文件，用于快速的随机处理。很多情况下需要有bai文件的存在，特别是显示序列比对情况下。比如samtool的tview命令就需要；gbrowse2显示reads的比对图形的时候也需要。

Usage: samtools index <in.bam> [out.index]

例子：

```
1   #以下两种命令结果一样
2   <span style="font-weight: inherit; font-style: inherit; color: #ff6600;">$
    samtools index abc.sort.bam</span>
3   $ samtools index abc.sort.bam abc.sort.bam.bai
```

5. faidx

对fasta文件建立索引,生成的索引文件以.fai后缀结尾。该命令也能依据索引文件快速提取fasta文件中的某一条（子）序列

```
1   Usage: samtools faidx <in.bam> [ [...]]
2
3   对基因组文件建立索引
```

```
4    $ samtools faidx genome.fasta
5    #生成了索引文件genome.fasta.fai,是一个文本文件，分成了5列。第一列是子序列的名称；第二
     列是子序列的长度；个人认为"第三列是序列所在的位置"，因为该数字从上往下逐渐变大，最后的
     数字是genome.fasta文件的大小；第4和5列不知是啥意思。于是通过此文件，可以定位子序列在
     fasta文件在磁盘上的存放位置，直接快速调出子序列。
6
7    #由于有索引文件，可以使用以下命令很快从基因组中提取到fasta格式的子序列
8    $ samtools faidx genome.fasta scffold_10 > scaffold_10.fasta
```

6. tview

tview能直观的显示出reads比对基因组的情况，和基因组浏览器有点类似。

```
01   Usage: samtools tview <aln.bam> [ref.fasta]
02
03   当给出参考基因组的时候，会在第一排显示参考基因组的序列，否则，第一排全用N表示。
04   按下 g ，则提示输入要到达基因组的某一个位点。例子"scaffold_10:1000"表示到达第
05   10号scaffold的第1000个碱基位点处。
06   使用H(左)J（上）K（下）L（右）移动显示界面。大写字母移动快，小写字母移动慢。
07   使用空格建向左快速移动（和 L 类似），使用Backspace键向左快速移动（和 H 类似）。
08   Ctrl+H 向左移动1kb碱基距离；  Ctrl+L 向右移动1kb碱基距离
09   可以用颜色标注比对质量，碱基质量，核苷酸等。30~40的碱基质量或比对质量使用白色表示；
10   20~30黄色；10~20绿色；0~10蓝色。
11   使用点号'.'切换显示碱基和点号；使用r切换显示read name等
12   还有很多其它的使用说明，具体按 ？ 键来查看。
```

7. flagstat

给出BAM文件的比对结果

```
01   Usage: samtools flagstat <in.bam>
02
03   $ samtools flagstat example.bam
04   11945742 + 0 in total (QC-passed reads + QC-failed reads)
05   #总共的reads数
06   0 + 0 duplicates
07   7536364 + 0 mapped (63.09%:-nan%)
08   #总体上reads的匹配率
09   11945742 + 0 paired in sequencing
10   #有多少reads是属于paired reads
11   5972871 + 0 read1
12   #reads1中的reads数
13   5972871 + 0 read2
14   #reads2中的reads数
15   6412042 + 0 properly paired (53.68%:-nan%)
16   #完美匹配的reads数：比对到同一条参考序列，并且两条reads之间的距离符合设置的阈值
17   6899708 + 0 with itself and mate mapped
18   #paired reads中两条都比对到参考序列上的reads数
19   636656 + 0 singletons (5.33%:-nan%)
20   #单独一条匹配到参考序列上的reads数，和上一个相加，则是总的匹配上的reads数。
21   469868 + 0 with mate mapped to a different chr
22   #paired reads中两条分别比对到两条不同的参考序列的reads数
23   243047 + 0 with mate mapped to a different chr (mapQ>=5)
```

#同上一个，只是其中比对质量>=5的reads的数量

7. depth

得到每个碱基位点的测序深度，并输出到标准输出。

```
1    Usage: bam2depth [-r reg] [-q baseQthres] [-Q mapQthres] [-b in.bed] <in1.bam>
     [...]
```

8. 其它有用的命令

reheader 替换bam文件的头

```
1    $ samtools reheader <in.header.sam> <in.bam>
```

cat 连接多个bam文件，适用于非sorted的bam文件

```
1    $ samtools cat [-h header.sam] [-o out.bam] <in1.bam> <in2.bam> [ ... ]
```

idxstats 统计一个表格，4列，分别为"序列名，序列长度，比对上的reads数，unmapped reads number"。第4列应该是paired reads中有一端能匹配到该scaffold上，而另外一端不匹配到任何scaffolds上的reads数。

```
1    $ samtools idxstats <aln.bam>
```

9. 将bam文件转换为fastq文件

有时候，我们需要提取出比对到一段参考序列的reads，进行小范围的分析，以利于debug等。这时需要将bam或sam文件转换为fastq格式。

该网站提供了一个bam转换为fastq的程序：http://www.hudsonalpha.org/gsl/information/softwar e/bam2fastq

```
1  $ wget http://www.hudsonalpha.org/gsl/static/software/bam2fastq-1.1.0.tgz
2  $ tar zxf bam2fastq-1.1.0.tgz
3  $ cd bam2fastq-1.1.0
4  $ make
5  $ ./bam2fastq <in.bam>
```

10. mpileup

samtools还有个非常重要的命令mpileup，以前为pileup。该命令用于生成bcf文件，再使用bcftools进行SNP和Indel的分析。bcftools是samtool中附带的软件，在samtools的安装文件夹中可以找到。

最常用的参数有2：

-f 来输入有索引文件的fasta参考序列；

-g 输出到bcf格式。用法和最简单的例子如下

```
1  Usage: samtools mpileup [-EBug] [-C capQcoef] [-r reg] [-f in.fa] [-l list] [-M
   capMapQ] [-Q minBaseQ] [-q minMapQ] in.bam [in2.bam [...]]
2
3  $ samtools mpileup -f genome.fasta abc.bam > abc.txt
4  $ samtools mpileup -gSDf genome.fasta abc.bam > abc.bcf
5  $ samtools mpileup -guSDf genome.fasta abc.bam | \
6   bcftools view -cvNg - > abc.vcf
```

mpileup不使用-u或-g参数时，则不生成二进制的bcf文件，而生成一个文本文件(输出到标准输出)。该文本文件统计了参考序列中每个碱基位点的比对情况；该文件每一行代表了参考序列中某一个碱基位点的比对结果。比如：

```
01  scaffold_1     2841    A      11       ,,,...,....      BHIGDGIJ?FF
02  scaffold_1     2842    C      12       ,$,,...,....^I.  CFGEGEGGCFF+
03  scaffold_1     2843    G      11       ,,,...,....      FDDDDCD?DD+
04  scaffold_1     2844    G      11       ,,,...,....      FA?AAAA<AA+
05  scaffold_1     2845    G      11       ,,,...,....      F656666166*
06  scaffold_1     2846    A      11       ,,,...,....      (1.1111)11*
07  scaffold_1     2847    A      11
    ,,+9acggtgaag.+9ACGGTGAAT.+9ACGGTGAAG.+9ACGGTGAAG,+9acggtgaag.+9ACGGTGAAG.+9ACGGTGAAG.+9ACGGTGAAG.+9ACGGTGAAG.+9ACGGTGAAG
    %.+....-..)
08  scaffold_1     2848    N      11       agGGGgGGGGG      !!$!!!!!!!!
09  scaffold_1     2849    A      11       c$,...,.....     !0000000000
10  scaffold_1     2850    A      10       ,...,....        353333333
```

mpileup生成的结果包含6行：参考序列名；位置；参考碱基；比对上的reads数；比对情况；比对上的碱基的质量。其中第5列比较复杂,解释如下：

1 '.' 代表与参考序列正链匹配。

2 ',' 代表与参考序列负链匹配。

3 'ATCGN' 代表在正链上的不匹配。

4 'atcgn' 代表在负链上的不匹配。

5 '*' 代表模糊碱基

6 '^' 代表匹配的碱基是一个read的开始；' ^'后面紧跟的ascii码减去33代表比对质量；这两个符号修饰的是后面的碱基，其后紧跟的碱基(.,ATCGatcgNn)代表该read的第一个碱基。

7 '$' 代表一个read的结束，该符号修饰的是其前面的碱基。

8 正则式' \+[0-9]+[ACGTNacgtn]+' 代表在该位点后插入的碱基；比如上例中在scaffold_1的2847后插入了9个长度的碱基acggtgaag。表明此处极可能是indel。

9 正则式' -[0-9]+[ACGTNacgtn]+' 代表在该位点后缺失的碱基；

pileup具体的参数如下：

```
01  #输入参数
02  -6 Assume the quality is in the Illumina 1.3+ encoding. -A Do not skip
     anomalous read pairs in variant calling.
03  -B Disable probabilistic realignment for the computation of base alignment
     quality (BAQ). BAQ is the Phred-scaled probability of a read base being
     misaligned. Applying this option greatly helps to reduce false SNPs caused by
     misalignments.
04  -b FILE List of input BAM files, one file per line [null]
05  -C INT Coefficient for downgrading mapping quality for reads containing
```

```
      excessive mismatches. Given a read with a phred-scaled probability q of being
      generated from the mapped position, the new mapping quality is about sqrt((INT-
      q)/INT)*INT. A zero value disables this functionality; if enabled, the
      recommended value for BWA is 50. [0]
06    -d INT At a position, read maximally INT reads per input BAM. [250]
07    -E Extended BAQ computation. This option helps sensitivity especially for MNPs,
      but may hurt specificity a little bit.
08    -f FILE The faidx-indexed reference file in the FASTA format. The file can be
      optionally compressed by razip. [null]
09    -l FILE BED or position list file containing a list of regions or sites where
      pileup or BCF should be generated [null]
10    -M INT          cap mapping quality at INT [60]
11    -q INT Minimum mapping quality for an alignment to be used [0]
12    -Q INT Minimum base quality for a base to be considered [13]
13    -r STR Only generate pileup in region STR [all sites]
14
15    输出参数
16    -D Output per-sample read depth (require -g/-u)
17    -g Compute genotype likelihoods and output them in the binary call format
      (BCF).
18    -S Output per-sample Phred-scaled strand bias P-value (require -g/-u)
19    -u Similar to -g except that the output is uncompressed BCF, which is preferred
      for piping.
20
21    Options for Genotype Likelihood Computation (for -g or -u):
22    -e INT Phred-scaled gap extension sequencing error probability. Reducing INT
      leads to longer indels. [20]
23    -h INT Coefficient for modeling homopolymer errors. Given an l-long homopolymer
      run, the sequencing error of an indel of size s is modeled as INT*s/l. [100]
24    -I Do not perform INDEL calling
25    -L INT Skip INDEL calling if the average per-sample depth is above INT. [250]
26    -o INT Phred-scaled gap open sequencing error probability. Reducing INT leads
      to more indel calls. [40]
27    -P STR Comma dilimited list of platforms (determined by @RG-PL) from which
      indel candidates are obtained. It is recommended to collect indel candidates
      from sequencing technologies that have low indel error rate such as ILLUMINA.
      [all]
```

11. 使用bcftools

bcftools和samtools类似，用于处理vcf(variant call format)文件和bcf(binary call format)文件。前者为文本文件，后者为其二进制文件。

bcftools使用简单，最主要的命令是view命令，其次还有index和cat等命令。index和cat命令和samtools中类似。此处主讲使用view命令来进行SNP和Indel calling。该命令的使用方法和例子为：

```
1   $ bcftools view [-AbFGNQSucgv] [-D seqDict] [-l listLoci] [-s listSample]
2           [-i gapSNPratio] [-t mutRate] [-p varThres] [-P prior]
3           [-1 nGroup1] [-d minFrac] [-U nPerm] [-X permThres]
4           [-T trioType] in.bcf [region]
5
6   $ bcftools view -cvNg abc.bcf > snp_indel.vcf
```

生成的结果文件为vcf格式，有10列，分别是：

1 参考序列名；

2 variant所在的left-most位置；

3 variant的ID（默认未设置，用'.'表示）；

4 参考序列的allele；

5 variant的allele(有多个alleles，则用','分隔);

6 variant/reference QUALity；

7 FILTers applied;

8 variant的信息，使用分号隔开；

9 FORMAT of the genotype fields, separated by colon (optional)；

10 SAMPLE genotypes and per-sample information (optional)。

例如：

```
1   scaffold_1      2847      .        A         AACGGTGAAG        194       .
    INDEL;DP=11;VDB=0.0401;AF1=1;AC1=2;DP4=0,0,8,3;MQ=35;FQ=-67.5    GT:PL:GQ
    1/1:235,33,0:63
2   scaffold_1      3908      .        G         A         111       .
    DP=13;VDB=0.0085;AF1=1;AC1=2;DP4=0,0,5,7;MQ=42;FQ=-63    GT:PL:GQ
    1/1:144,36,0:69
3   scaffold_1      4500      .        A         G         31.5      .
    DP=8;VDB=0.0034;AF1=1;AC1=2;DP4=0,0,1,3;MQ=42;FQ=-39    GT:PL:GQ
    1/1:64,12,0:21
4   scaffold_1      4581      .        TGGNGG    TGG       145       .
    INDEL;DP=8;VDB=0.0308;AF1=1;AC1=2;DP4=0,0,0,8;MQ=42;FQ=-58.5     GT:PL:GQ
```

```
      1/1:186,24,0:45
  5   scaffold_1     4644     .     G     A     195     .
      DP=21;VDB=0.0198;AF1=1;AC1=2;DP4=0,0,10,10;MQ=42;FQ=-87 GT:PL:GQ
      1/1:228,60,0:99
  6   scaffold_1     4827     .     NACAAAGA     NA     4.42     .
      INDEL;DP=1;AF1=1;AC1=2;DP4=0,0,1,0;MQ=40;FQ=-37.5       GT:PL:GQ
      0/1:40,3,0:3
  7   scaffold_1     4854     .     A     G     48     .
      DP=6;VDB=0.0085;AF1=1;AC1=2;DP4=0,0,2,1;MQ=41;FQ=-36    GT:PL:GQ
      1/1:80,9,0:16
  8   scaffold_1     5120     .     A     G     85     .
      DP=8;VDB=0.0355;AF1=1;AC1=2;DP4=0,0,5,3;MQ=42;FQ=-51    GT:PL:GQ
      1/1:118,24,0:45
```

第8列中显示了对variants的信息描述，比较重要，其中的 Tag 的描述如下：

```
01  Tag    Format    Description
02  AF1    double    Max-likelihood estimate of the site allele frequency (AF) of the
    first ALT allele
03  DP int    Raw read depth (without quality filtering)
04  DP4    int[4]    # high-quality reference forward bases, ref reverse, alternate
    for and alt rev bases
05  FQ int    Consensus quality. Positive: sample genotypes different; negative:
    otherwise
06  MQ int    Root-Mean-Square mapping quality of covering reads
07  PC2    int[2]    Phred probability of AF in group1 samples being larger
    (,smaller) than in group2
08  PCHI2 double    Posterior weighted chi^2 P-value between group1 and group2
    samples
09  PV4    double[4]    P-value for strand bias, baseQ bias, mapQ bias and tail
    distance bias
10  QCHI2 int    Phred-scaled PCHI2
11  RP int    # permutations yielding a smaller PCHI2
12  CLR    int    Phred log ratio of genotype likelihoods with and without the
    trio/pair constraint
13  UGT    string    Most probable genotype configuration without the trio constraint
14  CGT    string    Most probable configuration with the trio constraint
```

bcftools view 的具体参数如下：

```
01  Input/Output Options:
02  -A Retain all possible alternate alleles at variant sites. By default, the view
    command discards unlikely alleles.
03  -b Output in the BCF format. The default is VCF.
04  -D FILE Sequence dictionary (list of chromosome names) for VCF->BCF conversion
    [null]
05  -F Indicate PL is generated by r921 or before (ordering is different).
06  -G Suppress all individual genotype information.
07  -l FILE List of sites at which information are outputted [all sites]
08  -N Skip sites where the REF field is not A/C/G/T
09  -Q Output the QCALL likelihood format
10  -s FILE List of samples to use. The first column in the input gives the sample
    names and the second gives the ploidy, which can only be 1 or 2. When the 2nd
    column is absent, the sample ploidy is assumed to be 2. In the output, the
    ordering of samples will be identical to the one in FILE. [null]
11  -S The input is VCF instead of BCF.
12  -u Uncompressed BCF output (force -b).
13
14  Consensus/Variant Calling Options:
15  -c Call variants using Bayesian inference. This option automatically invokes
    option -e.
16  -d FLOAT When -v is in use, skip loci where the fraction of samples covered by
    reads is below FLOAT. [0]
17          当有多个sample用于variants calling时，比如多个转录组数据或多个重测序
18          数据需要比对到参考基因组上，设置该值，表明至少有该<float 0~1>比例的
19          samples在该位点都有覆盖才计算入variant.所以对于只有一个sample的情况
20          下，该值设置在0~1之间没有意义，大于1则得不到任何结果。
21  -e Perform max-likelihood inference only, including estimating the site allele
    frequency, testing Hardy-Weinberg equlibrium and testing associations with LRT.
22  -g Call per-sample genotypes at variant sites (force -c)
23  -i FLOAT Ratio of INDEL-to-SNP mutation rate [0.15]
24  -p FLOAT A site is considered to be a variant if P(ref|D)
25  -t FLOAT Scaled muttion rate for variant calling [0.001]
26  -T STR Enable pair/trio calling. For trio calling, option -s is usually needed
    to be applied to configure the trio members and their ordering. In the file
    supplied to the option -s, the first sample must be the child, the second the
    father and the third the mother. The valid values of STR are 'pair',
    'trioauto', 'trioxd' and 'trioxs', where 'pair' calls differences between two
    input samples, and 'trioxd' ('trioxs') specifies that the input is from the X
    chromosome non-PAR regions and the child is a female (male). [null]
27  -v Output variant sites only (force -c)
28
29  Contrast Calling and Association Test Options:
30  -1 INT    Number of group-1 samples. This option is used for dividing the
    samples into two groups for contrast SNP calling or association test. When this
    option is in use, the following VCF INFO will be outputted: PC2, PCHI2 and
    QCHI2. [0]
31  -U INT    Number of permutations for association test (effective only with -1)
    [0]
32  -X FLOAT Only perform permutations for P(chi^2)
```

使用bcftools得到variant calling结果后。需要对结果再次进行过滤。主要依据比对结果中第8列信息。其中的 DP4 一行尤为重要，提供了4个数据：1 比对结果和正链一致的reads数、2 比对结果和负链一致的reads数、3 比对结果在正链的variant上的reads数、4 比对结果在负链的variant上的reads数。可以设定（value3 + value4）大于某一阈值，才算是variant。比如：

```
1  $ perl -ne 'print $_ if /DP4=(\d+),(\d+),(\d+),(\d+)/ && ($3+$4)>=10 &&
   ($3+$4)/($1+$2+$3+$4)>=0.8' snp_indel.vcf > snp_indel.final.vcf
```

12. samtools rmdup

NGS上机测序前需要进行PCR一步，使一个模板扩增出一簇，从而在上机测序的时候表现出为1个点，即一个reads。若一个模板扩增出了多簇，结果得到了多个reads，这些reads的坐标(coordinates)是相近的。在进行了reads比对后需要将这些由PCR duplicates获得的reads去掉，并只保留最高比对质量的read。使用rmdup命令即可完成.

```
1  Usage:   samtools rmdup [-sS]
2  -s 对single-end reads。默认情况下，只对paired-end reads
3  -S 将Paired-end reads作为single-end reads处理。
4
5  $ samtools input.sorted.bam output.bam
```

发布日期：2014年01月26日 所属分类：Bioinformatics

标签：samtools 使用教程 参数说明

```
:scription
mplate having multiple segments in sequencing
ch segment properly aligned according to the alig
gment unmapped
xt segment in the template unmapped
:Q being reverse complemented
:Q of the next segment in the template being rev
e first segment in the template
e last segment in the template
:condary alignment
t passing filters, such as platform/vendor quality
CR or optical duplicate
pplementary alignment
```

被忽视的Samtools参数