

# TruSight Tumor 170 v1.0 本地应用程序

## 用户指南

简介	3
安装 TruSight Tumor 170 本地应用程序	10
在 Docker 中运行 TruSight Tumor 170 本地应用程序	11
分析方法	13
分析输出	16
生成 Picard 指标	22
卸载 TruSight Tumor 170 本地应用程序	25
修订历史记录	27
技术协助	28



本文档及其内容归 Illumina, Inc. 及其附属公司（“Illumina”）所有，并且仅供其客户用于与本文档内所描述的产品用途相关的合同用途，不得用于其他任何目的。在未获得 Illumina 的事先书面同意的情况下，不得出于任何其他目的使用或分发本文档及其内容，和/或以其他任何方式对其进行传播、披露或复制。Illumina 不通过本文档向任何第三方授权其任何专利、商标、所有权或习惯法权利或类似权利。

必须由具备资质且受过相关培训的人员严格明确遵照本文档中的说明操作，以确保本文档中所述产品的使用适当且安全。在使用此类产品之前，相关人员必须通读并理解本文档中的所有内容。

未能完整阅读并明确遵守本文档中包含的所有说明可能会导致产品损坏、对用户或其他人员造成人身伤害以及对其他财产造成损害，并且将导致产品适用的保证失效。

对于由不当使用本文档中描述的产品（包括其部件或软件）引起的任何后果，ILLUMINA 概不承担任何责任。

© 2019 Illumina, Inc. 保留所有权利。

所有商标均为 Illumina, Inc. 或其各自所有者的财产。有关特定的商标信息，请参见 [www.illumina.com/company/legal.html](http://www.illumina.com/company/legal.html)。

## 简介

TruSight™ Tumor 170 本地应用程序是基于 Docker 的软件包，可对使用 TruSight Tumor 170 测序面板制备的 DNA 和 RNA 文库产生的测序片段进行分析。DNA 文库会输出小型变异和拷贝数变异的变异检出文件。RNA 文库会输出融合和剪接变异检出文件。结果包括高可信度融合和剪接变异的易于理解的输出，以及已发布融合结果。

应用程序支持基于包含碱基检出文件 (BCL) 的 Illumina® 测序运行文件夹开始二级分析工作流程，或者基于 FASTQ 文件夹中的 FASTQ 文件开始该工作流程。每次分析运行都需要使用 TruSight Tumor 170 样品表。

本文档概述了有关 TruSight Tumor 170 本地应用程序工作流程、计算要求、安装、分析软件、分析方法、分析输出和可选 Picard 指标生成工作流程及指导。

## 工作流程要求

- ▶ 具备 Linux 操作系统和 Docker 软件容器平台的基本知识
- ▶ 与使用 Illumina TruSight Tumor 170 实验分析方法制备的文库兼容
- ▶ 与 NextSeq™ 或 HiSeq™ 2500 Rapid Run 系统兼容
  - ▶ NextSeq:
    - ▶ 如果只对 DNA 文库测序，则每次运行的样品不要超过 8 个。
    - ▶ 如果只对 RNA 文库测序，则每次运行的样品不要超过 16 个。
    - ▶ 如果要对 DNA 和 RNA 两种文库测序，建议使用下面的样品组合：

DNA（每次运行的样品数）	RNA（每次运行的样品数）
8	1-8
7	1-9
6	1-10
5	1-11
4	1-12
3	1-13
2	4-14
1	8-15

- ▶ HiSeq 2500 Rapid Run:
  - ▶ 在每次运行中测序 12 个文库（6 个 DNA 和 6 个 RNA），以实现每个样品的最高覆盖率。
  - ▶ 如果只对 DNA 文库测序，则每次运行的样品不要超过 6 个。
  - ▶ 如果只对 RNA 文库测序，则每次运行的样品不要超过 12 个。
- ▶ 测序设置
  - ▶ 2 x 101（建议）
  - ▶ 双末端测序（必要）
  - ▶ 测序运行应该有三个包含以下标签集之一的文库，以提供足够的多样性；不建议使用三个以下的文库。此外，建议将 UP 标签与 RNA 文库搭配使用。
    - 标签集 1：UP01 (D702, D503)、UP02 (D707, D504)、UP03 (D717, D509)
    - 标签集 2：UP04 (D706, D510)、UP05 (D712, D513)、UP06 (D724, D515)
    - 标签集 3：UP07 (D705, D501)、UP08 (D713, D502)、UP09 (D715, D505)
    - 标签集 4：UP10 (D703, D506)、UP11 (D710, D517)、UP12 (D701, D518)
  - ▶ 有关其他规格，请参见 TruSight Tumor 170 数据表

- ▶ SampleSheet.csv 指导
  - ▶ 有关 TruSight Tumor 170 本地应用程序样品表 (SampleSheet.csv) 示例, 请参见 Illumina [支持网站](#)。
  - ▶ 必须将样品表命名为 SampleSheet.csv
  - ▶ “Manifest (清单)” 部分必须定义 PoolIDNA 和 PoolRNA
  - ▶ 必须定义 “Reads (片段)” 部分
  - ▶ 每次分析运行的最大样品数为 16
  - ▶ 在 “Data (数据)” 部分, 必填的列和列标题如下
    - ▶ Sample ID
    - ▶ Sample\_Name
    - ▶ Sample\_Well
    - ▶ I7\_Index\_ID
    - ▶ index
    - ▶ I5\_Index\_ID
    - ▶ index2
    - ▶ Manifest
  - ▶ 如未注明, “Data (数据)” 部分的字段为可选字段
    - ▶ 列标题 Sample\_ID 下的 Sample ID (必填)
    - ▶ Sample\_Name
    - ▶ Sample\_Plate
    - ▶ Sample\_Well
    - ▶ Index\_ID
    - ▶ 列标题 I7\_Index\_ID 下的 I7\_Index\_ID (输入为 BCL 时, 是必填字段, 输入为 FASTQ 时, 是可选字段)
    - ▶ index
    - ▶ 列标题 I5\_Index\_ID 下的 I5\_Index\_ID (输入为 BCL 时, 是必填字段, 输入为 FASTQ 时, 是可选字段)
    - ▶ index2
    - ▶ Sample Project
    - ▶ Description
    - ▶ 列标题 Manifest 下的 PoolIDNA 或 PoolRNA (区分大小写的字符串, 必填)
    - ▶ 列标题 Pair\_ID 下的 Pair ID (必填)
      - ▶ 为每个样品分配一个整数值 (从 1 开始)。从值相同的 DNA 和 RNA 样品得到的结果在合并变异报告中会加以合并。必须为每个样品分配一个对 ID 值, 即使该样品并不成对也需如此。
      - ▶ 必须以字母数字字符开头
  - ▶ 标签指导
    - ▶ 每个样品标签对必须唯一
    - ▶ 每个样品标签对必须与 TruSight Tumor 170 实验分析方法兼容
- ▶ 使用样品表模板在样品表中配置的样品 ID 必须满足以下标准:
  - ▶ 少于 60 个字符
  - ▶ 仅使用字母、数字、下划线和短破折号

- ▶ 样品 ID 不重复
- ▶ 下划线和短破折号前后必须为字母数字字符（不能由多个下划线或短破折号组成字符串）
- ▶ 样品名称不能以下划线或短破折号开头或结尾
- ▶ 对于 DNA 分析
  - ▶ 建议至少使用 60 M 片段（30 M 测序簇）
  - ▶ 需要使用片段长度相同的双末端测序样品
- ▶ 对于 RNA 分析
  - ▶ 建议最少使用 16 M 片段（8 M 测序簇）
  - ▶ 需要使用片段长度相同的双末端测序样品

## 更多资源

访问 Illumina 网站上的 TruSight Tumor 170 [支持页](#)，查看相应文档、下载软件、获取培训资源和 Illumina 兼容产品的的相关信息。

有关安装该产品的信息，请参见支持页上的[下载](#)。

以下文档可从 Illumina 网站下载。

资源	描述
<a href="#">TruSight Tumor 170 本地应用程序客户发行说明</a> (文档号 1000000036401)	提供有关当前版本的软件增强功能、限制和注意事项的信息。
<a href="#">SampleSheet.csv</a> (文档号 1000000038524)	用于运行分析的模板。
<a href="#">TruSight Tumor 170 本地应用程序客户软件安装作业指导书</a> (文档号 1000000041749)	提供有关安装 TruSight Tumor 170 本地应用程序软件的指导。

## 工作流程图

图 1 从 BCL 开始的 TruSight Tumor 170 软件工作流程

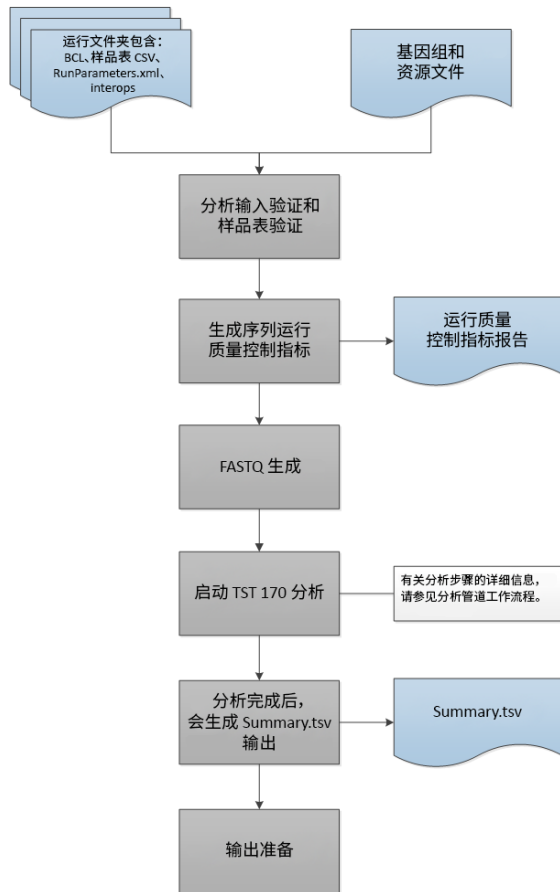


图 2 从 FASTQ 开始的 TruSight Tumor 170 软件工作流程

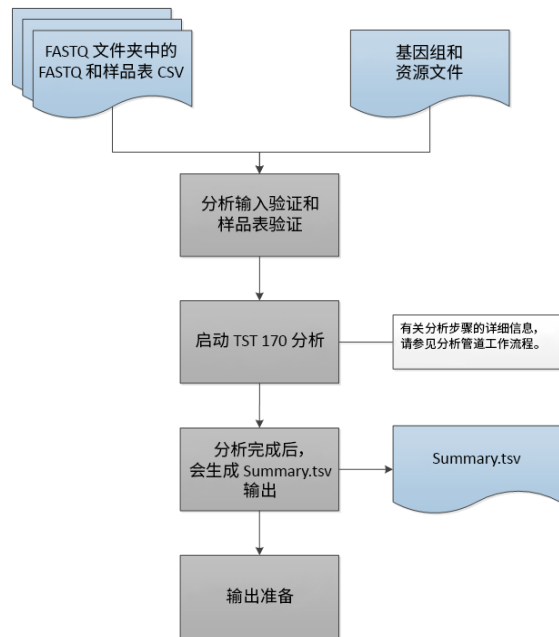


图 3 TruSight Tumor 170 DNA 工作流程

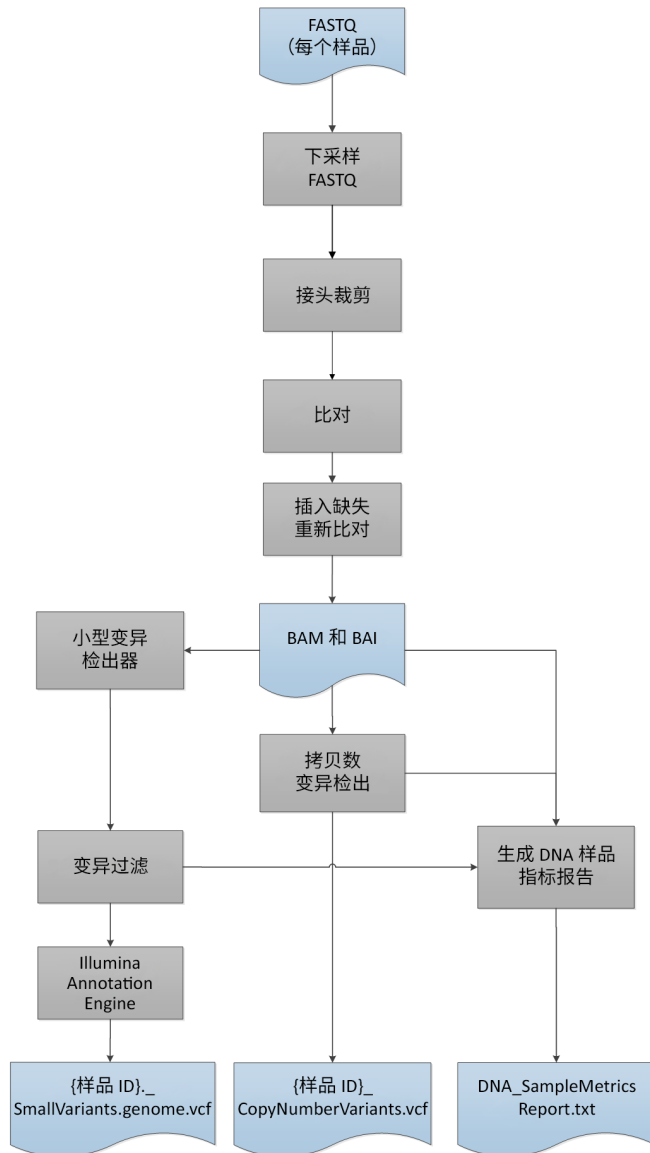
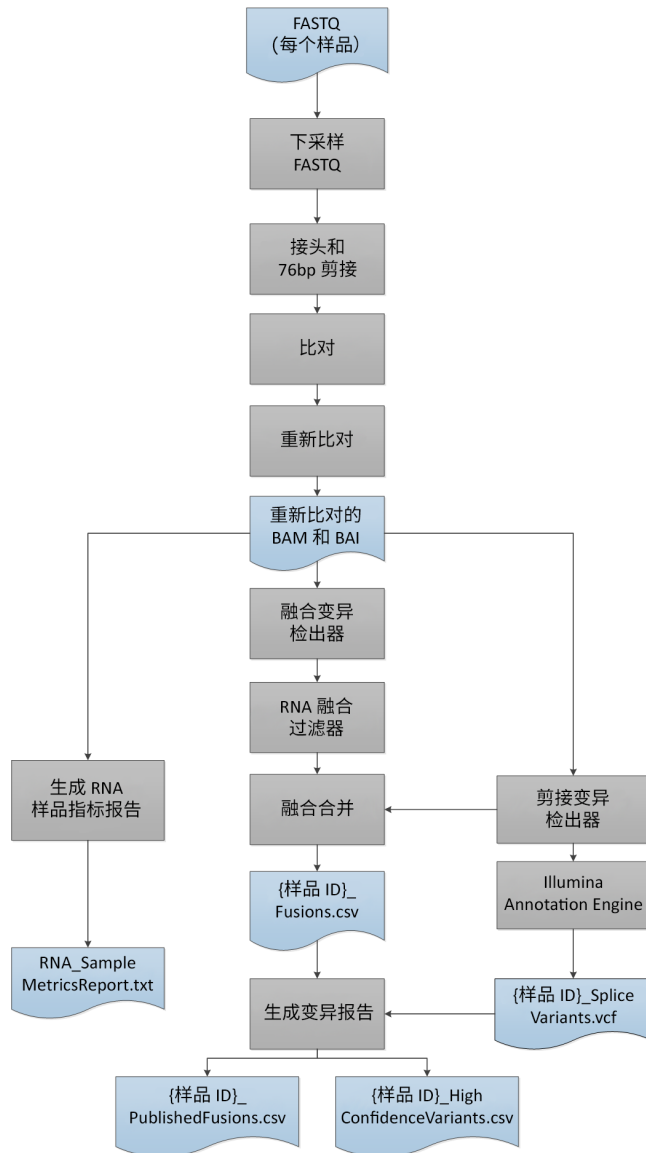




图 4 TruSight Tumor 170 RNA 工作流程



## 计算要求

下文说明了每次进行 TruSight Tumor 170 分析的最低操作要求。

## 软件

要使用 TruSight Tumor 170 本地应用程序，必须运行装有 Docker 1.12 或更高版本的 Linux CentOS 7.3 操作系统（内核版本 3.10.0-514.26.2.el7.x86\_64）。

## 硬件

要使用 TruSight Tumor 170 本地应用程序，硬件配置应如下所示：

- ▶ 最低：
  - ▶ 64 GB RAM

- ▶ 64 位 CPU，具有 16 个物理内核，32 个线程
- ▶ 1 TB 硬盘可用空间
- ▶ 64 GB IPC 共享内存
- ▶ 建议：
  - ▶ 64 GB RAM，2400 兆赫 DDR4 RDIMM ECC
  - ▶ 双 Intel Xeon 64 位 CPU；E5-2620 v4（发布于 2012 年第一季度）或更新。处理器必须支持 AVX 指令集。
  - ▶ 1 TB 硬盘可用空间
  - ▶ 硬件 RAID 控制器（1 GB 缓存）
  - ▶ 2 个 2 TB SATA 7.2k RPM 硬盘（企业级），采用 RAID 1 配置

## 安装 TruSight Tumor 170 本地应用程序

安装 TruSight Tumor 170 本地应用程序之前，请先完成先决任务。

### 先决条件

- ▶ 已安装并设置 Docker。有关详细信息，请访问 [www.docker.com](http://www.docker.com)。
- ▶ 具有 sudo 权限。如果您没有该权限，请联系系统管理员。
- ▶ 下载软件之前，您已查看并接受最终用户许可协议 (EULA)。您下载和使用该软件即表示您确认已查看并接受 EULA 的条款和条件。

在 [Illumina 支持网站](#) 上的 TruSight Tumor 170 本地应用程序支持页面中，提供了 TruSight Tumor 170 本地应用程序 README 和 EULA。

## 安装 TruSight Tumor 170 本地应用程序

- 1 从 [TruSight Tumor 170 支持页](#) 获取支持文档。
- 2 联系 Illumina 代表获得访问 Docker 映像和资源文件的权限。
- 3 使用下面的命令计算 md5 哈希值，以验证 Docker 映像文件：

```
md5sum {下载路径}/TruSight_Tumor_170_Local_App_{应用程序版本号}.tar
```

计算出的 md5sum 必须与 [支持页](#) 中提供的值一致。
- 4 使用下面的命令计算 md5 哈希值，以验证资源文件：

```
md5sum {下载路径}/TruSight_Tumor_170_Resources_{应用程序版本号}.tgz
```

计算出的 md5sum 必须与 [支持页](#) 中提供的值一致。
- 5 使用下面的命令解压缩资源文件：

```
tar -zxvf {下载路径}/TruSight_Tumor_170_Resources_{应用程序版本号}.tgz -C {基因组资源路径}
```
- 6 使用下面的命令加载 Docker 映像：

```
sudo docker load < {下载路径}/TruSight_Tumor_170_Local_App_{应用程序版本号}.tar
```
- 7 使用下面的命令列出 Docker 映像，以确认是否成功加载该映像：

```
sudo docker images
```
- 8 使用下面的命令检查您主机上的共享内存设置：

```
sudo cat /proc/sys/kernel/shmmax
```

值必须大于 63554432。

- 9 使用下面的命令检查 Docker 容器中的共享内存设置：

```
sudo docker run --entrypoint cat docker-oncology.dockerhub.illumina.com/tst170localapp:{应用程序版本号} /proc/sys/kernel/shmmax
```

值必须等于或大于 64 GB。

## 在 Docker 中运行 TruSight Tumor 170 本地应用程序

您可以基于包含 BCL 文件的 Illumina 测序运行文件夹运行工作流程分析，也可以基于 FASTQ 文件夹中的 FASTQ 文件来运行该分析。

运行 TruSight Tumor 170 本地应用程序时，请考虑以下事项：

- ▶ 等待所有正在运行的 TruSight Tumor 170 容器完成后再启动新分析。在启动 TruSight Tumor 170 本地应用程序前，运行以下命令来检查正在运行的容器：

```
sudo docker ps
```

- ▶ 命令行中的路径必须是绝对路径。
- ▶ 分析输出文件夹将保存到用户指定的位置。要运行多个分析，您必须为每个分析指定不同的文件夹位置，或在每次分析后清除输出文件夹中的所有文件。有关输出和报告的详细信息，请参见[分析输出（第 16 页）](#)。
- ▶ 如果发生“Access to the path... is denied. (访问...路径被拒)”错误，请使用下面的命令运行 docker 映像：

```
sudo docker run -t \  
-v /etc/localtime:/etc/localtime:Z \  
-v {RUN_PATH}:/data:Z \  
-v {GENOME_RESOURCE_PATH}:/genomes:Z \  
-v {ANALYSIS_OUTPUT_PATH}:/analysis:Z \  
docker-oncology.dockerhub.illumina.com/tst170localapp:{AppVersionNumber}
```

- ▶ 如果在 STAR 卸载基因组时发生错误，请检查是否存在挂起的共享内存，并视需要进行清除。

## 基于 BCL 文件

仅当使用 BCL 文件来运行 TruSight Tumor 170 本地应用程序时，才需要以下输入：

- ▶ <RUN\_PATH>：包含 BCL 和 SampleSheet.csv 的测序运行文件夹所在目录
- ▶ <GENOME\_RESOURCE\_PATH>：基因组和资源文件的位置
- ▶ <ANALYSIS\_OUTPUT\_PATH>：分析输出文件夹的位置

- 1 使用以下标准来运行 TruSight Tumor 170 本地应用程序：

```
sudo docker run -t \  
-v /etc/localtime:/etc/localtime \  
-v <RUN_PATH>:/data \  
-v <GENOME_RESOURCE_PATH>:/genomes \  
-v <ANALYSIS_OUTPUT_PATH>:/analysis \  
docker-oncology.dockerhub.illumina.com/tst170localapp:<AppVersionNumber>
```

分析输出文件夹中的分析日志和 Summary.tsv 输出会指出每个样品的分析运行及分析状态。有关详细信息，请参见[分析输出（第 16 页）](#)。

当分析完成时，分析日志会显示“Completed analysis (已完成分析)”。

## 基于 FASTQ 文件

使用 FASTQ 文件来运行 TruSight Tumor 170 本地应用程序时，需要以下输入：

- ▶ {FASTQ\_PATH}：包含 FASTQ 文件和 SampleSheet.csv 的文件夹所在目录
- ▶ {GENOME\_RESOURCE\_PATH}：基因组和资源文件的位置
- ▶ {ANALYSIS\_OUTPUT\_PATH}：分析输出文件夹的位置

若要使用 FASTQ 文件开始分析，FASTQ 文件必须与 TruSight Tumor 170 本地应用程序兼容。有关详细信息，请参见 [FASTQ 文件兼容性要求（第 12 页）](#)。

1 使用以下标准来运行 TruSight Tumor 170 本地应用程序：

```
sudo docker run -t \
-v /etc/localtime:/etc/localtime \
-v {FASTQ_PATH}:/data \
-v {GENOME_RESOURCE_PATH}:/genomes \
-v {ANALYSIS_OUTPUT_PATH}:/analysis \
docker-ontology.dockerhub.illumina.com/tst170localapp:{AppVersionNumber} -fastq
```

分析输出文件夹中的分析日志和 Summary.tsv 输出会指出每个样品的分析运行及分析状态。有关详细信息，请参见 [分析输出（第 16 页）](#)。

当分析完成时，分析日志会显示“Completed analysis（已完成分析）”。

## FASTQ 文件兼容性要求

从 FASTQ 文件开始时，FASTQ 文件必须满足以下要求。

- ▶ 使用 bcl2fastq 版本 2.17.1.14、2.18 或 2.19.1 生成。
- ▶ 使用以下接头序列：
  - ▶ Adapter AGATCGGAAGAGCACACGTCTGAACTCCAGTCA
  - ▶ AdapterRead2 AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
- ▶ 请勿将 FASTQ 文件与 UMI 配合使用。
- ▶ 运行 bcl2fastq 来生成 FASTQ 文件时，使用模板样品表 (SampleSheet.csv) 创建样品表。FASTQ 文件名包含 SampleSheet.csv 的 Sample\_ID 列中的唯一样品 ID。
- ▶ FASTQ 文件名必须包含样品 ID。默认情况下，bcl2fastq 会生成文件名中包含样品名（如果 SampleSheet.csv 中指定了样品名）的 FASTQ 文件。单独运行 bcl2fastq 时，会将 Sample\_Name 字段留空，以便 bcl2fastq 生成的 FASTQ 文件使用包含样品 ID 的预期文件名。
- ▶ 运行 bcl2fastq 时，务必使用最低硬件和软件规格中指定的内存设置。
- ▶ 将样品表的“Sample Project（样品项目）”列留空，以确保与 TruSight Tumor 170 本地应用程序工作流程兼容。
- ▶ 对于 NextSeq，请在运行 bcl2fastq 之前，先在样品表的“Data（数据）”部分内反向互补标签 2 序列。
- ▶ 如果样品计数为 4 个或以上，则将 --writing-threads 设置为 4；否则，将 --writing-threads 设置为等于样品计数。如果您指定的写入线程多于样品数，额外的线程不起作用，但可能会因位置切换而耗费时间。
- ▶ 使用以下实参并使用 bcl2fastq 2.17.1.14 和 2.18 生成 FASTQ 文件：
  - #{BCL2FASTQ\_PATH}：bcl2fastq 可执行文件的路径
  - #{OUTPUT\_DIRECTORY\_PATH}：输出生成的 fastqs 的绝对路径
  - #{RUN\_FOLDER\_PATH}：包含 BCL 的运行文件夹的绝对路径

`{SAMPLE_SHEET_PATH}`: 样品表的绝对路径 (通常包含在运行文件夹中)

```
{BCL2FASTQ_PATH} --output-dir '{OUTPUT_DIRECTORY_PATH}' --runfolder-dir '{RUN_
  FOLDER_PATH}' --sample-sheet '{SAMPLE_SHEET_PATH}' --ignore-missing-bcls --ignore-
  missing-filter --ignore-missing-positions --ignore-missing-controls --find-adapters-with-
  sliding-window --auto-set-to-zero-barcode-mismatches --adapter-stringency '0.9' --mask-
  short-adapter-reads '35' --minimum-trimmed-read-length '35' --loading-threads '4' --
  demultiplexing-threads '7' --processing-threads '32' --writing-threads '4'
```

- ▶ 使用以下实参并使用 `bcl2fastq 2.19.1` 生成 FASTQ:

`{BCL2FASTQ_PATH}`: `bcl2fastq` 可执行文件的路径

`{OUTPUT_DIRECTORY_PATH}`: 输出生成的 fastqs 的绝对路径

`{RUN_FOLDER_PATH}`: 包含 BCL 的运行文件夹的绝对路径

`{SAMPLE_SHEET_PATH}`: 样品表的绝对路径 (通常包含在运行文件夹中)

```
{BCL2FASTQ_PATH} --output-dir '{OUTPUT_DIRECTORY_PATH}' --runfolder-dir '{RUN_
  FOLDER_PATH}' --sample-sheet '{SAMPLE_SHEET_PATH}' --ignore-missing-positions --
  ignore-missing-controls --find-adapters-with-sliding-window --auto-set-to-zero-barcode-
  mismatches --adapter-stringency '0.9' --mask-short-adapter-reads '35' --minimum-
  trimmed-read-length '35' --loading-threads '4' --processing-threads '32' --writing-threads
  '4'
```

- ▶ 请勿合并 FASTQ 文件, TST170 分析管道预期每个样品都有单独的 FASTQ 文件 (FASTQ 文件数 = 泳道总数 × 每个泳道 2 个片段)。
  - ▶ 对于 NextSeq, 每个样品有 8 个 FASTQ 文件
  - ▶ 对于 HiSeq 2500 Rapid Run, 每个样品有 4 个 FASTQ 文件

## 分析方法

TruSight Tumor 170 工作流程会评估癌症相关基因中的 DNA 和 RNA 靶向变异。

## DNA 分析

### 输入验证

验证所需的分析输入是否存在及其格式是否正确。

### RunQC

仅当使用 BCL 文件运行 TruSight Tumor 170 本地应用程序来生成运行指标报告时, 才使用 RunQC。此报告使用测序运行文件夹中的 InterOp 文件提供运行指标和建议的值, 用于确定运行质量结果是否在可接受范围内。对于片段 1 和片段 2, 报告会提供  $\geq$  Q30 的碱基所占百分比, 这是一种质量分值 (Q-score) 测量机制。Q-score 是碱基检出出错概率的预测。

## FASTQ 生成

仅当使用 BCL 文件运行 TruSight Tumor 170 本地应用程序时, 才会生成 FASTQ 文件。该软件将对 BCL 文件进行逆多重分析, 并会生成 FASTQ 格式的中间分析文件。FASTQ 文件使用文本格式表示序列。其中包含每个样品的片段及关联的质量分值。将排除所有用于运行和簇且未通过过滤的对照品。每个 FASTQ 文件仅包含一个样品的片段, 该样品的名称包含在 FASTQ 文件名中。有关 FASTQ 文件要求的详细信息, 请参见[基于 FASTQ 文件 \(第 12 页\)](#)。

## 下采样

TruSight Tumor 170 分析管道会将每个样品下采样至 1 亿个 DNA 片段。此数字表示单一片段的总数（即所有泳道中 R1 + R2 的总和）。使用建议的测序配置或重混合时，样品的片段数可能会少于下采样限制。在这种情况下，FASTQ 文件将保持原样。

## 接头裁剪

对每个 DNA 文库对应的下采样 FASTQ 文件进行接头裁剪。

## 比对

Isaac Aligner 会根据人类 hg19 基因组对 DNA 片段执行比对。输入为经过接头裁剪的 FASTQ 文件，输出为 BAM 和 BAM 标签文件。在 Isaac 执行初始比对期间会提供 COSMIC 插入缺失作为先验。

有关 Isaac Aligner 的详细信息，请访问

<https://academic.oup.com/bioinformatics/article/29/16/2041/199472/Isaac-ultra-fast-whole-genome-secondary-analysis>。

## Isaac Aligner

Isaac Aligner 软件会执行以下步骤，使用长度为 36–150 碱基对且错误率很低的单端或双末端测序片段来比对 DNA 测序数据。

- ▶ **候选映射位置** — 使用基于 32-mer 种子的搜索确定全部相关候选映射位置。
- ▶ **映射选择** — 选择所有候选映射中的最佳映射。
- ▶ **比对分值** — 根据贝叶斯模型确定所选候选映射的比对分值。
- ▶ **比对输出** — 生成 BAM 文件格式的最终输出（文件中的内容已排序并标出了重复数据），以及重新比对的插入缺失及摘要文件。

## 插入缺失重新比对

重新比对检测到的插入缺失附近的片段，以删除比对物。

重新比对操作会识别所有片段中发现的疑似插入缺失，并根据这些靶点来重新比对片段。此步骤非常必要，因为在第一次比对过程中，有时会不正确地比对末端含插入缺失的片段。重新比对的 BAM 和 BAI 文件会保存到比对文件夹中。

## 小型变异检出

Pisces 会执行体细胞变异检出，以识别 DNA 样品中的低频变异。

有关详细信息，请参见 [github.com/Illumina/Pisces/wiki](https://github.com/Illumina/Pisces/wiki)。

## 变异过滤

DNA 变异过滤器软件可对小型变异检出 VCF 执行后处理，产生预期的背景修饰和质量分值调整结果。输入和输出均为 gVCF 文件。

## 注释

Illumina Annotation Engine 会对小型变异进行注释。输入和输出均为 VCF 文件。

## 拷贝数变异检出

CRAFT 拷贝数变异检出器会对实验分析方法内的靶扩增基因进行扩增、参考和缺失检出。这种检出根据计算得出的倍数变化值进行。倍数变化是指靶基因的标准化连锁区域计数中间值除以整个面板的连锁区域计数中间值。输入采用 BAM 格式，输出为 VCF 文件。缺失在 VCF 文件中会以 LowValidation 过滤进行标记，因为其仅使用 *in silico* 数据集进行过测试。

## DNA 样品指标报告

此组件用于执行 DNA 样品指标计算和报告。输入为 CRAFT 标准化 BinCount.txt 文件、BAM 文件和 gVCF 文件。输出为文本文件。

## RNA 分析

### 输入验证

验证所需的分析输入是否存在及其格式是否正确。

### RunQC

仅当使用 BCL 文件运行 TruSight Tumor 170 本地应用程序来生成运行指标报告时，才使用 RunQC。此报告使用测序运行文件夹中的 InterOp 文件提供运行指标和建议的值，用于确定运行质量结果是否在可接受范围内。对于片段 1 和片段 2，报告会提供  $\geq$  Q30 的碱基所占百分比，这是一种质量分值 (Q-score) 测量机制。Q-score 是碱基检出出错概率的预测。

## FASTQ 生成

仅当使用 BCL 文件运行 TruSight Tumor 170 本地应用程序时，才会生成 FASTQ 文件。该软件将对 BCL 文件进行逆多重分析，并会生成 FASTQ 格式的中间分析文件。FASTQ 文件使用文本格式表示序列。其中包含每个样品的片段及关联的质量分值。将排除所有用于运行和簇且未通过过滤的对照品。每个 FASTQ 文件仅包含一个样品的片段，该样品的名称包含在 FASTQ 文件名中。有关 FASTQ 文件要求的详细信息，请参见[基于 FASTQ 文件 \(第 12 页\)](#)。

## 下采样

TruSight Tumor 170 分析管道会将每个样品下采样至 3 亿个 RNA 片段。此数字表示单一片段的总数（即所有泳道中 R1 + R2 的总和）。使用建议的测序配置或重混合时，样品的片段数可能会少于下采样限制。在这种情况下，FASTQ 文件将保持原样。

## 接头裁剪

在每个 RNA 文库对应的下采样 FASTQ 文件上执行接头裁剪。随后将片段裁剪为 76 个碱基对，以便做进一步处理。

## 比对

STAR Aligner 会根据人类 hg19 基因组和基因转录比对 RNA 片段。比对步骤会识别要进行单端重新比对的候选，并通过 STAR 重新运行候选。对于每个文库会将两个 STAR 输出合并成一份 BAM 文件。输入为经过裁剪的 FASTQ 文件，输出为 BAM 文件。

## 融合检出

Manta<sup>1</sup> 会执行融合检出。输入为 BAM 文件，输出为 VCF 文件。

1. Chen X、Schulz-Trieglaff O、Shaw R 及其他人。Manta：针对种系和癌症测序应用快速检测结构变异和插入缺失。*Bioinformatics*（生物信息学）。2016;32:1220-1222。

## RNA 融合过滤

RNAFusionFilter 执行融合过滤。输入为 BAM 和 VCF 文件，输出为 CSV 文件。

## 融合合并

通过融合检出和 RNA 融合过滤确定的融合会与剪接变异检出期间执行的基因间检出相结合。此外，还会重新计算 RNA 融合过滤中的每个精确融合事件，以支持片段计数。输入为基因间剪接变异 TSV 文件、融合过滤 CSV 文件和 BAM 文件。输出为 CSV 文件。

## 剪接变异检出

使用 RNA 剪接变异检出软件执行剪接变异检出。输入为 BAM 文件和来自 STAR 的 SJ.out.tab 文件输出为 VCF 文件（此为最终输出）以及 TSV 文件（包含用于融合合并的基因间变异）。

## 注释

Illumina Annotation Engine 会对剪接变异进行注释。输入和输出均为 VCF 文件。

## RNA 样品指标报告

此组件用于执行 RNA 样品指标计算和报告。输入为 BAM 文件，输出为文本文件。

## 分析输出

分析运行完成时，TruSight Tumor 170 本地应用程序会在用户指定的位置生成一个分析输出文件夹，文件夹名为 TruSightTumor170\_Analysis\_{日期}\_{时间}。

要查看分析输出，请执行以下操作：

- 1 导航到包含该分析输出文件夹的目录。
- 2 打开该文件夹并选择要查看的文件。

## 文件夹

本部分介绍使用 BCL 或 FASTQ 工作流程执行分析过程中生成的各输出文件夹包含的内容。

### TruSightTumor170\_Analysis\_{日期}\_{时间}

- ▶ FASTQ — 包含原始样品 FASTQ 文件
- ▶ DNA\_IntermediateFiles
  - ▶ Alignment
    - ▶ 样品特定的 BAM 和 BAI 文件（iSAAC 输出）
    - ▶ 样品特定的重新比对的 BAM 和 BAI 文件（Hygea 输出）
    - ▶ DNA 探针 BED 文件（适用于 Picard）
    - ▶ DNA 靶点 BED 文件（适用于 Picard）
  - ▶ VariantCalling — DNA 区间文件（供 Pisces 使用）
  - ▶ Annotation — 包含样品特定的 JSON 文件
  - ▶ CNV



- ▶ **倍数变化文件** — 制表符分隔文件，包含 TruSight Tumor 170 面板所靶向的 59 个扩增基因中每个基因的基因名称、倍数变化和扩增的 Q-score。
- ▶ **标准化连锁区域计数文件** — 制表符分隔文件，包含用于扩增检出的每个目标基因组区域的标准化连锁区域计数，以及每个区域的元数据。
- ▶ **原始连锁区域计数文件** — 制表符分隔文件，包含用于扩增检出的每个目标基因组区域的原始连锁区域计数，以及每个区域的元数据。
- ▶ **DNA\_{样品 ID}** (每个样品)
  - ▶ {样品 ID}\_CopyNumberVariants.vcf
  - ▶ {样品 ID}\_SmallVariants.genome.vcf
- ▶ **DNA\_SampleMetricsReport.txt** 请参见 [DNA 样品指标报告 \(第 19 页\)](#)
- ▶ **Logs**
  - ▶ AnalysisLog-YYYYMMDD-HHMMSS .txt
  - ▶ **FASTQ\_Logs** — 包含以下文件：
    - ▶ 样品特定的 FASTQ 日志
    - ▶ FASTQ 组件日志
    - ▶ 中间 SampleSheet.csv (仅用于 BCL 的输入)
  - ▶ **DNA\_Logs** — 包含组件日志。
  - ▶ **RNA\_Logs** — 包含组件日志。
- ▶ **RNA\_IntermediateFiles**
  - ▶ **Alignment**
    - ▶ 样品特定的二次通过 BAM 和 BAI 文件 (嵌合体重新比对输出)
    - ▶ RNA 探针 BED 文件 (适用于 Picard)
    - ▶ RNA 靶点 BED 文件 (适用于 Picard)
  - ▶ **Annotation** — 包含样品特定的 JSON 文件
- ▶ **RNA\_{样品 ID}** (每个样品)
  - ▶ {样品 ID}\_Fusions.csv
  - ▶ {样品 ID}\_SpliceVariants.vcf
  - ▶ {样品 ID}\_HighConfidenceVariants.csv
  - ▶ {样品 ID}\_PublishedFusions.csv
- ▶ **RNA\_SampleMetricsReport.txt** — 请参见 [RNA 样品指标报告 \(第 22 页\)](#)
- ▶ **RunMetricsReport.txt** — 请参见 [运行指标选报告 \(第 17 页\)](#)
- ▶ **SampleSheet.csv** — 请参见 [文件名: SampleSheet.csv \(第 18 页\)](#)
- ▶ **Summary.tsv** — 请参见 [文件名: Summary.tsv \(第 18 页\)](#)

## 文件

本部分介绍分析过程中生成的摘要输出文件。

### 运行指标选报告

文件名: RunMetricsReport.txt

此报告包含测序运行指标和阈值指导。仅当输入是测序运行文件夹中的 BCL 时，才会生成 RunMetricsReport.txt。

运行指标	描述	预期计算值	阈值
片段 PF (%):	通过过滤器的簇的百分比	93.0	≥80.0
Q30+(R1): 片段 1 的 Q30 分值	片段 1 质量分值等于或高于 30 的碱基所占百分比	93.2	≥80.0
Q30+(R2): 片段 2 的 Q30 分值	片段 2 质量分值等于或高于 30 的碱基所占百分比	87.5	≥80.0

## Summary.tsv

文件名: Summary.tsv

系统会为每个分析生成一份制表符分隔的摘要报告，以显示每个已分析样品的分析状态。样品完成分析时，输出中的“Summary (摘要)”列会显示“Analysis Completed (分析已完成)”。发生错误时，“Summary (摘要)”列会显示失败或跳过的分析步骤。

## SampleSheet.csv

文件名: SampleSheet.csv

TruSight Tumor 170 特定样品表是逗号分隔文件，用于存储设置和分析测序实验所需的信息。文件包含样品列表以及标签序列，还包含要为分析中的每个样品执行的工作流程。有关填写样品表的信息，请参见[工作流程要求 \(第 3 页\)](#)。

## DNA 输出

### 文件

本部分介绍 DNA 样品分析过程中生成的输出文件。

### 小型变异基因组 VCF

文件名: {样品 ID}\_SmallVariants.genome.vcf

此文件包含 TruSight Tumor 170 DNA 面板中每个靶点坐标的小型变异检出。

执行变异检出和过滤器步骤期间，会应用以下变异检出过滤器。

过滤器名称	描述
低变异质量	过滤出变异质量分值 < 20 的项 (适用于参考和变异等位基因检出)。
低深度	过滤出变异检出的覆盖范围 < 100 且参考检出的覆盖范围 < 250 的项。
链偏向性	过滤出链偏向性分值 ≥ 0.5 的项 (仅适用于变异等位基因检出)。
低频变异	过滤出频率 < 0.026 的项 (仅适用于变异等位基因检出)。
插入缺失参考重复	过滤出重复 ≥ 8 的项 (仅适用于插入和缺失变异检出)。
调整后质量	过滤出基线质量分值 < 10 的项 (仅适用于变异等位基因检出)。
低变异支持	过滤出所含支持片段 < 7 的变异 (仅适用于变异等位基因检出)。

如果适用，请通过 Illumina Annotation Engine 并使用以下信息对变异检出进行注释：

- ▶ HGNC 基因
- ▶ 转录 (使用标准 Ensembl 转录)
- ▶ 外显子 (使用标准 Ensembl 转录)
- ▶ 结果 (使用标准 Ensembl 转录)
- ▶ HGVS 编码序列名称 (使用标准 Ensembl 转录)
- ▶ HGVS 蛋白质序列名称 (使用标准 Ensembl 转录)

► COSMIC ID

## 拷贝数变异 VCF

文件名: {Sample\_ID}\_CopyNumberVariants.vcf

包含 TruSight Tumor 170 面板所靶向的扩增基因 DNA 文库的 CNV 检出。CNV 检出会指出每个基因的倍数变化结果, 分为参考、缺失或扩增三类。VCF 的 QUAL 列中的值是 p-value 的 Phred 转换:

$$Q = -10 \log_{10}(p - value)$$

p-value 来自相对其余基因组的基因倍数变化之间的 t-test。Q-score 较高即表示 CNV 检出的可信度较高。

在 VCF 表示法中, <DUP> 表示检测到的倍数变化 (FC) 大于相应基因的预定义扩增临界值。<DEL> 表示检测到的倍数变化 (FC) 小于相应基因的预定义缺失临界值。不同基因的此临界值可能各不相同。

<DEL> 检出仅通过 *in silico* 数据集进行过验证。因此, 所有 <DEL> 检出在 VCF 中都具有 LowValidation 过滤器。

通过相应的倍数变化来报告每个拷贝数变异。此值由样品中基因的标准化片段深度与同一样品中二倍体区域的标准片段深度相比得出。给定肿瘤纯度, 便可根据报告的基因倍数变化推断出肿瘤中基因的拷贝数。

如果肿瘤纯度为 X%, 报告的倍数变化为 Y, 则拷贝数 *n* 可通过以下公式计算得出:

$$n = \frac{[(200 \times Y) - 2 \times (100 - X)]}{X}$$

例如, 在肿瘤纯度为 30% 的测试样品中, 倍数变化为 2.2x 的 MET 表示观察到 10 份 MET DNA 拷贝。

## DNA 样品指标报告

文件名: DNA\_SampleMetricsReport.txt

请查看此报告, 使用以下指标来评估 DNA 样品数据质量。

指标	描述	通过质量阈值	失败指示
Coverage_MAD	偏离每个 CNV 目标区域的标准化计数中间值的绝对偏离中间值。	≤0.2	拷贝数变异检出的可信度较低
Median_BinCount_CNV_Targets	每个 CNV 目标区域的原始片段计数 (按区域大小进行标准化) 的中间值。	≥1	拷贝数变异检出的可信度较低
Median_Insert_Size	从片段比对计算得出的片段大小中间值。	≥79	小型变异检出的可信度较低
PCT_ExonBases_100X	gVCF 中大于或等于 100x 覆盖范围的外显子碱基的百分比。	≥0.95	小型变异检出的可信度较低
PCT_Chimeric_Reads	嵌合体比对 (即片段的一半映射到一个位置, 另一半映射到一个非连续位置) 中的唯一片段数占映射的唯一片段总数的百分比。	不适用	不用于评估样品数据质量

## RNA 输出

### 文件

本部分介绍 RNA 样品分析过程中生成的顶层输出文件。

### 融合 CSV

文件名: {样品 ID}\_Fusions.csv

此文件包含 RNA 分析管道所识别的所有候选融合。此输出中将列出剪接变异检出器中的候选融合及相关支持信息，但并不表示具有高可信度。文件中的两个主要输出列 Filter 和 KeepFusion 用来说明候选融合。

下表介绍 Filter 列中的输出。如下所示，输出要么是可信度过滤器，要么只是参考信息。如果未触发以下任何过滤器，Filter 列将显示“PASS（通过）”。

过滤器	描述
Imprecise	低分辨率候选，非组装融合检出。（可信度过滤器）
RepeatOverlap	融合标记与重复区域重叠。仅用作非唯一映射融合候选的可信度过滤器，否则仅作参考信息之用。（可信度过滤器）
WeakBreakend	融合一侧的片段/比对证据薄弱。通常，此过滤表示片段仅由少数碱基对来重叠融合。或者，也可能表示同源（非唯一序列）太多。（可信度过滤器）
Homology	融合重叠群是另一个融合重叠群的子字符串。（可信度过滤器）
DuplicateContig	融合的两个重叠群相同。（可信度过滤器）
ContigIntragenic	半个重叠群的重新比对会生成映射到两端上（或在未注释的情况下，不超过 1kb）同一基因的比对。（可信度过滤器）
LowQ	融合支持片段（唯一）< 5（1600 万个片段中每 1000 万个片段便 +1）。（可信度过滤器）
LowDupReads	融合支持片段（重复）< 5。（可信度过滤器）
NonExonic	融合断点不属于外显子。（仅供参考）
LocalContigAlign	重叠群重新比较发现此重叠群的非融合比对。（仅供参考）
LowFusionRatio	少数有力证据片段与原株片段对比。（仅供参考）
NoReferenceReads	假定断点任一端上的所有片段都未标记为参考（结构正常）片段。（仅供参考）

当 RNAFusionFilter 分数  $\geq 0.45$  且未触发任何可信度过滤器时，输出的 KeepFusion 列值为 True。

有关每一列的详细信息，请参见输出中的标题。



#### 注意

使用 Microsoft Excel 查看此报告时，Excel 会将可转换为日期的基因（例如 MARCH1）自动转换为 dd-mm 格式（1-Mar）。

## 剪接变异 VCF

文件名：{Sample\_ID}\_SpliceVariants.vcf

包含 TruSight Tumor 170 面板所靶向的所有候选剪接变异（由 RNA 分析管道识别）。

可对每个变异检出应用以下过滤器。

过滤器名称	描述
LowQ	剪接变异分值小于通过质量分值的阈值 1。
PASS	剪接变异分值大于等于通过质量分值的阈值 1。

使用 Illumina Annotation Engine 注释每个剪接变异，在 VCF 中采集以下信息（如果可用）：

- ▶ HGNC 基因
- ▶ 转录
- ▶ 外显子
- ▶ 内含子
- ▶ 典型

## ▶ 后果

有关每一列的详细信息，请参见输出中的标题。

## 高可置信度变异

文件名：{样品 ID}\_HighConfidenceVariants.csv

包含高可置信度融合和剪接变异检出。

高可置信度融合是指使用通过过滤器靶向的融合（{样品 ID}\_Fusions.csv 中的 KeepFusions 列显示为 True）。总共有 55 个基因要进行融合检出。如果至少一个基因对的基因位于靶向基因列表中，则会报告融合。

▶ 报告的每个基因融合都包含以下信息：

- ▶ 断点 1
- ▶ 断点 2
- ▶ 基因 1 参考片段
- ▶ 基因 2 参考片段
- ▶ 分值

▶ “Gene Fusion（基因融合）”列中报告的基因 1/基因 2 融合的顺序不会指出融合的方向性。对于任何所需融合，请搜索两个方向的基因对。

▶ 分号表示具有重叠基因的融合。例如，基因 1;基因 2/基因 3 表示可能的融合事件基因 1/基因 3 和基因 2/基因 3。

▶ 融合变异依次按基因 1、基因 2、基因 1 断点和基因 2 断点升序排序。

高可置信度剪接变异检出为含通过过滤的面板（总共三个基因）所靶向的剪接变异。

▶ 报告的每个剪接变异都包含以下信息：

- ▶ 受影响的外显子
- ▶ 转录
- ▶ 断点起点
- ▶ 断点终点
- ▶ 剪接支持片段
- ▶ 参考片段
- ▶ 分值

▶ 剪接变异依次按基因、转录、断点起点和断点终点升序排序。

## 已发布的融合

文件名：{样品 ID}\_PublishedFusions.csv

包含来自 Mitelman 参考数据库（“Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer (2016).Mitelman F and Johansson B and Mertens F (Eds.)<http://cgap.nci.nih.gov/Chromosomes/Mitelman>”）的已发布的融合列表，这些融合为 TruSight Tumor 170 面板的靶向目标。

▶ 当 Observed=True 时，表示样品中检出具有或高或低可信度的融合。

▶ 当 Observed=False 时，表示样品中未检测到融合。

▶ 已发布的融合报告显示基因对中至少有一个 TruSight Tumor 170 面板所靶向的基因的 Mitelman 融合。

▶ 基因融合列不会指出融合的方向性。对于任何目标融合，请搜索基因 1/基因 2 及基因 2/基因 1。

对于每个融合，会采集以下信息：

▶ Mitelman ID - 一个融合可能会有多个 Mitelman ID，这是因为它们可能有多种临床关联或解释

- ▶ 断点 1: {样品 ID}\_Fusions.csv 中的基因 A 断点
- ▶ 断点 2: {样品 ID}\_Fusions.csv 中的基因 B 断点
- ▶ 融合支持片段: {样品 ID}\_Fusions.csv 中的 Alt Split + Alt Pair 值
- ▶ 基因 1 参考片段: {样品 ID}\_Fusions.csv 中的 Ref A Split + Ref A Pair 值
- ▶ 基因 2 参考片段: {样品 ID}\_Fusions.csv 中的 Ref B Split + Ref B Pair 值
- ▶ 分值: {样品 ID}\_Fusions.csv 中的分值
- ▶ 过滤器: 如果 Keep Fusion=True, 则为通过, 否则为 {样品 ID}\_Fusions.csv 中 “Filter (过滤器)” 列的值

融合依次按基因 1、基因 2、断点 1 和断点 2 排序。

## RNA 样品指标报告

文件名: RNA\_SampleMetricsReport.txt

请查看此报告, 通过评估以下指标来评定 RNA 样品数据质量。

指标	描述	评估的输出文件	质量阈值
Median_Insert_Size	从片段比对计算得出的片段大小中间值。	Fusions.csv、 HighConfidenceVariants.csv、 SpliceVariants.vcf、 PublishedFusions.csv	≥63
Median_CV_Coverage_1000x	所有基因中覆盖范围大于 1000 的目标区域覆盖 CV 值的中间值。	Fusions.csv、 HighConfidenceVariants.csv、 SpliceVariants.vcf、 PublishedFusions.csv	≤0.88
PCT_Chimeric_Reads	嵌合体比对 (即片段的一半映射到一个位置, 另一半映射到一个非连续位置) 中的唯一片段数占映射的唯一片段总数的百分比。	不用于评估样品数据质量	不适用
PCT_On_Target_Reads	映射到目标区域的片段占总 PF 片段数的百分比。	不用于评估样品数据质量	不适用

## 生成 Picard 指标

TruSight Tumor 170 本地应用程序输出可用于生成 Picard 指标。

有关 Picard HsMetrics 工具所生成 Picard 指标的说明, 请参见 [broadinstitute.github.io/picard/picard-metric-definitions.html#HsMetrics](http://broadinstitute.github.io/picard/picard-metric-definitions.html#HsMetrics)。

### 注意

计算出的 Picard 指标与通过 TruSight Tumor 170 本地应用程序所生成结果的质量之间并无关联。这些指标仅供您参考。

## 安装 Picard 并生成文件

请执行以下程序, 以安装支持环境, 并生成用于运行 Picard CollectHsMetrics 的文件。

### 注意

内存设置必须设为 -Xmx2g 来提升 Java 虚拟机 (JVM) 内存, 以在运行 Picard 工具时可容纳大型文件。

有关安装和设置适用于 Picard 的环境的详细信息, 请参见 [broadinstitute.github.io/picard/](http://broadinstitute.github.io/picard/)

1 使用以下命令行下载并安装 Picard 和支持环境:

```
#用于下载 Picard 2.9.4 可执行文件的 Linux 命令
```

```
wget https://github.com/broadinstitute/picard/releases/download/2.9.4/picard.jar
```

- 2 使用 Picard CreateSequenceDictionary 工具通过 hg19 genome.fa 参考基因组序列生成序列字典文件，其中：

- ▶ R= 参考基因组序列文件的绝对路径。必须指定 hg19 genome.fa 文件。
- ▶ O= 输出序列字典文件的路径。可以指定扩展名为 .dict 的文件名。默认使用输入参考的碱基名称及 .dict 扩展名。

例如

#生成参考序列的序列字典

```
java -Xmx2g -jar picard.jar CreateSequenceDictionary R=genomes/Homo_
sapiens/UCSC/hg19/Sequence/WholeGenomeFasta/genome.fa O=hg19.dict
```

- ▶ 有关 Picard CreateSequenceDictionary 工具的信息，请参见 [broadinstitute.github.io/picard/command-line-overview.html#CreateSequenceDictionary](https://broadinstitute.github.io/picard/command-line-overview.html#CreateSequenceDictionary)
- ▶ hg19 genome.fa 参考基因组序列文件包含在 TruSight\_Tumor\_170\_Resources\_{应用程序版本号}.tgz 中
  - ▶ 解压缩后，hg19 genome.fa 文件将位于基因组文件夹中：genomes/Homo\_sapiens/UCSC/hg19/Sequence/WholeGenomeFasta/genome.fa
  - ▶ 有关将文件解压缩到基因组文件夹的信息，请参见 [安装 TruSight Tumor 170 本地应用程序 \(第 10 页\)](#)。

- 3 检索靶点和探针 BED 文件的绝对目录路径，以与步骤 4 中的 INPUT 变量搭配使用。

- ▶ TruSight Tumor 170 靶点 BED 文件
  - ▶ 对于 DNA，位于分析输出文件夹下的 DNA\_IntermediateFiles/Alignment 中，文件名为 DNA\_PicardTarget.bed。
  - ▶ 对于 RNA，位于分析输出文件夹下的 RNA\_IntermediateFiles/Alignment 中，文件名为 RNA\_PicardTarget.bed。
- ▶ TruSight Tumor 170 探针 BED 文件
  - ▶ 对于 DNA，位于分析输出文件夹下的 DNA\_IntermediateFiles/Alignment 中，文件名为 DNA\_PicardProbe.bed。
  - ▶ 对于 RNA，位于分析输出文件夹下的 RNA\_IntermediateFiles/Alignment 中，文件名为 RNA\_PicardProbe.bed。

- 4 运行 Picard BedToIntervalList 工具，生成与 Picard CollectHsMetrics 兼容的区间文件（将用于生成 Picard 指标），其中：

- ▶ INPUT：靶点或探针 BED 文件的绝对路径。请参见步骤 3。
- ▶ SEQUENCE\_DICTIONARY：步骤 2 中生成的序列字典文件的绝对路径。
- ▶ OUTPUT：生成的输出区间文件的路径。

例如

#生成 DNA 靶点区间文件

```
java -Xmx2g -jar picard.jar BedToIntervalList INPUT=DNA_PicardTarget.bed SEQUENCE_
DICTIONARY=hg19.dict OUTPUT=DNA_PicardTarget.intervals
```

#生成 DNA 探针区间文件

```
java -Xmx2g -jar picard.jar BedToIntervalList INPUT=DNA_PicardProbe.bed SEQUENCE_
DICTIONARY=hg19.dict OUTPUT=DNA_PicardProbe.intervals
```

#生成 RNA 靶点区间文件

```
java -Xmx2g -jar picard.jar BedToIntervalList INPUT=RNA_PicardTarget.bed SEQUENCE_
DICTIONARY=hg19.dict OUTPUT=RNA_PicardTarget.intervals
```

### #生成 RNA 探针区间文件

```
java -Xmx2g -jar picard.jar BedToIntervalList INPUT=RNA_PicardProbe.bed SEQUENCE_
    DICTIONARY=hg19.dict OUTPUT=RNA_PicardTarget.intervals
```

## 运行 Picard CollectHsMetrics

使用单个命令行或运行 Bash 脚本都可生成 Picard 指标。

### 命令行

使用单个命令行运行 Picard CollectHsMetrics 可生成指标。

- 1 从 TruSight Tumor 170 实验分析方法应用程序输出中检索样品 BAM 文件的绝对目录路径，以便在步骤 4 中使用。
  - ▶ 对于 DNA 样品，Isaac 比对的 BAM 位于分析输出文件夹下的 DNA\_IntermediateFiles/Alignment 中，文件名为 {样品 ID}.bam。请勿使用重新比对的 BAM。
  - ▶ 对于 RNA 样品，STAR 比对的 BAM 位于分析输出文件夹下的 RNA\_IntermediateFiles/Alignment 中，文件名为 {样品 ID}.bam。
- 2 从基因组文件夹中检索 hg19 genome.fa 文件的绝对目录路径，以便在步骤 4 中使用。
  - ▶ hg19 genome.fa 文件位于基因组文件夹中：genomes/Homo\_sapiens/UCSC/hg19/Sequence/WholeGenomeFasta/genome.fa
- 3 检索步骤 4 安装 *Picard* 并生成文件 (第 22 页) 中生成的区间文件的绝对输出路径。
- 4 使用单个命令行运行 Picard CollectHsMetrics，其中：
  - ▶ {Picard CollectHsMetrics 命令} (例如 CollectHsMetrics for Picard v2.9.4)：根据 BAM 文件计算一组特定指标的 Picard 工具。
  - ▶ BI (或 BAIT\_INTERVALS)：包含探针相关信息的探针区间文件的绝对路径。请参见步骤 3。
  - ▶ TI (或 TARGET\_INTERVALS)：包含靶点区域相关信息的靶点区间文件的绝对路径。请参见步骤 3。
  - ▶ R：参考基因组序列文件的绝对路径。必须指定 hg19 genome.fa 文件。请参见步骤 2。
  - ▶ I：样品 BAM 文件的绝对路径。请参见步骤 1。
  - ▶ O：用于写入指标的输出文件路径。用户可指定文件名。
  - ▶ VALIDATION\_STRINGENCY=SILENT：使用该标志可将验证严格程度设置为 SILENT。当处理含有可变长度数据 (片段、质量、标记) 但无需解码的 BAM 文件时，使用此标志可提高性能。

#下面的命令行示例适用于使用 Picard 2.9.4 进行的 DNA 分析：

```
java -Xmx2g -jar picard.jar CollectHsMetrics BI=DNA_PicardProbe.intervals TI=DNA_
    PicardTarget.intervals R=genomes/Homo_
    sapiens/UCSC/hg19/Sequence/WholeGenomeFasta/genome.fa O=DNA-Picard-HsMetrics.txt
    I=DNA_IntermediateFiles/Alignment/dnasample.bam VALIDATION_STRINGENCY=SILENT
```

#下面的命令行示例适用于使用 Picard 2.9.4 进行的 RNA 分析：

```
java -Xmx2g -jar picard.jar CollectHsMetrics BI=RNA_PicardProbe.intervals TI=RNA_
    PicardTarget.intervals R=genomes/Homo_
    sapiens/UCSC/hg19/Sequence/WholeGenomeFasta/genome.fa O=RNA-Picard-HsMetrics.txt
    I=RNA_IntermediateFiles/Alignment/rnasample.bam VALIDATION_STRINGENCY=SILENT
```

### Bash 脚本

使用 Bash 脚本运行 Picard CollectHsMetrics，可定义在单次执行中生成指标所需文件的路径。

- 1 从 TruSight Tumor 170 实验分析方法应用程序输出中检索样品 BAM 文件的绝对目录路径，以便在步骤 4 中使用。



- ▶ 对于 DNA 样品，Isaac 比对的 BAM 位于分析输出文件夹下的 DNA\_IntermediateFiles/Alignment 中，文件名为 {样品 ID}.bam。请勿使用重新比对的 BAM。
  - ▶ 对于 RNA 样品，STAR 比对的 BAM 位于分析输出文件夹下的 RNA\_IntermediateFiles/Alignment 中，文件名为 {样品 ID}.bam。
- 2 从基因组文件夹中检索 hg19 genome.fa 文件的绝对目录路径，以便在步骤 4 中使用。
    - ▶ hg19 genome.fa 文件位于基因组文件夹中：genomes/Homo\_sapiens/UCSC/hg19/Sequence/WholeGenomeFasta/genome.fa
  - 3 检索步骤 4 安装 *Picard* 并生成文件 (第 22 页) 中生成的区间文件的绝对输出路径。
  - 4 使用以下模板创建 Bash 脚本。有关脚本所需变量的详细信息，请参见步骤 1、2 和 3。

```
#!/bin/bash
#picard.jar 可执行文件所在位置的路径
PICARDPATH=picard.jar
#基因组文件夹的路径
GENOMESPATh=genomes/Homo_sapiens/UCSC/hg19/Sequence/WholeGenomeFasta/genome.fa
#DNA 探针区间文件的路径
DNAPROBEINTERVALS=DNA_PicardProbe.intervals
#DNA 靶点区间的路径
DNATARGETINTERVALS=DNA_TargetProbe.intervals File
#BAM 文件的路径
DNABAMPATH=AnalysisOutput/DNA_IntermediateFiles/Alignment/dnasample.bam
#RNA 探针区间文件的路径
RNAPROBEINTERVALS=RNA_PicardProbe.intervals
#RNA 靶点区间的路径
RNATARGETINTERVALS=RNA_TargetProbe.intervals File
#BAM 文件的路径
RNABAMPATH=AnalysisOutput/RNA_IntermediateFiles/Alignment/rnasample.bam

java -Xmx2g -jar $PICARDPATH CollectHsMetrics BI=$DNAPROBEINTERVALS
TI=$DNATARGETINTERVALS R=$GENOMESPATh O=DNA-Picard-HsMetrics.txt I=$DNABAMPATH
VALIDATION_STRINGENCY=SILENT

java -Xmx2g -jar $PICARDPATH CollectHsMetrics BI=$RNAPROBEINTERVALS
TI=$RNATARGETINTERVALS R=$GENOMESPATh O=RNA-Picard-HsMetrics.txt I=$RNABAMPATH
VALIDATION_STRINGENCY=SILENT
```

## 卸载 TruSight Tumor 170 本地应用程序

- 1 使用以下命令删除 TruSight Tumor 170 本地应用程序 tar 文件：
 

```
rm -f TruSight_Tumor_170_Local_App_{应用程序版本号}.tar
```
- 2 使用以下命令列出 Docker 映像，以确定安装的映像 ID：
 

```
sudo docker images
```

- 3 使用以下命令根据映像 ID 删除 Docker 映像：

```
sudo docker rmi -f {映像 ID}
```

- 4 使用下面的命令删除资源 tarball（.tgz 文件和解压缩的基因组目录）：

```
rm -f TruSight_Tumor_170_Resources_{应用程序版本号}.tgz
```

```
rm -rf {基因组资源路径}
```

其中，{基因组资源路径} 是解压缩 \*.tgz 文件后基因组目录的路径。

## 修订历史记录

文档	日期	更改描述
文档号 1000000036413 v01	2019 年 7 月	更新了片段和测序簇的 DNA 分析要求。 删除了建议的其他内核版本。 指明了建议的硬件。 删除了命令行先决条件。
文档号 1000000036413 v00	2017 年 8 月	初始版本

## 技术协助

如需技术协助，请与 Illumina 技术支持部门联系。

网站：[www.illumina.com](http://www.illumina.com)  
 电子邮件：[techsupport@illumina.com](mailto:techsupport@illumina.com)

### Illumina 客户支持部门电话号码

地区	免费电话	区域电话
北美	+1.800.809.4566	
爱尔兰	+353 1800936608	+353 016950506
奥地利	+43 800006249	+43 19286540
澳大利亚	+1.800.775.688	
比利时	+32 80077160	+32 34002973
丹麦	+45 80820183	+45 89871156
德国	+49 8001014940	+49 8938035677
法国	+33 805102193	+33 170770446
芬兰	+358 800918363	+358 974790110
荷兰	+31 8000222493	+31 207132960
南韩	+82 80 234 5300	
挪威	+47 800 16836	+47 21939693
日本	0800.111.5011	
瑞典	+46 850619671	+46 200883979
瑞士	+41 565800000	+41 800200442
西班牙	+34 911899417	+34 800300143
新加坡	+1.800.579.2745	
新西兰	0800.451.650	
意大利	+39 800985513	+39 236003759
英国	+44 8000126019	+44 2073057197
中国	400.066.5835	
中国台湾	00806651752	
中国香港	800960230	
其他国家/地区	+44.1799.534000	

**安全数据表 (safety data sheet, 简称 SDS)** — 可通过 Illumina 网站 ([support.illumina.com/sds.html](http://support.illumina.com/sds.html)) 获取。

**产品文档** — 可通过 Illumina 网站下载 PDF 版本。请转到 [support.illumina.com](http://support.illumina.com)，选择一个产品，然后选择 Documentation & Literature (文档与文献)。



Illumina  
5200 Illumina Way  
San Diego, California 92122 U.S.A.  
+1.800.809.ILMN (4566)  
+1.858.202.4566 (北美以外地区)  
techsupport@illumina.com  
www.illumina.com

仅供研究使用，不可用于诊断过程。

© 2019 Illumina, Inc. 保留所有权利。

illumina®