****

**肿瘤产品转录本选择管理规程**

|  |  |
| --- | --- |
| 文件编号： | SOP-RDM-005 |
| 版本号： | A1 |
| 生效日期： | 2021-10-01 |
| 文件密级： | □普通 ■秘密 □机密 □绝密 |

**目录**

[1 目的 2](#_Toc82615115)

[2 适用范围 2](#_Toc82615116)

[3 职责 2](#_Toc82615117)

[4 术语和定义 2](#_Toc82615118)

[5 管理要求 2](#_Toc82615119)

[5.1 数据下载与准备 2](#_Toc82615120)

[5.2 转录本挑选原则 3](#_Toc82615121)

[5.3 转录本挑选流程 4](#_Toc82615122)

[5.4 挑选流程图 5](#_Toc82615123)

[5.5 转录本选择过程流程介绍 6](#_Toc82615124)

[6 相关文件 7](#_Toc82615125)

[6.1 无 7](#_Toc82615126)

[7 相关记录 7](#_Toc82615127)

[7.1 无 7](#_Toc82615128)

[8 引用标准及参考文件 7](#_Toc82615129)

[8.1 无 7](#_Toc82615130)

[9 附录 7](#_Toc82615131)

[9.1 附录A 修订记录 8](#_Toc82615132)

# 目的

在核酸变异注释到CDS或氨基酸时，会因转录本选取不同，而产生不同的注释结果。通常致病性突变区域较为保守，即使转录本不同，注释结果也可能相同，但转录本信息的差异，尤其是相同基因，不同产品间的差异，可能会造成不必要的麻烦。故为了统一不同产品间的转录本信息，并规范化转录本选择来源，本管理规程制定了肿瘤产品转录本选择规则。统一的转录本选择原则及通用的转录本数据库，不仅规范化了产品解读，同时为跨产品数据挖掘和数据库构建奠定了基础。

# 适用范围

本文件适用于肿瘤现有产品覆盖基因的注释或解读环节的转录本选择，以及后续研发和更新过程中转录本的选择规则和流程。产品设计或更新时，依此文件制定的规则选择解读时的转录本，但芯片捕获范围需选择全部转录本覆盖范围的合集。

# 职责

信息分析人员依据本规则确定的转录本列表进行信息分析，提供版本统一稳定的注释结果，并在转录本列表更新时更新分析流程。

遗传分析人员依据本规则确定的转录本列表进行解读，并在有疑义或有更新项时反馈更新问题，经与产品研发、信息研发讨论确定后对转录本列表进行更新。

# 术语和定义

注释：对于一个变异，通过注释可以添加上公共人群数据库频率信息、软件保守度和致病性预测信息，以及基本的基于某个转录本的编码碱基/氨基酸变化的信息。因为同一基因可对应多个转录本，同一变异不同转录本可对应不同的碱基／氨基酸变化。

转录本：转录本是基因序列通过转录形成的一种或多种可供编码蛋白质的成熟的mRNA。同一基因存在因剪切差异造成的多个转录本形式。

# 管理要求

## 5.1 数据下载与准备

### NCBI hg19(GRCh37) annotation文件

下载网址：https://www.ncbi.nlm.nih.gov/projects/genome/guide/human/index.shtml 根据页面显示，选择GRCh37的RefSeq Reference Genome Annotation数据，gff3格式。

### CLINVAR数据库收录转录本文件

文件地址：<https://ftp.ncbi.nlm.nih.gov/pub/clinvar/tab_delimited/> ，下载clinvar\_variant\_summary.txt文件

### LRG数据库Reference standard文件

文件地址：ftp://ftp.ncbi.nlm.nih.gov/refseq//H\_sapiens/RefSeqGene 以客人身份登陆后，下载LRG\_RefSeqGene 文件。

### NCBI RefSeqGene Project文件

文件地址：[ftp://ftp.ncbi.nlm.nih.gov/refseq//H\_sapiens/RefSeqGene](ftp://ftp.ncbi.nlm.nih.gov/refseq/H_sapiens/RefSeqGene) 以客人身份登录，下载

LRG\_RefSeqGene文件

## 5.2 转录本挑选原则

### 时珍知识库+龙舌兰所有基因

##### 第一优先级：NCBI hg19的最新注释版本文件中有唯一转录本的。

##### 第二优先级：最新CLINVAR数据库收录并荐有唯一转录本的。

##### 第三优先级：最新LRG数据库带有Reference standard标签并荐有唯一转录本的。

##### 第四优先级：最新HGMD数据库收录并荐有唯一转录本的。

##### 第五优先级：CDS编码区最长的转录本。

##### 特殊优先级：生信、遗传分析、产品部统一挑选确定的。

### WES所有基因（除去时珍知识库+龙舌兰所有基因）

##### 第一优先级：NCBI hg19的最新注释版本文件中有唯一转录本的。

##### 第二优先级：最新CLINVAR数据库收录并荐有唯一转录本的。

##### 第三优先级：最新LRG数据库带有Reference standard标签并荐有唯一转录本的。

##### 第四优先级：NCBI RefSeqGene Project最新版本推荐的RefSeq Select有唯一转录本的。

##### 第五优先级：CDS编码区最长的转录本。

##### 特殊优先级：生信、遗传分析、产品部统一挑选确定的。

## 5.3 转录本挑选流程

1. 时珍知识库+龙舌兰所有基因

#### 1）NCBI hg19的最新注释版本文件中是否是唯一转录本。

##### a) 如果候选转录本唯一，选择该转录本，结束挑选。

##### b) 如果候选转录本有多个，进入2）。

#### 2）最新CLINVAR数据库中是否为唯一推荐转录本。

##### a) 如果候选转录本唯一，选择该转录本，结束挑选。

##### b) 如果候选转录本有多个，进入3）。

#### 3）最新LRG数据库带有Reference standard标签是否为唯一转录本的。

##### a) 如果候选转录本唯一，选择该转录本，结束挑选。

##### b) 如果候选转录本有多个，进入4）。

#### 4）最新HGMD数据库是否为唯一推荐转录本。

##### a) 如果候选转录本唯一，选择该转录本，结束挑选。

##### b) 如果候选转录本有多个，进入5）。

#### 5）选择CDS编码区最长的转录本。

#### 6）如果特定变异所需的转录本与如上流程确定后的转录本存在差异，则选择特定变异所需转录本，并在数据库中记录，作为指定变异的最高转录本优先级。

1. WES所有基因（除去时珍知识库+龙舌兰所有基因）

#### 1）NCBI hg19的最新注释版本文件中是否是唯一转录本。

##### a) 如果候选转录本唯一，选择该转录本，结束挑选。

##### b) 如果候选转录本有多个，进入2）。

#### 2）最新CLINVAR数据库中是否为唯一推荐转录本。

##### a) 如果候选转录本唯一，选择该转录本，结束挑选。

##### b) 如果候选转录本有多个，进入3）。

#### 3）最新LRG数据库带有Reference standard标签是否为唯一转录本的。

##### a) 如果候选转录本唯一，选择该转录本，结束挑选。

##### b) 如果候选转录本有多个，进入4）。

#### NCBI RefSeqGene Project最新版本推荐的RefSeq Select是否为唯一转录本。

##### a) 如果候选转录本唯一，选择该转录本，结束挑选。

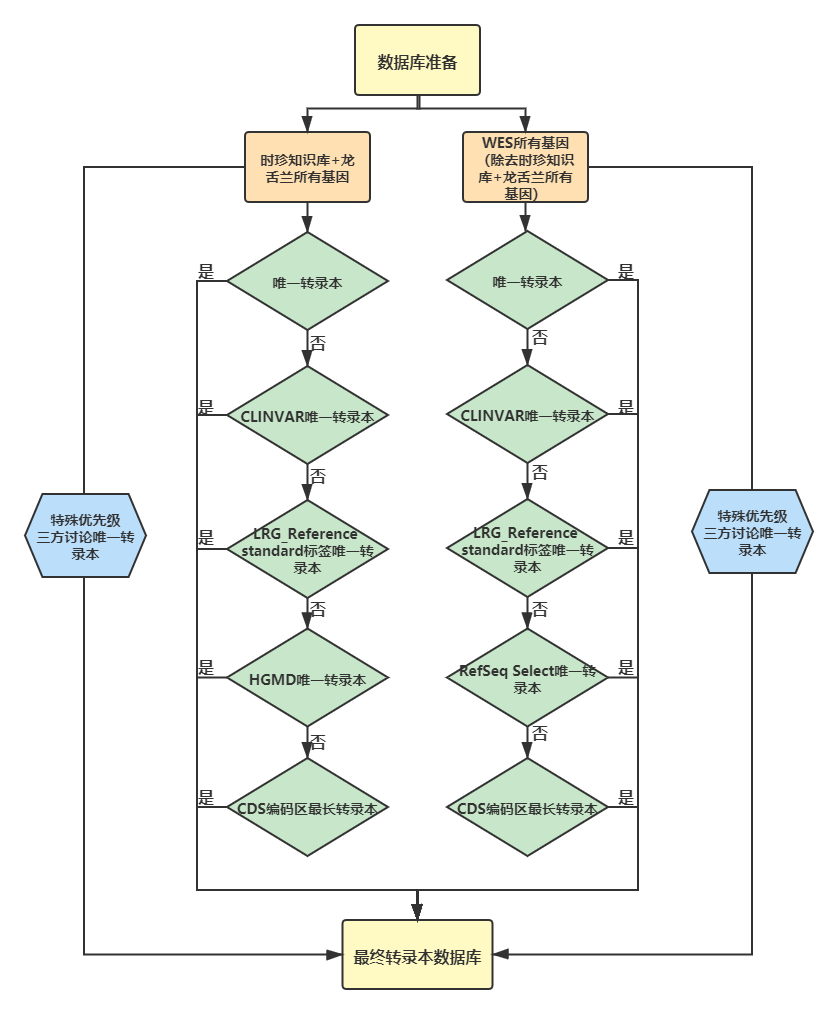
##### b) 如果候选转录本有多个，进入5)。

#### 选择CDS编码区最长的转录本。

#### 如果特定变异所需的转录本与如上流程确定后的转录本存在差异，则选择特定变异所需转录本，并在数据库中记录，作为指定变异的最高转录本优先级。

## 5.4 挑选流程图

根据5.3部分转录本挑选原则，可视化转录本挑选流程如图1所示。

图1. 转录本挑选流程

## 5.5 转录本选择过程流程介绍

与本文件同时存在的Get\_Best\_Transcript文件夹，即转录本选择过程的数据、流程包。文件目录结构如图2所示。

下载数据存放在Get\_Best\_Transcript/transcript中，readme文件中记录了对各个数据数为的前处理操作及对应的处理程序。Get\_Best\_Transcript/best\_transcript记录整合各个数据库结果得到最优转录本的过程，详细操作可见readme文件。best\_trans.v2.xls为688芯片最终结果。

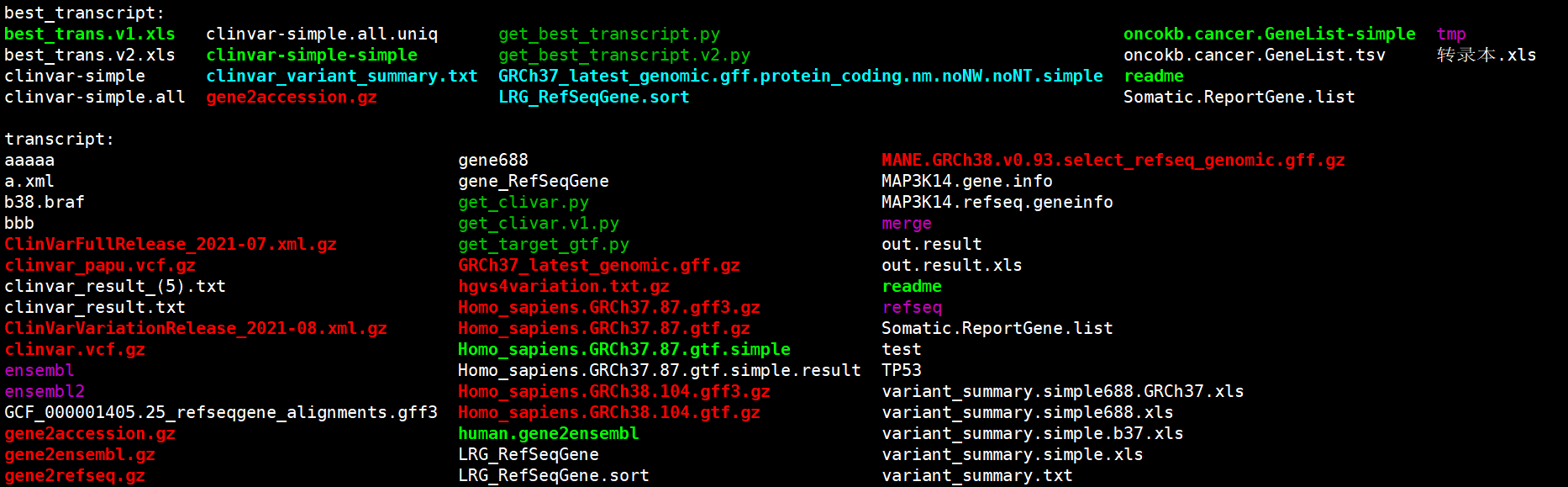


图2. 转录本选择流程目录结果

# 相关文件

## 6.1 无

# 相关记录

## 7.1 无

# 引用标准及参考文件

## 8.1 无

# 附录

附录A 修订记录。

## 9.1 附录A 修订记录

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **序号** | **版本号** | | **修订日期** | **修订内容摘要** | **修订人** | **审批人** |
| **前** | **后** |
| 1 | / | A0 | 20190225 | 文件新增 | 谭美华 | 蔡宇航 熊云  候裕梅  王春丽  石太平  袁玉英  黄辉  叶明芝 |
| **2** | A0 | A1 | 20211001 | 文件修订 | 邓彩萍  吕梦婷 | 吴仁化  高丽霞  王春丽  王喆 |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

