肿瘤产品转录本选择管理规程

Regulations for the selection and management of tumor product transcripts

文件编号：	
Document NO.:	TJ-SOP-RI-129
版本号：	
Version NO.:	A0
生效日期：	
Effective Date:	
文件密级：	□普通/Unclassified　□秘密/Secret
Classification:	▪机密/Highly Secret　□绝密/Top Secret

目录/Directory

1 目的/Objectives

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。

2 适用范围/Scopes

本文件适用于肿瘤现有产品覆盖基因的注释和解读环节的转录本选择，以及后续研发过程中涉及的转录本选择、更新原则。产品设计或更新时，依此文件制定的规则选择解读时时所用转录本，但芯片捕获范围需选择全部转录本覆盖范围的合集。

3 职责/Responsibilities

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。

4 术语和定义/Terms and Definitions

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。

5 管理要求/Management requirements

5.1 数据下载与准备

5.1.1 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。

5.1.2 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

5.1.3 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。

5.1.4 NCBI RefSeq Select 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，筛选 RefSeq Select。

5.2 转录本挑选原则

5.2.1 时珍知识库+龙舌兰所有基因

a) 第一优先级：NCBI hg19 的最新注释版本文件中有唯一转录本的。

b) 第二优先级：最新 CLINVAR 数据库收录并有唯一转录本的（不一定是当前转录本最新版本号）。

c) 第三优先级：最新 LRG 数据库带有 Reference standard 标签并有唯一转录本的（不一定是当前转录本最新版本号）。

d) 第四优先级：最新 HGMD 数据库收录并有唯一转录本的（不一定是当前转录本最新版本号）。

e) 第五优先级：CDS 编码区最长的转录本。

f) 特殊优先级：生信、遗传分析、产品部统一挑选确定的。

5.2.2 283\171\WES 产品剩余基因（除去时珍知识库+龙舌兰所有基因）

a) 第一优先级：NCBI hg19 的最新注释版本文件中有唯一转录本的。

b) 第二优先级：最新 CLINVAR 数据库收录并荐有唯一转录本的（不一定是当前转录本最新版本号）。

c) 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（不一定是当前转录本最新版本号）。

d) 第四优先级：NCBI RefSeq Select 最新版本带有 RefSeq Select 标签并荐有唯一转录本的。

e) 第五优先级：CDS 编码区最长的转录本。

f) 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

5.3.1 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。

5.3.2 283\171\WES 产品剩余基因（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）则最新 CLINVAR 数据库中是否为唯一推荐转录本。

a) 如果候选转录本和 1）的交集唯一，选择 NCBI 的转录本，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) 如果候选转录本和 1）的交集唯一，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) NCBI RefSeq Select 最新版本推荐的 RefSeq Select 是否为唯一转录本。

a) 如果候选转录本和 1）的交集唯一，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。

7) 如果基因筛选出的转录本为 hg38 (GRCh38)转录本，或者为 NP 格式的 XM 格式的（没有 NM 格式的）则沿用以前的，没有以前的，则删除。

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 5.4-1 所示。

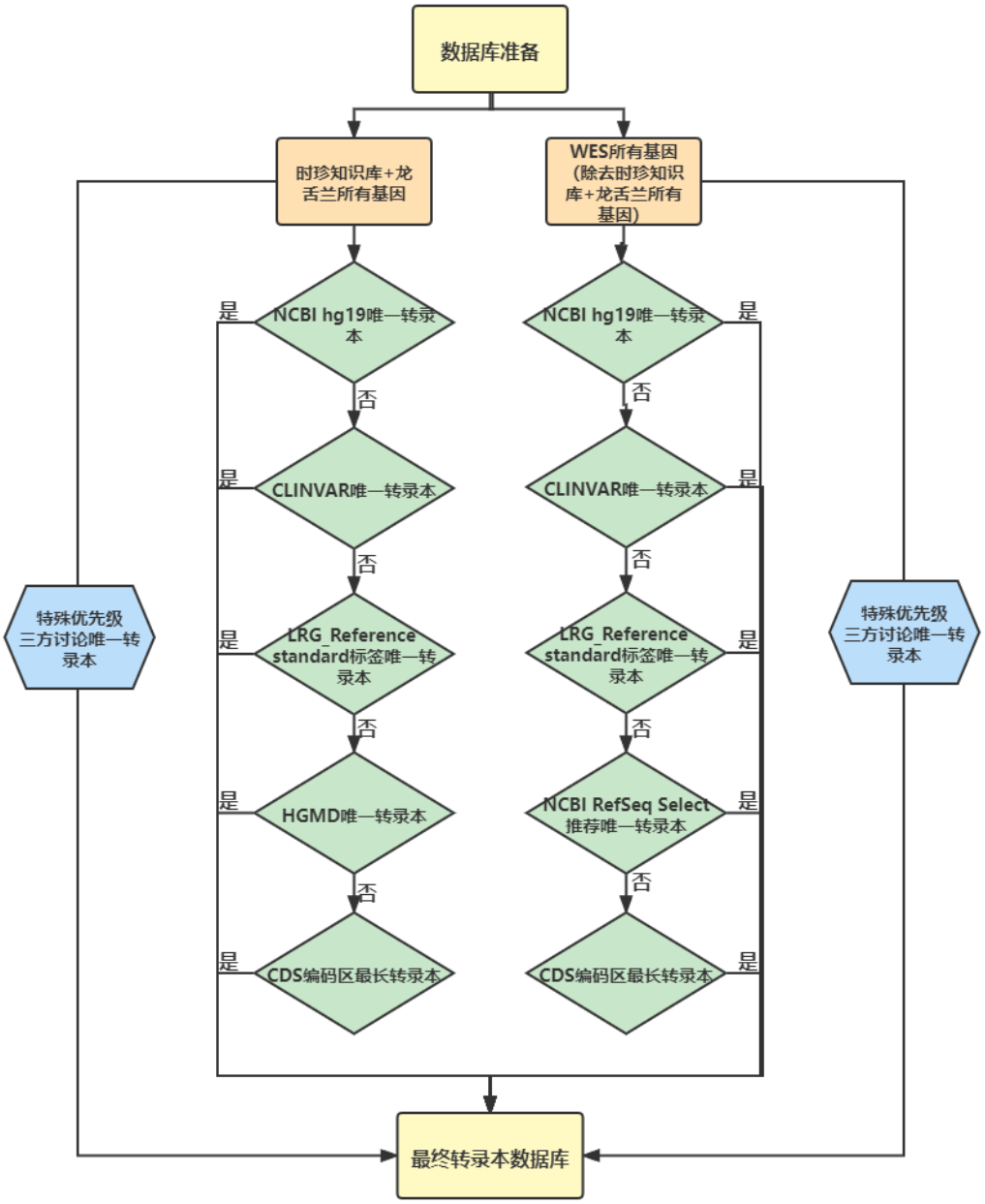

图 5.4-1 转录本挑选流程

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 5.5-1 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。

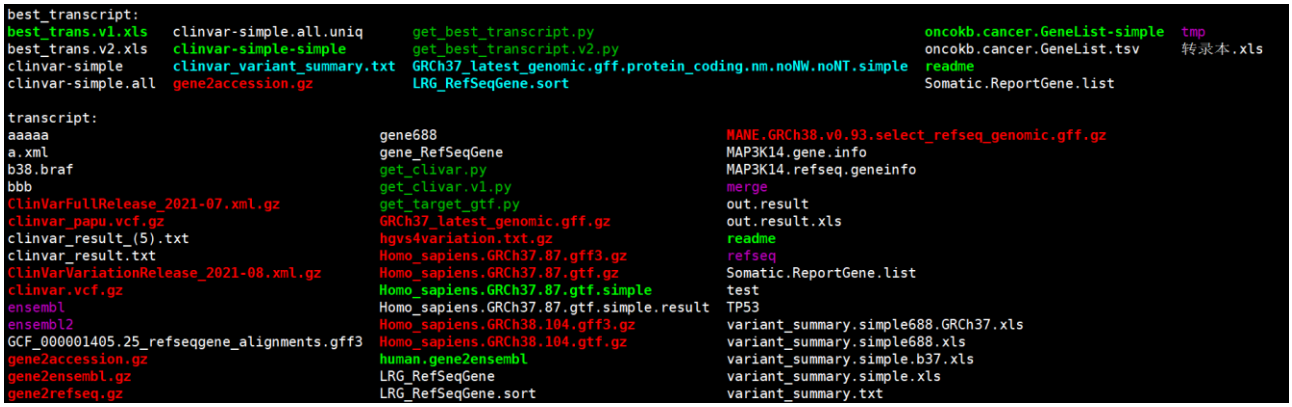

图 5.5-1 转录本选择流程目录结果

最终的代码上传到 gitlab（https://gitlab.genomics.cn/lvmengting/vep_annotation/-/tree/main）。

5.6 转录本人工校准过程

特殊优先级转录本人工筛选方法：与现行转录本不同基因核实 clinvar 数据库和时珍知识库的药物情况，如果没有涉及用药，则按照正常流程最终选取 CDS 区编码最长转录本；如果涉及用药，则进入人工调研流程。

人工调研基因，参考每个基因的各个数据库推荐转录本，对比各个转录本 CDS 区长以及核心区域长度，clinvar，Oncokb，CKB 等各个权威数据库推荐情况，将库内致病位点按照各个转录本重新注释结果，将注释结果逐一过文献。选出核心区域长度较长，权威数据库推荐度高以及文献中偏好性强的转录本（相同偏好性情况下可能是转录本有更迭的情况，这种情况下建议保留两个转录本），如果遇到特殊情况，如 MET 基因，则保留两个转录本。调研转录本文档格式如图 5.6-1。

KRAS 转录本选择

1、共核实 87 个突变，其中 86 个突变用 NM_033360 和 NM_004985 注释结果一致。

2、不一致位点为：KRAS g.25368455G>A，NM_033360 c.490C>T p.R164*、NM_004985 c.451-5610C>T。该位点为无义突变，无药。

3、转录本序列情况

NM_033360.4 CDS 区长 570bp，NM_004985.5 CDS 区长 567bp，其中核心区域一致，在 450bp 之后存在不一致。

4、各个数据库推荐情况

refseq	ensemble	clinvar	hgmd
NM_004985.5	NM_004985.5	NM_004985.5	NM_004985.5

总结：鉴于 NM_033360.4 和 NM_004985.5 在突变核心区域无区别，且各个权威数据库一致推荐 NM_004985.5，因此 NM 号选择 NM_004985.5。

图 5.6-1 特殊基因转录本人工选择情况

5.7 转录本用于各产品芯片覆盖度评估

分为知识库内位点评估和 clinvar 位点评估。

生信人员评估 HRR\283\Pancancer\WES 芯片上新转录本与旧转录本相比缺少的位点，发给时珍数据库人员进行位点临床意义过库评估，如评估后发现没有致病位点，则评估通过。

生信人员进行 HRR\283\Pancancer\WES 芯片上新转录本与旧转录本相比各自不同的位点，导出 clinvar 上注释出的临床意义，交由解读人员评估。HRR\283\Pancancer\WES 芯片旧转录本多出的位点如果没有致病位点，则通过。如果有致病位点，需评估位点是否在检测范围以及是否有药，如果致病位点没有药或者不在检测范围，则通过；如果致病位点有药，则综合评估新转录本注释出位点的临床意义情况，与产品经理及生信研发人员共同讨论决定。

覆盖度评估过程需撰写 TJ-R-RI-160 A0 项目升级验收报告—NM 号芯片覆盖度评估报告。并将终版转录本按照各个芯片进行拆分汇总存档。

该转录本选择方案目前适用于实体瘤，血液病肿瘤转录本选择按照产品固有的转录本。

5.8 转录本更新周期与记录

标准：年度更新。

当前转录本确定终版时间为 2022 年 1 月 26 日，执行记录如表 5.8-1。

表 5.8-1

转录本库（涉及产品）	更新人	更新时间	覆盖度评估人员	覆盖度评估时间	终版转录本确认时间
171 206 220 283 508 636 689 CRC_6 WES	邓彩萍、周丽颖、刘博、吕梦婷	2021/11/17	刘博、孟培、宋铁峰、周丽颖	2022/1/9	2022/1/26

转录本参考数据库的版本号如表 5.8-2：

表 5.8-2

转录本参考库	2022/1/26 此次参考的数据库版本
NCBI hg19	##gff-version 3 #!gff-spec-version 1.21 #!processor NCBI annotwriter #!genome-build GRCh37.p13 #!genome-build-accession NCBI_Assembly:GCF_000001405.25 #!annotation-date 10/22/2020 #!annotation-source NCBI Homo sapiens Updated Annotation Release 105.20201022 ##sequence-region NC_000001.10 1 249250621 ##species https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=9606
CLINVAR	variant_summary_2021-08.txt.gz
LRG	2021-8 月
HGMD	2021-4 月

NCBI RefSeq Select	

6 相关文件/Related Documents

无

7 相关记录/Related Records

TJ-R-RI-158 A0 肿瘤产品转录本更新记录表

Update Form for tumor product transcripts TJ-R-RI-158 A0

8 引用标准及参考文件/Reference Standards and Reference Documents

无

9 附录/Appendixes

附录 A 修订记录

Appendix A Amendment Record

序号 Number	版本号 Version NO.		修订日期 Date	修订内容摘要 Content of Revision	修订人 Revise	审批人 Approve
	前 Before	后 After				
1	/	A0		新增	吕梦婷 周丽颖	宋铁峰 刘博 吴仁花

终止符