



中华人民共和国国家医药

行业标准

XX/T XXXXX—XXXX

代替 XX/T

肿瘤体细胞变异解读规范和数据库建立的技术指南

Technical guideline for interpreting tumor somatic mutations and databases establishment

(点击此处添加与国际标准一致性程度的标识)

征求意见稿

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX – XX – XX 发布

XXXX – XX – XX 实施

国家药品监督管理局 发 布

目 次

前言 2

1 范围 3

2 规范性引用文件 3

3 术语和定义 3

4 缩略语 4

5 肿瘤体细胞变异的注释 4

6 肿瘤体细胞变异的致癌性解读 4

7 体细胞评价结果的数据构成 8

8 数据库的质量评价 9

9 参数调整 9

附录 A..... 11

附录 B..... 12

附录 C..... 13

参考文献 14

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由国家药品监督管理局提出。

本文件由全国医用临床检验实验室和体外诊断系统标准化技术委员会（SAC/TC136）归口。

本文件起草单位：

本文件主要起草人：

肿瘤体细胞变异解读规范和数据库建立的技术指南

1 范围

本文件提供了肿瘤体细胞变异的解读规范和数据库建立的技术指南。

本文件适用的体细胞变异类型为单核苷酸变异（SNV）和短片段的插入/缺失变异（Indels），不适用于胚系变异，不适用于融合、拷贝数变异（除了限制在单一基因内的缺失）或其他染色体重排。

本文件适用于对肿瘤基因体细胞序列变异致癌性的判断，是指导肿瘤精准用药的重要环节。肿瘤精准用药全流程包括以下5个步骤：

- a) 体细胞突变鉴定：体细胞突变鉴定识别癌症患者个体基因组在体细胞层面发生的变异，如单核苷酸变异（SNV）和短片段的插入/缺失变异（Indels）等。
- b) 突变注释：包括确定突变的类型、位置以及可能影响的基因或蛋白质等，以及通过比对公共数据库，如 ClinVar、dbSNP 或 HGMD，以了解突变的已知特性和相关研究。
- c) 突变评级：突变评级遵循 ACMG/AMP 等相关指南，将突变分为五类：致病性（P）、可能致病性（LP）、意义不明（VUS）、可能良性（LB）和良性（B）。评级基于多条证据，包括突变的健康人群频率、蛋白功能改变等。
- d) 临床意义：评估体细胞突变对患者疾病的具体影响，包括评估突变与疾病的相关性，与特定药物的敏感性或耐药性等。
- e) 临床决策：临床医生综合以上信息做出选择用药方案、监测疾病进展、预防性干预或遗传咨询等决策行为。

本文件目标在于指导检测机构对肿瘤体细胞突变的评级，规范体细胞评级中公共数据库或内部数据库的使用，提高肿瘤体细胞评级和精准用药的准确性及有效性。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和定义

下列术语和定义适用于本文件。

3.1

体细胞变异 somatic mutation

在个体的生长发育过程中，由于各种诱变因素的作用，如DNA复制错误、化学物质、环境因素、病毒感染或辐射暴露等，导致体细胞中发生的基因变化。

注：这类变异不同于胚系变异，它们仅在部分非胚系细胞中出现，并不通过生殖细胞传递给后代。

3.2

变异解读 mutation interpretation

检测机构依据指南规范对检出变异进行综合分析和评估，加权匹配到的证据，对变异的潜在影响进行临床判断，从而为诊断、治疗和预后评估提供科学依据。

4 缩略语

下列缩略语适用于本文件。

AMP: 美国分子病理学协会 (Association for Molecular Pathology)

ASCO: 美国临床肿瘤学会 (American Society of Clinical Oncology)

CAP: 美国病理学家协会 (College of American Pathologists)

CGC: 国际癌症基因组联盟 (Cancer Genomics Consortium)

ClinGen: 美国临床基因组资源中心 (Clinical Genome Resource)

EMBL-EBI: 欧洲分子生物学实验室 - 欧洲生物信息学研究所 (European Molecular Biology Laboratory - European Bioinformatics Institute)

HGVS: 人类基因组变异学会 (Human Genome Variation Society)

ID: 标识符 (Identifier)

MANE: NCBI和EMBL-EBI的匹配注释 (Matched Annotation from NCBI and EMBL-EBI)

NCBI: 国家生物技术信息中心 (National Center for Biotechnology Information)

OM: 中等致癌证据 (Oncogenic Moderate)

OP: 支持致癌证据 (Oncogenic Supporting)

OS: 强致癌证据 (Oncogenic Strong)

OVS: 非常强致癌证据 (Oncogenic Very Strong)

SBP: 支持良性证据 (Somatic Benign Supporting)

SBS: 强良性支持证据 (Somatic Benign Strong)

SBVS: 非常强良性证据 (Somatic Benign Very Strong)

VICC: 国际癌症变异释义联盟 (Variant Interpretation for Cancer Consortium)

5 肿瘤体细胞变异的注释

体细胞变异的注释应以HGVS的命名体系为基础。为了确保变异在不同基因组构建和转录本之间的正确映射，推荐为每个变异提供一个ClinGen等位基因注册表ID。此外，实验室应采用NCBI和EMBL-EBI的匹配注释选择转录本作为默认的报告转录本，并在MANE Plus临床转录本中报告变异，MANE Plus临床转录本包含了MANE选择转录本以外的其他临床相关的转录本变体，这样可以确保在临床决策中考虑到所有可能相关的变异形式。

6 肿瘤体细胞变异的致癌性解读

6.1 证据等级

每一种数据类型的升降级调整的等级应遵循表1。

注：体细胞变异致癌性判断基于ClinGen（Clinical Genome Resource，临床基因组资源中心）、CGC（Cancer Genomics Consortium，癌症基因组联盟）和VICC（Variant Interpretation for Cancer Consortium，癌症变异解读联盟）联合制定的标准操作规程^[1]，也可同时结合其他相关体细胞变异解读标准，如AMP、ASCO和CAP发布的序列变异解读和报告指南^[2]等。

表1 不同证据来源数据的证据等级对照表

证据来源	证据等级						
	SBVS	SBS	SBP	OP	OM	OS	OVS
	-8	-4	-1	+1	+2	+4	+8
人群频率数据	SBVS1	SBS1		OP4			
功能数据		SBS2				OS2	
预测数据			SBP2		OM1,OM2,OM4	OS1	OVS1
热点变异数据				OP3	OM3	OS3	
计算数据			SBP1	OP1			
其他数据				OP2			

6.2 证据类型

6.2.1 致癌性证据

6.2.1.1 致癌性人群频率证据

OP4: gnomAD^[3] 数据库中正常对照人群中未发现的变异（或频率极低）。

6.2.1.2 致癌性功能证据

OS2: 充分验证的体外或体内的功能研究支持该变异具有致癌效应。

注1：功能实验需要具有重复性与稳定性；

注2：若OS1适用，则只有在功能研究基于特定核苷酸变化时才能使用该证据。

6.2.1.3 致癌性预测证据

OVS1: 抑癌基因发生功能失活的变异，如无义变异、移码变异、经典 ± 1 或 ± 2 剪接突变，起始密码子变异等。

注1：具体操作可参考ClinGen工作组提出的PVS1决策树，来对变异的致癌性强弱进行升降级调整^[5]。

OS1: 具有与先前已确认的致癌变异相同的氨基酸变化，无论核苷酸的改变如何。需注意剪接影响的改变。

OM1: 位于关键和/或被充分认可的基因功能域，如酶的活性位点。

OM2: 已知的癌基因或抑癌基因中发生编码框内的缺失/插入或在抑癌基因中发生终止密码子丧失，引起的蛋白质长度变化。

OM4: 新的错义变异导致氨基酸变化，同一位置，另一种氨基酸变化已确认是致癌的，并且新的变异在氨基酸改变方面应该与先前被确认为致癌的变异相当或更为显著。

注2: 建议采用广受认可的氨基酸差异度量标准，如Grantham's距离、Epstein's差异系数或Miyata's距离。

示例: 以 p.Arg156His 突变为例，该突变已被证实与癌症相关。我们观察到另一种突变 p.Arg156Cys，从 Arg 到 Cys 的 Grantham's 距离（180），显著高于从 Arg 到 His 的 Grantham's 距离（29）。因此，可以认为 p.Arg156Cys 在氨基酸改变方面比 p.Arg156His 更显著。

6.2.1.4 致癌性热点变异证据

OS3: 变异位于热点区域，相同氨基酸位置上的体细胞变异在至少50个样本中观察到，并且相同氨基酸变化的样本数大于等于10。

OM3: 变异位于热点区域，相同氨基酸位置上的体细胞变异在少于50个样本中观察到，并且相同氨基酸变化的样本数大于等于10。

OP3: 变异位于热点区域，并且特定氨基酸变化的样本数少于10个。

6.2.1.5 致癌性计算证据

OP1: 所有使用的生物信息学计算工具均支持变异的致癌作用（保守性/进化性、剪接效应等）。

6.2.1.6 致癌性其他证据

OP2: 为恶性肿瘤已知的单一遗传病因的特定基因的体细胞变异。

6.2.2 良性证据

6.2.2.1 良性人群频率证据

SBVS1: 在人群数据库中任意人群中（如东亚人、非洲人、拉丁美洲人、南亚人、欧洲人（非芬兰人）等），次等位基因的频率 $\geq 5\%$ 。

SBS1: 在人群数据库中任意人群中（如东亚人、非洲人、拉丁美洲人、南亚人、欧洲人（非芬兰人）等），次等位基因的频率 $\geq 1\%$ 。

6.2.2.2 良性功能证据

SBS2: 充分验证的体外或体内功能研究未显示出致癌效应。

6.2.2.3 良性预测证据

SBP2: 同义变异，剪切预测算法预测它对剪切保守序列没有影响，也不会产生新的剪切位点，并且该核苷酸并不高度保守。

6.2.2.4 良性计算证据

SBP1: 所有使用的计算工具均表明该变异对基因或基因产物没有影响（保守性/进化、剪切效应等）。

6.3 证据加权规则

应参照ClinGen、CGC和VICC联合制定的SOP提出的一套证据加权规则进行操作，如表2和表3所示。

表2 各证据强度得分

证据强度	致癌性	良性
非常强	+8	-8
强	+4	-4
中等	+2	/
支持	+1	-1

表3 致癌性分类的得分范围

分类	得分范围
致癌	≥ 10
可能致癌	6-9
意义不明	0-5
可能良性	(-6)-(-1)
良性	≤ -7

部分证据不能被同时使用，此时，应制定相关的规定对证据进行取舍，如表4所示。

表4 不能共用的证据

证据 1	证据 2	不能共用的原因
OM1	OS1	两个证据等级都涉及从氨基酸水平评估变异的致癌性
OM1	OS3	
OM3	OM1	
OM3	OM4	
OM4	OS1	
OM4	OS3	
OM4	OM1	
OM2	OVS1	适用的变异类型不同

注1：如果变异适用OS1证据项，只有功能研究是基于特定核苷酸改变时，才能同时用OS2证据项。

注2：如果变异适用OS1证据项，只有当基于特定核苷酸改变可以观察到热点变异时，才能同时用

OS3证据项。

6.4 变异致癌性分类

应根据证据加权规则计算结果对应的评分（见表3），最终将变异分为5类，分别为致癌、可能致癌、意义不明、可能良性、良性。

6.5 知识库搭建的宏观要求

为了提升体细胞变异致癌性解读的准确性，各家单位可考虑搭建内部数据库，结合现有公共数据库的信息，为致癌性判断提供更多证据。

自行搭建数据库应至少考虑以下几个方面：主要功能或目的（如变异致癌性解读数据库、人群数据库、热点突变数据库、功能学验证数据库登），参考来源，数据采集方式，存储内容与形式，更新维护的方式和频率等。

数据库的评价应考虑以下几个方面：完整性（包括1，数据库覆盖的基因范围；2，数据库是否能覆盖所有证据类型），准确性（确保收录的信息准确无误），时效性（确保收录的信息是最新的）为了提升体细胞变异致癌性解读的准确性，我们引入了多角度的证据类型。不同变异匹配对应的证据需要有一个可靠的知识库支持。

在知识库的搭建过程中，为了不断完善知识库，需要选择合适的变异进行多轮测试，帮助修正知识库中的问题。

6.6 公共专家知识库和软件

各家单位在体细胞致病性评级实际操作中可根据自身需求和具体情况选择合适的公共专家知识库、自建知识以及预测软件，本文件不做硬性规定，详见附录B和附录C。

7 体细胞评价结果的数据构成

体细胞评价结果至少包括以下信息，见表5。

表 5 结果数据构成

序号	名称	信息内容
1	肿瘤类型(Tumor Type)	肿瘤的命名应符合取得国际公认的国际疾病肿瘤学分类(International Classification of Diseases for Oncology, ICD-O)、国家癌症研究所辞典(National Cancer Institute Thesaurus, NCIt)、统一医学语言系统(Unified Medical Language System, UMLS)或同类型肿瘤学分类标准和命名规则
2	基因(Gene)	基因应以 HGNC approved symbol 应为标准名
3	变异类型(Mutation Type)	应明确突变类型为单核苷酸变异（SNV）、短片的段插入/缺失突变(Indels)，具体详见附录 A
4	变异命名描述(Mutation Description)	对于单核苷酸变异（SNV）、短片的段插入/缺失突变 (Indels)，应按照人类基因组协会(Human Genome Variation Society, HGVS)命名规范进行统一的命名，包

		括碱基改变和氨基酸改变。
5	变异来源(Mutation Origin)	应明确证据适用胚系突变和/或体细胞突变
6	证据类型(Evidence type)	分为致癌性证据和良性证据
7	证据强度(Evidence strength)	分为非常强、强、中等和支持
8	证据等级(Evidence level)	分为非常强致癌证据 (Oncogenic Very Strong, OVS)、强致癌证据 (Oncogenic Strong, OS)、中等致癌证据 (Oncogenic Moderate, OM)、支持致癌证据 (Oncogenic Supporting, OP)、非常强良性证据 (Somatic Benign Very Strong, SBVS)、强良性支持证据 (Somatic Benign Strong, SBS) 和支持良性证据 (Somatic Benign Supporting, SBP)
9	证据来源(Source of evidence)	分为人群频率数据、功能数据、预测数据、热点变异数据、计算数据和其他数据
10	致癌性(Oncogenicity)	分为致癌、可能致癌、意义不明、可能良性、良性
11	提交者(Created user)	应记录提交人姓名或机构名称
12	提交时间(Created time)	应记录提交证据记录的日期和时间
13	审核者(Review user)	应记录审核人姓名或机构名称
14	审核时间(Review time)	应记录审核证据记录的日期和时间

8 数据库的数据管理

数据库构建时应确保数据的安全性、稳定性，并制定维护与升级方案和异常情况处置方案。

安全性方面，基因检测数据属于受检者的隐私，数据应保密，需要杜绝未经授权单位对数据的非法访问。数据库中的数据应在安全、稳定的介质中进行备份。

稳定性方面，应规定需要储存的数据内容。数据文件可以在本地或云端保存，如有本地存储，应说明本地存储所需要具备能力，和存储备份的时间表和建议。

数据库应具有完整的维护及升级方案，并针对可能发生的数据丢失、损坏、泄密等风险建立相应的预防措施和预防机制。

9 数据库的质量评价

9.1 评价数据集

ClinGen提供的10个基因的94个突变以及评价标准，草案另外提供30个突变作为测试数据集的补充（VCF格式）。

9.2 评估变异的致癌性

依据指南提供的分类标准和构建完成的数据库，对收集到的证据进行评分和分类，将变异分类为致癌性、可能致癌性、意义不明、可能良性或良性。

9.3 统计分析

使用统计学方法评估评级方案的一致性指标，确保评级的准确性和可靠性。

9.4 参数调整

根据构建的数据库和指南，使用基于半定量的评分系统，通过实际病例数据进行验证和参数调整，对打分系统进行迭代升级。

检测单位也可以使用实际病例数据进行验证和参数调整，对打分系统进行迭代升级，并进行环境/人员变量的外部测试。

AA

附 录 A
(资料性)
突变类型

本文件适用的突变类型, 见表A。

表 A. 适用变异类型

序号	变异类型	变异类型说明
1	WT	基因表达正常或基因为野生型
2	MUT	涵盖基因下所有变异类型。当证据适用于某基因或某外显子/内含子内全部变异时采用。
3	SNV+INDEL	涵盖基因下 SNV、INDEL 及其子类的所有突变类型
4	INDEL	小片段插入或缺失突变 (≤50bp)
5	DEL	小片段缺失突变 (≤50bp)
6	DUP	小片段重复突变 (≤50bp)
7	DELINS	小片段插入缺失突变 (≤50bp)
8	INS	小片段插入突变 (≤50bp)
9	SNV	单个碱基突变
10	ExonDEL	基因内一个或多个外显子缺失

BB

附 录 B
(资料性)
参考数据库

体细胞致病性评级常用数据库, 见表B。

表 B. 体细胞评级常用数据库

类别	名称	链接
人群数 数据库	gnomAD	https://gnomad.broadinstitute.org
	1000 Genomes Project	http://browser.1000genomes.org
	Exome Variant Server	http://evs.gs.washington.edu/EVS
	dbSNP	http://www.ncbi.nlm.nih.gov/snp
	dbVar	http://www.ncbi.nlm.nih.gov/dbvar
	ExAC	http://exac.broadinstitute.org
癌症数 数据库	cancerhotspots	https://www.cancerhotspots.org
	PeCanPIE	https://pecan.stjude.cloud/pie
	COSMIC	https://cancer.sanger.ac.uk/cosmic
	My Cancer Genome	http://www.mycancergenome.org
	Personalized cancer therapy, MD Anderson Cancer Center	https://pct.mdanderson.org
	cBioPortal, Memorial Sloan Kettering Cancer Center	http://www.cbioportal.org
	Intogen	https://www.intogen.org/search
	ClinicalTrials.gov	https://clinicaltrials.gov
	IARC (WHO) TP53 mutation database	http://p53.iarc.fr
	Pediatric Cancer Genome Project (St. Jude Children's Research Hospital-Washington University)	http://explorepccgp.org
	International Cancer Genome Consortium	https://dcc.icgc.org
	OncoKB	https://www.oncokb.org/
	The Clinical Knowledgebase (CKB)	https://ckb.jax.org/
序列数 数据库	NCBI Genome	http://www.ncbi.nlm.nih.gov/genome
	RefSeqGene	http://www.ncbi.nlm.nih.gov/refseq/rsg
	Locus Reference Genomic	http://www.lrg-sequence.org
	UCSC table browser	https://genome.ucsc.edu/cgi-bin/hgTables
	Ensemble BioMart	http://useast.ensembl.org/biomart/martview
其他疾 病/突变 数据库	ClinVar	http://www.ncbi.nlm.nih.gov/clinvar
	Human Gene Mutation Database	http://www.hgmd.org
	Leiden Open Variation Database	http://www.lovd.nl
	dbNSFP (compiled database of precomputed in silico prediction scores for nonsynonymous SNVs)	https://sites.google.com/site/jpopgen/dbNSFP
	Ensemble Variant Effect Predictor	http://www.ensembl.org/info/docs/tools/vep/index.html

附 录 C
(资料性)
参考预测软件

体细胞致病性评级常用预测软件, 见表C。

表 C. 体细胞评级常用预测软件

功能	预测工具	链接
错义 变异	CADD	https://cadd.gs.washington.edu/snv
	CHASMPPlus Cancer Type Specific	https://karchinlab.github.io/CHASMPplus
	CHASMPPlus Generic	https://karchinlab.github.io/CHASMPplus
	FATHMM	http://fathmm.biocompute.org.uk/index.html
	FATHMM MKL	http://fathmm.biocompute.org.uk/fathmmMKL.htm
	FATHMM XF Coding	http://fathmm.biocompute.org.uk/fathmm-xf/index.html
	Likelihood Ratio Test	http://www.genetics.wustl.edu/jflab/lrt_query.html
	Mutation Assessor	http://mutationassessor.org/r3/
	MutPred	http://mutpred.mutdb.org/index.html
	PhD-SNPg	https://snps.biofold.org/phd-snp/index.html
	Polyphen HDIV	http://genetics.bwh.harvard.edu/pph2/
	Polyphen HVAR	http://genetics.bwh.harvard.edu/pph2/
	PROVEAN	http://provean.jcvi.org/genome_submit_2.php?species=human
	REVEL	https://sites.google.com/site/revelgenomics
	SIFT	https://sift.bii.a-star.edu.sg/
	VEST-4	https://karchinlab.org/apps/appVest.html
剪接 位点 变异	MutSpliceDB	https://brb.nci.nih.gov/splicing
	spliceAI	https://spliceailookup.broadinstitute.org
	varSEAK	https://varseak.bio
	Human Splicing Finder	http://www.umd.be/HSF3
	MaxEntScan	http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html
	NetGene2	http://www.cbs.dtu.dk/services/NetGene2
	NNSplice	http://www.fruitfly.org/seq_tools/splice.html
	GeneSplicer	http://www.cbcb.umd.edu/software/GeneSplicer/gene_spl.shtml

参 考 文 献

- [1] Horak P, Griffith M, Danos AM, et al. Standards for the classification of pathogenicity of somatic variants in cancer (oncogenicity): Joint recommendations of Clinical Genome Resource (ClinGen), Cancer Genomics Consortium (CGC), and Variant Interpretation for Cancer Consortium (VICC). *Genet Med*. 2022; 24(5): 986-998.
- [2] Li MM, Datto M, Duncavage EJ, et al. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *J Mol Diagn*. 2017; 19 (1): 4-23.
- [3] Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020; 581(7809): 434–443.
- [4] UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res*. 2023; 51(D1): D523-D531.
- [5] Abou Tayoun AN, Pesaran T, DiStefano MT, et al. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat*. 2018 Nov;39(11):1517-1524.
-