



# A Tool for the Interpretation of Sequence Variants in Cancer Based on the Standard and Guidelines

Yue Hu, Yunfei Bai\*

Biological Sciences & Medical Engineering, Southeast University, Nanjing, China

## Email address:

hu\_yue1994@163.com (Yue Hu), whitecf@seu.edu.cn (Yunfei Bai)

\*Corresponding author

## To cite this article:

Yue Hu, Yunfei Bai. A Tool for the Interpretation of Sequence Variants in Cancer Based on the Standard and Guidelines. *Science Discovery*. Vol. 6, No. 2, 2018, pp. 124-129. doi: 10.11648/j.sd.20180602.18

Received: January 20, 2018; Accepted: March 2, 2018; Published: June 22, 2017

**Abstract:** With the application and development of next generation sequencing technology, tumor genome sequencing has become widely available in clinical and research setting. Many groups have published germline variant interpretation and classification systems or tools for use in clinical laboratory reporting. However, interpretation and classification for somatic variants remains plenty of challenges. The meaning of tumor somatic variants for the process of the cancers is uncertain. Based on the standards and guidelines for the interpretation and reporting of sequence variants in cancer, and combined with other evaluation systems and tools, we developed a set of somatic mutation evaluation tools, to automatically apply the part of standards and guidelines, and provide the help for the researchers and clinical scientists.

**Keywords:** Sequence Variants in Cancer, Standards and Guidelines, Next-Generation Sequencing, Somatic Variants Classification

## 基于分类指南的体细胞变异解读分类工具

胡月, 白云飞\*

生物科学与医学工程学院, 东南大学, 南京, 中国

## 邮箱

hu\_yue1994@163.com (胡月), whitecf@seu.edu.cn (白云飞)

**摘要:** 随着二代测序技术的持续发展, 癌症基因组测序越来越多的应用在临床研究领域中。针对遗传变异致病性的解读和评估已经存在众多分类系统和工具, 而对肿瘤体细胞变异的解读还存在着众多的困难, 大部分已知癌症相关基因中的改变对肿瘤形成和发展的意义都是未知的, 我们基于癌症序列变异解释和报告的标准和指南, 并结合其他评价系统和标准, 开发了一套针对体细胞变异评估的工具, 能够自动化应用部分标准和指南, 为临床科研人员提供帮助。

**关键词:** 癌症序列分析, 标准与指南, 高通量测序, 体细胞变异分类

## 1. 引言

高通量测序技术 (High-Throughput sequencing) 也称二代测序技术 (Next-generation sequencing technology),

能一次并行对几十万到几百万DNA分子进行序列测定。[1] 基因组高通量测序技术可用于全基因组、全外显子组, 或者是特定区域进行测序并检测序列变异, 主要包括单碱基变异 (SNVs)、拷贝数变异 (CNVs)、结构变异 (SVs) 和序列插入缺失 (Indels)。

近年来,随着高通量测序的快速发展,基因组二代测序开始逐步应用于临床分子诊断[2]。然而二代测序技术产生了大量的遗传变异数据,在数据解读的方面不断面临新的挑战[3],美国遗传学与基因组学学会ACMG联合分子病理学协会(AMP) 和美国病理学家协会(CAP) 2015年发表了关于孟德尔疾病的序列变异分类最新标准和指导方针[4],美国临床肿瘤学会(ASCO) 和美国病理学家协会(CAP)、分子病理协会(AMP) 发表了肿瘤及疑似肿瘤相关序列变异检测的检测标准[5]。此外,还有许多研究机构对序列变异建立自己的分类标准,;例如ENIGMA研究机构在ACMG分类标准的基础上,发布了专门针对BRCA1/2基因的变异分类标准[6]。

然而在应用指南或分类标准时,由于每个研究小组不同理解及对实现它们的标准算法的缺乏,可能出现模棱两可或互相矛盾的解释。本文旨在从癌症序列变异标准和指南分出发,开发出一套在分类指南的基础上,针对分类标准扩展成一组相关的规则并且分别加权,能够自动化应用分类标准的工具,以支持对变异分类更直观的理解,帮助研究人员评估解释变异的临床意义。

## 2. 数据库及分析工具

### 2.1. 基因组数据库

公共数据库收集了大量的变异数据,对评估解释变异是非常重要的证据资源。

#### 2.1.1. 人群数据库

The Exome Aggregation Consortium	<a href="http://exac.broadinstitute.org/">http://exac.broadinstitute.org/</a>
Exome Variant Server 1000 Genomes Project	<a href="http://evs.gs.washington.edu/EVS">http://evs.gs.washington.edu/EVS</a> <a href="http://browser.1000genomes.org">http://browser.1000genomes.org</a>
dbSNP	<a href="http://www.ncbi.nlm.nih.gov/snp">http://www.ncbi.nlm.nih.gov/snp</a>
dbVar	<a href="http://www.ncbi.nlm.nih.gov/dbvar">http://www.ncbi.nlm.nih.gov/dbvar</a>

#### 2.1.2. 疾病数据库

ClinVar	<a href="http://www.ncbi.nlm.nih.gov/clinvar">http://www.ncbi.nlm.nih.gov/clinvar</a>
OMIM	<a href="http://www.omim.org">http://www.omim.org</a>
Human Gene Mutation Database	<a href="http://www.hgmd.org">http://www.hgmd.org</a>
Catalog of Somatic Mutations in Cancer	<a href="http://cancer.sanger.ac.uk/cosmic">http://cancer.sanger.ac.uk/cosmic</a>
My Cancer Genome Orphanet	<a href="http://www.ncbi.nlm.nih.gov/dbvar">http://www.ncbi.nlm.nih.gov/dbvar</a> <a href="http://www.orpha.net/">http://www.orpha.net/</a>

#### 2.1.3. 其他数据库

CIViC	<a href="https://civic.genome.wustl.edu/">https://civic.genome.wustl.edu/</a>
KEGG	<a href="http://www.genome.jp/kegg/">http://www.genome.jp/kegg/</a>
OncoKB	<a href="http://www.genome.jp/kegg/">http://www.genome.jp/kegg/</a>
PharmGKB	<a href="https://www.pharmgkb.org/">https://www.pharmgkb.org/</a>

### 2.2. 生物信息分析工具

生物信息预测算法和软件常用于预测核苷酸的变化是否会影响对蛋白质结构和功能。

#### 2.2.1. 序列变异功能影响预测工具

SIFT	<a href="http://sift.jcvi.org">http://sift.jcvi.org</a>
PolyPhen2	<a href="http://genetics.bwh.harvard.edu/pph2">http://genetics.bwh.harvard.edu/pph2</a>
MutationAssessor	<a href="http://mutationassessor.org">http://mutationassessor.org</a>
MutationTaster	<a href="http://www.mutationtaster.org">http://www.mutationtaster.org</a>
GERP++	<a href="http://mendel.stanford.edu/sidowlab/downloads/gerp/index.html">http://mendel.stanford.edu/sidowlab/downloads/gerp/index.html</a>
FATHMM	<a href="http://fathmm.biocompute.org.uk">http://fathmm.biocompute.org.uk</a>

#### 2.2.2. 注释工具

大量的遗传变异数据中仅少数变异具有功能意义,选择不同的变异注释工具对下游分析至关重要。

SnPEff[7]是基于Java的分析工具,输入文件的格式为VCF,支持肿瘤突变分析,并且与GATK兼容。

Variant Effect Predictor (VEP) [8]可以对不同类型变异进行注释,包括SNPS, insertions, deletions, CNV和SV,也可以依据不同的数据库进行过滤。结果文件格式为VCF格式,包含功能注释信息,变异频率信息,预测软件预测结果信息及疾病数据库信息。

ANNOVAR[9]注释工具支持多个版本的基因组,除了人类基因组还包括小鼠等物种,支持三种不同形式的注释:

(1) 基于基因的注释: 可以自主选择RefSeq、UCSC、ENSEMBL、GENCODE等基因定义系统,注释信息包括变异类型,蛋白质一级结构改变等信息;(2) 基于区域的注释: 包括位点的保守性、表观遗传标记物靶向性等变异位点所处的基因组环境等信息;(3) 基于筛选的注释: 包括变异位点在不同群体中频率信息,预测位点对蛋白质三维结构影响等信息的注释。

## 3. 肿瘤变异评估标准

根据指导规范,对肿瘤细胞的体细胞变异划分为四个级别: 1类具有确定性临床意义的突变; 2类可能具有临床意义的突变; 3类临床意义不明确的突变; 4类良性或可能良性的突变。其中1类和2类确定具有或可能具有临床意义被细分为4个等级:

(A) 可以作为预测药物反应或耐药性的FDA批准的针对特定类型肿瘤治疗的突变,或者被包括在其他专业指南中如NCCN作为特定类型肿瘤的治疗、诊断或预后的突变;

(B) 可以作为预测药物反应或耐药性的基于充分研究和专家共识的治疗的突变,或者是基于充分研究和专家共识的具有特定疾病诊断、预后意义的突变;

(C) 可以作为预测药物反应或耐药性的FDA或专业协会批准的跨适应症治疗的突变,或是已经作为临床试验的入组参考标准,基于多项研究的具有特定疾病诊断和预后意义的突变;

(D) 基于临床研究、有案例报道的突变,或者有研究表明有主语疾病诊断和预后判断的突变。

根据指导规范,本工具从以下几个方面来对体细胞变异进行评估。

### 3.1. 生物标记物

生物标志物是指一种可客观检测和评价的特性,可作为正常生物学过程、病理过程或者治疗干预药理学反应的指示因子[10],根据潜在用途可将其分为四类:(1)用于疾病诊断;(2)用于检测疾病活动;(3)反应药物作用;(4)与药物代谢动力学有关[11]。

癌症生物标志物数据库(the Cancer Biomarkers database)[12]由 the European Union's Horizon 资助,收集了一千多条由精准肿瘤学领域的临床专家根据FDA准则、NCCN指南等权威指南筛选的生物标志物信息。此数据库也收集了通过临床或实验验证具有致癌性或经过生物信息工具预测为驱动基因的基因,和在癌症基因中观倡导的具有特定突变的临床/实验证实的致癌突变。

### 3.2. 变异类型

突变所处的参开基因功能区通常分为基因间区(intergenic)、外显子区(exonic)、内含子区(intronic)、3'非翻译区(3 prime UTR variant),5'非翻译区(5 prime UTR variant),上游区(upstream),下游区(downstream)等。

目前更多的测序分析集中在外显子区域,根据功能性大致可以将外显子区变异分为非同义突变(nonsynonymous SNV)、同义突变(synonymous SNV)、移码插入(frameshift insertion)、移码缺失(frameshift deletion)、终止子获得(stopgain)、终止子缺失(stoploss)、可变剪切体(splice variant)等。

### 3.3. 变异频率

正常人体基因组中90%以上的变异信息是高频率的(>5%),高频变异往往与疾病无关[13],因此需要对变异进行频率筛选,去掉高频变异。突变频率通常从群体遗传学数据库中获得,比较有代表性的数据有dbSNP、1000G数据库、ExAC。

千人基因组计划(1000 Genomes Projects)[14]由多家国际研究机构共同启动,包含了来自欧洲、非洲、东亚和美洲的14个人类群体1092人的罕见和常见遗传变异信息。

外显子组整合数据库(the Exome Aggregation Consortium, ExAC)[15]是Broad研究所整合了17个国际性人类外显子项目数据而成的数据库,包含近7万个无亲缘关系个体的上千万个变异信息其中很多为罕见变异。

单核苷酸多态性数据库(Single-Nucleotide Polymorphism database, dbSNP)[16]由美国国立生物技术信息中心(National Center for Biotechnology Information, NCBI)与人类基因组研究所(National Human Genome Research Institute)合作建立,为研究者提供丰富的单碱基替换以及短插入、删除多态性信息。

Exome Sequencing Project (ESP)[17]数据库来自国际心肺和血液研究所外显子组测序计划(NHLBI-ESP

project),包含了SNP、inDel和Y染色体上突变碱基的等位基因频率。

### 3.4. 遗传变异数据信息

ClinVar[18]由美国国立卫生研究院(National Institutes of Health, NIH)创建及维护,它通过标准的命名法来描述疾病,整合了十多个不同类型数据库,经专家评审逐步形成一个标准的、可信的、稳定的遗传变异—临床表型相关数据库,对变异有一套自己的分类系统,分为致病、可能致病、良性、可能良性、意义不明确五类。ClinVar支持科研人员将数据下载到本地分析,以便开展更为个性化的研究。

OMIM(Online Mendelian inheritance in man)[19]主要着眼于遗传性的基因疾病,包括所有已知的遗传病、遗传决定性状的基因,除了有各种疾病的临床特征、诊断、鉴别诊断等简略描述外,还提供已知致病基因的连锁关系、染色体定位、组成结构和功能等资料,还包括严格筛选出的相关参考文献。

### 3.5. 体细胞变异数据库信息

COSMIC(Catalogue Of Somatic Mutations In Cancer)[20]是最大最全的人类体细胞突变数据库,每三个月更新一次。数据库信息主要来自两方面,一是从已发表的文献中,二是从癌症基因组计划的全基因组重测序研究中,目前包含了超过五百万编码突变,两万篇文献信息,三万多个肿瘤样本全基因组测序数据。COSMIC对于突变微体的信息记录十分详细,包括文献出处、样品名称、组织类型、设计到的癌症种类和具体的突变内容等。

ICGC(The International Cancer Genome Consortium)[21]由世界各地的研究所组成,包含了来自76项目及21个肿瘤类型的近两万条肿瘤相关的癌症基因组信息。

CIViC[22]是一个由业内人士共同编辑整理的论坛,用于讨论和解读与基于已发表的临床相关的科学文献与研究的肿瘤突变信息,包括分子改变、药物、诊断、预后及其他医疗干预的相关性,目前CIViC包含332个基因,1740个位点,383种药物和207种疾病信息。

### 3.6. 生物信息分析工具预测

比较常见的蛋白质功能预测工具有SIFT[23]、PolyPhen2[24]、MutationTaster[25]、MutationAssessor[26]、CADD[27]、FATHMM[28]等,这些工具一部分是基于进化保守性,还有一部分是基于蛋白质结构域功能或是蛋白的空间结构。SIFT根据进化保守性来预测氨基酸变化对蛋白功能的影响,变异有害性越高,值越小,通常阈值为0.05;PolyPhen2基于蛋白结构同源性算法,其值越大变异有害性越大,通常阈值为0.03;MutationTaster综合考虑蛋白结构、数据库变异频率、OMIM和HGMD数据库突变信息;MutationAssessor则是基于进化保守性,阈值设定为0.85。

剪切位点突变危害分析工具dbSNV[29]是根据突变前后分值的变化来预测剪切位点的突变危害性,基于不同算法给出AdaBoost(ADA)和Random Forest(RF)分值,阈值取0.6。

生物信息分析工具虽常用以预测突变的致病性评估, 但是总体上看生物信息学算法的特异性并不是很高, 其结果只能作为参考, 不能作为评估的唯一证据[30]。

3.7. 生物信号通路相关基因

调控细胞正常生长、分化、增殖和凋亡的信号通路众多, 信号通路中若发生异常, 就会导致细胞癌变, 所以这些信号通路中的重要因子日益成为抗癌药物筛选和癌症治疗汇总的关键靶点。目前, 一些能够阻断致癌信号的分子已被成功应用于临床癌症治疗[31]。

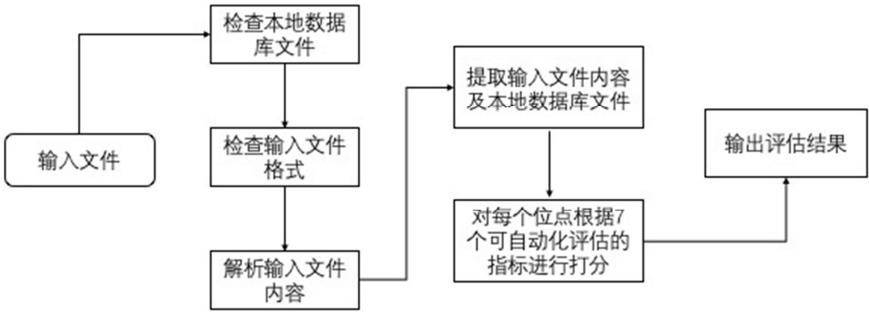


图1 评估流程。

4.1. 本地数据库文件

本地数据库中存放从各个公共数据库中下载整合的文件: (1) 与生物标记物相关文件来自Cancer Genome Interpreter, 最新版本更新至2017/11/02; (2) 从数据库中收集整合的癌症相关基因全称与缩写的文件; (3) 来自CIViC的数据集; (4) 来自遗传变异评估软件Intervar[32]本地资料库的kownGeneCanonical文件, 是基因外显子的第一个或最后一个或在基因的3'端最后50bp以内的位置信息; (5) 从OMIM获取的基因和表型对应信息文件; (6) OMIM中ID号与Orphanet中ID号对应文件; (7) OMIM中ID与MIM Entry Type、NCBI的Entrez Gene ID、Ensembl中ID 对应信息; (8) 经证实的跟疾病相关基因列表; (9) 疾病相关生物信号通路基因列表 (10) 来自Orphanet数据库的信息文件; (11) 在Clinvar和ExAC中经过RefGene注释的致病性位点所在基因列表文件; (12) 来自精密医学知识库 (the Precision Medicine KnowledgeBase, PMKB) 的临床癌症变异和解释信息文件。

4.2. 评估标准与结果

本工具采用对每个变异位点从各项评估标准打分的方式对变异致病性进行评估, 参考ACMG标准采用四级分类, 即确定具有临床意义的突变、可能具有临床意义的突变、意义不明确的突变、良性或可能良性的突变。

(1) 根据SIFT、PolyPhen2、MutTaster、CADD、MetaSVM等几大工具预测结果, 如果有三种及三种以上工具预测变异有害或是可能有害, 则得分为2分, 三种及三种以上工具预测变异为良性的, 致病性得分为0分, 三种以下软件预测变异有害或是可能有害, 得分为1分。

(2) 变异在ClinVar数据库中的临床意义如果是“致病 (Pathogenic)”或是“可能治病 (Likely Pathogenic)”则得

4. 评估流程及结果

本工具采用JAVA编程语言编写, JAVA是一种跨平台, 适合于分布式计算环境的面向对象编程语言, 具有可靠、安全、平台无关、高性能、多线程, 可移植性, 强大的兼容性等众多特性, 可以编写桌面应用、Web应用程序、分布式系统和嵌入式系统应用程序等各类系统及程序。

通过命令行传递参数, 如使用帮助 (-h)、输入文件 (-i)、输出文件 (-o) 等。

分为1分, 若是“良性 (Benign)”或“可能良性 (Likely Benign)”, 或者是“不确定性意义 (Uncertain significance)”则得分为0。

(3) 根据在三大群体数据库即ExAC、1000Genome Project、ESP6500中的最小等位基因频率 (MAF), MAF<0.005的得分为1分, 相反则得分为0。

(4) 变异位点的dbscSNV的ADA得分和RF得分都大于0.6, 并且变异所处的参考基因功能区域, 以及外显子区的变异类型包含“nonsynonymous、missense、nonsense、frameshift、splice、stopgain、stoplost、CNV、fusion”中一种, 或者变异所在基因在致病性位点所在基因列表文件中, 则得分为1分。

(5) 根据基因以及氨基酸水平改变查询在CGI数据库中的信息, 在CGI中注释为相关药物已经获得FDA或是European LeukemiaNet 批准的突变得分为2分; 若是早期临床实验或是已有临床案例证实的则得分为1分。

(6) 突变位点能够在Cosmic数据库中及ICGC中都找到的则为2分, 只在一个中有的则为1分, 都没有的则为0分。

(7) 突变所在基因在癌症相关基因列表中得分为2分, 在癌症相关信号通路基因列表中为1分。

综合以上各项得分, 若是总分为11分则评估为具有确定性临床意义的突变, 若是总分为7分以上则评估为可能具有临床意义的突变, 若得分为0则为不确定性意义的, 其它得分为良性或可能良性。

4.3. 测试集

我们使用了9例肺癌病人的全外显子测序数据, 及5例乳腺癌病人的BRCA1/2外显子测序数据进行测试。在9例肺癌样本中, 有7例找到了具有 / 可能具有临床意义的突变位点; 有2例乳腺癌样本中也找到了这样的位点。

表1 测试集部分结果数据。

Sample	Chr	Pos	Ref	Alt	Gene	Score
L1	7	55249071	C	T	EGFR	8
L2	7	55259515	T	G	EGFR	9
L3	17	7578526	C	A	TP53	7
L4	3	178916946	G	C	PIK3CA	9
L5	5	112170788	T	—	APC	7
L7	12	25398284	C	T	KRAS	9
L9	10	89692923	G	A	PTEN	9
L9	7	55259515	T	G	EGFR	9
B1	17	41276113	T	C	BRCA1	7
B3	13	32911775	C	T	BRCA1	9

## 5. 结论

标准和指南为体细胞变异的解读和分类从各层面提出了指导，然而各个实验室在应用指导时存在不同理解，本工具旨在帮助研究人员建立更统一的标准，更好的对体细胞突变进行解读。

本工具为自动化应用指南提供了一种方法，但目前还存在很多限制，比如在临床诊断、预后等很多方面的评估方法尚未明确，与药物相关的描述难以量化，针对不同来源、不同实验方法的数据评估结果是否存在差异未有统计，对结论打分权值的选择也在进一步优化中。下一步我们将收集更多信息，纳入更多的临床评估标准，对没有计算机基础的研究者，命令行应用也存在一定困难，为解决这一问题，网页版应用工具在开发中。

## 参考文献

- [1] Tsimberidou AM: Targeted therapy in cancer. *Cancer Chemother Pharmacol* 2015, 76:1113-1132.
- [2] Garraway L A, Verweij J, Ballman K V. Precision oncology: an overview. [J]. *Journal of Clinical Oncology Official Journal of the American Society of Clinical Oncology*, 2013, 31(15):1803-5.
- [3] Jones S, Anagnostou V, Lytle K, et al. Personalized genomic analyses for cancer mutation discovery and interpretation [J]. *Science Translational Medicine*, 2015, 7(283):283ra53.
- [4] Richards S, Aziz N, Bale S, et al. Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology [J]. *Genetics in Medicine Official Journal of the American College of Medical Genetics*, 2015, 17(5):405.
- [5] Li M M, Datto M, Duncavage E J, 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]. *Journal of Molecular Diagnostics Jmd*, 2017, 19(1):4-23.
- [6] Spurdle A B, Healey S, Devereau A, et al. ENIGMA-Evidence-based network for the interpretation of germline mutant alleles: An international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes [J]. *Human Mutation*, 2012, 33(1):2-7.
- [7] Cingolani P, Platts A, Wang L L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. [J]. *Fly*, 2012, 6(2):80.
- [8] McLaren W, Gil L, Hunt S E, et al. The Ensembl Variant Effect Predictor [J]. *Genome Biology*, 2016, 17(1):122. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from next-generation sequencing data *Nucleic Acids Research*, 38:e164, 2010.
- [9] Adzhubei I, Jordan DM, Sunyaev SR: Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* 2013, Chapter 7:Unit7.20.
- [10] Blackhall F, Peters S, Kerr K M, et al. Biomarkers [J]. *Annals of Oncology*, 2012, 23:ix73-ix94.
- [11] McDonagh E M, Whirl-Carrillo M, Garten Y, et al. From pharmacogenomic knowledge acquisition to clinical applications: the PharmGKB as a clinical pharmacogenomic biomarker resource [J]. *Biomarkers in Medicine*, 2011, 5(6):795.
- [12] Tamborero D. The Cancer Genome Interpreter: a tool to annotate the biological and clinical relevance of tumor alterations. *Basic & Clinical Pharmacology & Toxicology [serial online]*. n.d.;121:30.
- [13] Shen M R, Jones I M, Mohrenweiser H. Nonconservative amino acid substitution variants exist at polymorphic frequency in DNA repair genes in healthy humans [J]. *Cancer Research*, 1998, 58(4):604.
- [14] 1000 Genomes Project Consortium, Abecasis G R, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes [J]. *Nature*, 2012, 491(7422):56-65.
- [15] Lek M, Karczewski K J, Minikel E V, et al. Analysis of protein-coding genetic variation in 60,706 humans [J]. *Nature*, 2016, 536(7616):285.
- [16] Sherry S T, Ward M H, Kholodov M, et al. dbSNP: the NCBI database of genetic variation [C]// *Nucleic Acids Res*. 2001: 308-311.
- [17] Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G, Kang HM, Jordan D, Leal SM, Gabriel S, Rieder MJ, Abecasis G, Altshuler D, Nickerson DA, Boerwinkle E, Sunyaev S, Bustamante CD, Bamshad MJ, Akey JM; Broad GO; Seattle GO; on behalf of the NHLBI Exome Sequencing Project. 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*. 337(6069):64-9.
- [18] Landrum, M. J. et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res*. 44, D862-8 (2015).
- [19] McKusick V A. Mendelian Inheritance in Man and its online version, OMIM[J]. *American Journal of Human Genetics*, 2007, 80(4):588-604.

- [20] Bamford S, Dawson E, Forbes S, et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website [J]. *British Journal of Cancer*, 2004, 91(2):355.
- [21] Bamford S, Dawson E, Forbes S, et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website [J]. *British Journal of Cancer*, 2004, 91(2):355.
- [22] Griffith M, Spies N C, Krysiak K, et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer [J]. *Nature Genetics*, 2017, 49(2):170.
- [23] Sim N-L, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC: SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* 2012, 40:W452-W457.
- [24] Adzhubei I, Jordan DM, Sunyaev SR: Predicting functional effect of human missense mutations using Poly Phen-2. *Curr Protoc Hum Genet* 2013, Chapter 7:Unit7.20.
- [25] Schwarz JM, Cooper DN, Schuelke MSD: Mutation Taster 2: mutation prediction for the deep-sequencing age. *Nat Methods* 2014, 11: 361-362.
- [26] Reva B, Antipin Y, Sander C: Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res* 2011, 39-118.
- [27] Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J: A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014, 46:310-315.
- [28] Shihab H A, Gough J, Mort M, et al. Ranking non-synonymous single nucleotide polymorphisms based on disease concepts [J]. *Human genomics*, 2014, 8(1):11.
- [29] Liu X, Wu C, Li C, et al. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs [J]. *Human Mutation*, 2016, 37(3):235-241.
- [30] Sukhai M A, Craddock K J, Thomas M, et al. A classification system for clinical relevance of somatic variants identified in molecular profiling of cancer.[J]. *Genetics in Medicine*, 2015, 18(2): 515–522.
- [31] Hunter T. Signal Transduction in Cancer and the DNA Damage Response [J]. *Nature Cell Biology*, 2007, 9(11):1311-8.
- [32] Li Q, Wang K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines [J]. *American Journal of Human Genetics*, 2017, 100(2):267.