

专家共识

肿瘤基因诊断二代测序生物信息学分析平台本地化建设中国专家共识(2024年版)

中国抗癌协会肿瘤基因诊断专业委员会,中国抗癌协会整合肿瘤学分会

【摘要】随着肿瘤基因诊断对精准性要求的不断提升,二代测序(next-generation sequencing, NGS)技术在肿瘤基因诊断中的应用愈加广泛。为应对临床实践中的紧急需求,部分医疗机构计划自建肿瘤基因诊断数据分析平台。NGS的基因诊断数据分析依赖完善的生物信息学分析平台,而目前尚缺乏统一的本地化平台建设共识。为此,中国抗癌协会肿瘤基因诊断专业委员会和中国抗癌协会整合肿瘤学分会组织中国专家组经过深入讨论,制定了本专家共识,并提出15条建议,涵盖了平台建设的硬件需求、数据存储与管理、分析流程搭建与性能验证、分析流程管理、人员配置与培训管理,以及相关法律法规等方面。本专家共识旨在为医疗机构提供NGS生物信息学分析平台本地化建设指导,以期提高肿瘤基因诊断的准确性及效率,推动肿瘤基因诊断行业的标准化建设与可持续发展。

【关键词】肿瘤基因诊断;二代测序;生物信息学分析;本地化建设;共识

【中图分类号】 R730.4 **【文献标识码】** A **【文章编号】** 1674-5671(2024)05-0000-0633-12

DOI: 10.3969/j.issn.1674-5671.2024.06.01

Chinese expert consensus on the localization of bioinformatics analysis platform for tumor genetic diagnosis by using next-generation sequencing (2024 edition)

Tumor Gene Diagnosis Committee of the China Anti-Cancer Association; Integrative Oncology Branch of the China Anti-Cancer Association

【Abstract】 With the growing demand for precision in tumor gene diagnostics, next-generation sequencing (NGS) technology is increasingly being applied in tumor genetic diagnosis field. To meet pressing needs in clinical practice, some medical institutions are preparing to establish their own tumor gene diagnostic data analysis platforms. NGS-based gene diagnostic analysis relies on a robust bioinformatics platforms; however, there is currently no consensus on the construction of a unified localized platforms. In response, the Chinese experts from the Tumor Gene Diagnosis Committee of the China Anti-Cancer Association and the Integrative Oncology Branch of the China Anti-Cancer Association to develop this expert consensus and put forward 15 consensus guidelines after thorough discussions. The guidelines offer medical institutions detailed instructions for localizing NGS bioinformatics platforms, including hardware requirements, data storage and management, design and validation of analysis workflows, workflow management, personnel allocation and training, as well as relevant legal and regulatory considerations. The aim of this consensus is to enhance the accuracy and efficiency of tumor gene diagnostics, while fostering the standardization and sustainable development of the industry.

【Keywords】 Tumor gene diagnosis; Next-generation sequencing; Bioinformatics analysis; Localized construction; Consensus

肿瘤是全球范围内威胁人类健康的主要疾病之一。尽早发现和精准诊断,并通过全面分析肿瘤基因组指导个体化治疗策略,对提高肿瘤治疗效果和改善患者预后至关重要。随着高通量二代测序(next-generation sequencing, NGS)技术的迅猛发展,基因组学在医学领域的应用日益广泛,特别是在肿瘤基因诊断方面。靶向药物和靶向治疗的应用对肿瘤基因诊

断提出了更高的精准度要求。生物信息学分析作为NGS数据分析的必要方法和步骤,是肿瘤基因诊断不可或缺的关键环节。肿瘤临床检测中的生物信息学分析的核心任务是将原始的NGS序列数据解析为基因组变异结果,这与肿瘤管理和患者治疗方案的选择紧密相关,并直接影响患者预后^[1-2]。

目前,肿瘤基因诊断工作大多由医疗机构及商业

【基金项目】 国家自然科学基金资助项目(82270159, 82070147, 82472988)

公司在诊疗单位本地搭建配套的生物信息学分析平台开展。然而,肿瘤基因诊断NGS生物信息学分析平台的本地化建设尚未形成统一共识,导致各机构在平台的建立、实施和验证上存在较大差异,这可能影响结果的准确性及临床解读,进而影响肿瘤患者的治疗^[3]。随着肿瘤NGS技术的普遍应用,对肿瘤基因诊断的需求不断提高,制定完整的、可参考的NGS生物信息学分析平台本地化建设指南显得尤为重要^[2]。为此,中国抗癌协会肿瘤基因诊断专业委员会与中国抗癌协会整合肿瘤学分会,联合全国16个省市的病理学、检验学、分子生物学及生物信息学等领域的40余位专家,共同制定了本共识。本共识已在国际实践指南注册平台(international practice guideline registry platform, IPGRP)上注册,编号为PREPARE-2024CN561。共识的形成经过多轮专家研讨会讨论,并进行了对初拟推荐建议的投票决议,投票设同意、不确定、不同意3个选项,专家可在投票时提出修改意见。每次调查结束后,根据专家反馈意见进行修改或增补。选择“同意”的专家比例 $\geq 70\%$ 则认为该条推荐达成一致,纳入本共识。最终从生物信息学平台的本地化建设基本硬件需求、数据存储与管理、分析流程搭建、分析流程的性能验证、分析流程管理、人员配置与培训管理,以及相关法律法规等7个方面提出了15条建议,旨在推动肿瘤基因诊断的标准化和精准化。

1 基本硬件需求

肿瘤基因诊断NGS生物信息学分析本地化平台属于数据体量大、计算密集型系统。为满足不同的检测样本规模和报告周期需求,本地化平台应选择合适的服务器或包含多计算节点规格的服务器集群,合理规划部署计算资源和存储资源。如有可能,平台应建设独立的计算机器房,并配置相应的管理及灾备方案,确保干实验和湿实验分离,为计算硬件提供良好的运行环境。

选择服务器时,应兼顾成本预算、分析需求、报告周期、最大样本处理量及业务扩展等因素,综合评估以下关键参数:CPU核心数量、内存容量、存储空间和网络连接能力。非集群方案需配置备用分析服务器,其操作系统、数据库、软件环境和网络配置与主分析服务器保持一致,并定期测试,以确保在紧急情况下

能够快速启动并接管主服务器的工作。为简化网络布线并确保网络性能,建议备用服务器应与主服务器放置在同一机柜。此外,应选择专业成熟的设备供应商,确保其技术支持和服务对生物信息学分析本地化平台的支撑作用。

专家共识1:肿瘤基因诊断NGS生物信息学分析本地化平台的硬件建设,应根据本地临床分析需求选择高性能服务器或多节点计算机集群,确保分析流程的高效运行和业务扩展的灵活性。设备购置和部署应尽可能选择能提供全面、长期技术支持的供应商。

2 数据存储与管理

在临床肿瘤基因诊断检测中,NGS数据存储是数据管理和伦理规范中的核心环节。NGS数据存储对后续分析、解读以及复审至关重要。NGS数据存储管理主要包含数据的命名标准性、完整性、安全性、稳定性、易访问和可追溯性等方面。

2.1 数据存储命名标准

在肿瘤基因诊断的NGS生物信息学分析中,制定高效标准的文件命名规则能确保数据的有序性和唯一性。命名规则需包含样本ID、测序类型、日期、序列编号及数据类型,以便于数据的快速检索和追踪。文件名可采用缩写,确保与诊断项目关联,并记录在文档中供查阅。建议使用统一的分隔符(如下划线“_”、短横线“-”),避免中英文字符混用,同时避免空格、斜杠等特殊字符,以确保在不同操作系统中的兼容性。为确保项目分析周期内文件名的唯一性,可对测序批次及样本使用四位或以上数字编号。目录结构可按多级分类,将测序数据、分析结果和报告分别存放,测序数据目录下每个批次单独设立子文件夹。FASTQ文件命名应简洁,包含项目缩写、日期、编号和样本类型,例如“XXX2024090001N_1.fastq.gz”表示2024年9月XXX项目的肿瘤对照样本。分析目录按项目设立子文件夹,如“211124_NB551417_XXX_01”表示该批次下的第一次XXX分析。科学的文件命名和目录结构设计能减少命名错误,降低管理复杂性,确保NGS分析流程的高效性和规范性。

专家共识2:肿瘤基因诊断NGS生物信息学分析本地化平台建设应建立一个包含关键信息且简洁明了的命名规则,对文件进行唯一标识。可采用统一的

命名规则和多级目录,提高样本的识别追溯能力及访问检索效率。

2.2 数据存储完整性和安全性管理

确保数据的完整性是数据存储的核心要求。在数据存储过程中,应采用多层次的加密策略,特别是在数据传输和静止状态下应使用高级加密标准(如 advanced encryption standard, AES),以确保数据的完整性。此外,定期使用校验和哈希函数(如安全哈希算法 256 位, SHA-256)验证数据的完整性,以防止在存储和传输过程中发生数据篡改。在备份策略方面,采用“祖父-父-子”(Grandfather-Father-Son)备份策略,测序原始文件按照批次实时备份,分析结果每周备份,归档数据每月备份。建议使用支持断点续传和 MD5 校验的工具编写备份脚本并自动执行。

数据安全性同样至关重要。在肿瘤基因诊断 NGS 数据分析中,应遵循“最小权限”原则,为不同角色用户分配适当的访问权限,以确保数据安全。在对接临床样本信息系统时,必须遵循国家相关隐私保护法规,使用加密传输协议(如 TLS/SSL 协议)保障患者信息在传输过程中的安全。此外,应建立严格的数据脱敏和匿名化处理流程,在分析前去除样本身份识别信息,并使用安全信息和事件管理(security information and event management, SIEM)系统或用户行为分析工具监控异常行为,以及时响应潜在的安全威胁。

专家共识 3: 肿瘤基因诊断 NGS 数据存储和传输应采用高级加密标准,并使用校验工具验证数据完整性,定期执行数据备份。设置基于角色的访问控制管理权限,并对数据进行脱敏和匿名化处理,严格保障患者信息安全。

2.3 数据存储稳定性管理

生物信息学本地化平台可根据存储规模选择不同的存储方案。存储容量规划应满足系统安装、软件环境部署和公共数据库资源存放等基本要求,并按照测序仪器单次最大生产数据量,分析占用空间并结合年度实验频次,综合评估所需空间容量。本地化平台分析服务器存储和原始数据存储应满足至少 1 年的存储需求,原始数据备份至少存储 15 年时间。分析服务器多采用独立硬盘冗余阵列(redundant array of independent disks, RAID)的直接附加存储、网络附属存储(network

attached storage, NAS)或 IB 网络连接分布式存储。备份存储配置建议以 5 年为建设周期,按照年度数据生产量和预期增长速度规划容量。为避免硬盘故障或空间扩展性不足造成数据损失,建议数据备份选择 NAS、磁盘阵列存储服务器或分布式存储。本地化平台应定期对存储系统进行性能监控和及时优化,以满足数据分析需求。此外,建议增设异地备份,以降低自然灾害或火灾等不可抗拒因素造成的数据丢失风险。条件允许的单位,可将原始数据备份至国家基因组科学数据中心肿瘤基因诊断数据分中心进行长期保存。

对于配置独立运算服务器机房的生物信息学本地化平台,应部署不间断电源(uninterruptible power supply, UPS)系统,并监控机房的温湿度及电力供应稳定性,制定详尽的灾难恢复计划(数据恢复时间目标和数据恢复点目标等),确保数据存储的稳定性。

专家共识 4: 肿瘤基因诊断 NGS 生物信息学分析本地化平台需根据业务规模和增长预期,选择适合的存储解决方案,并定期监控和优化存储系统性能,确保存储配置的稳定性、可扩展性和灵活性以满足临床检测需求。NGS 原始数据应保存超过 15 年,条件允许的实验室,建议将原始数据上传至国家基因组科学数据中心肿瘤基因诊断数据分中心的相应数据库进行长期保存。

2.4 数据访问及追溯管理

肿瘤基因诊断 NGS 分析数据报告涉及病患敏感信息,数据管理视同为医疗机构信息管理内容,应参照相关规定设置数据访问规则。同时,NGS 测序原始文件和分析初始文件一般在检测实验室保存和备份,应保证其易访问和可追溯。建议本地化平台设置数据管理专员或小组,建立并维护一个高效的数据管理系统并制定标准操作规范,确保数据在整个存储周期内保持可用状态。数据管理系统须详细记录数据的生成时间、修改记录及访问记录。管理系统建议使用自动化日志记录工具,确保数据操作都有详细记录,包括操作时间、操作人员和操作内容。管理系统中的数据应进行版本控制,若分析结果数据被修改或更新,则需要保留旧版本的副本,便于数据回溯和比对。数据的调取及共享应符合相关法律法规、医学伦理和本地化平台所在单位管理规定。数据管理员应定期检查数据访问记录,确保所有访问和操作均符合安全和隐私规定。

专家共识5:肿瘤基因诊断NGS生物信息学分析本地化平台应建立高效的数据管理系统,详细记录数据的生成、修改和访问记录,实行数据版本控制,定期检查数据访问记录,确保数据的易访问性和可追溯性。

3 分析流程搭建

3.1 检测位点的确定

肿瘤基因诊断NGS检测在疾病诊断、分子分型、靶向治疗、预后预测等方面具有重要价值。不同癌种的检测位点存在差异,为确保目标位点被有效检出,分析流程应参考对应位点的指南和共识^[1,3-12],同时明确突变类型及风险分级。

3.2 分析流程及软件选择

肿瘤基因诊断NGS生物信息学分析涵盖多个关键步骤,包括原始数据拆分、数据清洗、数据比对、变异检测及过滤、变异分析等(图1)。在此过程中,需综合考虑NGS原始数据产生过程中的“湿实验”环节中的各种因素和检测目标,选择经广泛验证和测试的开源或商业软件^[2,7-8,13-14]。常用软件如表1所示。

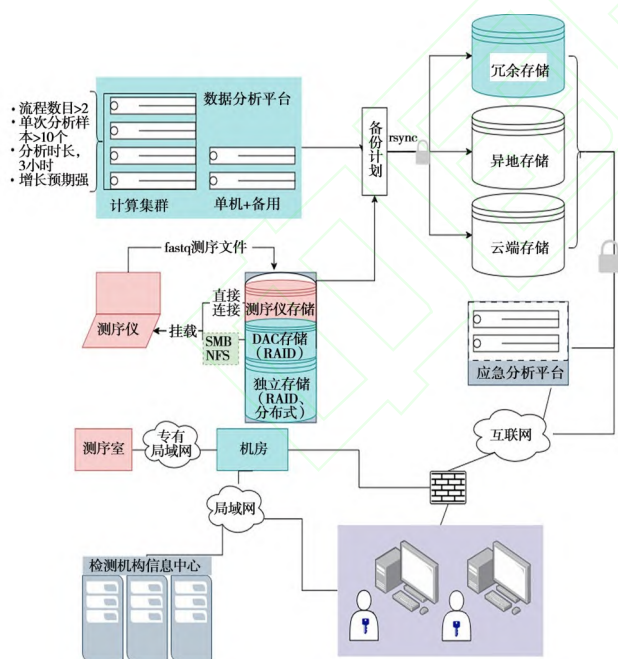


图1 肿瘤基因诊断NGS生物信息学分析流程图

Fig.1 Flow chart of NGS bioinformatics analysis for tumor gene diagnosis

3.2.1 数据拆分 NGS技术生成的原始数据通常是包含多个样本的混合数据,需根据样本的barcode或index拆分数据。拆分后的数据转换为FASTQ格式文

件,通常以压缩格式fastq.gz保存。这些文件包含读段(reads)的碱基序列、质量和标签等信息。

3.2.2 数据清洗和质量评估 原始FASTQ数据需经过质量评估和过滤,以去除低质量序列、接头序列以及过短序列等,确保后续分析的可靠性。可使用FastQC获取质量报告,使用fastp^[15]、Trimmomatic^[16]或Cutadapt^[17]等软件进行数据过滤。

3.2.3 序列比对及去重 序列比对是质量合格reads比对到基因组的过程。在肿瘤NGS分析中参考序列可选择较新的参考基因组,如GRCh38(hg38)。比对需根据数据或测序策略选择合适的软件,建议使用最新版本的BWA软件^[18],现阶段最新版本为0.7.18。序列比对结果采用经典的BAM/SAM或其压缩格式CRAM/SECRAM输出。比对后需采用SAMtools^[19-20]、Picard或Sambamba^[21]去重。

3.2.4 变异检测和分析 胚系突变的检测通常使用GATK HaplotypeCaller^[22-23]、SAMtools^[19-20]、Freebayes^[24]、Sentieon^[25]、DeepVariant^[26]等软件;体细胞突变的检测则使用GATK MuTect2^[22-23]、VarScan 2^[27]、VarDict^[28]、Sentieon TNseq^[29]、NeuSomatic^[30-31]和Strelka2^[32]等。体细胞变异检测建议采用配对样本,使用单样本时可利用公共或自建胚系数据库鉴定胚系变异并注明检测局限性。变异分析建议使用2种以上软件分别计算并比较结果,人工复核后出具报告。

3.2.5 复杂变异检测和分析 在肿瘤基因诊断中,复杂基因组变异的检测分析是指导肿瘤患者治疗和预后的关键组成部分。复杂变异包括多种类型,如拷贝数变异(copy number variation, CNV)、微卫星不稳定性(microsatellite instability, MSI)、肿瘤突变负荷(tumor mutation burden, TMB)、同源重组修复缺陷(homologous recombination deficiency, HRD)和基因融合检测(fusion)等。CNV分析按照检测方式可分为三类:单样本分析、配对样本分析及群体正常对照样本分析。单样本分析没有参考对照,仅能检测单样本的绝对拷贝数。配对样本分析时,可检测与对照相比较的相对拷贝数差异。在基于人群的研究中,通常使用基于正常样本组的分布分析检测肿瘤样本中的拷贝数变化。常用的CNV分析软件包括针对全外显子或者全基因组数据开发的CNVkit^[33]、VarScan 2^[27]、PureCN^[31]、Sequenza^[34]、TitanCNA^[35]、XHMM^[36]或CNVnator^[37]

表1 肿瘤基因诊断常用的分析软件

Tab.1 Commonly used analysis software for tumor gene diagnosis

| 功能 | 软件名 | 网址 |
|-----------|----------------------|---|
| 数据清洗和质量评估 | fastp | https://github.com/OpenGene/fastp |
| | Trimmomatic | https://github.com/timflutre/trimmomatic |
| | Cutadapt | https://github.com/marcelm/cutadapt |
| | FastQC | https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ |
| 序列比对 | BWA | https://github.com/lh3/bwa |
| PCR 去重 | SAMtools* | http://www.htslib.org/ |
| | Picard | https://broadinstitute.github.io/picard/ |
| | Sambamba | https://github.com/biod/sambamba |
| 变异检测和分析 | GATK HaplotypeCaller | https://gatk.broadinstitute.org/hc/en-us/articles/360037225632-HaplotypeCaller |
| | Sentieon | https://www.sentieon.com/ |
| | DeepVariant | https://github.com/google/deepvariant |
| | GATK MuTect2 | https://gatk.broadinstitute.org/hc/en-us/articles/360037593851-Mutect2 |
| | Sentieon TNseq | https://support.sentieon.com/manual/TNseq_usage/tseq/ |
| | NeuSomatic | https://github.com/bioinform/neusomatic |
| | Strelka2 | https://github.com/Illumina/strelka |
| | 复杂变异检测和分析 | CNVkit |
| 复杂变异检测和分析 | VarScan 2* | https://github.com/Jeltje/varscan2 |
| | PureCN | https://github.com/lima1/PureCN |
| | Sequenza | https://github.com/oicr-gsi/sequenza |
| | TitanCNA | https://github.com/gavinha/TitanCNA |
| | XHMM | https://github.com/RRafiee/XHMM |
| | CNVnator | https://github.com/abyzovlab/CNVnator |
| | DECoN | https://github.com/RahmanTeam/DECoN |
| | CoNVaDING | https://github.com/molgenis/CoNVaDING |
| | panelcn.MOPS | https://github.com/bioinf-jku/panelcn.mops |
| | ExomeDepth | https://github.com/vplagnol/ExomeDepth |
| | CODEX2 | https://github.com/yuchaojiang/CODEX2 |
| | MSI 分析 | MSIsensor |
| MANTIS | | https://github.com/TIGER-AI-Lab/Mantis |
| mSINGS | | https://bitbucket.org/uwlabmed/msings/src/master/ |
| 量化基因组瘢痕程度 | scarHRD | https://github.com/sztup/scarHRD |
| DNA 融合分析 | LUMPY | https://github.com/arq5x/lumpy-sv |
| | DELLY | https://github.com/dellytools/delly |
| | FACTERA | https://github.com/FredHutch/easybuild-life-sciences/tree/main/lh_easyconfigs/l/factera |
| | CREST | https://github.com/crest-lab/crest |
| RNA 融合分析 | STAR-Fusion | https://github.com/STAR-Fusion/STAR-Fusion |
| | Arriba | https://github.com/oicr-gsi/arriba |
| | CICERO | https://github.com/stjude/CICERO |
| | FusionCatcher | https://github.com/ndaniel/fusioncatcher |
| 质控参数设置 | fastp | https://github.com/OpenGene/fastp |
| 结果注释 | hgvs | https://hgvs.readthedocs.io |
| | mutalyzer | https://github.com/mutalyzer/mutalyzer |

等软件,或采用针对特定区域靶向测序(targeted next-generation sequencing, Tg-NGS)数据设计的DECoN^[38]、CoNVaDING^[39]、panelcn.MOPS^[40]、ExomeDepth^[41]或CODEX2^[42]等。CNV软件的检测性能取决于数据集,建议使用2种以上软件分别计算并比较结果,关键区段建议进行实验验证,人工复核后出具报告。

根据检测算法原理,常见的MSI分析有需要正

常对照的MSIsensor^[43]、MANTIS^[44]软件和不需要正常对照的mSINGS^[45]、MSI-Fone和MSI-ColonCore^[46]软件等。注意所有基于NGS的MSI检测平台在临床应用前,必须与多重荧光PCR毛细管电泳法或免疫组织化学法进行性能验证,测试并确定检测算法的临床检测准确性。TMB一般是指特定区域内的体细胞变异的个数,包含单核苷酸变异(single nucleotide variant,

SNV),多核苷酸变异(multiple nucleotide variant,MNV),插入/缺失(insertion and deletion,Indel),通常用每兆碱基突变个数表示。TMB检测受样本质量、肿瘤纯度和检测位点数目等因素影响,目前尚无特定成熟的检测软件,临床应用前应验证计算结果和全外显子测序的一致性。HRD临床检测所描述的肿瘤基因组特定改变,也被称为“基因组瘢痕”。杂合性缺失(loss of heterozygosity, LOH)、端粒等位基因不平衡(telomeric allelic imbalance, TAI)、大片段迁移(large-scale state transition, LST)等被作为基因组瘢痕标志物,以量化基因组瘢痕的程度,可使用scarHRD^[47]计算HRD分值。Fusion检测可大致分为依赖于基因组重排融合(DNA水平变化)与不依赖于基因组重排融合(RNA水平变化)两大类,建议同时检测DNA和RNA的融合以提高检测的准确性,其中DNA融合分析软件分析软件主要包括LUMPY^[48]、DELLY^[49]、FACTERA^[50]和CREST^[51]等,RNA融合分析软件包括STAR-Fusion^[52]、Arriba^[53]、CICERO^[54]和FusionCatcher^[55]等。

专家共识6:肿瘤基因诊断NGS生物信息学分析本地化平台流程搭建应根据样本类型、测序平台和检测目的,结合软件性能评估,合理选择软件组合,确保分析结果的可靠性和准确性。

3.3 流程质控参数设置

生物信息学分析流程包含多个环节,根据样本及软件应用场景,应在每个环节设置对应的分析参数^[56]。主要包括以下方面:(1)处理原始FASTQ数据时,需要设置一些关键参数。对于样本标签序列的碱基错配数,通常设置为0或1。碱基质量值阈应确保Q30碱基比率大于90%,Q20碱基比率大于95%。建议计算10×、20×和30×的覆盖度,通过变异系数(coefficient of variation, CV)来评估数据的均一性。一些软件(如fastp)支持自动检测和去除序列末端或中间的多聚N碱基,这些多聚物通常是由测序错误引起,而非源于样本本身^[57]。(2)数据比对及质控。设置参考基因组及索引文件,确保参考基因组文件和索引文件正确且一致。设置比对错配参数时,需要根据测序错误率适当调整,过高会降低比对率,过低会增加错误比对率。临床DNA测序多使用BWA比对,reads小于70 bp应采用BWA-backtrack策略,reads在70 bp以上的应采用BWA-MEM策略^[18,58]。建议计算30×、50×、100×和

500×覆盖度以评估测序覆盖的均一性,胚系检测50×覆盖率不低于95%,肿瘤检测500×覆盖率不低于95%。靶标区测序reads数目总占比不低于75%^[3]。(3)SNV和Indel分析。先根据样本类型明确目标检测限,在此基础上设置特定变异检测所需的最小测序深度,变异等位基因被识别为真实变异所需的最小变异频率(variant allele frequency, VAF)和支持reads数目。根据期望的Indel率,设置合理的间隙开放惩罚参数,对Indel丰富的样本应适当降低,复杂区域需要重比对并进行质量值矫正^[59]。(4)CNV检测时一般需要设置同一个体正常组织对照样本,根据数据集训练结果合理自定义拷贝数扩增和缺失的阈值。Tg-NGS捕获区域较小,受区域选择、GC偏好等因素影响较大,应避免不同平台测序数据的交叉使用。(5)TMB、HRD和MSI检测。需要考虑肿瘤细胞占比和Tg-NGS目标捕获区域大小的影响。计算TMB和HRD评分时,通常要求肿瘤不低于20%,TMB稳定性与Tg-NGS目标捕获区域大小相关,覆盖范围>1 Mb的结果趋于稳定,建议介于1.5~3.0 Mb^[8-9]。设计Tg-NGS检测HRD时,应优先选择中国健康人群频率在0.4~0.6的SNPs位点。MSI检测需要明确微卫星位点的覆盖情况,并根据抽样实验确定最低覆盖位点数目及各位点权重。

专家共识7:肿瘤基因诊断NGS生物信息学分析本地化分析应设置包含但不限于碱基质量、覆盖均一性、测序读长、变异位点的测序深度、变异频率、支持读数及测序平衡性等质控参数。各项参数应根据本地分析平台的样本特性、测序技术和目标检测限精细化调整。在检测CNV、TMB、HRD和MSI等复杂变异时,还需考虑检测覆盖度和肿瘤纯度对结果的影响。

3.4 分析结果报告前的质控过滤参数设置

在NGS变异检测分析中,分析参数可做第一层变异过滤,但仍可能存在假阳性结果,因此需通过设置过滤参数过滤假阳性结果^[56]。过滤参数包括突变频率、突变质量及背景噪音等。通常过滤策略应包含以下方面:(1)对照检测过滤策略。存在对照样本时,比对检测样本与对照样本的变异结果,排除在对照样本中也存在的背景性变异,该法可有效排除克隆性造血干扰。(2)批内过滤策略。同一批次多个样本数据联合分析,可反映背景噪音,但该法不适用于对敏感性要求高的热点变异。(3)基线过滤策略。利用公开

基线数据库(如 1000 Genomes 和 gnomAD),根据本地化平台的方法学和条件,积累并构建本地基因变异位点的基线数据库,对背景噪音信息进行过滤,将检测到的变异与基线数据库中的变异频率进行比较,过滤掉背景范围内的位点。(4)基因组特征区域过滤策略。识别并过滤在基因组重复序列区域、低复杂度区域及简单重复序列区域等处的变异,这些区域易产生假阳性。可使用公共数据库(如 UCSC、Ensembl)提供的基因组注释文件,标记重复序列、低复杂度区域等。(5)黑名单过滤策略。若一种方法、一个实验室或一批试剂总是在同一变异上产生高背景或者假阳性,可利用预先定义的“黑名单”区域过滤,提升特异性。(6)白名单过滤策略。针对已知的热点区域设置“白名单”,提升检测敏感性。

专家共识 8: 肿瘤基因诊断 NGS 生物信息学分析本地化平台流程需建立多种过滤策略,以降低假阳性结果。过滤策略包括但不限于对照检测过滤策略、批内过滤策略、基线过滤策略、基因组特征区域过滤策略、黑名单和白名单过滤策略。

3.5 分析结果注释

在进行 NGS 变异结果注释时,应选择合适的转录本。首选与检测的肿瘤类型和组织来源相关的经典转录本版本,其次建议采用 MANE 转录本^[60]。变异需要注释的包括但不限于描述性信息、证据支撑和功能频率数据库信息。其中,描述性信息主要包括碱基改变、蛋白改变、突变类型及外显子内含子信息等,该过程通常采纳 HGVS 标准命名规则。对于 Indel,存在 5'端对齐和 3'端对齐的表示方式,VCF 文件通常使用 5'端对齐表示,而 HGVS 命名系统使用 3'端对齐表示,这可能导致在自动化生成 HGVS 命名时出现错误,建议部署 hgvs、mutalyzer 等工具辅助校正。证据支撑和功能频率数据库信息主要包括支撑突变可信度的信息、人群频率及功能等突变背景信息等。其中,人群频率及功能数据库主要包括人群频率、肿瘤相关数据、突变功能预测性和热点标识等^[1,6,10-12,61]。

专家共识 9: 肿瘤基因诊断 NGS 生物信息学本地化分析变异注释应包括人群频率(特别关注 East Asia 人群)、生物功能、基因致病性等内容。报告基因优先选择经典转录本和 MANE 版本。注释应遵循 HGVS 标准,并校验 Indel 的对齐方式。

3.6 结果的规范性输出

肿瘤变异检测分析及注释结束后,需要把其中主要的信息解读后反馈给临床医生,该过程需要按照确定的模板进行结果输出,模板中应包含以下信息:(1)样本基本信息;(2)检测范围和检测内容;(3)检出的变异及其注释信息;(4)质控指标;(5)结果解读;(6)检测方法的使用范围和局限性。按照上述模板指定原则,对应具体的变异检测内容,需包含以下内容:(1)SNV、Indel 的描述包括染色体位置、参考碱基和变异碱基、基因名称和转录本号、蛋白质变化、测序深度、支持 reads,注释包括临床意义、功能注释、人群数据库频率等;(2)CNV 描述包括染色体区域、涉及基因、拷贝数变化类型、拷贝数、log2 值;(3)Fusion 描述包括融合基因名称、染色体位置、支持 reads(根据是否跨越融合断点分别列举);(4)微卫星不稳定性状态描述包括 MSI 状态,检测方法覆盖的 MSI 位点数目、测序质控合格的 MSI 位点数目、MSI 得分;(5)TMB 描述包括 TMB 值、Tg-NGS 目标捕获区域大小、纳入突变的数目、TMB 状态;(6)HRD 描述包括 HRD 状态、HRD 评分、样本纯度信息。HRD 评分由多个指标构成的,应当额外列出每项指标的具体数值。

专家共识 10: 肿瘤基因诊断 NGS 分析报告应包括样本基本信息、检测位点覆盖范围、HGVS 命名的变异注释信息以及必要的质控信息。应明确检测方法的适用范围和局限性。对于 MSI、TMB 和 HRD 等特殊指标,需要标注具体数值,以确保信息的全面性和对临床应用的指导性。同时,报告还需合理解读变异内容与肿瘤发生、治疗及预后的相关性,以辅助临床医生诊断治疗。

3.7 分析控制系统选择

科学的分析控制系统可以协助生物信息工作者进行多个层级的工作管道生命周期设计、执行和维护。分析控制系统将 workflow 定义为处理单元链,每个处理单元将不同的生物信息学软件封装成执行脚本,接受前一处理单元数据流的输入并输出到下一处理单元。NGS 检测报告相关的工作可以分为基因组比对、SNV 和 Indel 变异识别、结构变异识别、拷贝数变异识别、变异注释、报告变异过滤质控、报告生成等多个处理单元。处理单元里的软件工具应以开源软件为主,尽量避免从头开发。

在完成分析控制系统开发后,应当制定严格的版本控制规则,开发者须确保其能顺利部署并能适应不同的计算配置环境,确保整个生物信息学分析流程正确、稳定、高效地运行。目前已有超过350种计算数据相关的工作流控制系统。随着测序数据量和分析复杂性的增加,分析流程的可扩展性和可重复性尤为重要,选择工作流控制系统应着重比较各候选系统的易用性、灵活性及处理复杂依赖关系和数据流的能力、社区维护建设和文档说明完整性、学习便利性等关键特征,结合本地化平台的计算资源和人员配置等综合考虑。

专家共识 11:肿瘤基因诊断NGS生物信息学分析本地化平台应优先采用基于工作流系统的控制系统,并根据本地化平台的计算资源和发展规划进行开发。在临床应用前,需充分验证和优化生物信息学分析的全流程,以确保分析平台的稳定性和高效运行。

4 分析流程的性能验证

肿瘤基因诊断NGS生物信息学本地化平台分析流程的性能验证总体遵循NGS基因检测试剂盒产品的各项要求,需要在开展常规检测前完成检测流程确认,明确质控标准,建立标准操作程序,并且在检测样品类型变化、关键质控阈值变更时重新进行性能验证^[3,9,56,62]。流程性能验证应以分析方法学为中心,覆盖临床检测的全部变异类型(SNV、Indel、CNV和Fusion)中具有代表性的变异位点和预期可检测到的变异频率,同时根据临床预期用途,有针对性地对常见复杂生物标志物(如TMB、MSI和HRD等)开展性能验证。

性能验证样本包括计算机模拟样本、公共数据、标准品和真实临床样本。考虑到医疗机构的实际操作、周期和成本等因素,可采用多种形式相结合的方式评估,总体上由易到难、逐层递进确认不同样本类型、不同变异类型的性能。(1)计算机模拟样本具有低成本、操作简单、数据生成快速且方便的特点。计算机模拟软件主要包括从头模拟、测序数据编辑两种方法,推荐使用VarBen模拟插入变异^[5]。(2)肿瘤NGS最常见的公共数据是WGS和WES,大部分来自真实临床样本,使用符合预期用途的公共数据,可以混合多样本或将采样到指定深度,结合计算机模拟软件插入变异,是低成本、快速验证生物信息学分析流程的

有效捷径。(3)标准品比计算机模拟样本、公共数据更接近真实样本的特性,常用于评估生物信息学分析流程的精密度、分析敏感性、分析特异性。标准品也存在局限性,部分变异类型难以人工制备,需要用真实临床样本补充验证。(4)真实临床样本最接近日常检测样本,样本类型包括新鲜组织、FFPE样本、血液cfDNA、胸腔积液和脑脊液等,选择与预期用途一致的样本类型,经其他方法学确认变异信息后,可用于生物信息学分析流程的性能验证。

性能验证样本的选择应严格区分训练集和验证集,流程构建使用的训练集样本不可作为性能验证样本使用,避免因过度学习导致性能验证结果偏乐观的情况^[26]。样本数量的选择和性能验证的评估,应适配检测试剂同步进行,具体可参考《实验室自建肿瘤全景变异检测性能确认中国专家共识(2024版)》^[1]。对于检测低频变异的NGS实验,交叉污染是一个公认的问题,建议在性能验证期间和后续的临床检测过程中,每次检测都包含阴性和空白对照样本,用于确认湿实验中是否存在来自邻近样本的污染。除了分析性能,还应对肿瘤检测和变异信息的一致性进行回顾性样本监测。确保分析软件运行速度和分析结果能满足临床检测预期。

专家共识 12:肿瘤基因诊断NGS生物信息学分析本地化平台的流程性能评估应综合使用模拟样本、公共WES和WGS数据、标准品以及真实的临床样本,评估分析流程的准确性、精密度、分析敏感度和特异性,并包含所有临床变异(panel涉及的重点难点)检测指标。

5 分析流程管理

为应对我国肿瘤NGS检测产品分析流程的多样化和复杂性,医疗机构在自建生物信息学分析能力时需参考以下策略:(1)在分析流程的软件周期管理方面,随着肿瘤分子特征研究的深入和新疗法的出现,NGS生物信息学分析流程需要及时更新。建议详细记录所有软件和数据库的版本与功能,建立质量控制方案,确保分析结果的准确性与稳定性。(2)在流程迁移管理方面,医疗机构可根据业务发展和信息安全需求进行迁移,包括从单机版向集群版的迁移,以适应更高的安全标准和性能要求。(3)为支持多样化的计

算需求并解决流程兼容性问题,需要进行流程框架维护升级,其中,软件和数据库变动需进行性能验证并保存变更日志。(4)为应对突发故障,医疗机构还应建立异地应急分析预案。异地托管服务器的环境配置应与本地保持一致,数据传输和结果回传需加密,确保数据安全。待本地系统恢复后,应对分析结果进行比对,以确保数据一致性。这些策略为医疗机构在生物信息学分析能力建设中提供了参考,以有效应对流程管理中的多重挑战。

专家共识 13:肿瘤基因诊断NGS生物信息学分析本地化平台应实行严格的版本控制,以保证运行的稳定性。软件和数据库的升级维护需进行性能验证,并保存变更日志。为应对突发故障,医疗机构还应建立异地应急分析预案。

6 人员配置及培训管理

建议配置生物信息学分析人员、数据及系统管理人员、报告及临床解读人员。生物信息学分析人员应熟练掌握Linux操作系统和常用生物信息学分析工具,以及具备扎实的计算机编程能力。通常这类人员具有生物信息学、计算生物学等交叉学科教育背景,建议有医学和生物学背景。数据及系统管理人员需掌握分析硬件设备的搭建和维护技能,通常这类人员来自软件、计算机等专业方向。报告及解读人员需充分理解NGS分析结果,同时掌握肿瘤学、临床医学、遗传学、分子生物学、分子病理学等专业知识。

不同分工和专业背景的人员均需接受临床医学实验室相关技能、职业素养、数据隐私和安全及伦理培训。本地化平台应对人员培训内容建立考核,经考核合格者方可上岗,每年应进行一次复核考核,并定期进行专业知识培训。生物信息学分析人员应掌握基因组学和分子生物学相关的基本概念,各类诊断技术和生物信息学的临床应用场景,具备在临床背景下解读分析结果的能力,并通过日常学习和培训提高自身数据分析、编程和脚本编写、数据可视化、临床统计分析等技能。数据管理人员应掌握各类存储架构基本原理和优缺点及相关网络安全知识,能科学制定本地化平台数据存储备份计划,并追踪最新相关技术进展,为平台扩展和迭代做好技能储备。另外,平台所有人员应当提高自身与临床及报告医生、实验

人员的沟通能力,在临床实践和交流学习中建立分析思维,提高解决复杂生物学问题和排除技术问题的能力。

专家共识 14:肿瘤基因诊断NGS生物信息学分析本地化平台的团队成员应具备相应的专业背景和分析技能并定期进行培训考核。

7 相关法律法规

根据《中华人民共和国人类遗传资源管理条例》^[63],人类遗传资源信息是指利用人类遗传资源材料产生的数据等信息资料,肿瘤NGS数据也包含在内。根据《中华人民共和国生物安全法》^[64]规定,采集、保藏、利用、对外提供我国人类遗传资源,应当符合伦理原则,不得危害公众健康、国家安全和社会公共利益。该法的第五十七条规定,将我国人类遗传资源信息向境外组织、个人及其设立或者实际控制的机构提供或者开放使用的,应当向国务院科学技术主管部门事先报告并提交信息备份。国家基因组科学数据中心肿瘤基因诊断数据分中心是符合国内法律法规和标准的数据归档库,当平台需要进行数据交换和共享时,推荐使用该分中心作为渠道。

专家共识 15:肿瘤基因诊断NGS生物信息学分析本地化平台需遵守生物安全及人类遗传资源管理相关法律法规,服从相应主管部门的监管、审批和伦理审查,确保人类遗传资源信息的安全。

8 小结

本共识针对肿瘤基因诊断中NGS生物信息学分析本地化平台的建设,系统阐述了硬件需求、数据存储与管理、分析流程搭建与性能验证、分析流程管理、人员配置与培训管理以及相关法律法规等方面内容,基于国内外现有文献证据进行总结,旨在为生物信息学分析本地化建设提供指导,并为临床检验人员提供参考。鉴于检验诊断、生物信息学及计算机科学的迅速发展,以及不同地区、医疗环境和机构规模的差异,建议临床检验人员在应用本共识时,结合本单位实际条件与具体临床情境,因地制宜,实施平台建设。

利益冲突说明:本共识由专家组内部成员针对性讨论得出,讨论过程中,所有参与者均不存在利益冲突。

编写专家组成员

编写专家组学术顾问

李红乐 河南省肿瘤医院
邢金良 空军军医大学

编写专家组组长

魏冰 河南省肿瘤医院
黄金艳 浙江大学医学院附属第一医院
马杰 河南省肿瘤医院

执笔者(按姓氏拼音排序)

刘永晶 浙江大学医学院附属第一医院
王庭杰 河南省肿瘤医院
王瑜亮 浙江大学医学院附属第一医院
王志中 河南省肿瘤医院

编写专家组成员(按姓氏拼音排序)

常志力 南京世和基因生物技术股份有限公司
董磊 上海交通大学医学院附属瑞金医院
方欢 北京吉因加科技有限公司
方向东 中国科学院北京基因组研究所
付莎 中山大学孙逸仙纪念医院
傅新晖 中山大学附属第六医院
郭萍 广州燃石医学检验所有限公司
韩博 北京大学人民医院
何俊俊 浙江大学医学院附属第一医院
胡欣 复旦大学附属肿瘤医院
胡晓彤 浙江大学医学院附属邵逸夫医院

纪元 复旦大学附属中山医院
姜国忠 郑州大学第一附属医院
姜艳芳 吉林大学第一医院
李文斌 中国医学科学院肿瘤医院
卢红阳 浙江省肿瘤医院
卢仁泉 复旦大学附属肿瘤医院
芦小单 吉林省人民医院
陆元志 暨南大学附属第一医院
孟宏学 哈尔滨医科大学附属肿瘤医院
欧阳能太 中山大学孙逸仙纪念医院
钱旭 浙江省肿瘤医院
申鹏 南方医科大学南方医院
施锦绣 上海市生物医药技术研究院
苏丹 浙江省肿瘤医院
王秋实 陆军特色医学中心
夏艳 无锡臻和生物科技有限公司
许世朋 北京诺禾致源科技股份有限公司
严令华 常州桐树生物科技有限公司
叶丰 四川大学华西医院
叶庆 中国科学技术大学附属第一医院
尹文娟 浙江省肿瘤医院
于津浦 天津医科大学肿瘤医院
岳君秋 湖北省肿瘤医院
曾韬 浙江大学医学院附属第一医院
张岩 哈尔滨工业大学
赵征 陕西省肿瘤医院
周晓燕 复旦大学附属肿瘤医院
周永春 云南省肿瘤医院

参考文献

- [1] 中国抗癌协会肿瘤病理专业委员会分子病理协作组,中华医学会病理学分会分子病理学组.实验室自建肿瘤全景变异检测性能确认中国专家共识(2024版)[J].中华肿瘤杂志,2024,46(4):274-284.
- [2] 江苏省医学会病理学分会,江苏省医学会检验学分会,江苏省临床检验中心.肿瘤二代测序生物信息学分析规范化管理江苏专家共识[J].临床检验杂志,2023,41(8):561-568,596.
- [3] 《临床分子病理实验室二代基因测序检测专家共识》编写组.临床分子病理实验室二代基因测序检测专家共识[J].中华病理学杂志,2017,46(3):145-148.
- [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]. Genet Med, 2015, 17(5):405-424.
- [5] HORAK P, GRIFFITH M, DANOS A M, 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) [J]. Genet Med, 2022, 24(5):986-998.
- [6] 中国抗癌协会肿瘤标志物专业委员会,上海市抗癌协会肿瘤标志物专业委员会.基于中国人群的BRCA胚系突变筛查专家共识(2024年版)[J].中国癌症杂志,2024,34(2):220-238.

- [7] 二代测序临床报告解读肿瘤学专家组. 肿瘤二代测序临床报告解读共识[J]. 循证医学, 2022, 22(2): 65-79.
- [8] 中国抗癌协会骨肿瘤与骨转移癌专业委员会药物与精准治疗学组. 骨肿瘤二代测序及分子检测技术临床应用专家共识[J]. 中华骨与关节外科杂志, 2022, 15(12): 889-901.
- [9] 王平, 苟阳, 彭贤贵, 等. 专家共识解读: 二代测序如何辅助血液肿瘤的临床诊治[J]. 海南医学, 2019, 30(21): 2721-2724.
- [10] 中国临床肿瘤学会肿瘤标志物专业委员会, 中国肿瘤驱动基因分析联盟. 二代测序技术在肿瘤精准医学诊断中的应用专家共识[J]. 中华医学杂志, 2018, 98(26): 2057-2065.
- [11] 中国抗癌协会血液肿瘤专业委员会, 中华医学会血液学分会, 中华医学会病理学分会. 二代测序技术在血液肿瘤中的应用中国专家共识(2018年版)[J]. 中华血液学杂志, 2018, 39(11): 881-886.
- [12] 赵辰, 谢小雷, 冀维真, 等. 美国二代测序技术临床应用的共识声明、实践资源、技术标准和指南的概述[J]. 中华医学遗传学杂志, 2021, 38(6): 513-520.
- [13] 孙隽, 黄颐, 王小冬, 等. 遗传病二代测序临床检测全流程规范化共识探讨(3)——数据分析流程[J]. 中华医学遗传学杂志, 2020, 37(3): 345-351.
- [14] 中国抗癌协会肿瘤标志专业委员会遗传性肿瘤标志物协作组, 中国抗癌协会肿瘤病理专业委员会分子病理协作组. 肿瘤突变负荷检测及临床应用中国专家共识(2020年版)[J]. 中国癌症防治杂志, 2020, 12(5): 485-494.
- [15] CHEN S, ZHOU Y, CHEN Y, et al. fastp: an ultra-fast all-in-one FASTQ preprocessor[J]. *Bioinformatics*, 2018, 34(17): i884-i890.
- [16] BOLGER A M, LOHSE M, USADEL B. Trimmomatic: a flexible trimmer for Illumina sequence data[J]. *Bioinformatics*, 2014, 30(15): 2114-2120.
- [17] KECHIN A, BOYARSKIKH U, KEL A, et al. cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing[J]. *J Comput Biol*, 2017, 24(11): 1138-1143.
- [18] LI H, DURBIN R. Fast and accurate long-read alignment with Burrows-Wheeler transform[J]. *Bioinformatics*, 2010, 26(5): 589-595.
- [19] DANECHEK P, BONFIELD J K, LIDDLE J, et al. Twelve years of SAMtools and BCFtools[J]. *Gigascience*, 2021, 10(2): giab008.
- [20] LI H, HANDSAKER B, WYSOKER A, et al. The Sequence Alignment/Map format and SAMtools[J]. *Bioinformatics*, 2009, 25(16): 2078-2079.
- [21] TARASOV A, VILELLA A J, CUPPEN E, et al. Sambamba: fast processing of NGS alignment formats [J]. *Bioinformatics*, 2015, 31(12): 2032-2034.
- [22] MCKENNA A, HANNA M, BANKS E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data[J]. *Genome Res*, 2010, 20(9): 1297-1303.
- [23] DEPRISTO M A, BANKS E, POPLIN R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data[J]. *Nat Genet*, 2011, 43(5): 491-498.
- [24] GARRISON E, MARTH G. Haplotype-based variant detection from short-read sequencing [J/OL]. *arXiv: Genomics*, 2012. [2024-9-12]. <https://www.biorxiv.org/content/10.1101/120295v1>.
- [25] KENDIG K I, BAHETI S, BOCKOL M A, et al. Sentieon DNaseq Variant Calling Workflow Demonstrates Strong Computational Performance and Accuracy[J]. *Front Genet*, 2019, 10:736.
- [26] POPLIN R, CHANG P C, ALEXANDER D, et al. A universal SNP and small-indel variant caller using deep neural networks[J]. *Nat Biotechnol*, 2018, 36(10): 983-987.
- [27] KOBOLDT D C, ZHANG Q, LARSON D E, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing[J]. *Genome Res*, 2012, 22(3): 568-576.
- [28] LAI Z, MARKOVETS A, AHDESMANI M, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research[J]. *Nucleic Acids Res*, 2016, 44(11): e108.
- [29] FREED D, PAN R, ALDANA R. TNscope: Accurate Detection of Somatic Mutations with Haplotype-based Variant Candidate Detection and Machine Learning Filtering[J/OL]. *bioRxiv*, 2018. [2024-9-12]. <https://www.biorxiv.org/content/10.1101/250647v1>.
- [30] SAHRAEIAN S M E, LIU R, LAU B, et al. Deep convolutional neural networks for accurate somatic mutation detection[J]. *Nat Commun*, 2019, 10(1): 1041.
- [31] SAHRAEIAN S M E, FANG L T, KARAGIANNIS K, et al. Achieving robust somatic mutation detection with deep learning models derived from reference data sets of a cancer sample [J]. *Genome Biol*, 2022, 23(1): 12.
- [32] KIM S, SCHEFFLER K, HALPERN A L, et al. Strelka2: fast and accurate calling of germline and somatic variants[J]. *Nat Methods*, 2018, 15(8): 591-594.
- [33] TALEVICH E, SHAIN A H, BOTTON T, et al. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing[J]. *PLoS Comput Biol*, 2016, 12(4): e1004873.
- [34] FAVERO F, JOSHI T, MARQUARD A M, et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data[J]. *Ann Oncol*, 2015, 26(1): 64-70.
- [35] HA G, ROTH A, KHATTRA J, et al. TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data[J]. *Genome Res*, 2014, 24(11): 1881-1893.
- [36] FROMER M, MORAN J L, CHAMBERT K, et al. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth[J]. *Am J Hum Genet*, 2012, 91(4): 597-607.
- [37] ABYZOV A, URBAN A E, SNYDER M, et al. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing [J]. *Genome Res*, 2011, 21(6): 974-984.
- [38] FOWLER A. DECoN: A Detection and Visualization Tool for Exonic Copy Number Variants[J]. *Methods Mol Biol*, 2022, 2493: 77-88.
- [39] JOHANSSON L F, VAN DIJK F, DE BOER E N, et al. CoNVaDING: Single Exon Variation Detection in Targeted NGS Data [J]. *Hum Mutat*, 2016, 37(5): 457-464.
- [40] POVYSIL G, TZIKA A, VOGT J, et al. panelcn.MOPS: Copy-number detection in targeted NGS panel data for clinical diagnostics [J]. *Hum Mutat*, 2017, 38(7): 457-464.
- [41] PLAGNOL V, CURTIS J, EPSTEIN M, et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling[J]. *Bioinformatics*, 2012, 28(21): 2747-2754.

- [42] JIANG Y, WANG R, URRUTIA E, et al. CODEX2: full-spectrum copy number variation detection by high-throughput DNA sequencing[J]. *Genome Biol*, 2018, 19(1): 202.
- [43] NIU B, YE K, ZHANG Q, et al. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data[J]. *Bioinformatics*, 2014, 30(7): 1015-1016.
- [44] KAUTTO E A, BONNEVILLE R, MIYA J, et al. Performance evaluation for rapid detection of pan-cancer microsatellite instability with MANTIS[J]. *Oncotarget*, 2017, 8(5): 7452-7463.
- [45] SALIPANTE S J, SCROGGINS S M, HAMPEL H L, et al. Microsatellite instability detection by next generation sequencing [J]. *Clin Chem*, 2014, 60(9): 1192-1199.
- [46] ZHU L, HUANG Y, FANG X, et al. A Novel and Reliable Method to Detect Microsatellite Instability in Colorectal Cancer by Next-Generation Sequencing[J]. *J Mol Diagn*, 2018, 20(2): 225-231.
- [47] SZTUPINSZKI Z, DIOSY M, KRZYSTANEK M, et al. Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer [J]. *NPJ Breast Cancer*, 2018, 4: 16.
- [48] LAYER R M, CHIANG C, QUINLAN A R, et al. LUMPY: a probabilistic framework for structural variant discovery [J]. *Genome Biol*, 2014, 15(6): R84.
- [49] RAUSCH T, ZICHNER T, SCHLATTL A, et al. DELLY: structural variant discovery by integrated paired-end and split-read analysis [J]. *Bioinformatics*, 2012, 28(18): i333-i339.
- [50] NEWMAN A M, BRATMAN S V, STEHR H, et al. FACTERA: a practical method for the discovery of genomic rearrangements at breakpoint resolution [J]. *Bioinformatics*, 2014, 30(23): 3390-3393.
- [51] WANG J, MULLIGHAN C G, EASTON J, et al. CREST maps somatic structural variation in cancer genomes with base-pair resolution [J]. *Nat Methods*, 2011, 8(8): 652-654.
- [52] HAAS B J, DOBIN A, STRANSKY N, et al. STAR-Fusion: Fast and Accurate Fusion Transcript Detection from RNA-Seq [J/OL]. *bioRxiv*, 2017. [2024-9-12]. <https://www.biorxiv.org/content/10.1101/120295v1>.
- [53] UHRIG S, ELLERMANN J, WALTHER T, et al. Accurate and efficient detection of gene fusions from RNA sequencing data [J]. *Genome Res*, 2021, 31(3): 448-460.
- [54] TIAN L, LI Y, EDMONSON M N, et al. CICERO: a versatile method for detecting complex and diverse driver fusions using cancer RNA sequencing data [J]. *Genome Biol*, 2020, 21(1): 126.
- [55] NICORICI D, ŞATALAN M, EDGREN H, et al. FusionCatcher - a tool for finding somatic fusion genes in paired-end RNA-sequencing data [J/OL]. *bioRxiv*, 2014. [2024-9-12]. <https://www.biorxiv.org/content/10.1101/011650v1/>.
- [56] 肖林林, 胡婷婷, 魏取好, 等. 二代测序临床应用的质量控制 [J]. *临床检验杂志*, 2019, 37(10): 745-748.
- [57] VAN DER AUWERA G A, CARNEIRO M O, HARTL C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline [J]. *Curr Protoc Bioinformatics*, 2013, 43(1110): 11.
- [58] LI H, DURBIN R. Fast and accurate short read alignment with Burrows-Wheeler transform [J]. *Bioinformatics*, 2009, 25(14): 1754-1760.
- [59] PEI S, LIU T, REN X, et al. Benchmarking variant callers in next-generation and third-generation sequencing analysis [J]. *Brief Bioinform*, 2021, 22(3): bbaa148.
- [60] MORALES J, PUJAR S, LOVELAND J E, et al. A joint NCBI and EMBL-EBI transcript set for clinical genomics and research [J]. *Nature*, 2022, 604(7905): 310-315.
- [61] 中国抗癌协会肿瘤精准治疗专业委员会. 二代测序技术在消化系统肿瘤临床应用的中国专家共识 [J]. *肿瘤综合治疗电子杂志*, 2024, 10(1): 69-92.
- [62] 中国抗癌协会泌尿男生殖系肿瘤专业委员会, 中国临床肿瘤学会前列腺癌专家委员会. 中国前列腺癌患者基因检测专家共识 (2020年版) [J]. *中国癌症杂志*, 2020年30(7): 551-560.
- [63] 中华人民共和国中央人民政府. 中华人民共和国人类遗传资源管理条例 [EB/OL]. (2019-6-10) [2024-9-12]. https://www.gov.cn/zhengce/content/2019-06/10/content_5398829.htm.
- [64] 中华人民共和国中央人民政府. 中华人民共和国生物安全法 [EB/OL]. (2020-10-17) [2024-9-12]. https://www.gov.cn/xinwen/2020-10/18/content_5552108.htm.

[收稿 2024-10-09][编辑 罗惠予/游雪梅]