

BIOF3 组学数据分析

01 数据类型: WGS vs WES vs Panel

导出日期: 2026年5月12日

01 数据类型: WGS vs WES vs Panel

基因组测序有三种主要策略, 选择取决于研究目的和预算:

策略	覆盖范围	数据量/样本	适用场景
WGS	全基因组 ~3Gb	30-60x, ~90GB FASTQ	结构变异、非编码区、群体遗传
WES	外显子区 ~60Mb	100-200x, ~6GB FASTQ	肿瘤体细胞突变、遗传病诊断
Panel	几十~几百个基因	500-1000x, ~1GB	临床检测、已知热点突变

WGS vs WES 的分析差异

维度	WGS	WES
变异类型	SNV + Indel + SV + CNV	主要 SNV + Indel
参考区域	全基因组	需要 BED 文件定义 target 区域
覆盖度均匀性	好	受捕获效率影响, 边缘区域覆盖低
数据分析工具	GATK / DeepVariant	GATK + Mutect2 (肿瘤)
下游重点	群体遗传、GWAS、SV	肿瘤驱动突变、maftools、OncoKB

肿瘤 WES 的特殊性

肿瘤样本通常配对测序 (tumor + matched normal), 用 Mutect2 做体细胞突变检测。输出的 MAF 文件是 maftools 的标准输入。

```
# Mutect2 典型调用
gatk Mutect2 \
  -R reference.fa \
  -I tumor.bam \
  -I normal.bam \
  -normal normal_sample_name \
  -O somatic.vcf.gz

# VCF -> MAF 转换
vcf2maf.pl --input-vcf somatic.vcf.gz --output-maf somatic.maf \
  --tumor-id tumor --normal-id normal --ref-fasta reference.fa
```

下一步

- [02 比对与变异检测流程](#)
- [03 VCF 注释与可视化](#)
- [04 maftools 肿瘤突变分析](#)

参考资源

- [GATK Best Practices](#)
- [Mutect2 文档](#)
- [vcf2maf](#)



扫码关注微信公众号【生信F3】

获取文章完整内容，分享生物信息学最新知识。