

BIOF3 组学数据分析

03 VCF 注释与可视化

导出日期：2026年5月12日

03 VCF 注释与可视化

VCF 文件拿到手之后，第一步是“每个变异落在什么基因、什么区域、有什么功能影响”。本章用 R 的 `VariantAnnotation` 包在内置 chr22 VCF 上演示。

真实示例

配套脚本 [genome03_variant_anno_sci.R](#) 输出 6 张图：

```
Rscript scripts/genomics/genome03_variant_anno_sci.R
```

每张图看什么

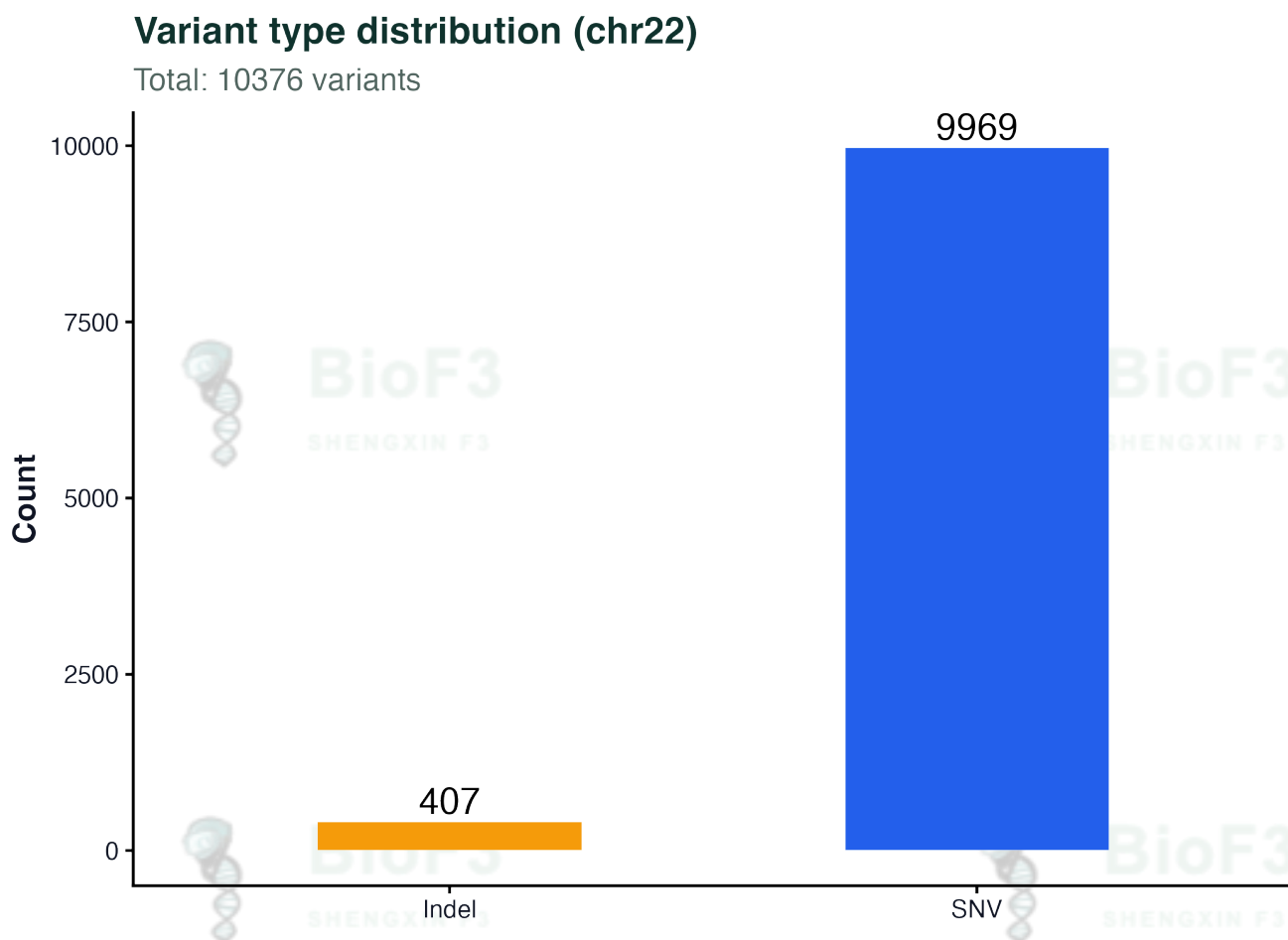


图 1：SNV vs Indel 的数量分布。WGS/WES 里 SNV 通常占 90%+。

Variant distribution by genomic region

Most variants fall in introns and intergenic regions

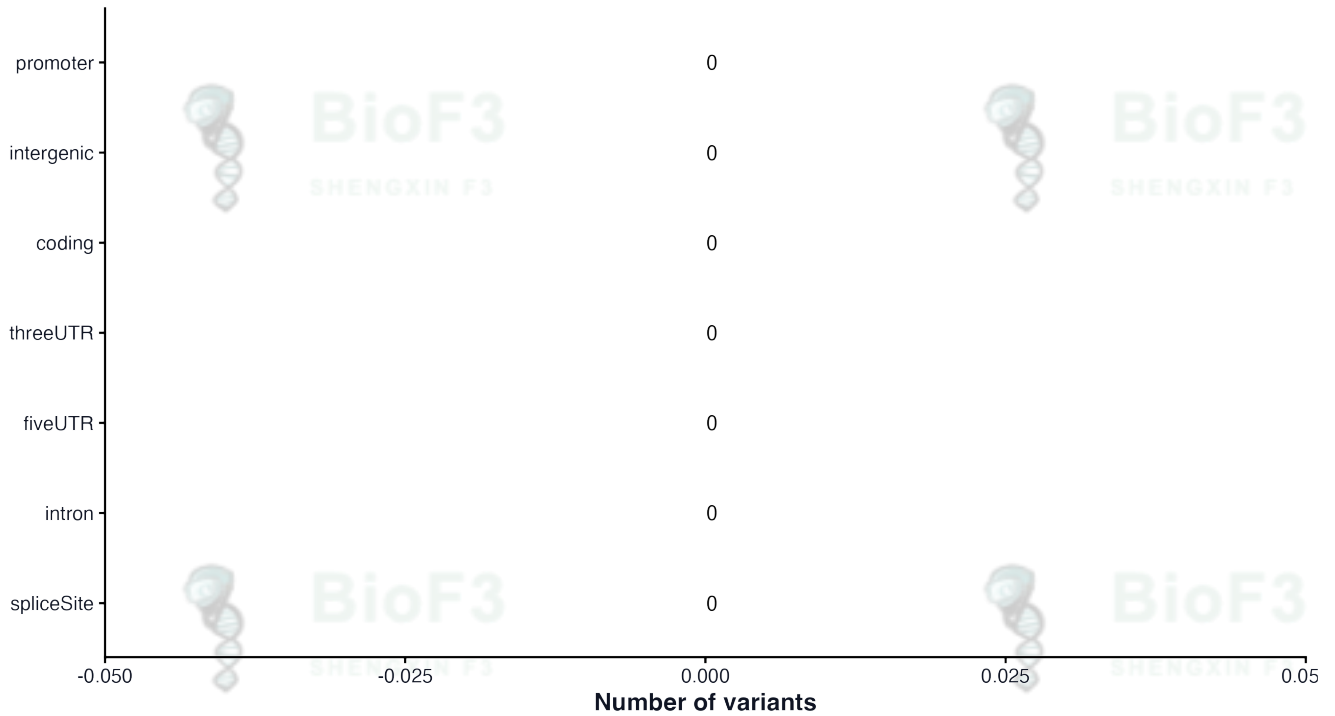


图 2：变异落在哪些基因组区域。大部分在内含子和基因间区（非编码区占基因组 98%+）。

Transition / Transversion (Ti/Tv = 3.05)

Expected Ti/Tv ~ 2.0-2.1 for WGS; lower in coding regions

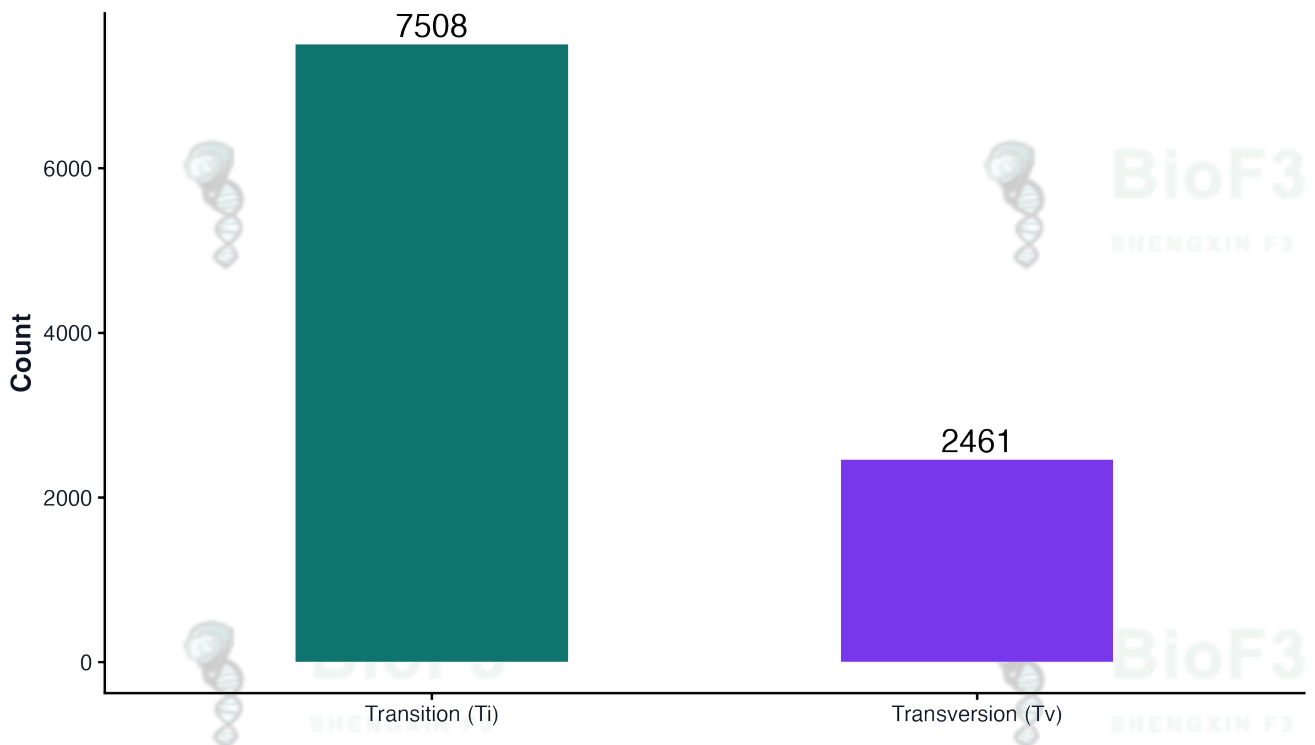


图 3：转换/颠换比。WGS 期望 ~2.0-2.1，WES 编码区 ~3.0。偏低可能说明假阳性多。

Site frequency spectrum (chr22)

Classic L-shaped distribution: most variants are rare

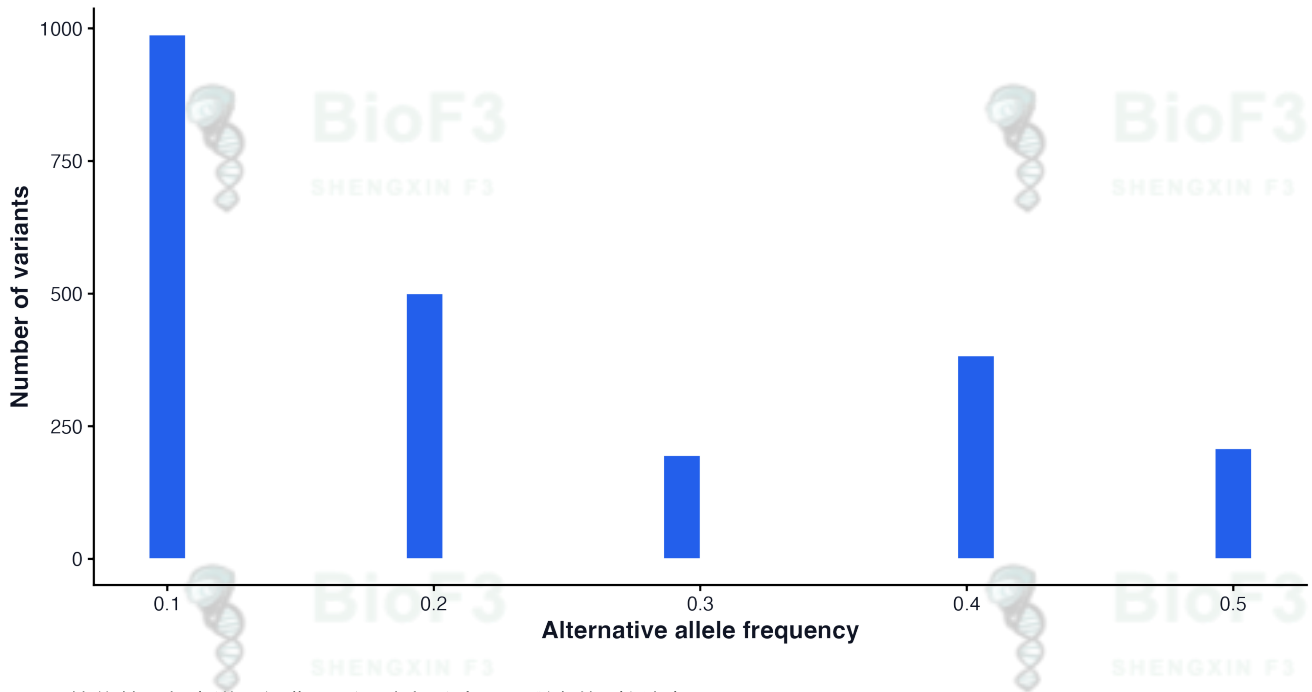


图 4：等位基因频率谱。经典 L 形：大部分变异是稀有的（低频）。

Variant density along chr22

Peaks may correspond to gene-dense regions or segmental duplications

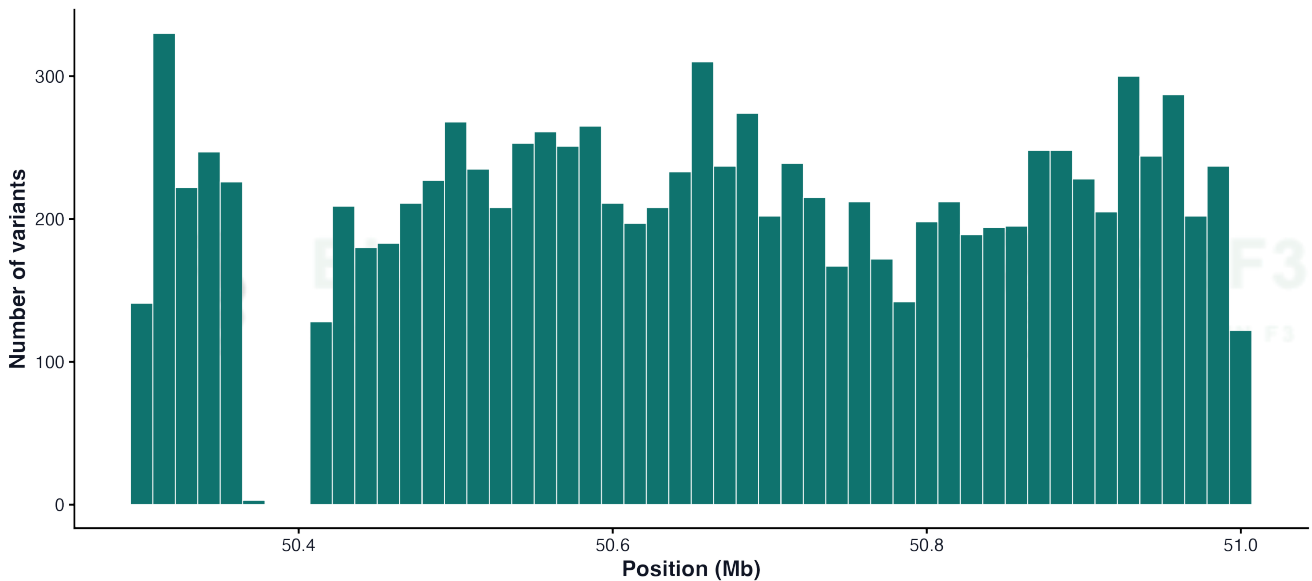


图 5：chr22 上的变异密度分布。某些区域密集可能对应基因密集区或重复序列。

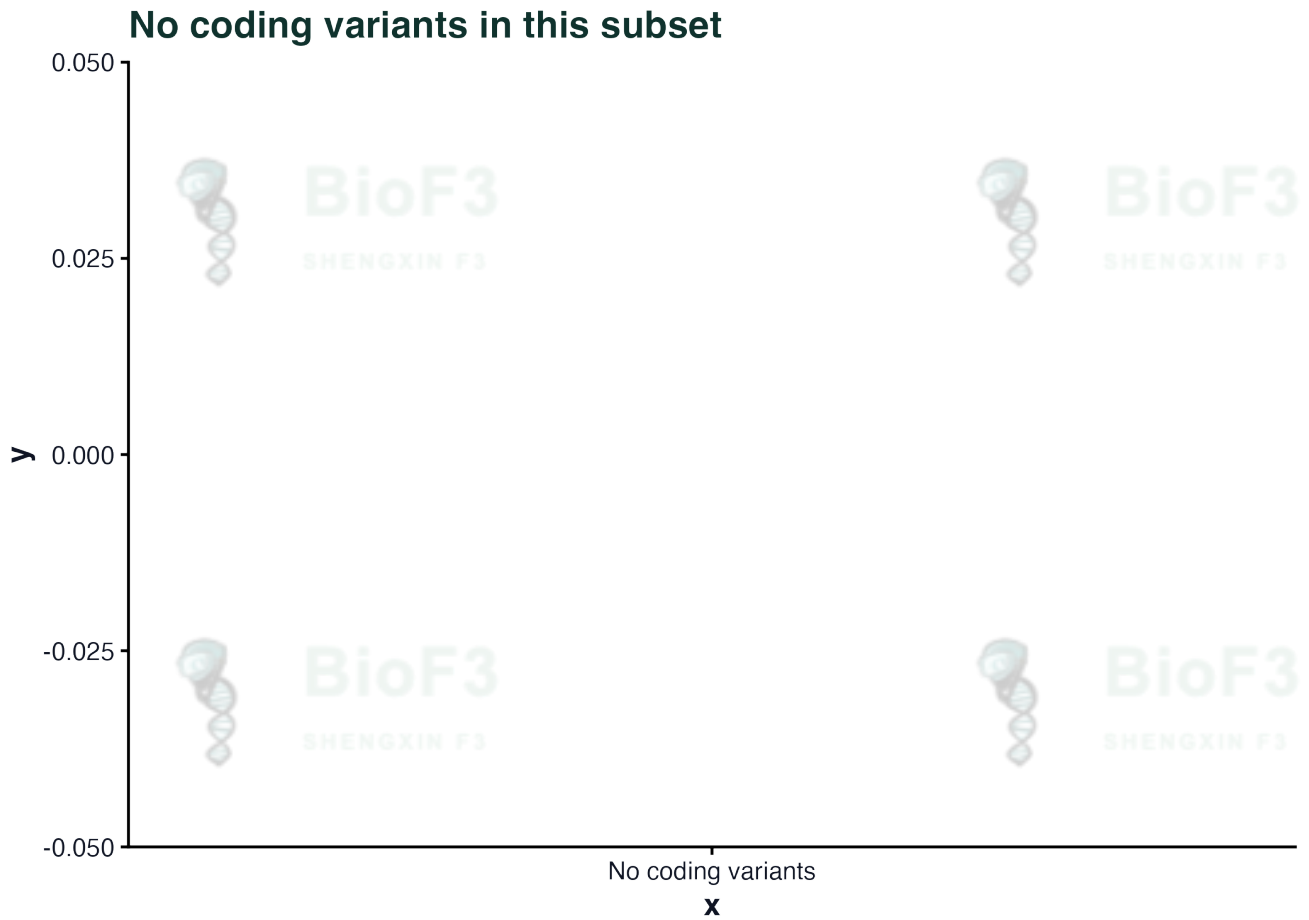


图 6：编码区变异的功能后果（同义/错义/无义）。

下载资源

genome03_variant_anno_sci.R
8 KB

下载 VCF 注释可视化完整脚本 ↗

参考资源

- [VariantAnnotation](#)
- [Ensembl VEP](#)
- [ANNOVAR](#)



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。

