

BIOF3 组学数据分析

06 群体遗传: PCA / 祖源 / LD

导出日期: 2026年5月12日

06 群体遗传: PCA / 祖源 / LD

群体遗传分析用大量个体的基因型数据回答"这些人从哪来、彼此什么关系、哪些位点受到选择"。

常用工具

工具	用途
PLINK 2	数据管理、QC、PCA、关联分析
ADMIXTURE	祖源成分估计 (K 群体)
vcftools	VCF 统计 (Fst、pi、Tajima's D)
EIGENSOFT	PCA + 群体结构

PCA 分析

```
# VCF -> PLINK 格式
plink2 --vcf cohort.vcf.gz --make-bed --out cohort

# LD pruning (去掉高 LD 的 SNP, 避免 PCA 被局部 LD 主导)
plink2 --bfile cohort --indep-pairwise 50 5 0.2 --out pruned
plink2 --bfile cohort --extract pruned.prune.in --make-bed --out cohort_pruned

# PCA
plink2 --bfile cohort_pruned --pca 10 --out cohort_pca
```

输出 cohort_pca.eigenvec 就是每个样本的 PC1~PC10 坐标, 用 R 画散点图。

ADMIXTURE 祖源分析

```
# 跑 K=2 到 K=6
for K in 2 3 4 5 6; do
  admixture --cv cohort_pruned.bed $K | tee log_K${K}.out
done

# 选最优 K: 看 CV error 最低的那个
grep "CV error" log_K*.out
```

LD 衰减

```
# 计算 LD ( $r^2$ ) 随距离的衰减  
plink2 --bfile cohort --ld-window-r2 0 --ld-window 1000 --ld-window-kb 500 \  
--out ld_decay
```

LD 衰减速度反映群体的有效群体大小和重组率。

参考资源

- [PLINK 2 文档](#)
- [ADMIXTURE](#)
- [1000 Genomes 群体结构教程](#)



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。

