

常见问题解析

1. 怎么解决输入标识符导致的匹配错误？

输入基因名或蛋白ID后，STRING提示“Multiple matches found”或匹配到非目标蛋白（如小鼠基因名匹配到人源同源基因），导致后续网络包含错误节点。解决办法如下：①在STRING映射页面（MAPPING界面）手动选择正确条目，优先选择“Identity>80%”且功能注释匹配的选项；②全部使用UniProt ID（如P04637）替代基因名，避免物种歧义；③对跨物种分析，先用BLAST筛选E-value<1e-10的序列再输入STRING。

2. 怎么解决置信度阈值设置不当导致假阳性/假阴性？

在互作分析过程中很容易出现网络过密（假阳性多）或过稀疏（遗漏真实互作），尤其对弱互作蛋白（如转录因子）影响显著。解决办法如下：①初步探索设阈值 ≥ 0.4 以观察网络密度，逐步提高至0.7来保留高置信互作以筛选核心互作；②在“Active Interaction Sources”中勾选特定来源，排除预测性证据；③导出分值在0.4-0.7的互作蛋白，用BioGRID或IntAct补充实验证据。

3. 当输入的蛋白数量过大时（大于50个）怎么办？

①设置“max interactors = 0”（仅显示输入蛋白间互作）或“=10”（每个蛋白最多10个直接互作）；②关闭“Second Shell”选项；③利用STRING内置“Cluster”功能按k-means划分模块；④对多结构域蛋白（如激酶），分别输入各结构域序列生成子网络。

参考文献

- [1] Göös H, Kinnunen M, Salokas K, et al. Human transcription factor protein interaction networks [J]. Nature communications, 2022, 13(1): 766.
- [2] Szklarczyk D, Kirsch R, Koutrouli M, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest [J]. Nucleic acids research, 2023, 51(D1): D638-d646.