



我国学者在合成生物学基因调控序列人工智能设计领域取得进展

日期 2023-11-24 来源: 信息科学部 作者: 赵瑞珍 刘屿 刘行健 【大中小】 【打印】 【关闭】

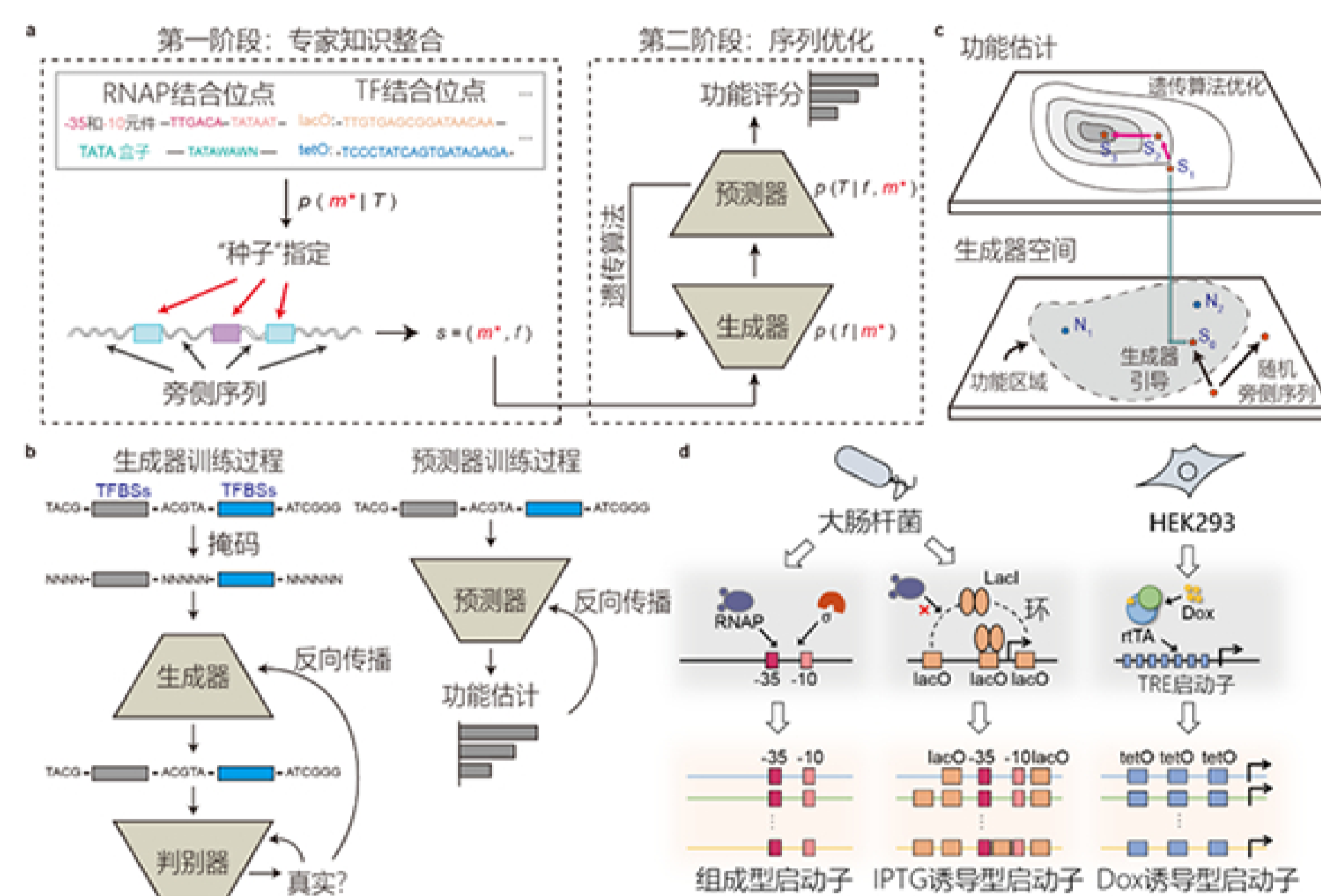


图 知识引导与数据驱动相融合的启动子辅助设计方法

在国家自然科学基金项目(批准号: 62250007、62225307、61721003)等资助下,清华大学自动化系汪小我教授团队提出了一种将专家知识与大数据学习相融合的合成启动子人工智能辅助设计方法DeepSEED。研究成果以“使用DeepSEED进行侧翼序列深度改造实现高效启动子设计(Deep flanking sequence engineering for efficient promoter design using DeepSEED)”为题,于2023年10月9日发表于《自然·通讯》(Nature Communications)期刊上。论文链接: <https://www.nature.com/articles/s41467-023-41899-y>。

人类专家擅长从小样本中发现强特征,而深度学习模型可以从大量数据中提取隐含的弱模式。基因启动子是决定基因在什么条件下以何种程度进行转录表达的DNA调控序列。设计具有特定功能的合成启动子是构建人工基因系统的基础,在代谢工程、基因治疗等领域有广泛应用。过去,对启动子序列的设计改造主要集中在修改具有显著模式特征的转录因子结合位点(TFBS)序列上,而忽视了连接不同TFBS之间的侧翼序列。尽管有证据表明侧翼序列对启动子活性具有一定影响,但其特征信号弱且规律复杂,难以被归纳总结为明确的设计和准则。

团队针对启动子侧翼序列维度高、模式复杂、在设计中被长期忽视的问题,提出了一种人工智能辅助的侧翼序列优化方法DeepSEED。如图所示,该方法基于知识引导与数据驱动融合的智能设计策略,一方面利用专家知识定义与启动子功能相关的强特征模式作为“种子”,另一方面利用深度学习从大量数据中学习侧翼序列满足的弱特征复杂约束,基于条件式生成模型和进化计算来优化侧翼序列与“种子”的匹配关系,以实现启动子序列的整体优化提高其转录调控活性。该方法成功应用于细菌和哺乳动物细胞内多类启动子的优化设计,细胞实验验证了其有效性,获得了一批序列多样性强、功能优于天然序列的全新启动子。

该研究提出的知识引导与数据驱动相融合的合成启动子智能设计策略,为解决功能启动子设计中维度高、样本小的核心难题提供了新思路,为合成生物学研究提供可用于侧翼序列优化和多类型启动子设计的人工智能辅助工具。

机构概况: 概况 职能 领导介绍 机构设置 规章体系 专家咨询 评审程序 资助格局 监督工作

政策法规: 国家科学技术相关法律 国家自然科学基金条例 国家自然科学基金规章制度 国家自然科学基金发展规划

项目指南: 项目指南

申请资助: 申请受理 项目检索与查询 下载中心 代码查询 常见问题解答 科学基金资助体系

共享传播: 年度报告 中国科学基金 大数据知识管理服务 优秀成果选编

国际合作: 通知公告 管理办法 协议介绍 进程简表

信息公开: 信息公开制度 信息公开管理办法 信息公开指南 信息公开工作年度报告 信息公开目录 依申请公开

相关链接 政府 新闻 科普