

上海市科学技术委员会

沪科指南〔2023〕14号

关于发布上海市2023年度“科技创新行动计划” 计算生物学重点专项申报指南的通知

各有关单位：

为深入实施创新驱动发展战略，加快建设具有全球影响力的科技创新中心，推动形成具有上海特色的计算生物学研发体系，根据《上海市建设具有全球影响力的科技创新中心“十四五”规划》《上海市计算生物学创新发展行动计划（2023-2025年）》，上海市科学技术委员会特发布2023年度计算生物学重点专项申报指南。

一、 征集范围

专题一、生物大数据共性算法

方向1：单细胞多组学整合计算方法

研究目标：面向配对和非配对单细胞多模态组学数据，发

展2种以上组学数据对齐及下游分析算法，性能相比国际主流算法提升10%以上。

研究内容：建立面向配对及非配对场景下的单细胞多组学整合AI计算框架，开发多模态整合场景下的单细胞多组学下游应用AI算法，实现单细胞去噪、聚类分析、多数据集整合、去批次效应、细胞类型识别、轨迹分析、细胞通讯，以及跨模态生成分析等。

经费额度：非定额资助，拟支持不超过1个项目，每项资助额度不超过200万元。

执行期限：2023年12月01日至2026年11月30日。

方向2：单细胞谱系追踪计算方法

研究目标：发展多模态数据整合的单细胞空间转录组谱系追踪算法，准确度比国际主流算法提升5%以上。

研究内容：针对静态条形码和基于CRISPR的动态条形码标记，建立单细胞转录组整合标记条形码的谱系追踪计算方法，实现单细胞转录组分析、条形码序列去噪、条形码进化树分析、细胞谱系进化树、细胞谱系发育图谱、谱系条形码与空间坐标多模态整合的空间转录组单细胞追踪算法等。

经费额度：非定额资助，拟支持不超过1个项目，每项资助额度不超过200万元。

执行期限：2023年12月01日至2026年11月30日。

方向3：样本特异网络标志物方法

研究目标：开发国际领先的样本特异网络标志物算法，复

杂生物学过程预测准确性和敏感性比国际最高水平提高至少5%，复杂疾病早诊准确率比国际最高水平提高至少10%。

研究内容：开发样本特异性的网络标志物算法，从而实现基于网络对于发育、疾病等复杂生物学过程的准确和鲁棒的预测。针对复杂疾病，开发样本特异性的动态网络标志物，从而实现基于动态网络对于复杂疾病的早期预警信号准确标识与预测。

经费额度：非定额资助，拟支持不超过1个项目，每项资助额度不超过200万元。

执行期限：2023年12月01日至2026年11月30日。

方向4：转录调控网络基础模型

研究目标：建立国际领先的解码转录调控网络的预训练大模型，完成2个以上模型泛化任务，性能相比国际主流算法提升10%以上。

研究内容：整合多模态高通量表观遗传、转录调控和转录组数据，基于深度神经网络构建解码转录调控网络的预训练大模型。根据不同应用场景泛化预训练大模型，解决生物学核心问题，包括但不限于预测细胞命运走向、鉴定细胞命运决定和疾病发生发展过程中的关键基因和调控因子。

经费额度：非定额资助，拟支持不超过1个项目，每项资助额度不超过200万元。

执行期限：2023年12月01日至2026年11月30日。

方向5：RNA转录后加工调控网络解析方法

研究目标: 建立模块化可拓展的RNA转录后加工智能分析平台, 发展RNA转录后加工调控网络解析和功能预测模型, 性能比国际主流算法提升10%以上。

研究内容: 建立RNA转录后调控解析方法, 系统解构RNA转录后加工多样性及其调控特征, 包括但不限于RNA编辑、修饰、剪接、加尾和二级/三级结构的时空变化及相互协作, 构建RNA转录后加工调控网络和生物学功能预测模型, 解析其在全局生命代谢进程中的功能和机制。

经费额度: 非定额资助, 拟支持不超过1个项目, 每项资助额度不超过200万元。

执行期限: 2023年12月01日至2026年11月30日。

方向6: 细胞器分辨率的细胞互作网络模型

研究目标: 发展高分辨率的细胞互作网络模型, 细胞与环境、细胞间及细胞器间互作的预测性能比国际主流算法提升10%以上。

研究内容: 通过高维度、多模态、亚细胞分辨率的空间多组学数据整合, 构建细胞与环境、细胞间及细胞器间互作动态网络; 开发预测细胞与环境、细胞间及细胞器间互作的人工智能新模型, 刻画互作的普适规律及其生物学作用, 解析互作与人类疾病的内在关联。

经费额度: 非定额资助, 拟支持不超过1个项目, 每项资助额度不超过200万元。

执行期限: 2023年12月01日至2026年11月30日。

方向7: 病原微生物时空变异演化网络模型

研究目标: 发展病原微生物基因变异大模型, 解析千万量级病毒、十万量级细菌全基因组演化网络, 揭示变异毒性及传播风险, 性能比国际主流算法提升10%以上。

研究内容: 设计病原微生物全基因组尺度的序列变异快速比对方法, 研发微生物全基因组的变异特征自动提取方法, 发展位点变异演化网络快速构建算法, 整合时空信息识别关键演化路径; 构建致病相关风险基因及位点大模型, 解析病原微生物变异位点的耐药风险及传播特征。

经费额度: 非定额资助, 拟支持不超过1个项目, 每项资助额度不超过200万元。

执行期限: 2023年12月01日至2026年11月30日。

专题二、蛋白质与核酸设计改造

方向1: 蛋白质与核酸静态结构预测

研究目标: 发展独立自主蛋白质与核酸结构预测及设计软件, 总体预测精度比国际同期最高水平提升5%以上, 蛋白质复合体结构预测精度比国际同期最高水平提升5%以上。

研究内容: 发展并完善独立自主的大分子结构预测及设计软件, 尤其是针对蛋白质侧链建模及点突变建模。面向蛋白质静态结构预测, 开发独立自主的算法软件; 开展蛋白质侧链建模研究, 开发先进的蛋白质侧链建模方法, 应用扩展到蛋白质相互作用、结构设计及进化研究上; 发展蛋白质复合体, 尤其是蛋白质与核酸复合体的结构预测算法, 填补国内与国际空

白；结合高通量核酸结构检测方法，开发核酸静态结构及其与不同生物大分子和小分子药物等在体内和体外不同条件下的结构预测模型和算法。

经费额度：非定额资助，拟支持不超过1个项目，每项资助额度不超过500万元。

执行期限：2023年12月01日至2026年11月30日。

方向2：蛋白质工程通用AI设计平台

研究目标：构建包含特色极限环境条件标注的蛋白数据库，建立蛋白质工程通用AI设计平台，指导蛋白质的改造和优化。

研究内容：建立具备亿级数据量且面向极端环境标注蛋白序列的核心壁垒蛋白数据库。通过学习蛋白的序列和结构信息，构建基于预训练的通用AI设计平台。在10个以上药用或者工程蛋白上，通过湿实验验证所建通用AI模型的能力：利用少于100个蛋白突变体实验数据，就能获得性能优越、满足应用需求的多点位改造工程蛋白。实现三款及以上蛋白的产业转化。

经费额度：非定额资助，拟支持不超过1个项目，每项资助额度不超过500万元。

执行期限：2023年12月01日至2026年11月30日。

专题三、重大疾病精准诊疗关键技术

方向1：精准医学大数据分析增强计算系统

研究目标：建立分析能力自我进化和知识自主挖掘的精准医学增强计算平台，推理长度、准确度以及分析效率达到世界领先，完成不少于10个重要疾病相关知识的长链自动推理验证。

研究内容：围绕精准医学大数据深度分析和智能推理，建立多组学数据分析能力自我进化的增强计算系统。研发知识和数据双驱动的自主挖掘系统，实现生物医学研究的自主规划和实现。以肿瘤等重要疾病为例，进行知识长链的自主推理和数据分析验证。

经费额度：非定额资助，拟支持不超过1个项目，每项资助额度不超过200万元。

执行期限：2023年12月01日至2026年11月30日。

方向2：智能健康数字孪生队列技术

研究目标：发展数字孪生模拟技术，建立超过1万人的多尺度多模态数字孪生队列，实现疾病发生发展的预防预警。

研究内容：基于真实世界临床疾病、健康人群、生物样本等多模态数据，引入GPT等新型人工智能模型和元宇宙技术，建立数字孪生队列。发展深度学习、因果推理、状态空间模型等，基于数字孪生队列进行癌症、病毒等高发、突发疾病的预警预防研究。

经费额度：非定额资助，拟支持不超过1个项目，每项资助额度不超过200万元。

执行期限：2023年12月01日至2026年11月30日。

方向3：疾病队列多层次数据集成共性技术

研究目标：开发3种以上面向复杂疾病的多层次数据集成分析算法，识别5个以上疾病的亚型关键靶点，在解释病人预后等关键临床指标差异方面提升10%以上。

研究内容：针对肿瘤、代谢性疾病、心脑血管疾病等复杂疾病队列产生的分子表型、影像表型、临床表型等跨尺度多层次数据，发展基于前沿统计和人工智能手段的集成分析技术，构建疾病演化的核心轨迹，开展复杂疾病的精准分型，并深入揭示不同层次关键表型信号之间相互偶联的底层机制，促进临床转化应用。

经费额度：非定额资助，拟支持不超过2个项目，每项资助额度不超过200万元。

执行期限：2023年12月01日至2026年11月30日。

方向4：后全基因组关联分析技术

研究目标：开发致病位点精细定位预测方法，比现有工具可靠性提升至少10%，实现遗传位点与疾病关联相关通路、细胞、组织的预测，比现有工具可靠性提升至少20%。

研究内容：针对复杂疾病全基因组关联研究发现的大量风险位点，开发基于多尺度多组学大数据集成的复杂疾病风险位点功能解析平台。结合器官组织及单细胞水平表观遗传组学、转录组、蛋白组等数据，实现非编码位点致病机制和协同作用的解析，预测精准性等指标超越国际主流方法。

经费额度：非定额资助，拟支持不超过1个项目，每项资助额度不超过200万元。

执行期限：2023年12月01日至2026年11月30日。

方向5：医学大数据的生物统计共性算法

研究目标：面向精准医学大数据，开发非匹配样本多组学

整合和驱动因子鉴定算法，准确度均达到世界领先，在统计基础理论和方法架构等层面产出原创新成果。

研究内容：突破传统组学研究范式，构建数据降噪大模型，开发多组学多模态对角整合模型和优化算法，基于深度学习、动力学模型和因果推理等实现驱动因子鉴定和标志物识别，解决精准医学多模态数据匹配样本缺乏、多变量实验干预困难等问题，为精准医学和疾病早筛提供理论和技术支撑。

经费额度：非定额资助，拟支持不超过1个项目，每项资助额度不超过200万元。

执行期限：2023年12月01日至2026年11月30日。

二、 申报要求

除满足前述相应条件外，还须遵循以下要求：

1. 项目申报单位应当是注册在本市的法人或非法人组织，具有组织项目实施的相应能力。

2. 研究内容已经获得财政资金支持的，不得重复申报。

3. 所有申报单位和项目参与人应遵守科研伦理准则，遵守人类遗传资源管理相关法规和病原微生物实验室生物安全管理相关规定，符合科研诚信管理要求。项目负责人应承诺所提交材料真实性，申报单位应当对申请人的申请资格负责，并对申请材料的真实性和完整性进行审核，不得提交有涉密内容的项目申请。

4. 申报项目若提出回避专家申请的，须在提交项目可行性方案的同时，上传由申报单位出具公函提出回避专家名单与理由。

5. 已作为项目负责人承担市科委科技计划在研项目2项及以

上者，不得作为项目负责人申报。

6. 项目经费预算编制应当真实、合理，符合市科委科技计划项目经费管理的有关要求。

7. 各研究方向同一法人单位限报2项。

8. 获资助的项目负责人及其所在单位应承诺将项目所产生的研究成果和数据资料等报送市科委

三、 申报方式

1.项目申报采用网上申报方式，无需送交纸质材料。申请人通过“中国上海”门户网站（<http://www.sh.gov.cn>）--政务服务--点击“上海市财政科技投入信息管理平台”进入申报页面，或者直接通过域名<https://czkj.sheic.org.cn/>进入申报页面：

【初次填写】使用“一网通办”登录（如尚未注册账号，请先转入“一网通办”注册账号页面完成注册），进入申报指南页面，点击相应的指南专题，进行项目申报；

【继续填写】使用“一网通办”登录后，继续该项目的填报。有关操作可参阅在线帮助。

2.项目网上填报起始时间为2023年09月7日9:00，截止时间（含申报单位网上审核提交）为2023年09月26日16:30。

四、 评审方式

采用第一轮通讯评审、第二轮见面会评审方式。

五、 立项公示

上海市科委将向社会公示拟立项项目清单，接受公众异议。

六、 咨询电话

服务热线：021-12345、8008205114（座机）、4008205114
（手机）

上海市科学技术委员会

2023年8月30日

（此件主动公开）