**一、项目名称**

面向多源异构数据的生物医学模式发现关键技术

**二、提名单位及提名等级**

中南大学，湖南省自然科学二等奖

**三、提名意见**

生物医学大数据对于研究人类健康、疾病机制、药物研发和临床治疗等方面具有重要的意义。以人工智能为代表的新一代信息技术的发展，为面向多源异构数据的生物医学模式发现创造了条件。该项目融合多源异构信息，采用人工智能新方法和新技术，针对药物组合协同及癌细胞药物敏感性预测、非编码RNA-疾病关联及功能预测、蛋白质-RNA相互作用热点残基识别问题，研究开发了一系列新的、高效的计算预测方法，在国内外获得了广泛的认可和应用。

该项目的重要发现在Nucleic Acids Research、Bioinformatics等国际权威期刊发表，5篇代表作SCI总他引661次，单篇最大他引370次。总影响因子35.3，单篇最高影响因子14.9。得到了国内外著名科学家、大学、制药企业和研究机构的高度评价和广泛应用。

**四、项目简介**

生物医学大数据对于研究人类健康、疾病机制、药物研发和临床治疗等方面具有重要的意义。以人工智能为代表的新一代信息技术的发展，为面向多源异构数据的生物医学模式发现创造了条件。本项目融合多源异构信息，采用人工智能新方法和新技术，针对药物组合协同及癌细胞药物敏感性预测、非编码RNA-疾病关联及功能预测、蛋白质-RNA相互作用热点残基识别问题，研究开发了一系列新的、高效的计算预测方法，提升了现有方法的性能或效率。5篇代表作SCI总他引661次，单篇最大他引370次。总影响因子35.3，单篇最高影响因子14.9。项目主要创新成果如下：

1. 构建了组合药物数据库及癌细胞药物敏感性预测方法。针对组合药物筛选困难、实验成本高、数据缺乏等问题，构建了组合药物领域的第一个综合型数据库—DrugCombDB（代表作1）,包含～44万个药物组合，～605万个癌细胞定量药物剂量响应与药物浓度信息。针对癌细胞系表达基因数据的复杂性和高维特征问题，提出了一种信号通路集成的深度神经网络模型pathDNN（代表作5），利用信号通道将细胞系表达基因与药物靶标基因联系起来，集成于DNN模型中，以预测药物与癌细胞系之间的相互作用。实验表明pathDNN具有较高的预测能力和良好的可解释性。该成果得到美国哈佛大学Cyril H. Benes教授、美国密苏里哥伦比亚大学Dong Xu教授等知名科学家的高度评价和应用，被GraphSynergy、MOOMIN、NEXGB等十余个模型用作基准数据集和比较方法。该成果能为加快癌症药物研发与个性化医疗提供重要支撑。
2. 研发了基于异构网络集成的lncRNA-疾病关联及功能预测方法。针对lncRNA序列保守性差、生物实验成本高等问题，融合多种异构数据（lncRNA-疾病关联数据，蛋白质相互作用数据、lncRNA表达谱数据、lncRNA-蛋白质相互作用数据等），构建全局异构网络，提出了基于Katz度量方法的lncRNA的功能预测模型KATZLGO（代表作2），研发了基于网络传播算法的大规模lncRNA-疾病关联预测方法LncRDNetFlow（代表作3），显著提升了预测的准确性和覆盖度。该成果得到日本东京大学Tatsuhiko Tsunoda教授、中国科学院深圳理工大学计算机科学与控制工程学院院长潘毅教授、密西根大学国家材料科学与工程研究中心主任张阳教授等国内外知名学者的好评和引用。该成果有助于更全面地理解疾病的发生机制，寻找新的疾病诊断标志物以及治疗靶点。
3. 提出了基于多特征融合的蛋白质-RNA相互作用热点残基识别方法。在蛋白质-RNA相互作用热点残基识别方面，首次结合序列特征、结构特征、网络特征和溶剂暴露特征，将梯度提升树、支持向量机和随机森林模型的预测结果进行加权平均，取得了更加准确的预测结果（代表作4）。该工作得到了美国橡树岭国家实验室生物部门主任Julie C Mitchell、英国剑桥大学David B. Ascher教授等知名专家的正面评价和应用。该成果能为创新药物研究和蛋白质功能设计提供支撑，具有广泛的应用前景。

**五、代表作目录**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 序号 | 代表作及论文名称/刊名/作者 | 年卷页码 | 发表时间（年月 日） | 通讯作者(含共同) | 第一作者(含共同) | 国内作者(排序) | 知识产权是否归国内所有 |
| 1 | DrugCombDB: a  comprehensive database of drug combinations toward the discovery of combinatorial therapy/Nucleic Acids  Research/刘辉、章文浩、邹博、王敬贤、邓媛媛、邓磊 | 2020年48卷D871-  881页 | 2020年  01月  08日 | 邓磊 | 刘辉 | 刘辉/1，章 文浩/2，邹博/3，王敬  贤/4，邓媛  媛/5，邓磊  /6 | 是 |
| 2 | KATZLGO: Large-Scale Prediction of LncRNA Functions by Using the KATZ Measure Based on Multiple Networks/IEEE/ACM  Transactions on Computational Biology and Bioinformatics/张祖平、张敬普、范超、唐勇军、邓磊 | 2019年16卷407-  416页 | 2019年  03月  01日 | 邓磊 | 张祖平 | 张祖平  /1，张敬普 /2，范超  /3，唐勇军  /4，邓磊/5 | 是 |
| 3 | Integrating Multiple  Heterogeneous Networks for  Novel LncRNA-Disease  Association  Inference/IEEE/ACM  Transactions on  Computational Biology and  Bioinformatics/张敬普、张祖平、陈志刚、邓磊 | 2019年16卷396-  406页 | 2019年  03月  01日 | 邓磊 | 张敬普 | 张敬普  /1，张祖平  /2，陈志刚  /3，邓磊/4 | 是 |
| 4 | Computational identification  of binding energy hot spots in  protein–RNA complexes  using an ensemble  approach/Bioinformatics/潘玉亮、王子祥、詹卫华、邓磊 | 2018年34卷1473-  1480页 | 2018年  05月  01日 | 邓磊 | 潘玉亮 | 潘玉亮 /1，王子祥 /2，詹卫华  /3，邓磊/4 | 是 |
| 5 | Pathway-Guided Deep Neural Network toward Interpretable and Predictive Modeling of Drug Sensitivity/Journal of  Chemical Information and Modeling/邓磊、蔡一灯、 章文浩、杨文艺、高波、刘辉 | 2020年60卷4497-  4505页 | 2020年  10月  26日 | 高波 、刘 辉 | 邓磊 | 邓磊/1，蔡 一灯/2，章文浩/3，杨  文艺/4，高  波/5，刘辉  /6 | 是 |

**六、主要完成人**

邓磊，张祖平，刘辉，张敬普，潘玉亮

**七、主要完成单位**

中南大学，常州大学

**八、主要完成人合作关系说明**

项目主要完成人邓磊、张祖平为中南大学计算机学院教授，对本项目重要科学发现做出了贡献。项目组第一完成人邓磊教授在“面向多源异构数据的生物医学模式发现方法”方面负责该项目的总体设计，制定研究总体方案、技术路线与实施计划。研发了组合药物数据库及癌细胞药物敏感性预测方法，基于异构网络集成的lncRNA-疾病关联及功能预测方法，基于多特征融合的蛋白质-RNA相互作用热点残基识别方法，提升了现有方法的性能或效率。是代表作1-5的第一或通讯作者；2项国家自然基金项目项目、1项湖南省自然科学基金项目的主持人。项目组第二完成人张祖平教授负责lncRNA-疾病关联及大规模lncRNA功能预测方向的技术方案制定及实施，是代表作2、3的第一作者或共同作者。

项目组第三完成人刘辉现为南京工业大学教授，在常州大学工作期间与中南大学邓磊教授课题组进行了深入合作，负责组合药物数据库及癌细胞药物敏感性预测技术方案制定及实施，是代表作1、5的第一作者或通讯作者。1项国家自然基金项目项目主持人。

项目组第四完成人张敬普为张祖平教授、邓磊教授共同指导的博士生，毕业后在河南城建学院担任讲师，博士期间开发了基于异构网络集成的lncRNA-疾病关联及功能预测模型，是代表作3的第一作者，代表作2的第二作者，1项国家自然科学基金的主要参与人。

项目组第五完成人潘玉亮博士毕业于同济大学，其攻读硕士期间在邓磊教授的指导下负责基于多特征融合的蛋白质-RNA相互作用热点残基识别技术研发，是代表作4的第一作者。