2020年湖南省科学技术奖励拟提名项目公示内容

（自然科学奖）

|  |  |
| --- | --- |
| **项目名称** | 生物网络特征发现与高效算法设计 |
| 提名意见 | 中南大学是本项目的第一完成单位。取得的主要创新成果包括：  1.发现了DNA测序读数重叠网络中节点度的小参数特征，设计了精确、快速的**单体型组装算法**，解决了大量长DNA测序读数的**单体型分型问题**，推动了下游遗传变异分析和应用研究。  2.首次提出了基于3-sigma法则的活性蛋白质检测算法，能够有效识别在整个细胞周期内表达水平低但变化较大的活性蛋白质，为时序网络分析、动态网络构建等研究提供了高质量的分析框架。  3.发现了关键蛋白质和蛋白质复合物在相互作用网络中的成簇特征，提出了基于边聚集系数的网络中心性，提高了基于网络水平的关键蛋白质识别的准确率，提出的网络功能模块预测方法能有效识别具有特定生物学意义的功能模块，被广泛应用在**医学、药学、植物学**等领域。  4. 发现了药物-疾病之间的不确定关系的隐含特征，设计了基于多相似性集成和矩阵填充的药物重定位算法，能够有效挖掘现有药物的新适应症并为新发疾病的治疗提供候选药物。  8篇代表性论文中的6篇以中南大学为第一单位分别发表在生物信息学领域国际权威期刊 Bioinformatics、Proteomics等上，其中4篇为ESI Top1%高被引论文。8篇论文SCI他引455次，研究成果得到了许多国内外知名学者的肯定和高度评价，并在生物、医学等领域得到广泛应用。  提名该项目为湖南省自然科学奖一等奖。 |
| 项目简介 | 生命系统各组成元件间通过协同合作共同行使功能，由**基因、RNA、蛋白质**等生命系统组成元件及其相互作用构成的复杂生物网络蕴含着探索生命机制和规律的重要信息。特别是随着高通量技术的迅猛发展，**复杂、异构、高维、高噪声**的生物医学大数据不断累积，挖掘其内在属性，发现与算法设计相关的关键特征，设计有效的算法，从而寻求对生命科学领域的难解科学问题的解答，成为计算科学与生命科学交叉领域所面临的巨大挑战。本项目围绕复杂生物网络的特征发现和高效算法设计开展了系统的研究，取得的主要科学发现如下：  1. 通过对DNA测序读数（DNA read）重叠网络结构特征的分析，发现了该网络顶点度的小参数特征，设计了启发式快速动态规划算法，首次实现了在大量长DNA测序读数上进行单体型高效组装，推动了下游基于单体型分型的遗传变异分析和应用。  2. 通过对时序下基因表达分布特征的分析，首次提出了基于3-sigma法则的活性蛋白质检测算法，能够有效识别在整个细胞周期内表达水平低但变化较大的活性蛋白质，为时序网络分析、动态网络构建等研究提供了高质量的分析框架。  3. 发现了关键蛋白质和蛋白质复合物在相互作用网络中的成簇特征，提出了基于边聚集系数的网络中心性，提高了基于网络水平的关键蛋白质识别的准确率，提出的网络功能模块预测方法能够有效识别具有特定生物学意义的功能模块，被广泛应用在医学、药学、微生物学、植物学等领域。  4. 通过分析异质性生物实体相互关联构成的复杂网络，发现了药物-疾病之间的不确定关系的隐含特征，设计了基于多相似性集成和矩阵填充的药物重定位算法，能够有效挖掘现有药物的新适应症并为新发疾病的治疗提供候选药物，为相关药物和疾病的进一步研究提供重要的实验参考。  项目遴选的8篇代表性论文Google scholar总引用870余次，SCI总引用591次（其中SCI他引455次，单篇最大SCI他引次数202次）。ESI Top1%高被引论文4篇。研究成果得到了国内外学者的高度评价和广泛应用，市场上占有份额最多的测序公司Illumina利用本项目提出的H-BOP算法进行单体型分型（Nature biotechnology，IF: 31.864）；美国物理学会院士、匈牙利科学院院士Albert-Laszlo Barabasi教授等利用本项目提出的基于3-sigma法则的活性蛋白质检测算法构建的网络作为其研究时序网络的来源（Science，IF: 41.037）；研发的复杂生物网络关键节点分析和模块化结构分析软件被下载使用3.6万余次，被广泛应用于医学、药物、植物学、微生物学等领域，例如：哺乳动物卵母细胞减数分裂过程中的关键基因研究（Nature，IF: 43.07）、注意力缺陷多动障碍症研究（Nature Genetics，IF: 25.455）。 |
| **客观评价** | 本项目遴选的8篇代表性论文，发表在生物信息学领域的国际权威期刊Bioinformatics、Proteomics、IEEE/ACM TCBB等上，SCI他引455次。许多国内外知名学者对我们的研究成果给出了高度评价，提出的单体型组装算法得到了德国马普所的高级研究员Felicity C. Jones等的肯定，指出了H-BOP算法解决了较长DNA片段的单体型组装问题，具有能充分利用DNA分子大范围内的变异信息进行单体型分型的优点。也有许多国内外知名学者直接应用我们提出的系列方法进行研究，在医学、生物学、药学、植物学、微生物学等领域取得重要研究成果。美国物理学会院士，匈牙利科学院院士Albert-Laszlo Barabasi教授利用本项目提出的基于3-sigma法则的活性蛋白质识别和动态网络构建方法所构建的网络作为其研究的时序网络的来源，其进一步研究成果发表在国际权威期刊SCIENCE上；宾夕法尼亚大学医学院的研究人员利用本项目提出的蛋白质复合物识别方法在注意力缺陷多动障碍症方面取得的重要研究成果，发表在遗传学最好的期刊NATURE GENETICS上；哈佛大学医学院、加利福尼亚大学伯克利分校、普林斯顿大学、橡树岭国家实验室等的研究人员在病毒学、微生物学和植物学等领域研究病毒的工作机制、微生物变异及拟南芥和杨树的细胞功能等都采用了本项目提出的方法并取得了重要研究成果。其中，代表性的评价与应用如下：  美国物理学会院士，匈牙利科学院院士Albert-Laszlo Barabasi教授等利用本项目提出的基于3-sigma法则的动态蛋白质网络构建方法所构建的动态蛋白质网络DPIN作为研究时序网络的来源（a biological network with time-varying protein-protein binding interactions [27]（文献[27]即代表性论文[2]）），其成果发表在国际顶级期刊（Science，IF: 41.037）上（见代表性引文[1]）。在论文致谢部分，Albert-Laszlo Barabasi教授等特别感谢了本项目完成人王建新教授提供动态蛋白质交互网络数据（We appreciate J. Wang and X. Peng for supplying the data of dynamic protein-protein interaction networks）。  市场上占有份额最多的测序公司Illumina的高级研究员Frank J. Steemers等提供一个简单、快速、自动化的单体型分型测序方法，发表在国际顶级期刊Nature biotechnology（IF: 31.864）上（见代表性引文[2]）。他们设计了单试管微颗粒长分子测序，获得通过重叠区能连接起来的长DNA读数集(islands)，利用本项目提出的H-BOP算法（The assignment of the haplotypes for the islands was performed by H-BOP）进行单体型分型，完成了N50为63kb的重叠读数集上的单体型分型任务。  德国马普所的高级研究员Felicity C. Jones等利用Illumina公司的技术对一只老鼠和一只小棘鱼的生殖细胞基因组进行测序，采用本项目提出的H-BOP算法进行单体型分型，精确重建了减数分裂基因重组热点图，相关成果发表在国际顶级期刊Nature Communications（IF: 11.878）上（见代表性引文[3]）。他们在论文中肯定了H-BOP算法的优点，即能够处理较长DNA片段，能充分利用DNA分子大范围内的变异信息进行单体型分型（HBOP is a single individual phasing algorithm that can take into account reads belonging to a longer DNA fragment and therefore capitalizes on the long-range information of the molecule during phasing）。  美国系统生物学研究所的首席科学家Glusman教授等在其综述中充分肯定了H-BOP的速度和精度（Genome Medicine, IF: 10.886）（见代表性引文[4]），详细介绍了ReFHap算法，指出该算法在基于forsmid测序技术产生的低覆盖度的长测序读数数据上生成了当时最好的单体型分析结果，并将H-BOP作为对RefHap在单体型分型的速度和精度改进的代表性算法进行了介绍（More recent algorithms claim improvements on parameters such as speed and accuracy (for example, H-BOP [38]) ）。在其全基因组单体型分型软件概览表Table 2中把H-BOP作为典型的基于测序读段的单体型分型方法列出。  英国医学研究理事会，分子生物实验室Melina Schuh 教授等利用本项目研发的ClusterViz软件对哺乳动物卵母细胞-siRNA网络进行聚类分析，找到了6个聚集度非常高的siRNA和phenotypes聚类（The clusters were obtained using ClusterViz (https://code.google. com/p/clustervizcytoscape/) of Cytoscape）。研究成果发表在了国际顶级期刊Nature（IF: 43.07）上（见代表性引文[5]）。  宾夕法尼亚大学医学院应用基因组学中心主任Hakon Hakonarson 博士等在遗传学国际顶尖期刊Nature Genetics（IF: 25.455）上发表论文（见代表性引文[6]），在注意力缺陷多动障碍症研究中，采用本项目组开发的ClusterViz软件中的FAG-EC算法进行聚类得到17个簇，发现了GRM受体基因在患病组内发生复制或删除。（We clustered this network of genes into 17 distinct modular clusters based solely on network topology using the ClusterViz plugin for the software using the FAG-EC algorithm with default parameters）。  瑞士弗里堡大学终身教授张翼成在其发表在Physics Reports 2018（IF: 28.295）的综述论文（见代表性引文[7]）将本项目提出的NC算法作为典型算法进行描述（The recent advances extend identification of essential nodes by using the topological features of protein–protein interaction (PPI) networks such as based on edge clustering coefficient [259,260]. (文献[259]即代表性论文[3]) )。  项目提出的两个药物重定位算法MBiRW和DRRS，受到研究人员的广泛关注和引用，被作为AI辅助药物重定位的典型算法进行比较和分析。Briefings in Bioinformatics、Bioinformatics等生物信息学领域的国际权威期刊上的十余篇研究论文都将其作为经典方法进行引用和描述，并进行对比分析。例如，南洋理工大学的Chee-Keong Kwoh教授在其发表在Briefings in Bioinformatics（IF: 9.101）的综述论文（见代表性引文[8]）中详细描述了MBiRW和DRRS两个算法，并将其作为典型算法进行比较分析（Luo et al. [176] simultaneously implemented random walk with different steps on the drug similarity network and disease similarity network to …... Besides, Luo et al. [177] constructed a low-rank matrix approximation to fill out the unknown entries in the drug–disease matrix, …...（文献[176]即代表性论文[5]、文献[177]即代表性论文[8]) )。  本项目的完成人王建新教授是医疗大数据应用技术国家工程实验室副主任、生物信息学湖南省重点实验室主任，担任ACM SGIBIO China主席和CCF生物信息学专业委员会主任，享受国务院政府特殊津贴，是国务院学位委员会第七届学科评议组成员、973前期研究专项首席科学家，担任国际期刊IEEE/ACM TCBB、Current Protein and Peptide Science、IJBRA编委，是FAW2015等8个国际会议的大会主席或程序委员会主席。 |
| **代表作及论文目录** | [1] Minzhu Xie, Jianxin Wang, Tao Jiang. A fast and accurate algorithm for single individual haplotyping. BMC Systems Biology, 2012, 6: S8.  [2] Jianxin Wang, Xiaoqing Peng, Min Li, Yi Pan. Construction and application of dynamic protein interaction network based on time course gene expression data. Proteomics, 2013, 13(2):301-312.  [3] Jianxin Wang, Min Li, Huan Wang, Yi Pan. Identification of Essential Proteins Based on Edge Clustering Coefficient. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2012, 9(4):1070-1080.  [4] Jianxin Wang, Jiancheng Zhong, Gang Chen, Min Li, Yi Pan. ClusterViz: a cytoscape APP for cluster analysis of biological network. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2015,12(4):815-22.  [5] Huimin Luo, Jianxin Wang, Min Li, Junwei Luo, Xiaoqing Peng, Fang-Xiang Wu, Yi Pan. Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm. Bioinformatics, 2016, 32(17): 2664-2671.  [6] Yu Tang, Min Li, Jianxin Wang, Yi Pan, Fang-Xiang Wu. CytoNCA: a cytoscape plugin for centrality analysis and evaluation of biological networks. BioSystems, 2015,127: 67-72.  [7] Minzh Xie, Qiong Wu, Tao Jiang. H-PoP and H-PoPG: heuristic partitioning algorithms for single individual haplotyping of polyploids. Bioinformatics, 2016, 32(24): 3735-3744.  [8] Huimin Luo, Min Li, Shaokai Wang, Quan Liu, Yaohang Li, Jianxin Wang. Computational Drug Repositioning using Low-Rank Matrix Approximation and Randomized Algorithms. Bioinformatics, 2018, 34(11):1904-1912. |
| **主要完成人情况** | **摘自“主要完成人情况表”中的部分内容，公示姓名、排名、行政职务、技术职称、工作单位、主要完成单位、对本项目的贡献**  王建新，排名1，教授/院长，主要完成单位：中南大学。对项目科学发现1、2、3、4都做出了重要贡献，提出了单体型组装的启发式快速动态规划算法；提出了基于3-sigma法则的活性蛋白质检测算法；提出了基于边聚集特性的网络中心性指标NC；主导研发了基于复杂生物网络的关键节点分析与识别平台CytoNCA和基于Cytoscape3.0的生物网络聚类分析平台ClusterViz；提出了基于集成的相似性度量和双向随机游走的药物重定位方法；提出了基于低秩矩阵填充的药物重定位方法。相关工作量占本人工作量的70%，是代表性论文1-8的主要作者。  谢民主，排名2，教授，主要完成单位：湖南师范大学。对项目科学发现1做出了重要贡献，提出了单体型组装的启发式快速动态规划算法。相关工作量占本人工作量的80%，是代表性论文1和7的主要作者。 |
| **主要完成单位情况** | **摘自“主要完成单位情况表”中的部分内容，公示单位名称、排名、对本项目的贡献**  本项目主要完成人依托中南大学，该项目在研期间，中南大学给予财政资助以便开展前期研究，提供场地及后勤保障措施保证项目的顺利实施，鼓励项目完成人积极参与申报国家及省部级基金，并对成功申报课题给予配套资金支持，组织和指导该项目的总结工作和成果推荐工作。 同时，中南大学积极扶持该项目学术团队成长，具有稳定的学术梯队，也为后续研究的开展奠定了人员基础。 |
| **主要完成人合作关系说明** | 本项目主要完成人之间**或有师承关系**、**或有长期合作关系**，具体合作关系如：中南大学王建新教授、湖南师范大学谢民主教授共同承担了“生物网络特征发现与高效算法设计”项目。其中谢民主为中南大学博士研究生，在攻读博士学位期间参与了该项目的研究工作，并在获得博士学位之后继续与王建新进行科研合作。 |