

青岛市科学技术奖项目公示内容

科技进步奖

一、项目名称

基于基因序列的遗传疾病信息挖掘技术及其应用

二、推荐单位（专家）意见（不超过 600 字符）

我单位认真审阅了该项目推荐书及其附件材料，确认真实有效，相关栏目符合填写要求。

按照要求，我单位及完成人所在单位均进行了公示，确认完成人、完成单位排序无异议。

该项目研究基于基因序列的遗传疾病信息挖掘技术及其应用。提出了基因序列搜索、单体型格局生成等关键技术，在全基因组序列中有效定位蛋白质调控序列的位置和长度，增加了等位基因补全率、降低了算法时间和空间复杂性；提出了基因通路、miRNA 和 SNP 标签选择等有效方法，基于 PAGIS 和 PADOG 识别基因通路，结合基因的重要性和特异性，考虑到通路中调控大量基因的重要性，提高了通路识别精度，并提出 SNP 标签集选择和预测方法，提高了 SNP 标签集的精度；设计了基于图模型的复杂疾病关键功能模块识别系统以及基于概率模型的癌症网络标志物确定方法及系统，利用基于图结构注意力的分子相似性搜索算法以及基于边驱动的标签传播算法的蛋白质功能模块识别方法，在保证搜索广泛性的同时，提高了分子相似性计算的准确度，能够在较大规模的 PPI 网络中获得更精确有效的蛋白质功能模块划分结果。获得授权发明专利 6 件，著作权 4 项，发表论文 7 篇，获得科技奖励 1 项。该项目技术居于国内领先水平。

该项目自 2017 年起由包括济南科虹生物技术有限公司、济南恒山生物工程有限公司、山东新岳腾电子科技有限公司等多家企业，分别在济南市、济宁市等进行了试应用，具有非常大的推广价值。同时，该项目取得的技术研究成果在山东省立第三医院、青岛市市立医院有良好的临床应用效果。

参照青岛市科学技术奖推荐条件，推荐该项目申报 2021 年度青岛市科学技术进步奖一等奖。

三、项目简介

该项目属于生物信息计算与深度挖掘领域。

2016 年 3 月国家发展改革委办公厅发布《关于第一批基因检测技术应用示范中心建设方案的复函》指出，大力发展基因检测技术、开展推广应用，有利于提高遗传性疾病、肿瘤、感染性疾病等重大疾病的防治水平。同年 6 月，国务院办公厅印发的《关于促进和规范健康医疗大数据应用发展的指导意见》突出强调了要集成基因组学、蛋白质组学等国家医学大数据资源。推进基因芯片与测序技术在遗传性疾病诊断、癌症早期诊断和病预防检测方面的应用，有利于加强人口基因信息安全管理，推动精准医疗技术发展。

新一代基因测序技术的飞速发展使得人类基因组计划提前完成，核酸数据库、基因的遗传、物理及转录表达图谱等已趋完整，为相关领域研究者提供了海量的高密度遗传多态性标志信息、基因序列、蛋白质及表型信息。但现有的生物信息计算与深度挖掘技术尚存在一些关键问题有待解决：（1）大多数研究聚焦于 DNA 甲基化位点的修饰变化，无法全面概括病症的表观遗传学机制；（2）缺少针对多基因组学数据整合分析中维度灾难、扩展性、归一化等问题的解决方法；（3）难以构建分析 SNP、基因和表型数据间关系的形式化模型，从而无法识别人类复杂疾病的致病基因及其发病机理。

针对这三个瓶颈问题，该项目依托国家自然科学基金、国家自然科学基金青年科学基金项目等多项研究课题，在生物信息计算与深度挖掘领域，研究了基因序列搜索、单体型格局生成等关键技术，提出了基因通路、miRNA 和 SNP 标签选择等有效方法，实现了基于图模

型的复杂疾病关键功能模块识别系统、基于概率模型的癌症网络标志物确定方法及系统。

该项目成果实现了蛋白质调控基因的恢复,可以在全基因组序列中有效定位蛋白质调控序列的位置和长度,能够在较大规模的蛋白质相互作用网络中获得更精确有效的蛋白质功能模块划分结果;增加了等位基因补全率、降低了算法时间和空间复杂性;能够准确有效的获取癌症网络标志物,对疾病进行分类,实现疾病的精准诊断和治疗。该成果为人类遗传学研究中复杂疾病相关基因定位提供新思路和新方法,解决了生物医学、遗传学等相关领域研究的热点和难点,是后基因组时代的重要课题。

依托相关研究成果获得获得授权发明专利 6 件,著作权 4 项,发表论文 10 篇,获得科技奖励 1 项。该技术成果与国内外同类技术指标相比,药物关联分类预测精度 0.9098,高于同类 SCI 期刊的 0.6011;肺癌识别准确率 85.7%,高于同类 SCI 期刊的 72%。通过对比发现,现有的研究成果中,对于药物关联分类预测、基于图模型的复杂疾病识别等技术还存在一定的局限性,预测的精确度与识别的准确率都没有达到理想效果。通过该项目中主要发明创新技术的研究,提高了算法识别过程的稳定性,能够在较大规模的网络中获得更精确有效的功能模块划分结果,降低了算法的时间和空间复杂性,高了药物关联分类预测精度。

四、客观评价

1、鉴定评价

2021 年 8 月 20 日,青岛智慧农大技术服务有限公司(青岛市科技成果标准化评价机构)对《基于基因序列的遗传疾病信息挖掘技术及其应用》成果进行评价:“该成果提出了基于密码子模板的真核生物功能基因序列搜索方法,实现了蛋白质调控基因的恢复;提出了基于遗传学规律的等位基因准确推断方法,增加了等位基因补全率、降低了算法时间和空间复杂性;提出了基于 PAGIS 和 PADOG 识别基因通路的方法,提供了一种基于概率模型的癌症网络标志物确定方法及系统,实现疾病的精准诊断和治疗;提出了一种新的基于图结构注意力的分子相似性搜索算法,在保证搜索广泛性的同时,提高了分子相似性计算的准确度;提出基于边驱动的标签传播算法的蛋白质功能模块识别方法,在较大规模的 PPI 网络中获得更精确有效的蛋白质功能模块划分结果”。“该成果技术成熟度为 9 级,技术创新度为 2 级,技术先进性为 6 级,项目成果整体水平达到国内领先水平”。

2、查新评价

2021 年 08 月,教育部科技查新工作站 SH02,项目名称“基于基因序列的遗传疾病信息挖掘技术及其应用”。查新结论为:“通过对所检国内相关文献的对比和分析,在国内公开发表的中文文献中,除本委托课题项目组成员发表的研究成果外,未见其他与本委托课题技术要点相符的报道”。

3、用户评价

(1) 济南科虹生物技术有限公司

利用本项目中基于 PAGIS 识别基因通路的方法,结合基因的重要性和特异性来识别通路,快速获取每一个基因的基因频度及基因出度。研究人员通过技术分析得出:该项目中相关技术能够对基因通路识别技术不断进行优化,提高了通路的识别精确度,使基因通路识别准确率达到 95%以上。

(2) 济南恒山生物工程有限公司

利用本项目中基于 PADOG 识别基因通路的方法,计算每一个信号道路的权重进行排序。公司研究人员通过系统测试:该类方法通过每一个基因的综合权重得到信号的通路权重,可以有效确定最大信号通路权重所对应的信号通路出现变化的概率最大。

(3) 山东新岳腾电子科技有限公司

利用本项目中基于基因通路识别小分子核糖核酸的方法,可以在短时间内识别基因通路并筛选出疾病潜在的 miRNA 标志物。研究人员利用该方法对复杂疾病标志物识别系统进行

测试与优化，通过实验测试，实施该类方法可以有效降低复杂度，且使得潜在 miRNA 的识别准确率能够达到 90%以上。

(4) 山东瑞佳园实业发展集团有限公司

利用本项目中“一种基于密码子模版的真核生物功能基因序列搜索方法”、“基于条件概率的基因型指派方法”等一系列基因定位系列方法，持续测试与优化复杂疾病相关基因区域或位点定位系统。通过实验数据分析，研究人员认为该类方法能够有效地在全基因组序列中定位出蛋白质调控序列的位置和长度，降低了算法的时间和空间复杂性，提高了系统中连锁和关联分析的数据准确性。

(5) 山东赛欧斯生物医学科技有限公司

利用本项目中基于边驱动的标签传播算法的蛋白质功能模块识别方法等对复杂疾病关键功能模块识别系统进行优化。研究人员认为该方法可获得更加准确的功能模块划分结果，对疾病的分类准确率可以达到 85.7%以上。

4、国内外学术性评价意见

(1) 英国雷丁大学，食品和营养科学系，人类营养和心血管代谢研究所，Karani Santhanakrishnan Vimalaswaran 教授评价该项目的 SNP 标签集的选择和预测方法为未来的重要研究方向，代表性论文 TagSNP-set selection for genotyping using integrated data. Future Generation Computer Systems 属于这方面研究成果。

(2) 德国遗传学与生物计量研究所，在遗传学领域具有突出的贡献的 Dörte Wittenburg 教授评价该项目的标记 SNP 集的选择方法为未来的重要研究方向，代表性论文 Tagging SNP-set selection with maximum information based on linkage disequilibrium structure in genome-wide association studies 属于这方面研究成果。

(3) 2021 年，菲律宾帝里曼大学，Francis George C. Cabarle 教授评价了细菌计算系统的三种规则，该系统具有强大的计算能力，代表性论文 Small universal bacteria and plasmid computing systems 属于这方面研究成果。

(4) 该项目代表性论文 Genome-wide differential-based analysis of the relationship between DNA methylation and gene expression in cancer 借助差异分析的全基因组层面的 DNA 甲基化与基因表达的关系研究，多位学者评价该方法能够为后续癌症的机理研究提供依据。

(5) 韩国崇实大学，Je-Keun Rhee 教授发表在 TOP 期刊 Briefings in Bioinformatics 上的工作肯定了该项目提出的肿瘤纯度计算方法，评价该方法具有更高的准确性，代表性论文 PEIS: a novel approach of tumor purity estimation by identifying information sites through integrating signal based on DNA methylation data 属于这方面研究成果。

5、重要科技奖励

2019 年 10 月，第五届中国“互联网+”大学生创新创业大赛，“胆石‘简’测——全球首创胆石病智能诊断辅助者（基于脉冲神经膜系统的胆石病自动识别方法）”获得银奖。该项目构建了高质量的胆结石数据集，利用 480 万张胆石病人的医学影像形成了人工智能学习的“知识库”，联合山东省立第三医院等 72 家医院、大学和医疗器械公司，形成了胆石病自动识别、结石 3D 重建与治疗方案推荐的全球首个胆石病智能诊断辅助系统。该系统能够在 3-5 秒实现对胆结石快速检测，在临床实验中为影像科医生节省超过 60%的诊断时间，大幅提高了医院接诊效率，有效缓解了看病难的问题。

五、主要完成人情况

姓名	排名	行政职务	技术职称	工作单位	完成单位	对本项目贡献
王淑栋	1	无	教授	中国石油大学 (华东)	中国石油大学 (华东)	提出了基于条件概率的基因型指派方法、等位基因准确推断方法、单体型格局生成方法。技术创新点 1 和 2 的主要贡献者。
王珣	2	无	副教授	中国石油大学 (华东)	中国石油大学 (华东)	提出了等位基因准确推断方法、特异性核苷酸序列模体识别和细胞核转运信号模体识别关键技术。技术创新点 2 和 3 的主要贡献者。
张媛媛	3	无	副教授	青岛理工大学	青岛理工大学	提出了标签 SNP 集选择方法。技术创新点 2 的主要贡献者。
张玉林	4	无	副教授	山东科技大学	山东科技大学	提出了单体型格局生成关键技术中构建合适形式化模型的方法，为研究分析 SNP 和基因之间的关系做出了创造性贡献。技术创新点 1 的主要贡献者。
王爽	5	无	讲师	中国石油大学 (华东)	中国石油大学 (华东)	提出了基于图结构注意力的分子相似性搜索算法。技术创新点 3 的主要贡献者。
刘文斌	6	无	教授	广州大学	广州大学	发明了癌症网络标志物确定方法及系统、基于 PAGIS 识别基因通路的方法、基于 PADOG 识别基因通路的方法、基于基因

						通路识别小分子核糖核酸的方法。技术创新点 2 和 3 的主要贡献者。
苏延森	7	无	副教授	安徽大学	安徽大学	发明了基于边驱动的标签传播算法的蛋白质功能模块识别方法。技术创新点 3 的主要贡献者。
宋弢	8	副院长	教授	中国石油大学 (华东)	中国石油大学 (华东)	设计了智慧医疗框架下的胆结石病自动识别方法，参与了基于脉冲神经膜系统的快速并行图像细化算法，并参与系统平台研发及推广应用工作。技术创新点 1 和 3 的主要贡献者。

六、完成人合作关系说明

(1) 王珣，本项目第 2 完成人，2018 年开始与本人合作“基于密码子模板的真核生物功能基因序列搜索方法与肺癌亚型分类”的研究工作。

(2) 张媛媛，本项目第 3 完成人，2017 年开始与本人合作进行“癌症性状相关基因的鉴定系统”的策划与研发。

(3) 张玉林，本项目第 4 完成人，2017 年开始与本人合作进行“基于机器学习的高精度追踪结缔组织的方法”的研发与推进。

(4) 王爽，本项目第 5 完成人，2020 年开始与本人合作推进“基于卷积神经网络的药物-蛋白质相互作用预测”的攻克工作。

(5) 刘文斌，本项目第 6 完成人，2003 年开始与本人合作参与“最大匹配问题的 DNA 表面计算模型”的设计工作。

(6) 苏延森，本项目第 7 完成人，2011 年开始与本人合作参与“拟南芥基因网络的构建与分析方法”的研究。

(7) 宋弢，本项目第 8 完成人，2015 年开始与本人合作参与“基于 DNA 计算的可逆门和可逆时序电路设计方法”的推进与应用。

七、推广应用情况

项目成果自 2017 年以来，完成新增销售收入 46915.88 万元，新增利润 9187.46 万元。其中 2020 年完成新增销售收入 9034.56 万元，新增利润 1845.77 万元；2021 年完成新增销售收入 8243.20 万元，新增利润 1552.03 万元。

序号	单位名称	应用的技术	应用对象及规模	应用起止时间	单位联系人/电话
1	济南科虹生物技术有限公司	基于基因序列的遗传疾病信息挖掘技术及其应用	复杂疾病关键功能模块识别系统,近五年新增销售额累计 3331.34 万元;新增利润为项目收入扣除项目成本、分摊的费用及相应税费所得,累计 1349.88 万元。	2017.02 至 2021.07	李建光 13854794071
2	济南恒山生物工程技术有限公司	基于基因序列的遗传疾病信息挖掘技术及其应用	复杂疾病标志物识别系统,近五年新增销售额累计 2801.95 万元;新增利润为项目收入扣除项目成本、分摊的费用及相应税费所得,累计 1170.04 万元。	2017.02 至 2021.07	王永斌 15554161555
3	山东新岳腾电子科技有限公司	基于基因序列的遗传疾病信息挖掘技术及其应用	癌症网络标志物识别系统,近五年新增销售额累计 1202.63 万元;新增利润为项目收入扣除项目成本、分摊的费用及相应税费所得,累计 614.47 万元。	2017.02 至 2021.07	杨洪伟 13553193831
4	山东瑞佳园实业发展集团有限公司	基于基因序列的遗传疾病信息挖掘技术及其应用	复杂疾病相关基因区域或位点定位系统,近五年新增销售额累计 2579.96 万元;新增利润为项目收入扣除项目成本、分摊的费用及相应税费所得,累计 1053.07 万元。	2017.02 至 2021.07	孙兰云 13953766627
5	山东省立第三医院	基于图模型的复杂疾病关键功能模块识别技术	2017 年 01 月至 2021 年 07 月多次用于患者的癌症诊断。经临床证明利用这些癌症网络标志物对疾病进行亚型分类,准确率在 95%以上,节省了医师资源和医疗成本。	2017.01 至 2021.07	李丕宝 13505314346
6	青岛市立医院	基于基因通路的复杂疾病标志物识别技术	自 2017 年 01 月至 2021 年 07 月辅助医生在基因治疗方面对癌症完成快速精准的基因诊断。截止至目前服务患者 1000 余人次,经专家鉴定,医生反馈,该系统能够准确有效的识别基因通路并快速高效定位潜在的疾病相关 miRNA。该方案在临床基因通路识别准确率达到 95%以上,潜在 miRNA 识别准确率达到 90%以上,节省了医师资源.和医疗成本,特此证明。	2017.01 至 2021.07	王子轩 13608979916

八、主要知识产权和标准规范等目录

知识产权(标准)类别	知识产权(标准)具体名称	国家(地区)	授权号(标准编号)	授权(标准发布)日期	证书编号(标准批准发布部门)	权利人(标准起草单位)	发明人(标准起草人)	发明专利(标准)有效状态
发明专利权	一种基于密码子模板的真核生物功能基因序列搜索方法	中国	CN107480473B	2021-02-26	4275110	中国石油大学(华东)	王珣、宋弢、朱虎	有效专利
发明专利权	一种基于PADOG识别基因通路的方法	中国	CN107220526B	2020-08-25	3953894	广州大学	刘文斌、沈良忠、咎乡镇	有效专利
发明专利权	基于边驱动的标签传播算法的蛋白质功能模块识别方法	中国	CN108388769B	2018-03-01	3717728	安徽大学	邱剑锋、张兴义、程凡、苏延森、张磊、王丛涛、聚秀琴	有效专利
发明专利权	一种基于PAGIS识别基因通路的方法	中国	CN107133492B	2017-09-05		温州大学	刘文斌、沈良忠、咎乡镇	有效专利
发明专利权	一种基于基因通路识别小分子核糖核酸的方法	中国	CN108319816B	2021-04-23	4377641	广州大学	刘文斌、崔振东、咎乡镇	有效专利
发明专利权	一种基于概率模型的癌症网络标志物确定方法及系统	中国	CN109101783B	2020-09-04	3973875	温州大学	杜玉改、刘文斌	有效专利
计算机软件著作权	家族基因型和表型数据仿真软件[简称:CPHIBD]	中国	2015SR047934	2015-01-25		中国石油大学(华东)		其他有效的知识产权

计算机软件著作权	基于条件概率的单体分型和多喂点连锁分析的IBD 计算软件[简称: PGP]	中国	2015SR047928	2015-01-25	0935014	中国石油大学(华东)		其他有效的知识产权
其他	TagSNP-set selection for genotyping using integrated data	中国	Future Generation Computer Systems	2021-02	DOI: 10.1016/j.future.2020.09.007	中国石油大学(华东)计算机科学与技术学院, 山东科技大学数学与系统科学学院, 青岛理工大学信息与控制工程学院, 中国石油大学(华东)地球科学与技术学院	Shudong Wang, Gao wei Liu, Xinze ng Wang, Yuan yuan Zhang, Sic heng He, Yulin Zhang	其他有效的知识产权
其他	Tagging SNP-set selection with maximum information based on linkage disequilibrium structure in genome-wide association studies	中国	Bioinformatics	2017-03-16	DOI: 10.1093/bioinformatics/btx151	中国石油大学(华东)计算机科学与技术学院	Shudong Wang, Sich eng He, Fayou Yuan, Xinji e Zhu	其他有效的知识产权

九、主要完成单位及创新推广贡献

中国石油大学(华东):

- (1)作为本项目的牵头单位, 并为项目研究提供技术支持;
- (2)参与制定了项目研究路线、研究计划、研究方案及项目鉴定;
- (3)负责等位基因推断关键技术和单体型格局生成关键技术的研究, 对创新点 1、2、3 有贡献;

(4)发表学术论文 5 篇，其中 SCI 收录 5 篇；

(5)授权发明专利 1 项，公开发明专利 1 项；

(6)获得科技奖励银奖 1 项；

(7)协助将研究成果进行应用推广，2017 年起协助项目由济南科虹生物技术有限公司、济南恒山生物工程有限公司、山东新岳腾电子科技有限公司等多家企业，分别在济南市、济宁市等进行了试应用，具有非常大的推广价值。同时，还协助该项目的技术研究成果在山东省立第三医院、青岛市市立医院进行推广应用。

青岛理工大学：

(1)参与项目研究工作，并为项目研究提供技术支持；

(2)负责标签 SNP 集选择、肿瘤纯度估计和多组学数据中挖掘肿瘤标记物的研究，对创新点 2 和 3 有贡献；

(3)发表与本项目相关的学术论文 3 篇；

(4)协助将研究成果在青岛市立医院和山东省立第三医院推广使用，取得了一定的经济效益。

山东科技大学：

(1)参与项目研究工作，并为项目研究提供技术支持；

(2)参与制定了项目研究路线、研究计划、研究方案及项目鉴定；

(3)负责等位基因推断关键技术和单体型格局生成关键技术的研究，对创新点 1 和 2 有贡献

(4)发表学术论文 2 篇，其中 SCI 收录 2 篇，参加国内学术会议 4 次；

(5)协助将研究成果进行应用推广，2017 年起协助项目由济南科虹生物技术有限公司、济南恒山生物工程有限公司、山东新岳腾电子科技有限公司等多家企业，分别在济南市、济宁市等进行了试应用，具有非常大的推广价值。同时，还协助该项目的技术研究成果在山东省立第三医院、青岛市市立医院进行推广应用。

广州大学：

(1)参与项目研究工作，并为项目研究提供技术支持；

(2)负责癌症网络标志物确定方法及系统、基于 PAGIS 识别基因通路的方法、基于 PADOG 识别基因通路的方法、基于基因通路识别小分子核糖核酸的方法的研究，对创新点 3 有贡献；

(3)授权与本项目相关的发明专利 2 项；

(4)协助将研究成果在青岛市立医院和山东省立第三医院推广使用，取得了良好的社会效益。

安徽大学：

(1)参与项目研究工作，并为项目研究提供技术支持；

(2)负责基于边驱动的标签传播算法的蛋白质功能模块识别方法的研究，对创新点 3 有贡献；

(3)授权与本项目相关的发明专利 1 项；

(4)协助将研究成果进行应用推广，协助项目分别在济南市、济宁市等进行了试应用。