

| | |
|------|---|
| 年份 | 2019 |
| 推荐奖种 | 医学科学技术奖 |
| 项目名称 | 恶性肿瘤风险非编码 RNA 的调控机制研究及大数据平台应用 |
| 推荐单位 | <p>推荐单位：哈尔滨医科大学</p> <p>推荐意见：</p> <p>恶性肿瘤已经成为严重危害人类生命健康的一类重大疾病。广泛存在的非编码 RNA 在肿瘤的发生发展和耐药形成过程中发挥重要的作用，解析非编码 RNA 在恶性肿瘤中的生物学功能及调控机制，对于探索癌症发病机理、发现癌症诊断与预后标记、发现药靶及药物作用通路具有重要意义。李霞教授研究团队近十年来一直致力于从非编码 RNA 的视角，解析肿瘤的发病机制、分子调控网络及药物靶点识别等，构建了大规模的癌症 ceRNA 竞争性调控网络，阐明了恶性肿瘤中 DNA 甲基化调控非编码 RNA 的作用模式，识别了大量非编码 RNA 区域中的恶性肿瘤风险非编码 RNA。除了功能机制研究，课题组还基于遗传谱、转录谱等多组学数据开发了恶性肿瘤精准诊断和预后分析的大数据平台，对预测癌症患病风险，提高诊断及预后预测准确率，实施个性化医疗等具有重大经济价值和社会效益。本项目发表代表性 SCI 论文 20 篇，其中影响因子>10 的 SCI 论文 7 篇，1 篇论文被评为 ESI 高被引论文，Google Scholar 总引用次数超过 700 次，Web of Science 总引用次数超过 400 次，获得软件著作权一项，编写国家级规划教材两部，开发的大数据在线平台已经得到超过 60 个国家的 10 万余次访问，培养了大批高级生物信息学人才。该项目按照要求进行了公示，且公示结果无异议，特推荐该项目申报。</p> |
| 项目简介 | <p>恶性肿瘤目前已经成为严重危害人类生命健康、制约社会经济发展的一类重大疾病。中国的恶性肿瘤发病率和死亡率一直居高不下，从 2010 年开始已经成为主要的致死原因。非编码 RNA（主要为 miRNA 和 lncRNA）在恶性肿瘤中发挥着重要的生物学作用，阐明非编码 RNA 在恶性肿瘤中的功能和调控机制一直是生物医学界研究的热点和重大挑战。课题组以识别恶性肿瘤风险非编码 RNA 并解析其在肿瘤进程中的功能和调控机制为研究目标，在 973 课题、863 课题、国家自然科学基金等 20 余项国家级和省部级课题的支持下，发表代表性 SCI 论文 20 篇，其中影响因子>10 分 SCI 文章 7 篇，累计影响因子 121.753，他引次数 457 次，单篇最高引用 95 次，包括 1 篇 ESI 高被引论文，获软件著作权 1 项，主编国家级规划教材 2 部和 1 本英文学术专著，获得的以下 4 项重要研究成果：1、开发了一套系统的方法全面识别癌症中具有调控活性的 lncRNA-miRNA-mRNA 竞争三元组。课题组在国际上首次：①构建了 12 种癌症的全局竞争活性谱，发现 lncRNA 的竞争活性具有高度癌症特异性；②识别了 9 种癌症的特异性竞争子网及风险通路；③发现一个由五个调控子组成的模块（H19、miR-17、miR-20、THBS1、MAP3K12），可以作为乳腺癌诊断风险分子标记物。2、揭示了 lncRNA 在恶性肿瘤中的 DNA 甲基化调控模式。课题组在国际上首次：①开发了一种重注释策略将来自 Infinium DNA 甲基化芯片的探针重新注释到 lncRNA 功能位点；②构建了 20 种不同癌症、4629 个肿瘤样本及 705 个正常样本的 DNA 甲基化谱；③识别出具有癌症状态特异、癌症亚型特异及预后相关甲基化模式的 lncRNA，可以作为潜在的癌症治疗靶点。3、系统分析了人类 lncRNA 上的 SNP 分布特征。课题组在国际上率先：①发现人类 lncRNA 具有较低的 SNP 密度，揭示风险 SNP 的致病机制与 lncRNA 功能有关；②发现 lncRNA 的进化保守区域上存在疾病风险 SNP；③发现 lncRNA 上的 SNP 对 RNA 二级结构具有显著的影响，疾病风险 SNP 可以通过改变 lncRNA 的二级结构导致疾病。4、课题组在恶性肿瘤相关非编码 RNA 研究领域，开发了多个具有国际影响力的大数据平台：① LincSNP 及其升级版 LincSNP2.0 数据库是国际上第一个 lncRNA 疾病风险 SNP 的数据库；②Lnc2Cancer 是国际上最大的癌症风险 lncRNA 数据库；③ Lnc2Meth 是国际上最全面的研究 lncRNA 上 DNA 甲基化模式的数据库；④LncACTdb 和 miRSponge 是研究 ceRNA 的重要数据库。这些数据库得到了超过 60 个国家的 10 万次访问，获得了 Nature Reviews Genetics 等杂志的高度评价。综上，课题组在恶性肿瘤非编码 RNA 相关的 ceRNA 调控、DNA 甲基化调控和</p> |

| | |
|--|--|
| | 风险 SNP 等方向完成了一系列高水平的研究工作，并建立了多个国际领先的大数据平台，为国内外科研人员提供便捷可靠的数据服务的同时，也极大的提高了我国在癌症非编码 RNA 研究领域的国际影响力。 |
|--|--|

知识产权证明目录

| 序号 | 类别 | 国别 | 授权号 | 授权时间 | 知识产权具体名称 | 发明人 |
|----|----------------|----|------------------|--------------------|--|------------|
| 1 | 中国计算机软件 著作权 | 中国 | 2017SR5041 69 | 2016 -01- 01 | 癌症风险 lncRNA 数据 平台[简称 Lnc2Cancer]V1.0 | 李霞，宁尚 伟 |

代表性论文目录

| 序号 | 论文名称 | 刊名 | 年, 卷 (期) 及页 码 | 影响 因子 | 通讯作 者(含共 同) | SCI 他引 次数 | 他引 总次 数 | 通讯作 者单位 是否含 国外单 位 |
|----|---|------------------------|--|------------|-------------------|-----------------|---------------|-------------------------------|
| 1 | Identification of lncRNA-associated competing triplets reveals global patterns and prognostic markers for cancer | NUCLEIC ACIDS RESEARCH | 2015, 43(7) :3478 -3489 | 11.5 61 | 郭政, 李霞 | 72 | 86 | 否 |
| 2 | Lnc2Cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers | NUCLEIC ACIDS RESEARCH | 2016, 44(D1)):D98 0-D98 5 | 11.5 61 | 王丽华, 李霞 | 78 | 95 | 否 |
| 3 | A novel reannotation strategy for dissecting DNA methylation patterns of human long intergenic non-coding RNAs in | NUCLEIC ACIDS RESEARCH | 2014, 42(13)):825 8-827 0 | 11.5 61 | 李霞 | 18 | 22 | 否 |

| | | | | | | | | |
|---|--|------------------------|-----------------------|--------|--------------|----|----|---|
| | cancers | | | | | | | |
| 4 | Lnc2Meth: a manually curated database of regulatory relationships between long non-coding RNAs and DNA methylation associated with human disease | NUCLEIC ACIDS RESEARCH | 2017, 46(D13-D138) | 11.561 | 宁尚伟, 金连弘, 李霞 | 1 | 1 | 否 |
| 5 | Identifying core gene modules in glioblastoma based on multilayer factor-mediated dysfunctional regulatory networks through integrating multi-dimensional genomic data | NUCLEIC ACIDS RESEARCH | 2015, 43(4):1997-2007 | 11.561 | 肖云, 李霞 | 9 | 11 | 否 |
| 6 | LincSNP 2.0: an updated database for linking disease-associate | NUCLEIC ACIDS RESEARCH | 2017, 45(D1):D74 | 11.561 | 李霞 | 10 | 12 | 否 |

| | | | | | | | | |
|----|--|---|-------------------------|--------|--------------|----|----|---|
| | d SNPs to human long non-coding RNAs and their TFBSs | | -D78 | | | | | |
| 7 | MSDD: a manually curated database of experimentally supported associations among miRNAs, SNPs and human diseases | NUCLEIC ACIDS RESEARCH | 2017, 46 (D1):D181-D185 | 11.561 | 张云鹏, 宁尚伟, 李霞 | 1 | 1 | 否 |
| 8 | miRSponge: a manually curated database for experimentally supported miRNA sponges and ceRNAs | DATABASE-THE JOURNAL OF BIOLOGICAL DATABASES AND CURATION | 2015, bav098 | 3.978 | 宁尚伟, 李霞 | 24 | 28 | 否 |
| 9 | A global map for dissecting phenotypic variants in human lincRNAs | EUROPEAN JOURNAL OF HUMAN GENETICS | 2013, 21 (10):1128-1133 | 3.636 | 李霞 | 10 | 10 | 否 |
| 10 | Co-LncRNA: investigating the lncRNA | DATABASE-THE JOURNAL OF | 2015, bav08 | 3.978 | 李永生, 徐娟, 李 | 26 | 28 | 否 |

| | | | | | | | | |
|----|--|-----------------------------------|------------------------|-------|---------|----|----|---|
| | combinatorial effects in GO annotations and KEGG pathways based on human RNA-Seq data | BIOLOGICAL DATABASES AND CURATION | 2 | | 霞 | | | |
| 11 | SNP@lincTFBS: An Integrated Database of Polymorphisms in Human LincRNA Transcription Factor Binding Sites | PLOS ONE | 2014, 9(7): e103851 | 2.766 | 王丽华, 李霞 | 10 | 12 | 否 |
| 12 | Integrating analysis reveals microRNA-mediated pathway crosstalk among Crohn's disease, ulcerative colitis and colorectal cancer | MOLECULAR BIOSYSTEMS | 2014, 10(9): 2317-2328 | 2.759 | 徐娟, 李霞 | 17 | 17 | 否 |
| 13 | Identification of a Core miRNA-Pathway Regulatory Network in Glioma | PLOS ONE | 2014, 9(7): e101903 | 2.766 | 康春生, 李霞 | 7 | 8 | 否 |

| | | | | | | | | |
|----|---|---------------------------|--------------------------------|-----------|----|----|----|---|
| | by Therapeutically Targeting miR-181d, miR-21, miR-23b, beta-Catenin, CBP, and STAT3 | | | | | | | |
| 14 | In silico detection and characteristics of novel microRNA genes in the Equus caballus genome using an integrated ab initio and comparative genomic approach | GENOMICS | 2009, 94(2) :125- 131 | 2.91 | 李霞 | 35 | 38 | 否 |
| 15 | LincSNP: a database of linking disease-associate d SNPs to human large intergenic non-coding RNAs | BMC BIOINFORM ATICS | 2014, 15, 52 | 2.21 3 | 李霞 | 26 | 29 | 否 |
| 16 | mirTarPri: Improved Prioritization of | PLOS ONE | 2013, 8(1): e5368 | 2.76 6 | 李霞 | 7 | 8 | 否 |

| | | | | | | | | |
|----|--|-------------------------|---------------------------------|-----------|------------|----|----|---|
| | MicroRNA Targets through Incorporation of Functional Genomics Data | | 5 | | | | | |
| 17 | Identifying the crosstalk of dysfunctional pathways mediated by lncRNAs in breast cancer subtypes | MOLECULAR BIOSYSTEMS | 2016, 12(3) :711- 720 | 2.75 9 | 肖云,李 霞 | 8 | 8 | 否 |
| 18 | Systematic analysis of human microRNA divergence based on evolutionary emergence | FEBS LETTERS | 2011, 585(1):240 -248 | 2.99 9 | 李霞 | 11 | 13 | 否 |
| 19 | Integrating Multi-Omics for Uncovering the Architecture of Cross-Talking Pathways in Breast Cancer | PLOS ONE | 2014, 9(8): e1042 82 | 2.76 6 | 田家玮, 李霞 | 10 | 13 | 否 |
| 20 | Genome-wide analysis of | FEBS | 2011, 278(6 | 4.53 | 李霞 | 17 | 17 | 否 |

| | | | | | | | | |
|--|--|---------|---------------|--|--|--|--|--|
| | clustering patterns and flanking characteristics for plant microRNA genes | JOURNAL |):929 -940 | | | | | |
|--|--|---------|---------------|--|--|--|--|--|

主要完成人和主要完成单位情况

| | |
|---------------|---|
| 主 要 完 成 人 情 况 | <p>姓名：李霞</p> <p>排名：1</p> <p>职称：教授</p> <p>行政职务：院长</p> <p>工作单位：哈尔滨医科大学</p> <p>对本项目的贡献：李霞，院长，博士生导师，龙江学者特聘教授，负责项目整体设计和实施。李霞教授课题组长期从事癌症非编码 RNA 相关的生物信息学和系统生物学研究，主持多项 863 项目、973 项目和国家自然科学基金项目，本项目中发表代表性 SCI 论文 20 篇。主要学术贡献包括：1) 建立了癌症 ceRNA 研究体系；2) 建立了癌症非编码 RNA 的 DNA 甲基化和遗传图谱；3) 开发了癌症 lncRNA 风险标记物识别方法；4) 识别了癌症 lncRNA 参与的风险通路。</p> |
| | <p>姓名：宁尚伟</p> <p>排名：2</p> <p>职称：副教授</p> <p>行政职务：大数据中心副主任</p> <p>工作单位：哈尔滨医科大学</p> <p>对本项目的贡献：宁尚伟，副教授，博士、临床医学博士后，硕士生导师，担任课题组组长 (PI)，在本项目中负责具体实施工作，共发表代表性 SCI 论文 10 篇。主要学术贡献包括：1) 建立了癌症 lncRNA 的分子图谱数据库；2) 开发了检测癌症患者风险 lncSNPs 数据库；3) 开发了基于多组学数据的癌症风险标记物识别算法。</p> |
| | <p>姓名：张云鹏</p> <p>排名：3</p> <p>职称：副教授</p> <p>行政职务：计算生物学教研室主任</p> <p>工作单位：哈尔滨医科大学</p> <p>对本项目的贡献：张云鹏，副教授，作为项目主要参加人，进行项目具体实施，对本项目主要学术贡献：1) 建立了癌症风险 lncRNA 大数据平台和遗传分子图谱数据库；2) 开发了检测癌症风险 ceRNA 的系列丰富；3) 挖掘和探索可作为癌症治疗靶点的非编码 RNA 调控风险通路。</p> |
| | <p>姓名：王鹏</p> <p>排名：4</p> <p>职称：副教授</p> <p>行政职务：计算生物学教研室副主任</p> <p>工作单位：哈尔滨医科大学</p> <p>对本项目的贡献：王鹏，副教授，作为项目参与人，负责项目的方法实施，发表论文 8 余篇。主要学术贡献：1) 建立了癌症非编码 RNA 注释数据库；2) 开发了检测癌症 ceRNA 的系列方法；3) 开发了基于甲基化重注释的癌症风险 lncRNA 识别</p> |

| | |
|----------|---|
| | <p>方法。</p> <p>姓名：智慧</p> <p>排名：5</p> <p>职称：副教授</p> <p>行政职务：教研室秘书</p> <p>工作单位：哈尔滨医科大学</p> <p>对本项目的贡献：智慧，副教授，长期从事复杂疾病系统生物学研究，作为项目参与人员，负责项目中数据收集、算法实现及生物学验证。主要学术贡献：1) 建立了癌症风险 lncRNA 的 DNA 甲基化数据库；2) 开发了癌症风险 lncRNA 的识别方法；3) 开发了基于甲基化重注释的癌症风险 lncRNA 标记物识别算法。</p> |
| 主要完成单位情况 | <p>单位名称：哈尔滨医科大学</p> <p>排名：1</p> <p>对本项目的贡献：哈尔滨医科大学作为项目的依托单位，完成了本项目的整体设计、实施和全部研究工作。主要包括：</p> <p>1、开发了一套系统的方法全面识别癌症中具有调控活性的 lncRNA-miRNA-mRNA 竞争三元组，发现一个由五个调控子组成的模块可以作为乳腺癌诊断风险分子标记物。</p> <p>2、揭示了 lncRNA 在恶性肿瘤中的 DNA 甲基化调控模式，识别出具有癌症状态特异、癌症亚型特异及预后相关甲基化模式的 lncRNA，可以作为潜在的癌症治疗靶点。</p> <p>3、系统分析了人类 lncRNA 上的 SNP 分布特征，发现 lncRNA 的进化保守区域上存在疾病风险 SNP，疾病风险 SNP 可以通过改变 lncRNA 的二级结构导致疾病。</p> <p>4、课题组在恶性肿瘤相关非编码 RNA 研究领域，开发了多个具体国际影响力的大数据平台，这些数据库得到了超过 60 个国家的 10 万次访问，获得了 Nature Reviews Genetics 等杂志的高度评价。本研究得到了 973 课题、863 课题、国家自然科学基金等 20 余项国家级和省部级课题的大力支持，发表代表性 SCI 论文 20 篇，其中影响因子 > 10 分 SCI 文章 7 篇，累计影响因子 121.753，他引次数 457 次，单篇最高引用 95 次，包括 1 篇 ESI 高被引论文，获软件著作权 1 项，主编国家级规划教材 2 部和 1 本英文学术专著。</p> |

