

浙江省生物信息学学会 Bioinformatics Society of Zhejiang Province

## 第八届学术年会

## 暨农业生物信息学专业委员会成立大会

暨第九届全国生物信息学与系统生物学学术大会——杭州卫星会议



## 2020年10月23-25日

浙江 临安 主办单位:浙江省生物信息学学会 承办单位:浙江农林大学农业与食品科学学院

承办单位介绍		•••••		•••••1
会议简介				3
会议组织信息				4
会议日程				5
会议费用		• • • • • • • • • • • • • • • • • • • •	•••••	9
交通出行		• • • • • • • • • • • • • • • • • • • •	•••••	10
报告人简介 …		• • • • • • • • • • • • • • • • • • • •	•••••	11
论文摘要		• • • • • • • • • • • • • • • • • • • •	•••••	21
"抗疫先锋"	事迹			68

### 承办单位介绍

#### 浙江农林大学

浙江农林大学位于杭州城西科创大走廊的西端,是国家生态文明教育基地, "一个读书做学问的好地方"。学校创建于1958年,是一所以农林、生物、环 境学科为特色,涵盖八大学科门类的多科性大学,为浙江省人民政府与国家林 业和草原局共建高校、浙江省重点建设高校。

学校拥有东湖等 3 个校区,占地面积 3200 余亩。设有 16 个学院(部)和1 个独立学院,66 个本科专业。现有省重点建设高校优势特色学科 1 个,省一流 学科 12 个;拥有林学一级学科博士后科研流动站,一级学科博士学位授权点 5 个,服务国家特殊需求博士人才培养项目 1 个;一级学科硕士学位授权点 16 个, 专业学位硕士授权类别 8 个。现有各类在校生 34300 余人,其中研究生 2300 余 人,留学生 1400 余人。

学校现有教职工 2360 余人,其中专任教师 1360 余人,高级职称 718 人。 拥有中国工程院院士 1 人、浙江省特级专家 2 人、各类国家级人才 31 人,省部 级人才 107 人。建有国家重点实验室等国家级平台 5 个,省部级创新平台 42 个。 获国家技术发明二等奖 1 项、国家科技进步二等奖 6 项、浙江省科学技术一等 奖 10 项、省哲学社会科学优秀成果奖一等奖 3 项。

先后与浙江、福建、贵州等 50 多个县市区开展多种形式的科技推广与社会服务。学校被评为全国科技特派员工作先进集体、国家级科技特派员创业培训基地、全国社会扶贫先进集体。与英国、美国、俄罗斯等 30 余个国家(地区)的 100 多所大学和科研机构,开展学术交流、联合办学、科学研究等活动。与塞尔维亚诺维萨德大学共建孔子学院,与加拿大大不列颠哥伦比亚大学、芬兰赫尔辛基大学等联合培养人才,国际知名度不断提高。





#### 农业与食品科学学院

农业与食品科学学院创建于 2007 年,设有作物学、园艺学、植物保护、食品科学与工程四个一级学科。其中,作物学和园艺学为省级一流学科(B类)和学校"高原学科"。

建有一支由院士领衔的师资队伍。现有教职工 143 人,其中专任教师 117 人,具有博士学位教师 113 人(96.6%),其中中国工程院院士 1 人、国家青年 专项人才等国家级人才 3 人、农业部"农业科研杰出人才及其创新团队"1 个、浙 江省钱江特聘教授、农业部现代农业产业技术体系岗位科学家等各类省级人才 28 人。

现有作物学、园艺学、食品科学与工程3个学术型硕士点,农艺与种业(作物、园艺、种业)、资源利用与植物保护(植物保护)、食品加工与安全3个专业学位硕士点。拥有园艺、农学、茶学、植物保护、食品科学与工程(粮油储检)、食品质量与安全(农产品加工与质量检测技术)3个本科专业方向。其中,园艺专业为国家级一流本科专业,农学专业为省级一流本科专业,农学、园艺、植物保护、食品科学与工程为教育部卓越农林人才教育培养计划试点专业,农学、园艺、植物保护、食品质量与安全为浙江省基层农技人员定向培养专业(事业编制),食品科学与工程(粮油储检)为浙江省粮食系统粮油储检人员定向培养专业(国企编制)。

学院高度重视人才培养工作。近3年,本科生获得各类科研创新训练项目 资助100余项;学生第一作者发表各类学术论文150余篇,其中SCI论文45篇, 一级期刊32篇;荣获"挑战杯"等各类省部级奖项60余项;本科毕业生就业率达 97%以上,高质量就业率达60%以上。学院设有"拜耳"、"勿忘农"、"萧农"等企 业奖学金。

学院建有浙江省农产品品质改良技术研究重点实验室、山区农业高效绿色 生产"2011 协同创新中心"、农业部农产品加工技术研发中心、植物生产类省级 重点实验教学示范中心等省级创新平台 4 个。实验室总面积 6000 余平方米,仪 器设备总值 5000 多万元。近 3 年,获批立项各级各类科研项目 356 项,其中国 家级项目 32 项,研究经费达 6400 余万元;发表学术论文 362 篇,其中被 SCI、 EI、ISTP 收录 129 多篇;授权专利 35 项。

学院积极开展社会服务和对外学术交流,近年来共派出省级科技特派员 141 人次,其中有 13 人次被授予"浙江省功勋科技特派员"、"浙江省优秀科技特派员" 荣誉称号;与美国、德国、英国等 18 个国家和地区的高校或科研机构建立了广 泛的学术交流和研究合作关系。

2

## 会议简介

生物信息学在我国起步于 20 世纪 90 年代初,是一门整合数学、物理、计 算机等理工学科和生命科学、医学、药学等学科的交叉学科,主要是通过结合 计算方法、高通量组学技术、生物学实验进行生物及医学问题的研究,目前生 物信息学已经由新兴学科成为一门成熟的尖端交叉学科。随着人工智能与大数 据时代的到来,以及基因组、转录组、蛋白质组、代谢组、表型组等各组学生 物信息技术的发展,精准农业、智慧育种,已成为农业生物信息领域的重要发 展趋势。如何发挥生物信息学与人工智能的优势特点,开展高水平的科学研究 与应用"是本次会议的核心议题。与会专家将结合自己的研究工作做大会报告, 通过大会报告与交流讨论,与会专家和学会会员从生物信息学的研究重点、生 物信息学与农学生态的交叉与应用等领域积极阐述自己的观点、想法与建议。

## 会议组织信息

大会主席: 陈铭教授

组委会主席:黄坚钦教授、刘庆坡教授

组织及程序委员会(按拼音顺序):包家立、陈欢、陈蔚青、代琦、邓祖 跃、丁维龙、董长征、樊龙江、甘卓慧、郭俊明、贺平安、黄方亮、黄晶、黄 有军、吉佔焘、姜俊平、郎秋蕾、李飞、李小波、廖奇、刘鹏渊、刘庆坡、刘 文斌、刘箴、罗洁、毛凌峰、邵朝纲、沈波、苏建忠、王卫凯、向太和、辛德 东、徐程、徐建红、姚玉华、颜成钢、叶子弘、张贵军、赵怀、钟伯雄、周猛、 朱峰

协办与支持单位:中国生物信息学学会(筹)、浙江大学生命科学学院、 浙江省疾病预防控制中心、江苏省生物信息学专业委员会、东南大学、江苏省 现代作物生产协同创新中心、中国科学院遗传与发育生物学研究所、内蒙古民 族大学、

企业支持单位:杭州遂真生物技术有限公司、杭州宝诚生物技术有限公司、 浙江天科生物技术股份有限公司、曙光信息产业股份有限公司浙江分公司、杭 州厚泽生物科技有限公司、杭州联川生物技术股份有限公司、杭州维讯生物科 技有限公司

4

## 会议日程

10月23日

时间	内 容	地点
15:00-18:00	报 到	酒店大厅
17:30-19:00	晚餐	六和厅
19:00-20:30	理事会会议	武肃阁

#### 10月24日

时间		内 容	地点
7:30-8:30		报 到	中都厅
上午	8:30-8:40	开幕式 <ol> <li>中国生物信息学学会(筹)理事长孙之荣致辞</li> <li>省市科协领导致辞</li> <li>浙江农林大学副校长吴家胜致辞</li> </ol>	中都斤 A 斤 主持人: 刘庆坡、廖奇
	8:40-9:10	<ul> <li>4. 浙江省生物信息学学会理事长陈铭致辞</li> <li>抗疫先锋业绩简介并颁发荣誉证书</li> <li>抗疫先锋报告:张严峻浙江省疾病预防控制中心</li> <li>报告题目:新冠病毒及其疫苗</li> </ul>	
	9:10-9:40	特邀报告:杨剑 西湖大学 报告题目:基因组变异与健康	
	9:40-10:00	集体合影 & 茶 歇	酒店北门外台阶
	10:00-10:45	全国生物信息学云论坛(第九讲)	江苏省

时间		内 容	地点
		大会报告:肖力宏 浙江农林大学	
	10:45-11:10	报告题目: Complete plastomes enhance the	
		understanding of evolution and phylogeography of	
		Eastern Asia - Eastern North American disjunctive	
		Carya (Juglandaceae)	
上		大会报告:马云龙 温州医科大学	中都厅A厅
午	11 • 10-11 • 35	报告题目: Integrative Genomics Analysis Reveals	主持人: 郭俊明
	11.10 11.55	a Novel 21q22. 11 Locus Contributing to	
		usceptibility of COVID-19	
	11:35-11:55	企业报告:毛凌峰 浙江宝诚生物技术有限公司	
		报告题目:纳米孔(第四代)测序技术在农业科学	
		研究中的应用进展	
山	12:00-13:30	午 餐	大堂西餐厅
- 午	12:00-13:00	农林生物信息学专业委员会成立大会	计正在该门
1		健康大数据与转化医学专业委员会筹备会议	以以向
	13:30-14:00	特邀报告:张忠华 青岛农业大学(在线报告)	
下 午		报告题目: Comparative genomics provide novel	
		insights into the evolution and genetic basis of	中御月 A 月 主持人
		agronomic traits in cucurbits	土时八: 茜右军 向大和
		大会报告:沈星星 浙江大学	央 1 千 1 円 八 仲
	14.00-14.20	报告题目:基因组数据能否更好地解析生物多样性	

时间		内 容	地点
		大会报告:刘庆坡 浙江农林大学	
	14:20-14:40	报告题目: Deleterious Variants and the "Cost of	中都厅A厅
		Domestication" in Plants	主持人:
	14.40-15.00	大会报告:张亮生 浙江大学	黄有军、向太和
	14.40 15.00	报告题目:花卉基因组及其应用	
	15:00-15:20	茶歇	
		大会报告:张文艺 西湖大学	
	15:20-15:40	报告题目: EDock: Blind Protein-ligand Docking by	
下		Replica-Exchange Monte Carlo Simulation	
午	15:40-16:00	大会报告:叶楚玉 浙江大学	
		报告题目: 孤儿作物稗的基因组学研究进展	山却臣▲臣
	16:00-16:20	大会报告:苏建忠 温州医科大学	中 仰月 A 月 → 桂 人
		报告题目: Grand H3 lysine 27 trimethylation domains	土付八:
		establish super-silencers involved in oncogene activity	<b>陈</b> 从、
		in cancer	
	16:20-16:40	企业报告:赵怀 杭州遂真生物技术有限公司	
		报告题目: AIGS 全自动核酸检测分析系统在新冠疫情	
		中的应用	

时间		内容	地点
下午	16:40-17:00		中都厅 A 厅
		优秀论文答辩	裁判长:钟伯雄
			裁判:所有理事
	17:00-17:30	闭幕式	
		1. "遂真杯"优秀研究生论文颁奖仪式	中都厅 A 厅
		2. 农林生物信息学专委会成立授牌仪式	主持人:徐 程
		3. 宣布下一届年会主办单位并介绍	
晚	17.20	昭 殇	シ和臣
上	17.30	· · · · · · · · · · · · · · · · · · ·	/ \ / \ / \ ]

10月25日

8:30     酒店大       上     9:00-10:30     参观       午     10:30-11:00     学会企	020	<b>洒亡土已在人</b>	前往浙江
	· 酒店入厅 集合	农林大学	
	9:00-10:30	参观考察	浙江大井十兴
	10:30-11:00	学会企业座谈会	浙江巜怀人子
	11:00—	返程	

## 会议费用

缴费时间请预注册	学生会员	其他会员	非会员
10月15日前	200 元	500 元	800 元
10月15日后及现场	500 元	800 元	1000 元

企业会员:本学会企业会员单位免费拥有1个展位(参会人员以学会其他会 员身份注册)及2名人员免注册费,会议手册A4宣传页,会议网站宣传。

**账户信息:**浙江省生物信息学学会 开户行:交通银行西湖支行 **账户号:** 331066130018170122500 汇款用途注明: 注册费 + 姓名单位 会场酒店: 临安青山湖中都大酒店

#### 会务组联系方式:

总协调人:	刘庆坡 18906513960	E-mail: liuqp@zafu.edu.cn
学会总务:	沈小仙 13666606799	E-mail: xxshen@zju.edu.cn
理事会议:	徐程 18057145588	E-mail: xuc123@zju.edu.cn
专委会筹:	刘庆坡 18906513960	E-mail: liuqp@zafu.edu.cn
接待安排 <b>:</b>	董辉 18968140682	E-mail: 107086708@qq.com
论文、注册	· 霍颖异 1381912893	9; 何磊 15268821020
会议网址:	http://www.zjbioinformatics	s.org/meetings/2020/



## 交通出行

方案一

自驾导航:杭州市临安区中都青山湖大酒店(浙江省杭州市临安区锦城街 道圣园路 88 号)

方案二

#### 杭州到临安农林大学站

高铁杭州东站:地铁1号线—地铁5 号线—地铁16号线(1小时34分钟, 票价10元)



杭州城站:地铁5号线—地铁16号线 (1小时36分钟,票价10元)



#### 临安农林大学站到中都青山湖大酒店

地铁农林大学站 B 出口打的约 16 元到中都青山湖大酒店 公交车 807 路到中都青山湖大酒店





简

介

#### 报告人简介: 张严峻 研究员 浙江省疾病预防控制中心微生物所所长

张严峻,研究员、主任技师,硕士导师,浙江省疾 病预防控制中心微生物所所长。现为浙江省"151"第二层 次人才、浙江省卫生高层次创新人才,浙江省疾控中心 高级专家,浙江省预防医学会出生缺陷防治委员会副主 委,浙江省预防医学会卫生检验专业委员会常务委员, 浙江省微生物学会理事,浙江省医学会医学病毒学分会



委员会委员,浙江省预防医学会食品卫生专业委员会委员。主持了国家重大传 染病专项子课题、浙江省重大科技专项、省部共建课题、浙江省创新团队课题 等研究。先后获浙江省科学技术二等奖、浙江省医药卫生科技创新一等奖等 11 项科研成果奖。第一作者或通讯作者发表 SCI 收录论文 10 多篇。作为主审出版 了《病原微生物与免疫学实验及学习指导》,作为副主编出版了《禽流感防控 手册》,参编出版了《手足口病》、《食源性疾病防制与应急处置》等著作。 以第一发明人获得的国家发明专利授权 5 项。

#### 报告人简介:杨剑 西湖大学终身教授

2003 年本科毕业于浙江大学生物科学系,2008 年于浙江大 学取得数量遗传学博士学位,同年赴澳大利亚昆士兰医学研究 所从事博士后研究工作。2012 年加入澳大利亚昆士兰大学担任 研究员,2014 年 1 月被聘为高级研究员和实验室主任。2014 年 12 月晋升为昆士兰大学副教授,2017 年晋升为教授。2020 年加入西湖大学。主要致力于统计遗传学、基因组学研究,以



及人类复杂性状和疾病(如:身高、肥胖和精神分裂)的大数据分析。2012年,杨剑获得澳大利亚百年学院劳伦斯创新奖。2015年获得澳大利亚科学院 Ruth Stephens Gani 奖章。2017年获得澳大利亚总理科学奖-Frank Fenner 年度生命科 学家奖。2018 和 2019 连续两年被列入科睿唯安"被高引科学家"。

#### 报告人简介:肖力宏 浙江农林大学特聘教授

肖力宏,博士,浙江农林大学校特聘教授,省部共建亚 热带森林培育重点实验室资源植物功能基因组学课题组组 长,硕士研究生导师。博士毕业于首都师范大学,师从匡廷 云院士和何奕騉教授。2010年至2014年,在首都师范大学 进行博士后和做研究和美国密苏里大学/USDA-ARS-PGRU 访问学者,2015年至2017年在中科院上海植物逆境研究中



心朱健康课题组担任副研究员,2017 年 8 月入职浙江农林大学林业与生物技术 学院。自2010 年开始,一直从事资源植物基因组和功能基因组学研究,先后参 与复苏植物牛耳草、(美国)山核桃等 10 种资源植物全基因组序列的测序组装 注释、全基因组进化和生物学问题相关分析,并通过各种组学整合研究构建调 控网络。同时以模式藓类植物小立碗藓为材料研究植物耐极端干旱的分子机制 及其在陆生植物中的进化,并建立了小立碗藓多基因 CRISPER 敲除系统。以第 一作者和通讯作者身份在 PNAS、Plant Cell & Environment、GigaScience 等刊物 发表学术论文 7 篇,另有多篇 SCI 论文在撰写之中。目前主要研究方向包括: 特色经济林木坚果品质、营养等相关基因的挖掘与应用,资源植物基因组学与 功能基因组学,植物逆境应答的分子机制及其进化,植物遗传转化及基因编辑, 特色经济林坚果品质营养和极端生境植物逆境应答的表观调控机制。

#### 报告人简介:马云龙 温州医科大学副研究员

马云龙博士,副研究员,温州医科大学引进学术骨干,担任浙江省生物信息学学会青年委员会常委,美国遗传学会会员,及中国计算机学会(CCF)会员。主要从事利用生物信息算法和医学统计方法从遗传学、基因组学和表观遗传学层次研究精神类疾病、癌症等复杂疾病的病因学机制。



目前研究方向:(1)挖掘精神分裂症和抑郁症等共发生疾病 的遗传和表观机制;(2)高通量循环肿瘤 ctDNA 甲基化数据中极小事件推断 算法开发与评估,并系统识别具有临床应用前景的食管癌诊断和预警相关分子 标志物;(3)单细胞转录组解析细胞异质性的计算方法。先后主持和参与国家 自然科学基金、中国博士后科学基金面上项目一等资助、国家重点研发计划"精

准医学研究"等在内的多项基金项目。至今在《Molecular Neurobiology》、

《Translational Psychiatry》、《Psychological Medicine》、《Computational and Structural Biotechnology Journal》、和《Addiction Biology》等国际期刊上累计发 表 31 篇 SCI 收录研究论文(其中第一和通讯作者 16 篇),影响因子 IF >120。 拥有 4 项肠道微生态生信分析的软件著作权及 2 项基于遗传突变位点疾病预测 的发明专利。参与英文书籍《Tobacco Smoking Addiction: Epidemiology, Genetics, Mechanisms, and Treatment》的编写,并参与本书中文版翻译。参加了 50 次以 上包括美国、加拿大等国内外学术会议和技能培训,并多次作大会发言。通过 大数据整合分析,发现多个与戒烟相关的突变位点,能够显著提高吸烟者戒烟 效率;此结果被国内外上百家媒体报道,包括 Nature 新闻部、美国洛杉矶时报 (美国第三大报业)、爱尔兰时报(爱尔兰主流大报)、New Scientist、MSN、 Daily Times、HealthDay News 等都作了特别报道。

15

#### 报告人简介:毛凌峰博士 杭州宝诚事业部经理

毛凌峰,浙江大学大学生物信息学博士,杭州宝诚 ONT 事业部经理,在 PNAS, Plant Biotechnology Journal、Database 等 SCI 杂志以第一作者发表多篇论 文,同时参与著作《生物信息学》(浙大出版社 2017)、《植物基因组学》(科 学出版社 2020)等,目前主要工作为纳米孔测序技术的研发应用及市场推广等。 在本次新冠疫情中,毛凌峰博士,协助杭州市疾控在 2 月 5 日全球第一个仅用 纳米孔测序技术完成新冠全基因组测序并上传 Gisaid 数据库(全球流感基因组 数据库),并带领 ONT 部门协助包括江苏省疾控、江西省疾控、浙江大学传染 病国家重点实验室等多家单位完成新冠的纳米孔全基因组测序和分析。

#### 报告人简介:张亮生 浙江大学教授

张亮生,复旦大学遗传学博士,浙江大学求是特聘教授,博士生导师。2015 年以来以第一作者或通讯作者在 Nature、 Molecular Plant、Plant Physiology、New Phytologist、Plant Journal、 Horticulture Research 等国际著名期刊发表论文 30 篇,影响因 子总和近 200。



#### 报告人简介:张忠华 青岛农业大学教授

青岛农业大学二级教授、博士生导师,长期从事黄瓜、 甜瓜等蔬菜作物功能基因组学研究,开发了农艺性状功能基 因研究的新方法,鉴定了果实风味、大小、性别决定等多个 重要性状的关键基因,并用于多个高产优质新品种选育,加 速了蔬菜作物遗传育种基础研究,为全基因组分子育种提供 了支持。入选国家"万人计划"科技创新领军人才,获第十



四届"中国青年科技奖"和国家自然科学二等奖(第 2)。中国园艺学会分子 育种分会秘书长, Journal of Integrative Agriculture (JIA)、Horticultural Plant Journal 等期刊编委,中央国家机关青联委员,全国青联常委。在 Nature、Science、 Cell、Nature Genetics 等期刊共发表 SCI 论文 44 篇,累计影响因子超过 500, 共被引用 8000 余次,其中(共同)第一或通讯作者在 Nature Genetics (3 篇)、 Nature Communications、PNAS 和 Plant Cell 等期刊发表 13 篇,编写了美国 Science Publishers 出版专著的两个章节。主持国家重点研发计划课题 1 项,2013 年获国家自然科学基金委"优秀青年"科学基金项目,主持国家自然科学基金 面上项目 3 项。

#### 报告人简介:沈星星 浙江大学"百人计划"研究员

2014 年获中山大学生物化学与分子生物学博士。 2014 年至 2019 年在美国范德堡大学 Antonis Rokas 教授 课题组进行博士后训练。目前以真核生物为研究对象, 利用大规模数据集以及综合多个学科手段(如系统发育 学、进化生态学、基因组学),从理论和方法出发,研究 系统发育学的热点和难点问题、探究生物多样性的演化



机制和生态因素。累计发表 SCI 期刊论文 20 篇,总影响因子超 200,以第一作 者身份在 Cell、Nature Ecology & Evolution、Science Advances、Systematic Biology、 Molecular Biology & Evolution 国际权威期刊发表 IF>10 论文 7 篇。曾多次受邀 于国内外高校、会议等做学术研究报告。其 Cell 研究成果被美国能源部办公室 特别报道、Nature Ecology & Evolution 研究成果被评为该期刊 2017 和 2018 年度 最具影响力论文之一。

课题组主页: https://xingxingshen.github.io/

#### 报告人简介:刘庆坡 浙江农林大学教授

刘庆坡,博士,教授,博士生导师,美国加州大学欧文 分校访问学者,为浙江省高校创新领军人才、浙江省高校中 青年学科带头人、浙江农林大学青年拔尖人才。现担任

《Journal of Genetics》的 Associate Editor、浙江省作物学会 副理事长、浙江省生物信息学会理事等。曾荣获浙江省优秀 教师、浙江省"三育人"先进个人、浙江农林大学"我心目



中的好老师"等荣誉。主要从事生物信息与作物逆境生理等方面的教学与科研工作。主持国家自然科学基金面上和省自然科学基金重点项目等课题 10 多项; 获浙江省自然科学学术奖 2 项; 在《Nature Plants》《Molecular Biology and Evolution》等国内外学术刊物上发表论文 54 篇。

#### 报告人简介:张文艺博士 西湖大学

张文艺博士毕业于东北师范大学信息科学学院。博士期 间在美国密歇根大学张阳实验室联合培养三年。主要从事蛋 白质功能位点预测以及蛋白质与配体分子对接算法研究。基 于 I-TASSER 开发了低精度蛋白质结构盲对接算法 EDock。 于 2019 年 10 月加入西湖大学黄晶实验室,担任博士后,从 事计算机辅助药物设计中核心算法的开发以及新型抑菌化合 物的设计研究。



#### 报告人简介: 叶楚玉 浙江大学副教授

博士毕业于北京林业大学,师从尹伟伦院士,曾在美国橡 树岭国家实验室(ORNL)交流学习两年。现为浙江大学作物 科学研究所副教授,作物所副所长、生物信息与大数据技术中 心主任,浙江大学求是青年学者(2014-2018)、仲英青年学者 (2020-2023)。研究方向为作物基因组学与生物信息学,以第 一作者或通讯作者在 Nature Ecology & Evolution、Molecular Plant、Briefings in Bioinformatics 等期刊发表多篇论文。



#### 报告人简介:赵怀博士 杭州遂真生物技术有限公司董事长

赵怀博士 杭州遂真生物技术有限公司董事长 CEO 浙江大学肿瘤学博士 从事生命科学仪器临床诊断试剂开发 20 年,曾任外商独资高科技企业常务副总 经理,首席运营官,研发中心主任等职。曾主持开发荧光定量 PCR、基因芯片 阅读仪等基因检测设备,多款定量基因检测试剂盒,并取得十余项国家二/三类 医疗器械认证,获得多国注册。

论 文 摘 要

## Systematic classification using supervised machine learning: learning from class imbalance

Solomon Shiferaw Beyene, Cong Feng, Yincong Zhou, Tianyi Ling, Blagoj Ristevski, Ming Chen\*

Department of Bioinformatics, College of Life Sciences, Zhejiang University, Hangzhou 310058

\*Correspondence: Ming Chen, E-mail: mchen@zju.edu.cn

Abstract: Supervised machine learning was used to classify regulatory riboswitches. Riboswitches are part of 5' UTR region of regulatory mRNA. It was discovered in several species but mainly in bacteria. They are very important in antibacterial drug design since its discovery in 2002. Classification and database were attempted since couple of years back, but with unsolved problems. These problems are: classification using class imbalance, besides some of previously used algorithms are computational costly, and existing database doesn't include up to dated riboswitches families also lack inclusiveness of different datum. To solve the challenges novel methodology was designed for classification and up-to-dated data collected for user friendly web server named as RibosLD. The new pipeline, includes SMOTE for balancing imbalanced classes was used and a total of 5460 k-mers  $(1 \le k \le 6)$  produced as feature for machine leaning analysis. A total of 156 feature was divided into training and test set, 70% and 30% respectively. Statistical analysis for parameters like accuracy, specificity, sensitivity and macro F-score showed significant differences at (p < 0.05, 0.01, 0.001). The result revealed impact of class imbalance in classification and algorithms performance. K-mers used as feature in machine learning identified to have biological function and even some are riboswitches motifs. Comprehensive riboswitches families were collected to develop riboswitches database hereafter named as RibosLD. The RibosLD contains most updated riboswitches. Overall, the database contains potential up-to-dated 39 riboswitches families' that includes more than thirty-seven thousand classes identified. RibosLD gives user interface with many options like browse, search and structural information.

#### Keywords: balanced, imbalanced, k-mers, machine learning, riboswitch.

**Reference :** Beyene SS, Ling T, Ristevski B, Chen M. (2020) A novel riboswitch classification based on imbalanced sequences achieved by machine learning. PLoS Comput Biol 16(7): e1007760.

#### Expansion of MIR482/2118 by a class-II transposable element in cotton

Enhui Shen<sup>1,2,†</sup>, Tianzi Chen<sup>3,†</sup>, Xintian Zhu<sup>1</sup>, Longjiang Fan<sup>1</sup>, Jie Sun<sup>4</sup>, Danny J. Llewellyn<sup>5</sup>, Iain Wilson<sup>5</sup> and Qian-Hao Zhu<sup>5,\*</sup>

- 1. Institute of Crop Sciences and Institute of Bioinformatics, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, China,
- 2. New Rural Development Institute, Zhejiang University, Hangzhou 310058, China,
- 3. Provincial Key Laboratory of Agrobiology, Institute of Crop Germplasm and Biotechnology, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China,
- 4. Key Laboratory of Oasis Eco-agriculture, College of Agriculture, Shihezi University, Shihezi, Xinjiang 832000, China
- 5. Black Mountain Laboratories, CSIRO Agriculture and Food, GPO Box 1700, Canberra, ACT2601, Australia

Some plant microRNA (miRNA) families contain multiple members generating identical or highly similar mature miRNA variants. Mechanisms underlying the expansion of miRNA families remain elusive, although tandem and/or segmental duplications have been proposed. In this study of two tetraploid cottons, Gossypium hirsutum and Gossypium barbadense, and their extant diploid progenitors, Gossypium arboreum and Gossypium raimondii, we investigated the gain and loss of members of the miR482/2118 superfamily, which modulates the expression of nucleotide-binding site leucine-rich repeat (NBS- LRR) disease resistance genes. We found significant expansion of MIR482/2118d in G. barbadense, G. hirsutum and G. raimondii, but not in G. arboreum. Several newly expanded MIR482/2118d loci have mutated to produce different miR482/2118 variants with altered target-gene specificity. Based on detailed analysis of sequences flanking these MIR482/2118 loci, we found that this expansion of MIR482/2118d and its derivatives resulted from an initial capture of an MIR482/2118d by a class-II DNA transposable element (TE) in G. raimondii prior to the tetraploidization event, followed by transposition to new genomic locations in G. barbadense, G. hirsutum and G. raimondii. The 'GosTE' involved in the capture and proliferation of MIR482/2118d and its derivatives belongs to the PIF/Harbinger

superfamily, generating a 3-bp target site duplication upon insertion at new locations. All orthologous *MIR482/2118* loci in the two diploids were retained in the two tetraploids, but mutation(s) in miR482/2118 were observed across all four species as well as in different cultivars of both *G. barbadense* and *G. hirsutum*, suggesting a dynamic co-evolution of miR482/2118 and its *NBS-LRR* targets. Our results provide fresh insights into the mechanisms contributing to *MIRNA* proliferation and enrich our knowledge on TEs.

# Keywords : miR482/2118, *NBS-LRR*, miRNA evolution, transposable element, gene duplication, cotton, Gossypium spp

Published on the Plant Journal ( $IF_5=6.629$ )

#### PTRN:基于文本挖掘和高通量测序的植物转录调控知识库

吴岩,陈铭\*

浙江大学生命科学学院,杭州,310058

\*E-mail: mchen@zju. edu. cn

转录因子作为一类重要的基因表达调控因子,通过对顺式作用元件的识别 与结合来调控基因的表达,在发育和应激反应中发挥着重要作用。对于植物来 说,重构转录调控网络一直是研究的热点,因此很多技术应运而生,与植物转 录调控有关的文献资源也日益增长。在这里,我们基于文本挖掘和高通量测序 分析建立了目前最为全面的植物转录调控知识库 PTRN,涵盖的主要内容和功能 如下: i)转录因子和基因(编码及非编码)之间的转录调控关系; ii)转录因子的靶 基因模块分析; iii)转录因子协同性分析; iv)转录因子所参与的代谢通路分析。

#### 关键字: 转录因子 靶基因 文本挖掘 高通量测序

#### Computational principles and practice for decoding immune contexture in the tumor microenvironment

Zicheng Zhang, Siqi Bao, Congcong Yan, Ping Hou, Meng Zhou, Jie Sun

School of Biomedical Engineering, Wenzhou Medical University, Wenzhou 325027, P.R. China

Tumor-infiltrating immune cells (TIICs) have been recognized as crucial components of the tumor microenvironment (TME) and induced both beneficial and adverse consequences for tumorigenesis as well as outcome and therapy (particularly immunotherapy). Computer-aided investigation of immune cell components in the TME has become a promising avenue to better understand the interplay between the immune system and tumors. In this study, we presented an overview of data sources, computational methods and software tools, as well as their application in inferring the composition of tumor-infiltrating immune cells in the TME. In parallel, we explored the future perspectives and challenges that may be faced with more accurate quantitative infiltration of immune cells in the future. Together, our study provides a little guide for scientists in the field of clinical and experimental immunology to look for dedicated resources and more competent tools for accelerating the unraveling of tumor-immune interactions with the implication in precision immunotherapy.

#### **Journal: Briefings in Bioinformatics (IF = 8.99)**

# EigenGWAS: An online visualizing and interactive application for detecting genomic signatures of natural selection

Guo-An Qi<sup>1,†</sup>, Yuan-Ting Zheng<sup>1,†</sup>, Feng Lin<sup>2</sup>, Xin Huang<sup>1</sup>, Li-Wen Duan<sup>1</sup>, Yue You<sup>1</sup>, Hailan Liu<sup>3</sup>, Ying Wang<sup>4</sup>, Hai-Ming Xu<sup>1,\*</sup>, Guo-Bo Chen<sup>2,\*</sup>

Institute of Bioinformatics and Institute of Crop Science, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, 310058 Zhejiang, China; Clinical Research Institute, Zhejiang Provincial People's Hospital, People's Hospital of Hangzhou

Medical College, Hangzhou, 310014 Zhejiang, China;

Maize Research Institute, Sichuan Agricultural University, Chengdu, 611130 Sichuan, China; Phase I Clinical Research Center, Zhejiang Provincial People's Hospital, People's Hospital of Hangzhou Medical College, Hangzhou, 310014 Zhejiang, China;

*†* These authors are contributed equally to this work.

\* To whom correspondence should be addressed.

Abstract: Detecting genetic regions under selection in structured populations is of great importance in ecology, evolutionary biology, and breeding programs. We recently proposed EigenGWAS, an *FFssss*-like linear regression approach, for detection of genomic regions under selection. Unlike *FFssss*, EigenGWAS does not require grouping information of the population, and can be considered an unsupervised genomic scanning approach for finding loci under selection. The EigenGWAS analysis identifies significant genome-wide loci under selection after correction for genetic drift. Here, we describe EigenGWAS implementation, including use of the online analysis platform. Our online computational tool accepts pedigree data in a standard binary format that can be easily converted from the original sequencing data. We applied the proposed method in various datasets and observed meaningful and biologically interpretable results.

Keywords: Evolution, Natural selection, Genomic adaptation, EigenGWAS

#### Computational Methods and Applications for Identifying Disease-Associated IncRNAs as Potential Biomarkers and Therapeutic Targets

Congcong Yan,<sup>1</sup> Zicheng Zhang,<sup>1</sup> Siqi Bao,<sup>1</sup> Ping Hou,<sup>1</sup> Meng Zhou,<sup>1</sup> Chongyong Xu,<sup>2</sup> and Jie Sun<sup>1</sup>

1.School of Biomedical Engineering, Wenzhou Medical University, Wenzhou 325027, P.R. China; 2.Department of Radiology, The Second Affiliated Hospital of Wenzhou Medical University, Wenzhou 3.25027, P.R. China

Long non-coding RNAs (lncRNAs) have been recognized as critical components of a broad genomic regulatory network and play pivotal roles in physiological and pathological processes. Identification of disease-associated lncRNAs is becoming increasingly crucial for fundamentally improving our understanding of molecular mechanisms of disease and developing novel biomarkers and therapeutic targets. Considering lower efficiency and higher time and labor cost of biological experiments, computer-aided inference of disease-associated RNAs has become a promising avenue for facilitating the study of lncRNA functions and provides complementary value for experimental studies. In this study, we first summarize data and knowledge resources publicly available for the study of lncRNA-disease associations. Then, we present an updated systematic overview of dozens of computational methods and models for inferring lncRNA-disease associations proposed in recent years. Finally, we explore the perspectives and challenges for further studies. Our study provides a guide for biologists and medical scientists to look for dedicated resources and more competent tools for accelerating the unraveling of disease-associated lncRNAs.

#### Journal: Molecular Therapy - Nucleic Acids (IF=7.032)

#### 剖析多个相关性状遗传结构的联合关联分析方法

林峰<sup>a.d</sup>,齐国安<sup>a</sup>,徐挺<sup>b</sup>,楼向阳<sup>c</sup>,洪永波<sup>d.\*</sup>,徐海明<sup>a.</sup> <sup>a</sup>浙江大学农业与生物技术学院作物科学研究所,杭州,310058

<sup>b</sup>浙江大学数学系,杭州,310058

<sup>c</sup> Department of Biostatistics, Colleges of PHHP and the College of Medicine,

University of Florida, Gainesville, FL 32610, USA

<sup>d</sup>中国水稻研究所水稻生物学国家重点实验室和超级稻研究浙江省重点实验室,杭州, 311401

全基因组关联研究(GWAS)已成为探索复杂性状遗传结构的有力工具。 如何通过多个遗传相关性状的联合分析,提高遗传位点的检测功效,揭示性状 遗传相关及基因间、基因与环境间相互作用的机制,是目前GWAS面临的一个 主要挑战。针对这些挑战,我们提出了一种基于混合线性模型的多性状联合关 联分析方法,该方法的关联模型包含了位点的上位性及其与环境互作效应,分 析方法利用了单个性状方差、成对性状间协方差信息;利用基于 Wilks 统计量转 换所得的 F 统计量检测每个 SNP 和成对互作 SNP 的显著性,然后根据显著的 SNP (QTS)构建全模型,采用马尔科夫链蒙特卡洛方法(MCMC)估计 QTS 的各 项遗传效应并检验其显著性。模拟结果表明,多性状 GWAS 方法可提高多效性 位点(影响多个性状)的检测功效,无偏估计 QTS 的效应。为演示该方法的应 用,我们分析了动脉粥样硬化多种族研究(MESA)样本的四个血脂相关性状以 及水稻永久 F<sub>2</sub>群体的两个产量相关性状。针对提出的方法,我们开发了相应的 分析软件包。

关键词:联合关联分析;混合线性模型;上位性;基因-环境互作;复杂性状

29

#### Epistasis and pleiotropy analysis dissect the genetic architecture of rice complex traits

Md. Asif Ahsan, Ming Chen

Department of Bioinformatics, College of Life Sciences, Zhejiang University, Hangzhou 310058, China.

Corresponding author: <u>asif@zju.edu.cn</u>, <u>mchen@zju.edu.cn</u>

*Motivation:* Genes involved in complex traits do not function alone. Combination of multiple genetic and environmental factors are responsible for complex traits. Single locus association or simple genotype-phenotype relationship may fail to explain such complex traits. Epistasis or gene-gene interaction is recognized to play a fundamental role in pinpointing the genetic architecture of complex traits. Besides, multi-trait epistasis analysis could help to find the pleiotropic genes that affecting more than one trait.

*Methods and Algorithms:* A two-step approach based on mixed linear model was developed for epistasis analysis that account population stratification and polygenic effect. To investigate the performance of our method, simulation studies were performed and compared with the classical and pc-linear approach. Two important rice agronomic traits–flowering time and plant height–were analyzed using the developed epistasis method. To characterize and validate the identified epistatic variants, various bioinformatics analysis such as gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis, protein-protein interactions (PPIs), subcellular location (SCL) and tissue-specific gene expression of the gene/protein were assessed. A pleiotropic interaction network and a genotype-phenotype map were also constructed using the candidate epistatic genes of both traits.

*Result:* Simulation studies show that our method gives better power and low false discovery rate compared to classical and pc-linear approach. Several novel genes were identified from the whole genome epistasis analysis of rice flowering time and plant height traits. Bioinformatics analyses also confirm the involvement of the identified candidate genes to the related traits. Seven pleiotropic genes were identified from the epistatic pleiotropic network and some of them were found in literature those validated by experimental analysis.

*Conclusion:* The findings of this study indicate the importance of epistasis and pleiotropy analysis in dissecting the genetic architecture of rice complex traits. EpiStructure is available at http://bis.zju.edu.cn/EpiStructure/.

Key words: Epistasis analysis, pleiotropy analysis, rice, flowering time, plant height

#### 抽穗期基因对水稻光合性状的遗传效应

刘蒙蒙1施俊1庄杰云2沈波1

1杭州师范大学生命与环境科学学院,浙江杭州,311121

2中国水稻研究所,浙江杭州,310006

摘要: 抽穗期与光合作用均是影响水稻产量的潜力因素,两者在控制水稻 生长发育以及开花繁殖上存在着一定的联系。抽穗期对于产量的控制可能与光 合作用途径密切相关。Hd1和 Ghd7均是控制水稻抽穗期和产量的多效基因,但 它们的影响作用机理尚未研究清楚。本研究应用从珍汕 97/密阳 46 重组自交系 中筛选构建的2套近等基因系 Z43 群体和 Z44 群体研究自然长日照条件下,Hd1 与 Ghd7 在水稻抽穗期间对光合性状的遗传调控作用,其中珍汕 97 的 Hd1为功 能型等位基因,密阳 46 的 Ghd7 为功能型等位基因。主要结果如下:

Z43 群体中 Hd1 座位与 PhyB 座位分离。实验检测到 Hd1 对光合相关性状的主效应,功能型 Hd1 降低水稻抽穗期间的净光合速率、气孔导度、胞间 CO2 浓度、蒸腾速率,同时提高叶绿素含量。但并未检测到 PhyB 对光合性状、叶绿素含量的显著效应。

Z44 群体中 Hd1 座位与 Ghd7 座位分离。实验检测到 Ghd7 对光合性状的主效应,功能型 Ghd7 降低水稻抽穗期间的光合速率、气孔导度、胞间 CO2 浓度和蒸腾速率,未检测到 Hd1 对这四种光合性状的主效应。同时检测到 Hd1 与 Ghd7 之间对光合性状的相互作用,功能型 Hd1 降低水稻抽穗期间的光合能力需要功能型 Ghd7。同时分别检测到 Hd1 与 Ghd7 对叶绿素含量的主效应,但两者之间没有相互作用,表明功能型 Hd1 可以提高叶绿素含量,功能型 Ghd7 可以降低叶绿素含量。

对两套近等基因系进行转录组分析研究,发现 Hd1 与 Ghd7 确实对光合作

31

用和叶绿素含量有着重要的调控作用。Hd1为功能型时,Ghd7引起 2836 个差 异基因,1916个下调;hd1为无功能型时,Ghd7引起 1097 个差异基因,621 个 下调。GO 功能和聚类分析显示功能型 Ghd7 能显著富集到光合作用以及叶绿体 发育相关的 GO 条目,功能型 Ghd7 下调大部分叶绿素合成相关的基因,但当 Hd1 为功能型时,功能型 Ghd7 下调的基因数远大于 hd1 为无功能时的基因数。 表明 Ghd7 主要对叶绿素合成起负调控作用,且在 Hd1 为功能型时,其调控作 用更强。

Ghd7 为功能型时, Hd1 引起 2082 个差异基因, 1346 个下调; ghd7 为无功能型时, Hd1 引起 564 个差异基因, 262 个下调。GO 功能和聚类分析显示当 Ghd7 为功能型时, 功能型 Hd1 能显著富集到光合作用以及叶绿体发育相关的 GO 条目, 功能型 Hd1 下调大部分叶绿素合成相关的基因; 而 ghd7 为无功能时, 功能型 Hd1 仅显著富集到 4 个相关 GO 条目, 引起的差异基因远远少于 Ghd7 为功能型时的差异基因数。表明功能型 Hd1 对光合作用的调控依赖于功能型 Ghd7。

#### 关键词:水稻,抽穗期,光合作用,Hd1,Ghd7

#### Identification of eRNAs-related prognostic signature for colon adenocarcinoma

Rong Hao<sup>1</sup>; Gao Liuying<sup>2</sup>; Hu Shiyun<sup>1</sup>; Yi Tianfei<sup>1</sup>; Li Yanguo<sup>1</sup>; Liao Qi<sup>1,\*</sup>

1. Department of Preventive Medicine, Zhejiang Provincial Key Laboratory of Pathological and Physiological Technology, Medical School of Ningbo University, Ningbo, China, 315211

2. The Affiliated People's Hospital of Ningbo University, Ningbo, China, 315040

\*Email: liaoqi@nbu.edu.cn

**Abstract:** Colon cancer is the third most malignancies worldwide<sup>[1]</sup>. Previous reports have shown that enhancer RNAs (eRNAs), a subclass of lncRNAs, could exert their effects in human tumors<sup>[2]</sup>. However, the potential function of eRNAs in colon cancer remains unexplored. In this study, we integrate the multi-omics data to identify the key eRNAs-related prognostic signature in colon adenocarcinoma (COAD). We identify a total of 8231 eRNAs-related genes (eRGs) in COAD and the functions of these genes mainly include regulation of GTPase activity, cell morphogenesis and cilium assembly. Compared with all genes, the eRGs have higher expression and lower variation. On the basis of tumor group and normal group, we found 261 differentially expressed eRGs which sequentially subjected to univariate cox, lasso and multivariate cox regression analysis. 12 eRGs are identified as prognostic signature for COAD. The final survival risk score model is established: risk

score=0.218\*ATP6V1C2+0.269\*GAS2+0.117\*HOXC6+0.207\*LRRN-0.305\*NEBL+0.129 \* NPSR1 + 0.290\*OSBPL3 +0.355\*PBX4+0.190\*SPTBN5+ 0.387\*TIMP-0.327 \* TMEM220 + 0.226 \*WDR72. According to the median risk score, the COAD patients are divided into high-risk group and low-risk group. High-risk group exhibits worse overall survival (OS) than that in low-risk group with the log-rank test  $p=1.225*10^{-09}$  and the area under the ROC curve (AUC) are 0.81, 0.77 and 0.78 in the 1-, 3- and 5-years, respectively. We also find that the risk score with the highest AUC is a novel and important independent prognostic factor for COAD patients. We
analyze the methylation status between high-risk group and low-risk group. 223 CpG sites are differentially methylated, including 19 sites significantly correlated with corresponding mRNA expression. For example, NPSR1 5'UTR hypomethylation (cg11310191) and 3'UTR hypermethylation (cg23448390) are associated with increased NPSR1 expression. The somatic mutation in high-risk group and low-risk group find that the high-risk group has more co-occurrence mutations. Interestingly, *BRAF* does not appear in the top 25 most frequently mutated genes in low-risk group. There are, however, three unique mutually exclusive mutations in high-risk group (*BRAF-APC*, *BRAF-TP53* and *BRAF- KRAS*) which associated with *BRAF* mutation. These three exclusive mutations probably influence the prognosis of COAD patients.

# Keywords: Colon adenocarcinoma; Enhancer RNAs; Gene expression; Multi-omics; Prognostic prediction

Reference:

- FREDDIE, BRAY, JACQUES, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. Ca A Cancer Journal for Clinicians, 2018,
- [2] LIU Y, DING M, GAO Q, et al. Current Advances on the Important Roles of Enhancer RNAs in Gene Regulation and Cancer [J]. Biomed Research International, 2018, 2018(1-6).

# Regulatory network analysis of mutated genes based on multi-omics data reveals the exclusive features in tumor immune microenvironment between left-sided and right-sided colon cancer

Tianfei Yi1,Yuwei Zhang1,Derry Minyao Ng1,Yang Xi3,Jianjiong Li2,Xiaoxiang Fan2,Yanguo Li4, Shiyun Hu1, Hao Rong1, Yangyang Xie2, Guofang Zhao2, Leyi Chen1, Chen Chen1, Shujing Ni1, Jiaying Mi2, Xiaoyu Dai2,\*, Qi Liao1,\*

Department of Preventative Medicine, Zhejiang Provincial Key Laboratory of Pathophysiology, Ningbo University School of Medicine, Ningbo, Zhejiang, 315211, China

Hua Mei Hospital, University of Chinese Academy of Science, Ningbo, 315000, China

Department of Biochemistry and Molecular Biology, Zhejiang Provincial Key

Laboratory of Pathophysiology, Ningbo University School of Medicine, Ningbo,

Zhejiang, 315211, China

Institute of Drug Discovery Technology, Ningbo University, Ningbo, Zhejiang, 315211, China \*Corresponding authors: Qi Liao, Email: liaoqi@nbu.edu.cn; Xiaoyu Dai, Email: daixiaoyu1968@163.com

#### Abstract

Left-sided colon cancer (LCC) and right-sided colon cancer (RCC) are two kinds of colon with distinct characteristics cancer in tumor immune microenvironment (TIME). Although existing studies have shown a strong association between gene mutation and TIME in colon cancer, whether the influence of mutations on TIME and regulatory mechanism of gene mutations are different between RCC and LCC is still unclear. In this study, we showed the fractions of naive and effector T cells such as CD8+ T cells are higher while those of regulatory T cells are lower in RCC than in LCC. Besides, a stronger association between gene mutation and immune cell infiltration was observed in RCC. Specifically, using multi-omics data, we demonstrated the mutations of most top mutated genes (TMGs)

including *BRAF*, *PCLO*, *MUC16*, *LRP2*, *ANK3*, *KMT2D*, *RYR2* made great contributions to an elevated fraction of immune cells by up-regulating immune-related genes directly or indirectly through miRNA and DNA methylation, whereas the effects of *APC*, *TP53* and *KRAS* mutations on TIME were reversed in RCC. Remarkably, we found the expression levels of several immune checkpoint molecules such as *PD-1* and *LAG3* were correlated with corresponding DNA methylation levels, which were associated with the mutations of TMGs in RCC. In contrast, the associations between gene mutations and TIME were less significant in LCC. We hope that our results will provide a deeper insight into the sophisticated mechanism underlying the regulation between somatic mutations and TIME and thus boost the discovery of differential immunotherapeutic strategies for RCC and LCC.

# Keywords: left-sided colon cancer (LCC), right-sided colon cancer (RCC), tumor immune microenvironment (TIME), gene mutation, multi-omics

# A comprehensive computational tool for functional interpreting single-cell RNA-seq data from pathway-oriented perspective

Yaru Zhang<sup>1</sup>, Yan Zhang<sup>1</sup>, Jun Hu<sup>3</sup>, Yunlong Ma<sup>1</sup>, Yukuan Huang<sup>1</sup>, Fulong Yu<sup>1</sup>, Guijun Zhang<sup>3</sup> and Jianzhong Su<sup>1,2</sup>

- 1. Institute of Biomedical Big Data, School of Biomedical Engineering, School of Ophthalmology & Optometry and Eye Hospital, Wenzhou Medical University, Wenzhou 325027, China.
- 2. Wenzhou Institute, University of Chinese Academy of Sciences, Wenzhou 325011, China.
- 3. College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China

Abstract Advanced in single-cell RNA-seq technologies enable us to uncover more refined and novel cell clusters, which have greatly improve our understanding of cellular states. However, biological interpretation of the clustering results remains a big challenge. Multiple evidences have demonstrated that pathway-based enrichment analysis benefits to elucidate biological implications and molecular mechanisms. However, no effective tool has been implemented for single-cell transcriptome data analysis based on prior biological pathway knowledge.

Here, we present scTPA (http://sctpa.bio-data.cn/sctpa), a web-based platform for pathway-based analysis of single-cell RNA-seq data in human and mouse [1]. scTPA incorporates seven widely-used gene set enrichment methods to estimate the pathway activation scores of single cells based on a collection of available biological pathways with different functional and taxonomic classifications. The clustering analysis and cell-type-specific activation pathway identification were provided for the functional interpretation of cell types from a pathway-oriented perspective. Furthermore, we also systematically evaluated the accuracy, stability, and scalability of these seven transformation tools with 32 real scRNA-seq datasets based on 16 scRNA-seq techniques [2]. Overall, our recent research tries to give an insight into exploration of transcriptional heterogeneity from pathway-oriented perspective and assist computational researchers to design high-quality researches.

# Keywords: scRNA-seq, web tool, transcriptional heterogeneity, pathway analysis score

## **References:**

- [1] Zhang Y\*, Zhang Y\*, Hu J\*, Zhang J, Guo F, Zhou M, Zhang G#, Yu F#, Su J#. scTPA: A web tool for single-cell transcriptome analysis of pathway activation signatures. *Bioinformatics*. 2020.
- [2] Zhang Y\*, Ma Y\*, Huang Y, Zhang Y, Jiang Q, Zhou M, Su J#. Benchmarking algorithms for pathway activity transformation of single-cell RNA-seq data. *Computational and Structural Biotechnology Journal* (Accepted)

## RiceNETDB: 水稻基因组尺度多层次网络重建的整合数据库

杨舒,陈铭\*

浙江大学生命科学学院,杭州,310058

\*E-mail: mchen@zju. edu. cn

理解代谢反应如何将有机体的基因型转化为表型是生物学上的一个重大挑战。全基因组关联研究(GWAS)使用统计方法,通过基因调控网络(GRNs)、蛋白相互作用(PPIs)等,将基因型与表型联系起来。不同水平的调控信息的整合有望为基因型-表型关系的研究提供一个良好的途径。在此,我们通过组学水平的调控信息重构和整合,将基因调控网络、蛋白相互作用和基因组规模的代谢网络合并成一个水稻的单一框架。首先,构建了一个功能基因、代谢物参与的全基因组代谢模型,再将基因调控信息和 microRNA-靶标相互作用整合到代谢模型中,最后,建立数据库以系统地存储和检索水稻基因组规模多层次网络。这为了解水稻基因型-表型关系及调控网络提供了参考。

## 关键字:水稻,基因调控网络,代谢网络

# Co-expression of Mitochondrial Genes and ACE2 in Cornea Involved in COVID-19

Jian Yuan, Ph.D. <sup>1,2,3,#</sup>, Dandan Fan, M.S. <sup>1,2,3,#</sup>, Zhengbo Xue, Ph.D. <sup>1,2,3</sup>, Jia Qu, M.D. <sup>1,2,\*</sup>, Jianzhong Su, Ph.D. <sup>1,2,3,\*</sup>

1.School of Ophthalmology & Optometry and Eye Hospital, Wenzhou Medical University, Wenzhou, 325027, China

2.National Clinical Research Center for Ocular Disease, Wenzhou, 325027, China

3. Institute of Biomedical Big Data, Wenzhou Medical University, Wenzhou, 325027, China,

#These authors contributed equally to the paper as first authors.

\*To whom correspondence should be addressed. Email: sujz@wmu.edu.cn

Word count: 2040

#### Abstract

#### Purpose

The Coronavirus disease 2019 (COVID-19) pandemic severely challenges public health and necessitates the need for increasing our understanding of COVID-19 pathogenesis, especially host factors facilitating virus infection and propagation. The aim of this study was to investigate key factors for cellular susceptibility to SARS-CoV-2 infection in ocular surface cell.

Methods

We combined co-expression and SARS-CoV-2 interactome network to predict key genes at COVID-19 in ocular infection based on the premise that genes underlying a disease are often functionally related and functionally related genes are often co-expressed.

Results

The co-expression network was constructed by mapping the well-known ACE2, TMPRSS2 and host susceptibility genes implicated in COVID-19 GWAS onto a cornea, retinal pigment epithelium and lung. We found a significant co-expression module of these genes in the cornea, revealing that cornea is potential extra-respiratory entry portal of SARS-CoV-2. Strikingly, both co-expression and interaction networks show a significant enrichment in mitochondrial function, which are the hub of cellular oxidative homeostasis, inflammation and innate immune response. We identified a corneal mitochondrial susceptibility module (CMSM) of 14 mitochondrial genes by integrating ACE2 co-expression cluster and SARS-CoV-2 interactome. Gene ECSIT, as a cytosolic adaptor protein involved in inflammatory responses, exhibits the strongest correlation with ACE2 in CMSM, which has shown to be an important risk factor for SARS-CoV-2 infection and prognosis.

#### Conclusions

Our co-expression and protein interaction network analysis uncover that the mitochondrial function related genes in cornea contribute to the dissection of COVID-19 susceptibility and potential therapeutic interventions.

Introduction

A novel coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) associated with severe human infected disease (COVID-19) outbreak starting from December 2019, in China, and the disease is quickly spreading worldwide.<sup>1</sup> Despite being primarily a respiratory virus, COVID-19 can also present with non-respiratory signs, including ocular symptoms as conjunctival hyperemia, chemosis, epiphora, increased secretions, ocular pain, photophobia and dry eye.<sup>2</sup> The presence of virus in tear, conjunctival swab specimens and animal models of infectious increasing clinical and scientific evidence that eye may serve as a potential site of virus replication.<sup>3, 4</sup> Moreover, immunohistochemical studies and single-cell RNA-sequencing datasets revealed both extra- and intra-ocular localization of SARS-CoV-2 entry factors, ACE receptor and TMPRSS2 protease in human eyes.<sup>5, 6</sup> Together, these results suggest that ocular surface cells are susceptible to infection by SARS-CoV-2. More a recent GWAS study identified 3p21.31 as a most significant genetic locus being associated with COVID-19 induced respiratory failure.<sup>7</sup> This locus covers a cluster of six genes consisting of SLC6A20, LZTFL1, CCR9, FYCO1, CXCR6, and XCR1, with the identified risk allele being associated with increased SLC6A20 and LZTFL1 expression. Of note, SLC6A20, LZTFL and FYCO1 is known to associate with eye development, electroretinography abnormal and anterior eye segment morphology. However, whether these key factors for cellular susceptibility to viral infection have correlation in ocular surface cell remains unclear. Herein, we constructed co-expression and interactome networks and mapped genes implicated by COVID-19 GWAS onto corneal co-expression network to infer the function of susceptibility gene.

## Methods

Dataset Summary and Quality Control

An overview of our strategy to inform corneal mitochondrial susceptibility module of SARS-CoV-2 infection in cornea is shown in Figure 1A. We began by collecting the several RNA-Seq datasets of the normal tissue of cornea (n = 19 samples),<sup>8</sup> retina (n = 310 samples)<sup>9, 10</sup> retinal pigment epithelium (RPE) (n = 207 samples),<sup>10</sup> iPSC–derived retinal (n = 5 samples)<sup>11</sup> and lung (n = 546 samples)<sup>12</sup> from the NCBI GEO database (GSE77938, GSE115828 and GSE141531) and GTEx release v8, respectively. Individual data sets underwent stringent quality control, outliers were removed as samples with standardized sample network connectivity Z scores < -2, as described,<sup>13</sup> and were removed. Quantile normalization was then used to transform the statistical distributions across samples to be same.

## Co-expression Analysis

In order to investigate the expression correlation of SARS-CoV-2 entry factors

and GWAS susceptibility genes, we calculated the Pearson correlation coefficient to evaluate the linear correlation between any two genes. To specifically characterize the biological pathways involved, we performed k-means analysis,<sup>14</sup> a popular unsupervised machine learning algorithm, to identify several co-expression clusters. The cluster eigengene was defined as the first principal component summarizing the expression patterns of all genes into a single expression profile within a given cluster. The cluster membership for each virus entry factors and disease susceptibility genes was determined by the correlation between the expression profile of a gene and the cluster eigengene of a module.

Gene Enrichment Analysis

GO and KEGG pathway enrichment analysis were performed by the Metascape gene enrichment analysis tool.<sup>15</sup> A Fisher's exact test was used to identify ACE2 cluster with significantly enriched in SARS-Cov-2 interactome<sup>16</sup> and mitochondrial genes.<sup>17</sup> The interactions between SARS-Cov-2 viral proteins and drug targets were visualized using Cytoscape version 3.6.1.<sup>18</sup>

Differential Expression Analysis

The transcriptome profiling of human keratoconus corneas was also downloaded from GEO database (GSE77938).<sup>8</sup> Only genes with TPM >1 were preserved in the down-stream analysis. The DESeq2<sup>19</sup> package was used to normalize expression levels and detect differential expressed genes (q value cutoff is 0.05). Statistical analysis was done using the R project for statistical computing (http://www.r-project.org).

Results

To understand the expression patterns of ACE2, TMPRSS2 and susceptibility genes in cornea, we firstly compared the expression level of ACE2 in cornea, retina, RPE and lung tissues based on bulk RNA sequencing. As expected from prior literature,<sup>20</sup> the cornea showed a higher ACE2 expression than lung both in terms of their TPM values (Fig 1A). ACE2 exhibits the highest co-expression correlation with TMPRSS2, SLC6A20C and LZTFL1 in the cornea compared to lung and RPE (Fig. 1B). To gain more insight into the biological network of genes associated with SARS-CoV-2 entry factors and susceptibility gene, we performed k-means clustering algorithm to identify genes associated with ACE2 on cornea datasets (Fig. 1C). We identified 26 co-expression modules ranging in size from 111 to 1798 genes. One cluster contained both ACE2, LZTFL1 and FYCO1 (cluster 7; 1434 genes). The strongest correlation with the eigengene (the principal component) of this ACE2 cluster was found for hub genes STK16 (r=0.95, p= $1.07 \times 10^{-9}$ ), a member of NAK family that activate the AP-2 scaffolding protein vital to viral entry and propagation.<sup>21</sup> TMPRSS2 belonged to a separate cluster 5 (603 genes) with hub gene SLC25A1 (r=0.91, p= $4.45 \times 10^{-8}$ ), which involved in TNF-a and IFN-g triggered inflammation.<sup>22</sup> Together these data suggest that the cornea may provide a susceptibility and entry portal for the SARS-CoV-2 entry.

To gain insights into the molecular functionality related to ACE2 cluster, we

integrated the ACE2 cluster and SARS-CoV-2 interactome and observed their shared a number of similarities. For instance, the ACE2 cluster and SARS-CoV-2 interactome eigengenes were highly correlated (r=0.56, p=0.014). Genes within ACE2 cluster were mainly related to mitochondrion functions such as mitochondrial inner membrane and mitochondrial electron transport (p=7.47×10<sup>-29</sup>) (Fig. 2A). We further evaluated shared GO enrichments to determine whether the similarity in the behavior meant that both clusters contained functionally related genes. ACE2 cluster and SARS-CoV-2 interactome were nominally enriched (p<0.001) for 41 and 35 GO terms, respectively. Of these, 17 terms were enriched in both modules. Furthermore, we observed a positive correlation in fold enrichments for shared terms (Fig. 2B). Most of the ontologies shared between the modules described cellular components, biological processes, and molecular functions pertinent to mitochondrion (Fig. 2B). In both the ACE2 cluster and SARS-CoV-2 interactome, the observed numbers of mitochondrial genes were significantly higher than the number what would be expected by chance ( $p=2.1 \times 10^{-14}$  and  $p=3.2 \times 10^{-6}$ , respectively; Fig. 2C). In addition, we found the observed number of ACE2 cluster that was significantly higher (p =0.005) associated with interaction protein of SARS-CoV-2 compared to randomness (Fig. 2C). Together, these results indicated that ACE2 cluster possess the key factors required for cellular susceptibility to SARS-CoV-2 infection in cornea. Therefore, the core 14 genes were co-occurred in all three datasets as the corneal mitochondrial susceptibility module (CMSM) of SARS-CoV-2 infection in cornea (Fig. 2D). Of the 14 CMSM genes, five genes have been shown to directly implicate in the function of electron transport. They include genes such as NDUFB9 a member of mitochondrial respiratory-chain complex 1, and NDUFAF1, NDUFAF2, ECSIT which are involved in mitochondrial respiratory-chain complex 1 assembly. Next, we studied the expression alteration of these genes in cornea from 23 keratoconus patients and 19 health controls. We found that ACE2 expression was significantly increased in keratoconus compared to control cornea ( $\log_2 FC=2.8$ , p=4.4×10<sup>-7</sup>; Fig. 2E). Other than the up-regulation of ACE2 in keratoconus, there were 10 of 21 genes related to SARS-CoV-2 infection was significantly upregulated in keratoconus patients (Fig. 2E, t-test p < 0.05). Based on the elevated expression of ACE2 and other susceptibility genes in keratoconus, we speculated that patients with keratoconus are more likely to be infected by the SARS-CoV-2.

Discussion

To control and mitigate the impact of the COVID-19 pandemic, it is vital to gain greater understanding of the routes and modes of transmission, including the role of the ocular surface. Using a cornea-relevant co-expression network to inform virus entry factors and GWAS interpretation, we were able to identify putative susceptibility genes for highly correlated with ACE2. Interestingly, we found that ACE2 co-expression cluster and SARS-CoV-2 interactome were both enriched for mitochondrial functions. As previously described, the ACE2 not only serves as a critical determinant of CoV-2 transmissibility but also regulates mitochondrial functions.<sup>23</sup> ACE2 overexpression regulates mitochondria-localized NADPH oxidase 4, which is known to produce reactive oxygen species (ROS) in the mitochondria.<sup>24</sup> Oxidative stress caused by ROS excessiveness was indicated as a major player in COVID-19 pathogenesis and severity.<sup>25</sup> Additionally, several lines of evidence have established a link between inflammation and oxidative stress,  $^{26-28}$  for example, TNF- $\alpha$ induces calcium-dependent increase in mitochondrial ROS.<sup>29</sup> Once SARS-CoV-2 enters the host cell, its RNAs such as ORF-9b also can directly manipulate mitochondrial function to release mitochondrial DNA (mtDNA) in the cytoplasm and activate mtDNA-induced inflammasome and suppress innate and adaptive immunity.<sup>30, 31</sup> Together, the proinflammatory cytokines, such as TNF-a, IL-1b, IL-6, IL-10 and CXCL-8, affect diverse physiological processes by driving cellular oxidative stress ROS generation. In turn, increased ROS production stimulates proinflammatory mediator release that contributes to mitochondrial dysfunction. As the corneal cell, especially corneal endothelial cells, we known. is а mitochondria-rich cells. Given highly exposed position, the cornea receives a significant amount of high-tension atmospheric oxygen and the ultraviolet range which result in the generation of ROS and subsequent oxidative stress. Moreover, reactive oxygen species (ROS) are a by-product of oxidative phosphorylation in mitochondria, which can subsequently result in further mitochondrial damage and a further increase in ROS. Overall, a vicious oxidation/ inflammatory cycle is more likely to have a potential impact of SARS-CoV-2 invasion and immune responses in corneal cells.

In our study, we predicted that the mitochondrial related CMSM gene set was vital susceptibility genes of COVID-19, including five genes involved in respiratory electron transport which are being targeted by metformin.<sup>32</sup> A favourable effect of metformin in patients with COVID-19 has been hypothesized as the drug might prevent virus entry into target cells via adenosine monophosphate-activated protein kinase activation and the phosphatidylinositol-3-kinase-protein kinase B-mammalian target of rapamycin signaling pathway.<sup>33</sup> Since metformin is found to have the properties of anti-inflammation and anti-oxidation,<sup>34</sup> it has also been used in the treatment of eye diseases including age-related macular degeneration, glaucoma and diabetic retinopathy.35-37 These may give insight into metformin may lower COVID-19 risk in eye infection. Notably, ECSIT, one of the CMSM gene, is a cytosolic adaptor protein involved in inflammatory responses and plays a regulatory role as part of the TAK1-ECSIT-TRAF6 complex that is involved in the activation of NF-κB by the TLR4 signal. In the previous study, treatment with drugs that inhibited NF-kB activation led to a reduction in inflammation and significantly increased mouse survival after SARS-CoV infection.<sup>38</sup> Additionally, ECSIT is also essential for the association of RIG-I-like receptors (RIG-I or MDA5) to VISA in innate antiviral responses.<sup>39</sup> Therefore, ECSIT may be used as a new drug target to protect against the development of severe forms of COVID-19 infection. Based on our results, we believe that significant insight into COVID-19 in cornea can be gained using

co-expression and interaction networks.

Acknowledgments

This work was supported by the Key Program of National Natural Science Foundation of China (81830027) to J. Qu; the National Natural Science Foundation of China (61871294), Zhejiang Provincial Natural Science Foundation of China (LR19C060001) to J. Su.

#### References

- 1. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine* 2020.
- 2. Wu P, Duan F, Luo C, et al. Characteristics of Ocular Findings of Patients With Coronavirus Disease 2019 (COVID-19) in Hubei Province, China. *JAMA ophthalmology* 2020;138:575-578.
- 3. Jiang R, Liu M, Chen Y, et al. Pathogenesis of SARS-CoV-2 in Transgenic Mice Expressing Human Angiotensin-Converting Enzyme 2. *Cell* 2020;182:50-58.e58.
- 4. Zhang X, Chen X, Chen L, et al. The evidence of SARS-CoV-2 infection on ocular surface. *The ocular surface* 2020;18:360-362.
- 5. Collin J, Queen R, Zerti D, et al. Co-expression of SARS-CoV-2 entry genes in the superficial adult human conjunctival, limbal and corneal epithelium suggests an additional route of entry via the ocular surface. *The ocular surface* 2020.
- 6. Zhou L, Xu Z, Castiglione G, Soiberman U, Eberhart C, Duh E. ACE2 and TMPRSS2 are expressed on the human ocular surface, suggesting susceptibility to SARS-CoV-2 infection. *The ocular surface* 2020.
- 7. Ellinghaus D, Degenhardt F, Bujanda L, et al. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *New England Journal of Medicine* 2020.
- 8. Kabza M, Karolak J, Rydzanicz M, et al. Collagen synthesis disruption and downregulation of core elements of TGF-β, Hippo, and Wnt pathways in keratoconus corneas. *European journal of human genetics : EJHG* 2017;25:582-590.
- 9. Orozco LD, Chen H-H, Cox C, et al. Integration of eQTL and a Single-Cell Atlas in the Human Eye Identifies Causal Genes for Age-Related Macular Degeneration. *Cell Reports* 2020;30:1246-1259. e1246.
- 10. Ratnapriya R, Sosina O, Starostik M, et al. Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. *Nature genetics* 2019;51:606-610.
- 11. Gao M, Lei X, Han F, et al. Patient-Specific Retinal Organoids Recapitulate Disease Features of Late-Onset Retinitis Pigmentosa. *Frontiers in cell and developmental biology* 2020;8:128.
- 12. Consortium G. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* 2015;348:648-660.
- 13. Oldham M, Langfelder P, Horvath S. Network methods for describing sample relationships in genomic datasets: application to Huntington's disease. *BMC systems biology* 2012;6:63.
- 14. Jain AK. Data clustering: 50 years beyond K-means. Pattern recognition letters 2010;31:651-666.
- 15. Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature communications* 2019;10:1-10.
- 16. Gordon DE, Jang GM, Bouhaddou M, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 2020;1-13.
- 17. Calvo S, Clauser K, Mootha V. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic acids research* 2016;44:D1251-1257.
- 18. Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 2011;27:431-432.
- 19. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 2014;15:550.
- 20. Sungnak W, Huang N, Béavin C, et al. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nature medicine* 2020;26:681-687.
- 21. Stebbing J, Krishnan V, de Bono S, et al. Mechanism of baricitinib supports artificial intelligence-predicted testing in COVID-19 patients. *EMBO molecular medicine* 2020;12:e12697.

- 22. Infantino V, Iacobazzi V, Menga A, Avantaggiati M, Palmieri F. A key role of the mitochondrial citrate carrier (SLC25A1) in TNFα- and IFNγ-triggered inflammation. *Biochimica et biophysica acta* 2014;1839:1217-1225.
- 23. Shi T, Yang F, Liu C, et al. Angiotensin-converting enzyme 2 regulates mitochondrial function in pancreatic β-cells. *Biochemical and biophysical research communications* 2018;495:860-866.
- 24. Kim S, Kim Y, Jeong K, et al. Angiotensin II-induced mitochondrial Nox4 is a major endogenous source of oxidative stress in kidney tubular cells. *PloS one* 2012;7:e39739.
- 25. Delgado-Roche L, Mesta F. Oxidative Stress as Key Player in Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) Infection. *Archives of medical research* 2020;51:384-387.
- 26. Naik E, Dixit V. Mitochondrial reactive oxygen species drive proinflammatory cytokine production. *The Journal of experimental medicine* 2011;208:417-420.
- 27. Khomich O, Kochetkov S, Bartosch B, Ivanov A. Redox Biology of Respiratory Viral Infections. *Viruses* 2018;10.
- 28. Shao H, Lan D, Duan Z, et al. Upregulation of mitochondrial gene expression in PBMC from convalescent SARS patients. *Journal of clinical immunology* 2006;26:546-554.
- 29. Dada LA, Sznajder JI. Mitochondrial Ca 2+ and ROS take center stage to orchestrate TNF-α–mediated inflammatory responses. *The Journal of clinical investigation* 2011;121:1683-1685.
- 30. Shi C, Qi H, Boularan C, et al. SARS-coronavirus open reading frame-9b suppresses innate immunity by targeting mitochondria and the MAVS/TRAF3/TRAF6 signalosome. *Journal of immunology* (*Baltimore, Md : 1950*) 2014;193:3080-3089.
- 31. Singh K, Chaubey G, Chen J, Suravajhala P. Decoding SARS-CoV-2 hijacking of host mitochondria in COVID-19 pathogenesis. *American journal of physiology Cell physiology* 2020;319:C258-C267.
- 32. Soberanes S, Misharin A, Jairaman A, et al. Metformin Targets Mitochondrial Electron Transport to Reduce Air-Pollution-Induced Thrombosis. *Cell metabolism* 2019;29:335-347.e335.
- 33. Sharma S, Ray A, Sadasivam B. Metformin in COVID-19: A possible role beyond diabetes. *diabetes research and clinical practice* 2020;164:108183.
- 34. Cameron A, Morrison V, Levin D, et al. Anti-Inflammatory Effects of Metformin Irrespective of Diabetes Status. *Circulation research* 2016;119:652-665.
- 35. Brown EE, Ball JD, Chen Z, Khurshid GS, Prosperi M, Ash JD. The common antidiabetic drug Metformin reduces odds of developing age-related macular degeneration. *Investigative ophthalmology* & visual science 2019;60:1470-1477.
- 36. Richards JE, Lin H-C, Nan B, et al. Targeting aging: Geroprotective Drug Metformin Reduces Risk of Adult-onset Open-angle Glaucoma. *Investigative Ophthalmology & Visual Science* 2014;55:1668-1668.
- 37. Munie M, Noorulla S, Rana S, et al. Effect of metformin on the development and severity of diabetic retinopathy. *Investigative Ophthalmology & Visual Science* 2014;55:1069-1069.
- 38. DeDiego M, Nieto-Torres J, Regla-Nava J, et al. Inhibition of NF-κB-mediated inflammation in severe acute respiratory syndrome coronavirus-infected mice increases survival. *Journal of virology* 2014;88:913-924.
- 39. Lei C, Zhang Y, Li M, et al. ECSIT bridges RIG-I-like receptors to VISA in signaling events of innate antiviral responses. *Journal of innate immunity* 2015;7:153-164.

## **Figure legend**



Figure 1. Expression of ACE2 across different eye tissues and its co-expression pattern in the cornea.

A. Overview of the approach and pipeline used to predict susceptibility module responsible for SARS-CoV-2 infection. B. Expression of ACE2 in cornea, retina, RPE and lung. C. Pearson's correlation analysis were estimated between ACE2 and TMPRSS2 and six susceptibility genes. The size of the circle scales with the

correlation magnitude. The darker the color, the larger the magnitude of correlation coefficient. Star sign (\*) indicates statistical significance. D. Eigengene correlation heatmap representing the strength and significance of correlations between cluster eigengenes. Pearson's correlation coefficient is used as the correlation descriptor (red and blue for positive and negative correlations, respectively).



## Figure 2. Identification of the Cornea Functional Module and Mitochondrial Susceptibility Genes

A. Clustered heat map of gene ontology (GO) terms among cluster genes. Color coding according to legend at the bottom, only gene ontology terms with FDR < 0.05 were considered. B. Gene ontology fold enrichments are correlated for GO terms shared between ACE2 cluster and SARS-CoV-2 interactome. C. Venn diagram depicting the overlap between ACE2 cluster, SARS-CoV-2 interactome and mitochondrial genes. *P*-value computed using Fisher's exact test. D. SARS-CoV-2 protein-protein interaction between 14 mitochondrial susceptibility genes (circles) and 6 SARS-CoV-2 proteins (red diamonds). Blue edge thickness proportional to interaction of gene expression. E. mRNA abundance change of ACE2, TMPRSS2 and susceptibility genes in keratoconus.

# DNN 在拟南芥 MYB 转录因子识别中的应用

马骥<sup>1</sup>,杨丙贤<sup>1</sup>

1 (浙江理工大学生命科学与医药学院, 杭州 311121)

摘要:目的:深度神经网络(deep neural network, DNN)具有强大的非线性 关系学习能力,在复杂的模式识别和分类问题上已经得到广泛的应用。转录因 子(transcription factor, TF)是一类能够与基因的启动子区结合,对基因的转录 起激活或抑制作用的 DNA 结合蛋白,MYB 家族是植物中最大的转录因子家族 之一。MYB TF 的鉴定属于模式识别问题,可以用 DNN 解决。方法:本研究使 用反向传播(backpropagation, BP)算法的多隐层 DNN,在阈值为 0.01,0.001 和 0.0001 的条件下,在模式植物拟南芥的 215 个 MYB TFs 和 1000 个随机挑选 的蛋白编码基因中以 7:3 的比例划分训练集和测试集,以蛋白氨基酸序列为基础 用主成分分析进行了特征提取,作为神经网络的输入进行了训练和预测。结果: 在阈值为 0.01 的条件下,双隐层 DNN 预测的真阳性率(True positive rate, TPR) 为 98.46%,假阳性率为 0.00;测试集数据在 HMMER 网站预测结果显示 TPR 为 98.46%,假阳性率为 0.01。结论: DNN 适用于 MYB 转录因子的识别问题求 解,在转录因子鉴定方面具有巨大的实用潜力。

## 关键词:深度神经网络;转录因子;MYB;反向传播算法

#### Predicting Arabidopsis MYB transcription factors Using a DNN

Ma Ji<sup>1</sup>, Yang Bingxian<sup>1</sup>

<sup>1</sup>(*Zhejiang Sci-Tech University college of life sciences and medicine*)

Abstract: Objective: Deep neural neural network (DNN) has a strong nonlinear relationship learning ability, and has been widely used in complex pattern recognition and classification problems. Transcription factor (TF) is a kind of DNA binding protein that can bind to the promoter region of a gene to activate or inhibit transcription. MYB family is one of the largest TF families in plants. The identification of MYB TFs is a pattern recognition problem, which can be solved by a DNN. Methods: The back propagation (BP) algorithm was used in this study, with thresholds of 0.01, 0.001 and 0.0001, 215 MYB TFs and 1000 randomly selected protein coding genes from Arabidopsis thaliana were divided into a training set and a test set at a ratio of 7:3. Based on amino acid (aa) sequences, feature extraction was performed by a principal component analysis (PCA) and was used as the DNN input. Results: With a threshold of 0.01, the true positive rate (TPR) of the prediction of a double-hidden-layer DNN is 98.46%, and the false positive rate is 0.00; the prediction results of the test set on HMMER website show that TPR is 98.46% and the false positive rate is 0.01. Conclusion: DNN is suitable for the identification of MYB TFs and has great practical potential in the TF identification problem.

# Key words: deep neural network; transcription factor; MYB; back propagation (BP) algorithm

引言

浅层人工神经网络(ANN)较浅的网络结构限制了神经网络学习复杂非线性关系的能力。当样本数量少,样本线性可分特征明显时,ANN 拟合的速度以及预测的精度均表现出色[1]。但当样本数量大,样本非线性特征较多时,深度神经网络(DNN)才是理想的选择。当样本数量充足时,相对于ANN,DNN 具有更强的非线性关系学习能力,预测精度更高,模型更为稳定[2]。目前,DNN 已证明可以应用于红外光谱定量分析[2]、地电厂数据处理[3]、智能用电负荷监测[4]、海面风场短时预报[5]、无人机森林步道识别[6]等领域,而且在预测精度上均较传统方法有较明显提高[2-6]。这表明DNN 在各种实际应用中的非线性模式识别问题上具有良好的泛化性能和实用意义。

转录调控因子(transcription factor, TF)是一类能够与基因的启动子区结合,对基 因的转录起激活或抑制作用的 DNA 结合蛋白[7]。典型的 TF 含有 DNA 结合域 (DNA binding domain)、转录调控区等功能区域,其中 DNA 结合区是 TF 的核心区域,其氨 基酸序列和三维结构具有较强的保守性。根据 TF DNA 结合区的特点,可以分成不同 的家族,比如碱性螺旋-环-螺旋(basic/helix-loop-helix, bHLH)家族[8]、MYB家族[9]、 WRKY 家族[10]、碱性亮氨酸拉链(basic region/leucine zipper motif, bZIP)家族[11]等。 其中, MYB 家族是植物 TF 中最大的家族之一, 在植物的转录调控中起着多方面的重 要作用。MYB TFs 的特征是含有保守的 MYB DNA 结合域,通常由 1-4 个不完全相同 的长度约为 52 个残基的氨基酸序列重复(R)组成。MYB TFs 根据含有 R 的数量分成 4 类: R2R3-MYB, 1R-MYB/MYB-related, 3R-MYB 和 4R-MYB, 大多数植物 MYB 基 因编码 R2R3-MYB TF[7]。目前, MYB TFs 的鉴定通常都是基于从全长转录组数据获 取蛋白氨基酸序列,然后根据 MYB 结合域的保守性,使用多序列比对软件比如 HMMER 进行相似程度打分并选出最大疑似候选,缺点是假阳性率较高[12]。目前,有 学者使用神经网络鉴定转录因子结合位点,发现在实际应用中,神经网络能成功识别 出不明显的模体,具有较高的分类精确度[13-15]。据作者所知,使用 DNN 进行基于氨 基酸序列的转录因子识别问题的求解尚未见报道。

DNN 在实际应用中对其预测性能影响较大的一个因素是输入数据的特征提取。如何用较少的维度准确而全面的概括研究对象的特征,是 DNN 发挥其优异性能的关键。 基于 DNA、RNA 或蛋白质序列进行分类的算法研究中,有的学者[16]基于 DNA 字符 串统计某段核苷酸序列出现的频率,出现的频率高,就认为可能是目标序列。此方法 的优点是能敏感地识别出保守片段,局限性在于当 DNA 序列中有突变或 DNA 序列保

51

守程度较低时,此种算法就难以识别。还有学者[17]使用感知器模型,对 2000 年全国 数学建模竞赛试题 A 题《DNA 序列分类》[18]进行了神经网络求解,其间采用了 DNA 片段 ATGC 的百分比作为分类的依据。在转录因子结合位点的预测中,有学者[19]统计 了蛋白-核酸复合物中的氨基酸侧链同核酸之间的作用对,通过分析 DNA 与转录因子 氨基酸残基作用的局部环境信息,发现一些三联或者五联残基片段总是结合 DNA。这 给出了用 TF 氨基酸序列中三联或五联残基片段出现频数作为特征向量进行神经网络 拟合的思路。

综上所述,由于 TF 的保守域存在于蛋白编码序列上,所以认为选用蛋白质的氨基 酸序列进行 TF 的预测是较为理想的方法。而接下来的问题就是如何将 TF 的氨基酸序 列中的特征提取出来。第一种想法是根据[17]的思路,计算蛋白序列中 20 个氨基酸出 现的频数,然后将这 20 个数字作为特征向量组成矩阵作为神经网络的输入。这种方法 是可行的,但是这种方法会遗漏氨基酸的顺序信息,因为保守区域出现的氨基酸前后 顺序是固定的。第二种想法是根据[19]的思路,以氨基酸序列中三联或五联残基片段出 现的频数作为特征向量进行神经网络拟合。这种方法考虑到了氨基酸的顺序信息,但 缺点是特征向量的维数将非常高(三联氨基酸残基共 20\*20\*20=8000 种,五联氨基酸 残基共 205=320 万种),能否适当的降维将是这种方法能否合理应用的关键。

拟南芥(Arabidopsis thaliana)是基因组研究最为透彻的模式植物之一,全基因组 共含有 48,359 蛋白编码基因[20],总共含有约 1770 个 TFs,可分成 51 个家族[21],其 中 MYB TF 共有 125 个。本研究拟采用拟南芥蛋白编码基因的氨基酸序列作为研究对 象,分别提取 20 种氨基酸残基的频数以及五联氨基酸残基的频数作为 DNN 的输入实 现对 MYB TF 的自动识别。

1 材料和方法

1.1 原始数据的获取和准备

拟南芥全基因组共 48,359 蛋白编码基因的氨基酸序列下载自 TAIR 网站[20]。拟南 芥 TF 的位点名称和所属家族的对应表下载自 AtTFDB 网站[21]。通过位点名称和 TF 家族的对应,确定了拟南芥全基因组每个蛋白编码基因是否是 MYB TF 基因,作为 DNN 的响应变量。使用自编的 perl 程序统计了拟南芥 215 个 MYB TFs 和 1000 随机蛋白编 码基因氨基酸序列中 20 种氨基酸的频数组成 1215 × 20 的矩阵作为 DNN 的输入来源 1。 使用自编的 perl 程序统计了拟南芥 215 MYB TF 氨基酸序列中所有五联残基的种类以 及频数,发现拟南芥 TF 氨基酸序列中总共有 33,885 种不同的五联残基;用主成分分

52

析(primary component analysis, PCA)对数据进行了降维,发现前 18 个主成分贡献之 和为 85.19%,前 38 个成分贡献之和为 94.96%。按照[22]的建议,选取累积贡献大于等 于 85%的前 18 个主成分作为特征向量组成 1215 x 18 的矩阵作为 DNN 的输入来源 2。

#### 2.2 DNN 的训练和预测应用

DNN 的建立采用 R 软件 neuralnet 包[23]的 neuralnet()函数, 建立 DNN 的算法是带 或 不 带 回 溯 [24] 的 反 向 传 播 (backpropagation, BP )、 弹 性 反 向 传 播 (resilient backpropagation, RPROP)[25]算法, 或新的全局收敛的 BP 训练算法[26]。将拟南芥 215 个 MYB TFs 和 1000 随机蛋白编码基因分别按 7:3 比例[1]分成训练集和测试集, 归一 化采用特征归一化法[1]:

$$x_{scaled} = \frac{x - \min(x)}{\max(x) - \min(x)}.$$
 (1)

归一化后使用输入来源 1 和 2 的训练数据集分别对 DNN 进行训练。DNN 的隐层 (hidden layer)数量范围为 2~5,每层所含神经元节点个数为 6~10。算法为带权重回 溯的弹性反向传播算法(rprop+),阈值(threshold)为 0.0001~0.01,学习率 (learningrate.factor)为 0.5~1.2,激活函数(act.fct)采用逻辑函数(Sigmoid),错误 函数(err.fct)采用误差平方和(sum of squared errors, SSE),输出采用非线性输出 (linear.output=False)。

将训练后的 DNN 应用于输入来源 1 和 2 的测试集数据分别进行预测,并与测试集 响应变量做比较。

#### 2 结果

2.1 使用 20 种氨基酸频数作为 DNN 的输入

用 perl 程序统计了拟南芥全基因组 48,359 蛋白编码基因氨基酸序列中 20 种氨基酸 的频数(图 1),发现氨基酸含量分布可分成三个阶梯,第一阶梯是亮氨酸和丝氨酸(频数=1.92~1.99M),其含量明显高于其他氨基酸;第二阶梯是谷氨酸、缬氨酸、赖氨酸、 甘氨酸和丙氨酸(频数=1.31~1.42M),第三阶梯是其余氨基酸(频数=0.26~1.13),含量最低的是色氨酸。



图 1 20 种氨基酸在拟南芥所有蛋白编码基因的频数条形图 Fig.1 A bar plot of the frequency of 20 amino acids of all protein coding genes in Arabidopsis thaliana

然后,根据 AtTFDB 下载的分类表格,按照位点名称将拟南芥全基因组蛋白编码 基因分成 MYB 相关和非 MYB 相关基因。其中编码 MYB 家族 TF 的基因有 215 个,非 MYB 基因有 48,144 个。215 的 MYB 基因中包括 135 个 MYB 家族 TF 基因以及 80 个 MYB-related 家族 TF 基因, R2R3-MYB TF 基因有 129 个,占 MYB 家族的 95.56%。

接着构建训练集和测试集。从 MYB 基因中随机选出 150 个作为训练,剩余 65 个 作为测试。从其余基因中随机选出 700 个作为训练,随机选出 300 个作为测试。这样, 训练集共 850 个样本,包括 150 个 MYB 基因;测试集共 365 个样本,包括 65 个 MYB 基因。然后,归一化数据,归一化能够消除输入数据量纲的影响,使输入数据在同一 个维度上进行比较,是影响神经网络收敛速度和输出结果非常关键的一步[1]。

接着进行神经网络的训练。训练数据共 21 个变量,包括 20 个氨基酸的频数以及 1 个是否是 MYB TF 的标记列。以前 20 列作为输入,最后一列作为输出训练 DNN,参数见方法部分的说明。训练结束后,将神经网络拟合的值四舍五入为整数,然后生成标记列与拟合值的二维列联表,即混淆矩阵(Confusion matrix)(表 1)。从表中可以看到, DNN 拟合的准确率(Accuracy)为 847/850=99.6%,表明神经网络拟合良好。

54

表 1 2 隐层 DNN 阈值为 0.01 的条件下在 850 个 20 种氨基酸频数训练数据上的拟合结果 Tab.1 Fitting results of the double-hidden-layer DNN on the 850-sample training data of the frequency of 20 kinds of amino acid with a threshold of 0.01

	DNN 拟:	合结果	
AtTFDB 注释			
	0 (Not_MYB)	1 (MYB)	
Not_MYB	700	0	
MYB	3	147	

接着用训练好的神经网络进行预测。输入数据是测试集 365 个样本的 20 个氨基酸 的频数,预测结果见表 2。表中数据显示真阳性率(True Positive Rate, TPR)为 49/65=75.4%,假阳性率(False Positive Rate, FPR)为 5/300=0.0167<0.05。数据表明神 经网络基本上能够根据输入数据识别出 MYB 和非 MYB 的蛋白序列。

表 2 2 隐层 DNN 阈值为 0.01 的条件下在 365 个 20 种氨基酸频数测试样本上的预测结

果

Tab.2 Predictions of the double-hidden-layer DNN on the 365-sample test data comprised of the frequency of 20 kinds of amino acid with a threshold of 0.01

	DNN 预测结果		
AtTFDB 注释			
	0 (Not_MYB)	1 (MYB)	
Not_MYB	295	5	
MYB	16	49	

2.2 使用五联残基频数作为 DNN 的输入

使用自编的 perl 程序统计后发现, 拟南芥 TF 氨基酸序列的 33,885 种五联残基中 频数最高的三种是 SCRLR, KSCRL 和 GRTDN。选取了频数最高的 400 种五联残基(频数范围 10~156),用 PCA 方法进行了降维(图 2),发现前 18 个主成分特征根的值(即主成分的方差贡献率)之和为 85.19%。按照主成分选取方差贡献率之和≥85%的原则[22],提取了前 18 个主成分作为 DNN 输入的特征向量。使用自编 perl 程序提取拟南 芥 48,359 个蛋白编码基因的最高 400 种五联残基组成 48,359 x 400 的矩阵,然后通过 PCA 的特征向量(400 x 18)转换成 48,359 x 18 的矩阵。



图 2 拟南芥 215 个 MYB 转录因子氨基酸序列中 400 个最高频数五联残基的主成分分 析碎石图

然后构建训练集和测试集,方法如上一节所述。接着进行神经网络的训练,训练数据共 19 个变量,包括 18 个主成分以及 1 个是否是 MYB TF 的标记列。训练结束后 DNN 拟合的结果如表 3,准确率为 100%。这表明五联残基 18 个主成分很恰当的提取 了 MYB 和非 MYB 蛋白序列的特征,而 DNN 学到了这个区别。

表 3 2 隐层 DNN 阈值为 0.01 的条件下在 850 个五联残基主成分训练数据上的拟合结果 Tab.3 Fitting results of the double-hidden-layer DNN on the 850-sample training data comprised of the frequency of penta residuals with a threshold of 0.01

	DNN 拟行	合结果
AtTFDB 注释		
	0 (Not_MYB)	1 (MYB)
Not_MYB	700	0
MYB	0	150

接着用训练好的神经网络进行预测。输入数据是测试集 365 个样本五联残基的 18 个主成分,预测结果见表 4。TPR 为 98.46%,假阳性率为 0。这表明上面训练的 DNN 具有很好的泛化能力,具有一定的解决真实问题的实际应用能力。

Fig.2 A scree plot of top 400 penta residues on the amino acid sequence of 215 MYB TFs in Arabidopsis thaliana

表 4 2 隐层 DNN 阈值为 0.01 的条件下在 365 个五联残基主成分测试样本上的预测结果 Tab.4 Predictions of the double-hidden-layer DNN on the 365-sample test data comprised of the frequency of penta residuals with a threshold of 0.01

	DNN 预测结果			
AtTFDB 注释				
	0 (Not_MYB)	1 (MYB)		
Not_MYB	300	0		
MYB	1	64		

2.3 预测样本蛋白序列在线 HMMER 预测

我们将用于 DNN 预测的 365 条蛋白序列 (包括 65 MYB 转录因子和 300 随机挑选 的拟南芥蛋白编码基因)上传到 EMBL-EBI 的 HMMER 网络服务器[27]进行了同源性 搜索,之后,我们根据每条蛋白序列的最佳匹配 (top hit)和 AtTFDB 的注释制作了混 淆矩阵 (表 5)。结果显示,HMMER 预测结果的 TPR 为 98.64% (65 个 MYB 转录因 子预测中了 64 个),假阳性率为 0.01。具体来看,65 个 MYB 转录因子中 HMMER 鉴 定 AT4G17780.1为 FBA\_3,为 F-box associated domain, DNN 预测的错误也是这个。HMMER 鉴定的 4 个假阳性是 AT4G02550.4, AT1G09710.4, AT5G17300.1 和 AT3G10113.1,在 AtTFDB 网站搜索结果均为 no result。

表 5 365 个测试样本在 HMMER 网络服务器在线预测的结果
Tab. 5 Predictions of the HMMER webserver on the 365-sample test set

	HMMER 预测结果		
AtTFDB 注释			
	0 (Not_MYB)	1 (MYB)	
Not_MYB	286	4	
МҮВ	1	64	

3 讨论

3.1 ANN 不适用于具有复杂输入的问题的求解

在标记拟南芥全基因组蛋白编码基因是否是 MYB TF 后,用 20 个氨基酸频数训练 神经网络时,我们最初采用了隐含层含 5 个神经元的单层人工神经网络(artificial neural

network, ANN)进行训练,函数采用 R 软件 nnet 包[28]的 nnet()函数,发现拟合数据的 准确率极低(1.88%),150 个 MYB TF 仅仅正确拟合了 11 个。将训练好的该网络用 于预测,发现 TPR 为 20.0%,假阳性率为 0.97,远大于 0.05 的标准。随后我们将神经 元的数量调整到 10 个,准确率更低了(1.76%)。导致这样结果的原因可能有二。其 一,nnet 采用的单层前馈神经网络没有使用 BP 算法,隐含层和输出层没有反馈,而我 们在后来用 neuralnet 建立的双隐层神经网络虽然也是前馈神经网络,但采用了 BP 算 法计算损失函数得到残差,然后把残差向后一层层传播调整权重,从而使权重能够符 合输出的要求。其二,nnet 能够成功的对 Iris 数据集[29]的三种花进行分类识别(表 6), 可能是因为输入较为简单(只有 4 个变量),而我们的输入有 20 个变量。而且,单层 神经网络只能解决线性可分(linearly separable)问题,这是因为单层神经网络是一种 线性划分输入数据空间的模型[30]。这是单层神经网络的局限性,而多层神经网络则可 以克服这一点。单层神经网络只适用于特定类型问题的求解,而多层神经网络没有这 样的局限性[30]。Iris 数据集中 Setosa 类和非 Setosa 类是线性可分的,而 Versicolor 和 Virginica 之间是有交集的(图 3),这导致 nnet 能够正确识别 Setosa 类 (TPR=25/25=100%),而在区分 Versicolor 和 Virginica 时则有 2 个错误(表 6)。



图 3 使用凸包(Convex Hull)[31]检查 Iris 数据集的三种花卉是否线性可分 Fig.3 The use of convex hull[31] to check whether the 3 species flowers are linearly separable

	ANN 预测结果			
品种	1 (versicolor)	2 (setosa)	3 (virginica)	
versicolor	23	0	2	
setosa	0	25	0	
virginica	0	0	25	

表 6 ANN 在 Iris 数据集上使用 4 个变量(萼片长宽、花瓣长宽)进行预测的结果 Tab.6 Predictions of ANN on Iris dataset using 4 variables (sepal length and width, petal length and width)

3.2 使用 2~5 隐层 DNN 在转录因子识别问题上的 TPR 无明显差异

BP 神经网络在解决非线性模式识别问题上的优势逐渐显现[32]。目前,BP 神经网 络已应用于 TF 结合位点的预测,人们已经开发了许多识别和预测 TF 结合位点的算法 和软件,但还有很多方面需要进一步加强和完善[32]。生物大分子系统的识别,例如 DNA-蛋白质的识别,是理解各种生物过程机制的最重要的问题之一。BP 神经网络至 今仍存在一个难以解决的问题是难以确定隐含层的数量[33]。在 BP 神经网络的众多参 数中,隐含层数量是其中非常重要的参数,它的设置对 BP 神经网络的性能影响很大。 目前理论上还不存在一种科学普遍的用于确定隐含层数量的方法,应用时只是凭借设 计者以往的经验以及借助多次实验进行确定[34]。隐含层数量太少,则网络所能获取的 用以解决问题的信息太少;若数目太多,不仅增加训练时间,更会导致学习时间过长, 且误差不一定最佳,并可能出现"过拟合"(Overfitting)问题。存在选择最佳隐含层数 量的参考公式[33]:

$$n_1 = \sqrt{n + m} + a,$$

(2)

其中 n1 为隐含层数量, n 为输入单元数, m 为输出单元数, a 为 1-10 之间的常数。 在实际问题中,通常的做法是枚举范围内的隐含层数量进行验证综合比较,然后确定 神经网络最终的最佳隐含层数量。本研究使用隐含层数量为 2, 3, 4, 5 的 BP 神经网络 (DNN),同时调整阈值,观察 DNN 在解决拟南芥 MYB 转录因子识别问题的性能优

劣。

训练集和测试集数据采用输入来源 1 的数据,在阈值为 0.01, 0.001 和 0.0001 的条件下,使用训练好的 ANN、隐层数量为 2, 3, 4, 5 的 DNN 在测试集上预测的 TPR 如图

4。当阈值为 0.0001 时, 5 隐含层的 DNN 预测 TPR 升高(从 75.38%上升到 87.69%), 预测 TPR 为全局最高(87.69%)。



图 4 在阈值为 0.01, 0.001 和 0.0001 下,不同隐层数量的 DNN(隐层数量为 1 时为 ANN) 在拟南芥 MYB 转录因子识别问题上的真阳性率(TPR)比较 Fig.4 Comparison of TPR of DNNs with different number of hidden layers on MYB TF

recognition in Arabidopsis thaliana at the threshold of 0.01, 0.001 and 0.0001

以隐层数量和阈值作为二因素进行方差分析发现隐层数量的主效应明显(p值 <0.05)(表7),阈值的主效应不明显,说明不同隐层数量的神经网络预测的 TPR 有明显不同,而不同阈值下的神经网络预测的 TPR 没有明显不同。进一步做不同隐层数量 TPR 之间的两两比较,发现隐层数为1的 ANN 与隐层数为2,3,4,5 的 DNN 预测的 TPR 均显著不同(p值<0.05)(表8),隐层数为2,3,4,5 的 DNN 预测的 TPR 没有显著差异。以上结果表明隐层数量在2~5 之间的 DNN 在阈值范围 0.0001~0.01 均能较好的识别拟南芥 MYB 转录因子。

表7 隐层数量和阈值两个因素对 TPR 效应的方差分析表

Tab.7 The ANOVA table of the effect of the number of hidden layers and threshold on the

u de positive fate					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
隐层数量	4	3404	851.1	47.24	0.0000132
阈值	2	3	1.3	0.07	0.933
Residuals	8	144	18		

#### 表 8 不同隐层数量的神经网络预测 TPR 的两两比较 t 检验 p 值表

		隐层数量		
隐层数量				
	1	2	3	4
2	0.00000094	-	-	-
3	0.00000128	0.34	-	-
4	0.00000128	0.34	1	-
5	0.00000094	1	0.34	0.34

Tab.8 The p-value table of t-test for the true positive rate of neural networks with different number of hidden layers

3.3 使用五联残基频数能更好的反映输入数据的特征

在本研究中,我们使用 20 种氨基酸频数作为输入数据的来源 1,还使用了五联残 基频数作为输入的来源 2。在这里,我们用隐层数量(=2,3,4,5,我们在这一步剔除了 单隐层的 ANN,只比较 DNN 预测的阳性率)和方法(20 种氨基酸频数或五联残基频 数)作为二因素进行方差分析,发现方法的主效应明显(p值<0.05)(表 9),隐层数 量的主效应不明显,这说明使用 20 种氨基酸频数和使用五联残基频数作为输入来源的 DNN 预测阳性率有显著不同,而不同隐层数量的 DNN 预测阳性率没有显著差异。以 上结果表明使用五联残基频数作为输入能使神经网络更加准确的区分拟南芥 MYB 转 录因子和非 MYB 基因,这也与我们最初的设想一致,即 20 种氨基酸频数忽略了 MYB 转录因子保守 DNA 结合域相邻残基之间具有明显的顺序性这一信息。

表9 隐层数量和方法两个因素对阳性率效应的方差分析表

Tab.9 The ANOVA table of the effect of the number of hidden layers and method on the

		1			
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
隐层数量	3	5.4	1.8	0.573	0.6705
方法	1	351.1	351.1	112.36	0.00179
Residuals	3	9.4	3.1		

positive rate

4 结论

本研究通过使用不同隐层数量的神经网络(ANN 和 DNN),使用不同的阈值(决

定神经网络停止训练的误差阈值),以拟南芥全基因组蛋白编码基因的氨基酸序列中 20 种氨基酸的频数以及五联残基的频数分别作为特征进行神经网络的训练和预测,发 现阈值和隐层数量在一定范围内,DNN的预测性能稳定;而用五联残基频数作为输入 训练 DNN,能够显著地提高 DNN 的预测准确度。以上结果显示,DNN 能够用于 MYB 转录因子识别的问题求解,具有良好的泛化性能和预测准确度。

另外,基于文献调研,DNN 在样本数量充足的前提下,才能展现出最优异的预测性能[2],所以我们在后续分析中,还将考虑大幅度增加用于训练和预测的样本数量,那时,我们将比较单独使用单个残基、单独使用五联残基、联合使用单个残基和五联 残基、乃至单独使用三联残基进行基于 DNN 的转录因子识别的真阳性率的变化。

参考文献

[1] Ciaburro G, Balaji Venkateswaran B. Neural networks with R [M]. Birmingham, UK: Packt Publishing. 2017.

[2] 张强,魏儒义,严强强,等. 深度神经网络在红外光谱定量分析 VOCs 中的应用 [J]. 光谱学与光谱分析. 2020. 4: 1099-1106.

[3] 汪凯翔,黄清华,吴思弘. 长短时记忆神经网络在地电场数据处理中的应用 [J]. 地球物理学报. 2020. 8: 3015-3024.

[4] 燕续峰, 翟少鹏, 王治华, 等. 深度神经网络在非侵入式负荷分解中的应用 [J]. 电力系统自动化. 2019. 1: 126-132.

[5] 王国松,王喜冬,侯敏,等. 基于观测和再分析数据的 LSTM 深度神经网络沿海风速预报应用研究 [J]. 海洋学报. 2020. 1: 67-77.

[6] 侯永宏,吕晓冬,陈艳芳,等.深度神经网络在森林步道视觉识别中的应用 [J]. 计算机科学与探索. 2019. 2: 263-274.

[7] Dubos C, Stracke R, Grotewold E, *et al.* MYB transcription factors in *Arabidopsis* [J]. Trends in Plant Science. 2010. 15(10): 573-581.

[8] Bailey PC, Martin C, Toledo-Ortiz G, *et al.* Update on the basic helix-loop-helix transcription factor gene family in *Arabidopsis thaliana* [J]. Plant Cell. 2003. 15(11): 2497-2501.

[9] Stracke R, Werber M, Weisshaar B. The R2R3-MYB gene family in *Arabidopsis thaliana* [J]. Current Opinion in Plant Biology. 2001.

[10] Zhang Y, Wang L. The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants [J]. BMC Evolutionary Biology. 2005.

[11] Jakoby M, Weisshaar B, Droge-Laser W, *et al.* bZIP transcription factors in *Arabidopsis* [J]. Trends in Plant Science. 2002. 7(3): 106-111.

[12] Riano DM, Ruzicic S, Dreyer I, *et al.* PlnTFDB: an integrative plant transcription factor database [J]. BMC Bioinformatics. 2007.

[13] Dai Z, Dai X, Wang J. Identifying Transcription factor binding sites based

on a neural network [C]. Third International Symposium on Neural Networks (ISNN 2006), part III. Berlin: Springer-Verlag, 2006.

[14] Schaffer JD, Riva A, Sichtig H. Evolving spiking neural networks for predicting transcription factor binding sites [C]. The 2010 International Joint Conference on Neural Networks. 2010.

[15] Sugimoto A, Sumi T, Kang J, *et al.* Novel strategy for discrimination of transcription factor binding motifs employing mathematical neural network [J]. Journal of the Physical Society of Japan. 2017. 86(7).

[16] 杨科利,李前忠,林昊. 预测酵母(Yeast)基因转录因子结合位点 [J]. 内蒙古大学学报(自然科学版).2006,(5):46-52.

[17] 刘于江. 人工神经网络与 DAN 序列分类问题 [J]. 赣南师范学院学报. 2004, 25(3): 52-54.

[18] 全国大学生数学建模竞赛[DB/OL].

http://www.mcm.edu.cn/html\_cn/node/616c965be296e1e56d58bb31cb242ac2.html.

[19] 汤丽华. 基于结构数据的转录因子结合位点分析 [D]. 东南大学. 2005.

[20] Berardini TZ, Reiser L, Li D, *et al.* The *Arabidopsis* information resource: making and mining the "gold standard" annotated reference plant genome [J]. Genesis. 2015. doi: 10.1002/dvg.22877.

[21] Yilmaz A, Mejia-Guerra MK, Kurz K, *et al.* AGRIS: *Arabidopsis* gene regulatory information server, an update [J]. Nucleic Acids Res. 2011. doi: 10.1093/nar/gkq1120.

[22] 吕书龙,梁飞豹,刘文丽,等. 应用统计分析与 R 语言实战 [M]. 北京: 北京大学出版社. 2017.

[23] Guenther F, Fritsch S. neuralnet: Training of Neural Networks [J]. R Journal. 2016. doi: 10.1007/BF00201417.

[24] Riedmiller M, Braun H. A direct adaptive method for faster backpropagation learning: The RPROP algorithm [C]. Proc. IEEE. Int. Conf. Neural Networks. 1993.

[25] Riedmiller M. Advanced supervised learning in multi-layer perceptrons from backpropagation to adaptive learning algorithms [J]. Int Journal of Computer Standards & Interfaces. 1994.

[26] Anastasiadis AD, Magoulas GD, Vrahatis MN. New globally convergent training scheme based on the resilient propagation algorithm [M]. Elseview Science Publishers B. V. 2005.

[27] Lopez R, Cowley A, Li W, *et al.* Using EMBL-EBI services via web interface and programmatically via web services [J]. Current protocols in bioinformatics. 2014. (48): 1-50.

[28] Venables WN, Ripley BD. Modern applied statistics with S [M], Fouth edition. Springer. New York. 2002. ISBN 0-387-95457-0.

[29] Fisher RA. The use of multiple measurements in taxonomic problems [J]. Annals of Eugenics. 7, Part II, 1936. 179-188.

[30] Phil Kim. MATLAB deep learning: with machine learning, neural networks and artificial intelligence [M]. Apress. 2017.

[31] Barber CB, Dobkin DP, Huhdanpaa H. The Quickhull algorithm for convex hulls [J]. ACM Transactions on Mathematical Software. 1996. 22(4): 469-483.

[32] 黄萍. 基于遗传神经网络的转录因子结合位点识别方法研究 [D]. 东北师范大学. 2009.

[33] 沈花玉, 王兆霞, 高成耀, 等. BP 神经网络隐含层单元数的确定 [J]. 天津理工大学学报. 2008.

[34] 王嵘冰,徐红艳,李波,等. BP 神经网络隐含层节点数确定方法研究[J]. 计算机技术与发展. 2018.

## 二甲双胍调节细胞自噬和焦亡发挥缺血再灌注脑损伤保护的机制研究

蒋霞<sup>1,2</sup>,袁钰萍<sup>1,2</sup>,邓祖跃<sup>1,2\*</sup>

1. 浙江省食品药品检验研究院, 浙江 杭州 310052310014;

2. 浙江工业大学药学院, 浙江 杭州

摘要:目的:研究二甲双胍调节细胞自噬和焦亡保护脑缺血再灌注损 伤(CIRI)的机制。方法:采用线栓法阻塞 SD 大鼠大脑中动脉, 2h 后再灌注 制备 CIRI 模型。分两组实验, 第一组: 采用三个剂量的二甲双胍(50、100、 200mg kg<sup>-1</sup> d<sup>-1</sup>, ig) 进行二甲双胍药效学和血糖评价。第二组:二甲双胍 对损伤脑组织的细胞自噬和焦亡影响,采用腺苷酸活化蛋白激酶(AMPK) 抑制剂 Compound C(Cc, 20mg kg<sup>-1</sup> d<sup>-1</sup>, ip), 自噬抑制剂三甲基腺嘌呤 (3-MA, 200 mmol /L 2 µL/只, i.c.v), Caspase-1 抑制剂 Z-YVAD-FMK (Z-YVAD, 8µg kg<sup>-1</sup> d<sup>-1</sup>, i.c.v) 分别阻断 AMPK 通路、自噬和细胞焦亡, 考察二甲双胍对这些通路和脑损伤的影响。采用 Zea-Longa 评分评价神经 功能,水迷宫实验考查大鼠记忆情况,TTC 染色检测脑梗死体积,HE 染 色观察受损脑组织病理变化, Western blot 测定受损脑组织 AMPK、 p-AMPK、微管相关蛋白1轻链3(LC3-I、LC3-II)、选择性自噬接头蛋 白(p62)、炎症小体 NLRP3、Pro-caspase-1、半胱氨酸蛋白酶-1(Caspase-1) 的表达。结果: 各剂量的二甲双胍均使 CIRI 大鼠的神经功能评分降低, 记忆力改善,脑梗死体积减小,脑神经元坏死数降低,对血糖没有影响。 二甲双胍预处理能显著激活脑中 AMPK 通路 (AMPK 表达升高) 和自噬活 性(LC3-II/LC3-I升高, p62的表达降低),该效应被 Cc 抑制,提示二 甲双胍所致自噬活性升高由 AMPK 所介导。用 3-MA 可阻断二甲双胍引起 的自噬活性和脑损伤缓解作用,对 AMPK 表达没有影响,说明二甲双胍通 过自噬发挥保护作用。CIRI 大鼠高表达的 Caspase-1 被 Z-YVAD 阻断,二 甲双胍与其联用后进一步加强这个作用。说明二甲双胍通过抑制炎症小体 表达来抑制细胞焦亡。结论:二甲双胍对缺血再灌注脑损伤有保护作用, 其作用机制包括激活 AMPK 通路, 促进细胞自噬, 通过自噬去除损伤的线 粒体,减少了损伤的线粒体引起的 NLRP3 炎症小体活化导致的细胞焦亡, 从而对缺血再灌注脑损伤起到保护作用。

## 关键词:二甲双胍;缺血再灌注脑损伤;自噬;细胞焦亡

[中图分类号] R915 [文献标识码] A [文章编号]

# Mechanism research of metformin protecting brain tissue from ischemia-reperfusion injury by regulating autophagy and pyroptosis

YUAN Yu-ping<sup>1,2</sup>, JIANG Xia<sup>1,2</sup>, LIU Lu<sup>2</sup>, CHEN Shu-huai<sup>2</sup>, DENG Zu-yue<sup>1,2\*</sup>, CHEN Xue-fang<sup>1</sup>

1. College of Pharmaceutical Science,Zhejiang University of Technology, Hangzhou 310014, China

2. Zhejiang Institute for Food and Drug Control, Hangzhou 310052, China;)

Abstract: Objective: To study the mechanism of metformin protecting brain tissue of cerebral ischemia-reperfusion injury (CIRI) by regulating autophagy and pyroptosis. **METHODS:** The CIRI model was prepared by perfusion of blood in the middle cerebral artery after 2 hours occlusion in SD rats. Two groups of experiments were performed. The first group, Metformin and blood glucose evaluation were performed using three doses of metformin (50, 100, 200 kg<sup>-1</sup> d<sup>-1</sup>, ig). In the second group of experiments, the effect of metformin on autophagy and pyroptosis in brain tissue of rats with ischemia-reperfusion injury. using adenosine-activated protein kinase (AMPK) inhibitor Compound C (Cc,  $20 \text{ kg}^{-1} \text{ d}^{-1}$ , ip), autophagy inhibitor trimethyladenine (3-MA, 200 mmol / L 2 µL, i.c.v), Caspase -1 inhibitor Z-YVAD-FMK (Z-YVAD, 8 µg kg<sup>-1</sup> d<sup>-1</sup>, i.c.v) blocked AMPK pathway, autophagy and pyroptosis, respectively, to investigate the effects of metformin on these pathways and brain damage. The Zea-Longa score was used to evaluate the neurological function. The water maze test was used to examine the memory of the rats. The infarct volume was detected by TTC staining. The pathological changes of the damaged brain tissue were observed by HE staining. Western blot was used to detect the expression of AMPK, p-AMPK, LC3-I, LC3-II, p62, NLRP3P, Pro-caspase-1, Caspase-1 of damaged brain tissue. **Results:** Each dose of metformin reduced the neurological function score in CIRI rats, improved memory, decreased cerebral infarction volume, decreased neuronal necrosis, and no effect on blood glucose. Pretreatment metformin

significantly activated AMPK pathway (increased AMPK expression) and autophagy activity (increased LC3-II/LC3-I, decreased expression of p62) in the rats brain, the effect was inhibited by Cc, It showed that metformin elevated autophagy activity by AMPK. 3-MA blocked the autophagic activity and brain damage alleviation induced by metformin, but no effect on the expression of AMPK, this indicated that metformin exerted protective effects through autophagy. The expression of Caspase-1, which are highly expressed in CIRI rats, was blocked by Z-YVAD, and this effect was further added by metformin. It is indicated that metformin inhibits cell death by inhibiting the expression of inflammasome. Conclusion: Metformin has protective effect on cerebral ischemia-reperfusion injury. its mechanism includes activation of the AMPK pathway, promotion of autophagy, removal of damaged mitochondria by autophagy, reduced pyroptosis caused by NLRP3 inflammasome which was activated damaged mitochondria, protection by and against ischemia-reperfusion brain damage.

Keywords: Metformin; Ischemia/reperfusion injury; autophagy; pyroptosis

# "抗疫先锋"事迹

一、个人事迹

1、王峰

无党派人士/主任技师,宁波市李惠利医院东部院区检验科副主任,浙 江省生物信息学学会会员、精准医学专委会委员。曾先后获"浙江省卫生 系统青年岗位能手"、"浙江省青年岗位能手"、"宁波市领军拔尖人才 培养工程第二层次培养对象"等荣誉称号。

医院的检验科对大多数人来讲,总是有些神秘,它和医院其他科室不同,是一个相对独立、封闭的部门。这场突如其来的疫情,因为"核酸检测"这个关键词,把检验科推到了战场的最前线,而核酸检测医生就被赋予了"排雷兵"的名号。

#### 2、侯国清

宁波市第一人民医院 抗疫诊疗专家组成员 浙江省优秀抗疫青年感染 科医师,宁波市战疫先锋。宁波市科技局新冠肺炎疫情防控科技成果示范 应用项目(202002N7033),第三负责人,经费 50万元已发表新冠肺炎相 关 SCI 文章 7 篇,共计影响因子 23.4 分;另有 2 篇文章修回中,其中一篇 为柳叶刀子刊。已发表文章列表:(1)第一作者,A COVID-19 Transmission within a family cluster by presymptomatic infectors in China, CID, IF:9.055; (2)第一作者,Epidemiologic and Clinical Characteristics of 91 Hospitalized Patients with COVID-19 in Zhejiang, China: A retrospective, multi-centre case series,QJM, IF: 2.649;(3)第一作者,Response letter to COIVD 19 Disease: Tackling a pandemic in 21st Century,QJM, IF: 2.649;(4)第一作者,

Response letter to Eosinophil count in severe coronavirus disease 2019 (COVID-19), QJM, IF: 2.649; (5)通讯作者, In-flight Transmission Cluster of COVID-19: A Retrospective Case Series, Infectious Disease, IF:2.494;(6)通

讯作者, COVID-19 with gastrointestinal symptoms as initial manifestations: a case report and literature review, Journal of International Medical Research, IF:1.351. (7) 第一作者, Duration of SARS-CoV-2 viral shedding during COVID-19 infection, Infectious Disease, IF:2.494

#### 3、杨乃彬

中共党员, 医学硕士, 毕业于温州医科大学, 宁波市第一医院感染科 主治医师。共青团宁波市委颁授的"在疫情防控和复工复产一线表现突出 的共青团员"现为浙江省生物信息学会第一届青年委员会委员。作为感染 领域的年轻医者和学者,杨乃彬医师在思想觉悟、一线抗疫、新冠临床研 究中均展现感染青年的不凡担当!鉴于突出表现,他被共青团宁波市委授 予"在疫情防控和复工复产一线表现突出的团员"荣誉称号,并被医院官 微公众号宣传报道。

#### 4、付丽云

中国科学院大学宁波华美医院主任医师 宁波市领军拔尖人才、无惧疫情,迎难而上,奋战在抗击新冠病毒肺炎疫情的第一线,守卫着全市人民的健康。分析总结当地 COVID-19 患者特点,以共同第一作者发表 SCI 收录论文1篇。

"戴口罩、戴帽子、戴手套、隔离服、鞋套……"这样一套繁琐的流程,付医生每天都在重复。作为一名中国共产党党员、传染病学专业的医生,付丽云在抗击新冠肺炎疫情上当仁不让。春节期间,付丽云进入隔离病房,开始了异常忙碌的临床工作,甚至用"连轴转"都不足以形容其中的强度与压力。被收治在隔离病房的新冠患者,因缺少特效药物的治疗,加上对新冠病毒的不了解,心理上或多或少都存在着焦虑,甚至有着恐惧的情绪,因此医生和患者的沟通,尤其是对病人情绪的安抚显得尤为重要。穿上
专业的防护设备,被裹成"粽子"的医护人员连续工作几小时下来,非常容易缺氧,但是对于各项临床工作,付医生依然一丝不苟。特别是在查房时, 她细致地了解每一位患者的心理状态,不厌其烦地一遍遍回答患者的疑问, 缓解他们的焦虑,帮助他们树立战胜病毒的信心。

疫情就是命令,医院就是战场。疫情来临的时候,我们就是战士,不 忘初心,医患同心,就一定能打赢这场战斗!

## 5、陈根浪

中共党员,1978年11月生,工学博士,教授,现任浙大宁波理工学院计算机与数据工程学院副院长、党委委员。

主动担当,深入一线抗疫情。由宁波市政府推广的宁波全域一码通, 即"甬行码",其雏形来自于陈根浪与团队的一次公益行动。针对社区人员 管控复杂,管控时间、人力成本高,社区和居民对便捷管理的迫切需要, 陈根浪与团队成员在线进行方案研讨,在2月2日至4日短短34个小时内, 发挥超强凝聚力战斗力,通过团队共同努力,针对新冠病毒突发疫情的"社 区疫情防控及风险分析系统 1.0"正式推出。

技术推动,科学治理显成效。带领团队一边跑现场调研社区新需求和 疫情防控新要求,一边每天工作 20 多个小时加急优化系统,先后完成 4 个版本升级改造,其中第四版在宁波市公安局牵头组织下升级成为"甬行 码",短短几天注册用户达 600 多万人,截至目前已覆盖 1000 多万人,为 宁波市新冠疫情精智防控作出重要贡献。浙江省科技信息研究院查新报告 表明"社区疫情防控及风险分析系统"是全国最早引入社区疫情信息化防 控的创新应用。产业服务,经济发展增助力。在浙大宁波理工学院、宁波 市公安局、教育局等领导下,"甬行码"为近千万人口大城市的疫情防控和 全面复工复产提供了有力保障,极大缓解了基层防控压力,有效支持了政

府科学决策。团队事迹被人民日报、今日头条、中国教育报、浙江新闻客户端、浙江在线、宁波日报等众多媒体广泛报道,并引起强烈社会反响。

## 6、张顺

1975年生,现任中国科学院大学宁波华美医院(宁波市第二医院)医 学实验部主任、浙江省消化系统肿瘤诊治及研究重点实验室副主任、宁波 市人类生物样本库副主任、宁波市肿瘤分子生物学重点实验室副主任。兼 任中国医药生物技术协会组织生物样本库分会委员、浙江省抗癌协会肿瘤 精准诊治专业委员会委员、浙江省免疫学会肿瘤免疫与生物治疗专业委员 会委员、浙江省数理医学学会新型生物标志物专业委员会常务委员、宁波 市医学会微生物与免疫分会委员、宁波市体外诊断产业技术创新联盟咨询 委员会委员。入选"宁波市领军和拔尖人才工程第二层次培养人员"、获 "宁波市优秀中青年卫生技术人才"、"十二五院士工作站建设先进个人" 等荣誉。张顺从业二十余年,主要从事医学检验、细胞治疗、基因检测、 样本库建设和临床科研等方面的工作。参与科研项目 20 余项,授权专利 7 项,发表论文 48 篇,其中 SCI 论文 20 篇。

2019 新冠肺炎疫情发生后, 医院成为浙江省新冠肺炎定点救治医院和 宁波市新冠肺炎集中收治医院。医学实验部作为拥有负压二级生物安全实 验室的检测部门, 日常即承担着 HIV 确证工作。临近农历新年, 面对未知 而危险的病毒, 张顺主任作为心胸实验支部书记, 深入贯彻落实党中央、 市委和院党委对新型冠状病毒疫情重要指示精神, 充分发挥党组织的战斗 堡垒和先锋模范作用, 临危受命, 守初心, 担使命, 做表率, 组织以党员 为核心, 积极分子为骨干的抗疫小分队, 建立起病毒核酸检测体系, 全力 以赴, 把人民群众生命安全和身体健康放在第一位, 为宁波人民的健康构 筑起一道坚实的防线。

二、会员单位事迹:

## 1. 理事单位: 杭州遂真生物技术有限公司

迅速自主研发了 AIGS 全自动核酸检测分析系统,并于第一时间驰援 武汉各大疫区医院,为武汉前线抗疫贡献自己的一份力量。此套系统能将 被检测样本放在全封闭自动检测仪中独立运行,全封闭体系很好防止了病 原的泄露,有效避免了扩增产物的交叉污染,因此也减少了检验人员暴露 的风险。加班加点、保质保量、争分夺秒以最快速度努力为全国多家检测 单位与医院提供 AIGS 全自动核酸检测分析系统与配套的新型冠状病毒 检测试剂盒,如湖北、浙江、广东、北京等疫情地区。(学会部"防疫" 工作信息第5期已报道)

## 2. 团体会员单位: 杭州厚泽生物

作为首批企业积极加入华大集团组织的百余家企业参与的"抗疫灭毒 联合行动",为抗击新冠病毒肺炎全力以赴!

随着将临床诊断病例数纳入确诊病例数,昨日(2月12日)湖北省新 增确诊14840例,全国新增确诊病例15153例。在NMPA应急审批7款试 剂盒后,科技部也于近日发布相关应急项目申报指南,面向社会广泛征集 新冠病毒的现场快速检测产品试剂盒。厚泽生物一直专注于诊断、基因检 测、蛋白等相关领域,服务于广大医院实验室,医学检验所和疾控、海关、 动物疫控等机构的实验室仪器构建,从全自动核酸提取仪,全自动核酸蛋 白分析仪,全自动移液工作站,荧光定量 PCR 仪到 INB-D200 光感测生物 标志物分析仪,可为实验室提供全套的先进检测设备。









生物样品快速制备系统

超微量紫外可见分光光度计

多波段酶标仪







免疫荧光定量检测仪

.

多功能恒温培养箱



板式移液指示器



高精度移液器系列



核酸提取耗材试剂系列

0-11



杭州遂真生物技术有限公司 专注研发型仪器试剂一体化生产商

地址:杭州市余杭区五常街道高順路8号 西溪软件团金牛座A座4楼 电话:+86-571-89170 228/8873 3231 传真:+86-571-8630 0671



# Qsep系列全自动核酸蛋白分析系统 测序,基因组学中的全方位质控解决方案

#### ☑ Qsep系列产品(不同通量可选)





Qsep 100



#### 🕑 仪器优势

- ★ 仪器灵敏度高:检测低至1pg/μL
- \* 仪器分辨率高:500bp以内可以达到1-4bp分辨率
- \* 检测通量灵活:可检测单个样本不浪费试剂

 $Qsep_1$ -Lite Qsep,

#### < 可检测样本类型

- \* 二代三代测序文库质控、片段化DNA/FFPE/mRNA/MicroRNA片段检测
- \* gDNA、Total RNA完整性质控(DQN/RQN), ct/cf DNA检测,蛋白检测
- ★ 普通PCR产物、高分辨率多重PCR产物检测 (多重病原菌检测、呼吸道病毒检测、致泻性大肠埃希氏菌检测、 结核分枝杆菌VNTR分型、SSR/STR分子标记等)

#### ☑ Qsep系列产品其他相关应用





131 单细胞cDNA质检结果



单细胞测序文库质控

# INB-D200生物标志物检测分析仪



该仪器采用光纤式金纳米颗粒等离子共振感测技 术,免标荧光标记,可直接检测血清、血浆、咽拭子、 肺泡灌洗液、唾液、骨关节滑液等样本,用于监测各 种不同分子(包括但不限于有机药物,寡核苷酸,蛋 白质和病毒)之间的实时相互作用。

#### 

替代ELISA、筛选抗体或核酸适配体、检测抗原抗体亲和力、即时蛋白检 测等,已应用于流感病毒、败血症、骨关节炎症等急性疾病的检测。



杭州厚泽生物科技有限公司 ——— 电话: 0571-81060645 地址: 杭州市西湖区振华路西城博司2幢1101室

邮箱: hz@houzebio.com 网址: www.houzebio.com



tRNA ,5.85











### 

扩增子





7.7 pM (100 pg/mL) 320 pM (4200 pg/mL)

SPR

FFPE









美国 Bio-Rad 分子生物学、蛋白质组学系列设备及耗材;

美国 Thermo Fisehr 超低温、细胞培养、安全柜、离心等细胞培养设备 及耗材;

ACEA 实时无标记细胞功能分析仪、流式细胞仪;

德国 GAS 气相离子迁移谱分析仪、Nicoya 分子相互作用仪、Infors 摇床等

Wiggens 磁力搅拌等实验室常规产品、MD 酶标仪、洗板机、药筛、 高内涵系统、超微量分光光度计、德国美天旎磁珠分选、流式细胞仪; Sysmex 五分类动物血细胞分析仪;

MP FastPrep 样品快速制备系统(核酸提取仪,均质器); Seahorse XF 细胞代谢分析仪。

Gold-SIM 金西盟气相液氮罐、渗透压仪、Telstar 安全柜、冻干系列

杭州宝诚生物技术有限公司

地址: 杭州市滨江区江南大道 143 号江南星座 B座5层

电话: 0571-87224666 Fax: 0571-87895217

维修热线: 400-883-1166



杭州维讯生物科技有限公司是一家专注于动态生物表型组系统研发及表型组图像分析、生物组织培养、技术 推广、生物技术培训、销售和服务的高科技公司。公司注册资本600万元,公司下设开放实验室,对外进行 科研合作和项目研发。全体员工均有高等教育背景,在SCI专业期刊上发表几十篇论文,申报及获批国家专利 若干项,并与德国、澳大利亚等相关研究所建立了紧密的合作关系。

产品服务:生物表型数据分析系统、生物自动化集成设备实验器材、铁皮石斛种苗、播种技术培训、实验试 剂、电子产品的销售和技术服务。

公司落户浙江省杭州市未来科技城(海创园),研发实验室及示范基地在余杭区闲林。



地址: 杭州市文一西路 1500 号 1 号楼 105 室

基地:杭州市佘杭区闲林街道云栖村胡家店 58 号76



箱式生物表型组系统。可实现同时进行植物培养和动态图像采集的功能。 NEW! 优惠服务!!! (在线咨询: 叠280681158)



药用植物组培生产及技术服务。结合智能控制系统和LED光源,采用 Ebb&Flow液体培养技术用于植物组培工厂的大规模产业化。在可控的LED光 源条件下,对温、光、pH、离子强度进行优化,大量繁殖植物胚状体或不定 芽,为本企业和其他企业提供优质原种种苗。

提供铁皮石斛种苗、种植技术培训等。







扫码关注



地址:杭州市文一西路 1500 号 1 号楼 105 室

基地:杭州市余杭区闲林街道云栖村胡家店 58 号



## 基因组学

简化基因组测序 基因组重测序 外显子组测序(遗传病&肿瘤) 遗传图谱 BSA性状定位/Graded-seq DNA甲基化芯片 GWAS、ChIP-seq SNP芯片

# 转录组学

转录组测序 绝对定量转录组测序 三代全长转录组测序 单细胞SMART-seq 外泌体RNA测序 降解组测序 基因表达谱芯片 全转录组测序

# 表观组学

全基因组甲基化测序 简化甲基化测序 miRNA芯片/测序 m<sup>6</sup>A mRNA/IncRNA测序 IncRNA芯片/测序 circRNA芯片/测序 ceRNA芯片 RIP-seq

## (代谢组学)

+

代谢全谱分析 脂质代谢组分析 靶向代谢组分析

# 蛋白质组学

iTRAQ定量蛋白检测 TMT标记定量蛋白检测 Label free蛋白定量检测 磷酸化定量蛋白分析 DIA蛋白定量检测

# 微生物组学

宏基因组测序 宏转录组测序 细菌基因组重测序 微生物多样性(16S/ITS) 细菌基因组完成图测序 功能基因组测序

# 专注基因技术的发明与创造,提供一流的服务与产品

● 助力2200⁺篇科研论文发表

# 10X Genomics

0X单细胞空间转录组测序 10X单细胞免疫组库测序 10X单细胞转录组测序 10X单细胞ATAC测序

Ribosome-Profiling

业务覆盖国内外数千家科研机构、医院及 药企集团,全力支持科学探索







杭州联川生物技术股份有限公司电话:0571-87662448网址:www.lc-bio.comLC-Bio Technologies (HangZhou) Co.,Ltd地址:杭州经济技术开发区6号大街260号中自科技园16幢4层





浙江省生物信息学学会是在遵守宪法、法律、法规和国家政策及社会道德风尚的前提下,团结、联合、组织浙江省生物信息 学、生命科学和计算机科学及其他相关专业人士,自愿组成并依法登记的学术性、公益性、非营利性法人社团。学会坚持学术民主 和开放的原则,团结广大生物学及计算机工作者,积极地开展相关学术活动,为促进科学技术的繁荣和发展、科学知识的普及和推 广、科技人才的成长和提高,以及提升我省科技创新能力做出贡献。

根据《社会团体登记管理条例》和有关规定,学会筹备组于2012年9月向省科协学会部提出筹建申请,2013年2月得到批复同 意。2013年3月向省民政厅提出筹建申请,2013年5月获得省民政厅同意筹备的批复。学会于2013年9月21日在杭州召开了全体会员 大会。会议审议并通过了学会章程、财务管理制度和会费收取标准,并选举产生了第一届学会理事会、常务理事和领导成员:浙江 大学陈铭教授为理事长,浙江工业大学丁维龙教授、浙江理工大学贺平安教授、宁波大学刘箴教授、杭州师范大学向太和教授、浙 江大学樊龙江教授、中国水稻所郭龙彪研究员和杭州博日生物科技有限公司副总经理赵怀为副理事长,浙江大学徐程副教授为秘书 长。

学会挂靠在浙江大学,秘书处与办公室设在浙江大学紫金港校区生命科学学院236。学会的业务主管单位是浙江省科学技术协会。

随着人工智能与大数据时代的到来,以及基因组、转录组、蛋白质组、 代谢组等各组学生物信息技术的发展,医疗人工智能、个体化医学、转化 医学,已成为生物信息领域的重要发展趋势。本学会目前学会拥有会员 300 余人,企业会员 11 家、下设 5 个委员会:

精准医学专业委员会(宁波大学医学院)、青年委员会(宁波大学医 学院)、生物信息学与人工智能专业委员会(浙江工业大学计算机学院)、 微生物组学专业委员会(浙江天科高新技术发展有限公司)、农林生物信 息学专业委员会(浙江农林大学)

欢迎新成员加入我们大家庭!



http://www.zjbioinformatics.org/
关注公众号:
浙江省生物信息学学会