



会议手册

CONFERENCE MANUAL

# 第三届生物医学数据挖掘与计算学术会议

在线会议 2021年11月26日-28日



# 目录

会议简介.....	1
会议组织机构.....	2
在线会议须知.....	3
腾讯会议使用说明.....	4
一、下载安装.....	4
二、登录注册&加入会议.....	4
三、会议控制.....	5
演讲者相关操作.....	6
在线会议简要日程.....	7
具体日程安排.....	8
大会报告嘉宾.....	11
分会场报告嘉宾.....	14
分组报告 1:深度学习前沿与生物信息分析.....	14
分组报告 2:精准医疗与健康大数据.....	17
分组报告 3:医学影像及类脑智能.....	20
分组报告 4:生物网络与系统生物学.....	23
分组报告 5:基因组数据分析与合成生物学.....	26
分组报告 6:蛋白质生物信息学.....	29
分组报告 7:组学数据分析.....	32
分组报告 8:人工智能前沿与生物信息学.....	35

## 会议简介

第三届生物医学数据挖掘与计算学术会议原定于2021年11月26日-28日在安徽合肥召开。由于疫情防控形势严峻，考虑到参会者的健康和安全，且为了避免潜在的问题和减少所有人的风险，会议组织委员会决定将此前预通知的线下会议方式调整为线上会议。

本次会议的主题为“深度学习与生物医学大数据”，旨在为深度学习、人工智能、机器学习、模式识别、生物信息学、计算生物学和医学信息学等领域的专家学者提供一个跨学科交流的平台，共同推动相关领域的发展与进步。

本次会议由中国生物信息学会生物医学数据挖掘与计算专业组（筹）与安徽省人工智能学会主办，安徽大学承办。会议将邀请国内智能计算与生物信息学等领域著名专家做大会特邀报告，并分多个主题对相关领域的最新研究进展和动向进行学术报告交流，欢迎业内专家学者参会交流讨论（人数不限）。

## 会议组织机构

**主办方:**中国生物信息学会生物医学数据挖掘与计算专业组（筹）、安徽省人工智能学会

**承办方:** 安徽大学

**大会主席:** 黄德双

**程序委员会主席:** 朱大铭、刘滨、苏延森

**程序委员会委员（按拼音排序）:** 蔡云鹏、陈伯林、陈伟、杜朴风、冯东雷、冯好娣、冯佳、冯伟兴、高军晖、龚乐君、郭菲、胡静、胡秀珍、黄德双、黄宏斌、江振然、姜伟、蒋庆华、蒋雪玲、孔韦韦、雷秀娟、李国亮、李兰婷、李敏、李政伟、林昊、林岩、刘滨、刘丙强、刘洪波、刘健、刘金星、刘琦、刘文斌、刘小平、刘学军、刘志宏、卢新国、陆慧娟、路永钢、罗飞、骆嘉伟、马小科、缪嘉嘉、彭玮、蒲莲容、乔少杰、邱春燕、邱望仁、苏延森、孙小强、汤斌华、唐华、涂仕奎、汪国华、王栋、王飞、王树林、王文辉、魏乐义、吴宏杰、夏俊峰、谢江、谢民主、邢立宁、熊毅、徐烽涛、徐磊、徐清华、许鹏、薛志东、业宁、叶明全、叶远浓、尤著宏、曾涛、张瀚、张力烨、张晓龙、张岩、张玉森、章乐、章文、赵方庆、赵琪、郑春厚、郑小琪、周雄辉、朱大铭、朱丽达、朱山凤、祝小雷、邹权

**组织委员会主席:** 郑春厚

**组织委员会委员（按拼音排序）:** 宾艳南、曹瑞芬、程娜、李晓艳、谭大禹、魏丕静、夏俊峰、岳振宇、郑春厚

## 在线会议须知

- 会议平台：腾讯会议
- 在线会议教室

11月 26 日		
测试以及备用会议室 腾讯会议ID: 515 8754 4686		
11月 27 日		
主会场 腾讯会议ID: 416 108 806	分会场 1 腾讯会议ID: 695 514 385	分会场 2 腾讯会议ID: 408 828 966
11 月 28 日		
分会场 1 腾讯会议ID: 619 820 596	分会场 2 腾讯会议ID: 544 657 400	

- 在线测试会议

请需要在线作报告的主讲嘉宾和分会场嘉宾务必进入测试会议室进行会前音视频测试和 PPT 共享调试；其他参会者也可以选择进入测试会议室来调试音视频以及举手提问。

每人测试时间不超过10分钟，如果在11月26日无法完成测试，可以在27日或28日进入备用会议室进行测试。

- 正式在线会议

大会报告的持续时间为50分钟，其中包括45分钟口头报告及5分钟问答；专题报告的持续时间为25分钟，其中包括20分钟的口头报告及5分钟的问答。

请报告嘉宾在各分会开始前至少提前10分钟进入对应的在线会议室并将演示文件放在电脑桌面并打开。

## 腾讯会议使用说明

### 一、下载安装

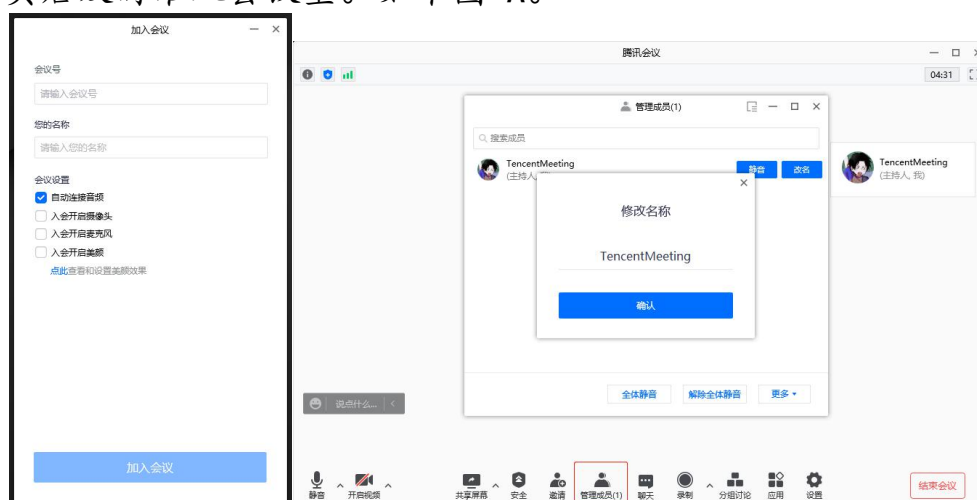
下载地址：<https://meeting.qq.com/download-center.html?from=1001>  
下载安装后请更新到最新版本，避免进入不到会议室。

移动设备使用支持 iOS、Android及微信小程序，请在移动端应用商店下载安装。友情提醒：经过测试，微信小程序有一定的不稳定性，为了保证会议质量，建议选择使用安装客户端的方式使用“腾讯会议”。

### 二、登录注册&加入会议

新用户可以用手机号发送验证码进行注册，也可以通过微信直接注册登录。

点击加入会议，输入对应“会议号”（即会议 ID）就可以加入该会议；同时请您修改备注名为“KN（主讲嘉宾）/SC(分场报告主持人)+真实姓名”，听众修改备注名为“LIS+真实姓名”，以便工作人员核实后及时准入会议室。如下图 A。

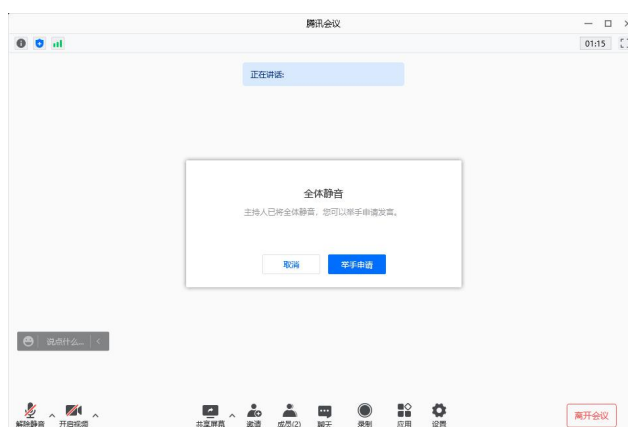


图A

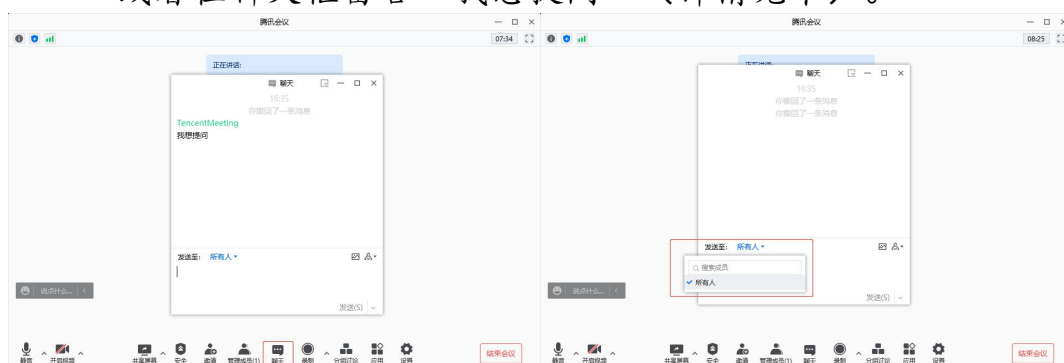
图B

您加入会议后，如果您想要变更当前展示的昵称，可以点击下方操作栏中的“成员/管理成员”，鼠标悬浮至您的姓名，点击“改名”，输入新昵称即可；如果您操作不方便，也可以与主持人沟通，让主持人帮您修改。如上图 B。

### 三、会议控制



- 主持人默认设置“全体静音”且不允许成员自行解除静音，您点击“解除静音”时，则会向主持人“举手申请”，请求解除静音，待主持人同意后，才会为您打开麦克风。
- “举手申请”一般适用于当您向演讲人提问时，请举手申请“解除静音”。
- 或者在聊天框留言“我想提问”（详情见下）。



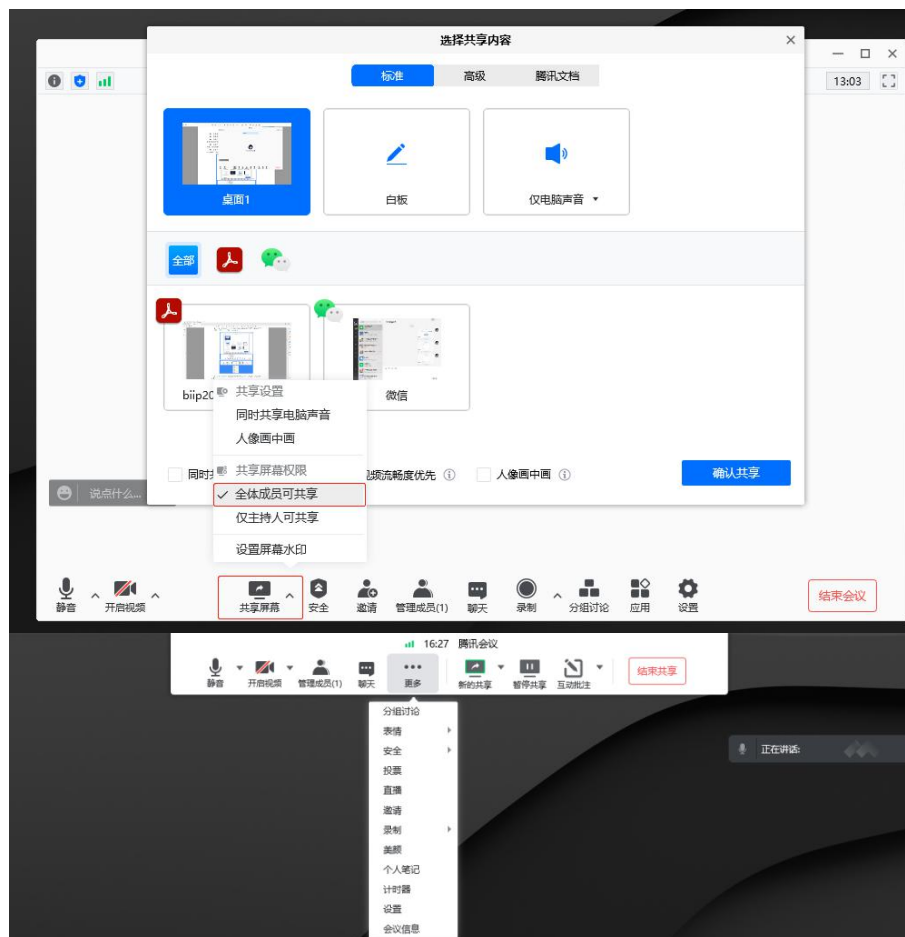
图C

图D

- 在会议中，点击“聊天”，即可唤出聊天窗口。
- 您可更换聊天模式，选择将信息私下发送给某位成员或所有人。
- 当您输入“我要提问”时，请务必留意须发送至“所有人”。

## 演讲者相关操作

- 点击会议中底部工具栏的“共享屏幕”按钮；
- 选择您希望共享的窗口或桌面；
- 勾选“同时共享电脑声音”（选填，勾选后即可在共享屏幕内容的同时，共享电脑的声音，当前仅支持 Windows）；
- 勾选“视频流畅度优先”（选填，勾选后，系统将为您减少其他性能的消耗，优先保障共享视频流畅度）；
- 点击“共享屏幕”按钮确定发起共享；选择要分享的 PPT（提前打开）——选择“分享”——选择“结束分享”。



- 当您共享屏幕后，屏幕共享菜单将会在3s后进入沉浸模式，自动隐藏在顶部；
- 您需要将鼠标放置桌面上方即可将其唤出。



## 在线会议简要日程

11月26日 测试会议					
10:00-17:00	11月26日 测试以及备用会议室 腾讯会议 ID: 515 8754 4686				
11月27日					
08:30-09:00	开幕及嘉宾致辞				11 月 27 日 主会场腾讯会议 ID: 416 108 806
09:00-11:30	大会报告				
11:30-14:30	休息				
14:30-15:45	分组报告 1 深度学习前沿与生物信息分析  主持人：苏延森	11 月 27 日 分会场1 腾讯会议ID: 695 514 385	14:30-15:45	分组报告 2 精准医疗与健康大数据  主持人：夏俊峰	11 月 27 日 分会场2 腾讯会议ID: 408 828 966
15:45-15:55	休息				
15:55-17:10	分组报告 3 医学影像及类脑智能  主持人：刘滨	11 月 27 日 分会场1 腾讯会议ID: 695 514 385	15:55-17:10	分组报告 4 生物网络与系统生物学  主持人：尤著宏	11 月 27 日 分会场2 腾讯会议ID: 408 828 966
11月28日					
09:00-10:15	分组报告 5 基因组数据分析与合成生物学  主持人：张法	11 月 28 日 分会场 1 腾讯会议ID: 619 820 596	09:00-10:15	分组报告 6 蛋白质生物信息学  主持人：熊毅	11 月 28 日 分会场2 腾讯会议ID: 544 657 400
10:15-10:25	休息				
10:25-11:40	分组报告 7 组学数据分析  主持人：朱大铭	11 月 28 日 分会场 1 腾讯会议ID: 619 820 596	10:25-11:40	分组报告 8 人工智能前沿与生物信息学  主持人：汪国华	11 月 28 日 分会场2 腾讯会议ID: 544 657 400

## 具体日程安排

11 月 26 日 测试会议		
10:00-12:00	在线测试	测试以及备用会议室 腾讯会议 ID: 515 8754 4686
12:00-14:00	休息	
14:00-17:00	在线测试	
2021 年 11 月 27 日 上午		
开幕及嘉宾致辞（主持人：郑春厚）		11 月 27 日 主会场 腾讯会议 ID: 416 108 806
08:30-08:40	安徽大学校领导致辞	
08:40-08:50	中国生物信息学学会候任理事长、清华大学孙之荣教授致辞	
08:50-09:00	大会主席黄德双教授致辞	
大会报告		
09:00-09:50	大会报告1：施蕴渝，主持人：黄德双	
09:50-10:40	大会报告2：李 霞，主持人：黄德双	
10:40-11:30	大会报告3：黄德双，主持人：郑春厚	
2021 年 11 月 27 日 下午		
分组报告 1:深度学习前沿与生物信息分析 主持人：苏延森		11 月 27 日 分会场 1 腾讯会议 ID: 695 514 385
报告时间	报告人信息	报告题目
14:30-14:55	章文，华中农业大学	The application of graph neural networks in bioinformatics
14:55-15:20	曾湘祥，湖南大学	面向药物发现的深度图学习
15:20-15:45	郭菲，中南大学	生物序列功能效应的混合神经网络方法研究
15:45-15:55	休息	
分组报告 2:精准医疗与健康大数据 主持人：夏俊峰		11 月 27 日 分会场 2 腾讯会议 ID: 408 828 966
报告时间	报告人信息	报告题目
14:30-14:55	林昊，电子科技大学	融合多种指标的糖尿病风险预测

## 2021年生物医学数据挖掘与计算学术会议

14:55-15:20	朱山凤, 复旦大学	HPOLabeler and HPOFiller: recent advances in the prediction of human protein-phenotype associations
15:20-15:45	董磊, 北京理工大学	Identification of driver genes in glioma with different PTEN status
15:45-15:55	休息	
<b>分组报告 3: 医学影像及类脑智能</b> 主持人: 刘滨		11 月 27 日 分会场 1 腾讯会议 ID: 695 514 385
报告时间	报告人信息	报告题目
15:55-16:20	蔡宏民, 华南理工大学	基于流形学习的脑网络分析方法研究
16:20-16:45	王珊珊, 中国科学院深圳先进技术研究院	智能快速医学成像与分析
16:45-17:10	张法, 中国科学院计算技术研究所	基于病理图像和基因数据融合的癌症生存期预测
<b>分组报告 4: 生物网络与系统生物学</b> 主持人: 尤著宏		11 月 27 日 分会场 2 腾讯会议 ID: 408 828 966
报告时间	报告人信息	报告题目
15:55-16:20	余国先, 山东大学	异质生物分子网络建模、融合与应用研究
16:20-16:45	李春权, 南华大学	复杂疾病的生物网络与通路的模型构建及分析
16:45-17:10	卢新国, 湖南大学	面向泛癌数据分析的癌症驱动基因挖掘方法
2021 年 11 月 28 日 上午		
<b>分组报告 5: 基因组数据分析与合成生物学</b> 主持人: 张法		11 月 28 日 分会场 1 腾讯会议 ID: 619 820 596
报告时间	报告人信息	报告题目
09:00-09:25	汪国华, 东北林业大学	基因组分析的生物信息学算法
09:25-09:50	朱大铭, 山东大学	基因组数据分析中的组合优化问题
09:50-10:15	汪小我, 清华大学	人工智能助力合成生物设计与构建
10:15-10:25	休息	

# 2021年生物医学数据挖掘与计算学术会议

<b>分组报告 6:蛋白质生物信息学</b> 主持人: 熊毅		11 月 28 日 分会场 2 腾讯会议 ID: 544 657 400
报告时间	报告人信息	报告题目
09:00-09:25	刘滨, 北京理工大学	基于自然语言处理的蛋白质序列分析
09:25-09:50	李敏, 中南大学	Computational methods and tools to predict protein functions and interactions
09:50-10:15	杨建益, 山东大学	trRosetta and its recent development for protein structure prediction
10:15-10:25	休息	
<b>分组报告 7:组学数据分析</b> 主持人: 朱大铭		11 月 28 日 分会场 1 腾讯会议 ID: 619 820 596
报告时间	报告人信息	报告题目
10:25-10:50	赵兴明, 复旦大学	Exploring human gut microbiome across the lifespan
10:50-11:15	张治华, 中国科学院北京基因所	Chromatin 3D architectures revealed by single cell Hi-C and ATAC-seq data
11:15-11:40	蔡禄, 内蒙古科技大学	A genome-wide analysis of the gene expression and alternative splicing events in a whole-body hypoxic preconditioning mouse model
<b>分组报告 8:人工智能前沿与生物信息学</b> 主持人: 汪国华		11 月 28 日 分会场 2 腾讯会议 ID: 544 657 400
报告时间	报告人信息	报告题目
10:25-10:50	薛宇, 华中科技大学	Artificial intelligence biology: a new paradigm in biology
10:50-11:15	刘琦, 同济大学	生物医药数据的度量、嵌入、迁移和联邦
11:15-11:40	李国亮, 华中农业大学	TAD boundary and strength prediction by integrating sequence and epigenetic profile

## 大会报告嘉宾



施蕴渝

中国科学技术大学生命科学学院

报告题目：人工智能与生物医药

**个人简介：**施蕴渝，生物物理学与结构生物学家。1997年当选为中国科学院院士。2009年当选为发展中国家科学院院士。中国科学技术大学生命科学学院教授。目前担任中国生物物理学会常务理事。中科院生物物理所生物大分子国家重点实验室学术委员会主任。1965年毕业于中国科学技术大学物理系生物物理专业。1965年-1970年卫生部中医研究院实习研究员。1970年-1978年中国科学技术大学助教，1978年-1984年中国科学技术大学讲师，1984年-1988年中国科学技术大学副教授，1988年-至今中国科学技术大学教授。合肥微尺度物质科学国家实验室（筹）研究员。1979年-1981年意大利罗马大学化学系，CNRS结构化学实验室进修。曾作为访问学者在荷兰格罗宁根大学物理化学系，法国CNRS酶学与结构生物学实验室，法国理论化学实验室进修或合作研究。

20世纪80年代开始从事生物大分子计算机分子动力学模拟，是国内此领域的开创者之一。她开展了与蛋白质分子设计及药物设计有关的基础理论与方法学研究。用分子动力学模拟加热力学积分方法模拟酶与底物，药物与受体的结合自由能；研究和发展了蛋白质及其周围溶剂的静电相互作用的理论模型和数值模拟方法；研究将分子动力学模拟与量子化学模拟结合模拟溶液中的有机反应及模拟酶反应。20世纪90年代中期之后主要从事生物大分子核磁共振波谱研究。她是国内生物大分子核磁共振波谱研究的先行者之一，她领导的研究组解析了系列重要蛋白质的溶液结构，研究揭示蛋白质相互作用及其功能意义。她在中国科学技术大学领导建立了中国科学院结构生物学重点实验室。领导了国家自然科学基金创新人才研究团队，培养了一批在计算生物学与结构生物学这些交叉学科领域工作的年轻学科带头人。

## 李霞

哈尔滨医科大学生物信息科学与技术学院

报告题目：生物医学大数据挖掘与精准医学



**个人简介：**李霞，博士、博士生导师、二级教授、国务院特殊津贴获得者、中国细胞生物学会功能基因组信息学与系统生物学分会会长、“头雁”团队带头人、“龙江学者”特聘教授、哈尔滨医科大学生物信息科学与技术学院院长，省级优秀中青年专家、省级领军人才梯队带头人、北京“百千万人才工程”入选者、五洲女子科技奖获得者、省生物医学工程学(生物信息学)一级学科重点学科带头人。主持承担了国家863、973、重大研发计划和国家自然科学基金等16项，提出了疾病生物医学组学大数据挖掘系列创新方法与技术，开发了多功能生物医学大数据系列分析平台，在国外著名生命科学杂志《Nature Communication》，《Cancer Research》，《Trends in Biochemical Science》，《Nucleic Acids Research》等上发表SCI论文200余篇，累计SCI影响因子2000余点，获教育部奖、省部级奖、中华医学奖等10余项。率先创建了国内一流且规模最大的生物医学信息学创新型人才培养团队，创办了我国第一个生物信息学本科专业，主编了第一部全国高等学校临床医学专业八年制卫生部规划教材《生物信息学》（第一、第二版、第三版主编），为生物医学信息学的学科发展和人才培养做出重要贡献。



## 黄德双

同济大学电子与信息工程学院

报告题目：Advances for Motifs Mining in  
Biological Binding Sequences

**个人简介：**黄德双，博士，同济大学教授、博士生导师，中国科技大学兼职教授、博士生导师，IEEE Fellow，国际模式识别学会(IAPR) Fellow，2000年度中科院“百人计划”入选者，中国计算机学会生物信息学专业委员会副主任委员。长期从事神经网络、模式识别与生物信息学方面的研究，在国内外等学术期刊上发表了超过230篇SCI论文，H因子73；曾荣获教育部和安徽省自然科学一等奖各1项、人工智能学会科技进步一等奖1项；担任国家科技创新 2030—新一代人工智能重大项目“面向复杂数据处理的新型神经网络模型研究”项目首席专家；担任期刊IEEE/ACM Transactions on Computational Biology & Bioinformatics等杂志编委。



## 分会场报告嘉宾

### 分组报告 1:深度学习前沿与生物信息分析



章文

华中农业大学信息学院

报告题目: The application of graph neural networks in bioinformatics

**摘要:** Graph neural network (GNN) is a series of emerging methods that apply deep learning on graphs. Recently, GNN has lots of successful applications in different fields: from social networks to biological networks. We review the recent advances of GNNs in bioinformatics, and present several applications of GNNs for link prediction problems, such as miRNA-disease associations and drug-cell line responses. At last, we summarize the future directions of GNN in bioinformatics.

**个人简介:** 华中农业大学信息学院教授, 博士生导师。中国计算机学会 (CCF) 高级会员, CCF YOCSEF武汉副主席, 中国计算机学会生物信息学专委会委员, 中国人工智能学会生物信息学与人工生命专委会常务委员, 湖北省生物信息学会理事。2003年在武汉大学数学与统计学院获得学士学位; 2006年在武汉大学数学与统计学院获得硕士学位; 2009年在武汉大学计算机学院获得博士学位。2007年9月到2008年8月, 获国家留学基金委资助, 在新加坡国立大学计算机学院开展博士生联合培养。2009年9月-2012年11月, 武汉大学计算机学院讲师; 2012年12月-2018年10月, 武汉大学计算机学院副教授、珞珈青年学者; 2018年11月-至今, 任职于华中农业大学信息学院。长期从事人工智能算法及其在生物医学大数据分析方面的研究, 成果发表于本领域权威期刊 PLOS Computational Biology, Bioinformatics, Briefings in Bioinformatics, IEEE/ACM transactions on computational biology and bioinformatics, Nucleic acids research, Information Science等和人工智能顶会IJCAI等。先后主持国家自然科学基金青年项目、面上项目和多个省部级科研项目。



## 曾湘祥

湖南大学信息科学与工程学院

报告题目：面向药物发现的深度图学习



**摘要：** Deep graph learning (graph neural network) has received great attention from artificial intelligence researchers in recent years. This report will introduce the latest progress of deep graph learning, and its applications in drug repurposing. Specifically, we built a comprehensive knowledge graph that includes 15 million edges across 39 types of relationships connecting drugs, diseases, proteins/genes, pathways, and expression from a large scientific corpus of 24 million PubMed publications. Using Amazon's AWS computing resources and a network-based, deep-learning framework, we identified 41 repurposable drugs (including dexamethasone, indomethacin, niclosamide, and toremifene) whose therapeutic associations with COVID-19 were validated by transcriptomic and proteomics data in SARS-CoV-2-infected human cells and data from ongoing clinical trials.

**个人简介：** 湖南大学岳麓学者特聘教授，博士生导师，国家优青，湖南省杰青，获2019年吴文俊人工智能优秀青年奖。2020年亚马逊机器学习研究奖，成果“面向药物发现的深度学习”入选2020世界人工智能大会卓越引领者奖(SAIL奖)榜单。曾获厦门大学教学成果奖特等奖，福建省教学成果奖二等奖。主持了科技部重点研发课题、国基金、以及来自华为、百度等公司的多个项目。



## 郭菲

中南大学计算机学院

报告题目：生物序列功能效应的混合神经网络方法研究

**摘要：**深度学习技术在处理大量生物医学数据的功能效应分析等方面具有快速精准的识别能力，并在各种生物信息学应用中取得了巨大成功。首先我们提出一个双通道深度神经网络（DeepAVP），可用于建模抗病毒肽的抗病毒性来源。通过将递归神经网络和卷积神经网络并列，然后处理原始的可变长度的肽链序列数据，从中提取序列和进化特征，鉴定抗病毒肽。该模型的重要特点是可以处理原始的变长肽链序列数据，将特征提取嵌入到神经网络模型中，可以在训练模型的同时对特征提取进行优化。另外我们提出了一个混合深度神经网络（DeepATT），用于建模DNA功能变异来源。设计了一个类别注意力层和一个类别全连接层，区分不同DNA功能并选择相应的有效特征，提供了对非编码基因组区域的新颖见解，有助于理解复杂的基因表达和发现性状相关遗传性变异的来源。我们将混合神经网络等深度学习技术应用于生物医学研究和疾病致病机理分析，致力于为重大疾病复杂机制解析提供研究范例。

**个人简介：**中南大学计算机学院教授、博士生导师。近年来主要研究生物信息学，突破海量生物数据的复杂关联关系和隐含生物特征的数据挖掘技术，建立高效精准的智能计算模型，实现生物分子识别及功能分析。作为负责人主持多项国家自然科学基金和国家重点研发计划子课题。围绕课题相关领域产出了一系列科研成果，发表相关学术论文90余篇，第一作者或通讯作者论文60余篇，包括生物信息学顶级期刊Bioinformatics、Briefings in Bioinformatics等，总引用2000余次，ESI高被引论文6篇。入选天津市“131”创新型人才培养工程（第三层次）；担任国际SCI期刊Current Proteomics编委。

## 分组报告 2:精准医疗与健康大数据

林昊

电子科技大学信息生物学中心

报告题目：融合多种指标的糖尿病风险预测



**摘要：**糖尿病目前被认为是一种全球性的慢性病。血糖长期过高可导致各种组织的慢性损伤，因此，早期发现和干预糖尿病是预防糖尿病或延缓糖尿病并发症发生的关键。然而，由于个人经济状况的限制和对自身健康状况了解有限，目前只有一半的糖尿病患者被确诊。随着体检数据的大量增长和人工智能技术的快速发展，研究人员希望利用体检数据建立糖尿病风险评估为临床指导和早期大规模筛查提供潜在的帮助。我们开发的糖尿病风险评估系统可以从不同层次级联的评估糖尿病风险。其中，糖尿病风险评估模型用于系统层面的筛查，模型可直接应用于各地区体检信息化决策系统，在每年的体检普查中根据风险评分提前筛查出糖尿病高风险人群，进行预防管控，做到糖尿病早发现、早预防、早治疗，从而降低国家和各地区医保成本；糖尿病风险评分卡将模型中各风险因素量化，可直接应用于临床，方便高风险人群掌握自身健康状况，制定合理的生活计划。随访满意度模型用于挖掘糖尿病人对病情控制满意度的关键因素，可用于糖尿病人的病情管理。此外，我们还研究了糖尿病人代谢产物，并基于代谢产物对糖尿病、糖尿病前期、正常人群进行了区分，为糖尿病的预防提供新思路。

**个人简介：**林昊教授长期从事生物医学信息学领域的研究。面向生命医学科学前沿，针对生物医学数据噪音高、特征提取不充分这一关键问题，开发了多种特征提取和优化算法，在生物关键大分子数据的整理、序列表示，临床数据分析与疾病风险预测领域取得若干研究成果。以第一作者或通讯作者在Information Fusion、Nucleic Acids Research、Clinical and Translational Medicine、Briefings in Bioinformatics、Bioinformatics等期刊上发表论文160余篇。迄今，有37篇论文引用超过100次，单篇最高引用达531次，超过40篇论文入选ESI高被引论文；构建生物信息在线服务52个(见<http://lin-group.cn>)，有90余个国家和地区的1.7万个独立IP超过3万次的访问。曾入选“爱思唯尔2020中国高被引学者”，斯坦福大学发布的“全球前2%顶尖科学家榜单”，四川省“杰出青年科技人才”，“Clarivate全球高被引学者”。获得国家自然科学基金面上项目、青年项目、重点项目的资助。



## 朱山凤

复旦大学类脑智能科学与技术研究院

报告题目：HPOLabeler and HPOFiller:  
Recent Advances in the Prediction of Human  
Protein-Phenotype associations

**摘要：** Exploring the relationship between human proteins and abnormal phenotypes is of great importance in the prevention, diagnosis and treatment of diseases. The human phenotype ontology (HPO) is a standardized vocabulary that describes the phenotype abnormalities encountered in human diseases. We have developed HPOLabeler for predicting HPO annotations of novel proteins and HPOFiller for identifying missing HPO annotations. HPOLabeler is able to use a wide variety of evidence, such as protein-protein interaction (PPI) networks, Gene Ontology, InterPro, trigram frequency and HPO term frequency, in the framework of learning to rank (LTR). We empirically evaluate HPOLabeler extensively through mainly two experiments of cross validation and temporal validation, for which HPOLabeler significantly outperformed all component models and competing methods including the current state-of-the-art method. In addition, according to the preliminary results of CAFA4 reported in ISMB2020 (July 2020), HPOLabeler achieved the first place in terms of F-max. On the other hand, the current HPO annotations of proteins are incomplete. We propose HPOFiller, a graph convolutional network (GCN)-based approach, for predicting missing HPO annotations. Empirical results of extremely stringent evaluation avoiding potential information leakage including cross-validation and temporal validation demonstrates that HPOFiller significantly outperforms all other state-of-the-art methods. The further examination offers literature evidence for highly ranked predictions. Finally using known disease-HPO term associations, HPOFiller could suggest promising, unknown disease-gene associations, presenting possible genetic causes of human disorders.

**个人简介：** 复旦大学类脑人工智能科学与技术研究院研究员，博士生导师。香港城市大学博士，日本京都大学博士后，美国伊利诺伊大学香槟分校访问学者。UniProt 国际科学顾问委员会委员，中国计算机学会生物信息专业委员会创始委员，中国人工智能学会生物信息与人工生命专业委员会创始委员、中国中文信息处理学会医疗健康与生物信息处理专业委员会创始委员，中国细胞生物学会生物信息与系统生物学分会理事，中国运筹学会计算系统生物学会分会理事。主持或完成四项国家自然科学基金项目，以及多个国内外企业研发项目。主要研究方向为人工智能与生物医学大数据挖掘，特别是生物医学文本挖掘、蛋白功能预测、宏基因组、药物发现、免疫信息学等。相关论文在生物信息、人工智能、数据挖掘等顶级国际会议和期刊发表，如 NeurIPS, KDD, ISMB, IJCAI, Bioinformatics, Nucleic Acids Research等。2014年-2021 年参加 BioASQ 大规模生物医学文本自动标注国际竞赛中取得七次第一名的好成绩。2017 年参加 CAFA 大规模蛋白功能自动标注国际竞赛，在全世界 50多个实验室中获得第一名。



## 董磊

北京理工大学生命学院

报告题目：Identification of driver genes in glioma with different PTEN status



**摘要：** Glioblastoma (GBM) is highly invasive and the deadliest brain tumor in adults. It is characterized by inter-tumor and intra-tumor heterogeneity, short patient survival, and lack of effective treatment. Prognosis and therapy selection is driven by molecular data from gene transcription, genetic alterations and DNA methylation. The four GBM molecular subtypes are proneural, neural, classical, and mesenchymal. More effective personalized therapy heavily depends on higher resolution molecular subtype signatures, combined with gene therapy, immunotherapy and organoid technology. We summarize the principal GBM molecular classifications that guide diagnosis, prognosis, and therapeutic recommendations.

**个人简介：** 医学博士，北京理工大学生命学院教授、博士生导师、副院长。第14批国家级青年人才。曾在美国埃默里大学、凯斯西储大学、阿肯色大学等科研院校学习和工作十年。从事幼年型粒-单细胞白血病（JMML）、实体肿瘤GBM发生发展的分子机制和相关药物开发的研究。研究领域涉及磷酸酶的结构与活化、及其功能调控、白血病和实体肿瘤的发生、发展机理等。主要研究内容是基因突变导致血液肿瘤的致病机制，以及针对肿瘤靶向蛋白的药物的开发和治疗。同行专家在Nature等杂志中发专题评论其“在骨髓的非血细胞中表达白血病相关的遗传突变足以在小鼠中引起白血病。而且，这一发现是治疗常发性致死性疾病的重大进展”。课题组以肿瘤、干细胞等研究为方向，交叉并集成医学与工科学科并利用北京理工大学传统生命学科优质资源，同时运用干细胞、高通量测序、生物化学和生物信息学等前沿技术，进行以下研究：

- 1.以突变磷酸酶细胞或动物为模型，解析调控细胞命运维持和转变（包括干细胞干性维持和分化）的分子机制；
- 2.通过遗传学、基因编辑/转录调控等手段，探究肿瘤发育的调控机制，及疾病发生机理；
- 3.通过化合物筛选、免疫疗法等技术，开发针对重大疾病的药物疗法和细胞疗法；
- 4.新发呼吸道病原体感染人体类器官芯片的开发和应用：针对新发呼吸道和消化道病原体防控的重大现实需求，开发和建立用于研究病原体感染和传播机制、重要靶点以及药物筛选的人体类器官芯片平台。a.开发和建立基于微流控技术的类器官芯片平台。b.研究新发病原体感染途径和机制。c.通过类器官模型进行药物筛选。

### 分组报告 3: 医学影像及类脑智能



蔡宏民

华南理工大学计算机科学与工程学院

报告题目：基于流形学习的脑网络分析方法研究

**摘要：**越来越多的研究发现与神经退行性疾病相关的病变脑区域不是随机出现的，其空间分布遵循大尺度的脑网络连接。在这种情况下，开发强大且通用的脑网络分析方法对了解整个大脑的神经病理传播机制至关重要。但是，脑疾病的研究与其他疾病的研究有所不同，我们难以获得大量的大脑组织进行深入的疾病机理探究。多模态脑网络的整合分析，能够非侵入式地全面分析复杂脑疾病的演变规律，反映大脑功能的运作，在应用上为脑疾病非侵入式早期诊断，重要生物标志物发现和个性化治疗等提供基础分析方法。在这种背景下，我们关注结构脑网络和功能脑网络的分析方法。在结构脑网络分析方面，我们提出了基于流形学习的全局谐波基识别方法和基于流形学习的局部谐波基识别方法，发现了一种新的脑网络标志物，为脑疾病的早期诊断提供新的思路。在分析功能脑网络时，我们提出了基于流形学习的大脑状态变化检测方法和基于流形学习的大脑意识状态解码方法，发现了新的与大脑状态相关联的几何模式，为理解脑网络功能动态提供了新的几何见解。

**个人简介：**华南理工大学计算机科学与工程学院教授、博士生导师。1997-2003年获得哈尔滨工业大学本科和硕士学位，2007年获得香港大学数学系博士学位。2016年破格晋升博士生导师，同年破格晋升教授。系京都大学客座教授(2012, 2019)，中国自动化学会生物信息学与人工生命专业委员会委员、常委会委员，中国计算机学会生物信息学专业委员会委员、常委委员；系多个国际会议如ISBI, ISBRA, ICIC, BIBM, GIW, MICCAI, BIBM, AAAI的PC Member；系国际会议如ICDKE 2012, ICBBB2021, ICBBB 2022的会议主席；系广东省转化医学眼科分会副主委、广东省精准医学应用学会-数字智能化分会副主任委员、广东省生物医学工程学会智能医学影像分会副主任委员。他长期参与生物医学图像和生物信息处理方面的研究工作，在医学图像分析与理解、生物信息分析、多源数据融合、模式识别和数据挖掘等领域积累了丰富的研究经验，在相关杂志如IEEE T-PAMI, IEEE T-CyberN, IEEE T-Image Proce., IEEE T-Medical Imaging, Neuroimage, Bioinformatics, Briefings in Bio. 累计发表论文100+篇，近5年发表论文IF总和：225.244，最高IF 16.389，申请和授权专利21项，建立了基因组学到医学影像全链条分析的若干关键理论和应用算法。

## 王珊珊

中国科学院深圳先进技术研究院

报告题目：智能快速医学成像与分析



**摘要：**针对磁共振数据采集慢、图像重建迭代久、临床诊断依赖人工等问题，讲者将从机器学习的基础理论出发，围绕“成像模型—重建算法—分析技术—转化应用”全链条，汇报其近几年在理论模型与算法、功能应用加速上开展的研究。例如，提出了联合稀疏自适应编码理论，克服并行MRI成像对灵敏度校正的依赖；国际公认首创复数卷积多源大数据快速MRI并行成像新框架，被国际同行评价为有潜力改革成像科学；构造类网络迭代细节优化成像新方法，实现高分辨细节成像，被医学物理著名期刊评为“年度亮点”工作；并且提出多种医学影像自动定量分析新方法。针对深度学习的全监督依赖问题，于Nature communication发表一个标注高效的深度学习开源框架，Annotation-efficient Deep Learning (AIDE)，具体提出一种跨模型标签自我校正机制，可以有效利用低质量的标签数据。这些标签或由预训练的低性能深度学习模型生成，或由无法保证质量的单个标注者提供。AIDE可用于处理具有稀缺或噪声标注的数据集，以统一框架处理医学图像分析常见的半监督学习、无监督域适应和噪声标签学习任务。

**个人简介：**双博士，博士生导师。研究方向机器学习、快速医学成像、放射组学等，迄今为止在Nature communications、IEEE Trans TMI/TBME/TIP、MICCAI等顶级期刊与会议发表英文学术论文50多篇，ESI 高被引2篇，PMB年度亮点文章一篇，发明授权专利9项，4项实现规模化产业应用。快速医学成像获广东省科技发明一等奖，影像组学定量分析工作获广东省科技进步一等奖。并获2018海外华人磁共振协会OCSMRM杰出研究奖（Outstanding Research Award），2020吴文俊人工智能优秀青年奖，2021年深圳市青年科技奖，澳大利亚John Make peace Bennett最佳博士论文提名奖，IEEE Senior member, OCSMRM BoT/Life member, Gordon Plenary Lecturer, NIBIB New Horizontal Lecturer. 为国际医学磁共振年会2018-2020基于机器/深度学习磁共振成像与分析的分会主席，ISMRM/ISICDM/MIDL/MICCAI workshop session/area chair, 曾4次获得ISMRM.E.KZavoisky奖金，为国际SCI期刊Magnetic resonance in medicine（SCI, JCR-1, Deputy editor）、Pattern Recognition (SCI, JCR-1, Associate Editor)和Biomedical Signal Processing and Control (SCI, JCR-2, Associate editor)编委；曾获IEEE-TMI、MRM和signal processing审稿杰出贡献奖。先后主持国家自然科学基金面上青年及省重点市级项目12项，曾受邀到世界顶级大学及会议如美国哈佛大学、加拿大魁北克大学、Gordon conference、ISMRM及SMRA等给大会或教育讲座。



## 张法

中国科学院计算技术研究所

报告题目：基于病理图像和基因数据融合的  
癌症生存期预测

**摘要：**生存分析是评估癌症治疗方案的关键。基于深度学习的方法整合病理图像和基因组数据已被用于预后预测。然而，现有方法大多是基于对病理图像块的分析，忽略了更大视场的形态结构信息和图像块之间的内在关系。同时，现有模型未能充分利用神经网络强大的表征学习能力，对病理图像和基因组数据无法进行有效的多模态特征融合。针对这一问题，本报告将介绍一种基于Transformer的融合网络，能够有效整合病理图像和基因组数据，有助于患者的预后预测。在TCGA的结直肠癌数据集上的实验结果验证了我们方法的有效性。

**个人简介：**博士，中国科学院计算技术研究所研究员，博士生导师。毕业于中国科学院计算技术研究所。现任中国计算机学会生物信息学专委会秘书长，中国生物物理学会冷冻电镜分会理事。主要从事生物信息学、生物医学图像处理和高性能计算等方面的研究工作，近年来在生物医学图像处理、影像基因组学、生物医学数据整合等方面取得了多项重要研究成果：研发了国内首款冷冻电镜三维重构的软件-AuTOM；结合人工智能技术开发了系列医学病理图像分类分级处理技术，相关研究成果已在多家科研机构 and 医院应用。率队获得2021年国际三维模型检索挑战赛（SHREC2021）冷冻电镜电子断层生物图像分类大赛全球第一名的成绩。作为项目负责人和主要参与人承担了多项科技部重点研发专项、国家自然科学基金重点和国际合作重大项目、中科院先导、中科院知识创新重点等项目。在Cell Research、Sciences Advances、Bioinformatics和SIAM Journal on Scientific Computing等国际著名期刊和ISMB、ICCV等顶级国际会议发表论文140余篇。



## 分组报告 4: 生物网络与系统生物学

余国先

山东大学软件学院

报告题目：异质生物分子网络建模、融合与应用研究



**摘要：**异质生物分子网络可以综合描述和建模多属性、弱关联、跨组学分子数据，在疾病关联分析和分子功能预测等领域应用广泛。现有方法在数据整合的粒度和自适应性等方面还存在较多局限性，亟需有效的智能自适应整合分析方法。本报告将介绍我们基于属性异质分子网络协同整合多组学数据预测lncRNA-disease关联和可变剪接异构体功能等方面的研究工作。

**个人简介：**山东大学软件学院教授，博士生导师，主要从事机器学习，数据挖掘及其在生物医学数据分析中的应用研究。在国内外重要会议和期刊(KDD, IJCAI, TKDE, TNNLS, Bioinformatics, TCBB, 中国科学-信息科学等)发表论文80余篇。主持(完成)国家自然科学基金3项(面上、应急管理、青年)，联合主持国基重点1项，担任IEEE/ACM Transactions 系列, Bioinformatics等多个国内外著名期刊审稿人。CCF生物信息学专委会委员, 人工智能与模式识别专委会委员；中国人工智能学会生物信息学与人工生命专委会委员, 机器学习专委会委员。



## 李春权

南华大学附属第一医院

报告题目：复杂疾病的生物网络与通路的模型构建及分析

**摘要：**细胞是由基因、蛋白质、代谢物等各种分子组成的整体。构成人体细胞复杂系统的代谢与信号通路、基因转录调控、蛋白质互作关系均可抽象成生物分子为节点的网络。分子与分子之间的相互作用关系所形成的生物分子网络构成了细胞生命运作系统，对人体生命系统的运作至关重要。尤其，越来越多的研究显示，癌症、心脑血管等复杂疾病通常不归因于单个基因突变或功能失调，而是由网络的功能异常所造成，属于“网络疾病”。尽管基于分子生物学的研究已经取得了重大进展，但并不足以解决人类复杂生命过程中和复杂疾病中出现的越来越多的问题。因此，突破单个基因和蛋白的研究束缚，构建全面的、精准的、完善的生物分子网络、模型和分析方法，对探究人体复杂生命系统的运转及复杂疾病预测、诊断、预后评估和治疗具有重要意义和价值。十几年以来，我们始终专注于生物分子网络模型构建、分析及基于网络模型的复杂疾病研究。通过结合生物信息学、计算系统生物学等技术方法，进行代谢与信号传导通路、基因转录调控网络的有向网络模型构建、数据库和软件平台开发，以及网络分析方法研究。尤其，近年来我们突破转录调控网络研究的瓶颈，建立了超级增强子、增强子、开放染色质介导的转录调控网络。进一步，开发了贯穿信号通路和转录调控网络的有向调控网络分析方法和工具。我们开发了SubpathwayMiner、SEanalysis、KnockTF、SEdb、TRlnc、ATACdb、LncSEA等多款医学大数据平台和分析软件，已广泛被十个国家的研究机构使用。这些研究对于更加精细的建立和理解生物网络，揭示生命系统的网络运作，以及复杂疾病的网络系统研究具有一定的价值。

**个人简介：**南华大学附属第一医院教授，博士研究生导师。研究领域为复杂疾病的生物信息学与大数据分析。曾获得省杰出青年科学基金，南华大学“青年创新”高层次人才和于维汉院士杰出青年培养基金。学术兼职包括中国计算机学会生物信息学专业委员会委员和中国抗癌协会肿瘤标志专业委员会肿瘤测序及大数据分析专家委员会委员等。多篇论文发表在《Nucleic Acids Research》、《Briefing in bioinformatics》、《Bioinformatics》和《Nat Commun》等高影响力杂志上，总影响因子>300。其中，IF>15的论文12篇，IF>10的论文15篇，中科院一区论文16篇。主持国家自然科学基金面上项目（2项）、国家自然科学基金青年基金项目和教育部长江学者青年项目等课题。研究成果分别获得省自然科学技术学术成果奖一等奖（排名第1）和中华医学科技奖三等奖（排名第5）。

## 卢新国

湖南大学信息科学与工程学院

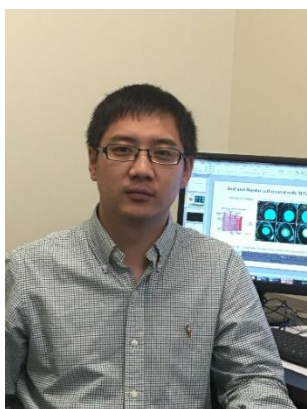
报告题目:面向泛癌数据分析的癌症驱动基因挖掘方法



**摘要:** 癌症是一种复杂的基因疾病,驱动突变为癌细胞提供了选择性优势,促进了癌症的发生。识别癌症驱动基因对研究癌症产生、发展和诊疗具有关键作用。由于癌症的高度异质性是由基因组、转录组、蛋白质组、表观遗传组等多组学共同决定。同时,泛癌数据是一种多类型、多平台和多层次的数据,这些数据具有异构、相关和互补等特性。在本报告中,将介绍我们在癌症驱动基因挖掘方面的最新研究成果,包括:(1)基于不同分子层次水平的组学数据融合方法、网络融合方法和癌症基因识别方法(2)结合复杂网络融合和矩阵分解方法提高药物和靶标关联预测问题(3)体细胞突变驱动挖掘方法及泛癌数据分析的初步研究成果。

**个人简介:** 博士,湖南大学信息科学与工程学院副教授,博士生导师。毕业于湖南大学信息科学与工程学院。中国计算机学会生物信息专委会委员,生物医学数据挖掘与计算专委会筹委会委员。主要从事多组学数据融合、肿瘤驱动基因识别、泛癌数据挖掘等研究工作。作为项目负责人和主要参与人承担了国家重点研发计划课题,国家社会科学基金重点项目、湖南省科技重大专项等国家级、省部级项目20余项。其中主持国家自然科学基金2项,教育部博士点基金、中国博士后基金、国家留学基金委留学基金等。在Bioinformatics, IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)等生物信息重要期刊上发表论文30余篇。主要研究成果曾荣获湖南省自然科学优秀学术论文三等奖。多次担任各种国际学术会议程序委员会委员,担任学术期刊Bioinformatics、TCBB、BMC Bioinformatics、BIBM、软件学报等重要会议与期刊论文评审人。

## 分组报告 5:基因组数据分析与合成生物学



汪国华

东北林业大学信息与计算机工程学院

报告题目：基因组分析的生物信息学算法

**摘要：**自 2001 年人类基因组计划完成以来，生命科学领域进入了基因组时代，利用大数据技术分析基因组数据和验证生物学新发现成为了生命科学领域崭新的发展模式。获得高质量基因组，分析整个基因组的功能和结构已经成为基因组学的核心问题，生命科学创新与发现的瓶颈已经由数据的产生转变为对数据的分析和利用。汪国华教授近十年一直致力于基因组学数据分析算法的研究，将主要介绍课题组在k-mer计数方法和基因组倍性估计、基于长读序列的基因拼接、结构变异检测和基因组相似性计算等多个方面的一系列进展。

**个人简介：**东北林业大学信息与计算机工程学院院长，教授，博士生导师。2009年获得哈尔滨工业大学计算机应用技术专业博士学位。2009年起在哈尔滨工业大学计算机科学与技术学院历任副教授，教授，博士生导师。2019年调入东北林业大学信息与计算机工程学院任院长，林木遗传育种国家重点实验室PI。2013年入选教育部“新世纪优秀人才支持计划”，约翰霍普金斯大学博士后。中国计算机学会生物信息专委会常务委员，人工智能学会生物信息学与人工生命专委会委员。主要研究方向为生物信息学、人工智能。任BMC Genomics期刊编委，已经在Nature Review Genetics, Nature protocols, Nature Communications, Nucleic Acids Research, Bioinformatics等期刊发表SCI检索国际期刊论文50余篇。作为负责人主持2项国家863, 4项国家自然科学基金项目。2018年，2019年十三五国家重点研发计划“数字诊疗装备研发”重点专项会议评审专家。

## 朱大铭

山东大学计算机科学与技术学院

报告题目：基因组数据分析中的组合优化问题



**摘要：**介绍基因组重组排序，基因组最长公共样本子序列等问题的建模背景，和人们针对组合问题所设计的求解算法，以及这些问题的计算复杂性研究进展。基因组重组排序在进化生物学研究中发挥着重要作用。

**个人简介：**朱大铭，1999年在中国科学院计算技术研究所获得博士学位，目前为山东大学计算机学院教授，博士生导师。长期从事生物信息学/计算生物学，算法与计算复杂性研究。擅长基因组数据分析的组合优化问题建模，基因组数据分析组合问题的精确算法和近似算法设计。发表学术论文150余篇，曾经获得教育部和山东省的科技奖励4次。





## 汪小我

清华大学自动化系

报告题目：人工智能助力合成生物设计与构建

**摘要：**基因编辑与DNA合成技术的快速发展，使得我们有可能按照功能需求从底层来重新设计和构建合成生物系统。如何开发与之匹配的定量模型和人工智能方法，理性设计满足性能需求的生物元件，预测合成基因回路的功能特性，是合成生物学领域亟待解决的瓶颈问题。我们近期尝试用利用计算仿真和机器学习来优化设计全新的基因调控元件。我们将人工智能对抗学习与合成生物学实验闭环耦合，采用数据与知识双驱动的策略解析基因调控复杂特性和编码模式，进而用机器学习方法设计获得全新的生物元件，完成从物理-虚拟-物理世界的迭代映射过程，从实践上探索了用人工智能设计基因元件的可行性。机器学习和定量模型有望帮助大幅提高实验筛选效率，并辅助理解基因调控的复杂规律，为推动合成生物系统更加高效可控的设计与构建奠定基础。

**个人简介：**清华大学长聘教授。于2003年和2008年在清华大学获学士和博士学位，并曾赴美国冷泉港实验室和加州大学伯克利分校访问学习，2008年起在清华大学自动化系任教至今。主要从事人工智能与生物医学的交叉研究，包括基因组大数据分析、合成生物智能设计等。曾担任中国生物工程学会青工委主任、中国人工智能学会生物信息学专委会副主任，并获全国百篇优博、国自然优青、教育部新世纪人才，中国自动化学会青年科学家奖等。

## 分组报告 6:蛋白质生物信息学

刘滨

北京理工大学计算机学院

报告题目：基于自然语言处理的蛋白质序列分析



**摘要：**蛋白质序列分析是生物信息学中的核心问题之一。生物序列和自然语言之间存在着相似性，例如在核糖体中，氨基酸组成具有特定结构和功能的蛋白质。与之相似，字词通过语法规则组成具有特定语义的句子。因此可以借鉴自然语言处理中，通过对句子进行句法分析获得其语义的研究思路，通过分析蛋白质序列预测其结构和功能。本报告将介绍研究组近年来在相关领域的研究进展。最后将对该未来发展方向进行展望和分析。

**个人简介：**北京理工大学教授、博士生导师。担任生物信息学学会生物医学数据挖掘与计算专业组（筹）副主任、中国计算机学会生物信息学专委会常务委员、中国人工智能学会生物信息学与人工生命专委会委员、中国自动化学会智能健康与生物信息专委会首届委员、中国运筹学会计算系统生物学分会常务理事。针对生物分子结构和功能精准识别难题，提出基于自然语言处理的生物序列分析算法，在Bioinformatics、Nucleic Acids Research、Briefings in Bioinformatics等期刊发表SCI论文100余篇。所开发的生物序列分析平台BIG累计被访问140万次。曾获爱思唯尔高被引科学家（2021）、北京杰青（2019）、国家优青（2018）、霍英东青年教师基金（2018）、广东省杰青（2016）、广东特支计划青年拔尖人才（2016）和深圳市青年科技奖（2016）。



李敏

中南大学计算机学院

报告题目：Computational methods and tools to predict protein functions and interactions

**摘要：** Proteins perform functions in various biological processes by interacting with proteins, DNA and RNA. They rarely act alone as their functions tend to be regulated. Identification of protein interactions can help understand how a protein performs its biological functions and design new antibacterial drugs. This talk will introduce some computational methods and tools for predicting protein functions, identifying protein-protein interaction sites, and discovering DNA, RNA and protein binding with intrinsically disordered regions (IDRs).

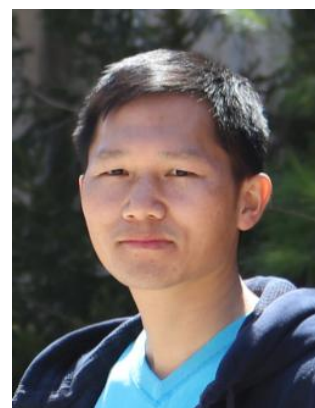
**个人简介：** 中南大学计算机学院教授、博士生导师、副院长，主要从事生物信息学与数据挖掘研究，主持国家自然科学基金重点项目、优秀青年项目、面上等国家和省部级科研课题十余项。担任湖南省人工智能与医学大数据国际联合实验室主任、生物信息学湖南省重点实验室副主任、中国计算机学会生物信息学专委会委员、中国人工智能学会-生物信息学与人工生命专业委员会常务委员等学术职务，担任《Current Protein & Peptide Science》、《International Journal of Data Mining and Bioinformatics》、《Interdisciplinary Sciences: Computational Life Sciences》等期刊编委，担任 ISBRA2019、ICPCSEE2017等国际会议的程序委员会主席。



## 杨建益

山东大学数学与交叉科学研究中心

报告题目: trRosetta and its recent development  
for protein structure prediction



**摘要:** trRosetta is a two-step approach [1,2] to protein structure prediction. Inter-residue geometries, including distance and orientations, are first predicted with the network ResNet, which are then converted into restraints to guide the energy minimization in Rosetta. trRosetta was shown to outperform other methods including AlphaFold1. With the inclusion of a couple of new features, we designed trRosettaX [3], an enhanced version of trRosetta. A new multi-scale network named Res2Net was used in trRosettaX to improve the prediction of inter-residue geometries. In addition, homologous templates are used automatically included into the network to increase the accuracy further. trRosettaX-based approach was ranked as one of the top server groups in CASP14's blind test.

**个人简介:** 2011年博士毕业于新加坡南洋理工大学, 2015-2021任职于南开大学, 2021年7月加入山东大学数学与交叉科学研究中心。主要研究蛋白质和RNA结构与功能预测算法, 开发了多个国际知名结构预测算法, 包括trRosetta和I-TASSER等, 在Nature Methods、Nature Protocols、PNAS、Advanced Science等期刊发表SCI论文50多篇。论文被SCI他引6000多次, 其中1篇论文入选ESI热点论文, 5篇论文入选ESI高被引论文。

## 分组报告 7:组学数据分析

赵兴明

复旦大学类脑智能科学与技术研究院

报告题目: Exploring human gut microbiome  
across the lifespan



**摘要:** Metagenomic sequencing is the main way to explore the microbiome, whereas the assembly of metagenome is the key for downstream analysis. In this talk, I'll present our recent work on identifying misassembled contigs and downstream binning of contigs, and the enterotypes of gut mycobiome across the human lifespan. I'll also show our work on the association between gut microbiome and autism.

**个人简介:** 复旦大学类脑智能科学与技术研究院教授, 计算神经科学与类脑智能教育部重点实验室副主任。目前主要从事人工智能与生物医学大数据交叉领域的研究, 上海市青年科技启明星和上海市浦江人才计划入选者。他是IEEE Senior Member、IEEE SMC Technical Committee on Systems Biology共同主席、ACM SIGBIO China副主席、IEEE SMC Shanghai Chapter主席、IAPR TC20委员, 并担任中国人工智能学会生物信息学与人工生命专委会等国内学术组织的主任、常务理事和委员, 担任了IEEE/ACM TCBB和IEEE SMC Magazine等国际期刊的副主编。承担了863、国家自然科学基金重大研究计划和重点项目、国家重点研发计划在内的多项科研课题, 在Cell Metabolism、Nature Communications、Genome Medicine、IEEE TPAMI和Nucleic Acids Research等国际期刊发表SCI论文100余篇, 曾获教育部高校优秀成果(自然科学)二等奖和吴文俊自然科学一等奖。



## 张治华

中国科学院北京基因组研究所  
(国家生物信息中心)

报告题目: Chromatin 3D Architectures Revealed by  
single cell Hi-C and ATAC-seq Data

**摘要:** The human genome has a dynamic, well-organized hierarchical 3D architecture, from arrangement of nucleosomes, megabase-sized topologically associating domains (TAD), and all the way to the chromosome territories. In the present work, we introduce two tools on predict nucleosome arrangement and TAD structures from single cell ATAC-seq and Hi-C data, respectively.

TADs are a key structure of the genome regulating nuclear processes, such as gene expression, DNA replication and damage repair. However, owing to a lack of proper computational tools, TADs have still not been systematically and reliably surveyed in single cells. We developed a new algorithm to decode TAD boundaries that keep chromatin interaction insulated (deTOKI) from ultra-sparse Hi-C data. By nonnegative matrix factorization, this novel algorithm seeks out for regions that insulate the genome into blocks with minimal chance of clustering. We found that deTOKI outperformed competing tools and that it reliably identified TADs with single-cell Hi-C (scHi-C) data. By applying deTOKI, we found that domain structures are prevalent in single cells. Further, although domain structures are highly dynamic between cells, TADs adhere to the ensemble, suggesting tight regulation of single-cell TADs.

Nucleosomes arrangement is at the base layer of chromatin hierarchical organization. However, it dynamic at single cell level remains unclear. We present a newly developed tool named deNOPA to predicted nucleosome positions with single cell ATAC-seq data. We showed that deNOPA outperformed over the stat-of-the-art tools. Last, we applied deNOPA on the heat shock response system of human cells, which is a classic gene regulation model, and found a nonlinear relationship between the nucleosome occupancy strength changes and the gene transcription regulation.

**个人简介:** Dr. ZHANG Zhihua is a principal investigator at Beijing Institute of Genomics (BIG), Chinese Academy of Sciences (CAS). ZHANG lab's research interests is in understanding the gene regulation in mammals. Particularly, focusing on

1. 3D Genome. We develop new algorithms and technologies to understand the mechanism and dynamic of genome architecture in 3D space. By decoding the 3D genome, we shall be better understanding the dynamic of gene regulatory network with all those genes and distal regulatory elements.

2. Artificial intelligence and biomedical big data. We are focusing on using artificial intelligence to deep mining biomedical big data to understand the relationship between genetic or epigenetic variations and complex diseases, such as cancer.

## 蔡禄

内蒙古科技大学生命科学与技术学院

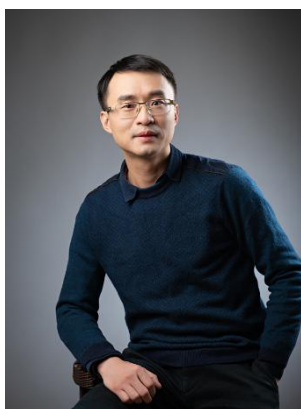
报告题目：A Genome-Wide Analysis of the Gene Expression and Alternative Splicing Events in a Whole-Body Hypoxic Preconditioning Mouse Model



**摘要：** Exposure to specific doses of hypoxia can trigger endogenous neuroprotective and neuroplastic mechanisms of the central nervous system. These molecular mechanisms, together referred to as hypoxic preconditioning (HPC), remain poorly understood. In the present study, we applied RNA sequencing and bioinformatics analyses to study HPC in a whole-body HPC mouse model. The preconditioned (H4) and control (H0) groups showed 605 differentially expressed genes (DEGs), of which 263 were upregulated and 342 were downregulated. Gene Ontology enrichment analysis indicated that these DEGs were enriched in several biological processes, including metabolic stress and angiogenesis. The Kyoto Encyclopedia of Genes and Genomes enrichment analysis showed that the FOXO and Notch signaling pathways were involved in hypoxic tolerance and protection during HPC. Furthermore, 117 differential alternative splicing events (DASEs) were identified, with exon skipping being the dominant one (48.51%). Repeated exposure to systemic hypoxia promoted skipping of exon 7 in *Edrf1* and exon 9 or 13 in *Lrrc45*. This study expands the understanding of the endogenous protective mechanisms of HPC and the DASEs that occur during HPC.

**个人简介：** 内蒙古科技大学生命科学与技术学院教授，博士生导师。内蒙古自治区“功能基因组生物信息学重点实验室”主任，“表观遗传学与生物信息学科技创新团队”负责人。开展生物信息学、表观遗传学等领域研究。主要关注基因表达在表观遗传学层面的调控机制，开展核小体定位、组蛋白修饰、RNA可变剪接、常见病转录组学机制等问题研究。获国家自然科学奖三等奖1项，内蒙古自治区自然科学奖一等奖1项，自治区科技进步奖一等奖、二等奖各1项，获自治区优秀教学成果奖一等奖2项。主持国家自然科学基金项目6项，发表SCI 收录50余篇。出版《表观遗传学前沿》学术专著1部，《生物信息学》教材3部。中国细胞生物学学会功能基因组信息学与系统生物学分会委员、中国生物工程学会生物信息学与计算生物学分会委员。

## 分组报告 8:人工智能前沿与生物信息学



薛宇

华中科技大学生命科学与技术学院

报告题目: Artificial intelligence biology: a new paradigm in biology

**摘要:** Recent advances in deep learning have provided a great opportunity to accurately infer the complex causality from big biological data, and boomed the establishment of a new interdisciplinary field named artificial intelligence biology (AIBIO). Here, we reported our recent progresses in AIBIO for PTM and COVID-19. First, we developed a new pipeline for computationally integrating circadian multi-omics data (iCMod) to identify 789 (~17%) normalized circadian p-sites (NCPs), and predicted 27 potential circadian kinases to participate in phosphorylating these sites, including 7 previously known to function in the clock. We screened the remaining 20 kinases and found an additional 3 to be involved in shaping the temporal pattern of the entire circadian molecular landscapes. Using the plasma proteomic data of COVID-19 patients, we developed a machine learning-based method named POC-19 to identify 11 proteins as biomarkers and a set of biomarker combinations, which were validated by independent cohorts and accurately distinguished and predicted COVID-19 outcomes, including non-survivors and survivors recovered from mild or severe symptoms. In addition, we constructed an open resource named iCTCF that contained chest computed tomography (CT) images, 130 clinical features and laboratory-confirmed clinical status. We show the utility of the database in the prediction of COVID-19 morbidity and mortality outcomes via a deep-learning algorithm named HUST-19 trained with data from 1170 patients and 19,685 manually labelled CT slices. In an independent validation cohort of 351 patients, the algorithm discriminated negative, mild and severe cases with areas under the receiving operating characteristic curve of, respectively, 0.944, 0.860 and 0.884. The open database may find further uses in the diagnosis and management of COVID-19 patients. Taken together, we anticipate that AIBIO can be a new paradigm in biology, and provide powerful methods for uncovering new biological findings.

**个人简介:** 薛宇, 华中科技大学生命科学与技术学院教授、人工智能生物学中心主任、湖北省“人工智能生物学”创新群体负责人。主要研究方向为蛋白质组信息学。近5年以唯一或共同通讯作者身份在Immunity、Nature Biomedical Engineering和Nature Communications等期刊上发表SCI论文25篇, 其中IF>10的论文8篇。先后入选教育部“新世纪优秀人才支持计划”、中组部“万人计划”青年拔尖人才和教育部“长江学者奖励计划”青年学者。2018年至今任湖北省生物信息学会秘书长。





## 刘琦

同济大学生命科学与技术学院

报告题目：生物医药数据的度量、嵌入、  
迁移和联邦

**摘要：**本报告主要分享课题组在应用人工智能技术进行生物医药数据分析的若干研究工作。具体关注于生物医药数据的度量、嵌入、迁移和联邦四大类AI技术。希望抛砖引玉，对于生物医药数据的分析和复杂疾病的精准治疗提供AI技术层面的一定启发和指导。

**个人简介：**同济大学生命科学与技术学院生物信息系长聘教授 (Tenured Professor)，博士生导师。长期致力于发展人工智能和生物组学交叉融合的研究范式，面向肿瘤精准用药、药物发现、肿瘤免疫治疗及基因编辑领域开发新的计算方法学和计算平台，推进人工智能和组学数据分析的结合及临床应用。在Nature Communications (2021, 2019, 2015), Science Advances (2020), Genome Biology (2018), Genome Medicine (2019), Chemical Science (2021), Nucleic Acids Research (2021, 2020) 等发表学术论文，受邀在Trends in Molecular Medicine (2019), Trends in Biotechnology (2016), Trends in Pharmacological Science (2017), WIREs Computational Molecular Science (2018) 等发表综述和评述论文。主持国家863基金及国家自然科学基金，参与国家重点研发计划精准医学及慢病专项等。任ELSEVIER出版社人工智能生命科学交叉领域杂志 Artificial Intelligence in the Life Sciences编委。其研究团队和国际制药企业、CRO公司以及互联网公司开展紧密合作，推动人工智能技术在生物医学领域的应用。其在人工智能和生物组学交叉领域的研究工作先后入选2019年、2020年中国生物信息学研究十大进展，曾获吴文俊人工智能自然科学技术奖三等奖、药明康德生命化学奖、华夏医学奖三等奖，中国发明协会发明创新奖二等奖等。先后入选上海市浦江人才，启明星人才，曙光人才，入选教育部“青年长江学者”。



李国亮

华中农业大学信息学院

报告题目：TAD boundary and strength prediction  
by integrating sequence and epigenetic profile

**摘要：** Topologically associated domains (TADs) are one of the important higher-order chromatin structures with various sizes in the eukaryotic genomes. TAD boundaries, as the flanking regions between adjacent domains, can restrict the interactions of regulatory elements, including enhancers and promoters, and are generally dynamic and variable in different cells. However, the influence of sequence and epigenetic profile-based features in the identification of TAD boundaries is largely unknown. In this work, we proposed a method called **pTADS** (prediction of TAD boundary and strength), to predict TAD boundaries and boundary strength across multiple cell lines with DNA sequence and epigenetic profile information. The performance was assessed in seven cell lines and three TAD calling methods. The results demonstrate that the TAD boundary can be well predicted by the selected shared features across multiple cell lines. Especially, the model can be transferable to predict the TAD boundary from one cell line to other cell lines. The boundary strength can be characterized by boundary score with good performance. The predicted TAD boundary and TAD boundary strength are further confirmed by three Hi-C contact matrix-based methods across multiple cell lines.

**个人简介：** 新加坡国立大学计算机专业博士，华中农业大学信息学院教授、副院长，中国计算机学会杰出会员，三维基因组学分析平台ChIA-PET Tool的开发者。以（共同）第一作者、通讯作者在Cell、Nature Genetics、Science Advances、Genome Biology、Nucleic Acids Research等杂志上发表论文30多篇，共发表SCI文章80多篇。主要研究方向为：生物信息学，生物大数据，三维基因组学，表观遗传学。