第七届“数学、计算机与生命科学交叉研究”

青年学者论坛

2019年5月18-19日，北京，中国科学院数学与系统科学研究院南楼一层报告厅

会 议 手 册

主办单位：

协办单位： NCMIS ORS

赞助单位： BerryGenomics 贝瑞和康 果壳生物 BIOGUOKE GPB Novogene 诺禾致源 PARATERA 并行

会议介绍

“数学、计算机与生命科学交叉研究”青年学者论坛旨在加强从事“数学、计算机与生命科学交叉研究”青年学者之间的联系，交流生命科学与计算生物学研究领域的最新成果，了解相关国际发展动态和研究热点，促进我国数学、计算机与生命科学交叉研究与应用实践的更好发展。前六届论坛已分别于2013年5月、2014年5月、2015年5月、2016年5月、2017年5月和2018年5月成功举办，受到了与会近千名青年学者的普遍好评。为进一步促进数学、计算机与生命科学交叉研究领域青年学者的交流，我们将于2019年5月18日-19日在北京举办第七届“数学、计算机与生命科学交叉研究”青年学者论坛。

会议将以国内在数学、计算机与生命科学交叉研究领域取得突出成果的青年科学家的学术报告为主，并邀请领域内的知名专家做特邀报告。报告的主题涵盖计算生物学和计算基因组学、单细胞测序数据分析与应用、癌症基因组学、表观基因组学、元基因组学、蛋白质组学以及系统生物学等研究领域。会议还特别安排了海报交流环节与特别论坛环节。

本届会议由中国科学院数学与系统科学研究院/国家数学与交叉科学中心与中国科学院遗传与发育生物学研究所主办，中国运筹学学会计算系统生物学分会、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、中国科学院青年创新促进会协办。会议组织者热诚欢迎从事相关研究的广大青年科学家与研究生参会。会议设有墙报介绍时间，将邀请部分墙报提交人通过幻灯的形式向全体参会者介绍墙报的内容，欢迎与会者积极提交墙报。

会议主席：

王秀杰　中国科学院遗传与发育生物学研究所

张世华　中国科学院数学与系统科学研究院

指导委员会（按照姓氏汉语拼音排序）：

主席：　陈润生　郭雷　马志明

委员：　敖平　卜东波　陈洛南　邓明华　冯建峰　高琳　巩馥洲　韩敬东　胡晓东　黄德双
李雷　李梢　李霞　李国君　李亦学　林魁　刘娟　吕金虎　沈百荣　孙啸
孙之荣　王亚东　魏冬青　吴家睿　闫桂英　于军　张成岗　张德兴　张立新　张奇伟
章祥荪　张学工　郑伟谋　周水庚　周天寿　邹秀芬

组织委员会（按照姓氏汉语拼音排序）：

主席：　张世华

委员：　付岩　何顺民　刘长宁　万林　王猛　王勇　吴凌云　肖景发　徐书华　章张

论坛会务组（按照姓氏汉语拼音排序）：

曹振　党大昌　董康宁　盖阔　何彬彬　李尚　刘星言　井方
沈群伦　史庆　徐泽千　闫秋鹏　闫晓芳　张驰浩　翟鹏龙　张彪

联系方式：

会务邮箱：cz@amss.ac.cn；咨询电话：010-82541288

会议目录

会议日程

2019 年 5 月 18 日 (9:00am – 5:30pm)

时间	报告人	报告题目
9:00 – 10:30　主持人：张世华		
9:00 – 9:10	**李霞教授、王秀杰研究员　开幕致辞**	
9:10 – 9:50	黄三文*	蔬菜育种的基因组路线图
9:50 – 10:30	汤 超*	Growth strategies of microbes on mixed carbon sources
10:30 – 10:50　茶歇、交流、海报布置		
10:50 – 12:10　主持人：高 歌		
10:50 – 11:30	李 霞*	数学在生物医学中的作用
11:30 – 12:10	汤富酬*	Decoding the gene regulation network in human germline cells by single-cell functional genomics approaches
12:10 – 1:30　午餐（物科餐厅四层）		
1:30 – 3:10　主持人：王秀杰		
1:30 – 1:55	梁承志	高质量基因组的构建与应用
1:55 – 2:20	李 雷	BAUM：基因组拼接的一个迭代方法
2:20 – 2:45	刘士勇	RNA 结构在编码潜能与 RNA-蛋白质识别中的重要作用
2:45 – 3:10	张 法	Pixer：一种基于深度分割网络的冷冻电镜颗粒图像自动挑选方法
3:10 – 3:30　茶歇、交流、海报陈述		
3:30 – 5:30　主持人：赵兴明		
3:30 – 3:55	王秀杰	An effective method to correct batch effects among single-cell RNA-sequencing data using machine learning approach
3:55 – 4:20	古 槿	肿瘤单细胞转录组计算分析的若干问题
4:20 – 4:55	秦丽璇	On the ‘off-label’ use of data normalization for sample classification in precision medicine
4:55 – 5:30	**墙报报告　主持人：蔡宏民**	
5:30　第一天会议日程结束		
5:40 – 8:00　冷餐招待会 (限教师、研究人员参加)（物科餐厅四层）		

会议日程

2019 年 5 月 19 日 (9:00am – 4:30pm)

时间	报告人	报告题目
9:00 – 10:20　主持人：王　勇		
9:00 – 9: 40	宿　兵*	Genetic mechanism of high altitude adaptation in Tibetans
9:40 – 10: 20	邢　毅*	Tracking intron splicing through space and time
10:20 – 10:40　茶歇、交流、海报陈述		
10:40 – 11:55　主持人：邢　毅		
10:40 – 11:05	张治华	High resolution genome 3D architectures as revealed by single cell level Hi-C and ATAC-seq data
11:05 – 11:30	唐忠辉	3D genome organization, function and evolution
11:30 – 11:55	杨建华	Decoding the regulation of m^{6}A methylation by modified histones from epigenomic and epitranscriptomic data
12:00 – 1:30　午餐（物科餐厅四层）		
1:30 – 2:45　主持人：万　林		
1:30 – 1:55	王吉光	癌症演化研究及在继发性胶质母细胞瘤中的应用
1:55 – 2:20	张　磊	Control of stem cells in both embryo and plant
2:20 – 2:45	冯桂海	Generation of uniparental mice from hypomethylated haploid ESCs with specific imprinting deletions
2:45 – 3:00　茶歇、交流、海报陈述		
3:00 – 3:50　主持人：郭安源		
3:00 – 3:25	张　勇	新基因起源：机制、功能、模式
3:25 – 3:50	姜　鹏	Big-data approaches to model cancer immune evasion and immunotherapy resistance
3:50 – 4:30　“如何将研究成果发表在高影响杂志”交流论坛（主持人：张世华） （嘉宾：邢　毅、王吉光、张磊）		
4:30　会议闭幕总结、会议结束		

*特邀报告

会议报告摘要

蔬菜育种的基因组路线图

黄三文

中国农业科学院农业基因组研究所

详情请见报告。

Growth strategies of microbes on mixed carbon sources

汤超

北京大学定量生物学中心

A classic problem in microbiology is that bacteria display two types of growth behavior when cultured on a mixture of two carbon sources: in certain mixtures the bacteria consume the two carbon sources sequentially (diauxie) and in other mixtures the bacteria consume both sources simultaneously (co-utilization). The search for the molecular mechanism of diauxie led to the discovery of the lac operon and gene regulation in general. However, questions remain as why microbes would bother to have different strategies of taking up nutrients and in the case of co-utilization what determines the partition and distribution of carbon sources in the cell. Here we show that diauxie versus co-utilization can be understood from the topological features of the metabolic network. A model of optimal allocation of protein resources quantitatively explains why and how the cell makes the choice when facing multiple carbon sources. When two carbon sources are being co-utilized, the model predicts the percentage of each carbon source in supplying the synthesis of every type of amino acid, which is quantitatively verified by experiments. Our work solves a long-standing puzzle and provides a quantitative framework for the carbon source utilization of microbes.

数学在生物医学中的作用

李霞

哈尔滨医科大学生物信息科学与技术学院

详情请见报告。

Decoding the gene regulation network in human germline cells by single-cell functional genomics approaches

汤富酬

北京大学生命科学学院 BIOPIC 中心

Human germline cells are crucial for maintenance of the species. However, the developmental trajectories and heterogeneity of human germline cells remain largely unknown. We performed single-cell RNA-seq and DNA methylome sequencing analyses of human germline cells in female and male human embryos spanning several critical developmental stages. We found that female fetal germ cells (FGCs) undergo four distinct sequential phases characterized by mitosis, retinoic acid signaling, meiotic prophase, and oogenesis. Male FGCs develop through stages of migration, mitosis, and cell-cycle arrest. Individual embryos of both sexes simultaneously contain several subpopulations, highlighting the asynchronous and heterogeneous nature of FGC development. Moreover, we observed reciprocal signaling interactions between FGCs and their gonadal niche cells, including activation of the bone morphogenic protein (BMP) and Notch signaling pathways. Our work provides key insights into the crucial features of human germline cells during their highly ordered mitotic, meiotic, and gametogenetic processes in vivo.

高质量基因组的构建与应用

梁承志

中国科学院遗传与发育生物学研究所

单分子长片段测序技术已成为当今构建高质量基因组的主流技术。然而，在复杂基因组中往往有很多复杂区域没有被组装出来，导致组装序列太碎片化。我们在水稻、苦荞、小麦、金鱼草中利用单分子测序结合大片段 DNA 文库、光学图谱、遗传图谱、HiC 等，构建了高质量的参考基因组，并对这些基因组的进化和某些特殊功能基因进行了初步研究。此外我们开发了一个利用单分子测序长片段进行基因组复杂区域组装的新方法 HERA。在玉米、苦荞和人基因组中与已发表版本进行对比，玉米的 Contig N50 从 1.3 Mb 提升至 61.2Mb，人的 Contig N50 从 8.3 MB 提升至 54.4 MB，苦荞基因组 Contig N50 达到了 27.85 Mb。我们已经利用 HERA 构建了多个其它物种的基因组。低成本的高质量基因组预示着后基因组时代的全面到来。

BAUM：基因组拼接的一个迭代方法

李雷
中国科学院数学与系统科学研究院

基因组拼接是计算生物学的基本科学问题，高质量的基因组是分子生物和医学研究的基础"地图"。二代测序技术准确性高，成本低，但是读长短，只有100~200bp。我们近几年开发了一个基因组拼接的一个迭代方法。首先通过设置自适应的映射准则，将短读序列映射到参照基因组上。然后根据有唯一映射位置的序列和它们的伙伴信息，采用局部的 OLC 方法做叠阵延拓，建立架构，用统计方法和序列比对连接相邻的 contigs，从而生成新的参照基因组。上述步骤可以进行迭代，其中序列映射和叠阵延拓可以并行实现，内存需求小。contig 的 N50 长度是衡量所拼接出的基因组的连续性的一个重要指标，对脊椎动物的基因组，BAUM 在保证正确性的前提下，可以将 N50 从现有方法的 30~50Kbp，提高到 200Kbp 以上。最近，BAUM 方法成功拼接了高原鼢鼠等物种的基因组。

RNA 结构在编码潜能与 RNA-蛋白质识别中的重要作用

刘士勇
华中科技大学物理学院

由于 RNA 可以折叠成各式各样的三维结构，蛋白质-RNA 之间形成的复合物也如蛋白质-蛋白质复合物一样复杂，因而蛋白质-RNA 复合物结构预测与蛋白质-蛋白质复合物结构预测一样仍然很困难。因此，我们开发了自由对接 3dRPC 与基于模板 PRIME 的方法。该工作首次揭示了蛋白质-RNA 复合物结构之间的序列-结构联系，并发现存在一个转变点。随后，我们发现 RNA 结构比对算法 SARA 的打分与 RNA 的长度相关，在某些情况下找不到模板。因此，我们系统地分析了 RNA 回转半径 Rg 与长度 N 的关系，满足 $Rg \propto N^{0.39}$ 指标定律，从而定义了一个新的 RNA 相似性打分函数 RMscore。基于 RMscore，我们开发了一个不依赖于 RNA 长度的结构比对算法 RMalign，进一步提高了 PRIME 算法。联合自由对接算法与基于模板的算法，我们提出了 P3DOCK 的算法。我们通过序列分析发现了一些识别 RNA 结合蛋白质(RBP)的重要特征，给实验生物学家提供了寻找 RBP 的高精度预测工具 RBPPred 与 Deep-RBPPred。在这些工作中，我们发现 RNA 的结构非常重要，进一步猜测 RNA 结构可能在编码潜能中发挥重要作用，利用 CTD 编码特征来描述 RNA 的折叠结构。研究表明，T2，C0 和 GC（CTD 编码的特征）在 RNA 编码潜能的预测上具有重要作用。CPPred 在人类，小鼠，斑马鱼和酿酒酵母测试集上，具有高的准确性，较目前发表的工具准确性有微弱的提高，然而，CPPRed 在这些物种的短的 RNA 序列上（sORF）具有特别的优势，比之前开发的工具有一个比较大的提升。

Pixer：一种基于深度分割网络的冷冻电镜颗粒图像自动挑选方法

张法
中国科学院计算技术研究所

冷冻电镜三维重构技术已成为确定生物大分子三维结构的第一选择。如何从冷冻电镜图像中准确、自动挑选出数以百万计的生物大分子颗粒图像一直是本领域急需解决的关键问题。本报告将首先介绍一种基于深度学习的全自动冷冻电镜颗粒图像挑选算法—Pixer，我们将图像分割的思想应用到颗粒挑选中，有效解决了训练样本不足的问题，其次将介绍分割网络在三维颗粒挑选中的最新应用。

An effective method to correct batch effects among single-cell RNA-sequencing data using machine learning approach

王秀杰
中国科学院遗传与发育生物学研究所
中国科学院大学

Single-cell RNA-sequencing technology has demonstrated great power in many aspects, and has been selected as one of the breakthroughs of the year 2018 by Science magazine. However, there are still many challenges in the analysis of single-cell RNA-sequencing data. One of such challenges is batch effect correction, which, if not dealt properly, could produce false cell-to-cell expression variability, thus jeopardize the analysis results to a large extent. Here, we present a new machine learning-based approach to correct batch effects among single-cell RNA-seq data. The method demonstrates higher accuracy as compared to other available batch effect correction methods, and is capable to produce more biologically meaningful results on multiple testing data sets.

肿瘤单细胞转录组计算分析的若干问题

古槿

清华大学自动化系

肿瘤是一种高度复杂的疾病，肿瘤微环境中存在免疫细胞、内皮细胞、成纤维细胞等有多种类型的细胞，肿瘤细胞也具有很高的异质性，多个克隆混杂、存在肿瘤干细胞等多种亚群。最新的高通量单细胞转录组学技术可以一次性检测数千至数万个细胞的基因表达情况，这使得在单细胞水平系统分析肿瘤的转录组特征成为可能。我们将以垂体瘤、肝癌等为例，并结合国际上的最新进展，探讨肿瘤单细胞转录组计算分析中的几个重要问题：1）数据主要噪声来源与处理方法；2）肿瘤细胞的转录异质性及其身份识别；3）肿瘤细胞转录组共性特征分析。

On the ‘off-label’ use of data normalization for sample classification in precision medicine

秦丽璇

Memorial Sloan Kettering Cancer Center

Data normalization is an important preprocessing step for genomics data containing unwanted data variation due to experimental handling. There has been a critical yet over-looked disconnection between the use of data normalization and the goals of subsequent analysis: on one hand, methods for data normalization that have been developed for group comparison frequently encounter ‘off-label’ use for other analysis goals such as sample classification; on the other hand, analysis are often performed on normalized data neglecting potential normalization ‘side-effects’ such as over-compressed data variability. A bridge between these two is made possible by a unique pair of microRNA array datasets on the same set of tumor tissue samples that were collected at Memorial Sloan Kettering Cancer Center. In this talk, I will illustrate the use of this dataset pair to study the impact of data normalization on the development of tumor sample classifiers, an important tool that is in dire need to tailor treatment choices for personalized medicine.

Genetic mechanism of high altitude adaptation in Tibetans

宿兵

中国科学院昆明动物研究所

Tibetans are well adapted to the hypoxic environments at high altitude, yet the molecular mechanism of this adaptation remains elusive. We reported comprehensive genetic and functional analyses of EPAS1, a gene encoding hypoxia inducible factor 2α (HIF-2α) with the strongest signal of selection in previous genome-wide scans of Tibetans. We showed that the Tibetan-enriched EPAS1 variants down-regulate expression in human umbilical endothelial cells and placentas. Heterozygous EPAS1 knockout mice display blunted physiological responses to chronic hypoxia, mirroring the situation in Tibetans. Furthermore, we found that the Tibetan version of EPAS1 is not only associated with the relatively low hemoglobin level as a polycythemia protectant, but also is associated with a low pulmonary vasoconstriction response in Tibetans. We propose that the down-regulation of EPAS1 contributes to the molecular basis of Tibetans’ adaption to high-altitude hypoxia.

Tracking intron splicing through space and time

邢毅

宾夕法尼亚大学费城儿童医院

详情请见报告。

High resolution genome 3D architectures as revealed by single cell level Hi-C and ATAC-seq data

张治华

中国科学院北京基因组研究所

The 3D genome architecture underlies many cellular processes in the nucleus. High-throughput chromosome conformation capture (3C) technologies, such as Hi-C, have made it possible to survey 3D genome structure. The sub-mage base size topological associating domains (TAD) have been observed from Hi-C. However, to accurately detect such domains relay on ultra-deep sequencing and sophistic normalization procedures, making it a major challenge to decode the genome architecture. Moreover, to the best of interests in the gene regulation, detecting enhancer-promoter interaction is much harder with solely Hi-C data. Previously, our lab developed CISD_loop and deDoc to predict high-resolution chromatin loop and TAD with low resolution MNase-seq and Hi-C data, respectively. However, for CISD_loop, the MNase-seq data is much less prevalence than ATAC-seq, while for deDoc, the data input still not in the single cell Hi-C level. Here, in this talk, we present two newly developed algorithms, adATAC and TOKI to predict high resolution genome 3D architecture with single cell level ATAC-seq and single cell level Hi-C data, respectively. Our algorithms may facilitate systematic investigations of chromosomal domains and loops on a larger scale than hitherto have been possible.

3D genome organization, function and evolution

唐忠辉
中山大学中山医学院

详情请见报告。

Decoding the regulation of m^{6}A methylation by modified histones from epigenomic and epitranscriptomic Data

杨建华
中山大学生命科学学院

N6-methyladenosine (m^{6}A) is the most prevalent internal post-transcriptional modification in human transcriptomes and has been shown to have important roles in various normal and pathological processes. However, the process by which m6A is deposited on mRNAs is largely unknown. Here we developed a serial of computational methods to decode the regulation of m6A methylation from epigenomic and epitranscriptomic data and demonstrated that histone H3 trimethylation at Lys36 (H3K36me3), a marker for transcription elongation, guides m6A deposition globally. Comparative analyses of ChIP-seq data for H3K36me3 and m6A-seq data revealed that majorities of m6A peaks overlapped with H3K36me3 sites and that the overlapping sites were enriched near stop codons. We also found that m6A sites identified from miCLIP-seq are enriched in the vicinity of H3K36me3 peaks and are reduced globally when cellular H3K36me3 is depleted. Furthermore, we show that a significant genome-wide correlation between chromatin binding of METTL14 to H3K36me3. Mechanistically, H3K36me3 is recognized and bound directly by METTL14, a crucial component of the m6A methyltransferase complex (MTC), which in turn facilitates the binding of the m6A MTC to adjacent RNA polymerase II, thereby delivering the m6A MTC to actively transcribed nascent RNAs to deposit m6A co-transcriptionally. The discovery of interplay between modified histones and RNA methylation represents a new regulatory layer, and an additional level of complexity, in the control of gene expression.

癌症演化研究及在继发性胶质母细胞瘤中的应用

王吉光
香港科技大学

理解肿瘤演化机制有助于制定更准确的肿瘤治疗方案。本研究以脑胶质瘤为例研究肿瘤在标准治疗下的演化过程。低级别神经胶质瘤会进展为继发性胶质母细胞瘤（sGBM），其治疗方案有限且机制尚不清楚。通过研究 188 名 sGBM 病人，构建此疾病体细胞突变的遗传图谱，我们发现相比其他类型的脑胶质瘤，sGBM 具有以下几种突变的富集，包括 TP53 突变、超突变、MET-外显子-14-跳跃等。这一研究揭示了一个低级别胶质瘤向高级别进展的重要途径，并为临床治疗提供了潜在的动态靶标。基于此靶标，合作者启动并顺利完成了 I 期临床试验，在至少两名晚期 sGBM 患者中该药物实现了缩减肿瘤体积，表明该疗法在精确治疗神经胶质瘤上的临床潜力。

Control of stem cells in both embryo and plant

张磊
北京大学定量生物学中心

Development and regeneration require plant and animal cells to make decisions based on their locations. In this talk, I will start with the dual role of Nanog during stem cell differentiation and reprogramming. A stochastic five-node network model shows the low-Nanog state can enhance cell differentiation through serving as an intermediate state to reduce the energy barrier of transition. Then I will present a mathematical model to study feedback of organs on shoot apical stem cells by auxin transport switch. We find that auxin transport from leaf primordia inhibits the establishment of polar auxin transport out of the meristem. In aberrant leaf development mutant and leaf removal plant, the inhibition from leaf primordia is interrupted and auxin transports out of the meristem, leading to enlarged stem cell and stem cell region. The joint work Qing Nie (UC Irvine), Chao Tang (PKU), Yuling Jiao (CAS).

Generation of uniparental mice from hypomethylated haploid ESCs with specific imprinting deletions

冯桂海
中国科学院动物研究所

Uniparental reproduction is widespread among lower vertebrates, but not in mammals. Deletion of the H19 imprinted region in immature oocytes produced bimaternal mice with defective growth, however bipaternal reproduction has not been previously achieved in mammals. We found that cultured parthenogenetic and androgenetic haploid embryonic stem cells (haESCs) display DNA hypomethylation resembling that of primordial germ cells. Through MII oocyte injection or coinjection of sperm into hypomethylated haESCs engineered with genetic deletions in specific imprinting regions, we obtained live uniparental mice. Deletion of three imprinted regions in parthenogenetic haESCs restored normal growth of fertile bimaternal mice, whereas deletion of 7 imprinted regions in androgenetic haESCs enabled production of live bipaternal mice that died shortly after birth. The phenotypic analyses of organ and body size of uniparental-derived mice support the genetic conflict theory of genomic imprinting. Taken together, our results highlight the factors necessary for crossing uniparental reproduction barriers in mammals.

新基因起源：机制、功能、模式

张勇

中国科学院动物研究所

新基因如何起源、进化是进化生物学的核心问题之一，其起源机制、分子功能及进化模式是该领域的主要研究方向。近年来，我们在这三方向及支撑性资源开发方面获得进展：1）我们发现一种在无脊椎和脊椎动物中保守的、LTR 类型逆转座子介导的新基因重复起源机制。该机制与哺乳动物中研究较多的 L1 机制相比，LTR 塑造了包含转座子和 mRNA 的嵌合基因，更可能发挥新功能。2）我们鉴定了 846 个灵长类特异蛋白编码基因，进而发现它们富集于精子发生、人脑发育等快速演化的过程。有意思的是精子发生中 DNA 去甲基化导致宽松的转录环境，从而使新基因易于在该器官起源；癌症与睾丸共享类似的表观调控，这些基因同样易于上调而促癌。3）我们发现新基因倾向于编码功能丢失突变。部分原因是新基因承担物种特异的功能，环境的变迁可能让这些功能不再被需要；因此新基因丢失的可能性更高。4）我们建立了基于全基因组共线性鉴定新基因的方法，同时针对新基因注释质量差的现状开发了目前仅存的人类新基因数据库 GenTree（http://gentree.ioz.ac.cn）。

Big-data approaches to model cancer immune evasion and immunotherapy resistance

姜鹏

National Cancer Institute

The rapid growth of big-data resources, catalyzed by breakthroughs in genomics technologies, has resulted in a paradigm shift in cancer research. I will introduce my recent works that integrated the vast amount of public data to model the cancer therapy efficacy. To predict immunotherapy response, we developed TIDE, a computational method to model two primary mechanisms of tumor immune evasion: the induction of T cell dysfunction in immune-hot tumors and the prevention of T cell infiltration in immune-cold tumors. TIDE repurposed many clinical data cohorts without immunotherapy to identify immune evasion signatures as surrogate immunotherapy biomarkers. Using pre-treatment RNA-Seq or NanoString tumor expression profiles, TIDE predicted the outcome of melanoma patients treated with first-line anti-PD1 or anti-CTLA4 more accurately than other biomarkers. TIDE also revealed new immunotherapy resistance regulators, such as SERPINB9, which hijacked the self-protection strategy of T cells for tumor immune evasion. Besides the immunotherapy focus, we also developed CARE, a computational method focused on targeted therapies, to identify synergistic drug combinations to overcome the resistance to primary treatments, using cell line compound screens. In summary, my recent works demonstrated that the integration of big public data is a cost-effective approach to rediscover new therapeutic knowledge.

海报摘要

1. Genome Warehouse: a centralized resource of genome assembly data

Meili Chen, Yingke Ma, Zheng Gong, Yiming Bao
(中国科学院基因组研究所)
Email: chenml@big.ac.cn

The Genome Warehouse (GWH; http://bigd.big.ac.cn/gwh) is a public repository housing genome-scale data for a wide range of species and delivering a series of web services for genome data submission, storage, release and sharing. For each species, GWH contains detailed genome-related information including species metadata, genome assembly, sequence data and the corresponding annotations. Particularly, to archive high-quality genome sequences and genome annotation information, GWH adopts a uniform standardized procedure for quality control. Since the availability of submission service online in July 2017, GWH has accommodated 256 genome submissions, viz., 98 animals, 30 plants, 3 fungi, 85 bacteria, 23 archaea, one virus, 13 metagenomes and 3 other species, showing the great promise to have more and more genome data submissions in the wake of high-throughput sequencing capability and large-scale sequencing-based projects. Until 11th May 2019, GWH has released 96 genome assemblies. Besides, GWH is also enriched by integrating 138 newly released genomes (61 animals and 77 plants) from NCBI and sequenced in-house. GWH provides friendly and interactive interfaces for data visulization, and will integrate popular online analysis tools to make genome analysis more simple and convenient.

2. A novel unsupervised learning model for detecting driver genes from pan-cancer data through matrix tri-factorization framework with pairwise similarities constraints

Jianing Xi[1,2,*], Ao Li[2], Minghui Wang[2] (1: School of Computer Science and Technology, Xidian University, 2: School of Information Science and Technology, University of Science and Technology of China)
Email: jnxi@xidian.edu.cn

Identifying cancer-causing mutated driver genes from passenger mutations is crucial to enhance the de- velopment of cancer diagnostics and therapeutics, and many previous effort s have been undertaken to identify cancer driver genes from somatic mutation data of specific types of cancers. However, many driver genes are underestimated when the mutation data of only specific cancers are investigated, which complicates the understanding of tumorigenesis. According to recent studies, cancers of disparate organs have many shared genomic mutations, and some driver genes that are not highly frequently mutated in patients of one cancer type may display considerable mutation frequencies across patients of mul- tiple cancer types. By taking into account both the similarities of mutation profiles of different cancer types and the information of gene interaction network, we propose a novel unsupervised learning model based on matrix tri-factorization by learning the similarities from pairwise constraints to detect driver genes from pan-cancer data. In the evaluation of known benchmarking genes, our model achieves bet- ter performance than those of the existing matrix factorization based methods which do not consider the pairwise similarities between cancers. Furthermore, the detection performance of our model is also largely increased (area under the precision-recall curve = 9.1% for Vogelstein genes) when compared with existing methods. Moreover, our model discovers some driver genes that have been reported in recent published studies, showing its potential for application in identifying driver gene candidates for further wet experimental verification.

3. Abnormalities in prefrontal cortical gene expression profiles relevant to schizophrenia in MK-801-Exposed C57BL/6 mice

赵佳璐，叶海虹（首都医科大学）
Email: yehh@ccmu.edu.cn

MK-801, a non-competitive NMDA receptor antagonist, disturbs NMDA receptor function in rodents and induces psychological and

behavioral changes similar to schizophrenia (SCZ). However, the effects of MK-801 treatment on gene expression are largely unknown. Here we performed RNA-sequencing on the prefrontal cortex of MK-801-exposed male mice in order to analyze gene expression and co-expression patterns related to SCZ and to identify mechanisms that underlie the molecular etiology of this disorder. Transcriptome analysis revealed that the differentially expressed genes were more often associated with biological processes that included postsynaptic transmission, immune system process, response to external stimulus and hemostasis. In order to extract comprehensive biological information, we used an approach for biclustering, called FABIA, to simultaneously cluster transcriptomic data across genes and conditions. When combined with analyses using DAVID and STRING databases, we found that co-expression patterns were altered in synapse-related genes and genes central to the mitochondrial network.

Abnormal co-expression of genes mediating synaptic vesicle cycling could disturb release, uptake and reuptake of glutamate, and the perturbation in co-expression patterns for mitochondrial respiratory chain complexes was extensive. Our study supports the hypothesis that research using MK-801-exposed male mice as an animal model of SCZ offers important insights into the pathogenesis of SCZ.

4. 元宝枫籽对小鼠肠道菌群生态的影响

孙朋浩[1], 薛玉环[1], 吴永继[1], 郑 伟[2], 任 玮[1], 朱晓岩[1], 赵善廷[1,*] (1:西北农林科技大学动物医学院神经生物学实验室, 2:西北农林科技大学学资源环境学院, 农业农村部西北植物营养与农业环境重点实验室)

Email: zhaoshanting@nwsuaf.edu.cn

元宝枫是我国特有树种，其籽含有丰富的不饱和脂肪酸和多种营养物质。本研究旨在探索元宝枫籽对小鼠肠道菌群组成结构及菌群相关代谢功能的影响。选取 8 周龄的昆明小鼠，元宝枫籽伺喂添加量为每天 1.5 克，连续伺喂两个星期后采集直肠粪便进行肠道菌群的高通量测序分析，并且全程记录体重变化。通过对小鼠肠道菌群的分析后发现：相比于对照组，实验组小鼠肠道菌群的α多样性有所上升；通过β多样性分析发现实验组数据与对照组数据发生显著性的分离（p<0.026; R^2= 0.09351）。同过对分类学水平数据的组间对比发现：相比于对照组，实验组中厚壁杆菌门和变形菌门的相对丰度增高而拟杆菌门的相对丰度下降；比较组间厚壁杆菌门和拟杆菌门相对丰度比值的差异变化，发现实验组 F/B 平均值要高于对照组（p=0.063）。通过“线性判别分析”（Linear Discriminant Analysis，LefSe）发现：相比于对照组，实验组中金黄色葡萄球菌属和费克蓝姆菌属的相对丰度要显著低于对照组，表明元宝枫籽对机会性致病菌有一定的抑菌作用。利用“通过重建未观察到的状态对社区进行系统发育研究”（Phylogenetic Investigationof Communities byReconstruction of Unobserved States，PICRUSt）方法进行差异功能基因预测发现：相比于对照组，实验组在多糖分解代谢功能水平上要低于对照组。结论：元宝枫籽可以影响小鼠肠道菌群的结构组成并提高个体肠道菌群的α多样性，对肠道菌群的分解代谢功能产生一定影响，同时对机会性致病菌有一定的抑菌作用

5. 异源肿瘤动物模型专属测序和药敏表型数据分析方法

戴文韬[1,2], 李全学[1,2], 刘继翔[1,2], 刘伟[1,2], 李亦学[1,2,*]，李园园[1,2,*] (1:上海生物信息技术研究中心,2:上海药物转化工程技术研究中心,＊:上海生物信息技术研究中心)

Email: yyli@scbit.org

异源肿瘤动物模型（patient‐derived xenograft ,PDX），是目前为止最佳的肿瘤临床前药物实验模型。在 PDX 模型的实际应用过程中，我们发现有两个关键问题对 PDX 模型数据质量有着重要影响：1）PDX 模型的肿瘤移植物测序，极有可能受到宿主遗传物质（DNA 和 RNA）污染，这对于下游数据分析影响极大。2）PDX 模型极好地重现了肿瘤内和肿瘤间的异质性，同时又具有并行的高通量药物实验能力，因此现有的基于细胞实验或临床队列的药敏表型数据分析方法均不适用于 PDX 模型药敏实验数据，无法充分发挥 PDX 模型优势。

针对以上两个问题，我们团队进行了积极探索，分别给出了如下解决方案：1）通过系统测评当前主流的去除宿主污染测序分析工具如 Xenome, Disambiguate 等，我们整合构建了适用范围广，效果好的肿瘤移植物去除宿主污染测序分析流程，如图 1。2）我们将 PDX 药物实验归纳为四种模式，并据此打造了一个 R 程序包 DRAP，这是第一个为 PDX 平台量身定制的药物反应分析和可视化计算工具（如图 2），可以充分发掘 PDX 模型在药物实验上的技术优势。

综上所述，我们整合了一套肿瘤移植物去除宿主污染测序分析流程，开发了第一个 PDX 模型专属药物反应分析和可视化计算工具 DRAP，以上工作极大地改善了 PDX 模型药敏实验的全链条数据质量，有助于促进 PDX 模型在药物开发和个性化癌症治疗中的应用。

6. CPPred: coding potential prediction based on the global description of RNA sequence

Xiaoxue Tong and Shiyong Liu*

Email: liushiyong@gmail.com

Recently, next-generation sequencing technology has generated thousands of novel transcripts. Previously developed methods CPAT, CPC2 and PLEK can distinguish coding RNAs and ncRNAs very well, but poorly distinguish between small coding RNAs and small ncRNAs. Herein, we report an approach, CPPred, which is based on SVM classifier and multiple sequence features including novel RNA features encoded by the global description. The CPPred can better distinguish not only between coding RNAs and ncRNAs, but also between small coding RNAs and small ncRNAs than the state-of-the-art methods due to the addition of the novel RNA features. Remarkably, we also reveal that the global description of encoding features (T2, C0 and GC) plays an important role in the prediction of coding potential.

7. P3DOCK: a protein-RNA docking webserver based on template-based and template-free docking

Jinfang Zheng[1], Xu Hong[1], Juan Xie, Xiaoxue Tong and Shiyong Liu* (School of Physics, Huazhong University of Science and Technology)
Email: liushiyong@hust.edu.cn

Protein-RNA interaction play an important role in the metabolism of organisms. The information of three-dimensional (3D) structures reveals that atomic interactions are particularly important. The calculation method for modeling a 3D structure of a complex mainly includes two strategies: free docking and template-based docking. In this paper, we compare the difference between the free docking and the template-based algorithm. And the results of these two methods indicate that the complementarity of these two methods. So, we combine these two methods. Based on the analysis of the calculation results, the transition point is confirmed and used to integrate two docking algorithms to develop P3DOCK. P3DOCK holds the advantages of both algorithms. The results of the three docking benchmarks show that P3DOCK is better than those two non-hybrid docking algorithms. And, the success rate of P3DOCK is also higher (0%-20%) than state-of-the-art hybrid and non-hybrid methods. Finally, the hierarchical clustering algorithm is utilized to cluster the P3DOCK's decoys. The clustering algorithm improves the success rate of P3DOCK. For ease of use, we provide a P3DOCK webserver, which can be accessed at www.rnabinding.com/P3DOCK/P3DOCK.html. An integrated protein-RNA docking benchmark can be downloaded from http://rnabinding.com/P3DOCK/benchmark.html.

8. CLIP1 and DMD are two novel RNA-binding proteins through computational prediction and experimental validation

Juan Xie, Xiaoli Zhang, Jinfang Zheng, Xu Hong, Xiaoxue Tong, Shiyong Liu* (School of Physics, Huazhong University of Science and Technology)
Email: liushiyong@gmail.com

Since RBPs play important roles in the cell, it is particularly important to find new RBPs. We performed iRIP-seq to verify two proteins, CLIP1 and DMD, predicted by RBPPred whether are RBPs or not. The experimental results confirm that these two proteins have RNA-binding activity in HeLa cell. By analyzing the experimental data, we identified significantly enriched binding motifs UGGGGAGG and CUUCCG for CLIP1 and DMD, respectively. The KEGG and GO analysis show that the CLIP1 and DMD share some biological processes and functions. In addition, we found that the SNPs between DMD and its RNA partners may associate with Becker muscular dystrophy, Duchenne muscular dystrophy, Dilated cardiomyopathy 3B and Cardiovascular phenotype. Among the seven cancers data, DMD and another 114 genes always co-mutate, and 24 of these 114 genes interact with DMD. These cancers may be associated with the mutations in both DMD and the genes it interacts with.

9. Evolution of avian limbs and digits

Wen Kang and Qi Zhou* (Life Sciences Institute, Zhejiang University)
Email: 11707043@zju.edu.cn

The variety of digits and limbs is closely related to the adaptation of avian. In ratites, a distinctive clade of flightless birds, the emu possesses a pair of vestigial wings with each having a single digit and greatly reduced forelimb musculature. While the ostrich and rhea possess strongly reduced zeugopod and autopod harboring three digits. Recent phylogenetic analysis indicated that each ratite species has lost flight independently, but the underlying molecular and developmental mechanisms are currently unknown. Meanwhile, in chicken and other flying birds, they also show instances of digit loss and have functional wings with three digits. Such cases also exist in hindlimb of the birds: while under different selective regimes, the adult emu, ostrich and chicken show limbs with three, two, and four digits respectively. This pattern has emerged through a process of digit loss during evolution, and the identities of avian digits still remain controversial. Here, we performed RNA-seq analysis on limbs and digits of four species (chicken, emu, ostrich and Chinese softshell turtle) in five different stages, to explore the molecular and cellular mechanisms associated with the evolution of limbs degeneration and digits loss in birds.

10. DiffMut-W:利用突变差异性分析评估基因致癌性的新工具

张浩洋 1, 杨跃东* (1:中山大学生物医学工程学院,*:中山大学数据科学与计算机学院)

Email: yangyd25@mail.sysu.edu.cn

寻找致癌基因是癌症基因组学的一个主要目标。传统的方法多以病人自身为参照，基于突变频率(MutSig, MutSigCV)，突变聚集性(OncodriveCLUST)以及突变功能性(OncodriveFML)比较寻找致癌基因。2016 年推出的 DiffMut 算法通过比较癌症人群与来自千人基因组的自然人群基因突变分布差异，取得了超越以往的表现。然而，该方法只考虑了两个群体之间突变数的分布差异，忽略了不同突变的功能差异。

因此，我们进一步将 Diffmut 扩展到 Diffmut-W，利用 dbNSFP 3.0 数据库提供的 24 种突变致病性评价体系作为权重构建突变矩阵，采用 Unidirectional Earth Mover’s Difference (uEMD)依次衡量不同基因突变在两人群中的分布差异(图一)。我们分析了 TCGA 提供的 33 种癌症突变数据，并在 Cancer Gene Census(CGC)的已知致癌基因列表上进行评估以选择 Diffmut-W 的最佳权重(图二)，并与其他方法在同一数据集的比较。结果显示，Diffmut-W 的平均 AUPRC 以及 致癌基因富集能力均高于其他方法 (DiffMut, OncodriveClust, OncodriveFML and SomInaClust) (图三)。此方法可以在 https://github.com/zhanghaoyang0/DiffMut-W 下载并使用。

11. GenTree, an integrated resource for analyzing the evolution and function of primate-specific coding genes

Yi Shao*[1], Chunyan Chen*[1], Hao Shen, Bin Z.He, Daqi Yu[1], Shuai Jiang, Shilei Zhao, Zhiqiang Gao, Zhenglin Zhu, Xi Chen, Yan Fu, Hua Chen, Ge Gao, Manyuan Long, and Yong E.Zhang[1] (1:Key Laboratory of Zoological Systematics and Evolution & State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, *:Equal contributions)

Email: zhangyong@ioz.ac.cn

The origination of new genes contributes to phenotypic evolution in humans. Two major challenges in the study of new genes are the inference of gene ages and annotation of their protein-coding potential. To tackle these challenges, we created GenTree, an integrated online database that compiles age inferences from three major methods together with functional genomic data for new genes. Genome-wide comparison of the age inference methods revealed that the synteny-based pipeline (SBP) is most suited for recently duplicated genes, whereas the protein-family–based methods are useful for ancient genes. For SBP-dated primate-specific protein-coding genes (PSGs), we performed manual evaluation based on published PSG lists and showed that SBP generated a conservative data set of PSGs by masking less reliable syntenic regions. After assessing the coding potential based on evolutionary constraint and peptide evidence from proteomic data, we curated a list of 254 PSGs with different levels of protein evidence. This list also includes 41 candidate misannotated pseudogenes that encode primate-specific short proteins. Coexpression analysis showed that PSGs are preferentially recruited into organs with rapidly evolving pathways such as spermatogenesis, immune response, mother–fetus interaction, and brain development. For brain development, primate-specific KRAB zinc-finger proteins (KZNFs) are specifically up-regulated in the midfetal stage, which may have contributed to the evolution of this critical stage. Altogether, hundreds of PSGs are either recruited to processes under strong selection pressure or to processes supporting an evolving novel organ.

12. Development and evolution of germ cells in birds

He Huang-yi, Li Jing and Zhou Qi (Life Sciences Institute, Zhejiang University)

Email: zhouqi1982@zju.edu.cn

The primordial germ cells (PGCs) are precursors to sperms and eggs and the only cell type which can transfer genetic or epigenetic information from generation to generation. After being specified by germ plasm (“preformation”) or induced from nearby cell signals (“epigenesis”), PGCs migrate and colonize the genital ridge by the bloodstream (birds) or the dorsal mesentery (mammals). Despite great advances have been made in tracing PGCs and culturing them in vitro, our understanding into the gene regulatory networks (GRNs) underlying their specification, migration and differentiation are restricted to few model organisms. In this project, gonad transcriptomes at single-cell level from HH25, HH34 and HH39 of both emu and chicken have been collected with 10X technology. Analysis of chicken male HH39 gonad has identified 5 major cell types, including germ cell, granulosa cell, sertoli cell, leydig cell and endothelial cell. Further analysis will be focused to characterize the dynamic transcriptome landscape of avian germ cells across different developmental stages and to explore the evolution of germline sex determination mechanisms between birds, and between birds and mammals.

13. C3: Connect separate Connected Components to form a succinct disease module

Jie Hu and Bingbo Wang* (School of Computer and Technology, Xidian University)
Email: bingbowang@xidian.edu.cn

Accurate disease module is helpful for understanding the molecular mechanism of disease causation and identifying drug target. However, for fragmentization of disease module in incomplete human interactome, how to determine connectivity pattern and detect a full neighbourhood of disease is an open problem. Here we develop a topology - based method to dissect the connectivity of intermediate nodes and edges and form a succinct disease module. By applying this Connect separate Connected Components (CCC, C3) method on a large corpus of curated diseases, we find that most Separate Connected Components (SCCs) formed by Disease-Associated Proteins (DAPs) can be connected into a well connected component as an observable module. This pattern also holds for altered genes from multi-omics data such as The Cancer Genome Atlas. Overall, C3 tool can inspire a deeper understanding of interconnectedness of phenotypically related genes, and can be used to detect a well-defined neighbourhood that drives complex pathological processes.

14. Long non-coding RNA CCTT is required for CENP-C targeting centromeres by RNA-centromeric DNA triplex in *trans*

Chong Zhang[1]*, Dongpeng Wang[2,3]*, Yajing Hao[2]*, Shuheng Wu[2,3], Xuemin Zhang[4], Yuanchao Xue[2], Jianjun Luo[2], Yan Teng[1#], Runsheng Chen[2#] and Xiao Yang[1#] (1:State Key Laboratory of Proteomics, Genetic Laboratory of Development and Diseases, Institute of Biotechnology, 2:Key Laboratory of RNA Biology, Institute of Biophysics Chinese Academy of Sciences, 3:University of the Chinese Academy of Sciences, 4:State Key Laboratory of Proteomics, Institute of Basic Medical Sciences, National Center of Biomedical Analysis, 27 Tai-Ping Road, * These authors contributed equally to this work)
Email: yangx@nic.bmi.ac.cn
rschen@ibp.ac.cn

Centromere is a unique region on the chromosome that is required to attach to the mitotic spindle and ensure chromosome segregation. Abnormal centromeres can lead to genomic instability and loss, leading to severe dysplasia and tumorigenesis. Although centromeres have been localized to specific chromosomal regions in many organisms and many centromere-associated proteins have been identified, the underlying problems with regulating the identification and formation of centromere regions remain unclear. Recent studies have indicated that RNA produced by centromere transcription plays an important role in centromere heterochromatin. Here, we use RIP-seq and irCLIP-seq to systematically search for long non-coding RNAs that may be involved in centromere formation and kinetochore assembly in humans. We found a long non-coding RNA CCTT, which plays a key role in the assembly of CENP-C at the metaphase centromere. Using the ChIRP-seq, we revealed that it can bind to centromeres in the form of RNA-DNA triplex via the 44-79 nt DNA binding region. Using irCLIP-seq and SHAPE-MaP, we show that it recruited CENP-C to the centromeric DNA to initiate kinetochore assembly through the 127-177 nt CENP-C binding region. In addition, the depletion of Lnc-CCTT leads to a significant reduction of CENP-C localization to the inner centromere and abnomal attachment of kinetochores to the mitotic spindle. Moreover, knockdown of Lnc-CCTT causes cell division arrest, chromosome bridges and genomic instability, including aneuploidy, chromosome bridges, binuclei and micronucleus. Furthermore, the overexpression of Lnc-CCTT can initiate the formation of ectopic centromeres. This work shows that human centromere is epigenetically regulated by long noncoding RNA CCTT.

15. Copy number analysis and inference of subclonal populations in cancer genomes using Sclust

Yupeng Cun[1,5]*, Tsun-Po Yang[1,4]*, Viktor Achter[2,]*, Ulrich Lang[2,3], and Martin Peifer[1,4] (1:Department of Translational Genomics, Center for Integrated Oncology Cologne–Bonn, Medical Faculty, University of Cologne, 2:Computing Center, University of Cologne, 3:Department of Informatics, University of Cologne, 4:Center for Molecular Medicine Cologne (CMMC), University of Cologne, 5:Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences)
Email: cunyupeng@mail.kib.ac.cn

The genomes of cancer cells are constantly reshaped during pathogenesis. This evolutionary process leads to the emergence of subclonal populations, which can limit therapeutic interventions by the emergence of drug-resistance mutations. Data derived from massively parallel sequencing can be used to infer these subclonal populations from tumor-specific point mutations. The accurate determination of copy number changes and tumor impurity is an indispensable requirement to reliably infer these subclonal populations by mutational clustering. This protocol describes a copy number analysis method together with a novel mutational clustering approach. The method is called Sclust. In a series of simulations and comparisons with alternative methods, we showed that Sclust accurately determines copy number states and subclonal populations. Performance tests showed that the entire method is computationally extremely efficient. In particular, copy number analysis and mutational clustering takes less than 10 minutes. Sclust is designed that even non-experts in computational biology or bioinformatics with basic knowledge of the Linux/Unix command line syntax should be able to carry out analyses with Sclust.

16. 黑叶猴适应喀斯特环境的分子机制研究

张立业, 李明* (中国科学院动物研究所)
Email: lim@ioz.ac.cn

石山叶猴种组(Trachypithecus spp., Colobinae, Cercopithecidae)是灵长类中一组快速进化的类群，其栖居在特殊的喀斯特环境中。为探究其适应性辐射机制，我们对乌叶猴属(Trachypithecus)一只雄性黑叶猴(T. francoisi) 进行 279.47×高深度测序、组装。同时对该属 19 个体进行 30×重测序探究叶猴种群结构、物种分化 (T. francoisi, T. leucocephalus, T. hatinhensis, T. phayrei and T. germaini)。

17. k-mer 频数对 LncRNAs 亚细胞定位预测的影响

尤晓庆，陈颖丽* (内蒙古大学物理科学与技术学院)
Email: stchenyl@imu.edu.cn

长链非编码 RNA 是指一类长度大于 200 个核苷酸、不编码蛋白质的非编码 RNA。有研究表明，在哺乳动物的基因组中，只有不到 2%的转录产物能够编码蛋白质，高达 98%为非编码 RNA。目前，许多长链非编码 RNA 的序列已知，但是其功能却所知甚少，为了解长链非编码 RNA 的功能信息，获得它的亚细胞位置是非常重要的。本文分别对 k-mer 频数，三维阅读框，柔性信息三个特征利用 SMOTE 进行数据集平衡并采用支持向量机的方法进行了分类预测，取得了较好的预测效果。

18. 基于序列柔性信息及 PseKNC 对大肠杆菌启动子的预测

郭东华，陈颖丽* (内蒙古大学物理科学与技术学院)
stchenyl@imu.edu.cn

在细菌中，启动子由 RNA 聚合酶核心酶与相应的 Sigma(σ)因子共同识别，根据分子量的不同，将与大肠杆菌 K-12 启动子结合的σ因子分为 7 种类型: σ19、σ24、σ28、σ32、σ38、σ54 和 σ70，每种σ因子所识别的序列都具有一定的特征。通过对σ38 的二联体保守性分析，σ38 启动子虽然属于σ70 家族，但是我们发现σ38 的二联体保守性位点不同于σ70，其在转录起始位点上游-35 位点附近无保守区域，因而需要进一步对σ38 启动子进行研究。本文利用位置关联权重矩阵结合 DNA 柔性参数及"PseKNC"网站对大肠杆菌σ38 启动子进行了预测。jackknife 检验下，平均预测结果分别为 82.45%、87%。

19. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice

Shuai Jin[1], Yuan Zong[1], Qiang Gao[1], Zixu Zhu, Yanpeng Wang , Peng Qin, Chengzhi Liang, Daowen Wang, Jin-Long Qiu, Feng Zhang, Caixia Gao* (State Key Laboratory of Plant Cell and Chromosome Engineering, Center for Genome Editing, Institute of Genetics and Developmental Biology, The Innovative Academy of Seed Design, Chinese Academy of Sciences; University of Chinese Academy of Sciences)

Cytosine and adenine base editors (CBEs and ABEs) are promising new tools for achieving the precise genetic changes required for disease treatment and trait improvement. However, genome-wide and unbiased analyses of their off-target effects in vivo are still lacking. Our whole genome sequencing (WGS) analysis of rice plants treated with BE3, high-fidelity BE3 (HF1-BE3), or ABE revealed that BE3 and HF1-BE3, but not ABE, induce substantial genome-wide off-target mutations, which are mostly the C>T type of single nucleotide variants (SNVs) and appear to be enriched in genic regions. Notably, treatment of rice with BE3 or HF1-BE3 in the absence of single-guide RNA also results in the rise of genome-wide SNVs. Thus, the base editing unit of BE3 or HF1-BE3 needs to be optimized in order to attain high fidelity.

20. METTL3-mediated N6-methyladenosine mRNA modification enhances long-term memory consolidation

Zeyu Zhang[1,5,*], Meng Wang[1,*], Dongfang Xie[1], Zenghui Huang[1,5], Lisha Zhang[1], Ying Yang[2], Dongxue Ma[1], Wenguang Li[1], Qi Zhou[3,4,5],Yun-Gui Yang[2,4,5] and Xiu-Jie Wang[1,4,5] (1:Key Laboratory of Genetic Network Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, 2:Key Laboratory of Genomic and Precision Medicine, Collaborative Innovation Center of Genetics and Development, Beijing Institute of Genomics, Chinese Academy of Sciences, 3:State Key Laboratory of Stem Cell and Reproductive Biology, Institute of Zoology, Chinese Academy of Sciences, 4:Institute for Stem Cell and Regeneration, Chinese Academy of Sciences, 5:University of Chinese Academy of Sciences,
* These two authors contributed equally to this work.)

The formation of long-term memory is critical for learning ability and social behaviors of humans and animals, yet its underlying mechanisms are largely unknown. We found that the efficacy of hippocampus-dependent memory consolidation is regulated by METTL3, an RNA N6-methyladenosine (m6A) methyltransferase, through promoting the translation of neuronal early-response genes. Such effect is exquisitely dependent on the m6A methyltransferase function of METTL3. Depleting METTL3 in mouse hippocampus reduces memory consolidation ability, yet unimpaired learning outcomes can be achieved if adequate training was given or the m6A methyltransferase function of METTL3 was restored. The abundance of METTL3 in wild-type mouse hippocampus is positively correlated with learning efficacy, and overexpression of METTL3 significantly enhances long-term memory consolidation. These findings uncover a direct role of RNA m6A modification in regulating long-term memory formation, and also indicate that memory efficacy difference among individuals could be compensated by repeated learning.

21. MSTD: an efficient method for detecting multi-scale topological domains from symmetric and asymmetric 3D genomic maps

Yusen Ye[1], Lin Gao[1,*], and Shihua Zhang[2,*]
(1:Xidian University, 2:Academy of Mathematics and Systems Science, CAS)
Email: zsh@amss.ac.cn

The chromosome conformation capture (3C) technique and its variants have been employed to reveal the existence of a hierarchy of structures in three-dimensional (3D) chromosomal architecture, including compartments, topologically associating domains (TADs), sub-TADs and chromatin loops. However, existing methods for domain detection were only designed based on symmetric Hi-C maps, ignoring long-range interaction structures between domains. To this end, we proposed a generic and efficient method to identify multi-scale topological domains (MSTD), including cis- and trans-interacting regions, from a variety of 3D genomic datasets. We first applied MSTD to detect promoter-anchored interaction domains (PADs) from promoter capture Hi-C datasets across 17 primary blood cell types. The boundaries of PADs are significantly enriched with one or the combination of multiple epigenomic factors. Moreover, PADs between functionally similar cell types are significantly conserved in terms of domain regions and expression states. Cell type-specific PADs involve in distinct cell type-specific activities and regulatory events by dynamic interactions within them. We also employed MSTD to define multi-scale domains from typical symmetric Hi-C datasets and illustrated its distinct superiority to the-state-of-art methods in terms of accuracy, flexibility and efficiency.

22. Learning common and specific patterns from data of multiple interrelated biological scenarios with matrix factorization

Lihua Zhang and Shihua Zhang* (Academy of Mathematics and Systems Science, CAS)
Email: zsh@amss.ac.cn

High-throughput biological technologies (e.g., ChIP-seq, RNA-seq and single-cell RNA-seq) rapidly accelerate the accumulation of genome-wide omics data in diverse interrelated biological scenarios (e.g., cells, tissues and conditions). Differential analysis and pattern identification are two common paradigms for exploring and analyzing such data. However, they are typically used in a separate or/and sequential manner. In this study, we propose a flexible non-negative matrix factorization framework CSMF to combine them into one paradigm to simultaneously reveal common and specific patterns from data generated under interrelated biological scenarios. We demonstrate the effectiveness of CSMF with four applications including pairwise ChIP-seq data describing the chromatin modification map on protein-DNA interactions between K562 and Huvec cell lines; pairwise RNA-seq data representing the expression profiles of two cancers (breast invasive carcinoma and uterine corpus endometrial carcinoma); RNA-seq data of three breast cancer subtypes; and single-cell sequencing data of human embryonic stem cells and differentiated cells at six time points. Extensive analysis yields novel insights into hidden combinatorial patterns embedded in these interrelated multi-modal data. Results demonstrate that CSMF is a powerful tool to uncover common and specific patterns with significant biological implications from data of interrelated biological scenarios.

23. Joint prediction of cis-regulatory DNA interactions across multiple tissues using single-cell chromatin accessibility data

Kangning Dong, Shihua Zhang* (Academy of Mathematics and Systems Science, CAS)
Email: zsh@amss.ac.cn

Chromatin accessibility of cis-regulatory DNA elements delineate the in vivo availability of binding sites to transcription factors (TFs) and it now can be measured in single-cell resolution. Recently, the single-cell chromatin accessibility data have been employed to connect regulatory DNA elements to target genes. However, existing method for interaction map reconstruction using single-cell chromatin accessibility data only works on cells sampled from the same condition. However, regulatory networks of different tissues cannot be directly compared due to the varied number of cells as well as data sparsity in different tissues. To this end, we develop JPRIM (Joint Prediction of cis-Regulatory Interactions Maps) to explore common and tissue-specific regulatory interactions across multiple tissues based on patterns of co-accessibility in single-cell data. We have applied JPRIM onto single-cell ATAC-seq datasets across 13 tissues of adult mice containing ~100,000 cells and ~400,000 potential regulatory elements. We find that the common interaction related genes are significantly enriched in housekeeping gene set and the tissue-specific genes show higher activity in their corresponding tissues. Furthermore, differential activity genes show significant relevance with tissue-specific biological functions and tissue-specific interactions are related to functional TFs.

附录

国家数学与交叉科学中心

国家数学与交叉科学中心(NCMIS)（以下简称“交叉中心”）成立于 2010 年 11 月 24 日，是根据 2010 年 3 月 31 日国务院第 105 次常务会议精神和中国科学院“创新 2020”组织实施方案的总体部署而成立的非法人机构，依托单位是中国科学院数学与系统科学研究院。交叉中心经费由“创新 2020”计划支持。

交叉中心设立理事会、学术委员会、国际咨询委员会与执行委员会，实行中心主任负责制。现有六个交叉研究部和一个分中心，分别是：数学与信息技术交叉研究部、数学与经济金融交叉研究部、数学与先进制造交叉研究部、数学与材料环境交叉研究部、数学与生物/医学交叉研究部、数学与物理/工程交叉研究部以及合肥分中心。

交叉中心的成立,旨在从国家层面搭建一个数学与其它学科交叉合作的高水平研究平台；通过体制机制创新，凝聚国内外数学及相关学科力量，协同攻关，成为国际一流的科学研究基地。

交叉中心将通过设立和组织重大研究专题、承担国家重大项目、组织数学与相关学科交叉论坛、设立交叉型研究生与博士后培养计划，开展数学及其与自然科学、工程技术与社会经济的交叉研究、合作交流与人才培养。

青年创新促进会概况

中国科学院青年创新促进会 2011 年 6 月 17 日宣布成立,全院 340 名优秀青年科技工作者成为首批会员，每年将在学术交流、科研活动、培训与个人补贴等方面获得 10 万元专项经费资助。

根据《中国科学院“创新 2020”人才发展战略》(科发党字〔2011〕1 号)，中科院成立“中国科学院青年创新促进会”（以下简称“促进会”），通过加强对青年科技人才的培养和支持，促进其提升科研活动组织能力和综合素质，拓宽学术视野，造就新一代学术技术带头人。

中科院院长白春礼在 当日举行的首届“中国科学院人才发展主题活动日”上介绍说，成立“青年促进会”是为了落实“创新 2020”相关人才举措，全面提升中科院 35 岁以下优秀青 年科技人才的创新能力、领导能力和交流合作能力，培养具有较高思想品德、善于把握科技前沿、能够带领团队进行自主创新的新一代学术技术带头人。

“这仅仅是中科院实施人才战略的众多举措之一，中科院将以高层次人才和青年人才为重点，全面提 升队伍整体创新能力。”白春礼说，为保障“创新 2020”战略目标的实现，中科院必须把人才工作放在各项工作的首要位置，“通过积极落实‘千人计划’、‘百人计划’、‘创新人才推进计划’等人才工程，加大引进高层次人才的力度和对国内尖子人才的支持加强对青年人才的扶持和培养。”

中科院副院长、人才工作领导小组组长詹文龙表示，中科院将突破人才发展的体制机制束缚，努力营造促进各类人才脱颖而出的环境和氛围，包括建立动态优化的流动机制、公平合理的收入分配秩序，完善各类人才表彰奖励机制，及时解决人才生活条件方面的后顾之忧。

贝瑞和康 宣传海报

BerryGenomics 贝瑞基因

Berry NGS 科技服务

基因组学综合解决方案

• 平台丰富 • 技术全面

满足不同物种、不同复杂程度的基因组研究

PacBio Sequel

- 超长读长轻松构建Mb级Contigs
- 大片段结构变异检测

BioNano Saphyr /Irys

- 光学图谱技术实现染色体臂的完整组装
- 超长读长检测大片段结构变异

Illumina NovaSeq

- 高准确率矫正PacBio数据
- 低成本开展大规模、超深度测序

10x Genomics

- 低成本构建Super-Scaffolds
- 高通量单细胞转录组测序

Hi-C 三维技术

- 无需群体数据、成本低、周期短
- 辅助搭建染色体水平基因组图谱

北京贝瑞和康生物技术有限公司
Berry Genomics Co., Ltd.
地址：北京市昌平区科技园区生命园路4号院5号楼

电 话：010-8440 9702
邮 箱：sequence@berrygenomics.com
网 址：www.berrygenomics.com

果壳生物 宣传海报

果壳生物
BIOGUOKE

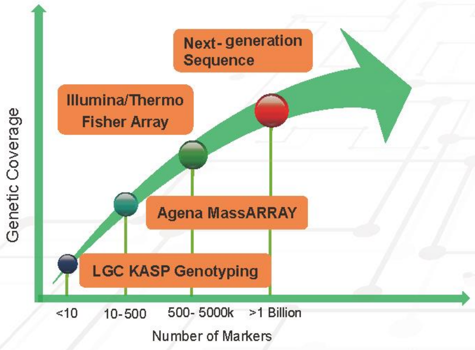
基因检测数字平台

Digital platform for gene detection

果壳生物助力健康大数据的挖掘，促进用户间个人基因组数据交换、个人基因组与其他数据交互，旨在形成个人及家庭精准健康管理新模式。同时，公司将依托大数据平台进行深度机器学习及算法模型的创新，基于多组学数据平台开发遗传病的辅助诊断系统、肿瘤早期诊断系统等，为医院各科室提供科学、精准的分子诊断方案和临床平行辅助诊断系统。真正做到**“未病先防，已病防变”**的健康管理方式。

五大平台

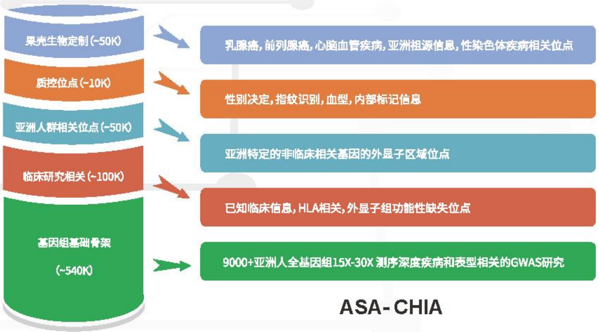
Flexible solutions for all approaches



明星产品 ASA-CHIA

MONGOLIA
CHINA
NORTH KOREA
SOUTH KOREA
JAPAN
MALAYSIA
SINGAPORE

ASA骨架人群分布图



果壳生物定制(~50K)	乳腺癌，前列腺癌，心脑血管疾病，亚洲祖源信息，性染色体疾病相关位点
质控位点(~10K)	性别决定，指纹识别，血型，内部标记信息
亚洲人群相关位点(~50K)	亚洲特定的非临床相关基因的外显子区域位点
临床研究相关(~100K)	已知临床信息，HLA相关，外显子组功能性缺失位点
基因组基础骨架 (~540K)	9000+亚洲人全基因组15X-30X 测序深度疾病和表型相关的GWAS研究

ASA-CHIA

合作单位

Cooperation with business units

西安交通大學 XIAN JIAOTONG UNIVERSITY

西湖大學 WESTLAKE UNIVERSITY

WIAS 浙江西湖高等研究院 WESTLAKE INSTITUTE FOR ADVANCED STUDY

北京大学第三医院 Peking University Third Hospital

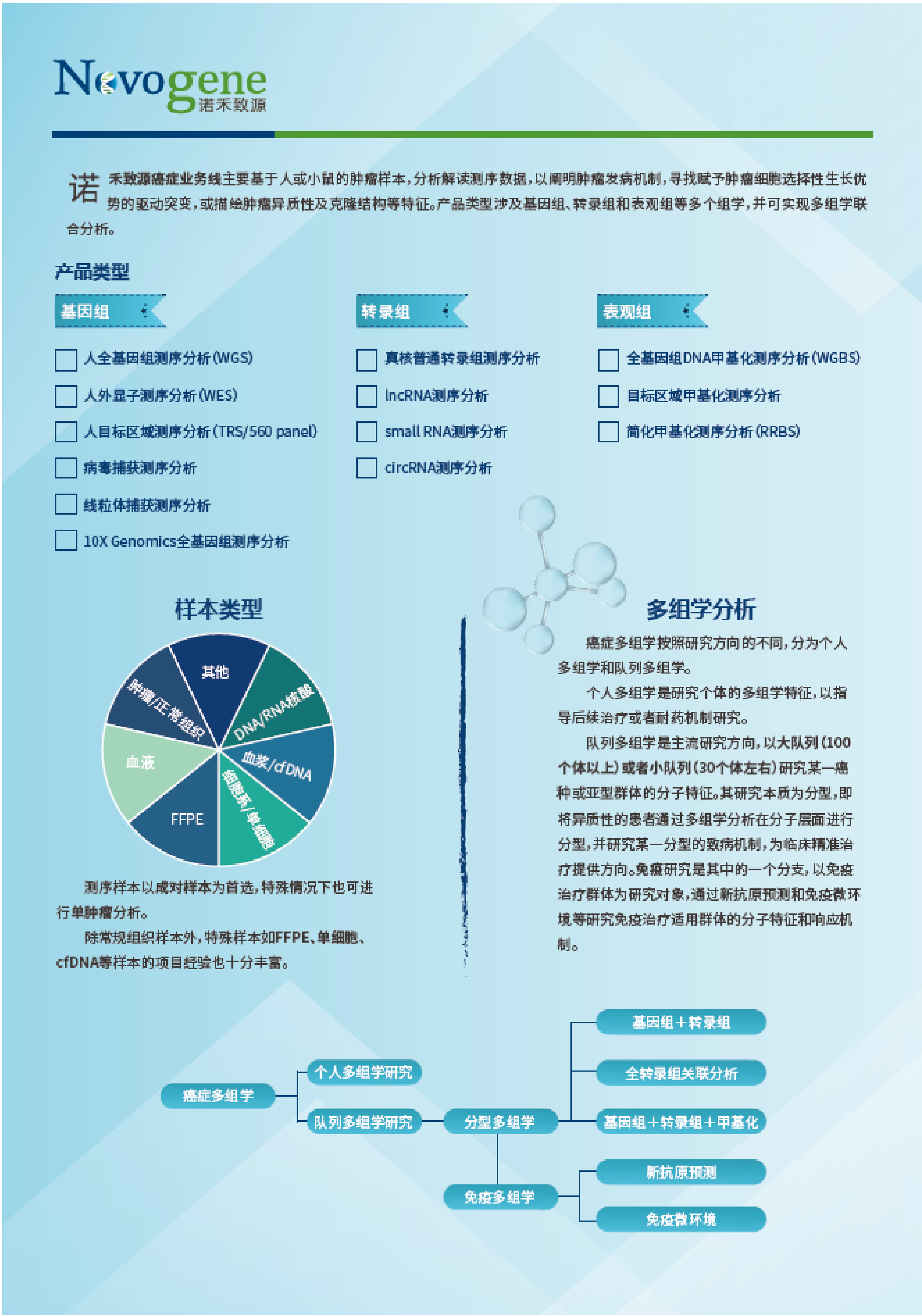
天津醫科大學 TIANJIN MEDICAL UNIVERSITY

中國科學院昆明動物研究所 KUNMING INSTITUTE OF ZOOLOGY .CAS

广西医科大学 Guangxi Medical University

北京果壳生物科技有限公司
电话：400 007 9358
邮箱：service@bioguoke.com
网址：www.bioguoke.com
地址：北京市昌平区中关村生命科学园生命园路8号院6号楼4层

诺禾致源 宣传海报



Novogene 诺禾致源

诺禾致源癌症业务线主要基于人或小鼠的肿瘤样本，分析解读测序数据，以阐明肿瘤发病机制，寻找赋予肿瘤细胞选择性生长优势的驱动突变，或描绘肿瘤异质性及克隆结构等特征。产品类型涉及基因组、转录组和表观组等多个组学，并可实现多组学联合分析。

产品类型

基因组

□ 人全基因组测序分析(WGS)

□ 人外显子测序分析(WES)

□ 人目标区域测序分析(TRS/560 panel)

□ 病毒捕获测序分析

□ 线粒体捕获测序分析

□ 10X Genomics全基因组测序分析

转录组

□ 真核普通转录组测序分析

□ lncRNA测序分析

□ small RNA测序分析

□ circRNA测序分析

表观组

□ 全基因组DNA甲基化测序分析(WGBS)

□ 目标区域甲基化测序分析

□ 简化甲基化测序分析(RRBS)

样本类型

其他

DNA/RNA核酸

血浆/cfDNA

细胞系/单细胞

FFPE

血液

肿瘤/正常组织

测序样本以成对样本为首选，特殊情况下也可进行单肿瘤分析。

除常规组织样本外，特殊样本如FFPE、单细胞、cfDNA等样本的项目经验也十分丰富。

多组学分析

癌症多组学按照研究方向的不同，分为个人多组学和队列多组学。

个人多组学是研究个体的多组学特征，以指导后续治疗或者耐药机制研究。

队列多组学是主流研究方向，以大队列（100个体以上）或者小队列（30个体左右）研究某一癌种或亚型群体的分子特征。其研究本质为分型，即将异质性的患者通过多组学分析在分子层面进行分型，并研究某一分型的致病机制，为临床精准治疗提供方向。免疫研究是其中的一个分支，以免疫治疗群体为研究对象，通过新抗原预测和免疫微环境等研究免疫治疗适用群体的分子特征和响应机制。

癌症多组学

个人多组学研究

队列多组学研究

分型多组学

基因组+转录组

全转录组关联分析

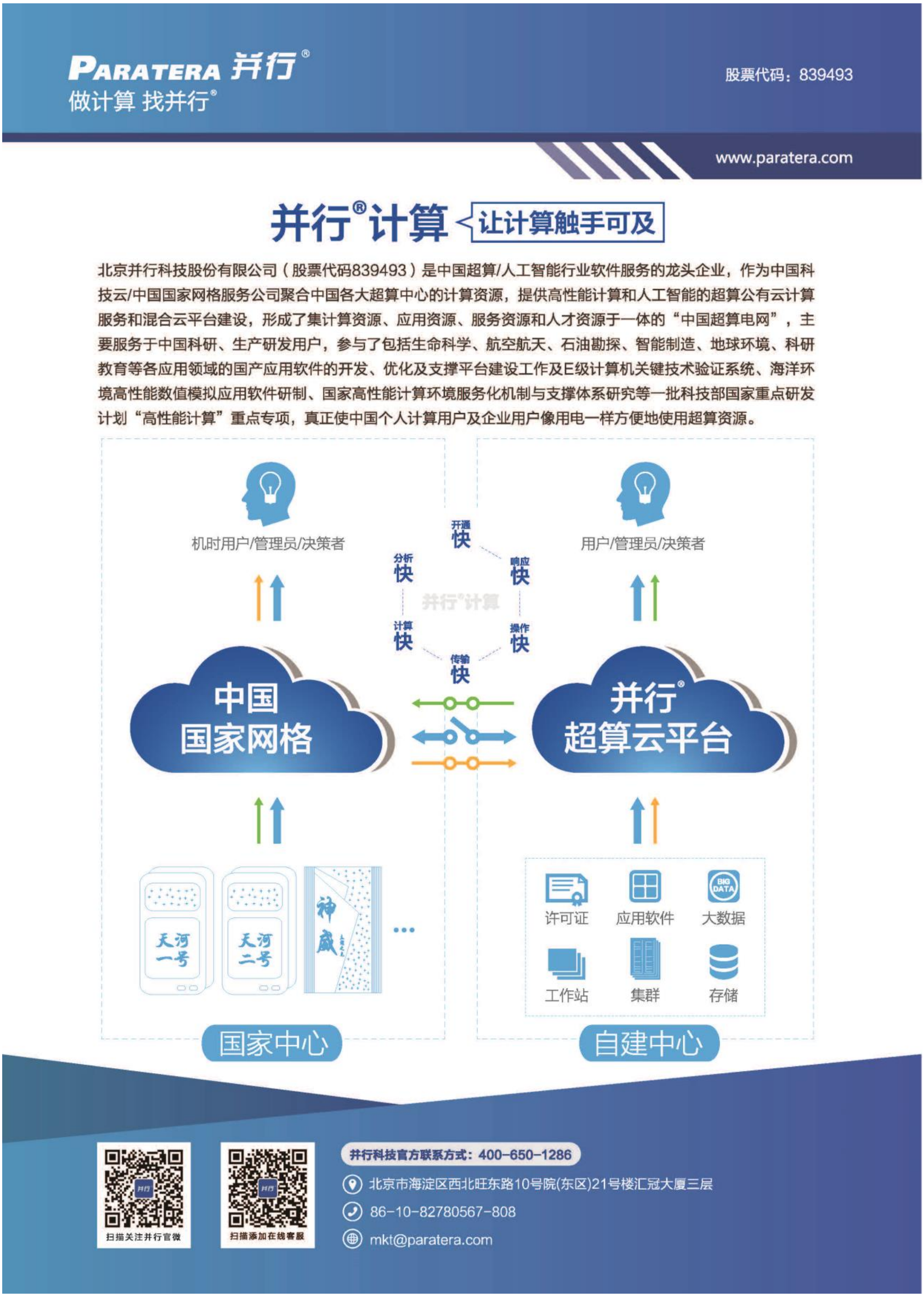
基因组+转录组+甲基化

免疫多组学

新抗原预测

免疫微环境

PARATERA 并行 宣传海报



PARATERA 并行®

做计算 找并行®

股票代码：839493

www.paratera.com

并行®计算 让计算触手可及

北京并行科技股份有限公司（股票代码839493）是中国超算/人工智能行业软件服务的龙头企业，作为中国科技云/中国国家网格服务公司聚合中国各大超算中心的计算资源，提供高性能计算和人工智能的超算公有云计算服务和混合云平台建设，形成了集计算资源、应用资源、服务资源和人才资源于一体的“中国超算电网”，主要服务于中国科研、生产研发用户，参与了包括生命科学、航空航天、石油勘探、智能制造、地球环境、科研教育等各应用领域的国产应用软件的开发、优化及支撑平台建设工作及E级计算机关键技术验证系统、海洋环境高性能数值模拟应用软件研制、国家高性能计算环境服务化机制与支撑体系研究等一批科技部国家重点研发计划“高性能计算”重点专项，真正使中国个人计算用户及企业用户像用电一样方便地使用超算资源。

机时用户/管理员/决策者

中国 国家网格

天河一号

天河二号

神威

国家中心

开通快

响应快

操作快

传输快

计算快

分析快

并行®计算

用户/管理员/决策者

并行® 超算云平台

许可证

应用软件

大数据

工作站

集群

存储

自建中心

扫描关注并行官微

扫描添加在线客服

并行科技官方联系方式：400-650-1286

北京市海淀区西北旺东路10号院(东区)21号楼汇冠大厦三层

86-10-82780567-808

mkt@paratera.com

Genomics, Proteomics & Bioinformatics 杂志海报

GPB

open access

IF 6.615
2018

Genomics, Proteomics & Bioinformatics

Special Issue

Single-Cell Omics Analysis

Guest Editors:

Luonan Chen
SIBS, CAS

Rong Fan
Yale University

Fuchou Tang
Peking University

To be published
Spring, 2020
Submission due
October 31, 2019

ELSEVIER

http://www.journals.elsevier.com/genomics-proteomics-and-bioinformatics

Journal of Genetics and Genomics 杂志海报

Founded in 1974

Monthly

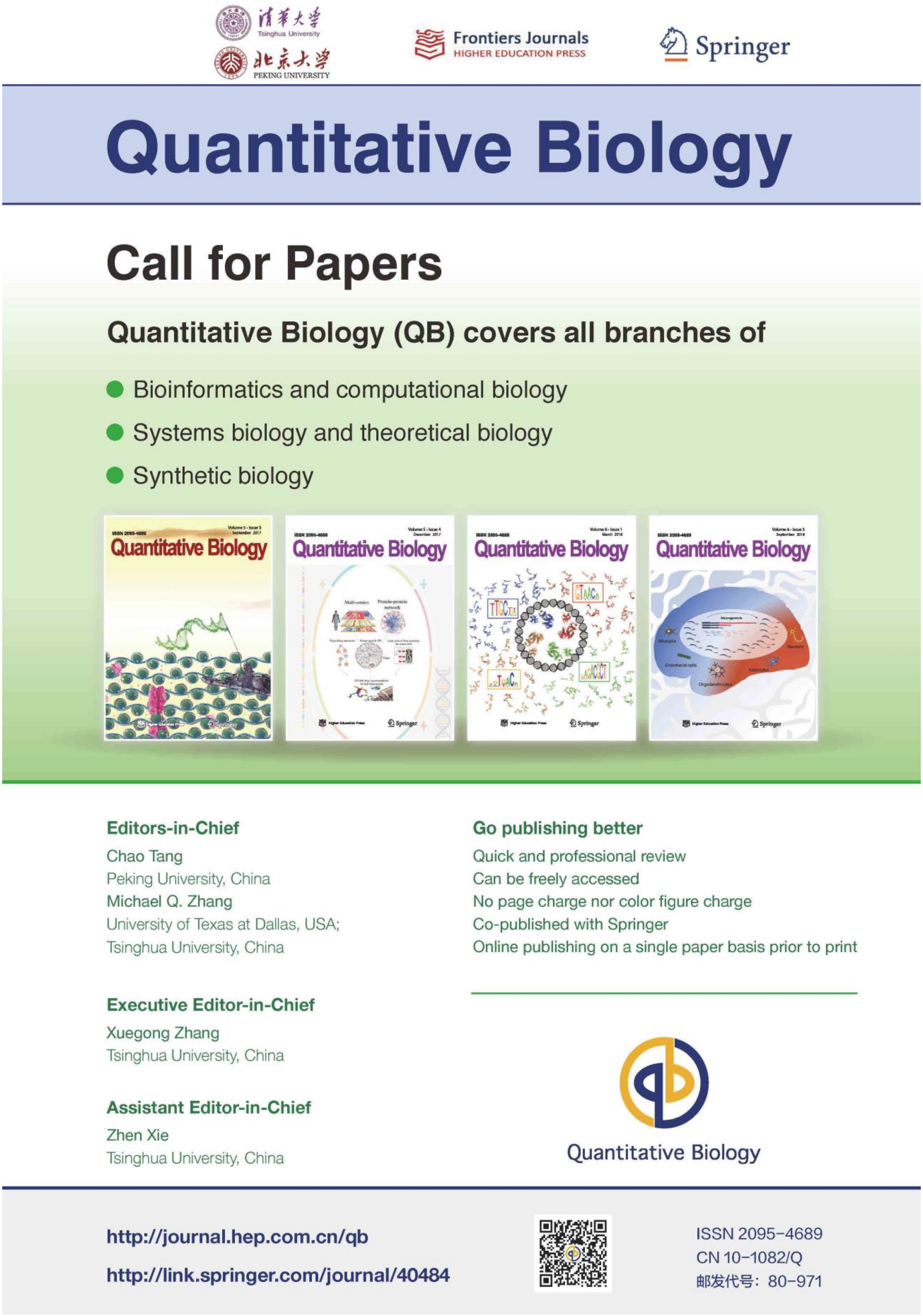
2017 Impact Factor: 4.066

Sponsored by Genetics Society of China and Institute of Genetics and Developmental Biology, Chinese Academy of Sciences

Published by Elsevier Ltd. and Science Press

wedsite: www.jgenetgenomics.org E-mail address: jgg@genetics.ac.cn

Quantitative Biology 杂志海报



Quantitative Biology

Call for Papers

Quantitative Biology (QB) covers all branches of

- Bioinformatics and computational biology
- Systems biology and theoretical biology
- Synthetic biology

Editors-in-Chief
Chao Tang
Peking University, China
Michael Q. Zhang
University of Texas at Dallas, USA;
Tsinghua University, China

Executive Editor-in-Chief
Xuegong Zhang
Tsinghua University, China

Assistant Editor-in-Chief
Zhen Xie
Tsinghua University, China

Go publishing better
Quick and professional review
Can be freely accessed
No page charge nor color figure charge
Co-published with Springer
Online publishing on a single paper basis prior to print

Quantitative Biology

http://journal.hep.com.cn/qb
http://link.springer.com/journal/40484

ISSN 2095−4689
CN 10−1082/Q
邮发代号：80−971

数学、计算机与生命科学
交叉研究青年学者论坛
2019年5月18-19日，北京

邀请信

尊敬的

为了加强国际学术交流，中国科学院数学与系统科学研究院/国家数学与交叉科学中心拟主办“数学、计算机与生命科学交叉研究青年学者论坛”。特别邀请您出席该次会议。

为加强从事 “数学、计算机与生命科学交叉研究”青年学者之间的联系，交流生命科学与计算生物学研究领域的最新成果，了解计算生物学国际发展动态和研究热点，促进数学、计算机与生命科学交叉研究与应用实践的迅速发展，我们特别组织“数学、计算机与生命科学交叉研究”青年学者论坛，该论坛的第六次活动将于 2019 年 5 月 18-19 日在北京召开。会议将邀请国内在“数学、计算机与生命科学交叉研究”取得突出成果的青年科学家 21 位作学术报告。报告的主题涵盖（但不局限于）计算生物学和计算基因组学、新一代测序数据分析与应用、癌症基因组学、表观基因组学、元基因组学、蛋白质组学以及系统生物学等研究领域。

将有一大批年富力强的计算生物学与生物信息学学者与研究生参会，交流前沿的学术工作。会议将围绕计算生物学与生物信息学的前沿课题，特别是生命科学领域与计算密切相关的基本问题，与数学、计算机以及其他学科背景的生物信息学工作者一起探讨、交流并建立联系。本次会议免注册费、食宿、交通费自理。

会议地点：北京
会议时间：2019 年 5 月 18－19 日，为期 2 天
举办单位：中国科学院数学与系统科学研究院
中国科学院遗传与发育生物学研究所
会议历史：这是该国际会议的第七届。
会议主题：生物大数据时代的机遇与挑战
会议主席：王秀杰（中国科学院遗传与发育生物学研究所）
张世华（中国科学院数学与系统科学研究院）

王秀杰 研究员
中国科学院遗传与发育生物学研究所遗传网络生物学所级重点实验室

张世华 研究员
中国科学院数学与系统科学研究院应用数学研究所

中国科学院数学与系统科学研究院
110102032974

数学、计算机与生命科学交叉研究
青年学者论坛

感谢您对本次会议的支持！

青年论坛会务组