

Introduction of Bioinformatics & AI for Biomedicine

陈铭

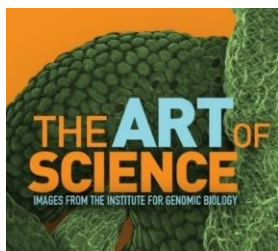
Introduction of Bioinformatics & AI for Biomedicine

Outline

1. Basic Concepts and Disciplinary Development of Bioinformatics: History and Trends
2. Research Fields and Core Content of Bioinformatics
3. Opportunities and Challenges in Bioinformatics: Big Data, Algorithms, and Integration
4. Artificial Intelligence (AI) and Bioinformatics
5. Talent Cultivation in Bioinformatics

§ 1: Basic Concepts and Disciplinary Development of Bioinformatics: History and Trends

Puzzle for Bio & Math



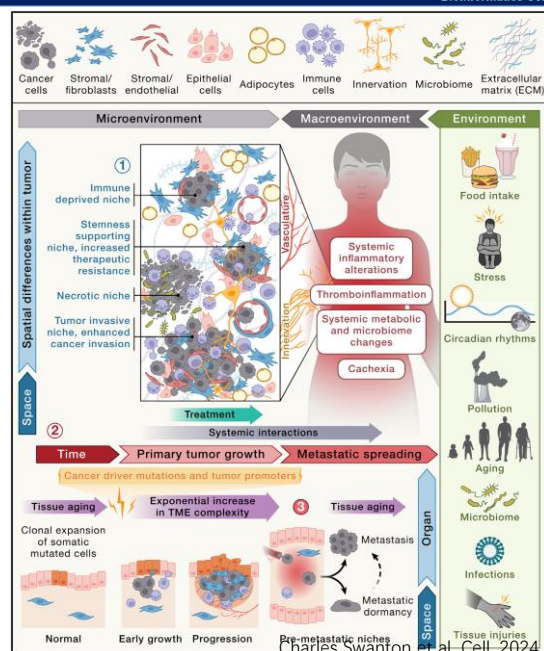
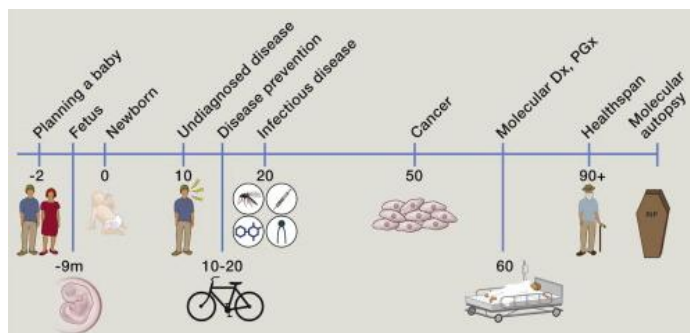
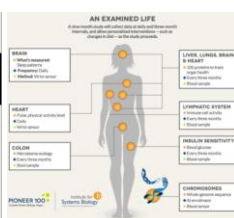
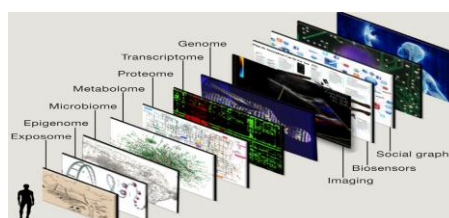
生命科学中的patterns



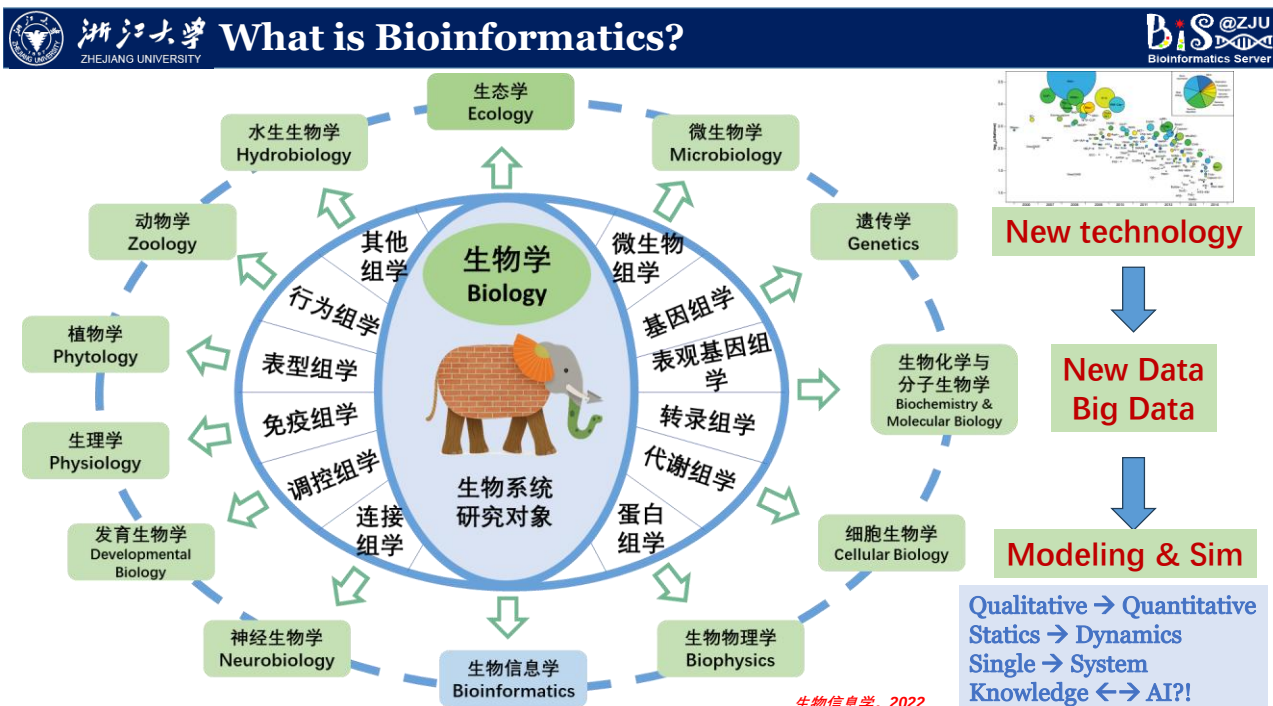
ScienceDaily (June 19, 2008)



Model to predict?

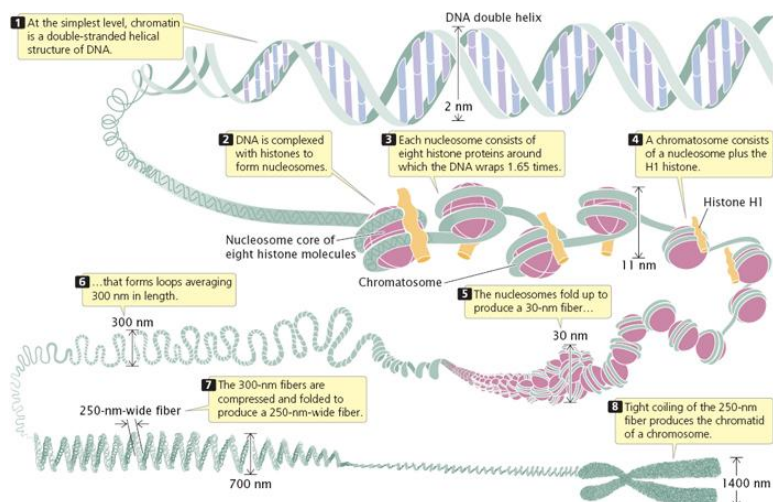


What is Bioinformatics?



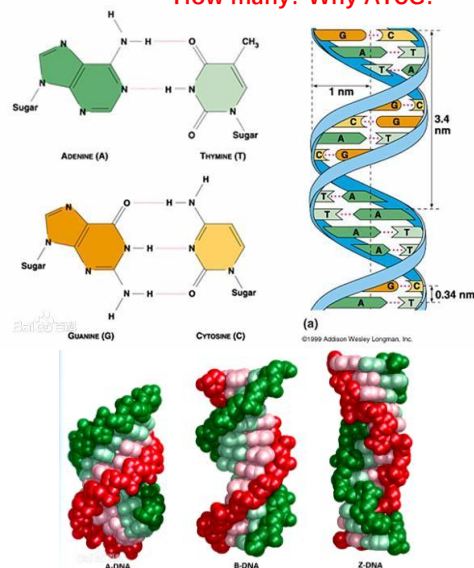
Why Bioinformatics?

Eukaryotic Genome Complexity



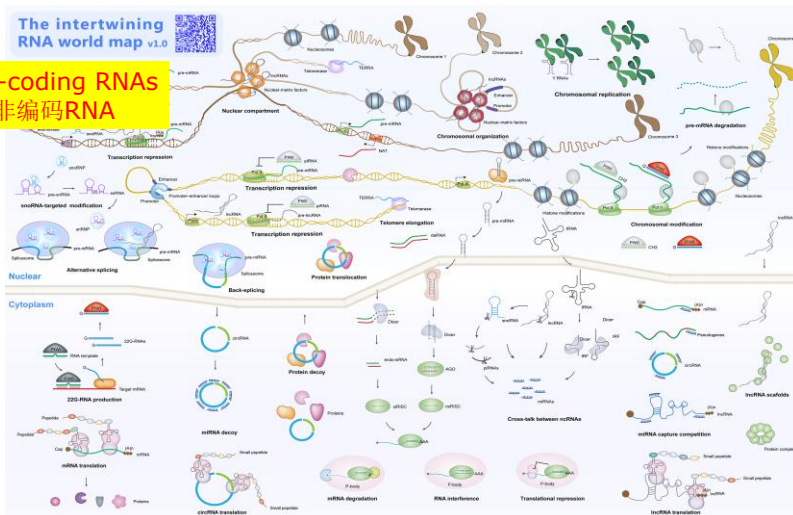
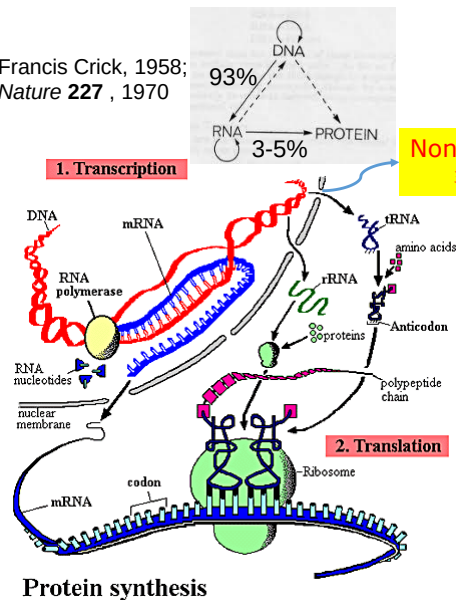
<http://www.whfreeman.com/>

How many? Why ATCG?



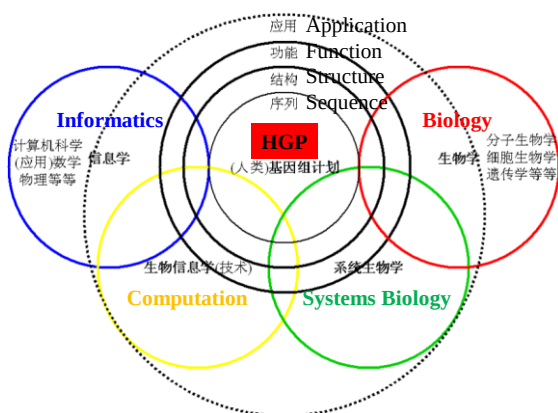
Central Dogma

Francis Crick, 1958;
Nature 227, 1970



Briefings in Bioinformatics, 2019

生物信息学是生物科学与信息科学的交叉学科，是利用计算机科学（信息学）的技术手段来研究生物学的的数据，如对生物数据进行获取(retrival)，存储(storage)，传输(transfer)，计算(manuipulation)，分析(analysis)，模拟(simulation)，预测(prediction)等等的一门新兴学科，是21世纪科学发展的热点之一。



Biology + Informatics = Bioinformatics

But

1 + 1 ≠ 2

Bioinformatics Positioning

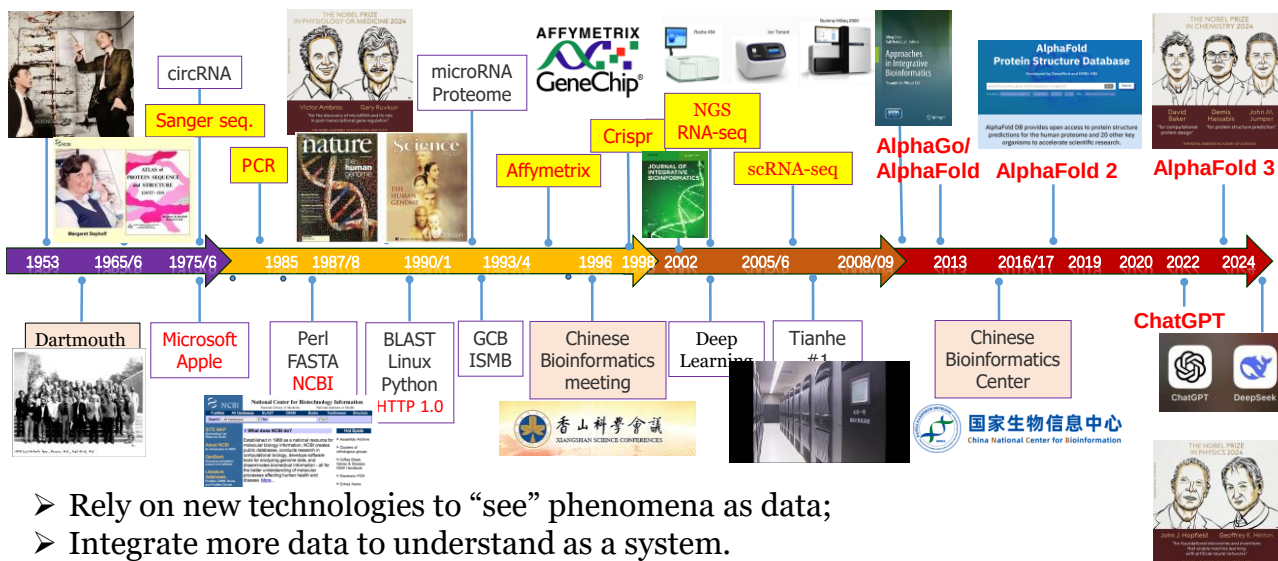
- Derived from Biology, Applied to Biology
- Omics-Specific Data Analysis
- Universal Methods and Technologies
- Integration of Multi-Omics to Grasp Holistic Characteristics and Discover New Laws

1953: Molecular Biology

1976: Genomics

2002: Post-genomics

2013: Big Data & AI

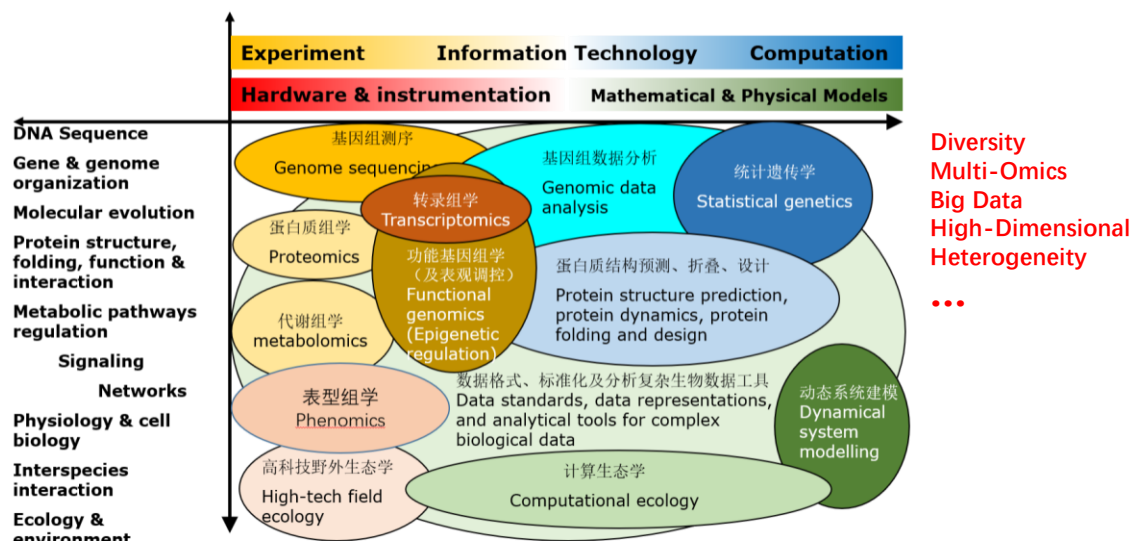


- Rely on new technologies to “see” phenomena as data;
- Integrate more data to understand as a system.

GWAS, AlphaFold, ChatGPT: Key Reflections for Us: ① Big Data, ② Effective Methods, ③ Integration!

§ 2: Research Fields and Core Content of Bioinformatics

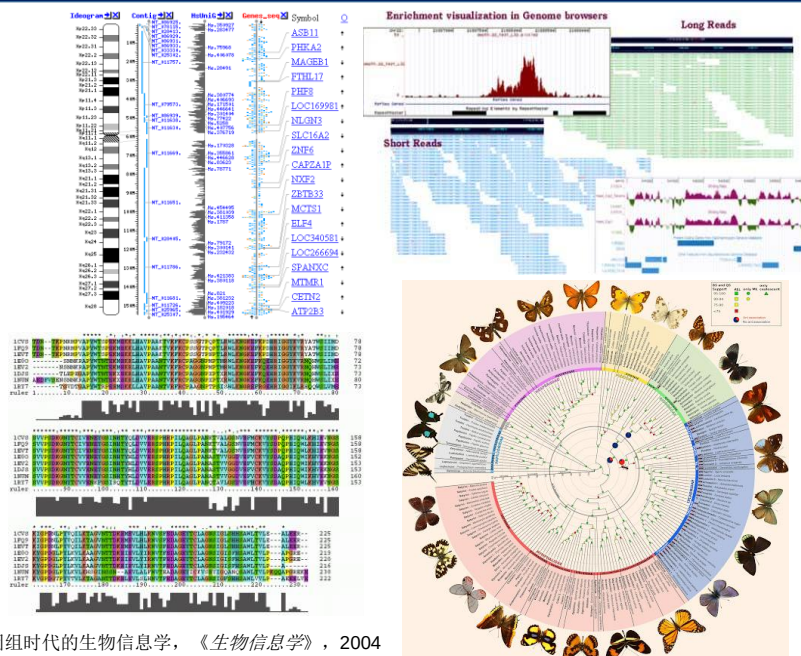
Bioinformatics & Omics



陈铭：整合生物信息学（2006）

Bioinformatics: Sequence analysis

- Database Platform
- Sequence Analysis
- Comparative Genomics (Metagenomics)
- Transcriptomics
- Transcriptional Regulation and Epigenetics
- Proteomics
- Metabolomics
- Phenomics
- Systems Biology
- Integrative Bioinformatics
- Synthetic Biology



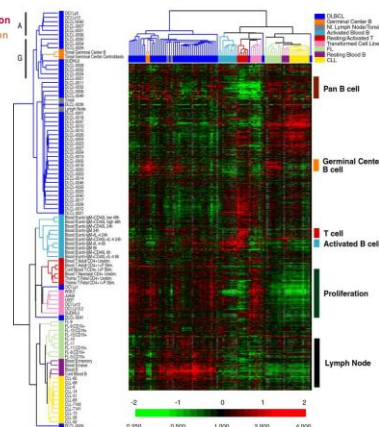
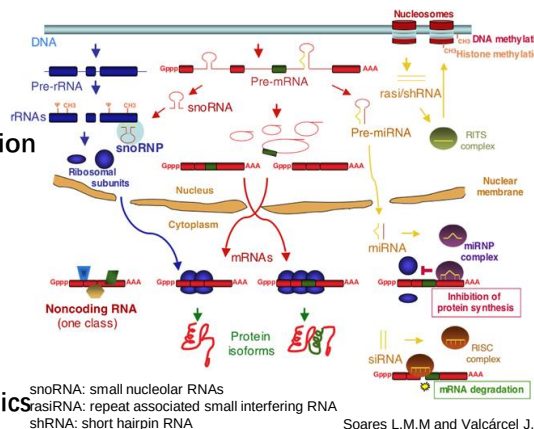
陈铭，后基因组时代的生物信息学，《生物信息学》，2004



- Database Platform
- Sequence Analysis
- Comparative Genomics (Metagenomics)
- Transcriptomics**
- Transcriptional Regulation and Epigenetics
- Proteomics
- Metabolomics
- Phenomics
- Systems Biology
- Integrative Bioinformatics
- Synthetic Biology

The **transcriptome** is the set of all RNA molecules, including mRNA, rRNA, tRNA, and other non-coding RNA produced in one or a population of cells.

生物多样性!
采样? 处理?



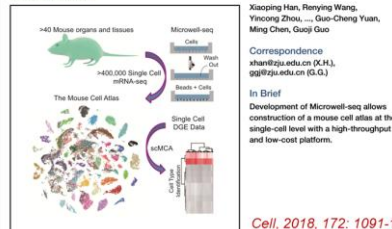
Single cell sequencing & webserver

Collaboration with Prof. Guo @ ZJU

Cell <http://bis.zju.edu.cn/MCA>

Mapping the Mouse Cell Atlas by Microvel-Seq

Graphical Abstract



nature

Human Cell Landscape, <http://bis.zju.edu.cn/HCL>



ARTICLE
<https://doi.org/10.1016/j.cell.2018.04.014> OPEN

A reference single-cell regulomic and transcriptomic map of cynomolgus monkeys

Jiao Q^{1,2}, Fei Yang^{1,2}, Tao Zhu^{1,2}, Yinghui Wang^{1,2}, Wen Fang¹, Yan Ding¹, Xue Zhao¹, Xianjie Qi¹, Qiangmin Xie¹, Ming Chen¹, Qiang Xu¹, Yicheng Xie^{1,2}, Yang Sun^{1,2,3} & Dijun Chen^{1,2,3}

Collaboration with Prof. Dijun Chen @ NJU

ARTICLE
<https://doi.org/10.1016/j.cell.2018.04.014> OPEN

墨西哥钝口螈正常与变态发育的单细胞转录组图谱

Construction of the axolotl cell landscape using combinatorial hybridization sequencing at single-cell resolution

Fang Ye^{1,2,3}, Guodong Zhang^{1,2}, Weigao Li^{1,2}, Haide Chen^{1,2}, Chenguang Yu^{1,2}, Lei Yang^{1,2}, Yuting Fu¹, Jiaqi Li¹, Sulei Fu¹, Zhongyi Sun¹, Lijiang Fei¹, Qile Guo¹, Jingjing Wang¹, Yanyu Xiao¹, Ximu Wang¹, Peijiang Zhang^{1,2}, Liang Ma¹, Dengqin Ge¹, Suibang Xu¹, Juan Caballero-Pérez^{1,2}, Alexandre Cruz-Ramirez^{1,2}, Yincang Zhou¹, Ming Chen¹, Ji-Feng Fei^{1,2,3}, Xiaoping Han^{1,2,3} & Guojin Guo^{1,2,3,4,5,6,7,8,9,10}

ARTICLE
<https://doi.org/10.1016/j.cell.2018.04.014> OPEN

Cell landscape of larval and adult *Xenopus laevis* at single-cell resolution

Yuan Luo^{1,2}, Liang Ma¹, Qile Guo^{1,2}, Weigao Li^{1,2}, Xing Fang^{1,2}, Lei Yang^{1,2}, Fanwei Ruan¹, Jingjing Wang^{1,2}, Peijiang Zhang^{1,2}, Zhongyi Sun¹, Haide Chen^{1,2}, Zhongliang Lin¹, Xumu Wang¹, Ximu Wang¹, Huiyu Sun¹, Xuan Fan¹, Yincang Zhou¹, Ming Chen¹, Wanhua Shang^{1,2,3}, Guojin Guo^{1,2,3,4,5,6,7,8,9,10} & Xiaoping Han^{1,2,3,4}

CELL Landscape

EVERY CELL MATTERS

Quick Access: <http://bis.zju.edu.cn/cellatla>

CELL Landscape Home Browse QSearch Annotation Help Download

Statistics

15 SPECIES, 6,052,867 CELLS, 18 PROJECTS, 7 TECHNOLOGIES

Cells Sequencing technology

Human Cell Landscape

Mouse Cell Landscape

NO Cell Landscape (NCL)

Scaphiopus Cell Landscape

Prochilopterus Cell Landscape

Adult Cell Landscape

Remora Cell Landscape

Scaphiopus Cell Landscape

Resources for other species

Nucleic Acids Research 2023



浙江大学
ZHEJIANG UNIVERSITY

Single-cell transcriptome analysis in Arabidopsis



Plant Communications
Research article

CellPress
Partner Journal

Single-cell transcriptome analysis dissects lncRNA-associated gene networks in Arabidopsis

Zhaohui He^{1,7}, Yangming Lan^{1,7}, Xinkai Zhou^{1,7}, Bianjiong Yu^{1,7}, Tao Zhu¹, Fa Yang¹, Liang-Yu Fu¹, Haoyu Chao², Jiahao Wang^{3,4}, Rong-Xu Feng⁵, Shimin Zuo^{3,4}, Wenzhi Lan¹, Chunli Chen⁶, Ming Chen^{2,*,}, Xue Zhao^{1,*,}, Keming Hu^{3,4,*,} and Dijun Chen^{1,*,}

¹State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing 210023, China

²Department of Bioinformatics, College of Life Sciences, Zhejiang University, Hangzhou 310058, China

³Jiangsu Key Laboratory of Crop Genomics and Molecular Breeding/Key Laboratory of Plant Functional Genomics of the Ministry of Education, College of Agriculture, Yangzhou University, Yangzhou 225009, China

⁴Co-Innovation Center for Modern Production Technology of Grain Crops of Jiangsu Province/Key Laboratory of Crop Genetics and Physiology of Jiangsu Province, Yangzhou University, Yangzhou 225009, China

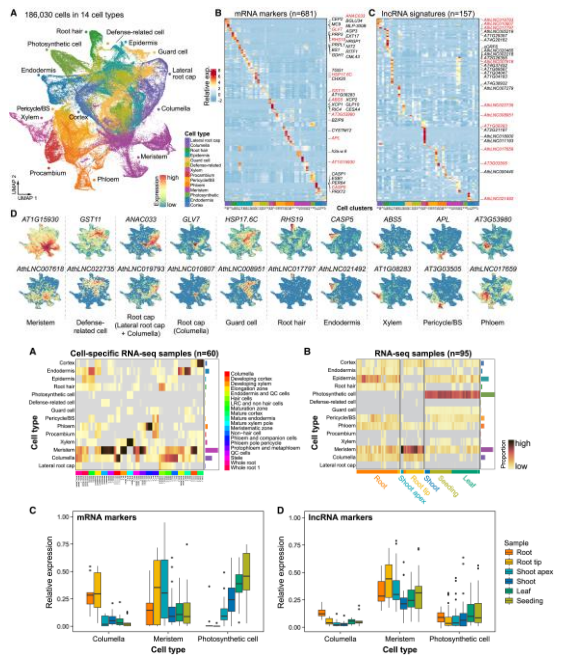
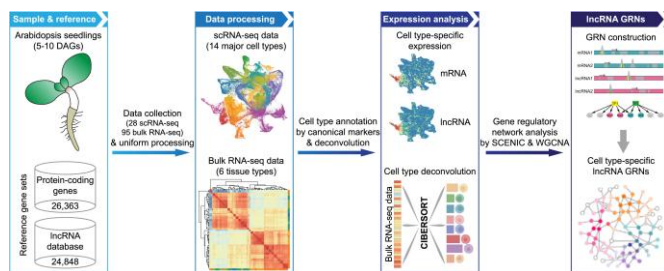
⁵Zhejiang Zhoushan High School, Zhoushan 316099, China

⁶National Key Laboratory for Germplasm Innovation and Utilization for Fruit and Vegetable Horticultural Crops, Hubei Hongshan Laboratory, Wuhan 430070, China

⁷These authors contributed equally to this article.

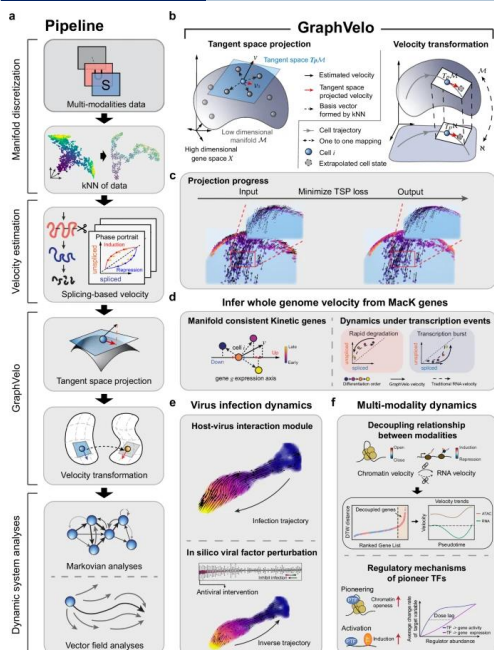
*Correspondence: Ming Chen (mchen@zju.edu.cn), Xue Zhao (zhaoxue@nju.edu.cn), Keming Hu (hukem@yzu.edu.cn), Dijun Chen (dijunchen@nju.edu.cn)

https://doi.org/10.1016/j.xpc.2023.100717



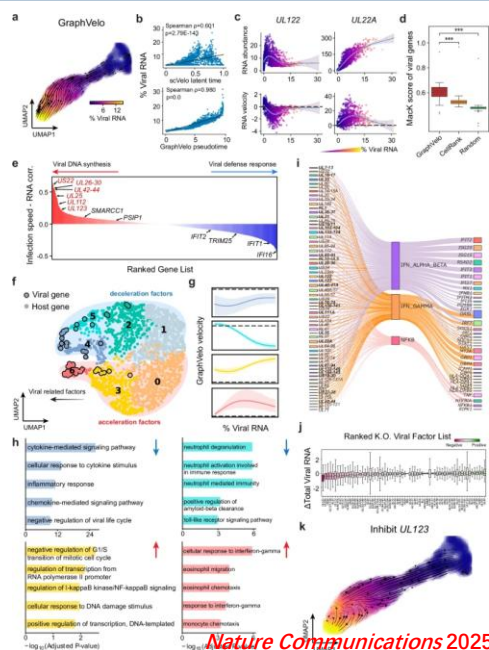
浙江大学
ZHEJIANG UNIVERSITY

Inference of multimodal velocities and molecular mechanisms for single cells



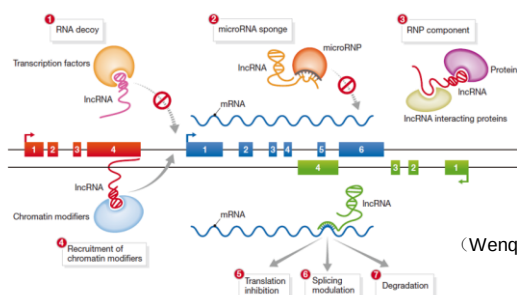
Refining RNA velocity by tangent space projection and transforming between representations using GraphVelo

Using GraphVelo velocities to infer host-virus infection trajectory and identify host-pathogen interactions

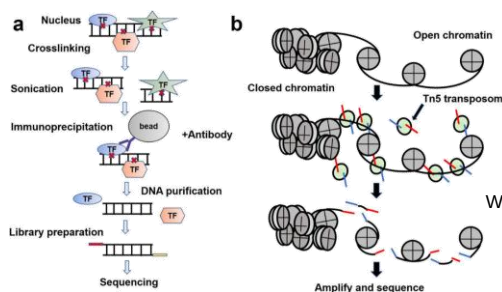


Nature Communications 2025

- Database Platform
- Sequence Analysis
- Comparative Genomics (Metagenomics)
- Transcriptomics
- **Transcriptional Regulation and Epigenetics**
- Proteomics
- Metabolomics
- Phenomics
- Systems Biology
- Integrative Bioinformatics
- Synthetic Biology

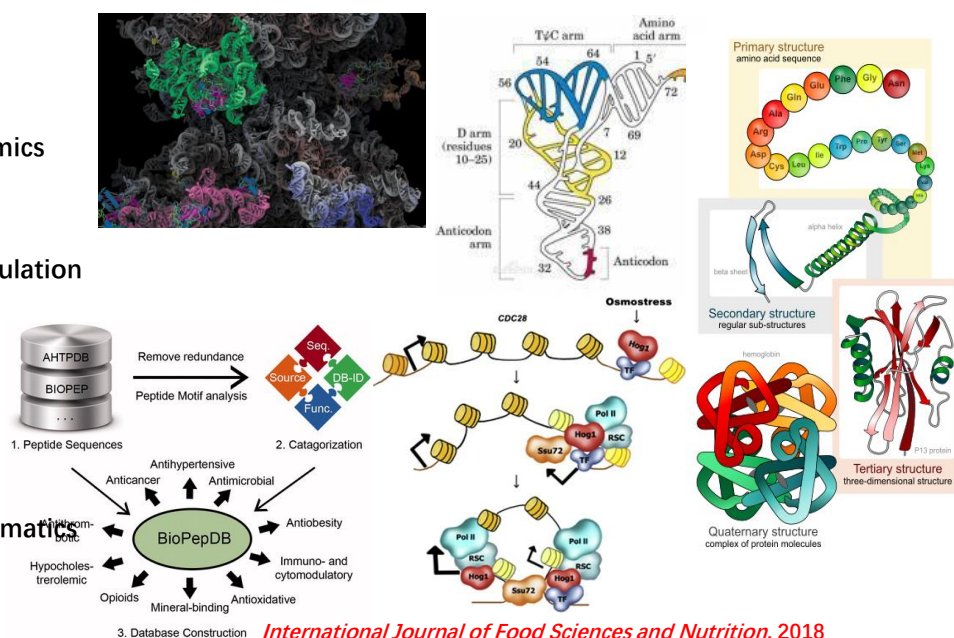


lncRNA的调控
(Wengian Hu, et al., EMBO reports. 2012)



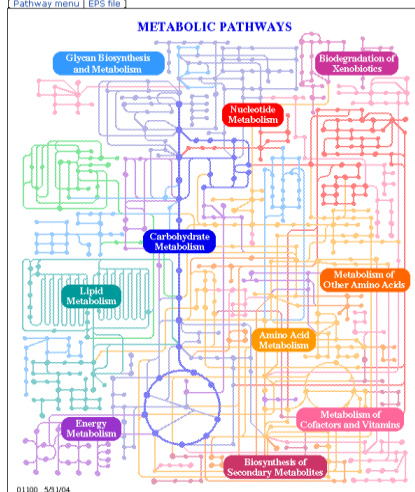
Workflows of ChIP-seq and ATAC-seq.
(Ma, et al., Mol Biomed. 2020)

- Database Platform
- Sequence Analysis
- Comparative Genomics (Metagenomics)
- Transcriptomics
- Transcriptional Regulation and Epigenetics
- **Proteomics**
- Metabolomics
- Phenomics
- Systems Biology
- Integrative Bioinformatics
- Synthetic Biology

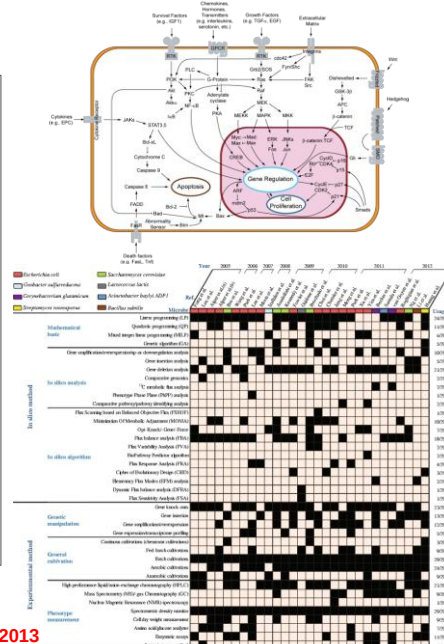


- Database Platform
- Sequence Analysis
- Comparative Genomics (Metagenomics)
- Transcriptomics
- Transcriptional Regulation and Epigenetics
- Proteomics
- Metabolomics**
- Phenomics
- Systems Biology
- Integrative Bioinformatics
- Synthetic Biology

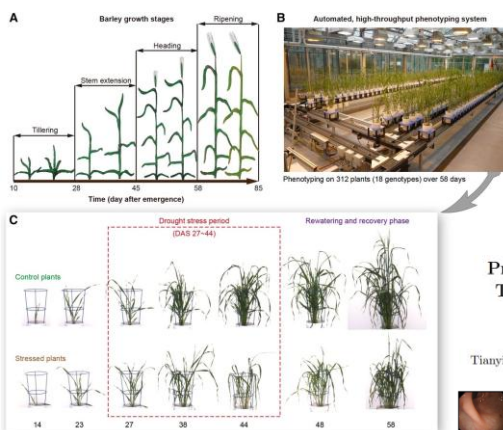
KEGG Metabolism - Reference pathway
Pathway menu | EPS file



Applied Microbiology and Biotechnology, 2013



- Database Platform
- Sequence Analysis
- Comparative Genomics (Metagenomics)
- Transcriptomics
- Transcriptional Regulation and Epigenetics
- Proteomics
- Metabolomics
- Phenomics**
- Systems Biology
- Integrative Bioinformatics
- Synthetic Biology



Dijun Chen et al. *Plant Cell* 2014

ARTICLE

10.1093/plcr/kdu003-018-0091-x

OPEN

The HTPmod Shiny application enables modeling and visualization of large-scale biological data

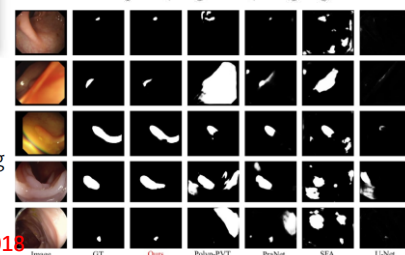
Dijun Chen^{1,2}, Liang-Yu Fu¹, Dahui Hu³, Christian Klukas^{2,4}, Ming Chen³ & Kerstin Kaufmann¹

Communication Biology, 2018

生物多样性!
常态? 动态?
关联分析!

Probabilistic Modeling Ensemble Vision Transformer Improves Complex Polyp Segmentation

Tianyi Ling^{1,2,†}, Chengyi Wu^{2,3,†}, Huan Yu^{2,4}, Tian Cai⁵, Da Wang¹,
Yincong Zhou², Ming Chen^{2,5,†}, Kefeng Ding^{1,5,†}

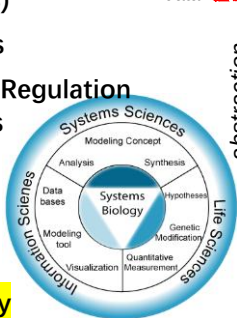




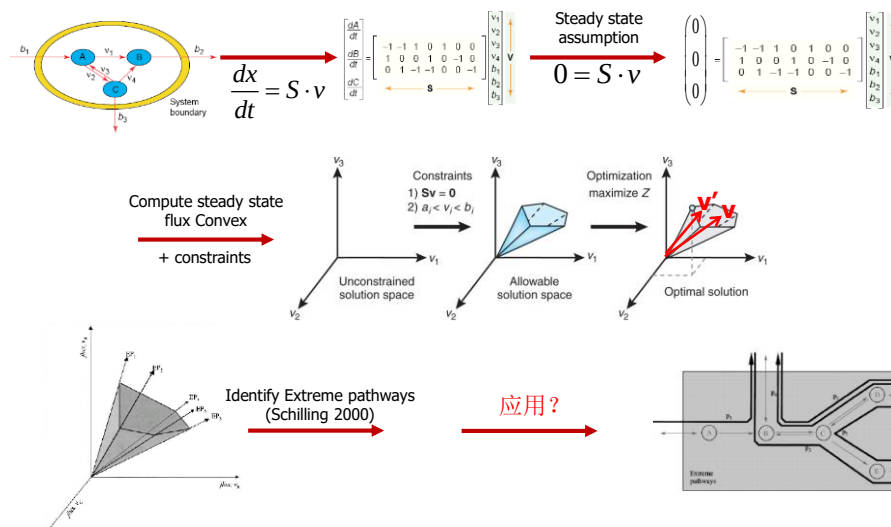
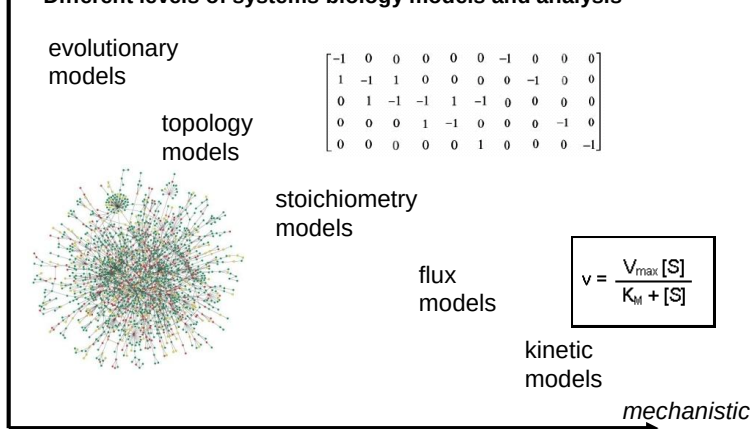
- Database Platform
- Sequence Analysis
- Comparative Genomics (Metagenomics)
- Transcriptomics
- Transcriptional Regulation and Epigenetics
- Proteomics
- Metabolomics
- Phenomics
- **Systems Biology**
- Integrative Bioinformatics
- Synthetic Biology

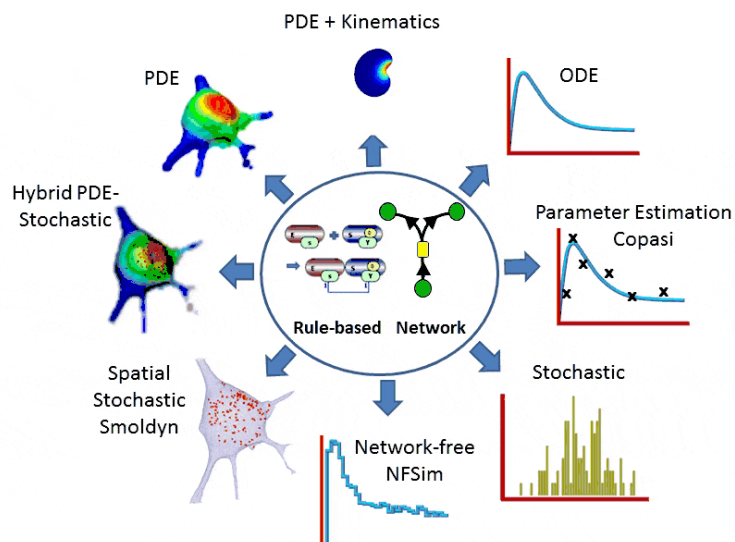
Leroy Hood, 1999

- The study of the mechanisms underlying complex biological processes as integrated systems of many interacting components. Systems biology involves:
- (1) collection of large sets of experimental data **数据**
 - (2) proposal of mathematical models that might account for at least some significant aspects of this data set **模型**
 - (3) accurate computer solution of the mathematical equations to obtain numerical predictions, **模拟**
 - (4) assessment of the quality of the model by comparing numerical simulations with the experimental data. **验证**



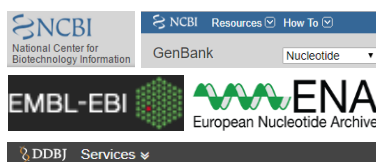
Different levels of systems biology models and analysis





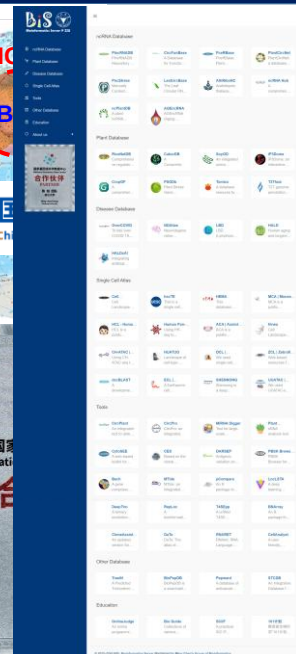
<https://vcell.org/>

- **Database Platform**
- **Sequence Analysis**
- **Comparative Genomics (Metagenomics)**
- **Transcriptomics**
- **Transcriptional Regulation and Epigenetics**
- **Proteomics**
- **Metabolomics**
- **Phenomics**
- **Systems Biology**
- **Integrative Bioinformatics**
- **Synthetic Biology**



Bioinformation and DDBJ Center

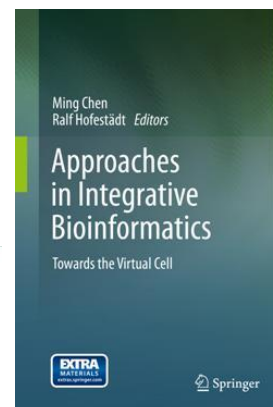
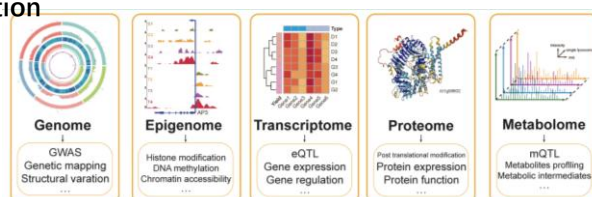
国家基因组科学数据中心
National Genomics Data Center



- Database Platform
- Sequence Analysis
- Comparative Genomics (Metagenomics)
- Transcriptomics
- Transcriptional Regulation and Epigenetics
- Proteomics
- Metabolomics
- Phenomics
- Systems Biology
- **Integrative Bioinformatics**
- Synthetic Biology

Research Domains of Integrative Bioinformatics

1. Biological Data Mining and Integration
2. Integration of Life Sciences and Bioinformatics Technologies
3. Integration of Disciplines and Talents



陈铭, 整合生物信息学, 《计算机教育》, 2006

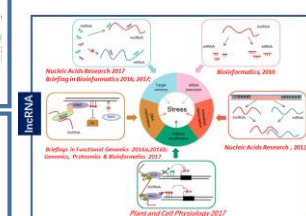
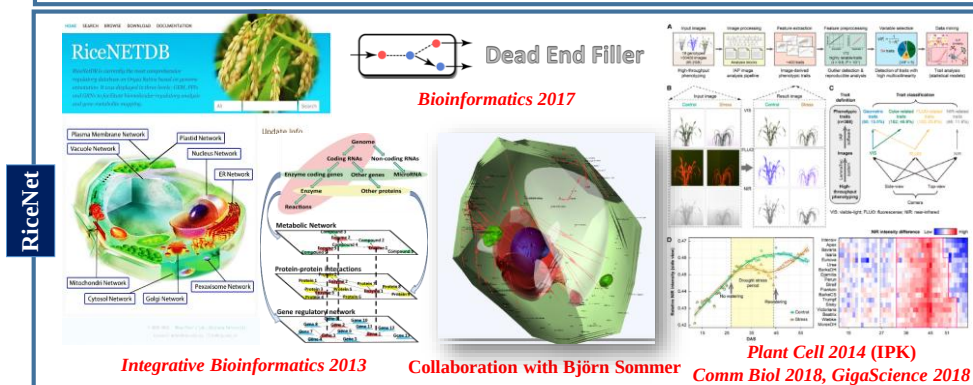
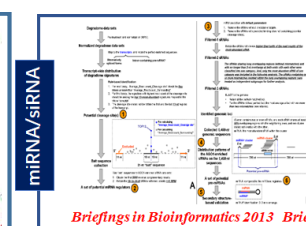
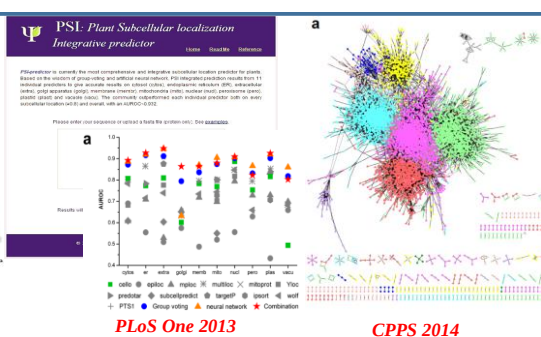
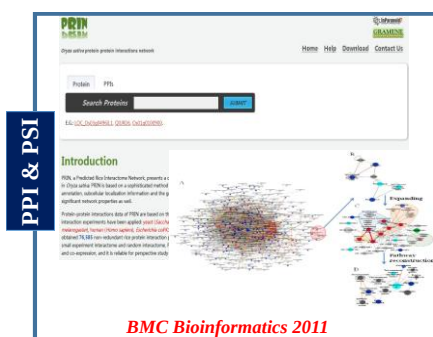
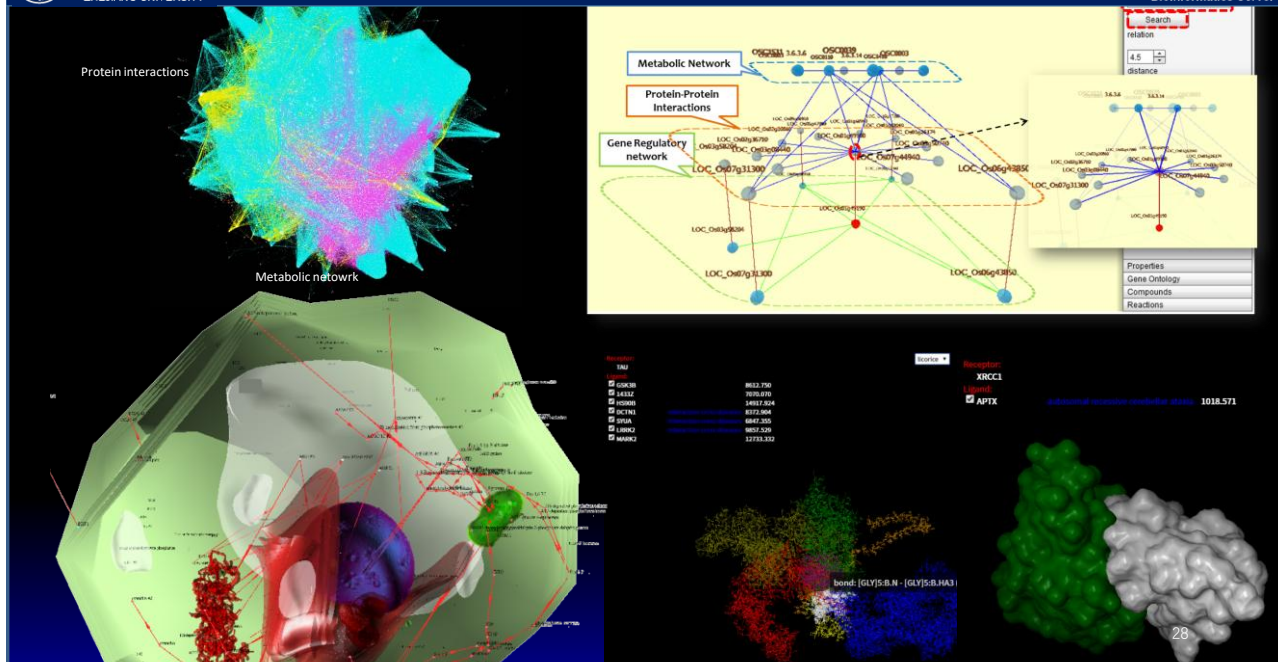


Chapter 11 Bridging Genomics and Phenomics

Dijun Chen, Ming Chen, Thomas Altmann, and Christian Klukas



Abstract Genomics and phenomics are two fundamentally important branches of biological sciences, and they stand at both ends of the multiple “omics” families. A central goal of current biology is to establish complete functional links between the genome and phenome, the so-called genotype–phenotype map. Recent advances in high-throughput and high-dimensional genotyping and phenotyping technologies enable us to uncover the casual networks inside the “black box” that lies between genotypes and phenotypes using the principles of genome-wide association studies (GWAS). Application of GWAS and analogous methodologies and incorporation of multiple omics data begin to unravel the contribution of genetic variation to phenotypic diversity. Integrating “omics” data at broad levels by using the systems-biology approach is paramount to further bridging the gaps between genomics and phenomics and eventually making accurate predictions of phenotypes based on genetic contribution.



- Molecular medicine 分子医学
- Personalized medicine 个性医学
- Preventative medicine 预防医学
- Gene therapy 基因治疗
- Drug development 药物研发
- Microbial genome applications 微生物基因组应用
 - Waste cleanup, Climate change Studies, Alternative energy sources, Biotechnology, Antibiotic resistance, Forensic analysis of microbes 法医检定, The reality of bioweapon creation
- Evolutionary studies 进化研究
- Crop improvement 作物改良
 - Insect resistance, Improve nutritional quality, Development of Drought resistance varieties
- Veterinary Science 兽医科学
- Comparative Studies 比较研究

精准医学
Precision
Medicine

精准药学
Precision
Pharmacy

智慧农学
精准育种
Smart Agriculture
Precision Breeding

Research Aligned with National Strategic Needs



On May 10, 2022, the National Development and Reform Commission (NDRC) issued the 14th Five-Year Plan for Bioeconomy Development (hereinafter referred to as the "Plan"). The Plan clearly proposes promoting the in-depth integration of advanced technologies such as genetic testing and biological genetics with disease prevention, conducting early screening for major diseases including genetic disorders, birth defects, tumors, cardiovascular diseases, and metabolic diseases, and **providing precise solutions and decision support for personalized treatment.**

China still faces numerous challenges in the innovative development of its bioeconomy. For instance, the capacity for original innovation remains relatively weak; **there is insufficient accumulation in basic life science theories, underlying key generic technologies, high-end instruments and reagents, and bioinformatics resources.** Additionally, the technology innovation system featuring enterprises as the mainstay, market orientation, and in-depth integration of industry, academia, and research is still imperfect.

§ 3: Opportunities and Challenges in Bioinformatics: Big Data, Algorithms, and Integration



浙江大学
ZHEJIANG UNIVERSITY

Bioinformatics: Challenges



Explosive Growth of Data Volume

- Breakthrough advancements in high-throughput sequencing technologies
- Expansion of abundant, heterogeneous biological data resources
- Proliferation of multi-omics technologies (genomics, transcriptomics, proteomics, etc.)

Imperatives for AI and Algorithm Innovation

- Rapid evolution of artificial intelligence methodologies for large-scale biological data processing
- Escalating demand for advanced algorithm applications (e.g., deep learning, graph neural networks) in complex biological scenario modeling

Need for Enhanced Interdisciplinary Collaboration

- Growing requirements for closer partnerships with computer science, mathematics, and statistics disciplines
- Urgency to deepen biological phenomenon interpretation through cross-disciplinary insights, while providing novel methodologies for other research fields

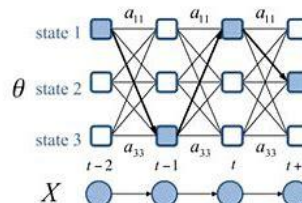
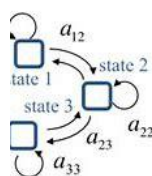
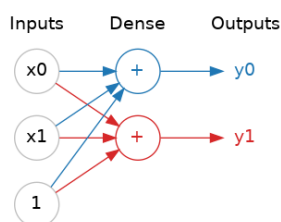
Big Data!

- 估计地球上生活着 4.6×10^{30} 个细菌细胞 (地球质量 2.9×10^{28} 克)
- 细菌的种数肯定超过 10^6 (每种动物包括昆虫都有特异的共生细菌)
- 细菌总含碳量约3500-5500亿吨, 是植物含碳量的60-100%
- 细菌总含氮量约850-1300亿吨, 总含磷量约90-140亿吨, 两者都比植物多10倍
- 单个细菌基因组大小: 0.5×10^6 - 1.4×10^7 bp (so far)
- 人体内外大约有100万亿个细菌分布, 占体重的1-3% (2斤左右)
- 人体细胞公约40-60万亿个, 平均直径10-20mm
- Even in *Escherichia coli*, for instance, there are 225,000 proteins, 15,000 ribosomes, 170,000 tRNA-molecules, 15,000,000 small organic molecules and 25,000,000 ions inside the a few μm cell.
- There are estimated 10^{14} - 10^{16} biochemical reactions in a cell.

部分数据来自郝柏林院士相关报告

Data Mining and Knowledge Discovery

- 归纳逻辑程序 (Inductive Logic Programming)
- 遗传算法 (Genetic Algorithm)
- 神经网络 (Neural Network)
- 统计方法 (Statistical Methods)
- 贝叶斯方法 (Bayesian Methods)
- 决策树 (Decision Tree)
- 隐马尔可夫模型 (Hidden Markov Model)
- 文本挖掘 (Text Mining)
- 网络挖掘 (Internet Mining)



Textbook
第二章
第三章

OXFORD
ACADEMIC

Bioinformatics

Article Navigation

JOURNAL ARTICLE

BNArray: an R package for constructing gene regulatory networks from microarray data by using Bayesian network

Xiaohui Chen, Ming Chen, Kaida Ning Author Notes

Bioinformatics, Volume 22, Issue 23, December 2006, Pages 2952–2954,
https://doi.org/10.1093/bioinformatics/btl491
Published: 27 September 2006 Article history

PDF Views Cite Permissions Share

Abstract

Summary: BNArray is a systemized tool developed in R. It facilitates the construction of gene regulatory networks from DNA microarray data by using Bayesian network. Significant sub-modules of regulatory networks with high confidence are reconstructed by using our extended sub-network mining algorithm of directed graphs. BNArray can handle microarray datasets with missing data. To evaluate the statistical features of generated Bayesian networks, re-sampling procedures are utilized to yield collections of candidate 1st-order network sets for mining dense coherent sub-networks.

Availability: The R package and the supplementary documentation are available at Author Website.

Contact: mchen@zju.edu.cn

Issue Section: APPLICATIONS NOTE > Gene expression

USCDornsife

Principal Investigator

Xiaohui Chen
Associate Professor of Mathematics
Email: xaohui@usc.edu

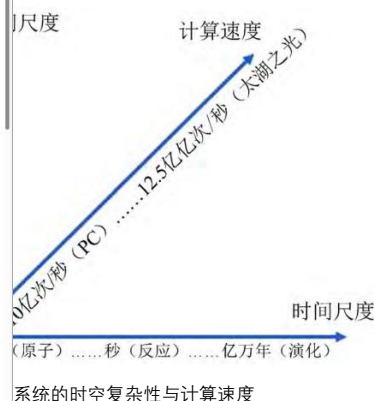
Research interest: high-dimensional statistics, machine learning, optimal transport

Experience:

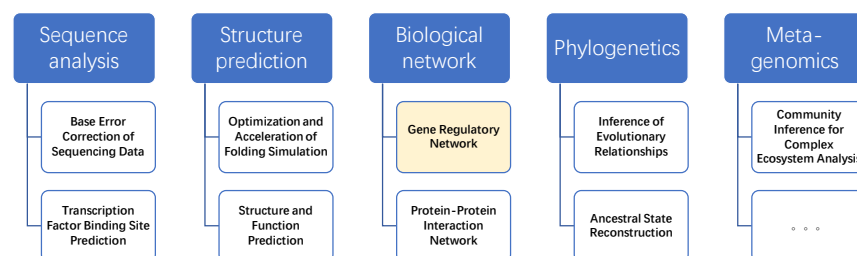
- Associate Professor of Mathematics, University of Southern California, 2023 – Present
- Adjunct Associate Professor, Department of Statistics, University of Illinois at Urbana-Champaign, 2023 – Present
- Associate Professor of Statistics, University of Illinois at Urbana-Champaign, 2019 – 2023
- Assistant Professor of Statistics, University of Illinois at Urbana-Champaign, 2013 – 2019
- Visiting Faculty, Institute for Data, Systems, and Society (IDSS), Massachusetts Institute of Technology, 2019 – 2020
- Affiliated Faculty, Discovery Partners Institute, University of Illinois, 2020 – 2023
- Affiliated Faculty, Computational Science and Engineering, University of Illinois at Urbana-Champaign, 2014 – 2023
- Affiliated Faculty, Informatics Programs, University of Illinois at Urbana-Champaign, 2014 – 2023
- Post-doc and research intern, Toyota Technological Institute at Chicago, 2011 – 2012

Education:

- Ph.D. in Electrical and Computer Engineering, University of British Columbia, 2013
- Advisor: Z. Jane Wang
- Master of Science in Bioinformatics, University of British Columbia, 2008
- Advisors: Richard Gottardo and Devin P. Hurnby
- Bachelor of Science, Zhejiang University, 2006



时代的整合生物信息学，《生物信息学》，2022



Advantages of Bayesian Methods:

- 1. Uncertainty Quantification:** Clarify the credibility of results through posterior probability (e.g., branch support rates of phylogenetic trees);
- 2. Multi-source Information Integration:** Naturally integrate prior knowledge (e.g., evolutionary conservation) with observed data (e.g., sequencing reads);
- 3. Dynamic Model Updating:** Continuously optimize inferences with the input of new data.

the integration of Bayesian methods and deep learning will become a key tool for solving biological complexity problems:

- Multi-Omics Data Integration
- Disease Risk Prediction
- Drug Discovery and Target Prioritization
- Bayesian Neural Networks (Bayesian + X)
- Bayesian Transformer

BIOINFORMATICS APPLICATIONS NOTE

Vol. 22 no. 23 2006, pages 2952–2954
doi:10.1093/bioinformatics/btl491

Gene expression

BNArray: an R package for constructing gene regulatory networks from microarray data by using Bayesian network

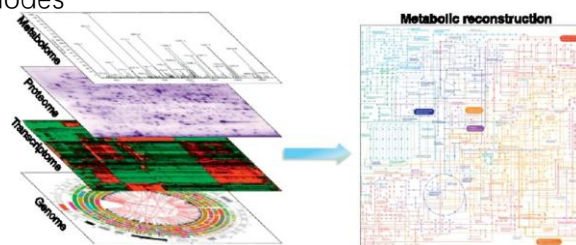
Xiaohui Chen¹, Ming Chen^{1,2,*} and Kaida Ning¹

¹Department of Bioinformatics and ²The National Key Laboratory of Plant Physiology and Biochemistry, College of Life Sciences, Zhejiang University, Hangzhou 310058, China

Received on June 14, 2006; revised on August 28, 2006; accepted on September 18, 2006
Advance Access publication September 27, 2006

Associate Editor: John Quackenbush

- BN: Boolean Network
- BN: Bayesian network, DBN: dynamical BNs
- Stochastic
- ODE: ordinary differential equation
- MCA: metabolic control analysis
- Stoichiometric analysis, elementary flux modes
- FBA: flux balance analysis
- **Petri net**
- S-Systems
- L-Systems
- Rule based
- Network reconstruction, visualization, comparison, structural & functional analysis...



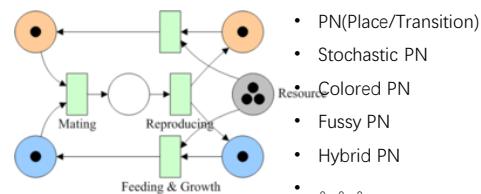
In Silico Biology 3 (2003) 347–365
IOS Press

347

Quantitative Petri Net Model of Gene Regulated Metabolic Networks in the Cell

Ming Chen* and Ralf Hofestädt

Bioinformatics / Medical Informatics, Technische Fakultät, Universität Bielefeld,
Postfach 10 01 31, D-33501 Bielefeld, Germany



it - Information Technology 2014; 56(2): 76–81 — DOI 10.1515/itit-2013-1021

DE GRUYTER OLDENBOURG

Special Issue

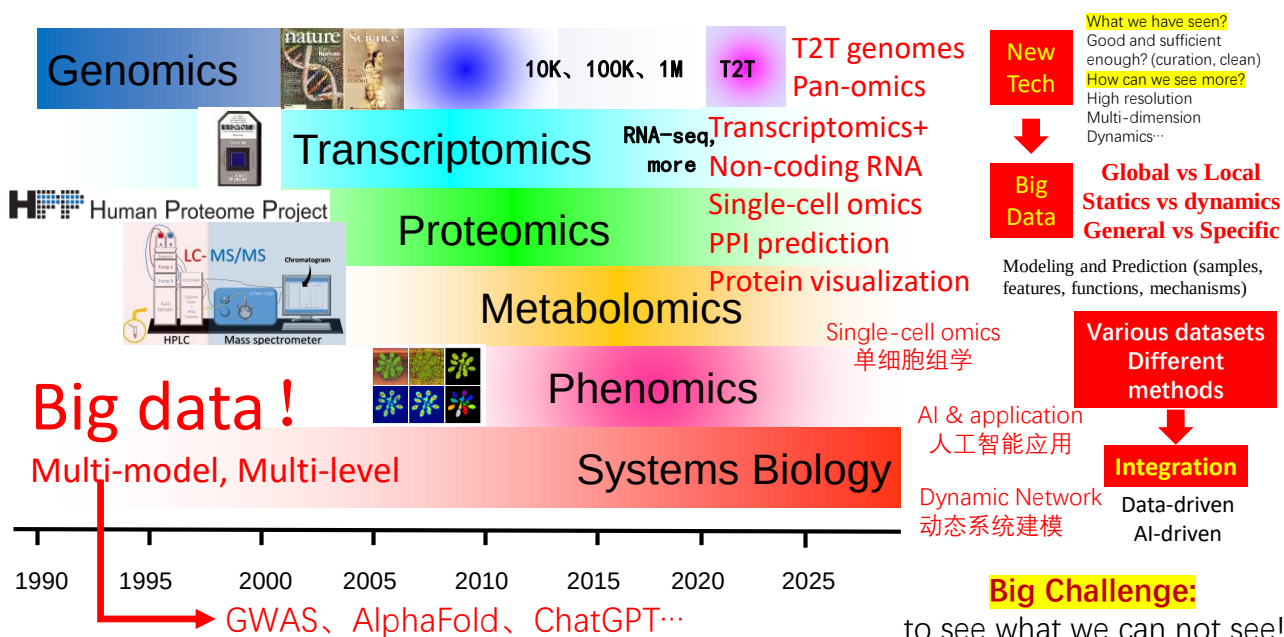
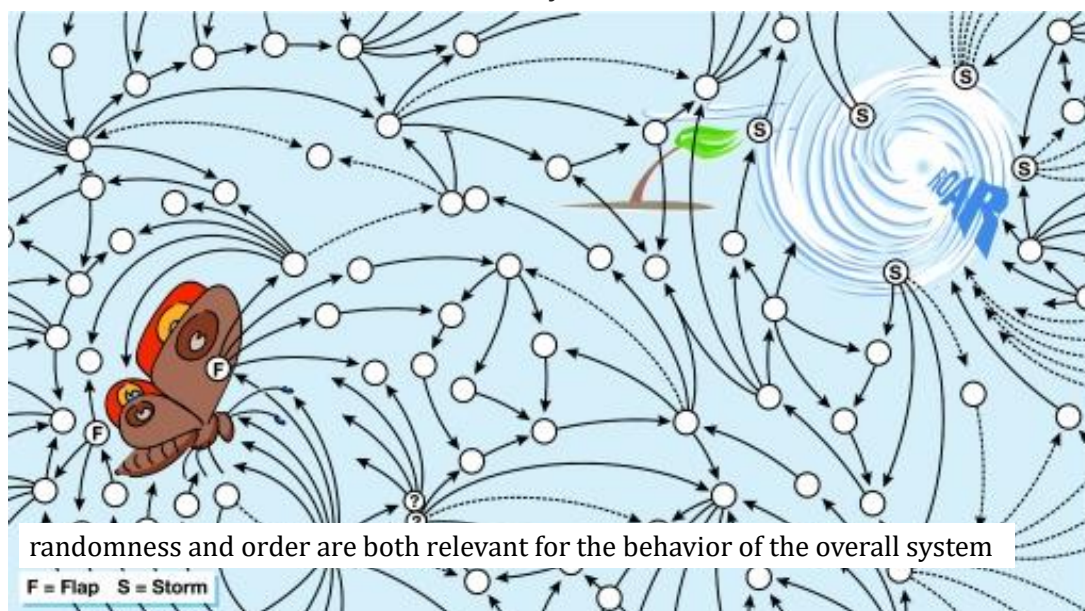
Ming Chen* and Ralf Hofestädt

Open problems in Petri net modeling and simulation of biological systems

- Structural analysis of Petri net models
- Multi-scale modeling
- Diversity
- Modularization
- Parameter settings and optimization in quantitative models
- Intelligence
- Data conversion and model construction
- Analysis and visualization of simulation processes
- Virtual cell modeling and computing power
- Cell-Cell Communication
- Software development

Ming Chen's Lab

Butterfly Effect



d Integrating 30 methods

Probability

Average rank

AUPR = 0.97



- 

PSI: Plant Subcellular localization Integrative predictor

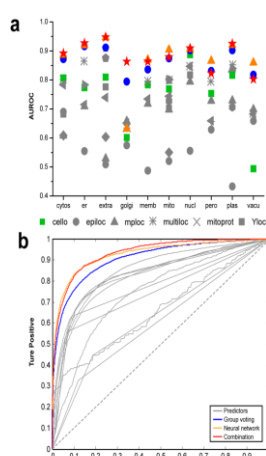
[Home](#)
[Read Me](#)
[Reference](#)

PSIpredictor is currently the most comprehensive and integrative subcellular location predictor for plants. Based on the wisdom of group-voting and artificial neural network, PSI integrated prediction results from 11 individual predictors to give accurate results on chloroplast (chl), endoplasmic reticulum (ER), extracellular (extra), golgi apparatus (golgi), membrane (membr), mitochondrion (mito), nuclear (nuc), peroxisome (pero), plastid (plast) and vacuole (vacu). The currently subformatted each individual predictor both on every subcellular location (a0-8) and overall, with an AUROC=0.932.

Please enter your sequence or upload a fasta file (protein only). See [examples](#)

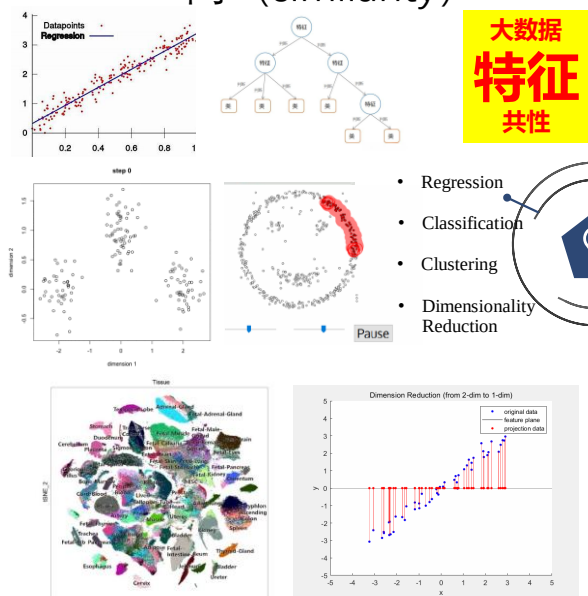
Results will be mailed if you enter an email address:

© 2012-2013 Ming Cheng's lab, Zhejiang University
Contact: mchen@zju.edu.cn



• 同 (similarity)

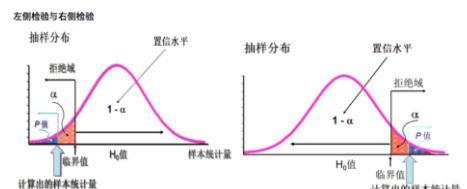
• 不同 (difference)



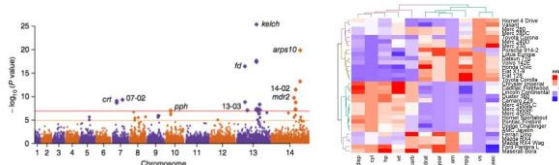
• Differential Analysis Method — Statistical Significance

P value

Analysis of Variance (ANOVA, F-test), t-test, Chi-square Test



• Biological Differences



Obj: Properties (Dimensions)

Integration...

Empirical Model
(Based solely on
key features)

Machine Learning
Model
(Focused on key
feature selection)

Deep Learning
Model
(Comprehensive
feature incorporation)

AI Model
(Multi-dimensional
detailed features)
(Global detailed features)

Autoregressive
LLM

AI-Enhanced Empirical Model
(Multi-modal correlation)
(Prediction & causality)
(Optimization & reasoning)

第20卷 第2期
2022年06月

生物信息学
Chinese Journal of Bioinformatics

Vol.20 No.2
Jun. 2022

DOI:10.12113/202110002

大数据时代的整合生物信息学

陈 铭

(浙江大学 生命科学学院, 生物信息学系, 杭州 310058)

摘 要: 随着生物数据测量技术的不断发展, 生物数据的类型、内容、复杂度不断增加, 生物信息学已进入大数据时代。面对大数据时代多模态、多层次、高维度、非线性的复杂生物数据, 生物信息学需要发展相应的方法和技术进行有效整合生物信息学研究与应用。本文对大数据时代整合生物信息学所涉及的数据整合、方法整合、系统整合及相关问题进行梳理和探讨。

关键词: 整合生物信息学; 生物系统; 组学; 大数据; 问题; 对策

Integrated "Roadmap"
Design for Diverse
Features

Models represented by
CNNs: Effective only for
capturing local details

Consideration of relationships
and contributions among
diverse features

Models represented by
CNNs: Enabling global
understanding yet limited to
single dimensions

Generative
Models:
Based on
probability
rather than
causality

Artificial General Intelligence
(AGI) Large Models (Multi-
task & Multi-modal):
Decomposable and
intervenable hierarchical
models capable of
addressing scenario-,
workflow-, and goal-specific
AI tasks

Supervised Learning → Reinforcement Learning → Self-Supervised Learning
→ ??? (Decomposition and integration are indispensable)

Interpretable?

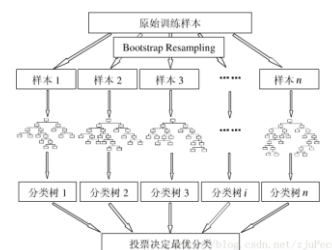
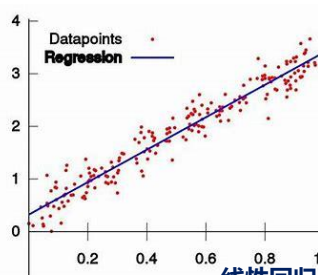
Checkpoint

"Prior Knowledge"

• 回归算法

回归：监督学习方法，对数值型连续随机变量进行建模和预测

Regression: A supervised learning method used to model and predict numerical continuous random variables.

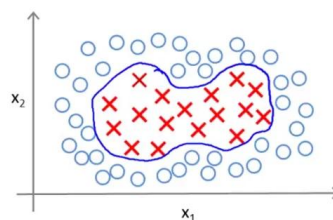
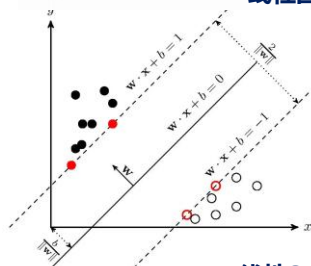


线性回归 和 随机森林

• 分类算法

分类：监督学习方法，对有标签样本学习，从而将未知类别的样本分类到预定义的种类或标签中

Classification: A supervised learning method that learns from labeled samples to classify unknown samples into predefined categories or labels.

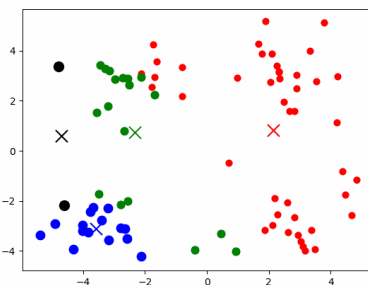


线性SVM 和 核SVM

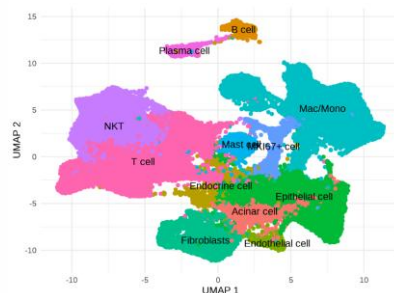
• 聚类算法

聚类：按照数据内在的相似性将未标注的数据集划分为多个类别，以寻找数据中隐含的范式；同一类别内的数据相似度较高，不同类别的数据相似度较低

Clustering: Partition unlabeled datasets into multiple categories based on the intrinsic similarity of data, to identify hidden patterns within the data; data points within the same category have high similarity, while those across different categories have low similarity.



K-means聚类



聚类应用：

基因表达谱的
样本分组
蛋白质序列的
家族分类等
单细胞聚类，
细胞类型注释

划分聚类Partitioning Clustering: 将数据集划分为不同的簇，如K-means算法、K-medoids算法等
层次聚类Hierarchical Clustering: 构建一棵聚类树来创建多层次的分层结构
密度聚类Density-Based Clustering: 将聚类定义为高密度区域之间的低密度区域
网格聚类Grid-Based Clustering: 将数据空间划分为网格，然后将数据分配到对应的网格中
混合聚类Hybrid Clustering: 将多个聚类算法结合起来，以便更好地捕捉不同数据结构

• 降维算法

Biological Big Data — Multi-modal, high-dimensional, and characterized by massive redundancy and noise (e.g., single-cell transcriptome expression matrices with tens of thousands of cells $\times$ tens of thousands of genes)

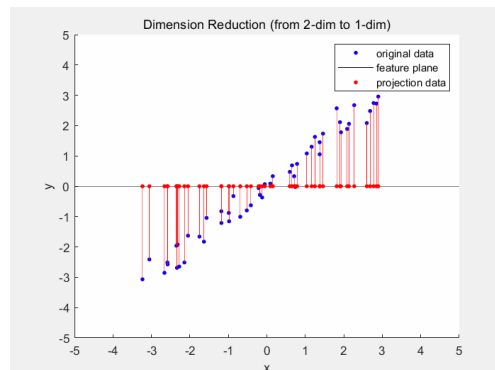
降维算法——从高维数据中提取出最具代表性的特征，保留主要信息的同时减少冗余信息和噪声，将数据降维到二维或三维还有助于直观的数据可视化

Dimensionality Reduction Algorithms — Extract the most representative features from high-dimensional data, reduce redundancy and noise while preserving key information. Reducing data to 2D or 3D also facilitates intuitive data visualization.

基于特征选取的方法：挑选最具信息量的特征以减少数据维度，如缺失值比率 (Missing Value Ratio)、随机森林 (Random Forest) 等

基于组分因素的方法：试图找到数据中的主要成分或因素以实现降维，如因索分析 (Factor Analysis)、主成分分析 (PCA) 等

基于映射的方法：将数据映射到新的低维空间中，如等距映射 ISOMAP、t-SNE 和 UMAP 等



PCA降维：将n维的向量映射到k维上，这k维是全新的正交特征称为主成分Principal Component，每个主成分都保留了原始数据一部分的变异

PCA Dimensionality Reduction: Map n-dimensional vectors onto k-dimensional space. These k dimensions are brand-new orthogonal features called **Principal Components**, and each principal component retains a portion of the variation present in the original data.

Explosive Growth of Data Volume

- Breakthrough advancements in high-throughput sequencing technologies
- Expansion of abundant, heterogeneous biological data resources
- Proliferation of multi-omics technologies (genomics, transcriptomics, proteomics, etc.)

Imperatives for AI and Algorithm Innovation

- Rapid evolution of artificial intelligence methodologies for large-scale biological data processing
- Escalating demand for advanced algorithm applications (e.g., deep learning, graph neural networks) in complex biological scenario modeling

Need for Enhanced Interdisciplinary Collaboration

- Growing requirements for closer partnerships with computer science, mathematics, and statistics disciplines
- Urgency to deepen biological phenomenon interpretation through cross-disciplinary insights, while providing novel methodologies for other research fields

Optimization of Data Governance and Analytical Pipelines

- Implement rigorous quality control protocols to ensure intrinsic data accuracy
- Establish unified data standardization frameworks for cross-study compatibility
- Develop standardized validation workflows to enhance data reliability and comparability across datasets

Advancement of Toolkits and Algorithm Innovation

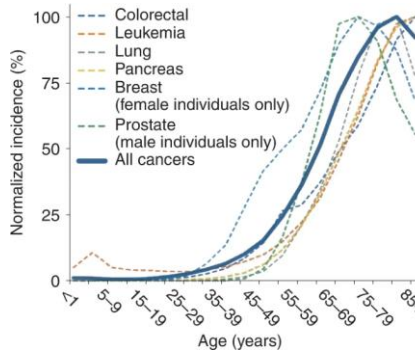
- Integrate newly developed cutting-edge tools with well-established classic analytical toolkits for complementary advantages
- Foster a culture of tool adoption and iterative improvement to drive technological innovation and field-wide progress

Development of Multi-Scale Data Integration and Cross-Analysis Frameworks

- Design sophisticated computational methods and specialized tools to enable the integration and cross-analysis of multi-scale data (e.g., multi-omics, phenotypic, and clinical data)
- Build cross-disciplinary teams to co-develop and validate integrated analysis pipelines for holistic biological insight mining

浙江大學 Cancer Hallmarks (2000)

BIS@ZJU
Bioinformatics Server



Cell

Volume 100, Issue 1, 7 January 2000, Pages 57-70

The Hallmarks of Cancer

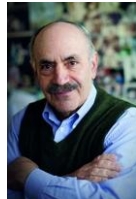
Douglas Hanahan¹ & Robert A. Weinberg²

¹ Department of Biochemistry and Biophysics and, Hormone Research Institute, University of California at San Francisco, San Francisco, California 94143, USA
² Whitehead Institute for Biomedical Research and, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA

Available online 11 April 2001.

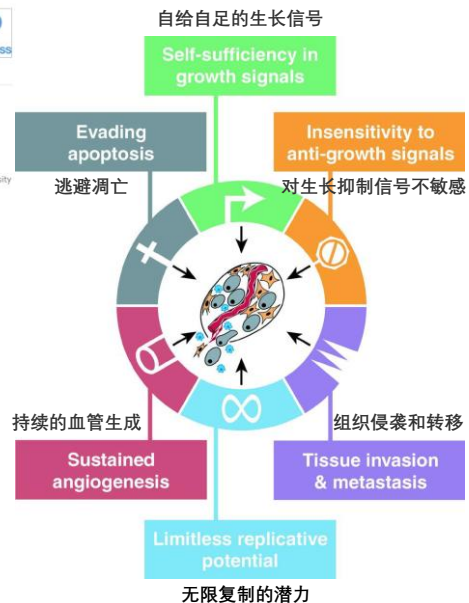
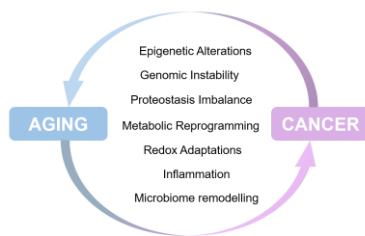


Douglas Hanahan (1951-), born in Seattle, Washington, USA, received a bachelor's degree in Physics from MIT (1976), and a Ph.D. in Biophysics from Harvard (1983).



Robert Allan Weinberg (1942-), born in Pittsburgh, Pennsylvania, USA, received his SB degree in biology from MIT (1964), PhD in biology from MIT (1969).

Nature Aging, 2022, 2: 365–366
Interactions of Cancer and Aging Hallmarks



浙江大學 Cancer Hallmarks (2011)

BIS@ZJU
Bioinformatics Server

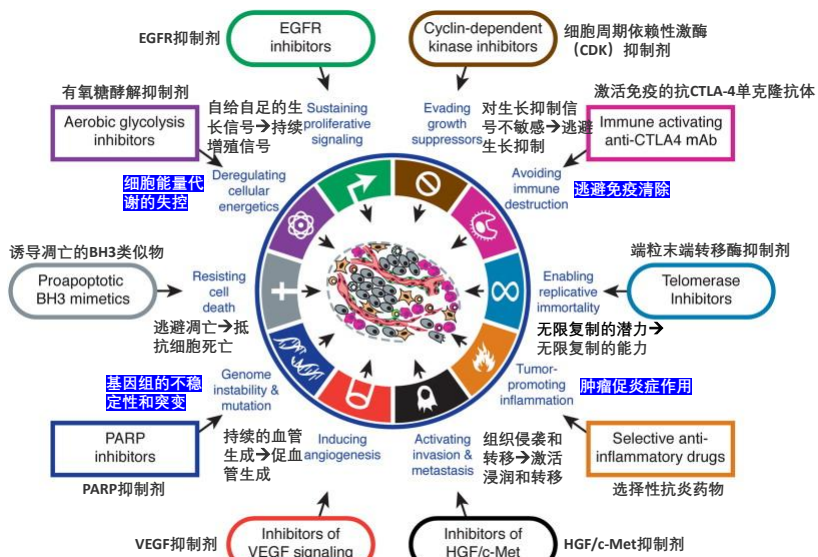
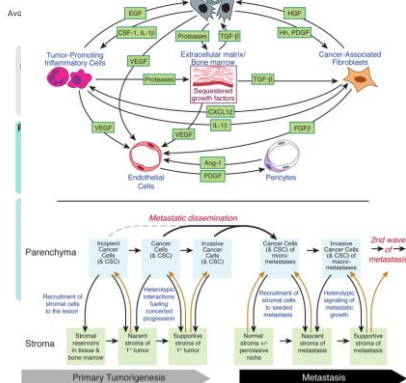
Cell

Volume 144, Issue 5, 4 March 2011, Pages 646–674

Hallmarks of Cancer: The Next Generation

Douglas Hanahan^{1,2} & Robert A. Weinberg³

¹ The Swiss Institute for Experimental Cancer Research (ISREC), School of Life Sciences, EPFL, Lausanne CH-1015, Switzerland
² The Department of Biochemistry & Biophysics, UCSF, San Francisco, CA 94158, USA
³ Whitehead Institute for Biomedical Research, Ludwig/MIT Center for Molecular Cancer Cells



10大类药物佐证“癌症十大标志性特征”

REVIEW | JANUARY 12 2022

Hallmarks of Cancer: New Dimensions

Douglas Hanahan

Check for updates

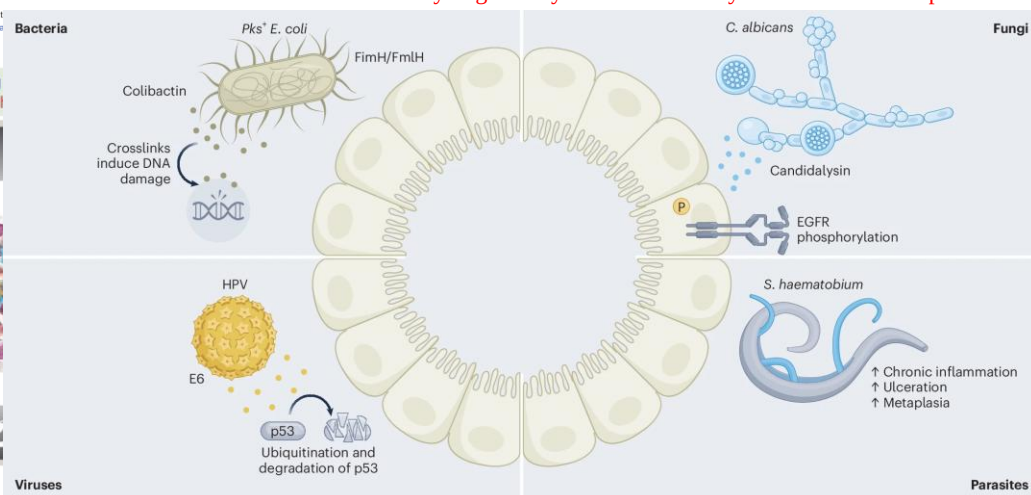
Author & Article Information

* Corresponding Author: Douglas Hanahan, Agora Translat
Switzerland, Phone: 41-21-545-1119, E-mail: douglas.hanahan@agora-translat.ch
Cancer Discov 2022;12:31-46

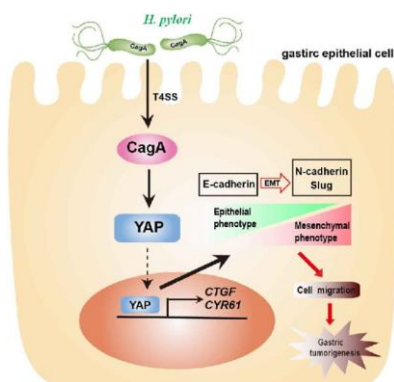
Gene, Cell Multi-cell, interaction Development, environment

2000 2011 2022

All these features do not exist in isolation but are interconnected and synergistically act to collectively drive tumor development.

From: Dohlman et al. The multi-kingdom cancer microbiome, *Nature Microbiology* 2025, 10: 2369–

浙江大学 Type IV secreted effector proteins (T4SEs)

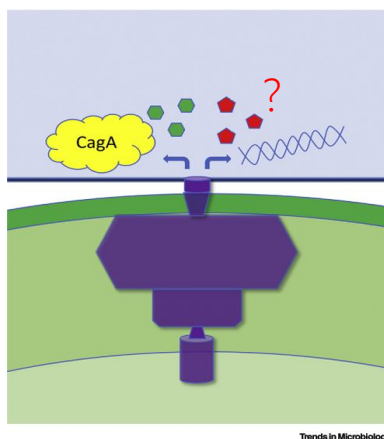


Li et al. JECRC 2018

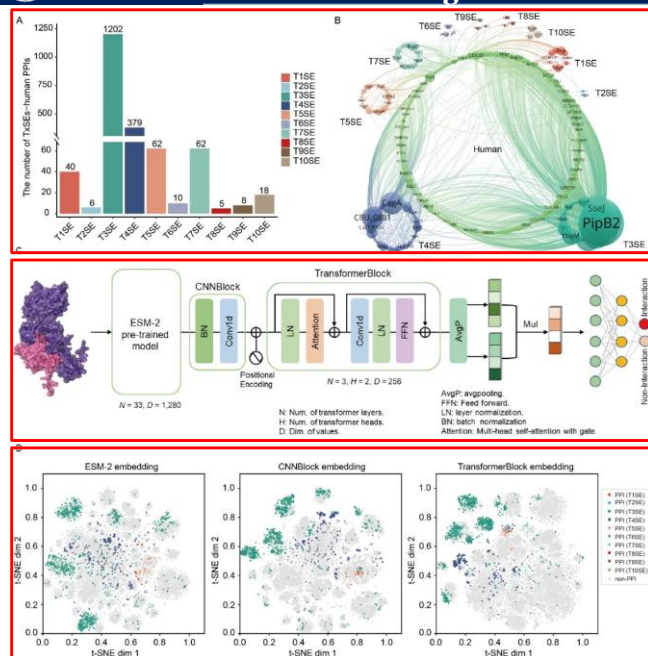
The proposed mechanism of *Helicobacter pylori* CagA-mediated promotion of gastric tumorigenesis. CagA + *H. pylori* delivers CagA into gastric epithelial cells via T4SS where CagA induces oncogene YAP expression, increases nuclear translocation of YAP that elevates downstream gene expression. The *H. pylori* CagA-mediated activation of YAP then leads to enhanced EMT program, thereby further promoting gastric tumorigenesis

Question:

systematically analyze *H. pylori* T4SEs (how many?)
their pathogenic mechanisms (how to interact?)

Cover TL, et al. *Trends Microbiol.* 2021

Tool	Method
S4TE	Motif searching
Burstein's	Voting algorithm
Lifshitz's	Hidden Semi-Markov Mode(HSMM)
Chen's	Genetic Screening
T4EffPred	SVM
T4SEpre	SVM
Wang's	SVM
PredT4SE Stack	Stacked generalization
Bastion4	Ensemble model
OPT4e	SVM
Tbooster	LR,RF and SVM
CNN-T4SE	CNN,voting
orgsissec	Phylogenetic profiles
T4SEfinder	CNN; PLM
Our method (T4SEpp)	Bi-conv; PLM; Ensemble



实验验证互作网络:

(A) Distribution of interaction counts between human proteins and bacterial effector proteins from the type I to type X secretion systems.

(B) Interaction network of human proteins and bacterial effector proteins.

HPINet 预测模型:

(C) Architecture of the HPINet model. Protein sequences are embedded using the ESM-2 model, followed by a convolutional module for extracting local features. A transformer module variant is employed to capture both global and local dependencies, using a gating mechanism to refine interaction predictions.

特征学习验证 (t-SNE):

(D) t-SNE visualization of feature embeddings generated at different stages of the model: ESM-2 embeddings (left), embeddings after the convolutional module (middle), and embeddings from the transformer module (right).

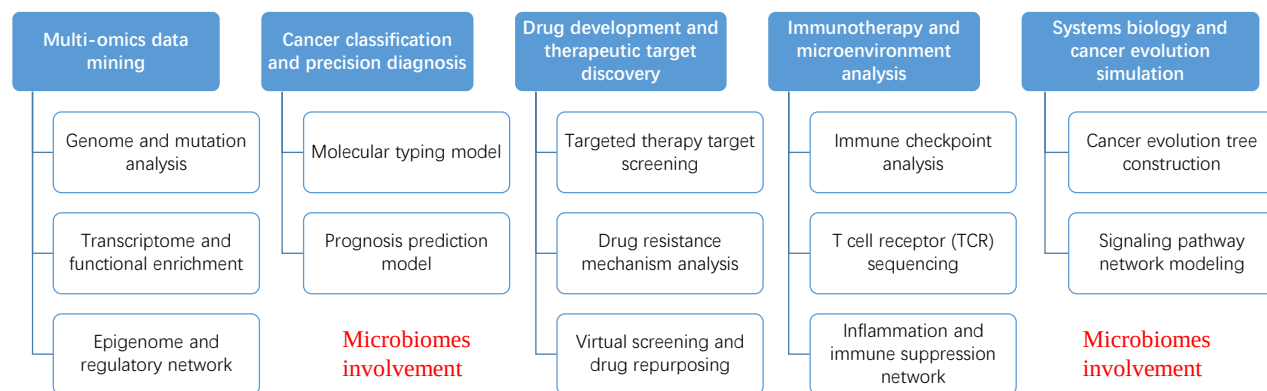
bioRxiv, 2025

The core role of bioinformatics in cancer research: By integrating multi-omics data (genome, transcriptome, proteome, metabolome, epigenome, etc.) and combining methods such as machine learning and systems biology, bioinformatics has become a core tool for analyzing cancer characteristics and driving precision medicine.

molecular mechanisms related to cancer characteristics

Precision cancer diagnosis and prognosis assessment

Drug development and therapeutic target discovery

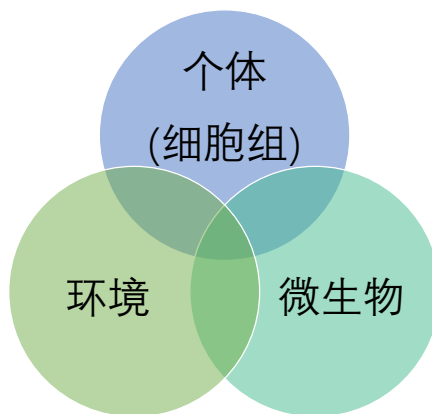


组学(Omics) → 整合生物信息学 → 解决大问题

基因组学 (Genomics)
表观基因组学 (Epigenomics)
转录组学 (Transcriptomics)
蛋白组学 (Proteinomics)
代谢组学 (Metabolomics)
脂类组学 (Lipidomics)
糖组学 (Glycomics)
免疫组学 (Immunomics)
微生物组学 (microbiomics)
表型组学 (Phenomics)
暴露组学 (Exposomics)
辐射组学 (Radiomics)
...
单细胞组学 (Single-cell omics)
文献组学 (Bibliomics)



小科技



Multiple Omics
Various Levels

Coding! or
Non-coding?!

人口 长寿与衰老
能源 生物能源
粮食 绿色农业 健康农业
环境 绿色发展



大科学
大工程

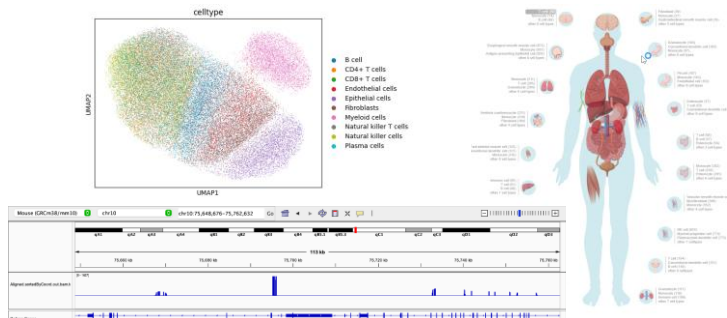
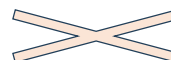
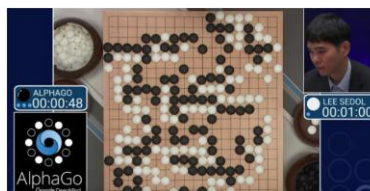
生物演化：从哪里来、
到哪里去？我是谁？
分子育种、智慧育种、
生物智造
生物安全、生态安全、
同一健康

MChen's Lab, Zhejiang University

§ 4: AI in Bioinformatics

Why AI in Bioinformatics?

- **Data Explosion:** We generate biological data (e.g., NGS, single-cell omics) faster than we can analyze them manually.
- **Complexity of Bio-data:** Biological systems are complex and non-linear, thus statistical methods struggle to find the rules.
- **Exploring Data with AI:** Artificial intelligence, particularly deep learning, can spot out hidden patterns within vast, noisy data and has become an effective tool in bioinformatics.



AlphaFold: CNN (ResNet)

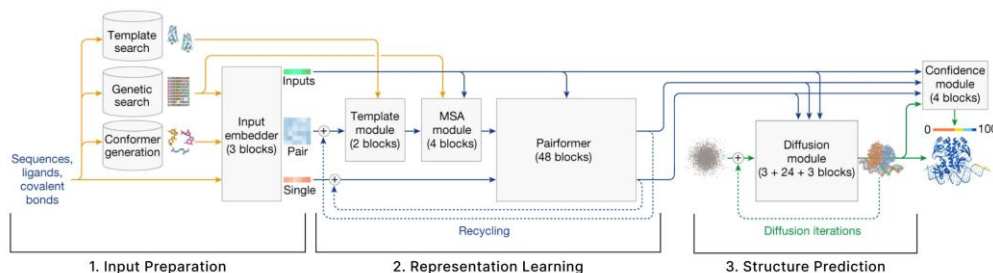
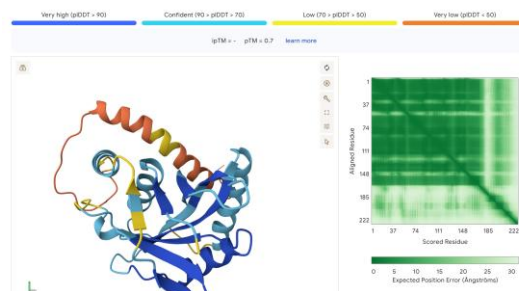
Predict inter-residue distance maps

AlphaFold 2: Transformer

Effectively solve the protein folding problem

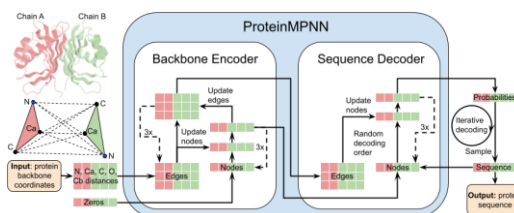
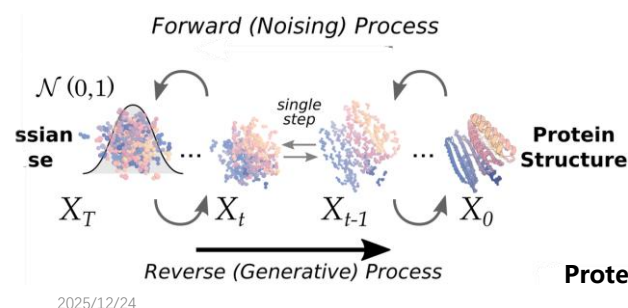
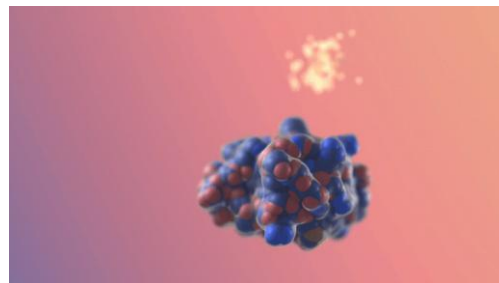
AlphaFold 3: Transformer + Diffusion

Predicts unified complexes of proteins, DNA, RNA, and ligands



RFdiffusion: The "Stable Diffusion" of Protein Backbones

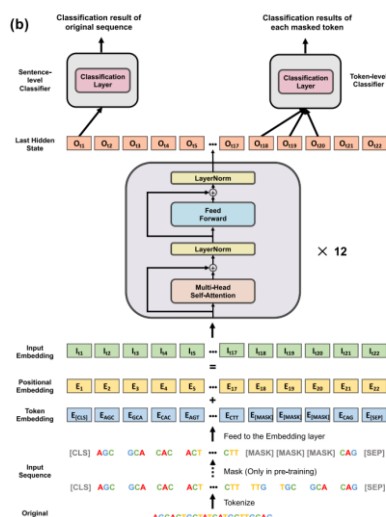
- ✓ Starts from random noise and denoises it into a structured protein
- ✓ From Prediction to Creation
- ✓ De Novo Design of novel proteins not found in nature
- ✓ Designing binders, enzymes, and symmetrical assemblies with high success rates



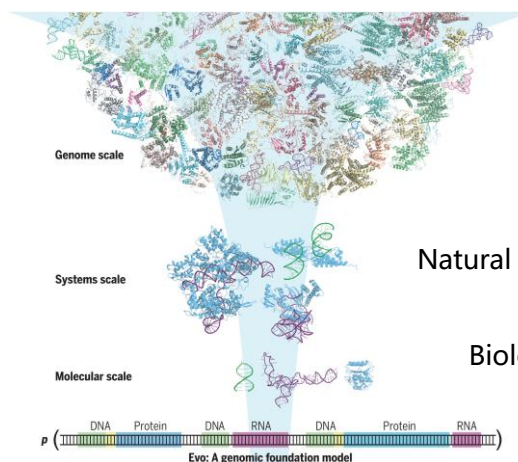
Protein Design Workflow: Rfdiffusion + ProteinMPNN

2025/12/24

62



2025/12/24



Sequence as Language (Bio-LLMs)

Natural Language Processing



Biological Sequences

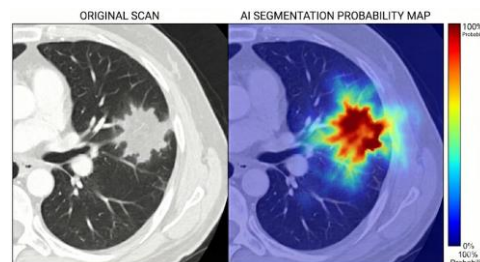
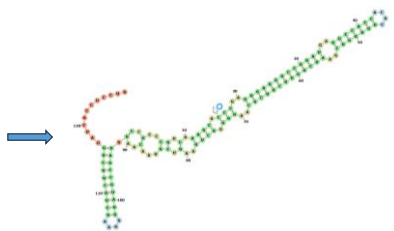
An emerging topic in bioinformatics

Evo

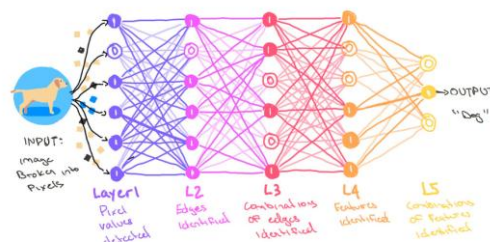
63

```

> 16s_H.volcanii_domain4
CCGCCGCUCAAAGCACCCGAGUG
AGGUCCGGAUGAGGCCACCACAC
GGUGGUCGAUCUGGGCUUCGC
AAGGGGCUUAAGUCGUAACAA
GGUAGCCGUAGGGGAUCUGCG
GCUGGAUACCUCCUG
    
```



- **Define a biological problem:** Sequence/structure processing? Omics data processing? Medical image analysis?
- **Construct a dataset:** Data collection (public or in-house), cleaning, normalization, and annotation (ground truth)
- **Build your model:** Discriminative or Generative? CNNs? RNNs? Transformers? Diffusion Models?

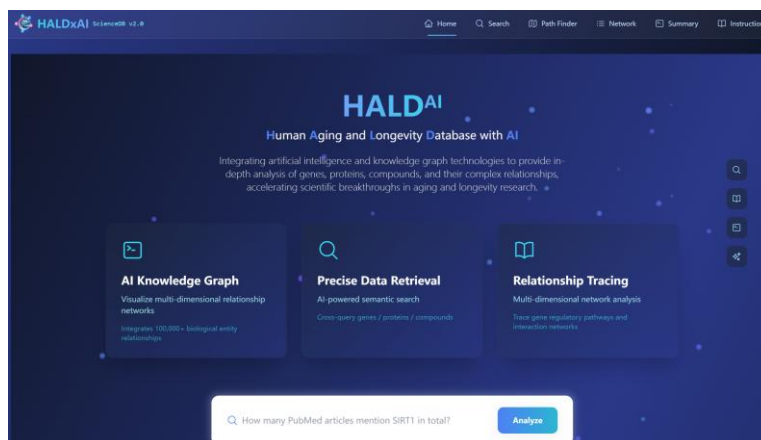


2025/12/24

64

Example: <https://bis.zju.edu.cn/haldxai/home>

- **Define a biological problem:** Integrating and mining biological factors related to aging and longevity
- **Construct a dataset:** Collecting entities from PubMed literature; Triplets in Knowledge Graph
- **Build your model:** DeepSeek large language model agent

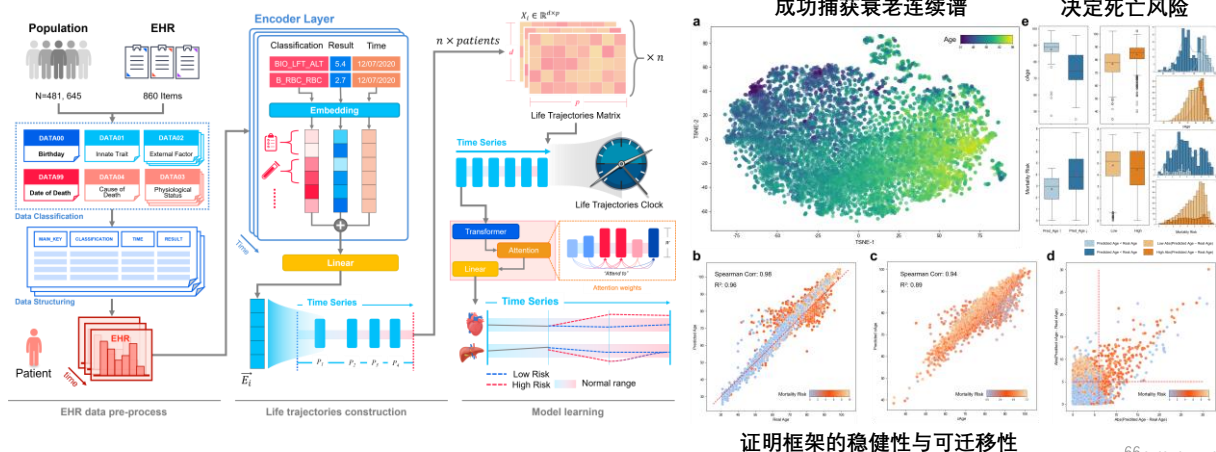


2025/12/24

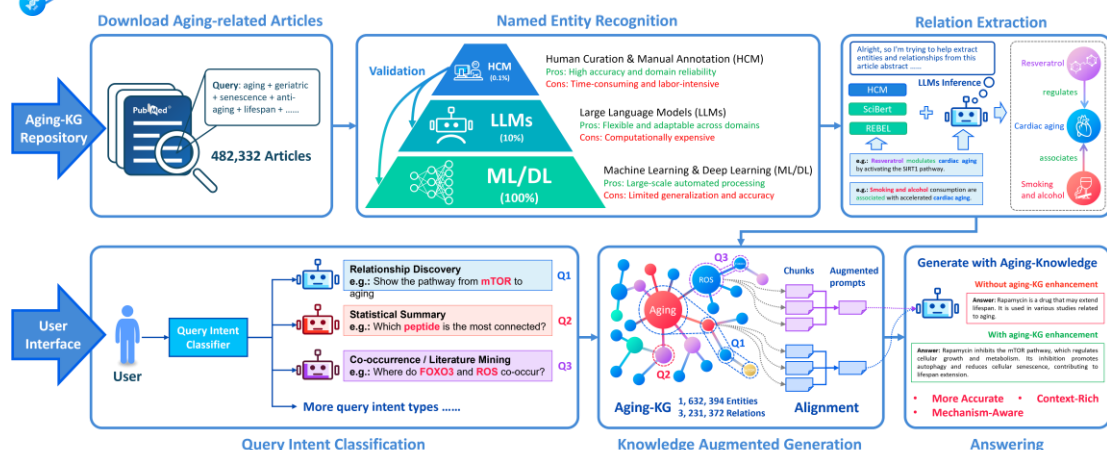
65



Transformer-Enhanced Time Series Model Utilizes Dynamic Health Trajectories for Mortality Risk Prediction

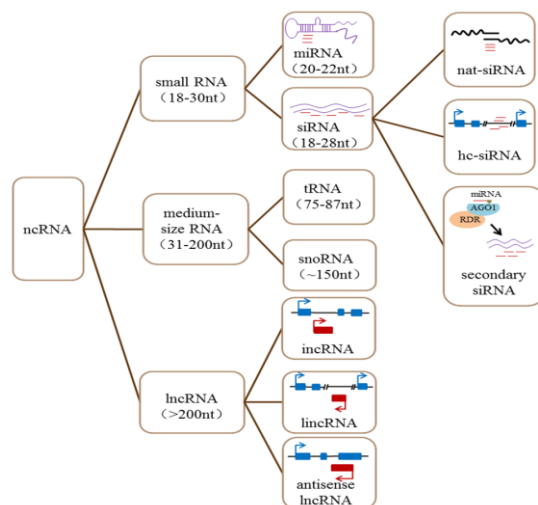


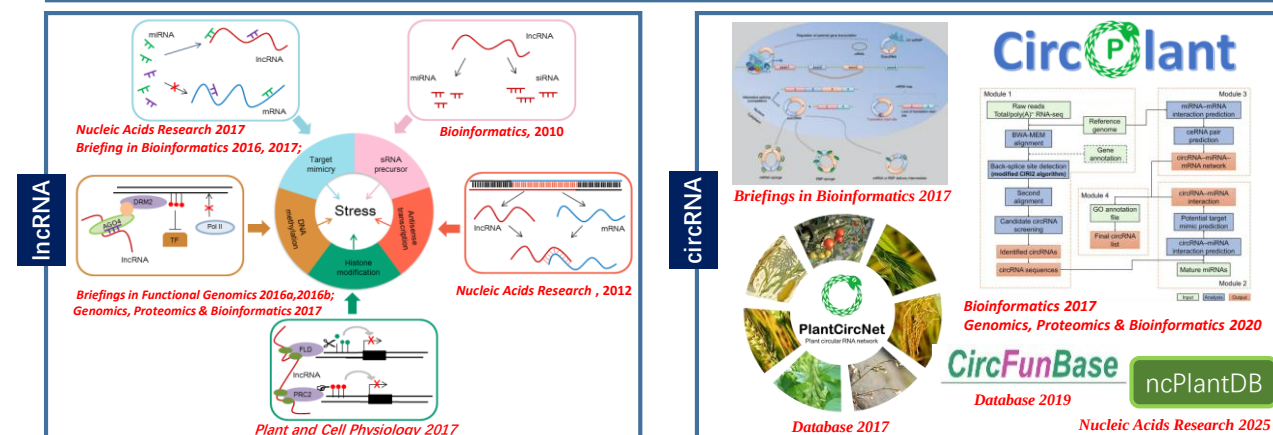
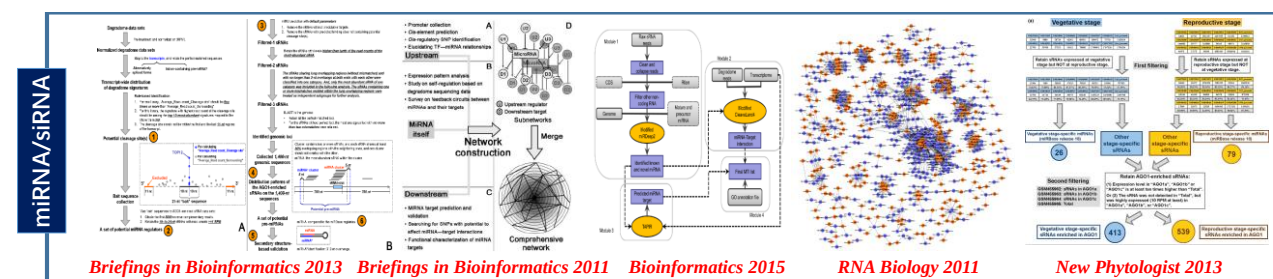
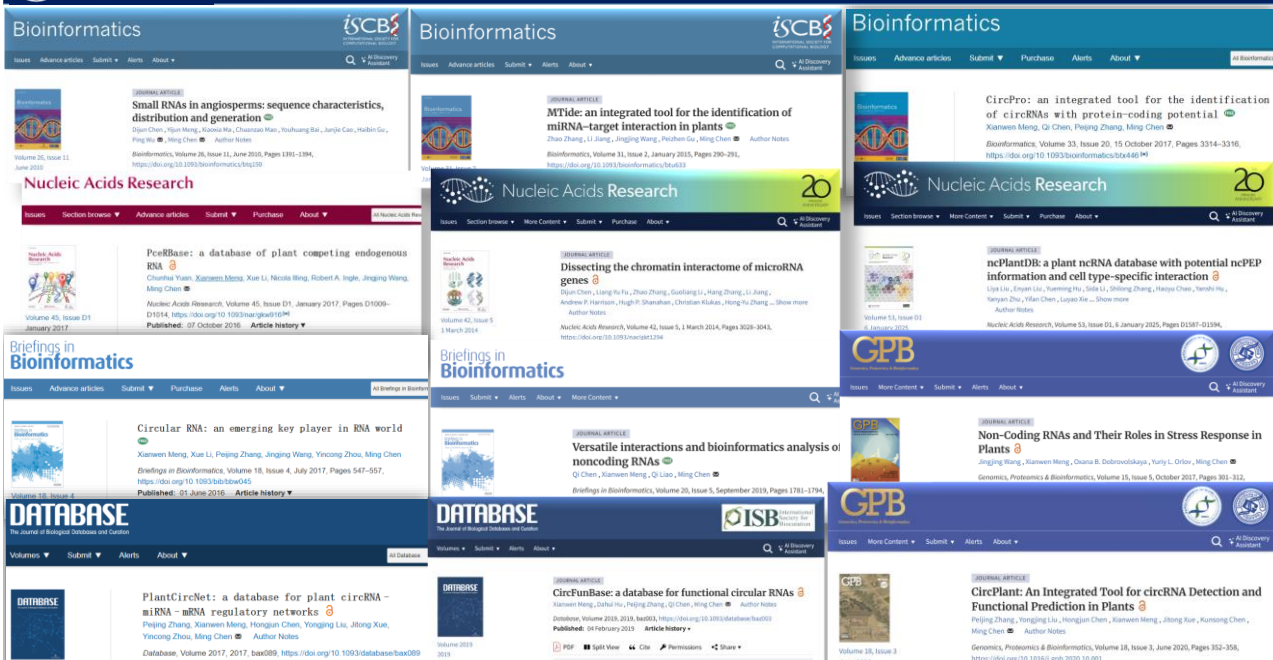
HALDxAI: An AI-Enhanced Aging Knowledge Graph for Intelligent Geroscience Discovery



• Integrative Approaches (Non-coding RNAs)

- ncRNAs, what are they? characteristics
- Biogenesis,
- Expressions,
- Interactions,
- Regulations,
- Even dynamic functions, 3D
- ...
- AI generation







GenNA can generate nucleotide sequences from natural-language prompts, including DNA/RNA descriptions, species names, gene names, and gene-function annotations (species and gene names being optional).



2025/12/24

72

§ 5: Talent Cultivation in Bioinformatics

人工智能时代生物信息学学科发展和人才培养模式研究

浙江大学生命科学院教授 陈 铭

【摘要】生物信息学作为生命科学与生物技术/信息技术研究领域的关键交叉学科，对生物经济和数字经济贡献日益显著。当前，生物信息学仍面临学科体系不健全、定位模糊以及交叉合作不充分等挑战。多模态高精度生物大数据的准确性、分析处理和共享整合问题也考验着生物信息学的发展。在建设科技强国的过程中，生物信息学是生物经济产业布局的关键环节。与此同时，人工智能技术的融入正引发生命科学研究范式的转变，促使生物信息学从认知科学向工程创造的 STEM 并轨模式方向发展。此外，生物信息学面临人才培养同质化和优秀青年人才“内卷”的困境，需要构建多层次培养体系和优化科研环境，培养具有战略眼光的科学家。由此，应加强顶层设计，完善学科体系与教学体系；建立多元化人才培养体系；全面推进“101 计划”；优化教育资源分配和教学模式创新。

【关键词】生物信息学 人工智能 人才培养 STEM 优化教育资源分配

【中图分类号】Q811.4/C961 【文献标识码】A

【DOI】10.16619/j.cnki.rmltssq.2024.16.002

[Abstract] As a key interdisciplinary field integrating life sciences, biotechnology, and information technology, bioinformatics has made increasingly remarkable contributions to the bioeconomy and digital economy. Currently, bioinformatics still faces challenges such as an **inadequate disciplinary system, ambiguous positioning, and insufficient interdisciplinary collaboration**. Meanwhile, issues regarding the accuracy, analytical processing, and shared integration of multi-modal, high-dimensional biological big data also pose significant tests to the development of bioinformatics. In the process of China building a world leader in science and technology, bioinformatics serves as a crucial link in the layout of the bioeconomy industry. At the same time, the integration of **artificial intelligence (AI) technology is triggering a paradigm shift in life science research, driving bioinformatics to evolve from a cognitive science toward a STEM-coexisting model that combines cognitive exploration with engineering innovation.**

Furthermore, bioinformatics is confronted with the dilemmas of homogenized talent training and intense internal competition ("involution") among outstanding young talents. To address these, it is necessary to establish a multi-level training system, **optimize the research environment, and cultivate scientists with strategic vision**. Therefore, we should strengthen top-level design to improve the disciplinary and educational systems; establish a diversified talent training framework; fully promote the "101 Plan"; and optimize the allocation of educational resources while innovating teaching models.

"The 101 Plan", a program for fundamental education and teaching reform, focusing on developing core courses, core textbooks, core faculty, and core practical projects. The aim is to cultivate a group of outstanding natural scientists, medical scientists, and social scientists who will have significant influence globally in the future.

33 top universities in Biology



生物科学“101”计划工作推进会



Core practical projects

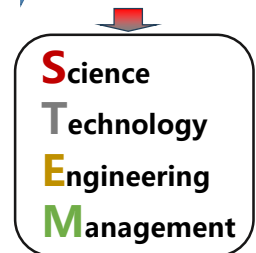
Virtual Teaching and Research section (MOE)



Core textbooks


 Top Scientists
 First front lecturers
Core Faculties

Core Courses



生物科学101计划

生物信息学课程
Bioinformatics

生物科学“101计划”

生物科学“101计划”核心课程及牵头专家

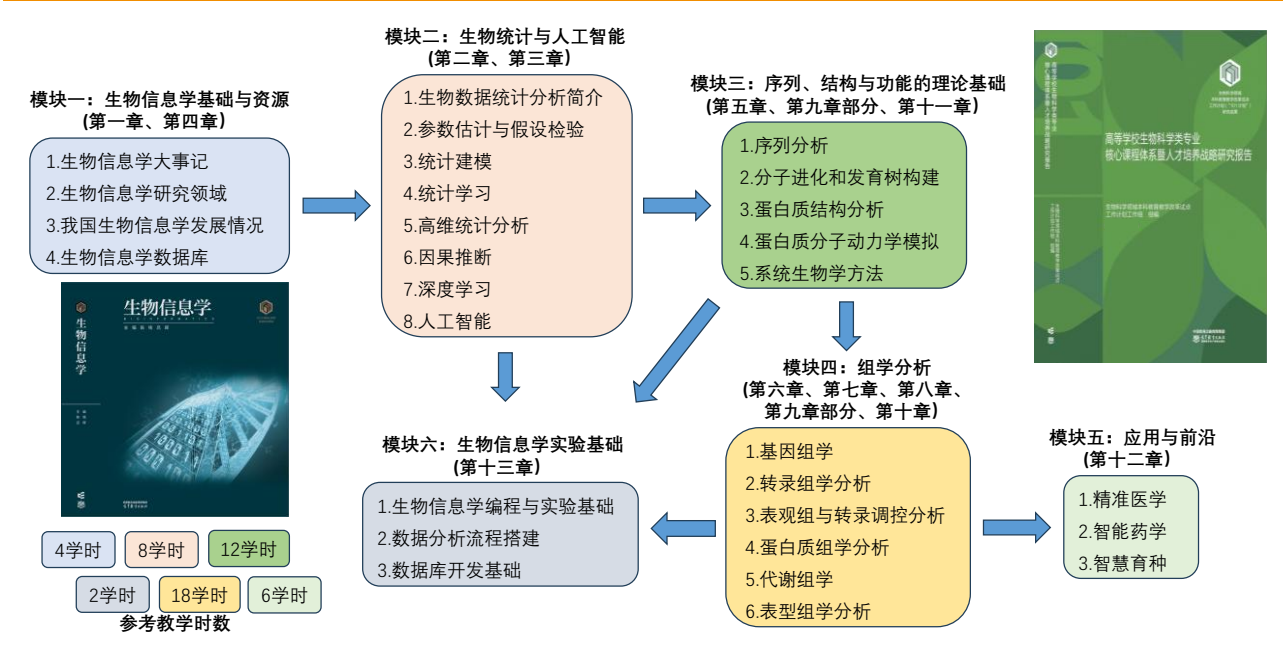
序号	核心课程	牵头专家	单位	序号	核心课程	牵头专家	单位
1	普通生物学	赵进东/金力	北京大学/复旦大学	7	神经生物学导论	饶毅/梅林/段树民	首都医科大学/浙江大学
2	生物化学	昌增益/施一公/王新泉	北京大学/西湖大学/清华大学	8	生物物理学	施一公/汤超/王宏伟	西湖大学/北京大学/清华大学
3	细胞与分子生物学	陈晔光/陈建国	清华大学/北京大学	9	普通生态学	方精云/唐志尧/王志恒	云南大学/北京大学
4	遗传学与发育生物学	许田/金力	西湖大学/复旦大学	10	生物信息学	陈铭/吕晖	浙江大学/上海交通大学
5	生理学原理	王世强/梅岩艾/朱景宁	北京大学/复旦大学/南京大学	11	免疫生物学	董晨	西湖大学/上海交通大学
6	微生物学	邓子新/陈峰/陈向东	上海交通大学/上海交通大学/武汉大学				



生物科学101计划

生物信息学课程Bioinformatics

教材与知识模块



生物科学101计划

生物信息学课程Bioinformatics

智慧课程：MOOC及题库

生物信息学

课程简介

课程介绍

知识图谱

教学团队

序号	章	节名	主讲	学校
第一章	生物信息学概念及发展历史	1.1生物信息学简介(一)	陈铭	浙江大学
		1.2生物信息学简介(二)		
第二章	生物统计学基础	2.1贝叶斯推理	蒋杭进	浙江大学
第三章	深度学习与人工智能	3.1AI概览：生物学视角	吕晖	上海交通大学
		3.2生物医学大数据分析扩散模型	吕晖	上海交通大学
第四章	生物信息资源	4.1中国生物信息学资源库	徐添翼	中国科学院大学
		4.2代表性生物资源数据库-GO&KEGG	凌毅	中国农业大学
第五章	序列比对与分析	5.1序列比对分析	冯聪	浙江大学
		5.2系统发育树的构建	王晶晶	浙江大学
第六章	基因组学及宏基因组学	6.1基因组学及宏基因组学	宁康	华中科技大学
第七章	转录组学	7.1转录组学	孙宝发	南开大学
		7.2非编码RNA	张霏婧	浙江大学
第八章	转录调控与表观遗传	8.1转录因子结合模体从头发现	张勇	同济大学
		8.2Hi-C数据分析要点		
第九章	蛋白质与代谢组学	9.1蛋白质分子动力学模拟	王勇	浙江大学
第十章	表型组学	10.1精细定位	陈超	中南大学
		10.2心理表型		
第十一章	系统生物学	11.1无标度网络	廖明帆	西北农林科技大学
		11.2系统生物学新理念新方法	王卓	上海交通大学
		11.3代谢网络重构与建模分析		
第十二章	生物信息学实验	12.1生物信息学实验基础	陈迪俊	南京大学

生物科学101计划

生物信息学 课程 Bioinformatics

核心课程知识图谱建设



共建设13个知识模块，101个知识点，239个知识关系 涉及到重点知识点49个，难点知识点24个，案例知识点52个

- 知识图谱
- 结构化的课程教学内容和设计帮助老师更好地去定位教学的重难点和薄弱点
 - 提供丰富的教学资源、习题库等，教学资源与知识点高度智能匹配，形成个性化的知识图谱，帮助教师做到“因材施教”

生物科学101计划

生物信息学 课程 Bioinformatics

《生物信息学》智慧课程

建设思路 | AI赋能 融合助力



多模态线上课程资源 人工智能 应用工具

《生物信息学》智慧课程

生物科学101计划

生物信息学课程

Bioinformatics

智慧课程：AI问答

基于通用大模型的AI问答

基于《生物信息学》专属知识库的AI问答设置

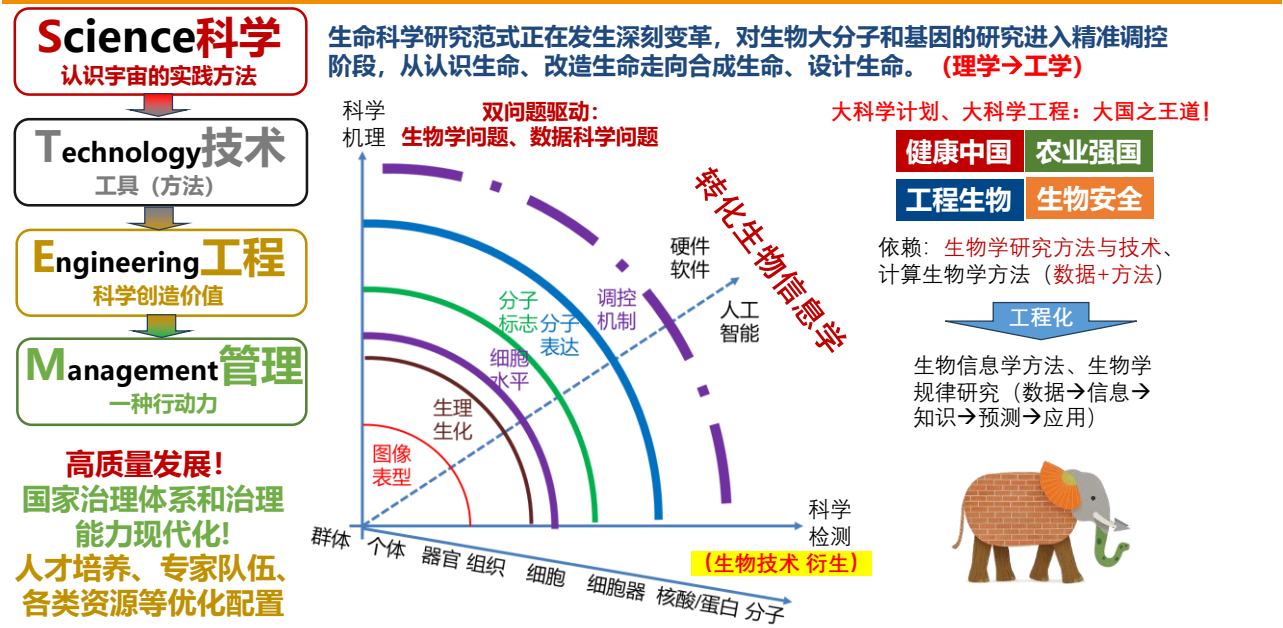


生物科学101计划

生物信息学课程

Bioinformatics

生物信息学人才培养模式探讨



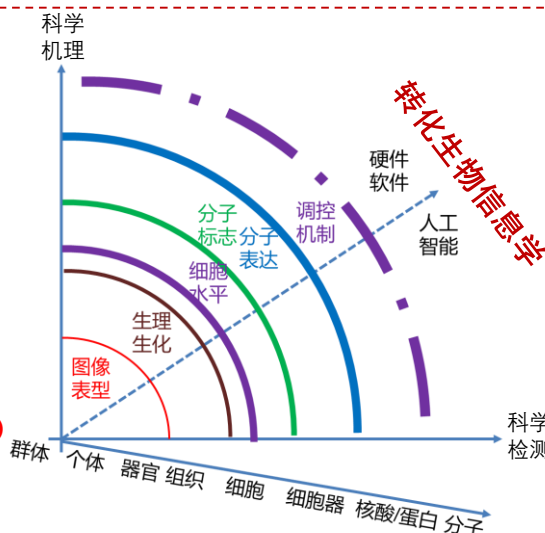
科学

认识宇宙的实践方法

是建立在可检验的解释和对客观事物的形式、组织等进行预测的有序知识系统，是已经系统化和公式化了的知识。根据这些（科学）系统知识所要反映对象的领域，主要可分为自然科学、社会科学、思维科学、形式科学和交叉科学。

Science
Technology
Engineering
Management

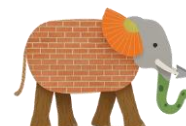
生命科学研究范式正在发生深刻变革，对生物大分子和基因的研究进入精准调控阶段，从认识生命、改造生命走向合成生命、设计生命。（理学→工学）



问题驱动：
生物学问题、数据科学问题

依赖：生物学研究方法与技术、
计算生物学方法（数据+方法）

生物信息学方法、生物学
规律研究（数据→信息→
知识→预测）

**技术**

工具（方法）

技术的特点 (1) 目的性 (2) 系统性 (3) 重复性 (4) 传承性 (5) 社会性 技术是人类伟大力量的展现，技术的先进与否，确定着一个社会、一个民族、甚至一个国家的兴衰。观其本质，技术的存在取决于人们的需要，并满足其需要。

系统性处理生物数据的能力和水平

Science
Technology
Engineering
Management

(生物技术 衍生)

整合新方法新技术：如面对大数据的压缩、整合方法、各组学的融合分析方法、各场景各类深度学习方法的整合使用、人工智能+人脑混合模型（人机整合）、全方位高精度分析方法。



生物
科学

101计划

生物信息学 课程
Bioinformatics

生物信息学人才培养模式探讨

工程
科学创造价值

是指以某组设想的目标为依据,应用有关的科学知识和技术手段,通过有组织的一群人将某个(或某些)现有实体(自然的或人造的)转化为具有预期使用价值的人造产品过程。有组织的科研:面向世界科技前沿、面向经济主战场、面向国家重大需求、面向人民生命健康。

Science
Technology
Engineering
Management

必由之路:科学探索→科学应用
→螺旋循环

从生物信息学到生物信息学技术、
生物信息工程到生物信息学管理

- 现有生物信息学专业定位:理学、工学
- 拓展生物信息学工程学的思考
- 推广生物信息学工程教育理念(工程思维)
- 加强生物信息学工程化的研究

大科学计划、大科学工程:大国之王道!

健康中国

农业强国

高质量发展

新质生产力



有组织的科研! 国家战略、国家需求、“系统”工程,各种整合!

大工程→大团队→大使命

生物信息学团队如何组织? 如何对接? 如何主动承担???

生物
科学

101计划

生物信息学 课程
Bioinformatics

生物信息学人才培养模式探讨

管理

一种行动力

是指一定组织中的管理者,通过实施计划、组织、领导、协调、控制等职能来协调他人的活动,使别人同自己一起实现既定目标的活动过程。是人类各种组织活动中最普通和最重要的一种活动。

安全、伦理、标准、宣传、应用

Science
Technology
Engineering
Management

高质量发展!
国家治理体系
和治理能力现代化!
人才培养、
专家队伍、
各类资源等
优化配置

从理论方法
(生物信息学)工程技术与应用
(生物信息技术、
生物信息工程)科学管理
(生物信息学
安全与管理)

➤ 标准化与规范化 (如数据层面)

顶层设计引领创新: 数据生产、
标准、命名、本体; 资源管理:
数据共享、数据保密、数据流失、
隐私保护等安全问题

Implementation Update: Enhancing Security Measures for NIH Controlled-Access Data

Notice Number: NOT-OD-25-083

Key Dates

Release Date: April 2, 2025

Related Announcements

- August 27, 2014 - NIH Genomic Data Sharing Policy. See notice NOT-OD-14-083.
- July 25, 2024 - Implementation Update for Data Management and Access.

Issued by

Office of the Director, National Institutes of Health (OD)

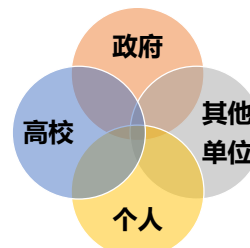
Purpose

NIH takes oversight of security including confidentiality, integrity, and availability update to enhance security measures focused on protecting data provided by NIH. Specifically, as of April 4, 2025, NIH is prohibiting access to NIH Controlled-Access data including China (including Hong Kong and Macao), Russia, Iran, North Korea, Cuba, and any other country or entity designated as a country of concern or country of interest.

Inquiries

Please direct all inquiries to: NIH Office of Science Policy gsd@mail.nih.gov

专业(方向)名称	所属学科门类及专业类	学位授予门类	年度计划招生人数	预计升 学人数	预计就 业人数
数学与应用数学 (生物统计学方向)	理学	理学类	25	17	8
数字治理	法学	法学类	20	20	0
工商管理 (数字经济管理方向)	管理学	工商管理类	40	20	20



生物
科学

101计划

生物信息学 课程
Bioinformatics

生物信息学人才培养模式探讨

教育

国之大计党之大计

吴岩副部长：实施基础学科教育教学改革试点工作计划（“101计划”），建好核心课程、核心教材、核心师资、核心实践项目，着力培养一批未来在世界上有重要影响力的杰出自然科学家、医学科学家、社会科学家。

以高等教育高质量发展全面服务支撑中国式现代化



学生
侧

- 1. 明确定位规划：**了解专业定位与培养目标，制定学习规划；结合国家需求、领域方向及拓展路径，明确职业发展目标。
- 2. 投身实践锻炼：**通过实验提升实践能力；参与科研项目，培养科研思维与合作精神。
- 3. 构建知识能力：**梳理知识体系，绘制知识图谱；推动向能力图谱转化，培育创新思维。
- 4. 自主探索前沿：**自主学习学科前沿动态；深入探索感兴趣的前沿领域。

教师
侧

- 1. 强化专业素养：**持续学习新知识新技能，紧跟学科前沿；通过学术交流，将科研成果融入教学。
- 2. 创新教学方式：**运用多元化教学方法；优化创新教学模式。
- 3. 聚焦课程资源：**参与并拓展“101计划”；整合各类教学资源。
- 4. 助力学生成长：**推动学生从知识图谱向能力图谱转化；因材施教，提供个性化指导。

生物
科学

101计划

生物信息学 课程
Bioinformatics

International Competition of Bioinformatics Engineering

