

DOI: 10.13376/j.cblls/2025004

文章编号: 1004-0374(2025)01-0026-11



江源, 博士, 馆员, 就职于中国科学院上海营养与健康研究所/生命科学信息中心, 主要从事生物领域的产业与技术情报研究工作, 先后参与中国科学院、上海市科委、苏州科技局等来源的课题。参与发表论文6篇, 参编生物制造领域著作1本。



毛开云, 中国科学院上海营养与健康研究所生命科学信息中心研究馆员, 主要从事生物领域的产业与技术情报研究、专利信息分析和知识产权分析评议工作。2016年获评全国专利信息实务人才(国家知识产权局)。先后主持和参与科技部、国家卫健委食品司、国家知识产权局、中国科学院、上海市科委等来源的课题, 主编《细胞治疗:技术与产业》等著作。

## 计算生物学前沿进展及其应用态势分析

江源<sup>1</sup>, 袁银池<sup>1</sup>, 李虹<sup>2,3</sup>, 刘樱霞<sup>1</sup>, 江洪波<sup>1,3</sup>, 毛开云<sup>1\*</sup>

(1 中国科学院上海生命科学信息中心, 中国科学院上海营养与健康研究所, 上海 200031;

2 中国科学院上海营养与健康研究所, 上海 200031; 3 中国科学院大学, 北京 100049)

**摘要:** 计算思维与技术已成为生命科学研究的关键因素, 促进了计算机科学与生物学间的融合。计算生物学推动科学家采用新方法建模、分析及解释, 利用计算工具深化科学知识。计算生物学研究从基因调控解码到细胞信号转导理解, 有望引领开创性发现。2024年, 计算生物学在分子(基因组、RNA、蛋白质)模型、细胞图谱和空间组学等领域获得了很大突破。同时, 研发投入的持续增加和个性化药物需求的不断攀升, 共同推动了相关市场的蓬勃发展。然而, 当前的计算生物学尚未构建起一个相对完备的研究体系。众多计算生物学的方法和理论尚待完善, 面对更为复杂的生物学问题, 科学家们仍在寻找合适的计算手段和方法以开展深入研究。

**关键词:** 计算生物学; 人工智能; 分子模型; 细胞图谱; 空间组学

**中图分类号:** Q-03; Q-33 **文献标志码:** A

收稿日期: 2025-01-03; 修回日期: 2025-01-15

基金项目: 上海市2024年度“科技创新行动计划”软科学研究项目(24692120900)

\*通信作者: E-mail: kymao@sinh.ac.cn

# Analysis of frontier progress and application trends in computational biology

JIANG Yuan<sup>1</sup>, YUAN Yin-Chi<sup>1</sup>, LI Hong<sup>2,3</sup>, LIU Ying-Xia<sup>1</sup>, JIANG Hong-Bo<sup>1,3</sup>, MAO Kai-Yun<sup>1\*</sup>

(1 Shanghai Information Center for Life Sciences, Shanghai Institute of Nutrition and Health, Chinese Academy of Science, Shanghai 200031, China; 2 Shanghai Institute of Nutrition and Health, Chinese Academy of Science, Shanghai 200031, China; 3 University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:** Advances in computational methods and technologies have become essential in life science research, driving the integration of computer science and biology. These developments have enabled researchers to adopt innovative approaches for modeling, analyzing, and interpreting complex biological data, significantly enhancing scientific understanding. From decoding gene regulation to elucidating cell signaling pathways, computational biology holds the promise of transformative discoveries. By 2024, significant breakthroughs have been achieved in molecular biology—encompassing genome modeling, RNA, and protein—as well as in the development of cell atlases and spatial omics. Concurrently, increased investment in research and development, coupled with the growing demand for personalized medicine, has spurred rapid growth in related markets. Nevertheless, computational biology has yet to establish a comprehensive and cohesive research framework. Many methodologies and theoretical foundations remain underdeveloped, presenting challenges as researchers confront increasingly complex biological problems. Ongoing efforts aim to refine computational approaches and identify novel strategies to enable deeper and more impactful investigations in the field.

**Key words:** computational biology; artificial intelligence; molecular model; cell atlas; spatial omics

## 1 概述

计算生物学作为一门前沿交叉学科,其发展可追溯至 20 世纪中叶,早期主要依赖简单的数学模型和统计学方法解析生物数据。随着 20 世纪 80 年代高通量检测技术的兴起,生物学数据量急剧增加,传统实验手段难以应对这一数据洪流。在此背景下,计算生物学凭借其强大的计算方法和定量分析工具,成为揭示生物体内复杂相互作用、精准预测与模拟生物过程的重要工具,并对多个学科领域产生了深远影响。如今,计算生物学已整合数学、统计学、计算机科学与传统生命科学的知识和技术,构建了一个强大的解析平台,用于深入探索复杂的生物现象。随着高性能计算 (high performance computing, HPC)、人工智能 (artificial intelligence, AI) 和机器学习 (machine learning, ML) 等技术的快速发展,计算生物学在多学科合作中不断创新,推动了基因组学、系统生物学和进化生物学等领域的突破性进展。同时,模型引导的研究方式增强了研究结果的可解释性,为实验设计提供了有力指导。计算生物学的应用范围广泛,涵盖医疗健康、环境保护等多个关键领域。例如在个性化医疗、生态系统健康评估及应对全球性挑战等方面,计算生物学所展现出的前沿技术尤为引人注目。未来,计算生物学将继续引领分子生物学研究,成为促进医药行业颠覆性创新发

展的重要制高点。

## 2 前沿进展

面向蛋白质设计的深度学习 (deep learning for protein design) 和细胞图谱 (cell atlases) 是 *Nature* 杂志 2024 年 1 月公布的年度前沿技术<sup>[1]</sup>,它们在 2024 年均取得了重要突破。在 AI 设计蛋白质方面,华盛顿大学 David Baker 团队和 Google DeepMind 分别开发了基于 AI 的蛋白质预测与设计新工具。在细胞图谱方面,人类细胞图谱 (Human Cell Atlas, HCA) 计划描绘了人体细胞的初步草图,人类肿瘤图谱网络 (Human Tumor Atlas Network, HTAN) 解析了肿瘤的起源、演化及细胞间相互作用。此外,还有多项关键进展,如首个解码 mRNA 序列大模型助力开发超高效疫苗,空间组学领域的创新技术提供了更加丰富的生物信息维度。总体来说,计算生物学在 2024 年取得了显著的进展。

### 2.1 分子模型

#### 2.1.1 基因组模型

基因组语言模型是挖掘海量基因组数据信息的一种强大新工具。它们通过深入学习 DNA 序列的复杂模式和相互依赖关系,有望显著增进我们对基因组功能及其进化历程的认识。随着大语言模型技术的兴起,科学家们积极探索其在生命科学领域的应用潜力。传统模型因受限于短 DNA 片段和有限

训练数据,泛化能力不足。美国 Arc 研究所与斯坦福大学合作开发了 Evo 模型,该模型基于 StripedHyena 框架构建,集多模态和多尺度分析于一体<sup>[2]</sup>。Evo 能够协同设计 DNA、RNA 及蛋白质分子,进而构建更为高阶的功能系统,并在分子、系统及基因组等多个层面执行预测与生成任务。通过利用 270 万个原始原核生物和噬菌体的基因组序列进行深度训练,Evo 包含 70 亿参数,处理上下文长度可达 131 千字节,实现了单核苷酸级别的精确分辨率。在 DNA、RNA 及蛋白质模态间,Evo 展现出卓越的零样本功能预测能力,其性能不仅与特定领域的语言模型旗鼓相当,甚至在部分场景下更为卓越。此外,Evo 还成功合成了具有功能活性的 CRISPR-Cas 分子复合物以及 IS200 和 IS605 转座系统,标志着语言模型在蛋白质-RNA 和蛋白质-DNA 协同设计领域的重大突破。通过全面整合全基因组信息,Evo 对核苷酸序列微小变化对生物体适应性的影响有了更为深入的理解,并能够生成长度超过 1 兆碱基、结构合理的 DNA 序列。尽管在生成大规模 DNA 序列方面已经展现出巨大的潜力,但 Evo 的性能仍需进一步的优化与微调,特别是在包含关键标记基因(如完整 rRNAs)的序列生成方面,仍面临一定的挑战。

从 DNA 序列预测分子表型是基因组学中的一个长期挑战,通常是由于注释数据有限以及无法在任务之间转移学习所致。英国伦敦 InstaDeep 开发了在 DNA 序列上进行预训练的基础模型 Nucleotide Transformer<sup>[3]</sup>。Nucleotide Transformer 的参数范围从 5 000 万到 25 亿,并整合了 3 202 个人类基因组和 850 个不同物种基因组信息。它能生成特定上下文的核苷酸序列,在低数据环境下也能预测准确。Nucleotide Transformer 微调成本低,适用于多种基因组学应用。在多个基因组学预测任务中表现出色,并且用不同物种的基因组训练的模型比只用人类似训练的模型表现更好。这表明,该模型已经学会了捕捉可能对不同物种具有功能重要性的基因组特征。预计在未来几年,基因组语言模型将取得迅猛进展。随着模型规模的扩大和更多样数据源的整合,它们将有助于揭示基因组的一些最深奥的秘密——从基因调控的规则到影响基因组结构的进化力量。

### 2.1.2 RNA模型

随着技术的不断成熟与应用范围的拓宽,mRNA 科学正逐步从实验室走向大众,开启个性化医疗与精准治疗的新时代。2023 年诺贝尔生理学或

医学奖授予了 mRNA 技术的两位奠基人——Katalin Karikó 和 Drew Weissman,以表彰他们对 mRNA 的机理研究与疫苗研发的贡献<sup>[4]</sup>。2024 年 4 月,普林斯顿大学突破性地推出 mRNA 非翻译区解码大模型 UTR-LM<sup>[5]</sup>。UTR-LM 利用先进的 Transformer 深度学习架构,这一架构在自然语言处理任务中展现出强大能力,被用于解析生物序列数据。通过整合包括基因表达谱、蛋白质结构预测以及 RNA-蛋白质相互作用等多模态信息,该模型实现了对 mRNA 非翻译区功能的深度理解与预测。在对少数物种进行模型训练后,UTR-LM 生成了数百个新的、经过优化的 mRNA 序列。这些序列在实验验证中展现出比现有疫苗序列高出 32.5% 的效率,这意味着它们可能引发更强的免疫反应,提供更持久的保护,或者在制备过程中成本更低、稳定性更高。

相同 pre-mRNA 能因组织差异进行可变剪接,产生多样转录组和蛋白质组表达谱。然而,当前算法不能预测组织特异性的可变剪接,亟须开发能够精准预测组织特异性可变剪接的算法工具。浙江大学良渚实验室推出了 SpliceTransformer (SpTransformer) 模型<sup>[6]</sup>。该模型基于 Transformer 深度学习架构,通过整合大量的转录组数据、基因组注释信息以及组织特异性表达模式,实现了对组织特异性剪接位点的预测。SpTransformer 不仅能够识别出哪些剪接事件是特定于某个或某些组织的,还能预测这些事件在不同组织中的发生概率,助力疾病遗传变异研究,并提供在线预测平台。另一方面,确定 RNA 三维结构对理解其功能和指导靶向药物开发至关重要。但 RNA 的结构灵活性导致通过实验确定的结构数据稀缺,使得计算预测工作变得困难。香港中文大学、哈佛大学、复旦大学及智峪生科合作开发了 RhoFold+ 模型<sup>[7]</sup>。RhoFold+ 采用端到端的深度学习策略,能够直接从 RNA 序列预测其三维结构,无须依赖任何已知的二级结构信息或同源比对结果。通过 RhoFold+,研究人员可以更快地解析 RNA 的结构特征,揭示其功能机制,特别是在设计针对 RNA 结构靶点的药物时,能够大大缩短研发周期,提高成功率。

### 2.1.3 蛋白质模型

蛋白质是生物体中关键的生物分子构建块,也是工业酶和众多医疗手段(包括药物和疫苗)的来源。长期以来,科学家面临的挑战是如何从线性的氨基酸序列预测蛋白质的最终结构,以及如何设计具有特定功能的蛋白质。Google DeepMind 的 Demis

Hassabis、John M.Jumpe 因对蛋白质结构的预测,与华盛顿大学的蛋白质设计先驱 David Baker 分享了 2024 年诺贝尔化学奖<sup>[8]</sup>。Google DeepMind 在 2018 年发布了首个基于 AI 的蛋白质结构预测工具 AlphaFold<sup>[9]</sup>;随后,在 2021 年又相继发表了关于 AlphaFold2<sup>[10]</sup> 和 AlphaFold-Multimer<sup>[11]</sup> 的研究论文。AlphaFold2 能根据氨基酸序列预测蛋白质三维结构,是 AlphaFold 的升级版;AlphaFold-Multimer 则进一步揭示蛋白质-蛋白质复合物结构及其相互作用机制。然而,扩大单一深度学习模型的预测范围至不同类型的复合物仍是一个巨大挑战。2024 年,Google DeepMind 发布了 AlphaFold3<sup>[12]</sup>,它不仅延续了 AlphaFold 系列在三维结构预测方面的卓越性能,更实现了对蛋白质与 DNA、RNA 以及小分子(如配体和离子)的相互作用的预测。在无需任何结构信息输入的情况下,AlphaFold3 在 PoseBusters 基准上的准确性比当前最佳的传统方法高出 50%。其免费平台 AlphaFold Server 可建模模拟蛋白质互作,预测翻译后修饰及核酸变化对蛋白质功能的影响。这种利用计算机解析蛋白质与其他分子复杂相互作用的能力,为疾病通路、基因组学、治疗靶点、蛋白质工程及合成生物学等领域带来了全新的见解。

David Baker 团队专注于计算机模拟设计全新蛋白质领域。2021 年,他们研发的“Rosetta 折叠”模型荣获 *Science* 杂志年度突破殊荣<sup>[13]</sup>。2023 年,基于“Rosetta 折叠”的深度学习模型“RFdiffusion”问世<sup>[14]</sup>,它不仅能够测试并融合多种结构元素,实现从零到一的蛋白质创造,还擅长设计氨基酸、寡聚体及具有潜在治疗或工业价值的复杂结构。2024 年,David Baker 团队推出 RoseTTAFold All-Atom 模型<sup>[15]</sup>,该神经网络模型在预测由蛋白质、核酸、小分子化合物及金属元素构成的复杂生物分子组装体结构方面展现出高精度。同时,David Baker 团队还精细调整 RFdiffusion 并推出了 Rfantibody<sup>[16]</sup>,通过输入抗原结合表位作为“热点”残基,能够高效生成针对特定抗原的单域抗体,实现了从头设计特定抗原单域抗体。当前,蛋白质设计领域正吸引着广泛的关注与投资,除了 David Baker 团队,还有诸多其他参与者。斯坦福大学 Possu Huang 团队发布了 Protpardelle 模型<sup>[17]</sup>,该模型不仅能构建蛋白质的基本结构框架,还能设计其关键功能部位——边缘上的特定原子簇。初创公司 Evolutionary-Scale 发布的 AI 算法 ESM3<sup>[18]</sup>,通过利用标注数据和整合 28 亿个蛋白质的多样化信息,为蛋白质设

计提供了有力支持。新兴 AI 药物公司(Cradle Therapeutics)<sup>[19]</sup>、老牌制药巨头(GlaxoSmithKline)<sup>[20]</sup>及科技巨头(IBM)等都在这一领域投入了大量资源,共同推动蛋白质设计技术的不断发展。

## 2.2 细胞图谱

得益于单细胞分析技术与空间组学等领域的进步,各类细胞图谱计划正迅速蓬勃发展。2024 年,人类细胞图谱、人类肿瘤图谱和哺乳动物衰老过程细胞图谱研究都有很大进展。

### 2.2.1 人类细胞图谱

人类细胞图谱(HCA)致力于深入揭示细胞基因表达模式、相互作用及其在疾病中的角色,目前正处于数据整合的关键期,预计 2026 年完成首版完整图谱<sup>[21]</sup>。2024 年 11 月 20 日,40 余篇结合新数据、新工具及综合分析的研究论文发布,涵盖骨骼形成、大脑成熟、胎盘发育、胸腺结构变化、肠道与血管细胞新状态、肺部病毒反应、基因变异影响等多个健康与疾病方面的发现<sup>[22]</sup>。图谱融合了发育、衰老的时间线索,组织结构的空间信息,以及健康和疾病的不同状态。其中,首份人类骨骼发育图谱<sup>[23]</sup>为妊娠初期骨骼发育提供了关键线索,并详细描述了相关细胞和通路机制。这些成果与 AI 新工具如 Scimilarity(寻找类似细胞)<sup>[24]</sup>、MultiDGD(多模态数据整合)<sup>[25]</sup>和 CellAgentChat(细胞相互作用建模)<sup>[26]</sup>协同,促进了细胞异质性分析的精细化,提升了多模态数据整合质量,揭示了细胞相互作用的深层机理,为精准医疗开辟了新路径。

### 2.2.2 人类肿瘤图谱网络

人类肿瘤图谱网络(HTAN)旨在构建人类癌症从癌前病变到晚期疾病演变过程中的细胞、形态和分子特征的 3D 图谱。2024 年 10 月 30 日,*Nature* 及其子刊发表了 HTAN 论文合集,全面解析肿瘤的起源、演化及细胞间相互作用。通过深入分析了近 2 000 名患者的超过 20 种类型肿瘤样本,揭示了多项创新发现<sup>[27]</sup>。例如,通过证据支持了结直肠癌可能由多种细胞协同驱动的模式,颠覆了此类癌症源自结肠黏膜单一细胞的传统认知<sup>[28]</sup>;在结直肠癌细胞向转移发展的进程中,其可塑性会呈现增强趋势<sup>[29]</sup>。这些研究还促进了新工具与新方法的开发,如美国范德比尔特大学医学中心利用一种多功能的单细胞 CRISPR 平台,开发了一种分子时钟方法,用于在体内记录细胞事件的发生时间和克隆性,并结合了细胞状态和谱系信息,为癌症发展建立一条精确的时间轴<sup>[30]</sup>。普林斯顿大学研发的 CalicoST

算法能够从空间分辨转录组学数据中推断等位基因特异性拷贝数异常并重建空间肿瘤进化<sup>[31]</sup>。

### 2.2.3 小鼠衰老细胞图谱

小鼠作为生物学研究中的经典模式动物，其衰老过程的研究对于理解人类衰老机制和开发抗衰老疗法具有重要意义。2024年，小鼠衰老细胞图谱的研究取得了显著突破。中国科学院动物研究所、华大生命科学研究院和中国科学院北京基因组研究所联合构建了全球首个高精度的泛器官衰老空间导航图(gerontological geography, GC)<sup>[32]</sup>。研究人员利用时空组学技术深入剖析了2~25月龄小鼠的心脏、肝脏、脾脏、肺、小肠等九个器官，绘制出一张包含72种细胞类型的多器官衰老时空图谱。通过对该图谱的研究，科研团队提出了组织结构熵这一新型生物学标志物，为衰老的预警与评估提供了全新视角。与此同时，洛克菲勒大学的曹俊越研究团队运用单细胞测序技术，深入分析了跨越五个生命阶段的小鼠所有主要器官中的超过2 100万个细胞，构建了迄今为止最大的哺乳动物衰老全景图谱(PanSci)<sup>[33]</sup>。这一图谱不仅揭示了3 000多种独特的细胞状态和200多种与衰老密切相关的细胞群，还深入探讨了生命进程中细胞动力学的器官、谱系和性别特异性变化。

这两项研究均采用了前沿技术手段，为构建全面、精准的小鼠衰老细胞图谱奠定了坚实基础。GG图谱侧重于小鼠多个组织器官的空间分布与细胞类型定位，而PanSci图谱则更侧重于小鼠生命五个阶段主要器官细胞的动态变化。在空间分辨率与细胞类型特异性方面，GG图谱通过空间转录组学技术实现了对小鼠组织器官中细胞类型的精确描绘，而PanSci图谱则通过单细胞测序技术揭示了不同细胞类型在衰老过程中的复杂变迁。这些研究不仅揭示了组织结构失序和细胞身份丢失是多器官衰老的共性特征，还发现了不同器官、不同细胞类型在衰老过程中的独特表现。这些发现为我们更深入地理解衰老的分子机制提供了宝贵线索，也为未来开发针对性的抗衰老疗法指明了方向。未来，随着技术的不断革新和研究的持续深入，衰老细胞图谱将在人类抗衰老事业中发挥越来越重要的作用，为延长人类健康寿命、提升生命质量贡献更多智慧与力量。

## 2.3 空间组学

### 2.3.1 空间转录组学

空间转录组学结合了单细胞测序和空间信息分

析，旨在探究基因在空间原位的表达水平和细胞的空间关系。自2020年荣获*Nature Method*年度技术以来<sup>[34]</sup>，空间转录组技术迭代迅猛，2024年更是见证了多项创新技术的问世，不仅提高了空间转录组技术的分辨率和精度，还拓展了其应用范围。德国马克斯-德尔布吕克中心推出开源平台Open-ST<sup>[35]</sup>，该平台能创建亚细胞精度的三维分子图谱，精准重建组织细胞在三维空间中的基因表达，展现了传统二维图谱难以捕捉的细节。比利时弗兰德生物技术研究院与鲁汶大学联合发布Nova-ST技术<sup>[36]</sup>，该技术采用密集纳米图案化设计，结合高密度分子索引(high-definition multimedia interface, HDMI)寡核苷酸，精准捕获组织样本中特定位置的mRNA分子。耶鲁大学樊荣团队开发Patho-DBiT技术<sup>[37]</sup>，该技术结合原位多聚腺苷酸化和计算创新，专门用于探测经过福尔马林固定和石蜡包埋处理(formalin fixed paraffin embedded, FFPE)样本中的多种RNA类型，实现空间全转录组测序。瑞典卡罗林斯卡研究所推出TRISCO技术<sup>[38]</sup>，解决了保持RNA完整性、实现均匀RNA标记和提高组织透明度等关键问题，专为大脑空间三维RNA成像设计。

### 2.3.2 空间蛋白质组学

空间蛋白质组学被*Nature Method*评为2024年度技术<sup>[39]</sup>。空间蛋白质组学凭借其卓越能力，在组织和器官切片上生成高质量且可重复使用的图像，从而深刻揭示蛋白质的组成及其精确的空间分布。这一技术已成为推动全球众多图谱项目发展的关键基石，诸如人类生物分子图谱计划(HuBMAP)与人类肿瘤图谱网络(HTAN)等前沿项目。空间蛋白质组学的持续进步离不开技术方法的突破，如新型超分辨率显微技术(SUM-PAINT)和分子像素化(MPX)方法。马克斯-普朗克生物化学研究所和哥廷根大学医疗中心联合发布一种新型高通量成像方法——SUM-PAINT<sup>[40]</sup>，它是首个实现快速、大规模蛋白质绘制与可视化的超分辨率显微技术，能以高于15 nm的分辨率对蛋白质进行多重分析。借助SUM-PAINT，研究人员绘制了包含30种不同蛋白质的单分子分辨率神经元细胞图谱，并在此过程中发现了全新的突触类型。瑞典皇家理工学院开发了一种不依赖光学设备的新方法——MPX<sup>[41]</sup>，它利用抗体-寡核苷酸偶联物和基于DNA的纳米级分子像素，实现单个细胞水平上的空间蛋白质组学研究。该方法在三维中实现了多路复用、高通量和空间分辨率的独特结合，为单细胞蛋白质组学研究引入了新的

空间维度。可以预见的是,空间蛋白质组学技术的临床转化将成为一大亮点,通过优化癌症治疗方案和指导个性化医疗实践,直接为患者带来福祉。与此同时,多组学技术的深度融合也将持续深化,形成多模态、多维度的研究平台,为科研人员提供更加全面和深入的生命科学信息。

## 2.4 其他领域

在生物力学建模与模拟领域,哈佛大学与 Google DeepMind AI 实验室合作开发了一种虚拟大鼠模型,准确地模拟了在真实大鼠身上观察到的神经活动<sup>[42]</sup>。瑞士洛桑联邦理工学院发布了新版果蝇生物力学模型 NeuroMechFly v2,能够模拟成年果蝇的具身传感运动控制<sup>[43]</sup>。在医学图像分析领域,美国斯米特心脏研究所开发了超声心动图视觉-语言基础模型 EchoCLIP,能够评估心脏功能,识别植入的心脏内装置,以及准确估计肺动脉高压、识别射血分数降低型心衰、评估左室肥厚与左心扩张严重程度、检测心包积液等<sup>[44]</sup>。此外,它还能通过计算超声心动图间的相似性,准确识别多个研究中的独特患者,以及识别随时间推移的临床变化。在疾病诊断领域,约翰霍普金斯大学医学院等机构联合开发了一种三维基因组分析技术来识别胰腺中的胰腺上皮内癌变,相关研究结果提供了迄今为止人类胰腺癌前病变最详细的三维蓝图<sup>[45]</sup>。在精准用药领域,中国科学院上海营养与健康所李虹研究员团队提出了新型深度学习模型 JointSyn,用于双视图联合学习预测药物组合的协同作用<sup>[46]</sup>。在基准数据集上,JointSyn 表现最优,证实双视角嵌入的有效性。此外,JointSyn 能利用少量实验测量微调模型,提升性能,适用于未见子集及独立数据集。在合成生物学领域,西湖大学研发了一种新的策略,即利用“三态门”电路/逻辑来设计基因线路(TriLoS)<sup>[47]</sup>。这种策略为人体细胞编写“代码”,使其能够开展智能生物计算和细胞疗法。在数据存储领域,美国北卡罗来纳州立大学和约翰·霍普金斯大学联合开发了一种新型 DNA 数据存储和计算技术<sup>[48]</sup>,利用聚合物结构“树状胶体”实现了超高数据密度的 DNA 存储和计算功能。这项突破性技术的核心是创建了一种独特的聚合物结构,能够从微观尺度以分层方式相互分支,形成纳米级纤维网络。在生态学领域,德国亥姆霍兹极地和海洋研究中心开发了一个状态依赖的动态迁徙模型,以计算迁徙鸟类从非繁殖地到繁殖地北迁时最大化其适应性的迁徙决策,预测迁徙鸟类对环境变化的适应能力<sup>[49]</sup>。

## 3 市场应用

全球计算生物学市场根据应用分类可细分为细胞和生物模拟、药物发现和疾病建模、临床试验、人体模拟软件等,其中细胞和生物模拟包括计算基因组学、计算蛋白质组学等。

### 3.1 全球计算生物学市场的蓬勃发展

全球计算生物学市场的快速增长反映了人工智能和机器学习技术在生命科学领域的广泛应用和深远影响。根据 Kings Research 发布的市场研究报告<sup>[50]</sup>,全球计算生物学市场在人工智能和机器学习技术的推动下,正展现出强劲的增长势头。预计从 2024 年的 70.8 亿美元增长到 2031 年的 210.5 亿美元,预测期内的复合年增长率为 16.85%。北美地区作为最大的市场,受益于有利的监管环境和对精准医疗的重视,其市场规模持续扩大。例如,2024 年 4 月, Xaira Therapeutics 完成了超 10 亿美元的种子轮融资,旨在利用 AI 重塑药物研发。欧洲地区以英国、法国和德国为代表,是全球第二大市场。亚太地区则是增长最快的区域,预计在预测期内将以 18.50% 的复合年增长率增长。例如,2024 年 7 月, Quantinuum 与新加坡合作,旨在增强量子计算在计算生物学中的应用。目前,国外市场已广泛涵盖各类场景,而国内则以 AI 制药为核心场景,生物模拟、疾病建模等场景开始起步<sup>[51]</sup>。

### 3.2 深度学习提升分子结构和相互作用预测

2024 年,多个团队发布了突破性工具,特别是在生物分子结构及其相互作用方式预测,以及微生物组和拉曼组数据分析方面。除了前面提到的 David Baker 团队和 Google DeepMind 更新了深度学习工具,实现对蛋白质等分子结构、受体-配体结构等全原子尺度的静态预测, Seed Health “CODA 平台”<sup>[52]</sup>、阿里云飞天实验室“LucaOne”<sup>[53]</sup>、Chai Discovery “Chai-1”<sup>[54]</sup>、Ginkgo Bioworks “ginkgo-AA-0-650m”<sup>[55]</sup>和星赛生物“拉曼智云系统”<sup>[56]</sup>等工具也相继发布(表 1),提升了生物分子结构预测的准确性与多样性。这些创新工具不仅为科学家和工程师提供了更为便捷、高效的研究手段,还极大地推动了生命科学和药物研发的快速发展。

### 3.3 生成式 AI 驱动药物发现与精准医疗

AI 技术在药物发现和疾病临床前药物开发中的应用正在逐步改变传统研发模式,加速药物研发进程,为精准医疗提供了新的可能性。通过整合多组学数据和 AI 算法,研究人员能够更快速地识别

表1 2024年计算生物学的产品进展

应用领域	发布时间	工具/模型	公司/机构	技术特点及其应用场景
细胞和生物模拟	3月7日	增强版RoseTTAFold All-Atom (RF-AA)	华盛顿大学	整合了多序列比对、模板生成和高级结构预测算法, 实现更高精度的分子结构和相互作用预测, 如蛋白质-小分子复合物的相互作用
	4月11日	计算生物学平台CODA	Seed Health	将表型数据结合微生物组、遗传、免疫、代谢组和蛋白质组数据, 发现和开发下一代精准益生菌和微生物组导向的干预措施
药物发现和疾病建模	5月8日	AlphaFold3	Google DeepMind、Isomorphic Labs	采用成对模块和直接生成三维坐标的扩散模块, 能以原子级别的精度预测蛋白质、DNA、RNA以及配体等生物分子的结构及其相互作用方式
	5月14日	LucaOne	阿里云飞天实验室	涵盖超16.9万个物种的数据, 对核酸、蛋白质的内部特征进行挖掘, 识别核酸与蛋白质之间的联系
	6月29日	拉曼智云	星寨生物	聚焦拉曼组数据类型, 为非标记、非侵入性的单细胞代谢组组分析提供云分析手段
	9月10日	Chai-I	Chai Discovery	多模态基础模型, 可预测蛋白质、小分子、DNA、RNA、共价修饰等
	9月17日	ginkgo-AA-0-650m和API “Vertex AI Model Garden”	Ginkgo Bioworks、Google Cloud	基于20多亿个专有Ginkgo蛋白质序列进行训练, 利用AI分析和理解复杂的蛋白质结构和相互作用
	2月27日	QIAGEN Biomedical KB-AI	QIAGEN	使用生成式AI提取超6.4亿个生物医学关系, 研究疾病机制, 识别出潜在的药物靶点, 进行药物重定位
	3月7日	增强版RFdiffuion All-Atom (RFdiffusionAA)	华盛顿大学	采用生成式AI系统, 从头设计全新蛋白质
	6月20日	DeepConformer	菲鹏	采用扩散模型作为生成式AI框架, 从蛋白质序列预测三维结构的多种动态构象和动态关联, 有助于药物靶点发现
	9月5日	AlphaProteo	Google DeepMind	基于给定的目标分子结构及其推荐的结合位点, 设计并生成相应的候选蛋白质
	10月27日	xTrim V3	百图生科	生命科学基础大模型, 参数规模达到2100亿, 覆盖生命科学7个模态, 并支持跨模态协作, 用于药物研发、精准医疗、生物制造等领域
11月18日	BioNeMo	NVIDIA	通过与GPU加速计算基础设施集成端到端平台, 加快药物研发流程的速度	

表1 2024年计算生物学的产品进展(续表)

应用领域	发布时间	工具/模型	公司/机构	技术特点及其应用场景
临床试验	5月30日	PathExplore™和IHC Explore1™	PathAI	AI病理学工具, 增强肿瘤表征和生物标志物识别, 解决早期临床环境中对最佳评分系统的需求
	6月27日	Stereo-seq FFPE	华大	构建全物种全转录组空间表达图谱; Stereo-seq FF V1.3能获得每个细胞内的更多高质量转录数据
	7月8日	REGLE	Google Research	无监督深度学习模型, 利用高维临床数据中的隐藏信息, 无需疾病标签, 且能整合专家知识, 提升计算效率
人体模拟软件	11月25日	Move PT	Academy Medtech Ventures	通过AI、计算机视觉和远程治疗监测来增强康复护理和患者预后
	12月2日	有享AI健康大模型	三生(中国)	集多模块于一体, 提供智能健康检测与评估, 助力个人成为自己的健康顾问与管理专家
	12月13日	知识驱动的生物标志物发现平台	菲鹏	基于AI和多组学技术, 整合涵盖逾3万种疾病、超过20万种生物标志物的知识网络, 构建“标志物筛选-机制解析-实验验证”智能研究服务体系

出具有潜力的药物靶点, 并优化药物设计流程。QIAGEN 的 Biomedical KB-AI 知识库<sup>[57]</sup> 和菲鹏的 DeepConformer 模型<sup>[58]</sup> 利用生成式 AI 提取生物学关系和预测蛋白质构象, 为深入研究疾病机制和药物靶点识别提供了新途径。Google DeepMind 的 AlphaProteo 系统<sup>[59]</sup> 和百图生科的 xTrimo V3 大模型<sup>[60]</sup> 在设计新型蛋白质和加速药物研发方面取得了显著成果, 而 NVIDIA 的 BioNeMo 开源平台<sup>[61]</sup> 则为药物研发和分子设计模型的创建与部署提供了端到端解决方案。

### 3.4 借助智能分析优化临床试验

通过智能化的临床试验工具, 研究人员能够更有效地解析复杂的疾病模式, 促进个性化治疗方案的发展。PathAI 推出的两款 AI 产品 PathExplore™ 免疫肿瘤学分析 (IOP) 与免疫组织化学 IHC Explore1™<sup>[62]</sup>, 增强了肿瘤表征能力, 使生物标志物识别更加准确。华大推出的 Stereo-seq FFPE 产品及其后续版本<sup>[63]</sup>, 为构建全物种全转录组空间表达图谱、助力临床科研理解疾病机制以及获取高质量转录本数据提供了有力支撑。Google Research 提出的 REGLE 模型<sup>[64]</sup> 则展示了如何利用无监督深度学习来揭示高维临床数据中的隐藏信息, 无须依赖疾病标签, 且可以整合专家定义的知识。

### 3.5 数字健康解决方案促进康复护理

AI 技术在人体模拟软件中的应用正推动数字健康和精准诊疗领域的快速发展。Academy Medtech Ventures 推出的 Move PT 平台<sup>[65]</sup> 结合了 AI、计算机视觉和远程监测, 改善了康复护理的效果。三生(中国)的有享 AI 健康大模型<sup>[66]</sup> 集成了多模块功能, 提供全面的健康检测与评估服务。菲鹏的知识驱动的生物标志物发现平台<sup>[67]</sup> 则聚焦于生物标志物的发现, 通过多维度证据链评估与智能因果推理相结合的知识驱动模式, 构建了一个从筛选到验证的完整服务体系, 加速了精准诊疗的实现。

## 4 展望

AlphaFold 系列革命性地将深度学习引入蛋白质结构预测领域, 其中 AlphaFold2 树立了结构生物学的新里程碑; AlphaFold3 再次突破界限, 在蛋白质-配体相互作用预测上展现出前所未有的精度与潜力。然而, AlphaFold3 目前仍局限于静态结构预测, 难以捕捉蛋白质的动态行为, 这对于理解如蛋白质折叠等复杂的细胞内过程构成了挑战。为了克服这一局限, 未来的研究或将致力于融合分子动力

学模拟技术, 以实现蛋白质动态的精准预测, 从而更全面地揭示生命的奥秘。此外, 尽管 AlphaFold3 在生物分子结构预测方面表现出色, 但在模拟真实细胞内环境方面仍有提升空间。因此, 未来的研究需要整合多尺度、多模态的数据资源, 如基因组学、转录组学和蛋白质组学, 以构建更加全面和真实的细胞模型。斯坦福大学 Charlotte Bunne 教授提出了 AI 虚拟细胞 (AIVC)<sup>[68]</sup>, 作为一种多尺度、多模态的大型神经网络模型, 能够模拟分子、细胞和组织的行为, 为生物医学研究开启了高保真模拟的新篇章。未来随着算法的不断优化、数据的日益丰富以及多学科交叉融合的深入, 蛋白质结构预测的研究领域将继续拓展至蛋白质动力学模拟、多蛋白相互作用网络解析等更为广阔的领域, 为人类揭示生命的更深层次奥秘提供强有力的支持。

在计算生物学研究中, AI 正发挥着核心作用, 高效处理包括基因组、蛋白质组及代谢组数据在内的海量信息, 深入挖掘潜藏规律与趋势, 并通过加速基因组测序、优化图像分析及提升基因诊断精确度等手段, 为计算生物学提供更精准的数据解读与诊断结果。与此同时, 数据共享与跨学科协作的趋势日益显著, 计算机科学家与生物学家携手合作, 共享资源、模型和算法, 共同推动研究进步。然而, 计算生物学也面临诸多挑战与机遇, 如研究模式的创新与整合、数据质量与技术挑战的应对、人才培养与科研合作的加强, 以及多学科协调合作机制的构建成为当前亟须解决的问题。为了真正超越传统研究模式, 需从生物学实际需求出发, 探求新方法、新手段和理论体系; 为应对数据与技术挑战, 需不断提升数据质量, 加强技术研发; 为弥补人才短缺, 需加强人才培养与跨学科素养提升, 并促进不同专业背景研究人员的交流与合作; 为提升研究水平和国际影响力, 则需集合多学科科研力量, 加强国际合作与交流, 共同推动计算生物学的发展。

随着高性能计算、机器学习和深度学习等技术的飞速发展, 计算生物学正逐步突破数据处理能力的界限, 能够深入分析更大规模、更复杂的基因组、转录组、蛋白质组等生物信息, 为揭示生物规律和构建模型提供了宝贵资源。该领域与物理学、化学、医学等多学科的深度交叉合作, 不断催生新理论、方法和技术, 推动其持续创新。在个性化医疗和精准治疗领域, 计算生物学通过分析个体基因和生物数据, 助力科学家和医生制定更为精确的治疗方案, 提升疗效并减少副作用。同时, 其在药物研发和疾

病预测上的潜力巨大, 通过模拟生物系统行为和功能, 加速新药筛选与优化, 降低研发成本, 并预测疾病趋势, 为预防和治疗提供科学依据。然而, 伴随发展而来的是数据隐私保护和基因编辑伦理规范等挑战, 要求在保障数据安全与隐私的前提下合理利用数据, 并加强伦理审查与监管, 确保技术应用的合理性、安全性和可控性。总之, 计算生物学作为一个活力四射、潜力无限的学科, 将在生命科学、医学等领域持续发挥关键作用, 取得更多突破性进展。

### [参 考 文 献]

- [1] Eisenstein M. Seven technologies to watch in 2024[EB/OL]. (2024-01-22)[2025-01-10]. <https://www.nature.com/articles/d41586-024-00173-x>
- [2] Nguyen E, Poli M, Durrant MG, et al. Sequence modeling and design from molecular to genome scale with Evo. Science, 2024, 386: eado9336
- [3] Dalla-Torre H, Gonzalez L, Mendoza-Revilla J, et al. Nucleotide Transformer: building and evaluating robust foundation models for human genomics. Nat Methods, 2024. <https://doi.org/10.1038/s41592-024-02523-z>
- [4] THE NOBEL PRIZE. The Nobel Prize in Physiology or Medicine 2023[EB/OL]. (2023-10-02)[2025-01-10]. <https://www.nobelprize.org/prizes/medicine/2023/press-release/>
- [5] Chu YY, Yu D, Li YP, et al. A 5' UTR language model for decoding untranslated regions of mRNA and function predictions. Nat Mach Intell, 2024, 6: 449-60
- [6] You NY, Liu C, Gu YX, et al. SpliceTransformer predicts tissue-specific splicing linked to human diseases. Nat Commun, 2024, 15: 9129
- [7] Shen T, Hu ZH, Sun SQ, et al. Accurate RNA 3D structure prediction using a language model-based deep learning approach. Nat Methods, 2024, 21: 2287-98
- [8] THE NOBEL PRIZE. The Nobel Prize in Chemistry 2024[EB/OL]. (2024-10-09)[2025-01-10]. <https://www.nobelprize.org/prizes/chemistry/2024/press-release/>
- [9] WIKIPEDIA. AlphaFold[EB/OL].(2025-01-04)[2025-01-10]<https://en.wikipedia.org/wiki/AlphaFold>
- [10] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature, 2021, 596: 583-9
- [11] Evans R, O'Neill M, Pritzel A, et al. Protein complex prediction with AlphaFold-Multimer. bioRxiv, 2021. doi: <https://doi.org/10.1101/2021.10.04.463034>
- [12] Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature, 2024, 630: 493-500
- [13] Science. 2021 Breakthrough of the Year[EB/OL]. (2021-12-16)[2025-01-10]. <https://www.science.org/content/article/breakthrough-2021>
- [14] Watson JL, Juergens D, Bennett NR, et al. *De novo* design of protein structure and function with RFdiffusion. Nature,

- 2023, 620: 1089-100
- [15] Krishna R, Wang J, Ahern W, et al. Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science*, 2024, 384: ead12528
- [16] Bennett NR, Watson JL, Ragotte RJ, et al. Atomically accurate *de novo* design of single-domain antibodies. *bioRxiv*, 2024[Epub ahead of print]
- [17] Chu AE, Kim J, Cheng L, et al. An all-atom protein generative model. *Proc Natl Acad Sci U S A*, 2024, 121: e2311500121
- [18] Hayes T, Rao R, Akin H, et al. Simulating 500 million years of evolution with a language model. *bioRxiv*, 2024[Epub ahead of print]
- [19] Techcrunch. Cradle builds out its protein-design AI platform (and wet lab) with \$73M in new funding[EB/OL]. (2024-11-26)[2025-01-13]. <https://techcrunch.com/2024/11/26/cradle-builds-out-its-protein-design-ai-platform-and-wet-lab-with-73m-in-new-funding/?guccounter=1>
- [20] GSK. Fine-tuning protein language models using deep mutational scanning data[EB/OL]. [2025-01-13]. <https://www.gsk.ai/blogs/fine-tuning-protein-language-models-using-deep-mutational-scanning-data/>
- [21] CNN. Atlas of cells offers a milestone leap in understanding of the human body[EB/OL]. (2024-11-21) [2025-01-10]. <https://edition.cnn.com/2024/11/21/science/atlas-cells-human-biology-body/index.html>
- [22] HUMAN CELL ATLAS. HUMAN CELL ATLAS ACHIEVES LEAP IN UNDERSTANDING OF THE HUMAN BODY[EB/OL]. (2024-11-20)[2025-01-10]. <https://www.humancellatlas.org/news/human-cell-atlas-achieves-leap-in-understanding-of-the-human-body/>
- [23] To K, Fei LJ, Pett JP, et al. A multi-omic atlas of human embryonic skeletal development. *Nature*, 2024, 635: 657-67
- [24] Heimberg G, Kuo T, DePianto DJ, et al. A cell atlas foundation model for scalable search of similar human cells. *Nature*, 2024. doi: 10.1038/s41586-024-08411-y.
- [25] Schuster V, Dann E, Krogh A, et al. multiDGD: a versatile deep generative model for multi-omics data. *Nat Commun*, 2024, 15: 10031z
- [26] Raghavan V, Li Y, Ding J. Harnessing agent-based modeling in CellAgentChat to unravel cell-cell interactions from single-cell data. *bioRxiv*, 2024[Epub ahead of print]
- [27] Nature. The Human Tumor Atlas Network (HTAN): exploring tumor evolution in time and space[EB/OL]. (2024-10-30)[2025-01-13]. <https://www.nature.com/collections/fihchcjhch>
- [28] Lu ZL, Mo SL, Xie D, et al. Polyclonal-to-monoclonal transition in colorectal precancerous evolution. *Nature*, 2024, 636: 233-40
- [29] Moorman A, Benitez EK, Cambulli F, et al. Progressive plasticity during colorectal cancer metastasis. *Nature*, 2024. doi: 10.1038/s41586-024-08150-0. [Epub ahead of print]
- [30] Islam M, Yang YL, Simmons AJ, et al. Temporal recording of mammalian development and precancer. *Nature*, 2024, 634: 1187-95
- [31] Ma C, Balaban M, Liu JX, et al. Inferring allele-specific copy number aberrations and tumor phylogeography from spatially resolved transcriptomics. *Nat Methods*, 2024, 21: 2239-47
- [32] Ma S, Ji ZJ, Zhang B, et al. Spatial transcriptomic landscape unveils immunoglobulin-associated senescence as a hallmark of aging. *Cell*, 2024, 187: 7025-44.e34
- [33] Zhang ZH, Schaefer C, Jiang WR, et al. A panoramic view of cell population dynamics in mammalian aging. *Science*, 2025, 387: eadn3949
- [34] Nature Method. Method of the Year 2020: spatially resolved transcriptomics[EB/OL]. (2021-01-06)[2025-01-13]. <https://www.nature.com/collections/dfibfggef>
- [35] Schott M, León-Periñán D, Splendiani E, et al. Open-ST: high-resolution spatial transcriptomics in 3D. *Cell*, 2024, 187: 3953-72.e26
- [36] Poovathingal S, Davie K, Borm LE, et al. Nova-ST: nano-patterned ultra-dense platform for spatial transcriptomics. *Cell Rep Methods*, 2024, 4:100831
- [37] Bai ZL, Zhang DY, Gao Y, et al. Spatially exploring RNA biology in archival formalin-fixed paraffin-embedded tissues. *Cell*, 2024, 187: 6760-79.e24
- [38] Kanatani S, Kreutzmann JC, Li Y, et al. Whole-brain spatial transcriptional analysis at cellular resolution. *Science*, 2024, 386: 907-15
- [39] 搜狐网.未来已来:空间蛋白质组学获评Nature Methods年度技术! [EB/OL]. (2024-12-10)[2025-01-13]. [https://www.sohu.com/a/834898359\\_100258692](https://www.sohu.com/a/834898359_100258692)
- [40] Unterauer EM, Shetab Boushehri S, Jevdokimenko K, et al. Spatial proteomics in neurons at single-protein resolution. *Cell*, 2024, 187:1785-800.e16
- [41] Karlsson F, Kallas T, Thiagarajan D, et al. Molecular pixelation: spatial proteomics of single cells by sequencing. *Nat Methods*, 2024, 21:1044-52
- [42] Aldarondo D, Merel J, Marshall JD, et al. A virtual rodent predicts the structure of neural activity across behaviours. *Nature*, 2024, 632: 594-602
- [43] Wang-Chen SB, Stimpfling VA, Lam TKC, et al. NeuroMechFly v2: simulating embodied sensorimotor control in adult *Drosophila*. *Nat Methods*, 2024, 21: 2353-62
- [44] Christensen M, Vukadinovic M, Yuan N, et al. Vision-language foundation model for echocardiogram interpretation. *Nat Med*, 2024, 30:1481-8
- [45] Braxton AM, Kiemen AL, Grahn MP, et al. 3D genomic mapping reveals multifocality of human pancreatic precancers. *Nature*, 2024, 629: 679-87
- [46] Li XL, Shen BH, Feng FYM, et al. Dual-view jointly learning improves personalized drug synergy prediction. *Bioinformatics*, 2024, 40: btae604
- [47] Shao JW, Qiu XY, Zhang LH, et al. Multi-layered computational gene networks by engineered tristate logics. *Cell*, 2024, 187: 5064-80.e14
- [48] Lin KN, Volkel K, Cao C, et al. A primordial DNA store and compute engine. *Nat Nanotechnol*, 2024, 19: 1654-64

- [49] Lisovski S, Hoye BJ, Conklin JR, et al. Predicting resilience of migratory birds to environmental change. *Proc Natl Acad Sci U S A*, 2024, 121: e2311146121
- [50] Kings Research. 计算生物学市场规模、份额、增长和行业分析[EB/OL].(2024-10)[2025-01-09]. <https://www.kingsresearch.com/zh/computational-biology-market-397>
- [51] 量子位智库. 计算生物学深度产业报告[EB/OL]. (2022-08-09)[2025-01-09]. <https://www.qbitai.com/2022/08/36776.html>
- [52] Epicos. Seed Health Launches Computational Biology Platform, Powered by Most Comprehensive Host-Microbiome Data Set, to Expand Pipeline Beyond the Gut[EB/OL]. (2024-04-11)[2025-01-09]. <https://www.epicos.com/article/823951/seed-health-launches-computational-biology-platform-powered-most-comprehensive-host>
- [53] He Y, Fang P, Shan YT, et al. LucaOne: generalized biological foundation model with unified nucleic acid and protein language. *bioRxiv*, 2024[Epub ahead of print]
- [54] Asif Razzaq. Chai-1 Released by Chai Discovery Team: A Groundbreaking Multi-Modal Foundation Model Set to Transform Drug Discovery and Biological Engineering with Revolutionary Molecular Structure Prediction[EB/OL]. (2024-09-10)[2025-01-09]. <https://www.marktechpost.com/2024/09/10/chai-1-released-by-chai-discovery-team-a-groundbreaking-multi-modal-foundation-model-set-to-transform-drug-discovery-and-biological-engineering-with-revolutionary-molecular-structure-prediction/>
- [55] Science AI. 新型蛋白质大语言模型即将登陆Google Cloud[EB/OL]. (2024-09-18)[2025-01-13]. [https://baijiahao.baidu.com/s?id=1810523949152578182&wfr=s\\_pider&for=pc](https://baijiahao.baidu.com/s?id=1810523949152578182&wfr=s_pider&for=pc)
- [56] 青岛高创科技资本运营有限公司. 园区企业星赛生物发布一站式AI云专家系统[EB/OL]. (2024-06-12)[2025-01-13]. <http://www.htcap.com/index.php?m=home&c=View&a=index&aid=472>
- [57] QIAGEN. QIAGEN launches AI-derived biomedical knowledge base to accelerate data-driven drug discovery[EB/OL]. (2024-02-27)[2025-01-09]. <https://corporate.qiagen.com/English/newsroom/press-releases/press-release-details/2024/QIAGEN-launches-AI-derived-biomedical-knowledge-base-to-accelerate-data-driven-drug-discovery/default.aspx>
- [58] 中国经济时报. 菲鹏发布国内首个知识驱动的生物标志物发现平台[EB/OL]. (2024-12-19)[2025-01-13]. <https://www.cet.com.cn/itpd/itxw/10149808.shtml>
- [59] Zambaldi V, La D, Chu AE, et al. De novo design of high-affinity protein binders with AlphaProteo. *arXiv*, 2024[Epub ahead of print]
- [60] 中国日报. 中国生物计算大会 百图生科发布全球首个2100亿参数生物大模型[EB/OL]. (2024-10-29)[2025-01-13]. <http://ex.chinadaily.com.cn/exchange/partners/82/rss/channel/cn/columns/sn19a7/stories/WS6720a49aa310b59111da08d8.html>
- [61] NVIDIA. NVIDIA Opens BioNeMo to Scale Digital Biology for Global Biopharma and Scientific Industry[EB/OL]. (2024-11-18)[2025-01-09]. <https://nvidianews.nvidia.com/news/nvidia-opens-bionemo-to-scale-digital-biology-for-global-biopharma-and-scientific-industry>
- [62] 动脉网. 医疗诊断技术研发商PathAI推出用于癌症研究的全新AI病理学工具[EB/OL]. (2024-05-30)[2025-01-13]. <https://www.vbdata.cn/newsDetail/dd8c7f8a1e8b11ef9dde00163e034e34>
- [63] 腾讯网. 华大时空转录组FFPE产品方案重磅发布: 编码及非编码RNA共捕获! 可实现真正的空间单细胞水平分析[EB/OL]. (2024-06-28)[2025-01-13]. [https://news.qq.com/rain/a/20240628A09M3O00?suid=&media\\_id=](https://news.qq.com/rain/a/20240628A09M3O00?suid=&media_id=)
- [64] Yun T, Cosentino J, Behsaz B, et al. Unsupervised representation learning on high-dimensional clinical data improves genomic discovery and prediction. *Nat Genet*, 2024, 56: 1604-13
- [65] 动脉网. 健康技术公司Academy Medtech Ventures推出改变康复和患者护理的AI平台Move PT[EB/OL]. (2024-11-21)[2025-01-13]. <https://www.vbdata.cn/newsDetail/7f04ea08a82211efb14600163e034e34>
- [66] 新京报. 三生(中国)发布AI健康大模型, 数智赋能行业发展[EB/OL]. (2024-12-13)[2025-01-13]. [https://baijiahao.baidu.com/s?id=1818312086188908671&wfr=s\\_pider&for=pc](https://baijiahao.baidu.com/s?id=1818312086188908671&wfr=s_pider&for=pc)
- [67] 新华报业网. 菲鹏发布国内首个知识驱动的生物标志物发现平台[EB/OL]. (2024-12-19)[2025-01-13]. <https://www.xhby.net/content/s6763e74be4b0595c185fac5c.html>
- [68] Bunne C, Roohani Y, Rosen Y, et al. How to build the virtual cell with artificial intelligence: priorities and opportunities. *Cell*, 2024, 187: P7045-63