



陈大明, 研究馆员, 长期从事战略研究、产业研究和科技信息分析, 主持多项国家科技重大专项、上海市软科学研究项目等, 为国家和区域生物科技发展和政策管理提供决策参考, 发表论文 100 余篇, 主编著作 8 部。

数字生命的演进历程及发展态势

张博文¹, 江源¹, 郑森予¹, 王子晨², 马雷², 毛开云¹, 熊燕¹, 陈大明^{1*}

(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 北京大学, 北京 100871)

摘要: 数字生命代表了生命科学与数字技术交织的前沿领域, 其核心在于构建跨越层级与时空的数字孪生体, 以此揭示生物系统的涌现规律及动态交互机制。本文通过梳理数字生命从理论假说到工程实现的演化历程, 阐明了其推动生命科学研究从被动“观测描述”向主动“预测设计”跨越的内在逻辑。本文重点分析了支撑这一范式变革的技术支柱, 包括基于高通量与超分辨成像的数据获取、利用大模型解析生物特征的解码技术、强调干湿闭环验证的活动模拟, 以及最终集成的生命数字孪生系统。目前, 随着人工智能虚拟细胞、数字器官及数字线虫等全生物体模型的问世, 通过计算手段预测并精准调控生命行为已初步成为现实。然而, 针对多源异构数据整合难、跨尺度模型解释性弱及伦理治理滞后等瓶颈, 未来亟需深化对生命涌现机制的理论认知, 开发内嵌生物物理约束的新型架构, 并建立标准化的协作与治理体系, 从而真正实现对生命全景的解析与重构。

关键词: 数字生命; 生物数字孪生; 跨尺度建模; 干湿结合

中图分类号: Q811.4; TP399 **文献标识码:** A

Evolutionary trajectory and development landscape of digital life

ZHANG Bo-Wen¹, JIANG Yuan¹, ZHENG Sen-Yu¹, WANG Zi-Chen², MA Lei²,
MAO Kai-Yun¹, XIONG Yan¹, CHEN Da-Ming^{1*}

(1 Shanghai Information Center for Life Sciences, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, Shanghai 200031, China; 2 Peking University, Beijing 100871, China)

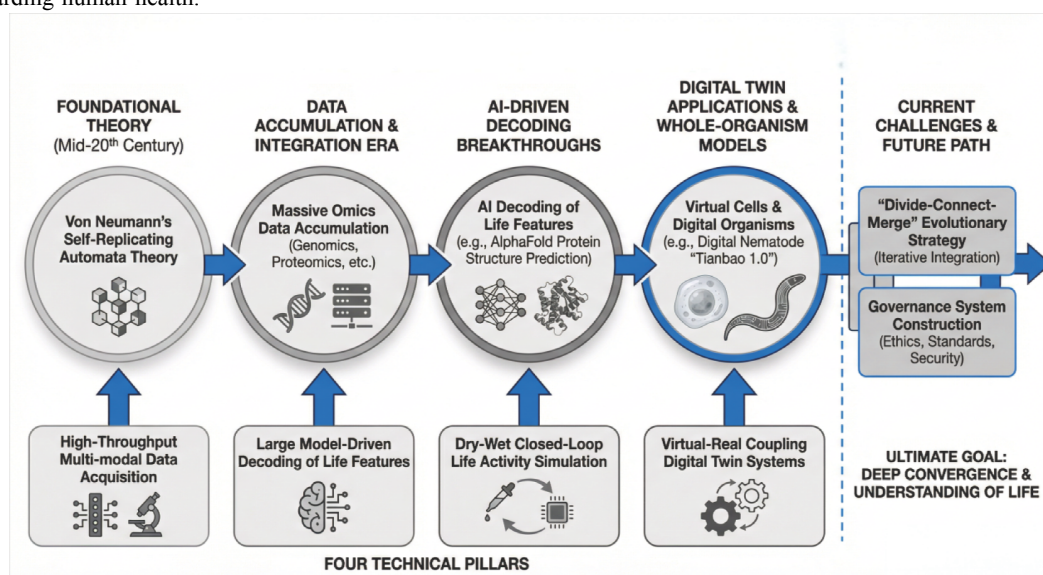
Abstract: Digital life, as a cutting-edge interdisciplinary field integrating life sciences and digital technologies, has become a pivotal research direction driving the paradigm shift of life science research. This paper aims to systematically trace the evolutionary trajectory of digital life from theoretical hypothesis to engineering implementation, deeply dissect its core technical system and current development status, identify the key bottlenecks restricting its in-depth development, and propose targeted development strategies and future prospects, thereby providing a comprehensive reference for academic research, industrial transformation and the construction of governance systems in this field. The study first combs the multi-

收稿日期: 2026-01-20; 修回日期: 2026-02-09

基金项目: 中国科学院学部咨询评议项目(3-ZXB2025025)

*通信作者: E-mail: chendaming@sinh.ac.cn

stage development process of digital life: starting with von Neumann's self-replicating automaton theory that laid the theoretical foundation of artificial life, it has experienced the stage of large-scale accumulation of life omics data driven by high-throughput sequencing and super-resolution imaging technologies, the stage of breakthroughs in biological feature decoding relying on synthetic biology and artificial intelligence (AI) models such as AlphaFold, and then entered the stage of practical application of digital twin technology, with the birth of a series of whole-organism models including AI virtual cells, digital organs, intelligent nematode "Tianbao 1.0", virtual rats and digital fruit flies. On this basis, the paper elaborates the four core technical pillars supporting the development of digital life: life data acquisition with high-throughput, ultra-high resolution and multi-modal characteristics, life feature decoding empowered by large models such as transformers and graph neural networks, life activity simulation emphasizing wet-dry closed-loop validation and multi-scale model iteration, and the integrated life digital twin system that realizes dynamic coupling between virtual and real entities. It also reveals the core connotation of digital life research in realizing the in-depth integration of spatial, temporal and logical dimensions, and constructing a digital system that conforms to the laws of real life and has controllability. In the research, it is found that digital life is currently in a critical period from concept verification to system engineering, and is facing prominent bottlenecks such as difficulty in integrating multi-source heterogeneous data, weak interpretability of cross-scale models, insufficient understanding of the dynamic causality of life systems, lagging ethical governance systems, and barriers in interdisciplinary collaborative innovation. In response to these problems, this paper holds that the future development of digital life should follow the progressive evolution path of "division-conjunction-integration". It is imperative to further deepen the theoretical understanding of the emergent mechanisms of life systems, develop new AI architectures embedded with biophysical constraints to improve the interpretability and predictability of models, establish standardized data platforms and model interfaces to break interdisciplinary barriers and promote open-source collaboration, and construct a comprehensive governance system covering intellectual property protection, ethical review and safety supervision. Only by overcoming the above technical and institutional bottlenecks can we truly realize the comprehensive parsing and reconstruction of the life panorama, promote the mature application of digital life in precision medicine, drug research and development, molecular engineering and other fields, and make digital life an important scientific infrastructure for revealing the mysteries of life and safeguarding human health.



Key words: digital life; biological digital twin; cross-scale modeling; integration of wet and dry experiments

“生命的本质?”这一命题,自薛定谔提出“生命以负熵为食”的经典论断以来,持续引发科学界的深入思考。伴随计算技术的发展,一个更具时代特色的问题浮出水面:“生命能否以数字化形式进行描述与重构?”早在20世纪中叶,约翰·冯·诺依曼(von Neumann)在其具有开创性意义的自复制自动机理论中,首次从数学和逻辑维度论证了生命系统所固有的可计算性本质——生命过程在理论层面能够被

抽象为一系列指令的存储、传递与执行^[1]。这一思想理念,历经数十年系统生物学、计算模型以及高性能计算的发展积淀,在21世纪迎来了新的发展契机。海量的多组学数据、超高分辨率成像技术以及呈指数级增长的计算能力,促使构建跨越分子、细胞、器官乃至个体层面的动态的、可计算的、可干预的数字化映射——数字生命体,从理论设想转变为工程现实。鉴于此,诸多学者已从数字孪生、生物建模等视

角对数字生命展开研究^[2,3],尽管不同研究对其定义存在一定差异,但其内涵均指向以多模态生物数据为基础,以人工智能与数字孪生技术为核心工具构建的可计算、可生成、可干预的生命系统数字化镜像。基于此,本文围绕以下核心定义展开深入探讨:通过构建跨层级数字化映射与多时间尺度融合的数字孪生体,揭示生物系统的涌现规律,最终实现对生命动态行为的预测、编程与精准调控。

当前,推进数字生命的研发不仅有助于深化对生命本质的认知,更是应对精准医疗、药物研发等领域严峻挑战的迫切需求。对其演进历程与发展态势进行系统性梳理,是因为支撑其发展的关键技术生态渐趋成熟。同时,人类基因组计划等大科学工程所积累的协作与治理经验为其发展提供了可资借鉴的范例。本文将梳理从“生命数据采集”到“生命数字孪生”的技术体系,审视人工智能虚拟细胞(AIVC)、数字器官(AIDO)等前沿成果,并剖析数据整合、模型验证以及伦理治理等方面存在的核心问题。最终,本文旨在描绘数字生命引领生命科学研究从被动的“观测描述”迈向主动的“预测设计”的范式变革路径,勾勒“分治-连接-统合”的未来发展图景,通过揭示该领域推动生命科学从成本高昂的物理试错走向数字设计与模拟的过程,呈现其在分子工程、细胞调控、临床诊疗等领域的应用状况,同时正视数据异构、跨尺度耦合、伦理治理等核心挑战,

最终描绘未来发展方向,为该领域的学术研究、产业转化与治理体系构建提供参考。

1 数字生命的发展历程

数字生命的发展并非一蹴而就,而是伴随着生命科学、计算科学和工程技术的协同演进,经历了从理论萌芽、数据积累、计算建模到智能融合的多个阶段(图1),其核心理念逐渐从对生命现象的静态描述转向对生命过程的动态模拟与主动设计,渐进衍生出数字生命的“大科学”研究设想。

在早期的理论萌芽与概念奠基阶段,“生命是什么”吸引众多科学家探索,从1674年列文虎克用显微镜观察原生动植物^[4],到1869年米歇尔分离出“核素”(核酸)^[5],再到DNA双螺旋结构^[6]的揭示以及“中心法则”^[7]的提出,信息流的探究为“生命作为信息处理系统”的数字化构想埋下伏笔。

技术创新则为构建数字生命的底层逻辑——即通过计算理解生命系统的多尺度行为——奠定了数据基础。在引物延伸法^[8]的基础上,测序技术^[9]的突破为人类基因组计划^[10]奠基,其后开发的单细胞RNA测序^[11]、空间转录组学^[12]等技术以及人类代谢组学计划^[13]、国际人类蛋白质组计划^[14]、人类细胞图谱计划^[15]、国际人类表型组计划^[16]、人类肿瘤图谱网络计划^[17]等标志着生命组学数据达到了前所未有的规模,肾脏成像组计划^[18]、脑多组学图谱

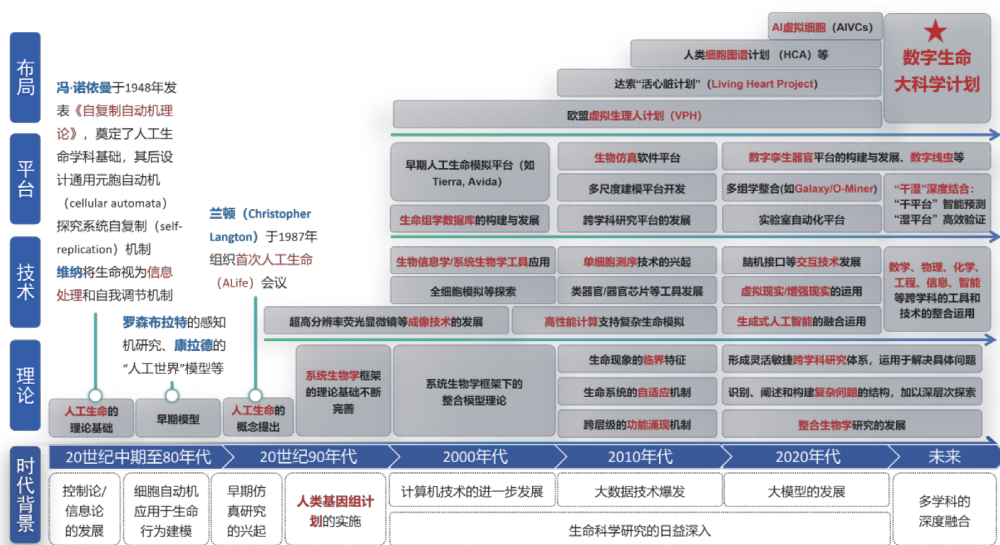


图1 数字生命的发展历程

Figure 1 The development process of digital life

计划^[19]等组织特异性的研究也由此展开。

对生物数据的深度“解码”则从“合成”与“预测”两条路径展开。在合成路径上,2002年人工合成病毒^[20]验证了生命可被“编写”,其后“辛西娅”人造细胞^[21]、人工合成酵母基因组计划^[22]的实施以及人造单染色体酵母^[23]的构建,验证了人类对生命数据的认知可转化为从头设计、构建与调试生命系统的工程能力。在预测路线上,AlphaFold^[24]等分子结构工具的突破,以及小鼠胚胎发育时空图谱^[25]、猕猴全细胞图谱^[26]、小鼠脑空间蛋白质组图谱^[27]、小鼠衰老全景图谱^[28,29]、小鼠全脑3D图谱^[30]等从细胞到器官层级的研究,再到中国免疫多组学图谱^[31]、果蝇发育3D全景图谱^[32]等功能层级的研究,再到Evo模型^[33]的发布,标志着数字生命研究从静态“读取”迈入动态“理解”的新阶段。

在“理解生命”的基础上,“构建与预测生命”的研究也在推进。2010年美国国家航空航天局(NASA)提出了数字孪生概念^[34],2014年达索系统启动了“活心脏计划”^[35]。近年来,研究进一步向个体层面的数字孪生拓展:2022年智能线虫“天宝1.0”以秀丽隐杆线虫为原型,结合强化学习与脑-体-环境闭环仿真,在功能层面实现了线虫运动控制,验证了神经控制策略可生成稳定的类生物运动行为^[36];2024年虚拟大鼠模拟了在真实大鼠身上观察到的神经活动^[37];2025年“数字果蝇”^[38]模拟了飞行、行走、越障等生物运动行为。与此同时,作

为生命基本单元的细胞成为虚拟化核心,2024年提出的AI虚拟细胞^[39]旨在建立可预测细胞状态、命运与响应的通用计算模型。2025年的肾脏单细胞基础模型Nephrobase Cell+^[40],实现了肾脏细胞的高精度解析与扰动预测。

2 数字生命的发展现状

数字生命研究的底层逻辑根植于对生命系统层级耦合与时序动态的深度解析,其核心在于实现从“观测描述”向“预测设计”的范式跃迁。这一跃迁的驱动力来源于“生命数据采集-生命特征解码-生命活动模拟-生命数字孪生”的体系化技术演进(图2)与递进逻辑。

在数据采集层面,高通量、跨时空、超高分辨成像等技术成功突破了生命系统在基因、分子、细胞、器官、个体等多个层级的观测瓶颈,实现了多模态数据的系统性获取。在特征解码层面,以大模型为代表的人工智能技术通过对多源异构数据的整合,解析生命系统的调控网络,识别其中的因果逻辑与反馈逻辑,进而揭示从分子相互作用到系统功能涌现的跨层级机制。在活动模拟层面,“干平台智能预测-湿平台高效验证-多尺度模型迭代”的研发范式,将计算预测与实验验证进行闭环结合,实现了对生命动态过程的可计算模拟。最终,在数字孪生层面,通过对结构嵌套、信息整合、反馈调控等机制进行建模,构建具备可计算、可模拟、可交互、可预测、可编

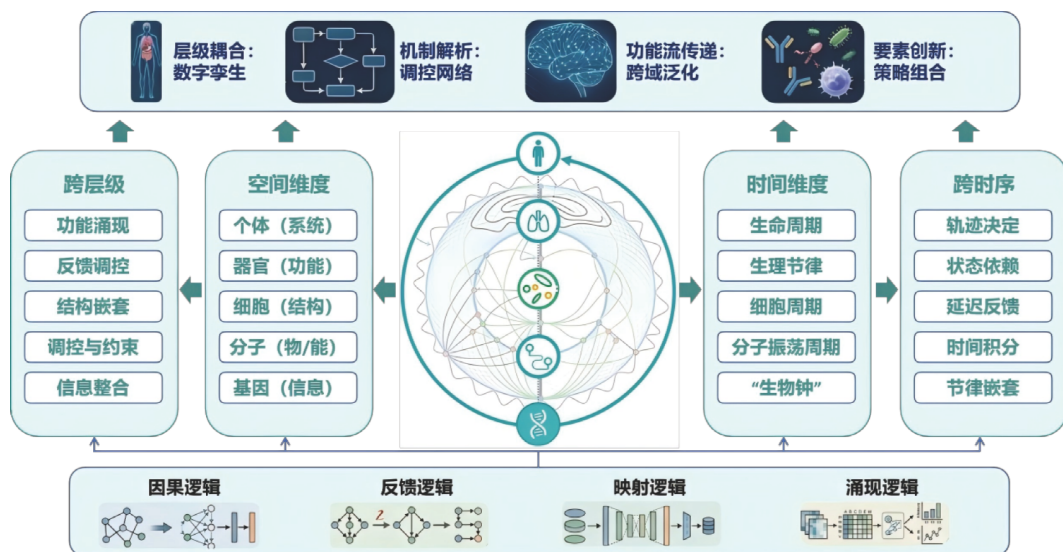


图2 数字生命的多维度整合

Figure 2 Multi-dimensional integration of digital life

程能力的数字生命体,完成从数据到知识、再到可操作模型的转化。

如图2所示,数字生命研究的核心在于实现空间、时间和逻辑三个维度的深度整合,从而构建出一个既符合真实生命规律,又具备可控性的数字体系。在具体机制上,空间维度构建了生命的物质载体,通过从基因、分子到细胞、器官再到个体的层级嵌套,打通了局部与整体的信息流动;时间维度则赋予了系统动态演变的灵魂,它利用分子振荡、细胞周期、生理节律直至生命周期的多重节律嵌套,配合时间积分与反馈机制,决定了数字生命的演化轨迹。而逻辑维度充当了联结两者的纽带,利用因果、反馈与涌现机制,将静态的空间结构与动态的时间演变耦合起来。空间提供“舞台”,时间提供“剧本”节奏,逻辑则制定“演出”规则,只有三者紧密协作,才能真正实现对数字生命的精准模拟与调控,满足跨越泛化等高阶研究需求。也因此,一系列可运行、可验证、可应用的生物数字孪生体得以研发。

2.1 生命数据采集的研究发展

数字生命的核心在于构建跨层级数字化映射、多时间尺度融合的数字生命孪生体,这一目标的实现离不开对生命系统全维度、高保真数据的捕获。数据采集端的技术突破提供了多模态、跨时空分辨率数据,构成了数字映射的底层“语料”(表1)。从高通量组学测序、超分辨生物成像到单细胞时空组学、多模态生理监测等技术的飞速发展,不仅突破了基因、分子、细胞、器官乃至个体等多个生命层级的观测瓶颈,更实现了对生命活动跨时空、多维度的系统性数据捕获。

这些数据库的构建源于数据采集技术的进步,包括基因组学领域的第三代测序技术等^[56]的突破、转录组学研究的全转录本图谱的构建^[57]、蛋白质组学和代谢组学的“功能”分子研究^[58]、脂类组学和糖组学对细胞膜结构和细胞间通讯的理解^[59]、免疫组学的免疫表型解析等^[60],以及基于多组学的“细胞状态图谱”(Cell State Atlas)研究^[61]等。

其中,单细胞技术已发展至利用微流控芯片、条形码系统和高通量测序的集成技术,将同一个细胞内的基因组、表观基因组、转录组,甚至蛋白质组信息同步捕获^[62],实现“分子条形码”技术数据采集(如SNARE-seq、SHARE-seq等)^[63]。由此,单细胞RNA测序(scRNA-seq)技术可用于捕获转录组信

息,为揭示细胞多样性、发育动态及疾病相关基因调控网络提供强有力工具^[64];单细胞蛋白质组学(如SCoPE-MS)技术则可用于单个细胞水平的蛋白质定量^[65]。在此基础上,单细胞建模技术应运而生,用于捕捉细胞内分子事件的随机性^[66],其应用场景包括预测干细胞分化路径、肿瘤细胞耐药机制以及免疫细胞激活过程^[67]等方面。

在时空动态模拟(spatial-temporal dynamics)的研究中,空间转录组学(Spatial Transcriptomics)技术中的Stereo-seq技术通过纳米孔阵列实现了亚细胞级别的基因表达定位^[68],能够在单细胞,甚至亚细胞水平绘制出组织切片的基因表达图谱,应用于癌症微环境研究^[69]、免疫治疗靶点发现^[70]等场景。空间组学技术与计算建模的结合,使得构建三维细胞模型成为可能^[71],细胞内基因表达的空间分布可映射到细胞器或亚区室^[72]。此外,基于物理引擎的细胞模拟平台(如CellSim、MCell)能够模拟分子扩散、结合与信号传播过程,实现对细胞内动态过程的可视化与量化分析^[73]。

生物医学成像技术极大地拓宽了数字生命的可视化边界。超分辨荧光显微镜(如SIM、STED、STORM、PALM)突破了传统光学显微镜约200 nm的衍射极限,将空间分辨率提升至约10 nm,甚至接近1 nm的水平,实现了纳米尺度下活细胞的动态可视化^[74]与毫秒级高速成像。MINIFLUX技术凭借其极高的定位效率与仅需极少光子即可实现纳米级分辨的优势,成功捕捉到DNA折纸结构中相距仅约6 nm的分子细节,分辨率已逼近原子级精度^[75]。在功能成像与多模态成像方面,磁共振成像(MRI)与正电子发射断层扫描(PET)正从传统的宏观结构观察逐步向微观功能解码转变,功能性MRI(fMRI)与深度学习的结合^[76]、PET技术与深度学习的融合,为“活体活检”提供了更高质量的数据支持^[77]。这些成像技术与无创或微创的生化检测等技术的融合应用,为数字表型学(Digital Phenotyping)^[78]研究提供了更丰富更多维的数据,从而推动了面向特定人群的个性化“动态健康地图”的开发^[79]。

在上述场景中,“人类细胞图谱计划”(Human Cell Atlas, HCA)已绘制涵盖十多类人体器官的细胞参考图谱^[80]。同时,人类胚胎头骨、髌膝肩等骨骼的多组学发育研究数据^[81]、人类神经类器官整合转录组图谱等数据,以及用于通过表达谱快速搜索

表1 数字生命代表性基础数据库概览
Table 1 Overview of representative digital life basic databases

数据库 (中文名)	数据库	核心功能	所属领域	数据规模	更新频率	主要应用案例
基因序列数据库	GenBank	存储海量核酸序列及注释信息,支持序列比对、检索与提交	核酸序列与基因组学	包含超过25亿条序列记录,数据量达TB级	每日更新	COVID-19 病毒变异株的早期序列鉴定与追踪 ^[41]
参考序列数据库	RefSeq	提供经过筛选、注释的高质量参考序列(核酸、蛋白质),作为基因组分析的基准	参考基因组与序列注释	覆盖超过1.2亿条蛋白质记录和4 000万条核苷酸记录	每日更新,定期发布大版本	临床基因检测中用于标准化的变异位点注释(如BRCA1基因突变比对) ^[42]
微生物基因组整合数据库	IMG(Integrated Microbial Genomes)	整合微生物(细菌、古菌、真菌等)基因组数据,支持微生物基因组功能注释与比较分析	微生物基因组学	收录超过150 000个微生物基因组和宏基因组数据集	定期更新(随JGI基因组数据库更新)	发现新型CRISPR-Cas系统或极端环境微生物的代谢途径挖掘 ^[43]
通用蛋白质数据库	UniProt	整合蛋白质序列、功能、结构及相关文献信息,是蛋白质研究的核心资源	蛋白质组学	UniProtKB包含超过2.5亿条序列,其中Swiss-Prot(人工注释)约57万条	每8周发布一次大版本	鉴定未知蛋白的功能结构域,例如分析肿瘤抑制蛋白p53的修饰位点 ^[44]
蛋白质结构数据库	PDB(Protein Data Bank)	存储实验测定的生物大分子(蛋白质、核酸等)三维结构数据	生物大分子结构学	包含超过215 000个实验测定的大分子结构	每周更新	药物设计中的靶点结构分析,如基于HIV蛋白酶结构的抗艾滋病药物开发 ^[45]
阿尔法折叠蛋白质结构数据库	AFDB(AlphaFold Database)	提供基于AI预测的蛋白质三维结构数据,补充实验的不足	计算结构生物学	包含超过2亿个蛋白质的预测结构,覆盖几乎所有已知序列的蛋白质	不定期重大更新(随DeepMind发布)	解决长期未知的核孔复合物结构组装问题;加速冷门疾病靶点研究 ^[46]
蛋白质相互作用数据库	StringDB	整合实验与预测的蛋白质-蛋白质相互作用数据,构建相互作用网络	蛋白质互作组学	覆盖超过14 000种生物,包含6 700万种蛋白质和200亿个互作关系	约每2年发布大版本更新	构建癌症信号转导网络,筛选关键枢纽蛋白(Hub genes)作为潜在药物靶点 ^[47]
京都基因与基因组百科全书	KEGG	存储基因、蛋白质、代谢物及代谢通路、信号通路等数据,支持通路分析	代谢组学与通路分析	包含超过600张手绘通路图及数百万基因条目	每周持续更新	解释转录组测序数据中的差异基因富集通路(如细胞凋亡通路) ^[48]
生物通路数据库	Reactome	以生物过程为核心,构建标准化的分子通路模型,解析细胞内分子互作机制	分子通路与细胞生物学	包含约2 600个人类生物通路和15 000个反应	每季度发布一次更新	系统性分析药物作用机制,例如他汀类药物对胆固醇合成通路的影响 ^[49]
基因表达数据库	GEO(Gene Expression Omnibus)	存储高通量基因表达数据(如芯片、RNA-seq),支持表达谱差异分析	转录组学	存储超过600万个样本和18万个数据集	每日实时更新	通过荟萃分析(Meta-analysis)挖掘不同癌症共有的生物标志物 ^[50]
DNA元件百科全书	ENCODE(Encyclopedia of DNA Elements)	解析人类基因组中功能性DNA元件(如启动子、增强子)的分布与功能	表观基因组学与功能基因组学	包含数千个ChIP-seq、DNase-seq等表观遗传学实验数据集	项目阶段性发布,数据持续维护	鉴定非编码区GWAS风险位点(SNP)对基因表达的调控机制 ^[51]
人类蛋白质图谱数据库	HPA(Human Protein Atlas)	绘制人类蛋白质在组织、细胞中的表达定位图谱,关联蛋白质表达与疾病	人类蛋白质组学	覆盖超过17 000种蛋白质在44种正常组织和20种癌症组织中的免疫组化图像	每年发布更新版本	确定特定癌症(如乳腺癌)中预后相关的蛋白标志物及其亚细胞定位 ^[52]
细胞系基因数据库	CellLineGene	整合不同细胞系的基因表达、突变及表型数据,支持细胞模型研究	细胞生物学与肿瘤研究	涵盖1 000+种人类癌细胞系的基因组、转录组及药物敏感性数据	每季度或半年更新(DepMap Portal)	筛选对特定基因敲除(CRISPR screen)敏感的癌细胞系,用于合成致死研究 ^[53]

表1 数字生命代表性基础数据库概览(续表)
Table 1 Overview of representative digital life basic databases(continued)

数据库 (中文名)	数据库	核心功能	所属领域	数据规模	更新频率	主要应用案例
人类细胞图谱数据库	Human Cell Atlas	构建单细胞分辨率的人类细胞图谱,解析细胞类型、状态及组织构成	单细胞组学与人体生物学	包含来自数千名供体的数千万个单细胞转录组数据	持续更新,随项目进展发布数据集	重新定义肺部等器官的细胞类型图谱,发现导致肺纤维化的罕见细胞亚群 ^[54]
人类表型组计划数据库	Human Phenome Project	整合人类表型数据与基因型数据,解析表型-基因型关联及疾病表型特征	人类表型组学与精准医学	定义了超过15 000个标准化表型术语(HPO)及其关联疾病	定期更新(通常每两月更新一次)	辅助罕见病诊断,通过表型特征匹配筛选致病基因变异 ^[55]

相似细胞的SCimilarity机器学习工具^[82]表明数据和工具条件日益成熟。

另外,欧盟启动了“生命时间”(LifeTime)计划,旨在通过整合多组学技术与前沿计算模型,实现对疾病发生发展过程的“数字孪生”级模拟,构建高分辨率的细胞图谱,实现对细胞状态随时间变化的追踪(trajec-tory tracing)。美国艾伦研究所也启动为期十年的“细胞景观”(CellScapes)计划,旨在揭示细胞内部复杂的物理化学原理,构建出能够预测细胞行为的动态图景^[83]。2025年4月,北京大学国家生物医学成像科学中心(NBIC)依托多模态跨尺度生物医学成像设施发起了“数字生命”大科学计划,标志着我国科学家在这一前沿领域迎接科学和技术双重挑战的决心。我国其他围绕数字生命启动的大科学计划或设施可参考表2。

2.2 生命特征解码的研究发展

当海量的组学和成像数据被采集后,核心挑战在于从异构数据中提取有生物学意义的特征信息,而现代AI技术特别是深度学习和大语言模型(LLM),正在从根本上重塑我们理解和操作组学数据的方式。

在蛋白质组学领域,生成式人工智能的介入带来了突破性进展,其中AlphaFold系列模型及其衍生工具已重塑了结构生物学的研究范式。相较于传统X射线晶体学或冷冻电镜高昂的时间与经济成本,基于深度学习的算法现已实现仅凭基因组序列即能在原子分辨率下还原蛋白质的三维折叠结构^[84]。除静态结构解析外,基于Transformer架构的“蛋白质语言模型”(protein language models, PLMs)亦成为前沿热点。以Meta AI发布的ESM-2模型为例,该模型通过在数亿级蛋白质序列上的自监督学习,解析了氨基酸排列隐含的进化规则。其优势在于无需依赖多序列比对(MSA),即可直接推断功能性突变效

应,甚至从头设计天然不存在的人造蛋白^[2]。这种将生物大分子序列视为一种自然语言的处理逻辑,极大拓展了对生命编码规则的认知维度。这种基础模型的迭代迅速转化为产业效能。美国医药企业利用AlphaFold技术重构研发管线,确立了数字生命解码的规模化应用范例。例如,谷歌旗下Isomorphic Labs依托AlphaFold 3平台构建了“分子生成-交互预测-动态优化”的闭环系统。该系统引入扩散模型架构,可在 10^{60} 数量级的化学空间内自主设计新药分子,并精准模拟药物与蛋白质(含天然无序蛋白)、核酸及其修饰离子间的动态复合物结构。通过快速预测结合亲和力及脱靶毒性,该平台将先导化合物的发现与验证周期从数月压缩至秒级,大幅降低了早期研发的试错成本,实现了计算模拟与实验验证的深度耦合。

随着单细胞测序与空间转录组技术的普及,单一组学数据的解释力日益受限,推动了能够兼容多种“组学语言”的多模态模型发展。在组学整合层面,深度学习模型正逐步取代传统的线性差异表达分析。例如,基于Transformer架构的DeePathNet模型通过引入通路感知(pathway-aware)的注意力机制,将基因组突变、转录组表达、蛋白质组丰度及代谢组图谱进行统一建模,动态捕捉跨层级特征在生物通路中的非线性相互作用^[85]。这种通过挖掘层级间协同作用的方法,有效克服了单一组学数据信噪比低的问题,为解析癌症预后及治疗反应提供了更为系统的视角。

在疾病分子机制的精细解码上,图神经网络(GNNs)凭借其处理非欧几里得拓扑数据的优势脱颖而出。以MOGONET模型为例,该算法利用图卷积网络(GCN)构建样本相似性网络,通过视图相关性发现算法(VCDN)融合跨组学标签分布,成功构建了反映分子间互作关系的图谱^[86]。在阿尔茨海

表2 数字生命相关国内外大科学计划/设施启动情况表
Table 2 Overview of the launch of major international and domestic large-scale scientific projects/infrastructure related to digital life

计划名称	发起方/主导方	核心目标与内容
人类基因组计划(Human Genome Project)	国际合作	测定人类基因组全序列,奠定生命组学研究基础
欧盟虚拟生理人计划(Virtual Physiological Human, VPH)	欧盟	构建跨尺度虚拟生理人模型,支撑数字孪生器官研发
达索“活心脏计划”(Living Heart Project)	达索系统(与美国 FDA 合作)	构建功能完整的计算机心脏模型,模拟血流动力学、电信号传导等生理机制
欧盟“生命时间”计划(LifeTime)	欧盟	整合多组学技术与计算模型,实现疾病发生发展的“数字孪生”模拟,构建高分辨率细胞图谱
美国艾伦研究所“细胞景观”计划(CellScapes)	美国艾伦研究所	揭示细胞内部物理化学原理,构建可预测细胞行为的动态图景
人类细胞图谱计划(Human Cell Atlas, HCA)	国际合作	构建单细胞分辨率的人类细胞图谱,解析细胞类型、状态及组织构成
国际人类蛋白质组计划(International Human Proteome Project)	国际合作	全面解析人类蛋白质的表达、定位及功能
国际人类表型组计划(International Human Phenome Consortium)	复旦大学主导的国际合作	整合人类表型与基因型数据,解析表型-基因型关联及疾病特征
人类肿瘤图谱网络计划(Human Tumor Atlas Network, HTAN)	国际合作	构建人类肿瘤多尺度图谱,解析肿瘤发生发展机制
中国脑多组学图谱计划(CBMAP)	中国国内团队	构建中国人群脑多组学整合图谱,解析脑功能与疾病关联
“数字肾脏”大科学计划	北京大学第一医院、北京大学国家生物医学成像中心	开展肾脏多组学研究,构建肾脏单细胞基础模型
“数字生命”大科学计划	北京大学国家生物医学成像中心(NBIC)	依托多模态跨尺度成像设施,推进数字生命多维度整合研究
多模态跨尺度生物医学成像设施(National Biomedical Imaging Center, NBIC)	中国国家重大科技基础设施	实现组织-细胞-亚细胞层级的结构、动态与定量测量
人类细胞谱系大科学研究设施(Cell Lineage Analysis Facility, CLAF)	中国国家重大科技基础设施	构建细胞谱系追踪技术体系,解析细胞发育与分化轨迹
人类器官生理病理模拟装置(Human Organ Physiology and Pathology Simulation Facility, HOPE)	中国国家重大科技基础设施	模拟器官生理功能与病理状态,支撑数字器官研发

默病及肿瘤分型研究中,该模型不仅实现了亚型的精准分类,更反向定位了关键调控节点,为临床制定基于分子分型的个体化治疗方案提供了依据。

此外,AI技术在空间数据解码方面展示了独特价值。针对空间转录组学面临的分辨率与成本制约,卷积神经网络(CNN)或生成对抗网络(GAN)等算法被应用于融合分析空间转录组与病理图像。这种“AI+空间组学”策略能够自动识别复杂细胞类型并高精度重建组织图谱^[69],显著提升了对组织微环境异质性的解析能力。

展望未来,信息解码技术有望被广泛应用于通过网络药理学预测潜在靶点的药物靶点发现、疾病早筛和生物学机制研究等领域。随着数字解码能力的成熟,技术应用正在倒逼监管创新。在此基础上,美国食品药品监督管理局(FDA)出台新规支持AI辅助下的类器官或计算机建模系统替代动物实验^[87],并新设AI审评通道。这种技术应用与监管创新的协同机制,意味着基于多组学数据的AI预测结果(如心脏毒性模拟或药代动力学参数等)有望直接作为新药临床试验申请(IND)的关键证据,推动数字生命解码技术从实验室辅助工具向临床转化核心驱动力的角色跨越。

2.3 生命活动模拟的研究发展

生命活动模拟是数字生命从微观向宏观跨越的关键节点。在这一层级,研究者不再局限于单个细胞的行为,而是关注细胞间的相互作用、组织结构形成与功能维持。早期的组织建模主要采用粗粒度方法,如有限元法(finite element method)或离散元法(discrete element method),用于模拟组织力学行为或血流动力学^[88]。然而,这些模型往往缺乏对细胞异质性和分子机制的深入刻画。随着系统论思想的深化,现代建模开始强调“系统性”与“预测性”。在此背景下,人工智能虚拟细胞(AIVC)成为了新一代建模的核心载体。AIVC是指利用人工智能技术,基于海量生物组学数据构建的数字化细胞模型。它通过整合多尺度数据,构建能够有效预测组织发育、疾病进展或治疗响应的动态系统。

目前生命活动模拟在细胞层面的研究正处于通过“细胞基础模型”重塑生物学范式的关键阶段,其核心进展在于利用深度学习架构(特别是Transformer)对海量单细胞数据进行语言建模学习。在该领域的开创性工作中,博德研究所提出的Geneformer通过

在约3 000万个单细胞转录组数据上进行掩码学习,证明了预训练模型在预测基因剂量敏感性及相关性方面的有效性^[89];与此同时,多伦多大学与微软研究院合作开发的scGPT进一步整合了生成式预训练架构,引入了基因表达值和批次校正机制,使其成为进行细胞类型注释、多组学整合及扰动预测等多种下游任务的通用基准^[90]。腾讯AI Lab开发的scBERT较早地探索了BERT架构在处理大规模稀疏单细胞数据中的应用,显著提升了细胞类型注释的精度^[91]。为了突破物种界限,斯坦福大学与Chan Zuckerberg Biohub推出的UCE(universal cell embedding)构建了一个跨物种的通用细胞嵌入空间,旨在捕捉生命演化中的保守特征^[92]。另外,虽然AlphaFold 3主要聚焦于生物分子结构预测,但其对蛋白质-DNA/RNA相互作用的高精度模拟能力,为构建包含物理与生化约束的完整“虚拟细胞”系统提供了至关重要的底层支撑^[93]。这些研究共同标志着生物计算正从单纯的数据观测迈向可预测、可模拟的“生命模拟”新阶段。

另外,在组织和器官层面的组织建模如类器官模拟技术也正蓬勃发展。例如,基于有限元分析(finite element analysis,FEA)的模型能够模拟心脏组织的电生理活动^[94],预测心律失常的发生机制。此外,数字模型还可以模拟药物在器官芯片上的代谢过程,为体外实验提供虚拟验证平台。随着模型模拟发展演进,整体技术方向正由小范围、低精度向大规模、高分辨率方向发展。例如,类器官技术的发展使得在体外培养具有三维结构和功能特性的微型器官成为可能,而数字孪生技术则进一步将类器官与计算模型结合,实现对器官发育与疾病的实时监测与干预^[95]。此外,空间预测与生成(spatial predictive & generative)模型正在兴起,其目标是根据输入的基因表达或环境条件,预测组织的空间结构与功能分布。这类模型通常基于深度学习架构,如卷积神经网络(CNN)或图神经网络(GNN),能够从高维空间数据中提取拓扑特征并进行生成式预测^[96]。

2.4 生命数字孪生研究发展

在数字生命的研究体系中,生命活动模拟主要由人工智能虚拟细胞(AIVC)等技术承载,专注于对生物过程本身进行可计算的再现,旨在探索和还原生命运作的底层机理。相比之下,生命数字孪生则代表了更高级别的功能跃迁。它在模拟的基础上,进一步引入了持续感知、状态实时更新与干预反馈

机制,使数字模型与实体生命体形成长期的动态耦合。可以说,生命数字孪生是数字生命研发的终极形态,其核心在于通过精准的虚实映射建模,实现对个体健康状态的精准预测、干预与优化。研究者通过医学影像(如MRI、CT)和组学数据重建实体器官的三维几何模型^[97],并将其与体内的生理参数(如血流、心电)进行关联。在此基础上,通过实时感知技术(如可穿戴式传感器或连续影像数据流)不断更新模型参数^[98],实现对实体生命状态的实时映射。

“人工智能虚拟类器官”(AIVOs)^[99]是涵盖“数据-模型-交互”三层架构的复杂数字系统,有望为每位患者提供一份可计算、可预测、可交互的“生命蓝图”。在精准医疗领域,数字孪生模型已被用于指导疾病治疗方案的选择和医疗科研领域。例如由复旦大学冯建峰教授团队研发的国际首个基于数据同化方法构建的数字孪生脑平台^[100]。该平台具备860亿神经元规模及百万亿突触连接,通过整合多模态神经影像数据建立动态拓扑基底模型^[101],可实现静息态与任务态脑活动的精准模拟。研究证实当模型神经元数量与突触结构接近真实人脑时,能复现临界现象与认知功能特征。在这些发展中,数字孪生模型与合成生物学的融合使得设计和构建人工生命体在湿实验层面得以验证,反过来又促进数字生命模型的迭代与优化^[102]。

随着生命模拟研究的不断深入,对人工智能虚拟细胞(AIVC)的研究已从单纯的形态模拟迈向了功能与机制的深层重构。传统的二维细胞培养往往难以反映体内复杂的微环境与异质性,而新兴的计算框架正在弥补这一鸿沟。例如,Toscano等^[103]开发的SimulCell平台展示了如何利用多智能体仿真技术,在硅基环境中精确复现实验性细胞培养群体的行为特征,包括细胞迁移、增殖动力学及对药物刺激的响应模式。这种高保真的数字孪生能力使得研究人员能够在虚拟空间中进行大规模参数扫描^[104],从而优化CAR-T细胞疗法的设计规格,解决工程化改造细胞在体内持久性和安全性方面的挑战。此外,针对精准医学的需求,数字器官模型(digital organoids)结合单细胞测序与AI算法^[105]能够有效表征肿瘤的遗传背景与分子机制,帮助筛选针对特定患者的最佳治疗方案。

达索的“活心脏”(Living Heart)计划通过与美国食品药品监督管理局(FDA)的深度合作(表2),构

建了一套基于虚拟临床试验的验证分层方法,为医疗器械审批提供了全新的数字证据来源。双方共同开发了首个功能完整的计算机心脏模型^[106],该模型不仅能够模拟血流动力学、机械运动和电信号传导等复杂生理机制^[107],还被用于评估药物对心脏组织的潜在影响,从而显著降低了新药研发中的临床试验风险与成本^[108]。北京大学围绕线虫模型在数字孪生领域有着典型的研发应用案例。2024年,北京大学未来技术学院马雷团队在前期工作的基础上,深入研究了线虫完整的神经系统连接图谱(Connectome)^[109],进一步提出BAAIWorm模型,通过数据驱动的高精度神经系统、身体与环境闭环耦合,实现了由生物物理真实神经网络直接驱动在线虫类真实运动,为系统性研究“脑-体-环境”交互机制及具身智能提供了可验证的计算生命平台^[110]。清华大学研发了一套果蝇活体动态成像配准算法^[111],专门攻克果蝇幼虫体节蠕动、成虫翅振行为导致的成像位移干扰问题。

3 数字生命的发展展望

数字生命领域正处于从概念验证迈向系统工程的关键节点。从冯·诺依曼的理论奠基,到人类细胞图谱的全面解析,再到人工智能虚拟细胞(AIVC)与数字器官(AIDO)等前沿概念的初步实践,学界正致力于将生命系统转化为可计算、可模拟的数字实体。以“生命数据采集-特征解码-活动模拟-数字孪生”为核心的技术发展路径,“干湿结合、跨域泛化”新范式正在形成。然而,从概念验证迈向具备稳健性、可靠性且能广泛应用于基础研究与临床实践的系统工程,仍面临诸多亟待突破的深层次挑战。这些挑战不仅源于技术瓶颈,更归因于生命系统内在的复杂性以及当前认知与建模能力的局限性。

首要挑战在于生命数据的固有属性:多源、异构、跨尺度且动态变化。构建真实数字孪生需整合基因组序列、单细胞转录组、蛋白质互作网络、器官影像及个体行为谱等多源异构数据。尽管当前生命数据采集技术在通量与分辨率方面取得显著进展,但数据标准化、标注一致性与平台互操作性仍存在明显短板。更为关键的难题在于如何将静态或短时序的“数据层”转化为驱动模型演化的“知识层”。这要求生命特征解码技术不仅能够识别多源异构生命数据的各类模式,更要理解数据背后所蕴含的从基

因到个体的空间层级结构关系,以及这些层级间物质、能量与信息流动的约束条件。缺乏对生物背景的深度嵌入,模型易陷入困境或产生统计合理但生物学不可解释的预测。

其次,模型可解释性与验证体系构成重大阻碍。基于深度学习的复杂模型虽在特征解码与活动模拟中展现出强大能力,但其决策过程难以追溯。当AIVC预测药物导致细胞凋亡时,科学家需明确被触发的通路及关键蛋白,而非仅仅获得高精度预测结果。这涉及将算法“特征空间”映射为可实验验证的生物学假设。跨层级分析维度,如功能涌现、反馈调控与信息整合,是理解复杂相互作用的关键。然而,验证跨尺度模型正确性极具挑战性。湿实验验证通常仅针对特定层级或时间点(如检测蛋白表达),而模型预测可能涉及跨越细胞周期乃至生理节律的长期系统行为。建立贯穿“干平台智能预测-湿平台高效验证-多尺度模型迭代”全流程的标准化基准测试与压力测试体系,是确保数字生命研究科学严谨性的必要举措。

第三,对生命系统动态因果性的理解仍存在不足。现有模型多擅长关联分析,难以区分共相关与因果关系。生命过程充满非线性时间积分、延迟反馈和状态依赖特性。例如,早期基因表达信号可能通过一系列中间事件,在数天后影响组织再生能力。构建真正具有预测力的数字孪生,需对跨时序因果链进行推理与模拟。这不仅需要更复杂的时间序列建模技术,更依赖于对生命系统时间维度(从分子振荡到生命周期)的节律嵌套与耦合的深刻理论认知。当前,融合因果推断与深度学习的新兴交叉学科,或为挖掘更具鲁棒性的因果机制提供新的工具。

最后,协同创新生态与治理框架构建是易被忽视的重要挑战。数字生命本质上是汇聚生物学、计算科学、工程学、物理学等多学科的“会聚”研究。然而,不同学科存在语言体系、研究文化与评价标准的差异。打破壁垒,构建标准化、可互操作的数据平台与模型接口,促进开源协作,是实现要素创新与跨域泛化目标的社会技术基础。同时,当生命体可被数字化、模拟与编程时,伦理、安全与监管问题日益凸显。数字孪生体涉及的个人隐私数据如何保护?基于虚拟模型的医疗决策责任如何界定?“数字基因”或“虚拟细胞”知识产权归属的界定问题,亟需提前谋划,构建涵盖知识产权、标准规范、伦理审查、安全

监管等多维度的综合治理体系,确保技术发展始终坚持以人为本,符合全人类共同的道德伦理标准。

展望未来,数字生命的发展将遵循“分治-连接-统合”的递进式演进路径。在近期发展阶段,研究重点将聚焦特定领域,如提升AIVC技术精准度,开发特定器官的AIDO模型,重点突破数据质量、模型优化与验证体系等关键难题。伴随各子领域模型的完善以及标准化接口的构建,层级耦合将成为核心工作内容。通过反馈逻辑、映射逻辑等基础机制,实现细胞模型与组织模型、器官模型与全身系统模型的有效衔接,逐步构建更为完善的数字孪生体系。从长远视角来看,未来数字生命的发展将步入“统合”发展阶段,即依托统一的智能计算架构,嵌入生物物理约束条件,实现空间维度、时间维度、跨层级与跨时序分析的深度融合,最终建成具备自主学习与进化能力的数字生命研究系统。

数字生命研究的根本目的并非替代或复制生物生命,而是为深入剖析生命复杂性、精准预测生命动态、负责任地干预生命过程提供全新的工具与平台。跨越从“描述”到“设计”的技术障碍任务艰巨,这不仅需要计算能力实现跨越式提升,更依赖于对生命本质的深刻认识、跨学科整合的系统性创新,以及全球范围内的协同治理。当这些条件渐趋成熟,数字生命必将发展成为揭示生命奥秘、保障人类健康的重要科学基础设施。

参考文献

- [1] Neumann J. The general and logical theory of automata [M]//Cerebral Mechanisms in Behavior-The Hixon Symposium. Jeffress LA, Ed. New York, NY, USA: John Wiley & Sons, 1951: 1-31.
- [2] Lin Z, Akin H, Rao R, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 2023, 379: 1123-30.
- [3] Song L, Segal E, Xing E. Toward AI-driven digital organism: a system of multiscale foundation models for predicting, simulating and programming biology at all levels. *arXiv*, 2024, doi: 10.48550/arXiv.2412.06993.
- [4] Kutschera U. Antonie van Leeuwenhoek (1632-1723): master of fleas and father of microbiology. *Microorganisms*, 2023, 11: 1994.
- [5] Dahm R. Friedrich Miescher and the discovery of DNA. *Dev Biol*, 2005, 278: 274-88.
- [6] Klug A. The discovery of the DNA double helix. *J Mol*

- Biol, 2004, 335: 3–26.
- [7] Cobb M. 60 years ago, Francis Crick changed the logic of biology. PLoS Biol, 2017, 15: e2003243.
- [8] Xue Y, Wang Y, Shen H. Ray Wu, fifth business or father of DNA sequencing? Protein Cell, 2016, 7: 467–70
- [9] Wikipedia. Sanger sequencing[EB/OL]. [2026-01-16]. https://en.wikipedia.org/wiki/Sanger_sequencing.
- [10] National Human Genome Research Institute. The Human Genome Project[EB/OL]. [2026-01-16]. <https://www.genome.gov/human-genome-project>.
- [11] Tang F, Barbacioru C, Wang Y, et al. mRNA-seq whole-transcriptome analysis of a single cell. Nat Methods, 2009, 6: 377–82.
- [12] Ståhl PL, Salmén F, Vickovic S, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science, 2016, 353: 78–82.
- [13] Wishart DS, Tzur D, Knox C, et al. HMDB: the human metabolome database. Nucleic Acids Res, 2007, 35: D521–26.
- [14] Cortals GL. Characteristics and future prospective of human proteome project. J Proteomics Bioinform, 2022, 15: 615.
- [15] Human Cell Atlas. About the Human Cell Atlas[EB/OL]. [2026-01-16]. <https://www.humancellatlas.org/learn-more/about-the-human-cell-atlas/#event-launch-of-the-human-cell-atlas>.
- [16] Secretariat of International Human Phenome Consortium. Report on the 3rd board meeting of the international human phenome consortium. Phenomics, 2022, 3: 77–82.
- [17] HTAN. Overview[EB/OL]. [2026-01-16]. <https://humantumoratlas.org/overview>.
- [18] 北京大学第一医院. 北大医院肾脏内科联合北京大学国家生物医学成像中心发起“数字肾脏”大科学计划[EB/OL]. (2025-03-14)[2026-01-16]. <https://www.pkufh.com/Html/News/Articles/60461.html>.
- [19] 科学网. CBMAP:中国脑多组学图谱计划[EB/OL]. (2025-10-15)[2026-01-14]. <https://wap.sciencenet.cn/home.php?mod=space&uid=571917&do=blog&id=1506005>.
- [20] Wimmer E. The test-tube synthesis of a chemical called poliovirus: the simple synthesis of a virus has far-reaching societal implications. EMBO Rep, 2006, Spec No: S3–S9.
- [21] Gibson DG, Glass JI, Lartigue C, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. Science, 2010, 329: 52–56.
- [22] Sliva A, Yang H, Boeke JD, et al. Freedom and responsibility in synthetic genomics: the synthetic yeast project. Genetics, 2015, 200: 1021–28.
- [23] Shao Y, Lu N, Wu Z, et al. Creating a functional single-chromosome yeast. Nature, 2018, 560: 331–35.
- [24] Senior AW, Evans R, Jumper J, et al. Improved protein structure prediction using potentials from deep learning. Nature, 2020, 577: 706–10.
- [25] Chen A, Liao S, Cheng MN, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. Cell, 2022, 185: 1777–92.e21.
- [26] Han L, Wei X, Liu C, et al. Cell transcriptomic atlas of the non-human primate *Macaca fascicularis*. Nature, 2022, 604: 723–31.
- [27] Chen H, Zhang Y, Zhou H, et al. Routine workflow of spatial proteomics on micro-formalin-fixed paraffin-embedded tissues. Anal Chem, 2023, 95: 16733–43.
- [28] 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.
- [29] Zhang ZH, Schaefer C, Jiang WR, et al. A panoramic view of cell population dynamics in mammalian aging. Science, 2025, 387: eadn3949.
- [30] Liu Y, Zhao S, Yun Z, et al. A mouse brain atlas based on dendritic microenvironments. Nat Neurosci, 2026, 29: 111–22.
- [31] Yin J, Zheng Y, Huang Z, et al. Chinese immune multi-omics atlas. Science, 2026, 391: eadt3130.
- [32] Wang M, Hu Q, Tu Z, et al. A *Drosophila* single-cell 3D spatiotemporal multi-omics atlas unveils panoramic key regulators of cell-type differentiation. Cell, 2025, 188: 4734–53.e31.
- [33] Nguyen E, Poli M, Durrant MG, et al. Sequence modeling and design from molecular to genome scale with Evo. Science, 2024, 386: eado9336.
- [34] Dassault Systèmes. Virtual Twin as a Service: How We Got Here[EB/OL]. [2026-01-16]. <https://www.3ds.com/consulting-services-value-engagement/virtual-twin-service/history-and-evolution>.
- [35] Sahal R, Alsamhi S, Brown KN. Personal digital twin: a close look into the present and a step towards the future of personalised healthcare industry. Sensors, 2022, 22: 5918.
- [36] 智源研究院. 最高精度智能线虫“天宝1.0”[EB/OL]. (2022-06-20)[2026-01-16]. <https://mp.weixin.qq.com/s/R8eCz6CF3jN8oe5fd9B6g>.
- [37] Aldarondo D, Merel J, Marshall JD, et al. A virtual rodent predicts the structure of neural activity across behaviours. Nature, 2024, 632: 594–602.
- [38] Vaxenburg R, Siwanowicz I, Merel J, et al. Whole-body physics simulation of fruit fly locomotion.

- Nature, 2025, 643: 1312–20.
- [39] Bunne C, Roohani Y, Rosen Y, et al. How to build the virtual cell with artificial intelligence: priorities and opportunities. *Cell*, 2024, 187: 7045–63.
- [40] Li C, Ziyadeh E, Sharma Y, et al. Nephrobase Cell+: multimodal single-cell foundation model for decoding kidney biology. *bioRxiv*, 2025, doi: 10.1101/2025.09.30.679471.
- [41] Sayers EW, Cavanaugh M, Clark K, et al. GenBank. *Nucleic Acids Res*, 2022, 50: 161–4.
- [42] O’Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database. *Nucleic Acids Res*, 2016, 44: 733–45.
- [43] Chen IA, Markowitz VM, Chu K, et al. IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res*, 2019, 45: 507–16.
- [44] The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*, 2023, 46: 2699.
- [45] Berman HM, Westbrook J, Feng Z, et al. The protein data bank. *Nucleic Acids Res*, 2000, 28: 235–42.
- [46] Varadi M, Bertoni D, Magana P, et al. AlphaFold protein structure database. *Nucleic Acids Res*, 2022, 52: 368–75.
- [47] Szklarczyk D, Kirsch R, Koutrouli M, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res*, 2023, 51: 638–46.
- [48] Kanehisa M, Furumichi M, Sato Y, et al. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res*, 2021, 49: 545–51.
- [49] Gillespie M, Jassal B, Stephan R, et al. The reactome pathway knowledgebase 2022. *Nucleic Acids Res*, 2022, 50: 687–92.
- [50] Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data. *Nucleic Acids Res*, 2013, 41: 991–5.
- [51] The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 2012, 489: 57–74.
- [52] Uhlén M, Fagerberg L, Hallström BM, et al. Proteomics. Tissue-based map of the human proteome. *Science*, 2015, 347: 1260419.
- [53] Barretina J, Caponigro G, Stransky N, et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 2012, 483: 603–7.
- [54] Regev A, Teichmann SA, Lander ES, et al. The human ccell atlas. *eLife*, 2017, 6: e27041.
- [55] Köhler S, Gargano M, Matentzoglou N, et al. The human phenotype ontology in 2021. *Nucleic Acids Res*, 2021, 49: 1207–17.
- [56] Vitorino R. Transforming clinical research: the power of high-throughput omics integration. *Proteomes*, 2024, 12: 25.
- [57] Subramanian I, Verma S, Kumar S, et al. Multi-omics data integration, interpretation, and its application. *Bioinform Biol Insights*, 2020, 14: 1177932219899051.
- [58] Ivanisevic T, Sewduth RN. Multi-omics integration for the design of novel therapies and the identification of novel biomarkers. *Proteomes*, 2023, 11: 34.
- [59] Gao Z, Lu Y, Li M, et al. Application of pan-omics technologies in research on important economic traits for ruminants. *Int J Mol Sci*, 2024, 25: 9271.
- [60] Xiao H, Rosen A, Chhibbar P, et al. From bench to bedside via bytes: multi-omic immunoprofiling and integration using machine learning and network approaches. *Hum Vaccin Immunother*, 2023, 19: 2282803.
- [61] Lv T, Zhang Y, Liu J, et al. Multi-omics integration for both single-cell and spatially resolved data based on dual-path graph attention auto-encoder. *Brief Bioinform*, 2024, 25: bbae450.
- [62] Wang X, Wu X, Hong N, et al. Progress in single-cell multimodal sequencing and multi-omics data integration. *Biophys Rev*, 2023, 16: 13–28.
- [63] Kim IS. Single-cell molecular barcoding to decode multimodal information defining cell states. *Mol Cells*, 2023, 46: 74–85.
- [64] Saliba AE, Westermann AJ, Gorski SA, et al. Single-cell RNA-seq: Advances and future challenges. *Nucleic Acids Res*, 2014, 42: 8845–60.
- [65] Manzoni C, Kia DA, Vandrovцова J, et al. Genome, transcriptome and proteome: The rise of omics data and their integration in biomedical sciences. *Brief Bioinform*, 2018, 19: 286–302.
- [66] Tian T. Stochastic modelling of gene regulatory networks[M]//Applied statistics for network biology: methods in systems biology. Dehmer M, Emmert-Streib F, Garber A, eds. Wiley - VCH Verlag GmbH & Co. KgaA, 2011.
- [67] Ma L, Wang L, Khatib SA, et al. Single-cell atlas of tumor cell evolution in response to therapy in hepatocellular carcinoma and intrahepatic cholangiocarcinoma. *J Hepatol*, 2021, 75: 1397–408.
- [68] Robles-Remacho A, Sanchez-Martin RM, Diaz-Mochon JJ. Spatial transcriptomics: Emerging tech-

- nologies in tissue gene expression profiling. *Anal Chem*, 2023, 95: 15450–60.
- [69] Jin Y, Zuo Y, Li G, et al. Advances in spatial transcriptomics and its applications in cancer research. *Mol Cancer*, 2024, 23: 129.
- [70] Zhang X, Cao Q, Rajachandran S, et al. Dissecting mammalian reproduction with spatial transcriptomics. *Hum Reprod Update*, 2023, 29: 794–810.
- [71] Vickovic S, Eraslan G, Salmén F, et al. High-definition spatial transcriptomics for *in situ* tissue profiling. *Nat Methods*, 2019, 16: 987–90.
- [72] Cao J, Spielmann M, Qiu X, et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, 2019, 566: 496–502.
- [73] Tapia JJ, Saglam AS, Czech J, et al. MCell-R: A particle-resolution network-free spatial modeling framework. *Methods Mol Biol*, 2019, 1945: 203–29.
- [74] Henriques R, Leterrier C, Weigel A. Decoding life's inner workings: Advances in quantitative bioimaging. *Open Biol*, 2023, 13: 230329.
- [75] Cheng Y, Li W, Jin T, et al. Frontiers and development in live-cell super-resolution fluorescence microscopy. *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi*, 2023, 40: 180–4.
- [76] Lang K, Fang J, Su G. NeuroFusionNet: Cross-modal modeling from brain activity to visual understanding. *Front Comput Neurosci*, 2025, 19: 1545971.
- [77] Lohmann P, Schäfer L, Krause S, et al. Advancements in noninvasive visualization of the immune environment in glioblastoma: A systematic review. *Neurooncol Adv*, 2025, 7: vdaf176.
- [78] Peng CK, Cui X, Zhang Z, et al. Wearable devices: Perspectives on assessing and monitoring human physiological status. *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi*, 2023, 40: 1045–52.
- [79] Langholm C, Kowatsch T, Bucci S, et al. Exploring the potential of Apple SensorKit and digital phenotyping data as new digital biomarkers for mental health research. *Digit Biomark*, 2023, 7: 104–14.
- [80] Rozenblatt-Rosen O, Stubbington MJT, Regev A, et al. The human cell atlas: From vision to reality. *Nature*, 2017, 550: 451–53.
- [81] To K, Fei L, Pett JP, et al. A multi-omic atlas of human embryonic skeletal development. *Nature*, 2024, 635: 657–67.
- [82] He Z, Dony L, Fleck JS, et al. An integrated transcriptomic cell atlas of human neural organoids. *Nature*, 2024, 635: 690–98.
- [83] Harris JA, Mihalas S, Hirokawa KE, et al. Hierarchical organization of cortical and thalamic connectivity. *Nature*, 2019, 575: 195–202.
- [84] Srivastava R. Advancing precision oncology with AI-powered genomic analysis. *Front Pharmacol*, 2025, 16: 1591696.
- [85] Cai Z, Poulos RC, Aref A, et al. DeePathNet: A transformer-based deep learning model integrating multiomic data with cancer pathways. *Cancer Res Commun*, 2024, 4: 3151–64.
- [86] Wang T, Shao W, Huang Z, et al. MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. *Nat Commun*, 2021, 12: 3445.
- [87] FDA. Roadmap to Reducing Animal Testing in Preclinical Safety Studies[EB/OL]. <https://www.fda.gov/media/186092/download>.
- [88] Post JN, Loerakker S, Merks RMH, et al. Implementing computational modeling in tissue engineering: Where disciplines meet. *Tissue Eng Part A*, 2022, 28: 542–54.
- [89] Theodoris CV, Xiao L, Chopra A, et al. Transfer learning enables predictions in network biology. *Nature*, 2023, 618: 616–24.
- [90] Cui H, Wang C, Maan H, et al. scGPT: Toward building a foundation model for single-cell multi-omics using generative AI. *Nat Methods*, 2024, 21: 1470–80.
- [91] Yang F, Wang W, Wang F, et al. scBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data. *Nat Mach Intell*, 2022, 4: 852–66.
- [92] Rosen Y, Roohani Y, Agarwal A, et al. Universal cell embeddings: A foundation model for cell biology. *bioRxiv*, 2023, doi:10.1101/2023.11.28.568918.
- [93] Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 2024, 630: 493–500.
- [94] Camps J, Berg LA, Wang ZJ, et al. Digital twinning of the human ventricular activation sequence to clinical 12-lead ECGs and magnetic resonance imaging using realistic Purkinje networks for *in silico* clinical trials. *Med Image Anal*, 2024, 94: 103108.
- [95] Lancaster MA, Knoblich JA. Organogenesis in a dish: Modeling development and disease using organoid technologies. *Science*, 2014, 345: 1247125.
- [96] Kopczweska K. Spatial machine learning: New opportunities for regional science. *Ann Reg Sci*, 2021, 68: 713–55.
- [97] Montgomery AJ, Litell J, Dang J, et al. Gaining consensus on expert rule statements for acute respiratory

- failure digital twin patient model in intensive care unit using a Delphi method. *Biomol Biomed*, 2023, 23: 1108–17.
- [98] Penverne Y, Martinez C, Cellier N, et al. A simulation based digital twin approach to assessing the organization of response to emergency calls. *NPJ Digit Med*, 2024, 7: 385.
- [99] Bai L, Su J. Artificial intelligence virtual organoids (AIVOs). *Bioact Mater*, 2025, 59: 45–68.
- [100] Lu W, Zeng L, Wang J, et al. Imitating and exploring the human brain's resting and task-performing states via brain computing: Scaling and architecture. *Natl Sci Rev*, 2024, 11: nwae080.
- [101] Lu W, Du X, Wang J, et al. Simulation and assimilation of the digital human brain. *Nat Comput Sci*, 2024, 4: 890–8.
- [102] Endy D. Foundation for engineering biology. *Nature*, 2005, 438: 449–53.
- [103] Toscano E, Cimmino E, Boccia A, et al. Cell populations simulated *in silico* within SimulCell accurately reproduce the behaviour of experimental cell cultures. *NPJ Syst Biol Appl*, 2025, 11: 48.
- [104] Weirauch U, Kreuz M, Birkenbihl C, et al. Design specifications for biomedical virtual twins in engineered adoptive cellular immunotherapies. *NPJ Digit Med*, 2025, 8: 493.
- [105] Yang Q, Li M, Xiao Z, et al. A new perspective on precision medicine: The power of digital organoids. *Biomater Res*, 2025, 29: 0171.
- [106] Armeni P, Polat I, De Rossi LM, et al. Digital twins in healthcare: Is it the beginning of a new era of evidence-based medicine? A critical review. *J Pers Med*, 2022, 12: 1255
- [107] De Benedictis A, Mazzocca N, Somma A, et al. Digital twins in healthcare: An architectural proposal and its application in a social distancing case study. *IEEE J Biomed Health Inform*, 2023, 27: 5143–54.
- [108] Sun T, He X, Song X, et al. The digital twin in medicine: A key to the future of healthcare? *Front Med (Lausanne)*, 2022, 9: 907066
- [109] Wang B. Recent advances in the study of spermatogenesis and fertilization in *Caenorhabditis elegans*. *Yi Chuan*, 2008, 30: 677–86.
- [110] Zhao M, Wang N, Jiang X, et al. An integrative data-driven model simulating *C. elegans* brain, body and environment interactions. *Nat Comput Sci*, 2024, 4: 978–90.
- [111] Fan J, Wang Y, Li L, et al. Prominent involvement of acetylcholine dynamics in stable olfactory representation across the *Drosophila* brain. *Nat Commun*, 2025, 16: 8638.