DOI: 10.13376/j.cbls/2025152
文章编号：1004-0374(2025)12-1549-14

冯桂海，博士，中国科学院动物研究所研究员。目前研究方向为跨物种的知识迁移模型构建。参与发表文章 70 余篇，被引 5 000 余次，以共同第一作者及通讯作者在 *Nature*、*Cell*、*Advanced Science*、*Trends in Biotechnology* 等杂志发表文章 20 余篇，作为主要作者的工作曾两次获评“中国生命科学十大进展”；入选中国科学院青促会优秀会员；作为课题负责人参与国家重点研发计划、中国科学院 A 类先导专项、基金委等项目；获 2020 年中国科学院杰出科技成就奖 (主要完成者) 及 2019 年度“全国妇幼健康科学技术奖”自然科学奖一等奖。

细胞基础模型的领域进展

马英克[1,2]，王思骐[3,4]，李　鑫[3,4]，冯桂海[3,4]*
(1 国家生物信息中心，北京 100101；2 中国科学院北京基因组研究所，北京 100101；
3 中国科学院动物研究所，北京 100101；4 北京干细胞与再生医学研究院，北京 100101)

摘　要：基于细胞基础模型表征能力对细胞状态进行预测和模拟，已成为当前生命科学领域的一种新型研究范式。该方法借助人工智能与大规模组学数据，旨在提取细胞相关特征并解码其状态调控网络，但在模型设计与构建方面其仍面临诸多挑战。本文系统梳理了细胞基础模型领域的关键进展，根据建模所使用数据类型，从单细胞组学、图像组学以及跨模态模型等方面，综述了各类具有代表性的模型及其建模思路。随后，本文概述了细胞基础模型在发展过程中于数据集构建、模型体系设计以及模型可解释性等方面面临的主要挑战。最后，从多个维度对未来发展方向进行了展望，期望推动跨物种、跨模态、跨尺度的细胞基础模型的构建，从而支撑生物基础研究与相关产业的创新发展。

关键词：细胞基础模型；单细胞组学；细胞表征；扰动预测；多模态融合；深度学习
中图分类号：Q811.4；TP18　　**文献标志码：**A

Advances in the field of cellular foundation models

MA Ying-Ke[1,2], WANG Si-Qi[3,4], LI Xin[3,4], FENG Gui-Hai[3,4]*
(1 China National Center for Bioinformation, Beijing 100101, China; 2 Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China; 3 Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China; 4 Beijing Institute for Stem Cell and Regenerative Medicine, Beijing 100101, China)

Abstract: Prediction and simulation of cellular states based on the representational capacity of cellular foundation models has emerged as a new research paradigm in contemporary life sciences. This approach leverages artificial intelligence and large-scale omics datasets to extract cellular characteristics and decode regulatory network, but challenges remain in the design and implementation of foundation models. This review systematically summarizes

收稿日期：2025-11-25；**修回日期：**2025-12-22
基金项目：中国科学院前瞻战略科技先导专项(XDA0460305)
***通信作者：**E-mail: fenggh@ioz.ac.cn

key advances in the field of cellular foundation models. According to the data types used for model construction, we introduce representative models and modeling strategies based on single-cell omics, image-based omics, and cross-modal models. We then outline the major challenges in the development of cellular foundation models, including issues related to dataset construction, model architecture design, and model interpretability. Finally, we discuss future directions from multiple perspectives, with the goal of promoting the development of cellular foundation models that span species, modalities, and scales, thereby providing support for basic biological research and innovation in related industries.

Key words: cellular foundation models; single-cell omics; cell embedding; perturbation prediction; multimodal integration; deep learning

细胞是生命的基本单元，解析细胞的功能与作用机制，是理解复杂生命系统的关键环节。传统“湿实验”方法能够提供由点至线的信息，逐步揭示细胞特征及其调控规律。但是，由于细胞调控网络的高度复杂性，全面系统地解析细胞行为对实验研究的精度与规模提出了较高要求。除“湿实验”工作外，科学家也一直尝试通过计算方法构建“虚拟细胞(Virtual cell)”模型，用以模拟和预测细胞在特定条件下的生理状态，尤其是对扰动的响应。早期的细胞模型多基于理化方程或随机模拟，通过刻画少数基因之间的相互作用来阐释细胞功能机制[1, 2]。此类模型结构相对简单，难以精准刻画如哺乳动物等复杂细胞系统的动态行为。

近年来，人工智能技术持续突破，组学测序技术也快速发展，尤其是单细胞组学数据呈指数级增长，为细胞模型的构建提供了新的契机。二者的结合催生了基于细胞基础模型(Cellular foundation models)解释并预测细胞状态这一全新研究范式。细胞基础模型是指基于海量细胞尺度数据预训练的深度学习模型，旨在学习细胞的通用表征和运行规律。理解其设计框架，可以类比自然语言处理(Natural language processing, NLP)领域的大语言模型。大语言模型可以通过学习海量文本中“词元(Token)”和“句子(Sentence)”之间的语法规则和语义逻辑，最终实现对语言的理解和生成。与之相似，在细胞基础模型中，可以将细胞内的基因、蛋白质等生物大分子视为语言模型中的“词元”，将细胞在特定时刻的整体状态视为“句子”。模型可通过对海量细胞数据进行自监督学习，捕捉基因共表达模式、调控网络等“细胞语法”，从而实现对细胞状态的理解、模拟与预测。

不同于AlphaFold[3]等专注于单一特定任务的人工智能模型，现有细胞基础模型普遍采用“预训练+微调”的建模和应用策略。此类模型通常利用大量无标签数据，通过自监督学习进行预训练(Pre-training)，并在此基础上结合少量标注数据进行微调(Fine-tuning)，以提升在特定任务中的表现。虽然同为预训练模型，现有细胞基础模型又不同于Evo[4, 5]、ESM[6, 7]等以序列为输入的分子尺度基础模型(表1)，细胞基础模型通常直接以能够代表细胞整体状态的高维数据为输入，例如单细胞转录组表达谱或反映单细胞形态特征的高分辨率图像，依托大规模神经网络的表示学习能力，从海量多组学、多模态数据中自主挖掘细胞的内在运行逻辑[21]。因其输入数据无需人工标注，不再严格依赖既有的生物学知识框架，这一思路标志着细胞研究从传统的“假设-检验”范式，走向“模式发现-规律涌现”的数据驱动范式，为探索未知生物学原理开辟了新路径。在此基础上，通过进一步融合多模态信息，对细胞分子层面的特征进行统一表示，“人工智能虚拟细胞(Artificial intelligence virtual cell, AIVC)”等概念[22]也应运而生。

凭借对细胞内在规律的深刻理解，细胞基础模型有望在生命科学研究与生物医药开发的多个关键环节起到重要作用[22, 23]。在基础研究领域，细胞基础模型不仅能够通过零/少样本(Zero/few-shot)学习解决罕见细胞类型的注释难题[24-26]，还能在缺乏时序数据的情况下，基于学到的细胞状态流形重构

表1 分子尺度代表性基础模型

分子序列类型	代表性模型
DNA	Evo系列[4, 5]、DNABERT系列[8, 9]、Nucleotide Transformer[10]、HyenaDNA[11]、Caduceus[12]、Grover[13]
RNA	RNA-FM[14]、SpliceBERT[15]、RiNALMo[16]、ERNIE-RNA[17]
蛋白质	ESM系列[6, 7]、ProtTrans[18]、Ankh[19]、ProGen2[20]

复杂的发育轨迹与分化路径，甚至通过跨物种迁移学习解析进化的保守性机制。在转化医学与药物研发中，细胞基础模型正逐步演变为高精度的“虚拟筛选实验室”[27]。它们能够在计算机中模拟基因敲除、过表达或药物处理等扰动对细胞组学 (如转录组、表观组及蛋白质组) 的系统性影响，预测药物反应的异质性与耐药机制 [28-30]，从而大幅缩减湿实验的筛选范围并降低研发成本。此外，结合空间组学与病理图像的基础模型，研究者正致力于从多模态视角解析肿瘤微环境的异质性与细胞间通讯网络，为精准医疗中的患者分层与个性化治疗方案制定提供量化依据 [31-33]。为全面呈现这一新兴领域的研究格局，本文系统梳理细胞基础模型的研究进展与关键技术，剖析其核心挑战，并对未来发展方向进行展望。

1 细胞基础模型的研究进展

本综述重点关注采用“预训练 - 微调”架构的细胞基础模型，即在深度神经网络架构上大规模预训练细胞数据，用以学习可泛化的细胞表征 (Embedding)。此类表征具有较强的可迁移性，既可以通过微调或零 / 少样本学习等方式高效适配多样化下游任务，也可以作为高质量输入特征，赋能轻量级专用模型，以较高精度解决细胞类型注释、扰动效应预测等具体问题 [34-36]。

为了增强模型的表征能力和生物学可解释性，研究者们常将多种形式的生物学先验知识引入细胞基础模型的构建中 [37]。一类常见的先验知识是分子互作网络，如基因调控网络和蛋白质相互作用网络，它们为模型提供了分子间潜在的调控与物理连接信息 [38]。另外，很多模型也整合了功能注释数据，如基因本体论和信号通路数据库，为基因提供了丰富的语义背景和功能分组信息 [39]。也有研究者把基因组结构信息整合到模型中，如三维基因组接触图谱或染色体上的线性邻近关系，帮助模型捕捉顺式调控元件与其靶基因之间的长程相互作用。此外，跨物种的序列保守性和进化同源性信息也常被用于提升模型的泛化能力 [40]。将这些结构化知识作为约束项或辅助输入融入细胞基础模型的深度学习框架，不仅有助于在数据稀疏的情况下获得更稳健的表征，还能显著提升模型输出的生物学合理性。

按照预训练阶段所使用的主要数据模态，现有模型大致可分为细胞组学模型、图像模型和跨模态融合模型 (图 1)。其中，细胞组学模型和图像模型对应模态特定的基础模型，跨模态融合模型则希望实现不同模态数据在统一表征空间中的融合和对齐。本章将围绕这三类方向，简要评述代表性模型及其最新进展。

1.1 细胞组学基础模型

组学检测借助高通量手段，从基因组、转录组、蛋白质组、表观组、代谢组等多个层面对细胞或组织器官进行整体性刻画 [41, 42]。对于细胞基础模型而言，需要从整体层面描述细胞状态。尤其是单细胞组学数据，凭借数据体量大、具有生物学意义的特征信息丰富等优势，成为构建细胞基础模型的首选数据类型 [43-45]。

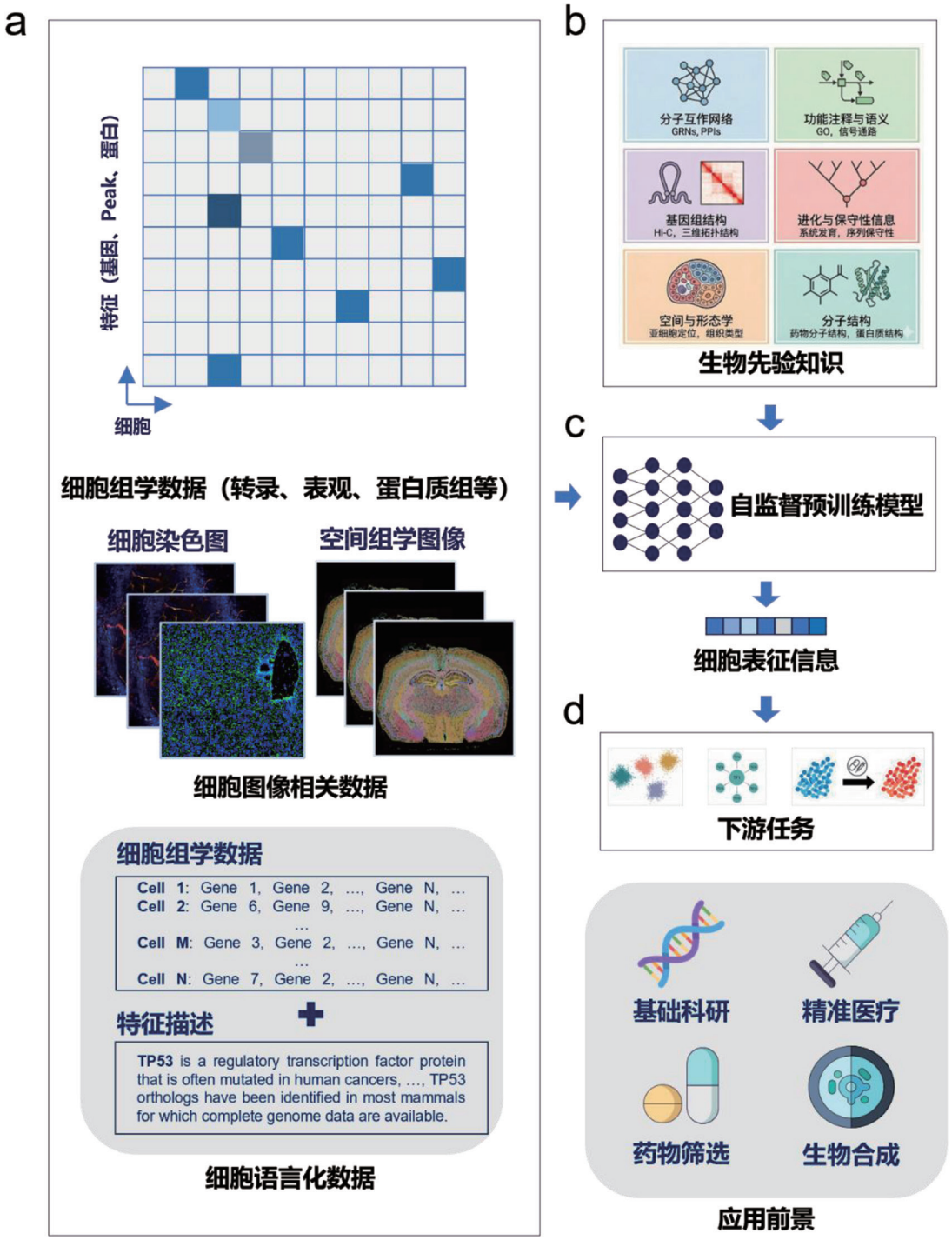
1.1.1 转录组基础模型

单细胞转录组 (Single-cell RNA-seq, scRNA-seq) 是当前技术最成熟、数据最丰富的单细胞组学数据类型，也是最直接定义细胞身份、描述细胞生理状态的数据模态之一。受自然语言处理领域基础模型范式的启发，研究者尝试将整个细胞转录组作为输入，将单个基因表达视为训练单位 (Token)，采用 Transformer 等架构 [46] 处理无标签转录组数据，通过自监督学习让模型理解基因共表达模式背后隐含的“转录语法”，从而实现对未见细胞状态的推断。然而，与离散的自然语言序列不同，scRNA-seq 数据既稀疏又为连续数值，这对数据编码方式和语义重构提出了额外挑战 [47]。

较早的代表性工作之一是 scBERT[26]。该模型采用 BERT (Bidirectional encoder representations from transformers) 双向编码器结构 [48]，通过掩码重建策略进行自监督训练。为适配 BERT 架构，scBERT 对连续的基因表达值进行离散化分箱，将其映射为类似“词频层级”的离散表示。BERT 擅长捕捉上下文信息，适用于细胞类型注释等判别型任务，但对未见细胞状态的生成和预测能力相对有限。

GeneFormer[27] 则将每个细胞中基因的绝对表达值，按该基因在整个样本中的平均表达值进行归一化，并依据归一化后的相对表达量对基因排序，以排序后的基因列表刻画细胞状态。这种基因排序在一定程度上具有技术噪声下的稳健性，有助于降低测序误差对细胞输入表示的影响。scGPT[24] 采用自回归生成式架构，通过生成被遮盖基因的表达值来学习细胞内的共表达规律。针对连续表达值，scGPT 采用“基因身份嵌入 + 表达量嵌入”的联合编码策略，以适应生成式预训练任务。

随后出现的一系列模型朝着更大规模预训练、



本图展示了构建细胞基础模型的核心要素与数据流动方向。(a)多模态数据输入：模型支持整合多种数据模态，包括高维稀疏的细胞组学数据(涵盖转录组、表观组、蛋白质组等矩阵信息)、包含空间与形态特征的细胞图像相关数据(如细胞染色显微图像、空间组学图像)，以及基于文本描述的细胞语言化数据(如对基因功能或突变信息的自然语言描述)。(b)生物学先验知识融合：为增强模型的可解释性，训练过程可以引入生物学先验知识，包括分子互作网络(如GRNs、PPIs)、功能注释与语义(如GO、信号通路)、基因组结构(如Hi-C、基因组三维拓扑结构)、进化与保守性信息(如系统发育树)、空间与形态学信息以及分子结构信息(如药物与蛋白质结构)。(c)预训练与表征学习：模型利用上述大规模无标注数据与先验知识进行自监督预训练，学习将复杂的细胞状态映射为低维、稠密的通用细胞表征信息。(d)下游任务与应用：习得的细胞表征可适配多种下游任务(如细胞类型聚类、调控关系推断、扰动效应预测等)，最终赋能基础科研、精准医疗、药物筛选及生物合成等关键领域。(注：图中空间组学图像素材源自10x Genomics网站)。

图1 细胞基础模型建模流程及应用前景

知识融合和跨物种训练方向发展。例如，scFoundation[25]在 5 000 万细胞上进行预训练，参数量约 1 亿；GeneCompass[40] 预训练细胞数达 1.2 亿；CellFM[49] 预训练细胞数为 1 亿，参数量达 8 亿。随着模型规模扩大，其在下游任务中的表现和泛化能力显著提升，部分模型甚至展现出零样本推理能力。GeneCompass 将多类生物学先验知识整合到基因嵌入中，引入生物学约束，使模型预测能力不再仅依赖统计拟合，从而提升了生物学机制层面的可解释性；同时尝试采用人 - 鼠跨物种数据联合预训练，以学习物种间保守的转录模式。

在扰动效应预测方面，最新的 STATE[28] 和 Tahoe-x1[29] 均引入专门的状态转换模块，直接学习“初始细胞状态 + 扰动因子”到“扰动后状态”的映射，使其在零 / 少样本细胞上下文的预测任务中，性能超过简单线性基线。STATE 的基础模型约含 6 亿参数，Tahoe-x1 参数量超过 30 亿，并在预训练阶段引入化学结构编码器，实现药物信息与细胞嵌入的深度融合，在相关任务中表现出更优性能。这两类模型在一定程度上具备对新扰动的零样本预测能力，也是 AIVC 能力的初步体现。

除了基于 Transformer 架构的模型，近期兴起的状态空间模型 (State space models, SSMs)，特别是 Mamba 架构 [50]，为处理大规模单细胞数据提供了新的视角。Mamba 通过引入选择性扫描机制，在保持线性计算复杂度的同时有效捕捉长程依赖，使其能够高效处理全基因组范围内的数万个基因表达数据。基于此架构的 GeneMamba 模型 [51] 创新性地引入了双向 Mamba 模块，以捕捉基因组中的双向上下文依赖。在输入数据上，GeneMamba 沿用了基于排序的离散化策略，但在训练目标中结合了下一个基因预测与基因通路对比损失，从而在学习细胞表征的同时强化了基因间的生物学功能联系，在多批次整合、细胞类型注释及基因相关性分析中展现了优异性能。scMamba[52] 则进一步探索了 Mamba 在处理未降维原始数据上的潜力，针对单核 RNA 测序 (snRNA-seq)，scMamba 摒弃了高变基因筛选的传统步骤，通过线性适配器策略直接对全量基因或基因组区域进行编码，利用双向 Mamba 模块捕捉稀疏数据中的全局依赖。这种设计不仅保留了完整的生物学信息，还在细胞类型注释、双细胞检测及数据插补等任务中表现出对低质量或高噪声数据的强大鲁棒性，证明了 SSM 架构在构建高效、全基因组尺度的细胞基础模型方面的巨大潜力。

除了架构上的创新，数据表示形式的转换也为细胞基础模型带来了新的思路。Cell2Sentence (C2S)[53] 提出了一种将单细胞数据直接适配到现有通用自然语言大模型的创新策略。在输入数据层面，C2S 直接将数值型的表达数据转化为大语言模型擅长处理的文本数据。在模型架构上，C2S 直接微调预训练好的通用自然语言大语言模型 (如 GPT-2、LLaMA 等)，利用这些模型在海量自然语言文本上习得的强大上下文理解和生成能力来解析细胞语言。在下游应用中，C2S 展现了独特的生成式能力，不仅能根据自然语言提示生成符合特定细胞类型特征的虚拟细胞，还能执行细胞类型注释等判别任务。更重要的是，它能够直接生成描述细胞生物学特征的自然语言文本摘要，为从单细胞数据中自动提取生物学知识提供了一个解决方案。

1.1.2 表观组基础模型

单细胞转座酶可及染色质测序 (Single-cell assay for transposase-accessible chromatin with high-throughput sequencing, scATAC-seq) 数据反映了细胞层面的染色质开放状态，是研究表观修饰的重要数据类型 [54]。基于该模态的基础模型可用于学习顺式调控元件 (Cis-regulatory elements, CREs) 与转录产物之间的调控语法，进而预测细胞分化轨迹、转录因子结合以及非编码区突变效应等 [55]。

为避免直接处理全基因组数百万潜在开放区域 (Peak) 所带来的巨大计算开销，EpiFoundation[56] 和 EpiAgent[57] 仅保留非零区域。EpiFoundation 在输入中引入染色体嵌入以标记 peak 的染色体位置信息；ChromFound 模型 [58] 采用类似的策略，使用染色体嵌入、位置嵌入、基因组坐标嵌入和连续的可及性嵌入的加和作为模型输入；EpiAgent 则采用词频 - 逆文档频率 (Term frequency-inverse document frequency, TF-IDF) 方法，对候选 CRE (Candidate cis-regulatory elements, cCREs) 按细胞类型特异性进行排序，由于输入长度限制，仅保留前若干高权重特征。GET 模型 [59] 采取另一种设计，将基因组划分为 2~4 Mb 的固定窗口，以窗口内转录因子基序评分和染色质开放性评分的组合作为输入向量，该表示在生物学上更具可解释性。

在模型架构上，EpiFoundation 采用 6 层仅有编码器的 Transformer 结构，预训练任务为“peak-基因对齐”，即利用细胞的 scATAC-seq 数据预测二值化的配对 scRNA-seq 数据，从而学习 peak 与基因之间的调控关系。模型可用于细胞类型注释和批

次校正，并在微调后用于基因表达预测。GET 为 12 层仅有编码器的 Transformer，预训练任务为“掩码区域预测”：随机遮盖基因组区域，根据上下文预测其特征，以学习转录调控语法及 CRE 共现关系。GET 可用于基因表达预测、零样本预测任意 DNA 序列的调控活性，以及识别长程增强子 - 启动子互作和转录因子协同作用。EpiAgent 为 18 层 BERT 编码器结构，预训练任务包括：细胞 -cCRE 对齐、信号重构以及“替换建模”等，用于支持数据插补、细胞类型注释和扰动效应预测。ChromFound 采用混合架构，结合了用于捕捉局部调控依赖的窗口分区自注意力机制和用于处理全基因组长程依赖的 Mamba 模块，用来完成零样本去批次效应和细胞类型注释、预测基因表达水平、推断增强子 - 基因的调控关系及模拟基因组扰动后的转录响应。

1.1.3 蛋白质组基础模型

蛋白质是细胞功能的直接执行者，从蛋白质层面构建细胞基础模型具有重要的科研和临床价值 [60]。但相比转录组数据，蛋白质数据获取成本更高、覆盖度更低。尽管 AlphaFold 系列模型已经在蛋白质静态结构预测方面取得了重大进展，但目前要模拟细胞内复杂蛋白质网络的动态变化，并在蛋白质层面准确预测扰动响应仍然非常困难。

ProteinTalks[30] 是该领域的开创性工作。研究者首先在约 1.6 万个癌细胞系样本中获得了约 3 800 万个药物扰动后的蛋白质测量值。在模型架构上，ProteinTalks 使用了神经常微分方程 (Neural ordinary differential equations, Neural ODEs)[61] 进行建模。这一设计使模型能够显式学习蛋白质网络随时间演变的连续轨迹，从而更精确地捕捉细胞对扰动的动态响应过程。

ProteinTalks 的两个预训练任务分别是预测未来时间点的蛋白质组状态和预测药物疗效，在此过程中学习细胞内蛋白质相互作用的内在规律。动态预训练赋予了模型较强的泛化能力和可解释性：不仅能够较准确地预测未见药物的疗效及其协同作用，还可以通过 SHAP 值分析反向推断导致耐药的关键蛋白，并展现出从细胞系泛化到类器官，再到临床患者数据的能力。

1.1.4 细胞组学信息预测模型

除了针对单一组学数据的表征学习，直接从 DNA 序列生成不同细胞状态的组学信息也逐渐建立，进而预测组学对应的细胞表型的“序列到功能”(Sequence-to-function, S2F) 模型 [62]。通过建立从基因组序列到基因表达、染色质状态及三维结构等组学表型的端到端映射，这类模型可以在不依赖实验数据的条件下预测序列变异的生物学效应。早期的 S2E 模型多数采用卷积神经网络 (Convolutional neural network, CNN) 架构，但受限于感受野小，难以有效捕捉调控序列之间的长程相互作用。随着深度学习技术的演进，基于 Transformer 架构的模型通过自注意力机制显著扩展了序列上下文窗口，使得对基因组长程调控逻辑的建模成为可能。

Enformer 模型 [63] 是 S2F 基础模型领域的里程碑式工作。该模型以约 200 kb 的长 DNA 序列作为输入，采用“卷积层 +Transformer”的混合架构，利用 Transformer 层极大地扩展了模型的感受野，能够有效整合远端调控信息。在输出端，Enformer 能够同时预测数千种跨细胞类型和组织的表观基因组及转录组图谱 (包括 CAGE、ChIP-seq、DNase-seq 等)。得益于对长程依赖关系的建模能力，Enformer 在增强子 - 启动子相互作用预测和非编码区变异效应预测等任务上展现出了优于传统 CNN 模型的性能，为解析非编码基因组的功能提供了有力工具。

近期提出的 AlphaGenome 模型 [64] 进一步将 S2F 模型的性能推向了新的高度。AlphaGenome 将输入序列长度扩展至 1 Mb，并引入了包含序列并行策略的 U-Net[65] 风格架构，结合 Transformer 模块，在大幅提升上下文窗口大小的同时实现了单碱基分辨率的精准预测。在目标数据模态上，AlphaGenome 不仅覆盖了基因表达和表观遗传修饰，还创新地整合了剪接、多聚腺苷酸化以及三维基因组接触图谱等多种复杂模态。通过在大规模人类和小鼠基因组数据上的预训练及蒸馏策略，AlphaGenome 在多项变异效应预测基准测试中展示了优异的性能，特别是在解析涉及复杂调控机制 (如剪接异常或致癌基因的 Neo-enhancer 形成) 的临床相关变异方面，展现出了超越现有模型的综合分析能力。

1.2 图像基础模型

高通量显微成像 (High-throughput microscopy, HTM) 是将显微镜、自动化机器人 (机械臂、自动载物台) 和图像分析软件整合在一起的成像系统 [66, 67]。该系统成像速率极高，每秒可采集上百张图像，一天即可产生 TB 级的细胞显微图像数据。显然，传统人工分析方法 [68] 及基于手工特征的计算机视觉算法 [69] 已难以应对这类海量数据；深度学习模型，尤其是 CNN 一度成为细胞图像分析的主流方法 [70]。但作为典型专用模型，CNN 通常针

对特定数据集和任务训练，泛化能力有限；同时，其局部感受野特性限制了对长程依赖关系的建模，不适用于需要同时考虑细胞微环境等上下文信息的任务。随着 Transformer 架构在 NLP 领域取得成功，基于 Transformer 的细胞图像基础模型逐渐兴起。通过在海量、多样化图像数据上进行自监督预训练，这类模型有望学习到通用的细胞形态特征，并具备一定的零 / 少样本泛化能力。

1.2.1 细胞结构感知基础模型

细胞结构感知是指对细胞及其亚细胞结构进行几何界定，例如识别细胞位置和形状，其核心任务包括实例分割、目标检测和细胞追踪。难点主要体现在：在高密度细胞簇中分离黏连个体、在染色不均或伪影干扰下精确勾勒目标边缘，以及处理病理图像中复杂的组织背景等。细胞结构感知基础模型通常采用编码器 - 解码器架构，输出为分割掩膜或边界框。

CellViT[71] 是一个用于病理组织图像中细胞核分割和分类的基础模型。该模型将 U-Net 架构中的下采样编码器替换为视觉 Transformer (Vision transformer, ViT)[72]，利用 ViT 的全局感受能力提取图像的高阶语义信息以定位目标，同时通过 ViT 编码器与上采样解码器之间的跳跃连接，保留编码器浅层中未被强烈压缩的细胞核纹理和边缘等高分辨率特征。经过解码器恢复空间分辨率后，模型对图像的每个像素进行分类，输出与原图同尺寸的二值掩膜，实现细胞核的像素级分割。

值得注意的是，CellViT 并未从头训练 ViT 编码器，而是采用迁移学习策略：一方面使用在 1.04 亿张病理图像上预训练的 ViT 权重作为“域内”初始化；另一方面也可采用自然图像分割基础模型 SAM (Segment anything model) 的编码器权重作为“域外”初始化 [73]。随后，在包含 19 种组织类型、约 20 万个标注细胞核的 PanNuke[74] 数据集上进行微调，从而获得较强的跨组织泛化能力。

除了 CellViT 之外，还有一类工作利用自然图像分割基础模型 SAM 的零样本能力，将其适配到细胞图像分割任务中。由于 SAM 的推理流程需要用户提供点或框作为提示，才能执行分割，在海量细胞图像分割任务中依赖人工提示并不现实。为此，CellSAM[75] 训练了一个轻量级目标检测器 CellFinder，自动生成细胞边界框作为提示，再交由 SAM 生成高质量分割掩膜。面对同样问题，SAMCell [76] 则选择对 SAM 进行微调，使其直接预测显微图像中每个像素到最近细胞边界的欧几里得距离图，并利用分水岭算法从距离图中恢复细胞实例。由于距离图相较于简单二值掩膜蕴含更丰富的拓扑信息，SAMCell 在需要分离黏连细胞的任务中表现更佳。

另一个基于 SAM 的显微图像分割模型 μSAM (Segment anything for microscopy)[77] 在保留 SAM 交互特性的同时，将其推广至三维和时序细胞图像的分割任务。理论上，基础模型可以在预训练时接收不同类型的图像输入，但不同成像方式的物理机制差异较大，导致图像视觉特征存在显著模态差异，细胞层面的模型通常面临不同程度的“模态鸿沟”。例如，μSAM 分别为光学显微镜和电子显微镜训练了两个通用模型，采用不同的权重以适配两类设备产生的图像数据。

1.2.2 细胞表型认知基础模型

细胞表型认知旨在从图像中提取高维、语义丰富的特征向量，用于表征和识别细胞状态，其核心任务包括表征学习、聚类和扰动预测。主要挑战在于：如何在从像素层面提取具有生物学意义的特征 (如细胞周期阶段、药物反应等) 的同时，有效去除批次效应和成像噪声。这类模型通常基于 DINO (Distillation with no labels)[78]、MAE (Masked autoencoder)[79] 等自监督学习框架，输出代表细胞状态的高维稠密向量。

Recursion Phenom 系列模型 [79] 是细胞级图像基础模型在工业界应用的代表。这些模型主要基于 MAE 架构，使用 ViT 作为骨干网络，自监督任务是通过随机遮盖大部分输入图像，训练模型重建缺失像素，从而学习细胞形态表征。其预训练数据规模极为庞大，包括公开数据集 RxRx1 (约 12.5 万张)、RxRx3 (约 220 万张)，以及内部构建的 RPI-52M (约 5 100 万张) 和 RPI-93M (约 9 300 万张)，这些图像来源于数以万计的基因敲除和化合物处理实验。研究发现，当将同一信号通路上游和下游基因，或同一蛋白质复合物成员基因视为“相关基因”，把针对这些相关基因扰动后获得的细胞图像输入 Phenom 时，模型生成的细胞表型嵌入向量会在特征空间中自然聚集在一起。这相当于模型仅通过观察细胞形态变化，就在一定程度上“重新发现”了已知的生物学通路结构。

DINO 模型 [78] 采用“教师 - 学生”式知识蒸馏 (knowledge distillation) 架构：学生模型通过对比自身输出与教师模型输出的差异来更新参数。与传统知识蒸馏不同的是，DINO 是自监督模型，并不存

在预先训练好的教师网络，而是通过对学生模型参数做指数移动平均 (Exponential moving average, EMA) 来得到教师参数，可视为对学生历史状态的一种平滑聚合。DINO 的另一个关键设计是，教师模型接收图像的全局视图，而学生模型仅看到局部视图。预训练目标要求：同一张图像的不同增强视图，分别经过教师与学生编码后，应得到一致的特征表示。通过强制对齐局部与全局视角下的特征，DINO 被迫学习图像内部的语义一致性，即相对稳定的高阶语义特征。

在荧光显微成像和高内涵筛选图像等数据上，预训练后的 DINO 展现出明显的“涌现”属性。例如，在完全没有输入时间序列标签的前提下，DINO 生成的细胞嵌入在隐空间中自动排布成环形流形，与细胞周期 (G_1-S-G_2-M) 的进程高度对应。这说明 DINO 仅凭形态学信息，就能够在一定程度上重新发现细胞周期这一基本生物学规律。

SubCell[80] 是一个用于蛋白质亚细胞定位识别和功能推断的图像基础模型。其预训练数据来自人类蛋白质图谱 (Human protein atlas, HPA)[81]，其中目标蛋白的亚细胞定位通过图像中的四色标记系统进行可视化：细胞核 (蓝色)、微管 (红色)、内质网 (黄色) 和目标蛋白 (绿色)，从而建立了细胞形态与蛋白质定位之间的对应关系。SubCell 采用多任务学习框架。首个任务为图块重构任务：随机遮盖图像中的一个图块，要求模型重构缺失区域，用于学习细胞纹理特征和局部连续性；随后的任务是细胞特异性对比任务：基于对比学习，最小化同一细胞不同增强视图 (旋转、裁剪、加噪等) 之间的特征距离，最大化不同细胞之间的特征距离，使模型学会识别细胞身份和整体形态，并对技术噪声具有鲁棒性；最后的任务是蛋白质特异性对比任务：最小化由同一种抗体染色的不同细胞之间的特征距离，使模型忽略个体间形态差异，专注于提取蛋白质空间分布模式。

研究表明，由 SubCell 提取的图像特征之间的距离与蛋白质 - 蛋白质相互作用发生的物理可能性高度相关。这意味着，深度学习模型已经能够仅依赖视觉数据跨越到分子互作层面的预测，为虚拟筛选提供了新的维度。

1.3 多模态融合模型的研究现状

传统单细胞组学技术虽然能够以高通量方式解析细胞的分子特征，但由于需要组织解离，细胞丢失了空间位置信息和微环境上下文。细胞功能不仅由其内部基因表达决定，还受到其在组织中的拓扑位置、与邻近细胞的相互作用以及组织形态特征的影响。空间转录组学技术 (Spatial transcriptomics, ST) 可同时获取同一组织切片的高分辨率病理图像和原位基因表达谱，为构建具备空间信息的细胞基础模型提供了全新的数据维度。

当前，细胞基础模型的研究呈现出从单一模态向多模态融合转变的趋势。多模态融合模型不再局限于学习基因间的共表达关系，而是试图在统一语义空间内对齐转录组学、形态学和拓扑学信息。本节介绍多模态融合基础模型的研究现状，涵盖图神经网络、Transformer 架构、对比学习以及大模型适配等多种技术路线，并重点讨论 Novae[82]、stFormer[83]、scGPT-spatial[31]、OmiCLIP[32]、ST-Align[84] 与 Nicheformer[33] 等代表性模型的架构思路、训练策略及下游应用。

1.3.1 基于图神经网络的空间表征学习模型

在空间转录组数据的建模中，将细胞或 spot 视为节点、将空间邻近关系视为边来构建细胞邻域图，是表征细胞微环境的直接方式。Novae[85] 是一类基于图神经网络的空间基础模型，重点解决多切片分析中的批次效应及通用表征学习问题。Novae 采用计算机视觉领域的 SwAV 自监督框架，并在图注意力网络 (Graph attention networks, GATs) 中通过聚合邻居节点特征来更新中心节点表示，从而具备处理非规则空间结构的能力。

在自监督训练过程中，Novae 借助最优传输 (Optimal transport) 方法，约束不同切片样本在表征空间中接近均匀分布，从而实现跨切片的批次效应校正。由于在包含 18 种组织、约 3 000 万个细胞的大规模数据上进行预训练，Novae 在跨组织空间分区和层级化多分辨率空间划分等任务中展现了较强的零样本泛化能力。

1.3.2 基于Transformer的空间交互模型

在空间转录组建模中，Transformer 由于其全局注意力机制，非常适合刻画细胞间通讯与空间长程依赖。Nicheformer[33] 是目前公开的规模最大的空间组学基础模型之一，其预训练语料包含人和小鼠共计 1.1 亿余个细胞，覆盖 73 种不同组织器官。Nicheformer 采用基因表达排序作为 Transformer 的输入，并通过同源基因映射实现人鼠跨物种联合训练。在下游任务上，Nicheformer 展现出较强的迁移能力，仅凭单个细胞的基因表达谱即可预测其在组织中的空间标签、周围细胞密度及邻域细胞组成。

stFormer[83] 通过引入交叉注意力机制来模拟细胞间信号转导。在结构上，stFormer 的编码器包含两个并行分支：自注意力分支处理中心细胞内部的基因表达数据，学习基因共表达关系和转录调控逻辑；交叉注意力分支则使用中心细胞的基因嵌入作为查询，将其空间邻域中其他细胞表达的配体基因嵌入作为键和值，从而在分子层面刻画微环境中配体 - 受体信号对中心细胞转录状态的调节。stFormer 在包含约 410 万个空间样本的 CROST 数据库上进行了预训练 [86]，在批次效应去除和计算模拟扰动等任务中取得了较好效果。

scGPT-spatial[31] 基于单细胞基础模型 scGPT，通过持续预训练注入空间信息，并引入混合专家 (Mixture of experts, MoE) 架构 [87] 以处理不同空间测序技术产生的数据。通过 spot 内掩码基因表达预测和 spot 间邻域重建两类预训练任务，scGPT-spatial 同时学习了 spot 内部的转录调控逻辑和微环境对细胞状态的影响模式。其预训练数据集包含由 Visium、Visium HD、MERFISH 和 Xenium 四种技术产生的 3 000 多万个 spots，覆盖 20 余种器官及多种疾病状态，使模型在多种下游任务中展现出优异的泛化性能。

1.3.3 图像-组学对齐模型

将 H&E 染色图像与基因表达谱在像素和分子两个层面进行对齐，是实现智能病理诊断的关键步骤。OmiCLIP 模型 [88] 基于 CLIP (Contrastive language-image pre-training) 的双塔架构：一侧使用基于 ViT 的视觉编码器处理 H&E 图像切片，另一侧将基因表达谱转化为文本描述输入文本编码器。模型通过最大化配对图像 - 转录组样本在隐空间中的余弦相似度，并最小化非配对样本的相似度，学习两种模态的联合表示，从而实现组织学图像与转录组数据之间的语义对齐。基于此，OmiCLIP 可在 Loki 平台上支持跨模态检索、多切片配准以及零样本的组织区域自动语义注释等任务。

ST-Align[84] 通过 spot 对齐、niche 对齐和 spot-niche 交互对齐三类层次化预训练任务，同时捕捉空间局部细节和整体结构，实现多尺度空间特征的深度融合。在模态融合方面，ST-Align 采用基于注意力的融合模块，利用交叉注意力机制实现图像特征与基因表达特征的深度对齐。在约 130 万对图像 - 转录组样本上完成预训练后，ST-Align 在空间聚类识别和基因表达预测任务中均优于相应单模态基线模型和简单的 CLIP 式对齐方法。

2 当前细胞基础模型面临的挑战

上述基于不同模态数据构建的细胞基础模型 (表 2)，已经在一定程度上学到细胞内部的调控逻辑，并可为多种下游任务提供有效表征。然而，相比 ChatGPT[93] 这样的通用语言模型或 AlphaFold 等专用模型，目前细胞基础模型的整体性能仍未达到理想水平，要实现性能的跨越式提升仍面临多重挑战。

从数据角度看，数据质量和体量无疑是决定模型性能的关键因素。现有细胞基础模型的训练数据主要为各类单细胞组学数据，这类数据普遍具有“高稀疏、高噪声、高维度”的特点 [43, 94]，一方面，有效表征细胞状态的信号被严重稀释；另一方面，每个细胞的高维特征又对训练数据体量提出了更高要求。目前部分最新模型的训练数据量已超过亿级，在现有测序技术条件下进一步指数级扩大量级的难度较大。此外，单细胞测序本质上获得的是细胞在某一时间点的“静态快照”，获取这些数据对细胞具有破坏性，难以获得严格配对的时序数据，这对模型从相关关系中推断因果关系提出了额外挑战 [95]。

从模型架构设计来看，当前细胞基础模型多直

表2 细胞尺度代表性基础模型

建模尺度	模态类型	代表性模型(文章中未详细介绍的模型用粗体展示)
细胞尺度	转录组	scBERT、Geneformer、scGPT、scFoundation、GeneCompass、CellFM、STATE、Tahoe-x1、GeneMamba、scMamba、Cell2Sentence (C2S)、**UCE**[89]、**xTrimoGene**[90]
	表观组	EpiFoundation、EpiAgent、ChromFound、GET
	蛋白质组	ProteinTalks
	多组学预测	Enformer、AlphaGenome
	图像	CellViT、CellSAM、SAMCell、μSAM、SubCell、Recursion Phenom、DINO (Cell)、**CytoImageNet**[91]
多细胞尺度	空间组+多模态	Nicheformer、Novae、stFormer、scGPT-spatial、ST-Align、**SToFM**[92]
	病理图像组+多模态	OmiCLIP

接借用自然语言处理或通用图像处理领域的常用架构，这在一定程度上导致组学数据在这些模型上的训练效率不高，难以充分利用其统计和生物学特征，并进一步抬高了对数据量的需求。更重要的是，这些架构在可解释性方面普遍存在不足，即决策过程难以溯源和结果可信度难以评估。而生物研究 (如致病基因发现、药物靶点设计等) 是客观事实导向的，需要提供分子通路层面的具体机制解释，现有模型体系在这一点上仍有较大提升空间 [96]。当前，提升模型可解释性的尝试主要集中在从模型内部参数或中间层表征中提取生物学知识。例如，GeneCompass 通过深度解析 Transformer 架构中的自注意力权重矩阵，识别关键转录因子与其靶基因之间的调控依赖关系，将抽象的数学权重转化为可验证的基因调控网络，验证了模型确实捕捉到了底层的转录调控语法而非简单的统计相关性。在图像模态方面，Recursion Phenom 系列模型通过掩码自编码器学习到的细胞表型特征向量在潜空间中呈现出高度结构化的分布，功能相关的基因或作用机制相似的化合物会自动聚类，这种生物学结构使得研究者能够通过特征空间的几何邻近性来推断未知扰动的功能机制。然而，现有的解释手段多停留在相关性层面，如何构建具备因果推理能力的可解释性框架，使模型能够生成可直接指导湿实验验证的机制假设，仍是未来亟待攻克的难题。此外，目前基础模型的评估主要依赖下游任务表现，而细胞基础模型可支持的下游任务类型十分多样。如果能够更加清晰地刻画不同基础模型所捕获的细胞特征，将有助于在设计下游任务时选择更合适的“底座模型”。

从泛化性能角度看，在批次效应去除、细胞类型注释等任务上，部分基础模型已经展现出较好的泛化能力，甚至可以实现零样本注释。然而，对于基因或小分子扰动引起的细胞状态变化预测等更具挑战性的任务，现有模型往往只具备有限的泛化能力。以基因扰动为例，真核生物的蛋白质编码基因数量有限，通常少于两万，而蛋白质长度一般为数百个氨基酸，现有针对特定物种的细胞基础模型多使用参考基因或蛋白质序列作为输入，并未系统融入变异信息。即便配合 ESM 等序列基础模型，对多样化突变的覆盖仍远远不足，难以期望模型准确预测未见基因扰动的效应 [97]。近期的独立基准测试显示，在预测基因扰动任务中，现有的细胞基础模型 (如 scGPT、Geneformer 等) 在零样本条件下的预测精度往往难以超越简单的线性模型或均值基线算法 [98, 99]。这反映出当前的预训练任务 (如掩码重建) 可能尚未深刻捕捉到基因调控网络中的因果逻辑，导致模型在面对全新类型的扰动或跨物种推断时泛化能力不足 [100]。此外，目前的评估标准尚未统一，过度依赖统计指标 (如 L2 距离或相关系数) 而忽视了生物学意义上的准确性 (如扰动方向或表型特异性基因的变化)。因此，如何构建涵盖因果推断的预训练任务，以及建立包含“简单基线模型”对照的严格评估体系，是提升细胞基础模型零样本预测可靠性亟待解决的瓶颈问题 [34]。相比之下，小分子药物的结构相对更简单，且可用于建模的化合物空间更加丰富，因此在小分子扰动效应预测任务上实现较好泛化在理论上更为可行 [28, 101]。

此外，由于生命科学问题本身的复杂性和多样性，要对细胞基础模型进行全面评估，必须依赖多种下游任务和场景。如何构建通用于不同细胞基础模型，并在社区中获得广泛认可的评测任务和评测数据集，需要整个领域的共同努力 [102]。

3 细胞基础模型的未来展望

通过梳理细胞基础模型的主要进展和当前挑战，可以发现大规模预训练模型在细胞状态表征和刻画等方面已经展现出重要潜力和应用前景。然而，要真正实现其在复杂生命过程解析和生物医学中的应用，仍需在模型性能提升等关键科学问题上取得突破。

不同于物理和化学等可用简洁公式刻画的学科，生命科学中的大量先验知识往往以自然语言形式存在，是在长期观察和实验基础上逐步总结出的经验规律。现有模型架构尚未充分考虑如何系统地将这些知识引入模型训练。未来的细胞基础模型有望从单纯的数据驱动逐步转向“数据 + 知识”并重的范式 [37, 39]。通过将分子层面的基本规律、基因调控机制、细胞谱系信息以及宏观进化理论等多层次生物学先验知识系统性整合进模型，不仅有助于在噪声较大或标注有限的场景下获得更加稳健的表征，也有望提升模型的泛化能力。知识图谱、因果网络等结构化知识表示可为模型提供可对齐的语义空间，使基础模型在“拟合数据”的同时，更好地利用既有机理，从而增强结果的可解释性和生物学可信度 [38, 103]。

近期关于 AIVC 的构想普遍将多模态信息融合视为构建虚拟细胞的关键特征 [22]。随着单细胞转录组、表观组、空间组学、高通量成像以及临床表型

等多源数据不断积累，如何在统一表征空间中整合这些信息，并通过多模态联合预训练实现模态间的相互校正和互补，而非简单对齐，将直接决定模型在真实生物和医学问题中的应用潜力。例如，空间组学为分子特征提供组织结构背景，成像数据为细胞状态附加形态学语境，而临床信息则将细胞层面的变化与个体预后联系起来。由此构建的多模态细胞基础模型，有望在鲁棒性和应用广度上显著优于单一模态模型 [104]。

数据多样性同样至关重要。目前高质量数据和实验验证主要集中在少数模式生物，而大量非模式生物蕴含的进化多样性尚未被充分利用 [105]。这些生物在遵循共同细胞运行规律的同时，又在生物分子、细胞类型和器官功能等层面展现出丰富差异。通过构建统一的跨物种基因与通路映射体系，引入序列和调控元件的保守性信息及系统发育信息，并在模型架构中同时刻画跨物种共享特征与物种特异特征，细胞基础模型有望实现更可靠的跨物种泛化，拓展模型适用的物种范围，进而推动进化研究和生物多样性保护等基础问题的深入研究 [4, 11]。

在模型设计层面，未来细胞基础模型需要逐步摆脱简单“照搬”自然语言处理或计算机视觉架构的做法，转向更加契合生物学数据特点的专门化设计。在模型结构上，需要更加重视多尺度图结构与稀疏高维特征的联合建模 [106]，将基因调控网络、细胞 - 细胞相互作用网络以及组织空间结构统一纳入考虑；在训练策略上，则需要探索更适合生物数据分布特性的自监督、弱监督和迁移学习范式，在控制模型复杂度并提升可解释性的前提下，充分挖掘海量未标注数据的价值。通过上述方向的持续优化，有望在稳步提升模型预测性能的同时，降低对超大规模算力的依赖，并增强方法在不同实验条件和技术平台间的可移植性。

此外，随着大语言模型技术的演进，后训练 (Post-training) 技术，如指令微调和人类反馈强化学习，将成为提升细胞模型可用性的关键。通过对齐生物学家的需求，未来的模型不仅能输出预测数值，还能给出推理过程 [107-110]。更进一步，结合 AI 智能体 (AI agents) 技术 [111, 112]，细胞基础模型有望从被动的分析工具进化为自主探索的科学助手。智能体可以根据模型预测结果提出科学假设，自主设计验证实验，甚至驱动自动化实验室完成闭环验证，从而加速生命科学的发现过程。

最后，进化信息是理解细胞功能的重要依据。未来的模型应更深度地融合比较基因组学数据，建模跨物种的序列和调控网络保守性。这不仅有助于提升模型在不同物种间的泛化能力，也能帮助我们追溯细胞类型和功能的起源与演化路径。

总之，未来的细胞基础模型将逐步从单一数据驱动的表征工具，演进为融合生物学知识、多模态数据和生物学场景化模型架构的跨模态、跨尺度、跨物种统一表征框架。结合不断发展的智能体技术，这类模型有望进一步加深我们对细胞系统运行机理的理解，为疾病机制解析、药物靶点发现和精准医疗实践提供更加坚实而灵活的计算支撑。

[参 考 文 献]

[1] Slepchenko BM, Schaff JC, Macara I, et al. Quantitative cell biology with the virtual cell. Trends Cell Biol, 2003, 13: 570-6

[2] Johnson GT, Agmon E, Akamatsu M, et al. Building the next generation of virtual cells to understand cellular biology. Biophys J, 2023, 122: 3560-9

[3] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature, 2021, 596: 583-9

[4] Brixi G, Durrant MG, Ku J, et al. Genome modeling and design across all domains of life with Evo 2. bioRxiv, 2025, https://doi.org/10.1101/2025.02.18.638918

[5] Nguyen E, Poli M, Durrant MG, et al. Sequence modeling and design from molecular to genome scale with Evo. Science, 2024, 386: eado9336

[6] Rives A, Meier J, Sercu T, et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proc Natl Acad Sci U S A, 2021, 118: e2016239118

[7] 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

[8] Zhou Z, Ji Y, Li W, et al. DNABERT-2: efficient foundation model and benchmark for multi-species genome. arXiv, 2023, https://doi.org/10.48550/arXiv.2306.15006

[9] Ji Y, Zhou Z, Liu H, et al. DNABERT: pre-trained bidirectional encoder representations from transformers model for DNA-language in genome. Bioinformatics, 2021, 37: 2112-20

[10] Dalla-Torre H, Gonzalez L, Mendoza-Revilla J, et al. Nucleotide Transformer: building and evaluating robust foundation models for human genomics. Nat Methods, 2025, 22: 287-97

[11] Nguyen ED, Poli M, Faizi M, et al. HyenaDNA: long-range genomic sequence modeling at single nucleotide resolution. arXiv, 2023, https://doi.org/10.48550/arXiv.2306.15794

[12] Schiff Y, Kao CH, Gokaslan A, et al. Caduceus: bi-directional equivariant long-range DNA sequence modeling. Proc Mach Learn Res, 2024, 235: 43632-48

[13] Sanabria M, Hirsch J, Joubert PM, et al. DNA language model GROVER learns sequence context in the human genome. Nat Mach Intell, 2024, 6: 911-23
[14] Chen J, Hu Z, Sun S, et al. Interpretable RNA foundation model from unannotated data for highly accurate RNA structure and function predictions. arXiv, 2022, https://doi.org/10.48550/arXiv.2204.00300
[15] Chen K, Zhou Y, Ding M, et al. Self-supervised learning on millions of primary RNA sequences from 72 vertebrates improves sequence-based RNA splicing prediction. Brief Bioinform, 2024, 25: bbae163
[16] Penić RJ, Vlašić T, Huber RG, et al. RiNALMo: general-purpose RNA language models can generalize well on structure prediction tasks. Nat Commun, 2025, 16: 5671
[17] Yin W, Zhang Z, Zhang S, et al. ERNIE-RNA: an RNA language model with structure-enhanced representations. Nat Commun, 2025, 16: 10076
[18] Elnaggar A, Heinzinger M, Dallago C, et al. ProtTrans: toward understanding the language of life through self-supervised learning. IEEE Trans Pattern Anal Mach Intell, 2022, 44: 7112-27
[19] Elnaggar A, Essam H, Salah-Eldin W, et al. Ankh: optimized protein language model unlocks general-purpose modelling. arXiv, 2023, https://doi.org/10.48550/arXiv.301.06568
[20] Nijkamp E, Ruffolo JA, Weinstein EN, et al. ProGen2: exploring the boundaries of protein language models. Cell Syst, 2023, 14: 968-78.e3
[21] Cui T, Tejada-Lapuerta A, Brbić M, et al. Towards multimodal foundation models in molecular cell biology. Nature, 2025, 640: 623-33
[22] 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
[23] Moor M, Banerjee O, Abad ZSH, et al. Foundation models for generalist medical artificial intelligence. Nature, 2023, 616: 259-65
[24] 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
[25] Hao M, Gong J, Zeng X, et al. Large-scale foundation model on single-cell transcriptomics. Nat Methods, 2024, 21: 1481-91
[26] 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
[27] Theodoris CV, Xiao L, Chopra A, et al. Transfer learning enables predictions in network biology. Nature, 2023, 618: 616-24
[28] Adduri AK, Gautam D, Bevilacqua B, et al. Predicting cellular responses to perturbation across diverse contexts with state. bioRxiv, 2025, https://doi.org/10.1101/2025.06.26.661135
[29] Gandhi S, Javadi F, Svensson V, et al. Tahoe-x1: scaling perturbation-trained single-cell foundation models to 3 billion parameters. bioRxiv, 2025, https://doi.org/10.1101/2025.10.23.683759
[30] Sun R, Qian L, Li Y, et al. A perturbation proteomics-based foundation model for virtual cell construction. bioRxiv, 2025, https://doi.org/10.1101/2025.02.07.637070
[31] Wang C, Cui H, Zhang A, et al. scGPT-spatial: continual pretraining of single-cell foundation model for spatial transcriptomics. bioRxiv, 2025, https://doi.org/10.1101/2025.02.05.636714
[32] Chen W, Zhang P, Tran TN, et al. A visual-omics foundation model to bridge histopathology with spatial transcriptomics. Nat Methods, 2025, 22: 1568-82
[33] Tejada-Lapuerta A, Schaar AC, Gutgesell R, et al. Nicheformer: a foundation model for single-cell and spatial omics. Nat Methods, 2025, 2525-38
[34] Wu J, Ye Q, Wang Y, et al. Biology-driven insights into the power of single-cell foundation models. Genome Biol, 2025, 26: 334
[35] Bian H, Chen Y, Luo E, et al. General-purpose pre-trained large cellular models for single-cell transcriptomics. Natl Sci Rev, 2024, 11: nwae340
[36] Baek S, Song K, Lee I. Single-cell foundation models: bringing artificial intelligence into cell biology. Exp Mol Med, 2025, 57: 2169-81
[37] Von Rueden L, Mayer S, Beckh K, et al. Informed machine learning – A taxonomy and survey of integrating prior knowledge into learning systems. IEEE Transac Knowl Data Engin, 2019, 35: 614-33
[38] Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics, 2018, 34: i457-66
[39] Ma J, Yu MK, Fong S, et al. Using deep learning to model the hierarchical structure and function of a cell. Nat Methods, 2018, 15: 290-8
[40] Yang X, Liu G, Feng G, et al. GeneCompass: deciphering universal gene regulatory mechanisms with a knowledge-informed cross-species foundation model. Cell Res, 2024, 34: 830-45
[41] Berger B, Peng J, Singh M. Computational solutions for omics data. Nat Rev Genet, 2013, 14: 333-46
[42] Manzoni C, Kia DA, Vandrovcova J, et al. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. Brief Bioinform, 2018, 19: 286-302
[43] Szalata A, Hrovatin K, Becker S, et al. Transformers in single-cell omics: a review and new perspectives. Nat Methods, 2024, 21: 1430-43
[44] Wang D, Bodovitz S. Single cell analysis: the new frontier in 'omics'. Trends Biotechnol, 2010, 28: 281-90
[45] Baysoy A, Bai Z, Satija R, et al. The technological landscape and applications of single-cell multi-omics. Nat Rev Mol Cell Biol, 2023, 24: 695-713
[46] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. arXiv, 2023, https://doi.org/10.48550/arXiv.1706.03762
[47] Lopez R, Regier J, Cole MB, et al. Deep generative modeling for single-cell transcriptomics. Nat Methods, 2018, 15: 1053-8

[48] Devlin J, Chang MW, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv, 2019, https://doi.org/10.48550/arXiv.1810.04805
[49] Zeng Y, Xie J, Shangguan N, et al. CellFM: a large-scale foundation model pre-trained on transcriptomics of 100 million human cells. Nat Commun, 2025, 16: 4679
[50] Gu A, Dao T. Mamba: linear-time sequence modeling with selective state spaces. arXiv, 2023, https://doi.org/10.48550/arXiv.2312.00752
[51] Qi C, Fang H, Hu T, et al. Bidirectional mamba for single-cell data: efficient context learning with biological fidelity. arXiv, 2025, https://doi.org/10.48550/arXiv.2504.16956
[52] Oh G, Choi B, Jin S, et al. scMamba: a pre-trained model for single-nucleus RNA sequencing analysis in neurodegenerative disorders. arXiv, 2025, https://doi.org/10.48550/arXiv.2502.19429
[53] Levine D, Rizvi SA, Lévy S, et al. Cell2Sentence: teaching large language models the language of biology. bioRxiv, 2024, https://doi.org/10.1101/2023.09.11.557287
[54] Buenrostro JD, Giresi PG, Zaba LC, et al. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods, 2013, 10: 1213-8
[55] Mitra S, Malik R, Wong W, et al. Single-cell multi-ome regression models identify functional and disease-associated enhancers and enable chromatin potential analysis. Nat Genet, 2024, 56: 627-36
[56] Wu J, Wan C, Ji Z, et al. EpiFoundation: a foundation model for single-cell ATAC-seq via peak-to-gene alignment. bioRxiv, 2025, https://doi.org/10.1101/2025.02.05.636688
[57] Chen X, Li K, Cui X, et al. EpiAgent: foundation model for single-cell epigenomics. Nat Methods, 2025, 22: 2316-27
[58] Jiao Y, Liu Y, Zhang Y, et al. ChromFound: towards a universal foundation model for single-cell chromatin accessibility data. arXiv, 2025, https://doi.org/10.48550/arXiv.2505.12638
[59] Fu X, Mo S, Buendia A, et al. A foundation model of transcription across human cell types. Nature, 2025, 637: 965-73
[60] Qian L, Sun R, Aebersold R, et al. AI-empowered perturbation proteomics for complex biological systems. Cell Genom, 2024, 4: 100691
[61] Chen TQ, Rubanova Y, Bettencourt J, et al. Neural ordinary differential equations. arXiv, 2018, https://doi.org/10.48550/arXiv.1806.07366
[62] Barbadilla-Martinez L, Klaassen N, Van Steensel B, et al. Predicting gene expression from DNA sequence using deep learning models. Nat Rev Genet, 2025, 26: 666-80
[63] Avsec Z, Agarwal V, Visentin D, et al. Effective gene expression prediction from sequence by integrating long-range interactions. Nat Methods, 2021, 18: 1196-203
[64] Avsec Ž, Latysheva N, Cheng J, et al. AlphaGenome: advancing regulatory variant effect prediction with a unified DNA sequence model. bioRxiv, 2025, https://doi.org/10.1101/2025.06.25.661532
[65] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. arXiv, 2015, https://doi.org/10.48550/arXiv.1505.04597
[66] Giuliano KA, Debiasio RL, Dunlay RT, et al. High-content screening: a new approach to easing key bottlenecks in the drug discovery process. SLAS Discov, 1997, 2: 249-59
[67] Mattiazzi Usaj M, Styles EB, Verster AJ, et al. High-content screening for quantitative cell biology. Trends Cell Biol, 2016, 26: 598-611
[68] Carpenter AE, Jones TR, Lamprecht MR, et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol, 2006, 7: R100
[69] Caicedo JC, Cooper S, Heigwer F, et al. Data-analysis strategies for image-based cell profiling. Nat Methods, 2017, 14: 849-63
[70] Chen C, Mat Isa NA, Liu X. A review of convolutional neural network based methods for medical image classification. Comput Biol Med, 2025, 185: 109507
[71] Horst F, Rempe M, Heine L, et al. CellViT: vision transformers for precise cell segmentation and classification. Med Image Anal, 2024, 94: 103143
[72] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale. arXiv, 2020, https://doi.org/10.48550/arXiv.2010.11929
[73] Kirillov A, Mintun E, Ravi N, et al. Segment anything[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE, 2023: 3992-4003
[74] Gamper J, Koohbanani NA, Graham S, et al. PanNuke dataset extension, insights and baselines. arXiv, 2020, https://doi.org/10.48550/arXiv.2003.10778
[75] Marks M, Israel U, Dilip R, et al. CellSAM: a foundation model for cell segmentation. Nat Methods, 2025: 2585-93
[76] Vandeloo AD, Malta NJ, Sanganeriya S, et al. SAMCell: generalized label-free biological cell segmentation with segment anything. PLoS One, 2025, 20: e0319532
[77] Archit A, Freckmann L, Nair S, et al. Segment anything for microscopy. Nat Methods, 2025, 22: 579-91
[78] Doron M, Moutakanni T, Chen ZS, et al. Unbiased single-cell morphology with self-supervised vision transformers. bioRxiv, 2023, https://doi.org/10.1101/2023.06.16.545359
[79] Kraus O, Kenyon-Dean K, Saberian S, et al. Masked autoencoders for microscopy are scalable learners of cellular biology[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2024: 11757-68
[80] Gupta A, Wefers Z, Kahnert K, et al. SubCell: vision foundation models for microscopy capture single-cell biology. bioRxiv, 2024, https://doi.org/10.1101/2024.12.06.627299
[81] Colwill K, Renewable Protein Binder Working Group, Graslund S. A roadmap to generate renewable protein binders to the human proteome. Nat Methods, 2011, 8: 551-8
[82] Blampey Q, Benkirane H, Bercovici N, et al. Novae: a graph-based foundation model for spatial transcriptomics

data. Nat Methods, 2025, 22: 2539-50
[83] Cao S, Yuan Y. stFormer: a foundation model for spatial transcriptomics. bioRxiv, 2024, https://doi.org/10.1101/2024.09.27.615337
[84] Lin Y, Luo L, Chen Y, et al. ST-Align: a multimodal foundation model for image-gene alignment in spatial transcriptomics. arXiv, 2024, https://doi.org/10.48550/arXiv.2411.16793
[85] Blampey Q, Benkirane H, Bercovici N, et al. Novae: a graph-based foundation model for spatial transcriptomics data. bioRxiv, 2024, https://doi.org/10.1101/2024.09.09.612009
[86] Wang G, Wu S, Xiong Z, et al. CROST: a comprehensive repository of spatial transcriptomics. Nucleic Acids Res, 2024, 52: D882-90
[87] Mu S, Lin S. A comprehensive survey of mixture-of-experts: algorithms, theory, and applications. arXiv, 2025, https://doi.org/10.48550/arXiv.2503.07137
[88] Radford A, Kim JW, Hallacy C, et al. Learning transferable visual models from natural language supervision. arXiv, 2021, https://doi.org/10.48550/arXiv.2103.00020
[89] Rosen Y, Roohani Y, Agrawal A, et al. Universal cell embeddings: a foundation model for cell biology. bioRxiv, 2024, https://doi.org/10.1101/2023.11.28.568918
[90] Gong J, Hao M, Cheng X, et al. xTrimoGene: an efficient and scalable representation learner for single-cell RNA-seq data. arXiv, 2023, https://doi.org/10.48550/arXiv.2311.15156
[91] Hua SBZ, Lu AX, Moses AM. CytoImageNet: a large-scale pretraining dataset for bioimage transfer learning. arXiv, 2021, https://doi.org/10.48550/2111.11646
[92] Zhao S, Luo Y, Yang G, et al. Stofm: a multi-scale foundation model for spatial transcriptomics. arXiv, 2025, https://doi.org/10.48550/2507.11588
[93] Achiam OJ, Adler S, Agarwal S, et al. GPT-4 technical report. arXiv, 2023, https://doi.org/10.48550/2303.08774
[94] Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol, 2019, 15: e8746
[95] Weinreb C, Rodriguez-Fraticelli A, Camargo FD, et al. Lineage tracing on transcriptional landscapes links state to fate during differentiation. Science, 2020, 367: eaaw3381
[96] Novakovsky G, Dexter N, Libbrecht MW, et al. Obtaining genetics insights from deep learning via explainable artificial intelligence. Nat Rev Genet, 2023, 24: 125-37
[97] Roohani Y, Huang K, Leskovec J. Predicting transcriptional outcomes of novel multigene perturbations with GEARS. Nat Biotechnol, 2024, 42: 927-35
[98] Wong DR, Hill AS, Moccia R. Simple controls exceed best deep learning algorithms and reveal foundation model effectiveness for predicting genetic perturbations. Bioinformatics, 2025, 41: btaf317
[99] Ahlmann-Eltze C, Huber W, Anders S. Deep-learning-based gene perturbation effect prediction does not yet outperform simple linear baselines. Nat Methods, 2025, 22: 1657-61
[100] Kedzierska KZ, Crawford L, Amini AP, et al. Zero-shot evaluation reveals limitations of single-cell foundation models. Genome Biol, 2025, 26: 101
[101] Hetzel L, Böhm S, Kilbertus N, et al. Predicting cellular responses to novel drug perturbations at a single-cell resolution. arXiv, 2022, https://doi.org/10.48550/arXiv.2204.13545
[102] Luecken MD, Buttner M, Chaichoompu K, et al. Benchmarking atlas-level data integration in single-cell genomics. Nat Methods, 2022, 19: 41-50
[103] Schölkopf B, Locatello F, Bauer S, et al. Toward causal representation learning. Proc IEEE, 2021, 109: 612-34
[104] Argelaguet R, Cuomo ASE, Stegle O, et al. Computational principles and challenges in single-cell data integration. Nat Biotechnol, 2021, 39: 1202-15
[105] Sebe-Pedros A, Ballare C, Parra-Acero H, et al. The dynamic regulatory genome of *Capsaspora* and the origin of animal multicellularity. Cell, 2016, 165: 1224-37
[106] Bronstein MM, Bruna J, Lecun Y, et al. Geometric deep learning: going beyond Euclidean data. IEEE Signal Process Mag, 2016, 34: 18-42
[107] Wei J, Wang X, Schuurmans D, et al. Chain of thought prompting elicits reasoning in large language models. arXiv, 2022, https://doi.org/10.48550/arXiv.2201.11903
[108] Ouyang L, Wu J, Jiang X, et al. Training language models to follow instructions with human feedback. arXiv, 2022, https://doi.org/10.48550/arXiv.2203.02155
[109] Thirunavukarasu AJ, Ting DSJ, Elangovan K, et al. Large language models in medicine. Nat Med, 2023, 29: 1930-40
[110] Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. Nature, 2023, 620: 172-80
[111] Wei J, Yang Y, Zhang X, et al. From AI for science to agentic science: a survey on autonomous scientific discovery. arXiv, 2025, https://doi.org/10.48550/arXiv.2508.14111
[112] Boiko DA, Macknight R, Kline B, et al. Autonomous chemical research with large language models. Nature, 2023, 624: 570-8