

科技新闻

·深度报道·

寻找细胞模拟的“AlphaFold时刻”： AI基础模型的突围与瓶颈

当AI细胞模型照进现实，生物医学领域即将迎来范式革命

文/Mitch Leslie

人类细胞是容纳数万亿分子的微观宇宙：约4200万种蛋白质与浩如烟海的碳水化合物、脂类、核酸共同构成生命基质。

细胞系统的复杂性已超越人类当前的认知极限。但许多学者坚信，凭借人工智能卓越的信息处理能力，或许能攻克这一难题。早在20多年前，科学家就已尝试建立方程系统模拟细胞活动机制；如今，研究已进阶至AI驱动的数字孪生细胞，这些系统通过吞噬海量数据实现自我演进。3年前ChatGPT的横空出世，为虚拟细胞研究人员注入强大动力。艾伦研究所AI科学家Kasia Kedzierska指出：“整个生物学界都在期待属于我们的‘ChatGPT时刻’”。

关于实现路径的时间表，学界尚未达成共识。高保真虚拟细胞的出现将为诸多领域带来革命性变革：制药企业能借助其快速筛选候选化合物，大幅降低研发成本与周期；它们可作为合成生物学的新型试验场，助力设计具有新功能的工程化细胞；基于患者个体分子特征定制的虚拟细胞，将为临床精准用药提供决策支持；研究人员还可将单元模型

整合为虚拟组织器官，深入探究肿瘤微环境等复杂生理机制。

陈-扎克伯格倡议组织(CZI)科学AI负责人Theofanis Karalatsos强调，面对分子数据库中海量异构信息的持续涌入，此类模型更具整合价值。他指出，AI赋能的细胞模拟“正在构建生命科学的统一认知图谱”。

与ChatGPT等生成式AI类似，AI细胞模型在诞生之初就承载着宏大愿景。弧光研究所计算生物学家Hani Goodarzi评论道：“每个新模型亮相时，总被塑造成划时代突破”，他本人也参与模型开发工作。2025年6月，他联合20余位科学家发起“虚拟细胞挑战赛”，旨在通过年度竞赛持续推动技术迭代。这项赛事借鉴了1994年启动的蛋白质结构预测竞赛的成功经验（该竞赛曾加速解决蛋白质折叠难题），试图在更复杂的细胞模拟领域建立突破机制。首届竞赛聚焦人类胚胎干细胞基因功能研究，要求参赛AI系统预测特定基因沉默后的细胞表型变化。

然而，并非所有科学家都对这些模型的前景持乐观态度。即便在竞赛中表现亮眼，这些模型

距离真正推动科学突破、赋能生物学研究，仍有漫长的路要走。微软研究院的Alex Lu表示：“虽然宣传热度很高，现有模型性能还未达到预期”。Alex Lu的研究聚焦于利用AI从庞大的生物数据中识别模式。事实上，有些模型的预测能力甚至不及简化版的模拟工具。

模型开发者则认为成功只是时间问题。在计算能力飞速发展的推动下，约25年前科学家开启了虚拟细胞构建的探索。早期研究采用基于复杂方程组的计算方法，试图重现代谢网络、蛋白质生物合成、DNA复制等核心细胞活动。

2012年，现任职于西奈山医学院的计算系统生物学家Jonathan Karr，与斯坦福大学Markus Covert实验室团队联合推出史上首个全细胞计算模型——生殖支原体的数字化孪生体。选择该物种的原因在于其具备当时已知最小细菌基因组（仅含500多个基因，显著少于常见大肠杆菌所含的4000多个基因）。该模型通过模拟1100个生化反应网络，动态计算700余种代谢物浓度变化，并整合染色体结构与核糖体等蛋白合成机器，初步实现了对真实细胞内部架构的数字重构。

2022年，伊利诺伊大学厄巴纳-香槟分校细胞生物物理学家Zan Luthey-Schulten团队发布的模型在基因数量上虽略有精简，但实现了技术跨越。该模型以实验室人工构建的极简细菌（仅保留493个必需基因）为原型，创新性地引入冷冻电子断层扫描实测数据建立三维模型，能模拟细胞

20 min 的生命进程。

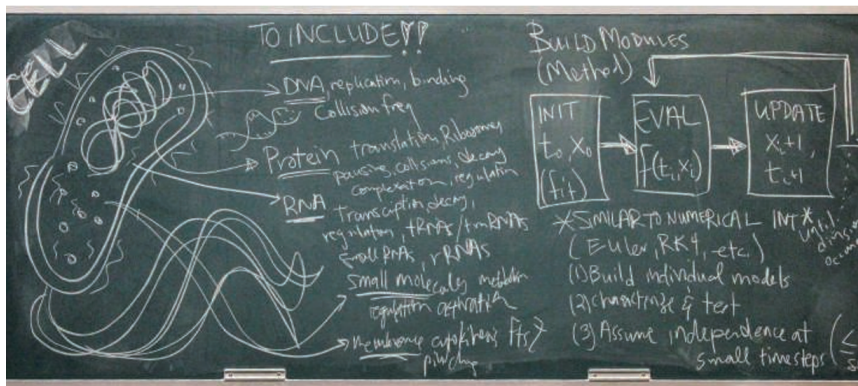
科研团队已相继开发出包括大肠杆菌在内的多种细菌全细胞计算模型。运行这些模型时,研究人员能以秒级精度模拟细胞在代谢活动、生长增殖与分裂过程中的完整生命周期。Karr 表示:“这些模型的突出优势在于预测短期细胞动力学行为,例如特定生物分子的浓度动态变化”。

但王波表示,传统计算模型仅能执行预设的生物规则,其底层数学方程对真实生物过程的还原度尚待验证。慕尼黑工业大学计算生物学家 Fabian Theis 进一步指出,现有模型难以解决许多关键科学问题:“在基因沉默效应评估等某些预测任务中,现有模型可谓完全失效”。

虚拟细胞挑战赛的设立,体现了科学界对新一代 AI 驱动模型实现突破的深切期待。与传统依赖固定方程的计算模型不同,这些新型 AI 通过分析海量实验数据,自主归纳细胞运行的内在规律。斯坦福大学生物物理学家 Steve Quake 在介绍其与 CZI 共同研发的 TranscriptFormer 模型时强调:“我们并未预设任何生物学规则”。

这些 AI 模型得以充分利用单细胞测序等技术产生的基因表达海量数据。以 CZI 建设的 CELLxGENE 数据库为例,该平台已整合超过 3500 万个人类及小鼠细胞的基因活动图谱。

与 GPT-5、Claude、Gemini 等通用人工智能系统相似,目前已公布的 AI 细胞项目大多采用基础模型架构,原则上无需重新训



Markus Covert 实验室构建的首个全细胞计算模型的技术框架(2012年拍摄)

(图片来源:《Science》)

练即可适应多种细胞过程研究。尽管现有 AI 细胞基础模型仅能理解部分生物学机制,执行细胞类型鉴定、基因活动扰动预测等有限功能,但科学家期望通过系统整合实现更大突破。

作为早期细胞基础模型的代表, Geneformer 由加州大学旧金山分校计算生物学家 Christina Theodoris 团队于 2023 年 5 月在《Nature》上提出。该研究旨在突破此前 AI 模型的局限——原有模型仅具备基因互作网络解析与调控效应预测的基础功能。Theodoris 团队曾在 2021 年《Science》发表重要成果:通过早期模型识别出可能抑制心脏瓣膜异常钙化的活性分子,该分子目前已成为多个研究团队的重点候选药物。但这项突破需要研究人员手动从实验数据构建基因网络供模型使用。

该团队基于来自近 3000 万个人类细胞的基因表达图谱对 Geneformer 进行预训练,使其获得基因互作关系,推断与遗传扰动效应预测的能力。随后采用健康人群与心肌病患者(以心肌收

缩功能衰退为特征)的心脏细胞转录组数据对模型进行微调,最终模型精准定位了多个关键基因——这些基因的失活操作可引导病变心肌细胞向正常表型逆转。

Geneformer 团队在验证环节取得了突破。他们通过 CRISPR 基因编辑技术,在体外培养的心肌病模型细胞中精准敲除 2 个预测基因,实验证实抑制任一基因均可增强心肌细胞收缩功能,为相关靶向药物研发提供了新思路。

当前多数 AI 细胞模型(包括 Geneformer)仅采用人类数据训练,仅有少数整合小鼠数据。而 Quake 与 Karaletsos 团队今年在 bioRxiv 发表的 TranscriptFormer 模型展现出更广阔的进化视角。该研究构建的 3 个版本中,最优版本整合了来自人类、小鼠、线虫、海绵及疟原虫等 12 个物种的 1.12 亿个细胞基因表达数据。

这种多元化训练策略显著提升了模型性能。研究数据显示,在罕见细胞类型鉴定任务中, TranscriptFormer 超越 2024 年王波团队发布的 scGPT 模型与

Geneformer模型。更令人惊喜的是,该模型对训练集外物种的细胞仍具备分类能力,可准确识别SARS-CoV-2感染的肺细胞,并预测药物细胞效应。Karaletsos指出,这些突破性表现均在没有额外训练的情况下实现,完美展现了“零样本学习”能力。

研究人员还可将细胞基础模型整合至“AI科学家”,该智能研究助手具备文献挖掘、假设生成、模拟实验与论文撰写的能力。哈佛大学计算生物学家Marinka Zitnik团队今年发布的600余种科研工具集中,TranscriptFormer已成为核心组件之一。

那么,究竟是什么在拖慢虚拟细胞的进展。首先,现有AI模型大多基于基因活性数据这一单一信息维度。马勤表示,科研人员正尝试引入细胞影像等多模态数据,以构建更强大的模型体系。Karr表示,虽然已出现CELLx-GENE等公共数据库,许多研究成果却依然未能进入统一的数据库,其他科学家难以便捷地获取并利用这些信息。此外,Karr认为该领域亟需建立类似蛋白质结构数据库(PDB)的标准化存储体系。这在很大程度上得益于多数资助机构与学术期刊要求研究人员必须将新解析出的结构提交至数据库,而细胞数据不仅缺乏此类政策支持,其复杂度也更高。正如Karr所言:“没有任何单一技术能够完整刻画细胞特征”。

为了改进现有模型,研究人员仍需找到一种可靠的方法来评估模型的性能,然而科学界对于评估方式尚未达成一致。当前开

发者在模型完成预训练后,通常会对模型进行进一步训练,以提升特定能力,然后再进行测试,这种实践与GPT-5等通用基础模型的开发范式一脉相承,即在向公众开放前需经历大量微调。但一些研究人员坚持认为,评估应该聚焦模型的零样本推理能力,如Theis表示:“如果它们真正掌握了生物学底层逻辑,这种认知应该在微调前就得以体现”。

近期多项评估研究却显示AI细胞模型的零样本表现不尽如人意。Alex Lu、Kedzierska及其团队将scGPT和Geneformer与简易AI模型及传统算法进行对比测试。令人意外的是,在细胞分类等任务中,更简单的方法表现优于两种基础模型,这一研究结果已于今年早些时候发表于《Genome Biology》。

Theodoris对此提出反驳,指出该研究为迁就传统方法的局限性而降低了任务难度,导致无法充分展现Geneformer的潜力。

更令人意外的是,即便经过微调,某些细胞AI模型依然难以比肩传统方法。伦敦大学学院Constantin Ahlmann-Eltze及其团队在2025年8月发表于《Nature Methods》的研究中报道,他们要求模型预测当某些基因活性被改变时,其他基因会如何响应。结果,一条基础的数学公式预测精度竟显著优于5种细胞基础模型和2种AI模型。Ahlmann-Eltze表示:“这些模型在论文中的表现原本令人印象深刻,实际对比结果却出乎意料”。

尽管现有研究仅涵盖有限模

型且标准不一,但Kedzierska认为领域内正逐步形成对模型功能与评估体系的共识。她特别强调CZI今年推出的基准测试工具,该工具首次实现了跨模型性能标准化比对。

Goodarzi团队发起的虚拟细胞挑战赛同样以此为目标。为了构建实验数据集,弧光研究所的研究人员利用CRISPR技术,在人类胚胎干细胞中逐一敲除300个基因,并追踪其对其他基因的影响。参赛团队需要用其中150个基因的数据训练模型,并可补充其他来源的同类数据。初赛要求模型预测另一组50个基因的沉默效应,结果将显示在排行榜上,为参赛者提供模型迭代的机会。最终排名将基于模型对第3组100个基因数据的预测表现。Goodarzi直言这项挑战难度极高:“我们预计不会出现压倒性胜利”。

Theis也指出,尽管已有竞赛尝试评估AI对细胞变化的预测能力,但“大规模竞赛的出现令人振奋,我们需要更完善的评估机制”。

比赛结果将在2025年12月公布,届时将为提升细胞基础模型性能提供宝贵指引。曾质疑现有模型的科学家也对未来充满信心。Kedzierska表示:“我不会放弃这些模型,每一次失败,其实都在提供线索,我们终将构建出真正优秀的模型”。

Karaletsos对此表示赞同,坦言当前TranscriptFormer仅仅是“智能模型演进之路的起点”。

(译自《Science》,2025,390(6772))

(科技新闻栏目编辑 黄文光)