

生命科学
全流程智能化引领者

A Pioneer in End-to-end
Intelligent Solutions in Life Sciences



BioHandler
汉赞迪科技



ProX 比邻星创投
VENTURES



智能技术 在合成生物学中的应用

*Application of Smart Technology
in Synthetic Biology*

● 白皮书



2023.08

前言：

合成生物学目前以势不可挡的趋势席卷各个产业，根据 BCC research 的数据估算，全球合成生物学市场预计将从 2021 年的 95 亿美元增长到 2026 年的 332 亿美元，在 2021-2026 年的预测期内，复合年增长率(CAGR)为 28.4%。政策机遇、市场机遇和资本激励的到来，合成生物学正在进入一个快速发展的新阶段。

随着合成生物学涉及的功能和潜在应用的不断拓展，运用合成生物学的复杂性和跨学科知识需求也在迅速增长。然而，生命系统极其精密，具有高度动态、灵活、非线性、不可预测等特征，如何对其进行精准预测和设计是合成生物学面临的核心挑战。

本文将简述合成生物学行业简介、智能化在合成生物学中的应用及展望，期望帮助研究人员高效率、低成本、多循环地完成“设计-构建-测试-学习”（DBTL），从而将合成生物各个环节的工作效率提升，成本降低，缩短研发周期并扩大研发可能。

目录

1 第三次生物技术革命：合成生物学简介	3
1.1 定义和发展历史	3
1.1.1 合成生物学定义	3
1.1.2 DBTL 循环	4
1.1.3 底层技术	5
1.1.4 合成生物学发展历程	8
1.2 政策与资本驱动加速合成生物学产业发展	9
1.2.1 政策端	9
1.2.2 资本端	10
1.3 合成生物学产业链剖析	13
1.4 他山之石：Ginkgo Bioworks 生物铸造厂模式	15
2 智能化加速合成生物学研究	18
2.1 合成生物学的智能化 - 加速科学 DBTL 循环	18
2.2 智能化在合成生物学中的应用 - 机器学习与高通量自动化平台介绍	19
2.2.1 以蛋白质定向进化过程示意机器学习与自动化平台结合的优点	20
2.2.2 SDLs 或自主实验-自动化平台趋势	26
2.3 海外合成生物学智能化应用发展现状 - 落地重点案例介绍	27
2.3.1 使用生物铸造厂（Biofoundry）全自动合成单转录物 TALENs	29
2.3.2 全自动算法驱动平台 BioAutomata	31
2.4 国内合成生物学智能化应用发展现状 - 国内落地的重点案例	33
2.4.1 各模块设备集成及功能	34
2.4.2 自动化方案的制定	35
2.4.3 酶的定向进化及筛选	37
2.4.4 iBioFoundry 建设中几点经验	38
3 未来合成生物学的智能化展望	40
3.1 机器学习方面	40
3.2 SDLs 自动化-自动实验室方面	41
3.3 实现 SDLs 的所需因素	43
4 参考文献	44

1 第三次生物技术革命：合成生物学简介

1.1 定义和发展历史

1.1.1 合成生物学定义

合成生物学（Synthetic Biology）是一个专注于生命系统和生物体的跨学科科学领域，汇聚并融合了生命科学、工程学、物理、化学和信息科学等诸多学科，是现代生物学最具发展潜力的领域之一。合成生物学的核心在于设计、改造、重建或制造生物元件、生物反应系统、代谢途径与过程，乃至创造具有生命活动能力的细胞和生物个体。这种以合成生物为工具进行物质加工与合成的生产方式，有望彻底变革未来医药、化工、食品、能源、材料、农业等传统行业，被誉为继 DNA 双螺旋结构发现（1953 年）和人类基因组测序（2003 年）之后的“第三次生物技术革命”。

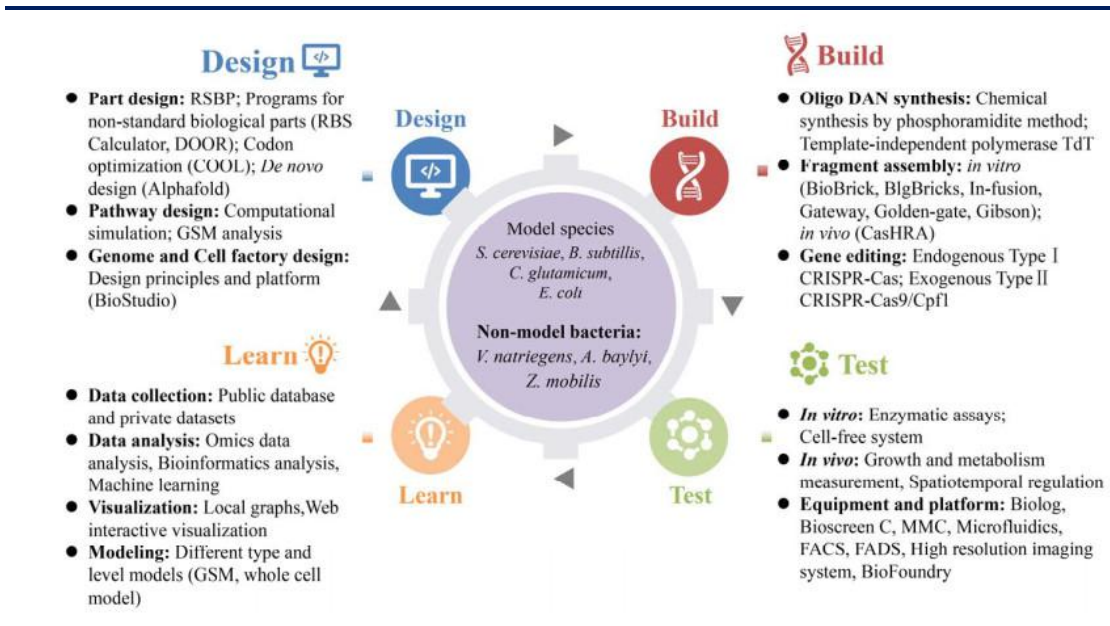
根据美国国家卫生研究院（National Institutes of Health, NIH）国家人类基因组研究所（National Genome Research Institute）所提供的定义：合成生物学是一门涉及生物改造的科学，使其赋有新能力，如合成特定物质、侦测环境等。广义来说，合成生物学结合了分子生物学、系统生物学，再透过生物工程的原理重新为生物系统设计新功能，让研究人员能够任意研究、改变、复制、甚至创造出复杂的生物路径、DNA 基因序列与生物系统，进而为医学、制造业及农业等方面提供有效的解决方案。

合成生物学的产生和快速发展，既是人类对生命现象深刻探索和系统认知之后合乎逻辑的必然结果，也是 20 世纪末和 21 世纪初学科交叉带来的技术与思维创新促使科学前沿迅速萌生的必然结果。合成生物学的核心科学基础，是它的工程科学内涵；但在一定意义上，它又是生物技术在基因组和系统生物学时代的延伸以及质的飞跃。

1.1.2 DBTL 循环

合成生物学的目的是设计符合标准的生物系统，而生命系统具有高度的复杂性，人工设计的基因线路需要海量工程化试错实验去实现预设功能，基于工程设计原则利用工程可预测性控制复杂系统构建的“设计-构建-测试-学习”循环（DBTL）逐渐成为合成生物学的核心策略。在生物制造领域，DBTL 循环四个阶段循环往复可以成功构建需要的细胞，生产出合适的产品。

图表 1. 合成生物学“设计-构建-测试-学习”研究策略涉及的主要关键技术



资料来源：文献《合成生物学时代基于非模式细菌的工业底盘细胞研究现状与展望》

1. 设计 (Design)

设计是合成生物学 DBTL 策略的基础，即在前期已有知识的基础上遵循一定的规则，利用现有的标准化生物元件对基因、代谢通路或基因组进行理性设计。其中标准化的生物元件，尤其是 DNA 水平描述的生物元件，在设计的过程中尤为重要。关键技术包括对生物元件、代谢通路、基因组及细胞工厂设计的软件及计算机模拟分析。

2. 构建 (Build)

构建过程包括 DNA 合成、大片段组装以及基因编辑。DNA 合成技术在合成生物学的发展过程中起着十分重要的支撑作用，其不同于体内扩增，不需要使用模板，可直接根据 DNA 序列进行化学合成。针对 DNA 大片段的拼接和组装，研究人员也开发了 BioBrick、BglBricks、In-Fusion、Gateway、Golden Gate、Start-Stop 及寡核苷酸连接介导的 DNA 组装(Oligonucleotides linkers mediated DNA assembly, OLMA)等多种组装方法。基因编辑技术从第一代锌指核酸酶(Zinc finger nucleases, ZFN)、第二代类转录激活因子效应物核酸酶 (TALEN) 发展到目前的第三代 CRISPR-Cas 技术。CRISPR-Cas 技术具有有效

率高、操作快捷、效果准确等优点，是目前基因编辑的主流技术，尤其是 II 型的 CRISPR-Cas9 或 CRISPR-Cas12a (Cpf1) 基因编辑技术，已在多种生物体系中得到广泛应用。在不同的工业底盘细胞中使用的构建方法存在一定差异性。

3. 测试 (Test)

无论是酶、报告基因或启动子、RBS 等单个生物元件，还是逻辑线路及模块化的代谢途径，在通过理性或非理性设计后，都会存在大量的突变体或候选目标，因此高效、准确和经济的检测方法对最佳生物元件及组合的选择至关重要。传统的检测方法无法满足合成生物学对大量定量化生物元件、逻辑线路及代谢与调控途径组合的需求，目前已经尝试开发利用多种高通量或自动化的筛选与检测技术来提高测试的效率。对于工业底盘细胞而言，使用高通量筛选技术快速筛选到相关工业性能提升的细胞工厂至关重要。

除了对已构建的细胞工厂进行生长测试外，对合成生物学 DBTL 整个闭环的试错性检测也十分重要。快速高效的自动化和工程化平台将是有效的解决手段。工程化平台可以快速大量试错，高效验证 DBTL 过程并快速积累数据和经验。目前包括美国 and 英国等多个国家已构建或在建的重大工程化平台约有 25 个，我国的天工所、国家蛋白质科学中心(上海)、武汉生物研究院及江南大学等研究机构，也都已构建了相关合成生物学自动化平台。目前天工所的平台已经实现了自动化单基因克隆，通量达到 300–600 个/d，多模块化复杂质粒组装也达到 100 个/d 的通量，组装正确率 > 90%，同时还实现了多种模式生物的自动化基因操作。

4. 学习 (Learning)

学习过程作为合成生物学 DBTL 中的重要一环，能为下一个循环改进设计提供指导，如基于系统生物学方法的组学技术进行“基因-RNA-蛋白-代谢-表型”不同层面的分析，构建基因型-表型和代谢调控网络的知识图谱等。学习这一过程涉及数据收集整合、数据分析、结果可视化和建模分析等。虽然由于研究积累的原因，非模式微生物相关的数据远少于模式微生物，但模式微生物相关数据的处理学习，也可以为非模式微生物提供很好的借鉴。收集整理的大量数据可以利用生物信息学和人工智能机器学习等相关技术进行分析以及构建数学模型，如基因组尺度的代谢网络模型和全细胞模型等。同时，这些数据也给依赖大数据的机器学习尤其是深度学习提供了可能。“学习”这一阶段总结的规律与结论，可以指导合成生物学 DBTL 其他阶段的模块，优化建立更为高效精简的合成生物学工作流程，用于不同底盘细胞的构建和性能优化。

1.1.3 底层技术

合成生物学的“工程学本质”在基因层、生命层和产业层三个层面实现。

- 基因层：基因组计划破译基因信息，使得元件标准化；使用基因编辑技术进行精确地增/删/改基因。
- 生命层：将预先设计的基因导入生命体，快速繁殖和培养出所需要的新物种。

- 产业层：不断实现更长的反应路径、更复杂的代谢调控、更高的生产效率。

合成生物学从概念向产业的转变，最主要得益于底层技术的创新。在 DBTL 的核心策略下，合成生物学的底层关键技术包括 DNA 测序、DNA 合成、基因编辑、DNA 组装，支撑了合成生物学产业的迭代与升级。

1. DNA 测序

基因测序指通过测序设备分析生物样本的基因组信息。高效、低成本的 DNA 测序是实现合成生物学快速发展的基础。快速、廉价和可靠的大规模基因组测序可以：

- 1) 提供有关自然界生物的信息，帮助合成生物学家从中寻找构建生物元件和装置的基因信息；
- 2) 验证制造的系统是否符合预期；
- 3) 促进构建的生命系统的快速检测和鉴定。从最初的 Sanger 测序发展到第四代纳米孔测序，DNA 测序技术不断迭代，测序成本下降显著，测试长度、速度均得到指数级提升。

2. DNA 合成

受制于技术限制，传统基因工程技术所操控的基因其实也是来自自然界，只是实现了物种、个体之间的跨越，目标基因及其产物的适用性有明显的局限。长链 DNA 合成技术的突破，基因可更根据需求主动合成，合成生物学才具备了创造非自然基因、获得新产物的可能。21 世纪以来，大规模、高精度、低成本的 DNA 合成技术发展推动了合成生物学的效率提升。

图表 2. DNA 合成技术特点对比

技术分类	技术原理	技术特点	发展限制因素
柱合成法	磷酸酰胺合成法	错误率低	大规模合成成本较高，通量小，合成中使用到有毒试剂
芯片合成法	喷墨法	品质较高、长链合成	合成长度短，错误率较高，单序列合成产量低、不利于组装拼接
	光化学法	品质一般，引物短	
	电化学法	品质较高、长链合成	
酶促合成技术	微阵列法	成本优势、试剂消耗小	早期商业化阶段
	酵母内 DNA 合成法	实现酵母体内合成	
	链接介导 DNA 合成法	简单易用、DNA 突变率较低	

资料来源：比邻星分析

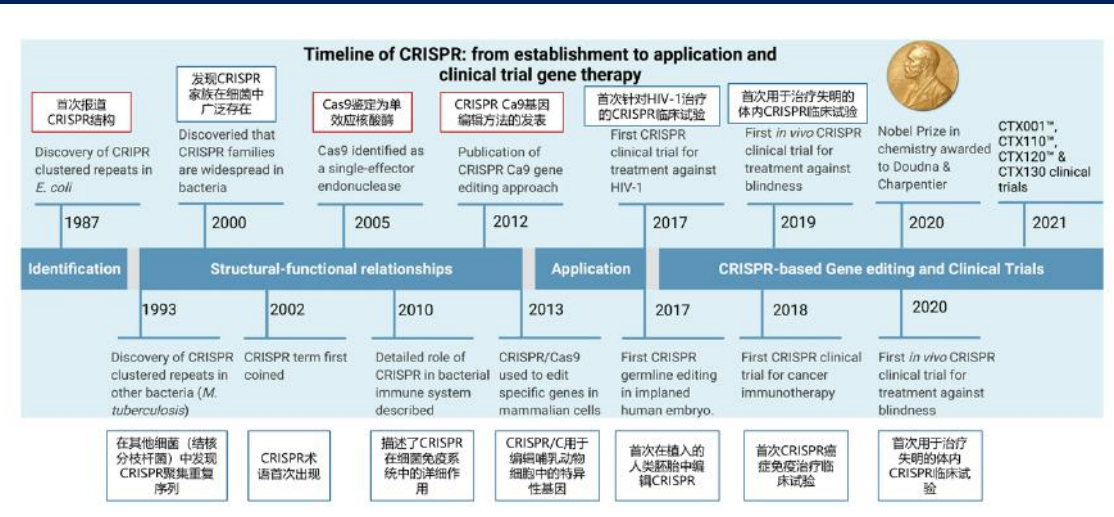
3. 基因编辑

基因组编辑技术，是指一种对目标基因进行编辑或修饰的基因工程技术，主要有以下 3 种技术：锌指蛋白核酸酶（ZFN）、类转录激活因子效应物核酸

酶（TALEN）以及 CRISPR/Cas 系统。基因编辑技术能够对合成生物的特定 DNA 序列进行改造。前两代技术构建周期长，步骤繁琐，难以进行高通量基因编辑，极大限制了其推广应用。第三代 CRISPR/Cas9 技术优势显著：

- 1) 构建简单方便快捷，适用于任何分子生物实验室；
 - 2) 用于基因组的点突变编辑优于 ZFN 或 TALEN；
 - 3) CRISPR/Cas9 精确的切口酶活性用于基因治疗安全性高于 ZFN 或 TALEN。
- 不过，大多数底盘细胞 CRISPR-Cas 技术的使用需要外源引入 Cas 蛋白，会引起细胞毒性，从而限制了 CRISPR 系统的应用。为了解决这个问题，研究人员在多个微生物中开发出了基于微生物自身内源 CRISPR-Cas 系统的基因组编辑技术。

图表 3. CRISPR 技术发展里程碑时间线



资料来源：文献《Strategies to overcome the main challenges of the use of CRISPR/Cas9 as a replacement for cancer therapy》

4. DNA 组装

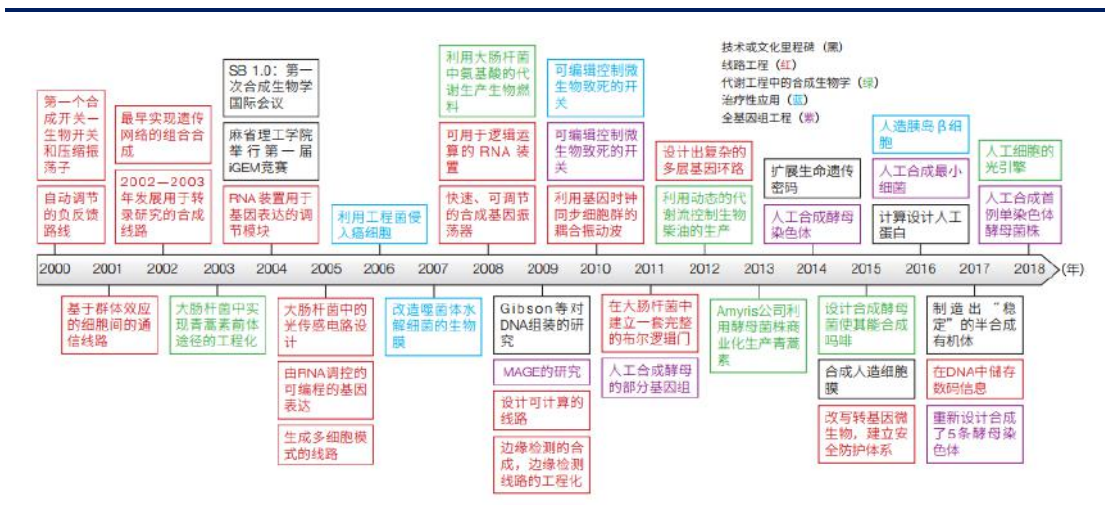
受限于 DNA 合成技术目前还无法直接准确地合成基因长度的 DNA 片段，DNA 片段还依赖于组装技术进行拼接。DNA 组装是利用分级的体外与体内组装技术的合理搭配，将分段合成的寡核苷酸片段装配成长片段 DNA，达到基因长度甚至基因组长度 DNA 序列的合成。常用的寡核苷酸组装方法有两种：连接酶组装法（ligase chain reaction, LCR）和聚合酶组装法（polymerase cycling assembly, PCA）。使用短初始片段组装染色体或基因组长度 DNA 所需的分层组装次数较多，过程中所需的克隆挑选和测序等质控成本也会相应增多，因此，低成本、自动化、一体化是 DNA 组装技术的未来发展方向。

1.1.4 合成生物学发展历程

合成生物学在本世纪初兴起，但其起源和发展由来已久。1910 年，法国科学家 Stephane Leduc 首先提出 synthetic biology，即合成生物学一词，但其含义与当前的合成生物学意义相差甚远。2000 年，基因开关及压缩震荡子两篇文章发表在学术期刊《Nature》上，常被认为拉开了合成生物学的时代序幕。同时，合成生物学的发展离不开上述底层技术的支持，进入二十一世纪以来，DNA 合成、基因编辑、基因测序等技术支撑着包括线路工程、代谢工程、治疗应用等一系列技术上的首次突破，基础研究持续向前推进。回顾合成生物学的发展历程，合成生物学行业可分为四个阶段：基础研究萌芽期、基础研究成熟期、应用开发期以及产业投资期，各期间合成生物学的技术难题不断突破，应用范围持续拓展。预期未来 10 年将是合成生物研究落地的 10 年。

- 基础研究萌芽期（2003 年以前）：技术发展初期，产生了许多具备领域特征的研究手段和理论，特别是基因线路工程的建立及其在代谢工程中的成功运用，代表成果为青蒿素前体在大肠杆菌中的合成。
- 基础研究成熟期（2004 - 2007 年）：领域有扩大趋势，基础研究进展快速，但是工程技术进步较缓慢，总体上处于工程化理念日渐深入、使能技术平台得到重视、工程方法和工具不断积淀的阶段，体现出“工程生物学”的早期发展特点。
- 应用开发期（2008 - 2013 年）：随着基因编辑、基因测序等领域涌现出的新技术，工程手段进一步成熟，合成生物学下游应用领域也开始不断拓展，合成生物学技术关注度日益提升，代表成果为 Amyris 公司利用酵母菌株商业化生产青蒿素。
- 产业投资期（2014 年至今）：合成生物学“设计-构建-测试-学习”等概念提出，工程化平台的建设和生物大数据的开源应用相结合，多学科融合程度进一步加深，资本加速入场，行业进入加速发展期。随着全球合成生物学的发展浪潮，我国合成生物学也同时进入飞速发展阶段。

图表 4. 合成生物学领域发展的代表研究成果



资料来源：文献《合成生物学：开启生命科学“会聚”研究新时代》

中国合成生物学虽起步较晚，但发展迅速，具备产业优势。我国在合成生物学关键核心技术领域论文发表和专利申请数量位居全球第二，尤其是下游应用端已处于领先地位。从产品优势看，与欧美等国家相比，国内发展合成生物学具备制造应用优势与成本优势，不少产品已占据全球原料市场排行榜前列。目前，国内最多的合成生物学企业汇聚长三角、京津冀、粤港澳，不少企业已经发展成为细分领域行业龙头。

1.2 政策与资本驱动加速合成生物学产业发展

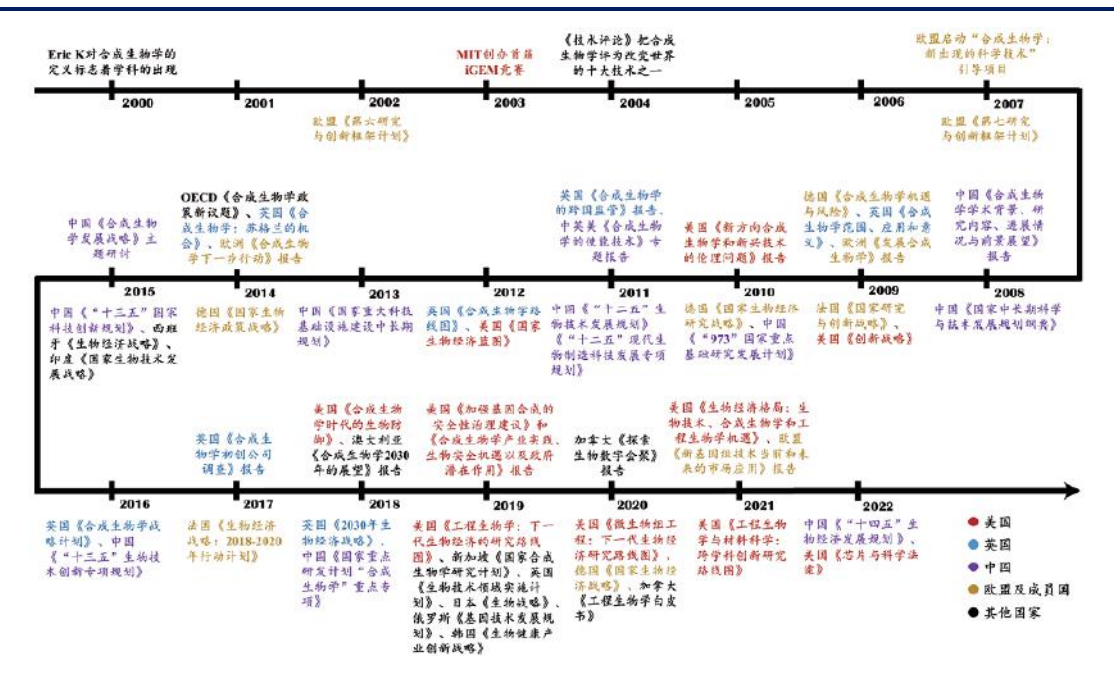
1.2.1 政策端

合成生物学产业具有清洁、高效、可再生等特点，能够减少工业经济活动对生态环境的影响，有望在医药、食品、能源、材料、化工、农业等众多领域引发生产制造模式的变革，因此备受政策支持。目前，合成生物学已进入全球共识、合作与竞争的快速发展时期，欧盟、美国、中国等国家/地区从学科发展、政策制定和战略布局等多维度促进合成生物学发展。

近年来，全球主要国家政府陆续出台合成生物学相关扶持政策，国际合成生物学科研究和产业发展十分迅猛。全球主要国家相继建立合成生物研究中心，形成了遍布全球的合成生物研究网络，以美国、英国为主导的发达国家在合成生物研究领域发展进程较快。欧盟最早通过第六研究框架计划从政策层面、以项目资助的方式促进合成生物学发展，法国、德国等成员国针对合成生物学及相关技术分别制定了针对本国的研究发展战略。英国政府于 2012 年和 2016 年相继发布《合成生物学路线图》和《英国合成生物学战略计划》，是首个在国家层面通过路线图方式推动合成生物学发展的国家。美国从多个维度来推动合成生物学的发展，自 2019 年开始连续 3 年发布了《工程生物学：下一代生物经济的研究路线图》《微生物组工程：下一代生物经济研究路线图》《工程生物学与材料科学：跨学科创新研究路线图》等合成生物学相关领域的研究路线图。

合成生物学是符合“双碳”目标的创新技术行业，其发展也受到我国国家政策支持。与传统技术路线相比，合成生物制造的环保、可持续优势明显，部分生物合成产品（如 L-丙氨酸、6-APA、洛伐他汀等）已具备低碳、低能耗、低成本的优势特征。具体到政策方针层面，“973”、“863”等国家重点基础研究发展计划建立了合成生物学专项；“十二五”期间，国家提出对生物制造技术的支持；到“十三五”时期，国家将合成生物技术列为引领产业变革的颠覆性技术之一；2022 年 5 月，国家发改委印发《“十四五”生物经济发展规划》，提出在医疗健康、食品消费、绿色低碳、生物安全等重点领域发展生物经济，生物经济成为“十四五”期间推动高质量发展的强劲动力。在政策的大力支持下，合成生物产业迎来了重要的发展机遇。

图表 5. 全球合成生物学领域发展和战略布局演进路径

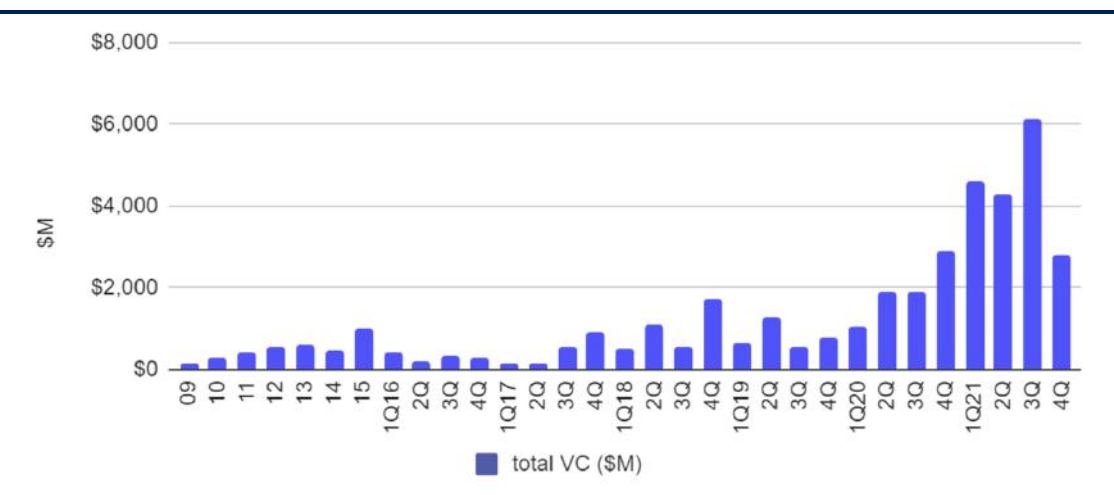


资料来源：文献《全球合成生物学发展现状及对我国的启示》

1.2.2 资本端

2021 年是合成生物学投资的重要转折点，这一年合成生物学独角兽 Zymergen 和 Ginkgo 相继上市，为后来者提供了 IPO 的参考标准。一级市场合成生物学赛道初创公司在 2021 年也迎来了最好的一年，根据 Synbiobeta 数据，2009-2020 年，全球合成生物学初创公司共计获融资额 215 亿美元，而在 2021 年达到 180 亿美元，单年获融资额占过去 12 年总融资额比例超 80%。

图表 6. 2009-2021 年全球合成生物学企业融资金额（单位：百万美元）



资料来源：Synbiobeta

2018 年以来，我国合成生物学融资不断扩大。据动脉网数据，即便是在一级市场相当低迷的 2022 年，合成生物学的热度也始终居高不下。2022 年，我国合成生物学领域共完成 54 起融资，是医疗板块融资次数最多的细分领域之一。以 2021 年为时间起点，到 2023 年初，我国合成生物学领域共完成了 103 起融资，共有多达 300 家投资机构完成加码。代表企业蓝晶微生物在 2022 年 1 月完成了金额 8.7 亿元 B3 轮融资，在 2023 年 2 月又完成了逾 4 亿元 B4 轮融资，至此公司的 B 系列融资总额已达 19 亿人民币。

投资标的方向上，国内的投资主要聚焦于生物医药领域应用和底层生物技术。融资阶段上，风口下的国内合成生物学正处于发展上升期，赛道内企业主要在成长期阶段，因此融资轮次目前集中在 A 轮及 A 轮以前。另一特点是，过去 2 年已经完成融资的合成生物学企业大都诞生于清华、西湖、中科院微生物所、深圳先进院等科研院校。在推动合成生物领域前沿成果从实验室到产业化这一阶段，国内不少科研院校已经形成了成熟经验。虽然合成生物学领域目前已诞生了凯赛生物、华恒生物等上市企业，但整体数量还相对较少。随着赛道内企业技术逻辑进一步成熟、核心产品进一步商业化落地，预期将有更多公司发展为行业独角兽。

图表 7. 2023 年合成生物学赛道融资事件（代表）

公司	融资时间	融资形式	融资规模	投资机构	公司简介
趣酶生物	2023/7/11	A 轮	近亿元	通德资本、京元通投资、厚实投资、康居创投、南岭创投、华垚股权	生物酶催化合成服务提供商，为工业客户提供“天然”生物基产品
微元合成	2023/7/06	Pre-A 轮	亿元	北京国管顺禧基金、北戴河新区高新技术产业基金、河南投资集团汇融基金	合成生物技术研发商，应用于医药、日化、农业、食品、饲料和材料等领域
微远生物	2023/6/30	种子轮	数千万元	大晶创投、藕舫天使、兴华鼎立、真石资本	以酶的智能进化和菌株的高通量筛选为技术平台
利夫生物	2023/6/25	B 轮	近 2 亿元	关子创投、中蓝创投、华盖资本和泽晖资本等	业内领先的全生物基新材料 FDCA 与下游应用产品研发和生产企业
CellX	2023/6/16	A+轮	数千万元	多个战略投资方	细胞培养肉为核心的新蛋白产品开发

昌进生物	2023/6/06	A+轮	-	益联资本、锦秋基金、盛山资本、惠远资本	食品合成生物学和微生物蛋白研发
和晨生物	2023/5/30	天使+轮	-	弘博资本，怀格资本，国元创新	专注于功能活性原料研发及产业化的合成生物学企业
周子未来	2023/4/26	A+轮	数千万元	启明创投	细胞培养肉领域领军企业
未米生物	2023/4/25	Pre-A 轮	数千万元	厚新健投	基因编辑和生物育种企业
零一生命	2023/4/20	B1 轮	亿元	松禾资本、云时资本、Green Future、旦恩资本等联合投资	人体大数据驱动的微生态高通量筛选及应用转化平台
光羽生物	2023/3/21	Pre-A 轮	近亿元	绿洲资本	光驱动合成生物
合生科技	2023/3/21	A 轮	-	蒙牛创投领投	合成生物研发
若弋生物	2023/3/18	Pre-A 轮	千万元	东久新宜资本领投	重组肉毒素药物企业
极麋生物	2023/3/8	天使+轮	千万元	十维资本	细胞培养肉企业
贻如生物	-	种子轮、天使轮	数千万元	种子轮由奇绩创坛领投；天使轮由线性资本领投	合成生物学生物材料企业
分子之心	2023/2/20	B 轮	超亿元	凯赛生物领投	AI 蛋白质设计平台企业
蓝晶微生物	2023/02/14	B4 轮	4 亿元	中平资本领投	合成生物学分子与材料创新企业
中农种源	-	种子轮	千万级	红杉中国种子基金、果壳	农业合成生物学企业
微构工场	2023/02/02	A+轮	3.95 亿元	中石油昆仑资本领投	生物降解材料 PHA、医药中间体四氢嘧啶、尼龙 56 前体戊二胺等高附加值产品研发和生产
赞倍司	2023/01/16	Pre-A 轮	数千万	-	植物基产品研发

资料来源：《合成生物学周报》

1.3 合成生物学产业链剖析

合成生物学产业链上游为合成生物使能技术（工具类公司），主要与 DNA&RNA 相关，提供包括测序、合成、基因编辑等底层工具和软件服务。中游为平台类公司，主要侧重对菌株的筛选与改造、培养成分开发等，旨在提供生物体设计和自动化等合成生物平台。下游则是应用产品型公司，面向广阔终端领域生产产品，是目前基数最大的一类型企业。

- 上游：合成生物学使能技术公司

上游的合成生物学使能技术公司提供包括测序、合成、基因编辑等底层工具和软件服务，为生物元件构建提供技术支撑。具备底层技术优势的公司和服务研发过程中积累了大量的 DNA 合成与生物元件设计方面的经验，构建的研发信息数据库能够为中游及下游企业提供更简便、准确的服务。例如，Twist Biosciences 可以根据客户选择的序列来合成 DNA；Desktop Genetics 正在构建一个用于 CRISPR 基因编辑的人工智能平台。

- 中游：合成生物学平台型公司

合成生物学产业链中游以合成生物学平台公司为主。实现菌种的选择和优化是合成生物行业发展的前提，平台型合成生物学公司专注于菌株改造、筛选等生物合成技术和工艺研发，并致力于构建自动化、工业化生产转化的合成生物学平台。例如，Ginkgo Bioworks 通过搭建高效的自动化平台，构建了微生物快速改造的技术护城河。MetaMixis 公司也设计用于发现生产各种化学品的新技术平台，使研发人员能够更快地对微生物进行编程，以生产用于香精或香料的化学物质。平台型公司通过与下游企业展开商业化合作，加速合成生物学产品产业化进程。

- 下游：合成生物学应用型公司

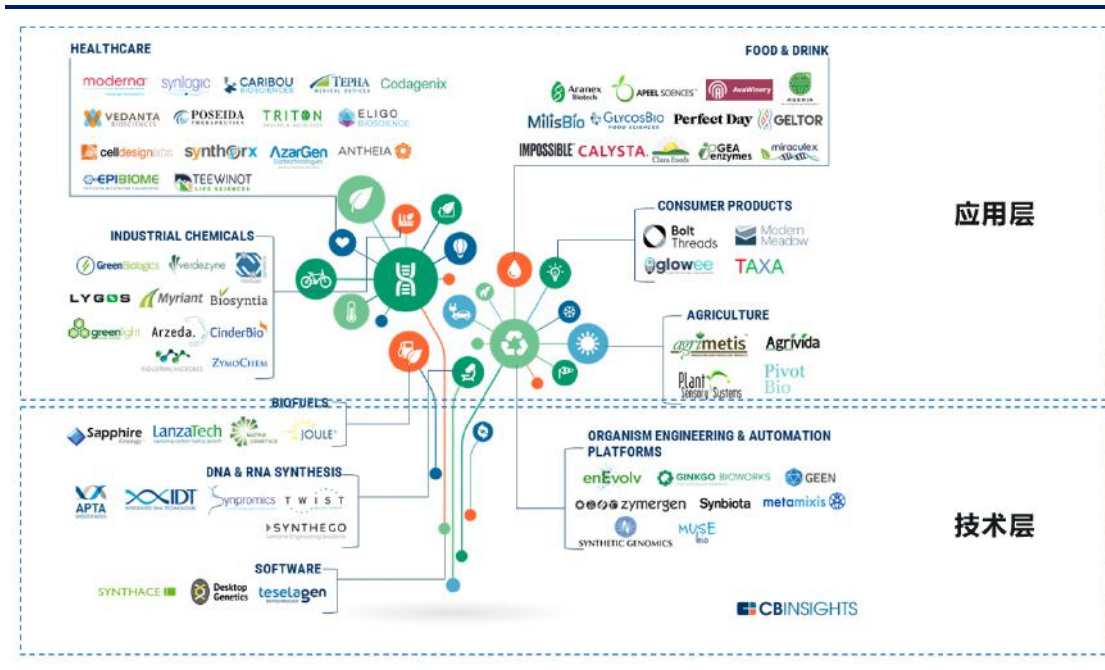
合成生物学产业链下游以产品型公司为主，主导产品的放大生产与下游市场应用。产品型公司利用合成生物学技术生产不同领域的目标产品，对传统石化基合成技术进行了补充或替代。合成生物学的应用领域丰富，市场延伸至医药健康、化工材料、能源、食品和农业等多个领域，并已在多个产品中体现出合成生物学的低碳、节能、经济的绿色产业特征。

1. 医疗保健：此类别包括将合成生物学技术应用于发现新药物、医疗设备或诊断方法的公司。例如，Poseida Therapeutics 使用基因编辑技术开发多发性骨髓瘤和前列腺癌等疾病的治疗方法。Antheia 正在改造酵母，将葡萄糖转化为阿片类药物。
2. 工业化学品：此类公司正在开发新的方法来生产用于香料、香精、药物和其他工业应用的化学品。例如，Lygos 对酵母进行了改造，使其能够从糖和二氧化碳中产生丙二酸。Green Biologics 对细菌进行工程改造以生产化学品，例如用于油漆、粘合剂、清洁剂和香料的丁醇。
3. 生物燃料：此类公司正在通过将通常无害的微生物设计到燃料生产实体中来开发

新的可再生能源。例如，LanzaTech 已使微生物能够将二氧化碳或甲烷等废气转化为燃料和化学品，已经与维珍航空合作生产低碳喷气燃料。

4. 农业：此类初创公司正在开发合成解决方案应用于植物和动物领域。例如，Agrivida 对酶进行工程改造，以提高动物营养产品的消化率和利用率。Plant Sensory Systems 对甜菜进行工程改造，以提高作物产量、抗虫性和糖产量。
5. 消费品：此类别包括开发皮革或照明结构等消费品的初创公司。例如，Modern Meadow 设计了能够产生胶原蛋白的细菌，然后将胶原蛋白捆绑在一起以生产皮革。
6. 食品和饮料：此类初创公司正在发明合成食品的生产和保存过程。例如，Ava Winery 正在通过复制流行葡萄酒的化学成分来开发不含葡萄的合成葡萄酒。对于配料，该团队还对酵母和玉米进行工程改造，以产生与葡萄酒中相同的某些化合物和氨基酸。结合起来，这些元素可以复制潜在任何葡萄酒的味道、风味和气味。

图表 8. 合成生物学商业形态与产业链上下游



资料来源：CB Insights、比邻星分析

我国合成生物学企业的产业链分布较完整，上中下游都已有企业布局。产业链上游企业主要提供合成生物学底层技术，代表企业包括：金斯瑞生物、博雅辑因、泓迅科技、擎科生物等。产业链中游企业借鉴国外的模式，通过整合科研院所以及社会各界的学术资源，通过自动化、机器学习以及大量生物数据的汇聚来提高研究的发现、通量和产量，从而构建技术研发平台，为下游企业提供基于合成生物学的解决方案，代表企业包括：恩和生物、弈柯莱生物、态创生物、惠利生物、衍进科技等。产业链下游企业专注于产品，在产品生产过

程中采用先进的生物工程技术，探索高效的生物合成方法，将合成生物制造应用于生物医药、农业、食品、化工、材料等领域，代表企业包括：川宁生物、瑞德林生物、凯赛生物、蓝晶微生物、华恒生物等。

1.4 他山之石：Ginkgo Bioworks 生物铸造厂模式

Ginkgo Bioworks 是一家全球平台型合成生物学龙头公司。2008 年，由麻省理工学院（MIT）教授 Tom Knight 和刚刚从 MIT 获得博士学位的 Jason Kelly、Reshma Shetty、Barry Canton 和 Austin Che 等 4 人联合成立。2021 年 9 月 Ginkgo Bioworks 成功上市，上市估值高达 150 亿美元。

Ginkgo Bioworks 处于合成生物学产业链的中游，并逐渐向上游拓展。公司业务以菌株改造及自动化平台为核心，连接并集成上游技术层公司提供的硬件与工具，创建平台为下游产品应用的客户编辑细胞。依托核心资产生物铸造厂 Foundry 和代码库 Codebase，主要业务分为两大部分：（1）细胞工程：采用定制软件、机器人自动化、数据分析技术，根据客户需求进行细胞编程。

（2）生物安全：提供核酸检测产品和服务，提供基因组测序和核酸疫苗生产改进服务。

目前为止，Ginkgo Bioworks 已经在多个垂直领域进行业务布局：

- 农业领域

2017 年 9 月，Ginkgo Bioworks 与 Bayer 合资成立独立运营子公司 Joyn Bio，合作开发新型固氮微生物。

2021 年 4 月，Ginkgo Bioworks 宣布与 Corteva Agriscience 达成合作，利用合成生物学开发全新的植物保护技术，目前该合作的核心内容是应对外来入侵物种和抗性有害生物。

- 食品领域

2016 年 10 月，Ginkgo Bioworks 与世界四大粮商的 Cargill 和 ADM 达成合作协议，为其定制开发微生物菌株。

2018 年 11 月，Ginkgo Bioworks 与生物技术公司 Glycosyn 宣布合作开发人乳低聚糖（HMO），作为培育健康肠道微生物群产品的核心成分。

2019 年 2 月，Ginkgo Bioworks 投入 9000 万美元成立 Motif Food Works，专注于利用微生物提供下一代蛋白质和食品成分，开发和生产不含动物成分的食品原料，以进入植物肉和替代乳制品市场。

- 医药领域

2018 年 9 月，Ginkgo Bioworks 与全球最大的工业大麻室内种植公司 Cronos Group 宣布一项金额高达 1.22 亿美元的合作计划，以生产多种独特的大麻化合物，为制药行业提供急需的治疗/医药消费品。

2019 年 6 月，Ginkgo Bioworks 与 Roche 达成新的合作协议投入 1600 万美元发现新的抗生素。

2019 年 6 月，Ginkgo Bioworks 对 Synlogic 进行 8000 万美元股权投资，加速推进工程益生菌创新疗法研发。

2020 年，Ginkgo Bioworks 宣布启动 Concentric 计划，提供大规模 COVID-19 现场测试服务。

- 环境治理领域

2020 年 10 月，Ginkgo Bioworks 投资 4000 万美金成立 Allonnia，开发能够断开化学键的微生物，以应对环境中的化学污染物。

图表 9. Ginkgo Bioworks 客户画像



资料来源：Ginkgo Bioworks、比邻星分析

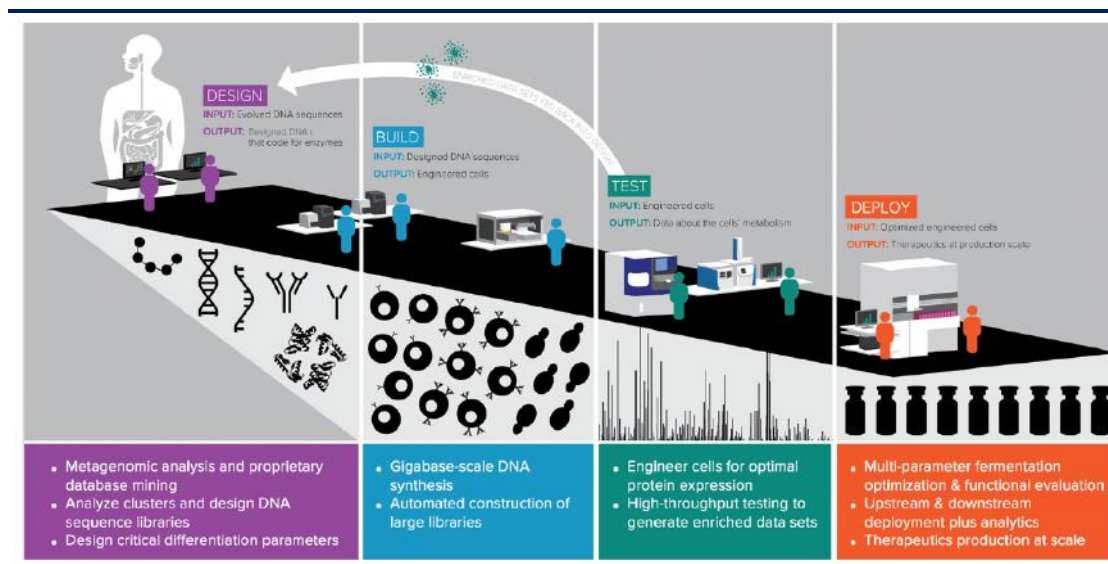
Ginkgo Bioworks 预计 2023 年总收入至少为 2.75 亿美元。其中，细胞工程收入至少为 1.75 亿美元；生物安全业务 2023 年收入至少为 1 亿美元。此外，Ginkgo 继续期待在 2023 年为其 Foundry 平台增加 100 个新的细胞编程项目（2023 年 Q1 增加了 13 个新的细胞编程项目）。

Ginkgo 商业上的成功主要基于其技术平台上的核心竞争力：

1. 公司借助自动化技术、集成软件和数据分析技术，搭建了生物铸造厂 Foundry，实现了生物工程的自动化和规模化。设计和合成目标 DNA 序列、插入细胞进行测试、筛选出最高效的菌株等操作均可在生物铸造厂高效完成。
2. 通过整合生物铸造厂实验原始数据和公开数据库，Ginkgo 形成了遗传代码库 (Codebase)，为合成生物学上游设计提供数据支撑。Codebase 库为生物学优化和设计提供了模块化的基因片段和可被重复使用的底盘菌株，并可开发新的生物学项目、测试和重新设计新的合成生物。“Foundry”和代码库“Codebase”两

大平台相辅相成，形成了 Ginkgo 在微生物快速构建上的领先优势。

图表 10. Ginkgo Bioworks 合成生物学平台：自动化的 Foundry 工作流程+独特的 Codebase 遗传库资源



资料来源：Ginkgo Bioworks、比邻星分析

2 智能化加速合成生物学研究

2.1 合成生物学的智能化 - 加速科学 DBTL 循环

近年来，基因编辑新工具和新技术取得突破性的发展，如第三代 CRISPR/Cas9 基因编辑技术以及酶定向进化等关键技术的广泛应用，彻底改变了在体内编辑 DNA 的能力，DNA 合成生产力几乎同步于摩尔定律提速，转录组学的的数据量每 7 个月可翻一番，蛋白质组学和代谢组学的高通量工作流程变得越来越可用，让我们能够以前所未有的速度对细胞进行生物工程改造。但生命体作为合成生物学的研究对象，具有高度动态、灵活、非线性、不可预测等特征，如何对其进行精准预测和设计是合成生物学面临的核心挑战。

- 机器学习提供所需的预测能力

虽然我们能够对目标细胞进行预期的 DNA 改变，但我们无法预测生物系统的 DNA 被改变后的表型。同时，在蛋白质定向进化中，许多新技术仅适用于某些特定场景，并且一个定向进化活动的成功可能很难转移到另一种目标蛋白上，导致这些新技术不具备通用性。

此外，我们缺乏从小规模实验中推断大规模行为的能力。例如在生物工程领域，一个关键瓶颈是设计能够可靠地将实验室结果（1-100 毫升）扩展到商业规模（100-10⁶升）的发酵系统。过去有大量未能从实验室走到大型商业生产的失败案例，即无法实现量产，这不仅带来了巨大的经济损失，还导致该领域后续的投资关注度大量减少。

机器学习可以在不需要详细理解机制的前提下，提供前文所述的挑战中需要的预测能力，通过机器学习模型，将训练数据的一组输入与一组输出进行统计学上的联系，可在抛开有所偏差的假设下，表达几乎所有的关系。机器学习可用于设计合成生物学系统，即可用于学习表观遗传回路，从而实现更稳定的回路。

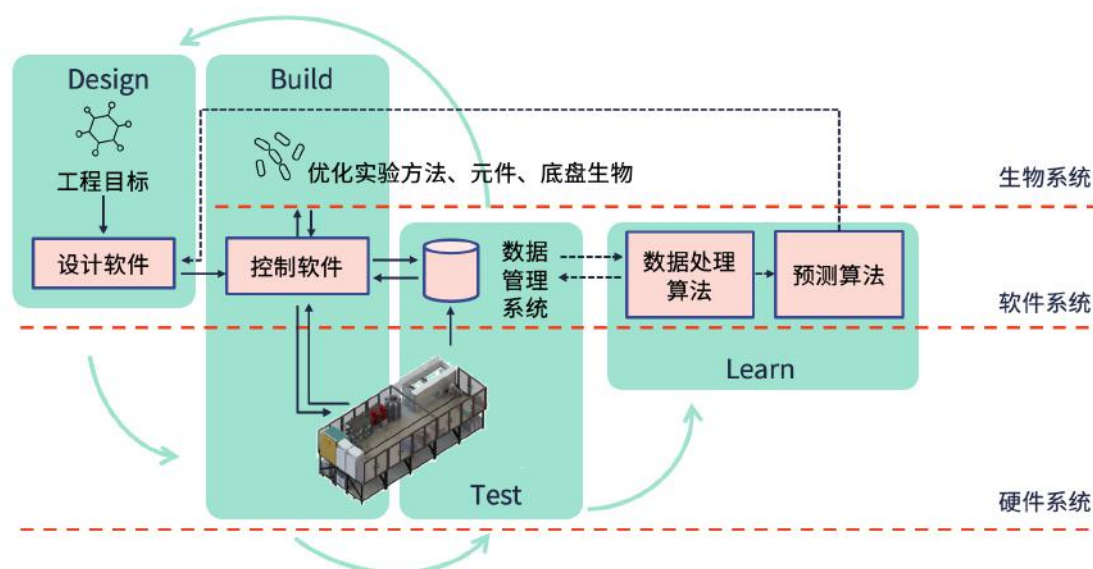
- 自动化技术保证有效的机器学习

机器学习的高速发展来源于日益增长的计算能力和大型数据库的可获得性，自动化能够产生足够数量和高质量的数据来训练有效的机器学习。目前许多研究常需从“整体”或“系统”层面切入，对基因片段、代谢网络、调控网络，甚至是细胞整体进行人工的设计与合成，而人工不断试错的实验模式导致项目的研发周期长、失败率高。然而大多数表型数据是在同一个实验室中产生和进行分析的，尽管有新的发展，但还不足以支持机器学习。同时，许多外部历史数据并不满足机器学习有效的要求（例如缺乏标准化的数据收集），因此在收集新数据时必须考虑到这些需求，将机器学习的算法与高通量、高效率、自动化的技术相结合，从而得到海量的有效数据得以训练模型。

- 机器学习和自动化相结合，构建 DBTL 循环

通过整合自动化设备、智能化控制软件以及数据管理系统，构建一套合成生物学自动化装置，实现从生物样本存取、DNA 元件组装、细胞筛选及培养到产物检测的全流程自动化操作，可以帮助工业及学术界大幅提高效率，同时，装置产生的大量高质量数据结合信息技术，有望高通量、低成本、多循环地实现合成生物学研究中“设计-构建-测试-学习（DBTL）”的自动化运行（图 1）。其中，自动化液体处理工作站、微流控或云实验室等加速了合成生物学在基础及诸多应用领域研究效率，提供具有广泛适用性的生物问题解决方案。

图表 11. 机器学习和自动化以不同的方式改进基本的合成生物学“设计-构建-测试-学习(DBTL)”循环。自动化可以通过液体处理工作站和微流控平台、高通量组学定量和实验数据分析快速实现合成生物学的构建和测试阶段。机器学习可以通过生成实验计划、智能选择样本进行量化、从实验数据中进行模型推断以及设计下一轮迭代的规则来驱动循环中的每个步骤。



资料来源：UIUC

2.2 智能化在合成生物学中的应用 - 机器学习与高通量自动化平台介绍

机器学习在合成生物学中的应用极为广泛，不止基因，还有基因调控元件识别、蛋白质结构和功能预测、蛋白质定向进化、代谢途径构建等——凡是可以从数据中提取路径模式并应用到新数据集的任务都可以应用机器学习。在生物信息学中，机器学习可以通过训练算法、建立模型、优化特征提取和分类而有效提高数据的解读效果。以下将会介绍机器学习在蛋白质定向进化中的应用，以及相关智能化平台优点：

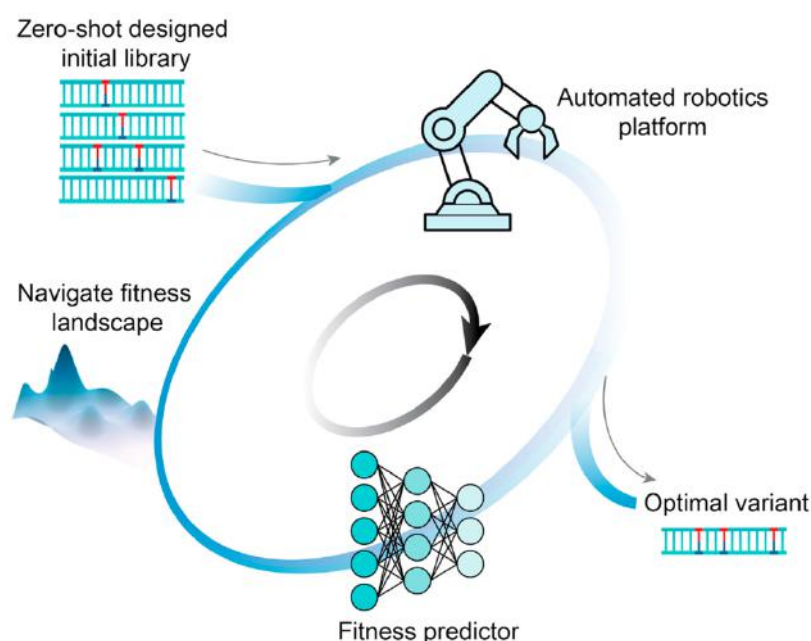
2.2.1 以蛋白质定向进化过程示意机器学习与自动化平台结合的优点

蛋白质工程是合成生物学重要应用之一，而定向进化（Directed Evolution, DE）是蛋白质工程中最成功、最强大的工具之一。随着机器学习（Machine Learning, ML）在蛋白质定向进化领域的出现，我们可以运用计算机评估变异并指导更有效的定向进化。自动化技术实现了在工业界、学术界中快速执行长时间的、复杂的实验，从而获得高通量的有效数据。

体外连续蛋白质进化大致涉及以下五个步骤：

1. 采用零样本预测器设计初始变体库；
2. 通过全自动化平台构建、表达和筛选初始变体库；
3. 用筛选出的高质量数据来训练机器学习模型，以预测适应度；
4. 根据预测适应度的结果，设计随后几轮的启动变体库；
5. 迭代执行以上步骤，直到识别出最佳变体。

图表 12. 体外连续蛋白质进化所涉及的步骤概述



资料来源：Cell Systems

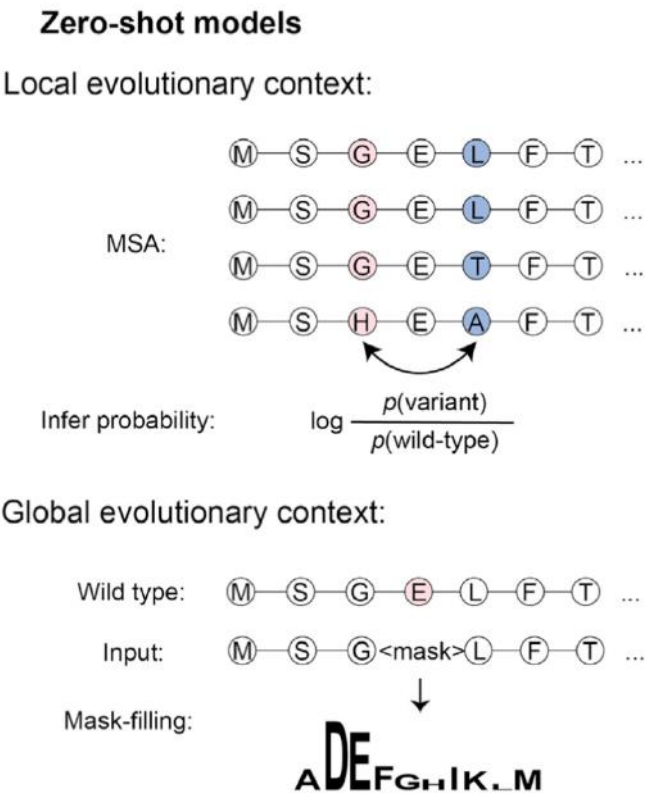
A. 使用零样本机器学习模型设计初始变体库

定向进化利用了各种基因多样化工具（例如随机诱变或基因重组）来探索蛋白质变体景观，以识别具有所需特性的变体。在机器学习引导的定向进化（MLDE）下，筛选量不再是最大的限制因素，而是需要最大限度地减少低适应度变体，即使用“零样本”预测模型设计初始库。此处零样

本意味着除了蛋白质和同源蛋白质的野生型序列之外，无需任何先验知识。

由于进行预测的信息类型不同，基于序列的零样本模型分为两种：局部进化和全局进化模型。局部进化模型，例如 EVmutation 和 DeepSequence，采用目标蛋白的多序列比对（MSA）数据；全局进化模型则采用包含数十亿个蛋白质序列的大型序列数据库，且不限目标蛋白质同源性。大多数全局模型是语言模型，例如进化规模建模（ESM）和 ProtTrans。

图表 13. 可以使用全局或局部进化模型的零样本变异效应预测器来设计知情变异库。全局模型根据所有蛋白质家族的序列进行训练，局部模型根据多个序列比对推断出的残基之间的相互依赖性来计算概率。



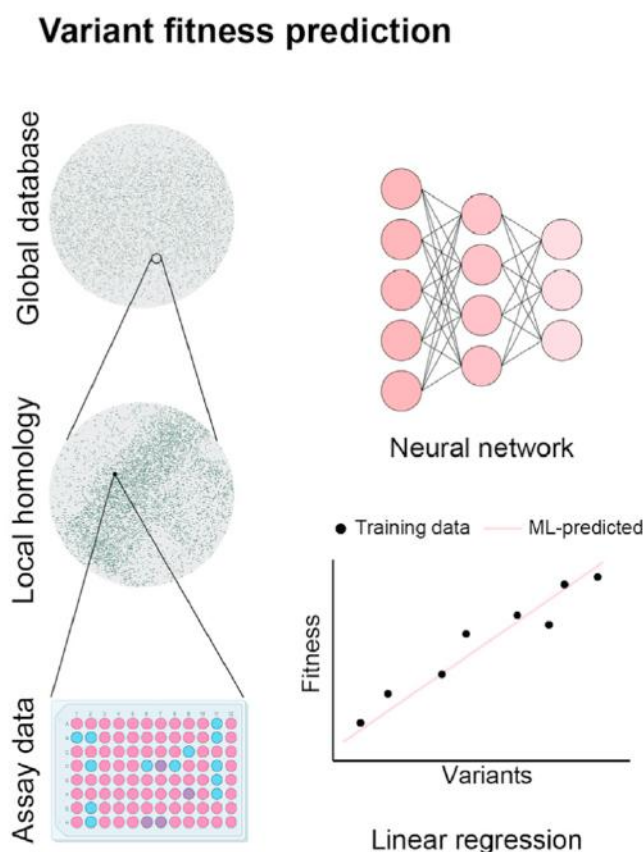
资料来源：Cell Systems

B. 开发预测变异特性的机器学习模型

在构建、表达和定量初始变体文库后，可获得基因型与表型对，用于机器学习模型的训练数据（如简单的线性回归或更复杂的神经网络），以预测未知变体的属性。由此，传统定向进化中较为费力且需要迭代的高通量筛选部分可用机器学习中变异景观的预测来代替，且只需要测试所需变异空间的一小部分。几项独立研究还证明，机器学习可以将低阶变异信息

泛化为高阶变异，在单点突变上训练的机器学习模型也可以捕获多点突变信息。

图表 14. 从预训练的全局语言模型获得的序列嵌入用于表示变体序列。可以通过同源序列进一步微调该序列，以增加任务特异性。由此获得的变量适应度数据可用于训练回归模型，如简单的线性回归或更复杂的神经网络。



资料来源：Cell Systems

机器学习训练需要大量数据，而想达到准确的预测结果也需要相当大的数据集。最近的研究趋势也在关注训练示例数量有限（例如少于 100 个）的“低 N”场景。理想低 N 设置下，研究人员不需要进行广泛筛选大量变体。低 N-机器学习模型可以从根本上避免对高通量筛选和选择方法的需求，这种工作流程可适用于所有蛋白质工程任务。

近年来，各种各样的机器学习模型问世，其中许多模型发展类似计算机科学界自然语言处理（NLP）领域，如 Doc2Vec、乘法长短期记忆

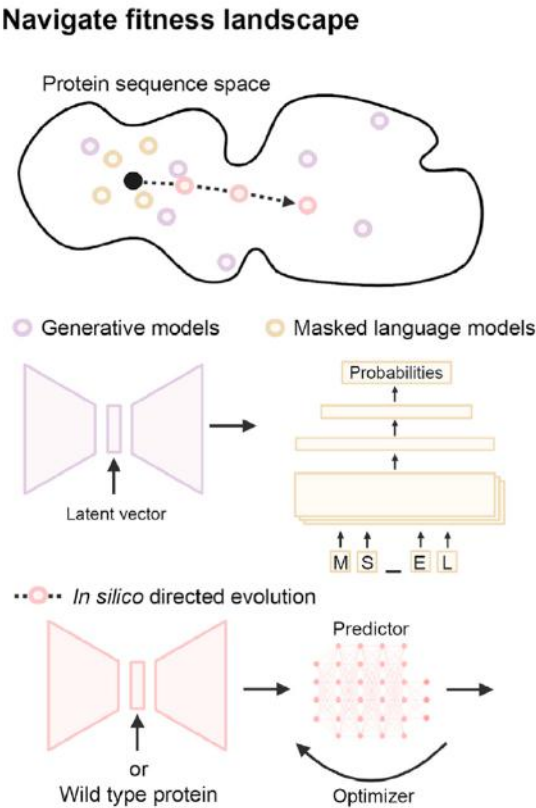
（mLSTM）和 transformer，其中主流 NLP 模型（使用监督数据来微调全局上下文模型）的性能不如无定向 NLP 模拟（增强方法）的机器学习模型。因此，在基于序列的模型开发针对蛋白质建模的生物学特异性解决方案时，需谨慎使用主流 NLP 模型。另外，近年来 AlphaFold、

RoseTTAFold 和 ESMFold 等蛋白质结构预测工具获得成功，让结构信息更容易获取，由此利用结构信息的新型数据模型可以将当前最先进的技术推向一个新的水平，实现更准确的低 N 或零样本预测。

C. 使用 MLDE 覆盖更广泛的领域

蛋白质的长度从数十个到数千个氨基酸不等。即使仅有 20 种氨基酸，使用高质量变异适应度的机器学习模型也不可能完整探索整个蛋白质景观。在筛选适合度和活动性时，可以优先考虑具有较高可行性的变体。两种常用的定向进化方法是位点饱和诱变（在主要结构模型中可能的位点上检查所有 20 个氨基酸）和深度突变扫描。但此类方法常因缺乏高通量筛选和选择方法而受到限制。近期，研究人员利用新的 MLDE 方法，使用各种深度生成模型来覆盖更广泛的领域，并帮助选择出功能性高阶（多突变）变体。

图表 15. 可以通过合并深度生成模型、掩码语言模型或计算机定向进化平台来定位蛋白质序列空间。后两种方法可以有效探索局部景观，而深度生成模型可以创建多样化/高度突变的功能变体，以帮助探索全局的蛋白质景观。



资料来源：Cell Systems

大多数 MLDE 研究是通过向野生型序列引入突变来进行的，但结合具有高维序列多样性的功能蛋白质则可为全局景观提供更多价值。生成模型可从训练数据中学习潜在的高维分布，之后帮助采样那些远离野生型序列的新序列，可以使进化速度更快、成本更低、简单容易。

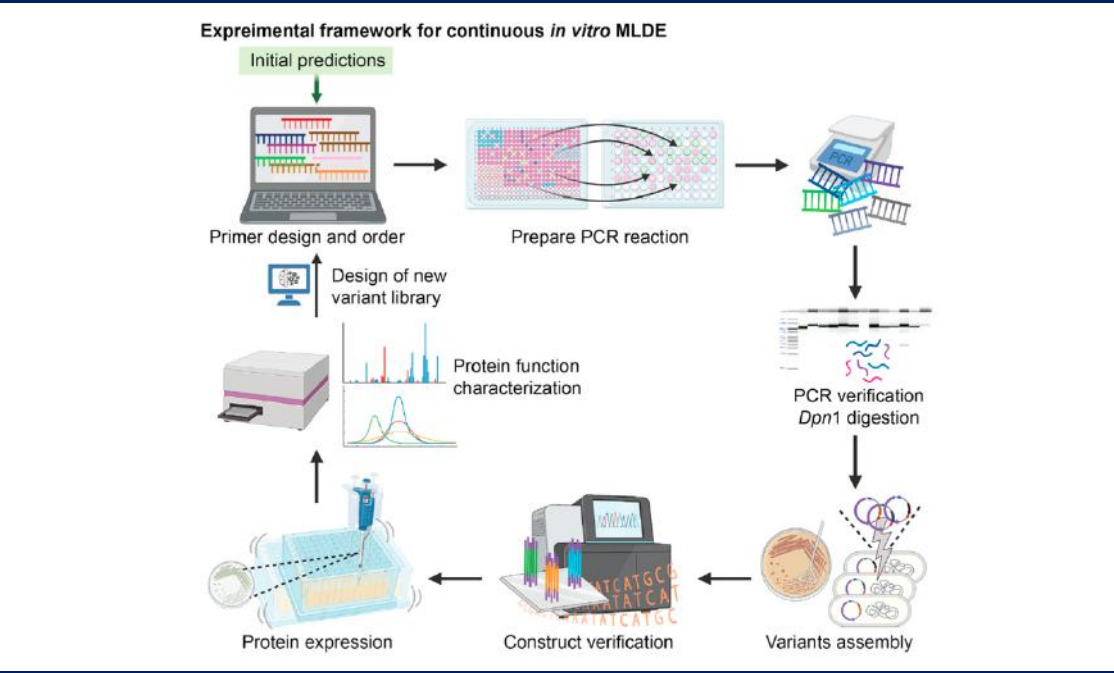
D. 确定后续实验测量的设计优先级

优先考虑后续实验测量的设计是体外连续蛋白质进化框架中的一个重要步骤，需通过设计一个获取函数来对所有变体进行排序，以确定在后续几轮模型训练中不断进化的子集，经常用抽样方法如贝叶斯优化或高斯过程来获得估算预测。深度学习模型和贝叶斯优化的结合是蛋白质工程的核心策略之一。而探索机器学习中的方差与偏差可能是一个具有挑战性但有趣的未来研究方向。

E. 开发自动化平台，以支持文库创建、蛋白质表达和分析筛选

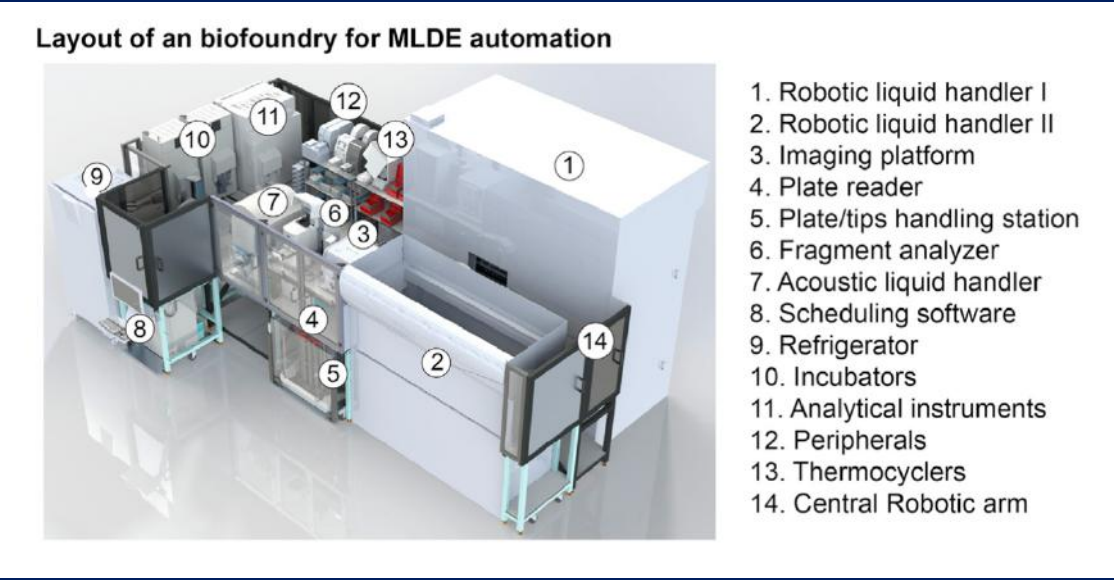
在定向蛋白质进化的湿实验室中，实验通常由熟练的研究人员以低通量的方式完成，但反复的人工试错导致实验周期长、失败率高。自动化平台提供了一个理想的集成环境，提供了用于高通量数据生成、采集和分析的硬件和软件，从而能提高通量、简化可重复的工作流程、减少周转时间并显著节省劳动力，更好地支持机器学习下的文库创建、蛋白质表达和分析筛选。实际应用中，爱丁堡基因组铸造厂和伊利诺伊州先进生物制造生物中心（iBioFAB）开发了自动化工作流程，每周可以执行数千次 DNA 组装。iBioFAB 和机器学习的集成又催生了一个名为 BioAutomata 的自动化闭环系统，用于路径工程。以上应用将在后半部分内容中详细展开。

图表 16. 连续 MLDE 的框架，所有实验都可以通过生物铸造厂实现自动化，并与机器学习相结合，创建闭环蛋白质进化平台。



资料来源：Cell Systems

图表 17. 自动化生物铸造厂的布局，其中包括通过中央机械臂连接的各种核心和外围仪器，并通过调度软件进行控制。



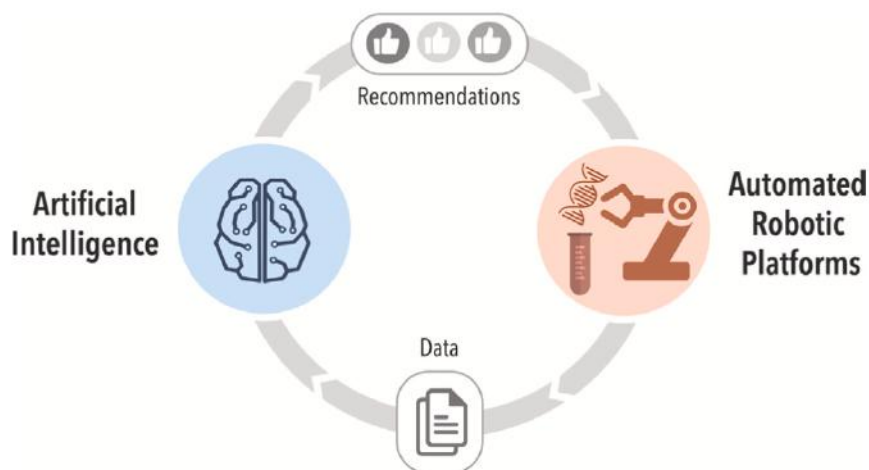
资料来源：Cell Systems

从自动化的角度，SDLs (Self-Driving Labs) 或自主实验 (Autonomous Experimentation) 提供了自动化概念的定义及未来平台的趋势。SDLs 在生物学中呈现出独特的机遇和挑战，尽管造价非常昂贵，但能实现重大的合成生物学进步，解决那些用现有的方法难以解决的生物学难题。

2.2.2 SDLs 或自主实验-自动化平台趋势

SDLs 或自主实验将自动化机器人平台、数据收集与人工智能相结合，通过处理这些数据来决定下一组要进行的实验，并决定接下来要测试的假设或结论。

图表 18. SDLs 工作流程图示



资料来源: ScienceDirect

合成生物学中的 SDLs 可以是一个 DNA 组装微流控芯片，它可以自动产生调控特定代谢途径的变体，以合成所需的代谢产物，并将它们转化到宿主中（如大肠杆菌、假单胞菌或酵母等），同时培养这个宿主并测量相应的代谢产物。这种自动化的实验设置将与 AI 推荐引擎相结合，引擎采用这些实验数据，提出调控不同代谢途径的变体，从而达到目标产物产量最大化的目的。AI 引擎提出的变体和基因编辑假设，会由下一个循环中的自动化微流控芯片进行测试。

SDLs (Self-driving labs) 的概念要求实验室中的各个环节都由自动化系统和人工智能来完成，包括实验计划的制定、实验的执行、数据的收集和分析等，而不需要人工进行主动的干预或操作。但就当前技术水平而言，可能无法达到完全自主的程度。为了描述当前的技术能力，激励现有体系逐步发展为完全自主的系统，我们提出了一套类似汽车自动驾驶概念的“自主等级”。从当下实际应用角度看，显示出三级或以上自主程度的系统就可以被认定为 SDLs，因为它们已经可以展示闭环的“设计-构建-测试-学习”循环。

图表 19. SDLs 各级别的人工介入的详细情况

Level	Description	Example
5	AI researcher	(To be achieved)
4	Highly-autonomous research	Adam, Eve [6, 7]
3	Conditional autonomy	SDL
2	Partial autonomy	Aquarium [5]
1	Research assistance	Pipetting robot
0	No autonomy	Electronic or paper notebook

资料来源: ScienceDirect

SDLs 是 3 级自主系统，其中自主级别描述了该系统与人工干预程度的独立程度。

- 0 级，所有实验设计、执行和数据收集都由人工处理。
- 1 级，一些重复性任务将分配给机器人。
- 2 级，需要人工对实验室自动化设备发出指令，连接调试各类设备的应用软件，人工处理海量的数据。
- 3 级，涉及闭环的“设计-构建-测试-学习”循环，为 SDLs 的最低要求。常规数据分析的工作和标记异常情况仍由人工处理。
- 4 级，涉及自动化机器人协议的执行和常规数据分析，如“Adam”和“Eve”，人工只参与设定目标和计划（即 SDLs 作为实验室助手）。
- 5 级，人工仅设定目标并接收结果（即 SDLs 作为调查者，人类作为管理者）。

2.3 海外合成生物学智能化应用发展现状 - 落地重点案例介绍

合成生物学的自动化循环，与轻工业中家用电器和电子等离散产品的生产制造有一定相似性，因此这些领域的智能制造、智能工厂理念被引入到合成生物学研究中。简单来说，即推进生产设备与生产线智能化，通过引进各类符合生产所需的智能装备（如液体处理工作站、PCR 仪、酶标仪、自动培养箱、离

心机等），建立基于信息化的车间级智能生产单元（如中央机械臂或传送带），提高精准制造、敏捷制造能力。更深入来说，是引入个性化定制与互联工厂理念，通过互联网平台开展大规模个性化定制模式创新，充分满足研究者多元化需求的同时，实现生命体的远程定制、异地设计、规模经济生产。该类用于合成生物学的自动化设施是工程化平台的核心，亦被称为生物铸造厂（BioFoundry）。

至今为止，美国政府已支持设立 3 个大型合成生物学研究中心，英国政府已经资助 6 个大型合成生物学研究中心。德国、荷兰、日本、新加坡、澳大利亚等国也在紧密跟进。在各大研究中心与学术机构中，一般都搭建有生物铸造厂作为核心。这些合成生物学自动化设施平台既用于加速学术研究，也用于推动产业发展。当然，许多企业也搭建了自己的自动化设施平台，如美国 Amyris 公司、Ginkgo Bioworks 公司、Zymergen 公司、Transcriptic 公司等。这些设施平台的规模不一，功能大都是帮助研究人员将特定的基因线路设计自动化装载到活细胞中，并辅以高通量测试。工作流程往往都依照“设计-构建-测试-学习”的循环来组织，以实现工程化的海量试错。

图表 20. 全球学术机构的知名合成生物学自动化设施平台

自动化设施	学术机构
深圳合成生物研究重大科技基础设施*	中国科学院深圳先进技术研究院
生物铸造厂	中国科学院天津工业生物技术研究所
规模化蛋白质制备系统	国家蛋白质科学中心（上海）
生物铸造厂*	天津大学化工学院
高通量筛选平台	武汉生物技术研究院
生物铸造厂（面向企业）*	华大生命科学研究院
生物铸造厂*	日本神户大学
生物铸造厂*	韩国先进科技学院
Synthetic Biology Foundry	新加坡国立大学
Agile BioFoundry	美国劳伦斯伯克利国家实验室
DAMP Lab	美国波士顿大学
Illinois Biological Foundry for Advanced Biomanufacturing (iBioFAB)	美国伊利诺伊大学厄巴纳-香槟分校
MIT-Broad Foundry	美国麻省理工学院
Center for Applied Synthetic Biology	加拿大康卡迪亚大学
London DNA Foundry	英国帝国理工学院
Earlham DNA Foundry	英国厄尔汉姆研究中心
Edinburgh Genome Foundry	英国爱丁堡大学
GeneMill	英国利物浦大学
SYNBIOCHEM	英国曼彻斯特大学
生物铸造厂（面向企业）*	法国农业科学研究院图卢兹中心
CompuGene	德国达姆施塔特工业大学
生物铸造厂*	荷兰代尔夫特理工大学
Novo Nordisk Foundation Center for Biosustainability (CFB)	丹麦技术大学
Australian Genome Foundry*	澳大利亚麦考瑞大学
生物铸造厂*	澳大利亚昆士兰大学

*表示在建或拟建

资料来源：中国科学院学报

其中，美国国防高级研究计划局（DARPA）资助的“生命铸造厂（Living Foundries）计划”是实施最早、规模最大的计划之一，自 2011 年 5 月启动以

来，已累计部署经费近 4 亿美元。实施以来，取得的主要进展包括：开发了新的合成生物学计算机软件系统，该软件系统将合成生物学设计时间从以往的 1 个月缩短至 1 天，并能实现端到端的监控；构建了大规模基因网络，以该网络为基础初步验证生物制造的正向工程能力；建立了大规模 DNA 组装新方法，将体外准确装配的 DNA 片段数从此前最高 10 个提高到 20 个的水平，错误率降低到原来的 1/4；将多种新生物制品的设计、工程化生产提速了 7.5 倍；目前已经实现了对乙酰氨基酚合成途径的设计和制备。

“生命铸造厂计划”虽已取得多项重要进展，可行性已得到初步验证，但其工程化应用仍存在诸多难点，面临的最大的技术挑战包括已知分子结构难以快速改进、某些分子结构无法合成、新分子结构难以设计等，总体上仍处于前沿探索阶段。未来一旦取得突破，可显著提升现有制造能力。

下文将从原理入手，介绍全球合成生物学研究具有代表性的自动化平台。

2.3.1 使用生物铸造厂（Biofoundry）全自动合成单转录物

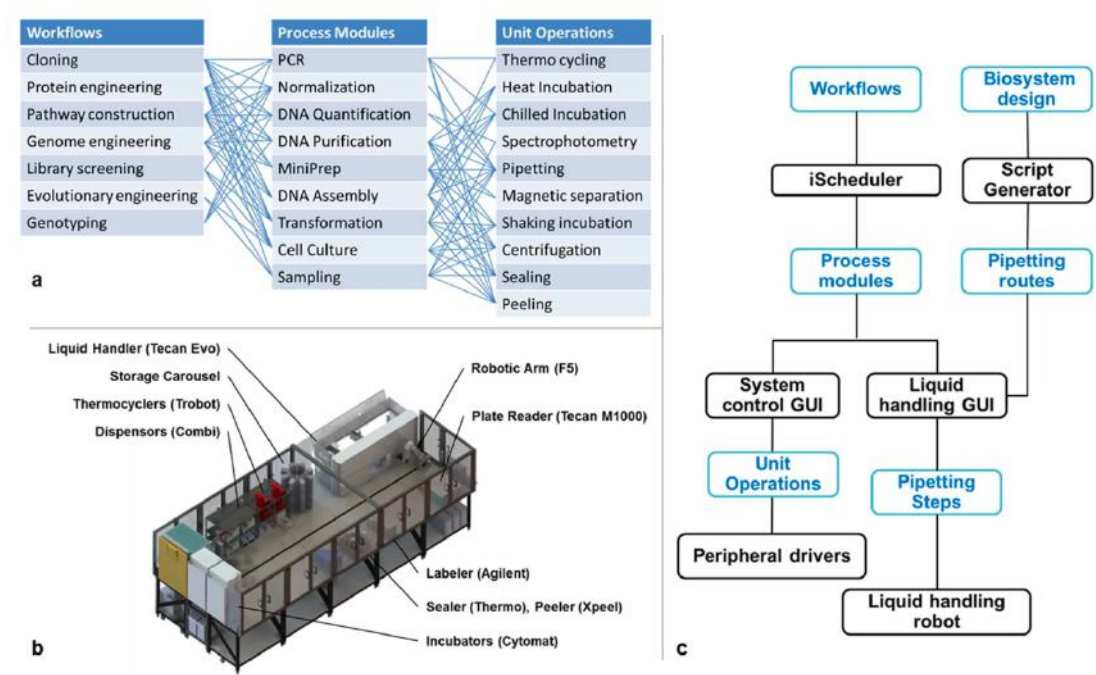
TALENs

转录激活因子样效应物核酸酶（TALENs）是一种具有广泛应用的基因组编辑工具。TALENs 以二聚体形式进行 DNA 切割，为每个目标基因组位点合成一对 TALENs。有两种传统方式，一种是在单独的载体上表达，另一种是单独合成，然后亚克隆到同一载体上，但这两种方法都不允许高通量的构建大规模应用的 TALENs 库。Huimin Zhao 等人提出了一种单步组装方案，在 P2A 自切割序列的帮助下，以单一转录格式合成和表达一对 TALENs。同时他们开发了一种在通用生物铸造厂中定制 TALENs 的全自动平台。使用本公开的系统和方法，每对能够以合理的材料成本以超过 96.2% 成功率合成四百对 TALENs。此平台为基于 TALENs 的全基因组研究以及生命科学中的许多其它应用打开了大门。

为实现 TALENs 的大规模应用，如基因筛选，Huimin Zhao 等人寻求完全自动化 TALENs 的合成过程。然而现有的集成平台均针对某些特定任务进行了深度定制，难以重新配置。如果构建专用于 TALENs 合成的深度定制系统，那就违背了高效与经济的目的。最终 Huimin Zhao 等成功实现了在 iBioFAB 上完成了组装工作。

iBioFAB 由组件仪器、自动机器人平台和模块化的计算框架组成，共有 20 台设备，每台设备负责一个操作单元，不同设备之间由机械臂进行连接，如移液和孵化模块会通过两个机械臂连接到其他过程模块如 DNA 组装和转化，然后进一步组织成通路构建和基因组工程等工作流程（图表 21，a、b）。除此之外，还开发了一个整体调度程序来协调单元操作，并允许对工作流程进行分层编程（图表 21，c）。

图表 21. iBioFAB 系统概述: a. 单位运作细目。b. iBioFAB 的硬件布局: iBioFAB 有两个机械臂, 5 米的轨道上集中了多向机械臂用于在仪器之间运输实验样本, 另一多向机械抓手在液体处理工作站内移动实验样本。c. iBioFAB 的控制层次结构: 过程模块是在系统控制 GUI 中开发的。iScheduler 负责工作流程级别的控制, 脚本生成器为液体处理 GUI 生成移液路线, 流程模块可以快速重新组合以组成不同的工作流程。



资料来源: ACS Synthetic Biology

iBioFAB 的配置可以执行通用的自动 DNA 组装工作流程, 其中可以使用 Golden Gate 方法按需制造各种 DNA 构建体。为了简化流程, Huimin Zhao 等人开发了 Script Generator 工具。其作为设计工具可以自动将 DNA 组装转换为混合匹配任意 DNA 部分的实验例程。脚本生成器随后为 iBioFAB 生成机器人命令来执行复杂的移液工作。为最大程度地减少吸头和时间消耗, 移液路线也经过了优化。对于同一基材, 尽可能进行组合吸取步骤, 之后将其分配到对应目的地, 吸头则按需从存储转盘装载到液体处理工作站。

通路优化工作流程是使用 iBioFAB 实施的, iBioFAB 已用于高通量 TALENs 合成和自动化酵母基因组工程。结合 iBioFAB 及贝叶斯优化算法, DBTL 循环的所有方面都实现了自动化。在每一轮运行中, 贝叶斯优化算法选择 46 个点进行评估 (实际测试选择的数量) 并将其提供给 iBioFAB 调度软件。然后, 软件从零件库中吸取要组装的正确 “零件”, 并使用 Golden Gate 组装来组装质粒, 然后在四个生物学重复中测量这些点的产量, 并将结果的平均值返回给算法以计算下一个要评估的点。贝叶斯优化算法首先均匀地探索整个景观, 并在后面几轮中逐渐不均匀, 以获得更多有关景观的信息, 从而促进

对特定区域的探索。在第二轮中，仍会做一些探索，而第三轮中的点几乎会收敛到一个被认为产量最高的特定区域。

已运行的实验证明，路径优化的后几轮具有更高的平均产量和更高的最大产量，这显示了贝叶斯优化算法在随后每一轮中找到更好突变体的有效性。

2.3.2 全自动算法驱动平台 BioAutomata

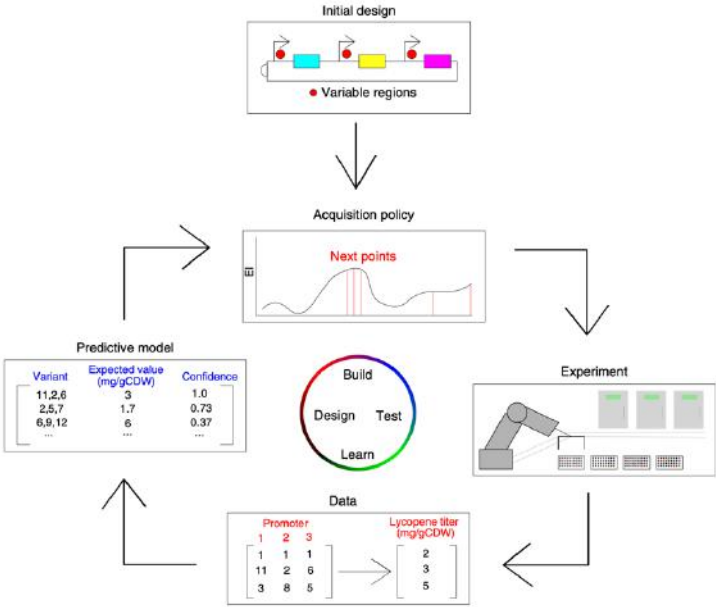
BioAutomata 是集成了自动化系统与机器学习算法，是以完全自动化的生物系统来设计的 DBTL 过程。该平台可以评估不到 1% 的可能变异，比随机筛选准确率高出 77%。其中配对预测模型和贝叶斯算法选择由伊利诺伊州先进生物制造生物铸造厂 (iBioFAB) 进行的实验。BioAutomata 擅长解决黑盒优化问题，这些待优化问题的实验成本高昂且操作复杂，更有趣的是，其成功并不依赖于对生物机制的广泛先验知识。

为通过学习组件的自动化和集成来执行多轮设计、构建、测试和学习 (DBTL) 循环，并以最少的人为干预实现该循环的迭代，会将 iBioFAB 与机器学习算法集成。BioAutomata 平台设计实验、执行实验并分析数据，以迭代方式优化指定的生物过程。BioAutomata 在最初生成的 (或可用的) 数据上训练概率模型，并决定要评估的优化空间的最佳点，即更有可能导致生物系统改进的点。该优化框架非常适合找到黑盒函数的最优值。

而 BioAutomata 普适性地用于路径优化的潜在挑战，包括难以在自动化平台上执行提取方法，或需要比生物读板仪更复杂的设备行使分析/定量方法，例如气相色谱-质谱 (GC-MS) 仪或液相色谱-质谱 (LC-MS) 仪。这些挑战可以克服，但必须依赖于建造更大规模、更复杂的生物铸造厂以整合上述仪器。

首选通常是优化生物合成基因的表达。通过结合贝叶斯优化算法和 iBioFAB 自动化系统以找到产生高效价的菌株。该数据上的每个点都表示在给定的每个基因的特定表达水平情况下，所需化学品的产量。初步设计和设置完成之后，研究人员角色从实验的驱动者转变为系统的监督者，让算法来驱动优化平台设计和执行实验，目的是让研究人员定义的目标函数达到最大化。

图表 22. 整体工作流程：设置初始参数、设计可变区的序列空间（例如组合途径组装中的启动子变体）并定义目标函数后，BioAutomata 选择哪些实验预计会带来最高的产量提高，执行这些实验，生成数据并从中学习，根据新提供的数据更新其预测模型。然后，它将决定执行下一个实验，以达到用户设定的目标，同时尽量减少实验数量和项目成本。



资料来源：Nature Communications

A. 全自动算法驱动平台 BioAutomata 介绍

优化的第一步是确定系统的初始设计、输入和输出以及目标函数。初始设置后，应选择预测模型和采集策略来估计当前可用数据的情况，并选择下一个要评估的点和要执行的实验。选择系统的所有要素（初始设计、采集策略、实验设置、数据采集和预测模型）后，BioAutomata 就可以开始优化。一，获取策略选择要评估的点。接下来，iBioFAB 执行实验来评估所选点的适合度，并将数据返回到预测模型。二，该模型将根据新提供的数据更新并训练预测模型。三，采集策略将在更新的预测模型的指导下选择下一步要评估的点。

B. 确定预测模型和数据采集策略

由于目标是找到黑盒函数的最大值，因此使用贝叶斯优化算法，该算法是解决此类问题的理想选择。它构建了一个概率模型，并使用该模型来决定下一步评估的位置，从而减少找到最大值所需的实验轮数。该算法考虑了每次评估的预期结果以及对该预期结果的置信度。

贝叶斯优化算法依赖于顺序性实验。每次评估一个点时，都会将结果提供给算法以更新先前的算法，并使用采集函数找到下一个要评估的点。采用并行和连续批次进行一些实验能减少实验轮数，从而减少整个项目耗时，是一种更有效的方式。最近，研究人员为多核并行处理应用程序开发

了贝叶斯优化的变体算法，它可以处理多个待评估的点，并且可以在任何给定时间获取任何待评估的结果，并返回下一个要评估的点，该算法考虑了每个待定点的可能结果，并根据所有可能的结果计算获取函数。在实验设置中，当使用并行实验进行评估时，待处理点会在后续批次中同时更新，而不是逐一更新。

在数学函数评估的情况下，如果实验中没有错误，则已评估点周围的置信度非常高。然而所有实验远非完美的数学计算，结果都会有一定误差，因此要对结果的置信度进行调整。

C. 贝叶斯优化算法的评估

为了说明 GP 的贝叶斯优化算法，实验定义了一个单变量函数，并尝试通过顺序采样找到最大值。设计时特意让该函数具有多个峰值和局部最优，以测试优化算法是否确实可以找到全局最大值。接下来，实验用具有三个输入和一个输出的类似三变量函数来说明优化方法，以模拟类似的多维优化问题。然后，实验通过在不同条件下运行多个模拟来测试贝叶斯优化算法。最后，实验设置优化每轮评估的点数，并在实验成本和时间之间进行权衡：随着每批规模的增加，实验成本也会增加，但实验的总轮数减少了，整个项目花费的时间也减少了。批量大小还受到实验条件的限制，特别是高通量生物实验的标准 96 孔格式。

D. 生物合成途径的自动优化

生物合成途径生产力和产量低的原因之一是通量不平衡，其中次优反应速率会导致反应中中间体分子的积累或耗尽，然而想要找到合成途径中每个步骤的复杂平衡是非常困难的。为了在各种不同情况下增加总通量，可通过微调并优化路径中每个步骤的通量来实现。这个问题的抽象概念可以用表达-生产函数来表示，其中最大通量通过合成途径中每个基因的特定表达水平来实现。据此可以得出实验设置，即可以调整通路中基因的表达（函数的输入）并定义想要的最大化输出，之后尝试将不同的基因表达水平作为输入，产量作为输出，从而找到与最高输出相对应的输入。

英美等国现有的生物铸造厂仍然存在一定的局限性，包括复杂线路设计能力不强、底盘细胞单一、大片段 DNA 的制造成本高、高通量测试手段少、与下游应用衔接不紧密等。许多研发需求仍未能满足，领域的发展仍面临障碍。

2.4 国内合成生物学智能化应用发展现状 - 国内落地的重点案例

与国外学术机构的工程化平台相比，国内以中国科学院深圳先进技术研究院的“合成生物研究重大科技基础设施”、中国科学院天津工业生物技术研究所的生物设计中心等为代表，也有自己的特点与创新。而下文将介绍浙江大学杭州国际科创中心下属的生物与分子智造研究院建设的合成生物学自动化装置

iBioFoundry，力争构建深度融合“BT+IT（生物经济+数字经济）”的高能级合成生物学研究平台。

iBioFoundry 采用模块化设计，利用轨道式机械臂整合各种外部设备，包含样本库、DNA 元件组装、细胞筛选及培养和分析检测四大功能模块。装置整体设计具备开放性，台面布局相对宽松，后期可以通过设备的位移、层叠或添加，实现设备布局的灵活变动或装置功能的升级。根据不同模块的实验操作需求和特点，装置搭配使用工业级和协作式机械臂，从而兼具工业机械臂的高效率、高稳定性和协作式机械臂操作灵活的特性。装置中机械臂可与所有外周设备直接交互，避免以往某些设计中个别设备需要通过移液工作站中转样本的情况，以此提高设备调用的灵活性。对于装置中经常需要进行耗材试剂补充或配件更换的设备，在设备底部增加滑轨或转盘以方便其使用和维护。同时为了保障实验环境的无菌需求，DNA 元件组装和细胞筛选及培养两大模块采用封闭式设计，配备全覆盖式 HEPA 空气过滤装置。

为满足不同科研任务需求，最大程度提升装置使用效率，iBioFoundry 设计具备多实验任务进程管理功能，能追踪记录全部实验流程的过程处理信息，实现所有样本的追踪溯源。装置的优势和特点包括：

1. 自动化无人值守工作；
2. 多个不同实验任务可同时运行；
3. 运行过程中可随时修改或添加新的实验任务；
4. 运行过程中可插入优先处理的紧急样本；
5. 运行过程中动态补充耗材、试剂；
6. 根据实验结果完成智能化逻辑判断；
7. 具有完备的设备故障恢复策略；
8. 实现远程和批量接收任务指令功能等。

2.4.1 各模块设备集成及功能

DNA 元件组装模块承担的主要实验任务包括基因片段的扩增、纯化、酶切、连接、转化以及各种生化反应体系的配制等。该模块实验操作通量大，样品板在各设备间移动频繁。工业级机械臂具有效率高、稳定性好的特性，能承担长时间、快频率的高负荷任务，因此该模块中配备了高性能的工业级 F7 机械臂来完成样本在各设备间的转移工作。整合的设备包括移液工作站、纳升级声波移液站、自动化封膜机和撕膜机、自动化 PCR 仪、自动化离心机和自动化耗材存储设备等。

细胞筛选及培养模块承担的主要实验任务包括感受态细胞制备、菌液自动化涂板、平板静置培养、克隆挑取、菌液振荡培养、诱导表达、细胞破碎及蛋白纯化等。此模块也是利用工业级 F7 机械臂来进行样本在各设备间的传递，整合的设备包括移液工作站、自动化正压过滤装置、大体积快速分液器、自动化

振荡培养箱及静置培养箱、自动化低温冰箱、高速冷冻离心机、酶标仪和自动化耗材存储设备等。

分析检测模块包括高效液相色谱仪、自动化核酸片段分析仪、流式细胞仪、荧光定量 PCR 仪四台分析检测设备和一台耗材存储堆栈，可满足核酸质检、克隆鉴定、基因表达分析、蛋白及细胞代谢产物分析等多种检测需求。不同于前述两个模块，分析检测模块采用开放式设计，可随时接收并快速分析检测样品，也便于耗材、试剂的随时添补及设备的日常维护。出于实验工作量需求 and 安全性考虑，分析检测模块选择使用协作式机械臂完成各设备及模块间的实验样本传送。

样本库模块用于菌种和合成生物学元件的存储和取用。此模块包含一台可存储超十万根 0.5ml 样品管的 -80°C 自动化生物样本存储设备 (Arktic, SPT labtech) 以及配套的扫码仪和开盖机。样本库模块与细胞筛选及培养模块共用一个工业级 F7 机械臂完成样本的传输。

iBioFoundry 实验运行案例：

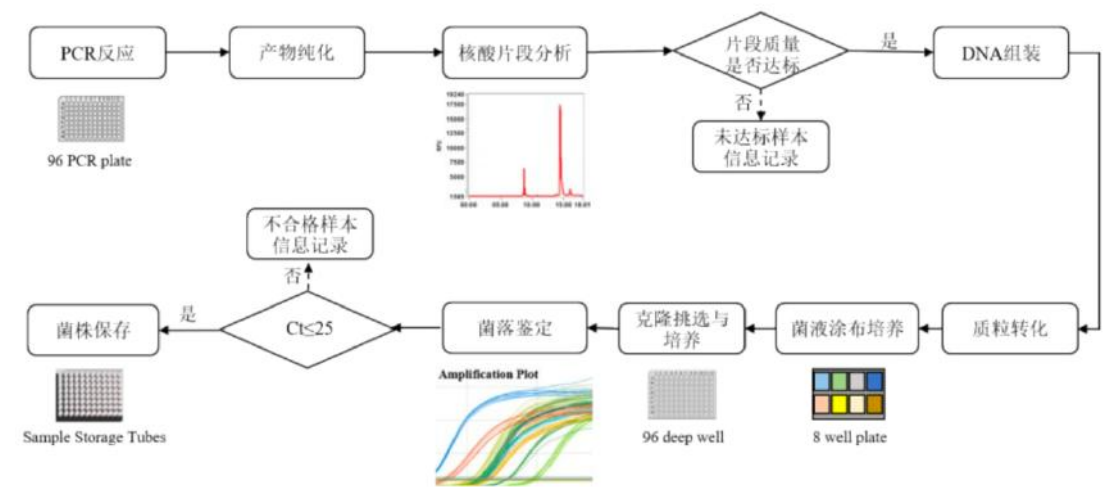
iBioFoundry 目前已成功实现大肠杆菌、酵母细胞、枯草芽孢杆菌、谷氨酸棒杆菌等多种工业微生物“人工细胞”的全自动构建。以下将以大肠杆菌工程菌的批量构建为例，介绍项目自动化方案的制定过程以及在 iBioFoundry 上运行具体实验的情况，并通过酶的定向进化及筛选案例就实验流程的智能化选择和实验通量进行分析。

2.4.2 自动化方案的制定

按流程顺序可将整个实验拆解成 9 个子任务，依次是：PCR 反应、PCR 产物纯化、核酸片段分析、DNA 组装、质粒转化、菌液涂布培养、克隆挑选与培养、菌落鉴定以及菌株保存。为了适应自动化操作，部分实验方法做如下调整：

1. 以磁珠纯化法代替切胶回收，利用移液工作站完成 PCR 产物的纯化；
2. 使用自动化核酸片段分析仪检验 PCR 产物的质量；
3. 利用荧光定量 PCR 仪对菌株构建的正确性进行验证。

图表 23. 工程菌批量构建实验流程



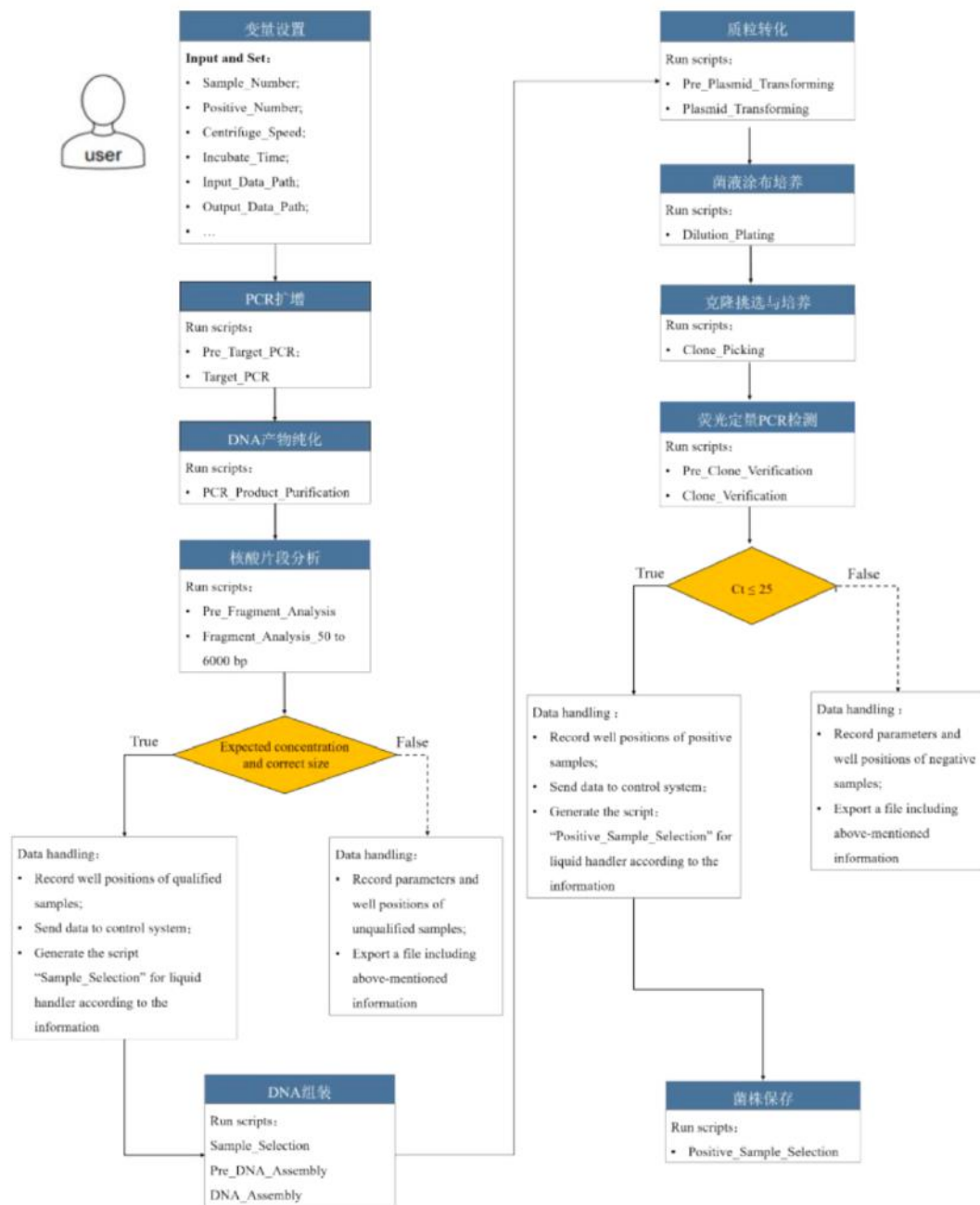
资料来源：Synthetic Biology Journal

图表 24. 子任务实验步骤和单机程序

	子任务名称	实验步骤	需编程的设备	程序名称
1	PCR 反应	PCR 反应体系配置 封膜 离心 PCR 扩增	移液工作站 自动化PCR 仪	Pre_Target_PCR Target_PCR
2	产物纯化	撕膜 磁珠纯化	移液工作站	PCR_Product_Purification
3	核酸片段分析	核酸片段检测体系配置 核酸片段分析	移液工作站 自动化核酸片段分析仪	Pre_Fragment_Analysis Fragment_Analysis_50 to 6000bp
4	DNA 组装	样品筛选及试剂分装 DNA 组装反应体系配制 封膜 离心 DNA 组装反应	移液工作站 纳升级移液工作站 自动化PCR 仪	Sample_Selection Pre_DNA_Assembly DNA_Assembly
5	质粒转化	撕膜 质粒转化体系配置 封膜 热激 孵育	移液工作站 自动化PCR 仪	Pre_Plasmid_Transforming Plasmid_Transforming
6	菌液涂布培养	8孔矩形培养板涂布 平板静置培养	移液工作站	Dilution_Plating
7	克隆挑选与培养	培养基分装 克隆挑选 封膜 菌液震荡培养	移液工作站	Clone_Picking
8	菌落鉴定	荧光定量PCR 反应体系配置 封膜 离心 荧光定量PCR 检测	移液工作站 荧光定量PCR 仪	Pre_Clone_Verification Clone_Verification
9	菌株保存	开盖 菌株挑选与保存 关盖 样本存储	移液工作站	Positive_Sample_Selection

资料来源：Synthetic Biology Journal

图表 25. 实验流程自动化程序编写示意图



资料来源: Synthetic Biology Journal

2.4.3 酶的定向进化及筛选

酶的定向进化通过改造酶的结构并利用高通量筛选手段,可以快速得到具有特定催化活性、稳定性或选择性的酶,以满足各种研究或应用需求。

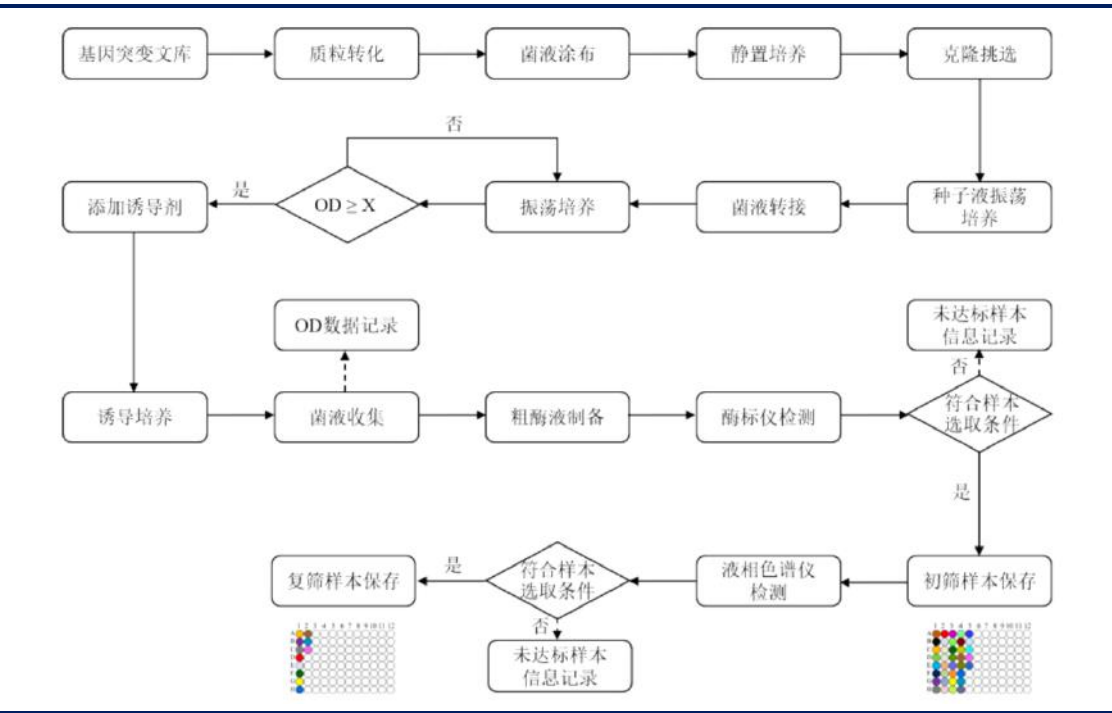
针对实验中大量突变子的操作需求,通常需要将高通量的培养、检测 and 数据分析手段相结合以提高研究效率。以下以酶标检测作为初筛方法,结合高效液相色谱分析作为复筛方法,对在 iBioFoundry 上完成酶定向进化的实验案例进行说明。

首先利用易错 PCR 或对关键位点进行随机突变等方法构建目标基因的突变文库，转入大肠杆菌后利用移液工作站的克隆挑选配件挑取菌落进行液体培养，根据 OD 值检测数据掌握菌体生长情况并择机诱导，诱导表达结束后测定菌液 OD 值。对于分泌表达的酶蛋白，孔板经离心取上清即得到粗酶液。如果是胞内蛋白，离心收集菌体后加入裂解液经振荡孵育破碎细胞，裂解液离心后取上清获得粗酶液。利用酶标仪检测酶活，根据检测结果筛选样本进行液相色谱复筛，最后根据检测数据筛选优势菌株保存，用于后续测序分析或进入下一轮的定向进化实验。

利用中控软件中的变量控制、逻辑判断、循环、数据提取、导入与导出等功能，可实现对实验流程的智能化控制。在本案例的实验流程中，有三个节点可以通过对实时检测数据的分析而实现对下游实验流程的智能选择，分别是：

1. 根据菌液 OD 值数据选择诱导时机；
2. 根据酶标检测数据筛选样本进入复筛；
3. 根据液相色谱检测数据筛选菌液样本保留备用。

图表 26. 酶的定向进化及筛选实验流程



资料来源：Synthetic Biology Journal

2.4.4 iBioFoundry 建设中几点经验

1. 耗材存储空间的分配

自动化耗材存储设备作为 iBioFoundry 中一个不可或缺的组成部分，其作用是存放实验中所需的如吸头、深孔板、微孔板等各类耗材。

iBioFoundry 装置内共有三台自动化耗材存储设备，分别位于 DNA 元件组装模块、细胞筛选及培养模块以和分析检测模块。

2. 多实验任务并行

多实验任务并行能大幅提高 iBioFoundry 的运行效率。本系统的中央控制软件具备任务的动态进程管理功能，可统筹计算并实现多个任务中各进程的最佳时间和设备资源调配，达到多任务同时运行的目标。在此基础上，iBioFoundry 还允许在实验进行中的任意时间点添加新任务，并按照实验需求优化时间和资源分配，同时保障多个实验的运行效率。

3. 实验流程标准化建设

自动化实验方案的制定过程中需要将具体实验流程拆解成多个子任务并编写相应的自动化程序。iBioFoundry 中实验流程的标准化，即将实验流程中各个子任务的自动化程序编写和测试规范化，并最终制定整个实验流程的标准操作文档。

3 未来合成生物学的智能化展望

合成生物学的核心逻辑是将工程原理（如计算设计、部件标准化、模块化和抽象化等）与生物铸造厂的流水线技术相结合，实现“合成生物学产业化”。然而只有一定范围的生物学问题可以通过工业化实施的“设计-构建-测试-学习”循环来解决，主要与某种发酵过程和个别酶有关。生物铸造厂可以加快微生物的初步开发（例如微生物组工程），但产品在农田或患者的肠道中就会严重受限。同理，植物和动物不会每 20 分钟繁殖一次，也不生活在发酵罐中。

合成生物学并不完全依赖于计算设计和标准化部件，甚至最具代表性的合成生物学成功案例也更多地归功于老式的试错、调试方法，而不是基于电子的工程学理论。单纯依仗以计算预测为中心的理解方式，会限制合成生物学的未来发展构想。合成生物学的本质是重新想象生命过程，甚至定义生命本身可能是什么。合成生物学取缔了通过建造复杂设备的工程原理应用于生物体的常规工作，转而通过利用生物进化创造的一切，以不同逻辑来解决多目标优化的挑战。

2018 年 6 月，来自全球各地 15 家学术机构的代表在伦敦参加了“全球生物铸造厂会议”。与会代表认为，生物铸造厂当前普遍面临着相同的挑战，包括昂贵的基础设施的维护费用和运营人员的费用，缺乏通用的设施软硬件标准和知识产权共享机制，生物元件材料难以跨国共享，潜在用户对自动化设施缺乏了解等。因此，合成生物学自动化平台的未来发展趋势包括：

1. 组成联盟，研发通用的操作系统、自动化协议与标准，推动行业整体发展；
2. 完善法律工具，如跨国、跨机构材料转移协议、知识产权共享协议等，为实验材料的共享降低成本，避免重复研发；
3. 围绕自动化设施提供云端服务，扩大用户群体，特别是中小企业用户；
4. 加强国际自动化设施之间的协同，如共同承担国际大科学计划的研究任务，相互分担调配任务，与耗材供应商共同协商定价等。

合成生物学自动化平台设施的建设，一旦走上高速发展的轨道，将为人类面临的医疗、能源和环境等重大问题提供全新解决方案，从根本上变革人类的生产和生活方式。

结合前文，以下将从机器学习和自动化两个方面观察未来发展趋势。

3.1 机器学习方面

目前，通过机器学习的研究方法构建基因序列与生物体具体功能间的因果链条还存在诸多问题。对于简单的线性模型来说，其模型参数代表了输入特征对预测结果的重要性。若输入特征相互独立，线性模型就具有很强的可解释性。然而，对于复杂模型，如神经网络，特征和目标之间存在非线性关系，模型的可解释性会变的极差。

另外，对一些生物网络，如基因调控网路，基因之间存在调控关系的混淆效应，需要借助因果推断，解除调控关系之间的混淆。随着基因组方面的对照实验数据不断丰富，因果关系模型在机器学习领域得到越来越多的关注，相信因果关系模型具备很大发展潜力。

另外，蛋白质等序列功能建模也是机器学习研究中越来越受欢迎的领域，然而要获取有标签的样本通常需要付出较大的代价。怎样获得有标签的样本数据，减少资源浪费，将筛选到无功能蛋白、启动子等序列样本的概率降低，同时怎样通过有限的样本数据构建筛选模型是机器学习技术在未来需要重点解决的问题。

3.2 SDLs 自动化-自动实验室方面

在解决重要且困难的生物学问题时，投资金额巨大的自主实验室（SDLs）方能真正体现价值。以下是一些尚未解决的合成生物学难题，涉及基础和实际重要性的主题：

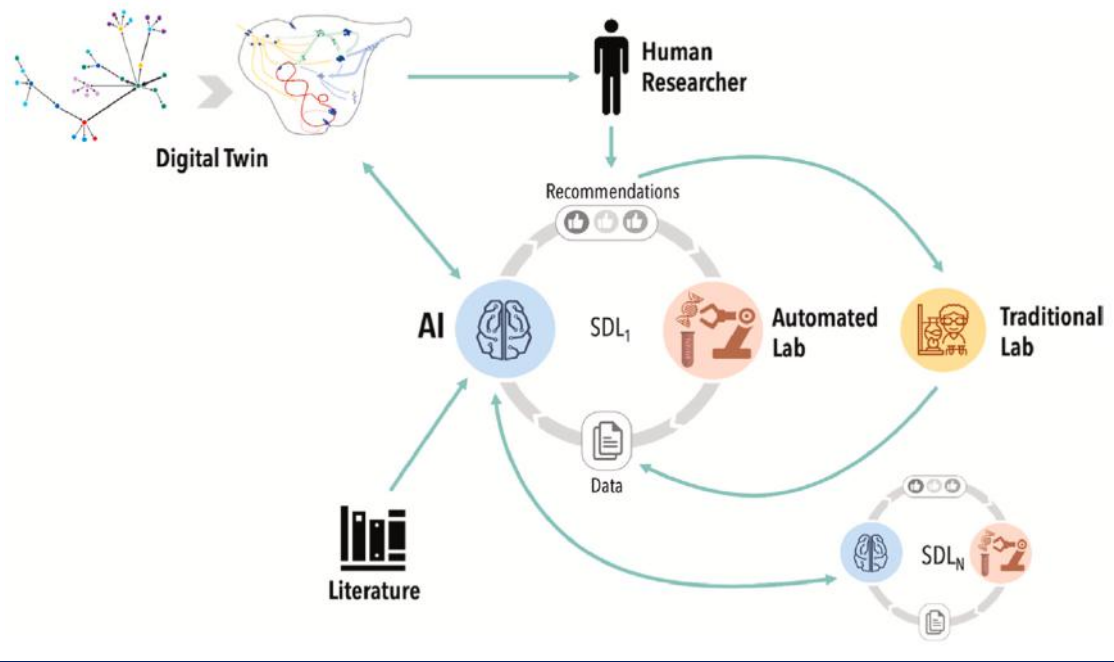
1. 针对生物工程微生物菌株 TRY（Titer，产量；Rate，速率；Yield，收率）的系统性提升是一个重要挑战。在开发商业可行性的过程中，一个重要障碍是达到经济可行的生物生产小分子的 TRY 水平。传统方法涉及基于菌株特定的深入代谢知识的试探性组合过程（即“拉-推-阻”方法），但在其他产品、途径和宿主中转移效果不佳；
2. 调控网络的建立。在预测一个生物体的新陈代谢时，了解它的调控机制可能才是最核心的，涉及到对其基因组大部分组成的机制理解；
3. 揭示基因型与表型之间的联系。这个挑战可以说是生物学的核心问题，我们仍然无法准确、定量地预测一个生物体在给定基因组情况下的行为；
4. 微生物群落的逆向设计。微生物群落展现了非凡的能力，从推动地球的生物地球化学循环到提高农作物产量。但我们目前缺乏设计满足特定规范的微生物群落的知识，例如，设计以稳定速率从废水中去除磷的微生物群落；
5. 探索地球以外的生物行为。了解生物系统对太空或其他行星/卫星环境的反应，对于推动太空探索和人类在地球之外的繁衍至关重要。然而，由于将人类和设备运送到轨道和轨道之外的物流成本非常高昂，因此太空中的劳动力和设备非常有限。

上述每个挑战都需要为相应的 SDLs 设置不同的机器设备。每个 SDL 的成本都与其功能直接相关：大范围使用复杂功能的 SDLs 成本高昂，而较简单功能的 SDLs 价格则可能相对较低。

SDLs 在科学领域最重要的影响或许来自于其能够自动构建科学知识的能力，科学知识指的能够解释和预测所研究系统的行为的事实、规律和理论的概括性体系。我们设想 SDLs 能够根据需要，从先前的知识和外部资源中提取并进行实验来改进这些知识。这种改进将体现为对机制的更深入理解和更强大的预测能力。我们也设想 SDLs 将以数字孪生的形式存储其积累的知识，其作用

将随着对所分析的生物系统的了解增加而发生演变。数字孪生是现实世界产品、系统、生命、社区甚至城市的虚拟复制品，并已成为工业界的关键资产。数字孪生的最初作用只是为了确定部件及其相互关系而提出的实验建议。一旦产生了足够的实验数据来识别这些关联，数字孪生的作用将成为设计实验，提出确定哪些关联是因果关系。当因果关系阐明，数字孪生的作用将变为设计实验，验证一个能够解释和定量预测这些因果效应的机制理论。一旦这个理论被校准，数字孪生的角色将转变为设计实验，构建满足预期规格的新生物系统（逆向设计）。这种构建科学知识的策略将涉及混合方法，即将纯粹的 SDLs 与人类、传统实验室和现有文献相结合。

图表 27. 我们设想 SDLs 将与其他 SDLs 和人类在一个网络中工作。SDLs 必须能够从现有的文献中获取信息，才能相对于当下科学知识的状态取得进展。SDLs 还应能够与其他 SDLs 进行通信，以便高效地划分待探索的科学相空间（即由所有可能的可配置实验参数选择组成的抽象空间）（例如 SDL1 关注一组启动子，SDL2 关注另一组启动子）。目前的技术限制了我们自动化所有实验的能力，因此 SDLs 应该能够产生明确的指令，供人类在传统实验室中遵循，并摄取所产生的数据。该网络运行的最终结果将是所研究的系统进行数字孪生。这个数字孪生可能最初是对系统部件及其连接的简单和定性描述，随着获得新信息的增加，它将发展成更复杂的机制模型和对所研究系统的定量预测模型，如整体细胞模型，具有准确的预测。人们将使用这些数字孪生来访问这个混合网络生成的科学知识，并提出自己的建议。



资料来源：ScienceDirect

3.3 实现 SDLs 的所需因素

完全实现 SDLs 需要许多项技术和社会的大幅度进步，差距涉及当前自动化技术、人工智能算法、数据管理以及重要的社会障碍的限制。

液体处理工作站应用于合成生物学自动化的过程正蓬勃发展，然而上文讨论的 SDLs 自动化的过程仅是合成生物学所需过程的一部分。典型的分子生物学过程，如细胞转化、菌落挑选、平板培养和生长，在液体处理工作站和其他仪器中虽然可行，但很难以 SDLs 所需的无缝方式链接在一起。最新的微流控技术通过将细胞和试剂封装在液滴中并对其进行精确操作，提供了实现这种无缝集成的机会。事实上，我们发现微流控平台已经被用于小型化生物反应，包括 DNA 合成和组装、转化、无细胞表达，以及荧光和质谱法进行表型筛选。通过将 these 功能与嵌入在芯片上的新型分子传感器、光学激活微小传感器、通过荧光监测自由基、通过光遗传学调节代谢或通过光操纵细胞等新技术相结合，可以实现真正的颠覆性功能。来自生物反应器的微流控采样还可以实现细胞在其生长环境中的实时传感和成像，实现连续的数据采集。这些微流控平台比移液工作站更经济实惠、使用的试剂更少，可以进行更多的实验，未来可期。

4 参考文献

- 王晓梅, 杨小薇, 李辉尚, 何微, & 辛竹琳. (2023). 全球合成生物学发展现状及对我国的启示. 生物技术通报, 39(2), 11.
- 杨永富, 耿碧男, 宋皓月, 胡蜜蜜, 何桥宁, & 陈守文等. (2021). 合成生物学时代基于非模式细菌的工业底盘细胞研究现状与展望. 生物工程学报, 37(3), 37.
- 赵国屏. (2018). 合成生物学:开启生命科学"会聚"研究新时代. 中国科学院院刊, 33(11), 1135-1149.
- 赵国屏. (2022). 合成生物学:从"造物致用"到产业转化. 生物工程学报, 38(11), 11.
- 张晓龙, 王晨芸, 刘延峰, 李江华, 刘龙, & 堵国成. (2021). 基于合成生物技术构建高效生物制造系统的研究进展. 合成生物学, 2(6), 13.
- 朱国广, 周新明. (2023). 合成生物学深度报告: 合聚万物, 成致未来. 东吴证券研究报告
- 刘宇腾. (2022). 合成生物学: 底层技术成熟促进行业高速发展. 东北证券研究报告
- 方亮, 屈昀, 陈乐益, 肖方, 郑仁福. (2022). 合成生物学行业概览. 兴业研究医药行业专题报告
- 方亮, 屈昀, 陈乐益, 肖方, 郑仁福. (2022). 合成生物学技术及应用. 兴业研究医药行业专题报告
- 合成生物学 2022 年专题信息合集 - 国家重点研发计划“合成生物学生物安全”项目组, 中国科学院上海营养与健康研究所, 生命健康科技智库, 上海市生物工程学会
- 选品、量产与定位, 合成生物学的下一步该如何走? -动脉网
- 4Q 2021 Synthetic Biology Venture Investment Report – Synbiobeta
- 卢挥, 张芳丽, & 黄磊. 合成生物学自动化装置 ibiofoundry 的构建与应用. 合成生物学, 1-14.
- Jinming, C., Bingzhao, Z., Yingfei, M. A., Xiongfei, F. U., Meng, W., & Chenli, L. . (2018). Engineering platforms for synthetic biology research. (11).
- Yu, T., Boob, A. G., Singh, N., Su, Y., & Zhao, H. (2023). In vitro continuous protein evolution empowered by machine learning and automation. Cell systems, S2405-4712(23)00115-1. Advance online publication.
- Martin, H. G., Radivojevic, T., Zucker, J., Bouchard, K., Sustarich, J., Peisert, S., Arnold, D., Hillson, N., Babnigg, G., Marti, J. M., Mungall, C. J., Beckham, G. T., Waldburger, L., Carothers, J., Sundaram, S., Agarwal, D., Simmons, B. A., Backman, T., Banerjee, D., Tanjore, D., ... Singh, A. (2023). Perspectives for self-driving labs in synthetic biology. Current opinion in biotechnology, 79, 102881.
- Gurdo, N., Volke, D. C., McCloskey, D., & Nikel, P. I. (2023). Automating the design-build-test-learn cycle towards next-generation bacterial cell factories. New biotechnology, 74, 1-15.
- Carbonell, P., Radivojevic, T., & García Martín, H. (2019). Opportunities at the Intersection of Synthetic Biology, Machine Learning, and Automation. ACS synthetic biology, 8(7), 1474-1477.
- Hamedirad, M., Chao, R., Weisberg, S., Lian, J., Sinha, S., & Zhao, H. (2019). Towards a fully automated algorithm driven platform for biosystems design. Nature communications, 10(1), 5150.
- Chao, R., Liang, J., Tasan, I., Si, T., Ju, L., & Zhao, H. (2017). Fully Automated One-Step Synthesis of Single-Transcript TALEN Pairs Using a Biological Foundry. ACS synthetic biology, 6(4), 678-685.
- Liu, L., Huang, Y., & Wang, H. H. (2023). Fast and efficient template-mediated synthesis of genetic variants. Nature methods, 20(6), 841-848.
- Enghiad, B., Xue, P., Singh, N., Boob, A. G., Shi, C., Petrov, V. A., Liu, R., Peri, S. S., Lane, S. T., Gaither, E. D., & Zhao, H. (2022). PlasmidMaker is a versatile, automated, and high throughput end-to-end platform for plasmid construction. Nature communications, 13(1), 2697.
- Smanski, M. J., Mead, D., Gustafsson, C., & Thomas, M. G. (2017). Meeting Report for Synthetic Biology for Natural Products 2017: The Interface of (Meta)Genomics, Machine Learning, and Natural Product Discovery. ACS synthetic biology, 6(5), 737-743.

公司简介

上海汉赞迪生命科技有限公司（BioHandler）是中国领先的生命科学全流程智能化解决方案提供商。作为“生命科学全流程智能化引领者”，汉赞迪业务涵盖研发、生产、销售和服务全生命周期，努力推动行业数智化转型升级。

汉赞迪依托资深创始团队十余年生命科学领域深耕与积淀，以模块化、智能化、简洁化的极致创新设计理念，设计开发具有自主知识产权的智能化产品。作为卓越的全球生命科学全流程产品和服务提供商，汉赞迪积极把握生物经济蓬勃发展机遇，以“让科技赞美生命，实现科学的承诺”为使命，用统一的生命科学语言对话科学家，精准地翻译和转化生命科学的需求。现已开发众多核心专利技术，积极创新营销管理与商业模式。

汉赞迪遵循“第一性原理，结构化思维”的方法论价值观；奉行“激情、创新、独立判断、利他主义”的组织素养；践行“简化已知，探索未知”的行事原则；坚持自主创新，科创报国。汉赞迪将致力于深度赋能生命科学领域，改变全球行业格局，打造卓越国际品牌，让中国智造享誉全球。

生命科学全流程智能化解决方案

智慧实验室解决方案

智慧实验室

定制化物联网

全流程自动化实验室

实验室自动化整合平台软件



合成生物学

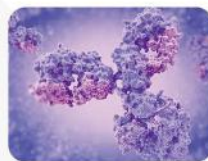


基因治疗和核酸药物



一体化整合平台

NGS建库	药物发现和筛选
PCR系统	核酸提取和分子检测
ELISA系统	质谱样品处理
分血系统	流式样本前处理
细胞株开发	类器官样品处理
菌株筛选	合成生物学筛选和检测



抗体药发现



细胞治疗

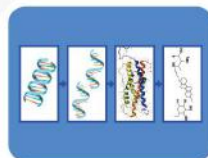
标准化核心单机

CANTUS系列液体处理系统

NEMO系列移液工作站

VOLA系列核酸提取仪

ARCO系列分液器



多组学

公司简介

比邻星创投专注于孵化、投资和发展出类拔萃并具有巨大潜力的新一代医学技术创新企业

由创投经验丰富的专业人士和产业专家于 2016 年底创立，专注于投资具有全球领先技术、有巨大发展潜力的医疗器械和生物技术企业。投资团队具有深厚医疗产业背景和丰富的临床转化经验，并具备深刻的行业洞察和广泛的国际医疗资源网络。基金在上海、苏州、台北、美国设有办公室。

目前，比邻星创投管理着多只人民币和美元基金，已经投出了健世科技、康沅生物、艾柯医疗、博辉瑞进、鹄远生物、睿笛生物、生命盾医疗、敦博医疗、纳米维景、昂泰微精、高诚生物、大湾生物、杰毅生物、德运康瑞、大橡科技等创新企业。

比邻星创投通过倾力汇聚各类资本，发掘全球领先医疗创新技术，打造有优质临床资源和产业资源的医疗创业生态系统；立志全方位陪伴被投资企业，将其发展成为具有强大竞争力和持续创新能力的伟大企业，共同为人类生命健康做出努力与突破，为社会创造长期价值。



Build Invest Grow

We together create breakthroughs in human health



投资成果

56

截至 2023 年 5 月
已投资 56 家企业

50+

支持了 50 多个
全球首创产品的诞生

300+

协助被投资企业引入
300 多位关键人才



白皮书电子版下载链接



BioHandler
汉赞迪科技



ProX 比邻星创投
VENTURES



上海汉赞迪生命科技有限公司

Shanghai Biohandler Life Sci-Tech Co., Ltd.

地址

上海总部：上海市闵行区东川路555号6号楼306B

北京公司：北京市海淀区永丰产业园中关村壹号A2座 602

广州办事处：广州市黄埔区南翔支路1号瑞粤汽车电子创新园A30

电话

86-21-34121223/131 6273 6602

网址

www.biohandler.com



比邻星创投

地址

上海市徐汇区吴兴路45号

网址

<https://www.proximavc.com>