

特约评述

DOI: 10.12211/2096-8280.2023-096

合成基因线路的工程化设计研究进展与展望

高歌^{1,2}, 边旗^{1,2}, 王宝俊^{1,2}⁽¹⁾ 浙江大学化学工程与生物工程学院, 浙江 杭州 310058; ⁽²⁾ 浙江大学杭州国际科创中心, 浙江 杭州 311200

摘要: 合成基因线路利用合成生物学的技术和方法, 将生物元件进行重新设计与构建, 使人工设计的生物分子线路在活细胞中行使特定生物功能, 在生物制造、医疗健康以及环境监测等领域具有巨大的潜力。但其工程化设计仍受到各种因素的制约, 包括正交元件数量有限、大规模线路组装困难、线路行为预测性低等。根据研究者们开发的各种调控元件工具箱和组装方法, 本文逐点阐述了工程化设计基因线路所需遵循的几个核心原则: 正交化、标准化、模块化与自动化。文章从DNA复制、转录和翻译层面介绍了正交基因元件库的构建和改造方法; 全面总结了基因元件的标准化定量表征方法与标准元件设计方法; 并介绍了本团队与其他团队在模块化基因线路设计方面的相关进展; 分别从软件、硬件和人工智能角度展示如何实现基因线路的自动化设计。最后, 本文探讨了基因线路设计的未来发展趋势, 指出需要进一步融合人工智能和自动化等信息技术来加速基因线路“设计-构建-测试-学习”循环的迭代, 提高线路设计的功能可预测性和复杂性, 高效设计出符合目标需求的人造生命体。

关键词: 合成生物学; 基因线路设计; 正交化; 标准化; 模块化; 自动化

中图分类号: Q81 文献标志码: A

Synthetic genetic circuit engineering: principles, advances and prospects

GAO Ge^{1,2}, BIAN Qi^{1,2}, WANG Baojun^{1,2}

⁽¹⁾ College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310058, Zhejiang, China; ⁽²⁾ ZJU-Hangzhou Global Scientific and Technological Innovation Center, Zhejiang University, Hangzhou 311200, Zhejiang, China

Abstract: Synthetic genetic circuits are engineered gene networks comprised of redesigned genetic parts for interacting to perform customized functions in cells. With the rapid development of synthetic biology, synthetic genetic circuits have shown significant application potentials in many fields such as biomanufacturing, healthcare and environmental monitoring. However, the efforts to scale up genetic circuits are hindered by the limited number of orthogonal parts, the difficulty of functionally composing large-scale circuits, and the poor predictability of circuit behaviors. A longstanding goal of synthetic biology research is to engineer complex synthetic biological circuits, using modular genetic parts, as we do with electronic circuits. Synthetic biologists have developed various genetic toolboxes

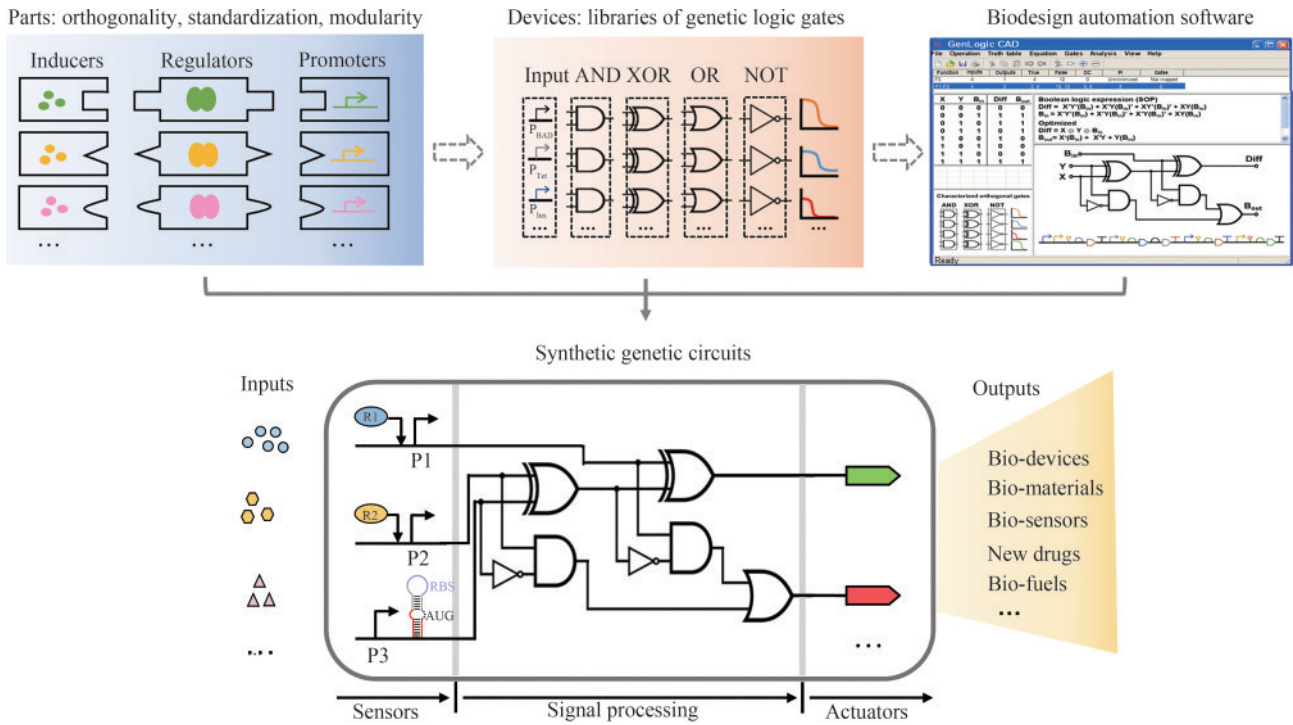
收稿日期: 2023-12-01 修回日期: 2024-04-10

基金项目: 国家重点研发计划“生物与信息融合(BT与IT融合)”重点专项(2023YFF1204500); 浙江省“尖兵”“领雁”研发攻关计划项目(2024C03011); 国家自然科学基金重点国际合作研究项目(32320103001); 国家自然科学基金面上项目(32271475); 中央高校基本科研业务费专项资金(226-2022-00214)

引用本文: 高歌, 边旗, 王宝俊. 合成基因线路的工程化设计研究进展与展望[J]. 合成生物学, 2025, 6(1): 45-64

Citation: GAO Ge, BIAN Qi, WANG Baojun. Synthetic genetic circuit engineering: principles, advances and prospects[J]. Synthetic Biology Journal, 2025, 6(1): 45-64

and functional assembly methods over the past few decades. Here we present an overview of the latest advances, challenges, and future prospects in genetic circuit engineering from four aspects corresponding to the four key engineering principles for circuit design, *i. e.* orthogonality, standardization, modularity, and automation. Firstly, the design and construction of orthogonal genetic part libraries are discussed in both prokaryotes and eukaryotes at the levels of DNA replication, transcription, and translation, respectively. Standardized characterization methods and the design of modular genetic parts are subsequently summarized. Furthermore, progress in developing modular genetic circuits are presented, providing new concepts and ways for engineering increasingly large and complex circuits. Finally, how to achieve automated design and building of genetic circuits are addressed from the advances in software, hardware and artificial intelligence, respectively, with an aim to replacing the presently time-consuming manual trial-and-error mode with the iterative “design-build-test-learn” cycle for improved efficiency and predictability of circuit design. The integration of these fundamental principles and the latest advances in information technology such as artificial intelligence and lab automation will accelerate the paradigm shift in genetic circuit engineering and synthetic biology research, making it feasible for designing synthetic lives to meet various customized needs.



Keywords: synthetic biology; genetic circuit design; orthogonality; standardization; modularity; automation

在过去二十几年中，合成生物学的基本原理和使能技术不断被拓展，在工业、农业、医疗保健、环境监测和科学研究等方面展现出了巨大潜力^[1-2]。合成基因线路作为合成生物学研究的一个重要方向，是利用合成生物学的技术和方法，将生物元件进行重新设计与构建，使人工设计的生物分子线路在活细胞中行使特定生物功能，具有

重大的生物学意义和应用价值。合成生物学基因线路设计主要受启发于电子线路设计中的工程化概念和原理，如正交化、模块化、标准化和自动化等指导原则，由此开发出了逻辑门^[3]、振荡器^[4]、双稳态开关^[5]、时序存储电路^[6]和模拟数字信号转换器^[7]等基因线路，它们处理来自输入生物传感器的信号，并控制执行器生成定制输出

响应^[8-9]。

然而，合成基因线路的工程化创建仍然受到各种因素的制约。基因线路受多种元素调节，其中多数依靠干扰转录启动来发挥作用，但用于调节翻译速率、蛋白质相互作用和DNA修饰的方法正在扩展^[10]。但是，这些新的调节方式在不同的环境背景中功能无法可靠预测，需要在每次更换使用背景时重新表征^[11]。这严重阻碍了复杂基因线路的构建进程和合成生物学的多场景应用。例如，在国际基因工程机器大赛（iGEM）标准生物部件数据库中所含部件虽然规模很大，但大多数未经统一标准表征，性能通常不可靠，从而降低了标准生物部件的实用性和通用性^[12]。因此，基因线路设计迫切需要运用工程化的设计原理和方法，在正交化、模块化、标准化和自动化这几大

核心指导原则下，逐步构建出高质量的大规模生物部件集合，最终实现可预测的基因线路精准设计。

1 合成基因线路的正交化设计进展

生物正交反应是指能够在生物体系中进行，且不会与天然生物化学过程相互干扰的一类化学反应^[13]。在基因线路设计过程中，为了精准地调控细胞内各组分的活性，应尽量减少元件之间的相互干扰，这就是基因线路中的“正交化”。随着合成生物学的快速发展，科学家们在遗传信息表达及生化代谢的不同过程中成功创建了正交元件（图1和表1）。

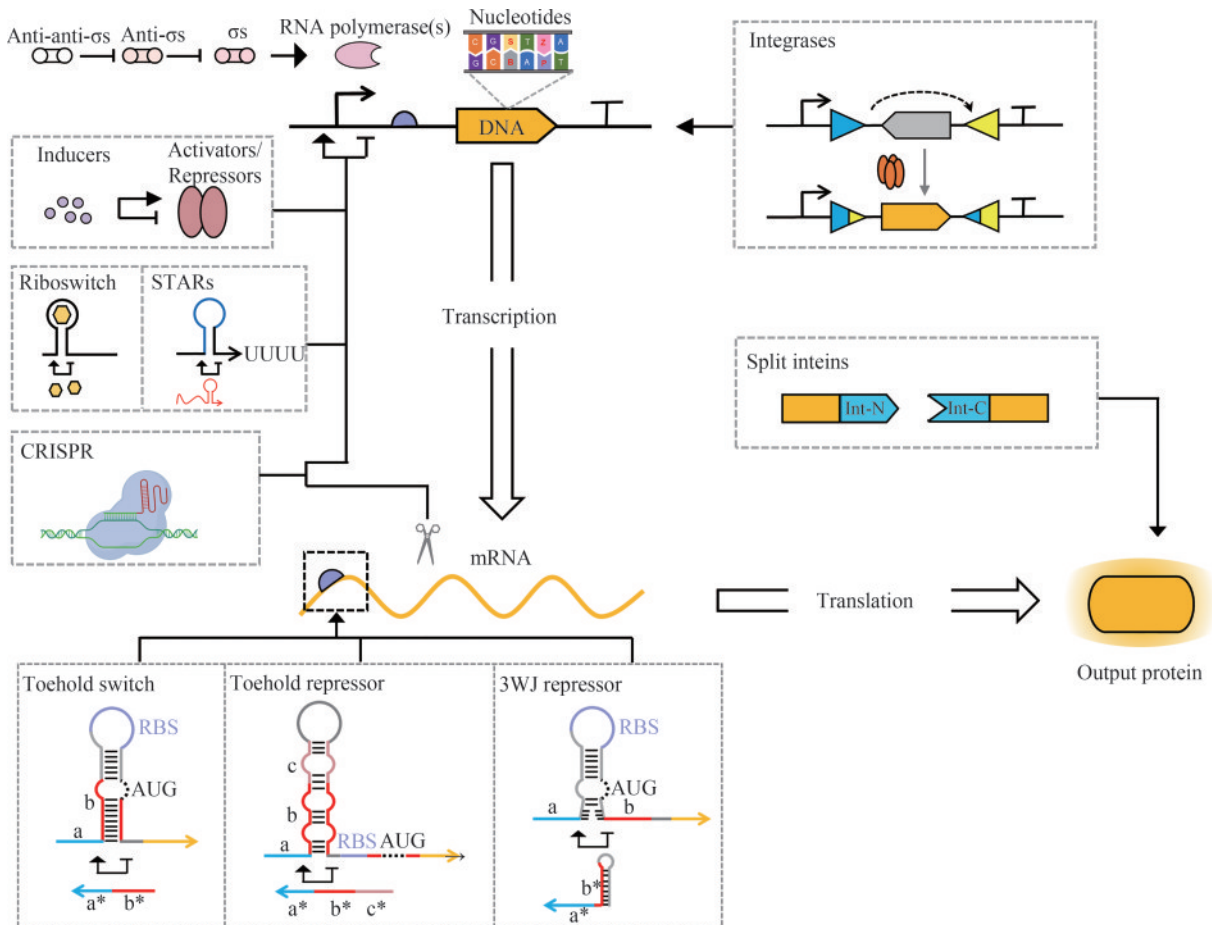


图1 合成基因线路中已经验证的用于基因表达控制的正交元件与调控工具

(合成基因元件可以在遗传信息表达的不同过程中发挥调控作用，包括DNA存储与复制^[14-20]、转录^[21-45]、翻译^[46-47]以及翻译后调控^[48-50])

Fig. 1 Validated orthogonal parts and tools for precise gene expression control in genetic circuit design

(Synthetic genetic parts can regulate various steps of gene expression, including DNA storage and replication^[14-20], transcription^[21-45], translation^[46-47], and post-translational regulation^[48-50].)

表1 典型正交基因元件库的设计与表征

Table 1 Design and characterization of the libraries of orthogonal genetic parts

| 元件名称 | 获得正交元件的方法 | 元件数量 | 正交元件数量 | 参考文献 |
|------------------|---------------------------------|------|--------|------|
| T7 RNAP 突变体 | 毒性降低的 T7 RNAP 突变体 | 4 | 4 | [21] |
| | 活性变高的 T7 RNAP 突变体 | 6 | 6 | [22] |
| ECF- σ 因子 | 可替换的 ECF- σ 因子用于同源启动子的激活 | 52 | 20 | [23] |
| LacI 突变体 | N 端序列突变的 LacI 与突变的 LacO 操纵子 | 5 | 5 | [26] |
| CI 突变体 | 基于噬菌粒的定向进化 | 12 | 6 | [27] |
| TetR 同系物 | 元件挖掘并鉴定 TetR 家族类似抑制子 | 20 | 17 | [28] |
| 可诱导表达系统 | 金属离子(由金属离子诱导的调控因子和相应的启动子) | 5 | 5 | [29] |
| | 小分子(插入基因组的小分子生物传感器) | 12 | 12 | [31] |
| | 代谢物(代谢的多样性) | 14 | 12 | [33] |
| 群体感应 | 对信号、遗传串扰优化后的群感调控因子和启动子 | 4 | 2 | [34] |
| | 突变 pLux 启动子序列 | 12 | 2 | [35] |
| | 对不同来源群体感应系统进行同源和非同源表征 | 6 | 3 | [36] |
| | 群感信号配体的筛选 | 10 | 6 | [39] |
| STARs | 目标 RNA 与小转录激活 RNA | 100 | 6 | [41] |
| CRISPRi | 高度非重复的超长 sgRNA 阵列 | 22 | 13 | [44] |
| CRISPRa | 修饰的 sgRNA 与 sigma 54 激活因子 | 5 | 5 | [45] |
| 核糖调控 | Toehold switches | 144 | 26 | [46] |
| | Toehold repressors | 95 | 15 | [47] |
| 断裂内含肽 | 元件挖掘并测试不同的断裂内含肽交叉活性 | 34 | 15 | [48] |
| 遗传密码子 | 筛选技术:tREX | 71 | 23 | [49] |

1.1 正交的DNA存储与复制

经典的中心法则中遗传信息通常是从 DNA 转录成 RNA, 然后再翻译成蛋白质进行表达。在该过程中, 除了经典的碱基互补配对, 核酸还可以通过(化学修饰主导的)表观遗传系统来扩展自身信息编码的能力。N6-甲基腺嘌呤(m6A)在原核生物中普遍存在, 但在真核生物基因组中很少发现, 将必要的甲基转移酶和转录因子转移到哺乳动物细胞中, 可实现高效和正交的信息存储和传播。Park 等^[14]使用 m6A 在哺乳动物细胞中建立了一个人工的正交表观遗传调节系统, 可以人为地修饰并控制基因的表达, 从头编程表观遗传功能, 从概念上验证“读-写”体系结构。非经典碱基可促使合成核苷酸引入 DNA, 通过增加序列多样性来扩展遗传密码。已有研究从经典的 4 个编码碱基增加到 6 个^[15]或者 8 个^[16], 这大大增加了可能的信息密度, 同时也降低了碱基与宿主成分相互作用的可能。

正交 DNA 复制系统是独立于宿主 DNA 复制,

主要由正交的低保真 DNA 聚合酶和正交 DNA 模板组成。Ravikumar 等^[17]在酵母体内建立了一种正交 DNA 复制系统(OrthoRep), 该系统基于依赖特殊的 TP-DNAI1 聚合酶引发自主复制的细胞质线性质粒, 具有不同于核基因组的基因转录系统。该实验室的研究者对 OrthoRep 的基因表达元件进行了改造, 获得了带有不同突变序列的启动子, 结合使用聚腺苷酸尾巴获得差异化表达强度的元件^[18]。然而, 酵母中由 OrthoRep 进化而来的基因表达元件在原核生物中可能不起作用, Tian 等^[19]依赖于稳定的正交线性质粒和正交易错的 DNA 聚合酶突变体开发了一种基于正交易错 DNA 复制机制的细菌连续进化系统(BacORep), 可用于开发细菌使用的基因表达元件、基因簇或蛋白质, 是持续进化的强大工具, 为合成生物学产生量身定制的各种元件。

重组酶系统只需要重组酶、重组方向性因子(recombination directionality factor, RDF)和特异性识别位点即可实现永久、可逆和可遗传的 DNA 重排, 且与其他调控手段相比, 能将代谢负荷最

小化^[51]。与酪氨酸重组酶相比，丝氨酸整合酶的识别位点简单，通常只需要整合酶蛋白和 *att* 位点 (<50 bp)，在基因线路中应用广泛。Yang 等^[20] 通过元件挖掘来识别假定的噬菌体整合酶，并表征获得了一个由 11 对整合酶和同源识别位点组成的正交文库，进一步增加了储存器的容量。

1.2 正交的转录调控

大多数基因线路是通过转录来控制基因的表达，该过程可以简化为 RNA 聚合酶 (RNAP) 沿着 DNA 链进行移动，包括 RNAP 复合物的形成、结合和分离，由启动子、转录因子 (激活因子和抑制因子)、RNAP 亚基与终止子组成。同时，转录因子的活性可以通过外源添加或培养过程中检测到的诱导剂进一步调节。此外，RNA 的高度可编程性为转录调控提供了新的方式。

1.2.1 蛋白转录调控

通过控制转录的第一步 (即利用独特的 RNAP 与特异性启动子序列结合) 对基因线路进行调控已被广泛应用。其中，T7 RNAP 对 T7 启动子序列具有高度特异性。Temme 等^[21] 鉴定了 4 个正交的 T7 RNAP 突变体，且每个突变体都有一组不同强度的同源启动子。Meyer 等^[22] 采用定向进化策略筛选到能够识别新启动子的 T7 RNAP 突变体，在体内和体外都达到了更高的活性。

细菌 σ 因子和 anti- σ 因子是 RNAP 的基本组成部分，它们在不同的条件下对启动子产生不同的影响。anti- σ 因子与其同源的 σ 因子结合，抑制其与 RNAP 结合。在外部刺激下，anti- σ 因子与其同源 σ 因子分离，从而允许 σ 与 RNAP 相互作用，并启动其同源启动子的转录^[52]。虽然许多刺激因素尚不清楚，但在大肠杆菌中，许多 σ 因子表现出相互的和宿主的正交性。其中，细胞质外功能 (extra cytoplasmic function, ECF) σ 因子最小，序列也最多样化。Rhodius 等^[23] 在大肠杆菌中研究了不同来源的 86 个 ECF- σ 因子及其启动子，发现 20 个 σ 因子-启动子对是相互正交的。同时，将 ECF 的 N 端和 C 端结构域重组获得了一个更大的嵌合 ECF 文库，且具有更好的正交响应范围。作者还确定了正交的 anti- σ 因子，提高了系统的协作

性。Bervoets 等^[24] 在大肠杆菌中异源表达枯草芽孢杆菌来源的 σ 因子，并创建了启动子文库，在保证正交性的同时拓展了转录起始强度的范围。本团队在大肠杆菌中异源表达铜绿假单胞菌来源的 σ^{54} 因子依赖性转录因子 (HrpR/HrpS) 及其启动子 (*hrpL*)，该转录调控模块与大肠杆菌的遗传背景和内源 σ^{70} 因子转录调控系统正交，被首次用于正交化基因逻辑与门的设计与构建^[25]。

许多最早的基因线路都是基于三种抑制因子^[4, 53-55]，即 LacI、 λ cI、TetR，它们通过与启动子附近的区域结合并阻断 RNAP 的进入，最终阻止转录起始而发挥作用。Zhan 等^[26] 对转录因子 (LacI) 和操纵子 (O_{lac}) 采用融合工程，产生了一系列全新的相互作用的特异性转录因子和操纵子，经表征具有良好的正交性。有研究者基于噬菌粒的方法获得了一组工程化的正交转录因子 (λ cI) 和启动子^[27]。Stanton 等^[28] 从不同来源的具有已知调节功能的 TetR 同源物列表中收集了 73 个抑制因子，并用体外法测定它们的操纵子序列，构建启动子，进一步筛选抑制因子和启动子的交叉反应，发现其中有 16 个被鉴定为既强烈抑制其同源启动子 (5~207 倍) 又不与其他启动子相互作用。

可诱导表达系统能够响应细胞内外的信号，也常常用于基因线路设计中。诱导剂可以是金属离子、小分子、光和环境因素 (如温度和 pH 值)。本团队开发了一系列对金属离子响应的新型小分子转录诱导系统，包括砷、汞、铜、锌离子，并构建了二输入和三输入与门生物传感器用于水质的检测^[29]。此外，基于汞生物传感器，本团队在全细胞体系和无细胞体系中系统地表征了 8 个报告基因，包括各种频谱正交的荧光蛋白、荧光素酶，有助于快速开发全细胞和无细胞生物传感器^[30]。Meyer 等^[31] 通过对调控蛋白和输出启动子进行随机突变和理性设计，优化了 14 个小分子诱导系统的响应曲线，获得了 12 个高灵敏度、低泄露且正交的高性能传感器。本团队通过对细胞内受体浓度的系统变化进行转录控制，开发了调节小分子转录诱导系统动态范围和敏感性的调控方法，可对小分子诱导的基因表达系统进行灵活的调整^[56]。由于转录抑制因子数量不多，其在合成生物学

中尚未被充分利用, Groseclose等^[32]设计了一组非天然抗抑制因子,对IPTG、果糖和D-核糖的反应中表现出反向变构活性,并将这些抗抑制因子与7个DNA识别区组装,创建一个抗诱导正交体系文库。除了外源添加,诱导剂也可以是细胞代谢物。Hanko等^[33]开发了一种通用的全基因组方法来识别基于转录因子的诱导型基因表达系统,并从*Cupriavidus necator* H16中鉴定和表征了14个调控因子,发现其中12个具有正交性。

细菌群体感应系统作为基因线路构建中常用的诱导系统和细胞间通信工具,得到了研究者的广泛关注和研究,文章《细菌群体感应系统在细胞间通讯中的应用及其合成生物学研究进展》^[57]、《细菌群体感应元件构建和工程应用》^[58]有详细的介绍,其中最常见的系统有*lux*、*las*、*rpa*和*tra*等。大多数通信系统都是基于自诱导物高丝氨酸内酯(HSL),化学结构相近,并且不同的通信系统之间也存在着一定的串扰^[34],因此提高群体感应系统的正交性对复杂线路的构建意义非凡。Grant等^[35]通过对pLux启动子进行突变,并结合数学模型,得到了两个正交的输入和输出。Kylilis等^[36]分别对6种群体感应系统进行同源和非同源表征,构建迄今最大的映射特征库,并开发一个工具软件可自动选择正交的通信通道,通过实验证明了在共培养中可同时使用三个通信通道。Wu等^[37]结合数学模型,对四种常见群体感应系统的启动子、调控因子和信号分子进行了组合和表征,筛选到17种变化倍数较高的组合,并计算串扰强度,为群感系统在混菌中的应用提供了理论基础。Du等^[38]利用生物小分子的化学多样性重新设计了一个遗传工具箱,获得6个正交性远超传统群体感应系统的细菌细胞信号通路。最近,Jonkergouw等^[39]基于配体的化学多样性,利用AI算法,挖掘和表征了不同类型的群体感应信号配体,获得了6个信号分子和启动子正交的群感系统。上述研究极大地扩展了合成生物学在多细胞生物工程中的能力,为细胞进行大规模生物计算提供了坚实的基础。

1.2.2 RNA转录调控

转录调控也可以通过RNA来实现,它们既可以形成二级结构来触发RNAP的解离,也可以引导

蛋白(CRISPR系统)来阻断或激活转录起始^[59]。RNA分子具有高度可编程性、可预测的碱基配对机制,以及比较明晰的热力学规律,这些特性使得其计算机设计和行为预测成为可能。

核糖开关(riboswitch)主要由感受外界配体的适配体和调控基因表达的表达平台两部分组成,是一种顺式调控元件。尽管利用核糖开关进行诱导调控和生物传感器具有巨大的潜力,但工程化合成核糖开关仍存在困难。为此,Espah Borujeni等^[40]开发和优化了一个生物物理模型,该模型可根据RNA适配体序列和结构生成优化的核糖开关。通过该法,研究者创建了62个合成的核糖开关,并用不同的RNA适配体结合各自的配体,将基因的激活表达提高了不同的倍数,最高可达383倍。

在反式作用的调控元件方面,基于RNA发夹结构的条件性形成,Chappell等^[41]创建了一种合成的sRNA转录激活因子,称作小转录激活RNA(STAR, small transcription activating RNA),由相互作用的目标RNA和STAR组成。目标RNA位于待调节基因上游,一旦转录,该目标RNA折叠形成内在的终止子发夹结构,阻止下游基因的转录;当STAR表达时,与目标RNA相互作用,防止发夹结构形成,继而允许下游基因转录。同时,作者构建了一个高性能和正交STAR文库,可实现约9000倍的基因激活。

CRISPR系统结合了RNA分子的可编程靶向性和蛋白质结构的鲁棒性,是一种强大的工具。在基因线路中,失去切割DNA功能的Cas9突变体dCas9蛋白可对转录进行调控,即dCas9靶向原核启动子区域时能够物理性阻碍RNAP的通过(CRISPRi),导致基因表达受到抑制,或者dCas9与转录激活因子融合时可靶向启动子和增强子区域(CRISPRa),导致基因表达上调。对于转录抑制,已有研究者创建了正交的gRNA间隔区-启动子文库^[41-42]。Reis等^[44]合理地设计了高度非重复的超长sgRNA阵列,可稳定共表达22个sgRNA,实现多达13个基因的同时正交抑制。

1.3 正交的翻译调控

类似于前文提到的基于RNA的转录调控,利

用反式RNA的链置换反应来破坏RNA的二级结构可影响核糖体与RBS的结合。基于简单的5'发夹结构设计的核糖调节因子动态调控范围低,且受到序列的限制^[60]。为了克服上述限制,Green等^[46]提出了一种新的核糖调控机制,称为支点开关(toehold switch),由触发RNA和开关RNA组成。触发RNA通过单链支点区域与具有翻译抑制发夹结构的开关RNA结合,导致开关RNA茎展开,从而暴露核糖体结合位点和起始密码子以激活输出基因的翻译。由于触发RNA可以是任意序列,作者创建了高度正交的支点开关文库。Kim等^[47]还开发了支点抑制系统,获得了15对相互正交的开关RNA和触发RNA,动态范围高达300倍。

1.4 正交的翻译后调控

除了转录和翻译调控外,还可以通过翻译后蛋白质的相互作用来设计基因线路。内含肽(intein)是一种可以实现蛋白质剪接的特殊多肽,介导将自身片段从肽段中“切除”,并将相邻肽段拼接在一起的“蛋白质剪接”过程。而断裂内含肽可以介导两个独立肽链的反式剪接,能帮助科学家在翻译后水平将不同肽段无缝组装成设计需要的功能蛋白。本团队建立了一个高通量内含肽筛选平台,同时测试了34种不同的内含肽元件,进而探索了用这些内含肽元件设计断裂内含肽的活性断裂位点,并进行了系统的正交性测试,获得了15对高度正交的断裂内含肽元件,构建出迄今为止最大规模的正交断裂内含肽元件库,该元件库可用来断裂正交转录因子,快速设计正交化和模块化的基因逻辑与门器件库和复杂逻辑计算线路^[48]。

遗传密码子拓展技术能够将非天然氨基酸(ncAA)特异性地引入目标蛋白,可丰富蛋白质的功能,为基因线路的设计提供了新契机。《基因密码子拓展技术的方法原理和前沿应用研究进展》^[61]一文中对氨酰-tRNA合成酶/tRNA配对的正交性以及改造进行了详细的论述,但正交对的数量仍有限。最近,Cervettini等^[49]开发了一种筛选技术tREX可快速确定tRNA体内氨酰化状态,在大肠杆菌中测试了243个候选tRNA,并鉴定出71个正

交的tRNA,且其中的23个与同源氨酰-tRNA合成酶正交,并优化了5个正交的tRNA-氨酰-tRNA合成酶对,该研究扩大了可用于遗传密码子拓展技术的正交对数量。Dunkelmann等^[50]基于缺乏N端结构域的PylRS基因(Δ NPylRS),从测试的88个 Δ NPylRS/^{ANPyl}tRNA元件中确定了18个相互正交的tRNA-aaRS组合,并生成了12组三正交对,最终在一个多肽中实现了三种不同的非天然氨基酸的整合。

2 基因元件与线路的标准化设计进展

基因元件是指天然存在的基本DNA功能片段,如启动子、核糖体结合位点、功能基因、终止子等。研究者根据动力学模拟参数以及载体和宿主的特性,在基因元件头尾两端加上特定的酶切位点,并采用统一的描述与分类方法使之标准化,从而更有效率地进行查询、设计与基因操作^[62]。随着人们对基因元件、遗传信息的开发研究不断深入,生物序列的不可见、易错配等问题给后续设计操作带来较大不便。因此,为了使生物合成更加简易快捷,需组建标准化的元件库。2003年,美国麻省理工学院建立了标准元件库(<http://parts.igem.org/Collections>),收集注册各种标准化生物元件,该元件库至今已包含启动子、核糖体结合位点、终止子等超过20 000种标准化生物元件^[12]。在常用的底盘细胞中,已构建了标准元件的标准化装配方法^[63],如BioBrick assembly、Gibson assembly、Golden Gate assembly以及ligase cycling reaction(LCR)。

同时,参照原核细胞的生物元件设计和构建方法,Guo等^[64]开发了酿酒酵母标准生物元件和标准化组装方法(YeastFab Assembly),可在一周内完成异种代谢途径从克隆到组装。Rajkumar等^[65]使用MoClo^[66]/Yeast Tool Kit^[67]的组装方法快速构建和表征了马克斯克鲁维酵母菌的标准元件库。Wei等^[68]以丝状真菌构巢曲霉为研究模型,建立了基于单细胞定量表征的丝状真菌启动子高通量筛选方法,筛选出93个与次级代谢相关的启动子,对其表达强度进行了精确评估,获得了相对表达强度高达37倍的天然启动子文库。

除去对不同底盘标准元件库的扩充,对单个标准元件的特性也设计出了元件内的分区调试方案。Redden等^[69]分析测试了酵母中的最小核心启动子序列,获得了一组非同源、纯合成的最小启动子,其中的一个116 bp的组成型启动子,强度接近常见的酵母启动子*GPD* (655 bp)。当这些最小启动子与最小终止子(35~70 bp)结合时^[70],表达盒的整体DNA负载可以减少80%~90%,提高了酵母中大规模合成系统的易用性,并进一步扩展了合成生物学标准元件库。从头合成启动子可以使人们获得具有新的结构特性和活性的启动子元件,但是这种从头合成的方式通常需要从海量的随机序列中筛选目标启动子,这需要耗费巨大的工作量,并且需要昂贵的高通量筛选设备^[71]。

同时,在基因元件的表征方面,目前缺少标准化的评定方法,这就造成已有的如Registry for Standard Biological Parts和SynBioHub^[72]基因元件存储库中的各元件之间无法进行准确的定量比较。研究者只能在同时表征大量基因元件的研究文献中获得具有可比性的数据,但归纳总结这类研究,并获得理想的数据十分低效耗时。为此,本团队对已经过实验验证和文献报道的标准基因元件及其表征数据进行了系统的分类,建立了BioPartsDB (<https://wanglab.net/biopartsdb/>)精选元件数据库^[73]。该数据库的展示方式与SynBioHub等存储库类似,基因元件目录按照作用机制(图2)分为启动子、核糖体结合位点、核糖调控因子、绝缘子等,按照其在基因线路中行使的功能分为输入、逻辑门、存储器等,同时也可以索引到该基因元件的原始出版物。在显示各基因元件类别的总览页面上点击任一类基因元件,会跳转到展示具体的表征数据、实验条件、核酸序列、相关基因元件及来源出版物的页面。该数据库目前涵盖了选自35篇出版物中的已有详细表征数据的417个基因元件。同时,组建国际合成生物联盟来推动合成生物系统设计构建过程中的数据共享和标准,及实验方案和元件的共享,开放标准、互通数据资源,将极大地促进基因元件和线路设计的标准化进程^[74]。

3 合成基因线路的模块化设计进展

合成生物学基因线路设计过程中,通常将整个复杂的线路分为几个小部分(模块),然后分别对它们进行表征,再组装在一起构成一个大的基因系统。每一个模块就像一个黑匣子,在最后组装时,研究者只需确认每个模块输入与输出的对应关系,而不需要清楚它们的内部结构^[75]。在理想情况下,模块间相互拼接时,各模块的输入与输出的对应关系不会发生改变,即模块化。

模块化、不受细胞类型影响且可编程的系统是合成生物学研究中的重要瓶颈问题,相关研究不断在扩充模块化的基因线路,使其不受转录、翻译资源影响,具备可预测性。本团队在标准化工作环境下充分测试基因线路各组成部件特性,利用铜绿假单胞菌中的异源 σ^{54} 依赖型转录因子和相关基因调控元件在大肠杆菌中率先设计并构建了模块化且正交化的逻辑与门和与非门控线路、模块化且放大倍数连续可调的转录信号放大线路[图3(a)和(b)]^[25, 76]。并通过结合现有的sgRNA支架设计和一种已被证明模块性较高的 σ^{54} 因子依赖型激活子PspF,在模式生物大肠杆菌(*Escherichia coli*)和固氮菌(*Klebsiella oxytoca*)中实现了类真核CRISPR激活,并且激活装置具有泄露表达水平低、激活倍率高等优异的调控性能^[45]。经过进一步的工程化改造和优化,其基因调控性能第一次达到了足以构建CRISPRa级联调控、CRISPRa正反馈调控和多通道基因调控的水平上。Li等^[77]在哺乳动物细胞中结合新设计的杂交启动子并通过关键转录起始元件的空间位阻发挥转录抑制作用,构建了包含26种可逆转录激活因子样效应抑制因子(transcription activator-like effector repressors, TALER)的文库,可以准确预测模块化组装的TALER级联和开关电路的行为。很多天然的转录因子已经被深入研究并应用于代谢通路动态调控和人工基因线路构建^[78]。然而,天然的转录因子的调控性能难以满足工程化设计的需求,传统筛选和定向进化方法可改善转录因子的性能,但相关实验周期长、工作量大。Hou等^[79]针对转录因子的协同性、超敏性需求,开发了一种模块化的协同性蛋白元件,将该元件与转

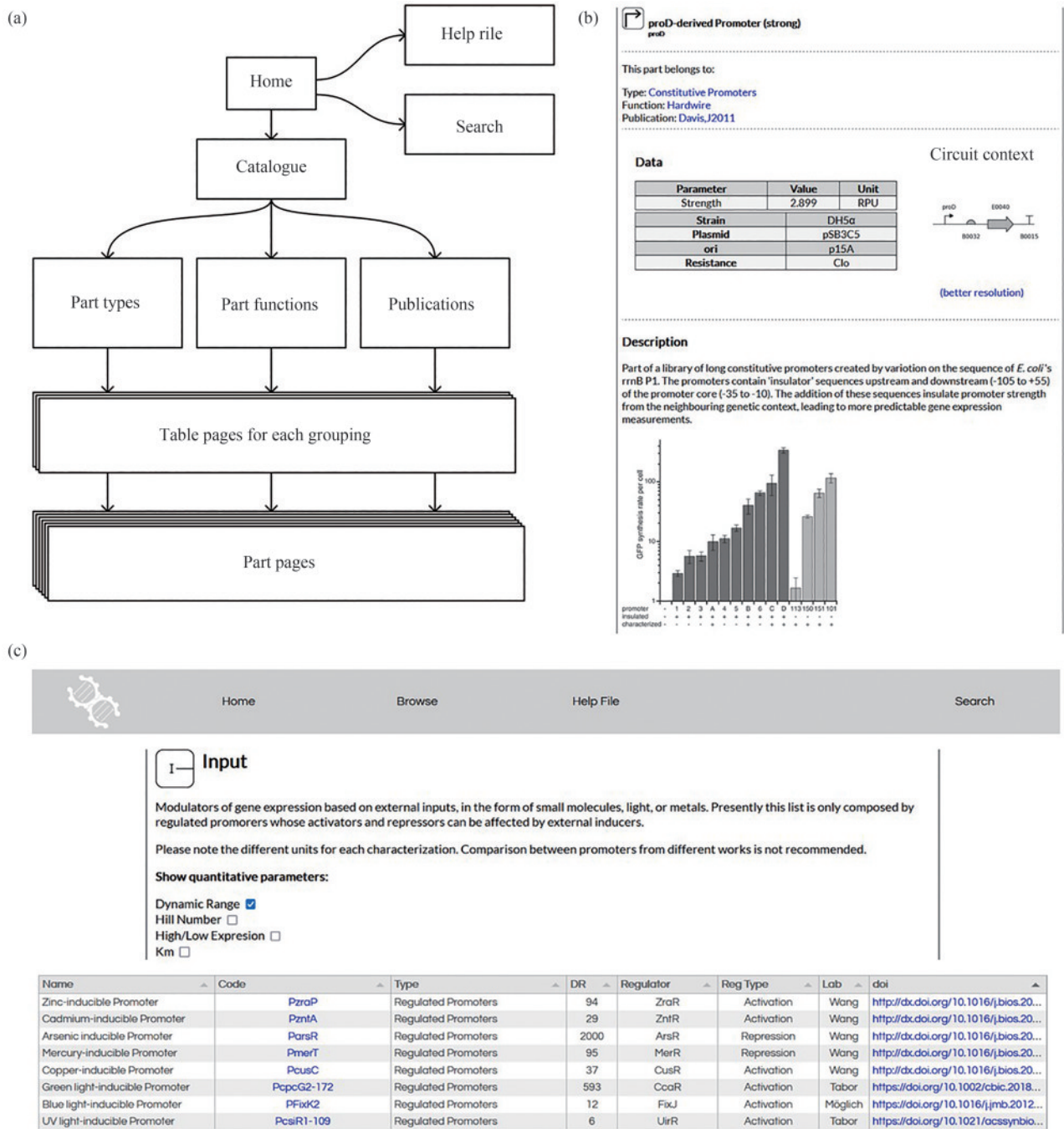


图 2 基因元器件精选数据库 BioPartsDB 的网站结构与内容设计示意图 [73]

(a) BioPartsDB 数据库的设计架构图，箭头代表数据库里的预期用户流，箭头往下越来越多的堆叠面板代表该部分的页面数量越多，包含的信息越详细；(b) 具体到某个基因元件数据的页面内容示例；(c) 某种类型的基因元件列表页面内容示例，显示元件的简要说明和关键性能数据信息

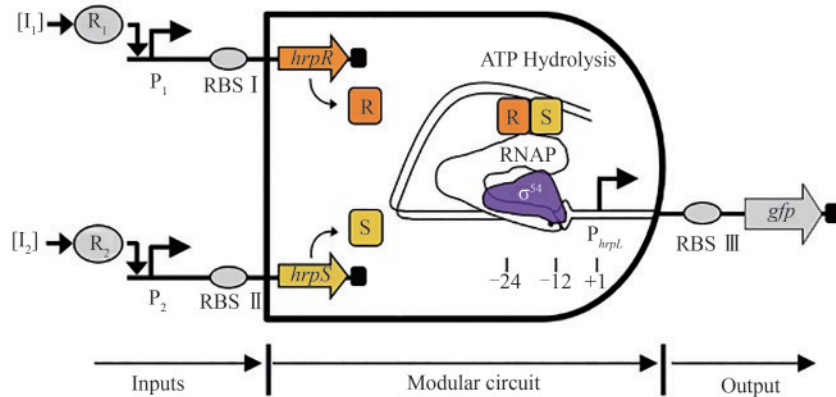
Fig. 2 Design portfolio and the web architecture of the BioPartsDB platform [73]

(a) A simplified diagram showing the information flow of the database platform. Arrows indicate the intended user browsing along the platform's webpages. Increasingly stacked panels indicate the higher number of pages in each section and consequently the more detailed level of information. (b) A web page with in-depth description of the information, performance, and characterization conditions for a specific genetic part. (c) A table for parts with a brief description and data of their key performance.

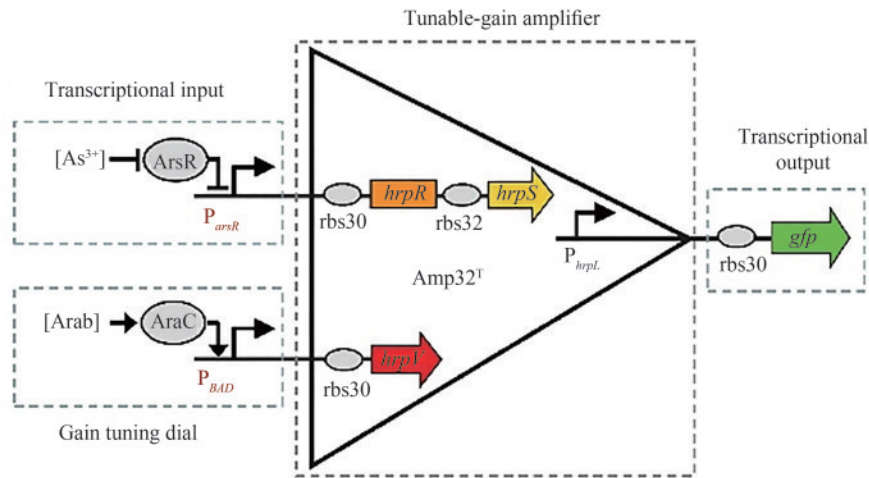
录因子融合后，可有效提高转录因子结合DNA的协同性，从而显著提升转录调控系统的超敏性和严谨性。该元件普遍适用于各种转录因子，在不同物种中也具有很高的通用性，可对遗传开关和人工调控网络实行“即插即用”式的改造，大幅改善人工系统的性能。

复杂基因线路中基因模块的行为通常是不可

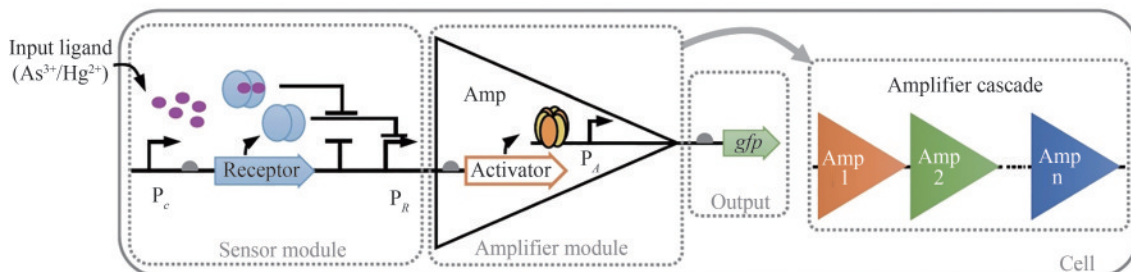
预测的^[80]。在组装模块为基因线路后，下游元件（例如调节蛋白的结合位点）会对上游模块产生负载，从而对预测的行为产生干扰^[81-83]。Ceroni等^[84]开发了一套用单位细胞产生的荧光量反映人工基因元件对细胞资源的占用情况的检测方法，该方法首次建立模型来计算可使用的细胞资源，有助于选择出最优的基因线路设计方案。细胞生



(a) 模块化且正交的逻辑与门和非门控线路设计^[25]
 (a) Design of a modular and orthogonal genetic AND gate for the logic integration of multi-input signals^[25]



(b) 模块化且放大倍数连续可调的转录信号放大线路设计^[76]
 (b) Design of modular and tunable genetic amplifiers for transcriptional signals^[76]



(c) 基于级联多层正交模块化的转录信号放大线路设计的领域目前最敏感的砷汞等重金属污染微生物传感器设计^[88]
 (c) Design of ultrasensitive bacterial sensors for arsenic and mercury by cascading multiple modular and orthogonal transcriptional amplifiers^[88]

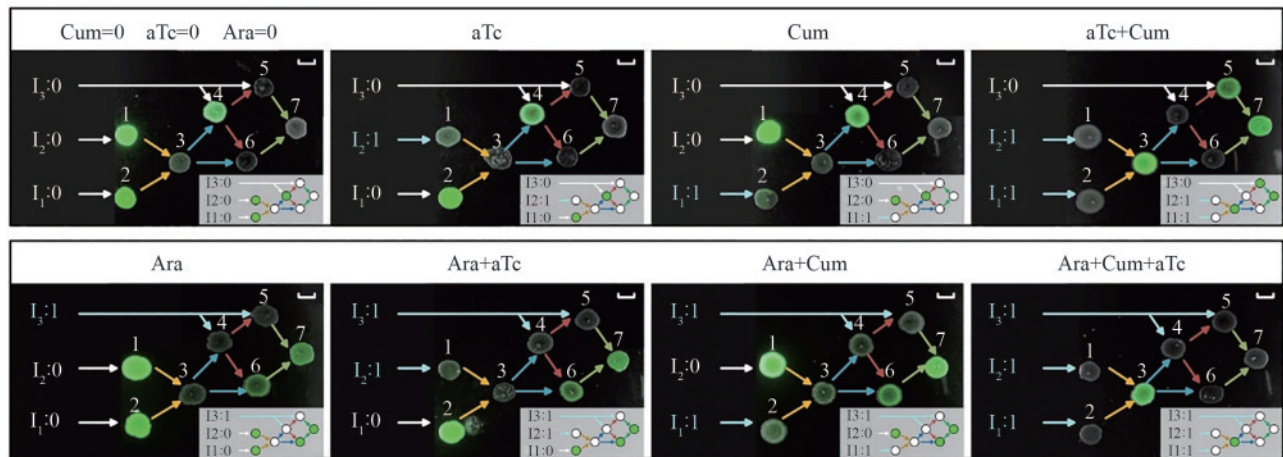
(d) 同时利用四个通信通道的分布式多细胞生物计算线路设计^[38](d) Engineered biocomputing circuit that simultaneously utilizes four communication channels^[38]

图3 合成基因线路的模块化设计

Fig. 3 Modular design of synthetic genetic circuits

长的内部环境是动态变化的，这使得构建基因线路或者网络时，在细胞中引入的元件难以实现稳定的输出，如众所熟知的启动子。MIT Voigt 实验室的研究人员运用控制理论设计了能够在任何拷贝数下实现目的基因同一水平表达的启动子元件，研究人员利用 transcription-activator-like effector (TALE) 构建了 iFFL 稳定的启动子 (iFFLs-stabilized promoter)。结果表明，即使基因的拷贝数由于基因组突变或者培养基条件改变发生了变化，这些 iFFL 稳定的启动子控制的目的基因在不同的基因组位点或者质粒上仍能实现接近相同的表达水平^[85]。本团队首次证明了人工合成核酸海绵可系统地调节基因线路中的基因表达，从而精确改变该线路中的基因表达泄露、输出幅度和诱导倍数、对小分子的响应灵敏度，并改善宿主细胞的生长速度^[86]。该基因调节方法简单、有效，可广泛用于多种应用领域相关的人工合成基因线路设计。Mishra 等^[87]设计了一种包含快速磷酸转移过程的负载驱动器，即使在存在大量负载的情况下也能响应时变输入信号，并且在酿酒酵母中测试了该模块的有效性。该负载驱动器可作为创建复杂的、更高级别的基因线路的基本构建块。本团队通过级联多层正交模块化的转录信号放大线路，首先通过调节细胞内受体蛋白的密度来提高传感器的灵敏度，其次通过串联多极转录放大器来促进输出模块的信号强度，最后结合转录调节和翻译后调节的方法来降低在无输入信

号时传感器的信号输出，从而增加输出信号的动态范围，创建了领域目前最敏感的基于大肠杆菌的砷和汞等重金属污染细胞传感器及创新的可视化输出模块 [图3(c)]^[88]。Liu 等^[89]揭示了 II 型 CRISPR 系统 crRNA-tracrRNA 配对的可编程性机制，利用重编程的 tracrRNA 劫持各种 RNA 分子激活 CRISPR 系统，并利用双 gRNA 介导的 CRISPR 激活 (CRISPRa) 系统，构建了一系列正交化的可编程 AND-gate 布尔逻辑门库，为编程生物学带来了更多可能性，包括重组细胞内的基因调控网络拓扑结构、开发新的 RNA 编辑器和 RNA 标记方法、设计复杂细胞生物计算线路等。

同时，单个细胞的负载有限，组合多种功能细胞，可以提高群体对复杂环境的适应性和鲁棒性，以完成更加复杂的工作。可通过多细胞分布式线路设计，让单个细胞里的线路模块化并进行多个细胞间的级联通信来提高线路的模块性和实现线路的规模化组装设计。相比于基于单种细胞的计算线路，分布式计算不仅可以降低每个细胞中合成线路的体积，减少代谢负担，而且可以利用细胞高度模块化的特性，在不同的细胞群体内重复使用相同的基因元件，降低对正交基因调控元件的依赖。为了演示更复杂的多细胞生物计算功能，研究人员设计了复杂的三输入 XOR-AND 逻辑门电路。XOR-AND 逻辑门电路分别部署在七个大肠杆菌菌株中，由四个通信通道协调。每个菌株包含一个 NOR 门 (cell-1 至 cell-6) 或

一个 Buffer 门 (cell-7), 这是已知第一个同时利用四个通信通道的生物计算线路 [图 3(d)]^[38]。虽然多细胞分布式计算在细菌、酵母和哺乳动物细胞中都已经实现, 但它们的规模却并未得到显著的提高。在细菌中最具有代表性的分布式计算线路仍然是基于多细胞或非门的 16 个 2 输入逻辑门以及更复杂一些的与门-异或门复合逻辑^[3, 38]。而在酵母与哺乳动物细胞中, 分布式计算线路的规模也与单细胞计算线路的规模相差不多, 最具代表性的是 1 位全加器程序^[90-91]。

不同类型的细胞共同组建成多细胞生物, 不仅如此, 多细胞生物中的每一个细胞之间都需要协调它们的行为, 这少不了细胞之间的通信。群体感应是很常见的一种细胞通信类型, 许多细菌合成自诱导物信号分子, 当诱导物信号分子超过临界浓度, 菌体内相关基因启动表达^[92]。其中诱导物可以在一个细胞中合成, 通过培养基扩散, 并在不同的细胞中进行基因调控。本团队通过偶联多细胞群体通信, 设计出了多输入逻辑门线路控制的高特异性环境监测细胞传感器^[29]。在信号整合或空间梯度感应上, 细胞通信也能驱动细胞图案的形成。研究人员对细胞间的信号系统进行了表征^[22], 这些系统可用于协调多细胞行为^[93-94]、稳定共培养种群以及在多细胞群体中执行任务分工。

目前多细胞分布式线路设计正在医药、制造、环保、能源等领域发挥着不可替代的作用。近几年, 合成基因线路广泛应用于疫苗研发, 现已研制出多款流感、新冠疫苗。合成生物学可以对细菌做一些定向的设计和改变, 包括控制细菌复制周期、调控细菌表面抗原、重新设计细菌调节免疫机制等, 进而缩短疫苗的研发周期, 提高疫苗的有效性^[95]。迄今为止, CAR-T 细胞疗法在根除血液恶性肿瘤 (例如: 白血病中的 CD19CAR) 方面取得了巨大成功。然而, 这种成功尚未外推到实体肿瘤, 主要受靶点多样性、肿瘤异质性和复杂的肿瘤微环境三个因素的影响。研究者基于合成生物学策略, 不断对 CAR-T 疗法进行升级。波士顿大学团队开发出了一种包含 11 种可编程合成转录因子 (synZiFTR) 的工具包。这些转录因子可以被一些获批临床使用的小分子药物激活,

然后通过精确调控基因, 启动特定的细胞程序, 即通过定时给药, 可实现按需激活细胞, 从而达到增殖、抗肿瘤活性等疗效^[96]。合成全细胞生物传感器 (synthetic whole-cell biosensor) 是以工程化细胞为主要感受元件, 可用于检测并报告某种特定物质、环境。与传统的化学和电子检测方法相比, 它们体积小、易生产、价格低、可降解, 因此在环境检测和临床诊断上有极高的应用价值。本团队研究提供了一套完整的合成全细胞传感器优化方案, 并构建了低价易用的砷离子传感平台。目前全世界有 50 多个国家、1.4 亿人口的饮用水中有超过 10 $\mu\text{g/L}$ 的砷污染, 本研究可以为砷污染地区提供低价、便捷的砷污染检测方法^[88]。

4 合成基因线路的自动化设计进展

目前大多数基因线路都是由研究人员手动设计并不断地进行试错, 需要消耗大量的时间和精力且容易出现错误。结合线路设计软件与高通量模块化的实验室设备, 可快速完成复杂基因线路的“设计-构建-测试-学习”这一闭环(图 4)。

MIT Voigt 团队开发的 Cello 系列软件: 研究人员通过把计算机辅助软件引入基因线路设计, 将工程思想与生物融合, 构建了第一代 Cello 平台^[99], 根据输入传感器、线路功能以及响应底盘细胞中的用户约束文件就可自动设计出编码所需线路的 DNA 序列。Cello 的基因元件库由原核抑制因子为基础的绝缘 NOT/NOR 门组成, 并经过标准化表征和测试。同时, Cello 支持使用合成生物学开放语言 (SBOL) 文件格式交换基因设计。该团队使用 Cello 在大肠杆菌中设计了 60 条线路, 其中 45 条在每个输出状态下能正常工作, 所有线路中 92% 的输出状态都能按照预期工作。并且, 用相似的策略在非模式原核生物拟杆菌中实现了线路的自动化设计^[100]。但真核生物转录和翻译机制复杂, 线路上的启动子受到多方面的内源影响, 并且组装时各转录模块的终止子也会对上下游基因的表达产生影响, 因此, 相似的策略不能直接应用。Chen 等^[101]通过对启动子进行模块化设计, 构建了一系列可调控的短启动子 (<120 bp), 利用原核 DNA 结合蛋白, 构建了 9 个响应函数几乎相

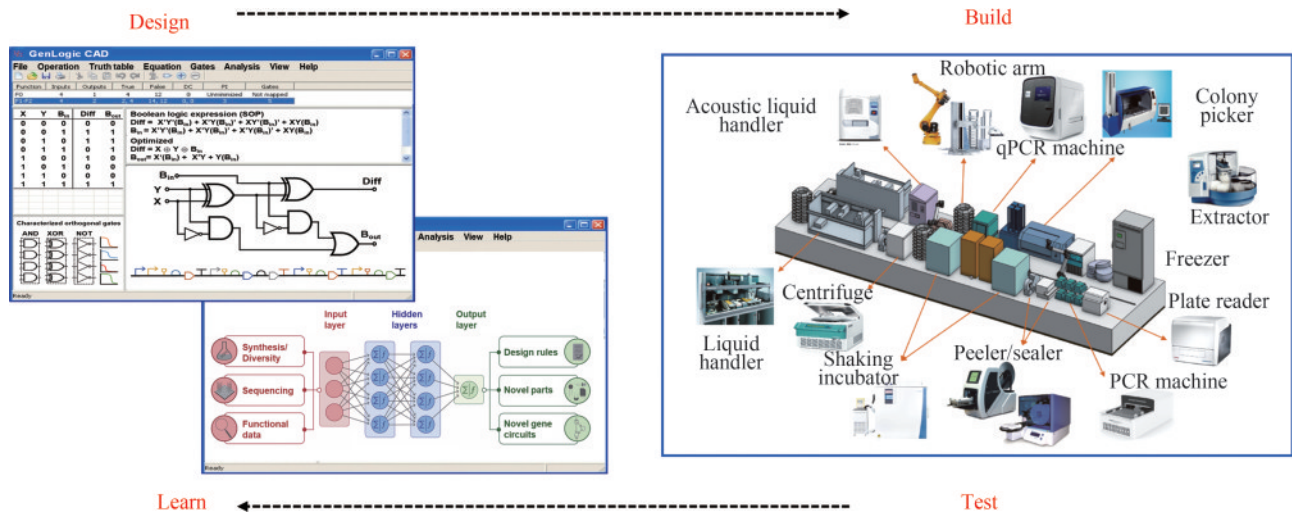


图4 合成基因线路的自动化“设计-构建-测试-学习”循环

(“构建与测试”的生物铸造厂设施摘自文献 [97]，“学习”的模型改编自文献 [98])

Fig. 4 An automated “design-build-test-learn” cycle for genetic circuit engineering.

(Automated instrumentation in biofoundries is adapted from reference [97]. The neural network-based deep learning model is adapted from reference [98].)

同、动态范围为400倍的NOT/NOR门。同时通过将核酶放置于启动子后以阻止从RNAP通读而产生的mRNA的核输出，从而达到相互转录隔离的目的。将上述逻辑门整合至酵母染色体中，表征其响应函数，所获得的数据用于创建用户约束文件，以此在Cello中构建3输入的复杂基因线路。其中，最复杂的线路长达33 kbp，包含11种调控蛋白及16个转录调控单元。

Cello 1.0虽然能够可靠地设计线路，但也有明显的局限性：逻辑门元件库类型单一，仅限于单一门类型；模拟能力受到限制；对于描述线路行为的硬件描述语言Verilog的支持有限。Jones等^[102]开发了Cello 2.0，能够灵活地描述逻辑门的结构及其代表动态行为的数学模型，描述基因组中逻辑门放置的新形式规则，新的图形用户界面，支持Verilog 2005语法，连接到SynBioHub元件库软件环境。这些特征将Cello的能力扩展到大肠杆菌之外的新生物体和更广泛的遗传背景，包括基因组。研究人员利用Cello 2.0自动化设计了三输入与门，并在酵母中得到了预期结果。

同时，其他的团队也开发了不少自动化设计工具，如iBioSim^[103-104]、j5^[105]、GenoCAD^[106]等，它们与Cello系列软件不同，不会对基因线路的运转方式或者运转结果进行预测，需要用户了解并输入有关线路结构的信息。

在过去的几年中，世界各地的研究机构一直致力于搭建生物铸造厂（biofoundry），以扩大其生物技术的开发能力。在2019年，来自全球8个国家的16所顶尖生物铸造厂成立了国际合成生物设施联盟（Global Biofoundry Alliance, GBA），至今已发展到35所，旨在全球推动合成生物设施建设为目标，共同应对全球性的科学挑战^[107]。

生物铸造厂集成了各种分子生物学设施，包括液体处理设备、PCR仪、自动培养箱、离心机、酶标仪等，设备间通过机器人或传送带串联，可对生物元件、模块及线路进行高通量表征。多个国家（如美国、英国等）搭建的生物铸造厂已投入使用，其通量是单人手动操作的10倍以上^[73, 108]，不仅可进行规模化工程，还可对多种表型进行功能定位和多重优化^[109]。基于国外已有的生物铸造厂，在国内政策的支持下，我国也有不少研究机构搭建了生物铸造厂，例如：中国科学院深圳先进技术研究院、中国科学院天津工业生物技术研究所（以下简称天津工生所）、天津大学化工学院、浙江大学杭州国际科创中心等。其中，天津工生所的合成生物学自动化平台已经实现了自动化单基因克隆，通量可达300~600个/天，复杂质粒的多模块化组装也达到100个/天，正确率达90%以上，实现了多种模式生物的自动化基因操作^[110]。深圳合成生物大设施的一期工程已投入

使用,可满足多样化的实验需求:针对天然产物的生物合成基因簇开展了一系列自动化、高通量的基因挖掘和工程化改造研究^[111-113];可应用于代谢途径的进化:2周内可实现挑菌、培养、突变库筛选、备选突变体产物提取等常规流程^[114]。浙江大学杭州国际科创中心已建成合成生物学自动化装置 iBioFoundry,该装置是目前合成生物学研究领域功能最全、集成度最高的一套全流程、高通量、自动化合成生物学科学装置,平均每天可以创建2000个人工细胞,将人工细胞构建效率提高2个数量级以上,致力于构建深度融合“BT+IT”的高能级合成生物学研究平台。具体关于生物铸造厂的构建、运行及建设可参考文章《合成生物学自动化装置 iBioFoundry 的构建与应用》^[115]。

人工智能在基因线路自动化设计也发挥着重要作用。迭代的“设计-构建-测试-学习”循环产生的实验数据适用于机器学习,同时,表征良好的基因线路元器件有助于人工智能的整合^[10]。在元件的设计中,Hu等^[116]通过多轮机器学习与机器实验迭代,开发了一种自动化蛋白质工程方法 BO-EVO,大幅提升了蛋白质多位点组合突变设计空间的探索效率,能够以小于1%的湿实验量寻找获得全局最优实验结果。Wang等^[117]利用深度学习技术并融合生物先验知识来建立调控元件的生成模型,该模型以天然启动子的序列特征为指导,可以捕获不同位置核苷酸之间的相互作用,并通过计算机设计新的合成启动子,经湿实验验证,大于70%的新启动子具有功能性。同时,基于 CO-BERTa 的新的深度学习模型、序列的卷积和递归神经网络模型已被分别用于密码子优化^[118]、识别蛋白质的潜在识别位点^[119]和基因的表达^[120]。在线路优化中,Valeri等^[121]将两种正交且协同的深度学习模型(STORM与NuSpeak)用于表征和优化 toehold switch。研究人员利用计算机视觉和自然语言处理技术,使用卷积过滤器、注意力机制、计算机诱变和迁移学习对模型进行优化,为 toehold 的选择和设计提供了序列到功能的深度学习框架,并增强了构建有效线路组件和精准诊疗的能力。Seak等^[122]使用基因线路来模拟人工神经网络,将生物计算算法的潜力提升到一个新的水平。人工智能最大的优势在于开发综合预测模型,

减少试错成本,这将加速基因线路设计研究和广泛应用。

5 合成基因线路工程化设计的挑战与展望

合成基因线路和电子线路都具有信号采集、信号处理和信号输出的过程,但两者在很多方面截然不同。电子线路的主体通常由绝缘材料组成,通过导线将各个电子元件进行连接,并且各个电子元件互不干扰。而合成基因线路的主体是活体细胞,由各种生物分子组成。与导线不同,合成基因线路中的信号载体是生物化学分子如RNA、转录因子蛋白等,可在细胞中或细胞间扩散和混合,容易造成串扰。并且生物系统的物理特性很难量化,当遗传元件用于新的细胞或者环境时,会出现意料之外的结果。此外,大规模的合成基因线路会影响宿主细胞内的资源竞争、宿主的生长等。这些差异铸就了合成基因线路的工程化设计需要探索新的途径,研究者们借鉴了电子线路发展的成功经验,提出了四大原则(正交化、标准化、模块化和自动化),开发了一系列全新的生物线路设计、构建、测试和学习方法。

虽然大量的研究工作致力于创建正交元件来减轻与宿主的相互作用,但尚未合并形成一个连续的和正交的中心法则。目前,细菌系统提供了大量的转录和翻译正交元件库,而正交DNA复制机制的研究相对滞后;相反,在真核生物中已开发了正交的DNA复制和转录方法,而正交翻译元件库与原核生物相比还相距甚远。未来需要继续开发正交化成分,支持现有的元件文库,以期实现一个完全正交的中心法则来实现复杂基因线路的构建和创新。

高性能的基因表达调控元件是高性能的基因线路的基础,但目前基因线路中的调控元件的标准化表征仍不清晰、不充分,应深入挖掘基因表达调控元件,并进行全面表征,以评价新的基因表达调控元件与现有元件的差异。另外需要建立新的元件改造策略和新的人工元件设计方法,突破天然元件的性能限制,如蛋白线路设计可以快

速响应信号输入, 或通过提高单位基因的信息处理能力来解决复杂大规模线路的细胞资源竞争及代谢负载问题, 为进一步构建高效的合成生物学系统打下坚实的基础。一个生命元件要发挥作用, 往往是在胞内和胞外环境的影响下, 通过多种相互作用实现的。而当前多数生命元件的表征, 基本上都是在严格控制的实验室条件下进行的, 各类环境影响和相互作用都未被计入考量。一旦添加了新的元件, 系统中原有元件便很有可能受到影响而需要重新设计。这样的情况下, 随着系统当中元件数量组合的增加, 设计的复杂性和难度也会随之爆发式地增长。

实现基因线路设计的可预测性, 仍需要开发出更多绝缘的基因模块, 以及跨物种的正交生物元件, 同时开发更精准的建模方法和设计框架来处理生物不确定性(例如模块之间的连通性、简单的物理过程行为和参数)和外界环境的干扰。此外, 开发细胞内基因线路内部的新型原位动态检测方法对提高基因线路的设计效率和后续排错至关重要, 例如运用RNA-seq、蛋白组学、代谢组学等高通量检测技术对线路进行表征, 有利于研究人员了解线路中每个元件和节点在细胞内原始工作状态下的性能和动力学特性, 继而可针对性地对基因线路进行优化。近年来, 人工智能、生物信息学等软件算法的进步也为高通量检测技术的发展带来新的机遇。通过更智能的信息化技术, 研究人员能够以更高效且自动化的方式对组学大数据进行分析, 快速获取检测结果。相信在不久的将来, 依靠高通量组学技术诊断基因线路运行情况将成为实验室中的常用手段, 极大地加速基因线路的“设计-构建-测试-学习”循环的迭代。

基因线路设计的自动化需要从硬件和软件两个方面着手。建造生物铸造厂是一个复杂的过程, 需要专门的设备和专业的人员, 且运行成本可能比标准实验室高一个数量级, 但其自动化有可能将采集结果的速度、再现性和可靠性提高几个数量级。当前, 大部分模型只关注基因调控环路中的逻辑操作, 虽然能够在一定程度上将少数基因元件进行模块化组装, 并在宿主细胞中实现特定功能, 但与生物细胞和生命体相比仍相距甚

远。人工智能技术结合了已知的生物物理、机器学习和强化学习模型, 通过机器学习来探索基础的规律, 并可以有效预测结构对宿主的影响以及反向影响。合成生物元件与合成基因线路的智能化设计将人工智能技术应用于构建全新的基因调控元件与线路, 利用机器学习并融合相关生物知识建立调控元件与线路的生成模型。通过在计算机中的算法寻优部分替代生物实验上的随机搜索, 可以大大提高实验成功率。理论上, 该方法可以产生数量远远超过天然启动子的全新元件, 并显著提高复杂基因线路设计的效率, 极大地丰富了可用于工程生物学研究的调控元件库与基因线路文库。

合成基因线路工程化设计研究已步入从“量的积累”转向“质的飞跃”发展轨道, 这不仅需要研究者在上述几方面推动技术变革, 更需要在理论层面加速合成生物学研究范式变革, 推动理论模型构建。未来, 在人工智能与大数据等驱动下构建完善的合成基因线路模型, 将推动设计符合目标需求的人造生命体成为可能, 这将为医药健康、农业环境和工业发酵等领域提供全新的解决方案, 并能产生颠覆性的技术成果。

参 考 文 献

- [1] MENG F K, ELLIS T. The second decade of synthetic biology: 2010—2020[J]. *Nature Communications*, 2020, 11(1): 5174.
- [2] WANG X Y, ZHOU N, WANG B J. Bacterial synthetic biology: tools for novel drug discovery[J]. *Expert Opinion on Drug Discovery*, 2023, 18(10): 1087-1097.
- [3] TAMSIR A, TABOR J J, VOIGT C A. Robust multicellular computing using genetically encoded NOR gates and chemical ‘wires’[J]. *Nature*, 2011, 469(7329): 212-215.
- [4] ELOWITZ M B, LEIBLER S. A synthetic oscillatory network of transcriptional regulators[J]. *Nature*, 2000, 403(6767): 335-338.
- [5] GARDNER T S, CANTOR C R, COLLINS J J. Construction of a genetic toggle switch in *Escherichia coli*[J]. *Nature*, 2000, 403(6767): 339-342.
- [6] ROQUET N, SOLEIMANY A P, FERRIS A C, et al. Synthetic recombinase-based state machines in living cells[J]. *Science*, 2016, 353(6297): aad8559.
- [7] RUBENS J R, SELVAGGIO G, LU T K. Synthetic mixed-signal computation in living cells[J]. *Nature Communications*,

- 2016, 7: 11658.
- [8] BRADLEY R W, BUCK M, WANG B J. Tools and principles for microbial gene circuit engineering[J]. *Journal of Molecular Biology*, 2016, 428(5 Pt B): 862-888.
- [9] GAO Y L, WANG L, WANG B J. Customizing cellular signal processing by synthetic multi-level regulatory circuits[J]. *Nature Communications*, 2023, 14(1): 8415.
- [10] XIANG Y Y, DALCHAU N, WANG B J. Scaling up genetic circuit design for cellular computing: advances and prospects [J]. *Natural Computing*, 2018, 17(4): 833-853.
- [11] ŞİMŞEK E, YAO Y, LEE D, et al. Toward predictive engineering of gene circuits[J]. *Trends in Biotechnology*, 2023, 41(6): 760-768.
- [12] PECCOUD J, BLAUVELT M F, CAI Y Z, et al. Targeted development of registries of biological parts[J]. *PLoS One*, 2008, 3(7): e2671.
- [13] COSTELLO A, BADRAN A H. Synthetic biological circuits within an orthogonal central dogma[J]. *Trends in Biotechnology*, 2021, 39(1): 59-71.
- [14] PARK M, PATEL N, KEUNG A J, et al. Engineering epigenetic regulation using synthetic read-write modules[J]. *Cell*, 2019, 176(1-2): 227-238.e20.
- [15] DHAMI K, MALYSHEV D A, ORDOUKHANIAN P, et al. Systematic exploration of a class of hydrophobic unnatural base pairs yields multiple new candidates for the expansion of the genetic alphabet[J]. *Nucleic Acids Research*, 2014, 42(16): 10235-10244.
- [16] HOSHIKA S, LEAL N A, KIM M J, et al. Hachimoji DNA and RNA: a genetic system with eight building blocks[J]. *Science*, 2019, 363(6429): 884-887.
- [17] RAVIKUMAR A, ARZUMANYAN G A, OBADI M K A, et al. Scalable, continuous evolution of genes at mutation rates above genomic error thresholds[J]. *Cell*, 2018, 175(7): 1946-1957.e13.
- [18] ZHONG Z W, RAVIKUMAR A, LIU C C. Tunable expression systems for orthogonal DNA replication[J]. *ACS Synthetic Biology*, 2018, 7(12): 2930-2934.
- [19] TIAN R Z, ZHAO R Z, GUO H Y, et al. Engineered bacterial orthogonal DNA replication system for continuous evolution [J]. *Nature Chemical Biology*, 2023, 19(12): 1504-1512.
- [20] YANG L, NIELSEN A A, FERNANDEZ-RODRIGUEZ J, et al. Permanent genetic memory with >1-byte capacity[J]. *Nature Methods*, 2014, 11(12): 1261-1266.
- [21] TEMME K, HILL R, SEGALL-SHAPIRO T H, et al. Modular control of multiple pathways using engineered orthogonal T7 polymerases[J]. *Nucleic Acids Research*, 2012, 40(17): 8773-8781.
- [22] MEYER A J, ELLEFSON J W, ELLINGTON A D. Directed evolution of a panel of orthogonal T7 RNA polymerase variants for *in vivo* or *in vitro* synthetic circuitry[J]. *ACS Synthetic Biology*, 2015, 4(10): 1070-1076.
- [23] RHODIUS V A, SEGALL-SHAPIRO T H, SHARON B D, et al. Design of orthogonal genetic switches based on a crosstalk map of σ s, anti- σ s, and promoters[J]. *Molecular Systems Biology*, 2013, 9: 702.
- [24] BERVOETS I, VAN BREMPT M, VAN NEROM K, et al. A sigma factor toolbox for orthogonal gene expression in *Escherichia coli*[J]. *Nucleic Acids Research*, 2018, 46(4): 2133-2144.
- [25] WANG B J, KITNEY R I, JOLY N, et al. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology[J]. *Nature Communications*, 2011, 2: 508.
- [26] ZHAN J, DING B, MA R, et al. Develop reusable and combinable designs for transcriptional logic gates[J]. *Molecular Systems Biology*, 2010, 6: 388.
- [27] BRÖDEL A K, JARAMILLO A, ISALAN M. Engineering orthogonal dual transcription factors for multi-input synthetic promoters[J]. *Nature Communications*, 2016, 7: 13858.
- [28] STANTON B C, NIELSEN A A, TAMSIR A, et al. Genomic mining of prokaryotic repressors for orthogonal logic gates[J]. *Nature Chemical Biology*, 2014, 10(2): 99-105.
- [29] WANG B J, BARAHONA M, BUCK M. A modular cell-based biosensor using engineered genetic logic circuits to detect and integrate multiple environmental signals[J]. *Biosensors & Bioelectronics*, 2013, 40(1): 368-376.
- [30] LOPRESIDE A, WAN X Y, MICHELINI E, et al. Comprehensive profiling of diverse genetic reporters with application to whole-cell and cell-free biosensors[J]. *Analytical Chemistry*, 2019, 91(23): 15284-15292.
- [31] MEYER A J, SEGALL-SHAPIRO T H, GLASSEY E, et al. *Escherichia coli* "Marionette" strains with 12 highly optimized small-molecule sensors[J]. *Nature Chemical Biology*, 2019, 15 (2): 196-204.
- [32] GROSECLOSE T M, RONDON R E, HERDE Z D, et al. Engineered systems of inducible anti-repressors for the next generation of biological programming[J]. *Nature Communications*, 2020, 11(1): 4440.
- [33] HANKO E K R, PAIVA A C, JONCZYK M, et al. A genome-wide approach for identification and characterisation of metabolite-inducible systems[J]. *Nature Communications*, 2020, 11(1): 1213.
- [34] SCOTT S R, HASTY J. Quorum sensing communication modules for microbial consortia[J]. *ACS Synthetic Biology*, 2016, 5(9): 969-977.

- [35] GRANT P K, DALCHAU N, BROWN J R, et al. Orthogonal intercellular signaling for programmed spatial behavior[J]. *Molecular Systems Biology*, 2016, 12(1): 849.
- [36] KYLILIS N, TUZA Z A, STAN G B, et al. Tools for engineering coordinated system behaviour in synthetic microbial consortia[J]. *Nature Communications*, 2018, 9(1): 2677.
- [37] WU S B, QIAO J J, YANG A D, et al. Potential of orthogonal and cross-talk quorum sensing for dynamic regulation in cocultivation[J]. *Chemical Engineering Journal*, 2022, 445: 136720.
- [38] DU P, ZHAO H W, ZHANG H Q, et al. *De novo* design of an intercellular signaling toolbox for multi-channel cell-cell communication and biological computation[J]. *Nature Communications*, 2020, 11(1): 4226.
- [39] JONKERGOUW C, SAVOLA P, OSMEKHINA E, et al. Exploration of chemical diversity in intercellular quorum sensing signalling systems in prokaryotes[J]. *Angewandte Chemie International Edition*, 2024, 63(2): e202314469.
- [40] ESPAH BORUJENI A, MISHLER D M, WANG J Z, et al. Automated physics-based design of synthetic riboswitches from diverse RNA aptamers[J]. *Nucleic Acids Research*, 2016, 44(1): 1-13.
- [41] CHAPPELL J, WESTBROOK A, VEROSLOFF M, et al. Computational design of small transcription activating RNAs for versatile and dynamic gene regulation[J]. *Nature Communications*, 2017, 8(1): 1051.
- [42] DIDOVYK A, BOREK B, HASTY J, et al. Orthogonal modular gene repression in *Escherichia coli* using engineered CRISPR/Cas9[J]. *ACS Synthetic Biology*, 2016, 5(1): 81-88.
- [43] NIELSEN A A K, VOIGT C A. Multi-input CRISPR/Cas genetic circuits that interface host regulatory networks[J]. *Molecular Systems Biology*, 2014, 10(11): 763.
- [44] REIS A C, HALPER S M, VEZEAU G E, et al. Simultaneous repression of multiple bacterial genes using nonrepetitive extra-long sgRNA arrays[J]. *Nature Biotechnology*, 2019, 37(11): 1294-1301.
- [45] LIU Y, WAN X Y, WANG B J. Engineered CRISPRa enables programmable eukaryote-like gene activation in bacteria[J]. *Nature Communications*, 2019, 10(1): 3693.
- [46] GREEN A A, SILVER P A, COLLINS J J, et al. Toehold switches: *de-novo*-designed regulators of gene expression[J]. *Cell*, 2014, 159(4): 925-939.
- [47] KIM J M, ZHOU Y, CARLSON P D, et al. *De novo*-designed translation-repressing riboregulators for multi-input cellular logic[J]. *Nature Chemical Biology*, 2019, 15(12): 1173-1182.
- [48] PINTO F, THORNTON E L, WANG B J. An expanded library of orthogonal split inteins enables modular multi-peptide assemblies[J]. *Nature Communications*, 2020, 11(1): 1529.
- [49] CERVETTINI D, TANG S, FRIED S D, et al. Rapid discovery and evolution of orthogonal aminoacyl-tRNA synthetase-tRNA pairs[J]. *Nature Biotechnology*, 2020, 38(8): 989-999.
- [50] DUNKELMANN D L, WILLIS J C W, BEATTIE A T, et al. Engineered triply orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids[J]. *Nature Chemistry*, 2020, 12(6): 535-544.
- [51] MERRICK C A, ZHAO J, ROSSER S J. Serine integrases: advancing synthetic biology[J]. *ACS Synthetic Biology*, 2018, 7(2): 299-310.
- [52] PAGET M S. Bacterial sigma factors and anti-sigma factors: structure, function and distribution[J]. *Biomolecules*, 2015, 5(3): 1245-1265.
- [53] GUET C C, ELOWITZ M B, HSING W, et al. Combinatorial synthesis of genetic networks[J]. *Science*, 2002, 296(5572): 1466-1470.
- [54] HASTY J, DOLNIK M, ROTTSCHÄFER V, et al. Synthetic gene network for entraining and amplifying cellular oscillations [J]. *Physical Review Letters*, 2002, 88(14): 148101.
- [55] HOOSHANGI S, THIBERGE S, WEISS R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102(10): 3581-3586.
- [56] WANG B J, BARAHONA M, BUCK M. Amplification of small molecule-inducible gene expression *via* tuning of intracellular receptor densities[J]. *Nucleic Acids Research*, 2015, 43(3): 1955-1964.
- [57] 李晓萌, 姜威, 梁泉峰, 等. 细菌群体感应系统在细胞间通讯中的应用及其合成生物学研究进展[J]. *合成生物学*, 2020, 1(5): 540-555.
- LI X M, JIANG W, LIANG Q F, et al. Application of bacterial quorum sensing system in intercellular communication and its progress in synthetic biology[J]. *Synthetic Biology Journal*, 2020, 1(5): 540-555.
- [58] 周爱林, 刘奕, 巴方, 等. 细菌群体感应元件构建和工程应用 [J]. *合成生物学*, 2021, 2(2): 234-246.
- ZHOU A L, LIU Y, BA F, et al. Construction and engineering application of bacterial quorum sensing elements[J]. *Synthetic Biology Journal*, 2021, 2(2): 234-246.
- [59] AMERUOSO A, GAMBILL L, LIU B Y, et al. Brave new 'RNA' world—advances in RNA tools and their application for understanding and engineering biological systems[J]. *Current Opinion in Systems Biology*, 2019, 14: 32-40.
- [60] ISAACS F J, DWYER D J, DING C M, et al. Engineered

- ribo regulators enable post-transcriptional control of gene expression[J]. *Nature Biotechnology*, 2004, 22(7): 841-847.
- [61] 付宪, 林涛, 张帆, 等. 基因密码子拓展技术的方法原理和前沿应用研究进展[J]. *合成生物学*, 2020, 1(1): 103-119.
- FU X, LIN T, ZHANG F, et al. Progress in the study of genetic code expansion related methods, principles and applications[J]. *Synthetic Biology Journal*, 2020, 1(1): 103-119.
- [62] CANTON B, LABNO A, ENDY D. Refinement and standardization of synthetic biological parts and devices[J]. *Nature Biotechnology*, 2008, 26(7): 787-793.
- [63] CASINI A, STORCH M, BALDWIN G S, et al. Bricks and blueprints: methods and standards for DNA assembly[J]. *Nature Reviews Molecular Cell Biology*, 2015, 16(9): 568-576.
- [64] GUO Y K, DONG J K, ZHOU T, et al. YeastFab: the design and construction of standard biological parts for metabolic engineering in *Saccharomyces cerevisiae*[J]. *Nucleic Acids Research*, 2015, 43(13): e88.
- [65] RAJKUMAR A S, VARELA J A, JUERGENS H, et al. Biological parts for *Kluyveromyces marxianus* synthetic biology[J]. *Frontiers in Bioengineering and Biotechnology*, 2019, 7: 97.
- [66] WEBER E, ENGLER C, GRUETZNER R, et al. A modular cloning system for standardized assembly of multigene constructs[J]. *PLoS One*, 2011, 6(2): e16765.
- [67] LEE M E, DELOACHE W C, CERVANTES B, et al. A highly characterized yeast toolkit for modular, multipart assembly[J]. *ACS Synthetic Biology*, 2015, 4(9): 975-986.
- [68] WEI P L, FAN J, YU J W, et al. Quantitative characterization of filamentous fungal promoters on a single-cell resolution to discover cryptic natural products[J]. *Science China Life Sciences*, 2023, 66(4): 848-860.
- [69] REDDEN H, ALPER H S. The development and characterization of synthetic minimal yeast promoters[J]. *Nature Communications*, 2015, 6: 7810.
- [70] CURRAN K A, MORSE N J, MARKHAM K A, et al. Short synthetic terminators for improved heterologous gene expression in yeast[J]. *ACS Synthetic Biology*, 2015, 4(7): 824-832.
- [71] YIM S S, AN S J, KANG M, et al. Isolation of fully synthetic promoters for high-level gene expression in *Corynebacterium glutamicum*[J]. *Biotechnology and Bioengineering*, 2013, 110(11): 2959-2969.
- [72] MCLAUGHLIN J A, MYERS C J, ZUNDEL Z, et al. SynBioHub: a standards-enabled design repository for synthetic biology[J]. *ACS Synthetic Biology*, 2018, 7(2): 682-688.
- [73] BUSON F, GAO Y L, WANG B J. Genetic parts and enabling tools for biocircuit design[J]. *ACS Synthetic Biology*, 2024, 13(3): 697-713.
- [74] HILLSON N, CADDICK M, CAI Y Z, et al. Building a global alliance of biofoundries[J]. *Nature Communications*, 2019, 10(1): 2040.
- [75] GRUNBERG T W, DEL VECCHIO D. Modular analysis and design of biological circuits[J]. *Current Opinion in Biotechnology*, 2020, 63: 41-47.
- [76] WANG B J, BARAHONA M, BUCK M. Engineering modular and tunable genetic amplifiers for scaling transcriptional signals in cascaded gene networks[J]. *Nucleic Acids Research*, 2014, 42(14): 9484-9492.
- [77] LI Y Q, JIANG Y, CHEN H, et al. Modular construction of mammalian gene circuits using TALE transcriptional repressors[J]. *Nature Chemical Biology*, 2015, 11(3): 207-213.
- [78] LEVINE M, CATTOGLIO C, TJIAN R. Looping back to leap forward: transcription enters a new era[J]. *Cell*, 2014, 157(1): 13-25.
- [79] HOU J R, ZENG W Q, ZONG Y Q, et al. Engineering the ultrasensitive transcription factors by fusing a modular oligomerization domain[J]. *ACS Synthetic Biology*, 2018, 7(5): 1188-1194.
- [80] PURNICK P E M, WEISS R. The second wave of synthetic biology: from modules to systems[J]. *Nature Reviews Molecular Cell Biology*, 2009, 10(6): 410-422.
- [81] CARDINALE S, ARKIN A P. Contextualizing context for synthetic biology: identifying causes of failure of synthetic biological systems[J]. *Biotechnology Journal*, 2012, 7(7): 856-866.
- [82] DEL VECCHIO D, NINFA A J, SONTAG E D. Modular cell biology: retroactivity and insulation[J]. *Molecular Systems Biology*, 2008, 4: 161.
- [83] LIU Q J, SCHUMACHER J, WAN X Y, et al. Orthogonality and burdens of heterologous AND gate gene circuits in *E. coli* [J]. *ACS Synthetic Biology*, 2018, 7(2): 553-564.
- [84] CERONI F, ALGAR R, STAN G B, et al. Quantifying cellular capacity identifies gene expression designs with reduced burden[J]. *Nature Methods*, 2015, 12(5): 415-418.
- [85] SEGALL-SHAPIRO T H, SONTAG E D, VOIGT C A. Engineered promoters enable constant gene expression at any copy number in bacteria[J]. *Nature Biotechnology*, 2018, 36(4): 352-358.
- [86] WAN X Y, PINTO F, YU L Y, et al. Synthetic protein-binding DNA sponge as a tool to tune gene expression and mitigate protein toxicity[J]. *Nature Communications*, 2020, 11(1): 5961.
- [87] MISHRA D, RIVERA P M, LIN A, et al. A load driver device for engineering modularity in biological networks[J]. *Nature Biotechnology*, 2014, 32(12): 1268-1275.
- [88] WAN X Y, VOLPETTI F, PETROVA E, et al. Cascaded amplifying circuits enable ultrasensitive cellular sensors for toxic

- metals[J]. *Nature Chemical Biology*, 2019, 15(5): 540-548.
- [89] LIU Y, PINTO F, WAN X Y, et al. Reprogrammed tracrRNAs enable repurposing of RNAs as crRNAs and sequence-specific RNA biosensors[J]. *Nature Communications*, 2022, 13(1): 1937.
- [90] REGOT S, MACIA J, CONDE N, et al. Distributed biological computation with multicellular engineered networks[J]. *Nature*, 2011, 469(7329): 207-211.
- [91] AUSLÄNDER D, AUSLÄNDER S, PIERRAT X, et al. Programmable full-adder computations in communicating three-dimensional cell cultures[J]. *Nature Methods*, 2018, 15(1): 57-60.
- [92] DAVIS R M, MULLER R Y, HAYNES K A. Can the natural diversity of quorum-sensing advance synthetic biology? [J]. *Frontiers in Bioengineering and Biotechnology*, 2015, 3: 30.
- [93] CURATOLO A I, ZHOU N, ZHAO Y, et al. Cooperative pattern formation in multi-component bacterial systems through reciprocal motility regulation[J]. *Nature Physics*, 2020, 16: 1152-1157.
- [94] 周楠, 夏婷颖, 黄建东. 合成生物学在探索生物图案形成基本原理中的应用与展望[J]. *合成生物学*, 2020, 1(4): 470-480. ZHOU N, XIA T Y, HUANG J D. Applications and prospects of synthetic biology in exploring the basic principles of biological pattern formation[J]. *Synthetic Biology Journal*, 2020, 1(4): 470-480.
- [95] KHAN M S, KIM E, MCPHERSON A, et al. Adenovirus-vectored SARS-CoV-2 vaccine expressing S1-N fusion protein [J]. *Antibody Therapeutics*, 2022, 5(3): 177-191.
- [96] LI H S, ISRANI D V, GAGNON K A, et al. Multidimensional control of therapeutic human cell function with synthetic gene circuits[J]. *Science*, 2022, 378(6625): 1227-1234.
- [97] 唐婷, 付立豪, 郭二鹏, 等. 自动化合成生物技术与工程化设施平台[J]. *科学通报*, 2021, 66(3): 300-309. TANG T, FU L H, GUO E P, et al. Automation in synthetic biology using biological foundries[J]. *Chinese Science Bulletin*, 2021, 66(3): 300-309.
- [98] CAMACHO D M, COLLINS K M, POWERS R K, et al. Next-generation machine learning for biological networks[J]. *Cell*, 2018, 173(7): 1581-1592.
- [99] NIELSEN A A K, DER B S, SHIN J, et al. Genetic circuit design automation[J]. *Science*, 2016, 352(6281): aac7341.
- [100] TAKETANI M, ZHANG J B, ZHANG S Y, et al. Genetic circuit design automation for the gut resident species *Bacteroides thetaiotaomicron*[J]. *Nature Biotechnology*, 2020, 38(8): 962-969.
- [101] CHEN Y, ZHANG S Y, YOUNG E M, et al. Genetic circuit design automation for yeast[J]. *Nature Microbiology*, 2020, 5(11): 1349-1360.
- [102] JONES T S, OLIVEIRA S M D, MYERS C J, et al. Genetic circuit design automation with Cello 2.0[J]. *Nature Protocols*, 2022, 17(4): 1097-1113.
- [103] MYERS C J, BARKER N, JONES K, et al. iBioSim: a tool for the analysis and design of genetic circuits[J]. *Bioinformatics*, 2009, 25(21): 2848-2849.
- [104] WATANABE L, NGUYEN T, ZHANG M, et al. iBioSim 3: a tool for model-based genetic circuit design[J]. *ACS Synthetic Biology*, 2019, 8(7): 1560-1563.
- [105] HILLSON N J, ROSENGARTEN R D, KEASLING J D. j5 DNA assembly design automation software[J]. *ACS Synthetic Biology*, 2012, 1(1): 14-21.
- [106] WILSON M L, HERTZBERG R, ADAM L, et al. A step-by-step introduction to rule-based design of synthetic genetic constructs using GenoCAD[J]. *Methods in Enzymology*, 2011, 498: 173-188.
- [107] HOLOWKO M B, FROW E K, REID J C, et al. Building a biofoundry[J]. *Synthetic Biology*, 2021, 6(1): ysaa026.
- [108] CHAO R, LIANG J, TASAN I, et al. Fully automated one-step synthesis of single-transcript TALEN pairs using a biological foundry[J]. *ACS Synthetic Biology*, 2017, 6(4): 678-685.
- [109] SI T, CHAO R, MIN Y H, et al. Automated multiplex genome-scale engineering in yeast[J]. *Nature Communications*, 2017, 8: 15187.
- [110] 崔金明, 张炳照, 马迎飞, 等. 合成生物学研究的工程化平台[J]. *中国科学院院刊*, 2018, 33(11): 1249-1257. CUI J M, ZHANG B Z, MA Y F, et al. Engineering platforms for synthetic biology research[J]. *Bulletin of Chinese Academy of Sciences*, 2018, 33(11): 1249-1257.
- [111] ZHANG S Y, ZHU J, FAN S, et al. Directed evolution of a cyclodipeptide synthase with new activities *via* label-free mass spectrometric screening[J]. *Chemical Science*, 2022, 13(25): 7581-7586.
- [112] GUO E P, FU L H, FANG X T, et al. Robotic construction and screening of lanthipeptide variant libraries in *Escherichia coli* [J]. *ACS Synthetic Biology*, 2022, 11(12): 3900-3911.
- [113] ZHANG J Z, LI T, HONG Z L, et al. Biosynthesis of hybrid neutral lipids with archaeal and eukaryotic characteristics in engineered *Saccharomyces cerevisiae*[J]. *Angewandte Chemie International Edition*, 2023, 62(4): e202214344.
- [114] DENG H X, YU H, DENG Y W, et al. Pathway evolution through a bottlenecking-debottlenecking strategy and machine learning-aided flux balancing[J]. *Advanced Science*, 2024: e2306935.
- [115] 卢挥, 张芳丽, 黄磊. 合成生物学自动化装置 iBioFoundry 的构建与应用[J]. *合成生物学*, 2023, 4(5): 877-891. LU H, ZHANG F L, HUANG L. Establishment of iBioFoundry for synthetic biology applications[J]. *Synthetic Biology Journal*, 2023, 4(5): 877-891.
- [116] HU R Y, FU L H, CHEN Y C, et al. Protein engineering *via* Bayesian optimization-guided evolutionary algorithm and robotic

- experiments[J]. *Briefings in Bioinformatics*, 2023, 24(1): bbac570.
- [117] WANG Y, WANG H C, WEI L, et al. Synthetic promoter design in *Escherichia coli* based on a deep generative network [J]. *Nucleic Acids Research*, 2020, 48(12): 6403-6412.
- [118] CONSTANT D A, GUTIERREZ J M, SASSTRY A V, et al. Deep learning-based codon optimization with large-scale synonymous variant datasets enables generalized tunable protein expression[EB/OL]. *bioRxiv*, 2023, 2023.02.11.528149. (2023-02-12) [2023-12-01]. <https://www.biorxiv.org/content/10.1101/2023.02.11.528149v1>.
- [119] ALIPANAHI B, DELONG A, WEIRAUCH M T, et al. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning[J]. *Nature Biotechnology*, 2015, 33(8): 831-838.
- [120] WANG M, TAI C, E W N, et al. DeFine: deep convolutional neural networks accurately quantify intensities of transcription factor-DNA binding and facilitate evaluation of functional non-coding variants[J]. *Nucleic Acids Research*, 2018, 46(11): e69.
- [121] VALERI J A, COLLINS K M, RAMESH P, et al. Sequence-to-function deep learning frameworks for engineered riboregulators[J]. *Nature Communications*, 2020, 11(1): 5058.
- [122] SEAK L C U, LO O L I, SUEN W, et al. Next-generation biocomputing: mimicking artificial neural network with genetic circuits[EB/OL]. *bioRxiv*, 2021, 2021.03.12.435120. (2021-03-12) [2023-12-01]. <https://www.biorxiv.org/content/10.1101/2021.03.12.435120v1>.



通讯作者: 王宝俊(1982—),男,浙江大学求是讲席教授,教育部长江学者讲席教授。研究方向为合成生物学和生物工程,长期从事合成生物使能技术、基因线路设计研究及其在生物传感、智能治疗和生物制造等领域的创新应用。
E-mail: baojun.wang@zju.edu.cn



第一作者: 高歌(1994—),女,博士,助理研究员。研究方向为合成生物学基因线路设计、肿瘤细菌疗法。
E-mail: gaoge@zju.edu.cn



共同第一作者: 边旗(1993—),女,博士,助理研究员。研究方向为合成生物学使能技术开发、代谢工程。
E-mail: bianqi@zju.edu.cn