

DOI: 10.13376/j.cblls/2018095

文章编号: 1004-0374(2018)08-0799-11

· 评述与综述 ·

## 通过合成基因线路实现细胞分类

胡冠华<sup>1,2</sup>, 王也<sup>1,2</sup>, 魏磊<sup>1,2</sup>, 汪小我<sup>1,2\*</sup>

(1 清华大学自动化系, 合成与系统生物学研究中心, 教育部生物信息学重点实验室, 北京100084; 2 北京信息科学与技术国家研究中心生物信息学研究部, 北京100084)

**摘要:** 合成生物学是生物学与工程学结合的新兴学科, 通过人工将生物元器件组合成线路引入细胞, 使细胞获得对信息进行处理并做特定输出的新功能。近年来, 针对疾病治疗的合成生物学研究发展迅猛, 基因线路的工程化特性使通过它对疾病进行更精确、灵活的干预成为可能, 在基因治疗中有广泛的应用前景。精确干预的前提是综合多种输入信号并识别出特定种类的细胞, 如特异性识别出癌细胞的溶瘤病毒。疾病的发生往往伴随着细胞内多种调控网络的改变, 从中提取出关键信号分子作为基因线路的输入至关重要。现综述哺乳动物细胞中对不同的信号输入进行感知的基因线路细胞分类器, 为未来模块化整合信号输入、设计基因线路提供新的思路。

**关键词:** 基因线路; 基因治疗; 细胞分类器; 合成生物学

**中图分类号:** Q819 **文献标志码:** A

## Cell type classifying by synthetic gene circuits

HU Guan-Hua<sup>1,2</sup>, WANG Ye<sup>1,2</sup>, WEI Lei<sup>1,2</sup>, WANG Xiao-Wo<sup>1,2\*</sup>

(1 Ministry of Education Key Laboratory of Bioinformatics, Center for Synthetic and System Biology, Department of Automation, Tsinghua University, Beijing 100084, China; 2 Bioinformation Division, BNRist, Beijing 100084, China)

**Abstract:** Synthetic biology is an arising interdisciplinary combining traditional biology and engineering, aiming at integrating biological components into gene circuits and delivering them into cells, enabling them to process information and perform certain function. Recently, gene therapy for diseases has developed rapidly. The engineering nature of genetic circuits has made it possible to intervene diseases with more spatiotemporal precision, giving gene circuits broad applications in gene therapy. Accurate intervention benefits from the integration of multiple input signals and the identification of specific types of cells, such as oncolytic viruses that specifically recognize cancer cells and kill them. Diseases are often accompanied by changes in a variety of regulatory networks within cells. So, it is crucial to extract key signal molecules from these cells as inputs for gene circuits. This paper reviews gene circuit cell classifiers that sense different signal inputs in mammalian cells, providing new ideas for modular integration of signal inputs and the design of gene circuits in the future.

**Key words:** gene circuits; gene therapy; cell type classifier; synthetic biology

合成生物学是一门将生命科学、信息科学和工程学结合的综合性的新型交叉学科。合成基因线路的设计与构建作为合成生物学的重要研究方向, 借鉴了电子电路的相关理念<sup>[1]</sup>, 将生物元器件标准化、模块化地组装<sup>[2-3]</sup>, 在宿主细胞中实现特定功能。随着分子生物学和生物信息学的飞速发展, 众多生物分子的功能被逐渐揭示, 为合成基因线路的设计

构建提供了潜在素材。同时, 对基因组 DNA 调控序列以及蛋白质互作机理理解的加深为理性组合这些素材提供了设计指导。早期的合成基因线路研究

收稿日期: 2018-05-10; 修回日期: 2018-07-18

基金项目: 国家自然科学基金项目(61773230, 61721003)

\*通信作者: E-mail: xwwang@tsinghua.edu.cn

聚焦于在原核细胞中实现程序化的运算功能,如振荡器<sup>[4-5]</sup>、双稳态开关<sup>[6-7]</sup>、逻辑门<sup>[8]</sup>、细胞通讯<sup>[9-10]</sup>等。随着人们对真核生物细胞,特别是哺乳动物细胞在转录、转录后、翻译和翻译后过程调控机制的深入理解和基因编辑等工具的发展,合成基因线路在医疗等领域展现出了广泛的应用前景,如细胞免疫疗法<sup>[11-12]</sup>、治疗肿瘤的溶瘤病毒<sup>[13]</sup>、基于细胞的代谢检测反馈<sup>[14-19]</sup>等。

合成基因线路的设计中主要有两种思路。一种旨在设计具有特定的、独立于宿主细胞本身功能的基因线路,如生物制造<sup>[20]</sup>等,在这种思路中宿主细胞被视为基因线路的承载者。而另一种思路着重于通过与宿主细胞进行互动,感知细胞内源的生物信号并对宿主细胞的生物功能进行干预,这种思路的基因线路更接近于细胞内的调控或信号通路,其目的是建立起内源信号与目标输出之间的关系<sup>[21-24]</sup>。后一种设计思路的一个重要应用是基因线路细胞分类器,在活细胞中识别、整合和处理信号输入,辨别不同的细胞类型,从而对特定细胞类型执行特定的功能,其大致由三个模块组成(图1)。(1)输入:通过基因元件感知细胞中特定信号分子的水平,例如利用细胞特异性启动子感知特定细胞中的转录信号;(2)运算:对不同信号的输入进行逻辑整合,如利用两个癌细胞特异性启动子分别驱动转录激活子不同结构域的表达,两者同时具有活性时激活下游报告基因,实现AND门(AND gate,又称与门、与逻辑)的功能;(3)输出:利用逻辑运算的结果驱动输出信号从而实现特定的生物功能,如标记或选择性杀伤。

哺乳动物基因线路细胞分类器靶向特定的细胞,在医疗特别是个性化医疗上有巨大的应用潜力,将来的设计可能不仅针对某一种疾病,还能对特定个人或者个人体内的特定细胞类型进行操作,实现

个性化精准治疗<sup>[25]</sup>。目前基因线路细胞分类器的发展难点集中在如何设计能整合多种输入且具有鲁棒性的系统来靶向特定的细胞,这依赖于对细胞内不同信号分子机制的理解,从而设计传感器实现对它们的感知。本文根据输入信号类型介绍哺乳动物细胞分类基因线路及其运算设计,并为以后理性设计综合各种信号输入的基因线路细胞分类器提供新的思路。

## 1 细胞分类基因线路的不同输入

### 1.1 转录调控信号检测

在哺乳动物细胞中,转录因子(transcription factor, TF)是一类被广泛研究的富含信息的内源信号分子,其可以结合到启动子上调控基因表达。同一启动子可能因为不同类型细胞中各种TF表达量的不同而呈现表达特异性<sup>[26]</sup>。这一性质被广泛应用于目前的多种基因线路细胞分类器中:利用组织特异性启动子启动基因的表达,从而在特定种类的细胞中表达适量蛋白。然而,基于单一启动子的信息决策容易受到基因表达噪声的影响,因而往往有一定假阳性,且面对众多类型的背景细胞时其特异性会受到限制。例如在一些将癌细胞特异性启动子驱动的毒性蛋白导入细胞中试图特异性杀伤癌细胞的研究中,这些特异性启动子可能特异性不足,导致效应蛋白在某些正常组织中表达,从而产生对正常细胞的杀伤<sup>[27-28]</sup>。因此,很多研究采用了组织特异性启动子与癌症特异性启动子结合的方法,通过蛋白质间的相互作用、蛋白质与RNA间的相互作用或RNA间的相互作用来实现两个启动子之间输入信息的运算和整合。基于蛋白质相互作用的信息处理是迅速高效的。例如在Nissim和Bar-Ziv<sup>[29]</sup>基于双杂交系统的基因线路(图2A)中,两种特异性启动子分别转录调控因子的N端和C端,只有两个

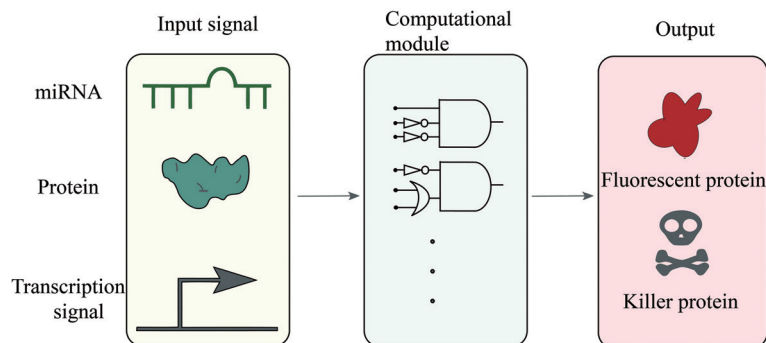
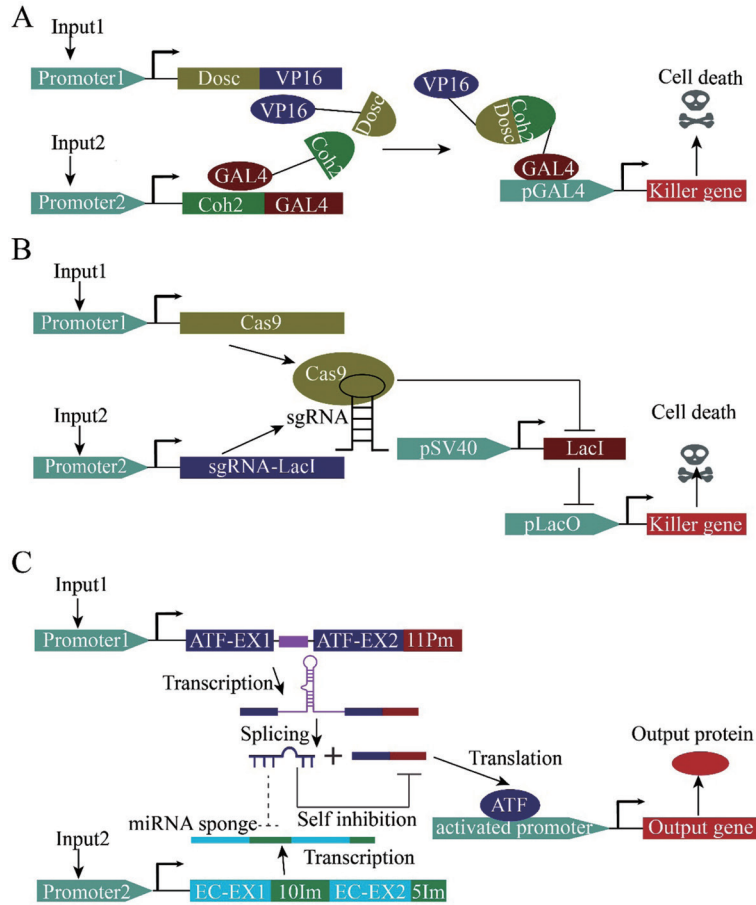


图1 合成基因线路组成(输入、运算、输出)



(A)利用蛋白质相互作用构建与逻辑。GAL4: 融合蛋白半乳糖调节上游元件(glucose-regulated upstream promoter element); VP16: 转录激活子(viral protein 16); DocS、Coh2: 两种细菌蛋白能高亲和性地结合, 都存在时能够组成一个具有功能的转录激活子GAL4-Coh2-DocS-VP16, 并结合至启动子pGAL4上, 激活促凋亡蛋白的表达, 最终导致肿瘤细胞的死亡。(B)利用Cas9与sgRNA相互作用构建与逻辑。pSV40: 猿猴病毒40启动子(simian virus promoter)。Promoter1启动表达Cas9蛋白, Promoter2启动表达靶向LacI的sgRNA, 两者都有活性时sgRNA靶向至转录抑制子LacI的基因, 招募Cas9蛋白, 阻断蛋白LacI的表达, 进而使LacI不能作用于操纵子LacO, 从而启动表达抗肿瘤的效应蛋白。(C)利用RNA竞争效应构建与逻辑。Pm: perfect match, 完全配对; Im: imperfect match, 不完全配对; EC: ECFP, 青色荧光蛋白。只有当Promoter2驱动的带有多不完全匹配靶点的mRNA被表达出来竞争性地结合miRNA, Promoter1驱动的转录激活子才能从其自抑制中解除出来激活下游信号表达。

图2 转录水平传感器

特异性启动子同时表达时, 才能形成完整的调控因子激活下游基因的表达, 这实际上实现了一个逻辑AND运算。相似地, 一种能自我切割的蛋白结构域内含肽(intein)也被应用到逻辑门的构建中。内含肽被分成两部分, 分别与两种蛋白质连接, 只有当两种蛋白质同时存在时, 它们才能发生反式剪切形成完整蛋白并发挥功能, 从而实现AND门<sup>[30]</sup>。Liu等<sup>[31]</sup>利用双特异性启动子和CRISPR/Cas9(clustered regular interspaced short palindromic repeats/CRISPR-associated protein)系统设计基因线路进行膀胱癌检测, 两个特异性启动子分别被癌细胞特异性启动子(人端粒酶逆转录酶启动子)和膀胱上皮

特异性启动子(hUP2)驱动, 分别表达Cas9蛋白和靶向乳糖操纵子调节基因(LacI)的非编码向导RNA(sgRNA), LacI表达被抑制后不能作用于乳糖操纵子调控序列(LacO), 最终使报告基因荧光素酶或者毒性蛋白表达(图2B)。RNA间的相互作用也能用于基因线路输入信号的整合。例如Nissim等<sup>[32]</sup>通过RNA竞争效应构建了不同启动子信号之间的与逻辑门。特异性启动子1驱动转录激活子(ATF)表达, 且该基因内含子中插入了miRNA, 可在转录剪切后加工出miRNA。在ATF基因的3'端非翻译区(3'UTR)有该miRNA完全配对的位点, 当该基因转录剪切后miRNA通过切割mRNA形成自抑制,

抑制 ATF 蛋白翻译。特异性启动子 2 驱动表达的 mRNA 上含有与 miRNA 不完全互补配对的位点, 通过形成 miRNA 分子海绵 (miRNA sponge) 产生竞争作用, 从而解除 ATF 基因的翻译自抑制。因此, 当两个基因同时表达时, ATF 才能翻译从而激活下游信号 (图 2C)。

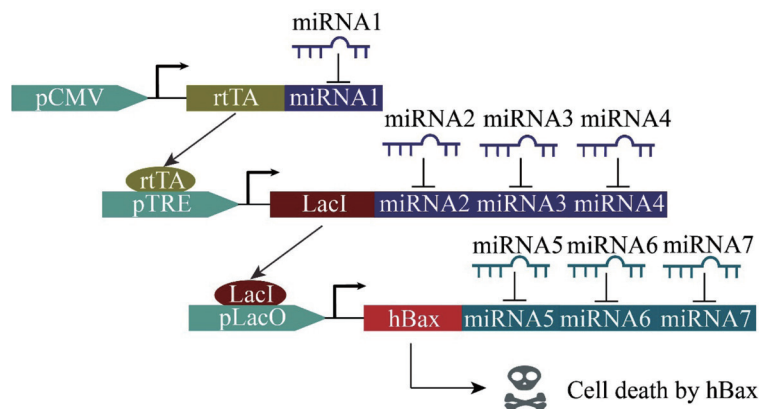
### 1.2 microRNA信号检测

microRNA (miRNA) 是一类短的非编码 RNA, 依靠碱基互补配对完全或不完全匹配地与靶 mRNA 结合调节基因的表达<sup>[33-36]</sup>。它们在不同的组织和细胞系中差异地表达<sup>[37-38]</sup>, 控制着众多正常的和病理性的生物过程<sup>[39-42]</sup>。有研究表明, 将组织特异性 miRNA 整合到基因线路中能够提高基因治疗的特异性<sup>[43-45]</sup>, 将外源基因的表达限制在特定的细胞类型中。同时, miRNA 相对较短, 容易被整合到基因线路中作为输入, 灵活性强。Xie 等<sup>[46]</sup>最先提出了以 miRNA 作为输入实现逻辑运算的设计原则, 靶向同一 mRNA 的 miRNA 之间构成 NOR 逻辑, 这些 miRNA 被作为输入信号或者被上游输入信号激活或者抑制, 可以实现不同信号间的任意逻辑运算整合。他们在后续研究中利用 miRNA 表达谱设计构建了基于特异性 miRNA 作为输入信号的肿瘤细胞基因线路分类器 (图 3), 可特异识别 HeLa 细胞中的内源 miRNA 并表达促凋亡蛋白 (hBax), 从而杀伤肿瘤细胞。这种方法可以分别感知 HeLa 细胞中高表达的 miRNA 及低表达的 miRNA 两类信号。高表达的 miRNA 通过抑制反四环素转录激活子 (reverse tetracycline transcriptional activator, rtTA)

的表达, 解除其对 hBax 的抑制, 低表达的 miRNA 直接抑制 hBax 表达, 这样这些 miRNA 之间形成逻辑门, 只有当 miRNA 符合 HeLa 细胞表达谱时, hBax 才能表达。很多研究基于该平台提出了优化方法。清华大学谢震课题组<sup>[47]</sup>利用模块化的转录激活抑制子 (TALER) 构建双稳态开关, 并利用内源 miRNA 来控制该开关, 从而增强了系统特异性。另一项研究利用重组酶控制抑制子和其所抑制基因的表达时序, 使得在瞬时转染中当抑制子积累到一定量时输出基因才能够表达, 增加了基因线路输出的动态范围, 减少了表达泄漏<sup>[48]</sup>。目前这种以 miRNA 作为信号输入的基因线路平台已经被广泛应用到了各种细胞分类问题上, 如特定神经细胞的标记、分化细胞与干细胞的分选等<sup>[48-49]</sup>。

### 1.3 蛋白质信号检测

蛋白质是细胞表型的主要决定者, 很多生物功能受蛋白质浓度和活性的调控, 感受胞内蛋白质水平的基因线路细胞分类器在生物医学上有广泛的应用前景。目前感应内源非转录因子的蛋白质水平主要通过蛋白质响应的 RNA 开关来实现, 这种 RNA 开关与蛋白质结合后发生构象改变, 从而影响基因表达。这种调控方式与依赖转录因子的调控系统相比, 可直接嵌入基因表达框或者调控区域发挥顺式作用, 结构更加简单, 易于构建。如 Culler 等<sup>[22]</sup>构建的核内蛋白传感器, 它利用了基因可变剪切机制, 在输出蛋白基因序列的一个中间外显子上插入了一个终止密码子, 同时在内含子中插入了待感应蛋白的适体。当无输入蛋白信号时, 中间外显子被



HeLa细胞中高表达的miRNA1、2、3与低表达的miRNA4、5、6之间形成了逻辑关系, 只有当miRNA1、2、3高表达, 而miRNA4、5、6低表达时, 符合这种表达谱的细胞中诱导凋亡蛋白hBax才能表达并杀死细胞。pTRE: 诱导型启动子; pCMV: 组成型启动子(CMV enhancer/chicken  $\beta$ -catenin promoter)。

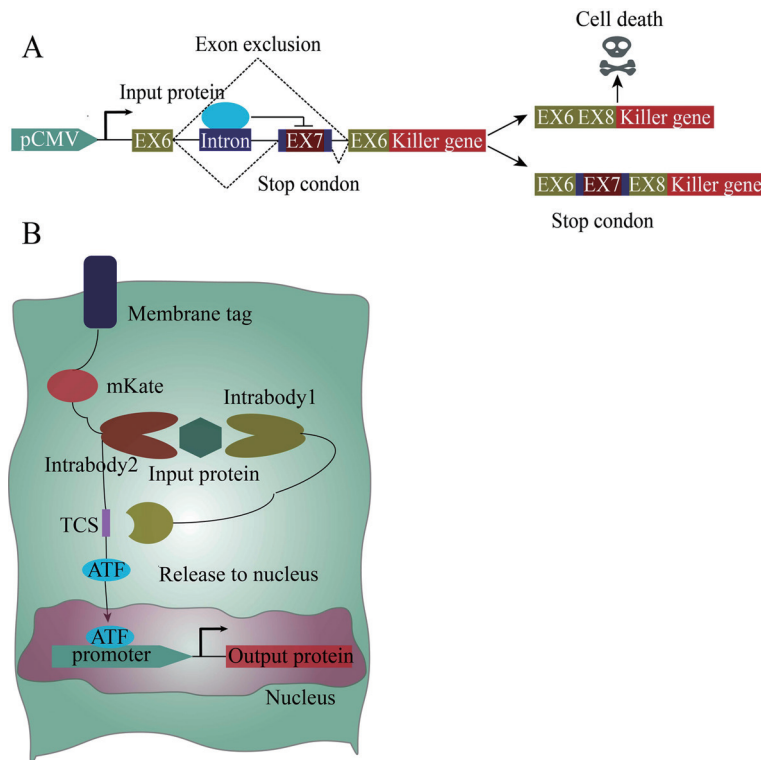
图3 感知miRNA的基因线路分类器

剪切入 mRNA, 目标蛋白无法正常翻译; 当输入蛋白存在且与适体结合时, RNA 二级结构发生变化, 从而通过空间位阻效应影响剪切模式, 导致中间外显子不能被剪切入 mRNA, 使目标蛋白表达 (图 4A)。他们同时也尝试利用该系统实现多信号整合, 在同时插入 MS2 和 NF- $\kappa$ B 适体后, 传感器在 MS2 和 NF- $\kappa$ B 单独存在时均表达上升, 且同时存在时表达上升量大于单独存在时的和, 证明适体间有协同作用, 但因单独表达泄露太大尚不足以构建与逻辑门。Kashida 等<sup>[49]</sup>将蛋白适体与短发夹 RNA (shRNA) 结合起来, 蛋白配体结合到 shRNA 上会影响 shRNA 与 Dicer 酶的结合从而使基因无法被沉默, 由此感知内源蛋白水平。RNA 适体只依赖 RNA 即可发挥作用, 而 DNA 的表达系统有造成宿主细胞基因组损伤等局限性, 一些基于 RNA 的基因线路也被开发出来。Kawasaki 等<sup>[50]</sup>利用 LIN28A 蛋白响应的 mRNA 基因线路区分了干细胞与分化后细胞。

核酶核糖开关也是一种基于 RNA 构象变化的信号感应传导系统, 在响应信号后诱导 mRNA 自切割实现基因调控<sup>[51]</sup>。Win 等<sup>[52]</sup>提出了从包含适

体的感知模块、包含核酶的功能模块和将两者耦合起来的传递模块来模块化设计核糖开关的原则。Kennedy 等<sup>[53]</sup>则提出了蛋白质响应核糖开关的体外筛选和构建方法, 通过传递模块的改造实现了输入 ON 和 OFF 开关; 且因核酶核糖开关不受上述 RNA 开关因作用机制导致的只能感应核内或胞质中的蛋白质的局限, 可实现对细胞总蛋白量的感应, 扩展了蛋白质传感器的使用范围。

2018 年, Siciliano 等<sup>[54]</sup>开发的基于胞内抗体的蛋白质检测系统也实现了对疾病蛋白分子的响应。该系统基于两个能与目标蛋白特异相互作用的胞内抗体。一个胞内抗体 N 端被连接在与膜锚定的 mKate (红色荧光蛋白) 上, C 端连接上烟草蚀刻病毒切割位点 (TCS) 和转录激活子, 这样转录激活子就只存在于细胞质中。另一个胞内抗体连接有烟草蚀刻病毒蛋白酶 (TEVp), 能特异识别 TCS 并切割。当目标蛋白存在时, 它能与两种胞内抗体同时相互作用, 将 TEVp 拉至 TCS 附近, TEVp 识别 TCS 并切割, 释放转录激活子, 这样转录激活子就可以入核激活目的基因表达 (图 4B)。



(A)基于可变剪接的蛋白质传感器。待感应蛋白存在时与内含子中的适体结合, 使基因剪切模式发生变化, 含有终止子的外显子7 (EX7)被排除在mRNA之外, 目标蛋白表达从而诱导细胞凋亡。(B)基于胞内抗体的蛋白质传感器。当目标蛋白存在时, 它能与两种胞内抗体同时相互作用, 将TEVp拉至TCS附近, TCS发生剪切, ATF被释放到胞内, 激活输出蛋白表达。ATF: 转录激活因子; mKate: 红色荧光蛋白; Intrabody: 胞内抗体; TEVp: 烟草蚀刻病毒蛋白酶; TCS: 酶识别及切割位点。

图4 基于胞内抗体的蛋白传感器

近年来 CRISPR/Cas9 技术迅猛发展, 提供了灵活的基因编辑及调控功能, 其功能依赖于 sgRNA 的引导, 因此将蛋白 RNA 适体与 sgRNA 结合起来成为构建蛋白质传感器的新思路。深圳市第二人民医院的黄卫人、蔡志明课题组<sup>[55]</sup>将响应蛋白信号的 RNA 适体放入 CRISPR 系统的 sgRNA 中, 当蛋白适体存在时诱导 sgRNA 构象变化从而与靶 DNA 结合发挥抑制或激活作用。他们设计了感知癌细胞中高表达的 NF- $\kappa$ B 和  $\beta$ -catenin 的基因线路, 特异性地诱导了癌细胞的凋亡。CRISPR 有激活和抑制两种系统, 且不同 sgRNA 对应的调节强度不同, 因此他们通过组合不同的 sgRNA 实现了多种感应两蛋白输入的逻辑结构。如通过构建响应 P53 蛋白与 HSF1 的强 CRISPR 抑制系统靶向同一目的基因, 只有 P53 和 HSF1 蛋白都不存在时目的基因才能表达, 实现了 NOR 逻辑。CRISPR 相关的 RNA 支架体系功能强大, 可以实现复杂的转录控制, 相关的蛋白质传感器具有广阔的应用前景。

#### 1.4 细胞表面抗原信号检测

近年来, 癌症基因治疗和免疫疗法发展迅速, 其中很大一部分以细胞表面抗原作为靶点。嵌合抗原受体 T 细胞 (chimeric antigen receptor-T cell, CAR-T) 就是其中一种<sup>[56]</sup>。该方法通过人为地将单链抗体 (single-chain variable fragment, scFv) 与跨膜区域和胞内信号转导区结合起来, 组装到患者的 T 细胞中, 使改造后的 T 细胞可以表达肿瘤抗原受体; 这种改造不依赖于天然的抗原提呈机制, 对特定肿瘤抗原具有高度亲和性, 并能进行特异性肿瘤细胞杀伤<sup>[57]</sup>。研究发现, 通过增加胞内信号转导区的共刺激分子, 可以增强 T 细胞的响应和杀伤性能<sup>[58-60]</sup>。CAR-T 方法虽然在临床上已经取得一定的治疗效果, 但一

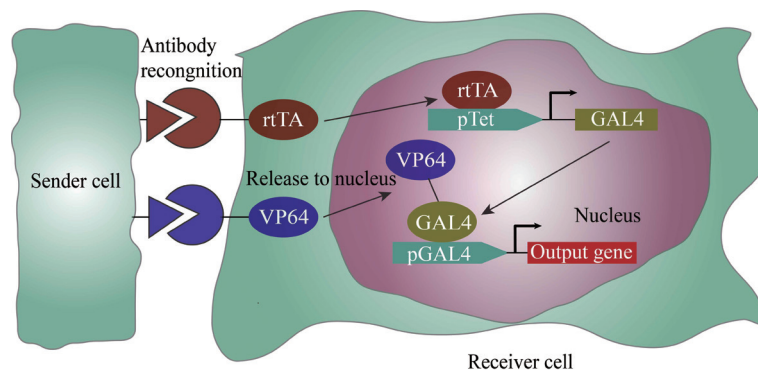
些肿瘤细胞缺少特定的表面抗原而从免疫疗法中逃逸是目前该方法面临的一大挑战<sup>[61]</sup>。目前已经有一些研究通过整合更多的 scFv 结构域实现对不同表面抗原的布尔逻辑整合从而提高疗法的有效性<sup>[62-64]</sup>, 如 Zah 等<sup>[62]</sup>设计的 OR 逻辑 CAR 能识别 B 淋巴细胞抗原 CD19 或 CD20 两种分子标记物, 减少了疗法的免疫逃逸。

Notch 是另一种可以进行工程化改造的信号转导工具。Notch 受体与其 Delta 配体结合使得膜蛋白水解释放 Notch 胞内转录调节结构域进入核内调控基因转录表达<sup>[65]</sup>。Notch 的胞内信号转导域和胞外配体区域都可进行工程化。Morsut 等<sup>[66]</sup>设计了人工合成 Notch (synotch), 保留跨膜自剪切区域, 并将胞外结构区域替换为单链抗体, 胞内区域替换成转录激活域。这种人工合成 Notch 在表达后可以特异性识别与其相对应的抗原细胞, 释放转录信号调节下游基因表达。由于大部分抗原在正常组织中也存在一定水平的表达, 针对表面抗原的方法会造成脱靶效应, 从而对非肿瘤组织造成损伤。针对该问题, 不同的合成 Notch 具有良好的正交性, 可以基于此设计具有 AND 逻辑功能的基因线路使得决策更加准确, 提高肿瘤细胞识别的特异性<sup>[67]</sup>(图 5)。

## 2 未来的发展方向

### 2.1 特异性启动子的人工优化设计

特异性启动子研究是基因治疗研究中的一个重要领域。如第二代条件复制溶瘤腺病毒由癌症特异性启动子驱动, 其溶瘤效率主要取决于特异性启动子的特异性和长度等<sup>[68-70]</sup>, 这种特异性启动子感知的转录信号本质上是细胞内的转录因子水平。然而, 有时在这种细胞分类问题中无法找到特异性足够强



pTet: 诱导型启动子, 在rtTA和四环素作用下启动; UAS: 上游激活序列; VP64: 转录激活域, VP16 的四聚衍生物。只有当两种细胞表面抗原同时存在时才能形成VP64-GAL4转录激活子激活输出蛋白的表达。

图5 由synotch组成的AND逻辑示意图

的启动子, 在这种情况下, 人工设计特异性启动子, 定制化地感知转录因子水平是解决该问题的一种潜在方法。

首先, 可以优化特异性启动子, 例如在癌细胞特异性启动子上增加厌氧响应元件、炎症因子响应元件<sup>[71]</sup>等。调节启动子活性的法则十分复杂, 转录因子调控元件 (TFRE) 的排列方向和顺序、空间位置以及它们对应 TF 的表达水平、活性和作用方式等都能影响启动子的活性<sup>[72-73]</sup>。因此, 很多工作通过构建启动子文库筛选得到特异性启动子。如 Roberts 等<sup>[74]</sup>通过表达谱分析和文献挖掘确定了在结肠癌细胞中高表达的基因, 然后通过生物信息学方法得到在这些基因启动子区富集的 TFRE, 将它们随机连接后连入 TATA 框 (TATA box) 上游, 构建启动子文库, 并通过报告基因筛选到了结肠癌细胞特异性启动子。而一些高通量检测启动子活性的方法, 如大规模平行报告系统 (massively parallel reporter assays, MPRA), 也使对包含大量启动子序列的文库进行筛选成为可能<sup>[75]</sup>。

随着对启动子活性调节法则的研究不断进展, 未来通过理解启动子结构如何影响基因表达从而理性设计启动子可能是该领域的研究方向。如 Angelici 等<sup>[76]</sup>发现, 在上游转录激活子与 TF 具有较强的协同作用时可以形成 AND 门, 而调整 TF 结合位点的长度、调整激活子结合域与 TATA box 的距离或者更换转录激活域蛋白可以调节协同作用的强弱, 使任意 TF 之间的 AND 门成为可能, 他们依此设计基因线路分类器实现了细胞分类。Smith 等<sup>[77]</sup>通过对大量人工设计序列进行系统性的测试发现, 启动子活性主要由不同 TFRE 的拷贝数决定。但一些 TFRE 需要形成与特定结构与蛋白质相互作用, 它们的位置也影响启动子活性<sup>[78-79]</sup>。天然启动子往往混合着对位置敏感和不敏感的 TFRE<sup>[80]</sup>, 因此, Brown 等<sup>[81]</sup>提出了鉴定并利用位置不敏感的 TFRE 来设计启动子, 简化了启动子的设计。

目前已经有一些研究尝试利用序列信息<sup>[82-84]</sup>、核小体结构<sup>[85]</sup>等构建数学模型来理解和解释启动子的表达特性。未来随着高通量数据如免疫共沉淀测序 (CHIP-seq)、染色质可接近性、染色质相互作用等数据的积累, 可以通过对 TFRE 组合的数据挖掘和机器学习等方法更好地理解启动子结构与基因表达之间的关系, 更好地设计满足需求的特异性启动子。另外, 人工智能领域以生成对抗网络 (GAN) 为代表的生成式模型技术发展非常快速, 值得特别

关注, 这些技术为设计和创造新的人工调控序列提供了新的机遇<sup>[86]</sup>。

## 2.2 拓展可感知的信号分子范围

哺乳动物细胞基因线路分类器依赖于对不同细胞状态的感知和决策, 特定的细胞状态可通过各种分子标记物体现出来。可作为分子标记物的化学物多种多样, 设计新的基因线路, 拓展可感知的信号分子类型, 也是未来基因线路分类器的一个发展方向。如在癌症中往往存在大量的基因突变, 设计基因线路靶向特定基因组 DNA 突变的细胞具有很大的挑战性。CRISPR 系统可以识别 DNA 序列, 但仅限于进行原位切割或招募特定蛋白, 无法将信息放大并传递到下游进行处理。Mircetic 等<sup>[87]</sup>利用 P53 突变后转录调节能力的变化构建了间接的突变传感器。近期也有研究设计了两个锌指蛋白 (zinc finger, ZF)-TF 组合, 两个 ZF 靶向目标 DNA 中两段相邻的 DNA 序列并分别连接转录因子的 DNA 结合域和转录激活域; 只有当目标 DNA 序列存在时, 两个 ZF-TF 组合才能结合到 DNA 上并经由内含肽介导的剪切释放出完整的转录因子触发下游的信号输出, 实现 DNA 序列的感知, 从而区分出腺病毒感染的细胞<sup>[88]</sup>。这种方法因 TF 起到转录调控作用需要积累到一定的量, 因此待感知的 DNA 需要在细胞中有多个拷贝。对基因组特定突变的感知需要对信号进行放大和传递, 仍需要更深入的研究。同时, mRNA 分子缺乏调控功能, 难以被整合到基因线路中作为输入。已经有研究在体外和原核生物中利用链置换反应实现了 mRNA 感知<sup>[89]</sup>, 但其在真核生物中的应用还有待研究。

## 2.3 系统整体设计

利用合成基因线路进行细胞分类依赖于不同特异性信号分子作为输入, 而这些输入在细胞群体中的表达存在异质性, 这被称为表达噪声。基因表达噪声可以影响分类基因线路的准确性和鲁棒性<sup>[88]</sup>。在上述的双启动子方法进行细胞分类的问题上, Morel 等<sup>[90]</sup>首次通过单细胞研究定量地揭示了细胞分类因细胞间表达的异质性存在特异性与敏感性之间的平衡, 并提出在具体问题中应通过调整基因线路参数来调节分类阈值, 以达到预期的分类效果。同时, 基因线路分类器可以抽象为信息传递的过程, 由人工设计的分类器接受上游信号, 进行逻辑运算后输出。人们可以通过借鉴信息论中的互信息、信道容量等模型定量地刻画信号传递的能力, 并通过设计提高系统的鲁棒性<sup>[91]</sup>。例如增加分类器的信号

输入的种类数可以提高信息量, 未来整合多种标志物分子信号输入的基因线路平台仍有待被开发。随着各种高通量技术的发展和数据的积累, 如何挑选出更好的标记物组合来更精确地定义和区分特定细胞状态也是一个重要问题, 通过计算方法分析大量数据, 总结出规律并辅助设计是未来解决该问题的新思路, 相关输入特征的选择及相应的基因线路的参数优化计算模型也有待提出。如 Mohammadi 等<sup>[92]</sup>提出了基于上述多 miRNA 输入的线路平台的两步优化算法, 第一步通过计算不同生化参数下基因线路的 ON/OFF ratio 优化全局的生化反应参数, 第二步利用遗传算法从不同细胞分类的 miRNA 表达谱数据中找出最优的 miRNA 组合及基因线路拓扑结构。该算法在仿真数据和实际分类问题数据中都取得了很好的效果, 并指导实验提高了细胞分类的准确性。

### 3 总结与展望

利用基因线路进行细胞分类有着重大的潜在应用价值。如何从数据中挖掘出信号分子标记物并根据其设计基因线路实现细胞分类仍需要系统的理论和实验方法。目前的基因线路细胞分类器大多停留在单一种类的输入信号上, 准确地进行细胞分类需要多种信号的整合, 设计和构建功能更加复杂的基因线路成为挑战, 应尽快提出相关理论并进行验证, 从而加速细胞分类器的设计和应用。

在设计过程中很多因素会影响分类器的性能, 如基因表达噪声<sup>[88]</sup>、宿主细胞与外源基因线路之间的相互作用<sup>[93-94]</sup>、下游元件对上游元件的负载效应等<sup>[95]</sup>。目前已经有一些理论被提出用来刻画上述问题, 如借鉴信息论中的理论描述基因线路的信道容量噪声传递<sup>[96-97]</sup>, 通过理解并应用自然界网络模体降低噪声并提高基因线路鲁棒性等<sup>[98]</sup>。相信随着合成生物学的发展, 理性设计复杂基因线路, 整合不同的信号输入来实现更精确的细胞分类将成为可能。

### [参 考 文 献]

- [1] Endy D. Foundations for engineering biology. *Nature*, 2005, 438: 449-53
- [2] Boettiger AN, Ralph PL, Evans SN. Transcriptional regulation: effects of promoter proximal pausing on speed, synchrony and reliability. *PLoS Comput Biol*, 2011, 7: e1001136
- [3] Mutalik VK, Guimaraes JC, Cambray G, et al. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Methods*, 2013, 10: 354-60
- [4] Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature*, 2000, 403: 335-8
- [5] Tiggles M, Marquez-Lago TT, Stelling J, et al. A tunable synthetic mammalian oscillator. *Nature*, 2009, 457: 309-12
- [6] Gardner TS, Cantor CR, Collins JJ. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 2000, 403: 339-42
- [7] Atkinson MR, Savageau MA, Myers JT, et al. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell*, 2003, 113: 597-607
- [8] Guet CC, Elowitz MB, Hsing WH, et al. Combinatorial synthesis of genetic networks. *Science*, 2002, 296: 1466-70
- [9] Bulter T, Lee SG, Woirl WW, et al. Design of artificial cell-cell communication using gene and metabolic networks. *Proc Natl Acad Sci USA*, 2004, 101: 2299-304
- [10] You LC, Cox RS, Weiss R, et al. Programmed population control by cell-cell communication and regulated killing. *Nature*, 2004, 428: 868-71
- [11] Abken H, Hombach A, Heuser C. Immune response manipulation: recombinant immunoreceptors endow T-cells with predefined specificity. *Curr Pharm Des*, 2003, 9: 1992-2001
- [12] Lesterhuis WJ, Haanen JB, Punt CJ. Cancer immunotherapy - revisited. *Nat Rev Drug Discov*, 2011, 10: 591-600
- [13] Cattaneo R, Miest T, Shashkova EV, et al. Reprogrammed viruses as cancer therapeutics: targeted, armed and shielded. *Nat Rev Microbiol*, 2008, 6: 529-40
- [14] Rossger K, Charpin-El-Hamri G, Fussenegger M. A closed-loop synthetic gene circuit for the treatment of diet-induced obesity in mice. *Nat Commun*, 2013, 4: 2825
- [15] Schukur L, Geering B, Charpin-El Hamri G, et al. Implantable synthetic cytokine converter cells with AND-gate logic treat experimental psoriasis. *Sci Transl Med*, 2015, 7: 201
- [16] Kemmer C, Gitzinger M, Daoud-El Baba M, et al. Self-sufficient control of urate homeostasis in mice by a synthetic circuit. *Nat Biotechnol*, 2010, 28: 355-60
- [17] Ye H, Xie M, Xue S, et al. Self-adjusting synthetic gene circuit for correcting insulin resistance. *Nat Biomed Eng*, 2017, 1: 0005
- [18] Shao J, Xue S, Yu G, et al. Smartphone-controlled optogenetically engineered cells enable semiautomatic glucose homeostasis in diabetic mice. *Sci Transl Med*, 2017, 9: eaa12298
- [19] Xie M, Ye H, Wang H, et al.  $\beta$ -cell-mimetic designer cells provide closed-loop glycemic control. *Science*, 2016, 354: 1296-301
- [20] Steen EJ, Kang YS, Bokinsky G, et al. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature*, 2010, 463: 559-U182
- [21] Auslander D, Auslander S, Charpin-El Hamri G, et al. A synthetic multifunctional mammalian pH sensor and CO<sub>2</sub> transgene-control device. *Mol Cell*, 2014, 55: 397-408



- [22] Culler SJ, Hoff KG, Smolke CD. Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. *Science*, 2010, 330: 1251-5
- [23] Kobayashi H, Kaern M, Araki M, et al. Programmable cells: interfacing natural and engineered gene networks. *Proc Natl Acad Sci USA*, 2004, 101: 8414-9
- [24] Slomovic S, Collins JJ. DNA sense-and-respond protein modules for mammalian cells. *Nat Methods*, 2015, 12: 1085-90
- [25] Re A. Synthetic gene expression circuits for designing precision tools in oncology. *Front Cell Dev Biol*, 2017, 5: 77
- [26] Hobert O. Gene regulation by transcription factors and microRNAs. *Science*, 2008, 319: 1785-6
- [27] Peng W, Chen J, Huang YH, et al. Tightly-regulated suicide gene expression kills PSA-expressing prostate tumor cells. *Gene Ther*, 2005, 12: 1573-80
- [28] Scaiewicz V, Sorin V, Fellig Y, et al. Use of H19 gene regulatory sequences in DNA-based therapy for pancreatic cancer. *J Oncol*, 2010, 2010: 178174
- [29] Nissim L, Bar-Ziv RH. A tunable dual-promoter integrator for targeting of cancer cells. *Mol Syst Biol*, 2010, 6: 444
- [30] Lienert F, Torella JP, Chen JH, et al. Two- and three-input TALE-based AND logic computation in embryonic stem cells. *Nucleic Acids Res*, 2013, 41: 9967-75
- [31] Liu Y, Zeng Y, Liu L, et al. Synthesizing AND gate genetic circuits based on CRISPR-Cas9 for identification of bladder cancer cells. *Nat Commun*, 2014, 5: 5393
- [32] Nissim L, Wu MR, Pery E, et al. Synthetic RNA-based immunomodulatory gene circuits for cancer immunotherapy. *Cell*, 2017, 171: 1138-50 e15
- [33] Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 2004, 116: 281-97
- [34] Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, 2005, 120: 15-20
- [35] Brennecke J, Stark A, Russell RB, et al. Principles of microRNA-target recognition. *PLoS Biol*, 2005, 3: e85
- [36] Guo H, Ingolia NT, Weissman JS, et al. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, 2010, 466: 835-40
- [37] Guo Z, Maki M, Ding R, et al. Genome-wide survey of tissue-specific microRNA and transcription factor regulatory networks in 12 tissues. *Sci Rep*, 2014, 4: 5150
- [38] Landgraf P, Rusu M, Sheridan R, et al. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, 2007, 129: 1401-14
- [39] Shenoy A, Belloch RH. Regulation of microRNA function in somatic stem cell proliferation and differentiation. *Nat Rev Mol Cell Biol*, 2014, 15: 565-76
- [40] Shivdasani RA. MicroRNAs: regulators of gene expression and cell differentiation. *Blood*, 2006, 108: 3646-53
- [41] Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer*, 2006, 6: 259-69
- [42] Hayes J, Peruzzi PP, Lawler S. MicroRNAs in cancer: biomarkers, functions and therapy. *Trends Mol Med*, 2014, 20: 460-9
- [43] Geisler A, Jungmann A, Kurreck J, et al. microRNA122-regulated transgene expression increases specificity of cardiac gene transfer upon intravenous delivery of AAV9 vectors. *Gene Ther*, 2011, 18: 199-209
- [44] Cawood R, Chen HH, Carroll F, et al. Use of tissue-specific microRNA to control pathology of wild-type adenovirus without attenuation of its ability to kill cancer cells. *PLoS Pathog*, 2009, 5: e1000440
- [45] Kelly EJ, Hadac EM, Greiner S, et al. Engineering microRNA responsiveness to decrease virus pathogenicity. *Nat Med*, 2008, 14: 1278-83
- [46] Xie Z, Wroblewska L, Prochazka L, et al. Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science*, 2011, 333: 1307-11
- [47] Li YQ, Jiang Y, Chen H, et al. Modular construction of mammalian gene circuits using TALE transcriptional repressors. *Nat Chem Biol*, 2015, 11: 207-13
- [48] Lapique N, Benenson Y. Digital switching in a biosensor circuit via programmable timing of gene availability. *Nat Chem Biol*, 2014, 10: 1020
- [49] Kashida S, Inoue T, Saito H. Three-dimensionally designed protein-responsive RNA devices for cell signaling regulation. *Nucleic Acids Res*, 2012, 40: 9369-78
- [50] Kawasaki S, Fujita Y, Nagaike T, et al. Synthetic mRNA devices that detect endogenous proteins and distinguish mammalian cells. *Nucleic Acids Res*, 2017, 45: e117
- [51] Yen L, Svendsen J, Lee JS, et al. Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. *Nature*, 2004, 431: 471-6
- [52] Win MN, Smolke CD. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc Natl Acad Sci USA*, 2007, 104: 14283-8
- [53] Kennedy AB, Vowles JV, d'Espaux L, et al. Protein-responsive ribozyme switches in eukaryotic cells. *Nucleic Acids Res*, 2014, 42: 12306-21
- [54] Siciliano V, DiAndreth B, Monel B, et al. Engineering modular intracellular protein sensor-actuator devices. *Nat Commun*, 2018, 9: 1881
- [55] Liu YC, Zhan YH, Chen ZC, et al. Directing cellular information flow via CRISPR signal conductors. *Nat Methods*, 2016, 13: 938
- [56] Davila ML, Riviere I, Wang XY, et al. Efficacy and toxicity management of 19-28z CAR T cell therapy in B cell acute lymphoblastic leukemia. *Sci Transl Med*, 2014, 6: 224ra25
- [57] Sadelain M, Riviere I, Brentjens R. Targeting tumours with genetically enhanced T lymphocytes. *Nat Rev Cancer*, 2003, 3: 35-45
- [58] Savoldo B, Ramos CA, Liu EL, et al. CD28 costimulation improves expansion and persistence of chimeric antigen receptor-modified T cells in lymphoma patients. *J Clin Invest*, 2011, 121: 1822-6
- [59] Song DG, Ye QR, Carpenito C, et al. *In vivo* persistence, tumor localization, and antitumor activity of CAR-engineered T cells is enhanced by costimulatory signaling

- through CD137 (4-1BB). *Cancer Res*, 2011, 71: 4617-27
- [60] Gandhi M, Jones K. Optimizing tumor-targeting chimeric antigen receptor T cells in B-cell lymphoma patients. *Immunotherapy*, 2011, 3: 1441-3
- [61] Chakravarti D, Cho JH, Weinberg BH, et al. Synthetic biology approaches in cancer immunotherapy, genetic network engineering, and genome editing. *Integr Biol (Camb)*, 2016, 8: 504-17
- [62] Zah E, Lin MY, Silva-Benedict A, et al. T cells expressing CD19/CD20 bispecific chimeric antigen receptors prevent antigen escape by malignant B cells. *Cancer Immunol Res*, 2016, 4: 498-508
- [63] Grada Z, Hegde M, Byrd T, et al. TanCAR: a novel bispecific chimeric antigen receptor for cancer immunotherapy. *Mol Ther Nucleic Acids*, 2013, 2: e105
- [64] Hegde M, Mukherjee M, Grada Z, et al. Tandem CAR T cells targeting HER2 and IL13Ralpha2 mitigate tumor antigen escape. *J Clin Invest*, 2016, 126: 3036-52
- [65] Bray SJ. Notch signalling: a simple pathway becomes complex. *Nat Rev Mol Cell Biol*, 2006, 7: 678-89
- [66] Morsut L, Roybal KT, Xiong X, et al. Engineering customized cell sensing and response behaviors using synthetic notch receptors. *Cell*, 2016, 164: 780-91
- [67] Roybal KT, Rupp LJ, Morsut L, et al. Precision tumor recognition by T cells with combinatorial antigen-sensing circuits. *Cell*, 2016, 164: 770-9
- [68] Xiong J, Sun WJ, Wang WF, et al. Novel, chimeric, cancer-specific, and radiation-inducible gene promoters for suicide gene therapy of cancer. *Cancer*, 2012, 118: 536-48
- [69] Viale DL, Cafferata EG, Gould D, et al. Therapeutic improvement of a stroma-targeted CRAd by incorporating motives responsive to the melanoma microenvironment. *J Invest Dermatol*, 2013, 133: 2576-84
- [70] Hogg RT, Garcia JA, Gerard RD. Adenoviral targeting of gene expression to tumors. *Cancer Gene Ther*, 2010, 17: 375-86
- [71] Cafferata EG, Lopez MV, Nunez FJ, et al. Application of synthetic tumor-specific promoters responsive to the tumor microenvironment. *Methods Mol Biol*, 2017, 1651: 213-27
- [72] Sharon E, Kalma Y, Sharp A, et al. Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat Biotechnol*, 2012, 30: 521-30
- [73] Weingarten-Gabbay S, Segal E. The grammar of transcriptional regulation. *Hum Genet*, 2014, 133: 701-11
- [74] Roberts ML, Katsoupi P, Tseveleki V, et al. Bioinformatically informed design of synthetic mammalian promoters. *Methods Mol Biol*, 2017, 1651: 93-112
- [75] Nguyen TA, Jones RD, Snavely AR, et al. High-throughput functional comparison of promoter and enhancer activities. *Genome Res*, 2016, 26: 1023-33
- [76] Angelici B, Mailand E, Haefliger B, et al. Synthetic biology platform for sensing and integrating endogenous transcriptional inputs in mammalian cells. *Cell Rep*, 2016, 16: 2525-37
- [77] Smith RP, Taher L, Patwardhan RP, et al. Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nat Genet*, 2013, 45: 1021-8
- [78] Swanson CI, Evans NC, Barolo S. Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. *Dev Cell*, 2010, 18: 359-70
- [79] Merika M, Thanos D. Enhanceosomes. *Curr Opin Genet Dev*, 2001, 11: 205-8
- [80] Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet*, 2012, 13: 613-26
- [81] Brown AJ, Gibson SJ, Hatton D, et al. In silico design of context-responsive mammalian promoters with user-defined functionality. *Nucleic Acids Res*, 2017, 45: 10906-19
- [82] Smith RP, Riesenfeld SJ, Holloway AK, et al. A compact, *in vivo* screen of all 6-mers reveals drivers of tissue-specific expression and guides synthetic regulatory element design. *Genome Biol*, 2013, 14: R72
- [83] Yuan Y, Guo L, Shen L, et al. Predicting gene expression from sequence: a reexamination. *PLoS Comput Biol*, 2007, 3: 2391-7
- [84] Irie T, Park SJ, Yamashita R, et al. Predicting promoter activities of primary human DNA sequences. *Nucleic Acids Res*, 2011, 39: e75
- [85] Curran KA, Crook NC, Karim AS, et al. Design of synthetic yeast promoters via tuning of nucleosome architecture. *Nat Commun*, 2014, 5: 2002
- [86] Killoran N, Lee LJ, DeLong A, et al. Generating and designing DNA with deep generative models. *ArXiv*, 2017, 1712.06148
- [87] Mircetic J, Dietrich A, Paszkowski-Rogacz M, et al. Development of a genetic sensor that eliminates p53 deficient cells. *Nat Commun*, 2017, 8:1463
- [88] Raj A, van Oudenaarden A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 2008, 135: 216-26
- [89] Green AA, Silver PA, Collins JJ, et al. Toehold switches: de-novo-designed regulators of gene expression. *Cell*, 2014, 159: 925-39
- [90] Morel M, Shtrahman R, Rotter V, et al. Cellular heterogeneity mediates inherent sensitivity-specificity tradeoff in cancer targeting by synthetic circuits. *Proc Natl Acad Sci USA*, 2016, 113: 8133-8
- [91] Bowsler CG, Swain PS. Environmental sensing, information transfer, and cellular decision-making. *Curr Opin Biotechnol*, 2014, 28: 149-55
- [92] Mohammadi P, Beerenwinkel N, Benenson Y. Automated design of synthetic cell classifier circuits using a two-step optimization strategy. *Cell Syst*, 2017, 4: 207-18. e14
- [93] Stanton BC, Nielsen AA, Tamsir A, et al. Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat Chem Biol*, 2014, 10: 99-105
- [94] Rhodius VA, Segall-Shapiro TH, Sharon BD, et al. Design of orthogonal genetic switches based on a crosstalk map

- of sigmas, anti-sigmas, and promoters. *Mol Syst Biol*, 2013, 9: 702
- [95] Del Vecchio D, Ninfa AJ, Sontag ED. Modular cell biology: retroactivity and insulation. *Mol Syst Biol*, 2008, 4: 161
- [96] Cheong R, Rhee A, Wang CJ, et al. Information transduction capacity of noisy biochemical signaling networks. *Science*, 2011, 334: 354-8
- [97] Uda S, Saito TH, Kudo T, et al. Robustness and compensation of information transmission of signaling pathways. *Science*, 2013, 341: 558-61
- [98] Zechner C, Seelig G, Rullan M, et al. Molecular circuits for dynamic noise filtering. *Proc Natl Acad Sci USA*, 2016, 113: 4729-34