



功能性增强子: 概念、筛选与验证

林虹利^{1†}, 陈斐^{1†}, 黄佳良^{1,2*}

1. 厦门大学生命科学学院, 细胞应激生物学国家重点实验室, 厦门 361102

2. 厦门大学健康医疗大数据国家研究院, 厦门 361102

† 同等贡献

* 联系人, E-mail: jhuang@xmu.edu.cn

收稿日期: 2025-04-19; 接受日期: 2025-05-14; 网络版发表日期: 2025-06-06

摘要 增强子作为基因组中的核心调控元件, 通过精确调控基因表达在生物发育和疾病进程中发挥关键作用. 基于DNA序列和表观遗传特征, 研究人员已鉴定出数百万个潜在增强子, 但其功能解析和机制研究仍然面临重大挑战. 近年来, 单细胞多组学、染色质构象捕获、空间组学和基因编辑等技术的突破性进展, 为深入解析增强子功能提供强有力的工具. 本文详细阐述功能性增强子概念的核心特征, 包括其对基因表达的调控、对细胞/组织及个体表型的影响, 以及在疾病发生发展中的关键作用; 接着系统梳理高通量计算和实验技术的最新进展, 这些技术显著地提升研究者对发育和疾病相关增强子调控网络的解析能力; 进而提出一套功能性增强子的完整筛选和验证策略, 并对该领域的未来发展方向进行展望.

关键词 增强子, 功能性增强子, 表观遗传, 染色质, 基因表达

人类基因组中仅约2%的序列是负责蛋白质编码的基因, 对于其余大量非编码序列的生物学功能, 人们仍然了解甚少. 其中, 增强子作为基因组中的顺式作用调控元件, 通过与转录因子结合, 以时空特异性方式调节目标基因的转录水平, 在发育和疾病发生过程中扮演重要角色^[1,2]. 随着“DNA元件百科全书”(Encyclopedia of DNA Elements, ENCODE)计划的完成, 研究人员已在人类基因组中成功鉴定出上百万个增强子, 然而绝大多数增强子的分子机制及其在病理生理方面的功能仍然未知^[3]. 研究表明, 一些增强子的异常(例如突变、失活或扩增等)能够驱动细胞、组织或个体水平的表型变化, 从而影响细胞命运决定或疾病发生发展, 研究者将这类增强子称为功能性增强子. 例如, 原癌基

因MYC的增强子异常与多种癌症的发生发展密切相关^[4], BCL11A基因的增强子在地中海贫血和镰刀型细胞贫血的治疗中展现出重要的临床应用价值^[5-8].

然而, 如何从浩瀚的基因组中精确鉴定出功能性增强子仍然面临诸多挑战, 特别是绝大部分增强子扰动(包括敲除、抑制或激活等)没有明显的可观测表型, 这与传统基因扰动存在显著差别. 首先, 增强子对基因表达的调控具有时空特异性, 其功能活性随细胞类型、发育阶段或病理生理状态呈现显著异质性, 使功能性增强子的准确识别变得更加困难. 其次, 许多增强子在基因组中以簇的形式存在, 被称为超级增强子^[9,10]或增强子簇^[4]. 多个增强子通过加和^[11]、协同^[12,13]、冗余^[14]、层次^[15]、嵌套^[16]或网络结构^[17]等不同机制,

引用格式: 林虹利, 陈斐, 黄佳良. 功能性增强子: 概念、筛选与验证. 中国科学: 生命科学

Lin H L, Chen F, Huang J L. Functional enhancers: concept, screening, and validation (in Chinese). Sci Sin Vitae, doi: [10.1360/SSV-2025-0113](https://doi.org/10.1360/SSV-2025-0113)

协调控制同一个靶基因的表达, 可以缓冲基因组中非编码调控区突变而导致的基因表达紊乱。最后, 尽管CRISPR基因编辑技术为增强子功能研究提供强大工具, 但全基因组范围的筛选仍然受限于高昂的实验成本^[18]。由于增强子调控机制的内在复杂性和相关技术瓶颈的存在, 目前仅有极少量增强子具有明确的生物学功能证据。

为更高效地发现和验证功能性增强子, 整合多种研究策略显得尤为关键。本综述首先阐述功能性增强子的基本概念, 并系统梳理该领域的核心分析与实验方法。在此基础上, 本文提出一套功能性增强子的筛选和验证策略, 涵盖从组学图谱绘制、计算筛选、实验筛选、功能验证到临床应用转化, 为相关研究工作提供参考和借鉴。

1 功能性增强子

1.1 功能性增强子概述

增强子作为非编码区域中的核心调控元件, 可独立于其与靶基因的距离和位置, 通过招募转录因子和辅因子, 显著增强靶基因的转录活性。增强子通常具有特定的表观遗传学特征, 如开放的染色质结构、特定的组蛋白修饰(如H3K27ac和H3K4me1)以及转录因子结合位点的富集^[19,20]。随着高通量测序技术及相关衍生技术的发展, 以及ENCODE^[3]和FANTOM^[21]等

国际计划的推进, 基于全基因组染色质可及性、组蛋白修饰、转录因子结合及染色质互作等图谱, 研究者已在人类基因组中鉴定出上百万个潜在的增强子序列。目前, 已有研究利用不同功能分析技术, 揭示部分功能性增强子的明确生物学功能, 即增强子的异常(例如突变、失活或扩增等)能够驱动细胞、组织或个体水平的表型变化, 从而影响细胞命运决定或疾病发生发展(表1)。

在细胞或组织水平, 功能性增强子在细胞命运决定和组织发育中发挥关键作用。例如, 基因座控制区(locus control region, LCR)作为经典的功能性增强子, 通过调控 β -珠蛋白基因簇的时空特异性表达, 控制红细胞分化进程^[22]。*MyoD*基因的肌肉特异性增强子在肌肉发育和再生过程中发挥重要作用^[23,24]。*TBX5*基因的心脏特异性增强子则通过精确调控其在心肌细胞中的表达, 参与心脏发育和功能维持^[25]。*ETV4*基因的增强子能促进黑色素瘤细胞的增殖和迁移^[26]。

在个体层面, 功能性增强子在发育和疾病中发挥关键作用。*Shh*基因的ZRS增强子精确调控四肢发育, 其点突变可导致先天性多指^[27,28]。*Sox9*基因的性别决定增强子在胚胎发育过程中通过多个增强子簇协同作用, 其缺失会引发性反转综合征^[29]。在肿瘤发生过程中, *MYC*基因的增强子异常激活, 驱动*MYC*致癌基因的持续高表达^[4]。而*BCL11A*基因的红系增强子则成为血液系统疾病治疗的重要靶点, 通过CRISPR技术破坏

表 1 代表性经典功能性增强子

Table 1 Representative typical functional enhancers

增强子	功能
SV40病毒增强子	增强兔 β -珠蛋白基因的表达 ^[1]
β -珠蛋白基因座控制区(LCR)	调控红细胞分化和发育 ^[22]
<i>MyoD</i> 肌肉特异性增强子	参与肌肉发育和分化 ^[23]
<i>Oct-4</i> 远端增强子	维持生殖细胞中的全能性或多能性 ^[30]
<i>TCF7L2</i> 增强子	调控异常增加2型糖尿病的风险 ^[31]
<i>TBX5</i> 增强子	该增强子上的突变导致心脏发育异常 ^[25]
<i>Six1</i> 增强子	调控异常导致相关组织的发育异常, 从而引发特定的疾病表型(如听力损失、肾脏缺失) ^[32]
<i>IRF4</i> 增强子	调控异常影响黑色素的合成 ^[33]
<i>BCL11A</i> 增强子	调控红细胞发育 ^[7]
<i>Nanog</i> 增强子	维持胚胎干细胞多能性和自我更新能力 ^[34]
<i>Shh</i> 肢体增强子(ZRS)	影响脊椎动物的肢体和手指发育 ^[27]
<i>Sox9</i> 增强子	缺失引发性反转综合征 ^[29]
<i>MYC</i> 血液增强子簇(BENC)	调控异常导致白血病 ^[4]
<i>AR</i> 增强子	在去势抵抗性前列腺癌(CRPC)中放大雄激素受体信号, 促进肿瘤耐药性 ^[35]

该增强子的调控功能, 可重新激活胎儿血红蛋白(HbF), 为镰状细胞贫血和 β -地中海贫血的基因治疗提供新的治疗策略^[5,6].

这些案例表明, 功能性增强子在发育调控和疾病机制中具有重要的作用. 因此, 系统筛选并探究功能性增强子的作用机制, 不仅有助于阐明基因调控的原理, 还有望为精准医疗提供新的治疗靶点.

1.2 疾病相关功能性增强子的作用机制

以疾病中的功能性增强子为例, 其序列中可能包含疾病相关的遗传变异, 导致增强子功能异常; 此外, 还可能通过体细胞结构变异(structure variation, SV)和染色体拷贝数变异(copy number variation, CNV)等方式促使癌基因的异常高表达, 进而参与癌症的发生发展. 近年来, 已有一系列研究证明, 存在不同的调控机制将这些增强子相关的遗传变异与疾病表型联系起来(图1).

(i) 遗传变异改变转录因子的结合状态(图1A). 增强子上的遗传变异可能会破坏转录因子在其上的结合位点序列, 使转录因子无法结合, 因而失去对靶基因的转录激活效应^[36]. 基因组上的突变可能产生新的转录因子结合位点, 进而出现异常的增强子活性, 导致与疾病或发育相关的基因异常激活^[37,38]. 此外, 位于转录因子结合位点的变异也可能导致转录因子结合的亲和力增加, 继而使增强子在非特异性的细胞类型中获得异常活性, 导致基因在非目的组织中的表达^[39,40].

(ii) 遗传变异改变染色质相互作用(图1B). 增强子通过不同的机制建立与靶基因之间的相互作用, 例如中介蛋白复合物、YY1、增强子RNA (enhancer RNA, eRNA), 以及CTCF和黏连蛋白介导的“环挤出”模型^[41,42]. 大多数的增强子-启动子连接位于相同的拓扑结构域(topologically associating domains, TAD)内, TAD边界使TAD内部相互作用在一定程度上绝缘于邻近区域, 并阻止异常的增强子-启动子激活. 位于TAD

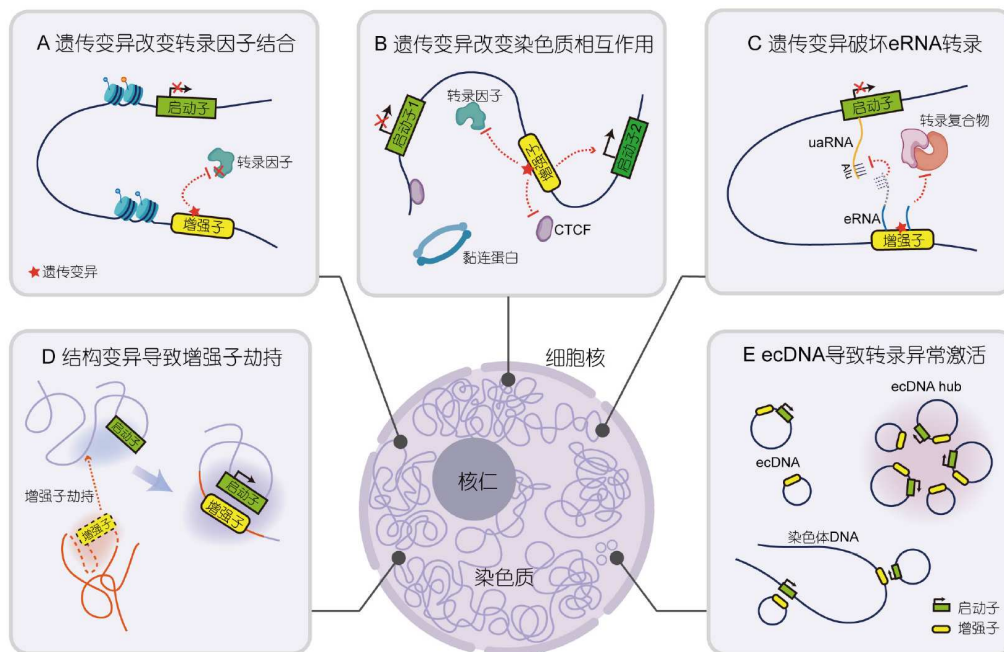


图 1 增强子相关的遗传变异导致疾病的机制. A: 增强子上的遗传变异导致转录因子结合异常, 引发靶基因表达失调; B: 增强子上的遗传变异改变染色质三维结构, 导致原本的靶基因失调, 还可能异常激活其他靶基因; C: 增强子上的遗传变异破坏eRNA的转录, 使其不能维持染色质环, 导致靶基因失调; D: 染色质结构变异导致癌基因劫持异位增强子, 并被其异常激活; E: ecDNA通过形成聚集的中心, 或作为可移动增强子被癌基因劫持, 导致癌基因异常高表达

Figure 1 Mechanisms by which enhancer-associated genetic variation leads to diseases. A: Genetic variations in enhancers lead to abnormal binding of transcription factors, leading to the dysregulation of target genes; B: genetic variations in enhancers alter the three-dimensional chromatin structure, leading to the dysregulation of the original target genes and potentially abnormally activating other target genes; C: genetic variations in enhancers disrupt the transcription of eRNAs, preventing them from maintaining chromatin loops and resulting in the dysregulation of target genes; D: structure variations lead to the hijacking of ectopic enhancers by oncogenes, resulting in their abnormal activation; E: ecDNAs lead to the abnormally high expression of oncogenes by forming ecDNA hubs or acting as mobile enhancers that are hijacked by oncogenes

边界的CTCF结合位点的变异可能导致TAD边界的消失^[43], 导致增强子或超级增强子与其他基因建立连接, 致使其表达异常升高^[44,45]. 此外, 增强子上的非编码变异还可以通过促进增强子与启动子间的物理相互作用, 使癌基因表达上调^[46].

(iii) 遗传变异破坏eRNA的转录(图1C). eRNA可以通过与蛋白复合体相互作用或者募集染色质结构相关蛋白, 从而维持^[47]或促进形成^[48]染色质环, 以上调靶基因的转录, 这是在某些癌症中癌基因受到异常驱动的原因. 此外, 增强子也可通过eRNA与启动子上游反义RNA (uaRNA)的相互作用促成增强子与靶基因之间的连接, 而这很大一部分是由于连接的序列上富集互补的Alu序列. 位于这些Alu序列上的突变破坏增强子-启动子RNA相互作用, 使靶基因失调而促进肿瘤发生^[49]. eRNA转录本是否完整也可能是使基因失调进而导致疾病的因素, 例如, 原本eRNA在转录过程中会折叠形成RNA环状结构而暂停转录, 此时该增强子仅部分被转录, 是一个无功能增强子, 但在特定细胞类型或者应激状态下, 完整的eRNA转录本被转录出来, 调控靶基因异常表达^[50,51].

(iv) 基因组上的结构变异(SV)导致增强子劫持事件(图1D). 结构变异例如染色体缺失、倒位、易位、融合等, 使癌细胞基因组发生重组, 导致原本位于远端甚至不同染色体上的增强子被劫持到癌基因附近, 形成新的染色质3D结构域或产生新的增强子-启动子连接, 从而异位激活癌基因表达^[52,53]. 此外, 结构变异使包含增强子的染色体片段发生局部重复, 增强子片段得以大量扩增, 导致基因表达被异常上调, 驱动癌症的发生和进展^[54].

(v) 染色体外DNA (extrachromosomal DNA, ecDNA)激活癌症相关基因的高表达(图1E). ecDNA是体细胞拷贝数扩增的最重要形式之一, 已在大多数人类癌症类型中被发现^[55]. 癌基因和其非编码调控元件被并置于ecDNA上, 凭借ecDNA的高染色质可及性和高扩增速率的特性, 驱动癌基因的异常高表达^[56,57]. 此外, 增强子劫持事件也常发生在ecDNA中: 一些缺乏关键增强子的ecDNA会劫持位于远端染色体上的增强子元件, 以增强自身转录^[58]; 富含增强子或超级增强子的ecDNA能够作为可移动增强子, 驻留在活性染色质区室内部并与多个染色体上的基因发生联系, 导致异位转录激活^[59]. 在一些癌症类型中, 多个ecDNA

可以与大量的转录因子复合物、RNA聚合酶等共同形成分子凝聚物, 该枢纽(ecDNA hub)以类似相分离的方式, 通过高度共享转录调控元件, 使ecDNA上的癌基因保持更高的转录水平^[59,60].

2 计算筛选

2.1 数据资源

近年来, 高通量测序技术的革新与成本的降低推动组学数据的爆发式增长. 目前, 研究者已积累大量复杂生命系统数据资源, 具有多层次、多尺度、多模态和高维度等特征. 主要包括: (i) 单细胞表观基因组和多组学图谱, 包括人类和小鼠胚胎及成体的正常组织^[61~67], 以及各种疾病的组织^[68,69]等; (ii) 染色质状态图谱, 包括人和小鼠的不同组织和细胞类型的染色质可及性、组蛋白修饰及转录因子结合位点信息等^[3,70~73]; (iii) 三维基因组图谱, 如人类大脑不同发育时期(妊娠中/晚期、婴儿期及成年期)^[74]、成年小鼠全脑^[75]等; (iv) 遗传变异信息, 如全基因组关联研究(genome-wide association studies, GWAS)发现的常见的疾病风险位点^[76,77], 基因表达数量性状位点(expression quantitative trait loci, eQTLs), 以及近期研究报道的eRNA关联变异位点(eRNA-QTLs)^[78]等. 这些全面的组学数据资源为深入研究增强子功能提供坚实的基础.

2.2 相关算法

上述丰富的数据资源推动生物信息学算法的创新, 为功能性增强子的高效筛选带来新的机遇(图2). 这些算法主要可分为以下几类.

2.2.1 时空特异性增强子的筛选

增强子的功能具有高度时空动态性, 其激活状态受细胞类型、发育阶段及空间位置的精确调控.

鉴定细胞类型特异性增强子可解析细胞身份建立的分子调控机制. ABC (activity-by-contact)模型^[79]通过整合细胞类型特异的表观遗传活性信号(如H3K27ac)和染色质三维互作数据(如Hi-C), 计算加权评分, 从而精准量化细胞类型特异的增强子-基因调控关系. SCENT^[80]和SCARlink^[81]算法分别采用泊松回归和正则化的泊松回归方法对单细胞多组学数据进行联

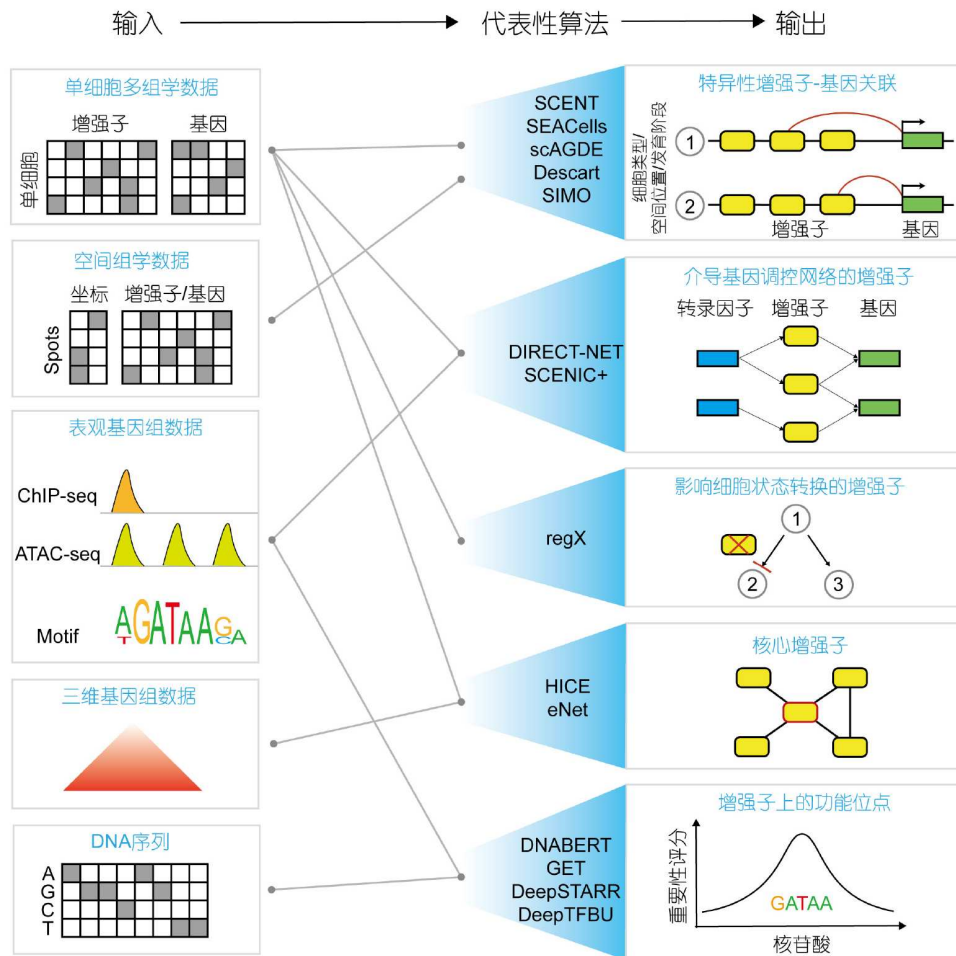


图 2 预测功能性增强子的代表性算法概览. 该图展示不同算法(中间)的核心输入数据类型(左侧)及其输出结果形式(右侧)
Figure 2 Overview of representative algorithms for predicting functional enhancers. This diagram illustrates the main input data types (left) and their output formats (right) for different algorithms (middle)

合建模, 成功实现细胞类型特异性增强子-靶基因关联的精准解析. 它们预测的增强子与GWAS和eQTL变异位点显著重叠, 证实其预测的可靠性. 针对单细胞数据 (尤其是scATAC-seq) 的固有稀疏性挑战, 研究人员开发多种解决方案. SEACells算法通过“元细胞”(meta-cell)构建策略, 在保留细胞异质性的同时有效提升信号信噪比, 基于该策略优化的跨组学分析成功揭示造血发育过程中关键增强子及其靶基因^[82]. 另一种解决方案则基于深度学习模型, 例如, scAGDE^[83]和scBas-set^[84]分别利用染色质可及性自编码器和图卷积网络, 以及基于深度卷积神经网络DNA序列模型, 有效克服scATAC-seq数据的高维稀疏性, 从而实现细胞类型特异的功能性增强子的高效识别. 与此同时, 基于调控

网络的功能性增强子筛选方法正在兴起, 其突破传统转录因子-基因调控网络的维度, 通过引入scATAC-seq数据包含的转录因子结合位点信息, 结合转录因子基序(motif)数据, 构建增强子驱动的基因调控网络(enhancer-driven gene regulatory network, eGRN). 其中, DIRECT-NET^[85], SCENIC+^[86]和SEanalysis 2.0^[87]等代表性算法已经展现出基于eGRN筛选细胞类型特异的功能性增强子的强大潜力.

在时间维度上, Price团队^[88]通过整合28种脑部疾病/性状的GWAS数据与人类大脑不同发育阶段的scATAC-seq和scRNA-seq数据, 成功识别出与脑部疾病显著相关的胎儿和成人脑细胞类型及其相关的功能性增强子, 发现ADHD遗传风险位点特异性富集于胎儿

星形胶质细胞中的增强子。Raychaudhuri团队^[89]则通过对类风湿性关节炎患者滑膜组织样本的单细胞多组学与GWAS数据联合分析,揭示了随免疫细胞状态动态变化的染色质开放区域(而非稳定区域)是自身免疫疾病遗传易感性的分子基础,并精确定位T细胞及髓系细胞的特定状态为遗传风险载体。此外,regX算法通过将基因上游调控和基因-基因互作整合到深度学习模型中,能够鉴定疾病和发育过程中影响细胞状态转换的关键转录因子和顺式调控元件,并在2型糖尿病中成功识别出与胰岛素分泌和 β 细胞凋亡等过程相关的转录因子和调控元件^[90]。

在空间维度上,得益于空间表观基因组学技术^[91-97]的快速发展,研究者逐渐认识到增强子活性还与细胞在组织中的空间定位密切相关,这为理解组织形态建成提供新的视角。然而,空间表观基因组学在功能性增强子筛选中的应用仍面临数据稀疏性的瓶颈。针对这一挑战,Descart算法通过整合空间定位信息与染色质可及性数据构建细胞间相关性图谱,实现数据去噪,成功捕捉到具有空间位置依赖性及细胞类型特异性的功能性增强子^[98]。值得注意的是,相较于快速积累的空间转录组数据,当前空间表观基因组学数据仍处于稀缺状态,而单细胞表观组学数据(如scATAC-seq)已形成规模化资源库。这一数据格局催生新型计算策略的开发。以SIMO算法为例,其通过深度学习框架整合空间转录组与单细胞多组学数据,实现空间特异性增强子的精准预测,有效突破数据稀缺与分辨率不足的双重限制^[99]。

2.2.2 进化保守性增强子的筛选

基因组调控元件的进化轨迹为功能注释提供天然筛选标记。在哺乳动物进化历程中,关键调控元件往往表现出高度的序列保守性。因此,从进化的视角来筛选功能性增强子,能够为传统的研究方法提供有力的补充。

Zoonomia计划近期的研究为此提供有力证据:通过对239种哺乳动物基因组的跨时间尺度比较分析,研究者以单碱基分辨率成功识别出在数百万年演化进程中保守度较高的基因组区域。研究结果表明,这种基于进化约束的分析方法能显著提升识别功能重要变异位点的准确性,这一优势在非编码区域尤为突出^[100,101]。然而,增强子的功能保守并不完全依赖于单

个碱基的保守性。有研究发现,即使个别碱基在进化过程中表现出一定程度的可变性,与增强子活性相关的序列模式在哺乳动物中仍然保持着高度的保守性^[102]。基于此,Pfenning团队^[103]开发一个计算框架TACIT,通过定量评估增强子活性保守性与特定表型特征之间的关联,精准筛选出与多个进化支系表型演化密切相关的候选功能性增强子。

除了在整个哺乳动物进化过程中保守的区域,科学家们还发现一类展现出独特进化模式的人类加速进化区(human accelerated regions, HARs):在大多数哺乳动物中高度保守,但在人类与黑猩猩分化后经历快速的核苷酸替换,并在神经发育相关基因附近的非编码区域中显著富集^[104]。这启发Shen团队^[105]深入探究HARs在基因调控和人类大脑发育中的关键作用。利用比较基因组分析,结合单细胞CRISPRi筛选和先导编辑等多种实验手段,他们成功筛选出多个功能性HARs,证实这些区域作为转录增强子特异性调控着人类大脑发育和功能相关基因的表达。

2.2.3 基于增强子互作筛选核心增强子

随着高通量测序技术的发展,基于增强子-增强子互作关系的核心增强子筛选策略逐渐成熟。早期研究主要依赖染色质构象捕获技术(如Hi-C, ChIA-PET等)产生的数据,通过构建增强子调控网络并计算网络中各节点的中心性指标,成功识别具有关键调控作用的核心增强子^[15,106]。这些研究为理解增强子间的调控机制奠定了重要基础。然而,传统Hi-C技术在分辨率方面存在一定局限性,难以精确捕获更细微的染色质互作关系。为克服这一技术瓶颈,研究者开发一系列基于单细胞测序数据的创新算法。其中,Cicero等算法利用scATAC-seq数据预测增强子-增强子互作关系,显著提高分析的分辨率和精确度^[107,108]。在此基础上,eNet算法通过整合单细胞多组学数据构建增强子网络,进一步提升核心增强子的识别效率^[17,109,110]。这些方法不仅揭示增强子之间复杂的协同调控关系,也为深入理解基因表达调控的层级性机制提供崭新视角。

2.2.4 增强子功能位点的精准定位

深度学习算法在功能基因组学领域取得重大突破,通过深度挖掘表观遗传特征、DNA序列信息以及三维染色质结构等多维数据中蕴含的潜在调控规律,

能够精确定位增强子的关键功能区域, 为后续实验筛选提供更精准的指导. 例如, 以DNABERT^[111]和GET^[112]等为代表的方法, 基于“预训练-微调”范式, 能够以细胞类型特异性的方式, 在单核苷酸分辨率水平预测转录因子结合位点和遗传变异效应. DeepSTARR基于UMI-STARR-seq数据, 通过深度学习揭示增强子功能的内在规律: 增强子中心区域的基序对活性贡献最大, 而同一基序在不同位置的功能效果会有差异, 这种差异主要受侧翼序列和相邻基序间距的影响^[113]. DeepTFBU则进一步聚焦于量化转录因子结合位点及其侧翼序列对增强子活性的贡献程度, 并能预测特定突变如何影响增强子功能^[114]. 这些技术进展将传统增强子预测的分辨率从数百碱基对显著提升至不足10个碱基对的功能区域, 显著提升功能元件筛选的精确度.

3 实验筛选

高通量技术的高速发展, 为增强子的大规模功能性筛选提供众多可用的工具. 通过对假定的增强子元件进行各类功能实验, 可以有效地筛选和鉴定出在发育或疾病等各种生物学过程中起到重要作用的功能性增强子.

3.1 高通量报告基因分析

增强子序列中包含的特定DNA基序, 通过与转录因子结合来实现基因表达调控的功能^[115]. 因此, 使用基于报告基因的检测方法可以测定假定增强子序列对异源报告基因表达的驱动能力, 达到直接检测增强子活性的目的(图3A).

MPRA和STARR-seq等高通量大规模平行报告分析方法使得在全基因组范围内直接定量增强子活性成为可能^[116-119]. 在以增强子为重点的众多报告实验设计中, 方法上的主要差异在于增强子与报告基因的相对位置, 以及报告体系是否整合在基因组中. 具体来说, MPRA利用微阵列合成增强子序列, 并使用条形码表征每个增强子, 因此通常用于较短序列(<200 bp)的分析; STARR-seq^[120]依据增强子不依赖方向性和相对位置发挥作用的特征, 将增强子序列置于启动子下游, 利用增强子自身转录本的丰度直接表征增强子的活性, 并且不再受限于条形码的数量, 能够大幅度地提高检测的通量. 此外, 上述检测方法均依赖于瞬时转染的

文库递送方式, 在应用方面受限于细胞类型, 而基于慢病毒感染的lentiMPRA技术^[121,122]不仅在重复性和适用性方面具有更大的优势, 还充分考虑基因组内的环境, 因此能更准确地预测调控元件的功能. 此外, AR-BITER系统^[123]实现基于腺相关病毒(adeno-associated virus, AAV)的体内高通量报告实验, 有效筛选出小鼠耳蜗基因的调控增强子, 为时空特异性的功能性增强子筛选提供强有力的工具.

基于报告基因的检测方法可以一定程度上反映增强子的活性, 因为合成的增强子序列具有一定的增强子功能, 例如包含转录因子结合位点或者具有可以被转录的特性^[124], 并且近期开发的scMPRA^[125]可以在单细胞水平实现报告基因和转录组的同时检测, 可进一步拓展到在组织中进行细胞类型特异性调控元件的活性检测. 但是, 这些方法通常只关注于序列本身的特征, 而忽略细胞内源环境下的其他特征对增强子活性的影响, 例如组蛋白修饰和染色三维结构等^[126]. 同时, 由于合成序列的长度限制, 通常只能探究增强子的部分核心序列, 而忽略侧翼序列对增强子活性的影响^[113,114].

3.2 基于CRISPR的高通量筛选

基于CRISPR的基因编辑技术是功能基因组学研究中不可或缺的工具^[127,128]. 利用Cas蛋白及其一系列变体, 对增强子位点进行敲除、抑制、激活等扰动, 从而实现对增强子在其内源环境中生物学功能的探究.

基于CRISPR基因编辑的高通量筛选技术已经被广泛应用于大规模的功能性增强子筛选中(图2B). 增强子不同于编码基因, Cas9切割造成的少量碱基缺失难以造成其功能异常, 因此, 除了在一些关键位点的研究中以高覆盖度的sgRNA靶向特定区域进行单次切割外^[129,130], 大多数对增强子的功能扰动通常基于CRISPRi/CRISPRa系统实现, 即使用失活的Cas9蛋白(dCas9)链接不同结构域进行表观基因组编辑, 以实现对整个增强子区域的活性抑制或激活^[18,131].

由于CRISPRi系统存在抑制效率有限且不均一、sgRNA靶向区域限制等问题, 目前也有研究使用双CRISPR系统(dual-CRISPR)进行基于增强子敲除的高通量筛选. 具体来说, 通过同时引入两条靶向增强子两端的gRNA (paired gRNAs, pgRNAs), 进行增强子片段

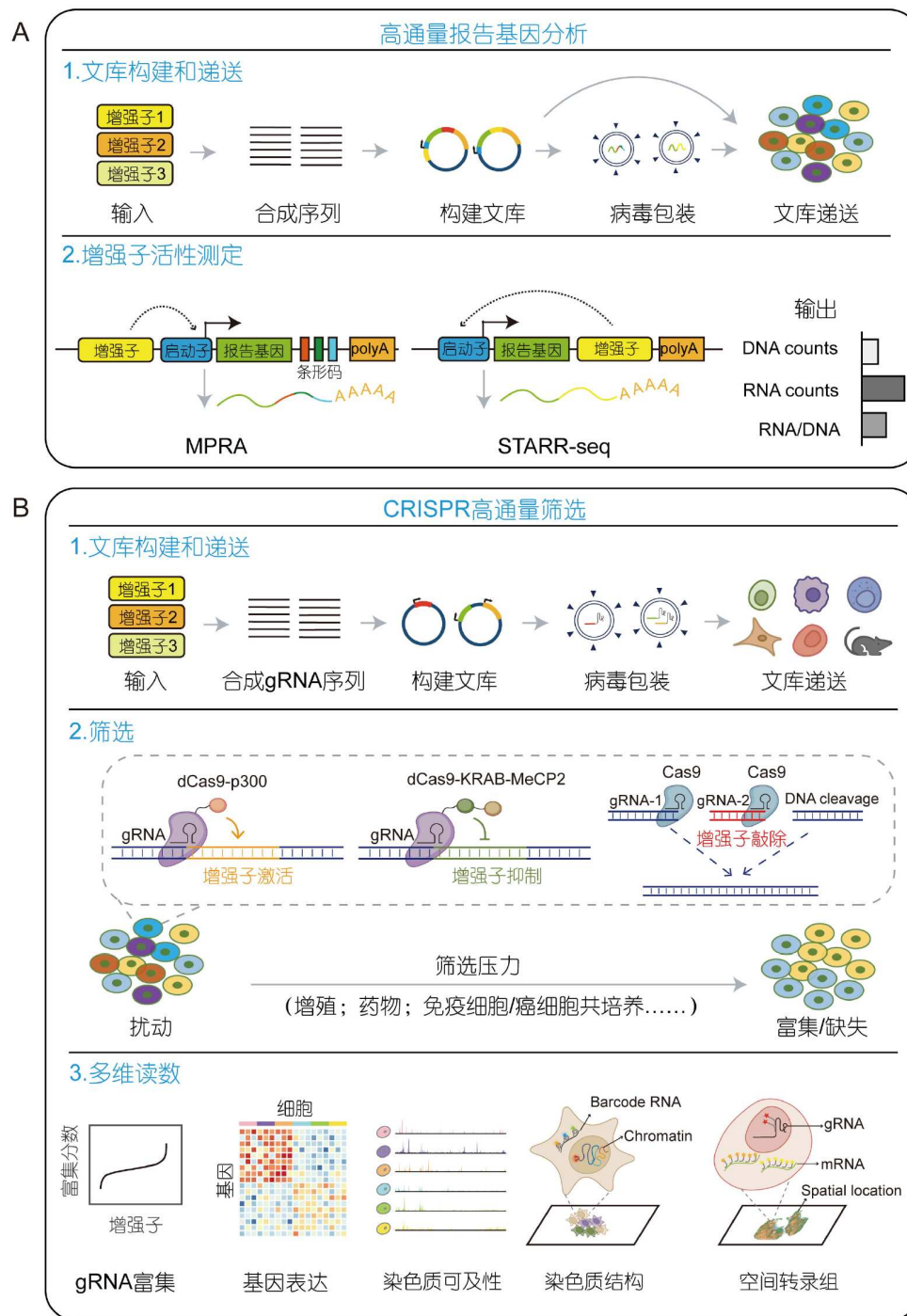


图 3 高通量筛选功能性增强子概览. A: 基于报告基因的高通量检测方法(MPRA和STARR-seq); B: 基于CRISPR的高通量筛选及其多维读数输出

Figure 3 Overview of high-throughput screening for functional enhancers. A: Reporter gene-based high-throughput detection methods (MPRA and STARR-seq); B: CRISPR-based high-throughput screening and multi-omic read-out

的敲除^[132-134]. 这种编辑方式能够充分破坏增强子上的
的转录因子结合位点, 进而大幅提升对于编辑事件的

响应.

体外CRISPR筛选技术在通量、可操作性、实验

成本和时间等方面具有较大的优势,但考虑到增强子的细胞类型特异性,以及自然生理环境的复杂性,在体内进行CRISPR筛选逐渐成为研究增强子在发育和疾病中生物学功能的强有力工具。例如,在研究造血系统发育调控时,通过在造血干细胞中靶向组织特异性或发育调节增强子,移植到小鼠体内提供分化环境,并检测产生的成熟细胞谱系,最终成功筛选出谱系特异性增强子^[135];在研究肢体发育相关的增强子ZRS上的序列变异时,使用构建的体内增强子报告模型,在小鼠受精卵中引入变异,并通过报告基因的表达判断其胚胎肢端的发育情况^[136]。

此外,碱基编辑器^[137]和先导编辑器^[138]等基因编辑技术的开发和广泛使用,为CRISPR高通量筛选展现出更广阔的前景。例如,在研究位于增强子等调控元件内的大量疾病相关遗传变异时,碱基编辑和先导编辑能够更精确、更灵活地引入十几个碱基甚至单碱基分辨率的扰动,直接形成“突变-调控元件-表型”的因果关系,在疾病致病机制研究方面展现出巨大的潜力。

3.3 多维读数的CRISPR筛选技术

传统的CRISPR筛选通常需要特定表型作为筛选的指标,因而只能提供单一或有限的表型读数。单细胞分辨率的多组学测序、空间成像和空间测序等高新技术的发展为探究增强子功能提供更多维、更有效的方案^[139]。

单细胞水平的单一组学和多组学技术已被广泛应用在多种细胞类型的检测中,将CRISPR筛选作为一个额外的维度引入,为深入解析增强子的调控机制提供新的机会(图3B)。目前,已有多种技术能够实现CRISPR筛选与单细胞技术的联合使用,其技术重点主要在于受扰动细胞的身份识别,例如直接捕获sgRNA本身^[140],或者利用标签化^[141]或自转录^[142]的mRNA转本来表征sgRNA。

根据所结合的单细胞技术的不同,以及目标的不同,与增强子筛选相关的单细胞CRISPR筛选技术可以分为以下几种:

(i) 单细胞筛选与基因表达: 增强子的功能一般通过调控其靶基因实现,因而确定增强子的靶基因是研究增强子功能的重点。通过在全基因组范围分析每个扰动对转录组的影响,可以鉴定出增强子的直接和间接靶基因,以及下游的其他转录程序,实现增强子-

基因-表型的关联^[143~145]。更进一步地,CRISPRai Perturb-seq采用组合的双向扰动,在同一细胞中同时对两个位点分别采用激活和抑制扰动,充分研究多个增强子如何相互作用来调节共享靶基因的表达^[146]。

(ii) 单细胞筛选与染色质可及性: 开放的染色质是活性增强子的重要特征。Perturb-ATAC和Spear-ATAC等方法^[147,148]在单细胞水平探究每个扰动导致的染色质变化,识别出影响染色质状态的关键转录因子,将转录因子-染色质状态-细胞表型进行关联。此外,靶向增强子的CRISPR筛选及其染色质状态检测,将有助于解析增强子在不同细胞类型和不同时间点的活性变化,以及增强子之间复杂的调控关系。

(iii) 单细胞筛选与三维基因组学: 增强子通过形成染色质环来发挥远端调控功能,并且受到拓扑结构域(topologically associating domains, TAD)的严格限制^[149]。通过在扰动细胞中检测染色质三维结构,可以高效识别染色质和核拓扑结构的调控因子。Perturb-tracing^[150]利用条形码RNA表征gRNA,并利用基于成像的3D基因组学技术检测单个细胞内不同尺度的染色质结构变化。这将有助于识别疾病发生发展中特异的基因组三维结构变化,为疾病特异性增强子的筛选提供新的表型策略,并为大规模探究多个增强子协同调控的机制提供新的可能。

(iv) 单细胞筛选与空间组学: CRISPR筛选结合单细胞空间转录组技术,在分析细胞内部转录程序的同时,进一步保留复杂组织的空间结构,尤其对探究癌症中肿瘤细胞与免疫细胞的相互作用和微环境调控具有重大意义。Perturb-map^[151]利用蛋白质标签来表征gRNA,结合多重成像与10×空间转录组学技术,实现CRISPR扰动、基因表达和空间位置信息的匹配; Perturb-FISH^[152]采用基于成像的空间转录组学技术,利用荧光原位杂交技术对单个细胞内的gRNA和mRNA进行检测,实现扰动与基因表达及肿瘤-免疫相互作用等空间表型的深度关联。这些技术将为深入探究增强子的时空特异性提供新的策略,使在复杂组织中直观地研究增强子在发育和癌症发生过程的调控作用成为可能。

(v) 单细胞筛选与多组学: 通过对扰动细胞的多组学数据进行联合分析,能够大规模探究增强子在发育或疾病中的作用机制。例如,MultiPerturb-seq^[153]同时检测单细胞转录组与表观组,不仅可以建立单个增

强子-基因-表型的因果关联, 还可以探究多个增强子内部的相互作用, 绘制增强子和基因的复杂调控网络; 此外, ECCITE-seq^[154]同时检测转录组、表面蛋白和克隆型特征, 提供更丰富的细胞层面的表型注释。

4 功能验证与机制解析

4.1 增强子的功能验证

大规模的功能实验为增强子的大批量筛选提供快速高效的方案, 然而, 这些方法仍具有一定的局限性, 需要用进一步的功能实验进行补充: 首先, MPRA鉴定所得的功能性增强子仅考虑了序列本身的信息, 需要进一步评估其在内源性基因组背景中的作用^[126]; 其次, CRISPR筛选具有较高的背景噪音^[155], 例如存在大量无功能的gRNA等, 需要对筛选出的功能性增强子进行进一步验证; 最后, 以gRNA的富集作为唯一读数的CRISPR筛选仅能对表型相关的位点进行筛选, 但无法同时探究其功能, 而多维读数的CRISPR筛选常常受限于技术和成本而无法扩大规模, 并且目前大多只能检测单个组学^[156,157], 因此, 还需要在单个增强子层面进行完整的功能机制探究。

对高通量筛选出的功能性增强子进行进一步验证和探究, 首先需要在对应的系统中进行单个增强子的编辑^[15,105]: 对于功能缺失型增强子, 可使用CRISPRi抑制增强子活性, 或者使用CRISPR-Cas9敲除整段增强子序列; 对于功能获得型增强子, 通常使用CRISPRa激活增强子区域, 或者使用CRISPR-Cas9敲入增强子、转录因子结合序列或特定突变。

为了将功能性增强子与表型直接联系起来, 需要对增强子编辑后的细胞株进行各种表型的检测。例如, 检测细胞的活力以表征细胞增殖表型^[158]; 检测IC₅₀以表征细胞的药物响应表型^[159,160]; 将免疫细胞与癌细胞进行共培养, 以检测癌细胞免疫逃逸功能^[161]等; 对于调控神经相关基因的增强子, 可以在干细胞中敲除该增强子并在定向分化诱导后检测神经元的分化程度^[105]; 对于影响肿瘤生长相关的增强子, 除了在相应癌细胞系中检测癌细胞的克隆增殖能力, 还可以进行增强子敲除后在体内检测肿瘤生长情况^[48]; 对于与发育相关的增强子, 通常需要在个体层面验证增强子与表型的关联, 例如对于多指畸形相关的增强子ZRS, 可在小鼠受精卵中引入相关的序列变异, 并判断其胚胎

肢端的发育情况^[136]。

4.2 增强子的作用机制

增强子通过调控其靶基因, 从而与各种表型产生联系^[162]。为找到增强子调控的靶基因, 对于单个增强子编辑后的细胞株, 一般通过qPCR对靶基因表达进行定量检测, 以验证增强子扰动对于基因表达水平的影响^[15]。此外, 由于一个增强子可能调控多个基因^[163], 为寻找更多直接靶基因或受到影响的下游基因, 可以对单个增强子扰动的细胞株进行RNA-seq以在全基因组范围内寻找发生转录水平变化的基因。

增强子发挥其转录调控作用通常包括几个方面: (i) 转录因子的结合。增强子序列上可能包含一个或多个转录因子结合位点, 增强子结合转录因子并招募转录复合物, 促进靶基因表达。通过检索增强子序列上的转录因子结合基序, 以及使用ChIP-seq^[164], CUT&Tag^[165]或FitCUT&RUN^[166]等实验验证相应蛋白的结合, 可以进一步确定增强子的核心活性区域。(ii) 具有染色质相互作用。增强子作为远端调控元件, 通过与启动子形成远程接触, 在物理上靠近靶基因促成调控。CTCF和黏连蛋白、多梳蛋白复合物、YY1、eRNA等都可能参与调节染色质的物理接触。通过Hi-C^[167]及其一系列衍生技术如HiChIP^[168], capture Hi-C^[169], ChIATAC^[170]等检测全局或与特定特征相关的染色质接触, 或利用4C^[171]检测与给定位点有接触的所有染色质区域, 可以直接识别增强子与靶基因之间或增强子簇内部的染色质相互作用。(iii) 增强子通过转录过程产生eRNA。eRNA可以作为调控因子直接调控靶基因的转录^[51], 也可以通过形成或维持染色质环而促进增强子对靶基因的激活^[172,173]。利用GRO-seq^[174], GRO-cap^[175], PRO-seq^[176]等检测新生转录本的方法可以检测到不稳定的eRNA转录本, 利用RIC-seq^[172]可以直接检测RNA-RNA相互作用, 继而确认eRNA在转录调控中发挥的作用。(iv) 多个增强子共同参与靶基因调控。多个增强子可能以增强子簇的形式, 通过不同的机制精确调控靶基因的表达。例如, 多个增强子可能在调控同一基因时发挥冗余作用, 因而敲除单个增强子的效应可能较轻微, 而对整个冗余增强子簇的敲除会显示出明显的表型^[14,133]; 超级增强子之间可能存在层次结构或者形成调控网络, 其中枢纽增强子负责协调其他增强子来共同调控靶基因, 因而

敲除枢纽增强子能够显著破坏超级增强子的功能^[15,110]; 在空间上聚集的超级增强子也可能通过形成相分离的凝聚物, 为转录调控提供相对隔离的反应环境和高浓度的转录相关因子, 以维持关键基因的稳健转录^[177]; 超长距离的增强子簇可能同时包含加和、协同等多种增强子相互作用的模式, 形成一个多层嵌套的复杂调控网络^[16].

4.3 增强子相关的新兴实验技术

近年来, 新兴实验技术的发展, 为功能性增强子的功能机制探究开辟新的方向.

增强子和启动子相互作用形成的染色质环可以作为识别增强子和靶基因功能连接的方法, Hi-C等染色体构象捕获技术通过交联DNA与蛋白质, 基于DNA打断和临近连接来识别物理空间上靠近的染色质相互作用. 在此基础上, 新一代单细胞三维基因组学技术如scSPRITE^[178], scNanoHi-C^[179]和Droplet Hi-C^[180]等可以高通量检测单细胞内的高阶染色质相互作用, 解决在复杂组织样本中存在的细胞异质性问题, 并且分辨出单个细胞内基因组结构随时间动态的变化; 同时结合转录组检测的HiRES^[181], scCARE-seq^[182], LiM-CA^[183], 以及在此基础上进一步提高通量的GAGE-seq^[184], Paired Hi-C^[180], dscHi-C multiome^[185]等方法, 能够将染色质结构的变化与特定的功能表型联系起来; 此外, MUSIC技术^[186]的发展使分析多重染色质相互作用(包括DNA-DNA, DNA-RNA或RNA-RNA)成为可能.

在eRNA探究方面, 新兴的scGRO-seq^[187]和Reap-TEC^[188]方法可以检测单细胞水平的eRNA转录, 为活性增强子的识别和功能机制解析开辟新途径. 在组蛋白修饰探究方面, TACIT和CoTACIT^[189]实现单个细胞内多种组蛋白修饰的同时检测, 对于探究发育过程中的细胞异质性和表观遗传机制具有重要价值; Si-Tomics^[190]技术可以在活细胞内模拟特定的组蛋白修饰分布, 并检测与其发生相互作用的蛋白质组与基因组, 从而揭示不同生理条件下的超级增强子介导的染色质动态调控. 此外, ENGRAM^[191]技术创新性地利用DNA序列记录多种细胞信号, 为研究增强子活性和调控过程提供新视角.

在疾病研究领域, 为了将癌症驱动突变与细胞行为和临床表型建立直接的因果联系, DEFND-seq^[192]等

技术在单细胞水平同时捕获DNA变异和RNA表达, 而ASAP-seq及其扩展的DOGMA-seq^[193]和GoT-ChA^[194]等多模态检测技术的发展, 能够将基因型与染色质状态、基因表达、表面蛋白表达等功能表型联系起来, 用以识别癌细胞特异的转录调控机制.

5 临床应用

随着对增强子在疾病发生发展中关键作用的深入认识, 针对增强子的治疗策略已成为临床转化研究的重要方向. 基于前期对功能性增强子的系统筛选和功能验证, 多种治疗策略已进入正式应用或临床试验阶段. 这些策略主要包括利用CRISPR技术精准编辑增强子区域、干预转录因子与增强子的相互作用、调控增强子的表观遗传修饰状态, 以及靶向超级增强子依赖的转录调控通路. 这些多层面的治疗策略为疾病的精准治疗提供了丰富的干预靶点.

(i) CRISPR基因编辑技术通过精准靶向疾病相关的增强子区域, 可直接调控基因表达模式(图4A). 例如, 利用CRISPR-Cas9靶向*BCL11A*基因的红系特异性增强子区域, 对其进行切割, 通过非同源末端连接(non-homologous end joining, NHEJ)修复途径引入缺失突变(indels), 从而破坏增强子的调控功能. 这导致*BCL11A*在红细胞中的表达降低, 解除了其对胎儿血红蛋白(HbF) γ -珠蛋白基因的转录抑制, 使成人红细胞重新高效表达HbF^[8]. HbF能够替代镰刀型细胞贫血(SCD)中突变的 β -珠蛋白(HbS), 抑制红细胞镰变, 同时弥补地中海贫血患者因 β -珠蛋白合成不足导致的贫血问题^[5]. 这种策略通过激活内源性HbF的生理补偿机制, 显著改善溶血、缺氧等症状. 相较于直接编辑编码基因可能带来的广泛基因组影响, 增强子编辑因其高度的细胞类型特异性而更具治疗优势. 目前, 基于该原理的CRISPR疗法Casgevy已成功获批上市^[5,6].

(ii) 在转录因子-增强子相互作用层面, 多种抑制剂的开发为临床治疗提供新的机会. 例如, BET抑制剂通过靶向溴结构域蛋白, 抑制其与乙酰化组蛋白(如H3K27ac)结合, 阻断增强子驱动的基因表达(图4B). pelabresib通过靶向溴结构域蛋白来抑制炎症细胞因子基因的表达, 在联合ruxolitinib治疗骨髓纤维化患者时展现出显著疗效^[195](NCT04603495). 在临床前研究中, 采用纳米乳剂eNano-JQ1进行眼内给药, 有效缓解

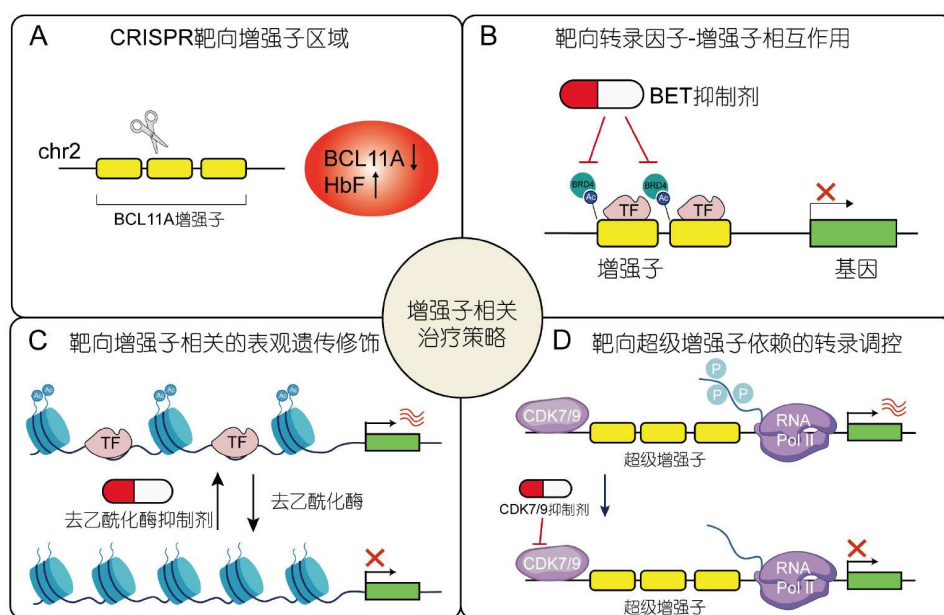


图 4 增强子相关的治疗策略。A: 利用CRISPR-Cas9靶向增强子区域, 代表案例如靶向*BCL11A*的增强子抑制其表达, 从而使成人红细胞重新高效表达HbF; B: 靶向转录因子-增强子相互作用, 例如BET抑制剂竞争性结合, 阻断BRD4蛋白与乙酰化组蛋白的结合, 进而抑制致癌基因的异常转录; C: 靶向增强子相关的表观遗传修饰, 如去乙酰化酶抑制剂通过恢复染色质可及性, 重新激活抑癌基因的表达; D: 靶向超级增强子依赖的转录调控, 如CDK7和CDK9抑制剂, 使RNA聚合酶II的磷酸化与转录起始或延伸受阻, 从而抑制相关基因的表达

Figure 4 Enhancer-related therapeutic strategies. A: Targeting enhancer regions using CRISPR-Cas9. A representative example involves targeting the enhancer of *BCL11A* to suppress its expression, thereby reactivating high-efficiency fetal hemoglobin production in adult erythrocytes; B: targeting TF-enhancer interactions. For instance, BET inhibitors competitively bind to block the interaction between BRD4 protein and acetylated histones, consequently inhibiting abnormal transcription of oncogenes; C: targeting enhancer-associated epigenetic modifications. Histone deacetylase inhibitors restore chromatin accessibility to reactivate tumor suppressor gene expression; D: targeting super-enhancer-dependent transcriptional regulation. For example, CDK7 and CDK9 inhibitors impede phosphorylation of RNA polymerase II and block transcriptional initiation or elongation, thereby suppressing expression of associated genes

小鼠创伤性增生性玻璃体视网膜病变的进展^[196]; 使用BRD4抑制剂和RET抑制剂的联合治疗在体外和体内实验中显示出抑制雌激素及其受体 α (estrogen and estrogen receptor α , ER α)阳性乳腺癌的协同效果^[197]。此外, 靶向蛋白降解剂AU-15330通过降解SWI/SNF复合物的ATPase亚基, 降低增强子区域的染色质可及性, 阻断驱动前列腺癌细胞增殖的关键转录因子的结合, 在前列腺癌模型中展现出显著的抗肿瘤活性^[198]。

(iii) 在表观遗传修饰方面, 研究主要集中于调节增强子的组蛋白修饰和DNA甲基化(图4C)。例如, 组蛋白去乙酰化酶(histone deacetylase, HDAC)抑制剂AR-42在2型神经纤维瘤病患者中展现出良好的安全性和初步的抗肿瘤活性^[199](NCT02569320)。DNA甲基转移酶(DNA methyltransferases, DNMT)抑制剂guadecitabine (SGI-110)在复发/难治性外周T细胞淋巴瘤患者中显示出可接受的客观缓解率和毒性特征, 为后续联合

治疗策略的开发提供重要依据^[200]。

(iv) 超级增强子依赖的转录调控机制为肿瘤治疗提供新靶点(图4D)。CDK7/9抑制剂通过靶向作用于转录起始和延伸阶段的关键调控激酶——CDK7和CDK9, 能够有效抑制由超级增强子驱动的癌基因表达。例如, CDK7抑制剂SY-5609在晚期实体瘤和胰腺导管腺癌的临床试验中显示出乐观趋势(NCT04247126), 而CDK9抑制剂Dinaciclib则在复发/难治性慢性淋巴细胞白血病患者中取得良好效果^[201](NCT01580228)。

综上, 针对增强子的多维度干预策略正推动疾病治疗模式向精准化方向革新, 这些策略通过不同分子机制重塑病理状态下的基因表达网络。然而, 增强子靶向治疗从实验室到临床仍面临着诸多挑战, 包括靶向特异性的优化、给药系统的改进以及潜在副作用的评估等。未来, 随着新型递送载体的开发和基因编辑技

术的完善, 结合人工智能辅助药物设计等创新方法, 有望克服这些技术瓶颈, 推动更多增强子靶向治疗策略进入临床应用。

6 总结与展望

作为基因表达调控的核心元件, 功能性增强子的

精准识别与机制解析是揭示发育程序与疾病机理的关键突破口^[88,110,202,203]。近年来, 随着单细胞多组学、空间表观基因组学、三维基因组学及基因编辑技术的快速发展, 功能性增强子的筛选与验证策略取得突破性进展。因此, 本文总结并提出一个多层次、系统化的功能性增强子的研究框架(图5)。首先, 通过高通量测序绘制染色质可及性、组蛋白修饰、转录因子结合及

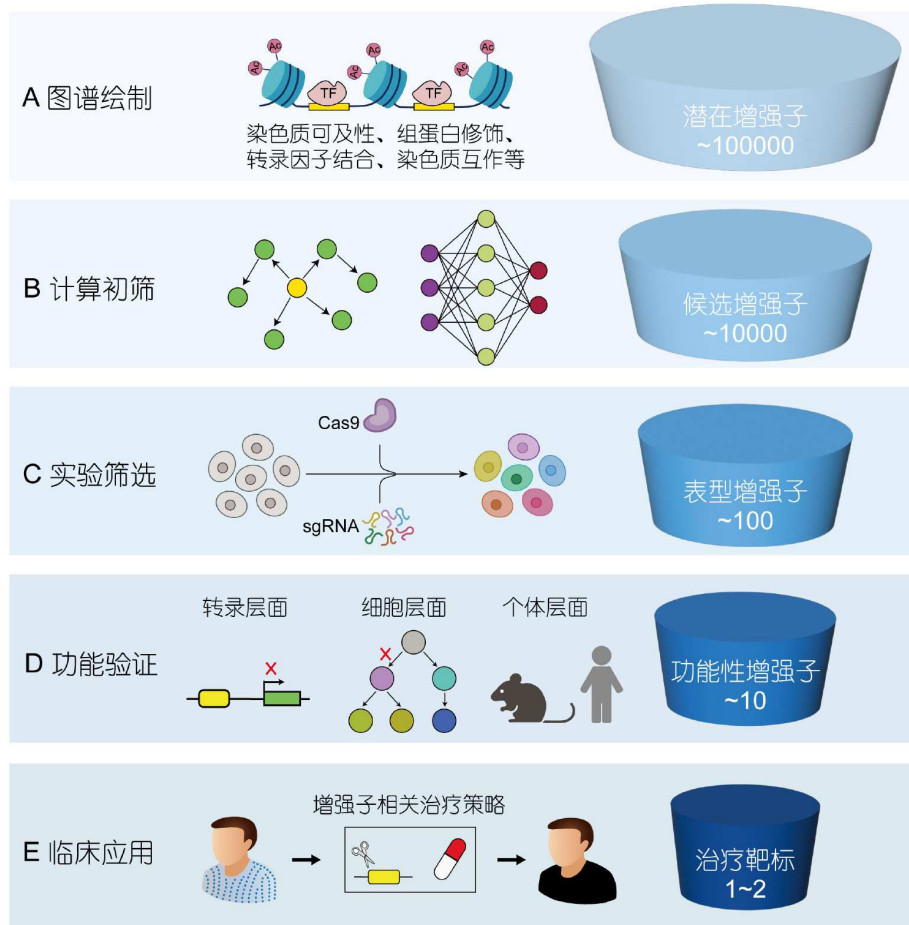


图 5 功能性增强子筛选与验证的研究策略。A: 通过分析染色质可及性、组蛋白修饰、转录因子结合及染色质互作等特征, 鉴定具有活跃染色质特征的基因组区域, 作为潜在增强子; B: 利用计算模型整合增强子图谱, 解析增强子调控网络复杂性, 初步筛选在调控网络中发挥关键作用的调控序列, 作为候选增强子; C: 通过表型导向的高通量实验, 筛选其扰动能显著影响细胞或个体表型的DNA元件, 作为表型增强子; D: 对转录调控、细胞功能及个体水平进行系统性功能验证, 确定能特异性调控基因表达并引起显著表型变化的调控元件, 作为功能性增强子; E: 评估功能性增强子在疾病治疗中的应用价值, 鉴定在疾病进展中发挥关键调控作用且具有良好可干预性的增强子, 作为治疗靶标

Figure 5 Research strategy for functional enhancer screening and validation. A: Identify genomic regions with active chromatin features as potential enhancers by analyzing features such as chromatin accessibility, histone modifications, transcription factor binding and chromatin interactions; B: use computational models to integrate enhancer landscapes, resolve the complexity of enhancer regulatory networks, and preliminarily screen regulatory sequences that play key roles in the regulatory network as candidate enhancers; C: screen for DNA elements whose perturbation significantly affects cellular or organismal phenotypes through phenotype-oriented high-throughput experiments as phenotypic enhancers; D: conduct systematic functional validation at transcriptional regulation, cellular function, and organismal levels to identify regulatory elements that can specifically regulate gene expression and cause significant phenotypic changes as functional enhancers; E: evaluate the therapeutic potential of functional enhancers and identify enhancers that play crucial regulatory roles in disease progression and possess good interventional properties as therapeutic targets

染色质互作等多维组学图谱^[64,95,116,204];其次,应用计算模型系统解析增强子的调控机制,包括其时空特异性^[82,98,112,205,206]、进化保守性^[103,105]及调控网络结构^[16,17,86,90,109]等特征,从而筛选具有重要调控潜力的候选增强子;接下来,结合表型相关的高通量实验筛选技术^[18,129,145,150,207,208],进一步确定表型增强子;随后,通过多样化的功能验证实验,鉴定关键功能性增强子并阐明其作用机制;最终,部分功能性增强子有望向临床转化,为疾病治疗提供新靶点。这种层层递进的研究策略不仅显著提升筛选效率、优化研究成本,更有望发现更多类似*BCL11A*增强子等具有重要临床应用价值的功能性调控元件^[5,7,8]。

尽管已取得显著进展,功能性增强子研究仍然面临着多维度的挑战。在数据资源层面,尽管单细胞测序和空间表观基因组学技术日新月异,但受限于技术成本和实验复杂度,高质量数据的获取仍然是一个瓶颈。在计算预测方面,现有的算法模型在准确性和泛化能力上仍有待提升。在实验验证层面,现有高通量筛选方法虽然可以在体外大规模评估增强子活性,

但往往难以完整模拟体内复杂的调控网络和动态变化。在临床转化层面,尽管许多研究证实疾病相关的遗传变异大部分位于增强子等非编码区域,但将研究发现转化为临床应用仍面临诸多技术难题,如靶向递送系统的特异性和安全性、治疗效果的持久性等。

展望未来,随着新型技术的不断成熟和自动化实验平台的推广应用,有望实现更多维度、更大规模、更低成本的数据获取,为增强子图谱的绘制提供更全面的支持。通过整合多组学数据并设计更有效的架构,特别是结合大语言模型在生物序列分析中的应用,有望构建出更精准的预测模型^[209~212]。在实验研究方向上,新型的体内验证技术结合实时监测技术,将有助于更准确地解析增强子在生理和病理条件下的调控机制^[136,191,213]。随着基因编辑技术的进步和药物递送系统的优化,未来有机会开发出更多基于增强子调控的精准治疗方案。可以预见,随着跨学科技术的深度融合与临床转化路径的清晰化,功能性增强子研究将为发育生物学和疾病治疗带来颠覆性突破,最终实现从“读懂基因组暗物质”到“改写生命调控密码”的跨越。

参考文献

- Banerji J, Rusconi S, Schaffner W. Expression of a β -globin gene is enhanced by remote SV40 DNA sequences. *Cell*, 1981, 27: 299–308
- Lin X, Chen J D, Wang C Y, et al. Cooperation of MLL1 and Jun in controlling H3K4me3 on enhancers in colorectal cancer. *Genome Biol*, 2023, 24: 268
- Abascal F, Acosta R, Addleman N J, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*, 2020, 583: 699–710
- Bahr C, von Paleske L, Uslu V V, et al. A *Myc* enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. *Nature*, 2018, 553: 515–520
- Fu B, Liao J, Chen S, et al. CRISPR-Cas9-mediated gene editing of the *BCL11A* enhancer for pediatric β^0/β^0 transfusion-dependent β -thalassemia. *Nat Med*, 2022, 28: 1573–1580
- Frangoul H, Altshuler D, Cappellini M D, et al. CRISPR-Cas9 gene editing for sickle cell disease and β -thalassemia. *N Engl J Med*, 2021, 384: 252–260
- Canver M C, Smith E C, Sher F, et al. *BCL11A* enhancer dissection by Cas9-mediated *in situ* saturating mutagenesis. *Nature*, 2015, 527: 192–197
- Bauer D E, Kamran S C, Lessard S, et al. An erythroid enhancer of *BCL11A* subject to genetic variation determines fetal hemoglobin level. *Science*, 2013, 342: 253–257
- Hnisz D, Abraham B J, Lee T I, et al. Super-enhancers in the control of cell identity and disease. *Cell*, 2013, 155: 934–947
- Cheng H, Dou X, Han J D J. Understanding super-enhancers. *Sci China Life Sci*, 2016, 59: 277–280
- Hay D, Hughes J R, Babbs C, et al. Genetic dissection of the α -globin super-enhancer *in vivo*. *Nat Genet*, 2016, 48: 895–903
- Fulton R, Van Ness B. Selective synergy of immunoglobulin enhancer elements in B-cell development: a characteristic of kappa light chain enhancers, but not heavy chain enhancers. *Nucleic Acids Res*, 1994, 22: 4216–4223
- Guerrero L, Marco-Ferreres R, Serrano A L, et al. Secondary enhancers synergise with primary enhancers to guarantee fine-tuned muscle gene

- expression. *Dev Biol*, 2010, 337: 16–28
- 14 Osterwalder M, Barozzi I, Tissi res V, et al. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature*, 2018, 554: 239–243
 - 15 Huang J, Li K, Cai W, et al. Dissecting super-enhancer hierarchy based on chromatin interactions. *Nat Commun*, 2018, 9: 943
 - 16 Lin X, Liu Y, Liu S, et al. Nested epistasis enhancer networks for robust genome regulation. *Science*, 2022, 377: 1077–1085
 - 17 Hong D, Lin H, Liu L, et al. Complexity of enhancer networks predicts cell identity and disease genes revealed by single-cell multi-omics analysis. *Brief Bioinf*, 2023, 24: bbac508
 - 18 Klann T S, Black J B, Chellappan M, et al. CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat Biotechnol*, 2017, 35: 561–568
 - 19 Zhang T, Zhang Z, Dong Q, et al. Histone H3K27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells. *Genome Biol*, 2020, 21: 45
 - 20 Buenrostro J D, Wu B, Chang H Y, et al. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *CP Mol Biol*, 2015, 109: 21–29
 - 21 Nobusada T, Yip C W, Agrawal S, et al. Update of the FANTOM web resource: enhancement for studying noncoding genomes. *Nucleic Acids Res*, 2025, 53: D419–D424
 - 22 Grosveld F, van Assendelft G B, Greaves D R, et al. Position-independent, high-level expression of the human β -globin gene in transgenic mice. *Cell*, 1987, 51: 975–985
 - 23 Goldhamer D J, Brunk B P, Faerman A, et al. Embryonic activation of the *myoD* gene is regulated by a highly conserved distal control element. *Development*, 1995, 121: 637–649
 - 24 Wang R, Chen F, Chen Q, et al. MyoD is a 3D genome structure organizer for muscle cell identity. *Nat Commun*, 2022, 13: 205
 - 25 Smemo S, Campos L C, Moskowitz I P, et al. Regulatory variation in a *TBX5* enhancer leads to isolated congenital heart disease. *Hum Mol Genet*, 2012, 21: 3255–3263
 - 26 Zhang J, Wang Q, Qi S, et al. An oncogenic enhancer promotes melanoma progression via regulating *ETV4* expression. *J Transl Med*, 2024, 22: 547
 - 27 Lettice L A, Devenney P, De Angelis C, et al. The conserved sonic hedgehog limb enhancer consists of discrete functional elements that regulate precise spatial expression. *Cell Rep*, 2017, 20: 1396–1408
 - 28 Kvon E Z, Kamneva O K, Melo U S, et al. Progressive loss of function in a limb enhancer during snake evolution. *Cell*, 2016, 167: 633–642.e11
 - 29 Gonen N, Futtner C R, Wood S, et al. Sex reversal following deletion of a single distal enhancer of *Sox9*. *Science*, 2018, 360: 1469–1473
 - 30 Yeom Y I, Fuhrmann G, Ovitt C E, et al. Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development*, 1996, 122: 881–894
 - 31 Savic D, Ye H, Aneas I, et al. Alterations in *TCF7L2* expression define its role as a key regulator of glucose metabolism. *Genome Res*, 2011, 21: 1417–1425
 - 32 Sato S, Ikeda K, Shioi G, et al. Regulation of *Six1* expression by evolutionarily conserved enhancers in tetrapods. *Dev Biol*, 2012, 368: 95–108
 - 33 Praetorius C, Grill C, Stacey S N, et al. A polymorphism in *IRF4* affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. *Cell*, 2013, 155: 1022–1033
 - 34 Blinka S, Reimer Jr. M H, Pulakanti K, et al. Super-enhancers at the *Nanog* locus differentially regulate neighboring pluripotency-associated genes. *Cell Rep*, 2016, 17: 19–28
 - 35 Takeda D Y, Spis k S, Seo J H, et al. A somatically acquired enhancer of the androgen receptor is a noncoding driver in advanced prostate cancer. *Cell*, 2018, 174: 422–432.e13
 - 36 Song H, Liu Y, Tan Y, et al. Recurrent noncoding somatic and germline *WT1* variants converge to disrupt MYB binding in acute promyelocytic leukemia. *Blood*, 2022, 140: 1132–1144
 - 37 Mansour M R, Abraham B J, Anders L, et al. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science*, 2014, 346: 1373–1377
 - 38 Li Z, Abraham B J, Berezovskaya A, et al. APOBEC signature mutation generates an oncogenic enhancer that drives *LMO1* expression in T-ALL. *Leukemia*, 2017, 31: 2057–2064
 - 39 Jindal G A, Bantle A T, Solvason J J, et al. Single-nucleotide variants within heart enhancers increase binding affinity and disrupt heart

- development. *Dev Cell*, 2023, 58: 2206–2216.e5
- 40 Lim F, Solvason J J, Ryan G E, et al. Affinity-optimizing enhancer variants disrupt development. *Nature*, 2024, 626: 151–159
- 41 Schoenfelder S, Fraser P. Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet*, 2019, 20: 437–455
- 42 Yang J H, Hansen A S. Enhancer selectivity in space and time: from enhancer-promoter interactions to promoter activation. *Nat Rev Mol Cell Biol*, 2024, 25: 574–591
- 43 Ge X, Huang H, Han K, et al. Outward-oriented sites within clustered CTCF boundaries are key for intra-TAD chromatin interactions and gene regulation. *Nat Commun*, 2023, 14: 8101
- 44 Lupiáñez D G, Kraft K, Heinrich V, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, 2015, 161: 1012–1025
- 45 Hnisz D, Weintraub A S, Day D S, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*, 2016, 351: 1454–1458
- 46 Tian J, Lou J, Cai Y, et al. Risk SNP-mediated enhancer-promoter interaction drives colorectal cancer through both *FADS2* and *AP002754.2*. *Cancer Res*, 2020, 80: 1804–1818
- 47 Grossi E, Raimondi I, Goñi E, et al. A lncRNA-SWI/SNF complex crosstalk controls transcriptional activation at specific promoter regions. *Nat Commun*, 2020, 11: 936
- 48 Jiao W, Chen Y, Song H, et al. HPSE enhancer RNA promotes cancer progression through driving chromatin looping and regulating hnRNPU/p300/EGR1/HPSE axis. *Oncogene*, 2018, 37: 2728–2745
- 49 Liang L, Cao C, Ji L, et al. Complementary Alu sequences mediate enhancer-promoter selectivity. *Nature*, 2023, 619: 868–875
- 50 Watts J A, Grunseich C, Rodriguez Y, et al. A common transcriptional mechanism involving R-loop and RNA abasic site regulates an enhancer RNA of *APOE*. *Nucleic Acids Res*, 2022, 50: 12497–12514
- 51 Wan M, Liu Y, Li D, et al. The enhancer RNA, AANCR, regulates APOE expression in astrocytes and microglia. *Nucleic Acids Res*, 2024, 52: 10235–10254
- 52 Weischenfeldt J, Dubash T, Drinas A P, et al. Pan-cancer analysis of somatic copy-number alterations implicates *IRS4* and *IGF2* in enhancer hijacking. *Nat Genet*, 2017, 49: 65–74
- 53 Xu J, Song F, Lyu H, et al. Subtype-specific 3D genome alteration in acute myeloid leukaemia. *Nature*, 2022, 611: 387–398
- 54 Herranz D, Ambesi-Impiombato A, Palomero T, et al. A NOTCH1-driven *MYC* enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. *Nat Med*, 2014, 20: 1130–1137
- 55 Bailey C, Pich O, Thol K, et al. Origins and impact of extrachromosomal DNA. *Nature*, 2024, 635: 193–200
- 56 Morton A R, Dogan-Artun N, Faber Z J, et al. Functional enhancers shape extrachromosomal oncogene amplifications. *Cell*, 2019, 179: 1330–1341.e13
- 57 Wu S, Turner K M, Nguyen N, et al. Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature*, 2019, 575: 699–703
- 58 Helmsauer K, Valieva M E, Ali S, et al. Enhancer hijacking determines extrachromosomal circular MYCN amplicon architecture in neuroblastoma. *Nat Commun*, 2020, 11: 5823
- 59 Zhu Y, Gujar A D, Wong C H, et al. Oncogenic extrachromosomal DNA functions as mobile enhancers to globally amplify chromosomal transcription. *Cancer Cell*, 2021, 39: 694–707.e7
- 60 Hung K L, Yost K E, Xie L, et al. ecDNA hubs drive cooperative intermolecular oncogene expression. *Nature*, 2021, 600: 731–736
- 61 Zhang K, Hocker J D, Miller M, et al. A single-cell atlas of chromatin accessibility in the human genome. *Cell*, 2021, 184: 5985–6001.e19
- 62 Domcke S, Hill A J, Daza R M, et al. A human cell atlas of fetal chromatin accessibility. *Science*, 2020, 370: eaba7612
- 63 Sun K, Liu X, Lan X. A single-cell atlas of chromatin accessibility in mouse organogenesis. *Nat Cell Biol*, 2024, 26: 1200–1211
- 64 Li M, Jiang Z, Xu X, et al. Chromatin accessibility landscape of mouse early embryos revealed by single-cell NanoATAC-seq2. *Science*, 2025, 387: eadp4319
- 65 Liu Y, Luo X, Sun Y, et al. Comparative single-cell multiome identifies evolutionary changes in neural progenitor cells during primate brain development. *Dev Cell*, 2025, 60: 414–428.e8
- 66 Ye F, Zhang S, Fu Y, et al. Fast and flexible profiling of chromatin accessibility and total RNA expression in single nuclei using Microwell-seq3. *Cell Discov*, 2024, 10: 33

- 67 Smith R J, Zhang H, Hu S S, et al. Single-cell chromatin profiling of the primitive gut tube reveals regulatory dynamics underlying lineage fate decisions. *Nat Commun*, 2022, 13: 2965
- 68 Sundaram L, Kumar A, Zatzman M, et al. Single-cell chromatin accessibility reveals malignant regulatory programs in primary human cancers. *Science*, 2024, 385: eadk9217
- 69 Xiong X, James B T, Boix C A, et al. Epigenomic dissection of Alzheimer's disease pinpoints causal variants and reveals epigenome erosion. *Cell*, 2023, 186: 4422–4437.e21
- 70 Moore J E, Pratt H E, Fan K, et al. An expanded registry of candidate *cis*-regulatory elements for studying transcriptional regulation. *bioRxiv*, 2024, 2024: 629296
- 71 Liu B, He Y, Wu X, et al. Mapping putative enhancers in mouse oocytes and early embryos reveals TCF3/12 as key folliculogenesis regulators. *Nat Cell Biol*, 2024, 26: 962–974
- 72 Taing L, Dandawate A, L'Yi S, et al. Cistrome Data Browser: integrated search, analysis and visualization of chromatin data. *Nucleic Acids Res*, 2024, 52: D61–D66
- 73 Choi E, Machado C R L, Okano T, et al. Joint-specific rheumatoid arthritis fibroblast-like synoviocyte regulation identified by integration of chromatin access and transcriptional activity. *JCI Insight*, 2024, 9: e179392
- 74 Heffel M G, Zhou J, Zhang Y, et al. Temporally distinct 3D multi-omic dynamics in the developing human brain. *Nature*, 2024, 635: 481–489
- 75 Liu H, Zeng Q, Zhou J, et al. Single-cell DNA methylome and 3D multi-omic atlas of the adult mouse brain. *Nature*, 2023, 624: 366–377
- 76 Buniello A, MacArthur J A L, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*, 2019, 47: D1005–D1012
- 77 Huang D, Zhou Y, Yi X, et al. VannoPortal: multiscale functional annotation of human genetic variants for interrogating molecular mechanism of traits and diseases. *Nucleic Acids Res*, 2022, 50: D1408–D1416
- 78 Chen W, Wang Z, Wang Y, et al. Enhancer RNA transcriptome-wide association study reveals a distinctive class of pan-cancer susceptibility eRNAs. *Adv Sci*, 2025, 12: e2411974
- 79 Fulco C P, Nasser J, Jones T R, et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat Genet*, 2019, 51: 1664–1669
- 80 Sakaue S, Weinand K, Isaac S, et al. Tissue-specific enhancer-gene maps from multimodal single-cell data identify causal disease alleles. *Nat Genet*, 2024, 56: 615–626
- 81 Mitra S, Malik R, Wong W, et al. Single-cell multi-ome regression models identify functional and disease-associated enhancers and enable chromatin potential analysis. *Nat Genet*, 2024, 56: 627–636
- 82 Persad S, Choo Z N, Dien C, et al. SEACells infers transcriptional and epigenomic cellular states from single-cell genomics data. *Nat Biotechnol*, 2023, 41: 1746–1757
- 83 Hao G, Fan Y, Yu Z, et al. Topological identification and interpretation for single-cell epigenetic regulation elucidation in multi-tasks using scAGDE. *Nat Commun*, 2025, 16: 1691
- 84 Yuan H, Kelley D R. scBasset: sequence-based modeling of single-cell ATAC-seq using convolutional neural networks. *Nat Methods*, 2022, 19: 1088–1096
- 85 Zhang L, Zhang J, Nie Q. DIRECT-NET: an efficient method to discover *cis*-regulatory elements and construct regulatory networks from single-cell multiomics data. *Sci Adv*, 2022, 8: eabl7393
- 86 Bravo González-Blas C, De Winter S, Hulselmans G, et al. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat Methods*, 2023, 20: 1355–1367
- 87 Qian F C, Zhou L W, Li Y Y, et al. SEanalysis 2.0: a comprehensive super-enhancer regulatory network analysis tool for human and mouse. *Nucleic Acids Res*, 2023, 51: W520–W527
- 88 Kim S S, Truong B, Jagadeesh K, et al. Leveraging single-cell ATAC-seq and RNA-seq to identify disease-critical fetal and adult brain cell types. *Nat Commun*, 2024, 15: 563
- 89 Gupta A, Weinand K, Nathan A, et al. Dynamic regulatory elements in single-cell multimodal data implicate key immune cell states enriched for autoimmune disease heritability. *Nat Genet*, 2023, 55: 2200–2210
- 90 Xi X, Li J, Jia J, et al. A mechanism-informed deep neural network enables prioritization of regulators that drive cell state transitions. *Nat Commun*, 2025, 16: 1284

- 91 Lu T, Ang C E, Zhuang X. Spatially resolved epigenomic profiling of single cells in complex tissues. [Cell](#), 2022, 185: 4448–4464.e17
- 92 Deng Y, Bartosovic M, Kukanja P, et al. Spatial-CUT&Tag: spatially resolved chromatin modification profiling at the cellular level. [Science](#), 2022, 375: 681–686
- 93 Deng Y, Bartosovic M, Ma S, et al. Spatial profiling of chromatin accessibility in mouse and human tissues. [Nature](#), 2022, 609: 375–383
- 94 Zhang D, Deng Y, Kukanja P, et al. Spatial epigenome-transcriptome co-profiling of mammalian tissues. [Nature](#), 2023, 616: 113–122
- 95 Jiang F, Zhou X, Qian Y, et al. Simultaneous profiling of spatial gene expression and chromatin accessibility during mouse brain development. [Nat Methods](#), 2023, 20: 1048–1057
- 96 Russell A J C, Weir J A, Nadaf N M, et al. Slide-tags enables single-nucleus barcoding for multimodal spatial genomics. [Nature](#), 2024, 625: 101–109
- 97 Guo P, Mao L, Chen Y, et al. Multiplexed spatial mapping of chromatin features, transcriptome and proteins in tissues. [Nat Methods](#), 2025, 22: 520–529
- 98 Chen X, Li K, Wu X, et al. Descart: a method for detecting spatial chromatin accessibility patterns with inter-cellular correlations. [Genome Biol](#), 2024, 25: 322
- 99 Yang P, Jin K, Yao Y, et al. Spatial integration of multi-omics single-cell data with SIMO. [Nat Commun](#), 2025, 16: 1265
- 100 Kuderna L F K, Ulirsch J C, Rashid S, et al. Identification of constrained sequence elements across 239 primate genomes. [Nature](#), 2024, 625: 735–742
- 101 Sullivan P F, Meadows J R S, Gazal S, et al. Leveraging base-pair mammalian constraint to understand genetic variation and human disease. [Science](#), 2023, 380: eabn2937
- 102 Kaplow I M, Schäffer D E, Wirthlin M E, et al. Inferring mammalian tissue-specific regulatory conservation by predicting tissue-specific differences in open chromatin. [BMC Genomics](#), 2022, 23: 291
- 103 Kaplow I M, Lawler A J, Schäffer D E, et al. Relating enhancer genetic variation across mammals to complex phenotypes using machine learning. [Science](#), 2023, 380: eabm7993
- 104 Capra J A, Erwin G D, McKinsey G, et al. Many human accelerated regions are developmental enhancers. [Phil Trans R Soc B](#), 2013, 368: 20130025
- 105 Cui X, Yang H, Cai C, et al. Comparative characterization of human accelerated regions in neurons. [Nature](#), 2025, 640: 991–999
- 106 Madsen J G S, Madsen M S, Rauch A, et al. Highly interconnected enhancer communities control lineage-determining genes in human mesenchymal stem cells. [Nat Genet](#), 2020, 52: 1227–1238
- 107 Pliner H A, Packer J S, McFaline-Figueroa J L, et al. Cicero predicts *cis*-regulatory DNA interactions from single-cell chromatin accessibility data. [Mol Cell](#), 2018, 71: 858–871.e8
- 108 Gao V R, Yang R, Das A, et al. ChromaFold predicts the 3D contact map from single-cell chromatin accessibility. [Nat Commun](#), 2024, 15: 9432
- 109 Lin H, Ye X, Chen W, et al. Modular organization of enhancer network provides transcriptional robustness in mammalian development. [Nucleic Acids Res](#), 2025, 53: gkae1323
- 110 Shu M, Hong D, Lin H, et al. Single-cell chromatin accessibility identifies enhancer networks driving gene expression during spinal cord development in mouse. [Dev Cell](#), 2022, 57: 2761–2775.e6
- 111 Ji Y, Zhou Z, Liu H, et al. DNABERT: pre-trained bidirectional encoder representations from transformers model for DNA-language in genome. [Bioinformatics](#), 2021, 37: 2112–2120
- 112 Fu X, Mo S, Buendia A, et al. A foundation model of transcription across human cell types. [Nature](#), 2025, 637: 965–973
- 113 de Almeida B P, Reiter F, Pagani M, et al. DeepSTARR predicts enhancer activity from DNA sequence and enables the *de novo* design of synthetic enhancers. [Nat Genet](#), 2022, 54: 613–624
- 114 Li J, Zhang P, Xi X, et al. Modeling and designing enhancers by introducing and harnessing transcription factor binding units. [Nat Commun](#), 2025, 16: 1469
- 115 Ouyang Z, Liu F, Li W, et al. The developmental and evolutionary characteristics of transcription factor binding site clustered regions based on an explainable machine learning model. [Nucleic Acids Res](#), 2024, 52: 7610–7626
- 116 Deng C, Whalen S, Steyert M, et al. Massively parallel characterization of regulatory elements in the developing human cortex. [Science](#), 2024, 384: eadh0559
- 117 Patwardhan R P, Lee C, Litvin O, et al. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. [Nat](#)

- [Biotechnol](#), 2009, 27: 1173–1175
- 118 Klein J C, Agarwal V, Inoue F, et al. A systematic evaluation of the design and context dependencies of massively parallel reporter assays. [Nat Methods](#), 2020, 17: 1083–1091
 - 119 Du A Y, Zhuo X, Sundaram V, et al. Functional characterization of enhancer activity during a long terminal repeat's evolution. [Genome Res](#), 2022, 32: 1840–1850
 - 120 Arnold C D, Gerlach D, Stelzer C, et al. Genome-wide quantitative enhancer activity maps identified by STARR-seq. [Science](#), 2013, 339: 1074–1077
 - 121 Inoue F, Kircher M, Martin B, et al. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. [Genome Res](#), 2017, 27: 38–52
 - 122 Agarwal V, Inoue F, Schubach M, et al. Massively parallel characterization of transcriptional regulatory elements. [Nature](#), 2025, 639: 411–420
 - 123 Zhao S, Yang Q, Yu Z, et al. Deciphering enhancers of hearing loss genes for efficient and targeted gene therapy of hereditary deafness. [Neuron](#), 2025
 - 124 Tippens N D, Liang J, Leung A K Y, et al. Transcription imparts architecture, function and logic to enhancer units. [Nat Genet](#), 2020, 52: 1067–1075
 - 125 Zhao S, Hong C K Y, Myers C A, et al. A single-cell massively parallel reporter assay detects cell-type-specific gene regulation. [Nat Genet](#), 2023, 55: 346–354
 - 126 Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human epigenomes. [Nature](#), 2015, 518: 317–330
 - 127 Jinek M, Chylinski K, Fonfara I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. [Science](#), 2012, 337: 816–821
 - 128 Cong L, Ran F A, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems. [Science](#), 2013, 339: 819–823
 - 129 Korkmaz G, Lopes R, Ugalde A P, et al. Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. [Nat Biotechnol](#), 2016, 34: 192–198
 - 130 Sanjana N E, Wright J, Zheng K, et al. High-resolution interrogation of functional elements in the noncoding genome. [Science](#), 2016, 353: 1545–1549
 - 131 Fei T, Li W, Peng J, et al. Deciphering essential cistromes using genome-wide CRISPR screens. [Proc Natl Acad Sci USA](#), 2019, 116: 25186–25195
 - 132 Zhu S, Li W, Liu J, et al. Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library. [Nat Biotechnol](#), 2016, 34: 1279–1286
 - 133 Li Y, Tan M, Akkari-Henić A, et al. Genome-wide Cas9-mediated screening of essential non-coding regulatory elements via libraries of paired single-guide RNAs. [Nat Biomed Eng](#), 2024, 8: 890–908
 - 134 Liu Y, Ding B, Zheng L, et al. Regulatory elements can be essential for maintaining broad chromatin organization and cell viability. [Nucleic Acids Res](#), 2022, 50: 4340–4354
 - 135 Li K, Liu Y, Cao H, et al. Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing. [Nat Commun](#), 2020, 11: 485
 - 136 Kvon E Z, Zhu Y, Kelman G, et al. Comprehensive *in vivo* interrogation reveals phenotypic impact of human enhancer variants. [Cell](#), 2020, 180: 1262–1271.e15
 - 137 Porto E M, Komor A C, Slaymaker I M, et al. Base editing: advances and therapeutic opportunities. [Nat Rev Drug Discov](#), 2020, 19: 839–859
 - 138 Chen P J, Liu D R. Prime editing for precise and highly versatile genome manipulation. [Nat Rev Genet](#), 2023, 24: 161–177
 - 139 Wang C, Liu Q, Fan X, et al. Single-cell technologies: current and near future. [Sci China Life Sci](#), 2025, 68: 1–4
 - 140 Replogle J M, Norman T M, Xu A, et al. Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. [Nat Biotechnol](#), 2020, 38: 954–961
 - 141 Adamson B, Norman T M, Jost M, et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. [Cell](#), 2016, 167: 1867–1882.e21
 - 142 Datlinger P, Rendeiro A F, Schmidl C, et al. Pooled CRISPR screening with single-cell transcriptome readout. [Nat Methods](#), 2017, 14: 297–301
 - 143 Morris J A, Caragine C, Daniloski Z, et al. Discovery of target genes and pathways at GWAS loci by pooled single-cell CRISPR screens. [Science](#), 2023, 380: eadh7699
 - 144 Li Y, Huang Z, Xu L, et al. UDA-seq: universal droplet microfluidics-based combinatorial indexing for massive-scale multimodal single-cell

- sequencing. *Nat Methods*, 2025, doi: 10.1038/s41592-024-02586-y
- 145 Xie S, Duan J, Li B, et al. Multiplexed engineering and analysis of combinatorial enhancer activity in single cells. *Mol Cell*, 2017, 66: 285–299. e5
- 146 Pacalin N M, Steinhart Z, Shi Q, et al. Bidirectional epigenetic editing reveals hierarchies in gene regulation. *Nat Biotechnol*, 2025, 43: 355–368
- 147 Rubin A J, Parker K R, Satpathy A T, et al. Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks. *Cell*, 2019, 176: 361–376.e17
- 148 Pierce S E, Granja J M, Greenleaf W J. High-throughput single-cell chromatin accessibility CRISPR screens enable unbiased identification of regulatory networks in cancer. *Nat Commun*, 2021, 12: 2969
- 149 Li A, Zeng G, Wang H, et al. DeDoc2 identifies and characterizes the hierarchy and dynamics of chromatin TAD-like domains in the single cells. *Adv Sci*, 2023, 10: e2300366
- 150 Cheng Y, Hu M, Yang B, et al. Perturb-tracing enables high-content screening of multi-scale 3D genome regulators. *Nat Methods*, 2025, 22: 950–961
- 151 Dhainaut M, Rose S A, Akturk G, et al. Spatial CRISPR genomics identifies regulators of the tumor microenvironment. *Cell*, 2022, 185: 1223–1239.e20
- 152 Binan L, Jiang A, Danquah S A, et al. Simultaneous CRISPR screening and spatial transcriptomics reveal intracellular, intercellular, and functional transcriptional circuits. *Cell*, 2025, doi: 10.1101/2023.11.30.569494
- 153 Yan R E, Corman A, Katgara L, et al. Pooled CRISPR screens with joint single-nucleus chromatin accessibility and transcriptome profiling. *Nat Biotechnol*, 2024, doi: 10.1038/s41587-024-02475-x
- 154 Mimitou E P, Cheng A, Montalbano A, et al. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat Methods*, 2019, 16: 409–412
- 155 Li W, Köster J, Xu H, et al. Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biol*, 2015, 16: 281
- 156 Cheng J, Lin G, Wang T, et al. Massively parallel CRISPR-based genetic perturbation screening at single-cell resolution. *Adv Sci*, 2023, 10: e2204484
- 157 Meyers S, Demeyer S, Cools J. CRISPR screening in hematology research: from bulk to single-cell level. *J Hematol Oncol*, 2023, 16: 107
- 158 Fulco C P, Munschauer M, Anyoha R, et al. Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science*, 2016, 354: 769–773
- 159 Lu B, Chen S, Guan X, et al. Lactate accumulation induces H4K12la to activate super-enhancer-driven RAD23A expression and promote niraparib resistance in ovarian cancer. *Mol Cancer*, 2025, 24: 83
- 160 Yu J, Chen M, Sang Q, et al. Super-enhancer activates master transcription factor NR3C1 expression and promotes 5-FU resistance in gastric cancer. *Adv Sci*, 2025, 12: e2409050
- 161 Michelatti D, Beyes S, Bernardis C, et al. Oncogenic enhancers prime quiescent metastatic cells to escape NK immune surveillance by eliciting transcriptional memory. *Nat Commun*, 2024, 15: 2198
- 162 Luo Z, Du Y, Li S, et al. Three distinct *Atoh1* enhancers cooperate for sound receptor hair cell development. *Proc Natl Acad Sci USA*, 2022, 119: e2119850119
- 163 Wu C, Huang J. Enhancer selectivity across cell types delineates three functionally distinct enhancer-promoter regulation patterns. *BMC Genomics*, 2024, 25: 483
- 164 Barski A, Cuddapah S, Cui K, et al. High-resolution profiling of histone methylations in the human genome. *Cell*, 2007, 129: 823–837
- 165 Kaya-Okur H S, Wu S J, Codomo C A, et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun*, 2019, 10: 1930
- 166 Wang X, Wang W, Wang Y, et al. Antibody-free profiling of transcription factor occupancy during early embryogenesis by FitCUT&RUN. *Genome Res*, 2022, 32: 378–388
- 167 Lieberman-Aiden E, van Berkum N L, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 2009, 326: 289–293
- 168 Mumbach M R, Rubin A J, Flynn R A, et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods*, 2016, 13: 919–922

- 169 Mifsud B, Tavares-Cadete F, Young A N, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet*, 2015, 47: 598–606
- 170 Chai H, Tjong H, Li P, et al. ChIATAC is an efficient strategy for multi-omics mapping of 3D epigenomes from low-cell inputs. *Nat Commun*, 2023, 14: 213
- 171 Zhao Z, Tavoosidana G, Sjölander M, et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet*, 2006, 38: 1341–1347
- 172 Cai Z, Cao C, Ji L, et al. RIC-seq for global in situ profiling of RNA-RNA spatial interactions. *Nature*, 2020, 582: 432–437
- 173 Akiki R M, Cornbrooks R G, Magami K, et al. A long noncoding eRNA forms R-loops to shape emotional experience-induced behavioral adaptation. *Science*, 2024, 386: 1282–1289
- 174 Core L J, Waterfall J J, Lis J T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, 2008, 322: 1845–1848
- 175 Kruesi W S, Core L J, Waters C T, et al. Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. *eLife*, 2013, 2: e00808
- 176 Kwak H, Fuda N J, Core L J, et al. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science*, 2013, 339: 950–953
- 177 Sabari B R, Dall'Agnese A, Boija A, et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science*, 2018, 361: eaar3958
- 178 Arrastia M V, Jachowicz J W, Ollikainen N, et al. Single-cell measurement of higher-order 3D genome organization with scSPRITE. *Nat Biotechnol*, 2022, 40: 64–73
- 179 Li W, Lu J, Lu P, et al. scNanoHi-C: a single-cell long-read concatemer sequencing method to reveal high-order chromatin structures within individual cells. *Nat Methods*, 2023, 20: 1493–1505
- 180 Chang L, Xie Y, Taylor B, et al. Droplet Hi-C enables scalable, single-cell profiling of chromatin architecture in heterogeneous tissues. *Nat Biotechnol*, 2024
- 181 Liu Z, Chen Y, Xia Q, et al. Linking genome structures to functions by simultaneous single-cell Hi-C and RNA-seq. *Science*, 2023, 380: 1070–1076
- 182 Qu J, Sun J, Zhao C, et al. Simultaneous profiling of chromatin architecture and transcription in single cells. *Nat Struct Mol Biol*, 2023, 30: 1393–1402
- 183 Wu H, Zhang J, Jian F, et al. Simultaneous single-cell three-dimensional genome and gene expression profiling uncovers dynamic enhancer connectivity underlying olfactory receptor choice. *Nat Methods*, 2024, 21: 974–982
- 184 Zhou T, Zhang R, Jia D, et al. GAGE-seq concurrently profiles multiscale 3D genome organization and gene expression in single cells. *Nat Genet*, 2024, 56: 1701–1711
- 185 Wu H, Wang M, Zheng Y, et al. Droplet-based high-throughput 3D genome structure mapping of single cells with simultaneous transcriptomics. *Cell Discov*, 2025, 11: 8
- 186 Wen X, Luo Z, Zhao W, et al. Single-cell multiplex chromatin and RNA interactions in ageing human brain. *Nature*, 2024, 628: 648–656
- 187 Mahat D B, Tippens N D, Martin-Rufino J D, et al. Single-cell nascent RNA sequencing unveils coordinated global transcription. *Nature*, 2024, 631: 216–223
- 188 Oguchi A, Suzuki A, Komatsu S, et al. An atlas of transcribed enhancers across helper T cell diversity for decoding human diseases. *Science*, 2024, 385: eadd8394
- 189 Liu M, Yue Y, Chen X, et al. Genome-coverage single-cell histone modifications for embryo lineage tracing. *Nature*, 2025, 640: 828–839
- 190 Qin F, Li B, Wang H, et al. Linking chromatin acylation mark-defined proteome and genome in living cells. *Cell*, 2023, 186: 1066–1085.e36
- 191 Chen W, Choi J, Li X, et al. Symbolic recording of signalling and cis-regulatory element activity to DNA. *Nature*, 2024, 632: 1073–1081
- 192 Olsen T R, Talla P, Sagatelian R K, et al. Scalable co-sequencing of RNA and DNA from individual nuclei. *Nat Methods*, 2025, 22: 477–487
- 193 Mimitou E P, Lareau C A, Chen K Y, et al. Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nat Biotechnol*, 2021, 39: 1246–1258
- 194 Izzo F, Myers R M, Ganesan S, et al. Mapping genotypes to chromatin accessibility profiles in single cells. *Nature*, 2024, 629: 1149–1157
- 195 Rampal R K, Grosicki S, Chraniuk D, et al. Pelabresib plus ruxolitinib for JAK inhibitor-naïve myelofibrosis: a randomized phase 3 trial. *Nat*

- Med, 2025
- 196 Liao M, Zhu X, Lu Y, et al. Multi-omics profiling of retinal pigment epithelium reveals enhancer-driven activation of RANK-NFATc1 signaling in traumatic proliferative vitreoretinopathy. *Nat Commun*, 2024, 15: 7324
- 197 Zheng Z, Xia L, Hu G, et al. Super-enhancer-controlled positive feedback loop BRD4/Era-RET-ER α promotes ER α -positive breast cancer. *Nucleic Acids Res*, 2022, 50: 10230–10248
- 198 Xiao L, Parolia A, Qiao Y, et al. Targeting SWI/SNF ATPases in enhancer-addicted prostate cancer. *Nature*, 2022, 601: 434–439
- 199 Welling D B, Collier K A, Burns S S, et al. Early phase clinical studies of AR-42, a histone deacetylase inhibitor, for neurofibromatosis type 2-associated vestibular schwannomas and meningiomas. *Laryngoscope Investig Oto*, 2021, 6: 1008–1019
- 200 Wong J, Gruber E, Maher B, et al. Integrated clinical and genomic evaluation of guadecitabine (SGI-110) in peripheral T-cell lymphoma. *Leukemia*, 2022, 36: 1654–1665
- 201 Ghia P, Scarfò L, Perez S, et al. Efficacy and safety of dinaciclib vs ofatumumab in patients with relapsed/refractory chronic lymphocytic leukemia. *Blood*, 2017, 129: 1876–1878
- 202 Chen H, Tu S, Yuan C, et al. HyperChIP: identification of hypervariable signals across ChIP-seq or ATAC-seq samples. *Genome Biol*, 2022, 23: 62
- 203 Huang J, Liu X, Li D, et al. Dynamic control of enhancer repertoires drives lineage and stage-specific transcription during hematopoiesis. *Dev Cell*, 2016, 36: 9–23
- 204 Kai Y, Li B E, Zhu M, et al. Mapping the evolving landscape of super-enhancers during cell differentiation. *Genome Biol*, 2021, 22: 269
- 205 Wang C, Zhang S, Chen L. Large-scale determination and characterization of cell type-specific regulatory elements in the human genome. *J Mol Cell Biol*, 2017, 9: 463–476
- 206 Avsec Ž, Agarwal V, Visentin D, et al. Effective gene expression prediction from sequence by integrating long-range interactions. *Nat Methods*, 2021, 18: 1196–1203
- 207 Gasperini M, Hill A J, McFaline-Figueroa J L, et al. A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell*, 2019, 176: 1516
- 208 Schraivogel D, Gschwind A R, Milbank J H, et al. Targeted Perturb-seq enables genome-scale genetic screens in single cells. *Nat Methods*, 2020, 17: 629–635
- 209 Shao B, Yan J. A long-context language model for deciphering and generating bacteriophage genomes. *Nat Commun*, 2024, 15: 9392
- 210 Dalla-Torre H, Gonzalez L, Mendoza-Revilla J, et al. Nucleotide transformer: building and evaluating robust foundation models for human genomics. *Nat Methods*, 2025, 22: 287–297
- 211 Benegas G, Batra S S, Song Y S. DNA language models are powerful predictors of genome-wide variant effects. *Proc Natl Acad Sci USA*, 2023, 120: e2311219120
- 212 Cao Z J, Gao G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nat Biotechnol*, 2022, 40: 1458–1466
- 213 Wang H, Nakamura M, Abbott T R, et al. CRISPR-mediated live imaging of genome editing and transcription. *Science*, 2019, 365: 1301–1305

Functional enhancers: concept, screening, and validation

LIN HongLi^{1†}, CHEN Fei^{1†} & HUANG JiaLiang^{1,2*}

¹ State Key Laboratory of Cellular Stress Biology, School of Life Sciences, Xiamen University, Xiamen 361102, China

² National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen 361102, China

† Contributed equally to this work

* Corresponding author, E-mail: jhuang@xmu.edu.cn

Enhancers are core regulatory elements in the genome that precisely modulate gene expression, shaping biological development and disease progression. Millions of putative enhancers have been nominated based on DNA sequence and epigenetic features, yet elucidating their functions and mechanisms remains a major challenge. Recent breakthroughs in technologies such as single-cell multi-omics, chromatin conformation capture, spatial omics, and genome editing have provided powerful tools to dissect enhancer function. In this review, we elaborate on the concept of functional enhancers—from their role in gene regulation to their impacts on cellular, tissue, and organismal phenotypes, as well as their contributions to disease. We systematically summarize the recent advances in high-throughput computational and experimental technologies that have significantly improved the identification and interpretation of enhancer-mediated regulatory networks in development and diseases. Finally, we propose a strategy for screening and validating functional enhancers from vast candidate sequences in the human genome and discuss future directions in enhancer research.

enhancers, functional enhancers, epigenetics, chromatin, gene expression

doi: [10.1360/SSV-2025-0113](https://doi.org/10.1360/SSV-2025-0113)