

增强子对基因表达的调控:从新技术到新机制

武成一, 陈 斐, 洪丹妮, 朱 明, 童梦莎, 黄佳良*

(厦门大学生命科学学院, 细胞应激生物学国家重点实验室, 细胞信号网络协同创新中心, 福建 厦门 361102)

摘要: 全基因组研究已经鉴定了人类基因组中数以百万计的增强子调控元件, 然而它们的生物学功能和潜在作用机制在很大程度上仍然未知. 本文回顾了应用于鉴定、描述和验证增强子新兴高通量技术, 并进一步讨论对增强子潜在调控机制的生物学理解, 特别是关于增强子簇如何控制基因表达的研究, 如超级增强子的功能层次结构等. 最后总结了增强子在发育、进化和疾病发生过程中调控基因时空表达的重要作用, 并对该领域一些未解决的关键问题和未来的研究方向进行探讨和展望.

关键词: 增强子; 基因表达调控; 高通量技术; 功能层次结构

中图分类号: Q 31; Q 33

文献标志码: A

文章编号: 0438-0479(2022)03-0496-10

人类基因组中仅约 2% 的序列是负责蛋白质编码的基因, 人们对于剩余大量的非编码序列的生物学功能依然了解甚少. 增强子调控元件是一段特定的非编码 DNA 序列, 它们通过与转录因子(TF)结合, 对细胞内基因表达进行顺式调控^[1]. 第一个被确定的增强子是猴空泡病毒 SV40 基因组中一段长 72 bp 的序列, 它可以将宫颈癌 HeLa 细胞中报告基因的转录提高数百倍^[1]. 增强子通常具有特定的染色质特征, 例如 DNA 序列保守性、染色质可及性、H3K27ac 和 H3K4me1 等组蛋白修饰、可与启动子形成染色质三维环状结构等^[2]. 随着“DNA 元件百科全书”(ENCODE)和“表观基因组学路线图”(Roadmap Epigenomics)等计划的实施, 研究人员找到大量(在每类细胞中超过 10 万个)增强子调控元件^[3-4]. 然而, 如何有效地注释这些增强子的生物学功能依然面临巨大挑战.

增强子增加了细胞内基因转录调控的复杂性, 在发育和疾病发生过程中扮演重要角色^[5]. 目前, 利用高通量研究方法研究增强子的主要工作包括以下几方面: 1) 增强子的预测, 利用 DNA 序列保守性、H3K27ac 和 H3K4me1 等组蛋白修饰、染色质可及性和 TF 结合图谱等多组学图谱, 预测具有细胞特异性

的潜在增强子. 例如哺乳动物胚胎着床前染色质可及性等动态图谱可用于预测这一生物过程中潜在的增强子^[6]. 2) 靶标基因的鉴定, 基于染色质线性距离^[7]、与基因表达相关性^[8]、染色质三维结构^[9]和表达数量性状位点^[10]等特征和技术手段, 寻找增强子潜在下游靶标基因. 3) 增强子对基因表达的影响, 利用聚簇规律间隔短回文重复序列及其相关蛋白 9 (CRISPR-Cas9) 基因编辑等方法, 检测增强子对下游靶标基因表达的影响. 例如, Fulco 等^[11]发现利用增强子自身 H3K27ac 强度、增强子与潜在靶标基因的距离、增强子与潜在靶标基因的染色质相互作用等因素进行建模, 可以准确预测出增强子对基因表达的影响. 4) 增强子与发育和疾病的关系, 例如 VISTA 数据库 (<https://enhancer.lbl.gov/>) 收集了 1 654 个经验证在小鼠发育与疾病模型中具有体内活性的增强子^[12], 发现增强子的冗余性有利于稳定哺乳动物发育过程中的表型^[13], 且大多数与疾病相关的序列变异(即单核苷酸多态性)富集在增强子等调控元件中^[14].

本文回顾近年来增强子调控基因表达研究中的部分最新进展, 介绍多组学、单细胞及基因编辑等可应用于增强子研究的相关技术, 阐述增强子对基因表达的调控机制, 最后对增强子在发育、进化和疾病中

收稿日期: 2021-11-01 录用日期: 2022-02-11

基金项目: 国家自然科学基金(32070635, 31871317); 福建省自然科学基金(2020J01028)

*通信作者: jhuang@xmu.edu.cn

引文格式: 武成一, 陈斐, 洪丹妮, 等. 增强子对基因表达的调控: 从新技术到新机制[J]. 厦门大学学报(自然科学版), 2022, 61(3): 496-505.

Citation: WU C Y, CHEN F, HONG D N, et al. Enhancer in regulation of gene expression: from new technology to new mechanism[J]. J Xiamen Univ Nat Sci, 2022, 61(3): 496-505. (in Chinese)



的重要功能进行概述。

1 增强子研究的新技术

1.1 多组学技术

1.1.1 染色质开放区域检测技术

染色质可及性是指细胞核内染色质中的 DNA 能够与生物大分子进行物理接触的程度,其图谱可用于识别基因组上的调控区域。染色质可及性以及核小体的定位传统上是通过核酸酶或限制性内切酶的消化等方法来检测的,这里简要描述几种常用测序(seq)实验技术的原理。1) DNase-seq^[15],通过一种可以优先在开放染色质中引入双链断裂的核酸内切酶 DNase I,消化基因组后对 DNA 片段进行测序。2) MNase-seq^[16],是一种测量核小体占位的技术,MNase 同时具备核酸外切酶和内切酶的活性,可以将碱基对逐个切下以切割和消化裸露的 DNA 区域,直至遇到核小体或 DNA 结合蛋白等阻滞物。3) ATAC-seq^[17],利用一种经过工程学改造的超高活性 DNA 转座酶(Tn5),同时切割和标记开放的或核小体缺失的染色质区域,具有实验方案简单和所需检测样品细胞少的优势,因此得到广泛应用。除了分析染色质可及性图谱之外,ATAC-seq 也可用于检测 TF 印迹和绘制核小体定位图。

1.1.2 TF 结合和组蛋白修饰检测技术

染色质可及性是高度动态变化的,通过染色质结合因子和核小体之间的相互竞争来调节。具有调控作用的 DNA 元件(如增强子、启动子、绝缘子等)与 TF 协同调节基因的表达,其中活性增强子附近的核小体通常含有特异组蛋白修饰。染色质免疫共沉淀(ChIP)-seq 用于在全基因组水平检测染色质结合因子如 TF 或组蛋白修饰在基因组上的位置^[18]。对于组蛋白修饰 ChIP-seq 实验,染色质可以从甲醛固定细胞或非固定细胞(天然染色质)中分离出来,并通过超声波或 MNase 消化作用将组蛋白修饰区域片段化至 100~500 bp。对于 TF 结合 ChIP-seq 实验,染色质需要通过甲醛固定交联以保留蛋白质与染色质的相互作用。这些 DNA 片段被特异性抗体以及 Protein A/G 偶联琼脂糖珠或磁珠捕获后进行建库和测序。近年来,研究人员已经开发出更多高效衍生方法,其中核酸酶靶向切割和释放(CUT&RUN)技术是一种以简单、可靠且成本较低的对染色质结合蛋白进行高分辨率分析的技术^[19]。此技术利用 Protein A/G-MNase 融合蛋白结合目标抗体,激活 MNase 对目标蛋白结

合位点周围的染色质进行切割、捕获。与传统的 ChIP-seq 相比,该技术不需要甲醛交联,具有更高的信噪比且所需的样品量少,已应用于单细胞水平的检测。Skene 等^[20]还推出了靶向剪切及转座酶(CUT&Tag)技术,与 CUT&RUN 相比,CUT&Tag 使用了 Tn5 代替 MNase。这两种方法低成本、高效率、高可重复性和高通量的特点,使其逐渐成为常规表观基因组分析的首选方法。

1.1.3 三维基因组技术

越来越多的证据表明,基因组的三维结构在不同尺度呈现层次结构,并在功能上与基因表达调控紧密关联^[21]。染色体在细胞核中占据的位置称为染色体区域,这些区域被划分为染色体室,并进一步划分为拓扑关联域(TAD),以及增强子-启动子相互作用介导的染色质环。在三维基因组结构研究中最有影响力的检测技术是高通量染色体构象捕获(Hi-C)技术^[22],该技术可对整个基因组的 DNA 相互作用进行分析,分辨率范围为数个碱基到数兆个碱基。根据 DNA 被甲醛交联在一起的概率,Hi-C 技术可以测量两个 DNA 片段在三维空间中物理结合的总体平均频率。利用限制性内切酶将交联的 DNA 消化并剪切,之后将相互作用的片段连接在一起并用生物素标记,从而对选择性纯化并形成嵌合 DNA 分子的基因组文库进行测序。Hi-C 的衍生技术可以检测由感兴趣的蛋白质所驱动或者所选定基因组位置的特异性 DNA 相互作用,包括 ChIP-loop、ChIA-PET、HiChIP、PLAC-seq 和 Capture-Hi-C 等^[23]。Hi-C 及其衍生技术的分辨率很大程度上受限于限制性内切酶位点的分布。Micro-C 技术引入了双重交联,并用 MNase 消化取代了 Hi-C 技术中使用的限制性内切酶,这不仅能产生非常均匀的切割片段从而提高局部分辨率,还可以保留核小体定位的信息^[24]。也有研究发现增强子和启动子的功能连接性可以通过它们自身转录的 RNA 相互作用来实现^[25]。RNA 原位构象测序(RIC-seq)技术可以用于对分子内和分子间的 RNA-RNA 相互作用进行全局分析,有助于生成人类细胞中 RNA 的三维相互作用图谱^[26]。

目前利用 Hi-C 等技术检测全基因组染色质三维结构图谱在技术上依然存在困难,而且非常昂贵,因此刺激了计算分析和数学建模的发展,结合验证实验,有助于实现对染色质结构和功能的理解。Huang 等^[27]构建贝叶斯累加回归树 BART 计算模型,在概念上首次论证了利用组蛋白修饰数据可以有效地预测染色质三维结构。Marco 等^[28]开发了层次化隐马尔

可夫模型的计算方法,能够跨不同尺度系统地注释染色质状态,对增强子进行功能注释,从而更好地筛选出细胞中关键的增强子。

1.2 单细胞组学技术

1.2.1 单细胞多组学技术

在过去很长一段时间内, RNA-seq 技术已被广泛用于研究群体水平的基因表达模式. 单细胞 RNA-seq (scRNA-seq)^[29] 的出现为在单细胞水平上探索基因表达谱提供了前所未有的机会, 并为研究细胞的基因表达异质性和动态变化提供更加深入的见解. 单细胞 ATAC-seq(scATAC-seq)可以在单细胞分辨率下检测染色质可及性,从而帮助人们了解细胞间的谱系关系^[30]. 除此之外,研究人员还开发了许多方法来同时获得单细胞的染色质可及性和基因表达水平,例如 Paired-seq^[31]、SNARE-seq^[32]和 SHARE-seq^[33]等。

1.2.2 单细胞多组学图谱

Li 等^[34]利用 scATAC-seq 捕获了成年小鼠同形皮层、嗅球、海马以及脑核区 45 个区域共计超过 80 万个细胞的染色质可及性,构建了超过 160 多种细胞类型的顺式作用元件调控图谱,为全面分析哺乳动物大脑的基因调控程序奠定了基础,有助于解释人类各种神经系统疾病. Domcke 等^[35]对发育时间 72~129 d 人类胎儿的 15 个器官 121 个样本进行 scRNA-seq 和 scATAC-seq,获得了 400 万个单细胞的基因表达和染色质可及性数据,为更好地理解人类发育中的罕见和常见疾病提供了信息资源,同时也为开发潜在的治疗方法提供了支持. Ranzoni 等^[36]对 8 000 多个来自胎儿肝脏和骨髓的血细胞进行 scRNA-seq 和 scATAC-seq,确定了肝脏和骨髓的造血干细胞之间的转录和功能差异. Trevino 等^[37]等对不同时期的人类胎儿大脑皮层组织进行 scRNA-seq 和 scATAC-seq,通过整合分析,完成了人大脑皮层发生过程中各细胞类型的单细胞基因表达和调控图谱,发现了神经元和胶质细胞发育分化的路径以及不同的 TF 在这些过程中的调控机制. 这些单细胞多组学图谱有助于在单细胞分辨率下进一步揭示发育或疾病过程中细胞命运调控的机制。

1.3 基因编辑技术

1.3.1 CRISPR-Cas9 基因编辑技术

随着 CRISPR 和 Cas 的发现,CRISPR-Cas 工具也逐渐成为各个领域广泛使用的基因编辑技术. II 型 CRISPR-Cas 系统由单个导向 RNA (sgRNA) 和 DNA

内切酶 Cas9 组成,sgRNA 引导 Cas9 到特定的 DNA 序列并进行切割,从而实现便捷有效的基因编辑^[38-39]. 此外,由 Cas9 两个核酸酶结构域内残基的突变体产生的失活 Cas9 (dCas9),虽然不能切割 DNA,但是保留了以特定序列方式结合 DNA 的能力. dCas9 可将转录激活因子招募到转录起始位点,促进基因转录激活 (CRISPRa)^[40],也可以结合转录抑制因子 Kruppel 相关盒 (KRAB) 特异性地抑制基因表达 (CRISPRi)^[41]. 此外还有与各种效应域融合的 dCas9,大大方便了对真核生物转录组和表观基因组的操纵. 例如,原位 CRISPR 亲和纯化调节元件 (CAPTURE) 方法可以识别位点特异性染色质调控复合物和远程 DNA 相互作用^[42],使用融合生物素的 dCas9,可以与链霉素结合从而分离完整的调控复合物. 在非编码调控元件的功能研究上, Li 等^[43]设计了以增强子为靶标的体内 CRISPR 双效应表观遗传编辑系统 enCRISPRa 和 enCRISPRi,用于在体外、异种移植和体内高效分析增强子功能. 在染色质空间构象与疾病的研究中, Lupiáñez 等^[44]利用 CRISPR-Cas 系统构建了特定基因位点染色体重排的小鼠模型,发现染色体重排破坏了编码基因及其增强子相对于 TAD 边界的正常拓扑结构,导致远端调控结构重新连接而致病。

1.3.2 CRISPR-Cas9 文库筛选技术

传统的利用 RNA 干扰途径进行基因敲除和大规模筛选的方式,存在广泛脱靶效应和基因敲除不完全等问题. 而 CRISPR-Cas9 为高通量筛选提供了更高效便捷的平台,通过设计 sgRNA 文库,可以根据特定的表型筛选重要基因,成为全基因组功能分析的强有力工具^[45]. 新开发的 Perturb-seq 将 CRISPR-Cas9 与 scRNA-seq 结合起来,在单个样品中平行生成数千个基因扰动,并测定每个细胞的扰动和表型,从而准确鉴定受扰动影响的单个基因靶点、基因特征和细胞状态,实现了大规模的基因组功能筛查^[46]. 基于 CRISPR-Cas9 系统的全基因组筛选技术对筛选疾病发生中起重要作用的非编码调控元件具有重要意义^[47]. 为了有效地预测增强子与基因之间的连接,研究人员将 CRISPRi、RNA 荧光原位杂交 (FISH) 和流式细胞术结合起来开发了 CRISPRi-FlowFISH,该方法通过扰动非编码元件并量化其对基因表达的影响,可以系统地大规模预测增强子调控基因^[11]. 此外,新开发的 CRISPRpath 方法^[48]将传统的仅限单个基因的增强子筛选方法进一步扩展,通过对人类诱导多能干细胞中的 6 个基因进行干扰和切割,可以同时表征

参与同一生物学途径中多个靶基因的调控元件,并区分出不同增强子在功能强度上的差异. Fei 等^[49]分别设计了针对叉头框蛋白 A1(FOXA1)或 CCCTC 结合因子(CTCF)结合位点的全基因组 CRISPR-Cas9 敲

除筛选文库,并在乳腺癌和前列腺癌细胞系中进行大规模筛选,筛选并探究超过 1 万个与 FOXA1 或 CTCF 结合的顺式作用元件的功能及特征. 增强子研究中的相关技术总结于表 1.

表 1 增强子研究中的相关技术

Tab. 1 Related technologies of enhancer research

技术	类型	方法	特点
多组学技术	染色质开放区域检测	DNase-seq ^[15]	检测染色质开放区域,也可以推测核小体位置,需要样品数较多
		MNase-seq ^[16]	测量核小体占位,需要样品数较多
		ATAC-seq ^[17]	检测染色质开放区域,实验简单,需要样品数较少
	TF 结合和组蛋白修饰检测	ChIP-seq ^[18]	利用甲醛固定细胞,无偏检测结合在 DNA 序列上的靶蛋白
		CUT&RUN ^[19]	无需甲醛交联,所需样品较少,实验重复性高且信噪比低
		CUT&Tag ^[20]	与 CUT&RUN 相比,使用 Tn5 同时切割和标记开放染色质,简化了实验步骤
三维基因组	Hi-C ^[22]	检测全基因组染色体片段间的交互作用,分辨率受限于限制性内切酶位点的分布	
	Micro-C ^[24]	MNase 取代了 Hi-C 中的限制性内切酶,产生均匀的 DNA 切割片段,提高分辨率	
	RIC-seq ^[26]	检测全基因组范围内的 RNA-RNA 相互作用	
单细胞组学技术	转录组基因组	scRNA-seq ^[29]	在单细胞水平对全转录组进行扩增与测序
		scATAC-seq ^[30]	在单细胞水平对细胞染色质开放区域进行检测
基因编辑技术	CRISPR-Cas9 基因编辑	CRISPR-Cas9 ^[39]	由 sgRNA 引导的序列特异性 DNA 双链切割,编辑效率高,操作便捷
		CRISPRa ^[40]	将 dCas9 与转录激活因子融合,位点特异性地激活基因转录
		CRISPRi ^[41]	将 dCas9 与转录抑制因子 KRAB 融合,在空间上干扰转录酶结合并抑制基因表达
	CRISPR-Cas9 文库筛选	CRISPR-Cas9 screen ^[45] Perturb-seq ^[46]	利用 CRISPR-Cas9 基因敲除文库进行高通量的全基因组筛选结合 CRISPR-Cas9 与 scRNA-seq,在单个样品中平行生成数千个基因扰动
		CRISPRi-FlowFISH ^[11]	结合 CRISPRi、RNA-FISH 和流式细胞术,对非编码元件进行扰动

2 增强子相关新机制

2.1 增强子靶标基因的预测方法

对增强子位置的识别只是解析其生物学功能的第一步. 在此基础上,鉴定出每一个增强子直接作用的基因靶点,是理解增强子生物学效应不可或缺的部分. Hi-C 技术直观地反映了有物理接触的基因组位置信息,特别是利用 Promoter Capture Hi-C 特异性捕获增强子和启动子相互作用的片段,可以达到更高的分辨率^[50]. 此外,利用多个样本 ATAC-seq 或 H3K27ac 信号和基因表达量相关性的方法也被广泛应用于预

测增强子的靶标基因^[51],甚至应用于单细胞水平^[8]. 还有研究人员设计了一种称为 ABC (activity-by-contact)的方法构建了人类全基因组增强子-靶基因映射图谱^[11],此方法综合考虑了增强子本身的表观修饰强度以及与靶基因三维连接的强度,已应用于对全基因组关联研究位点的功能性注释中^[52].

2.2 增强子簇

许多增强子于基因组中协同调节同一个靶基因,并以簇的形式存在. 2008 年“阴影增强子”的概念被提出^[53],阴影增强子代表在空间和时间上表现出部分或完全冗余的调控模式的多个增强子. 研究表明阴影增强子可能提供了一种重要的机制,用于缓冲与人类疾

病相关的基因非编码调控区突变而导致的基因表达紊乱,提高动物发育过程中基因表达的精确性和表型稳健性^[13,54].增强子簇中通常具有异常高水平的增强子活性和增强子相关染色质特征,如 H3K27ac 的富集,TF、辅助激活因子和染色质的结合与占用等,因此在研究中也鉴定为超级增强子(SE)^[55].SE 通常与细胞命运决定和肿瘤发生发展过程密切相关.目前 SE 是利用生物信息学方法通过对基因组中距离在 12.5 kb 以内的增强子进行线性聚类而识别的.

2.3 增强子之间的层次结构

增强子簇在细胞命运决定和疾病过程中扮演重要角色,然而它是如何发挥作用的尚不清楚,其中一个关键问题是,增强子簇内多个增强子元件是如何相互协调以实现目标基因表达的精确调控的^[56].大量研究证明增强子的调控机制具有高度的多样性和异质性.例如:Hay 等^[57]对 α -珠蛋白增强子元件的解析表明, α -珠蛋白基因的活性随着增强子的数量增加呈线性增加的状态;Bahr 等^[58]证明了调节 MYC 基因的增强子簇的加和性协同作用;Shin 等^[59]发现乳清酸蛋白基因(*Wap*)增强子簇中最远端的增强子似乎起到最关键的作用.增强子在基因激活方面也表现出时间上的差异,小鼠 *Fgf5* 增强子簇中的大多数增强子元件会在不同时间点促进 *Fgf5* 的表达,从而诱导小鼠干细胞的分化^[60].也有研究提出动态“转录枢纽”的概念,即多个增强子及其目标启动子、大量 TF、转录辅助因子和 RNA 聚合酶 II 同时参与相同的微环境,形成一个转录共激活聚合物^[61],为多个增强子对启动子的调控提供潜在的模型.

Huang 等^[62]利用 CRISPR-Cas9 基因编辑技术,在造血干细胞体外分化为成红细胞的模型中,解析了两个 SE:1)造血干细胞特异性转录因子 GATA2 基因的 SE 通过 3 个增强子不同分工,精确调控 GATA2 在分化过程中的动态表达,从而影响干细胞的维持和分化;2) *SCL25A37* SE 中,增强子 E3 不仅对基因表达影响最大,而且还决定了 E1 和 E2 的功能,表明增强子在功能上存在层次结构.Huang 等^[63]还基于染色质三维结构和组蛋白修饰等表观组学图谱,在人白血病细胞 K562 中系统解析了全基因组范围内的 SE,发现在 SE 内部存在一个重要的枢纽增强子,负责协调多个增强子共同调控基因表达,由此提出一个包含枢纽增强子的模型,阐述增强子之间功能层次结构的分子机制.Kai 等^[64]的最近一项研究还发现增强子在形成时间上同样存在层次结构,增强子簇中的元件在分化过程中的不同时间点出现,通过与不同的细胞特异

性 TF 结合来驱动细胞的分化过程.

3 增强子调控基因表达的功能

3.1 发育

单细胞受精卵通过一系列发育途径的选择过程发展成为一种复杂的多细胞动物.发育过程中增强子可以精确控制基因的时空表达,使得基因在物种特异性遗传程序的控制下表现出时空受限的表达模式,调控从生殖层形成到细胞分化的发育过程.调节细胞特异性基因的元件中,最典型的例子是基因座控制区(LCR)^[65],LCR 通常是基因组上具有调节能力的 DNA 序列,比如人类 β -珠蛋白 LCR 由多个红系特异性的增强子组成,可以在不同发育时期通过染色质环状结构与珠蛋白基因结合来促进基因特异性的表达.关于增强子在肢体发育中的作用也开展了大量研究,这里以 *Shh* 和 *HoxD* 基因位点为例.远端骨骼的发育模式部分由 *Shh* 基因所控制,而 *Shh* 基因的表达由一个被称为“极性活化带调控元件”(ZRS)的增强子驱动^[66],ZRS 可以与 TF ETS1 结合,从而驱动 *Shh* 的转录.在 ZRS 缺失的突变体中,*Shh* 转录信号消失,导致小鼠远端骨骼的缺失,这表明 ZRS 对于远端肢体发育具有关键作用.*HoxD* 在四足动物四肢的构建中起着至关重要的作用,控制 *HoxD* 的增强子多达 7 个,其中位于 3'-端的两个增强子控制 *HoxD* 表达的早期阶段,随后染色质发生拓扑变化,使得部分增强子与 5'-端增强子一起调控 *HoxD* 后期表达^[67].近期一些研究还探究了增强子在血液细胞发育过程中的作用.Huang 等^[62]利用人类造血干细胞体外分化模型,结合 RNA-seq 和 ChIP-seq 等技术检测细胞的转录组和表观基因组,利用 MAnorm^[68] 分析了增强子上 H3K27ac 组蛋白修饰信号的差异,系统绘制了人类造血干细胞分化过程中增强子激活或失活的动态图谱,发现不同的 TF 组合以高度特定的方式合作调控增强子的时空活动,比如造血干细胞和祖细胞(HSPC)特有的 RUNX1-FLI1-PU.1-ETS 模块,血液细胞中的 GATA1 和 TAL1 模块,以及无处不在的“管家”E2F-SP1 和激活蛋白 1(AP1)模块等,进一步阐述了增强子在人类造血干细胞分化过程中的调控网络.Cai 等^[69]还发现分化后期的红细胞中细胞特异性基因表达依赖于远端增强子的程度比在胚胎时期红细胞中的更强.这些特异性基因主要通过远端增强子所结合的 Myb 以及 Myb 介导的 Gata1 占位来调控,而胚胎时期红细胞特异性基因主要通过 Gata1 在近端启动

子区域的占用来调节,说明细胞特异性调节模式随个体发育而改变。

3.2 进化

除在发育调控中的核心作用外,增强子还是产生进化变异的热点^[70],因为它们拥有细胞特异性,允许在特定组织环境中调节靶基因表达,而不影响其他基因功能,还可以通过冗余性来缓冲致死风险从而促进遗传变异的积累。在哺乳动物基因组中包含数百个连续的保守碱基序列,可以作为远距离增强子发挥作用,称为“超保守”增强子。这些超保守调控元件对维持正常发育过程有着不可或缺的作用^[71]。TF 和增强子的相互作用在动物发育过程中决定了细胞特性,尽管 TF 家族在动物界高度保守,但控制基因表达的增强子在很大程度上具有物种特异性^[72],在哺乳动物基因组中,只有 1% 的人类组织特异性增强子是保守的^[72]。在更大进化距离上有相似性的增强子更为罕见,即使是超保守增强子,其功能也并不意味着需要在进化中通过序列的完美保守来维持^[73]。虽然泛后生动物中增强子保守性似乎很少见,但有研究发现增强子存在一种古老、保守但灵活的调控机制^[74]。人类与海绵动物的胰岛增强子没有显著的序列一致性,但它们通过一套相似的 TF 结合基序来驱动相似的基因表达模式。这种保守的 TF 结合基序在进化过程中保持了稳定的增强子功能,同时这些增强子可以通过扩展整合新的 TF 结合基序以及丢掉其他基序来完成进化。

3.3 疾病

随着研究的不断深入,发现 SE 驱动基因转录调控是一种参与癌症发生发展的独特转录调控机制,在乳腺癌^[75]、结直肠癌^[76]和急性淋巴细胞白血病^[77]等多种癌症中驱动或抑制癌症的发生,对肿瘤细胞的生长、耐药性以及免疫逃逸等能力产生重要的影响。SE 介导癌症发生的机制主要有以下几种:1) 基因组上的碱基突变导致新的 TF 结合位点出现,进而出现新的 SE,例如在一部分 T 细胞急性淋巴细胞白血病(T-ALL)病例中,由于 *TAL1* 上游非编码区域发生突变,产生了新的 MYB 结合位点,MYB 蛋白结合在此新位点上并招募组蛋白乙酰转移酶(CBP)形成新的 SE,激活相邻的原癌基因 *TAL1* 的表达,导致肿瘤的发生^[77];2) SE 中某些关键位点的突变导致致癌基因或抑癌基因的表达异常,从而影响癌症的发生和发展,例如位于神经母细胞瘤致癌基因 *LMO1* 内的 SE 中的单碱基突变 G > T,使得原有的 TF GATA3 的结合

位点被破坏,导致 SE 驱动的 *LMO1* 表达下降,影响了神经母细胞瘤的易感性^[78];3) SE 通常调控其所在 TAD 内的基因,由碱基突变等原因导致 TAD 边界发生改变,造成染色质结构的重组,使得原本不受 SE 调控的基因出现异常表达,例如在一些 T-ALL 病例中,TAD 边界发生突变导致 SE 异常驱动原癌基因 *TAL1* 和 *LMO2* 的表达^[79]。SE 功能层级的紊乱通常也伴随着基因的表达异常。例如研究发现位于 *MYC* 基因下游 1.7 Mb 的血液增强子簇(BENC)在 *MYC* 基因的表达调控中发挥着重要作用,BENC 是一段由 9 个具有选择性活性的增强子组成的 SE,在不同的造血系统中每个增强子的活性不同,从而使多个增强子能够形成不同的组合,实现谱系特异性的基因调控,而 BENC 的调控异常使 *MYC* 基因的表达失调,导致白血病的发生^[58]。

近年来针对各类疾病的基因编辑疗法正在快速发展,其中靶向增强子位点的基因编辑疗法已被成功应用到镰刀型贫血症和 β 地中海贫血症中,这类疾病是由于编码 β -珠蛋白的基因突变使其无法与 α -珠蛋白形成正常的血红蛋白 HbA 导致的。利用 CRISPR-Cas9 对患者自身 CD34⁺ 细胞中 *BCL11A* 增强子上的 GATA1 结合位点进行切割,由于增强子的时空特异性,对该增强子的切割不影响其他细胞类型中 *BCL11A* 的功能,仅特异性地抑制了红系细胞中 *BCL11A* 的表达,恢复了出生后被抑制的 γ -珠蛋白的表达,与 α -珠蛋白重新形成胎儿血红蛋白 HbF^[47,80]。经靶向增强子编辑后的 CD34⁺ 细胞回输到患者体内后,能有效恢复患者体内的 HbF 水平,使患者不再需要输血治疗^[81]。相比于靶向基因序列,对增强子进行编辑改变了对基因的表达调控,且没有造成基因突变,增强子的时空特异性也大大降低了基因编辑的风险,使其成为各类疾病中都具有巨大潜力的基因编辑疗法新靶点。

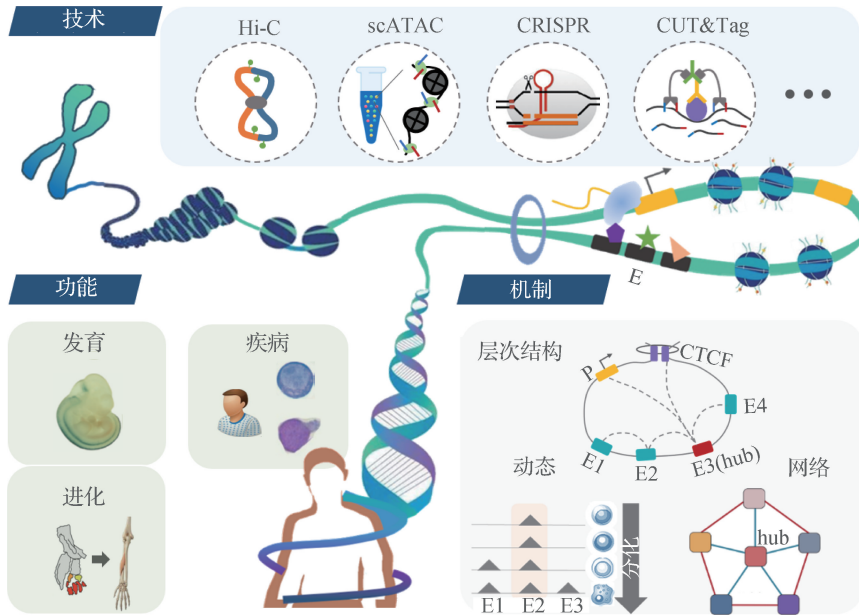
4 总结与展望

增强子是基因转录的关键调节元件,研究人员在其效应和机制研究方面取得了巨大进展,它们的功能与 TF、辅助因子的募集以及增强子与启动子相互作用的形成密切相关。近年来,二代测序的发展也极大地扩展了人们探索全基因组组成的知识和技能。本文从不同方面回顾了增强子的历史、研究技术、重要性、机制和研究进展(图 1)。目前,如何有效注释基因组中大量增强子的生物学功能及其作用机制是该领域的

研究热点.

目前仍有许多关于增强子的未解之谜,如:增强子是如何选择目标基因的?多个增强子如何在调节网络中协调,这些增强子中的冗余有多广泛?增强子与启动子活性的机制差异是什么?三维环状染色质

结构是否是基因调控的决定因素?增强子激活靶基因启动子的确切机制是什么?研究人员需要在开发新方法和积累不同细胞类型和组织的数据方面做出更多努力,为该领域的研究提供更多的线索,以期进一步了解增强子在基因调控中的作用和机制.



P 表示启动子;E 表示增强子;hub 表示枢纽增强子.

图 1 增强子对基因表达调控的相关技术、机制与功能

Fig. 1 Related technology, mechanism and function of enhancer in regulation of gene expression

参考文献:

[1] BANERJI J, RUSCONI S, SCHAFFNER W. Expression of a β -globin gene is enhanced by remote SV40 DNA sequences[J]. Cell, 1981, 27(2): 299-308.

[2] GASPERINI M, TOME J M, SHENDURE J. Towards a comprehensive catalogue of validated and target-linked human enhancers[J]. Nat Rev Genet, 2020, 21(5): 292-310.

[3] The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome[J]. Nature, 2012, 489(7414): 57-74.

[4] BERNSTEIN B E, STAMATOYANNOPOULOS J A, COSTELLO J F, et al. The NIH roadmap epigenomics mapping consortium[J]. Nat Biotechnol, 2010, 28(10): 1045-1048.

[5] SHLYUEVA D, STAMPFEL G, STARK A. Transcriptional enhancers: from properties to genome-wide predictions [J]. Nat Rev Genet, 2014, 15(4): 272-286.

[6] WU J, XU J, LIU B, et al. Chromatin analysis in human early development reveals epigenetic transition during

ZGA[J]. Nature, 2018, 557(7704): 256-260.

[7] MOORE J E, PRATT H E, PURCARO M J, et al. A curated benchmark of enhancer-gene interactions for evaluating enhancer-target gene prediction methods[J]. Genome Biology, 2020, 21(1): 17.

[8] PLINER H A, PACKER J S, MCFALINE-FIGUEROA J L, et al. Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data[J]. Molecular Cell, 2018, 71(5): 858-871.

[9] RAO S S P, HUNTLEY M H, DURAND N C, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping[J]. Cell, 2014, 159(7): 1665-1680.

[10] GTEx Consortium. Genetic effects on gene expression across human tissues[J]. Nature, 2017, 550(7675): 204-213.

[11] FULCO C P, NASSER J, JONES T R, et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations[J]. Nat Genet, 2019, 51(12): 1664-1669.

[12] VISEL A, MINOVITSKY S, DUBCHAK I, et al. VISTA

- enhancer browser; a database of tissue-specific human enhancers[J]. *Nucleic Acids Res*, 2007, 35(Sup. 1):D88-D92.
- [13] OSTERWALDER M, BAROZZI I, TISSIÈRES V, et al. Enhancer redundancy provides phenotypic robustness in mammalian development[J]. *Nature*, 2018, 554(7691): 239-243.
- [14] MAURANO M T, HUMBERT R, RYNES E, et al. Systematic localization of common disease-associated variation in regulatory DNA[J]. *Science*, 2012, 337(6099):1190-1195.
- [15] BOYLE A P, DAVIS S, SHULHA H P, et al. High-resolution mapping and characterization of open chromatin across the genome[J]. *Cell*, 2008, 132(2): 311-322.
- [16] SCHONES D E, CUI K R, CUDDAPAH S, et al. Dynamic regulation of nucleosome positioning in the human genome[J]. *Cell*, 2008, 132(5):887-898.
- [17] BUENROSTRO J D, GIRESI P G, ZABA L C, et al. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position[J]. *Nature Methods*, 2013, 10(12):1213-1218.
- [18] BARSKI A, CUDDAPAH S, CUI K R, et al. High-resolution profiling of histone methylations in the human genome[J]. *Cell*, 2007, 129(4):823-837.
- [19] SKENE P J, HENIKOFF S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites[J]. *eLife*, 2017, 6:e21856.
- [20] SKENE P J, HENIKOFF J G, HENIKOFF S. Targeted *in situ* genome-wide profiling with high efficiency for low cell numbers[J]. *Nat Protoc*, 2018, 13(5):1006-1019.
- [21] LUO X, LIU Y T, DANG D C, et al. 3D genome of macaque fetal brain reveals evolutionary innovations during primate corticogenesis[J]. *Cell*, 2021, 184(3): 723-740.
- [22] LIEBERMAN-AIDEN E, VAN BERKUM N L, WILLIAMS L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome[J]. *Science*, 2009, 326(5950):289-293.
- [23] KEMPFER R, POMBO A. Methods for mapping 3D chromosome architecture[J]. *Nat Rev Genet*, 2020, 21(4):207-226.
- [24] HSIEH T H S, CATTOGLIO C, SLOBODYANYUK E, et al. Resolving the 3D landscape of transcription-linked mammalian chromatin folding[J]. *Molecular Cell*, 2020, 78(3):539-553.
- [25] SARTORELLI V, LAUBERTH S M. Enhancer RNAs are an important regulatory layer of the epigenome[J]. *Nat Struct Mol Biol*, 2020, 27(6):521-528.
- [26] CAI Z K, CAO C C, JI L, et al. RIC-seq for global *in situ* profiling of RNA-RNA spatial interactions[J]. *Nature*, 2020, 582(7812):432-437.
- [27] HUANG J, MARCO E, PINELLO L, et al. Predicting chromatin organization using histone marks[J]. *Genome Biology*, 2015, 16:162.
- [28] MARCO E, MEULEMAN W, HUANG J, et al. Multi-scale chromatin state annotation using a hierarchical hidden Markov model[J]. *Nat Commun*, 2017, 8:15011.
- [29] TANG F, BARBACIORU C, WANG Y, et al. mRNA-Seq whole-transcriptome analysis of a single cell[J]. *Nature Methods*, 2009, 6(5):377-382.
- [30] BUENROSTRO J D, WU B, LITZENBURGER U M, et al. Single-cell chromatin accessibility reveals principles of regulatory variation[J]. *Nature*, 2015, 523(7561): 486-490.
- [31] ZHU C, YU M, HUANG H, et al. An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome[J]. *Nat Struct Mol Biol*, 2019, 26(11):1063-1070.
- [32] CHEN S, LAKE B B, ZHANG K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell[J]. *Nat Biotechnol*, 2019, 37(12):1452-1457.
- [33] MA S, ZHANG B, LAFAVE L M, et al. Chromatin potential identified by shared single-cell profiling of RNA and chromatin[J]. *Cell*, 2020, 183(4):1103-1116.
- [34] LI Y E, PREISSL S, HOU X, et al. An atlas of gene regulatory elements in adult mouse cerebrum [J]. *Nature*, 2021, 598(7879):129-136.
- [35] DOMCKE S, HILL A J, DAZA R M, et al. A human cell atlas of fetal chromatin accessibility[J]. *Science*, 2020, 370(6518):eaba7612.
- [36] RANZONI A M, TANGHERLONI A, BEREST I, et al. Integrative single-cell RNA-seq and ATAC-seq analysis of human developmental hematopoiesis [J]. *Cell Stem Cell*, 2021, 28(3):472-487.
- [37] TREVINO A E, MÜLLER F, ANDERSEN J, et al. Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution[J]. *Cell*, 2021, 184(19):5053-5069.
- [38] JINEK M, CHYLINSKI K, FONFARA I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity[J]. *Science*, 2012, 337(6096):816-821.

- [39] CONG L, RAN F A, COX D, et al. Multiplex genome engineering using CRISPR/Cas systems[J]. *Science*, 2013, 339(6121): 819-823.
- [40] MALI P, AACH J, STRANGES P B, et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering [J]. *Nat Biotechnol*, 2013, 31(9): 833-838.
- [41] GILBERT L A, HORLBECK M A, ADAMSON B, et al. Genome-scale CRISPR-mediated control of gene repression and activation[J]. *Cell*, 2014, 159(3): 647-661.
- [42] LIU X, ZHANG Y, CHEN Y, et al. *In situ* capture of chromatin interactions by biotinylated dCas9 [J]. *Cell*, 2017, 170(5): 1028-1043.
- [43] LI K L, LIU Y X, CAO H, et al. Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing[J]. *Nat Commun*, 2020, 11(1): 485.
- [44] LUPIÁÑEZ D G, KRAFT K, HEINRICH V, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions [J]. *Cell*, 2015, 161(5): 1012-1025.
- [45] SHALEM O, SANJANA N E, HARTENIAN E, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells[J]. *Science*, 2014, 343(6166): 84-87.
- [46] DIXIT A, PARNAS O, LI B Y, et al. Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens [J]. *Cell*, 2016, 167(7): 1853-1866.
- [47] CANVER M C, SMITH E C, SHER F, et al. BCL11A enhancer dissection by Cas9-mediated *in situ* saturating mutagenesis[J]. *Nature*, 2015, 527(7577): 192-197.
- [48] REN X J, WANG M C, LI B K, et al. Parallel characterization of *cis*-regulatory elements for multiple genes using CRISPRpath [J]. *Sci Adv*, 2021, 7(38): eabi4360.
- [49] FEI T, LI W, PENG J Y, et al. Deciphering essential cistromes using genome-wide CRISPR screens [J]. *Proc Natl Acad Sci USA*, 2019, 116(50): 25186-25195.
- [50] DRYDEN N H, BROOME L R, DUDBRIDGE F, et al. Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C [J]. *Genome Research*, 2014, 24(11): 1854-1868.
- [51] CORCES M R, GRANJA J M, SHAMS S, et al. The chromatin accessibility landscape of primary human cancers [J]. *Science*, 2018, 362(6413): eaav1898.
- [52] NASSER J, BERGMAN D T, FULCO C P, et al. Genome-wide enhancer maps link risk variants to disease genes [J]. *Nature*, 2021, 593(7858): 238-243.
- [53] HONG J W, HENDRIX D A, LEVINE M S. Shadow enhancers as a source of evolutionary novelty [J]. *Science*, 2008, 321(5894): 1314.
- [54] KVON E Z, WAYMACK R, GAD M, et al. Enhancer redundancy in development and disease [J]. *Nat Rev Genet*, 2021, 22(5): 324-336.
- [55] HNISZ D, ABRAHAM B J, LEE T I, et al. Super-enhancers in the control of cell identity and disease [J]. *Cell*, 2013, 155(4): 934-947.
- [56] BLOBEL G A, HIGGS D R, MITCHELL J A, et al. Testing the super-enhancer concept [J]. *Nat Rev Genet*, 2021, 22(12): 749-755.
- [57] HAY D, HUGHES J R, BABBS C, et al. Genetic dissection of the α -globin super-enhancer *in vivo* [J]. *Nat Genet*, 2016, 48(8): 895-903.
- [58] BAHR C, VON PALESKE L, USLU V V, et al. A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies [J]. *Nature*, 2018, 553(7689): 515-520.
- [59] SHIN H Y, WILLI M, YOO K H, et al. Hierarchy within the mammary STAT5-driven *Wap* super-enhancer [J]. *Nat Genet*, 2016, 48(8): 904-911.
- [60] THOMAS H F, KOTOVA E, JAYARAM S, et al. Temporal dissection of an enhancer cluster reveals distinct temporal and functional contributions of individual elements [J]. *Molecular Cell*, 2021, 81(5): 969-982.
- [61] TSAI A, ALVES M R, CROCKER J. Multi-enhancer transcriptional hubs confer phenotypic robustness [J]. *eLife*, 2019, 8: e45325.
- [62] HUANG J, LIU X, LI D, et al. Dynamic control of enhancer repertoires drives lineage and stage-specific transcription during hematopoiesis [J]. *Dev Cell*, 2016, 36(1): 9-23.
- [63] HUANG J L, LI K, CAI W, et al. Dissecting super-enhancer hierarchy based on chromatin interactions [J]. *Nat Commun*, 2018, 9(1): 943.
- [64] KAI Y, LI B E, ZHU M, et al. Mapping the evolving landscape of super-enhancers during cell differentiation [J]. *Genome Biology*, 2021, 22(1): 269.
- [65] LI Q L, PETERSON K R, FANG X D, et al. Locus control regions [J]. *Blood*, 2002, 100(9): 3077-3086.
- [66] KVON E Z, KAMNEVA O K, MELO U S, et al. Progressive loss of function in a limb enhancer during snake evolution [J]. *Cell*, 2016, 167(3): 633-642.
- [67] ANDREY G, MONTAVON T, MASCREZ B, et al. A switch between topological domains underlies *HoxD* genes collinearity in mouse limbs [J]. *Science*, 2013, 340

- (6137):1234167.
- [68] SHAO Z, ZHANG Y J, YUAN G C, et al. MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets[J]. *Genome Biology*, 2012, 13(3):R16.
- [69] CAI W, HUANG J, ZHU Q, et al. Enhancer dependence of cell-type-specific gene expression increases with developmental age[J]. *Proc Natl Acad Sci USA*, 2020, 117(35):21450-21458.
- [70] SEBÉ-PEDRÓS A, BALLARÉ C, PARRA-ACERO H, et al. The dynamic regulatory genome of capsaspora and the origin of animal multicellularity[J]. *Cell*, 2016, 165(5):1224-1237.
- [71] DICKEL D E, YPSILANTI A R, PLA R, et al. Ultraconserved enhancers are required for normal development[J]. *Cell*, 2018, 172(3):491-499.
- [72] VILLAR D, BERTHELOT C, ALDRIDGE S, et al. Enhancer evolution across 20 mammalian species[J]. *Cell*, 2015, 160(3):554-566.
- [73] SNETKOVA V, YPSILANTI A R, AKIYAMA J A, et al. Ultraconserved enhancer function does not require perfect sequence conservation[J]. *Nat Genet*, 2021, 53(4):521-528.
- [74] WONG E S, ZHENG D W, TAN S Z, et al. Deep conservation of the enhancer regulatory code in animals[J]. *Science*, 2020, 370(6517):eaax8137.
- [75] DENG R, HUANG J H, WANG Y, et al. Disruption of super-enhancer-driven tumor suppressor gene *RCAN1*. 4 expression promotes the malignancy of breast carcinoma[J]. *Molecular Cancer*, 2020, 19(1):122.
- [76] COHEN A J, SAIKHOVA A, CORRADIN O, et al. Hotspots of aberrant enhancer activity punctuate the colorectal cancer epigenome[J]. *Nat Commun*, 2017, 8:14400.
- [77] MANSOUR M R, ABRAHAM B J, ANDERS L, et al. Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element[J]. *Science*, 2014, 346(6215):1373-1377.
- [78] OLDRIDGE D A, WOOD A C, WEICHERT-LEAHEY N, et al. Genetic predisposition to neuroblastoma mediated by a LMO1 super-enhancer polymorphism[J]. *Nature*, 2015, 528(7582):418-421.
- [79] HNISZ D, WEINTRAUB A S, DAY D S, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods[J]. *Science*, 2016, 351(6280):1454-1458.
- [80] BAUER D E, KAMRAN S C, LESSARD S, et al. An erythroid enhancer of *BCL11A* subject to genetic variation determines fetal hemoglobin level[J]. *Science*, 2013, 342(6155):253-257.
- [81] FRANGOUL H, ALTSHULER D, CAPPELLINI M D, et al. CRISPR-Cas9 gene editing for sickle cell disease and β -thalassemia[J]. *N Engl J Med*, 2021, 384(3):252-260.

Enhancer in regulation of gene expression: from new technology to new mechanism

WU Chengyi, CHEN Fei, HONG Danni, ZHU Ming, TONG Mengsha, HUANG Jialiang*

(State Key Laboratory of Cellular Stress Biology, Innovation Center for Cell Signaling Network, School of Life Sciences, Xiamen University, Xiamen 361102, China)

Abstract: Genome-wide studies have identified millions of putative enhancers in the human genome; however, their biological functions and underlying mechanisms remain largely unknown. Here, we review emerging high-throughput technologies used for identifying, characterizing and validating enhancers. We further discuss our biological understanding of the potential regulatory mechanisms of enhancers, with an emphasis on how enhancer clusters control gene expression, such as functional hierarchy of super-enhancers. Finally, we summarize the important roles of enhancers in regulating spatiotemporal gene expression during development, evolution and disease occurrence. We also provide thoughts on key unanswered questions and future directions in this field.

Keywords: enhancer; gene expression regulation; high-throughput technology; functional hierarchy

(责任编辑:曾礼娜)