



何江平, 2014—2019 年就读于中国科学院广州生物医药与健康研究院, 在陈捷凯课题组攻读博士学位, 目前任生物岛实验室副研究员。博士期间, 主要研究转座元件的表观遗传调控机制及其在细胞命运决定中的功能。利用生物信息学方法系统性解析了小鼠胚胎干细胞中所有转座元件的表观遗传修饰图谱, 发现不同转座元件存在多种不同的表观遗传修饰调控模式, 包括抑制型的 H3K9me3 和激活型的 H3K27ac、H3K4me1 等, 并进一步筛选到了调控胚胎干细胞往 2 细胞样细胞重编程的新型调控因子。此外, 还研究了转座元件 LTR6B 在人定型内胚层分化及转座元件 RLTR13B2 在小鼠上胚层干细胞往胚胎干细胞重编程等细胞命运决定中的功能及机制。博士论文《转座元件表观遗传调控与细胞命运决定》获得 2020 年中国科学院优秀博士学位论文。截止目前, 以第一作者(含共同)在 *Nature*、*Cell*、*Nature Cell Biology*、*Nature Communications* 和 *Protein & Cell* 等杂志发表论文 8 篇, 以其他作者身份发表论文 10 篇, 累积引用超过 600 次。

转座元件、表观遗传调控与细胞命运决定

何江平^{1,2}, 陈捷凯^{1,2}

1. 生物岛实验室细胞命运与谱系中心, 广州 510530
2. 中国科学院广州生物医药与健康研究院, 广州 510530

摘要: 转座元件是哺乳动物基因组内含量最多的元素。尽管转座元件的存在对基因组稳定性具有潜在的危险, 但它们同时还是潜在的基因调控序列、蛋白质编码序列和染色质结构序列, 并参与物种进化过程。因此, 基因组中转座元件的有害性和有益性保持着谨慎的平衡, 并且这种平衡主要由表观遗传修饰来调控。本文详细介绍了异染色质类型表观遗传修饰如 H3K9me3 和 DNA 甲基化在转座元件沉默中的功能; 转座元件作为增强子元件富集激活型表观遗传修饰如 H3K4me1 和 H3K27ac, 以及作为转录因子结合靶点、染色质构象锚点等方式参与基因表达调控的模式; 从体内胚胎发育到体外细胞命运转变, 阐述了转座元件在细胞命运决定中的潜在功能及作用方式; 最后, 对转座元件领域研究存在的挑战及潜在解决方法提出了见解。总之, 本文对转座元件与表观遗传、基因表达调控以及细胞命运决定等方面的研究及存在的问题进行了较全面的综述, 旨在为相关领域的研究人员提供参考。

关键词: 转座元件; 表观遗传调控; 细胞命运决定

收稿日期: 2021-03-25; 修回日期: 2021-07-23

基金项目: 国家重点研发计划(编号:2019YFA0110200)和中国科学院基础前沿科学研究计划 0-1 项目(编号:ZDBS-LY-SM007) [Supported by the National Key R&D Program of China (No. 2019YFA0110200) and the Frontier Science Research Program of the CAS (No. ZDBS-LY-SM007)]

作者简介: 何江平, 博士, 研究方向: 转座元件的表观遗传调控及功能。E-mail: he_jiangping@grmh-gdl.cn

通讯作者: 陈捷凯, 研究员, 博士生导师, 研究方向: 表观遗传与细胞命运决定。E-mail: chen_jiekai@gibh.ac.cn

DOI: 10.16288/j.ycz.21-113

网络出版时间: 2021/8/19 16:23:56

URI: <https://kns.cnki.net/kcms/detail/11.1913.R.20210819.1034.002.html>

Epigenetic control of transposable elements and cell fate decision

Jiangping He^{1,2}, Jiekai Chen^{1,2}

1. Center for Cell Lineage and Atlas (CCLA), Bioland Laboratory (Guangzhou Regenerative Medicine and Health Guangdong Laboratory), Guangzhou 510530, China

2. Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China

Abstract: Transposable elements (TEs) are the most prevalent elements in mammalian genomes. Although potential risks for genome stability, they are a pool of potential regulatory sequences, chromatin control elements, protein-coding genes, and substrates for evolutionary processes. Consequently, a delicate balance is maintained between the potential benefits and deleterious aspects of TEs, and this balance is mediated by the epigenetic regulatory system. In this review, we introduce the role of heterochromatin associated epigenetic modifications such as histone 3 lysine 9 trimethylation (H3K9me3) and DNA methylation in the silencing of TEs as well as epigenetic modifications such as histone 3 lysine 4 monomethylation (H3K4me1) and histone 3 lysine 27 acetylation (H3K27ac) in activation of TEs. Further, we elaborate the functions of TEs as binding sites of transcription factors and as anchors of chromosomal conformation in regulation of gene expression. We introduce the impact of TEs on the process of cell fate determination including natural embryonic development *in vivo* and artificial cell fate transition *in vitro*. We discuss the main challenges associated with computational TEs analysis and TEs functions exploration, as well as the different experimental and computational strategies in studying these processes. In all, this article provides a comprehensive review of the research advances and existing problems in study of transposable elements in epigenetic regulatory mechanisms, gene transcriptional regulation, and cell fate determination, thereby providing some references for researchers in the fields.

Keywords: transposable elements; epigenetic regulation; cell fate decision

转座元件(transposable elements, TEs)是一类在基因组内可以自由移动的DNA,最早由美国遗传学家芭芭拉·麦克林托克(Barbara McClintock)在玉米(*Zea mays*)基因组中发现,并证明TEs能通过“跳跃”调控玉米粒的颜色^[1,2]。继此项开创性的工作之后,研究者们发现TEs几乎在所有真核细胞基因组中都存在,并在生命演化过程中扮演着非常重要的角色。

在人类(*Homo sapiens*)基因组中,总共含有30亿对碱基,但只有约2%能够编码蛋白质,而接近一半的序列由TEs组成,共有超过300万份不同的拷贝,小鼠(*Mus musculus*)基因组也基本类似。TEs可以扩增并插入到基因组新的位置,调控基因表达,为物种进化提供原动力,是哺乳动物基因组非常重要的组成部分。而TEs在基因组中“随意”移动则会造成大量遗传突变威胁生命安全,因而基因组进化出了一系列调控机制以限制TEs的活性。本文主要综述了TEs在哺乳动物基因组中的主要类别、TEs

在哺乳动物中的表观遗传调控模式及在胚胎发育等细胞命运决定中的功能及作用机制。

1 转座元件

基因组中TEs的分类非常庞大而复杂。根据转座方式的不同可以分为两大类:DNA转座元件(DNA transposons)和逆转座元件(retrotransposons)。逆转座元件在转座过程中需要先转录成RNA,由RNA逆转录成cDNA再发生转座。逆转座元件转座后原DNA序列仍然保留,通过逆转录的cDNA插入到新的基因组位点,所以这类转座元件是通过“复制-粘贴”的方式完成“跳跃”。哺乳动物基因组中大部分转座元件为逆转座元件,在人和小鼠基因组序列中占比超过40%^[3]。与逆转座元件不同,DNA转座元件则不需要RNA的介导,直接通过转座酶将原位点DNA序列切下来插入到新的位点即可发生转座,

通过“剪切-粘贴”的方式完成。人类基因组中大约有 500,000 个 DNA 转座元件, 约占基因组序列 3% (图 1A)。在灵长类基因组中, 大部分 DNA 转座元件通常已经失去了转座的能力; 相反, 活跃的逆转座元件几乎在所有的灵长类基因组中都有被发现^[4]。逆转座元件根据其两端是否有长末端重复序列(long terminal repeat, LTR)可分为 LTR 和非 LTR 转座元件。LTR 转座元件因其在序列上与外源性逆转录病毒非常相似, 因此又常被称为内源性逆转录病毒(endogenous retrovirus, ERV)。进化学上认为, 这主要是在进化过程中外源逆转录病毒感染宿主或其祖先后, 这些逆转录病毒的序列保留在宿主基因组中, 并获得生殖系传递, 所以这些序列一直保留至今。因此, 完整的 ERVs 元件和外源性逆转录病毒一样, 主要由两端的 LTR 以及中间编码的 *gag*、*pol* 以及 *env* 编码序列组成(图 1B)。但在进化过程中, 由于突变等因素, 这些 ERVs 元件大多被截断掉, 导致大部分是非完整性的, 很多甚至只保留了单独的 LTR 序列。在人类基因组中, 大约有 650,000 个 ERVs 元件, 但其中只有 2000 个保留着几乎完整的 ERV 序列^[3]。人类基因组中有超过 300 万个非 LTR 转座元件, 占基因组含量约 30%, 而 LTR 转座元件只占基因组含量约 8%^[3]。非 LTR 转座元件中, 又可分为长的散在元件(long interspersed nuclear elements, LINEs)和短的散在元件(short interspersed nuclear elements, SINEs)。其中, LINEs 包含编码转座所需要的两个蛋白编码框 ORF1 和 ORF2 (图 1B), 但与 ERVs 元件一样, 大部分 LINEs 元件在进化过程中被截断进而失去了完整的编码序列。人类基因组中有约 950,000 个 LINEs 元件, 但其中只有不到 1% 具有完整的蛋白编码序列^[5]。与 LINEs 不同, SINEs 不包含转座所需要蛋白的编码序列(图 1B), 它们需要借助 LINEs 编码的蛋白进行转座。所以 ERVs、LINEs 常被称为自主性转座元件, 而 SINEs 由于不能编码转座所需要的蛋白, 又被称为非自主性转座元件。ERVs、LINEs 和 SINEs 根据序列的不同, 又可进一步分为 1000 多种不同的家族, 散在分布在基因组每一条染色体上。

与小鼠基因组相比, 除 LINE、SINE、DNA 和 LTR 转座元件外, 人类基因组中存在一类特有的被称为 SVA(SINE-VNTR-Alu)的转座元件。SVA 是人类科(hominoid)物种进化过程中活跃的 TEs, 大约在

2500 万年前产生, 在人基因组中有约 3000 个拷贝^[6]。经典的 SVA 全长约 2 kb, 序列由(CCCTCT)_n、Alu 样区域、串联重复区域(variable number of tandem repeats, VNTR)、HERV-K10 样区域以及 Poly-A 尾组成。SVA 为非自主转座元件, 目前的研究证据提示其转座可能依赖于 L1 的转座机器^[6], 其中, KRAB 锌指蛋白 ZNF91/93 和 SVA/L1 在进化过程中互相博弈, 抑制进化过程中新生 SVA/L1 的转座活性^[7]。此外, 基因组中还存在其他 KRAB-ZNF 蛋白与转座元件的共进化^[8,9]。

TEs 是物种进化的主要动力之一。科学家们通过比较鼠-人之间序列的同源性, 发现存在大量的 TEs 序列在真哺乳亚纲形成之前就已经存在^[10,11]。并且, 在人基因组中发现的数千个保守的调控元件中, 其中大部分来源于不同类别的 TEs^[12]。除存在保守的调控序列来源于 TEs 外, 同时还存在大量物种特异性的调控元件, 其中很多也来源于 TEs。例如, 通过比较有袋类动物(如负鼠)和真哺乳亚纲动物(人、狗、小鼠和大鼠等)基因组, 发现至少有 16% 的真哺乳亚纲动物特异性的保守基因调控元件来源于 TEs^[13]。同样与小鼠基因组比较, 人类基因组中也存在许多特异性的调控元件由 TEs 构成, 反之亦然^[10,14]。暗示这些非保守 TEs 的获得, 可能是新物种形成的主要原因之一, 并且 TEs 是不同物种进化的主要痕迹之一。

1.1 转座元件与表观遗传调控

基因的表达主要受表观遗传修饰调控, TEs 亦是如此。TEs 由于转座特性会导致基因组不稳定性, 所以其表观遗传调控一直以来都是科学研究的重要问题之一。基因组中大部分转座元件通常会被表观遗传所沉默, 包括转录水平异染色质的修饰以及转录后水平 RNA 的修饰。DNA 甲基化和组蛋白 H3K9me3 是负责基因表达沉默的两种最主要的表观遗传修饰, 通常高富集这些修饰的染色质会处于沉默状态, 如形成异染色质。尽管 TEs 上的表观遗传调控机制还不是非常清楚, 但可以确定的是这两种修饰在 TEs 表达调控上都起着非常重要的主导作用^[15,16]。

DNA 甲基化通常被认为是体细胞中控制转座元件活性的主要因素^[11]。然而, 在早期胚胎发育过

程中, 基因组中发生整体 DNA 去甲基化, 因此会导致部分转座元件活化而表达^[17]。然而基因组会以其他的表观遗传方式来代偿 DNA 甲基化以控制 TEs 的表达, 防止转座元件过度活跃而导致基因组不稳定, H3K9me3 介导的异染色质化是其中最主要的方式。锌指蛋白(Zinc-finger proteins, ZFPs)是已知可以直接结合到 TEs 上并招募 *Trim28*, *Trim28* 能够进一步募集组蛋白 H3K9me3 的甲基转移酶 *Setdb1* 使转 TEs 异染色质化而被沉默^[9,18-20]。能够招募 *Trim28* 的 KRAB-ZFP 蛋白还包括 *Zfp809*、*Yy1*、*Zfp819* 以及多能性因子 *Zfp42* 等^[18,20-22], 这些锌指蛋白都参

与了部分 TEs 的沉默。有研究表明, 与 H3K9me3 修饰相关的其他表观遗传因子如 *Suv39h1/2*、*G9a*、*Daxx/Atrx* 等都与 TEs 沉默有关^[17,23,24]。不同 H3K9me3 甲基转移酶调控的 TEs 也有所不同, 其中 *Setdb1* 主要调控 ERVs 元件, 而 *Suv39h1/2* 主要调控 LINEs 元件^[15,23]。近期研究还发现, 具有 m⁶A 修饰的 TEs 来源的 RNA 及其相关的甲基转移酶 *Mettl3* 以及识别因子 *Ythdc1* 等在对 TEs 的沉默起着至关重要的作用^[25-27] (图 2A)。除 H3K9me3 外, H4K20me3、H4R3me2 以及组蛋白变体 H3.3 都被发现在 TEs 区富集并参与了转座元件的转录调控^[28,29]。此外, 许

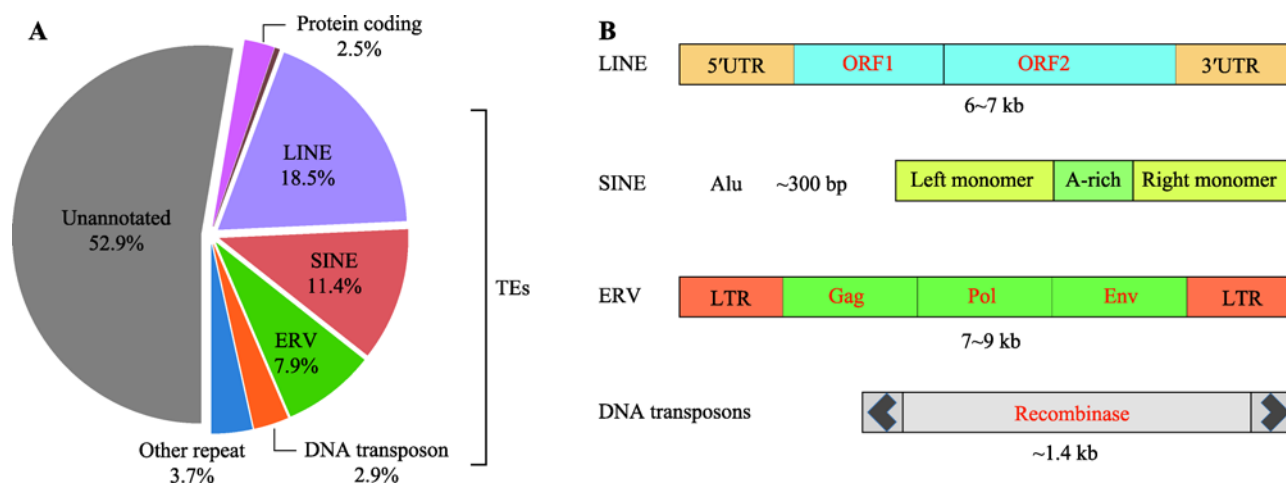


图 1 小鼠基因组中转座元件分类及含量

Fig. 1 Classification and content of transposable elements within the mouse genome

A: 小鼠基因组中不同序列占基因组的比例。基因注释来源于 GENCODE v20, TEs 注释来源于 RepeatMasker。B: 不同种类转座元件结构示意图。

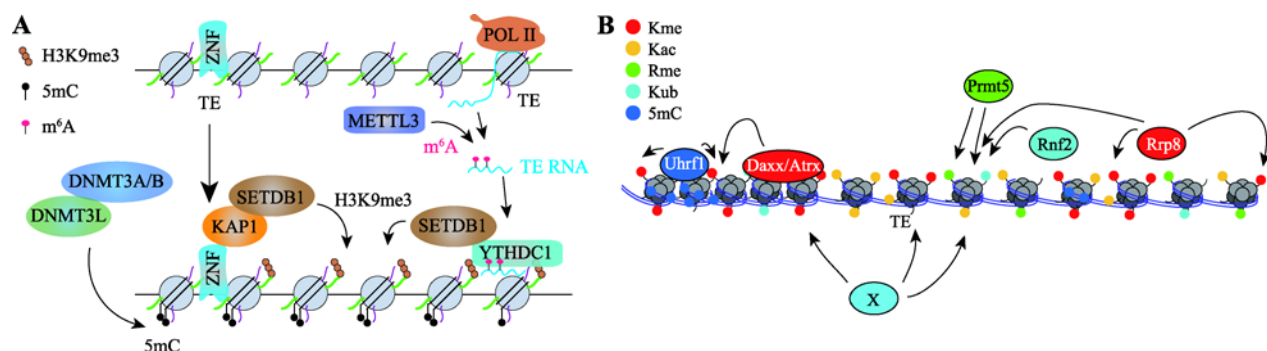


图 2 小鼠基因组中转座元件主要表观遗传沉默模式

Fig. 2 Mechanism of transposable elements silencing

A: 小鼠基因组中逆转座元件的表观遗传沉默主要方式。H3K9me3 和 DNA 甲基化占主导作用, RNA 的 m⁶A 修饰通过其书写器蛋白 METTL3、阅读器蛋白 YTHDC1 招募 SETDB1, 指导 TEs 区 H3K9me3 异染色质化沉默。B: 参与沉默 TEs 表观遗传修饰及各类表观遗传因子。

多其他表观遗传因子如 *Uhrf1*、*Kdm1a*、*Rif1*、*Sumo2*、*Chaf1a/b*、*Tet2*、胞嘧啶脱氨脱氨酶 APOBEC3 家族以及 RNA 结合蛋白 TRIM33 等在转座元件调控中起着非常重要的作用^[22,30-37]。除以上提到的表观遗传修饰外,本课题组的研究还发现不同的 TEs 之间表观遗传调控方式存在明显差异^[38],基因组中 TEs 的表观遗传调控非常精密而复杂(图 2B)。

除大部分 TEs 被表观遗传修饰沉默外,也存在少部分 TEs 在特定的细胞类型中被特定的表观遗传修饰所激活。组蛋白修饰 H3K27ac、H3K4me1 和 H3K4me3 通常被认为是基因激活的标签,其中 H3K27ac 和 H3K4me1 是活性增强子的表观修饰标记, H3K4me3 是活跃启动子区表观遗传修饰,这些修饰同样也在部分 TEs 染色质区域富集从能激活其表达。比如在小鼠滋养外胚层干细胞(trophoblast stem cell, TSCs)中,转座元件 RLTR13D5 上就显著富集 H3K4me1 和 H3K27ac 修饰,并且激活 RLTR13D5 作为增强子进一步招募 TSCs 核心转录因子 CDX2、EOMES 和 ELF5 的结合^[39]。类似的,在人胚胎干细胞中,转座元件 LTR77 等富集 H3K4me1 修饰^[40];在小鼠胚胎干细胞中,转座元件 RLTR9、RLTR13 等富集 H3K27ac 和 H3K4me1 修饰被激活并调控基因表达^[41]。

1.2 转座元件与基因转录调控

TEs 可以通过多种方式调控基因表达^[42,43]。总体而言,大多数 TEs 在基因组内呈散在分布,但对部分 TEs 而言,它们的分布并非完全随机。人类基因组中约 25% 的基因启动子由 TEs 构成^[44],并且基因的表达与其附近 TEs 的分布及含量有显著联系,其中基因附近 SINEs 的含量与基因表达呈正相关关系,而 LINEs 的含量与基因表达呈负相关关系^[45]。此外,TEs 也可通过作为增强子调控基因表达,并对多种生物学功能如胚胎干细胞多能性维持、胎盘发生、神经发生、免疫反应以及生殖系细胞形成等过程具有重要调控作用^[14,46,47]。并且,TEs 也可在转录后水平调控基因表达,如影响基因的可变剪接、RNA 编辑、细胞内定位等^[48,49]。此外,TEs 也可作为许多转录因子如多能性因子 POU5F1、SOX2、NANOG、ZFP42,中内胚层相关转录因子 FOXA2、GATA4、SOX17 等提供结合位点进而调控基因表

达^[50,51]。值得一提的是,TEs 对基因组三维结构的维持与构建也发挥着重要的作用,如 LINE1、B1/Alu (属于 SINEs 家族)转座元件对三维基因组 A/B 区隔化起到重要的调控作用^[52]。同样,转座元件 MERV1 和 HERVH 分别为小鼠早期胚胎和人心肌细胞发育过程中拓扑结构域(Topologically associating domain, TAD)的形成提供锚点,以完成对基因的区块化调控^[53,54]。此外,许多 TEs 还能作为染色质高级结构塑造因子 CTCF 的结合位点,形成增强子/绝缘子-启动子相互作用套环,实现针对单个基因的精确调控^[55-57]。同时,TEs 还可通过反式作用如通过 lncRNA、miRNA 等非编码 RNA 的形式调控基因表达^[48,50]。总之,TEs 通过影响染色质修饰、转录因子结合、RNA 编辑以及染色质高级结构等,对基因的表达起到重要的调控作用。

1.3 表观遗传、转座元件与细胞命运决定

从原核生物到真核生物,从单细胞生物到多细胞生物,这是物种进化上两次重大的突破。不同细胞命运状态的形成是多细胞生物形成的基础,细胞命运决定是多细胞生物发育过程中必须经历的生物事件。多细胞生物个体中所有的细胞均来源于同一个受精卵细胞,随着发育的过程,逐渐形成多种命运状态不同的细胞,这一过程究其根本原因是表观遗传修饰导致的遗传信息非均一性展示。

表观遗传修饰在细胞命运决定过程中起着根本性调控作用。在发育过程中,不同细胞命运形成过程发生不同细胞命运决定,表观遗传修饰对正常发育、不同细胞谱系细胞的形成起着非常重要的作用。例如,组蛋白修饰 H3K9me3 甲基转移酶 *Setdb1* 敲除的小鼠早期胚胎致死^[58],同样 DNA 甲基转移酶 *Dnmt3a/b* 也是小鼠正常发育所必需,敲除后致死^[59],其他表观遗传因子如 *Kdm2a*、*Kdm2b*、*Kdm1a* 等敲除亦是胚胎致死^[60-62]。这些表观遗传因子敲除致死通常是因为特定谱系细胞命运决定,在发育过程中不能在特定的时间点和特定的组织中精确完成或者维持,在动物表型上,则会进一步导致胚胎发育异常从而致死。作为发育/分化过程相对应的体细胞重编程和转分化,同样也伴随着大量细胞命运决定的过程。研究表明,许多表观遗传修饰参与了体细胞重编程过程中细胞命运决定,如异染色质修饰

H3K9me3 的擦除是体细胞重编程末期非常关键的节点,许多未能完全重编程的细胞中,其多能性基因如 *Oct4*、*Nanog*、*Sox2* 以及 *Dppa5a* 等位点都存在较高的 H3K9me3 修饰,并且敲低 H3K9me3 甲基转移酶 *Setdb1* 基因可以显著促进未完全重编程的细胞继续重编程^[63]。此外, DNA 去甲基化酶 *Tet1/2* 基因也能够调控体细胞重编程效率^[64,65],其他表观遗传因子如 *Kdm1a*、*Kdm2a/b*、*Kdm5b*、*Dot1l*、*Chaf1a/b*、*Ncor2* 和 *Rybp* 等均能显著影响体细胞重编程过程^[66-73]。此外,以上表观遗传因子在 TEs 活性调控上也起着至关重要的作用^[3,38,74],提示表观遗传修饰可能通过调控 TEs 对细胞命运决定和维持起着及其重要的作用。

作为表观遗传修饰主要锚点的 TEs 也参与细胞命运决定调控。在早期胚胎发育过程中,小鼠特有转座元件 MERVL 和 LINEs 参与早期胚胎发育过程中 2-细胞期到囊胚的细胞命运转变的调控^[75-77]。在神经发育过程中,转座元件 L1 参与了神经祖细胞向神经元的细胞命运转变^[78,79]。在体细胞重编程过程中,部分转座元件在重编程中后期被激活^[80]。类似的,人原始态(naïve)胚胎干细胞中也存在部分活化的 TEs,如进化过程中较为年轻的人猿科类特有的转座元件 SVA、HERVK 和 HERVH 等,并且这些 TEs 的激活与沉默,通过改变染色质开放状态、组蛋白修饰和 DNA 甲基化修饰等,改变基因调控网络,直接参与了人 naïve 胚胎干细胞命运状态的维持和决定^[81-83]。在人体细胞重编程过程中,转座元件 HERVH 也会被激活,并且其过度激活将导致获得的诱导多能性干细胞分化缺陷^[84]。在免疫系统中,ERVs 的激活也直接参与了 naïve T 细胞向 Th2 和 Th1 的细胞命运转变,ERVs 的激活直接促进了 naïve T 细胞向 Th1 细胞命运的转变,而抑制向 Th2 细胞命运的转变^[85]。

2 转座元件研究领域存在的主要挑战

随着测序技术和系统生物学的发展,全基因组/转录组的深度测序、染色质修饰表征及表达谱分析等生物信息学分析成为驱动转座元件领域研究发展的有力工具,这对理解非编码 RNA、TEs 和宿主基因之间的关系有重要意义。然而,针对 TEs 的生物

信息学分析和功能研究同样也面临着诸多挑战,如测序数据的匹配。TEs 由于其多拷贝重复序列的特性,同一家族不同 TEs 拷贝间序列高度相似,因此在序列分析上难以确定测序数据来源的准确位点。TEs 来源的测序片段通常能够匹配到基因组多个不同的位置,这种序列称为多次匹配序列。通常处理这类多次匹配序列的策略大致可以分为 3 种:第一种为去除所有多次匹配的序列,只保留单次匹配的数据,常规针对基因表达调控的分析采用的即是这种策略;第二种是保留所有多次匹配的序列,并将所有可匹配位点记录下来用于后续的分析计算;第三种策略与第二种策略类似,保留了所有多次匹配序列,但针对每条序列只保留其中一个最佳可匹配位点,如果存在多个最佳匹配位点时,则随机保留其中一个用于后续的分析计算。这 3 种策略各有利弊,其中策略一得到的匹配结果最准确,但大部分 TEs 来源的信息被去除,这对进化上年轻的 TEs 尤为明显,由于它们突变较少,很难找到可靠的突变位点以区分不同的拷贝;策略二保留了重复序列信息,但会导致部分位点重复序列信号过度放大,产生过多的假阳性;策略三在保留了重复序列信息的同时也没有过度放大重复序列的信号,值得注意的是,虽然这种策略在整个家族层面比较真实的反应了 TEs 的整体水平,但由于是随机分配,所以这种方法计算的单个拷贝位点信号同样并不是非常准确的^[86,87]。无论何种分析方法,都无法准确衡量 TEs 单个拷贝在测序数据中的信号程度。解决这一问题的根本办法或许还得在测序技术上获得突破,使得测序技术能够更长更准确的读取碱基序列。另一个主要的挑战是 TEs 在基因组中的注释问题。TEs 在基因组中含量有数百万条不同的拷贝,如何准确识别每一个拷贝在基因组中的具体位置对 TEs 调控基因表达、染色质高级结构、TEs 来源的嵌合转录本等研究非常重要。目前不同基因组版本中 TEs 位置信息变动较大,TEs 的位置也是染色体组装的主要困难之一^[86]。其次,不同个体之间 TEs 的多态性对 TEs 的分析也会存在较大的影响,例如不同品系小鼠中 TEs 在基因组中的位置信息就存在很大差异^[88]。个体之间的表观遗传变异性是科学家们需要解决的一个重要问题,TEs 在不同个体之间插入的多态性也许能够部分地解释这种变异性,因此 TEs 各拷贝

在基因组中位置的准确注释就显得及其重要。这一问题的解决同样很大程度上依赖于测序技术的改进。

如何准确鉴定细胞命运决定相关的 TE 还存在一定的挑战。无论是发育还是疾病等一系列细胞命运决定过程中,细胞都存在非常大的异质性。单细胞技术是解析细胞命运决定过程中最有利的技术,通过单细胞测序技术解析 TE 的动态变化将对鉴定一系列潜在细胞命运决定相关 TE 提供极大的帮助^[89,90]。同时,实验上针对 TE 的功能验证也存在巨大挑战。尽管基因编辑技术的发展给基因功能的研究带来了极大的便利,但由于 TE 多拷贝的特性,这些技术一般难以针对 TE 进行编辑。尽管近年来也有研究报道利用改进的基因编辑技术可以针对 TE 进行整体调控^[77,91,92],但一般仅适用于非常保守的 TE 元件,并且存在效率低以及高脱靶性的问题。针对 TE 的基因编辑主要存在两个难点:其一,由于其在基因组中数目非常大,难以同时靶向所有/大部分拷贝位点,只靶向个别/部分位点时,对整体水平影响不大,或其他位点可能存在补偿效应;其二,TE 在基因组中散在分布,作为染色质的骨架,多个位点同时编辑时会影响染色质整体结构,从而导致细胞死亡。总之,针对 TE 的生物信息学分析及生物学功能实验验证目前在技术上还存在一定的挑战。

3 结语与展望

TE 是基因组中含量最多的元素。尽管 TE 的转座特性会给基因组完整性带来潜在的危险,但它们同时又是基因表达调控序列,并且是物种进化的主要动力和产物。基因组中 TE 的有害性和有益性保持在严格的平衡状态,这种平衡主要由表观遗传系统控制,特别是染色质层面上的表观遗传修饰。异染色质的形成是抑制转座元件表达的主要机制,本文重点总结了近年来关于 H3K9me3 修饰及其相关机制介导的异染色质在 TE 沉默中的关键作用,介绍了基因组中部分处于“激活”状态的 TE 在基因表达调控与细胞命运决定过程中的功能。

尽管前人的研究已经付出了很多的努力,但仍存在许多悬而未决的问题,比如:不同 TE 是如何产生的,物种特异性的 TE 进化在进化过程中的作

用是什么?不同 TE 间的具体调控机制是什么?不同细胞命运状态下是如何精确区分并激活/抑制特异的 TE?这些 TE 在细胞命运转变过程中具有什么功能?个体间 TE 插入位点的多态性和表观遗传变异性的关系是什么?越来越多的 TE 被检测到存在多样的表观遗传修饰和特定的表达模式,这背后的生物学意义是什么?直至目前,这些问题都尚不清楚,对细胞命运转变过程中 TE 的生理功能及调控机制的了解仍然十分匮乏。尽管人们还不能全面认识 TE 的精细调控机制及在细胞命运转变过程中的规律和功能,但随着测序技术、单细胞技术、生物信息分析方法和基因编辑技术的不断发展,相信未来该领域会取得新的突破。同时,研究 TE 在细胞命运转变过程中的变化情况及分子调控机制,将为细胞命运转换过程揭示一个全新的、重要的遗传因素。从新的角度来定义并研究细胞状态与命运决定,也许将为揭示影响细胞命运决定因素的研究提供新的突破口和原创性的发现,并且有助于人们对细胞命运转变更深层次的理解,以及将加深人们对 TE 在再生医学与人类疾病中的认识。

参考文献(References):

- [1] McCLINTOCK B. Chromosome organization and genic expression. *Cold Spring Harb Symp Quant Biol*, 1951, 16: 13–47. [\[DOI\]](#)
- [2] Cui XK, Cao XF. Overview of the function of transposable elements in higher plants. *Prog Biochem Biophys*, 2015, 42(11): 1033–1046.
崔颢奎, 曹晓凤. 高等植物转座元件功能研究进展. *生物化学与生物物理进展*, 2015, 42(11): 1033–1046. [\[DOI\]](#)
- [3] Hutchins AP, Pei DQ. Transposable elements at the center of the crossroads between embryogenesis, embryonic stem cells, reprogramming, and long non-coding RNAs. *Sci Bull (Beijing)*, 2015, 60(20): 1722–1733. [\[DOI\]](#)
- [4] Pace JK, Feschotte C. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome Res*, 2007, 17(4): 422–432. [\[DOI\]](#)
- [5] Beck CR, Garcia-Perez JL, Badge RM, Moran JV. LINE-1 elements in structural variation and disease. *Annu Rev Genomics Hum Genet*, 2011, 12: 187–215. [\[DOI\]](#)
- [6] Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. *Nat Rev Genet*, 2009, 10(10): 691–703. [\[DOI\]](#)

- 691–703. [DOI]
- [7] Jacobs FMJ, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, Paten B, Salama SR, Haussler D. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature*, 2014, 516(7530): 242–245. [DOI]
- [8] Wang JL, Wang J, Tian CY. Evolution of KRAB-containing zinc finger proteins and their roles in species evolution. *Hereditas(Beijing)*, 2016, 38(11): 971–978.
王进龙, 王建, 田春艳. KRAB 型锌指蛋白的进化及在物种演化中的功能. *遗传*, 2016, 38(11): 971–978. [DOI]
- [9] Imbeault M, Helleboid PY, Trono D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature*, 2017, 543(7646): 550–554. [DOI]
- [10] Silva JC, Shabalina SA, Harris DG, Spouge JL, Kondrashovi AS. Conserved fragments of transposable elements in intergenic regions: evidence for widespread recruitment of MIR- and L2-derived sequences within the mouse and human genomes. *Genet Res*, 2003, 82(1): 1–18. [DOI]
- [11] Friedli M, Trono D. The developmental control of transposable elements and the evolution of higher species. *Annu Rev Cell Dev Biol*, 2015, 31: 429–451. [DOI]
- [12] Xie XH, Kamal M, Lander ES. A family of conserved noncoding elements derived from an ancient transposable element. *Proc Natl Acad Sci USA*, 2006, 103(31): 11659–11664. [DOI]
- [13] Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, Garber M, Gentles AJ, Goodstadt L, Heger A, Jurka J, Kamal M, Mauceli E, Searle SMJ, Sharpe T, Baker ML, Batzer MA, Benos PV, Belov K, Clamp M, Cook A, Cuff J, Das R, Davidow L, Deakin JE, Fazzari MJ, Glass JL, Grabherr M, Grealley JM, Gu WJ, Hore TA, Huttley GA, Kleber M, Jirtle RL, Koina E, Lee JT, Mahony S, Marra MA, Miller RD, Nicholls RD, Oda M, Papenfuss AT, Parra ZE, Pollock DD, Ray DA, Schein JE, Speed TP, Thompson K, VandeBerg JL, Wade CM, Walker JA, Waters PD, Webber C, Weidman JR, Xie XH, Zody MC, Broad Institute Genome Sequencing Platform; Broad Institute Whole Genome Assembly Team; Graves JAM, Ponting CP, Breen M, Samollow PB, Lander ES, Lindblad-Toh K. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature*, 2007, 447(7141): 167–177. [DOI]
- [14] Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet*, 2017, 18(2): 71–86. [DOI]
- [15] Groh S, Schotta G. Silencing of endogenous retroviruses by heterochromatin. *Cell Mol Life Sci*, 2017, 74(11): 2055–2065. [DOI]
- [16] Karimi MM, Goyal P, Maksakova IA, Bilenky M, Leung D, Tang JX, Shinkai Y, Mager DL, Jones S, Hirst M, Lorincz MC. DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell Stem Cell*, 2011, 8(6): 676–687. [DOI]
- [17] Rowe HM, Trono D. Dynamic control of endogenous retroviruses during development. *Virology*, 2011, 411(2): 273–287. [DOI]
- [18] Tan XY, Xu XB, Elkenani M, Smorag L, Zechner U, Nolte J, Engel W, Pantakani DVK. Zfp819, a novel KRAB-zinc finger protein, interacts with KAP1 and functions in genomic integrity maintenance of mouse embryonic stem cells. *Stem Cell Res*, 2013, 11(3): 1045–1059. [DOI]
- [19] Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T, Maillard PV, Layard-Liesching H, Verp S, Marquis J, Spitz F, Constam DB, Trono D. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature*, 2010, 463(7278): 237–240. [DOI]
- [20] Wolf D, Goff SP. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature*, 2009, 458(7242): 1201–1204. [DOI]
- [21] Schoorlemmer J, Pérez-Palacios R, Climent M, Guallar D, Muniesa P. Regulation of mouse retroelement MuERV-L/MERVL expression by REX1 and epigenetic control of stem cell potency. *Front Oncol*, 2014, 4: 14. [DOI]
- [22] Yang BX, Farran CAE, Guo HC, Yu T, Fang HT, Wang HF, Schlesinger S, Seah YFS, Goh GYL, Neo SP, Li YH, Lorincz MC, Tergaonkar V, Lim TM, Chen LY, Gunaratne J, Collins JJ, Goff SP, Daley GQ, Li H, Bard FA, Loh YH. Systematic identification of factors for provirus silencing in embryonic stem cells. *Cell*, 2015, 163(1): 230–245. [DOI]
- [23] Bulut-Karslioglu A, De La Rosa-Velázquez IA, Ramirez F, Barenboim M, Onishi-Seebacher M, Arand J, Galán C, Winter GE, Engist B, Gerle B, O'Sullivan RJ, Martens JHA, Walter J, Manke T, Lachner M, Jenuwein T. Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Mol Cell*, 2014, 55(2): 277–290. [DOI]
- [24] He QY, Kim H, Huang R, Lu WS, Tang MF, Shi FT, Yang D, Zhang XY, Huang JJ, Liu D, Zhou SY. The Daxx/Atrx complex protects tandem repetitive elements during DNA hypomethylation by promoting H3K9 trimethylation. *Cell*

- Stem Cell*, 2015, 17(3): 273–286. [DOI]
- [25] Xu WQ, Li JH, He CX, Wen J, Ma HH, Rong BW, Diao JB, Wang LY, Wang JH, Wu FZ, Tan L, Shi YG, Shi Y, Shen HJ. METTL3 regulates heterochromatin in mouse embryonic stem cells. *Nature*, 2021, 591(7849): 317–321. [DOI]
- [26] Chelmicki T, Roger E, Teissandier A, Dura M, Bonneville L, Rucli S, Dossin F, Fouassier C, Lameiras S, Bourc'his D. m⁶ A RNA methylation regulates the fate of endogenous retroviruses. *Nature*, 2021, 591(7849): 312–316. [DOI]
- [27] Liu J, Dou XY, Chen CY, Chen C, Liu C, Xu MM, Zhao SQ, Shen B, Gao YW, Han DL, He C. N⁶-methyladenosine of chromosome-associated regulatory RNA regulates chromatin state and transcription. *Science*, 2020, 367(6477): 580–586. [DOI]
- [28] Elsässer SJ, Noh KM, Diaz N, Allis CD, Banaszynski LA. Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. *Nature*, 2015, 522(7555): 240–244. [DOI]
- [29] Kim S, Günesdogan U, Zylicz JJ, Hackett JA, Cougot D, Bao S, Lee C, Dietmann S, Allen GE, Sengupta R, Surani MA. PRMT5 protects genomic integrity during global DNA demethylation in primordial germ cells and preimplantation embryos. *Mol Cell*, 2014, 56(4): 564–579. [DOI]
- [30] Esnault C, Heidmann O, Delebecque F, Dewannieux M, Ribet D, Hance AJ, Heidmann T, Schwartz O. APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. *Nature*, 2005, 433(7024): 430–433. [DOI]
- [31] Richardson SR, Narvaiza I, Planegger RA, Weitzman MD, Moran JV. APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. *eLife*, 2014, 3: e02008. [DOI]
- [32] Esnault C, Millet J, Schwartz O, Heidmann T. Dual inhibitory effects of APOBEC family proteins on retrotransposition of mammalian endogenous retroviruses. *Nucleic Acids Res*, 2006, 34(5): 1522–1531. [DOI]
- [33] Isbel L, Srivastava R, Oey H, Spurling A, Daxinger L, Puthalakath H, Whitelaw E. Trim33 binds and silences a class of young endogenous retroviruses in the mouse testis; a novel component of the arms race between retrotransposons and the host genome. *PLoS Genet*, 2015, 11(12): e1005693. [DOI]
- [34] Ancelin K, Syx L, Borensztein M, Ranisavljevic N, Vassilev I, Briseño-Roa L, Liu T, Metzger E, Servant N, Barillot E, Chen CJ, Schüle R, Heard E. Maternal LSD1/KDM1A is an essential regulator of chromatin and transcription landscapes during zygotic genome activation. *eLife*, 2016, 5: e08851. [DOI]
- [35] Ishiuchi T, Enriquez-Gasca R, Mizutani E, Bošković A, Ziegler-Birling C, Rodriguez-Terrones D, Wakayama T, Vaquerizas JM, Torres-Padilla ME. Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat Struct Mol Biol*, 2015, 22(9): 662–671. [DOI]
- [36] Li PS, Wang L, Bennett BD, Wang JJ, Li JL, Qin YF, Takaku M, Wade PA, Wong JM, Hu G. Rif1 promotes a repressive chromatin state to safeguard against endogenous retrovirus activation. *Nucleic Acids Res*, 2017, 45(22): 12723–12738. [DOI]
- [37] Guallar D, Bi XJ, Pardavila JA, Huang X, Saenz C, Shi XL, Zhou HW, Faiola F, Ding JJ, Haruehanroengra P, Yang F, Li D, Sanchez-Priego C, Saunders A, Pan F, Valdes VJ, Kelley K, Blanco MG, Chen LY, Wang HY, Sheng J, Xu MJ, Fidalgo M, Shen XH, Wang JL. RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in pluripotent stem cells. *Nat Genet*, 2018, 50(3): 443–451. [DOI]
- [38] He JP, Fu XL, Zhang M, He FF, Li WJ, Abdul MM, Zhou JG, Sun L, Chang C, Li YH, Liu H, Wu KX, Babarinde IA, Zhuang Q, Loh YH, Chen JK, Esteban MA, Hutchins AP. Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells. *Nat Commun*, 2019, 10(1): 34. [DOI]
- [39] Chuong EB, Rumi MAK, Soares MJ, Baker JC. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat Genet*, 2013, 45(3): 325–329. [DOI]
- [40] Xie MC, Hong CB, Zhang B, Lowdon RF, Xing XY, Li DF, Zhou X, Lee HJ, Maire CL, Ligon KL, Gascard P, Sigaroudinia M, Tlsty TD, Kadlecsek T, Weiss A, O'Geen H, Farnham PJ, Madden PAF, Mungall AJ, Tam A, Kamoh B, Cho S, Moore R, Hirst M, Marra MA, Costello JF, Wang T. DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat Genet*, 2013, 45(7): 836–841. [DOI]
- [41] Sundaram V, Choudhary MNK, Pehrsson E, Xing XY, Fiore C, Pandey M, Maricque B, Udawatta M, Ngo D, Chen YJ, Paguntalan A, Ray T, Hughes A, Cohen BA, Wang T. Functional cis-regulatory modules encoded by mouse-specific endogenous retrovirus. *Nat Commun*, 2017, 8: 14550. [DOI]
- [42] Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, Mager DL, Feschotte C. Ten things you

- should know about transposable elements. *Genome Biol*, 2018, 19(1): 199. [DOI]
- [43] Wei LY, Cao XF. The effect of transposable elements on phenotypic variation: insights from plants to humans. *Sci China Life Sci*, 2016, 59(1): 24–37. [DOI]
- [44] Jordan IK, Rogozin IB, Glazko GV, Koonin EV. Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet*, 2003, 19(2): 68–72. [DOI]
- [45] Jjingo D, Huda A, Gundapuneni M, Mariño-Ramírez L, Jordan IK. Effect of the transposable element environment of human genes on gene length and expression. *Genome Biol Evol*, 2011, 3: 259–271. [DOI]
- [46] Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science*, 2016, 351(6277): 1083–1087. [DOI]
- [47] Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu XY, Chan YS, Ng HH, Bourque G. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet*, 2010, 42(7): 631–634. [DOI]
- [48] Elbarbary RA, Lucas BA, Maquat LE. Retrotransposons as regulators of gene expression. *Science*, 2016, 351(6274): aac7247. [DOI]
- [49] Chen LL, DeCerro JN, Carmichael GG. Alu element-mediated gene silencing. *EMBO J*, 2008, 27(12): 1694–1705. [DOI]
- [50] Feschotte C. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet*, 2008, 9(5): 397–405. [DOI]
- [51] Ito J, Sugimoto R, Nakaoka H, Yamada S, Kimura T, Hayano T, Inoue I. Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses. *PLoS Genet*, 2017, 13(7): e1006883. [DOI]
- [52] Lu JY, Chang L, Li T, Wang T, Yin YF, Zhan G, Han X, Zhang K, Tao YB, Percharde M, Wang L, Peng Q, Yan PX, Zhang H, Bi XJ, Shao W, Hong YT, Wu ZY, Ma RZ, Wang PZ, Li WZ, Zhang J, Chang Z, Hou YP, Zhu B, Ramalho-Santos M, Li PL, Xie W, Na J, Sun YJ, Shen XH. Homotypic clustering of L1 and B1/Alu repeats compartmentalizes the 3D genome. *Cell Res*, 2021, 31(6): 613–630. [DOI]
- [53] Kruse K, Díaz N, Enriquez-Gasca R, Gaume X, Torres-Padilla ME, Vaquerizas JM. Transposable elements drive reorganisation of 3D chromatin during early embryogenesis. *bioRxiv*, 2019, doi: <https://doi.org/10.1101/523712>. [DOI]
- [54] Zhang YX, Li T, Preissl S, Amaral ML, Grinstein JD, Farah EN, Destici E, Qiu YJ, Hu R, Lee AY, Chee S, Ma KY, Ye Z, Zhu Q, Huang H, Fang RX, Yu LQ, Belmonte JCI, Wu J, Evans SM, Chi NC, Ren B. Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat Genet*, 2019, 51(9): 1380–1388. [DOI]
- [55] Diehl AG, Ouyang NX, Boyle AP. Transposable elements contribute to cell and species-specific chromatin looping and gene regulation in mammalian genomes. *Nat Commun*, 2020, 11(1): 1796. [DOI]
- [56] Cao YQ, Chen GY, Wu G, Zhang XL, McDermott J, Chen XW, Xu C, Jiang QL, Chen ZX, Zeng YY, Ai DS, Huang Y, Han JDJ. Widespread roles of enhancer-like transposable elements in cell identity and long-range genomic interactions. *Genome Res*, 2019, 29(1): 40–52. [DOI]
- [57] Schmidt D, Schwalie PC, Wilson MD, Ballester B, Gonçalves A, Kutter C, Brown GD, Marshall A, Flicek P, Odom DT. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell*, 2012, 148(1–2): 335–348. [DOI]
- [58] Dodge JE, Kang YK, Beppu H, Lei H, Li E. Histone H3-K9 methyltransferase ESET is essential for early development. *Mol Cell Biol*, 2004, 24(6): 2478–2486. [DOI]
- [59] Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, 1999, 99(3): 247–257. [DOI]
- [60] Andricovich J, Kai Y, Peng WQ, Foudi A, Tzatsos A. Histone demethylase KDM2B regulates lineage commitment in normal and malignant hematopoiesis. *J Clin Invest*, 2016, 126(3): 905–920. [DOI]
- [61] Kawakami E, Tokunaga A, Ozawa M, Sakamoto R, Yoshida N. The histone demethylase Fbx11/Kdm2a plays an essential role in embryonic development by repressing cell-cycle regulators. *Mech Dev*, 2015, 135: 31–42. [DOI]
- [62] Wang J, Hevi S, Kurash JK, Lei H, Gay F, Bajko J, Su H, Sun WT, Chang H, Xu GL, Gaudet F, Li E, Chen TP. The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nat Genet*, 2009, 41(1): 125–129. [DOI]
- [63] Chen JK, Liu H, Liu J, Qi J, Wei B, Yang JQ, Liang HQ, Chen Y, Chen J, Wu YR, Guo L, Zhu JY, Zhao XJ, Peng TR, Zhang YX, Chen S, Li XJ, Li DW, Wang T, Pei DQ. H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nat Genet*, 2013, 45(1): 34–42. [DOI]

- [64] Chen JK, Guo L, Zhang L, Wu HY, Yang JQ, Liu H, Wang XS, Hu X, Gu TP, Zhou ZW, Liu J, Liu JD, Wu HL, Mao SQ, Mo KL, Li YY, Lai KY, Qi J, Yao HJ, Pan GJ, Xu GL, Pei DQ. Vitamin C modulates TET1 function during somatic cell reprogramming. *Nat Genet*, 2013, 45(12): 1504–1509. [DOI]
- [65] Sardina JL, Collombet S, Tian TV, Gómez A, Di Stefano B, Berenguer C, Brumbaugh J, Stadhouders R, Segura-Morales C, Gut M, Gut IG, Heath S, Aranda S, Di Croce L, Hochedlinger K, Thieffry D, Graf T. Transcription factors drive Tet2-mediated enhancer demethylation to reprogram cell fate. *Cell Stem Cell*, 2018, 23(5): 727–741.e9. [DOI]
- [66] Cheloufi S, Elling U, Hopfgartner B, Jung YL, Murn J, Ninova M, Hubmann M, Badeaux AI, Ang CE, Tenen D, Wesche DJ, Abazova N, Hogue M, Tasdemir N, Brumbaugh J, Rathert P, Jude J, Ferrari F, Blanco A, Fellner M, Wenzel D, Zinner M, Vidal SE, Bell O, Stadtfeld M, Chang HY, Almouzni G, Lowe SW, Rinn J, Wernig M, Aravin A, Shi Y, Park PJ, Penninger JM, Zuber J, Hochedlinger K. The histone chaperone CAF-1 safeguards somatic cell identity. *Nature*, 2015, 528(7581): 218–224. [DOI]
- [67] Zhuang Q, Li WJ, Benda C, Huang ZJ, Ahmed T, Liu P, Guo XP, Ibañez DP, Luo ZW, Zhang M, Abdul MM, Yang ZZ, Yang JY, Huang YH, Zhang H, Huang DH, Zhou JG, Zhong XF, Zhu XH, Fu XL, Fan WX, Liu YL, Xu Y, Ward C, Khan MJ, Kanwal S, Mirza B, Tortorella MD, Tse HF, Chen JY, Qin BM, Bao XC, Gao SR, Hutchins AP, Esteban MA. NCoR/SMRT co-repressors cooperate with c-MYC to create an epigenetic barrier to somatic cell reprogramming. *Nat Cell Biol*, 2018, 20(4): 400–412. [DOI]
- [68] Onder TT, Kara N, Cherry A, Sinha AU, Zhu N, Bernt KM, Cahan P, Marcarci BO, Unternaehrer J, Gupta PB, Lander ES, Armstrong SA, Daley GQ. Chromatin-modifying enzymes as modulators of reprogramming. *Nature*, 2012, 483(7391): 598–602. [DOI]
- [69] Li HH, Lai P, Jia JP, Song YW, Xia Q, Huang KM, He N, Ping WF, Chen JY, Yang ZZ, Li J, Yao MZ, Dong XT, Zhao JC, Hou CH, Esteban MA, Gao SR, Pei DQ, Hutchins AP, Yao HJ. RNA helicase DDX5 inhibits reprogramming to pluripotency by miRNA-based repression of RYBP and its PRC1-dependent and -independent functions. *Cell Stem Cell*, 2017, 20(4): 462–477.e6. [DOI]
- [70] Zhou ZW, Yang XJ, He JP, Liu J, Wu F, Yu SY, Liu YT, Lin RX, Liu H, Cui YB, Zhou CH, Wang XS, Wu J, Cao ST, Guo L, Lin LH, Wang T, Peng XZ, Qiang BQ, Hutchins AP, Pei DQ, Chen JK. Kdm2b regulates somatic reprogramming through variant PRC1 complex-dependent function. *Cell Rep*, 2017, 21(8): 2160–2170. [DOI]
- [71] Wang T, Chen KS, Zeng XM, Yang JG, Wu Y, Shi X, Qin BM, Zeng LW, Esteban MA, Pan GJ, Pei DQ. The histone demethylases Jhdm1a/1b enhance somatic cell reprogramming in a vitamin-C-dependent manner. *Cell Stem Cell*, 2011, 9(6): 575–587. [DOI]
- [72] Smith ZD, Sindhu C, Meissner A. Molecular features of cellular reprogramming and development. *Nat Rev Mol Cell Biol*, 2016, 17(3): 139–154. [DOI]
- [73] Hörmanseder E, Simeone A, Allen GE, Bradshaw CR, Figlmüller M, Gurdon J, Jullien J. H3K4 methylation-dependent memory of somatic cell identity inhibits reprogramming and development of nuclear transfer embryos. *Cell Stem Cell*, 2017, 21(1): 135–143.e6. [DOI]
- [74] Deniz Ö, Frost JM, Branco MR. Regulation of transposable elements by DNA modifications. *Nat Rev Genet*, 2019, 20(7): 417–431. [DOI]
- [75] Percharde M, Lin CJ, Yin YF, Guan J, Peixoto GA, Bulut-Karslioglu A, Biechele S, Huang B, Shen XH, Ramalho-Santos M. A LINE1-nucleolin partnership regulates early development and ESC identity. *Cell*, 2018, 174(2): 391–405.e19. [DOI]
- [76] Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono D, Pfaff SL. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*, 2012, 487(7405): 57–63. [DOI]
- [77] Jachowicz JW, Bing XY, Pontabry J, Bošković A, Rando OJ, Torres-Padilla ME. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat Genet*, 2017, 49(10): 1502–1510. [DOI]
- [78] Muotri AR, Chu VT, Marchetto MCN, Deng W, Moran JV, Gage FH. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature*, 2005, 435(7044): 903–910. [DOI]
- [79] Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu YL, Lovci MT, Morell M, O'Shea KS, Moran JV, Gage FH. L1 retrotransposition in human neural progenitor cells. *Nature*, 2009, 460(7259): 1127–1131. [DOI]
- [80] Friedli M, Turelli P, Kapopoulou A, Rauwel B, Castro-Díaz N, Rowe HM, Ecco G, Unzu C, Planet E, Lombardo A, Mangeat B, Wildhaber BE, Naldini L, Trono D. Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. *Genome Res*, 2014, 24(8): 1251–1259. [DOI]
- [81] Theunissen TW, Friedli M, He YP, Planet E, O'Neil RC,

- Markoulaki S, Pontis J, Wang HY, Iouranova A, Imbeault M, Duc J, Cohen MA, Wert KJ, Castanon R, Zhang ZZ, Huang YM, Nery JR, Drotar J, Lungjangwa T, Trono D, Ecker JR, Jaenisch R. Molecular criteria for defining the naive human pluripotent state. *Cell Stem Cell*, 2016, 19(4): 502–515. [DOI]
- [82] Grow EJ, Flynn RA, Chavez SL, Bayless NL, Wossidlo M, Wesche DJ, Martin L, Ware CB, Blish CA, Chang HY, Pera RAR, Wysocka J. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature*, 2015, 522(7555): 221–225. [DOI]
- [83] Wang JC, Xie GC, Singh M, Ghanbarian AT, Raskó T, Szvetnik A, Cai HQ, Besser D, Prigione A, Fuchs NV, Schumann GG, Chen W, Lorincz MC, Ivics Z, Hurst LD, Izsvák Z. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature*, 2014, 516(7531): 405–409. [DOI]
- [84] Ohnuki M, Tanabe K, Sutou K, Teramoto I, Sawamura Y, Narita M, Nakamura M, Tokunaga Y, Nakamura M, Watanabe A, Yamanaka S, Takahashi K. Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. *Proc Natl Acad Sci USA*, 2014, 111(34): 12426–12431. [DOI]
- [85] Adoue V, Binet B, Malbec A, Fourquet J, Romagnoli P, van Meerwijk JPM, Amigorena S, Joffre OP. The histone methyltransferase SETDB1 controls T helper cell lineage integrity by repressing endogenous retroviruses. *Immunity*, 2019, 50(3): 629–644.e8. [DOI]
- [86] Goerner-Potvin P, Bourque G. Computational tools to unmask transposable elements. *Nat Rev Genet*, 2018, 19(11): 688–704. [DOI]
- [87] Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet*, 2011, 13(1): 36–46. [DOI]
- [88] Nellåker C, Keane TM, Yalcin B, Wong K, Agam A, Belgard TG, Flint J, Adams DJ, Frankel WN, Ponting CP. The genomic landscape shaped by selection on transposable elements across 18 mouse strains. *Genome Biol*, 2012, 13(6): R45. [DOI]
- [89] He JP, Babarinde IA, Sun L, Xu SY, Chen RH, Wei YJ, Li YH, Ma G, Zhuang Q, Hutchins AP, Chen JK. Unveiling transposable element expression heterogeneity in cell fate regulation at the single-cell level. *bioRxiv*, 2020, <https://doi.org/10.1101/2020.07.23.218800>. [DOI]
- [90] Shao WQ, Wang T. Transcript assembly improves expression quantification of transposable elements in single-cell RNA-seq data. *Genome Res*, 2021, 31(1): 88–100. [DOI]
- [91] Niu D, Wei HJ, Lin L, George H, Wang T, Lee IH, Zhao HY, Wang Y, Kan YN, Shrock E, Lesha E, Wang G, Luo YL, Qing YB, Jiao DL, Zhao H, Zhou XY, Wang SQ, Wei H, Güell M, Church GM, Yang LH. Inactivation of porcine endogenous retrovirus in pigs using CRISPR-Cas9. *Science*, 2017, 357(6357): 1303–1307. [DOI]
- [92] Fuentes DR, Swigut T, Wysocka J. Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation. *eLife*, 2018, 7: e35989. [DOI]

(责任编辑: 陆发隆)

中国科学院广州生物医药与健康研究院陈捷凯课题组简介

中国科学院广州生物医药与健康研究院陈捷凯课题组成立于 2013 年,课题组长为陈捷凯研究员。课题组专注于细胞命运决定机理的研究,以干细胞为模型(如类器官、重编程等),主要研究表观遗传调控的特异性机制,及其与信号转导、转录因子、非编码遗传信息之间的联系。除了干细胞实验平台和生物化学平台之外,课题组擅长通过计算生物学在表观遗传组和单细胞转录组等大数据中挖掘重要生物学机理,并根据生物学问题开发新算法和新软件,在转座元件和细胞谱系方面开发了一系列工具。课题组近年来取得了一系列研究成果,在 *Nature*、*Cell*、*Cell Stem Cell*、*Nature Cell Biology*、*Nature Communications*、*National Science Review* 和 *Cell Reports* 等国际知名期刊发表论文数十篇。与此同时,课题组还承担了科技部、国家自然科学基金委及中国科学院等一系列重大课题。课题组网站:
http://www.gibh.cas.cn/sourcedb_gibh_cas/zw/zjrc/201202/t20120210_3438472.html。

