

· 研究进展 ·

## “细胞编程和重编程的表观遗传机制” 重大研究计划结题综述

张 儒<sup>1</sup> 赵天宇<sup>2</sup> 田艳艳<sup>2</sup> 高绍荣<sup>1</sup> 杜生明<sup>2</sup> 谷瑞升<sup>2\*</sup>

(1. 同济大学 生命科学与技术学院, 上海 200092;

2. 国家自然科学基金委员会 生命科学部, 北京 100085)

**[摘要]** 国家自然科学基金委员会重大研究计划“细胞编程和重编程的表观遗传机制”历时8年于2017年完成结束评估。本文概述了该计划取得的研究进展,特别是在发现新的表观遗传调控因子和染色质重塑因子并揭示其生物学功能和作用机制,解析干细胞自我更新、体细胞重编程的调控机理及创建半克隆新方法,揭示表观遗传调控细胞分化转分化、个体发育及疾病发生发展的规律,以及从全基因组的层次上认识表观遗传网络形成和运行的机制等方面取得一系列突破性研究成果,产生了重要国际影响。文中展望了表观遗传学研究的发展态势,提出了未来关注的重点学术方向。

**[关键词]** 体细胞重编程;表观遗传;染色质重塑;干细胞;基金委重大研究计划

表观遗传学是自20世纪80年代后期逐渐兴起的一门新学科,研究在DNA序列不变的前提下,引起可遗传的基因表达或细胞表型变化的分子机制。表观遗传调控是生命现象中一种普遍存在的基因表达调控方式,是调控生长、发育、衰老与疾病发生的重要机制之一。表观遗传调控特别在干细胞维持和自我更新与分化,个体的衰老和发育异常,如肿瘤、糖尿病、精神病及神经系统疾病等复杂疾病的发生发展以及植物杂种优势等起着重要作用,而且生命个体对环境因素(包括营养、物理化学因素、甚至心理因素等)发生有序应答在很大程度上依赖于表观遗传调控网络的有效运行。

为了提升我国生命科学基础研究的创新能力,促进学科交叉,培养创新人才,实现在表观遗传学研究领域跨越发展,国家自然科学基金委员会(以下简称“基金委”)于2008年启动了重大研究计划“细胞编程和重编程的表观遗传机制”(以下简称“计划”)。该计划遵循基金委重大研究计划“有限目标、稳定支持、集成升华、跨越发展”的总体思路,重点围绕DNA甲基化和去甲基化的分子机制及生物学意义、

细胞重编程的表观遗传机制和细胞重编程过程中核染色质和非编码核酸的高级结构及动态变化等3个核心科学问题开展深入研究,共资助培育项目68项,重点支持项目23项,集成项目56项,总经费约1.9亿元。

在指导专家组、管理工作组和承担项目科学家的共同努力下,本计划全面完成了预定的各项科学目标,取得了丰硕的学术成果。在该计划资助下,我国科学家在国际一流期刊发表论文815篇,全部论文引用总计超过1.8万次,单篇引用最高1200余次(统计时间截至2017年10月)。30纳米染色质结构等成果已入选最新版的国际主流教科书 *Lehninger Principles of Biochemistry* 和 *Fundamentals of Biochemistry*;申报国际国内专利37项;荣获多项重要奖项,其中包括国家自然科学基金二等奖2项、国家科技进步一等奖1项、国家科技进步二等奖4项,并有4项研究成果入选中国科学十大进展。在本计划的支持下,我国在表观遗传和细胞命运决定领域培养出一批具有国际水准的优秀科学家。本计划实施期间,专家指导组成员或项

收稿日期:2018-04-18;修回日期:2018-05-14

\* 通信作者,Email: rsgu@nsfc.gov.cn

目承担人中有 5 人当选“中国科学院院士”，15 人获得基金委“国家杰出青年科学基金”，16 人获得基金委“优秀青年科学基金”，9 人获基金委“创新研究群体”学术带头人，培养出站博士后 84 名，毕业博士生 441 名，毕业硕士生 255 名，为我国表观遗传研究领域创新能力的全面提升和可持续发展提供了重要的人才保证。该计划实施，使我国科学家在短时间内成功跻身细胞重编程、细胞核移植及半克隆技术的国际前沿，特别在单倍体干细胞研究上处于国际引领地位。该计划的主要进展和亮点如下。

### 1 发现表观遗传调控新机制，在染色质高级结构和动态调控上取得重要成果

DNA 甲基化修饰、组蛋白修饰、核小体的组装、组蛋白密码的建立与识别、高级染色质结构的建立、维持及转换均是表观遗传调控的重要手段。本计划积极推动与结构生物学交叉，在发现与鉴定 DNA 甲基化修饰、组蛋白修饰的新分子方面取得了一系列突破性研究成果。

(1) 解析出 30 nm 染色质纤维高级结构及其动态调控机制。我国科学家利用自主建立的染色质体外组装和冷冻电镜技术，首次解析了 30 nm 染色质纤维的高精度三维冷冻电镜结构(11 埃)，发现 30 nm 染色质纤维是以 4 个核小体为结构单元形成

的左手双螺旋结构<sup>[1]</sup>，解决了分子生物学领域一个 30 多年悬而未决的重大科学问题(图 1)，利用单分子磁镊技术对 30 nm 染色质纤维结构建立和调控的动力学过程进行了深入研究，进一步描绘了 30 nm 染色质纤维三维结构的动态调控，首次揭示“四核小体串珠”是 30nm 染色质纤维折叠过程中一个重要的中间结构，基因转录过程中组蛋白分子伴侣 FACT 可以负调控“四核小体”结构。该成果先后入选最新版的世界著名的生物化学教科书 *Fundamentals of Biochemistry: Life at the Molecular Level* 和 *Lehninger Principles of Biochemistry*。

(2) 揭示出 DNA 甲基化与组蛋白甲基化修饰互动调控的新机制。确定了从头甲基化转移酶 DNMT3A 自抑制形式的 DNMT3A-DNMT3L 和活化形式的 DNMT3A-DNMT3L-H3 复合物的晶体结构，发现 DNMT3A 的 ADD 结构域与催化结构域(catalytic domain, CD)相互作用，通过阻断 DNA 结合力，抑制了 CD 的酶活性。组蛋白 H3(而非 H3K4me3)可以破坏 ADD-CD 互作，诱导 ADD 结构域大幅度移动，由此解除 DNMT3A 的自抑制。这一研究结果揭示出了 DNA 甲基化的另一个调控层面，确定了 DNMT3A 被激活的前提条件，并强有力地证实了在整个哺乳动物基因组中 H3K4me3 与 DNA 甲基化之间的负相关性<sup>[2]</sup>。这一研究提供了

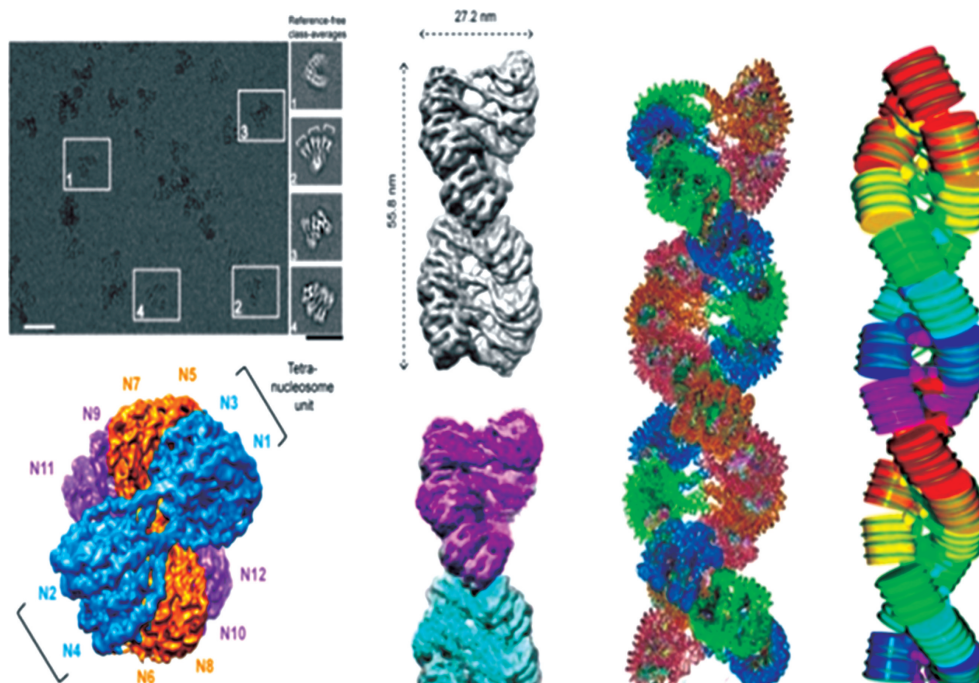


图 1 利用冷冻电镜三维重构技术解析 30nm 染色质左手双螺旋高清晰三维结构

DNA 从头甲基化转移酶在最初的基因组定位后的自抑制和组蛋白 H3 诱导其活化的一些新见解。

发现 H3K9 甲基化修饰与 DNA 甲基化修饰的互动机制,并提出 DNA 维持性甲基化的新模型:通过协同结合甲基化组蛋白及半甲基化 DNA, UHRF1 能够更有效的结合至 DNA 复制叉,进而招募 DNMT1 来完成 DNA 甲基化的维持功能。在小鼠模型中进一步证明了 UHRF1 通过识别 H3K9 甲基化来介导组蛋白修饰与 DNA 甲基化互动的的作用,并且明确了 H3K9 甲基化在哺乳动物细胞 DNA 甲基化修饰中起着辅助性但不是决定性的作用,表明不同物种中 DNA 甲基化修饰与组蛋白修饰互动的程度不同并且是通过不同的机制来实现的。发现 UHRF1/2 可以负调控 DNA 起始性甲基化,提出了 UHRF1/2 一方面促进 DNA 维持性甲基化一方面抑制起始性甲基化来维持细胞 DNA 甲基化稳态的假说。发现半甲基化 DNA(由 DNA 复制后形成)结合表观遗传调控因子 UHRF1 并引起 UHRF1 从闭合到开放的构象改变,从而激活 UHRF1 对组蛋白甲基化修饰的识别,确保 UHRF1 能够精确的定位到基因组的特定区域发挥维持 DNA 甲基化的功能。

**(3) 发现 DNA 甲基化酶和去甲基化酶的活性调控新机制。**哺乳动物基因组的胞嘧啶上会产生甲基化修饰,称为 5-甲基胞嘧啶(5mC,即第 5 种碱基),而 TET 蛋白是哺乳动物细胞中的一种氧化酶,可以执行 DNA 去甲基化功能。哺乳动物 TET 蛋白在受精卵表观遗传重编程、多能干细胞分化、骨髓造血等关键生命过程中扮演着至关重要的角色,其失活也与多种疾病,尤其是血液肿瘤的发生密切相关,以 TET 蛋白为中心的 DNA 去甲基化研究是近年表观遗传学最活跃的领域之一。前期研究发现, TET 蛋白在去甲基化过程中,将 5mC 氧化为 5hmC(5-羟甲基胞嘧啶,第 6 种碱基)后,可继续催化 5hmC 转化为 5-fC(5-醛基胞嘧啶,第 7 种碱基)和 5-caC(5-羧基胞嘧啶,第 8 种碱基)。其中,5hmC 在细胞内相对稳定存在,且其含量远远高于 5fC 和 5caC。基因组中 5hmC 稳定存在,且其含量远远高于 5fC 和 5caC。但这一现象一直没有合理的生物学解释。我国科学家利用结构生物学、生物化学和计算生物学等研究方法,揭开了这一谜底,发现了 5mC 在 TET 蛋白催化口袋中的取向使得它很容易被催化活性中心俘获并被氧化为 5hmC。5hmC 和 5fC 由于已经有氧的存在,其在催化口袋中被限制

住,不容易发生进一步的氧化反应,导致 TET 蛋白对这两种碱基活性降低。在这样的催化能力差异下,TET 会很顺利将 5mC 氧化产生 5hmC,一旦 5hmC 产生,TET 将不容易使其进一步氧化为 5fC 和 5caC,导致细胞内 5hmC 相对稳定,并且其含量远远高于 5fC 和 5caC。这一研究揭示了 TET 蛋白底物偏好性的机制,为基因组中 5-羟甲基胞嘧啶稳定存在提供了分子水平的解释。在特定的基因区域,TET 蛋白可能被特定的调控因子激活,会跨越能垒阻碍产生高活性的 TET,连续氧化为 5fC 和 5caC。这一发现解决了困扰表观遗传学领域的一个难题,也为揭示其他蛋白质逐步催化反应的分子机制提供了新思路和新方法<sup>[3]</sup>。

锌指结构域蛋白 SALL4A 倾向于结合含有 5hmC 修饰的 DNA,而 SALL4 是早期胚胎发育过程中的一个重要基因,它的突变会导致常染色体显性遗传的 Duane-radial ray 综合症。*Sall4* 基因敲除的小鼠胚胎在围着床期即停止发育,并很快死亡。我国科学家研究发现,在小鼠胚胎干细胞中,SALL4A 蛋白主要定位于增强子,其与染色质的结合在很大程度上依赖于 TET1 蛋白。进一步分析基因组上 SALL4A 结合位点的胞嘧啶修饰状态发现,这些位点上缺乏稳定的 5hmC,却富集了进一步氧化的产物 5fC 和 5caC,提示 SALL4A 可能促进 5hmC 的进一步氧化。敲除 *Sall4* 导致在原先的 SALL4A 结合位点上积累较高水平的 5hmC,因为敲除 *Sall4* 降低了 TET2 的稳定结合,不利于 5hmC 的进一步氧化。这一工作丰富了对 TET 家族蛋白调控的 DNA 氧化和去甲基化过程的理解,并提出了 5mC 的协同性递进氧化概念,促进了对 DNA 甲基化的动态性及其在胚胎干细胞功能及重编程中作用的理解。

**(4) 发现组蛋白修饰的新密码与解读机制。**组蛋白修饰是表观遗传调控基本机制之一,被认为构成一类“组蛋白密码”,调控着遗传信息在染色质层面的解读,在基因表达和细胞命运决定等过程中发挥着重要作用。近年来,众多新型组蛋白修饰被不断发现,其中一大类是组蛋白赖氨酸乙酰化修饰,如乙酰化(ac)、丙酰化(pr)、丁酰化(bu),和巴豆酰化(cr)等。其中组蛋白巴豆酰化是一类从酵母到人类都保守存在的修饰密码,与活跃转录和基因激活密切相关,调控着基因表达及配子成熟等生物学过程。自 2011 年组蛋白巴豆酰化被国外实验室报道以来,组蛋白巴豆酰化的产生、消除和识别机制研究成为

了领域热点。在本计划支持下,我国科学家发现新型组蛋白巴豆酰化修饰阅读器,解析了两类组蛋白巴豆酰化修饰阅读器 YEATS 和 DPF 结构域,阐明了该结构域通过特异读取组蛋白巴豆酰化密码促进基因转录的分子细胞机制,开启了代谢与表观遗传研究的新方向,深化了人们对巴豆酰化修饰生物学的理解和认识。

我国科学家还发现和证明 KIAA1718(KDM7A)是 H3K9 和 H3K27 的具有双活性的组蛋白去甲基化酶。用线虫 *ceKDM7A* 研究发现,该酶通过结合 H3K4 三甲基而去除 H3K9 和 H3K27 二甲基,揭示了与转录抑制相关的甲基化(H3K9 和 H3K27 二甲基)和与转录激活相关的甲基化(H3K4 三甲基)不在一起的一种机理。获得了含有不同组蛋白甲基化修饰肽段的 6 种 *ceKDM7A* 蛋白质共结晶,从结构上解释了该酶催化的特异性和底物识别机理。发现 PHF8 是 H3K9 的组蛋白去甲基化酶,该酶位于核仁区,调控 rRNA 转录。

(5) 在染色质高级结构建立、维持及转换研究上获得重要进展。揭示出着丝粒染色质建立过程及核小体组装机制。通过系统研究组蛋白变体 CENP-A 在着丝粒染色质建立过程中的识别、定位及其核小体组装等的作用,获得了组蛋白伴侣 HJURP 与着丝粒特征组蛋白变体 CENP-A-H4 复合体的晶体结构,揭示了 CENP-A 与 HJURP 的特异识别机制,发现 CENP-A 关键残基 Ser68 对于 HJURP 的识别起到重要的调控作用,发现该 Ser68 的磷酸化修饰调控 HJURP 与 CENP-A 的特异性识别,以及 CENP-A 在细胞周期中的动态组装。这些结果为理解着丝粒区核小体的组装机理提供了强有力的证据。

揭示了组蛋白变体在染色质组装中的作用。组蛋白 H3 家族包括 H3.1、H3.3 和着丝粒特异的 CenH3,它们在从果蝇到人类和植物中都非常保守。在组蛋白变体 H3.3 识别和装配的分子机制研究中,解析了 DAXX-H3.3-H4 复合体的晶体结构,揭示了组蛋白变体 H3.3 被其特异性分子伴侣 DAXX 和 HIRA 复合物识别的分子机制,为勾画 H3.3 的存储途径和理解 H3.3 的特异识别和组装机制奠定了基础。发现了 H3.3 可以和 H2A.Z 协同作用,动态调控 Enhancer 和 Promotor 区域染色质结构,从而维持干细胞自我更新和促进干细胞神经定向分化。在对拟南芥组蛋白 H3.3 和 H3.1 及它们的突变蛋白在核仁 rDNA 上精细的细胞生物学动态分析

中,提出和证实了一个模型:即处于组蛋白 H3.3 核心区域的 87 位和 90 位氨基酸介导了核小体的组装,而位于 N 端的 31 位和 41 位氨基酸则介导了核小体的去组装。

## 2 体细胞重编程的调控机制获得新进展,动物克隆和生殖技术获得重大突破

体细胞重编程一直是生命科学研究领域的热点之一。早期研究表明,体细胞重编程可以通过体细胞核移植技术实现。2006 年日本科学家山中伸弥利用 OSKM 四个转录因子将体细胞重新诱导成类似于胚胎干细胞的一种细胞类型,即诱导多能干细胞(iPS)。这种具有多能干细胞状态的 iPS 细胞的建立与发现,不仅开创了体细胞重编程的新方法,更无疑是生命科学历史上的一个里程碑事件。体细胞重编程对再生医学及药物的开发利用,以及治疗人类各种遗传性和功能性疾病都起到了巨大的推动作用。

(1) 成功建立单倍体胚胎干细胞及半克隆技术,获得了“人造精子”。通过将精子注射到去核卵母细胞中,构建了孤雄发育的单倍体胚胎并从中提取建立了孤雄单倍体胚胎干细胞系(ahES cells)。这些细胞具有分化形成三胚层细胞以及生殖细胞的能力。并且在注射入 MII 卵母细胞质内后可以替代精子使卵子受精,进一步发育成健康可育的个体一半克隆小鼠。将 PGK 启动子控制的 NEO 抗性基因电转进 ahES 细胞系 AHGFP-4 中,并进行 G418 药物筛选,建立了转基因单倍体干细胞系,随后挑取了三个转基因系进行 ICAI 实验,发现三个系都能发育至胎儿期,其中两个系分别获得 7 只和 1 只健康存活转基因动物<sup>[4]</sup>。这一研究提供了一种新的研究生殖和基因印记等基础问题的模型和一种新的获得转基因动物的方式,同时也为辅助生殖新技术的开发提供了新的思路。除此之外,利用单倍体胚胎干细胞创制出哺乳动物异种杂合二倍体胚胎干细胞<sup>[5]</sup>,通过细胞融合技术将小鼠孤雄(雌)和大鼠孤雌(雄)单倍体胚胎干细胞融合,从而绕开了小鼠和大鼠的精卵融合后无法发育的生殖隔离障碍,获得了异种杂合二倍体胚胎干细胞。这是首例人工创建的、以稳定二倍体形式存在的异种杂合胚胎干细胞,为进化生物学、发育生物学和遗传学等研究提供了新的模型和工具。

单倍体细胞的“受精”能力随着细胞的传代逐渐丢失,特别是经过基因编辑后,这些细胞很难获得健

康半克隆小鼠。我国科学家通过将调控雄性印记基因 H19 和 Gtl2 表达的 H19-DMR 和 IG-DMR 敲除后获得了能稳定产生半克隆小鼠的“类精子细胞样”的单倍体细胞(“人造精子”),并证明它们能稳定和高效支持获得遗传修饰的半克隆小鼠,这些“人造精子”细胞携带 CRISPR-Cas9 文库,能够一步产生大量的携带不同突变基因的小鼠,从而为开展小鼠个体水平的遗传筛选和遗传修饰提供了技术支持<sup>[6]</sup>。进一步研究证明,卵子来源的单倍体细胞在去除 H19-DMR 和 IG-DMR 后,同样具备了“人造精子”的能力,从而在哺乳动物中实现了高效的孤雌发育。

(2) 首创极体基因组移植预防线粒体遗传疾病技术。我国科学家成功地将极体作为受体基因放入到健康供体卵母细胞的胞质中,实现了基因组转移—线粒体置换。将干细胞治疗的细胞替代技术推进到细胞器置换技术,为难治性疾病的治疗提供了新的策略和路径。极体移植防止线粒体病技术比国际上通常使用的原核移植(PNT)、母系纺锤体移植(MST)等技术有五大先进性:降低线粒体 DNA 带进子代细胞的携带量;与纺锤体移植相比,减少了将染色体遗留在细胞内的风险(而极体中染色体都存在);避免应用细胞支架抑制因素从受精卵或卵母细胞中取出原核或纺锤体;避免应用更传统的操作方法,可以减少损伤患者细胞核和供体卵细胞的机会,因此能极大的提高效率;可以同时实施第一极体移植、第二极体移植、原核移植、纺锤体移植,因此能成倍提高成功率。极体移植预防线粒体置换技术的发明使得我国在该领域进入国际前沿。

(3) 提出谱系决定因子诱导成体细胞转变成 iPS 细胞的“seesaw 模型”。传统的 iPS 技术是通过

向成体细胞中导入多潜能性相关基因而建立多潜能性。研究中发现,谱系决定因子可以替代多潜能性基因,诱导成体细胞转变成 iPS 细胞。基于这一新发现,我国科学家提出了“seesaw 模型”,从新的角度解释了细胞命运决定的理论。“多潜能性”是各种不同的分化谱系之间达到平衡所形成的细胞状态,通过不同细胞谱系之间的平衡而诱导重编程发生可能是一个普适的原理:(1) 发现作为谱系分化因子的整个 GATA 家族蛋白的六个成员都能替换 Yamanaka 四因子中的 OCT4。此外,证实了 GATA 家族的所有成员都能够抑制外胚层分化基因的表达,验证了之前的“跷跷板”模型。随后进一步发现 GATA 转录因子通过中间桥梁因子 Sall4 来建立起多能性状态所需的干性网络;(2) 在之前建立的完全使用小分子化合物实现体细胞重编程为多潜能干细胞(CiPS)的方法中,发现在小分子诱导重编程过程中也存在着动态“seesaw”模型—与传统转录因子诱导重编程截然不同,在重编程早期,小分子启动 XEN 相关基因的表达,使“seesaw”倾斜并使细胞进入 XEN-like 状态;在重编程后期,通过切换至小鼠 ES 细胞培养条件诱导 Sox2 的表达,将倾斜的“seesaw”重新恢复到平衡位置,从而使细胞进入到多潜能状态<sup>[7]</sup>。

### 3 发育与疾病相关的表观遗传机制研究取得丰硕成果

(1) 发现体细胞向肝细胞转分化的关键因子,生物人工肝临床应用展现曙光。我国科学家利用慢病毒系统在成体小鼠鼠尾成纤维细胞中过表达肝脏细胞 Foxa3、Hnf1a、Gata4 三个转录因子,成功地

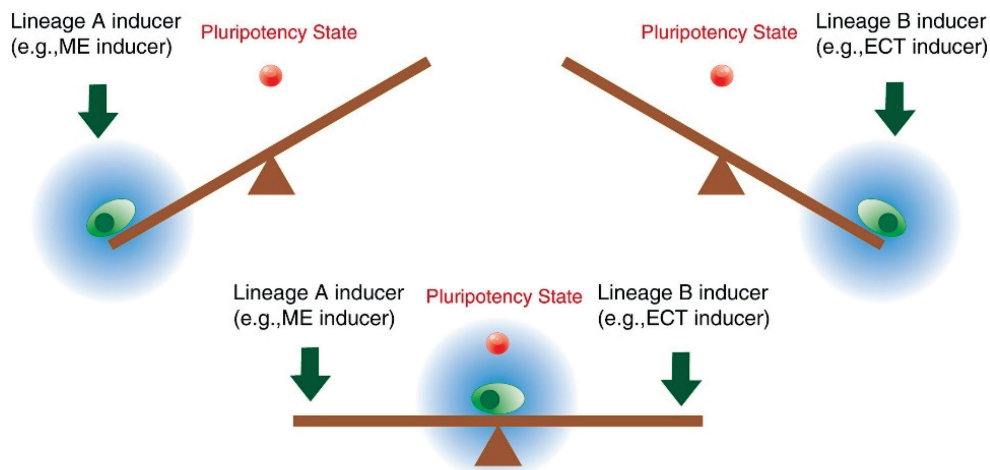


图2 小分子诱导重编程过程中的动态“seesaw”模型

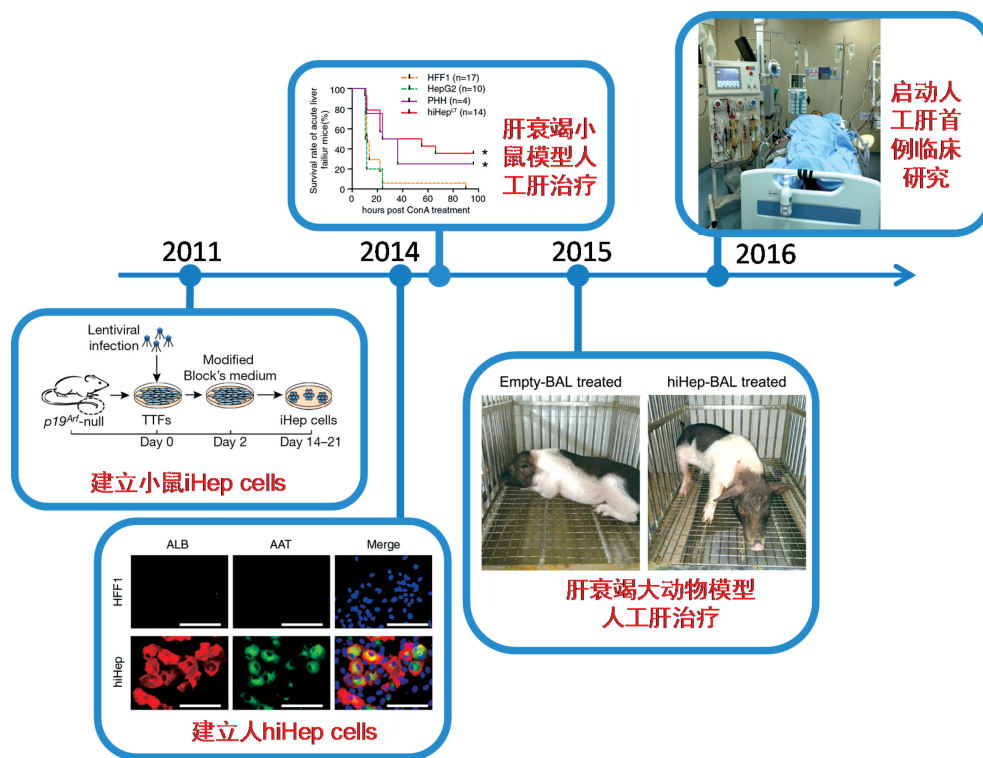


图3 iHep 的获得及临床转化研究

将 p19 失活的小鼠鼠尾成纤维细胞转分化为肝细胞样细胞 (iHep), 同时发现 p53 的激活是去分化重编程的关键抑制机制。iHep 细胞不仅能整合到小鼠肝脏中, 还在体内发挥了肝实质细胞的功能。此外, 移植 iHep 细胞的 *Fah*<sup>-/-</sup> 小鼠和 NOD/SCID 小鼠均没有肿瘤形成, 进一步说明了 iHep 细胞的安全性<sup>[8]</sup>。利用 FOXA3、HNF1A、HNF4A 三个转录因子和 SV40 Large T 可将人胚胎成纤维细胞诱导转分化为可增殖的功能肝细胞 (hiHep)。hiHep 细胞具有与人类原代肝细胞相似的基因表达谱, 也获得了肝脏的体外功能, 尤为重要, hiHep 具有良好的胆汁排泄能力, 可应用于药物开发过程中药物胆汁排泄能力的评估。进一步研究已成功将产生的功能细胞 (hiHep) 扩增至临床数量级, 与生物人工肝装置结合, 成功救治了急性肝衰竭的小型猪。在 2016 年 1 月, 项目科学家与专科医院合作进行了第一例急性肝衰竭患者的救治实验, 成功救治一位罹患乙肝 40 年并突发急性肝衰竭的患者。患者在接受 hiHep 生物人工肝治疗后各项肝功能指标均明显好转, 且无任何不良反应。这一病例标志着新型 hiHep 生物人工肝第一例临床治疗成功完成, 而顺利推进 hiHep 生物人工肝的临床应用将极大地促进我国科研成果转化的进程。

(2) 揭示出表观遗传调控发育及疾病发生的新机制。我国科学家在表观遗传调控发育及疾病发生取得诸多学术成果。例如, 发现了组蛋白的修饰形式与基因的转录调控密切相关, 在神经发生中扮演着重要角色。在胚胎干细胞神经分化过程中, 抑制组蛋白去乙酰化酶活性可以显著抑制胚胎干细胞的神经分化, 并促进中内胚层分化, 提示组蛋白去乙酰化在神经发育中有重要作用。发现 HDAC1 特异地结合到 *Nodal* 基因区域, 并且组蛋白去乙酰化酶抑制剂能显著上调 *Nodal* 的表达, *Nodal* 信号的活化促进中内胚层命运的表型、抑制神经命运决定, 表明 HDAC1 通过拮抗 *Nodal* 维持上胚层阶段的神经命运决定。AF9 基因对于人胚胎干细胞神经分化和神经发育相关基因的转录激活是充分且必要的, DNA 羟氧化酶 TET2 能够与 AF9 相互作用, 并且 AF9 和 TET2 共同定位于 5-hmC 阳性的神经细胞中。AF9 通过识别含有 AAC 的作用元件并结合靶基因启动子区域, 并招募 TET2 结合到 AF9 与 TET2 的共同下游基因, 介导 5mC 向 5hmC 的转变并激活这些神经发育关键因子的表达, 从而促进人胚胎干细胞的神经命运决定。

我国科学家发现淋巴细胞特异性转录因子

Aiolos 通过改变 p66Shc 基因染色质高级结构阻止增强子与启动子之间的相互作用,抑制 p66Shc 基因转录,促进肿瘤细胞逃逸失巢凋亡发生远端转移,提出实体肿瘤细胞通过“co-opt”淋巴细胞转录调控因子,模仿淋巴细胞的低粘附性和失巢凋亡抵抗特性,实现远端转移的新观点,证实染色质高级结构这一表观遗传调控在肿瘤发生中的重要作用。在 DNA 甲基转移酶(DNMT)家族中,Dnmt3a 和 Dnmt3b 行使从头甲基化的修饰作用。Dnmt3a 的突变常见于多种血液系统恶性肿瘤,相反,Dnmt3b 在血液系统中的突变鲜见报道。利用 MLL-AF9 诱导的小鼠 AML 模型研究 Dnmt3b 在 AML 中发挥的作用,通过增加细胞中白血病干细胞(LSC)的数量和促进细胞周期的活跃,Dnmt3b 的缺失加速 MLL-AF9 白血病进展。Dnmt3b 的缺失能够与 Dnmt3a 缺失协同促进白血病发展,本研究对 DNA 甲基化转移酶在白血病发生中的作用提供了新的见解。

一直以来人们认为端粒的调控和细胞代谢相关的线粒体功能调控是两条独立的细胞衰老调控通路。TIN2 蛋白能被召集到端粒中,与多个端粒调控因子,比如 TPP1 相互作用。TPP1 通过与 TIN2 的 N 末端发生反应从而调控 TIN2 的定位。我国科学家们的研究发现,端粒体蛋白 TIN2 不仅仅对端粒功能的实现具有重要的作用,TIN2 还可以在线粒体中进行翻译后修饰,并且能调控线粒体氧化磷酸化过程。通过 RNAi 敲除减少 TIN2 的表达,就会抑制糖酵解的发生和氧自由基的生成,并增加了细胞中 ATP 水平和耗氧量。而 TIN2 在端粒和线粒体间的定位则依赖于 TPP1 的调控。这些研究结果表明在端粒蛋白和代谢调控之间存在直接的关联,是端粒蛋白继调控癌症和衰老的作用之外的又一重要作用机制。

颅面骨由头部神经脊细胞发育分化而成,涉及神经脊细胞的命运决定、迁移和分化等重要过程,其中任何环节的异常都会导致颅面畸形。BMP 配体表达于咽区来自内胚层的咽囊,而其拮抗因子 Noggin3 表达于咽囊旁侧的软骨前体细胞中。我国科学家的研究表明,miR-92a 在斑马鱼咽部表达,主要通过维持咽部骨形成蛋白 BMP 的信号强度而保障咽部软骨的正常发育。该研究不仅阐明了 miR-92a 在软骨发育中的重要作用,还进一步说明 BMP 信号在咽部软骨形成过程中必须受到严格的调控,

其活性高于或低于生理水平,都会导致严重的发育缺陷,这一项研究结果揭示了导致颅面畸形的新分子机制。

增强子是一类控制基因表达的重要调控元件,我国科学家发现在肿瘤发生过程中由 H3K4me3 和 H3K27Ac 所共同标记的增强子可处于“过度活化态”,同时发现两个潜在的肿瘤抑制因子 RACK7/KDM5C 复合体是该过度活化态的负调控因子,RACK7 是一个潜在的染色质阅读器,KDM5C 是一种组蛋白去甲基化酶,二者协同作用通过控制活化增强子 H3K4me1 与 H3K4me3 的动态,充当了活化增强子的“刹车”。丧失这样的增强子监管机制可以导致细胞行为改变,有可能促成了肿瘤发生<sup>[9]</sup>。

#### 4 深化对表观遗传信息网络起源与进化的认识

(1) 揭示植入前胚胎中全基因组水平组蛋白修饰的分布及变化规律。表观遗传重塑对受精过程中高度特化配子的重编程和胚胎发育都十分重要。这些表观修饰的变化是胚胎基因组激活及第一次细胞谱系分化的关键。组蛋白的转录后修饰直接调控了基因表达的激活和沉默。早期利用抗体免疫荧光染色的方法发现,大部分的组蛋白修饰在植入前胚胎的发育过程中都发生了明显的变化。而一些调节组蛋白修饰的酶的异常表达或缺失会导致胚胎发育异常甚至植入前胚胎的死亡,表明组蛋白修饰的变化在早期胚胎发育的过程中起了很重要的作用。但是在植入前胚胎中这些组蛋白修饰在基因组上是如何分布及变化的,这些变化如何调控胚胎基因的表达以及第一次细胞命运的分化依然不清楚。我国科学家通过改进适用于低起始量细胞的 ULI-NchIP (ultra-low-input micrococcal nuclease-based native ChIP) 技术,利用极少量的细胞检测了小鼠植入前胚胎发育各个时期的组蛋白 H3K4me3 和 H3K27me3 修饰变化情况,这两个修饰分别对应基因的激活和沉默,这一研究率先从全基因组水平上揭示了小鼠植入前胚胎发育过程中的组蛋白 H3K4me3 和 HK27me3 修饰建立过程,并发现宽的(broad)H3K4me3 修饰在植入前胚胎发育过程中对基因表达调控发挥重要作用。研究发现,基因组激活前(2 细胞前),不转录的卵子以及基因组激活前的 1 细胞、2 细胞早期,H3K4me3

呈现一种非经典形式(non-canonical H3K4me3),并且大量出现在非启动子区(non-promoter region),比如基因间区(intergenic region)。通过在卵子中过表达去甲基化酶 KDM5B 来去除 H3K4me3,发现了这种非经典的 H3K4me3 可能对于卵子的基因组沉默(而不是激活)是必需的<sup>[10]</sup>。这些研究展示了受精后及胚胎发育早期组蛋白修饰的变化细节,以及染色质开放程度对基因表达的调节和表观遗传信息是怎样在亲代和子代之间传递的,对研究胚胎发育异常、提高辅助生殖技术的成功率具有重要意义。

(2) 构建全基因组甲基化图谱,揭示表观遗传修饰的遗传与进化规律。DNA 甲基化作为重要表观遗传机制调控基因的表达,从而影响一系列的生物学过程,如细胞命运决定、发育和组织、器官的稳态维持。DNA 甲基化以多种修饰方式(5-methylcytosine (5mC), N6-methyladenine (6mA) 和 N4-methylcytosine(4mC)等)广泛存在于细菌、真核生物中。5mC 和其去甲基化过程中的衍生物 5hmC 在哺乳动物基因组 DNA 中被认为是碱基甲基化形式。与之不同的是,6mA 以较高丰度存在于原核生物及一些低等的真核生物。尤其在细菌中,

6mA 修饰在 DNA 复制、修复、基因表达调控及宿主-病原体相互拮抗等方面发挥重要的功能。我国科学家发现了果蝇基因组中存在 6mA 修饰<sup>[11]</sup>,并且证明该修饰在胚胎发育的早期阶段受到去甲基化酶 DMAD(果蝇 Tet 同源蛋白)的精确调控,揭示了真核生物 DNA 新修饰形式。应用 Methy-Seq 技术构建了一张昆虫的单碱基分辨率甲基化谱——家蚕丝腺甲基化谱。

我国科学家利用 methylC-seq 的方法绘制了斑马鱼配子和早期胚胎单碱基分辨率全基因组图谱,研究了斑马鱼亲代 DNA 甲基化图谱遗传到子代中的规律<sup>[12]</sup>,发现父源的 DNA 稳定保持着精子的甲基化图谱,而母源 DNA 抛弃卵子甲基化图谱,并重编程为精子图谱用于调控胚胎的早期发育。在随后的对斑马鱼甲基化遗传的研究证明:除了 DNA 序列被遗传外,表观遗传信息(DNA 甲基化图谱)也可以完整地遗传到子代中。表观遗传信息也可以遗传到子代,意味着在调控动物发育、表型、甚至疾病等方面,表观遗传信息的变异可能和遗传信息的变异一样起重要作用,引发了关于表观遗传信息是否对进化起驱动作用的新思考(拉马克主义)。同时,他们在哺乳动物 DNA 甲基化重编程的研究中发现,

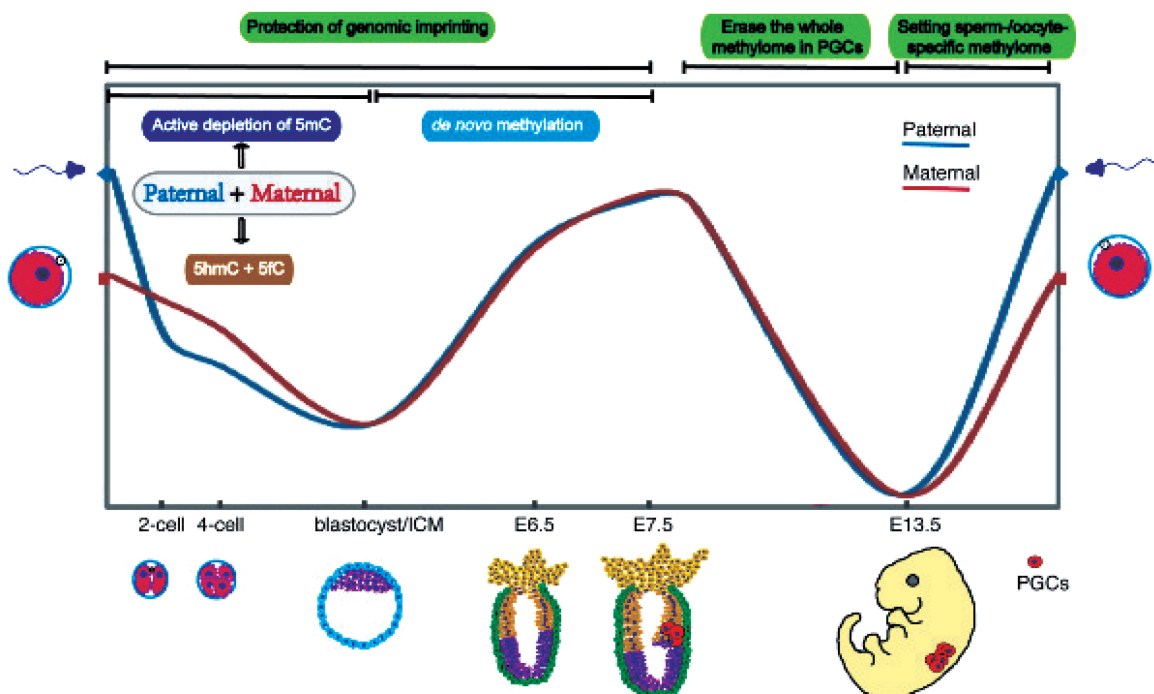


图 4 不同物种全基因组甲基化图谱



在哺乳动物早期发育,甲基化的氧化产物在父源和母源基因组中都存在,因此无论父源还是母源 DNA 都存在主动去甲基化的方式<sup>[13]</sup>。

### (3) 建立单细胞转录组测序及分析强大工具。

单细胞 RNA 测序 (single-cell RNA-sequencing, SCRS) 是分析单个细胞或微量 RNA 中基因组表达的一个强有力的技术。与微阵列技术相比, SCRS 能检测出更多的转录组,灵敏度更高;既能分析同一基因的多个转录本及其对应的蛋白类型,也能检测已知基因中新的剪接点;还具有准确度高、噪音低等优点。SCRS 有助于精准展示细胞编程重编程分化过程中的细节变化。近年来研究人员开始利用这一技术来克服研究中起始样本量少的瓶颈。我国科学家利用自主研发的单细胞转录组分析技术,同时运用加权基因共表达网络分析(WGCNA),揭示出了成年小鼠前脑神经发生区域 CD133+/GFAP-室管膜(E)细胞的分子特征,发现室管膜 CD133+/GFAP-休眠细胞独特基因网络中的重要枢纽基因,包含较多的免疫应答基因以及血管生成因子受体编码基因。给予血管内皮生长因子(VEGF)可激活侧脑室以及第四脑室 CD133+室管膜神经干细胞(NSCs),加上碱性成纤维生长因子(bFGF)则诱导了随后的神经谱系分化和迁移。研究结果表明了中枢神经系统整个脑室表面都存在有休眠的室管膜神经干细胞,并揭示出了在损伤之后可让它们激活的丰富信号<sup>[14]</sup>。他们还通过膜片钳和单个神经元细胞转录组检测相结合,揭示出神经元成熟的分子标记及其与能量代谢的关系。

## 5 展望及未来关注的研究方向

表观遗传学自 20 世纪 80 年代末期开始兴起。进入 21 世纪以来随着技术手段的进步,这一新兴学科呈现出前所未有的快速发展态势。越来越多的生物物理学家、发育生物学家、化学家、生物信息学家与遗传学家一起投身于表观遗传学的探索当中。他们着眼国际学科前沿和发展动态,及时将相关学科的最新研究思想和技术手段应用到表观遗传学研究,极大地提高了人们对表观遗传体系的全面认识。经过本重大研究计划的推动,我国学者在表观遗传学的分子基础以及细胞重编程理论创新都取得了长

足的进步,为实现我国表观遗传学研究的全面跨越式发展做出了突出贡献。

当前细胞重编程研究正处于快速发展的时期,在很多方向上都出现了原理和技术上的重大突破,热点层出不穷,变化日新月异,除了 Yamanaka 因子组合 OSKM 以外,科学家已经找到了更多的多能性重编程因子,如 Nanog, Prdm14, Sall4, Esrrb, Utl1, Tet2, Glis1 等。为了防止这些外源性基因整合入目标细胞的基因组,提高系统的安全性,科学家在利用非病毒载体、RNA、可穿膜蛋白质以及化学小分子实现体细胞重编程和细胞转分化方面也取得了长足的进步。同时,细胞重编程过程中的表观遗传学研究进展非常迅速,最新的灵长类体细胞核移植的成功正是借助了表观遗传因子的使用。然而,现有的研究大多集中于关键因子的发现和功能验证,对重编程细胞中染色体动态以及细胞表观遗传组学变化及其如何调控转录组变化、决定细胞命运转变的机制等依然知之甚少。在这一大背景下,我国科学家应该抓住机遇继续发挥学科交叉的优势,将在研究生命大分子结构和单细胞组学,基因编辑过程中所积累的技术、策略和经验应用到细胞重编程和表观遗传学研究中。紧盯生命科学的前沿问题来开展研究,发展新技术、新方法,推动表观生物学研究从单分子走向多分子,从单一类型修饰走向多种类型修饰,从定性走向定量,从简单体系走向复杂体系,提高细胞重编程过程的效率和质量。在生命科学理论的发展上和技术革新上实现质的飞跃,为人类疾病的治疗带来美好的前景,开辟新的方向。今后关注和重点加强研究方向有:

(1) 继续加强细胞染色质高级结构的研究,特别是解析细胞不同分化状态,不同细胞类型的染色质高级结构组成方式和动态变化。我国在染色质高级结构特别是核小体 30nm 高级结构的总体研究水平已经处于世界前列,30nm 染色质是以 4 个核小体为结构单元扭曲形成的,这种结构单元中的空隙刚好为表观遗传调控提供了一个窗口,为研究表观遗传调控的机理、解释表观遗传的基本问题提供了可能。加强这个方向的研究以保持我国在该领域的优势地位。

(2) 拓展对新发现的表观遗传修饰机制与功能的研究。我国科学家在发现新的表观调控因子方面

取得了重要进展,例如新发现的 RNA m6A 修饰。下一步需要加强这个方向的投入,同时深入研究新发现的表观遗传修饰的机制与功能,研究 DNA 表观修饰和 RNA 表观修饰之间可能的相互作用,加强表观遗传调控细胞分化、组织稳态以及器官发育与再生机理的研究。发展新的模式生物,从物种进化的角度研究表观遗传因子对基因组稳定性和生物体性状的调控机理。

(3) 开辟新技术、新理论和新方法。发展和优化表观基因组编辑技术以及表观转录组单细胞表观遗传信息测定技术,拓展表观遗传网络的生物信息学算法,在单细胞水平上研究表观遗传组学和转录组学之间的网络互作。加强表观遗传调控机制在相关重大疾病中致病机理的研究,加强与临床医学的交叉合作。

(4) 在单细胞水平上,重点研究 OSKM 及其他多能性因子在人源细胞重编程过程中对于细胞表观遗传组学,转录组学的动态影响。提高细胞重编程的效率,提升重编程细胞的质量和重编程体系的稳定性、均一性,为基于细胞重编程的再生医学以及 iPSC 和转分化细胞的临床应用打下基础。

(5) 加强对表观遗传对环境信号的响应及记忆研究,深入揭示其作用机制。生命机体内的表观遗传体系可以使得拥有相同基因组的细胞在不同环境条件下呈现不同的表观基因组、转录组,从而分化出不同的形态功能。目前发现组蛋白修饰具有继承性,受到表观遗传修饰酶反馈调节,预示着生物存在着表观遗传的响应和记忆体系,然而目前对其认识还不多。

**致谢** 本文是依据“细胞编程与重编程的表观遗传学机制”重大研究计划的总结报告,成果报告和战略研究报告整理而成,在此感谢所有参加重大研究计划工作和材料准备的专家。由于篇幅所限,文中仅为枚举,未能列出所有优秀成果,敬请谅解。文中所列的未来重点加强的方面和关注方向,为一己之见,仅供参考。

## 参 考 文 献

- [1] Song F, Chen P, Sun DP, et al. Li. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science*, 2014, 344 ( 6182 ): 376—380.
- [2] Guo X, Wang L, Li J, et al. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature*, 2015, 517(7536): 640—644.
- [3] Hu L, Lu J, Cheng J, et al. Structural insight into substrate preference for TET-mediated oxidation. *Nature*, 2015, 527 (7576): 118.
- [4] Li W, Shuai L, Wan H, et al. Androgenetic haploid embryonic stem cells produce live transgenic mice. *Nature*, 2012, 490(7420): 407.
- [5] Li X, Cui XL, Wang JQ, et al. Generation and application of mouse-rat allodiploid embryonic stem cells. *Cell*, 2016, 164(1): 279—292.
- [6] Zhong C, Yin Q, Xie Z, et al. CRISPR-Cas9-mediated genetic screening in mice with haploid embryonic stem cells carrying a guide RNA library. *Cell Stem Cell*, 2015, 17: 221—232.
- [7] Zhao Y, Zhao T, Guan J, et al. XEN-like state bridges somatic cells to pluripotency during chemical reprogramming. *Cell*, 2015, 163(7): 1678—1691.
- [8] Huang P, He Z, Ji S, et al. Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature*, 2011, 475(7356): 386.
- [9] Shen H, Xu W, Guo R, et al. Suppression of enhancer overactivation by a RACK7-histone demethylase complex [J]. *Cell*, 2016, 165(2): 331—342.
- [10] Zhang BJ, Zheng H, Huang B, et al. Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature*, 2016, 537(7621): 553—557.
- [11] Zhang GQ, Huang H, Liu D, et al. N 6-methyladenine DNA modification in *Drosophila*. *Cell*, 2015, 161(4): 893—906.
- [12] Jiang L, Zhang J, Wang JJ, et al. Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. *Cell*, 2013, 153(4): 773—784.
- [13] Wang L, Zhang J, Duan JL, et al. Programming and inheritance of parental DNA methylomes in mammals. *Cell*, 2014, 157(4): 979—991.
- [14] Luo Y, Coskun V, Liang A, et al. Single-cell transcriptome analyses reveal signals to activate dormant neural stem cells. *Cell*, 2015, 161(5): 1175—1186.

## Review of the achievements of major research plan on “Epigenetic Mechanisms Underlining Cell Programming and Reprogramming”

Zhang Ru<sup>1</sup>      Zhao Tianyu<sup>2</sup>      Tian Yanyan<sup>2</sup>      Gao Shaorong<sup>1</sup>  
Du Shengmin<sup>2</sup>      Gu Ruisheng<sup>2\*</sup>

(1. School of Life Science and Technology, Tongji University, Shanghai 200092;

2. Department of Life Sciences, National Natural Science Foundation of China, Beijing 100085)

**Abstract** The NSFC major research plan of epigenetic mechanism underlining cell programming and reprogramming has experienced 8 years and terminated in 2017. This paper summarized the research progress of this plan and presented a series of breakthrough research achievements, such as discovering new epigenetic regulation factors and chromatin remodeling factors, interpreting their biological function and acting mechanism, giving new insights into the regulation mechanism of stem cell self-renewal and somatic cell reprogramming, creating novel methods of haploid embryonic stem cell induction and semi-cloning, revealing the epigenetic regulation mechanism of cell differentiation and trans-differentiation, individual development and disease incidence, and understanding epigenetic network and its operation in whole genome level. This paper also prospected the development of epigenetic research and put forward some key academic fields that should be pay more attention to in future.

**Key words** somatic cell reprogramming; epigenetic inheritance; chromatin remodeling; stem cells; Major Research Plan of NSFC