

2

基因调控研究进展

2.1 与多种疾病相关的miRNA

为了寻找疾病诊断和治疗的新方法，生物学家与各大公司展开了激烈竞争，希望在破译RNA“疾病密码”的角逐中处于领先地位。

生物学家发现，面对RNA分子，随着研究的深入，更广阔的未来世界呈现于眼前。RNAi现象的发现得到2006年诺贝尔奖的垂青。RNAi是指在进化过程中高度保守的、由dsRNA诱发的、同源mRNA高效特异性降解的现象。简单来说，就是使用短链RNA对特异基因进行靶向沉默。RNAi现象的发现也让人们认识到，细胞本身就具有利用双链microRNA（miRNA）对基因表达进行调控的功能。

miRNA是一种进化上十分保守的、长度为18~25个核苷酸的小分子。最近，miRNA由于在疾病发生中所具有的令人吃惊的重要作用而成为生物医药领域研究的焦点。一系列的研究都证实，miRNA与多种疾病相关，如癌症、糖尿病、心脏病等，同时也为我们提供了一个发掘疾病相关靶点的有效手段。

耶鲁大学（Yale University）发育生物学学家Frank Slack曾对癌症相关的miRNA进行研究，并发现这些小分子与阿茨海默病（Alzheimer's disease, AD）和寿命长短之间具有相关性。当机体内的miRNA表达下调时，就会更容易罹患癌症和其它一系列疾病。这一现象能够使我们通过监控miRNA表达来实现疾病的早期诊断。此外，来自病毒的miRNA有利于病原体在宿主体内定居，因此这些miRNA也可以作为抗病毒药物的新靶点，从而有

助于新型药物的开发。

那些较早发现miRNA及其相关作用的生物学家，现正致力于将他们的发现转变成成为新型治疗方法，同时还将获得巨大的经济效益。许多生物学家已经独立地或者和生物技术公司联合成立了公司，进行这一方面的研究与开发。几年前，以不同类型的RNA分子，尤其是以RNAi技术或其它反义技术为基础的公司，现在都纷纷扩展至miRNA的领域中来。

目前，主要有两种以miRNA为基础的治疗策略可供考虑：一是将可以促进健康的小分子注入体内，二是利用小分子将导致疾病发生的基因进行靶向沉默。与其它新的治疗方法类似，在将miRNA用于人体试验之前，首先需要解决该方法本身存在的一些问题。目前，将小分子准确注入靶向细胞，仍然是一个技术难题。此外，还须确保miRNA作用的目的基因，在其表达情况被改变后，不会出现不利于机体的影响。很多miRNA都具有多个靶点，单独一个miRNA就可能调控数十或上百个基因。

此外，从生物学家在动物试验结果来看，哪怕提高或降低某一个miRNA的表达水平都会显著地影响下游基因的表达。这也是我们在进行miRNA治疗研究时需要慎之又慎的又一主要原因。许多研究人员都相信，经过一段时间，miRNA最终将给疾病治疗领域带来新的曙光。

将miRNA领域的新发现应用于临床

公司名称	地点	研究重点	成立年份
Asuragen	美国得克萨斯州奥斯汀	癌症诊断	2006
Crogen Pharmaceuticals	美国宾夕法尼亚州费城	癌症	2004
Miragen Therapeutics	美国科罗拉多州玻尔	心血管及心肌疾病	2007
Regulus Therapeutics	加利福尼亚州卡尔斯巴德	病毒性疾病及癌症	2007
Rosetta Genomics	以色列雷霍沃特 以及新泽西州泽西市	癌症	2000

心脏病患者的曙光

研究人员已经从约25000个人类基因中鉴别出了约500个能编码miRNA的基因，此外，他们还不断鉴别出新的miRNA编码基因。来自加州大学（University of California）Gladstone心血管疾病研究所的Deepak Srivastava，目前正在研究心脏发育和相关基因传导的通路。他发现当信号通路被阻断时，就会导致先天性或成人心脏病的发生。

其实早在1993年，首次在线虫中发现miRNA后，Srivastava就开始寻找心脏中高水平或低水平表达的miRNA。他先后在小鼠、果蝇、乃至人体内进行了研究，共鉴定出约10个相关miRNA。然后，他开始对这些小分子一一进行深入研究，以确定这些小分子表达水平的改变是否与心脏功能或发育过程的改变相关。Srivastava等人建立了敲除了某个特定miRNA——miR-1-2的小鼠模型，并以此为研究对象进行研究。相关结果发表于2007年4月的《细胞》（*Cell*）杂志上。

尽管通常miR-1-2在心肌中高表达，但Srivastava的研究小组并不认为miR-1-2敲除小鼠会与正常小鼠有很大的区别。这是因为在另一条染色体上还有一个与miR-1-2功能相同的“双胞胎”DNA序列，而研究人员并未将其敲除。当研究人员敲除上述序列的一个拷贝后，小鼠体内的miRNA表达量下降了50%，而由其导致的相关改变也十分显著：一半的基因敲除小鼠死于心脏出现的漏洞，其它小鼠的心肌也出现了致命的损伤（*Science*, 27 April 2007, p. 530）。

虽然，Srivastava表示这在当时确实是个令人惊讶的结果，但在今天看来，个中道理显而易见，这是因为研究人员明白了miRNA是主调控基因（master regulator）的调控因子，即miRNA是整个基因表达通路的调控者。在这项试验中，小鼠心肌出现的损伤基本涵盖了人类最常见的心肌损伤。不过到目前为止，还没有实验能够证实人类心肌损伤也是由miRNA缺失而导致。Srivastava目前正打算研究miRNA调控心脏四个腔室形成的机制。

尽管Srivastava的研究只证实了miRNA表达水平的差异可以诱发先天性心脏疾病，但事实上，这些小分子对成年人的心脏同样具有一定的影响。德克萨斯西南医学中心（Texas Southwestern Medical Center）分子生物学家Eric Olson及其博士后同事Eva van Rooij利用小鼠对miRNA与心肌肥大等心脏疾病进行研究。心肌肥大患者的心脏腔壁会增厚，心脏往往进行着代偿性泵血，并最终导致心脏功能衰竭。目前对于这种疾病还没有非常有效的治疗方法。

这两位研究人员在小鼠体内鉴定出了一种miRNA——miR-208，其DNA编码序列位于编码肌球蛋白（myosin）的基因内。myosin的功能是帮助心脏进行收缩以及在应激状态下维持正常的功能。最近，他们发现许多myosin基因中都包含编码相互进行调控的miRNA的序列，这些miRNA互为调控因子，从而保障心脏的健康。van Rooij指出，这种miRNA间的相互作用和联系还没有其他研究小组进行过类似报道。

“通过操控miR-208等miRNA，我们的确可以开始用一种全新的方式去攻克疾病。” William Marshall预测。

在心脏病学领域进行miRNA研究有一个优势，那就是心脏病医生都有向靶向器官直接给药的经验，比如，将药物注射到冠状动脉。因此，部分研究人员预测，同时将miRNA或miRNA抑制剂一起注射到心脏应该是可行的。Markus Stoffel是瑞士联邦技术研究所的分子生物学家，他认为：“我们将会心脏病领域看到首批以miRNA为基础的治疗。”

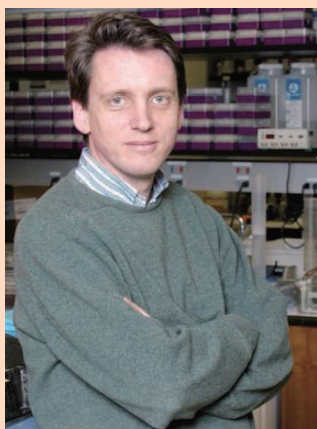
靶向治疗

Markus Stoffel等研究人员希望，miRNA不会重演上世纪90年代反义技术用于临床治疗研究中的让人失望的结局。像miRNA治疗一样，反义技术同样是通过调控基因表达发挥作用，但反义技术是以mRNA为靶点的，而mRNA的功能是将基因翻译成蛋白质。因此，反义技术是一种阻断蛋白质合成的方法。然而，约翰霍普金斯大学（Johns Hopkins University）遗传学和分子生物学家Joshua Mendell发现，反义RNA在抑制mRNA方面的表现并不是十分突出。Joshua Mendell目前的研究重心是miRNA与癌症。

对于抑制miRNA过表达的问题，研究人员目前采用的是Stoffel在2005年设计的一个方案。Stoffel对糖尿病和新陈代谢过程中的miRNA进行研究时发现，他需要一个可以将RNA序列导入细胞的方法，从而沉默miRNA。为此，Stoffel和来自马萨诸塞州Alnylam Pharmaceuticals（一家RNAi生物技术公司）的Muthiah Manoharan一起，开发出了一种名为“antagomirs”的可以沉默miRNA的化合物。之所以这样命名，是因为这些“antagomirs”会拮抗miR或miRNA的表达。这些被称作“antagomirs”的试剂是与胆固醇分子偶联的寡聚核苷酸，它有助于沉默子（文后小词典1）进入细胞。当研究人员将antagomirs注射进小鼠的末梢静脉后，antagomirs会随血液循环进入小鼠全身，并且改变小鼠体内多种器官的miRNA的表达水平。但antagomirs无法通过血脑屏障（文后小词典2），因此研究人员将其直接注入小鼠大脑，这样antagomirs便可以进入大脑细胞了。Stoffel指出，在他们将antagomirs注入小鼠体内的第三天，小鼠体内的靶向miRNA消失了，而且在数周内都没有再出现。他推测这是由于antagomirs可以在细胞内存留一段时间，从而抑制了新的miRNA在细胞内的表达。

Srivastava表示，现在还不确定是否会继续使用antagomirs作为沉默miRNA的方法，因为antagomirs中的胆固醇会损害机体肝脏。尽管在动物实验中，这种方法出现明显的副作用，但是Srivastava仍然怀疑这种方法是否适用于人体治疗。

与上述治疗策略刚好相反的是，可以通过促进miRNA的表达而不是抑制表达来对疾病进行治疗。这一治疗策略同样在应用上存在着巨大挑战。Stoffel表示：“当我们将miRNA进行过表达时，其作用并非像我们所预计的那样。”目前考虑的策略包括：使细胞摄入合成的RNA前体（pre-RNA），然后在细胞内被加工成为成熟的miRNA。但是，目前还没有明确的方法将pre-RNA导入细胞。若希望研制出提高编码miRNA基因的活性的小分子药物，无疑还有很长一段路要走。



分子生物学家Markus Stoffel是较早发现小RNA与糖尿病之间关联的科学家之一。

从胰腺到肝脏

Stoffel在他的研究中还鉴定出了可能在糖尿病发病机制中发挥重要作用的miRNA。Stoffel对小鼠胰腺的胰岛素分泌细胞—— β 细胞进行分析，以寻找高表达的miRNA。结果，Stoffel等人一共筛选出10个相关的miRNA，他们将目光集中在miR-375—— β 细胞中表达水平最高的miRNA上。Stoffel等人的研究发现表明，miR-375可以提高胰岛在应激状态下的调节能力，以适应相应的机体状态，如怀孕、肥胖等，从而促使机体分泌更多的胰岛素。

miRNA领域的研究者还在考虑如何利用这些小分子对抗病毒感染。杜克大学（Duke University）的分子病毒学家Bryan Cullen对入侵人体的病毒如何与机体内存在的RNA干扰机制相互作用进行了研究。目前，他的研究重点又转向了由病毒产生的miRNA是否有利于这些入侵的病原体在宿主内的存活的研究。Bryan Cullen认为：“很多病毒，比如疱疹病毒家族中的成员，都会产生许多miRNA，而其中Epstein-Barr病毒（EB病毒）产生的miRNA数目最多，达到23个。”

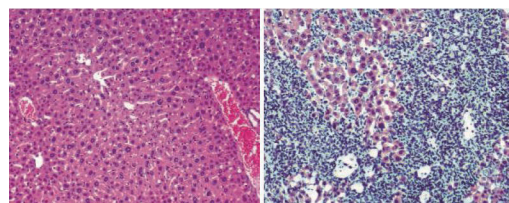
法国和德国的基础研究表明，令这些病毒的一些miRNA基因失活，可以降低病毒对人体的损害性。“但是我们还不明确这一现象的相关机制。”Cullen说。Cullen与Regulus公司合作，想要明确miRNA对于携带它们的病毒而言重要性到底有多大。Alnylam公司的首席执行官 John Maraganore说：“Regulus公司的第一个项目是沉默人类肝脏中一个miRNA，这个被沉默的miRNA有可能具有促进HCV（丙型肝炎病毒）复制的功能。

寻找癌症发生的第一步

miRNA与疾病之间的关系，首先是在癌症中发现的。对人类肿瘤组织进行广泛扫描分析后发现，miRNA在癌细胞中的表达与在相同器官的正常组织内的表达有所不同。此外，通过对癌症患者体内的miRNA表达情况进行回顾性研究，表明miRNA的表达模式有一些过表达，而另一些则不表达或低表达，并与疾病预后相关。

通过动物模型的研究，科学家们发现，在某些癌症病例中，miRNA处于不表达或低表达水平，如能提高miRNA的表达水平，就可以抑制癌细胞的生长。但是首先需要保证的是这些提高了表达水平的miRNA不会对正常组织（已经有足够的miRNA表达）造成不良影响。

在小鼠中的研究发现，一些miRNA与癌症发生有关。哥伦比亚俄亥俄州立大学（Ohio State University）的Carlo Croce是研究肿瘤与miRNA领域的先驱者之一，其他研究者也发现如果人为的上调或下调特定的miRNA将会诱导或促进肿瘤的发生。Carlo已经成立了一家名为Crogon的医药公司，提供与miRNA相关的肿瘤疾病的诊断、预测和治疗产品。于2000年成立于以色列的Rosetta基因组公司是第一批miRNA公司，目前他们也把所有的精力投入miRNA相关的肿瘤研究中。



癌症快速诊断

小鼠免疫细胞中某种miRNA的过表达可以导致小鼠罹患白血病；体内不含多余miRNA的小鼠（左图）的肝组织（粉色）是正常的，而miRNA过表达的小鼠（右图），白血病细胞（蓝色）进入了小鼠肝组织。

为了确定miRNA在肿瘤中所扮演的角色，迫切需要经过基因改造的动物，它们生来就缺乏特定的miRNA。来自约翰霍普金斯大学的Mendell说：“我们需要做更多的工作来弄清miRNA在肿瘤中的作用及其作用机制。”

这项工作听起来容易做起来却困难重重。比如，Mendell曾经发现在肿瘤细胞中通常都具有活性的癌基因myc的产物蛋白，它可以下调许多miRNA。但这远远不是事实的全部，因为miRNA的基因网络是超乎想象的复杂。尽管像Mendell说的那样，一些癌基因的表达产物会对miRNA进行调控，但反过来许多miRNA也可以调控癌基因的表达。结果，在一系列的相互作用后，miRNA对肿瘤起了显著的影响。

充分利用miRNA治疗肿瘤以后，许多研究者又把目光转向了研究miRNA在诊断方面的潜力，他们期望可以实现早期诊断或取代无效的传统检测方法。过往，在2%到4%的癌症病例中，转移肿瘤的检测是难点，但现在则有望通过miRNA的表达模式的变化进行相关检测。2005年，Dana Farber癌症研究所Todd Golub等人在《自然》(Nature)上发表文章，报道了他们对依照形态学无法分类的17种肿瘤进行分类的研究工作。他们以肿瘤的miRNA表达模式为基础，对上述17种肿瘤中的12种进行了准确分类，这12种肿瘤按照肿瘤来源来看，差别十分微小。采用传统的以mRNA为基础的基因表达鉴

别方法，仅仅能准确鉴别其中一种肿瘤。

罗赛塔基因组学协会(Rosetta Genomics)带头人称，他们计划今年晚些时候出售三个miRNA为基础的癌症诊断技术；其中一个技术可以识别未知来源的肿瘤，另外两个可以帮助临床医生区分不同的肺癌。2006年，位于德克萨斯州奥斯汀市的Asuragen公司成立了，现在，该公司的重心是进行将人体体液内的miRNA作为癌症诊断标志的研究。耶鲁大学(Yale University)的Slack将是这一研究的合作者之一。公司高层表示将把癌症作为研究重点，包括肺癌、前列腺癌、大肠癌、乳腺癌和胃癌等。他们还表示说，miRNA为基础的诊断手段的发展还需要若干年的时间。由于许多miRNA在不同癌症中发挥作用时有所重合，因此将它们进行商品化发展的工作，还不至于让人望而却步。最近，Asuragen公司刚刚结束了第二轮的基金申请和募集，共获得1850万美元的资助。

Atlas Venture是一家位于马萨诸塞州的风险投资公司，他旗下的风险投资家Bruce Booth也认为miRNA是当今医学研究中的“热点”，因此，他选择资助Miragen公司的成立和及其进一步发展。当然，到目前为止，还没有人能够肯定地说靶向miRNA研究是否会给患者带来福音——但是如果将来得到的答案是肯定的，那么，这些小分子将极大地改变当前的医疗面貌。

全新视角看待我们的身体

从心脏到血液、到胰腺、再到其它器官，科学家们努力寻找那些可以让我们保持健康或使我们罹患疾病的miRNA。生物技术公司不断涌现，希望将这些研究发现转变成为可以对疾病进行诊断、治疗、预测的新方法。

原文检索：<http://baike.baidu.com/view/206870.htm>
www.sciencemag.org

 YORK/编译

小词典

1 沉默子: 沉默子 (silencer) 是参与基因表达负调控的一种元件, 是在研究T淋巴细胞的T抗原受体 (TCR) 基因表达调控时发现的。TCR有 α/β 、 γ/δ 两种, 编码 α 链和 δ 链是同一基因座的两个等位基因, α 链恒定区Ca基因3' 端下游的。增强子在各种T细胞中都有可能表达。因此, 在T细胞分化时, 需要通过各种正负调控机制进行精确的选择, 方能正确地表达。已知有3个负调控元件, 一个紧接在Ca基因的3' 下游, 另一个则在Co增强子的上游。对 α 基因专一的沉默子的作用是阻止ja基因在 γ/δ 型 T 细胞中参与重排, 使 ja 基因只参与 α/β 型T细胞中的重排。这说明 α 沉默子对在 α 和 δ 两个等位基因中选择。等位基因在转录和重排时起作用, 从而使 δ 等位基因沉默。这些负调控元件不受距离和取向的限制。

2 血脑屏障: 血脑屏障 (blood-brain barrier, BBB) 是指脑毛细血管阻止某些物质 (多半是有害的) 进入脑循环血的结构。血液中多种溶质从脑毛细血管进入脑组织, 有难有易; 有些很快通过、有些较慢、有些则完全不能通过。这种有选择性的通透现象使人们设想可能有限制溶质透过的某种结构存在, 这种结构可使脑组织少受甚至不受循环血液中有毒物质的损害, 从而保持脑组织内环境的基本稳定, 这对维持中枢神经系统正常生理状态具有重要的生物学意义。

2.2

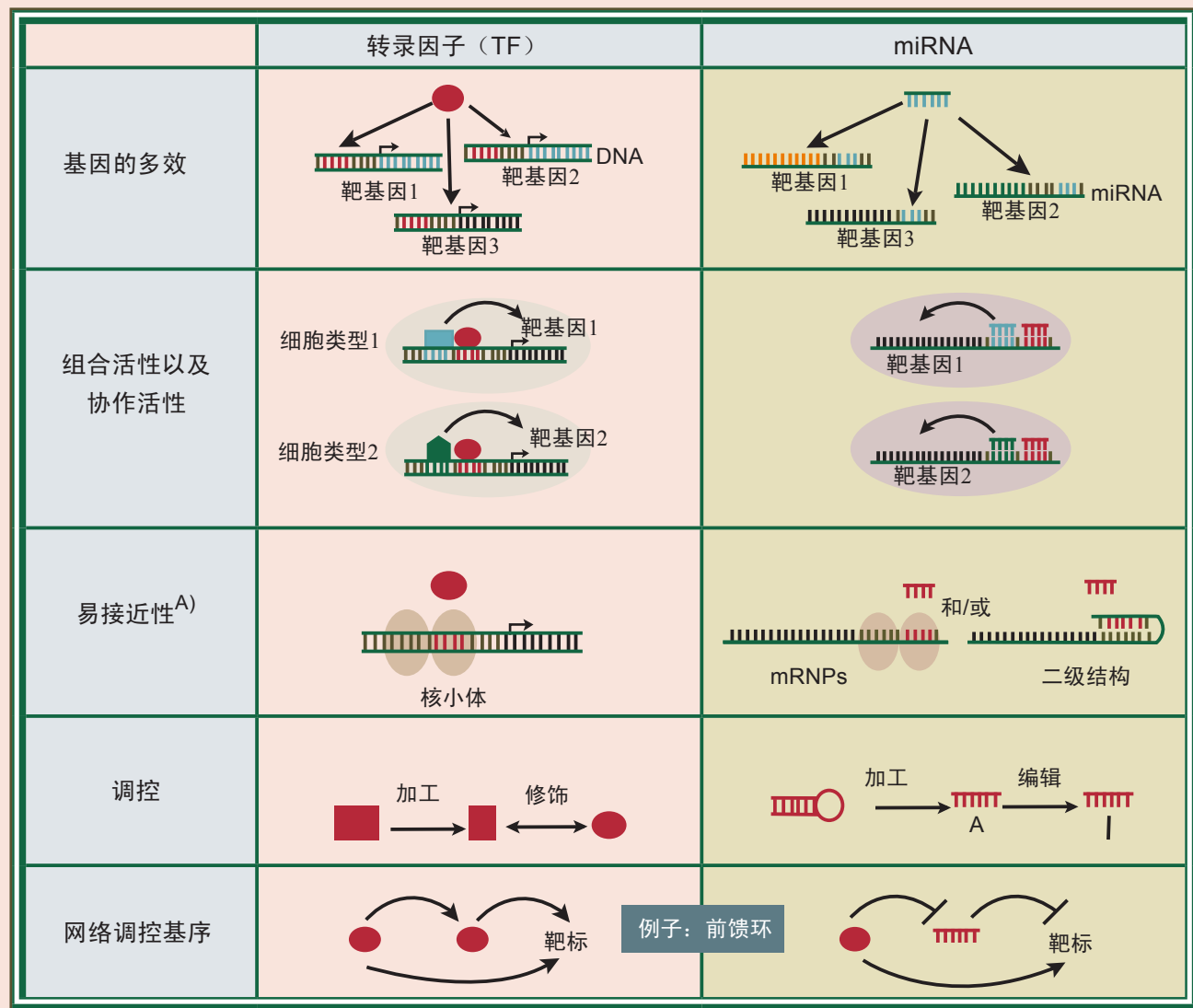
转录因子及小分子RNA

介导的基因调控

细胞的特性是由编码在其基因组中的遗传信息所决定的。分子生物学目前所面临的一大挑战是理解这些遗传信息如何动态地介导细胞的不同形态以及不同状态。在多细胞物种中, 基因组信息的基因表达调控因子, 可以以多种形式及调控方式呈现, 其中以转录因子 (TF) 与miRNA最为引人注目。

TF与miRNA是是多细胞物种中最大的基因表达调控的反式作用因子家族, 它们有着共同的调控逻辑^[1] (图1)。TF与miRNA的组合表达调控装置造就了精确的细胞形态分类。通过结合不同的顺式作用因子、单个转录因子及miRNA就可以调控, 即使没有几百个, 也至少有几个靶标基因。而且大多数情况下, 同一个基因通常由几个反式作用因子共同调控 (图1)。许多TF结合其同源DNA序列或者招募其它转录因子协作调控基因的表达^[2, 3]。与此类似, miRNA也有着相同的协作调控行为, 这种现象已获得体外报告基因实验证实^[4]。因此, “协作作用”是解释TF及miRNA组合表达模式的基础机制 (图1)。

结合位点的亲和力大小为基因表达调控提供了另一个角度（图1）。在体内，对转录因子结合位点的结合依赖于核小体对位点的覆盖率，这个过程由核小体的定位及重修饰机制来调控^[5]。同样，miRNA识别位点的结合能力是由具有RNA识别基序（RNA recognition motif, RRM）的RNA结合蛋白家族决定^[6]。其它蛋白家族也有可能参与调控miRNA与靶标基因的结合。除了蛋白的调控，miRNA结合位点的亲和力也会受到miRNA靶标序列的二级结构影响^[7]。因此，仅仅有miRNA及其靶标基因的共表达，不一定总是会产生功能结合效应——这点尤其在用计算机预测miRNA/靶标的相互作用时要注意。



^{A)}易接近性（accessibility）是指抗原分子的特殊化学基团与淋巴细胞表面相应的抗原受体相互接触的难易程度。

图1 TF及miRNA介导的基因表达调控中某些共享原则示意图

阻遏作用的重要性

众所周知，TF在转录水平上对基因表达进行正负调控，而miRNA普遍通过阻遏作用来调控基因的表达^[8,9]。阻遏作用是细胞特异性基因表达调控程序中的一个重要机制。由普遍表达的TF所介导的广泛的、非细胞特异性的转录激活事件，通过细胞特异性的转录抑制事件来获得其特异性，从而将某些基因限制在更小范围的特定细胞中表达^[1]。更有甚者，在很多细胞发育过程中，TF的激活效应是通过“双重负调控”调控系统介导完成的，这种作用通过阻断基因转录抑制因子的表达来激活基因的转录^[1,10]。相比较而言，miRNA的阻遏作用在细胞特异性基因表达程序中的调控更为灵活。

调控网络及修饰

TF对细胞特异性基因表达的调控，无论是单枪匹马行动，还是组合进行，都取决于其本身在细胞中特异性的表达。而其本身的这种表达模式又取决于更上游的调控程序。因此，细胞的发育过程是一系列阶段性调控过程连续作用的结果^[11]。最终，基因转录调控程序被置入一个精确的调控网络中，这个调控网络包含了许多微小的特异反馈基序（feedback motif）或前馈基序（feedforward motif），这些基序赋予了调控网络系统信号的放大、减弱、持续性，以及来回变动的特殊性能^[12]。TF基因本身会受到某些miRNA的调控，同时miRNA的细胞特异性表达模式大部分又由TF在转录水平介导。因此，在基因调控网络中，miRNA与TF是相互

联系的^[13,14]，在只有TF的调控系统中，主要的基序是前馈和反馈回路基序。

除了在基因表达水平上会受到调控，TF也很明显地在翻译后水平受到调控。例如在蛋白的磷酸化、组装，以及定位过程中，miRNA调控的基因表达也常遭受转录后水平的调节（图1）。例如miRNA从其前体转变为成熟体的过程在多种细胞中都受到调控^[15,16]。又比如通过RNA编辑（RNA editing, 是指在mRNA水平上改变遗传信息的过程）来修饰miRNA进而调节其活性是某些细胞特定的行为模式^[17]。参与miRNA调控作用的某些蛋白的翻译后修饰提高了miRNA调控功能的复杂性。

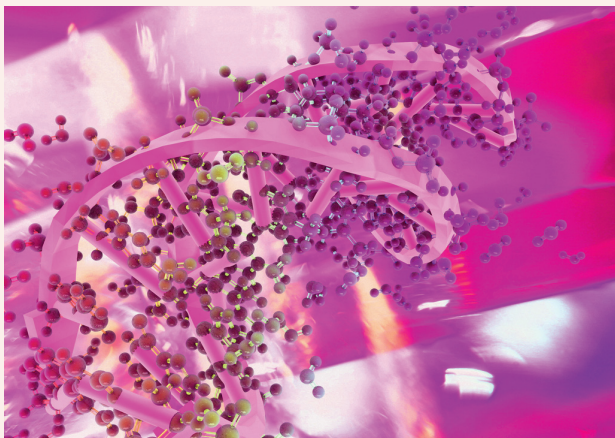
表型图谱

上面提及的内容似乎说明TF与miRNA行使的基因表达调控过程是相似的（图1）。但是，miRNA与TF在基因表达调控中真的同等重要么？已经有充足的遗传证据表明，无论是单细胞物种还是多细胞物种，TF在细胞发育和维持内环境的稳态中都起着非常重要的作用^[11]。敲除掉某些特定动物的miRNA亦会导致表型的显著差异^[18]。

敲除秀丽隐杆线虫（*Caenorhabditis elegans*）80%以上的miRNA位点，结果显示只有不到10%的miRNA的敲除会产生明显的发育或形态缺陷^[19]。相比较而言，利用RNAi技术对线虫中

TF基因进行阻断，结果显示30%的TF基因的缺失会导致十分显著的表型变化^[20]。亲缘关系相近的miRNA家族成员的冗余可能掩盖了其关键作用^[19]。然而，以miRNA *Isy-6*在两种相似的化学感应神经元细胞中表现出不同的功能为例^[21]，miRNA可能更倾向于被用来调控某些细胞的最终分化过程。为了支持这个观点，miRNA往往会在正常发育以及疾病（如癌症）发生过程中呈现出一种高度细胞特异性的表达模式^[22, 23]。有没有miRNA本身的一些固有特点可以解释，相对于TF来说，为何miRNA参与的调控表型变化更有限？

miRNA介导的基因表达调控的特异性



TF靶向的基因调控区域通常非常复杂，并且可以覆盖到几万个碱基，而miRNA靶向的基因3'端非翻译区(3' UTR)平均长度小于1000个碱基^[1]。miRNA这个调控特点与转录因子相比，不仅大大缩小了其靶向基因序列搜索的范围，同时也减少了进化的可能性，减少了新的调节因子出现的可能性，而这恰恰是物种进化的主要驱动力^[24]。由于这种依赖miRNA的基因表达调控机制具有与生俱来的调控序列长度的限制，所以会影响该机制所调控基因的进化。

另一方面，miRNA介导的基因表达调控的固有特性也可能为其提供了某些独特的进化“机遇”。为什么这样说呢？原则上，在限定的进化时间内，基因的表达会因为其启动子内出现一个新的抑制性TF结合位点而被抑制，但这个新的TF结合位点可能又要受到转录激活机制的影响。在进化过程中，与基因在转录阶段进化出新的miRNA相比，在启动子中进化出同时需要几个不同转录因子共同起作用才能抑制转录的机制要困难得多，也就是更加不可能发生进化。因此，通过获得miRNA结合位点，或者重新进化出一个新的miRNA基因，基因表达调控程序的进化获得了一个全新的、独立的、更易运行的机制，并获得了多样性。总的来说，实际上miRNA的调控作用可以仅仅是针对转录调控程序的结果作进一步的修饰，因为它们是对mRNA进行作用，而且miRNA作用区域范围的狭窄性可以解释，相比较转录因子的调控而言，为何它所调控的表型图谱更为简单。

速度和可逆性是另外两个miRNA所介导的基因表达调控的独特之处。为使基因在转录水平受到抑制，一个更精细的调控机器必须定位在细胞器——细胞核内，但是，这与产生这些调控蛋白的空间——细胞质完全分离。另外，这些已经转录生成的mRNA的稳定性亦会降低其调控的速度。相反，miRNA可以在蛋白合成的场所——核糖体中快速关闭蛋白的合成。此外，miRNA的小巧及非翻译特性，使得其产生速度远远快于转录调控蛋白，因而大大降低了阻遏反应时间。最后，miRNA所抑制的靶标基因比在转录调控中受到抑制的基因重新激活的速度要快，因为miRNA靶标基因的重新激活只需要简单地将其mRNA的定位改变到另外一个具活性的核糖体上就可以了^[25-27]。

miRNA介导的基因表达调控与TF介导的基因表达调控在概念上还有一个重要区别，那就是miRNA可以在同一细胞中根据不同的亚细胞定位快速地改变局部的基因表达情况。例如，高分化的神经元细胞只需要在突触局部表达调节基因，而不是在全细胞水平进行表达^[28]。TF不能提供这种亚细胞特异性的调控方式。相反，miRNA介导的基因表达调控却具备这种潜能，因为miRNA是在核糖体中发挥作用的，而核糖体散落分布在各种亚细胞器中，包括神经突触。有报道表明，miRNA确实特异地在神经突触的功能中发挥作用^[26,27]。总之，miRNA调控的速率、可逆性以及局限性注定了miRNA要介导快速的、适应性的调控模式来维持基因表达的动态平衡，并且在特定的环境、营养及神经信号等因素下迅速作出反应。

综上所述，TF与miRNA介导的基因表达调控有许多相似之处，同时也应用于许多不同的情形之下。然而，miRNA调控作用的特异性使其应用于更多特定的调控情境中，这个观点可以在将来的研究中通过对miRNA基因的敲除进行详细而全面的分析。

原文检索：

www.sciencemag.org

<http://baike.baidu.com/view/207994.htm>

 小强/编译

参考文献

1. Hobert, *Trends Biochem. Sci.* 29, 462 (2004).
2. M. Ptashne, A. Gann, *Genes and Signals* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 2002).
3. Grimson et al., *Mol. Cell* 27, 91 (2007).
4. P. Saetrom et al., *Nucleic Acids Res.* 35, 2333 (2007).
5. M. J. Buck, J. D. Lieb, *Nat. Genet.* 38, 1446 (2006).
6. M. Kedde et al., *Cell* 131, 1273 (2007).
7. H. Robins, Y. Li, R. W. Padgett, *Proc. Natl. Acad. Sci. U.S.A.* 102, 4006 (2005).
8. S. Vasudevan, Y. Tong, J. A. Steitz, *Science* 318, 1931 (2007).
9. R. S. Pillai, S. N. Bhattacharyya, W. Filipowicz, *Trends Cell Biol.* 17, 118 (2007).
10. S. Gray, M. Levine, *Curr. Opin. Cell Biol.* 8, 358 (1996).
11. E. H. Davidson, *Genomic Regulatory Systems* (Academic Press, San Diego, CA, 2001).
12. U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman & Hall/CRC, Boca Raton, FL, 2006).
13. R. J. Johnston Jr., S. Chang, J. F. Etchberger, C. O. Ortiz, O. Hobert, *Proc. Natl. Acad. Sci. U.S.A.* 102, 12449 (2005).
14. J. Tsang, J. Zhu, A. van Oudenaarden, *Mol. Cell* 26, 753 (2007).
15. J. M. Thomson et al., *Genes Dev.* 20, 2202 (2006).
16. E. J. Lee et al., *RNA* 14, 35 (2007).
17. Y. Kawahara et al., *Science* 315, 1137 (2007).
18. R. W. Carthew, *Curr. Opin. Genet. Dev.* 16, 203 (2006).
19. E. A. Miska et al., *PLoS Genet* 3, e215 (2007).
20. www.wormbase.org
21. R. J. Johnston, O. Hobert, *Nature* 426, 845 (2003).
22. J. Lu et al., *Nature* 435, 834 (2005).
23. E. Wienholds et al., *Science* 309, 310 (2005).
24. G. A. Wray et al., *Mol. Biol. Evol.* 20, 1377 (2003).
25. S. N. Bhattacharyya, R. Habermacher, U. Martine, E. I. Closs, W. Filipowicz, *Cell* 125, 1111 (2006).
26. S. I. Ashraf, A. L. McLoon, S. M. Sclarsic, S. Kunes, *Cell* 124, 191 (2006).
27. G. M. Schratt et al., *Nature* 439, 283 (2006).
28. K. C. Martin, M. Barad, E. R. Kandel, *Curr. Opin. Neurobiol.* 10, 587 (2000).
29. I thank R. Mann, C. Desplan, N. Rajewsky, and members of my laboratory for comments on the manuscript. I regret that space limitations prevented a more extensive coverage of the literature. Work in my laboratory is funded by the NIH and the Howard Hughes Medical Institute.

科研综述 研究前沿 热点话题
 技术方法 专题译述 生命百态
 会议展览 教学视频



www.LifeOmic.com

2.3

RNA——真核基因组中的重要分子

过去数年研究表明，所有经研究的真核生物基因组几乎都是被整个进行转录的，随之产生了数量相当庞大的非蛋白编码RNA（ncRNA）。同时，越来越多证据表明，这些ncRNA中的许多分子都具有基因调节功能。

RNA是染色体的有机组成部分，此外，它们在染色体结构的形成中具有十分重要的作用^[1,2]。随着研究的深入，人们开始认识到，染色体的结构以及基因表观记忆是由RNA介导的生物学过程调控的。尽管其中的具体机制尚不明了，但是已经可以肯定的是，在该过程中包括组蛋白修饰复合体及DNA甲基转移酶被募集至特异位点的过程^[3]。人们早已熟知长ncRNA在动物体内具有调控剂量补偿和基因组印记的功能^[4]；然而事实上，

这些长ncRNA在生物体的发育过程中，在表观遗传学方面的控制作用上扮演着更为多面的角色^[3]。例如，最近一项研究发现231个与人HOX基因簇相关的长ncRNA在人的发育过程中有相应表达^[5]，研究人员对其中一个名为HOTAIR的基因进行了深入的研究：它由HOXC位点序列转录而来，可以通过募集Polycomb复合体，从而反过来抑制HOXD簇的基因表达^[5]（图1）。其它ncRNA也具有类似的功能，例如目前已证实由珠蛋白和抗原受体位点转录而来的基因间转录产物与复杂的表观遗传学现象相关^[6,7]。

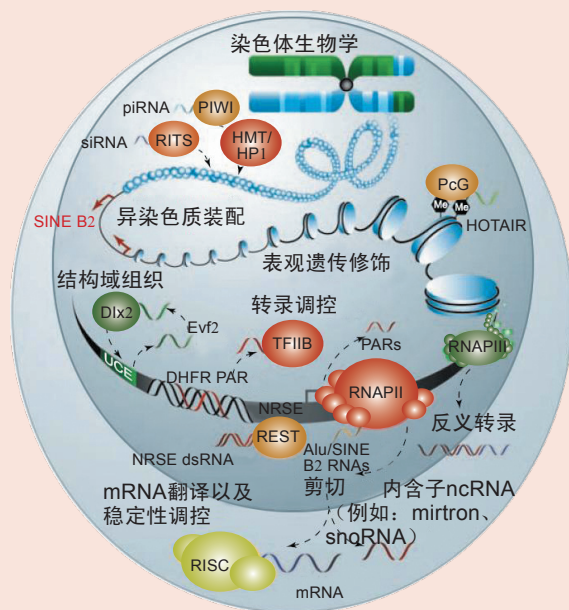


图1 最近研究发现的由ncRNA对真核细胞基因表达及细胞生物学功能方面多个水平的调控机制

dsRNA: 双链RNA; HMT: 组蛋白甲基转移酶; HP1: 异染色质蛋白1; PAR: 启动子相关RNA; PcG: Polycomb群蛋白质; RISC: RNA诱导的沉默复合体; RITS: RNA诱导的转录基因沉默; siRNA: 小干扰RNA; TFIIB: 转录因子IIB; UCE: 超保守元件。

更多ncRNA调控模式图例请参见[3,58]。

小ncRNA通过RNAi途径参与异染色质的形成^[8]，包括Piwi-相互作用RNA（piRNA）^[9]，这些小分子可以引导从果蝇到脊椎动物等生物体内的PIWI家族蛋白质对转座子活性进行调控^[10]。然而，在果蝇的某些末端着丝粒区域，PIWI是常染色质形成所必需的，因此piRNA也可能具有调控常染色质形成的功能^[11]。

更高水平的生物学现象与功能，如细胞核形成、染色质动力学等，也同样受到ncRNA的调控。例如，裂殖酵母中动粒和着丝粒异染色质的形成是依赖细胞周期调控的着丝粒重复源RNA以及RNAi途径的，而动粒组装及染色体分离需要外染色体的核糖核酸酶活性^[12~15]。这些发现表明，在细胞有丝分裂所必经的上述过程之间有以RNA为基础的机制相连。在四膜虫（见文后小词典）内，RNA通过RNAi依赖性Polycomb复合体的募集及组蛋白甲基化从而指导异染色质的形成以及DNA清除过程^[16]。在果蝇内，RNAi途径和定向组蛋白修饰都会对其核仁的形成进行相应调控^[17]。

同样地，长ncRNA在纤毛虫分化过程中指导程序性全基因组的基因重排^[18]。在哺乳动物中，长ncRNA的转录有助于很多生物学过程的进行，

包括T细胞受体重组^[7]、端粒维持^[19,20]、X染色体配对所需的剂量补偿^[21]以及无活性X染色体在核仁外周的定位^[22]。

来自基因组序列中重复元件的ncRNA也可以调控染色质的功能性形成。在小鼠的器官发生过程中，在RNA聚合酶（RNAP）II和RNAPIII的催化下，逆转录转座的SINE B2序列的双向转录将有关的生长激素定位至细胞核中，并在细胞核内确定异染色质和常染色质的边界，并调控相关基因^[23]（图1）。由于所转录的重复序列在基因组内广泛存在，因此有研究者推断这可能是所有真核生物在全基因组水平上对染色质行程进行调控的一个保留的策略^[23~25]。此外，上述结果也表明，基因组的相当一部分序列事实上都是具有功能活性的，来自转座子的序列有可能并非是中性的标志^[26]。

转录



非编码RNA可以通过与转录因子如RNAP或DNA序列本身的相互作用对转录进行调控（图1）。这些小双链RNA NRSE通过激发REST转录机制，来引导包含有上游保留下来的NRSE序列的神经元基因的激活^[27]，这些序列的一部分同时也是miRNA的靶标^[28]。与之类似，ncRNA Evf-2由高度保守的与Dlx-5/6位点相关的增强子转录而来，并且与Dlx-2转录因子相互作用而激活增强子^[29]。的确，从这些研究发现看，在人类基因组中高度保守的遗传元件（基因序列：在很多时候作为增强子行使功能，并在脊椎动物的进化过程中保持不变）^[30]是在既定的调控

方式下进行转录的，这些序列的异常表达很有可能导致机体的病理变化，如癌症的发生^[31]。

此外，包括miRNA在内的启动子导向的序列特异性RNA^[32]，具有诱导^[32,33]或抑制转录的功能^[34,35]。对转录的抑制功能机制包括表观修饰与靶向启动子相关的低拷贝RNA^[34,35]。启动子相关转录产物在人^[37,38]、拟南芥（*Arabidopsis*）^[39]、酵母^[40]等基因组中的广泛存在，提示我们这可能是一个普遍的生物学现象。

RNAPIII可以产生各种类型的调节RNA。在小鼠中，热休克可以诱导RNAPIII催化一个B2 SINE序列元件转录产生一种可以在特定位点对RNAPII产生抑制作用的RNA分子^[41]。这些人类细胞在热休克状态下所同样具有的特点与功能，来自于Alu序列元件的ncRNA——灵长类专有的SINE，它在我们人类的基因组中占10.5%，而且研究也证实SINE作为基因表达转录抑制因子的功能高度保守^[42]。

RNA加工与翻译



miRNA通过介导翻译抑制或降解mRNA（图1）从而调控动植物体内大部分生物学过程，对调节通路产生影响^[28]。然而，miRNA在应激等情况下，也可以促进基因表达的激活，这一点取决于与哪种调节因子相关联^[43]，甚至还可以在细胞周期停滞中产生翻译激活效应^[44]。

miRNA的形成及其靶标特异性可被组织特异性的A-I 编辑（A-I editing）^[45]以及RNA结合蛋白^[46,47]调控。此外，miRNA可能来自于同一发夹结构的正义链和反义链或者同一基因位点的正义和反义转录产物^[48,49]，从而使得单个基因位点也可以转录出具有不同靶标的多种miRNA。

很多miRNA来自蛋白编码基因的内含子，很多情况下是在酶切机制下形成，而不是通过经典的Drosha途径^[50~52]形成，因此，与小核仁RNA一样，其存在很可能比我们预期的要广泛得多。

其它类别的ncRNA也可在转录后发挥作用。在最近的一项研究中，研究者发现了大量RNAPIII转录的短ncRNA与蛋白质编码基因互补，进一步证明了RNAPIII转录产物对RNAPII进行调控的顺义-反义调节网络的存在（图1）^[53]。

非编码RNA的世界

上述研究事例证实了RNA可以以多种机制在各种水平上对基因表达进行调控（图1）。ENCODE报告指出，经分析的人基因组中至少有93%的核苷酸在不同细胞内被转录^[54]，在小鼠^[55]及其它真核生物中的研究也得出了类似结果。这表明有远远超过约1.2%的编码蛋白数量具有生物学意义的RNA的存在。RNA中具有短开放阅读框（ORF）的RNA有可能具有编码蛋白质的功能^[56]，但是很多目前已经注释的ORF并不是保守序列或者可能是错误的，因此意味着人基因组内蛋白质编码基因要大大少于我们的预期^[57]。

一直以来关于这些ncRNA是否具有功能都存在着争议。在某些情况下，可能只有转录产物或仅



仅转录过程，或者两者同时都具有一定的生物学意义。但是，已有大量研究结果表明，相当数目的ncRNA确实具有实质上的生物学功能。例如，有研究发现很多基因序列转录产生在发育过程中具有调控功能的酶切转录产物^[5,55,58]，而且至少存在一些反义和基因间的ncRNA在转录中具有功能^[5,41,53,59]。在大脑的特定区域有大量ncRNA表达，并表现出精确的细胞定位^[60]。还有研究发现在细胞内存在新的

ncRNA区域，表明ncRNA可能对细胞生物学也具有影响。

相关的比较分析表明ncRNA启动子通常比那些蛋白质编码基因在进化上更为保守^[55]，而ncRNA序列、二级结构、酶切位点基序都有可能作为纯化筛选的策略目标^[63,64]。此外，很多ncRNA进化很快，有些经历了正向挑选，如本文中所述的表达于人大脑的HAR1 RNA，这一分子包含有哺乳动物体内的保守序列，但在人—黑猩猩分离开来之后经历了最快的分支^[66]。

基于RNA具有的多种生物学功能，有研究者认为ncRNA可能是真核生物进化过程中一类非常重要的作用底物。为证实这一观点，研究者对多种个体发育过程中器官内的调节性RNA进行研究，从原虫独特的发育途径^[18]到多细胞动物体内对进化枝特异的发育调节因子的调控路径^[5,52]都有所涉及。此外，还有越来越多的证据表明RNA同样具有在亲代与子代之间传递遗传信息的功能，在小鼠^[67]和植物^[68]内可以介导非孟德尔式遗传模式的表观遗传学的改变。

尽管我们都很清楚建立大规模研究ncRNA功能的方法的必要性，但是，我们只需要借助genome browser就可以轻松地看到与大多数我们感兴趣的具有调节功能的基因相关的非编码表达序列标签。例如，最近一项关于与肿瘤抑制因子基因相关的ncRNA研究中，发现了一个与p15基因互补的RNA，它可以改变组蛋白甲基化过程从而沉默p15基因的表达。而这一功能无疑在细胞分化和肿瘤发生中具有非常重要的作用^[59]。

最近，研究逐步证实：ncRNA与癌症^[69,70]、冠心病、糖尿病^[71,72]等多种疾病相关。因此，对于ncRNA功能的阐明无疑将有利于我们对这些复杂疾病的认识与治疗。同时，对ncRNA研究的不断深入，也改变了我们对于多细胞有机体遗传程序的认识。尤其值得我们注意的是，在复杂系统的遗传信息传递过程中，调控环节似乎占据了主导地位^[3]。

原文检索：www.sciencemag.org

 YORK/编译

小词典



四膜虫 (Tetrahymena)：原生动物门寡膜纲膜口目四膜科四膜虫属的通称。已知有10余种。体长40~60微米，呈倒卵形或梨形。口位于腹面前方正中，体表被以纵纤毛带，口后纤毛带一般为2条。胞肛和2个伸缩泡孔均位于细胞后端。无性生殖为横分裂，有性生殖为接合生殖。合子核分裂分化产生新的大小核，两细胞分开、分裂。

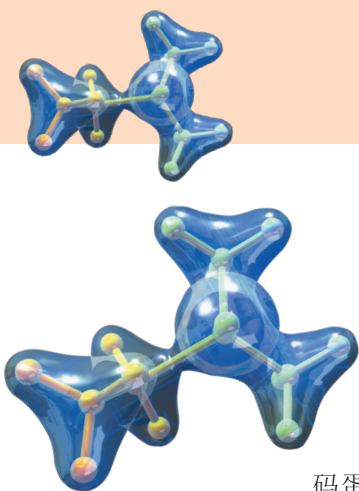
世界性分布，主要产自淡水，也有的生活于咸水或温泉中。四膜虫能在无菌的液体培养基中生长繁殖，长期以来用它为材料做了大量营养生长和药物学方面的研究，是真核细胞基因工程研究的理想材料。

(<http://zhidao.baidu.com/question/14300896.html>)

参考文献

1. J. A. Nickerson, G. Krochmalnic, K. M. Wan, S. Penman, *Proc. Natl. Acad. Sci. U.S.A.* 86, 177 (1989).
2. Rodriguez-Campos, F. Azorin, *PLoS ONE* 2, e1182 (2007).
3. J. S. Mattick, *J. Exp. Biol.* 210, 1526 (2007).
4. P. K. Yang, M. I. Kuroda, *Cell* 128, 777 (2007).
5. J. L. Rinn et al., *Cell* 129, 1311 (2007).
6. H. L. Ashe, J. Monks, M. Wijgerde, P. Fraser, N. J. Proudfoot, *Genes Dev.* 11, 2494 (1997).
7. Abarrategui, M. S. Krangel, *EMBO J.* 26, 4380 (2007).
8. M. Buhler, D. Moazed, *Nat. Struct. Mol. Biol.* 14, 1041 (2007).
9. Brower-Toland et al., *Genes Dev.* 21, 2300 (2007).
10. Aravin, G. J. Hannon, J. Brennecke, *Science* 318, 761 (2007).
11. H. Yin, H. Lin, *Nature* 450, 304 (2007).
12. M. Buhler, W. Haas, S. P. Gygi, D. Moazed, *Cell* 129, 707(2007).
13. H. Murakami et al., *PLoS ONE* 2, e317 (2007).
14. H. D. Folco, A. L. Pidoux, T. Urano, R. C. Allshire, *Science* 319, 94 (2008).
15. E. S. Chen et al., *Nature* 451, 734 (2008).
16. Y. Liu et al., *Genes Dev.* 21, 1530 (2007).
17. J. C. Peng, G. H. Karpen, *Nat. Cell Biol.* 9, 25 (2007).
18. M. Nowacki et al., *Nature* 451, 153 (2008).
19. S. Schoeftner, M. A. Blasco, *Nat. Cell Biol.* 10, 228 (2008).
20. M. Azzalin, P. Reichenbach, L. Khoraiuli, E. Giulotto, J. Lingner, *Science* 318, 798 (2007).
21. N. Xu, M. E. Donohoe, S. S. Silva, J. T. Lee, *Nat. Genet.* 39, 1390 (2007).
22. L. F. Zhang, K. D. Huynh, J. T. Lee, *Cell* 129, 693 (2007).
23. V. V. Lunyak et al., *Science* 317, 248 (2007).
24. K. Noma, H. P. Cam, R. J. Maraia, S. I. Grewal, *Cell* 125, 859 (2006).
25. A. Willoughby, A. Vilalta, R. G. Oshima, *J. Biol. Chem.* 275, 759 (2000).
26. M. Pheasant, J. S. Mattick, *Genome Res.* 17, 1245(2007).
27. T. Kuwabara, J. Hsieh, K. Nakashima, K. Taira, F. H. Gage, *Cell* 116, 779 (2004).
28. V. Makeyev, T. Maniatis, *Science* 319, 1789 (2008).
29. J. Feng et al., *Genes Dev.* 20, 1470 (2006).
30. S. Stephen, M. Pheasant, I. V. Makunin, J. S. Mattick, *Mol. Biol. Evol.* 25, 402 (2008).
31. A. Calin et al., *Cancer Cell* 12, 215 (2007).
32. R. F. Place, L. C. Li, D. Pookot, E. J. Noonan, R. Dahiya, *Proc. Natl. Acad. Sci. U.S.A.* 105, 1608 (2008).
33. L. C. Li et al., *Proc. Natl. Acad. Sci. U.S.A.* 103, 17337(2006).
34. J. Han, D. Kim, K. V. Morris, *Proc. Natl. Acad. Sci. U.S.A.* 104, 12422 (2007).
35. Martianov, A. Ramadass, A. Serra Barros, N. Chow, A. Akoulitchev, *Nature* 445, 666 (2007).
36. M. Ohno, T. Fukagawa, J. S. Lee, T. Ikemura, *Chromosoma* 111, 201 (2002).
37. P. Kapranov et al., *Science* 316, 1484 (2007).
38. M. G. Guenther, S. S. Levine, L. A. Boyer, R. Jaenisch, R. A. Young, *Cell* 130, 77 (2007).
39. J. A. Chekanova et al., *Cell* 131, 1340 (2007).
40. C. A. Davis, M. Ares Jr., *Proc. Natl. Acad. Sci. U.S.A.* 103, 3262 (2006).
41. C. A. Espinoza, J. A. Goodrich, J. F. Kugel, *RNA* 13, 583(2007).
42. P. D. Mariner et al., *Mol. Cell* 29, 499 (2008).
43. K. Leung, P. A. Sharp, *Cell* 130, 581 (2007).
44. S. Vasudevan, Y. Tong, J. A. Steitz, *Science* 318, 1931(2007).
45. Y. Kawahara et al., *Science* 315, 1137 (2007).
46. M. Kedde et al., *Cell* 131, 1273 (2007).
47. S. R. Viswanathan, G. Q. Daley, R. I. Gregory, *Science*, 21 February 2008 (10.1126/science.1154040).
48. D. M. Tyler et al., *Genes Dev.* 22, 26 (2008).
49. Stark et al., *Genome Res.* 17, 1865 (2007).
50. J. G. Ruby, C. H. Jan, D. P. Bartel, *Nature* 448, 83 (2007).
51. K. Okamura, J. W. Hagen, H. Duan, D. M. Tyler, E. C. Lai, *Cell* 130, 89 (2007).
52. E. Berezikov, W. J. Chung, J. Willis, E. Cuppen, E. C. Lai, *Mol. Cell* 28, 328 (2007).
53. Pagano et al., *PLoS Genet.* 3, e1 (2007).
54. E. Birney et al., *Nature* 447, 799 (2007).
55. P. Carninci et al., *Science* 309, 1559 (2005).
56. M. C. Frith et al., *PLoS Genet.* 2, e52 (2006).
57. M. Clamp et al., *Proc. Natl. Acad. Sci. U.S.A.* 104, 19428(2007).
58. K. V. Prasanth, D. L. Spector, *Genes Dev.* 21, 11(2007).
59. W. Yu et al., *Nature* 451, 202 (2008).
60. T. R. Mercer, M. E. Dinger, S. M. Sunkin, M. F. Mehler, J. S. Mattick, *Proc. Natl. Acad. Sci. U.S.A.* 105, 716(2008).
61. M. Sone et al., *J. Cell Sci.* 120, 2498 (2007).
62. Royo et al., *Mol. Biol. Cell* 18, 2817 (2007).
63. E. Torarinsson et al., *Genome Res.* 18, 242 (2008).
64. J. Ponjavic, C. P. Ponting, G. Lunter, *Genome Res.* 17, 556 (2007).
65. K. C. Pang, M. C. Frith, J. S. Mattick, *Trends Genet.* 22, 1 (2006).
66. K. S. Pollard et al., *Nature* 443, 167 (2006).
67. M. Rassoulzadegan et al., *Nature* 441, 469 (2006).
68. C. J. Hale, J. L. Stonaker, S. M. Gross, J. B. Hollick, *PLoS Biol.* 5, 2156 (2007).
69. R. Lin, S. Maeda, C. Liu, M. Karin, T. S. Edgington, *Oncogene* 26, 851 (2007).
70. D. S. Perez et al., *Hum. Mol. Genet.* 17, 642 (2007).
71. N. Ishii et al., *J. Hum. Genet.* 51, 1087 (2006).
72. M. Broadbent et al., *Hum. Mol. Genet.* 17, 806 (2007).

2.4

miRNA——基因表达调节
的多面手

miRNA通常可以阻遏翻译、降解含有互补靶序列的信使RNA。某些miRNA在细胞或组织中的特异表达，可能与细胞特性的形成与维持有关。最近有研究表明，组织特异性miRNA可能在基因表达调节网络的多个层面上发挥作用。如靶定百余个与当前分化状态不一致的效应基因，或是调控用于调节转录或选择性剪切的通用因子的表达水平。

在形态结构和行为习性的复杂性上差距甚远的各种生物体中，其负责编码蛋白质的基因数目差距却非常小，这不禁让人感到奇怪。如果要为这一悖论提供可能的解释，那么也许就是因为基因表达调节网络精巧度的提高^[1]可以对转录水平和mRNA前体（pre-mRNA）的选择性剪切进行调节^[2]。最近，研究人员认为miRNA可能在生物体复杂性增大的过程中发挥了作用^[3]。某些miRNA的确在胚胎发育期间具有细胞或组织表达特异性，因而很可能在细胞分化过程中扮演重要角色^[4]（O. Hobert的文章中也提及这一问题^[5]）。个别组织特异性miRNA，如miR-1和miR-2，它们分别在肌肉细胞和神经元中表达，已被证明能够激活相应细胞类型的分化^[6-8]。

靶定基因群

miRNA通过几种不同的分子机制发挥其生物学作用。单个miRNA可以抑制某个特定发育阶段所需表达的一大类mRNA。像这样一类被miRNA作用，功能相关且受控于基因表达网络的效应基因可以归为“基因群”（gene battery）。在哺乳动物非神经元细胞中，导入miR-124就会首先减少多种非神经元mRNA的表达量，如编码细胞增殖或是神经干细胞作用所需蛋白的mRNA，导入miR-124后可以观察到它们数量的减少^[7,9,12]。相反，如果在初级神经元中除去miR-124就会导致大量非神经元靶标mRNA的累积^[13]。有报道称，miR1-9也以相似的模式发挥作用。因此，在发生分化的细胞中，miRNA可以有效地除去祖细胞中留下的非必需mRNA。另外，miRNA还可以只在翻译水平上调节其表达量，而不影响部分靶标mRNA的稳定性。



靶定转录调节因子

miRNA也可以调控关键转录调节因子的表达，如首个发现的、能够抑制转录因子lin-14表达的miRNA lin-4^[16,17]。另一个例子是miR-124，它的靶标mRNA是羧基端小结构域磷酸酶1（small C terminal domain phosphatase 1, SCP1/CTDSP1），SCP1/CTDSP1是RE1-沉默转录阻遏物（RE1-silencing transcription repressor, REST/NRSF）的组成之一^[6]（详见图1），而REST在非神经元细胞中阻遏大量神经元特异基因的转录^[13]。因此，在分化的神经元中，miR-124可以通过减少SCP1的表达量，解除SCP1对神经元转录程序的阻遏作用^[6]。REST阻遏复合物也可以抑制miR-124基因的表达，从而在miR-124和REST复合物间建立起一个双重负反馈环路^[6,13]（详见图1）。与此类似，miR-1、miR-133以及其它在肌肉特异表达的miRNA都可以调节肌肉发育过程中几个重要转录因子的表达^[4,8]。

靶定选择性剪切的调节因子

最近的研究发现了miRNA的另一种活性，即通过靶定选择性剪切（alternative splicing）的通用调节因子，从而诱导基因表达发生大规模变化^[7,18]。在早期肌肉细胞前体中，多聚嘧啶区域结合蛋白1（polypyrimidine tract binding protein 1, PTBP1/PTB/hnRNP-I）以及它的同源物PTBP2（nPTB/brPTB/PTBLP）都是选择性剪切的阻遏物，能够抑制多个肌肉特异的外显子拼接起来形成成熟mRNA。但当肌管分化发生时，PTBP1和PTBP2的蛋白表达量就会下降，此时拼接受到抑制的外显子就能够成功拼接成为肌肉特异的成熟mRNA。这种剪切转换的完成至少受到了miRNA的部分调节：前文提到的miR-133会强烈抑制PTBP2的产生，而miR-1及其序列类似物miR-206也会在肌肉发育过程中令PTBP1和PTBP2的表达量下调^[18]。

在神经系统发育过程中，PTBP1和PTBP2的表达量同样受到miRNA的调节^[7,19]。PTBP1可以抑制神经元特异的可变外显子拼接，在神经前体细胞以及其它多种非神经细胞中都有表达。在正处于分化过程的以及成熟的神经元中，PTBP1的表达量下降，致使多个神经元特异的可变外显子可以拼接在一起形成成熟mRNA^[19]。与肌肉细胞中的调节类似的是，神经元中PTBP1表达量的下降也是由miR-124介导产生的，它能够与位于PTBP1 mRNA 3' UTR的保守和非保守的同源靶位点结合^[7]（图1）。



PTBP1除了会抑制上述可变外显子剪切外，还会阻遏PTBP2 pre-mRNA的第10个外显子并入PTBP2成熟mRNA^[7,19,20]。当第10个外显子没有加入成熟mRNA的组成时，PTBP2 mRNA进行读码翻译时会提前遇到终止密码子，最终导致该蛋白在无义介导的mRNA降解机制（nonsense-mediated decay machinery）的作用下被降解^[7,19,20]。在神经元分化的过程中，miR-124的表达量增加，随之降低了PTBP1的表达量，从而使得正确剪切的PTBP2 mRNA得以累积，继而导致PTBP2蛋白表达量的大量上调。值得注意的是，尽管PTBP1和PTBP2是密切相关的同源物，但PTBP1却具有更强的遏制神经元特异可变外显子拼接的作用^[7,19]。因此，PTBP1的表达量下降而PTBP2的表达上调就会发生从非神经元剪切模式到神经元特异性选择剪切模式的变换，使得神经元特异的蛋白亚型得以形成。有趣的是，PTBP2 mRNA含有保守的miR-124结合位点，因此miR-124也能够抑制PTBP2的表达，但比不上对PTBP1的抑制有效^[7]。这一调节机制可用于解释PTBP2表达的减缓。

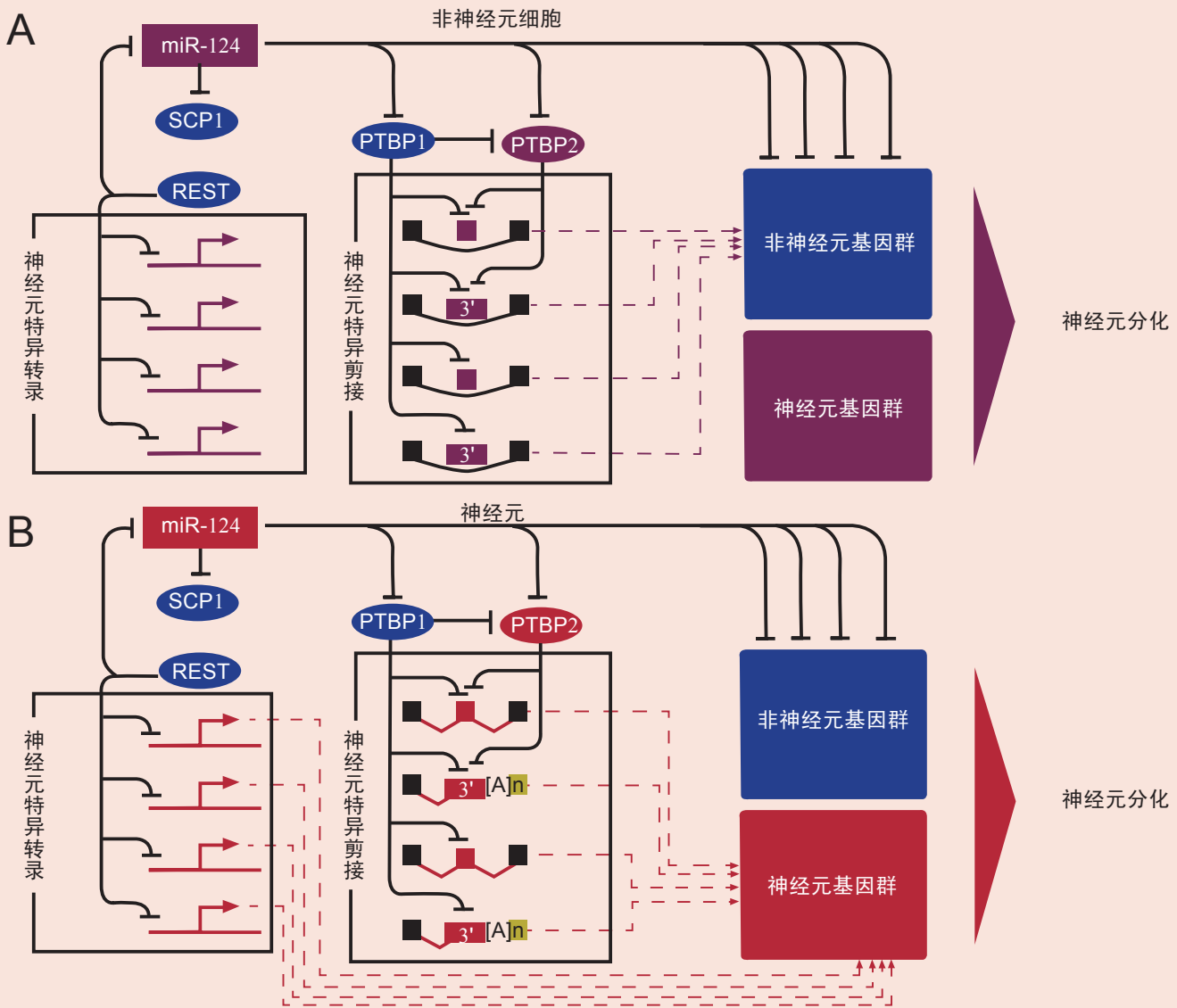


图1 miR-124调控着一个庞大的基因表达调节网络。

图中以黑色实线标示活化的相互调节作用，灰色实线标示的为非活化的相互作用，细线标示的为弱的相互调节作用，从基因群输至细胞转录本的用虚线标示。非神经元基因以深蓝色（表达的）或浅蓝色（不表达的）标示；神经元特异的表达或不表达基因则以相应的深红色与粉红色标示。上图已经经过简化处理，没有涵盖其它miRNA可能发挥的作用，也没有对近期发现miRNA在激活翻译上所起的作用^[15]加以说明。另外，PTBP1和PTBP2也可以激活一类选择性剪切的外显子，其作用机制有待进一步研究^[19]。（A）在非神经元细胞或神经祖细胞中，miR-124不表达或低表达，因此阻遏神经元特异的转录和选择性剪切的因子得以有效表达，非神经元的基因群也同样处于表达状态。（B）在正处于分化状态的神经元细胞中，miR-124的表达量增加，受其调控的阻遏因子的表达量因此发生下调，相应的神经元特异蛋白得以表达，非神经元基因群此时在miR-124的直接调控下表达量下调。

伴随着基因表达调节网络复杂性的增大，后生动物（metazoans）大大提高的细胞分化多样性必定要求某种精巧调控机制的出现，以阻止时间或空间上相邻的基因表达程序发生相互干扰。上文提及的这些例子表明，至少有个别miRNA在这一调控机制中扮演了重要的角色，它们在调节系统的各个

层面上——从基因群到转录与选择性剪切的调节因子，均有效地重新编写了细胞特异的调节网络。随着其它miRNA作用靶点的确定，这样的多重调节的范例很可能会不断涌现。

原文检索：www.sciencemag.org

 Sirius/编译

参考文献

1. M. Levine, R. Tjian, *Nature* 424, 147 (2003).
2. T. Maniatis, B. Tasic, *Nature* 418, 236 (2002).
3. R. Niwa, F. J. Slack, *Curr. Opin. Genet. Dev.* 17, 145 (2007).
4. Y. Zhao, D. Srivastava, *Trends Biochem. Sci.* 32, 189 (2007).
5. Hobert, *Science* 319, 1785 (2008).
6. J. Visvanathan, S. Lee, B. Lee, J. W. Lee, S. K. Lee, *Genes Dev.* 21, 744 (2007).
7. E. V. Makeyev, J. Zhang, M. A. Carrasco, T. Maniatis, *Mol. Cell* 27, 435 (2007).
8. J. F. Chen et al., *Nat. Genet.* 38, 228 (2006).
9. L. P. Lim et al., *Nature* 433, 769 (2005).
10. J. Giraldez et al., *Science* 312, 75 (2006); published online 15 February 2006 (10.1126/science.1122689).
11. E. H. Davidson, *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution* (Academic Press, San Diego, CA, 2006).
12. X. Cao, S. L. Pfaff, F. H. Gage, *Genes Dev.* 21, 531 (2007).
13. C. Conaco, S. Otto, J. J. Han, G. Mandel, *Proc. Natl. Acad. Sci. U.S.A.* 103, 2422 (2006).
14. F. V. Karginov et al., *Proc. Natl. Acad. Sci. U.S.A.* 104, 1929 (2007).
15. S. Vasudevan, Y. Tong, J. A. Steitz, *Science* 318, 1931 (2007); published online 29 November 2007 (10.1126/science.1149460).
16. R. C. Lee, R. L. Feinbaum, V. Ambros, *Cell* 75, 843 (1993).
17. B. Wightman, I. Ha, G. Ruvkun, *Cell* 75, 855 (1993).
18. P. L. Boutz, G. Chawla, P. Stoilov, D. L. Black, *Genes Dev.* 21, 71 (2007).
19. P. L. Boutz et al., *Genes Dev.* 21, 1636 (2007).
20. R. Spellman, M. Llorian, C. W. Smith, *Mol. Cell* 27, 420 (2007).
21. We thank S. Oliferenko, S. Buchanan, B. Friedman, E. Morimoto, and B. McCallum for critical comments. This effort was supported by NIH (grant 2R01NS043915-27;T.M.) and Leukemia and Lymphoma Society (E.V.M.).

2.5

通过RNAPII在启动子附近区域的停顿所介导的基因转录调控

最近我们的研究发现，果蝇和许多哺乳动物的多个基因进行转录时，RNAPII会在启动子附近的某些特定区域发生停顿。这种停顿往往是发生在RNAPII复合体形成和转录起始之后。而这个转录早期延伸的时期也是许多基因调控手段发生效用的时期，在基因调控方面有着重要意义。

大部分真核生物的mRNA都是在转录水平上初次受到调控的。这些基因的转录过程可以分为几个不同的时期：第一，RNAPII在基因启动子附近形成转录起始复合物；第二，转录起始；第三，RNAPII从启动子区域和启动子附近的停顿位点脱落；第四，RNA链的延伸；最后为转录终止^[1]。以上五个过程每一个都可以被分成若干个不同的可以被调控的生化过程。为了弄清特定基因的表达如何被调控，很重要的一点是要找出这些生化过程中哪些是转录的限速步骤，以及信号介导的激活因子和抑制因子如何在限速过程中发挥作用。

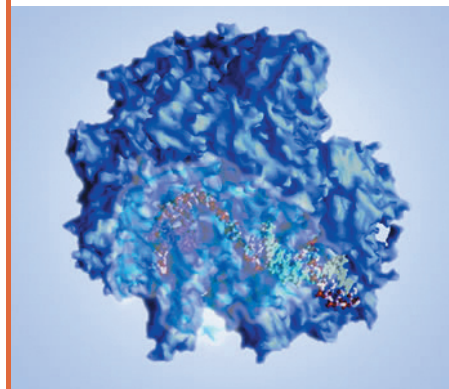
真核生物基因转录调控的机理研究主要集中在转录起始前的转录起始复合物的形成以及转录起始这两个过程；但是，RNA链延伸过程中的转录调控例子在过去25年里被反复提起。Chambon等人早期的一个令人瞩目的研究成果表明，参与人类 β -珠蛋白(β -globin)基因转录的RNAPII在红细胞成熟 β -珠蛋白基因被关闭之后仍然会停留在该基因的5'端^[2]。另外，一系列研究表明：在非诱导型果蝇热激蛋白70(Heat Shock Protein 70, Hsp70)^[4]以及其它一些果蝇和哺乳动物基因中^[5,6]，RNAPII会停留在转录起始位点下游20至

50个碱基处^[3]。这种在真核生物中发现的RNAPII在基因启动子附近区域的停顿现象与原核生物中的基因转录调控机制是相似的^[7]，这种现象被认为是一种限制基因转录速度的调控手段^[4,8]。在多细胞动物中已经发现存在这种停顿现象的基因模型。然而，在单细胞动物酿酒酵母(*Saccharomyces cerevisiae*)中，基因转录的调控大多发生在更早的时期，也就是RNAPII在基因启动子附近形成转录起始复合物的时期^[9,10]。因而，认为在没有被激活的基因启动子附近接合上PolIII是很少见的例外情况。

RNAPII在基因启动子附近区域停顿现象的概述

虽然在酿酒酵母(*S. cerevisiae*)中RNAPII在基因启动子附近区域的浓度与mRNA的浓度相关^[11]，但是最近一些基因组范围的研究发现，在果蝇和哺乳动物细胞中情况并不总是这样^[12-15]。这些研究是运用染色质免疫沉淀分析法(Chromatin Immunoprecipitation Assay)和基因芯片技术(Genomic Microarray Technologies, ChIP-chip)共同测定RNAPII在基因附近的浓度。这些研究发现20%~30%的基因会在5'端聚集，其中包括有表达的基因，而这些基因的表达可能是可检测的，也可能是不可检测的。这种RNAP在未完全转录的基因附近聚集的现象说明，在这些基因表达的过程中，RNAPII复合体形成之后的转录过程存在着基因转录的限速步骤。虽然染色质免疫沉淀分析法(ChIP)可以检测到基因全长附近RNAPII的浓度，但这种技术不能确定RNAPII是否正在实行基因转录的任务。也就是说，在以上研究中检测到的5'端的浓度梯度分布的RNAPII可能处于转录起始前的状态，也可能处于转录起始后的停顿状态。

在最近的全基因组范围研究中，有三个研究另外也做了高锰酸盐印迹实验(permanganate footprinting)。这种实验方法可以发现由RNAPII转录生成的转录泡(transcription bubble)，或者可以检测短的RNA产物。在多种基因中，这种短小的RNA都可以作为RNAPII已经跨过了转录起始阶段的证据^[13-15]。虽然这些研究中证实的，具有转录过程中RNAP停顿现象的基因主要是一些低表达或者检测不到表达的基因，但必须强调的一点是，高水平表达的基因也完全有可能受停顿现象的调控。因此，转录起始之后，RNAPII在启动子附近区域出现停顿的现象被认为是一种广泛存在的转录调控机制。

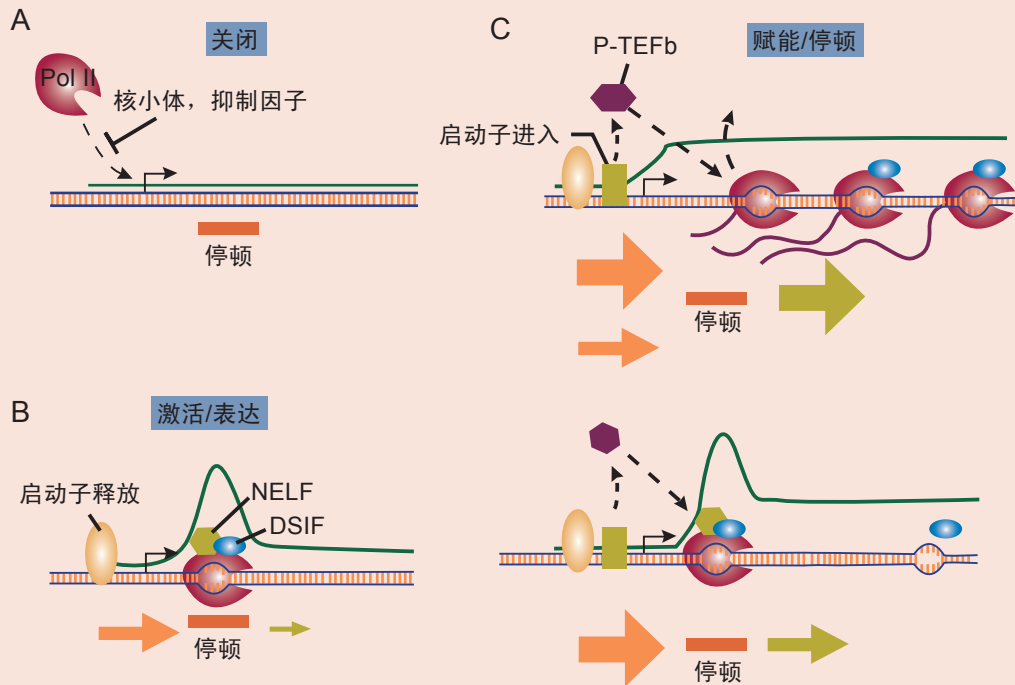


RNAPII在启动子附近的停顿是一种调控机制

体外实验证明：即便在辅助因子不存在的情况下，转录也会在RNAPII启动子附近区域停顿。转录早期RNA延伸的过程中，DNA模板和新生的RNA序列被认为是影响转录复合体结构改变的重要因素^[16,17]。这种构象改变可能是RNAPII转录复合物在完成长距离转录时不会脱离DNA模板和新生RNA所必须的。在体内，这种固有的停顿现象所出现的频率和范围仍有待进一步鉴定，但这种停顿现象出现的位置与一些已经知道的停顿相关分子的作用区域是相一致的。例如：咪喃核糖苯并咪唑敏感性可诱导因子（DRB Sensitivity Inducing Factor, DSIF）和延伸抑制因子（Negative Elongation Factor, NELF）^[18]，它们在停顿过程中起到进一步稳定RNAPII的作用。现在，这些停顿相关因子已经成为

研究的热点^[1,6,19]。

RNAPII要进入启动子附近的停顿位点，所必需的一个条件是：转录起始复合物必须先进入启动子区域并且启动转录。RNAPII从停顿位点释放需要RNAPII从启动子附近移出以留出足够的空间给另一个RNAPII复合体结合上去并且启动转录。RNAPII进入和释放速度的比率决定了它在一个基因中停顿的时间（图1）。高的进入率低的释放率会使其在启动子附近的停顿时间延长，在ChIP中就可以看到RNAPII在基因5'端的浓度比在基因3'端的浓度高（图示1 B和C图中下面一幅）。当RNAPII进入停顿位点的比率小于或者等于释放的速率时，RNAPII的有效停顿时间就会缩短，结果是更多的RNAPII通过停顿位点（图示1 C图中上面一幅）。



图示1 Pol II在停顿位点进入和释放的调控

(A) Pol II不能进入启动子区域，基因转录关闭。（B）在促进因子作用下Pol II在启动子附近停顿的状态。NELF、DSIF起稳定Pol II的作用。（C）基因转录的全面激活需要促释放因子，同样，一种因子也可能含有一种或两种活性域，可以依次通过可逆的构象改变及接合起到调控的作用。

由RNAPII进入和脱离停顿位点速率的不同而进行的转录调控，在果蝇Hsp70的调控机制中得到了更好的证明。在GAGA因子作用下，这个基因可以保持未激活状态和较高的停顿率；而促进RNAPII从停顿位点的释放和基因的活化则需要结合热激因子（Heat Shock Factor, Hsf）^[6]。与此相似，在哺乳动物和果蝇中，调节因子相互配合共同调节基因表达中的限速步骤的例子并不少见^[5, 20~22]。最近的这些基因组范围研究带给我们的启示是：激活停顿位点的激活子可能与多种辅助因子一起广泛作用于基因的启动子，使转录成为可能。多种途径联合起来调控转录机制使得细胞可以受多种信号协同刺激，在需要的时候迅速上调基因的表达。

那些激活RNAPII复合体从停顿位点释放的激活因子是通过招募其它因子直接调节转录复合物，或者也可能是通过调节染色质的环境来执行自己

的功能的，例如改变染色质环境，使得越过核小体的转录成为可能。最主要的一个帮助RNAPII转录复合物从停顿位点释放的辅助因子是有激酶活性的转录延长促进因子b（Positive Transcription Elongation Factor b, P-TEFb）^[19~23]。它可以使转录复合物中的PolIII、NELF和DSIF等多个位点磷酸化，并且对NELF或者DSIF依赖性的转录延长阻碍物的释放有重要作用（图示1）。在这个磷酸化过程中，NELF从转录复合物中分离下来，但是被修饰的DSIF仍然留在转录复合物上继续促进转录的进行。不出所料，细胞进化出了多种途径，从而把P-TEFb带给需要表达的基因。这个过程可以直接通过细胞与促进转录复合物在停顿位点脱离的激活因子相互作用而完成，也可以通过细胞与激活过程中所招募而来的蛋白间的相互作用而完成^[6,19]。



停顿现象是与RNA转录后处理过程相联系的

转录复合体在基因启动子附近的停顿现象有可能是为了协调RNA链的延伸和前体mRNA（pre-mRNA）的修饰这两个过程^[1]。实际上，转录复合体在启动子附近的停顿和pre-mRNA的加帽过程是同时发生的。pre-mRNA的加帽过程无论对RNA分子的稳定还是后续的mRNA处理过程都很重要^[1]。另外，Pol II从停顿状态的结构换成转录延伸结构时，其最大的亚基的C末端结构域（C-terminal domain, CTD）会分步骤发生磷酸化，从而引起与pre-mRNA修饰有关系的辅助因子或者蛋白的构象的改变^[1]。



停顿不只是抑制：也使转录成为可能

很多证据表明，启动子附近出现的转录复合体停顿现象对整个基因的激活有重要作用。对果蝇Hsp70基因和人类FOS和Myc的研究表明：除去与停顿相关的序列会导致转录因子的趋近性降低和基因的激活不完全^[24~26]。停顿中的Pol II是怎样帮助转录调控因子靠近基因的启动子的机理还不清楚。但是，可以推测，Pol II可能是通过阻止核小体（nucleosome）遮挡DNA上的结合位点或者是通过招募一些其它的蛋白分子修饰启动子附近的染色质来发挥作用的。在果蝇中，通过基因组范围的研究确定的能够使Pol II发生停顿的基因，都可以快速对细胞的调节信号和发育信号做出反应^[14,15]。所以，在基因激活前发生停顿的现象很有可能是为了

使细胞能够快速地控制发育事件^[15,27]。

转录的调控是一个复杂的过程，它包括对启动因子的招募、转录的起始、转录的停顿以及RNA链的延长等过程的调控。许多基因组范围的研究发现了很多可能存在Pol II在启动子附近区域发生停顿的基因，从而为我们提供很多可供研究的基因模型，去深入研究这种转录调控机制。未来，这个领域的研究将会聚焦于启动子特异性结合蛋白如何影响转录停顿、转录起始及RNA链延伸之间转换。研究结果将会为揭示细胞信号传导对基因转录的影响机制起到重要作用。

原文检索：www.sciencemag.org



小强/编译

参考文献

1. M. Levine, R. Tjian, *Nature* 424, 147 (2003).
2. R.J.Sims3rd,R.Belotserkovskaya,D.Reinberg,*GenesDev.* 18,2437(2004).
3. P.Gariglio,M.Bellard,P.Chambon, *NucleicAcidsRes.* 9,2589(1981).
4. Although a large fraction of promoter-proximal PolII can Elongate in nuclear run-on assays,some PolII may beBacktracked and cannot readily elongate.Therefore,theBroader term of stalling,which includes arres tand pause,Has also been used.
5. A.E.Rougvie,J.T.Lis, *Cell* 54,795(1988).
6. D.L.Bentley, *Curr.Opin.Genet.Dev.* 5,210(1995).
7. Saunders,L.J.Core,J.T.Lis, *Nat.Rev.Mol.CellBiol.* 7,557(2006).
8. J.W.Roberts etal., ColdSpringHarb.Symp.*Quant.Biol.*63,319(1998).
9. J.Lis, ColdSpringHarb.Symp.*Quant.Biol.* 63,347(1998).
10. M.Keaveney,K.Struhl, *Mol.Cell* 1,917(1998).
11. M.Ptashne,A.Gann, *Nature* 386,569(1997).
12. F.Robert etal., *Mol.Cell* 16,199(2004).
13. T.H.Kim etal., *Nature* 436,876(2005).
14. M.G.Guenther,S.S.Levine,L.A.Boyer,R.Jaenisch,R.A.Young, *Cell* 130,77(2007).
15. G.W.Muse etal., *Nat.Genet.* 39,1507(2007).
16. J.Zeitlinger etal., *Nat.Genet.* 39,1512(2007).
17. M.Pal,D.McKean,D.S.Luse, *Mol.Cell.Biol.* 21,5815(2001).
18. Ujvari,M.Pal,D.S.Luse, *J.Biol.Chem.* 277,32527(2002).
19. Y.Yamaguchi etal., *Cell* 97,41(1999).
20. B.M.Peterlin,D.H.Price, *Mol.Cell* 23,297(2006).
21. J.Blau etal., *Mol.Cell.Biol.* 16,2044(1996).
22. T.Sawado,J.Halow,M.A.Bender,M.Groudine,*GenesDev.* 17,1009(2003).
23. Y.V.Wang,H.Tang,D.S.Gilmour, *Mol.Cell.Biol.* 25,3543(2005).
24. N.F.Marshall,D.H.Price, *Mol.Cell.Biol.* 12,2078(1992).
25. L.S.Shopland,K.Hirayoshi,M.Fernandes,J.T.Lis,*GenesDev.* 9,2756(1995).
26. Krumm,L.B.Hickey,M.Groudine, *GenesDev.* 9,559(1995).
27. J.Fivaz,M.C.Bassi,S.Pinaud,J.Mirkovitch, *Gene* 255,185(2000).
28. X.Wang,C.Lee,D.S.Gilmour,J.P.Gergen, *GenesDev.*21,1031(2007).

2.6

三维空间上的基因表达调控



通过对染色体空间结构的分析，研究人员发现染色体之间通过相互作用形成了一个复杂的三维网络。这种染色体之间的相互作用能够在多个水平对基因表达进行调节，例如通过相距很远的增强子和抑制子进行远距离控制、协调相关基因的表达以及进行表观遗传修饰。目前所面临的挑战是阐明基因座之间是通过何种机制接近并相互作用的，以及弄清这种瞬时关联有何功用。

在低等生物的基因组中，例如酵母，一个基因和其调控元件在基因组上组成一个连续的片段，称为“表达调控单元”（regulatory expression unit）。但是在人类和老鼠等更为复杂的基因组中，基因和其调控元件可以相距几十万个碱基^[1,2]。长期以来，人们一直认为空间上相距很远的基因组元件可以通过染色体空间构型的变化结合到一起，从而实现调控元件对基因表达的调节（图1）。

研究人员以前大多通过显微镜来研究染色体的组织结构，最近，染色体三维构象捕获技术（chromosome conformation capture, 3C）也越来越广泛地应用于染色体结构的研究^[3]。3C是一项分子技术，它通过甲醛交联以及基因座位特异性PCR来检测基因组中基因座之间的物理性接触。3C技术弥补了传统镜检法的一些不足。镜检法至少一个细胞就能得到其染色体相关信息，但是图像的分辨率却很低；而3C技术能得到高分辨率的结果，但需要比较多的细胞样本。研究人员通过3C技术和显微镜检同时证实远距离的染色体相互作用在生物体内广泛存在，这表明看似分散的基因组调控元件之间很可能存在高度的关联关系。

表达单元的空间组装

研究人员对许多基因组元件之间的空间关联都进行了深入研究，增强子与其靶基因之间的相互作用就是一个典型例子。现在以β球蛋白基因座为例具体说明。β球蛋白基因座含有好几个β球蛋白样基因，这些基因都受到一个位于β球蛋白基因启动子上游10至60kb的同一个顺式作用元件——基因座控制域（locus control region, LCR）的调节，该元件能与活化的球蛋白基因发物理性关联^[4]。此外，越来越多的元件被发现可以通过远距离成环而相互作用，例如α球蛋白基因座^[5,6]和白介素基因簇^[7]，甚至在单一基因内部也可以成环^[8,9]。



已经有研究报道称分布在不同染色体上的基因座具有高度的关联性。与嗅觉受体基因的调控方式类似，这种关联性可以体现为增强子与其可能的靶基因之间的顺式相互作用^[10]。但是在其他情况下，这些元件需要在一个更高的水平上进行调节以同时调控多个基因座的表达（图1, B）。老鼠的2型辅助T淋巴细胞相关细胞因子基因与 γ 干扰素基因分别位于11和10号染色体上^[11]。这两个基因座的表达是相互排斥的，因此它们之间的相互作用可能会引发或加强与原来相反的表现遗传状态。

X染色体之间的特异顺式关联是哺乳动物X染色体失活过程中的一个关键环节。雌性细胞携带两个X染色体，但是其中一条处于沉默状态，这样就能够保证雌性细胞中X相关基因的表达水平与雄性细胞相当。X染色体的失活最先发生在X失活中心。目前，有研究发现，在X染色体开始失活的发育阶

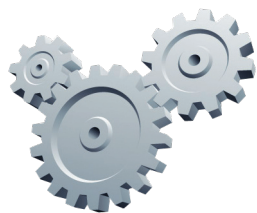
段，两个X失活中心会发生瞬时的相互作用^[12, 13]。失活中心的基因突变分析结果表明，它们之间的相互作用与X染色体失活过程密切相关。X染色体的这种配对现象是生物体计算细胞内X染色体数目的一个精妙机制，该机制同时也确保了两条X染色体中只有一条会处于失活状态。

目前的观察发现向研究人员提出了一个有趣的问题：在复杂的基因组中“表达调控单元”究竟由什么组成？虽然简单基因组中基因和其调控元件簇呈线性排列，但在更为复杂的基因组中，表达单元可以由空间上相距很远的基因簇和调控元件组成（图1）。组成表达单元的基因和调控元件可以通过重新组合以在更多的层次上调节基因表达。例如在印记基因座中，母系于父系等位基因会与不同调控元件关联，从而组成不同的调控单元^[14]。

染色体间的瞬时互动



利用3C或基于3C的改良技术在许多观察中都发现了远距离染色体互作现象。好几篇综述都认为3C技术在使用上相对比较简单，但是对其结果进行合理解释与分析则不是轻易能够完成的^[15, 16]。特别需要注意的是，尽管许多由3C技术发现的染色体互作现象都得到了镜检的证实，但我们还是很难由3C信号得出准确的染色体互作的实际发生率。在许多情况下，染色体之间发生相互作用的几率很低（同一时刻只有不到10%的细胞能观察到该现象）。之所以会这样，是因为染色体的构象在细胞中是动态变化的。因此，目前广泛使用刚性成环模型来描述这种染色体之间的关联作用可能是不正确的。尽管该方法看起来似乎没问题，但该模型没有能够反映染色体远距离互作的高度瞬时特性。



染色体互作的功能

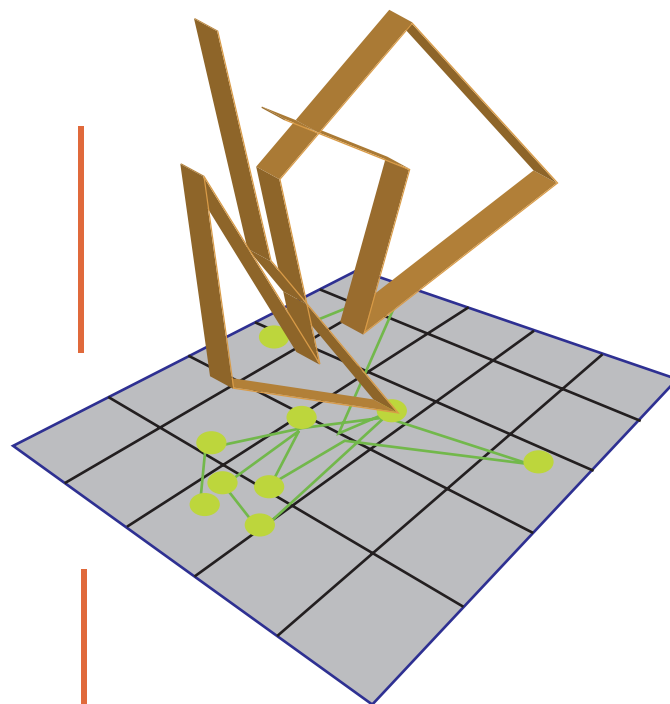
虽然我们观察到基因座之间存在特异性的关联作用，但这并不意味着我们已经了解了这种现象背后所蕴藏的功能影响。研究人员想了许多方法和手段来了解基因座远距离相互作用与基因表达之间的因果关系，而较常用的两种方法是下调介导相互作用的蛋白表达（例如转录因子）和去除某些调控元件。另外还有一种有效的方法是利用可视化的RNA产物对基因进行原位杂交以分析基因座的定位，从而知道基因座互作是否在单个细胞水平上与基因转录相关联。有一点我们必须注意，即虽然已经发现了一些基因座互作与基因的转录相关，但是敲除互作调控元件对基因表达并无影响^[10, 17]。或许这些现象可以说明染色体之间的相互作用可能与基因转录无关，但这同时也反映了目前我们对于该领域知识的匮乏。

染色体之间的关联作用究竟通过何种方法影响基因的表达？增强子与启动子之间的相互作用能够帮助组成转录复合物的各组分结合到启动子区域。与增强子结合的酶会在基因组的帮助下与启动子复合物接触并对其进行修饰，例如磷酸化或甲基化修饰，从而调节启动子的活性。另外，以X失活中心之间的相互作用为代表的基因组互作还能够允许发生作用的两个成员各自协调装配不同的蛋白复合物。由于基因座之间的关联具有瞬时性，因此发生关联的两个基因座应该会获得一些特异而又稳定的标记，比如DNA甲基化，从而在接下来染色体不再关联的阶段指导蛋白复合物的装配。

基因座是如何关联的？

研究人员提出了一些模型^[18]以解释不同的基因组元件之间是如何接触的（图2）。被动扩散模型认为基因座所具有的运动能力为元件之间的随机碰撞提供了前提条件，而这种基因调控元件之间的随机碰撞是否会转化为真正具有生理意义的关联作用取决于这两个蛋白复合物之间的亲和力和特异性。尽管被动扩散模型对染色体如何发生关联做出了解释，但这一过程肯定还需要一个主动的过程以介导基因座互相接触和关联。例如，研究人员认为增强子会主动沿着染色质纤维移动直至遇到一个能与它结合的启动子。最近，人们发现基因座能够以一种肌动蛋白依赖性的方式，沿着一条快速定向轨道穿越细胞核^[19,20]。目前科学界对于细胞核肌动蛋白和肌浆球蛋白的功能作用还存在争议，但是最近的研究发现强烈表明，它们在帮助染色体远距离相互作用的过程中扮演了极其重要的角色：这两种蛋白不仅能够将处于不同位置的基因座运输到一起，而且还能将基因座运送到一些重要的亚细胞核区域，比如富含RNA聚合酶的转录工厂区域（图1, C）。

基因组中当然也包含一些调节其它基因座互作的调控元件。这类元件被称为绝缘体，因为它们位于增强子和启动子之间时能够抑制增强子对启动子的激活作用。我们暂时还不知道绝缘体具体的作用机制，但它们也参与了远距离调控元件的相互作用^[21]。实验表明绝缘体会使染色体形成环状结构从而以某种方式帮助增强子与其靶基因结合。



展望

目前，研究人员正努力绘制染色体互作的完整图谱。目前，一些经过改进的3C技术已能够借助于芯片技术或新开发的高通量测序技术对基因组互作进行高通量的检测^[22]。4C法（也称为3C芯片法或3C循环法）能够在染色体范围内鉴定出与研究人员所关注的某一基因座位发生物理性近距离接触的区域^[23,24]。而5C（也称为3C碳拷贝法）法则更进一步不再局限于单一位点，而且能够绘制出我们所感兴趣的染色体区域的密集互作网络图谱^[25]。以上方法让我们更深入的了解基因组的空间组织结构，但所有这些在本质上还只是描述性信息。我们需要利用新的方法来探索染色体关联如何影响基因表达调控。这些方法包括染色体基因座的时序成像，控制基因座移动和亚细胞核定位的分子和遗传机理，同时也包括对于介导染色体关联的复合物性质所进行的生化研究。总而言之，这些新的技术手段一定能为我们展示出基因三维表达调控的全新画面。

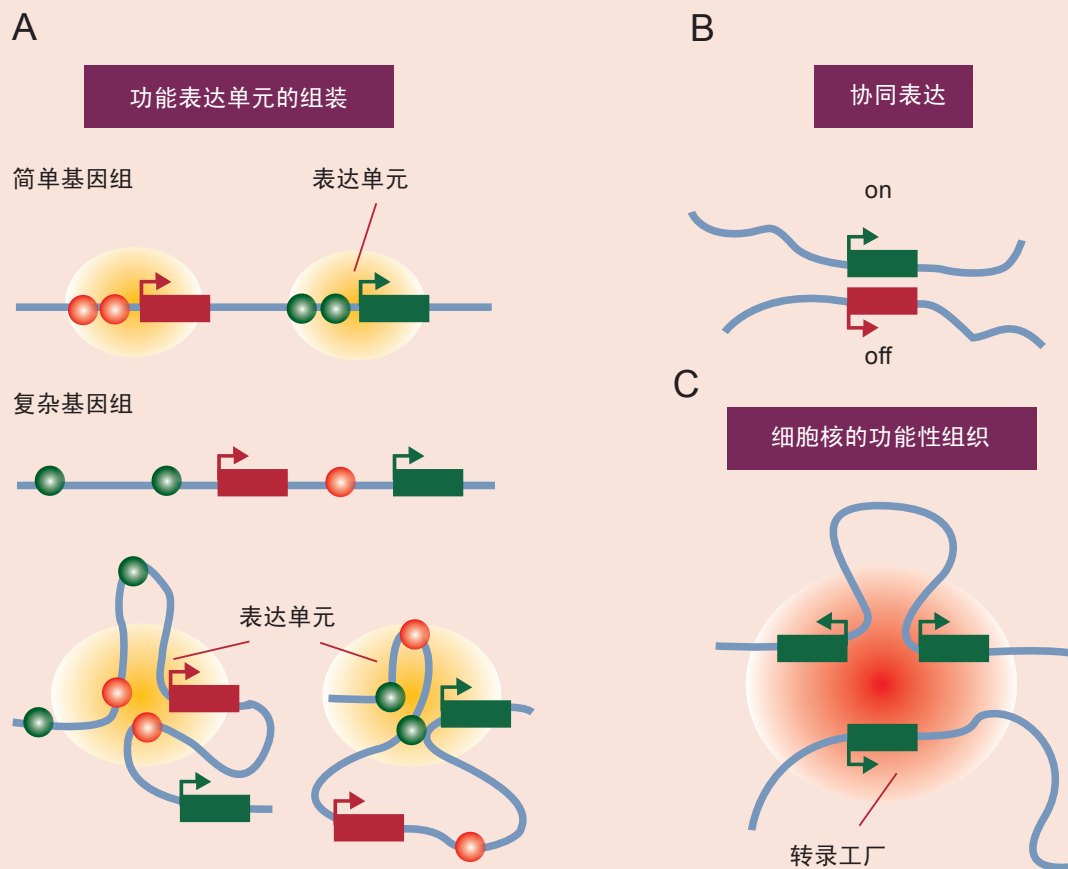


图1 表达单元的空间转配

(A) 简单基因组中的线性表达单元及复杂基因组中的空间装配表达单元。(B) 基因的协调表达。(C) 基因在转录工厂等亚细胞核区域的共定位。圆形代表调控元件；长方形代表基因。箭头代表转录方向。

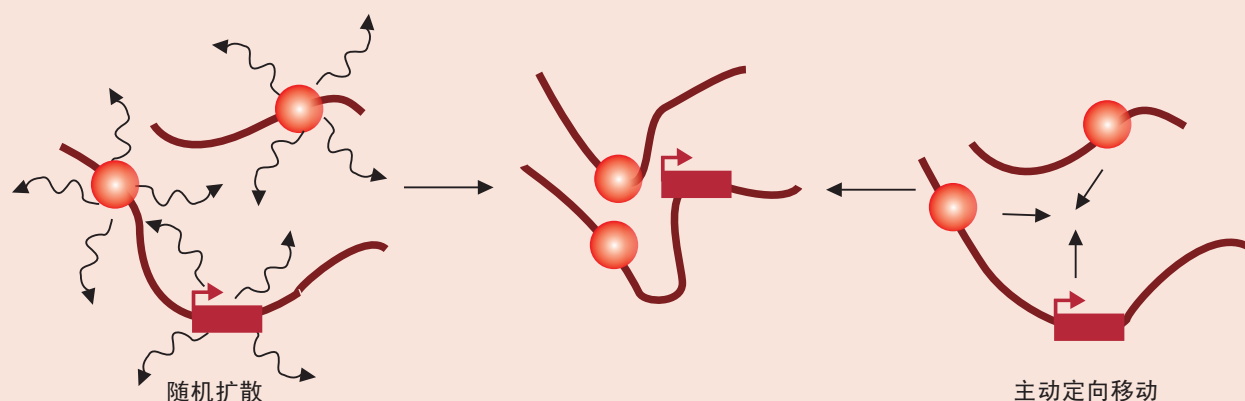



图2基因座相接触的被动和主动模型

圆形和长方形分别代表调控元件和基因。波浪箭头代表随机扩散，直箭头代表主动定向移动。

原文检索: www.sciencemag.org

 知易行难/编译

参考文献

1. D. A. Kleinjan, V. van Heyningen, *Am. J. Hum. Genet.* 76, 8 (2005).
2. G. West, P. Fraser, *Hum. Mol. Genet.* 14, R101 (2005).
3. J. Dekker, K. Rippe, M. Dekker, N. Kleckner, *Science* 295, 1306 (2002).
4. Tolhuis, R. J. Palstra, E. Splinter, F. Grosveld, W. de Laat, *Mol. Cell* 10, 1453 (2002).
5. Vernimmen, M. De Gobbi, J. A. Sloane-Stanley, W. G. Wood, D. R. Higgs, *EMBO J.* 26, 2041 (2007).
6. G. L. Zhou et al., *Mol. Cell. Biol.* 26, 5096 (2006).
7. G. Spilianakis, R. A. Flavell, *Nat. Immunol.* 5, 1017 (2004).
8. J. A. Grass et al., *Mol. Cell. Biol.* 26, 7056 (2006).
9. H. Jing et al., *Mol. Cell* 29, 232 (2008).
10. S. Lomvardas et al., *Cell* 126, 403 (2006).
11. C. G. Spilianakis, M. D. Lalioti, T. Town, G. R. Lee, R. A. Flavell, *Nature* 435, 637 (2005).
12. N. Xu, C. L. Tsai, J. T. Lee, *Science* 311, 1149 (2006).
13. C. P. Bacher et al., *Nat. Cell Biol.* 8, 293 (2006).
14. S. Kurukuti et al., *Proc. Natl. Acad. Sci. U.S.A.* 103, 10684 (2006).
15. J. Dekker, *Nat. Methods* 3, 17 (2006).
16. M. Simonis, J. Kooren, W. de Laat, *Nat. Methods* 4, 895 (2007).
17. S. H. Fuss, M. Omura, P. Mombaerts, *Cell* 130, 373 (2007).
18. J. D. Engel, K. Tanimoto, *Cell* 100, 499 (2000).
19. C. H. Chuang et al., *Curr. Biol.* 16, 825 (2006).
20. M. Dunder et al., *J. Cell Biol.* 179, 1095 (2007).
21. J. A. Wallace, G. Felsenfeld, *Curr. Opin. Genet. Dev.* 17, 400 (2007).
22. B. Wold, R. M. Meyer, *Nat. Methods* 5, 19 (2008).
23. M. Simonis et al., *Nat. Genet.* 38, 1348 (2006).
24. Z. Zhao et al., *Nat. Genet.* 38, 1341 (2006).
25. J. Dostie et al., *Genome Res.* 16, 1299 (2006).

2.7

复杂的核糖开关

代谢产物结合核糖开关（riboswitch）利用简单的生化机制，便可以行使基因调控的功能，这一功能长久以来都被认为是蛋白的“专利”。尽管现代核糖开关可能是保存至今唯一的原始遗传物质，但是它们所具有的功能却不容小觑，其传感及调控功能在许多方面与蛋白相比毫不逊色。



核糖开关存在于mRNA中，它们与小分子相互结合从而调节基因表达^[1]。大部分核糖开关只有一个识别靶向配体的结合位点或核酸适体。核糖开关的核酸适体一般位于基因表达区域附近，当核酸适体与代谢产物结合时，会改变自身结构，从而行使基因调控的功能。核糖开关在所有基因表达开关中大概是最为古老的一员^[2~4]，但是由于其结构非常简单——只由四个核苷酸组成，因此可以推断核糖开关在结合多种代谢产物、行使复杂基因调控功能方面的能力是较为有限的。但是，新的研究发现开始不断揭示这些来源古老、化学组成简单的RNA分子如何形成复杂的基因调控系统机制。

精确的化学传感

细胞内，核糖开关必须从哪怕是无数个代谢产物中，特异性地识别其靶向配体。对十分相似的化合物进行精确鉴别可以有效避免错误的代谢调控。研究人员采用人工合成的核酸适体证明了RNA确实具有上述高特异性识别的功能。那些经人工合成的核酸适体可以以高亲和力和特异性与许多不同的配体结合^[5]，它们可以在细胞内发挥“设计师”基因调控元件的作用^[6]，但是与天然核酸适体相比，这些人工分子更小、结合力较弱且特异性较低^[4]。当然，这种区别并不奇怪：合成的核酸适体是在实验室内仅用几周时间就制备出来的，而那些天然核酸适体则经历了数十亿年的进化演变。

对核糖开关核酸适体进行原子结构研究，发现这些分子通过很多的氢键与配体结合，并相互接触以稳定附近RNA之间的相互作用，从而进一步提高与配体结合时的亲和力。有些核酸适体会形成口袋结构，将整个配体都包绕进去，这种类型的核酸适体配体结合，是以诱导契合机制为基础的。这些口袋结构会改变自身形状，以便精确地与其配体结合，使范德华力发挥作用，增强两

者的结合力。

核糖开关的核酸适体通常带负电荷，可以与焦磷酸硫胺素（TPP）和葡萄糖胺6磷酸（GlcN6P）结合。与TPP^[7,8]和GlcN6P^[9,10]结合的核糖开关可以富集带正电的镁离子，使核酸适体的磷酸根与配体结合^[11]。SAM-I核糖开关与S-腺苷甲硫氨酸（SAM）结合，可以有效鉴别SAM及其代谢副产物S-腺苷高半胱氨酸^[12]。SAM和S-腺苷高半胱氨酸十分相似，只相差一个甲基团和一个正价电荷，而且正是这一个正价电荷的差别决定了核糖开关SAM-I与SAM的特异性结合。SAM-I核糖开关中的羰基团（而不是磷酸根）与正电荷基团相结合。此外，SAM发生构象折叠，以便形成另一个正电胺基团（图1）。

在一些有机体中，识别SAM的分子是完全不同的RNA结构^[12,13]，而且这些RNA与SAM的结合方式也各有不同^[14]。研究者发现，多种核糖开关可以与同一个配体结合，这一现象说明，RNA是一种在代谢传感基因调控过程中具有多向性的媒介。因此，采用大量不同核酸适体识别多种配体，使得核糖开关可以检测多种代谢产物，有时多种核酸适体结构可以共同识别同一种配体。

发生构象改变的RNA

核糖开关可以利用配体结合的能量稳定mRNA改变了的构象。大部分核糖开关在进化过程中都形成了参与反馈环构成的功能：当某种代谢产物的数量达到一定水平时，核糖开关就会关闭合成这些产物的酶类基因的表达，以避免不必要的浪费^[1]。在革兰氏阳性菌（G+）中，基因表达区域通常以相互不能相容的发夹结构的形成为基础，调控结果要么是发生转录（反终止因子stem）要么终止转录（终止因子stem）^[12]。当代谢产物与核糖开关结合时，可以阻止反终止因子stem形成，而促使终止因子的形成，从而“关闭”基因表达。革兰氏阴性菌（G-）中的基因表达区域一般对mRNA与核糖体的结合进行调控^[12]，核糖体结合位点有时位于核糖开关的核酸适体内。这种类型的核糖开关往往是控制“关闭”的，这是由于位于配体结合位点的核苷酸是无法同时发挥其通常所具有的促使蛋白质合成起始的功能。基因表达的配体依赖性激活由核糖开关进行调控，这是通过改变上述遗传物质作用机制来实现的。

那些通过传感GlcN6P来控制基因表达的核糖开关不是通过改变构象而是作为自杀性RNA来行使功能^[9,10]。GlmS核酶只有在与GlcN6P结合时，才可以催化RNA的剪切，这一特殊的性质显得十分有

趣，因为所有其它的核糖开关只对其表达区域的构象进行调控。GlcN6P位于核酶剪切位点，是RNA剪切的辅助因子，该化合物中的胺基团参与质子传递的催化过程^[10]。与之相反，诸如葡萄糖6磷酸这样的化合物不含胺基团，没有激活核酶剪切的功能。这一更为复杂的核糖开关通过两种机制对其十分相似的靶向配体进行区分，其一是通过不同的结合亲和力，其二则是如上所述，需要配体具有催化RNA剪切的的功能。

尽管细菌、真菌和植物内与TPP结合的核糖开关核酸适体非常相近，但它们在mRNA中的定位却各不相同^[15]，所采用的基因调控机制也相差甚远^[15-20]。在真菌内，配体与一些TPP核糖开关结合，通过控制碱基配对的“茎”结构的形成，对pre-mRNA剪切位点的结合过程进行调控（图2A）^[17]。真菌粗糙链孢霉（*Neurospora crassa*）中的两个TPP核糖开关调控mRNA的5' UTR内的剪切，该区域内TPP调控的剪切可以反应上游开放阅读框（uORF）是否被保留。这些uORF可以作为诱饵把核糖体“骗”到这一区域，抑制主要ORF的表达。多细胞植物中的TPP核糖开关调控着mRNA 3' UTR^[18,19]，这种核酸适体的模块性使得它们可以适应有机体mRNA中不同位点的多种基因调控机制^[15]。

多层次的核糖开关

任何被仅有一个核酸适体和基因调控区域的核糖开关所控制的基因都需要一个刺激因子的调控，其基因表达的幅度达到10%到90%^[21]。采用这种标准方式的细胞，必须能够耐受某些代谢产物巨大的浓度变化幅度。但是，核糖开关至少有两种策略（机制）来缩小这一动态学变化的幅度。第一种策略是，两种或更多的核糖开关可以前后排列，形成串联结构，对同一种配体进行检测。如果这种“串联”的核糖开关被浓度基本相同的代谢产物激活，且如果单独激活任何一个开关都会抑制基因的表达，那么系统只需要一种代谢产物的浓度变化幅度在40%的因子就可以使基因表达发生同样的改变。由于增加第三个核糖开关对开关反应性的提高并不明显，因此，最常见的“串联”核糖开关就是由两个连续的核糖开关组成^[22]。串联排列的tRNA感受基因控制元件被称为T-box RNA，这种T-box RNA也是很常见的^[23]，它们的作用是增加基因对tRNA浓度微小变化的反应性。

另一个更加精妙的可以减小动力学变化幅度的策略是，将多个核酸适体整合在同一个核糖开关内。这种结构在许多甘氨酸核糖开关中可以见到。甘氨酸核糖开关包含两个甘氨酸适体和一个基因表达区域^[12,24]。如果开关内两个核酸适体可以协同作用，也就是任一核酸适体与配体结合都可以相应提高配体与另一个核酸适体的亲和力，那么甘氨酸浓度变化因子只变化9个百分点，基因表达变化幅度就可以达到10%到90%。这一串联的核糖开关结构使得RNA可以与蛋白因子竞争，感应更小浓度变化的代谢产物，发挥更为敏感的基因开关作用。

拥有不同配体特异性的多层次的核糖开关允许细胞用多种不同的仅由RNA介导的化学刺激信号^[22,25]来控制同一基因。在克劳氏芽孢杆菌中，合成甲硫氨酸的基因metE受感受SAM和辅酶B₁₂^[22]的串联排列的核糖开关控制。这两个开关感受双信号，一起控制基因的表达。因为SAM由甲硫氨酸合成，所以metE基因的mRNA携带有一个SAM核糖开关，在辅酶B₁₂充足时，这个开关可以用来下调metE基因的表达。当有足够浓度的辅酶B₁₂时，B₁₂核糖开关就会抑制metE基因的表达，因为细菌被B₁₂诱导、表达出了一个更有效率的甲硫氨酸合成酶。尽管串联排列的核糖开关在细菌中并不经常出现，当细胞面临环境生存压力等挑战时，这种独特的核糖开关组织结构也是会被迫采用的。

动力学与热力学

核糖开关利用三维折叠形成了一个具有多种核酸适体、表达系统的体系，但还存在第四维的因素——时间，这更增加了核糖开关的复杂程度和调控力度。被核酸适体结合的配体通常用自身的解离常数 (K_D) 来评价。 K_D 值可以在皮摩尔数量级来测量一些核糖开关核酸适体的浓度，因此核糖开关可以用来检测极低浓度的代谢产物。实际上，一些核糖开关表现出有足够的时间来与他们所处的环境保持平衡，并对与核糖开关核酸适体的 K_D 值匹配的配体的浓度做出反应^[26]。然而，一些控制转录终止的核糖开关却没有那么充足的时间，他们合成以后只有很短的时间来进行充分的折叠，这样才能迅速地起作用^[27,28]。

在这些情况下，是配体结合的速度和RNA多聚酶的活性来决定什么样的代谢产物的浓度最终来调控基因的表达。这是核糖开关的优势，所以它们在漫长的进化过程中超过了蛋白因子。细胞可以通过

突变其核酸适体或蛋白受体来增强或减弱它们与配体的亲和力来改变胞内代谢产物浓度的感受阈值。基于核糖开关的动力学因素，细胞还可以仅仅通过精确的突变核糖开关，改变其折叠的速度或者改变合成它的RNA多聚酶的活性的办法来改变核糖开关的动力学强度，并以此法来改变胞内代谢产物浓度的感受阈值。

在上述情况下，调控基因表达的代谢产物浓度是由配体结合的速度及RNAP作用的速度决定的。这是核糖开关在进化过程中形成的与蛋白因子竞争的优势特性。细胞可以通过使核糖开关的核酸适体或蛋白受体发生突变，来改变所需结合的代谢产物浓度，从而加强或减弱与配体结合的结合力。由于有了动力学驱动的核糖开关，细胞就可以通过发生分子的突变来调节相应的基因表达，因为这些突变可以改变核糖开关折叠构象或RNAP催化核糖开关合成过程的速度。

核糖开关的发现

随着越来越多DNA序列数据库的建立和新的生物信息学工具的出现，研究人员在细菌中发现了越来越多的核糖开关类别^[12,29,30]。此外，研究人员在真核生物上投入的精力，也让我们相信会有更多真核生物体内的核糖开关被发现。目前，还没有确切的真核核糖开关被发现的报道。在漫长的进化过程中，为了获得基因调控的主导地位，蛋白质在进化树上所处的位置似乎与那些具有某些相似功能的RNA之间的距离太远。已有研究发现，真核生物也会表达许多非编码RNA，核糖开关很有可能同样控制着这些RNA的表达和活性。如同在细菌研究中发现的那样，如果不能找到一个发挥信号传导功能的蛋白因子的存在，那么就很有可能是由RNA构成的更为复杂的传感物质在行使它们的职责了。

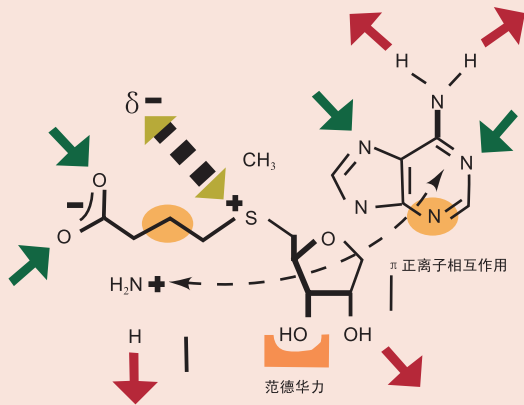


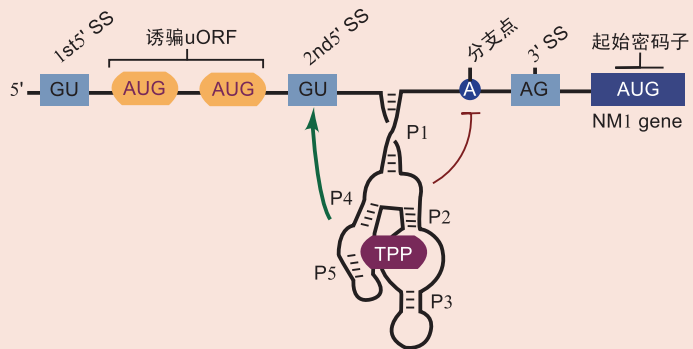
图1 SAM-I适体与SAM特异性结合示意图^[31]

红色和绿色箭头分别表示氢键离子供体和受体。SAM的阴影部分是结合位点，但并不与RNA直接结合，原因可能是其修饰构象阻碍了可以稳定配体构象的 π 正离子相互作用。

图2 控制剪切的真核核糖开关示意图

(A) 真菌粗糙链孢霉 (*N. crassa*) 中NMT1基因内的核糖开关。

真菌TPP核糖开关：选择性剪切



植物TPP核糖开关：选择性加工

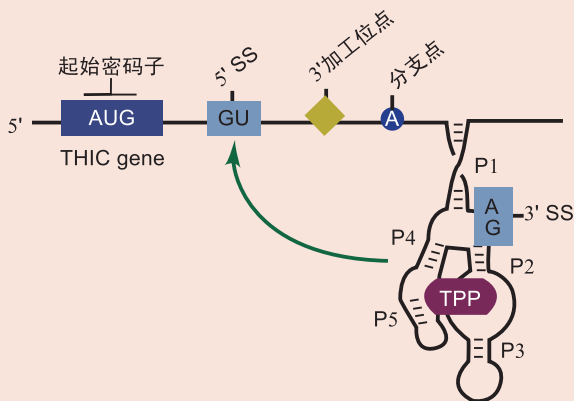


图2 控制剪切的真核核糖开关示意图

(B) 植物拟南芥 (*Arabidopsis thaliana*) 中 THIC 基因内的核糖开关。绿色和红色线分别表示激活或抑制剪切元件的TPP诱导的构象折叠改变。SS表示剪切位点。

原文检索: www.sciencemag.org

YORK/编译

背景介绍

最近发现,某些依赖代谢物调节的基因转录产物的5' UTR存在特征性结构——核糖开关。核糖开关可以以特异性结合代谢物,通过构象变化,在转录或翻译水平上调节基因表达。核糖开关广泛存在于G⁺及G⁻细菌的代谢相关基因中,在真菌、植物中也有发现。核糖开关调节维生素、氨基酸、核苷酸等基础代谢过程,而调节基因表达不需要任何蛋白因子作为中介,在进化上可能是RNA世界遗留的分子化石。核糖开关可用于研究基因功能,开发新型药物及进行基因治疗。

参考文献

1. R. L. Coppins, K. B. Hall, E. A. Groisman, *Curr. Opin. Microbiol.* 10, 176 (2007).
2. Nahvi et al., *Chem. Biol.* 9, 1043 (2002).
3. G. Vitreschak, D. A. Rodionov, A. A. Mironov, M. S. Gelfand, *Trends Genet.* 20, 44 (2004).
4. R. R. Breaker, in *The RNA World*, R. F. Gesteland, T. R. Cech, J. F. Atkins, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, ed. 3, 2006), pp. 89–107.
5. S. E. Osborne, A. D. Ellington, *Chem. Rev.* 97, 349 (1997).
6. J. P. Gallivan, *Curr. Opin. Chem. Biol.* 11, 612 (2007).
7. S. Thore, M. Leibundgut, N. Ban, *Science* 312, 1208 (2006); published online 3 May 2006 (10.1126/science.1128451).
8. Serganov, A. Polonskaia, A. T. Phan, R. R. Breaker, D. J. Patel, *Nature* 441, 1167 (2006).
9. D. J. Klein, A. R. Ferré-D' Amaré, *Science* 313, 1752 (2006).
10. J. C. Cochrane, S. V. Lipchock, S. A. Strobel, *Chem. Biol.* 14, 97 (2007).
11. T. E. Edwards, D. J. Klein, A. R. Ferré-D' Amaré, *Curr. Opin. Struct. Biol.* 17, 273 (2007).
12. J. E. Barrick, R. R. Breaker, *Genome Biol.* 8, R239 (2007).
13. R. T. Fuchs, F. J. Grundy, T. M. Henkin, *Nat. Struct. Mol. Biol.* 13, 226 (2006).
14. S. D. Gilbert, R. P. Rambo, D. Van Tyne, R. T. Batey, *Nat. Struct. Mol. Biol.* 15, 177 (2008).
15. N. Sudarsan, J. E. Barrick, R. R. Breaker, *RNA* 9, 949 (2003).
16. T. Kubodera et al., *FEBS Lett.* 555, 516 (2003).
17. M. T. Cheah, A. Wachter, N. Sudarsan, R. R. Breaker, *Nature* 447, 497 (2007).
18. S. Bocobza et al., *Genes Dev.* 21, 2874 (2007).
19. Wachter et al., *Plant Cell* 19, 3437 (2007).
20. M. T. Croft, M. Moulin, M. E. Webb, A. G. Smith, *Proc. Natl. Acad. Sci. U.S.A.* 104, 20770 (2007).
21. R. Welz, R. R. Breaker, *RNA* 13, 573 (2007).
22. N. Sudarsan et al., *Science* 314, 300 (2006).
23. Gutierrez-Preciado, R. A. Jensen, C. Yanofsky, E. Merino, *Trends Genet.* 21, 432 (2005).
24. M. Mandal et al., *Science* 306, 275 (2004).
25. D. Stoddard, R. T. Batey, *ACS Chem. Biol.* 1, 751 (2006).
26. R. Rieder, K. Lang, D. Graber, R. Micura, *ChemBio* 8, 896 (2007).
27. J. K. Wickiser, W. C. Winkler, R. R. Breaker, D. M. Crothers, *Mol. Cell* 18, 49 (2005).
28. W. J. Greenleaf, K. L. Frieda, D. A. N. Foster, M. T. Woodside, S. M. Block, *Science* 319, 630 (2008); published online 2 January 2008 (10.1126/science.1151298).
29. Z. Weinberg et al., *Nucleic Acids Res.* 35, 4809 (2007).
30. M. D. Kazanov, A. G. Vitreschak, M. S. Gelfand, *BMC Genomics* 8, 347 (2007).
31. R. K. Montange, R. T. Batey, *Nature* 441, 1172 (2006).

2.8

真核生物转录通路的进化

生物的转录通路在漫长的进化过程中所发生的渐变,是科学家研究生命多样性的一个重要线索。过去十年的动物实验研究结果揭示了有关基因调节的小分子的改变如何对生物体形态学和生理学产生巨大影响,以及环境的选择压力会对这些分子的改变产生怎样的影响。最近开始的全基因组序列研究,尤其是对单细胞酵母进行的研究,证实了转录通路发生改变的广泛存在,表明即使亲缘关系很近的有机体在对其基因进行调节时,采用的调节通路也大有区别。

在真核生物中，每个基因的转录是由一系列顺式调节序列控制的。这些序列通常位于基因编码序列的附近。机体内每个基因相关的顺式调节序列都具有基因转录的时间特异性和空间特异性。这一现象是通过序列特异性DNA结合蛋白即转录调节子（transcription regulator, TR）的研究得以证实的。TR可以识别相应要转录的基因序列，其自身也在机体生命历程中具有特异性的时间与空间表达以及激活。TR通过与顺式调节序列相互作用，从而选择有待转录的基因。当然，在实际的基因转录和调节过程中，还有许多其它步骤；但是顺式调节序列和TR对基因的识别是基因调节中一个关键性的基础步骤。

生物转录调节的一些特性有助于我们了解其在生物进化过程中所扮演的角色^[1~4]。顺式调节序列通常为5至10个核苷酸组成的具有简并性的短序列，它们的位置因所调控基因的不同而有所不同。不同的顺式调节序列通常在位置上相互靠近，TR则往往与这些相近位点协同结合。这种协同结合是一种联合调控的形式——即通过多个TR而不是单个TR对基因表达进行调控。顺式调节序列往往成簇形成一个调节模块，每个模块对机体内的基因表达进行时间和空间特异性的独立调控。

TR也是以模块的形式存在，在实验室内可以将不同TR的片段和“零件”重新组合，制备出新的调节模块。基因突变可以改变TR的DNA结合特异性，以及它们的协同蛋白，乃至其对基因转录的作用（激活或者抑制）。这是由于很多非常关键的由TR参与的蛋白质与蛋白质相互作用，其作用力不强，又不具有作用的特异性。因此，相关基因即使发生很小的改变，都可能对基因调节产生很大的影响。

上述基因转录调节的特点使得我们比较容易理解，为何一些简单的突变就会产生出新的基因表达模式。在过去十年中，科学家对动物体内的某些单个基因进行研究，发现了许多由于顺式调节序列发生改变而对机体形态学和生理学特点产生影响的例子^[5,6]。这些形态学和生理学改变包括：人类的乳糖耐受^[7]、鱼类的骨架结构^[8]、果蝇中的表皮毛^[9]以及色素沉着基因^[10]。尽管研究者更多地强调顺式调节序列的作用，但事实上，TR的改变也会导致机体表型的改变^[11,12]。

上述动物实验从生物物种内和种群间在基因调节上的不同给予了阐释。对转录调节在进化中的改变进行研究的一个方法，对转录通路，特别是包含有若干TR和很多靶向基因（例如那些与TR结合并接受其调节的基因）的大的转录通路进行分子水平的研究。然后，科学家会对两个或以上种群的转录通路进行比较。这种研究方法的优点是，可以获得所研究的转录通路的全貌信息，不会因为发生于机体局部的转录调节而得出有失偏颇的结论。当然，这个优势同时也是其局限性所在，该方法往往很难分辨转录调节的变化到底是否会使有机体受益。

这一基因组研究方法被用于对亲缘关系接近的酵母、果蝇和哺乳动物中转录通路的比较。通常在研究中，这种方法会与生物信息学、转录谱分析、全基因组染色质免疫沉淀等方法结合使用。该方法既证实了先前研究者的推测，也衍生出新的相关论点。

首先，高水平转录通路的改变可发生在相对很短的进化历程中^[13,14]。尽管定向进化同源性的（从共同的始祖基因分化而成）TR所具有的DNA结合特异性，在种属间并没有太大差别，但是受TR直接调控的基因在不同种属间却有很大不同。例如，研究者对小鼠和人类肝细胞内多达4000个基因与四个肝特异性TR的结合情况进行了比较，结果发现，只有不到三分之二的基因同时是小鼠和人的相应TR的靶基因^[15]。对三种亲缘关系相近的酵母中的两个TR进行全基因组研究，据估计，在一种酵母中存在的TR-靶基因结合关系中，只有三分之一在另两个酵母种中仍然存在^[16]。尽管这种不同可能部分是由于发生在某些种属顺式调节序列的丢失和获得，但是这一结果仍然向我们阐明了，分子水平的改变（例如TR活性的改变）是如何在生物进化中起作用的。

另一项对三个亲缘关系很远的酵母所做的转录通路研究，通路中包括TR Mcm1及其辅助因子，同样证实了上述结论^[17]。存在于酿酒酵母中的仅15%的Mcm1与靶向基因的相互作用关系在另外两个酵母种中得以保存。Mcm1在一些辅助因子作用下与相应DNA序列结合，对一个种属中的多个基因进行调节。而上述实验中所观察到的基因转录在进化过程中发生的变化要归因于在进化中顺式调节序列的丢失和获得



的频繁发生，以及新的Mcm1辅助因子的形成代替了原有的辅助因子。

其二，基因组研究方法表明，在不同种属间相同类别的基因表达是由不同的机制进行调节的。早期对果蝇所做的研究发现，稳定性选择可以维持某一单独基因的表达模式，但是仍然会在基因调节机制方面发生相当的变化（如[18]所述）。对酵母的全基因组研究补充了上述观点，揭示了在不同种属中存在着共同表达的成组基因（即在环境改变或发生其它干扰时，基因以组为单位进行表达），但是负责调控这些基因的TR却因种属不同而不同。例如，在酿酒酵母中，半乳糖可以通过TR Gal4诱导半乳糖代谢酶类的转录。在另一种真菌——白色念珠菌中，同样的酶类也是由半乳糖诱导产生，但并非通过Gal4的调控，而是通过另一个TR来识别顺式调节序列的；而Gal4调节的是糖酵解酶^[19]。

另一个支持上述观点的实验是对真菌内的接合型调节进行的研究：在此过程中， α 特异性基因（在 α 细胞内转录，而在 a 细胞内不转录）表达的调节从转录激活因子变成了转录抑制因子^[20]。由于激活因子和抑制因子是在不同细胞类型中表达，转录通路的总体机制得以保留。而发生在进化过程中对不同基因进行调控的TR的改变，有可能是通过一种中间状态实现的。也就是说，TR靶基因受到双重调控，从而通过这种中间状态保留了共表达（图1）。研究者认为，真菌内

的核糖体基因转录调控的改变也是通过上述过剩的中间状态实现的^[14]。

目前，由于靶基因的总体表达模式从表面上看呈现相反的状态，因此还不能证实这些基因转录通路的改变是否会给机体带来益处。有可能许多发生改变的转录通路只是一种中性的进化现象，并非为适应环境而发生的利于机体的生物学过程，仅仅是在两种模式中任选其一而已^[21]。

第三，基因组研究方法所得出的结论，与有关TR与多个DNA序列的协同结合方式利于转录通路的改变这一推论是一致的。以一个最简单的模式来说明这个问题：两个TR分别为A和B，都可以与某DNA序列结合，而是否结合取决于两种蛋白质各自的浓度大小、蛋白质与DNA间作用力的大小以及两种蛋白质A和B间的相互作用力。对于A和B蛋白质而言，上述各种参数中有任何一个下降，都可能会导致另一个蛋白与DNA的结合。这种调节模式可以使得在不破坏调节通路的基础上，通过不同TR与DNA结合能力大小的改变，来最终决定由哪种TR对基因进行调控，从而促进基因转录调节的改变。如图1所示（右边的通路），如果A-B间相互作用力更强，那么B就不会再与顺式调节序列结合。这种改变就使得顺式调节序列需要选择另一个TR，C的表达与B有部分一致。如果由于点突变的发生，A-C间相互作用得到加强，这个基因的调控就会从A-B转变为A-C，但这种转变过程

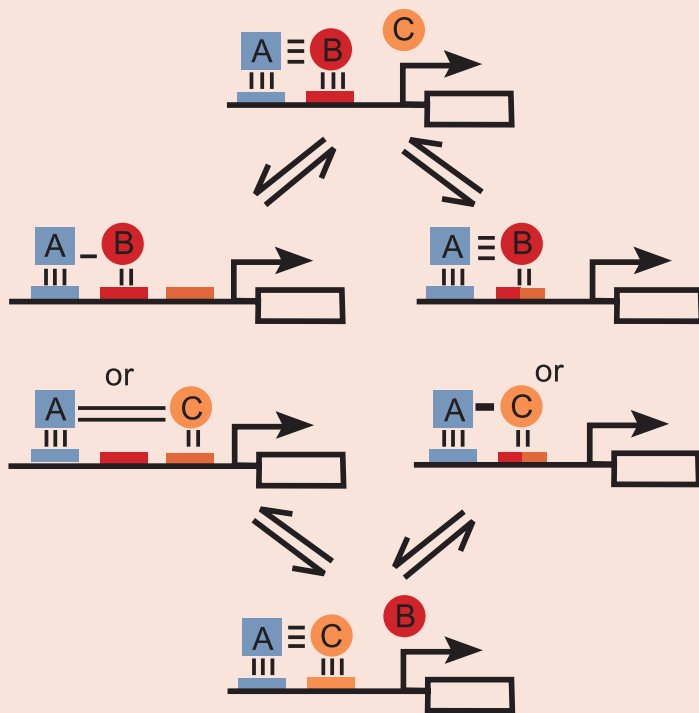


图1 转录通路的联合调控机制

图中的两个调控通路可以阐释调控同一个基因的TR如何转换为另一个TR的。在两条通路中，都存在着调节因子B和C可能都为冗余的中间状态。小黑线代表蛋白质与蛋白质间及蛋白质与DNA间的相互作用，黑线的数目表示相互作用力的大小。在任一时间，每个处于共表达环境中的基因都可能具有不同的调控状态（仅受B调控，或仅受C调控，或同时受B和C的调控）。左边通路可表示真菌内核糖体基因及半乳糖代谢基因的调控通路^[14,19]。右边的通路可表示 α 特异性调控基因通路^[20]。

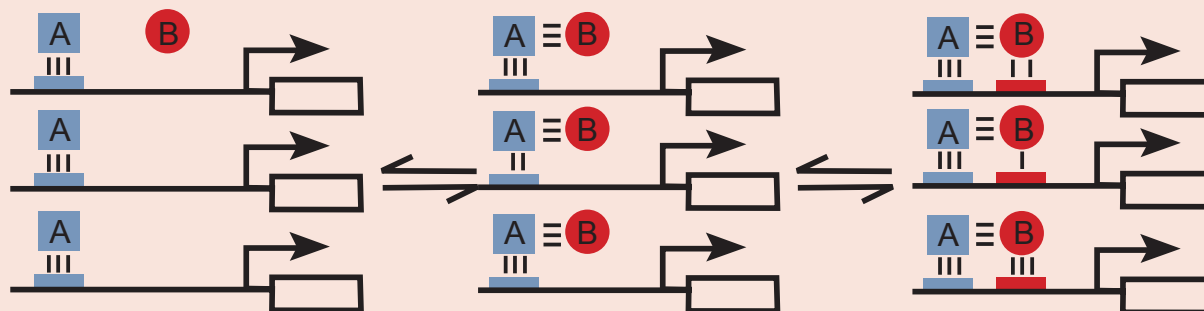


图2 多个基因调控通路改变的可能机制

在此模式中，通过TR因子A和B的相互作用，促进了基因间相互作用的优化。

并不会破坏基因原有的调控通路。上述例子只是协同结合所带来的众多效应之一。如果发生协同结合的TR数目增加，那么该系统内可能发生的调控改变就会相应增多。

一些实验结果已经对上述论点给予了直接的支持。比如，在真菌内的接合型通路改变已经继上述调控模式之后被报道^[20]。进一步的证据来自对酿酒酵母进行的全转录通路网络的分析^[22]。在这项研究中，在调控基因的TR的数目和该基因的顺式调节序列与TR结合的非特异性之间，存在着强相关性。这种非特异性表明，多个TR与DNA的结合这一机制使得某一种TR-DNA间相互作用变得不再那么重要。同时也表明，通过对相互作用因子系统的刺激性进化，冗余的中间状态（如上文所述）可以大大促进这些系

统发生转录改变^[23]。最后，认为进化过程中转录调节由于TR与DNA的协同结合而发生的中性改变，有可能利于复杂调节通路的形成^[24]。当然，我们要强调的是，TR对基因其它形式的联合调控（如两个与某DNA序列结合的TR相互独立地对靶基因进行调控）同样可以利于转录通路的改变。

我们认为，协同结合对于在所有共表达基因的调节通路中发生的改变具有尤其重要的意义。例如，在某个基因处已经存在一个TR的情况下，如果可以获得两个TR间蛋白质与蛋白质的相互作用，就可以发生基因调控通路的改变（图2）。此后，随着最佳的顺式调节序列的逐渐形成，新的转录调控通路便可以通过一个又一个靶基因得到提高。这一推论可以帮助我们理解为何在整个共表达基因内都会出现调节通路的改变。

原文检索：www.sciencemag.org

筱玥/编译

参考文献

1. S. B. Carroll, J. K. Grenier, S. D. Weatherbee, *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* (Blackwell, Malden, MA, ed. 2, 2005).
2. E. H. Davidson, *Genomic Regulatory Systems: Development and Evolution* (Academic Press, San Diego, 2001).
3. G. A. Wray et al., *Mol. Biol. Evol.* 20, 1377 (2003).
4. M. Kirschner, J. Gerhart, *Proc. Natl. Acad. Sci. U.S.A.* 95, 8420 (1998).
5. G. A. Wray, *Nat. Rev. Genet.* 8, 206 (2007).
6. B. Prud'homme, N. Gompel, S. B. Carroll, *Proc. Natl. Acad. Sci. U.S.A.* 104 (suppl. 1), 8605 (2007).
7. S. A. Tishkoff et al., *Nat. Genet.* 39, 31 (2007).
8. M. D. Shapiro et al., *Nature* 428, 717 (2004).
9. P. McGregor et al., *Nature* 448, 587 (2007).
10. S. Jeong et al., *Cell*, 132, 783 (2008).
11. M. Ronshaugen, N. McGinnis, W. McGinnis, *Nature* 415, 914 (2002).
12. R. Galant, S. B. Carroll, *Nature* 415, 910 (2002).
13. P. Gasch et al., *PLoS Biol.* 2, e398 (2004).
14. Tanay, A. Regev, R. Shamir, *Proc. Natl. Acad. Sci. U.S.A.* 102, 7203 (2005).
15. D. T. Odom et al., *Nat. Genet.* 39, 730 (2007).
16. R. Borneman et al., *Science* 317, 815 (2007).
17. B. Tuch, D. J. Galgoczy, A. D. Hernday, H. Li, A. D. Johnson, *PLoS Biol.* 6, e38 (2008).
18. M. Z. Ludwig, C. Bergman, N. H. Patel, M. Kreitman, *Nature* 403, 564 (2000).
19. M. Martchenko, A. Levitin, H. Hogues, A. Nantel, M. Whiteway, *Curr. Biol.* 17, 1007 (2007).
20. E. Tsong, B. B. Tuch, H. Li, A. D. Johnson, *Nature* 443, 415 (2006).
21. M. Lynch, *Nat. Rev. Genet.* 8, 803 (2007).
22. Y. Bilu, N. Barkai, *Genome Biol.* 6, R103 (2005).
23. E. S. Haag, M. N. Molla, *Evol. Int. J. Org. Evol.* 59, 1620 (2005).
24. E. Zuckerkandl, *J. Mol. Evol.* 44 (suppl. 1), S2 (1997).