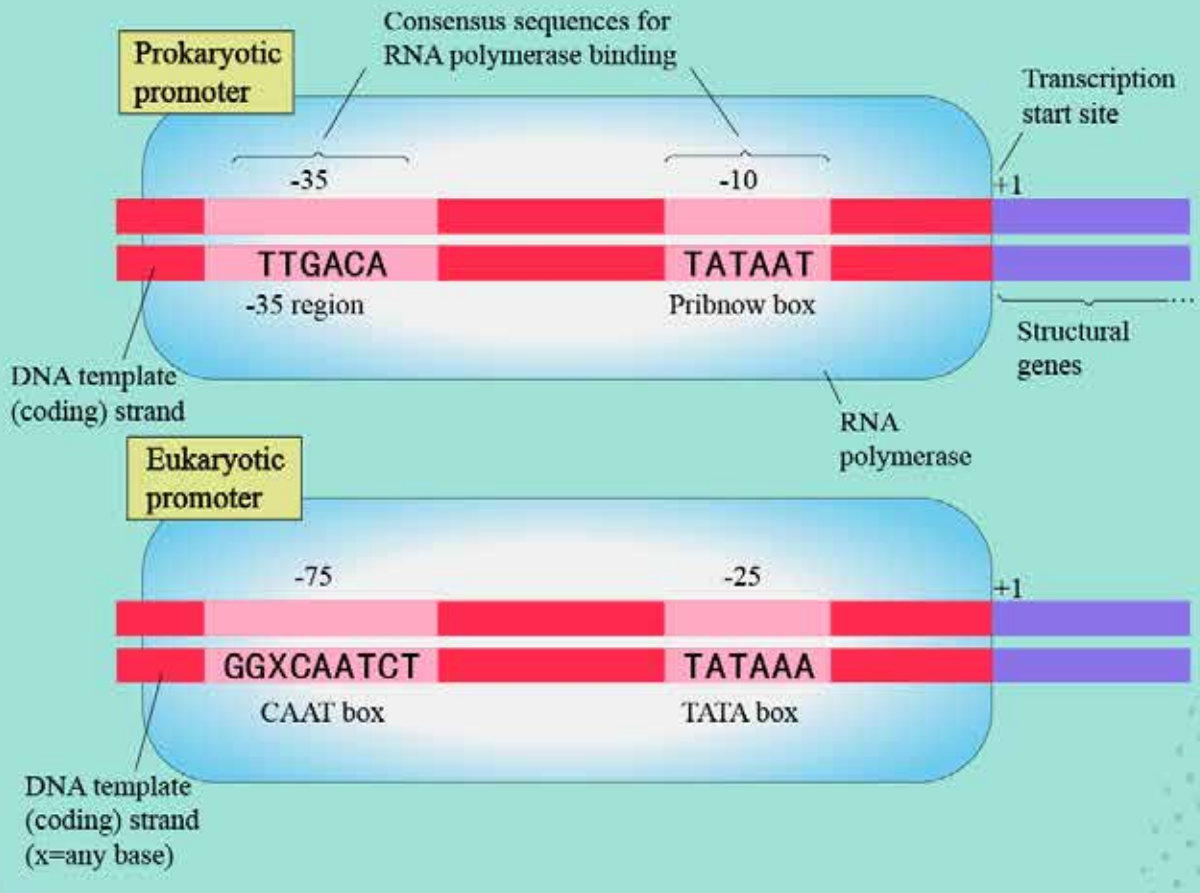
2019年6月刊
总116期

生命奥秘

2019/6
LIFEOMICS

从结构到功能——盘点2018启动子研究热点



无奇不有
生命世界
你读生命
走进科学

目录 CONTENTS

专题：从结构到功能——盘点2018启动子研究热点

前言 01

一、启动子的结构 02

1. 原核生物启动子 02

2. 真核生物启动子 02

二、启动子的分类 05

1. 按照生物种类分类 05

2. 按照转录模式分类 05

3. 按照结构特点分类 06

4. 按照启动效率分类 06

5. 按照来源分类 06

三、启动子的分析研究方法 11

1. 启动子序列和功能预测 11

2. 启动子与蛋白质相互作用分析 13

3. 启动子克隆 14

4. 启动子活性检测 14

5. 启动子结构观察 14

四、启动子功能实现的调控因素 16

1. DNA甲基化调控 16

2. 增强子等顺式作用元件与基因调控 16

3. 单核苷酸多态性影响启动子功能 18

五、展望 20

本刊文章主要由国外网站文章编译而成，如有版权问题，请版权所有人与本刊联系。
凡本刊所载文章，版权归作者本人和本刊所有，如需转载，请注明作者及出处“生命奥秘”。
本刊提供的任何信息都不能作为医疗凭证和依据，仅供科研参考。

专题

从结构到功能——盘点2018启动子研究热点

叶景佳，女，博士，研究方向：肿瘤生物治疗及发病机制

前言

为了适应环境，维持生存和繁衍，原核生物和真核生物各自都有一套独特的调节基因表达的机制。启动子是调控基因表达的顺式作用元件（*cis*-acting element），通过启动子进行的转录起始调控正是基因表达的关键，而启动子也由于这方面的强大功能，成为生物工程的重要工具。近年来随着以基因技术、冷冻电镜技术等为代表的研究方法的突破和完善，各个相关领域尤其是生物信息学领域的快速发展，以及不同领域之间交叉融合的研究趋势，研究人员在启动子的研究和应用方面取得了不少优秀成果。本文从启动子的结构、分类、研究方法和功能调控等方面进行梳理，介绍近年来关于启动子研究的热点和重要突破，着重介绍2018年在高水平刊物上发表的关于真核启动子的结构、功能及相关应用的最新研究。

一、启动子的结构

所谓启动子是指能确保转录精确而有效起始的DNA序列。

1. 原核生物启动子

在原核生物中，启动子存在两段共同序列——中央大约位于-10bp处的TATA区，及-35bp附近的TTGA区，这两段都是RNA聚合酶的结合位点。它们之间的距离在16-19bp之间，过短或过长都会影响启动子的活性。

2. 真核生物启动子

在真核生物中，情形更复杂一些。真核生物的基因调控，既有类似于原核细胞的对环境变化作出瞬时反应的调控，又有能在特定时间、特定细胞中激活特定基因的发育调控。根据RNA聚合酶的种类，真核生物启动子可分为rRNA启动子（I型）、mRNA启动子（II型）和tRNA启动子（III型）三类。凡是编码功能蛋白的核基因都通过RNA聚合酶II进行转录，其调控过程除了受到以启动子和增强子为代表的顺式作用元件的影响，还受到“反式作用因子”（*trans*-acting factor）的调控。由于转录主要在常染色质上进行，所以染色质的结构状态、DNA链的空间结构变化都会对调控作用产生影响。此外，RNA聚合酶的作用也不可忽视。

简单来讲，真核基因的启动子的TATA序列位于约-25~-35bp处，在-70~-80bp附近含有CCAAT序列（CAAT box），在-80~-110bp附近含有GCCACACCC或GGGCGGG序列（GC box）。TATA区主要负责使转录精确地起始，CAAT区和GC区主要控制转录起始频率。由于不是所有启动子都包含这三种序列，而且其间的距离也大于RNA聚合酶所能覆盖的范围，因此为了了解转录因子结合到启动子DNA序列的分子基础，我们需要对转录起始复合物的结构进行详细的鉴定和描述。2018年12

月刊登于（科学）《*SCINECE*》杂志上的一篇文章中，作者利用冷冻电镜技术，对一般转录因子IID（TFIID）的结构进行了分析，推导出了TFIID特异性结合到启动子区域的机制模型：

真核生物转录起始前复合物（transcription preinitiation complex, PIC）包括一般转录因子和RNA聚合酶Ⅱ（PolⅡ）。一般转录因子中的TFIID负责初始识别核心启动子，是参与核心启动子识别过程中最关键的转录因子之一。人TFIID是由TATA盒结合蛋白（TATA box binding-protein, TBP）和13个进化保守的TBP相关因子（TAF1到TAF13）组成的三叶（A、B和C叶）复合体，其中6个TAF存在于两个拷贝中。

在文中，作者报告了TFIID的冷冻电镜结构——其中叶片C的分辨率为4.3Å、叶片B的分辨率为4.5Å，叶片A的分辨率则为9.8Å，同时作者还结合化学交联质谱和结构预测生成了一个TFIID进化保守核心的完整结构模型。TFIID建立在TAF的二聚体支架上，其中心含有叶片C上的TAF6二聚体，该二聚体连接到叶片A和B。叶片A和B均围绕TAF4、-5、-6、-9、-10和-12组织，同时还含有行使独特功能的其它亚基（图1）。叶片A含有TAF11和TAF13，可作用于TBP使其保持被抑制的状态，直到TFIID与启动子结合，此时TBP被加载到DNA上。叶片B包含TAF8，后者延伸开来将叶片B和C牢牢地固定在一起。 叶片B将TAF4置于适当位置以稳定上游DNA的结合并募集TFIIA。叶片C除了含有TAF6和TAF8外，还含有TAF1、-2和-7，它们都可与下游核心启动子序列结合。

通过对冷冻电镜图像进行计算检索，作者描述了在TFIIA和启动子DNA均存在的情况下apo-TFIID（载脂蛋白-TFIID）和TFIID的五种结构状态，包括apo-TFIID的两种主要状态——标准态和扩展态，以及在TFIIA和核心启动子DNA均存在的情况下观察到的另外三种状态——扫描态、重排态和结合态。叶片A携带TBP，从原来靠近标准态叶片C的位置向靠近扩展态叶片B的位置移动了150 Å，其携带的TBP处于抑制状态，且仅在启动子结合的情况下释放。识别不同的TFIID状态有助于获得TFIID启动子结合的机制模型（图1）。研究者设想TFIID首先通过TAF1和TAF2结合下游启动子核心元件。这种结合和叶片A的灵活附着有助于将上游DNA定位在TBP附近。然后TBP寻找TATA盒或其序列变体。TBP与上游启动子核心序列的结合受到TFIIA与叶片B内的TAF4和TAF12相互作用的促进。当TBP最终结合到启动子上时，它就从叶片A上释放出来，开放TFIIB的结合位点，然后可以募集Pol II。

TFIID的结构也可以让研究者推断出与转录激活因子作用的调节区域的位置，以及与活性染色质标志（负责招募和调节TFIID功能）结合的位点。

该研究建立了一个机制模型，该模型可说明TFIID如何防止TBP与基因启动子序列以外的DNA非特异性结合，从而防止异常的PIC装配和错误的转录起始。该模型还可呈现TFIID如何将TBP加载到无TATA的启动子上，以及激活因子和染色质标记如何引导TFIID的招募和PIC的组装。在此基础上，我们可以对TBP结合启动子序列过程中的调节因子以及PIC装配过程中的细节进行更深入的研究。

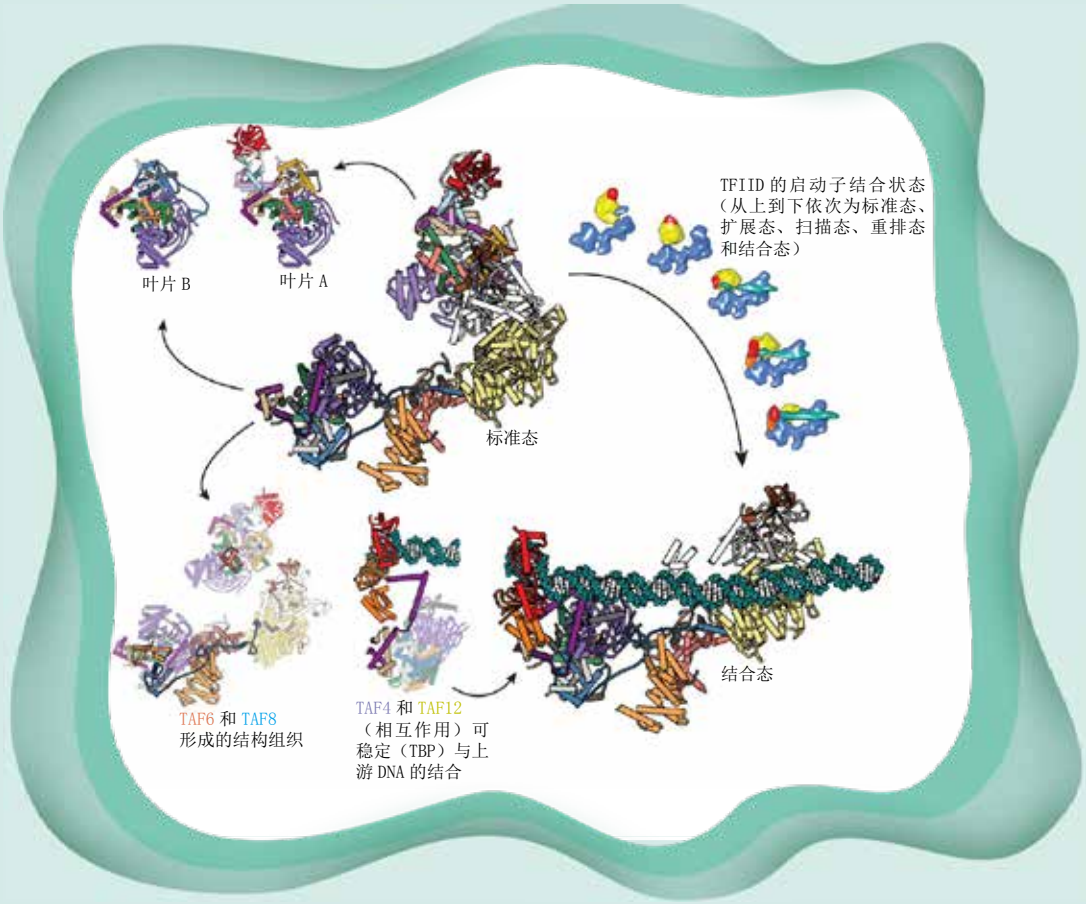


图1 人类TFIID的结构。apo-TFIID的结构以标准态显示，而与启动子结合的TFIID的结构以结合态显示。TFIID中的叶片A和B拥有类似的架构，包含组蛋白折叠域，其组织方式类似于组蛋白八聚体。TFIID的TAF6亚基使TAF的核心二聚化。TAF8亚基将叶片B和C牢固地系在一起。在启动子结合过程中观察到的五种状态的TFIID，引导研究者获得了TBP加载到启动子DNA上的机制模型。

图片来源：Patel AB, *et al.* (2018) Structure of human TFIID and mechanism of TBP loading onto promoter DNA. *Science*, 21: 362(6421).

二、启动子的分类

依照不同的标准和研究目的，可将启动子划分成不同的类型。

1. 按照生物种类分类

按照来源生物种类，可将启动子分为原核生物启动子和真核生物启动子。对应不同类型的RNA酶，又可将真核生物启动子分为rRNA启动子（I型）、mRNA启动子（II型）、tRNA启动子（III型）三种。编码功能蛋白的基因都通过RNA聚合酶II转录表达。

2. 按照转录模式分类

根据转录模式的不同，还可将启动子分为组成型启动子、组织特异性启动子和诱导型启动子。组成型启动子又称非特异型启动子，这类启动子调控的基因表达水平大致恒定，在不同部位和组织间差异不明显，也不受时间限制。如花椰菜花叶病毒（CaMV）35启动子，已被广泛应用于植物基因工程。组织特异性启动子又称器官特异型启动子，这类启动子可以调控基因在特定的组织或器官中表达，其表达受到一些组织特异性表达的调控元件以及增强子和沉默子等元件的共同调控。诱导型启动子可以在诱导物刺激下大幅提高所调控基因的转录水平，诱导物可以是环境、理化因子及微生物等。这几类启动子都被广泛应用于基因工程。

3. 按照结构特点分类

若按照结构特点来区分，那么自然界中还存在一些其它类型的启动子，如双向启动子(bidirectional promoter)——位于两个相邻且转录方向相反的基因之间的一段DNA序列，它在真核生物中广泛分布，也称为头对头基因。大多数双向启动子结构中缺少TATA盒，但GC含量较高，还有丰富的CpG岛，在基因工程研究中也有一定的应用。

4. 按照启动效率分类

按照启动效率的高低，启动子还能被分成强启动子和弱启动子。在实际应用中启动效率并非总是越强越好，有时强启动子会干扰其它蛋白表达，影响生物体正常代谢。如果在研究中需要调控蛋白表达的产量，那么诱导型启动子则可以部分实现这个功能。为了对表达产物进行更为精细的调控，我们还可以通过人工合成启动子文库来实现。

5. 按照来源分类

除此之外，从启动子的来源出发，还可将启动子分为内源启动子和外源启动子。在细胞或生物中引入外源启动子以表达各种基因，已成为一种常规操作。如上文所述，为实现对基因表达水平的精细调控，或实现多个基因表达强度的最佳组合，研究者可以根据需要，对不同来源的启动子进行设计加工，常用方法有引入突变或随机序列、将不同来源的调控元件和启动子序列进行重组以及增加或删除特定的调控元件等。这些外源启动子或人工合成启动子在基础研究和商业领域都显示了巨大的应用潜力。

以玉米为例，通过生物技术开展昆虫防治并改善抗除草剂性状已大大提高了玉米的产量。其它方面的应用还包括改善种子营养成分、工业蛋白质生产、治疗性生产、抗病性和非生物胁迫抗性等，并且还能提高产量。生物技术还能拓展控制植物生长和繁殖的重要机制的相关基础研究。不少转基因优化策略都与启动子结构息息相关。下文将重点介绍截止2018年，在玉米转基因技术应用中，启动子的选择和使用情况。

上世纪末到本世纪初，多个国家开展了针对玉米基因资源和玉米基因组学的研究，并取得了丰硕成果。这些数据和成果不但扩展了人们对基因组水平上的基因活动的理解，也是基因芯片技术的基础，还为基因组水平上的启动子开发提供了支持。有趣的是，尽管玉米中存

在多种具有活性的启动子，但实际可在玉米中应用的启动子只占一小部分——大多数转基因研究以及大多数商业状况都依赖于20多年前开发的一小部分启动子。目前使用的启动子中有许多是异源的，一些来自植物病原体，如花椰菜花叶病毒（CaMV）35S启动子，一些来自单子叶植物，如水稻Actin1、玉米衍生的Emu或甘蔗泛素启动子。它们的应用范围包括检测一些促进启动子活性的成分，使用异源启动子开发一些性状，以及将玉米启动子用于发育过程研究等。

在基础研究领域，转基因技术极大地扩展了研究玉米生物学的工具箱。CaMV 35S启动子是实验中最早使用的启动子之一。与双子叶植物的情况相反，研究人员发现，在性状基因启动子附近加入的5’端非翻译区内含子，可大大提高转基因的效能。表1列举了包括CaMV 35S启动子在内的可应用于玉米基础研究的各种启动子。

到目前为止，尽管在商业应用中使用的启动子列表相对有限，但是公开表征的启动子越来越多。继续应用基因组水平分析有望进一步丰富这一列表。在基因工程的重组DNA时代，将转基因元件分为启动子、编码序列和终止子的做法十分便利。高质量基因组信息的出现和DNA合成技术的改进将这种研究开发模式转变为一项信息驱动的事业。而且过去在分子克隆中需要考虑的是否存在可用的特定限制性内切酶位点等已不再是限制。

表1. 应用于玉米基础生物学研究的启动子特征描述

研究目的	性状启动子	标记启动子	说明
利用基因枪技术制备转基因植物	CaMV 35S	CaMV 35S	启动子还含有玉米5'端非翻译区的Adh1内含子
利用基因枪技术制备转基因植物	ZmAdh1	CaMV 35S	启动子还含有玉米5'端非翻译区的Adh1内含子
利用基因枪技术制备转基因植物	/	CaMV 35S	启动子还含有玉米5'端非翻译区的Adh1内含子
利用基因枪技术制备转基因植物	CaMV 35S	ZmALS	pALGenLUC不能产生稳定的转基因
内含子介导的转基因表达增强	CaMV 35S	/	玉米ADH1内含子的使用及其在不同构建体之间的变化
烟草中ZmRab17启动子的分析	ZmRab17	NOS	ABA诱导启动子
ZmPPDK1基因的分析	ZmPPDK1	/	显示两种不同的表达程序调节基因活性

瞬时启动子分析	Zm13	/	花粉特异性启动子，也检测了ZM13终止子
水稻Zmubi1启动子的分析	ZmUbi1	ZmUbi1	/
CaMV 35S双增强子的表征	小麦 α-amylase	CaMV 35S	启动子活性随叶片年龄、叶片位置和事件的变化而变化
原生质体和BMS细胞中的启动子分析	ZmADH1和CaMV 35S	/	报告了内含子增强、TMV-ω增强以及Hvα-淀粉酶5’-和3’-非翻译区可以促进蛋白质的产生
转化技术发展	CaMV 35S	CaMV 35S	启动子还含有玉米5'端非翻译区的Adh1内含子
转化技术发展	ZmAdh1	CaMV 35S	启动子还含有玉米5'端非翻译区的Adh1内含子
转化技术发展	ZmMFS14	CaMV 35S	绒毡层特异性启动子
转化技术发展	Zm13	CaMV 35S	花粉特异性启动子
启动子活性对内含子的要求	ZmGAPA1	/	报告内含子增强，启动子在幼苗中产生0.5%的mRNA
玉米转化的构建	ZmUbi1	/	高利用率玉米泛素启动子的确定
玉米转化技术	CaMV 35S	CaMV 35S	建立了农杆菌介导的玉米转化
微管蛋白合成	MFS14	CaMV 35S	探讨BMS细胞中a、b微管蛋白的平衡
微管蛋白合成	CaMV 35S	CaMV 35S	启动子还含有玉米5'端非翻译区的Adh1内含子

B-Peru等位基因分析	B-Peru	CaMV 35S	鉴定了一种糊粉特异性启动子，并对其调控元件进行了表征
质体靶向	CaMV 35S	/	维管束鞘细胞中优势性状蛋白的积累
绿色荧光蛋白（GFP）表达	ZmUbi1	CaMV 35S	经修饰的GFP表达
P1-rr启动子分析	ZmP1-rr	CaMV 35S	不同事件之间的表达谱不同
选择性标记的开发	ZmUbi1	eCaMV 35S	/
Mu 转座	CaMV 35S	ZmUbi1	描述了MuDR元素
P1-rr启动子分析	多种P1-rr	CaMV 35S	表征了P1-rr启动子。活性主要定位于花卉组织
表位标记分析	ZmUbi1	ZmUbi1	检测了蛋白质和RNA标记在转化细胞系中的有效性
ZmPEPC启动子特征	PPCZm1	CaMV 35S	启动子具有叶肉特异性
转基因活性定位分析	5126 或 MS45	CaMV 35S、OsAct1或 ZmUbi1	描述了两种花药特异性启动子的使用
GUS 表达	Pi-rr、TaPeroxidase和Rsyn7	CaMV 35S	描述了如何利用MAR影响转基因表达
玉米adh-1 MAR分析	ZmUbi1	CaMV 35S	ZmAHD1 MAR影响了ZmUbi活性概况
玉米MEG1基因的分析	ZmMEG1	/	显示ZmMEG1在胚乳基底转移区具有活性
玉米和烟草中的启动子分析	ZmHyPRP	OsAct1	表征了一个早期胚胎特异性启动子

转基因烟草中的启动子分析	ZmPul1	NOS	显示玉米Zpu1启动子在烟草中具有种子优先活性
GFP作为组织标记物marker	Glb-1 和 27、22 或 19 kDa Zn	ZmUbi1	包括定量和空间表达数据
检查了分枝控制机制	ZmPIN1a 或 ZmDR5rev	2X CaMV 35S	将各个基因的gDNA序列进行翻译融合
描述可视化质体的构建	ZmUbi1、ZmC4PepC或ZmRBCS1	/	发现荧光蛋白的玉米密码子优化版本在稳定的转基因中表现更好
ZmDSUL表征	ZmDSUL 或 CamV 35S	CaMV 35S	描述了一种雌性配子体特异性启动子
ZmFEA3表征	ZmFEA3、ZmWUS1、pYABBY14和pOp	/	描述了三种分生组织特异性启动子
CENH3 tailswap技术	ZmCENH3	ZmUbi1	展示了玉米中CENH3 tailswap技术的应用
玉米单倍体诱导QTL	ZmMTL	ZmUbi1	证明MATRILINEAL（一种精子特异性磷脂酶）是花粉特异性的

介绍了在转基因环境中玉米启动子在玉米植株、玉米细胞培养或异源植物中的应用实例。重点关注目前应用于转基因研究的启动子的选择和用于驱动植物选择标记基因表达的启动子。

表格来源：Nuccio ML. (2018) A Brief History of Promoter Development for Use in Transgenic Maize Applications. *Methods Mol Biol*, 1676: 61-93.

三、启动子的分析研究方法

随着生命科学的发展和研究手段的不断进步，越来越多的技术方法可被应用于启动子相关的分析和研究。下文便列举并介绍几种具有代表性的启动子研究方法。

1. 启动子序列和功能预测

生物信息学的发展使得利用软件对启动子的结构特征、关键位点及其结合因子进行分析和预测成为可能。另外，各类在线数据库也提供了海量基因序列，运用专门启动子分析软件，可为启动子研究提供非常有价值的参考。

常用的预测软件包括：预测真核Pol II启动子序列的Promoter Scan（网址https://www-bimas.cit.nih.gov/molbio/proscan/analysis.html）和Neural Network Promoter Prediction（网址https://fruitfly.org/seq_tools/promoter.html）。识别并标记核酸序列中CpG岛的EMBOSS Cpgplot（网址https://www.ebi.ac.uk/Tools/seqstats/emboss_cpgplot/）和newcpgreport（网址：http://emboss.bioinformatics.nl/cgi-bin/emboss/newcpgreport）等。此外，也不乏一些针对特定物种或对象的启动子预测和分析软件：如进行人类DNA中第一外显子和启动子预测的FirstEF（网址http://rulai.cshl.org/tools/FirstEF/）、预测脊椎动物DNA序列中Pol II启动子转录起始位点的Promoter 2.0（网址http://www.cbs.dtu.dk/services/Promoter/），以及查找人类核心启动子的Human Core-Promoter Finder by Michael Zhang（网址http://rulai.cshl.org/tools/genefinder/CPROMOTER/human.htm）等。

随着时间推移，有一些网站因人员等问题不再提供相关服务，如分析转录因子结合位点的Transcription Element Search System TESS等。也有一些网站将相关的工具和信息整理汇总，方便使用者查找，如Bioinformatics Sequence and Genome Analysis（Second Edition）（网址http://www.bioinformaticsonline.org/links/ch_09_t_6.html）和PROMOTERS & TERMINATORS（网址http://molbiol-tools.ca/Promoters.htm ）等。表2列举了一些网站提供的启动子及转录因子结合位点相关的数据库。

利用生物信息学的方法，在对启动子区域和序列、转录因子结合位点等特征进行分析，获得潜在的启动子结构后，再通过结构观察、

活性分析和功能验证（缺失突变转化分析、启动子与蛋白质相互作用分析）等，我们可进一步明确启动子的结构和功能，分析转录因子与启动子的相互作用，了解启动子的作用机制。

表2. 启动子及转录因子结合位点数据库

名称	说明	链接
真核生物启动子数据库(Eukaryotic Promoter Database, EPD)	主要收集了真核生物RNA 聚合酶II 启动子的相关数据，这些启动子都是已经被注释过的，并通过了实验验证。目前共收录了来自多个物种的4,806条启动子信息。	https://epd.epfl.ch//
植物DNA 顺式作用调控元件(plant cis-acting regulatory DNA elements, PLACE)	数据库内是从已发表文献中搜集的植物DNA顺式作用元件的基序(motif)信息。该数据库标注了每个基序和相关文献摘要，但是仅包含维管植物。	http://www.dna.affrc.go.jp/htdocs/PLACE/
植物顺式作用调控元件(Plant-CARE)数据库	不仅收录了植物顺式作用元件，而且收录了植物增强子(enhancer)和抑制子(inhibitor) 。	http://bioinformatics.psb.ugent.be/webtools/plantcare/html/
转录调控区数据库(TRRD)	主要收集真核基因转录调控区域的结构和功能组织信息，但只包含实验确认的信息。	http://www.bionet.nsc.ru/trrd/

注：根据网上相关数据库网络信息整理。

2. 启动子与蛋白质相互作用分析

酵母单杂交系统、免疫共沉淀、启动子芯片、凝胶电泳和DNaseI足迹法等方法都可用于启动子与蛋白质相互作用分析。

酵母单杂交技术是由酵母双杂交技术发展而来的，基本原理为：真核生物基因的转录起始需转录因子参与，而转录因子通常包含DNA结合结构域（DNA-binding domain, BD）和转录激活结构域（activation domain, AD）。用于酵母单杂交系统的酵母GAL4蛋白是一种典型的转录因子，其DNA结合结构域可激活酵母半乳糖苷酶的上游激活位点（UAS），而其转录激活结构域可激活RNA聚合酶。利用这两种结构域可独立发挥作用的特点，研究者可将GAL4的DNA结合结构域置换为文库蛋白编码基因，同时将已知的顺式作用元件构建到报告基因启动子上，只要表达的文库蛋白能与顺式作用元件结合，就可通过GAL4的转录激活结构域激活RNA聚合酶，启动下游报告基因的转录。由于酵母属于真核生物，通过这个方法检测到的蛋白具有天然构象，灵敏度很高。但有时由于插入的靶元件与酵母内源转录激活因子可能发生相互作用，或插入的靶元件不需要转录激活因子便可直接激活报告基因的转录，所以会产生一定数量的假阳性结果。此外，如果酵母表达的融合蛋白有细胞毒性，或不能稳定地表达，或发生错误折叠等，也有可能无法顺利结合到靶元件上，从而产生假阴性结果。

免疫共沉淀（Co- Immunoprecipitation, Co-IP）是依据抗原和抗体的特异性结合以及细菌的Protein A或G可特异性地结合到免疫球蛋白的Fc片段上这一原理而开发的。在含有抗原蛋白的溶液中加入抗目标蛋白的抗体，孵育后再加入可与抗体特异结合的结合于Agarose珠上的Protein A或G，形成“目标蛋白—抗性蛋白抗体—Protein A或G—Agarose珠”复合物，而后经低速离心或变性聚丙烯酰胺凝胶电泳等方法将复合物分离并回收。然后由免疫印迹或质谱来检测目的蛋白。该方法常用于测定两种目标蛋白质是否在体内结合，也可用于确定一种特定蛋白质新的作用搭档。上文所提到的TFIID最初就是利用免疫共沉淀技术从Hela细胞的核提取物中分离得到的。

启动子芯片是将已知启动子序列的寡核苷酸固定到芯片上，可检测单个转录因子与大量已知基因启动子相互作用的情况的一种芯片。结合染色质免疫共沉淀技术，我们可以确定特定转录因子的靶基因群。这种方法具有操作简便和高通量的特点，市场上已有商业化产品。

此外，电泳迁移率分析法（electrophoretic mobility shift assay, EMSA）——利用蛋白质与核酸结合的复合物在凝胶电泳中迁移速率远远小于游离核酸的这一原理，可对启动子及其作用蛋白进行分析和鉴定。DNase I足迹法（DNase I footprinting）则利用当DNA片段与蛋白质结合时可受到结合蛋白的保护而免于被DNase I降解的现象，将DNA—蛋白质复合物标记后用DNase I处理，产生一系列不同长度的片段，当电泳分离后显影，图谱上间断的区域则提示该位点存在结合蛋白。该方法常和EMSA法联合应用，可鉴定启动子的各种调控因子。

3. 启动子克隆

为进行启动子的活性和功能分析，将其克隆至合适的载体上使其表达“报告基因”是一种广泛使用的操作。目前基因克隆技术已经十分成熟，用于克隆的启动子片段可以通过PCR法、酶切法和基因文库法等获取。当需要对启动子活性进行调控或构建不同数量级表达强度的启动子组合时，我们可通过人工构建启动子文库来进行选择和优化。人工合成启动子文库的设计策略有理性设计、非保守序列随机化、易错PCR引入突变，以及结合调控原件进行组合优化等。在实现基因的适量表达及优化代谢途径等方面有着巨大的应用潜力。

4. 启动子活性检测

要验证已经克隆的启动子的活性，需要通过转基因技术使其进入细胞并表达报告基因——如荧光素酶基因等，然后检测和报告基因瞬时或稳定表达的情况来评估启动子活性。常用的将外源DNA导入细胞的方法有利用脂质体、磷酸钙、病毒或细菌介导、显微注射、电穿孔和基因枪等方法。我们可根据实验目的和转染对象进行选择。瞬时表达操作简单快速、安全有效；稳定表达通常需要通过抗性标记筛选出阳性克隆，耗时久，比较适合一些需要诱导表达基因或历时较长的研究。

5. 启动子结构观察

冷冻电镜技术的出现，使得生物大分子高分辨率（原子水平分辨率）的结构研究成为可能，从而推动了结构生物学的快速发展。2017年的诺贝尔化学奖即授予研发该技术的三位科学家Jacques Dubochet、Joachim Frank和Richard Henderson。这是一种通过快速冷冻蛋白质，获取大量二维图像，并在计算机程序的帮助下生成清晰三维图像的研究方法（图2）。它不仅可以解析生物分子的高分辨率结构，还能捕捉到蛋白分子的动态变化。第一章里所介绍的TFIID的结构分析以及结合到启动子区域的机制模型都基于这一技术。过去十多年其分辨率不断提高。以TFIID的结构解析为例，早期研究（1999-2004）中的冷冻电镜结构分辨率为20-30 Å，而近期研究中叶片C分辨率已达到4.3Å。

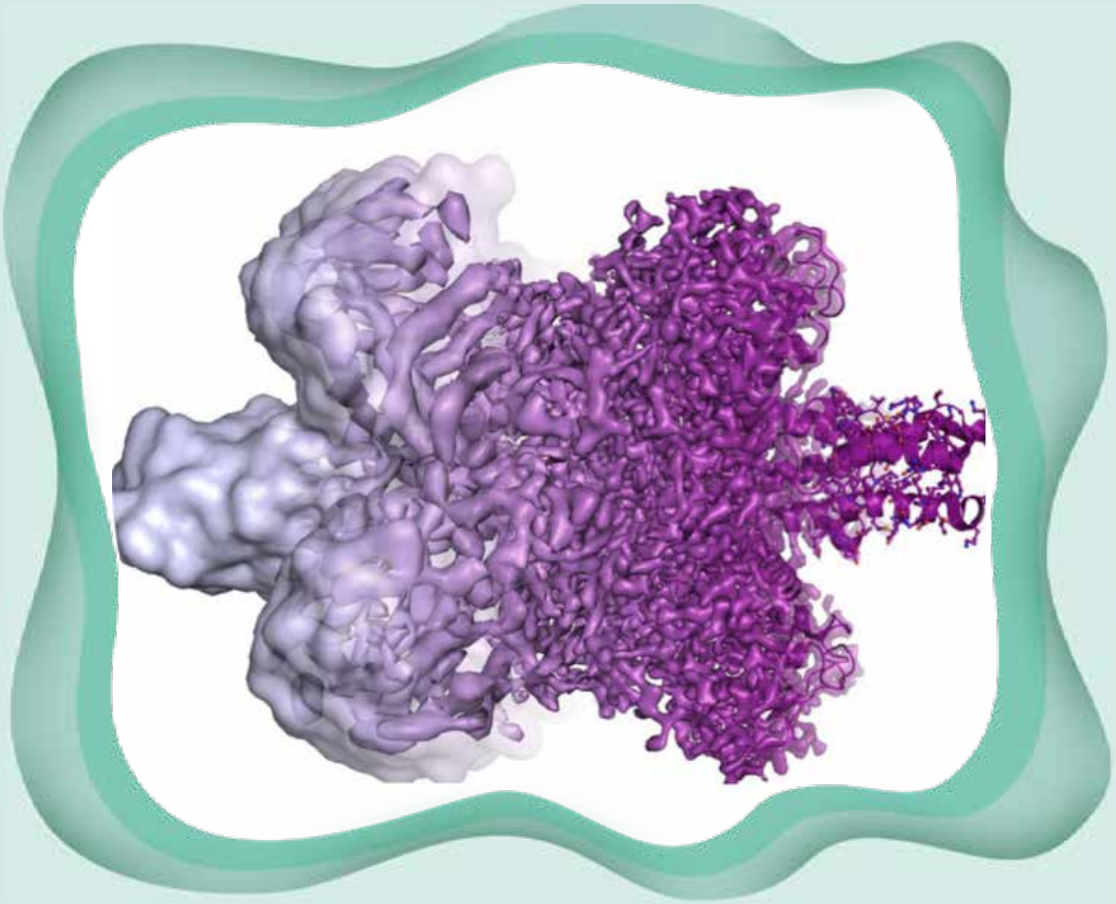


图2. 冷冻电镜技术是一种通过快速冷冻蛋白质，获取大量二维图像，并在计算机程序的帮助下生成清晰三维图像的研究方法。

图片来源：

https://www.nobelprize.org/uploads/2018/07/NobelPrizeLessons_2017_Chemistry_Slideshow.pdf

四、启动子功能实现的调控因素

启动子的识别与分析是对其进行表达调控研究的前提和基础，在生物体内其功能的实现还受到各种调控因素的作用，这些因素主要包括DNA甲基化调控、增强子和沉默子等元件的协同作用，以及启动子多态性的影响等。

1. DNA甲基化调控

DNA甲基化（DNA methylation）是一种基因修饰途径，常见于基因的5'-CG-3'序列。在甲基转移酶的作用下，CG核苷酸中的胞嘧啶被选择性地添加甲基，在启动子序列中DNA甲基化常出现在CpG岛。DNA甲基化可通过引起染色质结构变化、干扰转录因子与启动子识别位点结合等方式影响启动子功能，进而影响基因表达。研究表明，肿瘤抑制基因被启动子甲基化沉默是人类肿瘤中的常见现象，研究这类现象在不同肿瘤中的作用机制及其与患者预后的关系，可以为肿瘤的诊断和治疗提供新的思路和途径。

2. 增强子等顺式作用元件与基因调控

真核生物的基因表达受到各个水平上的调控，在转录水平上主要受到特定的顺式作用元件（如增强子等）和反式作用因子（如一些蛋白质或激素蛋白质分子复合物等）的调控。关于增强子对基因表达的调控，研究者提出了各种不同的机制，包括增强子追踪、链接、成环或动员至转录工厂（transcription factories）。但Eileen E. M. Furlong等人认为这些机制尚不足以解释在活细胞、组织和胚胎中观察到的转录动力学及转录调控的精确度。他们在最近发表的一项研究中提出了一种动态三维中心（Hub）模型，该模型结合了新出现的证据，并整合了经典模型的各种元素。作者认为基因的激活在许多位点上可分成两个步骤：增强子和启动子接近（局部压实）准备初期表达，然后细微的拓扑变化触发激活（图3）。

根据他们提出的这种增强子—启动子作

用的“中心和聚集”模型，增强子不需要直接接触目标启动子，而只需在100至300 nm范围内接近启动子。这个多介质的联合或聚合复合物、起始前复合物和Pol II都可以作为增强子与目标启动子之间的桥梁（图3）。

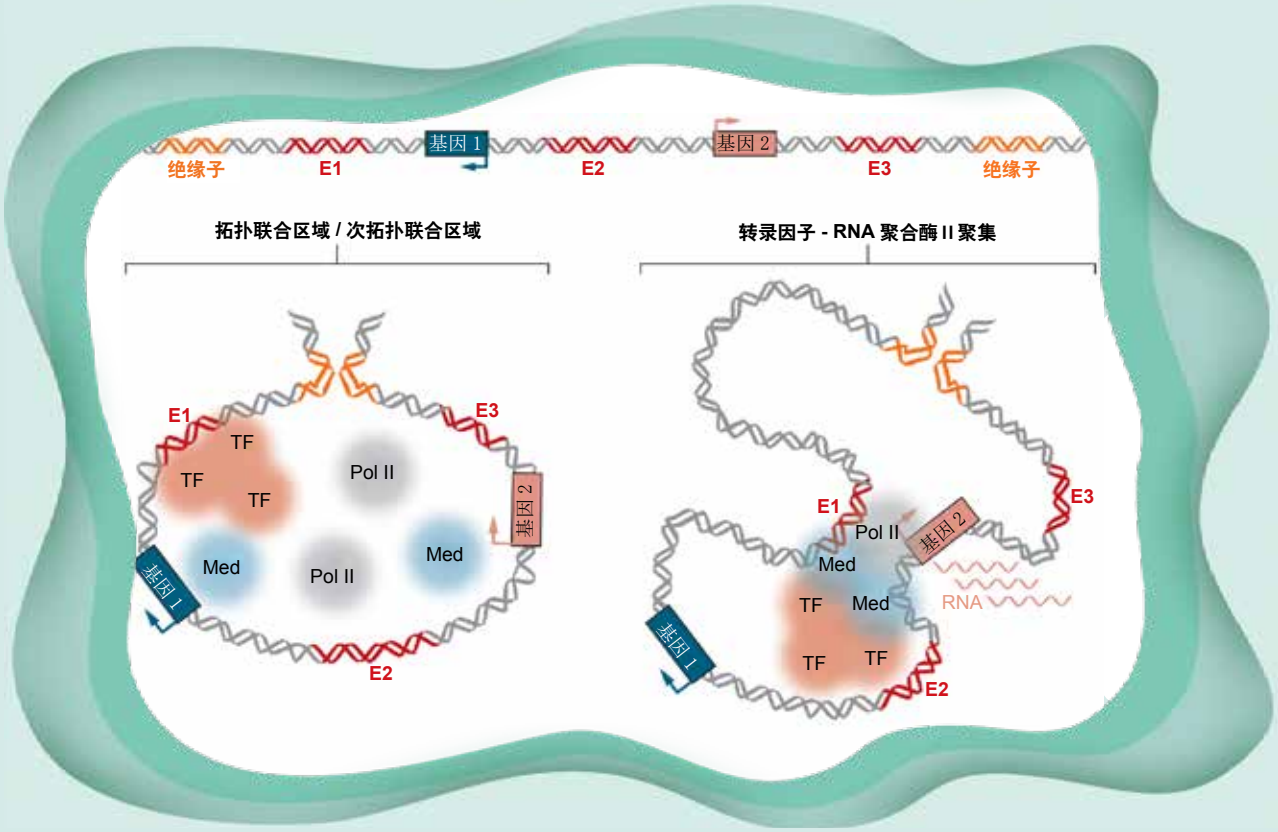


图3. Hub及凝聚模型。预形成的拓扑结构（在hub或微环境中）提高了局部转录因子TF、共激活因子和Pol II的浓度，不同的增强子（E）可动态共享共同资源。E1: 增强子1，E2: 增强子2，E3: 增强子3，TF：转录因子，Med：介导物。

图片来源：Furlong EEM *et al.*, (2018) Developmental enhancers and chromosome topology. *Science,* 361: 1341-1345.

3. 单核苷酸多态性影响启动子功能

单核苷酸多态性（Single Nucleotide Polymorphism, SNP）主要是指在基因组水平上由单个核苷酸的变异所引起的DNA序列多态性。SNP在人类基因组中广泛存在，既可能在基因序列内，也可能在基因以外的非编码序列上。发生在非编码区的启动子序列上的SNP有可能通过改变转录因子结合位点而影响所调控基因的表达。

以往的研究发现，启动子序列多态性与多种疾病的发生相关，研究相应基因的启动子多态性可以有助于疾病发生机制的研究和基因诊疗手段的开发。Joon-Yong An等人发现在自闭症谱系障碍（autism spectrum disorder, ASD）中，非编码基因组的突变对疾病发生的风险有一定的影响。特别是来自启动子的进化保守区域、远端启动子或转录因子结合位点的核苷酸突变。这些突变可能通过与启动子区域的增强子元件相互作用来干扰基因转录，而不是直接干扰转录起始。

资讯 · 频道
www.LifeOmics.com

特约编辑招聘启事

为了及时收集生命科学最新资讯、提高《生命奥秘》办刊质量，现面向从事生命科学或对这学科有浓厚兴趣的科研人员、学生诚聘特约编辑（兼职）。

职位职责：

独立完成《生命奥秘》专题的策划：对基因组学、蛋白组学、生物信息学和细胞生物学等学科的发展以及生物医学领域相关技术（例如基因诊断技术、干细胞和克隆技术、生物芯片技术等）的应用进行翻译及深入评述。

选题要求内容新颖、评述精辟、注重时效和深入浅出。尤其欢迎以自身系统研究为基础的高水平译述与评论，结合所从事的科研工作提出自己的见解、今后设想或前瞻性展望。

要求：

1.具备基因组学、蛋白组学、生物信息学、细胞生物学等生命科学学科背景；
2.具备良好的生命科学前沿触觉；
3.具备较高的外文文献翻译、编译水平；
4.具备较强的选题策划、资料搜集、组织能力，以及专业稿件撰写能力；
5.具有高级职称；或者拥有（正在攻读）该领域的最高学位。

有意者请将个人简历发送至 editor@lifeomics.com

五、展望

基因的表达是一个经历多层次调控、受到多因素影响的复杂过程。在过去相对于人们对于编码序列的关注程度，作为非编码区序列的启动子获得的关注和投入相对较少。但数十年间，基因工程技术的发展已极大地推动了启动子结构、功能和应用的研究。随着人们对基因表达调控的认识不断深入、新的研究方法和手段不断更新，关于启动子的认识越来越全面和深入，其应用也会更广泛和高效。

现在人们可以在多个层面上对启动子的调控进行深入分析，并在原子水平上探索其结构和细微变化。新的研究方法和工具可以产生和分析大量数据，为启动子调控基因表达的过程提供更加广泛的研究基础和更多新的研究线索。

在未来，可以期待随着包括生物信息学、数据处理技术和显微技术等在内的多个领域的技术不断发展，对启动子的研究既可以聚焦至特定位点，也可以拓展至启动子以外，各项技术的交叉融合与联合应用，将会促进启动子的研究在各个领域获得突破性的进展：包括对启动子等非编码区功能进行精细标注、研究启动子与其它顺式作用元件和各种转录调控因子相互作用的新机制，以及相关分子组合与重排的各种细节；整合比较基因组学、功能基因组学等多个领域的工作，从大规模数据分析中提取线索，进行启动子相关疾病发生发展机制的研究及基因诊疗手段的开发；完善人工合成启动子的设计，通过启动子调控构建标准化的生物系统；而且在最新研究的基础上开发新的技术手段、研制新的设施设备，发掘启动子的新功能和新应用。上述所有相关研究的前景从未显得如此美好和令人振奋过。

图片来源：

https://www.nobelprize.org/uploads/2018/07/NobelPrizeLessons_2017_Chemistry_Slideshow.pdf

参考文献：

1. Patel AB, *et al.* (2018) Structure of human TFIID and mechanism of TBP loading onto promoter DNA. *Science*, 21: 362(6421).

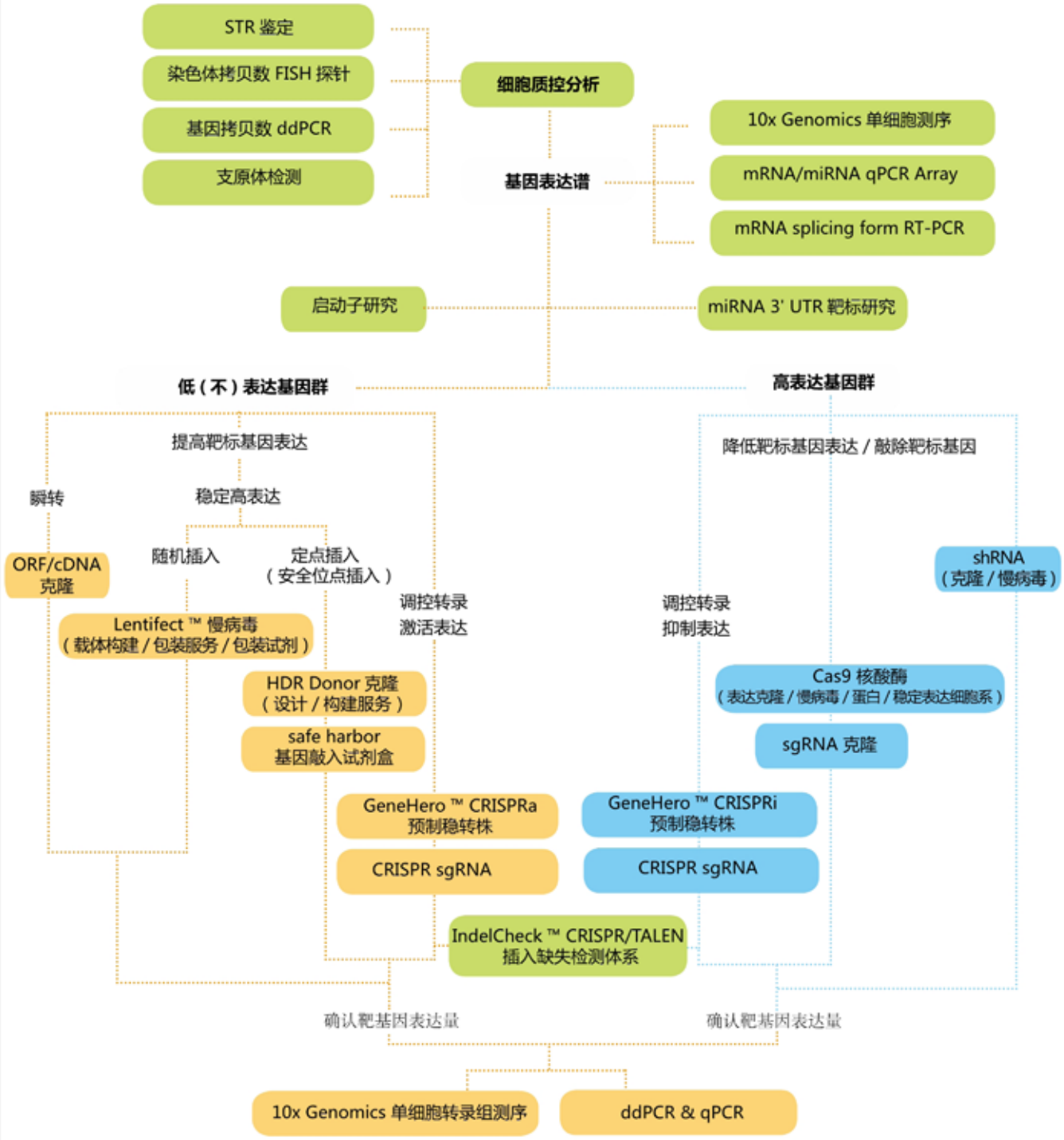
2. Nuccio ML. (2018) A Brief History of Promoter Development for Use in Transgenic Maize Applications. *Methods Mol Biol.* 1676:61-93.

3. Furlong EEM *et al*., (2018) Developmental enhancers and chromosome topology. *Science*, 361: 1341-1345.

4. An YJ *et al.*, (2018) Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder. *Science*, 362(6420).

基因功能研究产品和服务配套图

* 颜色框可以点击进入详细网页

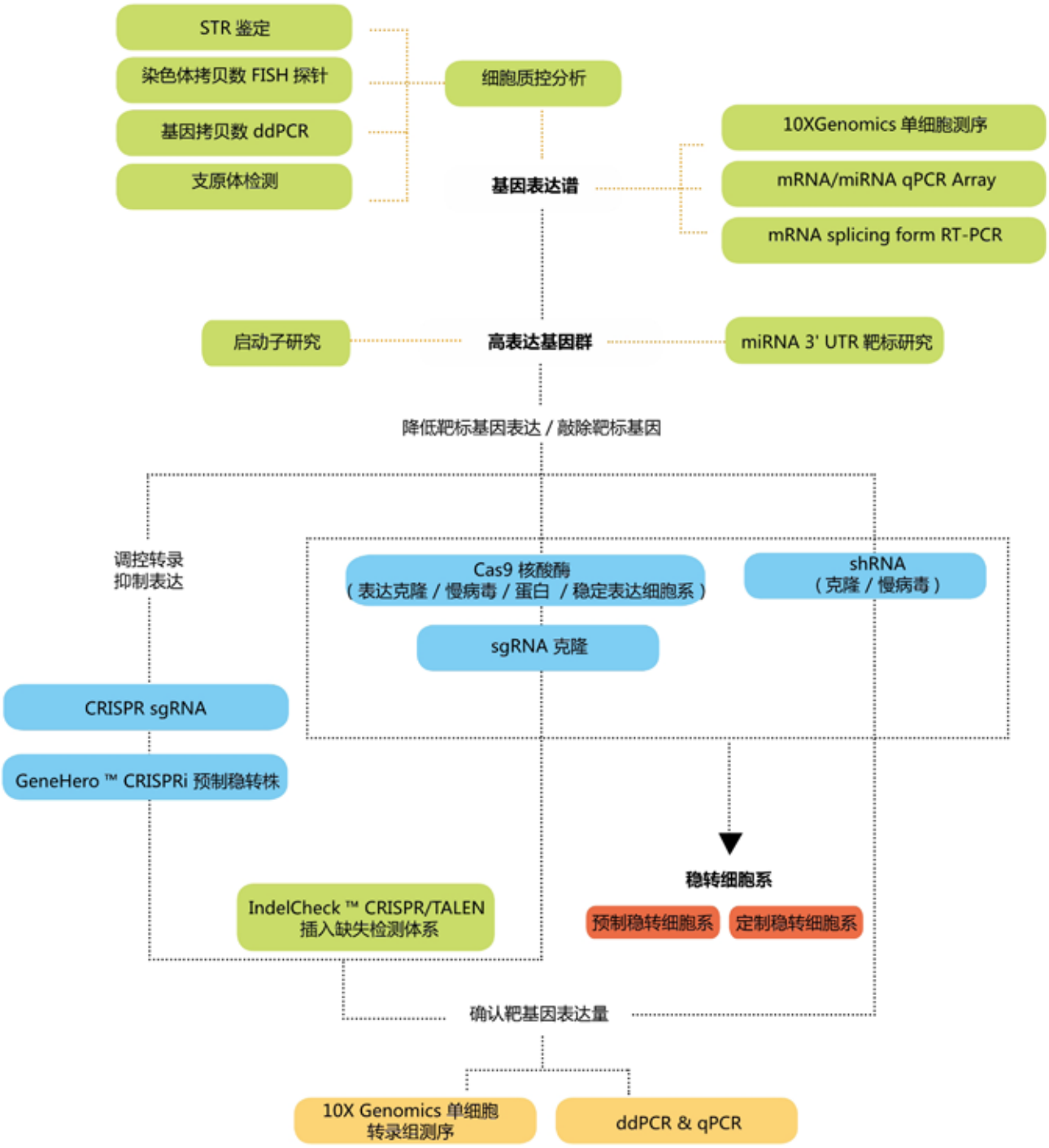


- 相关产品及服务：【转染试剂】【AAV 包装服务】【荧光素酶检测试剂盒】【细胞系构建服务】
- 客户发表文章超过 100,000 篇，【点击查看】

复能基因 FulenGen　易锦生物 iGeneBio　扫一扫，关注官方微信

高表达基因功能研究产品和服务配套图

* 颜色框可以点击进入详细网页

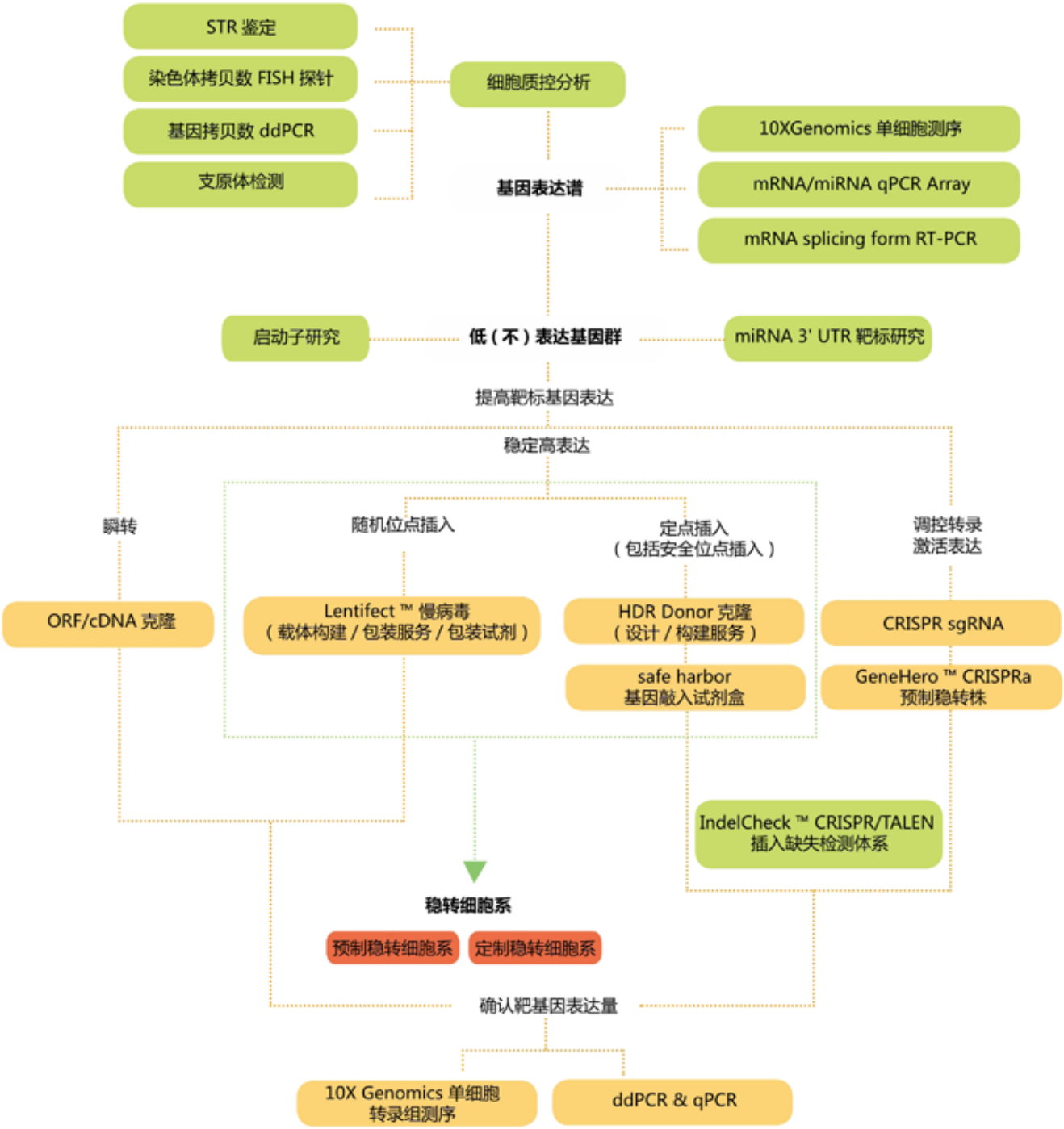


- 相关产品及服务：【转染试剂】【AAV 包装服务】【荧光素酶检测试剂盒】【细胞系构建服务】
- 客户发表文章超过 100,000 篇，【点击查看】

复能基因 FulenGen　易锦生物 iGeneBio　扫一扫，关注官方微信

低（不）表达基因功能研究产品和服务配套图

* 颜色框可以点击进入详细网页



- 相关产品及服务：【转染试剂】【AAV 包装服务】【荧光素酶检测试剂盒】【细胞系构建服务】
- 客户发表文章超过 100,000 篇，【点击查看】

复能基因 FulenGen　易锦生物 iGeneBio　扫一扫，关注官方微信

合办专题专刊
网站广告合作
邮件群发推广
请致电(020)32051255

www.LifeOmics.com