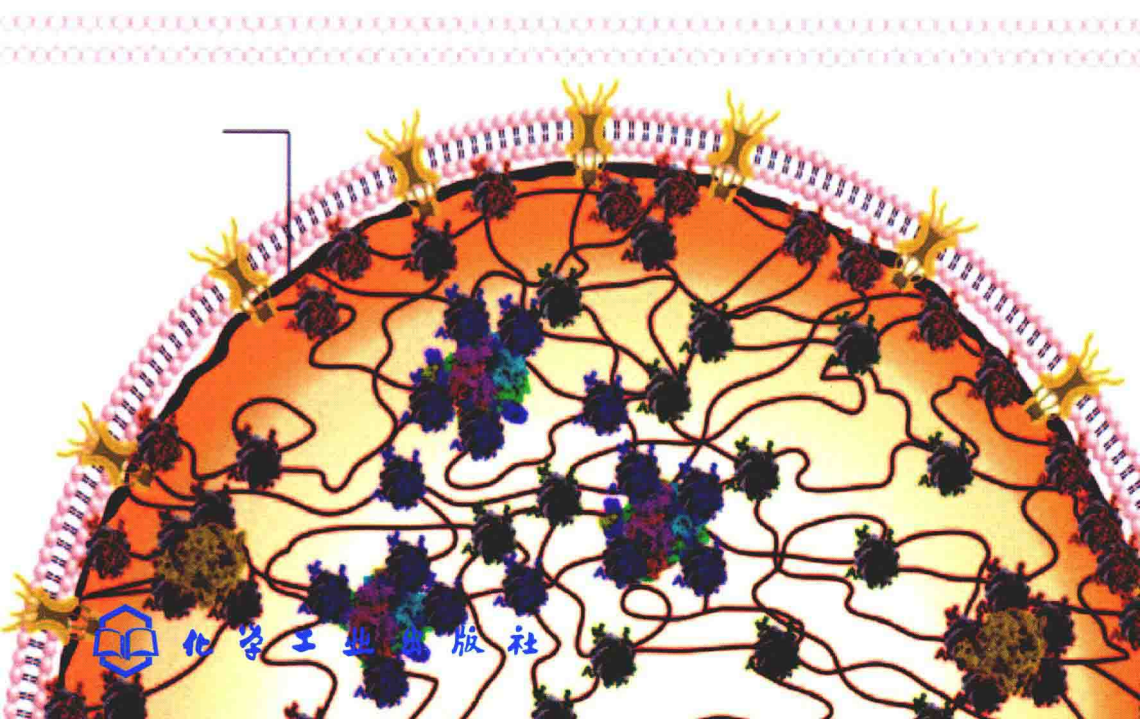


Mechanisms of Gene Regulation

基因调控机制

[芬兰] C. 卡尔伯格 (Carsten Carlberg) 著
F. 美恩 (Ferdinand Molnár)

秦玉琪 钟耀华 译



化学工业出版社

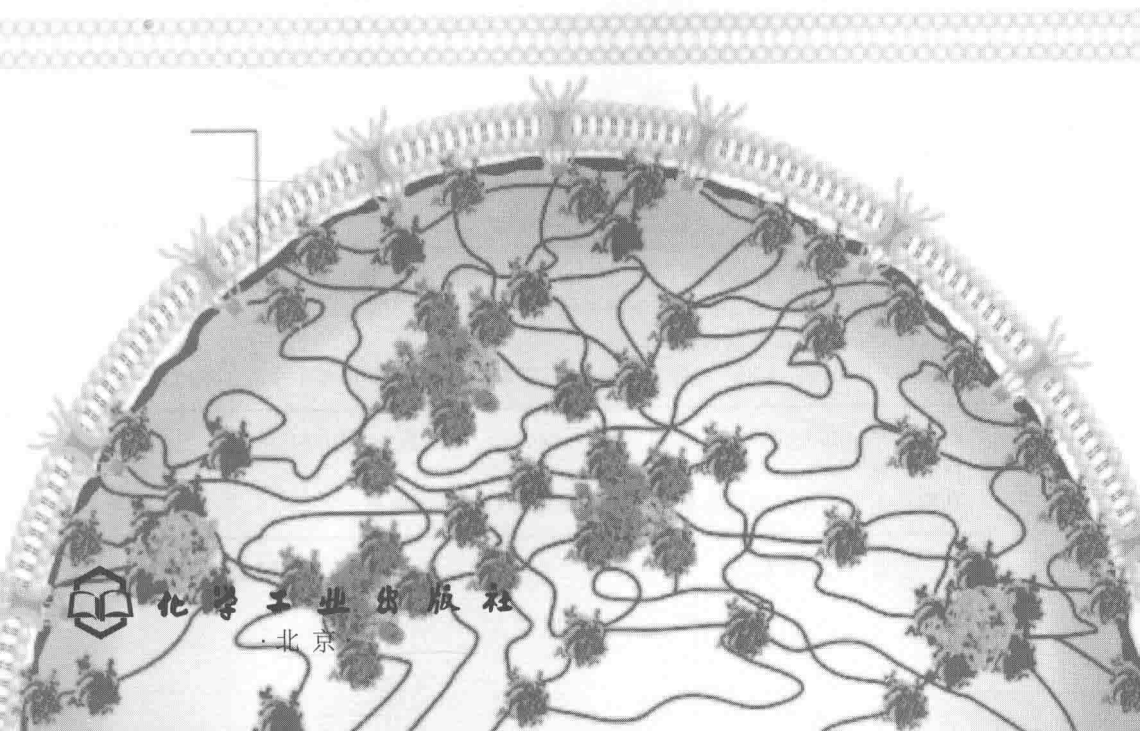
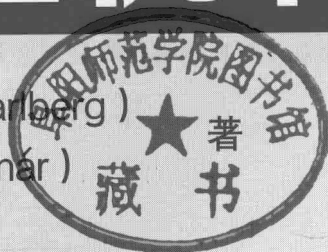
Mechanisms of Gene Regulation

基因调控机制

[芬兰] C. 卡尔伯格 (Carsten Carlberg)

F. 美恩 (Ferdinand Molnár)

秦玉琪 钟耀华 译



化学工业出版社

· 北京

图书在版编目 (CIP) 数据

基因调控机制/(芬)卡尔伯格,(芬)美恩著;秦玉琪,
钟耀华译. —北京:化学工业出版社,2015.8
书名原文: Mechanisms of Gene Regulation
ISBN 978-7-122-24670-7

I. ①基… II. ①卡…②美…③秦…④钟…
III. ①基因表达调控-研究 IV. ①Q786

中国版本图书馆 CIP 数据核字 (2015) 第 166067 号

Translation from English language edition:
Mechanisms of Gene Regulation
Carsten Carlberg, Ferdinand Molnár
Copyright © 2014 Springer Science+Business Media Dordrecht.
All Rights Reserved.

本书中文简体字版由授权化学工业出版社独家出版发行。
未经许可,不得以任何方式复制或抄袭本书的任何部分,违者必究。

北京市版权局著作权合同登记号: 01-2015-4812

责任编辑:傅四周
责任校对:宋 玮

文字编辑:焦欣渝
装帧设计:张 辉

出版发行:化学工业出版社(北京市东城区青年湖南街13号 邮政编码100011)
印 刷:北京永鑫印刷有限责任公司
装 订:三河市宇新装订厂
710mm×1000mm 1/16 印张14 彩插2 字数218千字 2016年1月北京第1版第1次印刷

购书咨询:010-64518888(传真:010-64519686) 售后服务:010-64518899

网 址: <http://www.cip.com.cn>

凡购买本书,如有缺损质量问题,本社销售中心负责调换。

定 价: 75.00 元

版权所有 违者必究

翻译和校对人员名单

秦玉琪	山东大学
钟耀华	山东大学
侯运华	齐鲁工业大学
郑 凯	齐鲁工业大学
李亚楠	山东大学
钱远超	山东大学
张秀君	山东大学
吕珊珊	山东大学
闫孟节	山东大学
张 静	山东大学

译者序

本书是由两位芬兰生物学家卡斯滕·卡尔伯格和费迪南·美恩共同完成的分子生物学专著。该书刚刚面世即受到各国生物研究人员的高度关注，这是因为自20世纪50年代分子生物学开创以来，基因表达调控的研究一直是生物学相关学科的核心内容。更重要的是，该书不仅对经典的转录机器、转录调控、信号转导及染色质修饰等进行了精细深入的阐释，而且着重定位于描述后基因组时代的基因调控。该书在一定程度上代表了当前基因表达调控研究的水平和动向，在国际学术界有较大影响。

后基因组学时代，科学研究快速发展，已经明显改变了生物科学的几乎所有领域，尤其是对产生生命现象（表型）的机制的认识。反观国内分子生物学教材或基因调控相关著作，明显缺乏对最新成果的总结及生物学规律的跟进理解，这使得国内生物研究者（包括研究生）依赖于对英文期刊文献的浏览，容易导致知识的片段化。“它山之石，可以攻玉”，更何况科学没有边界。因此，本书的翻译出版能够为我国生物学领域的研究人员、研究生及高年级本科生提供一本全面的最新的学习基因表达调控的书籍。

就本书译校而言，我们本着实事求是的态度、认真负责的精神，反复校对近十遍，以求如实地体现作者的意图和科学的本质。但由于生物学英文专有名词与中文的差异，以及新的英文专有名词不断出现，书中个别词汇上的翻译可能会有不当，但我们尽力在适当位置（如加括弧）将原来英文名词一并列出，希望读者能够更加容易辨别。此外，个别图的质量达不到满意的程度，这是因为英文原版教材也是如此，但我们尽最大可能在不改变原意的情况下，将不清晰的地方适当描述出来。

为了使本书早日与广大读者见面，化学工业出版社的领导给予了大力支持，相关编辑室和负责出版发行的同志也多方协助，在此一并表示感谢！

由于译校者水平有限，存在译文错误及不妥之处在所难免，敬请读者批评指正。

山东大学生命科学院
国家糖工程技术中心

秦玉琪 钟耀华

2015年5月

前言

本书主要讲述真核生物基因的调控。基因表达塑造了细胞和组织的表型；无论在健康还是疾病研究领域，基因表达的调控几乎是所有生物学过程的基础。因此，不仅生物学家和生物化学家应该深入探索基因调控的原理，生物医药领域的学生也应该了解此研究领域，从而获得良好的专业学科基础。对转录因子及其活性改变过程的完整认知是现代生命科学研究的一个基本目标。

对整个人类基因组序列（以及其他真核生物基因组序列）的利用，以及第二代测序技术的发展，已经明显改变了生物科学的几乎所有领域。例如，由 DNA 元件百科全书计划提供的基因组范围内的组蛋白修饰定位和转录因子结合位点定位的研究，极大地促进了我们对基因调控的理解。因此，本书着重定位于描述后基因组时代的基因调控。

不同于 Ben Lewin 的《基因》一类的经典分子生物学书籍，本书的目的是将当前对基因表达调控机制的认知加以浓缩概述，汇聚精华。鉴于“一图胜千言”，本书中使用了大量的图示。

本书的内容以 2001 年以来一直由 C. Carlberg 教授在东芬兰大学教授的课程“基因表达的调控机制”为基础，分为 4 大部分，13 个章节。在总体介绍之后的三部分，分别从转录因子、染色质和非编码 RNA 的视角对基因调控加以概述。

本书的内容适用于生命科学的硕士研究生，但同样适用于其他生物、医学学科的学生以及博士研究生。本书有以下四点目标，学生应该：

1. 对基因结构、染色质构成、转录因子以及它们的调控机制有深入理解。
2. 识别影响基因表达的主要成分、机制和过程，以及调控基因表达复杂性的多个层次。
3. 有能力分析转录因子、它们的共调控因子以及它们表达相关的非编码 RNA，以此来判断它们在真核生物基因表达调控中的影响。
4. 应用基因调控中的知识来设计、开展和分析实验，比如：定量 PCR、微阵列和染色质免疫共沉淀。

我们希望读者能够喜欢这本书，并对基因表达调控机制的研究产生浓厚的兴趣。

2013 年 8 月于 Kuopio

Carsten Carlberg 和 Ferdinand Molnár

致 谢

感谢 Reinhard Bornemann 教授和 Magdalena Schröder 博士对本书的审阅并提出建设性的意见。

缩略词

1,25(OH) ₂ D ₃	1,25-dihydroxyvitamin D ₃	1,25-二羟维生素 D ₃
3C	chromosome conformation capture	染色体构象捕获
AID	activation-induced cytidine deaminase	激活诱导胞嘧啶核苷脱氨酶
AP1	activator protein 1 (JUN-FOS heterodimer)	激活蛋白 1 (JUN-FOS 异二聚体)
APOL	apolipoprotein	载脂蛋白
AR	androgen receptor	雄激素受体
ATF3	activating transcription factor 3	激活转录因子 3
atRA	all- <i>trans</i> retinoic acid	全反式维甲酸
BAF	BRG1- or hBRM-associated factors	BRG1 或 hBRM 相关因子
bp	base pair	碱基对
BRE	TF II B-binding element	TF II B 结合元件
BrUTP	5-bromouridine 5'-triphosphate	5'三磷酸-5-溴尿核苷
CAR	constitutive androstane receptor	组成型雄甾受体
CREBBP	CREB1-binding protein, also known as CBP	CREB1 结合蛋白, 也称 CBP
CEBP	CCAAT-binding protein	CCAAT 结合蛋白
CHD	chromodomain-helicase-DNA-binding	克罗莫结构域-解旋酶-DNA 结合
ChIP	chromatin immunoprecipitation	染色质免疫沉淀
CREB1	cAMP response element-binding protein	cAMP 应答元件结合蛋白
CTCF	CCCTC-binding factor	CCCTC 结合因子
DBD	DNA-binding domain	DNA 结合结构域
DGCR8	DiGeorge syndrome critical region gene 8	迪格奥尔格综合征决定区域基因 8
DHS	DNAse I hypersensitive site	DNA 酶 I 超敏感位点
DNMT	DNA methyltransferase	DNA 甲基转移酶
DPE	downstream promoter element	下游启动子元件
DR	direct repeat	正向重复
DVL	disheveled	散乱蛋白
EP300	E1A-binding protein p300	E1A 结合蛋白 p300
ENCODE	encyclopedia of DNA elements	DNA 元件百科全书
ER	estrogen receptor	雌激素受体
eRNA	enhancer RNA	增强子 RNA
ES cell	embryonic stem cell	胚胎干细胞
EZH2	enhancer of zeste homolog 2	zeste 基因增强子同源物 2
FAIRE	formaldehyde-assisted isolation of regulatory elements	甲醛辅助的调控元件分离方法
FRAP	fluorescence recovery after photobleaching	光脱色恢复技术
FXR	farnesoid X receptor	类法尼醇 X 受体
GCRP	G-coupled cell surface receptor protein	G 蛋白耦连的细胞表面受体蛋白
GLI	glioma-associated oncogene homolog	神经胶质瘤相关的原癌基因同源基因

GO	gene ontology	基因本体论
GR	glucocorticoid receptor	糖肾上腺皮质激素受体
GSK3	glycogen synthesis kinase 3	糖原合成激酶 3
HAT	histone acetyltransferase	组蛋白乙酰转移酶
HBB	β -globin	β -球蛋白
HDAC	histone deacetylase	组蛋白去乙酰化酶
HDM	histone demethylase	组蛋白去甲基化酶
HMG	high-mobility group protein	高迁移率族蛋白
HMT	histone methyltransferase	组蛋白甲基转移酶
HP1	heterochromatin protein 1	异染色质蛋白 1
HSP	heat-shock protein	热激蛋白
ICR	imprinted control region	印记控制区
NFKBI	inhibitor of NFKB	NFKB 抑制子
IKBK	NFKBI kinase	NFKBI 激酶
IGF2	insulin-like growth factor 2	胰岛素样生长因子 2
IL	interleukin	白细胞介素
INO80	inositol requiring	肌醇依赖
iPS cell	induced pluripotent stem cell	诱导多能干细胞
IRF	interferon-regulatory factor	干扰素调控因子
ISWI	imitation SWI	模拟 SWI
JAK	janus kinase	janus 激酶
JmjC	Jumonji domain-containing	含 Jumonji 结构域
kb	kilo base pairs (1000 bp)	千碱基对 (1000 bp)
LAD	lamina-associated domain	核纤层相关结构域
LBD	ligand-binding domain	配体结合结构域
LCR	locus control region	位点控制区
LINE	long interspersed element	长散在元件
LSD1	lysine specific demethylase 1	赖氨酸特异的去甲基化酶 1
LXR	liver X receptor	肝脏 X 受体
MAP	mitogen-activated protein	丝裂原活化蛋白激酶
MAR	matrix attachment region	核基质结合区
MBD	methyl-DNA-binding domain protein	甲基-DNA 结合域蛋白
MBP	methyl-binding protein	甲基结合蛋白
MDM2	murine double minute-2	鼠双微体-2
MED	mediator	中介复合物
MECP2	methyl-CpG-binding protein 2	甲基-CpG-结合蛋白 2
MLL	mixed lineage leukemia	混合型白血病
mRNA	messenger RNA	信使 RNA
miRNA	micro RNA	小 RNA
ncRNA	non-coding RNA	非编码 RNA
NEMO	NFKB essential modifier	NFKB 必需修饰因子
NFKB	nuclear factor κ B	核因子 κ B
NICD	NOTCH intracellular domain	NOTCH 胞内结构域
nt	nucleotides	核苷酸
PHD	plant homeodomain	同源异型结构域
PKA	cAMP-dependent protein kinase	cAMP 依赖的蛋白激酶
Pol II	RNA polymerase II	RNA 聚合酶 II
PPAR	peroxisome proliferator-activated re- ceptor	过氧化物酶体增生物激活受体

PRC	Polycomb repressive complex	多梳抑制复合物
pre-miRNA	precursor miRNA	前体 miRNA
pri-miRNA	primary miRNA	初始 miRNA
PTCH	patched receptor	补丁受体
PXR	pregnane X receptor	孕烷 X 受体
PWM	position weight matrix	位置权重矩阵
qPCR	quantitative PCR	定量 PCR
RAR	retinoic acid receptor	维甲酸受体
RE	response element	响应元件
RISC	RNA-induced silencing complex	RNA 诱导沉默复合体
RNAi	RNA interference	RNA 干扰
ROR	RAR-related orphan receptor	RAR 相关孤儿受体
rRNA	ribosomal RNA	核糖体 RNA
RSC	remodels the structure of chromatin	染色质结构重塑子
RUNX1	runt-related transcription factor 1	runt-相关的转录因子 1
RXR	retinoid X receptor	维甲酸 X 受体
SINE	short interspersed element	短散在序列
siRNA	small interfering RNA	小干扰 RNA
snRNA	small nuclear RNA	小核 RNA
snoRNA	small nucleolar RNA	核仁小分子 RNA
SP1	specificity protein 1	特征蛋白 1
SREBF1	sterol regulatory element-binding transcription factor 1	甾醇调控元件结合转录因子 1
SRF	serum response factor	血清效应因子
SWI/SNF	switching/sucrose nonfermenting	转换/酵母蔗糖非发酵酶
T ₃	triiodothyronine	三碘甲状腺氨酸
TAF	TBP-associated factor	TBP-相关因子
TATA box	TATAWADR core DNA sequence	TATAWADR 核心 DNA 序列, TATA 框(盒)
TBP	TATA box-binding protein	TATA 框结合蛋白
TET	ten-eleven translocation	10-11 易位
TFF	trefoil factor 1	三叶因子 1
TLF	TBP-like factor	TBP-样因子
TLR	Toll-like receptor	Toll-样受体
TNF	tumor necrosis facto	肿瘤坏死因子
TRBP	transactivation-response RNA-binding protein	反式激活应答 RNA 结合蛋白
TSS	transcription start site	转录起始因子
TR	thyroid hormone receptor	甲状腺激素受体
tRNA	transfer RNA	转运 RNA
TP53	tumor protein p53	肿瘤蛋白 p53
UTR	untranslated region	非翻译区
VDR	vitamin D receptor	维生素 D 受体

目 录

第一篇 概述	1
第 1 章 介绍：什么是基因表达？	3
1.1 分子生物学的中心法则	3
1.2 基因转录的复杂性	7
1.3 转录调控元件	8
1.4 染色质活动在基因调控中的作用	11
1.5 基因表达程序	12
1.6 主要概念	14
拓展阅读	15
第 2 章 染色质的功能	16
2.1 常染色质和异染色质	16
2.2 核小体	19
2.3 染色质开放和封闭时的组蛋白修饰	22
2.4 组蛋白修饰的基因组效应	25
2.5 染色质结构	28
2.6 表观遗传信号对健康和疾病的影响	30
2.7 主要概念	32
拓展阅读	33
第二篇 转录因子篇	35
第 3 章 基础转录机器	37
3.1 核心启动子	37
3.2 TATA 框	42
3.3 核心启动子元件	44
3.4 TFIIID 作为多蛋白复合物的范例	46
3.5 基因组范围内识别核心启动子的方法	47
3.6 中介复合物	49

3.7 主要概念	52
拓展阅读	53
第 4 章 转录因子	54
4.1 位点特异性转录因子	54
4.2 转录因子结构域	55
4.3 二聚体化的转录因子复合物:核受体异源二聚体的例子	59
4.4 生物信息学方法识别转录因子结合位点	61
4.5 转录因子表达谱	64
4.6 转录因子的分类	66
4.7 主要概念	68
拓展阅读	70
第 5 章 信号转导与基因调控	71
5.1 潜在转录因子的激活	72
5.2 转录因子网络	74
5.3 转录因子调控的程序性细胞分化	76
5.4 NF κ B 信号途径	79
5.5 炎症反应中的转录因子	83
5.6 感应细胞压力: p53 途径	85
5.7 主要概念	88
拓展阅读	89
第 6 章 基因表达开关:核受体的例子	90
6.1 核受体超家族	90
6.2 核受体的分子间相互作用	93
6.3 核受体的生理作用	96
6.4 核受体及其配体	98
6.5 核受体和辅因子的相互作用	100
6.6 主要概念	102
拓展阅读	103
第 7 章 基因组作图	104
7.1 进化足迹法	104
7.2 ENCODE 计划	107

7.3	ENCODE 计划的数据挖掘	109
7.4	整合 ChIP-seq 数据	111
7.5	转录因子的其他 DNA 结合模式	113
7.6	转录因子与靶基因的匹配	115
7.7	主要概念	118
	拓展阅读	119
第三篇 染色质篇		121
第 8 章 染色质修饰因子		123
8.1	组蛋白密码模型	124
8.2	组蛋白修饰酶	128
8.3	组蛋白标记的全基因组分析和它们的修饰酶	132
8.4	染色质修饰和疾病	135
8.5	主要概念	137
	拓展阅读	139
第 9 章 基因组印记		140
9.1	绝缘子	140
9.2	基因组调节器 CTCF	142
9.3	基因组印记	144
9.4	绝缘子功能模型	147
9.5	CTCF 介导的染色质结构的遗传	148
9.6	主要概念	149
	拓展阅读	151
第 10 章 表观基因组		152
10.1	DNA 甲基化	152
10.2	基因组范围的 DNA 甲基化	154
10.3	DNA 甲基化、异染色质和基因沉默	156
10.4	健康与疾病中的表观基因组学	160
10.5	主要概念	163
	拓展阅读	164
第 11 章 染色质重塑		165
11.1	染色质重塑的作用	165

11.2	ATP-依赖的重塑复合物	167
11.3	启动子上的核小体定位	170
11.4	先锋因子	173
11.5	转录动力学和噪声	175
11.6	主要概念	177
	拓展阅读	178
第 12 章	染色质结构	180
12.1	细胞核的组织	180
12.2	高级 DNA 环结构的形成	182
12.3	细胞核内的分区	185
12.4	转录工厂	186
12.5	核定位与疾病	188
12.6	主要概念	188
	拓展阅读	189
第四篇	非编码 RNA 篇	191
第 13 章	调控 RNA	193
13.1	非编码 RNAs	193
13.2	miRNA 及其功能	195
13.3	长链非编码 RNA	199
13.4	eRNA	201
13.5	miRNA 和转录因子介导的基因调控	203
13.6	主要概念	204
	拓展阅读	206

第一篇 概述

第1章 介绍：什么是基因表达？

摘要：人类基因组（参见专栏 1.1）是由大约 20000 个蛋白编码基因和大约相同数量的非编码 RNA 基因（ncRNA）组成的，这些非编码 RNA 被用作核糖体的结构骨架（核糖体 RNA，rRNA），或蛋白质翻译中的衔接分子（转运 RNA，tRNA），或作为 mRNA 稳定性、蛋白质翻译和染色质密度的调控因子（长链非编码 RNA 和 miRNA）（详见 13.2 节）。当一个基因被转录为 RNA 时，我们称之为“表达”。在一个特定的组织或细胞类型中，只有大约一半的基因被转录。此外，每个组织的表型都是由具有自身特色的基因转录模式确定的。当细胞暴露于不同信号时，如处于不同营养或环境压力时，其表型会随之改变。

组织和信号特异性的基因表达是控制细胞一般属性并对环境扰动作出应答的重要机制。后者在健康和疾病情况下都适用。因此，对基因表达的详细了解不仅有助于理解疾病状态下的调控异常，而且对疾病的治疗也有所帮助。

在本章中，我们将概述基因表达的核心特征。我们将会介绍到，基因表达是受到严格控制的，特别是在基因转录水平。基因表达至少在三个水平上受到控制：DNA 编码、表观遗传密码和转录因子程序。

关键词：蛋白编码基因，转录，基因组，RNA，聚合酶，定量 PCR，微阵列，染色质，核小体，增强子，启动子，基因座控制区，表观遗传学，转录起始位点，基因表达

1.1 分子生物学的中心法则

“分子生物学的中心法则”显示了从 DNA 到 RNA 再到蛋白质的遗传信息的一个明确方向（图 1.1）。这意味着，除了少数例外，如反转录病毒的 RNA 基因组的反转录，基因组 DNA 储藏着所有原核和真核生物的遗传信息。因此，基因被定义为可以转录成 RNA 的基因组 DNA 区域。在这个传统观点中，RNA 指的是 mRNA，即用于蛋白质翻译的模板。然而，现在已知有一些其他