



Precision  
Medicine

精准医学出版工程

精准医学基础系列

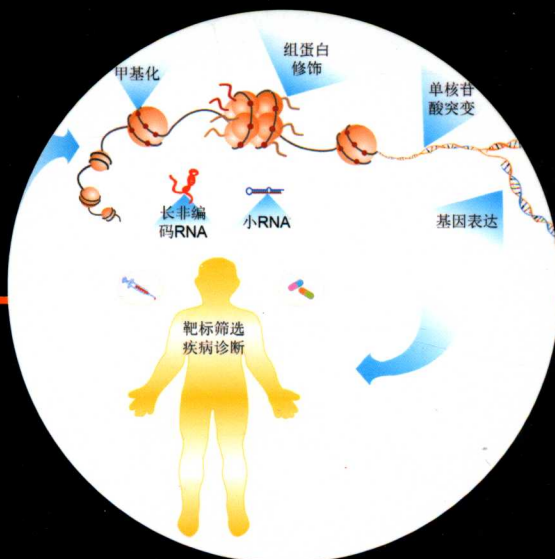
总主编 詹启敏

“十三五”国家重点图书出版规划项目

# 转录组学与精准医学

## Transcriptomics and Precision Medicine

方向东 胡松年 等  
编著



上海交通大学出版社  
SHANGHAI JIAO TONG UNIVERSITY PRESS



国家出版基金项目  
NATIONAL PUBLICATION FOUNDATION

“十三五”国家重点图书出版规划项目

Precision  
Medicine

精准医学出版工程

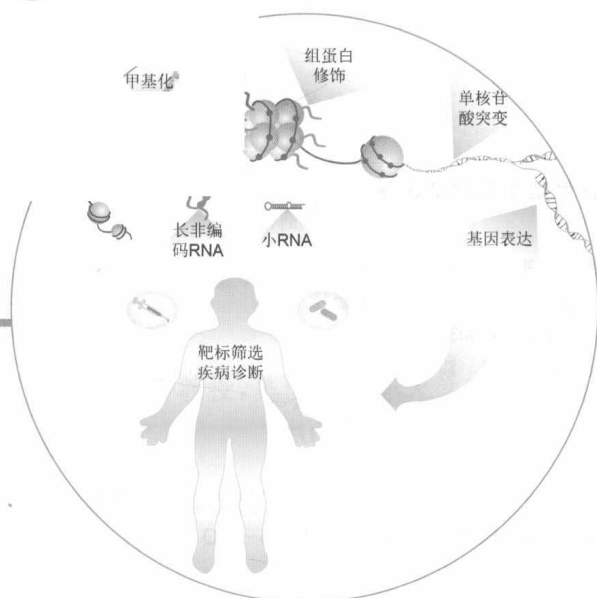
精准医学基础系列

总主编 詹启敏

# 转录组学与精准医学

## Transcriptomics and Precision Medicine

方向东 胡松年 等  
编著



上海交通大学出版社  
SHANGHAI JIAO TONG UNIVERSITY PRESS

## 内容提要

转录组是生命组学研究中重要的组成部分,狭义转录组是指一种细胞中可直接参与翻译蛋白质的 mRNA 的总和,广义转录组是指特定细胞中所有转录本的总和,即所有按基因信息单元转录和加工的 RNA 分子。转录组学是后基因组时代生命科学领域的研究热点,也是快速发展并不断完善中的一门新兴学科,在精准医学研究和个体化医疗实践中占有非常重要的地位。本书首先介绍了转录组学的基本理论、研究方法和研究内容;其次分别阐述了转录组学在出生缺陷与生殖健康、恶性肿瘤、心脑血管疾病、血液系统疾病、免疫和代谢系统疾病、神经系统疾病以及药物研发与临床用药指导中的应用;再次介绍了重要模式动物的转录组研究及其在精准医学领域的转录组学应用;最后系统整理了转录组学数据资源、研究工具、知识库和转录组学相关的疾病数据库等。

本书希望能为正在或者将要从事转录组学和精准医学工作的科研人员、研究生和高年级本科生以及有兴趣了解这些研究内容的其他人员提供参考。

## 图书在版编目(CIP)数据

转录组学与精准医学/方向东等编著. —上海:上海交通大学出版社,2017  
精准医学出版工程  
ISBN 978-7-313-18175-6

I. ①转… II. ①方… III. ①基因转录—应用—医学 IV. ①Q753②R

中国版本图书馆 CIP 数据核字(2017)第 235939 号

## 转录组学与精准医学

编 著: 方向东 胡松年等  
出版发行: 上海交通大学出版社  
邮政编码: 200030  
出 版 人: 谈 毅  
印 制: 苏州市越洋印刷有限公司  
开 本: 787mm×1092mm 1/16  
字 数: 428 千字  
版 次: 2017 年 12 月第 1 版  
书 号: ISBN 978-7-313-18175-6/Q  
定 价: 268.00 元

地 址: 上海市番禺路 951 号  
电 话: 021-64071208  
经 销: 全国新华书店  
印 张: 26.5  
印 次: 2017 年 12 月第 1 次印刷

版权所有 侵权必究

告读者: 如发现本书有印装质量问题请与印刷厂质量科联系  
联系电话: 0512-68180638

## 编 委 会

### 总主编

詹启敏(北京大学副校长、医学部主任,中国工程院院士)

### 编 委

(按姓氏拼音排序)

- 陈 超(西北大学副校长、国家微检测系统工程技术研究中心主任,教授)  
方向东(中国科学院基因组科学与信息重点实验室副主任、中国科学院北京基因组研究所“百人计划”研究员,中国科学院大学教授)  
郜恒骏(生物芯片上海国家工程研究中心主任,同济大学医学院教授、消化疾病研究所所长)  
贾 伟(美国夏威夷大学癌症研究中心副主任,教授)  
钱小红(军事科学院军事医学研究院生命组学研究所研究员)  
石乐明(复旦大学生命科学学院、复旦大学附属肿瘤医院教授)  
王晓民(首都医科大学副校长,北京脑重大疾病研究院院长,教授)  
于 军(中国科学院基因组科学与信息重点实验室、中国科学院北京基因组研究所研究员,中国科学院大学教授)  
赵立平(上海交通大学生命科学技术学院特聘教授,美国罗格斯大学环境与生物科学学院冠名讲席教授)  
朱景德(安徽省肿瘤医院肿瘤表观遗传学实验室教授)

### 学术秘书

张 华(中国医学科学院、北京协和医学院科技管理处副处长)

# 《转录组学与精准医学》

## 编 委 会

### 名誉主编

陈润生(中国科学院生物物理研究所研究员,中国科学院院士)

### 主 编

方向东(中国科学院基因组科学与信息重点实验室副主任、北京基因组研究所“百人计划”研究员,中国科学院大学教授)

胡松年(中国科学院基因组科学与信息重点实验室主任、研究员,中国科学院大学教授)

### 副主编

赵方庆(中国科学院北京生命科学研究院研究员)

雷红星(中国科学院北京基因组研究所研究员,中国科学院大学教授)

李全贞(中南大学湘雅医院教授)

### 编 委

(按姓氏拼音排序)

陈敬洲(中国医学科学院阜外医院研究员)

崔庆华(北京大学医学部医学信息学系研究员)

杜政霖(中国科学院北京基因组研究所高级工程师)

何顺民(中国科学院生物物理研究所副研究员)

和夫红(中国科学院北京基因组研究所副研究员)

黄向阳(中南大学湘雅二医院风湿免疫科主任,教授、主任医师)

李川昀(北京大学分子医学研究所研究员)

李 默(北京大学第三医院研究员)

李 伟(温州医科大学检验医学院、生命科学学院副教授)

林 强(中国科学院北京基因组研究所副研究员)

刘红星(河北燕达陆道培医院病理和检验医学科主任,副主任医师)

乔 杰(北京大学第三医院院长,中国工程院院士)

渠鸿竹(中国科学院北京基因组研究所副研究员,中国科学院大学副教授)

石桂秀(厦门大学附属第一医院教授、主任医师)

石乐明(复旦大学生命科学学院、复旦大学附属肿瘤医院教授)

王久存(复旦大学生命科学学院教授)

王前飞(中国科学院北京基因组研究所研究员)

伊成器(北京大学生命科学学院教授)

于 军(中国科学院北京基因组研究所研究员)

郁 颖(复旦大学生命科学学院副研究员)

张 伟(中南大学临床药理研究所、中南大学湘雅医学检验所教授)

张昭军(中国科学院北京基因组研究所副研究员,中国科学院大学副教授)

张治华(中国科学院北京基因组研究所研究员)

赵洪军(中南大学湘雅医院副主任医师)

赵文明(中国科学院北京基因组研究所生命与健康大数据中心副主任,高级工程师)

赵 屹(中国科学院计算技术研究所副研究员)

左晓霞(中南大学湘雅医院风湿免疫科主任,教授、主任医师)

## 主编简介



方向东，1969年出生。第一军医大学临床医学本科、免疫学硕士、人体解剖学和组织胚胎学博士，现任中国科学院基因组科学与信息重点实验室副主任、中国科学院北京基因组研究所“百人计划”研究员，中国科学院大学教授。从事医学遗传学、基因组学和转化医学研究20余年，带领研究组整合分析多组学数据，筛选遗传性血液病、肿瘤等疾病的诊疗靶点，为临床制订更加高效、特异的干预策略提供理论基础和实验依据。曾为美国华盛顿大学医学院医学遗传系助理教授。作为国家“十

三五”重点研发计划“精准医学研究”重点专项“精准医学大数据处理和利用的标准化技术体系建设”项目的首席科学家，还担任中国转化医学联盟理事、国家科学技术部“人类遗传资源管理”专家、国家卫生计生委临床遗传专科委员会和中国遗传学会遗传咨询分会专家委员；人民卫生出版社系列期刊管理委员会和中国医药生物技术协会生物医学信息技术专业委员会常务委员；中国卫生信息学会健康医疗大数据政府决策支持与标准化、中国医药生物技术协会生物诊断技术、中国生物工程学会计算生物学与生物信息学、中国生理学会血液生理等专业委员会委员。同时担任《发育医学电子杂志》主编、*Genomics Proteomics Bioinformatics* 副主编、《遗传》杂志编委。发表SCI收录论文60余篇；申请国家发明专利10项，授权5项；获得新药证书1项。

## 主编简介



胡松年，1969 年出生。中国农业大学理学博士，中国科学院基因组科学与信息重点实验室主任、中国科学院北京基因组研究所研究员，中国科学院大学教授。主要研究方向为基因组结构与功能解析，包括复杂基因组的从头测序、基因组的精细注释、基于组学大数据的知识整合型数据库开发。作为主要负责人之一先后参与了“人类基因组 1% 计划”“家猪基因组计划”和“水稻基因组计划”等大型高等动植物的基因组测序和分析工作。承担了多项国家“863”、“973”、国家自然科学基金等项目的研究工作。作为第一作者或通讯作者，已先后在 *Science*、*Nature Communications*、*Nucleic Acids Research*、*New Phytologist* 等国际知名杂志上发表多篇高质量的论文。作为第一作者在 *Science* 发表的水稻基因组文章，已被引用 2 600 余次，因此获得“2002 年度求是杰出科技成就集体奖”和“2003 年度中国科学院杰出科技成就集体奖”。



“精准”是医学发展的客观追求和最终目标，也是公众对健康的必然需求。“精准医学”是生物技术、信息技术和多种前沿技术在医学临床实践的交汇融合应用，是医学科技发展的前沿方向，实施精准医学已经成为推动全民健康的国家发展战略。因此，发展精准医学，系统加强精准医学研究布局，对于我国重大疾病防控和促进全民健康，对于我国占据未来医学制高点及相关产业发展主导权，对于推动我国生命健康产业发展具有重要意义。

2015年初，我国开始制定“精准医学”发展战略规划，并安排中央财政经费给予专项支持，这为我国加入全球医学发展浪潮、增强我国在医学前沿领域的研究实力、提升国家竞争力提供了巨大的驱动力。国家科技部在国家“十三五”规划期间启动了“精准医学研究”重点研发专项，以我国常见高发、危害重大的疾病及若干流行率相对较高的罕见病为切入点，将建立多层次精准医学知识库体系和生物医学大数据共享平台，形成重大疾病的风险评估、预测预警、早期筛查、分型分类、个体化治疗、疗效和安全性预测及监控等精准防治方案和临床决策系统，建设中国人群典型疾病精准医学临床方案的示范、应用和推广体系等。目前，“精准医学”已呈现快速和健康发展态势，极大地推动了我国卫生健康事业的发展。

精准医学几乎覆盖了所有医学门类，是一个复杂和综合的科技创新系统。为了迎接新形势下医学理论、技术和临床等方面的需求和挑战，迫切需要及时总结精准医学前沿研究成果，编著一套以“精准医学”为主题的丛书，从而助力我国精准医学的进程，带动医学科学整体发展，并能加快相关学科紧缺人才的培养和健康大产业的发展。

2015年6月，上海交通大学出版社以此为契机，启动了“精准医学出版工程”系列图

书项目。这套丛书紧扣国家健康事业发展战略,配合精准医学快速发展的态势,拟出版一系列精准医学前沿领域的学术专著,这是一项非常适合国家精准医学发展时宜的事业。我本人作为精准医学国家规划制定的参与者,见证了我国“精准医学”的规划和发展,欣然接受上海交通大学出版社的邀请担任该丛书的总主编,希望为我国的“精准医学”发展及医学发展出一份力。出版社同时也邀请了刘彤华院士、贺福初院士、刘昌效院士、周宏灏院士、赵国屏院士、王红阳院士、曹雪涛院士、陈志南院士、陈润生院士、陈香美院士、金力院士、周琪院士、徐国良院士、董家鸿院士、卞修武院士、陆林院士、乔杰院士、黄荷凤院士等医学领域专家撰写专著、承担审校等工作,邀请的编委和撰写专家均为活跃在精准医学研究最前沿的、在各自领域有突出贡献的科学家、临床专家、生物信息学家,以确保这套“精准医学出版工程”丛书具有高品质和重大的社会价值,为我国的精准医学发展提供参考和智力支持。

编著这套丛书,一是总结整理国内外精准医学的重要成果及宝贵经验;二是更新医学知识体系,为精准医学科研与临床人员培养提供一套系统、全面的参考书,满足人才培养对教材的迫切需求;三是为精准医学实施提供有力的理论和技术支撑;四是将许多专家、教授、学者广博的学识见解和丰富的实践经验总结传承下来,旨在从系统性、完整性和实用性角度出发,把丰富的实践经验和实验室研究进一步理论化、科学化,形成具有我国特色的“精准医学”理论与实践相结合的知识体系。

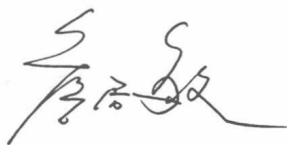
“精准医学出版工程”是国内外第一套系统总结精准医学前沿性研究成果的系列专著,内容包括“精准医学基础”“精准预防”“精准诊断”“精准治疗”“精准医学药物研发”以及“精准医学的疾病诊疗共识、标准与指南”等多个系列,旨在服务于全生命周期、全人群、健康全过程的国家大健康战略。

预计这套丛书的总规模会达到 60 种以上。随着学科的发展,数量还会有所增加。这套丛书首先包括“精准医学基础系列”的 11 种图书,其中 1 种为总论。从精准医学覆盖的医学全过程链条考虑,这套丛书还将包括和预防医学、临床诊断(如分子诊断、分子影像、分子病理等)及治疗相关(如细胞治疗、生物治疗、靶向治疗、机器人、手术导航、内镜等)的内容,以及一些通过精准医学现代手段对传统治疗优化后的精准治疗。此外,这套丛书还包括药物研发,临床诊疗路径、标准、规范、指南等内容。“精准医学出版工程”将紧密结合国家“十三五”重大战略规划,聚焦“精准医学”目标,贯穿“十三五”始终,力求打造一个总体量超过 60 本的学术著作群,从而形成一个医学学术出版的高峰。

本套丛书得到国家出版基金资助,并入选了“十三五”国家重点图书出版规划项目,体现了国家对“精准医学”项目以及“精准医学出版工程”这套丛书的高度重视。这套丛书承担着记载与弘扬科技成就、积累和传播科技知识的使命,凝结了国内外精准医学领域专业人士的智慧和成果,具有较强的系统性、完整性、实用性和前瞻性,既可作为实际工作的指导用书,也可作为相关专业人员的学习参考用书。期望这套丛书能够有益于精准医学领域人才的培养,有益于精准医学的发展,有益于医学的发展。

此次集束出版的“精准医学基础系列”系统总结了我国精准医学基础研究各领域取得的前沿成果和突破,内容涵盖精准医学总论、生物样本库、基因组学、转录组学、蛋白质组学、表观遗传学、微生物组学、代谢组学、生物大数据、新技术等新兴领域和新兴学科,旨在为我国精准医学的发展和实施提供理论和科学依据,为培养和建设我国高水平的具有精准医学专业知识和先进理念的基础和临床人才队伍提供理论支撑。

希望这套丛书能在国家医学发展史上留下浓重的一笔!



北京大学副校长  
北京大学医学部主任  
中国工程院院士  
2017年11月16日

# 序

近年来,精准医学研究发展迅速,但迄今,还没有系统总结精准医学基础研究各领域相关成果的系列图书。因此,上海交通大学出版社推出“精准医学出版工程·精准医学基础系列”图书,对于精准医学基础领域的研究和应用,对于精准医学的发展具有重要意义。

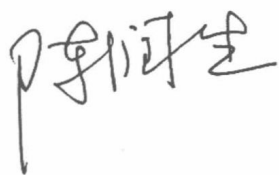
在“精准医学出版工程”中由方向东研究员、胡松年研究员主编的《转录组学与精准医学》一书,内容丰富、实用、严谨,视角新颖。转录组学是后基因组时代生命科学领域的研究热点,也是快速发展并不断完善中的一门新兴学科,在精准医学研究和个体化医疗实践中占有非常重要的地位。狭义转录组是指一种细胞中可直接参与翻译蛋白质的 mRNA 的总和;而广义转录组则是指特定细胞中所有转录本的总和,即所有按基因信息单元转录和加工的 RNA 分子,包括 miRNA、lncRNA 等具有重要调控功能的非编码 RNA。

《转录组学与精准医学》包含 4 个主要模块:①理论部分;②转录组学在精准医学中的应用;③重要模式动物的转录组学研究;④转录组学数据资源、研究工具、知识库和转录组学相关的疾病数据库等。各个模块主题明确,模块内相互补充,模块间逻辑自然,体系科学严谨,具有系统性、完整性、实用性和前瞻性的特点。

《转录组学与精准医学》一书的主编方向东和胡松年研究员,来自中国科学院北京基因组研究所。作为国内最早开展大规模基因组学研究的国立机构的优秀科研骨干,曾经参加过人类基因组、HapMap、ENCODE、家猪基因组和水稻基因组等国际组学计划,拥有丰富的多组学数据整合分析的工作经验。目前正在主持和参与国家“十三五”规划“精准医学研究”重点专项项目、中国科学院“中国人群精准医学研究计划”重点项

目和国家“863”计划“组学大数据中心和知识库构建与服务技术”项目等重要研究计划。其他作者,无论来自高校、科研院所,还是来自生物科技公司,都是该领域非常出色的,能够代表这一领域国内的最高水平。

鉴于此,我认为《转录组学与精准医学》一书与国家“十三五”规划紧密契合,体现了国内外医学领域的最新动向和发展趋势,对于精准医学研究领域人员,特别是从事精准医学基础领域研究的专家学者具有重要的学术参考价值。



中国科学院生物物理研究所

中国科学院院士

2017年11月于北京

# 前言

《转录组学与精准医学》是“精准医学出版工程·精准医学基础系列”中的一个分册,涵盖了转录组学的基本概念、研究内容及其在医学领域的主要应用,同时还简要介绍了模式动物转录组学的研究进展,并提供了转录组学应用于精准医学的相关数据和软件资源。希望通过本书,向科研工作者展示转录组学应用于精准医学的重要性和迫切性;向医疗工作者传播转录组学在精准医疗中的应用模式和诊疗价值,努力推动转录组学研究成果及时纳入精准医学的评价和治疗体系,为临床提供个体化诊断及精准药物治疗的有力工具,通过“交叉融合、协同创新”,共同促进中国精准医学研究的协调发展。

本书分为四个部分,共计 10 章及附录。第一部分为理论部分,包括转录组学基础和转录组学的研究方法 2 章,主要介绍转录组学的基本理论、研究方法和研究内容。第二部分为转录组学在精准医学中的应用,共计 7 章,按照临床常见疾病及应用,分别阐述转录组学在出生缺陷疾病和生殖健康、恶性肿瘤、心血管疾病、血液疾病、自身免疫病和代谢性疾病、神经精神类疾病及药物研发与临床用药指导中的应用。第三部分为重要模式动物的转录组研究,是第 10 章综述了小鼠、大鼠、斑马鱼、恒河猴、小型猪等模式动物在精准医学领域的转录组学应用。第四部分为附录,系统整理了转录组学数据资源、研究工具、知识库和转录组学相关的疾病数据库等。

转录组学成果应用于精准医学是崭新的研究领域,本书的写作既是探索,也是尝试。在创新性基础上,我们努力强调实用性、引导性,突出理论与实践相结合等特点,希望能为正在或者将要从事转录组学和精准医学工作的科研人员、研究生和高年级本科生以及有兴趣了解这些研究内容的其他人员提供参考。

本书由中国科学院北京基因组研究所方向东、胡松年研究员主持编著,编写工作得到诸多科研院所、高等院校和医疗机构的大力支持和帮助。衷心感谢我国理论生物学和生物信息学研究的拓荒者——陈润生院士为本书作序!转录组学概念、原理和技术部分内容主要由中国科学院北京基因组研究所、生物物理研究所、运城学院、计算技术研究所和北京生命科学研究院等科研单位的专家执笔。转录组学的医学应用部分内容主要由北京大学和北京大学第三医院、中国科学院北京基因组研究所、中国医学科学院阜外医院、首都医科大学附属潞河医院、河北燕达陆道培医院、中南大学湘雅医院、复旦大学、四川大学华西医院、厦门大学附属第一医院、大连医科大学、温州医科大学、全国科学技术名词审定委员会事务中心等单位的专家完成。其中第1章由胡松年、梁浩、刘万飞执笔,第2章由卜德超、崔东亚、方双桑、高远、何顺民、胡松年、李溪远、林强、刘万飞、孙亮、王佳佳、赵方庆、赵屹执笔,第3章由白青云、方向东、贺宁、李默、李永君、乔杰、任云晓、施剑、杨亚东、伊成器、张娜、赵惠聆执笔,第4章由丁楠、董绪浓、方向东、李永君、渠鸿竹、肖茹丹、谢兵兵、杨琼、杨亚东、张璐、张倩、赵学彤执笔,第5章由陈晨、陈敬洲、崔庆华、方向东、郭思思、刘天龙、马伟、王晨、谢兵兵执笔,第6章由陈雪、方向东、李艳明、刘红星、施剑、王晨、杨亚东、张璐、张阳执笔,第7章由陈偲均、丁伟峰、郭慕瑶、黄向阳、姜帅、李全贞、李伟、李媛、刘晶、刘庆梅、牛振民、濮伟霖、渠鸿竹、石桂秀、石祥广、王久存、宣景秀、张鸽、张辉、张治华、赵洪军、郑佳文、左晓霞执笔,第8章由雷红星执笔,邢培琪绘制图7-1,第9章由方向东、李品、王海、王幼鸿、杨琼、张伟、张昭军执笔,第10章由方向东、和夫红、李川昀、申晴、石乐明、宋利璞、王海、王恒涛、王前飞、徐嘉悦、于军、郁颖、张璐、张昭军执笔,附录由陈晨、杜政霖、贺宁、杨亚东、赵文明执笔。杨亚东对本书文字进行了校对工作。在此衷心感谢各位专家的辛苦和付出!

本书引用了一些作者的论著及其研究成果,在此表示衷心的感谢!

转录组学与精准医学的发展都极为迅速,目前涉及内容已极为宽泛。作为精准医学基础系列丛书,本书虽试图在写作内容上尽可能涵盖相关领域的各个方面,但是也难以管中窥豹、挂一漏万。书中如有错谬之处,还期待得到广大读者的反馈信息,以使本书臻于完善。

编著者

2017年11月于北京

# 目 录

<b>1</b>	<b>转录组学基础</b> .....	001
1.1	转录组学的概念 .....	001
1.2	转录组学研究的发展历史 .....	002
1.3	转录组学研究的实验技术方法 .....	005
1.3.1	基于芯片的实验技术方法 .....	005
1.3.2	基于测序的实验技术方法 .....	007
1.4	不同平台转录组学研究方法比较 .....	016
	参考文献 .....	017
<b>2</b>	<b>转录组学的研究方法</b> .....	021
2.1	转录组学数据质量控制 .....	021
2.2	mRNA 表达量研究 .....	024
2.3	mRNA 差异表达分析 .....	026
2.3.1	mRNA 差异表达分析概述 .....	026
2.3.2	多细胞 RNA-Seq 差异表达分析 .....	027
2.3.3	单细胞 RNA-Seq 差异表达分析 .....	028
2.4	mRNA 可变剪接 .....	033
2.5	基因融合鉴定 .....	036



2.6	RNA 编辑鉴定 .....	039
2.6.1	RNA 编辑概述 .....	039
2.6.2	RNA 编辑与疾病的关系 .....	040
2.6.3	RNA 编辑的鉴定 .....	041
2.7	环形 RNA 鉴定及其功能预测 .....	043
2.7.1	环形 RNA 的分类 .....	044
2.7.2	环形 RNA 的生化特性 .....	046
2.7.3	环形 RNA 的早期研究 .....	048
2.7.4	环形 RNA 的生成机制 .....	049
2.7.5	环形 RNA 的功能 .....	051
2.7.6	环形 RNA 的识别方法 .....	054
2.8	piRNA 鉴定及其功能预测 .....	057
2.8.1	piRNA 的早期研究 .....	057
2.8.2	转座子相关的 piRNA .....	059
2.8.3	mRNA 相关的 piRNA .....	061
2.8.4	Piwi 蛋白参与癌细胞的发生 .....	062
2.8.5	癌症相关的 piRNA .....	064
2.8.6	piRNA 的研究方法 .....	068
2.8.7	piRNA 功能研究的现状与展望 .....	070
2.9	lncRNA 的功能及其鉴定 .....	070
2.9.1	lncRNA 的分子作用模式 .....	071
2.9.2	lncRNA 的功能机制 .....	072
2.9.3	lncRNA 与疾病发病的关系 .....	073
2.9.4	lncRNA 的生物信息学鉴定 .....	074
2.10	3' poly(A) 长度变化分析 .....	075
	参考文献 .....	078