



“十二五”普通高等教育本科国家级规划教材



国家卫生和计划生育委员会“十二五”规划教材  
全国高等医药教材建设研究会“十二五”规划教材

全 国 高 等 学 校 教 材

供8年制及7年制（“5+3”一体化）临床医学等专业用

# 生物信息学

## Bioinformatics

第2版

主 编 李 霞 雷健波  
副主编 李亦学 李劲松

Medical science  
foundation

Medical professional  
attitude, behavior and ethics

Clinical skills

# MEDICAL ELITE EDUCATION

Information management capacity

Critical thinking

Group health and health system

Communication skills



人民卫生出版社  
PEOPLE'S MEDICAL PUBLISHING HOUSE





“十二五”普通高等教育本科国家级规划教材



国家卫生和计划生育委员会“十二五”规划教材  
全国高等医药教材建设研究会“十二五”规划教材  
全国高等学校教材

供8年制及7年制(“5+3”一体化)临床医学等专业用

# 生物信息学

Bioinformatics

第2版

主 编 李 霞 雷健波

副 主 编 李亦学 李劲松

编 者 (以姓氏笔画排序)

王 举 (天津医科大学)

邹凌云 (第三军医大学)

朱 浩 (南方医科大学)

沈百荣 (苏州大学)

许丽艳 (汕头大学)

张 岩 (哈尔滨医科大学)

李 瑛 (吉林大学)

陈小平 (中南大学)

李 霞 (哈尔滨医科大学)

赵雨杰 (中国医科大学)

李冬果 (首都医科大学)

徐良德 (哈尔滨医科大学)

李亦学 (同济大学)

崔庆华 (北京大学)

李劲松 (浙江大学)

雷健波 (北京大学)

吴忠道 (中山大学)

魏冬青 (上海交通大学)

学术秘书

王 宏 (哈尔滨医科大学)

人民卫生出版社

图书在版编目(CIP)数据

生物信息学 / 李霞, 雷健波主编. —2 版. —北京: 人民卫生出版社, 2015

ISBN 978-7-117-20453-8

I. ①生… II. ①李…②雷… III. ①生物信息论—高等学校—教材 IV. ①Q811.4

中国版本图书馆 CIP 数据核字(2015)第 051271 号

|       |  |                                 |
|-------|--|---------------------------------|
| 人卫社官网 | <a href="http://www.pmph.com">www.pmph.com</a>   | 出版物查询, 在线购书                     |
| 人卫医学网 | <a href="http://www.ipmph.com">www.ipmph.com</a> | 医学考试辅导, 医学数据库服务, 医学教育资源, 大众健康资讯 |

版权所有, 侵权必究!

生物信息学

第 2 版

主 编: 李 霞 雷健波

出版发行: 人民卫生出版社(中继线 010-59780011)

地 址: 北京市朝阳区潘家园南里 19 号

邮 编: 100021

E - mail: [pmph@pmph.com](mailto:pmph@pmph.com)

购书热线: 010-59787592 010-59787584 010-65264830

印 刷: 北京铭成印刷有限公司

经 销: 新华书店

开 本: 850×1168 1/16 印张: 33

字 数: 908 千字

版 次: 2010 年 8 月第 1 版 2015 年 6 月第 2 版  
2015 年 6 月第 2 版第 1 次印刷(总第 2 次印刷)

标准书号: ISBN 978-7-117-20453-8/R·20454

定 价: 108.00 元

打击盗版举报电话: 010-59787491 E-mail: [WQ@pmph.com](mailto:WQ@pmph.com)

(凡属印装质量问题请与本社市场营销中心联系退换)

为了贯彻教育部教高函[2004-9号]文,在教育部、原卫生部的领导和支持下,在吴阶平、裘法祖、吴孟超、陈灏珠、刘德培等院士和知名专家的亲切关怀下,全国高等医药教材建设研究会以原有七年制教材为基础,组织编写了八年制临床医学规划教材。从第一轮的出版到第三轮的付梓,该套教材已经走过了十余个春秋。

在前两轮的编写过程中,数千名专家的笔耕不辍,使得这套教材成为了国内医药教材建设的一面旗帜,并得到了行业主管部门的认可(参与申报的教材全部被评选为“十二五”国家级规划教材),读者和社会的推崇(被视为实践的权威指南、司法的有效依据)。为了进一步适应我国卫生计生体制改革和医学教育全方位深入推进,以及医学科学不断发展的需要,全国高等医药教材建设研究会在深入调研、广泛论证的基础上,于2014年全面启动了第三轮的修订改版工作。

本次修订始终不渝地坚持了“精品战略,质量第一”的编写宗旨。以继承与发展为指导思想:对于主干教材,从精英教育的特点、医学模式的转变、信息社会的发展、国内外教材的对比等角度出发,在注重“三基”、“五性”的基础上,在内容、形式、装帧设计等方面力求“更新、更深、更精”,即在前一版的基础上进一步“优化”。同时,围绕主干教材加强了“立体化”建设,即在主干教材的基础上,配套编写了“学习指导及习题集”、“实验指导/实习指导”,以及数字化、富媒体的在线增值服务(如多媒体课件、在线课程)。另外,经专家提议,教材编写委员会讨论通过,本次修订新增了《皮肤性病学》。

本次修订一如既往地得到了广大医药院校的大力支持,国内所有开办临床医学专业八年制及七年制(“5+3”一体化)的院校都推荐出了本单位具有丰富临床、教学、科研和写作经验的优秀专家。最终参与修订的编写队伍很好地体现了权威性、代表性和广泛性。

修订后的第三轮教材仍以全国高等学校临床医学专业八年制及七年制(“5+3”一体化)师生为主要目标读者,并可作为研究生、住院医师等相关人员的参考用书。

全套教材共38种,将于2015年7月前全部出版。



# 全国高等学校八年制临床医学专业国家卫生和计划生育委员会 规划教材编写委员会

## 名誉顾问

韩启德 桑国卫 陈 竺 吴孟超 陈灏珠

## 顾问(按姓氏笔画排序)

马建辉 王 辰 冯友梅 冯晓源 吕兆丰 闫剑群 李 虹

李立明 李兰娟 杨宝峰 步 宏 汪建平 张 运 张灼华

陈国强 赵 群 赵玉沛 郝希山 柯 杨 桂永浩 曹雪涛

詹启敏 赫 捷 魏于全

## 主任委员

刘德培

## 委 员(按姓氏笔画排序)

丁文龙 于双成 万学红 马 丁 马 辛 丰有吉 王 杉

王兰兰 王宁利 王吉耀 王宇明 王怀经 王明旭 王建安

王建枝 王庭槐 王海杰 王家良 王鸿利 尹 梅 孔维佳

左 伋 冯作化 刘艳平 江开达 安 锐 许能锋 孙志伟

孙贵范 李 和 李 霞 李甘地 李明远 李桂源 李凌江

李继承 杨 恬 杨世杰 吴 江 吴忠道 何 维 应大君

沈 铿 张永学 张丽霞 张建中 张绍祥 张雅芳 陆 林

陈 红 陈 杰 陈孝平 陈建国 欧阳钦 尚 红 罗爱静

金征宇 周 桥 周 梁 赵旭东 药立波 柏树令 姜乾金

洪秀华 姚 泰 秦 川 贾文祥 贾弘禔 贾建平 钱睿哲

徐志凯 徐勇勇 凌文华 高兴华 高英茂 诸欣平 黄 钢

龚启勇 康德英 葛 坚 雷健波 詹希美 詹思延 廖二元

颜 虹 薛辛东 魏 泓

## 教材目录

|    | 学科名称            | 主审             | 主编      | 副主编              |
|----|-----------------|----------------|---------|------------------|
| 1  | 细胞生物学(第3版)      | 杨恬             | 左 伋 刘艳平 | 刘 佳 周天华 陈誉华      |
| 2  | 系统解剖学(第3版)      | 柏树令 应大君        | 丁文龙 王海杰 | 崔慧先 孙晋浩 黄文华 欧阳宏伟 |
| 3  | 局部解剖学(第3版)      | 王怀经            | 张绍祥 张雅芳 | 刘树伟 刘仁刚 徐 飞      |
| 4  | 组织学与胚胎学(第3版)    | 高英茂            | 李 和 李继承 | 曾园山 周作民 肖 岚      |
| 5  | 生物化学与分子生物学(第3版) | 贾弘禔            | 冯作化 药立波 | 方定志 焦炳华 周春燕      |
| 6  | 生理学(第3版)        | 姚 泰            | 王庭槐     | 闫剑群 郑 煜 祁金顺      |
| 7  | 医学微生物学(第3版)     | 贾文祥            | 李明远 徐志凯 | 江丽芳 黄 敏 彭宜红 郭德银  |
| 8  | 人体寄生虫学(第3版)     | 詹希美            | 吴忠道 诸欣平 | 刘佩梅 苏 川 曾庆仁      |
| 9  | 医学遗传学(第3版)      |                | 陈 竺     | 傅松滨 张灼华 顾鸣敏      |
| 10 | 医学免疫学(第3版)      |                | 曹雪涛 何 维 | 熊思东 张利宁 吴玉章      |
| 11 | 病理学(第3版)        | 李甘地            | 陈 杰 周 桥 | 来茂德 卞修武 王国平      |
| 12 | 病理生理学(第3版)      | 李桂源            | 王建枝 钱睿哲 | 贾玉杰 王学江 高钰琪      |
| 13 | 药理学(第3版)        | 杨世杰            | 杨宝峰 陈建国 | 颜光美 臧伟进 魏敏杰 孙国平  |
| 14 | 临床诊断学(第3版)      | 欧阳钦            | 万学红 陈 红 | 吴汉妮 刘成玉 胡申江      |
| 15 | 实验诊断学(第3版)      | 王鸿利 张丽霞<br>洪秀华 | 尚 红 王兰兰 | 尹一兵 胡丽华 王 前 王建中  |
| 16 | 医学影像学(第3版)      | 刘玉清            | 金征宇 龚启勇 | 冯晓源 胡道予 申宝忠      |
| 17 | 内科学(第3版)        | 王吉耀 廖二元        | 王 辰 王建安 | 黄从新 徐永健 钱家鸣 余学清  |
| 18 | 外科学(第3版)        |                | 赵玉沛 陈孝平 | 杨连粤 秦新裕 张英泽 李 虹  |
| 19 | 妇产科学(第3版)       | 丰有吉            | 沈 铿 马 丁 | 狄 文 孔北华 李 力 赵 霞  |

|    | 学科名称           | 主审      | 主编      | 副主编             |
|----|----------------|---------|---------|-----------------|
| 20 | 儿科学(第3版)       |         | 桂永浩 薛辛东 | 杜立中 母得志 罗小平 姜玉武 |
| 21 | 感染病学(第3版)      |         | 李兰娟 王宇明 | 宁 琴 李 刚 张文宏     |
| 22 | 神经病学(第3版)      | 饶明俐     | 吴 江 贾建平 | 崔丽英 陈生弟 张杰文 罗本燕 |
| 23 | 精神病学(第3版)      | 江开达     | 李凌江 陆 林 | 王高华 许 毅 刘金同 李 涛 |
| 24 | 眼科学(第3版)       |         | 葛 坚 王宁利 | 黎晓新 姚 克 孙兴怀     |
| 25 | 耳鼻咽喉头颈外科学(第3版) |         | 孔维佳 周 梁 | 王斌全 唐安洲 张 罗     |
| 26 | 核医学(第3版)       | 张永学     | 安 锐 黄 钢 | 匡安仁 李亚明 王荣福     |
| 27 | 预防医学(第3版)      | 孙贵范     | 凌文华 孙志伟 | 姚 华 吴小南 陈 杰     |
| 28 | 医学心理学(第3版)     | 姜乾金     | 马 辛 赵旭东 | 张 宁 洪 炜         |
| 29 | 医学统计学(第3版)     |         | 颜 虹 徐勇勇 | 赵耐青 杨士保 王 彤     |
| 30 | 循证医学(第3版)      | 王家良     | 康德英 许能锋 | 陈世耀 时景璞 李晓枫     |
| 31 | 医学文献信息检索(第3版)  |         | 罗爱静 于双成 | 马 路 王虹菲 周晓政     |
| 32 | 临床流行病学(第2版)    | 李立明     | 詹思延     | 谭红专 孙业桓         |
| 33 | 肿瘤学(第2版)       | 郝希山     | 魏于全 赫 捷 | 周云峰 张清媛         |
| 34 | 生物信息学(第2版)     |         | 李 霞 雷健波 | 李亦学 李劲松         |
| 35 | 实验动物学(第2版)     |         | 秦 川 魏 泓 | 谭 毅 张连峰 顾为望     |
| 36 | 医学科学研究导论(第2版)  |         | 詹启敏 王 杉 | 刘 强 李宗芳 钟晓妮     |
| 37 | 医学伦理学(第2版)     | 郭照江 任家顺 | 王明旭 尹 梅 | 严金海 王卫东 边 林     |
| 38 | 皮肤性病学          | 陈洪铨 廖万清 | 张建中 高兴华 | 郑 敏 郑 捷 高天文     |



## 第三版序言

经过再次打磨,备受关爱期待,八年制临床医学教材第三版面世了。怀纳前两版之精华而愈加求精,汇聚众学者之智慧而更显系统。正如医学精英人才之学识与气质,在继承中发展,新生方可更加传神;切时代之脉搏,创新始能永领潮头。

经过十年考验,本套教材的前两版在广大读者中有口皆碑。这套教材将医学科学向纵深发展且多学科交叉渗透融于一体,同时切合了环境-社会-心理-工程-生物这个新的医学模式,体现了严谨性与系统性,诠释了以人为本、协调发展的思想。

医学科学道路的复杂与简约,众多科学家的心血与精神,在这里汇集、凝结并升华。众多医学生汲取养分而成长,万千家庭从中受益而促进健康。第三版教材以更加丰富的内涵、更加旺盛的生命力,成就卓越医学人才对医学誓言的践行。

坚持符合医学精英教育的需求,“精英出精品,精品育精英”仍是第三版教材在修订之初就一直恪守的理念。主编、副主编与编委们均是各个领域内的权威知名专家学者,不仅著作立身,更是德高为范。在教材的编写过程中,他们将从医执教中积累的宝贵经验和医学精英的特质潜移默化地融入到教材中。同时,人民卫生出版社完善的教材策划机制和经验丰富的编辑队伍保障了教材“三高”(高标准、高起点、高要求)、“三严”(严肃的态度、严谨的要求、严密的方法)、“三基”(基础理论、基本知识、基本技能)、“五性”(思想性、科学性、先进性、启发性、适用性)的修订原则。

坚持以人为本、继承发展的精神,强调内容的精简、创新意识,为第三版教材的一大特色。“简洁、精练”是广大读者对教科书反馈的共同期望。本次修订过程中编者努力做到:确定系统结构,落实详略有方;详述学科三基,概述相关要点;精选创新成果,简述发现过程;逻辑环环紧扣,语句精简凝练。关于如何在医学生阶段培养创新素质,本教材力争达到:介绍重要意义的医学成果,适当阐述创新发现过程,激发学生创新意识、创新思维,引导学生批判地看待事物、辩证地对待知识、创造性地预见未来,踏实地践行创新。

坚持学科内涵的延伸与发展,兼顾学科交叉与融合,并构建立体化配套、数字化的格局,为第三版教材的一大亮点。此次修订在第二版的基础上新增了《皮肤性病学》。本套教材通过编写委员会的顶层设计、主编负责制下的文责自负、相关学科的协调与磋商、同一学科内部的专家互审等机制和措施,努力做到其内容上“更新、更深、更精”,并与国际紧密接轨,以实现培养高层次的具有综合素质和发展潜能人才的目标。大部分教材配套有“学习指导及习题集”、“实验指导/实习指导”以及“在线增值服务(多媒体课件与在线课程等)”,以满足广大医学院校师生对教学资源多样化、数字化的需求。

本版教材也特别注意与五年制教材、研究生教材、住院医师规范化培训教材的区别与联系。①五年制教



材的培养目标:理论基础扎实、专业技能熟练、掌握现代医学科学理论和技术、临床思维良好的通用型高级医学人才。②八年制教材的培养目标:科学基础宽厚、专业技能扎实、创新能力强、发展潜力大的临床医学高层次专门人才。③研究生教材的培养目标:具有创新能力的科研型和临床型研究生。其突出特点:授之以渔、评述结合、启示创新,回顾历史、剖析现状、展望未来。④住院医师规范化培训教材的培养目标:具有胜任力的合格医生。其突出特点:结合理论,注重实践,掌握临床诊疗常规,注重预防。

以吴孟超、陈灏珠为代表的老一辈医学教育家和科学家们对本版教材寄予了殷切的期望,教育部、国家卫生和计划生育委员会、国家新闻出版广电总局等领导关怀备至,使修订出版工作得以顺利进行。在这里,衷心感谢所有关心这套教材的人们!正是你们的关爱,广大师生手中才会捧上这样一套融贯中西、汇纳百家的精品之作。

八学制医学教材的第一版是我国医学教育史上的重要创举,相信第三版仍将担负我国医学教育改革的使命和重任,为我国医疗卫生改革,提高全民族的健康水平,作出应有的贡献。诚然,修订过程中,虽力求完美,仍难尽人意,尤其值得强调的是,医学科学发展突飞猛进,人们健康需求与日俱增,教学模式更新层出不穷,给医学教育和教材撰写提出新的更高的要求。深信全国广大医药院校师生在使用过程中能够审视理解,深入剖析,多提宝贵意见,反馈使用信息,以便这套教材能够与时俱进,不断获得新生。

愿读者由此书山拾级,会当智海扬帆!

是为序。

中国工程院院士

中国医科学院原院长

北京协和医学院原院长

劉德培

二〇一五年四月

## 主编简介

李霞, 博士、教授、博士研究生导师, 哈尔滨医科大学生物信息科学与技术学院院长, 龙江学者特聘教授, 北京“百千万人才工程”入选者, 享受国务院特殊津贴。从事生物信息学、计算系统生物学等本科、研究生教学工作 30 余年, 主持创建的我国领先生物信息学人才培养和教育教学体系成为全国生物信息学教育模板, 培养了大批既具有扎实生物医药知识, 又具有很强理工科学思维和实践能力的现代紧缺人才, 为推动我国生物医学教育和科技发展做出了突出贡献, 先后获得黑龙江省教学名师、优秀中青年专家、优秀科技工作者、优秀共产党员等荣誉称号。

李霞教授是我国重大疾病生物信息学与计算系统生物学研究的开创者之一, 在复杂疾病治疗靶标与风险标志物筛选、重大疾病通路重构与子网识别、非编码基因(RNA)介导的疾病发生机理研究、新一代测序技术与复杂疾病分析、面向转化医学的重大疾病分析平台构建等领域做出了开创性的研究工作, 科研成果处于国内前列。主持国家 863 课题、973 课题、国家自然科学基金重大研究计划等国家级课题 15 项, 于国际著名学术期刊发表高水平 SCI 论文 130 余篇, 荣获中华医学科技奖、黑龙江省政府科技奖、中国女医师协会基础医学科技奖等科研奖励 20 余项。

雷健波, 美国生物医学信息学博士。华西医科大学临床医学毕业, 原北京协和医院临床医生, 获美国哥伦比亚大学工程学院计算机硕士(M.S)和医学院生物医学信息学硕士(M.A.), 美国德州大学医学部生物医学信息学博士(PhD), 现任北京大学医学信息学中心副教授, 硕士生导师。

独特的国内外跨学科(临床, 计算机, 医学信息学)的学习、研究和背景, 主持过国内新一代电子病历(EMR), 医院信息系统(HIS), 临床路径(CP)的开发, 以及用于新药创制的国家级临床和标本资源库的建设等。2010 年 4 月以人才引进到北京大学, 创立北京大学医学信息学中心, 任代主任, 常务副主任, 负责创建新学科“医学信息学”。主要的研究领域包括: 电子病历系统和个人健康档案、临床决策支持、医学自然语言处理、移动医疗、健康信息搜索和消费者健康信息学、移动医疗、医疗卫生大数据、医疗信息系统易用性等。

现任欧美同学会留美分会副会长, 中国卫生信息学会卫生信息学教育专业委员会副主委, 全国高等医药教材建设研究会“十二五”规划本科教材《卫生信息学概论》主编等。



李霞



雷健波



## 副主编简介



李亦学

李亦学, 博士, 教授, 博士生导师, 上海生物信息技术研究中心主任, 同济大学生命科学与技术学院教授, 上海交通大学生命技术学院生物信息学与生物统计学系主任, 上海科技大学生命科学与技术学院教授, 中科院上海生化细胞所研究员, 中科院系统生物学重点实验室副主任, 国家重大科技专项首席科学家, 享受国务院特殊津贴, 曾任“十五”国家 863 计划生物信息技术主题专家组组长, “十一五”国家 863 计划生物医药技术领域专家组专家。

从事生物信息学、比较基因组学、疾病生物标志物识别、基因调控网络构建、蛋白质组学等方面研究多年。在国际著名杂志 *Nature*, *Science*, *Nature Genetics*, *Nature Biotechnology*, *Nature Communications*, *Genome Research*, *Molecular Biology & Evolution*, *Molecular Systems Biology* 等共发表 SCI 论文 200 余篇, 先后获教育部自然科学一等奖, 上海市自然科学一等奖, 全国五一国际劳动奖章等奖项。



李劲松

李劲松, 浙江大学教授、博士生导师, 浙江省“千人计划”入选专家。1984 年毕业于浙江大学生物医学工程专业, 1997 年获日本京都大学医学博士学位。现任浙江大学生物医学工程与仪器科学学院常务副院长、生物医学工程研究所所长, 兼任国家自然科学基金委学科评审专家、国家留学基金委评审专家, 中国医院协会信息管理专委会 (CHIMA) 副主委, 中华医学会医学信息学分会医院信息化学组副组长等职。研究方向为: 生物医学信息学, 数字医疗技术与系统, 医学知识库及数据挖掘, 本体理论及语义技术的医学应用。出版学术著作《生物医学语义技术》(主编), 《数字医学概论》(副主编), 《数字化医院建设理念与实践》(副主编), *New Fundamental Technologies in Data Mining* (Chapter 8) 等。



在我讲生物信息学课的时候,学生们总是让我推荐好的教材。在过去近五年的时间里,这本由李霞任主编、李亦学和廖飞任副主编的《生物信息学》作为全国高等医药教材建设研究会的规划教材不仅为医药领域,也为生命科学的其他领域的广大使用者提供了切实的帮助。几年过去了,随着以DNA测序技术为代表的组学技术持续发展,使得数据获取成本不断降低、获取效率不断提高,导致以基因组为代表的组学数据迅速增加,因此对生物信息学的需求,无论是深度上、还是广度上都大大增加了。为了适应该领域科学研究的快速发展,并满足教学改革的需要,本书主要编审者对本教材进行了第2版修订。

总体上,本教材的第2版紧密跟踪国际、国内在组学和生物信息学领域的发展,关注该领域的前沿,并尽可能地把它介绍给读者。作者也注意研究与学习当前大数据领域在理论和技术上的成就,并融汇到本教材当中。具体来说,本教材对序列比对、表达谱分析、分子进化、芯片数据处理等经典的生物信息学内容做了精炼和集中。同时,增加了专门介绍生物信息学研究的前沿与热点的内容,用以启发读者的兴趣和开拓读者的眼界;增加了RNA序列与结构特征分析及非编码核酸,特别是长非编码核酸的研究进展,使生物信息学研究对象更加丰富、全面;增加了对新一代测序技术的介绍及由此带来的生物信息分析技术的发展;新版教材特别讨论了生物信息学与相关学科,如:转化医学、医学信息学和信息技术等的关联。这都表明,新版教材的宗旨不仅着重于生物信息学的内涵,也着重于生物信息学的前沿和发展。

现在,组学和生物信息学的成果已逐渐进入国民经济及人类健康的很多领域,像育种、分子诊断等。在不少教学和科研单位生物信息学已成为了一门基础学科和不可或缺的技术手段和研究工具。

新版教材在李霞和李亦学等教授的努力下不仅保留了第1版的优势与特点:集中、全面、易懂、实用,又增加了前沿与进展。相信这本教材一定能为教学和科研的发展提供更好的服务。

陈润生

2015年4月16日

2010年,在国内资深生物学家与医学专家的倡导下,经全国高等医药教材建设研究会、卫生部教材办公室组织有关专家论证,由全国十三所大学生物信息学领域专家和一线教师编写出版了《生物信息学》。教材使用覆盖面广泛,包括长学制基础医学及临床医学、五年制生物医学工程、生物信息学等专业学生,生命科学领域研究学者、教师,临床医生、及生物信息学从业人员。5年过去了,伴随着新一代测序技术的深入发展,“大数据”、“组学”研究成为当今生命科学领域的热点。及时、充分地补充相关基础理论知识及实践操作方法,是当下必行之路。经全国医药教材研究会及卫生部论证后,决定编写《生物信息学》第2版。

《生物信息学》第2版教材仍坚持“三基”、“五性”原则,并力求在内容和形式上有所创新。使教材具备相对系统、全面的生物信息学知识体系;突出实用性,以临床实际问题作为编写出发点;简化算法流程,突出应用软件和网络资源;贴合前沿技术、方法,增加国内外研究热点知识;主要培养长学制学生运用生物信息学方法解决临床问题、进行科研设计的能力。

《生物信息学》第2版精简基础内容,合并上一版内容相近或有较大关联的章节;结合实际应用,增加前沿新知识、新技术章节。全书共分为三篇十五章。第一篇生物信息学基础,含DNA、RNA和蛋白质序列信息资源、序列比对、序列特征分析、分子进化分析、基因芯片数据分析五章,均系生物医学相关领域发展过程中形成的基础生物信息数据及分析方法,合并第一版“双序列比对”、“多序列比对”两章为“序列比对”,合并“序列特征分析”、“表达序列分析”两章为“序列特征分析”;第二篇功能基因组信息学,含蛋白质组与蛋白质结构分析、基因注释与功能分类、转录调控的信息学分析、生物分子网络与通路和计算表观遗传学五章,均系功能基因组研究中颇具特色的生物信息学方法,合并第一版“蛋白质分析与蛋白质组学”、“蛋白质结构分析”两章为“蛋白质组与蛋白质结构分析”;第三篇生物信息学与人类复杂疾病,含复杂疾病的分子特征与计算分析、非编码RNA与复杂疾病、新一代测序技术与复杂疾病、药物生物信息学、生物信息学相关学科进展五章,均系近年发展起来的与复杂疾病有关的重要生物信息学方法,新增非编码RNA与复杂疾病的生物信息学研究、新一代测序、“组学”研究等前沿热点。

《生物信息学》第2版各章相对独立,都反映了学科研究领域某一方向的最新成果与发展趋势。为适应不同读者的需要,各章布局统一:第一节是引言,以简明易懂的语言介绍该章主要内容,包括能解决什么问题和解决问题的思路;后续各节介绍基本概念和常用生物信息学方法,着重于生物医学实际应用、操作方法和生物医学意义的解释;各章最后附小结、习题及主要参考文献。

本书是在第1版基础上修订完成的,修订过程中借鉴了第1版作者的论著和成果,在此致以谢意!教材编委是来自于全国15所高校相关研究方向的专家,每一章都凝聚了独特的学术思想、研究心得和研究成果。他们在百忙之中精心组织素材,斟字酌句编写,付出了大量心血。在此我们对全体编委的无私奉献深表谢意!同时,哈尔滨医科大学生物信息科学与技术学院的老师和研究生们也做了大量协助工作,在此一并致谢!同时,哈尔滨医科大学生物信息科学与技术学院王宏、张绍军、肖云、宁尚伟、徐娟、张云鹏、智慧、刘洪波、许超汉、李永生、白静,中南大学李曦,中山大学李学荣、吕志跃,上海交通大学徐沁等专家学者和青年教师参与了大量前沿章节的材料汇总、讨论和编写工作,在此一并致谢!

第2版教材得到国家863项目、973课题和黑龙江省生物医学工程重点学科经费资助,特此鸣谢!

本书修订过程中,尽管我们努力跟踪学科新发展、新技术,并尽力纳入到教材中来,以保持先进性和实用性,但时间紧迫,直至完稿,仍觉有许多不足之处,希望学术同仁不吝赐教,以便再版时改正。

李 霞

2015年4月10日



# 目 录

|  |   |
|--|---|
| 绪论   | 1 |
| INTRODUCTION TO BIOINFORMATICS                               | 1 |
| 第一节 生物信息学的兴起   | 1 |
| Section 1 The Rise of Bioinformatics                         | 1 |
| 第二节 生物信息学的内涵及其在生命科学中的应用                                      | 2 |
| Section 2 Bioinformatics and Its Application in Life Science | 2 |
| 一、生物信息学的内涵   | 3 |
| 二、生物信息学在现代生物医学中的应用   | 4 |
| 第三节 大数据时代的生物信息学与医学   | 5 |
| Section 3 Bioinformatics and Medicine: The Era of Big Data   | 5 |
| 一、人类基因组计划  | 5 |
| 二、组学与生物信息学   | 6 |
| 三、大数据时代的生物信息学与医学   | 8 |

## 第一篇 生物信息学基础 FOUNDATION OF BIOINFORMATICS

|   |    |
|---|----|
| 第一章 生物序列资源                                    | 14 |
| CHAPTER 1 BIOLOGICAL SEQUENCE RESOURCES       | 14 |
| 第一节 引言  | 14 |
| Section 1 Introduction                        | 14 |
| 第二节 NCBI 数据库与数据资源                             | 15 |
| Section 2 NCBI Data Sources                   | 15 |
| 一、NCBI 序列数据库概述                                | 15 |
| 二、NCBI 中的重要子库介绍                               | 16 |
| 第三节 UCSC 基因组浏览器与数据资源                          | 20 |
| Section 3 UCSC Genome Browser and Data Source | 20 |
| 一、UCSC 概述                                     | 20 |
| 二、UCSC 基因组浏览器                                 | 21 |
| 三、UCSC 中的数据资源和常用工具                            | 23 |
| 第四节 EMBL-EBI 数据库与数据资源                         | 25 |
| Section 4 EMBL-EBI Data Sources and Tools     | 25 |
| 一、EMBL-EBI 数据库概况                              | 25 |
| 二、EMBL 基因组和核酸序列资源                             | 26 |
| 三、UniProt 蛋白质数据资源                             | 29 |

|  |    |
|--|----|
| 四、Biomart 数据检索平台 .....                                     | 31 |
| 第五节 重要的非编码基因数据库 .....                                      | 33 |
| Section 5 The Important Non-coding Gene Databases .....    | 33 |
| 一、ENCODE 数据库与数据资源 .....                                    | 33 |
| 二、microRNA 数据资源 miRBase .....                              | 34 |
| 小结 .....   | 35 |
| Summary .....  | 36 |
| <br>   |    |
| 第二章 序列比对 .....   | 37 |
| CHAPTER 2 SEQUENCE ALIGNMENT .....                         | 37 |
| 第一节 引言 .....   | 37 |
| Section 1 Introduction .....                               | 37 |
| 一、同源、相似与距离 .....   | 37 |
| 二、相似与距离的定量描述 .....   | 38 |
| 三、算法实现的比对 .....  | 40 |
| 四、序列比对的作用 .....  | 42 |
| 第二节 比对算法概要 .....   | 42 |
| Section 2 Alignment Algorithms .....                       | 42 |
| 一、替换计分矩阵 .....   | 42 |
| 二、双序列全局比对 .....  | 46 |
| 三、双序列局部比对 .....  | 48 |
| 四、多序列全局比对 .....  | 48 |
| 五、多序列局部比对 .....  | 52 |
| 六、比对的统计显著性 .....   | 53 |
| 第三节 数据库搜索 .....  | 54 |
| Section 3 Database Search .....                            | 54 |
| 一、经典 BLAST .....   | 54 |
| 二、衍生 BLAST .....   | 56 |
| 三、BLAT .....   | 57 |
| 四、RNA 序列搜索 .....   | 58 |
| 五、数据库搜索的统计显著性 .....  | 59 |
| 第四节 比对软件、参数与数据资源 .....                                     | 60 |
| Section 4 Alignment Software, Parameter and Resource ..... | 60 |
| 一、参数选择的一般原则 .....  | 60 |
| 二、主要比对软件 .....   | 61 |
| 三、EBI 中的序列比对工具 .....                                       | 62 |
| 四、UCSC 中的 BLAT 比对工具 .....                                  | 62 |
| 第五节 比对技术的发展 .....  | 63 |
| Section 5 Advances of Alignment Techniques .....           | 63 |
| 一、glocal 比对 .....  | 63 |
| 二、全基因组比对 .....   | 64 |
| 小结 .....   | 66 |
| Summary .....  | 66 |



|   |     |
|---|-----|
| 第三章 序列特征分析  | 69  |
| CHAPTER 3 SEQUENCE CHARACTER ANALYSIS                       | 69  |
| 第一节 引言  | 69  |
| Section 1 Introduction                                      | 69  |
| 第二节 DNA 序列特征分析  | 70  |
| Section 2 DNA Sequence Character Analysis                   | 70  |
| 一、DNA 序列的基本信息   | 70  |
| 二、DNA 序列的特征信息   | 72  |
| 三、基因组结构注释分析   | 73  |
| 第三节 蛋白质序列特征分析   | 76  |
| Section 3 Protein Sequence Character Analysis               | 76  |
| 一、蛋白质序列的基本信息分析  | 76  |
| 二、蛋白质序列的特征信息分析  | 79  |
| 三、蛋白质序列的功能信息分析  | 82  |
| 第四节 RNA 序列与结构特征分析   | 85  |
| Section 4 Analysis of RNA Sequence and Structure Characters | 85  |
| 一、RNA 的序列特征   | 85  |
| 二、RNA 的结构特征   | 85  |
| 三、RNA 二级结构预测方法  | 86  |
| 四、RNA 结构预测的在线资源与软件  | 90  |
| 第五节 表达序列特征分析  | 91  |
| Section 5 Analysis of Expressed Sequence Characters         | 91  |
| 一、表达序列的获取和数据库资源   | 92  |
| 二、表达序列标签分析方法  | 97  |
| 三、GEO 数据处理与分析   | 100 |
| 四、SAGE 数据分析   | 101 |
| 小结  | 102 |
| Summary   | 102 |
| 第四章 分子进化分析  | 105 |
| CHAPTER 4 MOLECULAR EVOLUTION ANALYSIS                      | 105 |
| 第一节 引言  | 105 |
| Section 1 Introduction                                      | 105 |
| 第二节 系统发生分析与重建   | 105 |
| Section 2 Phylogeny Reconstruction                          | 105 |
| 一、核苷酸置换模型及氨基酸置换模型   | 105 |
| 二、系统发生树的基本概念及搜索方法   | 110 |
| 三、分子钟假说   | 113 |
| 第三节 核苷酸和蛋白质的适应性进化   | 113 |
| Section 3 Adaptive Evolutions of Nucleotide and Protein     | 113 |
| 一、中性与近中性理论  | 113 |
| 二、微观适应性进化的检验方法  | 114 |
| 三、宏观适应性进化的检验方法  | 115 |