

公共选修课课程介绍

课程名称	生物信息与大数据		总学分：2 总学时：34			其中	理论	34	
							实验	0	
开课院部	生命科学与技术学院	教研室	微 生 物 与 合 成 生物学	教学 对象	不限				
教材名称	生物信息学	主编	陈 铭	年 版	2022 第 4 版	出版社名 称/自编	科学出版社		

课程简介（课程的背景、教学目的、主要内容等，不少于 600 字，）：

课程背景

随着生命科学的迅速发展和高通量测序技术的普及，生物信息学已成为现代生物学、医学及制药领域的核心学科之一。生物信息学不仅推动了基因组学、转录组学、蛋白质组学等多组学数据的深入解析，还在精准医学、新药研发、疾病诊断和个体化治疗等方面发挥着重要作用。同时，人工智能、大数据和云计算等技术的快速进步，使得生物数据的存储、处理和分析能力显著增强，极大地提升了生物信息学的应用价值。因此，掌握生物信息学的基本理论和数据分析方法，已成为生命科学研究者和制药行业从业者的必备技能。

主要内容

本课程结合中国药科大学的学科特色，以新药研发为核心背景，系统介绍生物信息学的基本概念、研究方向、数据分析方法及其在医药领域的应用。课程内容涵盖生物信息学基础、大数据管理、Linux 操作系统、生物信息学数据库、测序技术、生物数据分析方法以及数据可视化等多个方面。通过理论讲解与实践操作相结合的方式，使学生能够在短时间内掌握生物信息学的核心技能，并能够在实际科研中运用所学知识解决生物医学问题。

教学目的

总目标：

本课程面向生命科学、医学、药学及相关交叉学科的学生，旨在系统讲解生物信息学的基本理论、核心方法和数据分析技能。通过理论讲授与实践操作结合，培养学生掌握计算工具、生物数据处理和可视化分析的能力，并能独立完成转录组数据分析等核心任务。同时，课程融入国家生物信息战略，增强学生的科研诚信意识、数据安全意识和科技自立自强精神，为未来科研与行业应用奠定基础。

课程子目标：

1. 掌握生物信息学基本技术（基础知识与技能）
 - 了解生物信息学的基本概念、学科定位及其在生命科学中的重要性。
 - 熟悉基因组学、转录组学、蛋白质组学等核心研究领域，并掌握常用数据分析方法。
 - 掌握计算机基础知识，包括计算机系统原理、Linux 操作、R 语言及数据可视化工具的使用。
2. 提升生物数据分析能力（方法与实践）
 - 掌握生物数据的处理与存储方式，包括数据库检索、序列数据格式、云计算环境的使用。
 - 能够独立完成 RNA-seq 数据的分析流程，包括数据预处理、差异表达分析、功能富集分析等。
 - 具备数据可视化能力，能够对一些基本的可视图表的结果进行解读。
3. 强化科研思维与问题解决能力（综合应用）
 - 通过案例分析和实际操作，培养学生的数据驱动科研能力，提高数据建模、分析和决策能力。
 - 能够运用多组学数据整合方法，解决复杂生命科学问题，并理解人工智能在生物信息学中的应用。
 - 具备批判性思维，能够独立分析实验数据，并对结果进行科学解释。
4. 增强国家科技自立自强意识与数据安全伦理意识（思政目标）
 - 结合我国生物信息学发展历程，引导学生认识生物数据主权的重要性，增强科技自立自强意识。
 - 通过国家生物信息战略、数据安全法规案例，让学生理解数据合规、伦理道德和科研诚信的核心价值。
 - 引导学生关注精准医学、新药研发等领域的社会价值，培养其责任感和科研使命感，促进科技报国精神的形成。

讲授提纲（每一章节的名称）

第 1 章：绪论

1. 生物信息学概述
2. 生物信息学的发展历程
3. 当代生物信息学的研究方向

第 2 章：生物信息学大数据与 Linux 基础

1. 计算机组成原理（2 学时）

- 计算机基本结构（CPU、内存、硬盘、总线）
- 存储系统与文件系统（RAM、ROM、SSD、HDD、文件格式）
- 进程与线程基础概念

2. 生物信息学与大数据（1 学时）

- 什么是大数据，生物信息学中的大数据类型
- 云计算与分布式计算（HPC、高通量计算、云计算、存储架构）
- 云计算的基本服务类型及其在生物信息学中的应用

3. Linux 基础知识（4 学时）

- Linux 系统简介（Unix/Linux 发展历史、Linux vs. Windows）
- Linux 文件系统结构（/home, /bin, /etc, /dev, /tmp, /var）
- 基本 Linux 命令（文件操作、目录管理、权限管理、进程管理）
- 用户管理与权限控制（sudo、chmod、chown）
- 文本编辑工具 Vim（基本使用方法、文件编辑、查找替换）
- Shell 脚本入门（变量、条件语句、循环）

第 3 章：生物信息学数据库与检索

1. 生物医学数据库概述

- 数据库的基本结构与作用
- 生物医学数据库的特点
- 常用数据库介绍（Ensembl、GEO、GenBank、PubMed）

2. 数据库检索方法与数据格式

- 核酸序列数据库的检索方法
- 常见数据格式（FASTA、FASTQ、GenBank、GFF）

3. 文献检索

- 学术论文检索
- 专利文献检索

第 4 章：测序技术的发展

1. 一代测序技术（Sanger 测序）

2. 二代测序技术（Illumina 测序）

3. 三代测序技术（Nanopore、PacBio 测序）

第 5 章：二代测序数据分析实践

1. 二代测序数据处理流程概述

- 质量控制 (QC)、修剪、比对、定量
- 二代测序数据处理的整体流程框架
- 各步骤的重要性及对后续分析结果的影响

2. 数据的下载与软件安装的准备工作

- 公开数据库下载二代测序数据 (如 GEO、SRA)
- Linux 环境下安装必要的软件 (FastQC、Trimmomatic、HISAT2、StringTie)

3. 质量控制 (Quality Control)

- FastQC 软件的功能及结果解读
- 质控指标分析 (序列质量得分、GC 含量等)

4. 序列修剪 (Trimming)

- Trimmomatic 软件的使用流程
- 低质量序列与接头序列的去除

5. 序列比对 (Mapping)

- HISAT2 软件在有参考基因组比对中的应用
- 关键比对参数 (比对率、覆盖度、错配率)

6. 基因表达定量 (Quantifying)

- StringTie 在转录本定量中的应用
- 生成表达矩阵的操作步骤
- 结果格式 (TPM、FPKM、Counts)

7. 实践环节

- 学生独立完成完整的二代测序上游分析

第 6 章：生物信息学与图形可视化展示

1. 数据可视化的重要性

- 可视化在生物数据分析中的作用
- 不同类型数据适用的可视化方法

2. 常见的生物信息可视化图形类型

- 统计学基础可视化 (箱线图、条形图、散点图、折线图)
- 组学数据常见可视化 (热图、PCA 分析、韦恩图、火山图)

- 临床数据常见可视化（生存曲线图、蜂群图）
3. 生物信息可视化工具
 - R 语言的 ggplot2、ComplexHeatmap、pheatmap
 - 其他可视化软件（Cytoscape、Graphpad）
 4. 实践环节
 - 使用 ggplot2 绘制各种常见图形
 - 美化图像（图形美化、颜色搭配等）
 - 利用 pheatmap/ComplexHeatmap 制作基因表达热图
 - 生存分析曲线绘制（Kaplan-Meier Plot）

第 7 章：R 语言初识：R 语言在转录组学数据分析中的应用

1. R 语言介绍（安装、配置、使用）
2. RNA-seq 数据下游数据的格式解析与导入
3. 差异表达分析（DESeq2、edgeR、limma）
4. PCA、GSEA、GO/KEGG 功能富集分析
5. 结果可视化（热图、火山图）

第 8 章：国内生物信息中心发展与数据安全

1. 我国生物信息中心的发展历程（1 学时）
 - 生命科学进入大数据时代（基因组测序、蛋白质组学、单细胞测序）
 - 国家生物信息中心（NGDC）的建立及发展历程
 - 科学数据管理的现状与挑战
2. 生命科学大数据的管理与整合（1 学时）
 - 生物大数据的存储与计算挑战（存储成本、数据格式、计算资源需求）
 - 国家级生物信息平台（NGDC、中国科学院生物信息中心）
 - 数据整合与共享机制（FAIR 原则）
3. 生物遗传资源与科学数据的安全管理（1 学时）
 - 生物数据安全的基本概念（数据合规、数据主权、隐私保护）
 - 国家政策法规介绍（人类遗传资源管理条例、生物安全法、数据安全法）
 - 科学数据安全管理的实践案例（国家基因组数据汇交与保护）

考核方式或评分标准（笔试、论文、实际操作考察等）：

序号	成绩类别	考核方式	考核要求	评价权重（%）
1	过程性考核 （平时成绩）	考勤 10%、课堂表现 10%、 小组任务 10%和作业 10%	课堂提问、讨论、小组案例分析作业、编程作业	40%
2	期中作业	项目报告	代码编写、数据分析报告撰写	10%
3	期末成绩	客观题、项目报告	对重点知识点进行客观题考察、完成期末项目分析报告	50%

任课教师简介（不少于 50 字）：

夏云，工学博士，中国药科大学生命科学与技术学院副教授，独立 PI，研究生导师。研究领域涵盖生物医学大数据整合、新药发现、肿瘤免疫及生物标志物识别。团队开发诊断与预后模型，构建肿瘤与免疫分析数据库，并申请多项国内外专利。2024 年在 Nature Immunology（封面文章）、Briefings in Bioinformatics、Thyroid、iMeta 等期刊发表论文。曾任职于华中科技大学科学技术发展院，具专利代理师资格，擅长生物医药领域的专利挖掘与成果转化。