

# 天津医科大学理论课教案首页

(共 4 页、第 1 页)

课程名称：生物信息学	课程内容/章节：基因组功能的高级注释
教师姓名：伊现富	职称：讲师
教学日期：2014 年 4 月 9 日 10 时 -12 时	
授课对象：生物医学工程学院 2011 级生信班（本）	听课人数：27
授课方式：理论讲授	学时数：2
教材版本：生物信息学（自编教材）	

## 教学目的与要求（分掌握、熟悉、了解、自学四个层次）：

- 掌握序列标识的含义和制作工具。
- 熟悉变异位点注释结果的解析；基因集的富集分析及其结果解析；box plot 的含义及其绘制。
- 了解变异位点注释的内容和常用工具；基因集富集分析的常用工具。
- 自学变异位点注释、基因集富集分析、序列标识制作等工具的使用方法。

## 授课内容及学时分配：

- (5') 回顾与导入：回顾基因组注释的基础知识，介绍功能注释的主要内容。
- (25') 变异位点的注释：介绍变异位点注释的内容、步骤及相关的注释工具，讲解对注释结果的解析。
- (20') 基因集富集分析：介绍基因集富集分析的用途，讲解常用的 DAVID 工具及其结果的解析。
- (25') 序列标识：讲解序列标识的含义，介绍常用的 WebLogo 及其使用方法并讲解对结果的解析。
- (20') box plot：介绍 box plot 及相关概念，讲解绘制 box plot 的主要步骤。
- (5') 总结与答疑：总结授课内容中的知识点与技能，解答学生疑问。

## 教学重点、难点及解决策略：

- 重点：序列标识的含义；解决策略：通过制作过程的演示和对结果的解读来加深学生的理解。
- 难点：注释分析结果的解析；解决策略：通过对实例的分析帮助学生掌握解析结果的基本原则和主要步骤。

## 专业外语词汇或术语：

单核苷酸变异 (SNV)	富集分析 (enrichment analysis)
基因集 (gene set)	序列标识 (sequence logo)
GO (gene ontology)	箱线图 (box plot)

## 辅助教学情况：

- 多媒体：变异位点注释、基因集富集分析的实例；序列标识和 box plot 的示意图；DAVID、WebLogo 等工具的界面；绘制 box plot 的演示视频。
- 板书：box plot 的主要绘制步骤。
- 操作演示：序列标识的制作。

## 复习思考题：

- 以变异位点的注释结果为例，论述如何解析一张表。
- 以 DAVID 富集分析结果为例，论述如何解析一张表。
- 简述序列标识的含义，能解读实际的序列标识图。
- 以 box plot 为例，论述如何解析一张图。

## 参考资料：

- 李霞，李亦学，廖飞。生物信息学，人民卫生出版社，2010 年。
- 朱玉贤，李毅，郑晓峰。现代分子生物学（第 3 版），高等教育出版社，2007。
- 维基百科

主任签字：

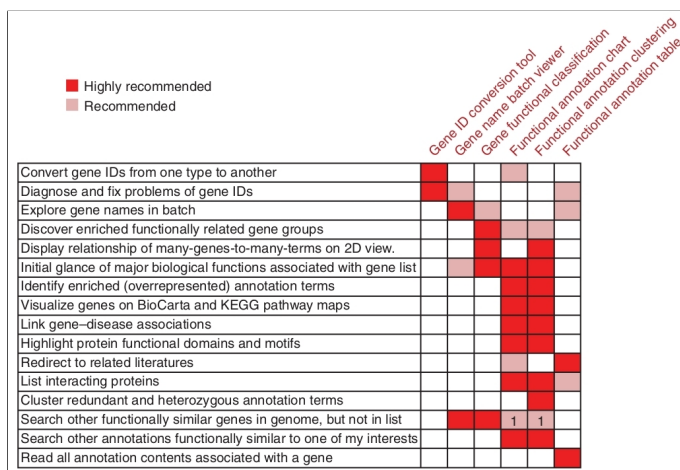
年 月 日

教务处制



## 2. DAVID 分析工具 (根据任务选择工具)

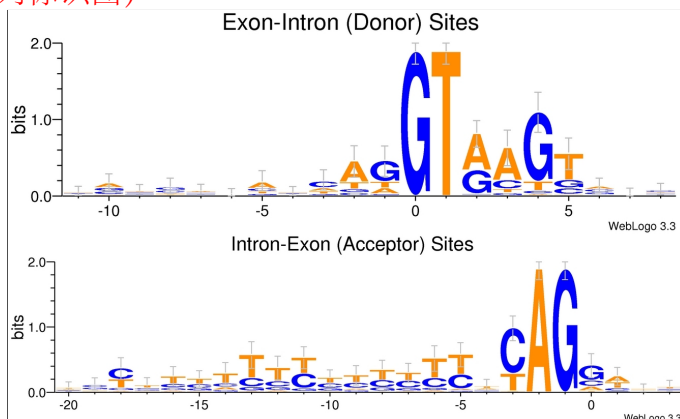
- Gene Name Batch Viewer
- Gene ID Conversion Tool
- Gene Functional Classification Tool
- Functional Annotation Tool
  - Functional Annotation Clustering: 根据注释信息聚类注释项目
  - Functional Annotation Chart: 根据注释信息进行富集分析
  - Functional Annotation Table: 以表格形式呈现注释信息



## 四、序列标识 (25 分钟)

### 1. 图形含义 (以“GT-AG 规则”为例讲解序列标识图)

- 数据: 多序列比对信息
- 横轴: 序列的坐标位置
- 纵轴: 比特, 计量单位
- 字符堆叠的总高度: 此位置的保守性
- 每个字符的高度: 出现的相对频率



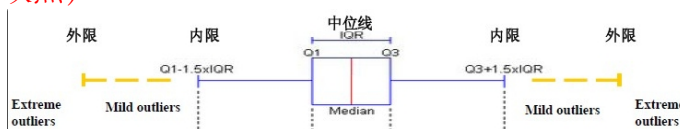
### 2. 制作工具 (演示 WebLogo 的使用)

- WebLogo
- enoLOGOS
- Skylign

## 五、box plot (20 分钟)

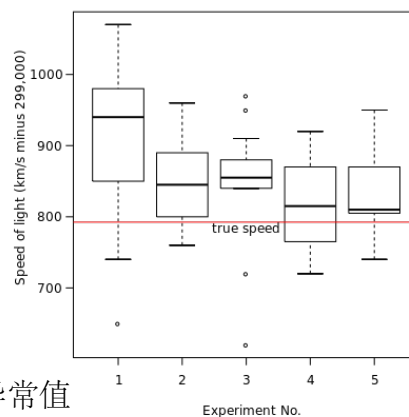
### 1. box plot 简介 (通过实例和示意图讲解其优缺点)

- box plot, Box-whisker Plot, 箱线图
- 1977, 美国, 约翰·图基 (John Tukey)
- 显示一组数据分散情况的统计图
- 可以粗略看出数据分布的离散程度
- 适合用于几个样本的比较
- 不能提供数据分布偏态的精确度量



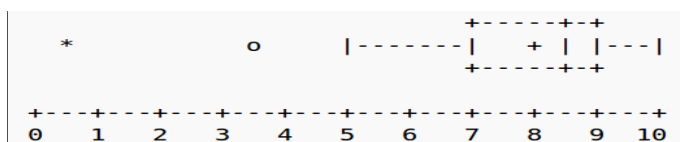
### 2. 相关概念 (通过实例帮助学生理解记忆)

- 最小值 min, 最大值 max, 中位数 median
- 下四分位数 Q1, 上四分位数 Q3
- 四分位数差  $IQR = Q3 - Q1$
- 内限:  $Q3 + 1.5IQR$ ,  $Q1 - 1.5IQR$
- 外限:  $Q3 + 3IQR$ ,  $Q1 - 3IQR$
- 异常值 (outliers): 处于内限以外的数据
- 温和的异常值 (mild outliers): 在内限与外限之间的异常值
- 极端的异常值 (extreme outliers): 在外限以外的异常值



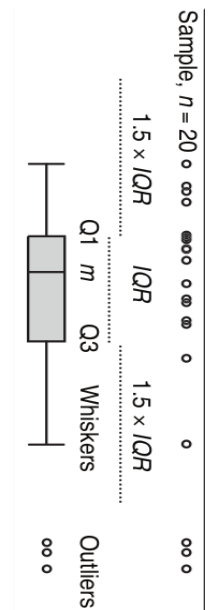
### 3. 图解概念

- $min = 0.5, max = 10$
- $Q1 = 7, Q3 = 9, IQR = 2$
- $median = 8.5, mean = 8$



## 4. 绘图步骤 (通过观看视频学习绘图的具体步骤)

- 绘制数轴。
- 计算上四分位数 ( $Q_3$ )，中位数，下四分位数 ( $Q_1$ )。
- 计算四分位数差 ( $IQR$ )。
- 绘制箱线图的矩形，上限为  $Q_3$ ，下限为  $Q_1$ 。在矩形内部中位数的位置画一条横线 (中位线)。
- 在  $Q_3 + 1.5IQR$  和  $Q_1 - 1.5IQR$  处画两条与中位线一样的线段，这两条线段为异常值截断点，称为内限；在  $Q_3 + 3IQR$  和  $Q_1 - 3IQR$  处画两条线段，称为外限。
- 在非异常值的数据中，最靠近上边缘和下边缘 (即内限) 的两个数值处画横线，作为箱线图的触须。
- 从矩形的两端向外各画一条线段直到不是异常值的最远点 (即上一步的触须)，表示该批数据正常值的分布区间。
- 温和的异常值用空心圆表示；极端的异常值用实心点 (一说用星号 \*) 表示。



## 5. 绘图工具: BoxPlotR, ECplot, R, ...

## 六、总结与答疑 (5 分钟)

### 1. 知识点

- 变异位点的注释: 用途, 注释工具
- 基因集富集分析: 功能, 分析工具
- 序列标识: 含义, 制作工具
- box plot: 理解, 绘制

### 2. 技能

- 解析表格: 行列, 缩写, 数值
- 解析图片: 数据, 横纵轴, 图元素, 元素大小、颜色