

# 天津医科大学理论课教案首页

(共 3 页、第 1 页)

课程名称：生物信息学	课程内容/章节：4.2 基因组结构注释分析之基因识别	
教师姓名：伊现富	职称：讲师	教学日期：2015 年 5 月 21 日 8 时 -9 时
授课对象：生物医学工程学院 2012 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）：

- 掌握基因识别的方法和策略。
- 熟悉原核基因和真核基因的结构特点。
- 了解基因识别的分析工具。
- 自学基因识别分析工具的使用方法。

授课内容及学时分配：

- (3') 引言与导入：回顾中心法则，阐释核酸序列携带的两类遗传信息。
- (15') 基因识别：介绍基因和基因识别的基本概念，回顾原核基因和真核基因的结构特点并进行比较，讲解基因识别的主要方法与策略，介绍基因识别的常用工具。
- (2') 总结与答疑：总结授课内容中的知识点，解答学生疑问。

教学重点、难点及解决策略：

- 重点：基因识别的方法和策略。
- 难点：基因预测中“信号”特征和“内容”特征的区别。
- 解决策略：通过示意图和实例帮助学生理解，通过对比加深记忆。

专业外语词汇或术语：

基因识别 (gene prediction/finding)  
间接识别法 (extrinsic approach)  
从头计算法 (*ab initio* approach)

辅助教学情况：

- 多媒体：原核基因和真核基因的结构，基因识别的策略。
- 板书：查找数据库和分析工具的主要策略。

复习思考题：

- 比较原核基因和真核基因结构的异同。
- 简述基因识别的三大类方法。
- 比较基因预测中的信号与内容。
- 论述基因识别的主要策略。

参考资料：

- 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。
- 李霞，李亦学，廖飞。生物信息学，人民卫生出版社，2010。
- 王明怡，杨益，吴平。生物信息学（中译本，第 2 版），科学出版社，2004。
- 维基百科。

主任签字：

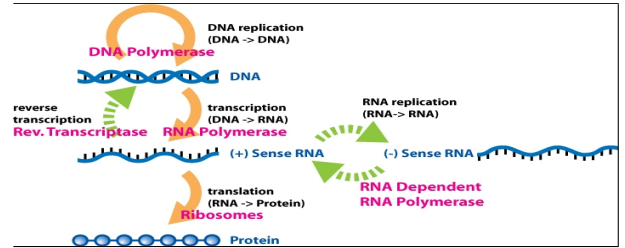
年 月 日

教务处制

## 一、引言与导入 (3 分钟)

### 1. 分子生物学的中心法则：DNA 转录成 RNA，RNA 翻译成蛋白质。

- DNA：携带最原始的决定个体性状的遗传信息
- RNA：参与遗传信息的表达和调控
- 蛋白质：执行特定的生物功能从而决定最终的表型



### 2. DNA 携带两类遗传信息

- 功能序列：具有功能活性的 DNA 序列，遗传的基本单位
- 调控信息：特定的 DNA 区域，能被功能性蛋白质分子特异地识别结合

## 二、基因识别 (15 分钟)

在介绍基本概念的基础上，通过比较原核基因和真核基因的异同，讲解基因识别的主要策略及各种方法在原核和真核基因识别中的具体应用。

### 1. 基本概念

- 基因：产生一条多肽链或功能 RNA 所需的全部核苷酸序列 (**强调既包括编码区，也包括非编码区**)
- 基因识别：识别 DNA 序列上具有生物学特征的片段



### 2. 基因结构 (教学重点：通过示意图形象化展示、比较原核和真核的基因结构)

- 共同点：都包括编码区和非编码区
- 原核基因：连续基因
- 真核基因：不连续性



### 3. 【重点】识别方法

- 间接识别法：mRNA/蛋白质序列  $\Rightarrow$  DNA 序列
- 从头预测法：基因预测，基于“信号”和“内容”两类特征
- 比较基因组学的方法：比较相关物种的 DNA 序列

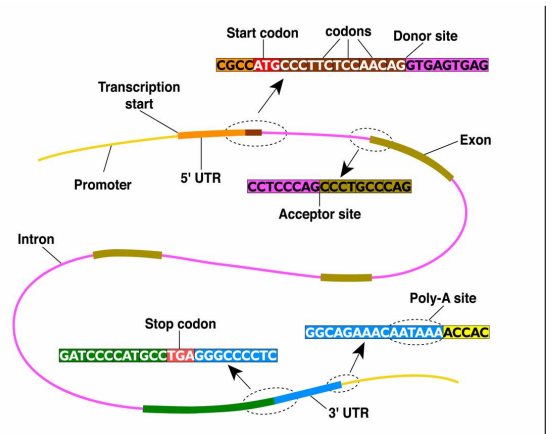
### 4. 基因预测 (教学难点：通过实例讲解并比较“信号”和“内容”)

#### • 【难点】“信号”和“内容”

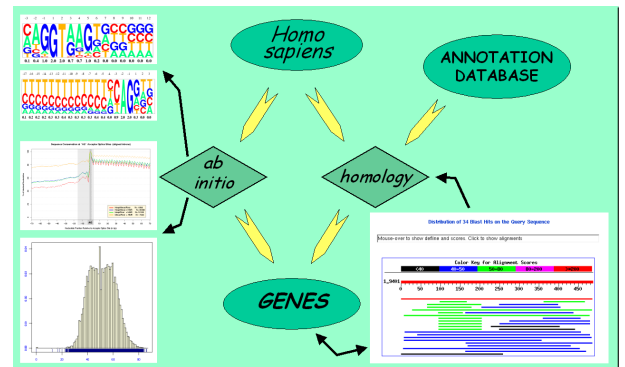
- 共同点：都包括编码区和非编码区
- 信号：不连续的局部序列模体，一般都有一致性序列；如启动子，剪接供体和受体位点，起始和终止密码子，polyA 位点
- 内容：不同长度的扩展序列，没有一致性序列，但具有把自己与周围 DNA 区分开来的保守特征；如密码子使用偏好性，双联密码子出现频率，基因组等值区

#### • 原核基因

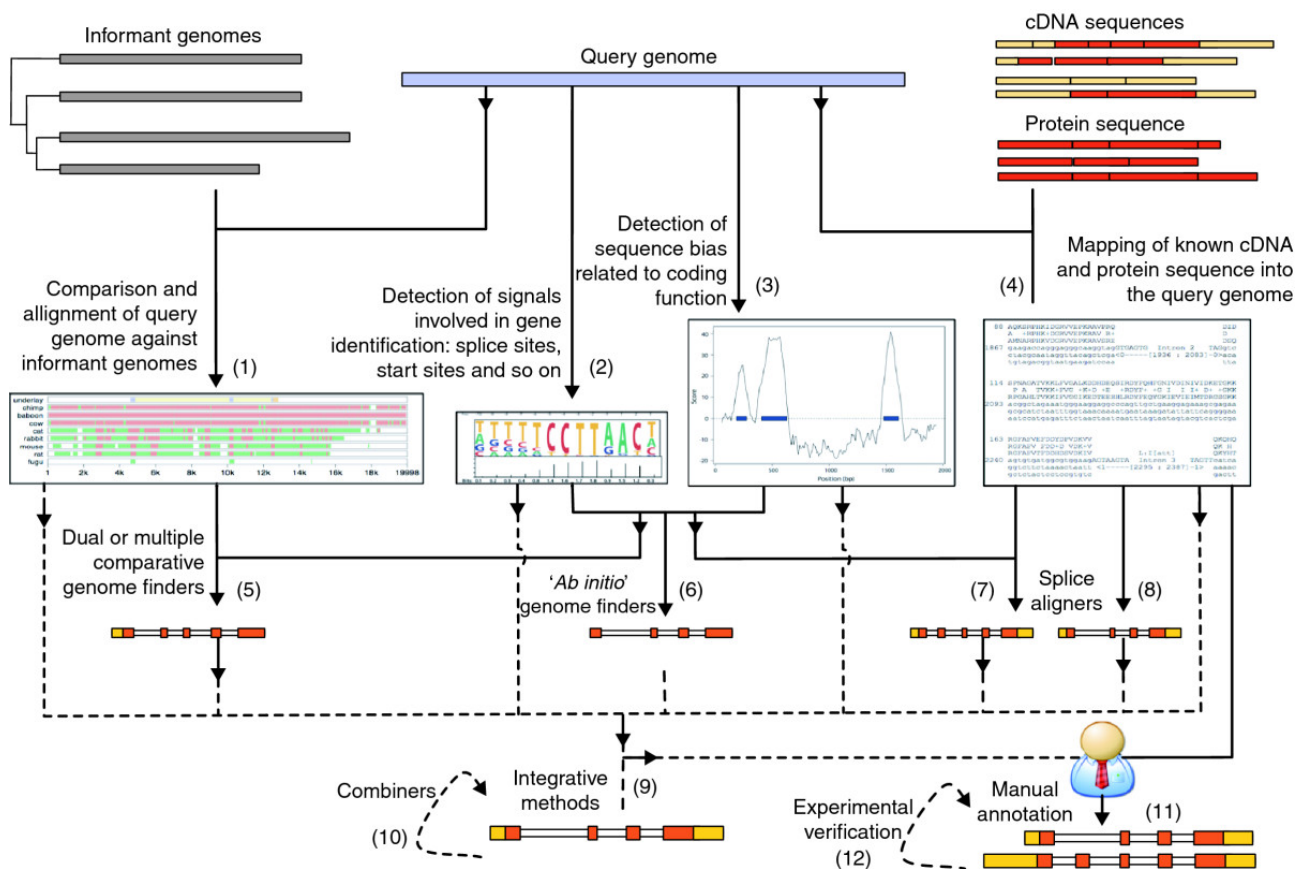
- 信号：启动子序列，转录因子结合位点
- 内容：连续的开放阅读框，统计学特征
- 总结：信号容易识别，内容容易判别，预测能达到相对较高的精度



- 真核基因
  - 信号：启动子区特征序列，供体和受体位点，起始和终止密码子，polyA 序列；确定外显子的边界，识别编码区域
  - 内容：密码子使用偏好性，双联密码子出现频率，基因组等值区；区分外显子、内含子和基因间区域
  - 总结：信号复杂，内容难判别，预测相当有挑战性；联合信号和内容检测以及同源性搜索，提高识别效率



## 5. 【重点】识别策略



## 6. 识别工具（强调分析工具的适用范围）

- 识别原核基因：GeneMarkS, Glimmer
- 识别真核基因：GENSCAN

## 三、总结与答疑（2 分钟）

- 原核和真核的基因结构，基因识别方法
- 基因识别方法
- 基因预测中的信号与内容
- 基因识别策略