

天津医科大学理论课教案首页

(共 4 页、第 1 页)

课程名称：系统生物学 课程内容/章节：转录组学 (RNA-Seq 概述) / 第 3 章

教师姓名：伊现富 职称：讲师 教学日期：2016 年 9 月 27 日 13:30-15:30

授课对象：生物医学工程与技术学院 2013 级生信班 (本) 听课人数：28

授课方式：理论讲授 学时数：2 教材版本：系统生物学，第 1 版

教学目的与要求 (分掌握、熟悉、了解、自学四个层次)：

- 掌握基因表达谱、转录组学等基本概念，RNA-Seq 流程中的关键步骤，RNA-Seq 的专用术语。
- 熟悉转录组学的研究内容和主要方法，RNA-Seq 的常见应用。
- 了解常见的组学，RNA-Seq 的主要技术。
- 自学转录组学研究的其他技术。

授课内容及学时分配：

- (5') 引言与导入：回顾第二代测序的主要技术和基本流程。
- (30') 转录组学：介绍常见的组学，讲解转录组学的基本概念和研究内容与方法。
- (30') RNA-Seq 简介：介绍 RNA-Seq，讲解 RNA-Seq 的主要技术和常见应用。
- (30') RNA-Seq 流程：介绍 RNA-Seq 的基本流程，总结实验操作与数据分析的关键步骤，讲解 RNA-Seq 的专用术语。
- (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：RNA-Seq 流程中的关键步骤，RNA-Seq 的专用术语。
- 难点：RNA-Seq 的专用术语。
- 解决策略：通过实例讲解和比较类比帮助学生理解、记忆。

专业外语词汇或术语：

基因表达 (gene expression)	Elements, ENCODE)
基因表达谱 (gene expression profile)	RPKM (Reads Per Kilobase Million)
转录组 (transcriptome)	FPKM (Fragments Per Kilobase Million)
DNA 元件百科全书 (Encyclopedia of DNA	TPM (Transcripts Per Kilobase Million)

辅助教学情况：

- 多媒体：常见组学介绍，转录组学的研究方法，RNA-Seq 的常见应用。
- 板书：RNA-Seq 的主要流程，RPKM/FPKM/TPM 的计算方法。

复习思考题：

- 什么是转录组学？
- 总结转录组学的研究内容和方法。
- 列举 RNA-Seq 的常见应用。
- 总结 RNA-Seq 的基本流程。
- 解释并计算 RPKM/FPKM/TPM。

参考资料：

- 维基百科等网络资源。

主任签字：

年 月 日

教务处制

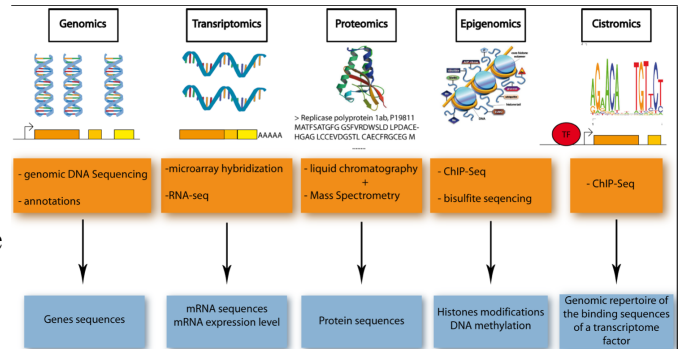
一、引言与导入 (5 分钟)

1. 二代测序技术: Roche/454, Illumina/Solexa, ABI/SOLiD
2. 二代测序流程: 实验流程, 分析流程 (QC, Mapping, Annotation)

二、转录组学 (30 分钟)

1. 组学 (结合分子生物学中的中心法则等知识点进行讲解)

- Genomics \Leftarrow DNA
- Transcriptomics \Leftarrow RNA
- Proteomics \Leftarrow Protein
- Epigenomics \Leftarrow Methylation/Histone
- Cistomics \Leftarrow Cistron(DNA-Protein)
- Metabolomics \Leftarrow Small molecules



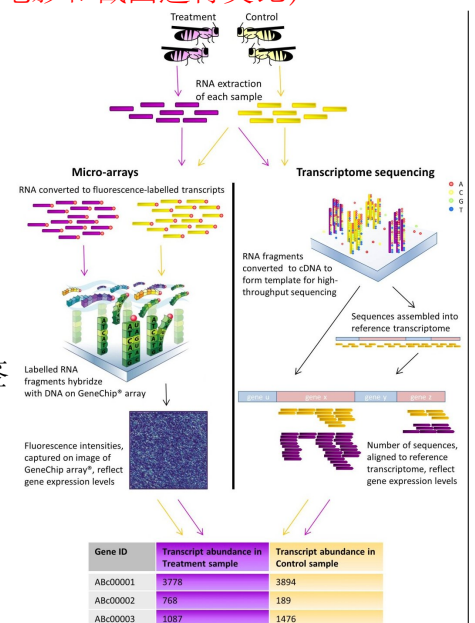
2. 转录组学

(1) 基本概念

- 基因表达: 基因中的信息合成基因产物 (蛋白质 + RNA) 的过程
- 基因表达谱: 是否表达, 表达丰度, 表达差异, ……
- mRNA 丰度: 每个细胞里每一种 mRNA 分子的平均数
- 转录组
 - 广义: 相同环境 (或生理条件) 下的在一个细胞或一群细胞中所能转录出的所有 RNA 的总和
 - 狭义: 细胞所能转录出的所有 mRNA

(2) 转录组学 (时空的连续性 vs. 快照, 动 vs. 静; 以电影和截图进行类比)

- 定义: 对转录水平上发生的事件及其相互关系和意义进行整体研究的一门学科
- 研究内容
 - 对特定细胞的转录与加工机制进行研究
 - 对转录物编制目录便于进一步归类研究
 - 绘制动态的转录物图形
 - 转录物调控网络
- 研究方法 (Microarray vs. RNA-Seq)
 - EST: expressed sequence tag, 表达序列标签
 - SAGE: serial analysis of gene expression
 - MPSS: massive parallel signature sequencing
 - Microarray
 - RNA-Seq: RNA sequencing



三、RNA-Seq 简介 (30 分钟)

1. 基本概念

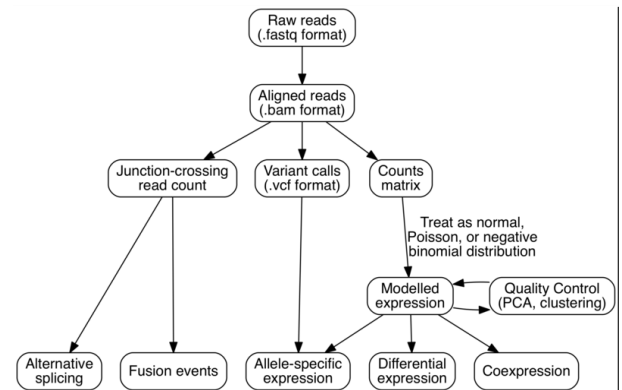
- RNA: mRNA(intron, polyA), rRNA+tRNA(95%), ncRNA
- RNA-Seq: RNA \Rightarrow cDNA \Rightarrow DNA 测序
- RNA-Seq 技术: Poly(A) library, reverse transcription, size selection, ……

2. 转录本组装 (以拼图进行类比)

- genome-guided: easier, computationally cheaper
- de novo: difficult

3. 应用 (结合实例进行讲解)

- Gene expression: Measuring mRNA concentration levels
- Differential expression: DGE
- Absolute quantification of transcripts
- SNV discovery: RNA-Seq vs. DNA-Seq
- Post-transcriptional edits
- Fusion gene detection
- Alternative transcript usage
- Coexpression networks
-



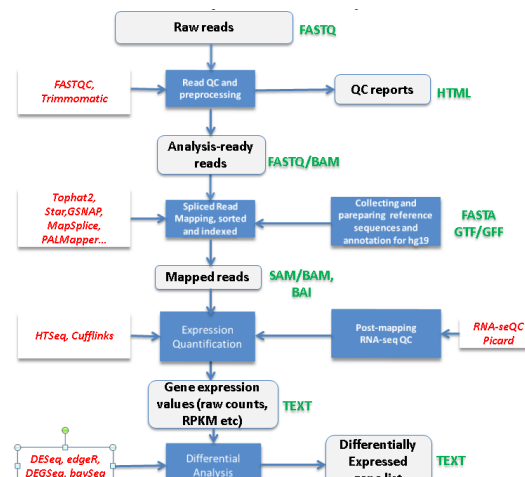
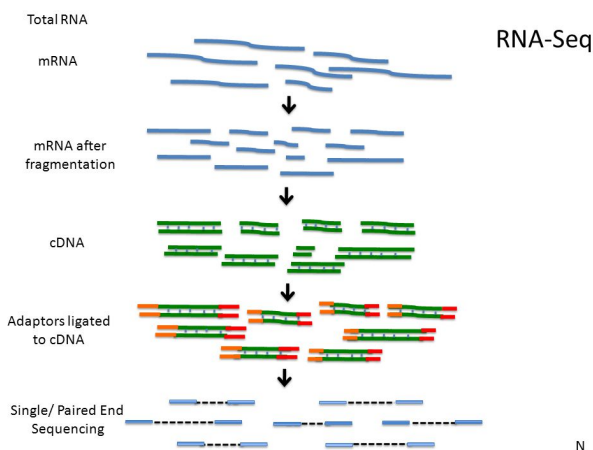
4. 数据库

- ENCODE: Encyclopedia of DNA Elements
- TCGA: The Cancer Genome Atlas

四、RNA-Seq 流程 (30 分钟)

1. 【重点】主要流程 (与外显子组测序进行比较)

- 实验: Purify RNA \Rightarrow Deplete rRNA/Select mRNA \Rightarrow Fragment \Rightarrow cDNA \Rightarrow Add adapter \Rightarrow Size selection \Rightarrow PCR \Rightarrow Sequencing
- 分析: QC \Rightarrow Pre-processing \Rightarrow Mapping \Rightarrow Assembly \Rightarrow Counting \Rightarrow Normalizing \Rightarrow DE \Rightarrow Annotation \Rightarrow Visualization



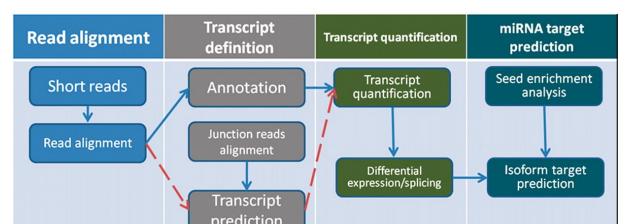
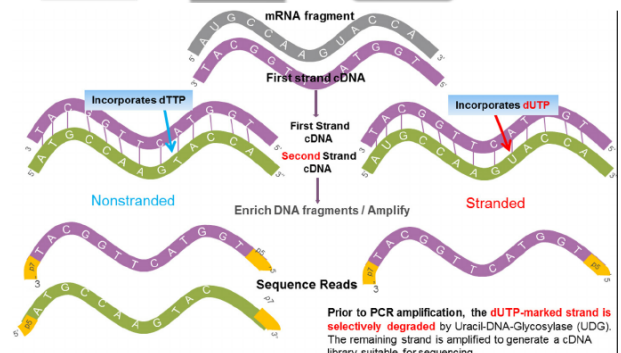
2. 相关技术

- Single cell RNA sequencing
- Stranded sequencing
- miRNA sequencing

3. 【重点, 难点】专用术语 (通过简单的例子进行概念讲解, 并应用公式进行实际演算; 强调每种方法的适用范围、优缺点及相互之间的关系)

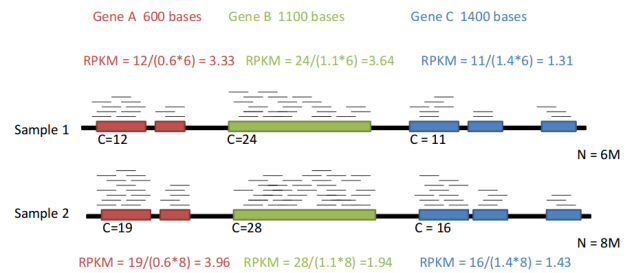
- RPKM: Reads Per Kilobase Million
- FPKM: Fragments Per Kilobase Million
- TPM: Transcripts Per Kilobase Million
- RPKM vs. FPKM: SE vs. PE

$$TPM_i = \left(\frac{FPKM_i}{\sum_j FPKM_j} \right) \cdot 10^6$$



$$RPKM = \frac{\text{total exon reads}}{\text{mapped reads (millions)} * \text{exon length (KB)}}$$

$$FPKM = \frac{\text{total fragments}}{\text{mapped reads (millions)} * \text{exon length (KB)}}$$



Normalization: different goals

- **R/FPKM:** (Mortazavi et al. 2008)
 - **Correct for:** differences in sequencing depth and transcript length
 - **Aiming to:** compare a gene across samples and diff genes within sample
- **TMM:** (Robinson and Oshlack 2010)
 - **Correct for:** differences in transcript pool composition; extreme outliers
 - **Aiming to:** provide better across-sample comparability
- **TPM:** (Li et al 2010, Wagner et al 2012)
 - **Correct for:** transcript length distribution in RNA pool
 - **Aiming to:** provide better across-sample comparability
- **Limma voom (logCPM):** (Law et al 2013)
 - **Aiming to:** stabilize variance; remove dependence of variance on the mean

RPKM / FPKM / TPM

- **RPKM** (Reads per kilobase of transcript per million reads of library)
 - Corrects for total library coverage
 - Corrects for gene length
 - Comparable between different genes within the same dataset
- **FPKM** (Fragments per kilobase of transcript per million reads of library)
 - Only relevant for paired end libraries
 - Pairs are not independent observations
 - RPKM/2
- **TPM** (transcripts per million)
 - Normalises to transcript copies instead of reads
 - Corrects for cases where the average transcript length differs between samples

五、总结与答疑 (5 分钟)

1. 知识点

- 组学：常见组学
- 转录组学：基本概念，研究内容，研究方法
- RNA-Seq：基本概念，常见应用，主要流程，专用术语

2. 技能

- 掌握 RNA-Seq 数据分析的基本流程
- 掌握 RPKM/FPKM/TPM 的计算方法