

天津医科大学实验课教案首页

(共 2 页、第 1 页)

课程名称：系统生物学		实验名称：实验一 测序数据的质量控制与预处理	
教师姓名：伊现富	职称：讲师	教学日期：2017 年 3 月 8 日 13:30-16:30	
授课对象：生物医学工程与技术学院 2014 级生信班（本）		实验人数：30	
实验类型（验证型、综合型、设计型、创新型）：验证型		实验分组：一人一机	
学时数：3		教材版本：系统生物学实验讲义（自编教材）	

实验目的与要求：

- 掌握二代测序数据的 FASTQ 格式。
- 熟悉 FastQC 和 FASTX-Toolkit 等工具的使用方法。
- 熟悉 Galaxy 的使用方法。
- 了解 FastQC 输出结果的含义。

实验内容及学时分配：

- (10') FASTQ 格式：回顾存储二代测序数据的 FASTQ 格式。
- (10') 质控与预处理：回顾二代测序数据分析流程中质控与预处理。
- (10') FastQC 和 FASTX-Toolkit：回顾 FASTQC 和 FASTX-Toolkit 的基本用途。
- (120') 实验操作：对以 FASTQ 格式存储的二代测序数据进行质量控制与预处理。

主要仪器和实验材料：

- 实验材料：以 FASTQ 格式存储的二代测序数据。
- 主要仪器：联网的计算机。
- 分析工具：Galaxy, FastQC, TASTX-Toolkit。

实验重点、难点及解决策略：

- 难点：FASTQ 格式；解决策略：通过实例进行讲解。
- 重点：FastQC 和 FASTX-Tollkit 的使用；解决策略：根据资料进行学习，通过练习熟练掌握。

思考题：

- 解释存储二代测序数据的 FASTQ 格式。
- 列举二代测序数据质控和预处理的主要内容。
- 如何对二代测序数据进行质控和预处理？

参考资料：

- FastQC
- FASTX-Toolkit
- Galaxy

主任签字：

年 月 日

教务处制

一、FASTQ 格式 (10 分钟)

Extension from traditional FASTA format

Each block has 4 elements (in 4 lines):

- Sequence name (read name, group, etc...)
- Sequence
- + (optional: sequence name again)
- Associated quality scores (phred-scaled) : different encoding possible

Example record:

- @FCD19MJACXX:2:1101:1735:1993#GTTTCGACA/1
- NGAGGCTGAGGCGGGCAGAGGTCAGGAGATCGAGACCATC
- +
- BP\cccc\cecheheeZbe_cZbd_dbbdd\Xab_`b



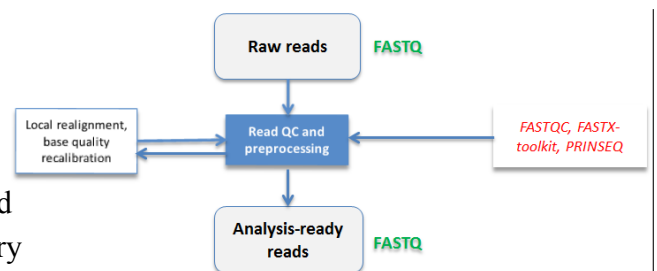
二、质控与预处理 (10 分钟)

1. 质量控制

- Identify poor/bad sample
- Identify contaminants

2. 预处理

- Trimming: remove bad bases from read
- Filtering: remove bad reads from library



三、FastQC 和 FASTX-Toolkit (10 分钟)

1. FastQC: quality control tool

- Basic statistics
- Per base sequence quality
- Per tile sequence quality
- Per sequence quality content
- Per base sequence content
- ...

2. FASTX-Toolkit: FASTA/FASTQ preprocessing

- Remove linker/adaptor sequences
- Trim low quality reads at the end of the read
- Filter sequences based on quality
- ...

四、实验操作 (120 分钟)

1. Upload data to Galaxy (比较导入数据的不同方法; 注意参数的设定)
2. Checking read quality with FastQC (理解输出结果中每一部分的含义)
3. Convert FASTQ quality to sanger (为什么要进行质量编码的转换)
4. Preprocessing with FASTX-Toolkit (结合 FastQC 的结果进行预处理)
5. Clean adapter containing reads from FASTQ data (结合 FastQC 的结果进行预处理)
6. 探索 “NGS: QC and manipulation” 中的其他工具 (同一个任务可以选用不同的工具)