

天津医科大学实验课教案首页

(共 2 页、第 1 页)

课程名称：生物信息学	实验名称：基于 Galaxy 的基因组数据处理	
教师姓名：伊现富	职称：讲师	教学日期：2013 年 9 月 18 日 14 时 -16 时
授课对象：生物医学工程学院 2010 级生信班（本）	实验人数：23	
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：2	教材版本：生物信息学实验讲义（自编教材）	

实验目的与要求：

- 掌握基因组注释中常用的 BED 格式。
- 掌握基因组坐标的逻辑运算模式。
- 掌握 Galaxy 的基本使用方法。

实验内容及学时分配：

- (5') BED 格式：回顾 BED 格式使用的坐标系统及其每一列的含义。
- (5') 逻辑运算模式：回顾交集、减法、联合等逻辑运算模式。
- (5') Galaxy 简介：简单介绍 Galaxy 分析平台的主界面、工具集及学习资料。
- (85') 实验操作：寻找人类基因组中 22 号染色体上含有 SNP 数目最多的外显子。

主要仪器和实验材料：

- 实验材料：人类基因组（hg19）中 22 号染色体（chr22）上的外显子和 SNP。
- 主要仪器：联网的计算机。
- 分析工具：Galaxy。

实验重点、难点及解决策略：

- 难点：基因组坐标的联合运算；解决策略：通过实例进行讲解。
- 重点：Galaxy 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：

- BED 格式使用的哪一类坐标系统？
- BED 格式每一列的含义是什么？
- 如何进行基因组坐标的联合操作？

参考资料：

- Galaxy

主任签字：

年 月 日

教务处制

一、BED 格式 (5 分钟)

1. 坐标系统: 0-based (与 1-based 的区别)
2. 每列含义: chrom, chromStart, chromEnd, name, score, strand

二、逻辑运算模式 (5 分钟)

1. intersect, 交集: 保留重叠的坐标
2. subtract, 减法: 去除重叠的坐标
3. join, 联合: 根据坐标重叠把两组记录对应起来 (与交集的关系)

三、Galaxy 简介 (5 分钟)

1. 主界面: 顶部是刊头, 左侧栏是工具菜单, 中间是工作区, 右侧栏是历史面板
2. 工具集: Get Data, Operate on Genomic Intervals, Join, Subtract and Group, Filter and Sort
3. 学习资料: Galaxy 101, Galaxy Screencasts and Demos, Learn Galaxy, Galaxy Wiki

四、实验操作 (85 分钟)

寻找人类基因组 (hg19) 中 22 号染色体 (chr22) 上含有 SNP 数目最多的外显子。

1. 获取数据 (选择正确的格式, 把结果导出到 Galaxy 中)
 - 外显子数据: Get Data, UCSC Main, hg19, chr22, RefSeq Genes, BED 格式
 - SNP 数据: Get Data, UCSC Main, hg19, chr22, dbSNP137, BED 格式
2. 提取含有 SNP 的外显子: Operate on Genomic Intervals, Join (注意数据集的顺序)
3. 对外显子上的 SNP 进行计数: Join, Subtract and Group, Group (注意选择正确的列)
4. 对 SNP 数目进行排序: Filter and Sort, Sort (注意选择正确的列)
5. 筛选至少含有 10 个 SNP 的外显子: Filter and Sort, Filter (学习编写筛选表达式)
6. 附加外显子的原始信息: Join, Subtract and Group, Join two Datasets (注意数据集的选择, 同时根据每一列的含义选择正确的列)