# 《生物信息学概论》课程作业安排

主讲教师：李梢

助教：祖松鹏, 张鹏

目录

[1.作业管理方式： 1](#_Toc430724126)

[1.1 提交方式 1](#_Toc430724127)

[1.2 代码规范 1](#_Toc430724128)

[2.习题课互动形式 2](#_Toc430724130)

[3.作业内容安排 2](#_Toc430724131)

## 1.作业管理方式：

### 1.1 提交方式

利用目前较为流行的版本管理体系Github：作业在网络学堂布置，学生提交代码到Github；

### 1.2 代码规范

1. 制定简单的R和Python语言基本代码规范，使其养成按照规范编写代码的习惯；
2. 代码需满足我们布置时所必需的接口，以便进行代码检测；
3. 作业不允许抄袭；

## 2.习题课互动形式

1. 在保证每组人数相当的条件下，自由分为三组。
2. 互动以文献报告交流的方式为主，第2-5次课，每组每次15分钟，即每次习题课的一节课时间用来文献报告与交流。
3. 文献主要分为三个方面：
   1. Genomic 与Phenotype 关系
      1. GWAS Study
      2. ENCODE
      3. Integrating multiple genomic data
   2. Sequence Analysis
      1. 多序列比对： ProfileHMM
      2. Motif Discovery: Gibbs Sampler
      3. Sequence Analysis : Euler Graph
   3. Network 方面
      1. Tang Chao：网络动力学模拟
      2. Barabasi: Network Control
      3. Shao Li:TCM Network
4. 细节要求
   1. 。。。
   2. 由组内同学推荐代表来报告，或者教师选择。。。
   3. 每次展示，建议组内同学轮流，或者几个人一起
   4. 课程的期末报告，结合组内阅读，每人撰写review（推荐英文）—— 大作业。
   5. 评分：。。。

## 3.作业内容安排

### 3.1第一节课

1. 建立Github 帐号并熟悉Github 平台；

2. 熟悉R或Python语言，并了解基本代码规范；

3. 布置各组第二次习题课的文献阅读材料。

### 3.2第二节课（作业1）

1. 给定两段较长的DNA序列，编写S-W比对算法，进行序列比对，并给出比对分数的显著性；

2. 给定某一蛋白质，利用blastp，完成多物种间同一蛋白质的序列比对，从而构建物种进化树；

3. 布置文献阅读报告。

### 3.3第三节课（作业2）

1. 编写主成分分析的程序，在给定的一组病人中，利用其mRNA表达数据，miRNA数据，进行聚类分析。

2. 在上述聚类中，尝试不同的聚类策略，不同的Kernel-PCA降维方法，对结果的影响，并给出聚类图谱。

3. 布置文献阅读报告。

### 3.4第四节课（作业3）

1. 提供疾病或药物数据，利用网络方法，预测疾病治病基因等。

2. 布置文献阅读报告。

### 3.5第五节课

1. 布置课程总结与建议报告。
2. 。。。