

Título do curso Métodos em Bioinformática

RIB0307

2º Semestre

Quinta-feira 14:00 - 18:00 (14: 00-15: 00 T / 15:00 - 18:00 P)

Local: LED Sala 3 (FMRP)

Instrutor: Professor Houtan Noushmehr (houtan@usp.br)

Descrição do Curso

Este curso se concentrará na análise e visualização de dados de microarray . O objetivo geral é apresentar aos alunos as várias técnicas e questões envolvidas na análise de dados de expressão multivariada de alta dimensionalidade. Exemplos serão retirados da literatura. Além disso, os alunos visualizarão os resultados em um moderno software de criação de scripts estatísticos. A R. Topics inclui:

- * Utilizando pacotes e bibliotecas para executar funções e plotar dados em R.
- * Familiarizar-se com questões de inferência estatística e métodos estatísticos modernos para análise de dados biológicos.
- * Detectar e atribuir fontes de variabilidade de dados (separar sinal de ruído) e normalizar essas diferenças com uma introdução a alguns métodos comumente empregados.
- * Determinar genes diferencialmente expressos com testes estatísticos relevantes e controlar a descoberta de falsos positivos.
- * Uma introdução aos métodos lineares e não lineares de redução de dimensões.
- * Uma introdução ao reconhecimento de padrões comuns (clustering) e técnicas de classificação.

Horário de palestras

<u>Palestra</u>	<u>Tópico</u>
Palestra 1	Introdução à tecnologia de expressão gênica / Introdução ao R
Aula 2	Introdução ao R (cont.)
Aula 3	Visualizações de dados, outliers e dados ausentes
Aula 4	Replicar análise e poder
5ª Palestra	Normalização e Biocondutor
Aula 6	Deteção de genes diferencialmente expressos
Aula 7	Vários ajustes de teste
Aula 8	Redução de dimensionalidade e análise de cluster
Aula 9	Classificação
Palestra 10	Análise do caminho

Distribuição de pontos

<u>atribuições</u>	<u>Pontos</u>
Trabalhos de laboratório	40 (5 / laboratório)
Tarefas de casa	105 (35 / atribuição)
Projeto final	65
Total	210

Políticas e Procedimentos do Curso

A primeira parte da aula será dedicada aos tópicos das palestras. Muitos dos tópicos virão diretamente da literatura. A segunda parte da aula exigirá a conclusão de uma tarefa de laboratório. Essas são tarefas relativamente curtas, projetadas para ajudar os alunos a se familiarizar com a linguagem R e implementar análises básicas de microarray . Muitas das funções serão dadas nas notas de aula, para que o laboratório se concentre em utilizar este código para resolver vários problemas. Cada aluno será responsável pela conclusão de um projeto final.

Livro (s) de texto

Não há texto obrigatório para este curso, no entanto, algumas boas referências para tópicos específicos de aula estão listadas abaixo.

A Análise de Dados de Expressão Gênica , Giovanni Parmigiani (Editor), Elizabeth S. Garrett (Editora), Rafael A. Irizarry (Editor), Scott L. Zeger

Análise Estatística de Dados de Microarray de Expressão Gênica , TP Speed (Editor)

Um Guia de Biólogos para Análise de DNA Microarray Data , Steen Knudsen (Autor)

Estatística Introdutória com R , Peter Dalgaard

Obtenção de conjuntos de dados (alternativa ao acesso ao site do curso)

Faça o download dos arquivos de texto dos seguintes sites Golub et al. Dados AML / ALL (excluir chamadas de A / P) -dados do Affymetrix http://www.broad.mit.edu/cgi-bin/cancer/publications/pub_paper.cgi?mode=view&paper_id=43 links: conjunto de dados de trem (texto) ; Teste o conjunto de dados (texto)

Eisen et al. Dados de linfoma de grandes células B (valores de NA) - dados de cDNA

<http://rana.lbl.gov/EisenData.htm>

links: dados de expressão gênica

Alon et al. Dados cancro do cólon - Affymetrix dados

<http://www.sph.uth.tmc.edu/hgc/Downloads.asp>

links: GECOLON.DAT

Spellman et al. yeast cell cycle data (missing values)) - cDNA data

<http://genome-www.stanford.edu/cellcycle/data/rawdata/>

links: Tab delimited data

or

Download zip files from Bioconductor and install packages

Bioconductor (<http://www.bioconductor.org/>)

Under <Software>

<Experimental Data>

yeastCC: Spellman et al. yeast cell cycle data - cDNA data

golubEsets: Golub et al. AML/ALL data - Affymetrix data

colonCA: Alon et al. Colon Cancer data - Affymetrix data

in R, access data objects as follows:

load package from local drive (*all commands below are given in R*)

```
> library(yeastCC)
```

```
> library(Biobase)
```

```
> library(annotate)
```

```
> data(yeastCC)
```

```
> dat <- exprs(yeastCC)
```