

## Trabalho prático de

### Extração de Conhecimento de Bases de Dados Biológicas

O trabalho consta da análise de um conjunto de dados usando o programa R e os packages do Bioconductor (em grupos de 3 elementos). A constituição dos grupos deverá ser enviada por email para o docente das práticas até ao dia **2 de maio.**([scorreia@deb.uminho.pt](mailto:scorreia@deb.uminho.pt) ou [vvieira@ceb.uminho.pt](mailto:vvieira@ceb.uminho.pt))

Os dados a analisar deverão ser provenientes de um conjunto de dados de expressão genética, que poderá escolher de entre dados publicados na literatura ou procurar em bases de dados apropriadas (e.g. GEO do NCBI ou ArrayExpress do EBI). São fornecidos pelos docentes exemplos de conjuntos de dados já analisados em anos anteriores que servem para perceber o tipo de dados (estes não poderão ser escolhidos este ano).

Deverá desenvolver scripts em R/ Bioconductor para carregar os dados, pré-processá-los/ fazer a sua filtragem (se necessário) e fazer o conjunto de análises que lhe pareçam apropriadas. Deverá incluir, como requisito mínimo pelo menos dois dos três tipos de análise:

- análise de expressão diferencial e de enriquecimento
- clustering de genes e/ou amostras
- análise preditiva (e.g. classificação de amostras)

A apresentação dos resultados deverá incluir:

- explicação dos dados, sua origem e relevância
- apresentação do pipeline de análise e ferramentas usadas
- apresentação dos principais resultados obtidos
- discussão (podendo relacionar-se com a publicação de onde os dados originaram)

Deverá submeter no sistema de e-learning o relatório em formato PDF e ainda um ficheiro com os comandos R (extensão R) que realizou de forma a que a análise se possa reproduzir.

O relatório deverá ser submetido no **e-learning** até ao dia **23 de maio de 2017**. Na última aula, cada um dos grupos terá no máximo 15 minutos para efectuar a apresentação/discussão do seu trabalho.

Exemplos de datasets do GEO (não poderá usar estes):

**GDS3715** – Insulin effect on skeletal muscle

<https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS3715>

**GDS1429** - Various anti-cancer drugs effect on liver: time course and dose-response

<https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS1429>

**GDS5218** - Resistance exercise effect on skeletal muscles of young and old adults

<https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS5218>

**GDS1686** - Sleep deprivation effect on the brain: time course

<https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS1686>

**GDS5393** - Bipolar disorder and response to lithium: blood

<https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS5393>

**GDS4758** - Postmortem Alzheimer's disease brains: Hisayama study

<https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS4758>

**GDS5473**- Short-term fasting effect on skeletal muscle: time course

<https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS5473>

**GSE61326** - Dietary vitamin D3 supplementation effect on hippocampus: dose effect

<https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GSE61326>