



# Análise computacional e bioinformática de variantes em doença genética

Apresentação do curso

Luís Vieira, José Ferrão, Hugo Martiniano e Daniel Sobral

# Objectivos

- Aprender a utilizar diversas ferramentas computacionais e de bioinformática para processar e analisar dados de sequenciação (NGS) em contexto de doença genética. Pretende-se especificamente que os formandos:
  - Identifiquem e interpretem as variantes de linha germinativa em função das características clínicas e fenótipo do doente
  - Adquiram autonomia na análise de dados de variantes humanas, compreendendo as diferentes etapas do processo e os métodos informáticos utilizados
  - Adquiram “boas práticas” na análise dos dados e aprendam a olhar de forma crítica para os resultados obtidos
  - Relevem a importância da análise de métricas de qualidade nas diferentes etapas do processo
  - Obtenham confiança na utilização de sistemas de base Unix e da linha de comandos.

# Funcionamento do curso

- O curso baseia-se no estudo do **exoma clínico de 3 casos de doença genética**, iniciando-se com a análise do *raw data* da sequenciação e terminando com a priorização das variantes (SNPs e pequenos indels) anotadas
- Os formandos devem organizar-se em grupos de 2, sendo que a cada grupo será atribuído 1 “caso clínico”, que poderá ser o caso índice ou um trio
- Cada grupo efectuará o processamento e análise dos dados para o caso clínico respectivo, tendo como objectivo obter no final um diagnóstico molecular
- No último dia, cada grupo fará uma breve apresentação dos resultados obtidos.



# Calendarização

## Dia 1: 30 Janeiro

Programa	Início	Fim
<b>Apresentação e introdução ao curso</b>	9.00h	9.15h
<b>(TP) Tutorial Unix e linha de comandos</b>	9.15h	10.30h
<b>Intervalo (coffee-break)</b>	10.30h	11.00h
<b>(TP) Tutorial Unix e linha de comandos</b>	11.00h	12.30h
<b>Intervalo (almoço)</b>	12.30h	14.00h
<b>(T) Tecnologias e aplicações de sequenciação em genética humana</b>	14.00h	15.00h
<b>Visita ao laboratório de sequenciação</b>	15.00h	16.00h

## Dia 2: 31 Janeiro

Programa	Início	Fim
<b>(T) Análise e interpretação de métricas de qualidade</b>	9.00h	9.30h
<b>(P) Avaliação de qualidade de amostras reais</b>	9.30h	10.30h
<b>Intervalo (coffee-break)</b>	10.30h	11.00h
<b>(TP) Mapeamento de leituras em genoma de referência</b>	11.00h	12.30h
<b>Intervalo (almoço)</b>	12.30h	14.00h
<b>(TP) Mapeamento de leituras em genoma de referência</b>	14.00h	16.00h

# Calendarização

## Dia 3: 1 Fevereiro

Programa	Início	Fim
(TP) Análise de qualidade e visualização do mapeamento	9.00h	10.30h
Intervalo (coffee-break)	10.30h	11.00h
(TP) Recalibração de valores de qualidade e marcação de duplicados	11.00h	12.30h
Intervalo (almoço)	12.30h	14.00h
(T) Chamada, filtragem e avaliação de qualidade de variantes	14.00h	14.30h
(P) Chamada e filtragem de variantes	14.30h	16.00h

## Dia 4: 2 Fevereiro

Programa	Início	Fim
(P) Análise de qualidade das variantes	9.00h	10.00h
(T) Anotação e priorização de variantes	10.00h	10.30h
Intervalo (coffee-break)	10.30h	11.00h
(P) Anotação e priorização de variantes	11.00h	12.30h
Intervalo (almoço)	12.30h	14.00h
(P) Análise e discussão de resultados	14.00h	15.00h
(T) Desenho de pipelines para análise reprodutível e controlada de dados de sequenciação	15.00h	15.30h
Conclusões e encerramento do curso	15.30h	16.00h