

## Dia 2 - hands on!

### Módulo 1 – Introdução ao formato e lógica de arquivos RNA-seq

#### Objetivos:

- Conhecer os principais arquivos utilizados na análise de RNA-seq.
- Entender a estrutura dos dados de entrada e saída.
- Interpretar uma matriz de contagens (input para DESeq2).

#### Formatos importantes

Tipo de arquivo	Extensão	Função
Sequência crua	<code>.fastq(.gz)</code>	Armazena leituras com qualidade
Alinhamento	<code>.sam</code> , <code>.bam</code>	Mapeamento das leituras no genoma
Anotação	<code>.gtf</code> , <code>.gff3</code>	Coordenadas de genes, éxons, etc.
Contagens	<code>.txt</code> , <code>.tsv</code> , <code>.csv</code>	Matriz com contagem de reads por gene

Leitura recomendada:

<https://www.ebi.ac.uk/training/materials/introduction-to-rna-seq-materials/intro-fundamentals/rna-seq-file-formats/>

---

title: "Controle de Qualidade com FastQC e MultiQC"

output: html\_document

---

#### ## Objetivo

Avaliar a qualidade de dados RNA-seq brutos usando FastQC e MultiQC, e aprender a interpretar os gráficos para decidir sobre o pré-processamento (como trimming de adaptadores).

---

#### ## O que é FastQC?

O **FastQC** é uma ferramenta de QC que analisa arquivos `.fastq` e gera relatórios gráficos com os seguintes indicadores:

### ### 1. **Per base sequence quality**

- Mostra a qualidade de cada base ao longo da leitura (Phred Score).
- Valores acima de 28 são considerados bons.
- Uma queda de qualidade nas últimas bases é comum.

### ### 2. **GC content**

- Avalia a distribuição de conteúdo GC.
- Deve ser aproximadamente normal e próxima ao valor esperado do organismo (ex: ~50% em humanos).
- Um pico secundário pode indicar contaminação.

### ### 3. **Adapter content**

- Detecta a presença de adaptadores (sequências técnicas).
- Presença alta indica a necessidade de **trimming** com ferramentas como Trimmomatic ou fastp.

---





## ## Rodando FastQC e MultiQC no Galaxy

### ### Etapas:

1. **Acesse:** <https://usegalaxy.org>
2. **Importe** arquivos `.fastq.gz` do seu experimento.
3. **Procure** por “FastQC” na barra de ferramentas à esquerda.
4. **Execute FastQC** para cada amostra (pode selecionar várias de uma vez).
5. Após completar, **rode** o “MultiQC” para compilar todos os relatórios em um só.

---

## ## Como interpretar os resultados

-  Se a **qualidade por base** for boa (>28) até o final, **não é necessário trimming**.
-  Se houver **queda acentuada nas últimas bases**, **recomenda-se trimming** das extremidades 3'.
-  Se houver **presença de adaptadores (>10-15%)**, **trimming com adaptador é essencial**.
-  Se o **GC Content** tiver dois picos ou desvio do esperado, verifique contaminação.

---

## ## Exemplo de saída do MultiQC

```
``{r echo=FALSE, out.width="90%"}
```

knitr::include\_graphics("multiqc\_example.png")

## Conclusão

- FastQC + MultiQC são ferramentas rápidas e visuais para QC de RNA-seq.
- Sempre revise os gráficos antes de iniciar o alinhamento.
- Dados de baixa qualidade devem ser tratados com trimming apropriado para evitar viés nos resultados downstream.

### SEGUIR EMENDA + EXERCÍCIOS DO GITHUB

[https://github.com/Kur1sutaru/transcriptomics\\_rsg\\_course](https://github.com/Kur1sutaru/transcriptomics_rsg_course)