

Explorando a Expressão Gênica

um guia prático com dados de RNA-Seq - Dia 1

Juliana Costa Silva^{1 2 3}

¹Departamento de Informática - Universidade Federal do Paraná - Campus Curitiba

²Escola de Tecnologia da Informação - Universidade Positivo - Campus Londrina

³Departamento de Computação - Universidade Estadual de Londrina

November 5, 2024

Apresentação

O mini-curso

Introdução

Dados de expressão, como registrar expressão de genes?

Análises com dados de RNA-Seq (expressão)

Expressão

Apresentação

Apresente-se

O mini-curso

Conteúdo

Programação

Introdução

Dados de expressão, como registrar expressão de genes?

Análises com dados de RNA-Seq (expressão)

Mapeamento

Contagem

Expressão

Apresentação

Apresente-se

O mini-curso

Introdução

Dados de
expressão, como
registrar
expressão de
genes?

Análises com
dados de
RNA-Seq
(expressão)

Expressão

Apresentação Apresente-se

Juliana Costa Silva

- **Graduação: Tecnologia em Análise e Desenvolvimento de Sistemas** pela UTFPR – Universidade Tecnológica Federal do Paraná (Cornélio Procópio) - 2011 a 2014;
- **Mestrado: Bioinformática** pela UTFPR – Universidade Tecnológica Federal do Paraná (Cornélio Procópio) - 2015 a 2017;
- **Doutorado: Doutorado em Informática** pela UFPR - Universidade Federal do Paraná (Curitiba) - 2020 a 2024.

Por favor, apresente-se.

Apresentação

Apresente-se

O mini-curso

Introdução

Dados de expressão, como registrar expressão de genes?

Análises com dados de RNA-Seq (expressão)

Expressão

- Seu nome?
- Sua instituição de ensino?
- Sua formação?
- Por que escolheu este mini-curso?



Apresentação

O mini-curso

Conteúdo

Programação

Introdução

Dados de
expressão, como
registrar
expressão de
genes?

Análises com
dados de
RNA-Seq
(expressão)

Expressão

O mini-curso

Conteúdo

Programação

Apresentação

O mini-curso

Conteúdo

Programação

Introdução

Dados de expressão, como registrar expressão de genes?

Análises com dados de RNA-Seq (expressão)

Expressão

- Conceito de expressão diferencial de genes;
- Como são gerados os dados de expressão;
- Métodos para identificar genes diferencialmente expressos;
- Análise de expressão com linguagem R;
- consexpressionR: pacote R para análise de expressão diferencial de genes.

[Apresentação](#)[O mini-curso](#)[Conteúdo](#)[Programação](#)[Introdução](#)[Dados de expressão, como registrar expressão de genes?](#)[Análises com dados de RNA-Seq \(expressão\)](#)[Expressão](#)

Programação

- 1 O que é expressão de genes?
- 2 Dados de expressão, como registrar expressão de genes?
- 3 RNA-Seq: dados e etapas de análise;
- 4 Tipos de experimentos e perguntas.

Programação

Utilizando o pacote experimental consexpressionR.

O que vou precisar?

- Computador com R e RStudio instalados;
 - Instalação do consexpressionR e suas dependências;
 - Conjunto de dados;
- 1 Dados de contagem, como importar?
 - 2 Métodos de análise de expressão: paramétricos e não paramétricos;
 - 3 Metodologia consexpressionR;
 - 4 Prática e perguntas.

Apresentação

O mini-curso

Introdução

Dados de
expressão, como
registrar
expressão de
genes?

Análises com
dados de
RNA-Seq
(expressão)

Expressão

Introdução

Apresentação

O mini-curso

Introdução

Dados de expressão, como registrar expressão de genes?

Análises com dados de RNA-Seq (expressão)

Expressão

- ❶ O que é expressão de genes?
- ❷ Dados de expressão, como registrar expressão de genes?
- ❸ RNA-Seq: dados e etapas de análise;
- ❹ Tipos de experimentos e perguntas.

Por que estudar expressão de genes?

Para compreender os mecanismos que geram algumas características fenotípicas, é necessário compreender o complexo mecanismo de regulação gênica.



Genótipo

+



Ambiente

=



Fenótipo

Apresentação

O mini-curso

Introdução

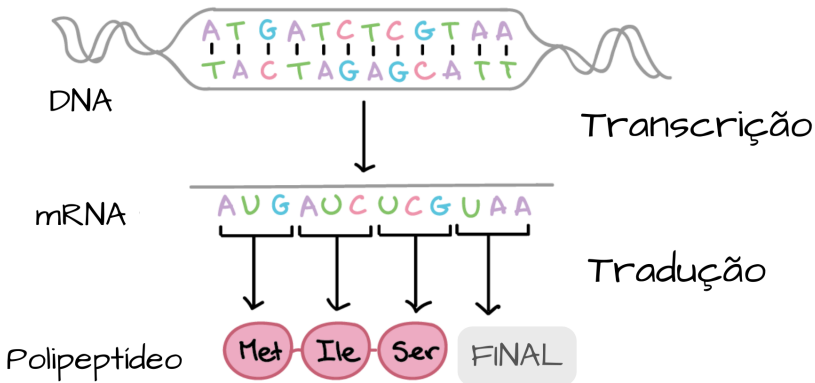
Dados de expressão, como registrar expressão de genes?

Análises com dados de RNA-Seq (expressão)

Expressão

Dogma Central

O DNA através do processo de transcrição faz o RNA, que na maior parte das vezes, gera apenas uma proteína através do processo de tradução.



De modo geral análises de expressão buscam identificar como a atividade transcricional de um gene interfere no fenótipo de um organismo.

Técnicas populares para identificar expressão

- 1 qRT-PCR;
- 2 MicroArray;
- 3 RNA-Seq;
- 4 Single-Cell RNA-Seq;

Apresentação

O mini-curso

Introdução

Dados de
expressão, como
registrar
expressão de
genes?

Análises com
dados de
RNA-Seq
(expressão)

Expressão

Dados de expressão, como registrar expressão de genes?

qRT-PCR: reação em cadeia da polimerase da transcrição reversa em tempo real (do inglês *real-time reverse transcription polymerase chain reaction*) [4].

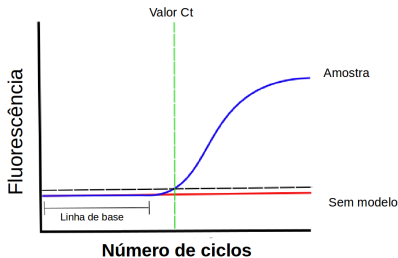


Fig. 3: Nível de limiar em uma curva de amplificação qRT-PCR. A linha vermelha indica o nível de fluorescência do controle utilizado (*Sem modelo*), a linha azul indica a curva de amplificação da sequência de interesse (*Amostra*), a linha tracejada em verde é o limiar dos ciclos da PCR (C_t). Fonte: Adaptado de [1]

O que é RNA-Seq?

Apresentação

O mini-curso

Introdução

Dados de expressão, como registrar expressão de genes?

Análises com dados de RNA-Seq (expressão)

Expressão

- Este nome foi dado a tecnologias de sequenciamento chamadas de "nova geração" (*Next-generation sequencing*) aplicadas aos transcriptomas, ou seja, às regiões do DNA transcritas

Utilidades

- Analisar expressão diferencial de diferentes condições ambientais ou tecidos;
- Analisar diferentes isoformas (*splicing* alternativo);
- Identificar regiões de transcrição em um genoma;
- Identificar moléculas de RNA que participam de processos regulatórios;
- Entre outros...

Apresentação

O mini-curso

Introdução

Dados de
expressão, como
registrar
expressão de
genes?

Análises com
dados de
RNA-Seq
(expressão)

Mapeamento
Contagem

Expressão

Análises com dados de RNA-Seq (expressão)

Mapeamento

Contagem

Descreveremos um experimento típico de análise de expressão diferencial de genes. Considere o seguinte cenário:

- Dados de RNA-Seq;
- Organismo investigado possui genoma sequenciado;
- Organismo investigado possui genoma anotado;

Os passos das análises de expressão dependem do objetivo do estudo, organismo investigado, disponibilidade de dados e amostras, entre outros. Desse modo, para outros cenários esses passos podem ser diferentes.

Análises RNA-Seq - sequenciamento

Apresentação

O mini-curso

Introdução

Dados de expressão, como registrar expressão de genes?

Análises com dados de RNA-Seq (expressão)

Mapeamento
Contagem

Expressão

Queremos avaliar a variação de expressão entre duas condições, controle e tratado (com alguma perturbação específica).

São sequenciadas amostras nas condições descritas.

Os *reads* resultantes do sequenciamento são alinhados com o genoma de referência ou transcriptoma.

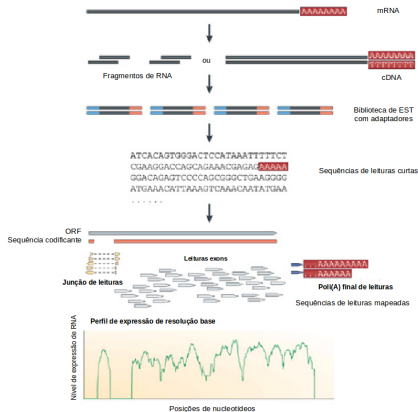


Fig. 4: Exemplo de um experimento típico de RNA-Seq. **Fonte:** Adaptado de [7]

Análises RNA-Seq - o experimento

Os *reads* são mapeados para identificar de qual região do genoma foram transcritos.

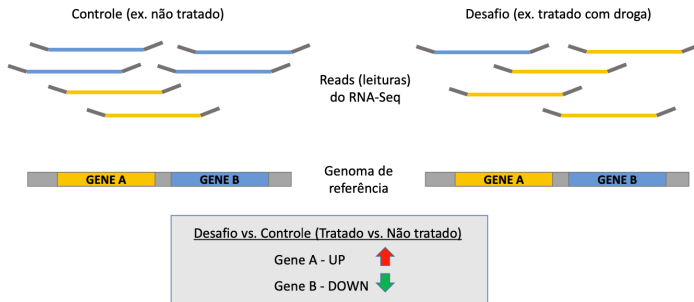


Fig. 5: Exemplo de *reads* gerados em um experimento típico de RNA-Seq, onde o objetivo é comparar expressão de genes na situação controle e tratado. **Fonte:** [6].

Os *reads* podem ser gerados de duas formas *single-end* e *paired-end*.

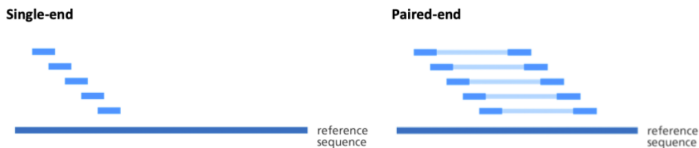


Fig. 6: Exemplo de *reads* gerados em um experimento típico de RNA-Seq, com bibliotecas *single-end* e *paired-end*. **Fonte:** [6].

Análise computacional (RNA-Seq)

Apresentação

O mini-curso

Introdução

Dados de expressão, como registrar expressão de genes?

Análises com dados de RNA-Seq (expressão)

Mapeamento
Contagem

Expressão

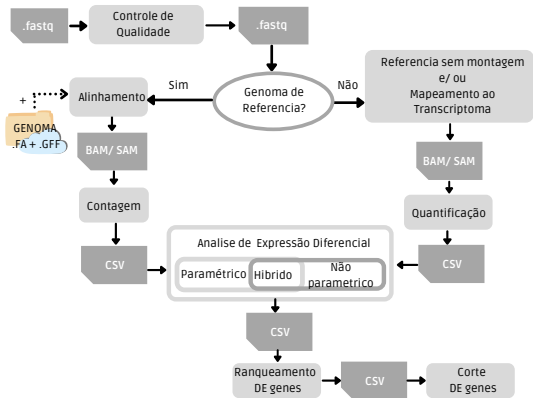


Fig. 7: Esquema com os principais passos da análise de expressão em massa com dados de RNA-Seq. **Fonte:** Autoria própria

Arquivos FASTA são utilizados como saída de sequenciadores e também é o formato em que comumente são disponibilizados genomas e transcriptomas.

```
>Sequence_1 assembly1
CCCTAAACCCTAAACCCTAAACCCTAAACCCTCTGAATCCTTAATCCCTAAATCCCTAAAT
CTTTAAATCCTACATCCATGAATCCCTAAATACCTAATTCCCTAAACCCGAAACCGGTTT
CTCTGGTTGAAAATCATTGTGTATATAATGATAATTTTATCGTTTTTATGTAATTGCTTA
TTGTTGTGTGTAGATTTTTTAAAAATATCATTTGAGGTCAATACAAATCCTATTTCTTGT
GGTTTCTTTCTTCACTTAGCTATGGATGGTTTATCTTCATTTGTTATATTGGATACAA
GCTTTGCTACGATCTACATTTGGGAATGTGAGTCTCTTATTGTAACCTTAGGGTTGGTTT
ATCTCAAGAATCTTATTAATTGTTTGGACTGTTTATGTTTGGACATTTATTGTCATTCTT

>Sequence_2
CCCTAAACCCTAAACCCTAAACCCTAAACCCTCTGAATCCTTAATCCCTAAATCCCTAAAT
CTTTAAATCCTACATCCATGAATCCCTAAATACCTAATTCCCTAAACCCGAAACCGGTTT
CTCTGGTTGAAAATCATTGTGTATATAATGATAATTTTATCGTTTTTATGTAATTGCTTA
TTGTTGTGTGTAGATTTTTTAAAAATATCATTTGAGGTCAATACAAATCCTATTTCTTGT
GGTTTCTTTCTTCACTTAGCTATGGATGGTTTATCTTCATTTGTTATATTGGATACAA
GCTTTGCTACGATCTACATTTGGGAATGTGAGTCTCTTATTGTAACCTTAGGGTTGGTTT
ATCTCAAGAATCTTATTAATTGTTTGGACTGTTTATGTTTGGACATTTATTGTCATTCTT
```

Fig. 8: Exemplo de duas sequências em um arquivo FASTA: Cada sequência é representado por um marcador de início da sequência > + ID da sequência, na linha abaixo ao ID a sequência.

Sequenciadores podem gerar dados em diversos formatos, o formato FASTQ informa a sequência e a qualidade de cada par de base sequenciado, o valor de qualidade pode ser utilizado para "limpeza", onde excluimos os *reads* de baixa qualidade.

Formato FASTQ.

```
@SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36
GGGTGATGGCCGCTGCCGATGGCGTCAAATCCCACC
+SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII9IG9IC
```

Fig. 9: Read em um arquivo FASTQ: Cada read é representado por 4 linhas: ID, sequência, ID, Qualidade da sequência.

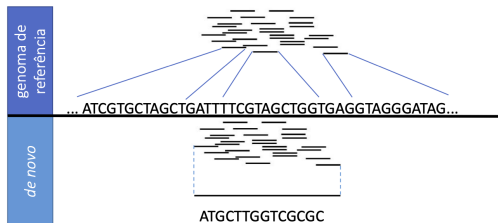


Fig. 10: Esquema de processo de mapeamento. O mapeamento tem como objetivo recuperar a região do genoma, ou transcrito que originou cada um dos reads gerados pelo sequenciador.

Análises RNA-Seq - contagem

Apresentação

O mini-curso

Introdução

Dados de expressão, como registrar expressão de genes?

Análises com dados de RNA-Seq (expressão)

Mapeamento

Contagem

Expressão

Para mapeamentos de organismos com genoma anotado, é necessário fornecer o arquivo de anotação, formato GFF ou GTF.

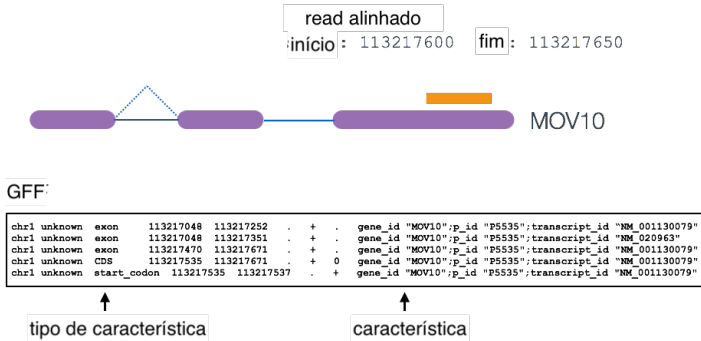


Fig. 11: Esquema mostrando região indicada no mapeamento, em laranja o read posicionado na região onde foi mapeado em relação ao gene MOV10. Abaixo o arquivo de anotação que indica qual região do genoma representa qual gene. **Fonte:** Adaptado de[3].

Apresentação

O mini-curso

Introdução

Dados de expressão, como registrar expressão de genes?

Análises com dados de RNA-Seq (expressão)

Mapeamento
Contagem

Expressão

gene	UHR1	UHR2	UHR3	UHR4	Brain1	Brain2	Brain3	Brain4
ENSG00000242268.2	0	0	0	1	0	0	2	1
ENSG00000167578.12	22	37	16	39	110	93	76	65
ENSG00000270112.2	0	1	1	0	28	20	16	24
ENSG00000078237.4	24	20	23	21	8	7	11	9
ENSG00000269416.1	33	25	29	22	1	1	1	0
ENSG00000268332.1	7	6	8	5	22	16	25	12

Fig. 12: Tabela de contagem. **Fonte:** [3].

Apresentação

O mini-curso

Introdução

Dados de
expressão, como
registrar
expressão de
genes?

Análises com
dados de
RNA-Seq
(expressão)

Expressão

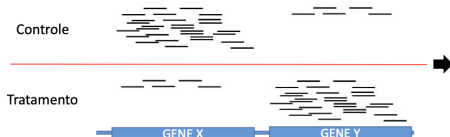
Expressão

Identificando a Expressão Diferencial

[Apresentação](#)[O mini-curso](#)[Introdução](#)[Dados de expressão, como registrar expressão de genes?](#)[Análises com dados de RNA-Seq \(expressão\)](#)[Expressão](#)

Para identificar genes diferencialmente expressos é necessário calcular a relevância da diferença de contagem.

Se um gene é mais extenso em quantidade de pares de base, teremos mais reads mapeados na região dele, o que não implica necessariamente em uma expressão diferencial.



	Controle	Tratamento	FoldChange	log ₂ FC
Gene X	28	5	0,18	-2,49
Gene Y	5	28	5,60	2,49

Apresentação

O mini-curso

Introdução

Dados de expressão, como registrar expressão de genes?

Análises com dados de RNA-Seq (expressão)

Expressão



S. A. Bustin, V. Benes, J. A. Garson, J. Hellemans, J. Huggett, M. Kubista, R. Mueller, T. Nolan, M. W. Pfaffl, G. L. Shipley, J. Vandesompele, and C. T. Wittwer.

The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments.

Clinical Chemistry, 55(4):611–622, 4 2009.



I. DGLab.

Genótipo e fenótipo: entenda o que são e como se relacionam, 2021.



hbctraining.

Introduction to rna-seq using high-performance computing - archived, 2018.



P. R. S. d. Ladeira, C. Isaac, and M. C. Ferreira.

Reação em cadeia da polimerase da transcrição reversa em tempo real.

Revista de Medicina, 90(1):47, 3 2011.



J. Ponomarenko, S. Bonnin, L. Cozzuto, and T. Hermoso.

Bioinformatics for Biologists, 2021.



P. R. Sanches.

Conceitos fundamentais sobre a metodologia para análise de rna-seq, 2022.



Z. Wang, M. Gerstein, and M. Snyder.

RNA-Seq: a revolutionary tool for transcriptomics.

Nature Reviews Genetics, 10(1):57–63, 1 2009.