

Contagem das leituras mapeadas no genoma de referência

Dr. Pablo Rodrigo Sanches

Departamento de Genética – FMRP/USP

psanches@usp.br

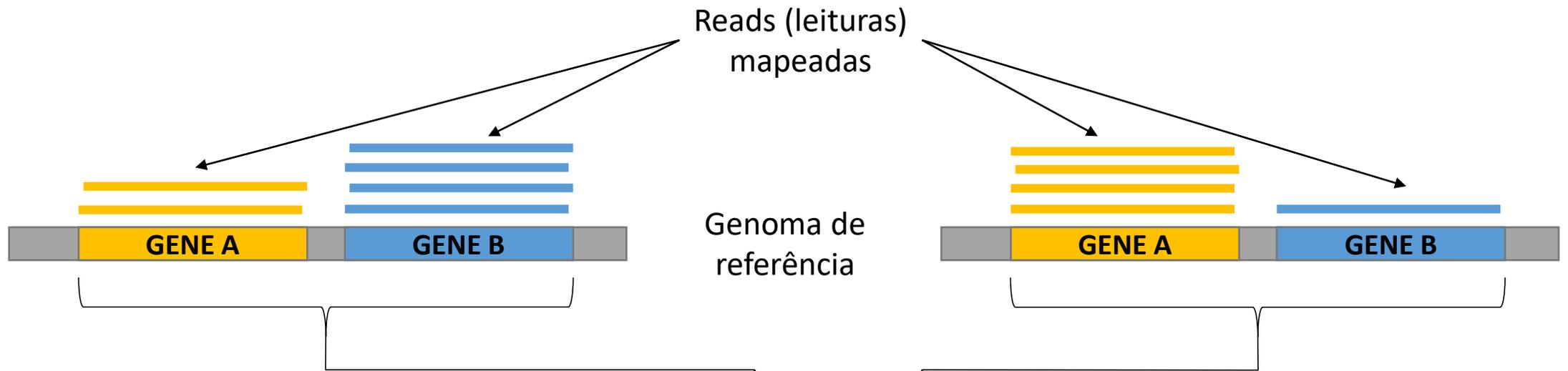
Roteiro de análise

1. StringTie (quantificação da transcrição)
2. Obter arquivo de contagem das leituras mapeadas por Gene

Introdução

Controle (ex. não tratado)

Desafio (ex. tratado com droga)



Reads (leituras)
mapeadas

Genoma de
referência

Contagem
das reads

	Controle	Desafio
Gene A	2	4
Gene B	4	1

Renomear arquivos mapped.bam no Galaxy (saída do software RNA STAR)

The screenshot displays the Galaxy web interface. On the left, a sidebar lists tool categories such as 'GENERAL TEXT TOOLS', 'GENOMIC FILE MANIPULATION', and 'COMMON GENOMICS TOOLS'. The main content area features a large graphic for the 'JXTX James P. Taylor Foundation' with text explaining its mission and a 'Donate Now' button. On the right, a 'History' sidebar shows a list of datasets. The dataset '124: RNA STAR on data 106, data 105, and other s: mapped.bam' is highlighted with a red box, and an arrow points to its 'Edit attributes' button. The URL at the bottom is https://usegalaxy.org/datasets/edit?dataset_id=bbd44e69cb8906b511acdcc987ab09f6.

Arquivo .bam no Galaxy

The screenshot displays the Galaxy web interface. The main area shows a list of datasets with columns for name, size, and other attributes. The left sidebar contains a 'Tools' menu with categories like 'GENERAL TEXT TOOLS', 'GENOMIC FILE MANIPULATION', and 'COMMON GENOMICS TOOLS'. The right sidebar shows a 'History' panel with a search box and a list of recent jobs, including RNA STAR and Trimmomatic runs.

Dataset Name	Size	Other Attributes
HWI-ST697:173:D1N1UACXX:2:2109:2484:26168	345	Supercontig21, 350, 0, 12M, *
HWI-ST697:173:D1N1UACXX:2:2215:20629:9058	419	Supercontig21, 357, 0, 11S9M, =
HWI-ST697:173:D1N1UACXX:2:2305:19059:3686	355	Supercontig21, 822, 0, 8M10S, =
HWI-ST697:173:D1N1UACXX:2:1211:1655:40186	419	Supercontig21, 1127, 1, 8M11S, =
HWI-ST697:173:D1N1UACXX:2:1108:8261:38044	163	Supercontig21, 1739, 60, 9M18774N11M, =
HWI-ST697:173:D1N1UACXX:2:1310:13300:86402	419	Supercontig21, 1801, 1, 18M, =
HWI-ST697:173:D1N1UACXX:2:1311:5125:43902	419	Supercontig21, 1809, 0, 10M268404N40M, =
HWI-ST697:173:D1N1UACXX:2:1311:12408:32050	419	Supercontig21, 1813, 0, 7M269900N43M, =
HWI-ST697:173:D1N1UACXX:2:2213:9496:34003	393	Supercontig21, 1831, 0, 12M, *
HWI-ST697:173:D1N1UACXX:2:2206:16370:89706	419	Supercontig21, 2803, 0, 3S16M, =
HWI-ST697:173:D1N1UACXX:2:1314:11983:57787	419	Supercontig21, 3594, 0, 9S10M, =
HWI-ST697:173:D1N1UACXX:2:2314:13926:9017	89	Supercontig21, 3769, 0, 12M, *
HWI-ST697:173:D1N1UACXX:2:1108:2208:24362	419	Supercontig21, 3900, 0, 12S7M, =
HWI-ST697:173:D1N1UACXX:2:1101:10540:19823	99	Supercontig21, 4396, 60, 2S11M229649N37M, =
HWI-ST697:173:D1N1UACXX:2:1108:7798:9177	163	Supercontig21, 4396, 60, 5S11M229649N34M, =
HWI-ST697:173:D1N1UACXX:2:1205:3211:48668	99	Supercontig21, 4396, 60, 9S11M229649N30M, =
HWI-ST697:173:D1N1UACXX:2:1304:20379:29169	99	Supercontig21, 4396, 60, 4S11M229649N35M, =
HWI-ST697:173:D1N1UACXX:2:2103:1699:13193	99	Supercontig21, 4396, 60, 5S11M229649N34M, =
HWI-ST697:173:D1N1UACXX:2:2111:17386:80876	163	Supercontig21, 4396, 60, 3S11M229649N36M, =
HWI-ST697:173:D1N1UACXX:2:2215:2781:15226	163	Supercontig21, 4396, 60, 2S11M229649N37M, =
HWI-ST697:173:D1N1UACXX:2:2215:2881:65232	163	Supercontig21, 4412, 60, 14M216471N9M, =
HWI-ST697:173:D1N1UACXX:2:1311:2286:53130	419	Supercontig21, 4475, 3, 13S9M, =
HWI-ST697:173:D1N1UACXX:2:1308:13748:31466	329	Supercontig21, 4526, 0, 1S13M, *
HWI-ST697:173:D1N1UACXX:2:2206:16370:89706	419	Supercontig21, 5315, 0, 11S8M, =
HWI-ST697:173:D1N1UACXX:2:1104:9639:20843	345	Supercontig21, 5780, 0, 12M, *
HWI-ST697:173:D1N1UACXX:2:1311:12408:32050	419	Supercontig21, 6907, 0, 7M264806N43M, =
HWI-ST697:173:D1N1UACXX:2:1211:3095:11645	419	Supercontig21, 7365, 0, 9M11S, =
HWI-ST697:173:D1N1UACXX:2:2215:18513:69680	419	Supercontig21, 7568, 1, 13S10M, =
HWI-ST697:173:D1N1UACXX:2:1312:11366:84109	393	Supercontig21, 7813, 0, 13M, *
HWI-ST697:173:D1N1UACXX:2:2106:3714:12029	99	Supercontig21, 7944, 60, 50M, =
HWI-ST697:173:D1N1UACXX:2:2315:8406:67892	99	Supercontig21, 7944, 60, 50M, =
HWI-ST697:173:D1N1UACXX:2:1315:18038:38905	163	Supercontig21, 8041, 60, 12S7M, =
HWI-ST697:173:D1N1UACXX:2:2106:3714:12029	147	Supercontig21, 8120, 60, 50M, =
HWI-ST697:173:D1N1UACXX:2:2315:8406:67892	147	Supercontig21, 8124, 60, 46M, =
HWI-ST697:173:D1N1UACXX:2:1304:10338:46154	329	Supercontig21, 8253, 1, 12M, *
HWI-ST697:173:D1N1UACXX:2:1209:13620:8867	419	Supercontig21, 8606, 1, 9M59502N41M, =
HWI-ST697:173:D1N1UACXX:2:1108:14361:36323	329	Supercontig21, 8785, 0, 12M, *
HWI-ST697:173:D1N1UACXX:2:2315:4225:78082	163	Supercontig21, 8867, 3, 14M32619N9M, =
HWI-ST697:173:D1N1UACXX:2:2315:4225:78082	419	Supercontig21, 8867, 3, 19M4S, =
HWI-ST697:173:D1N1UACXX:2:1113:16281:38206	163	Supercontig21, 9184, 60, 6S15M, =
HWI-ST697:173:D1N1UACXX:2:2315:17657:47026	163	Supercontig21, 9568, 60, 7S17M1S, =
HWI-ST697:173:D1N1UACXX:2:1113:8025:23752	419	Supercontig21, 9571, 3, 9M190983N10M, =
HWI-ST697:173:D1N1UACXX:2:1206:10901:47802	163	Supercontig21, 9571, 60, 9M13S, =
HWI-ST697:173:D1N1UACXX:2:2206:16370:89706	419	Supercontig21, 10259, 0, 13S6M, =
HWI-ST697:173:D1N1UACXX:2:2305:19059:3686	355	Supercontig21, 10450, 0, 11M7S, =

StringTie – Via Galaxy

The screenshot displays the Galaxy web interface for the StringTie tool. The main configuration area is titled "StringTie transcript assembly and quantification (Galaxy Version 2.2.1+galaxy0)".

Input options:

- Short reads: [Dropdown menu]
- Input short mapped reads *
Arquivos mapped.bam

Specify strand information *

- Reverse (RF) Fita "reverse"

Use a reference file to guide assembly?

- Use reference GTF/GFF3 Use reference GTF

Reference file

- Use a file from history Use a file from history

History Panel:

- Search datasets
- SRT-Aluno
- 242 GB, 186, 55
- 200: RNA STAR on data 3, data 2, and others: mapped.bam
- 196: RNA STAR on data 3, data 2, and others: mapped.bam
- 192: RNA STAR on data 3, data 2, and others: mapped.bam
- 188: RNA STAR on data 3, data 2, and others: mapped.bam
- 184: RNA STAR on data 3, data 2, and others: mapped.bam
- 180: RNA STAR on data 3, data 2, and others: mapped.bam
- 176: RNA STAR on data 3, data 2, and others: mapped.bam
- 172: RNA STAR on data 3, data 2, and others: mapped.bam
- 168: RNA STAR on data 3, data 2, and others: mapped.bam
- 164: RNA STAR on data 3, data 2, and others: mapped.bam
- 160: RNA STAR on data 3, data 2, and others: mapped.bam
- 156: RNA STAR on data 3, data 2, and others: mapped.bam
- 195: RNA STAR on data 3, data 2, and others: splice junction s.bed
- 194: RNA STAR on data 3, data 2, and others: log
- 193: RNA STAR on data 3, data 2, and others: reads per gene
- 192: RNA STAR on data 3, data 2, and others: mapped.bam
- 191: RNA STAR on data 3, data 2, and others: splice junction s.bed
- 190: RNA STAR on data 3, data 2, and others: log
- 189: RNA STAR on data 3, data 2, and others: reads per gene

StringTie – Via Galaxy

The screenshot shows the Galaxy web interface for the StringTie tool. The browser address bar displays the URL: `usegalaxy.org/?tool_id=toolshed.g2.bx.psu.edu%2Frepos%2Fucsc%2Fstringtie%2Fstringtie%2F2.2.1%2Bg08xversion=2.2.1%20galaxy0`. The Galaxy logo and navigation menu are visible at the top.

Tools Panel (Left): A search bar contains the text "strin". Below it are buttons for "Upload Data" and "Show Sections". A list of tool categories is shown, including "StringTie merge transcripts", "StringTie transcript assembly and quantification", "Filter mapped reads on MD tag string", "Convert gffCompare annotated GTF to BED for StringTie results", "Search ChEMBL database for compounds which are similar to a SMILES string", and "WORKFLOWS".

Main Configuration Area:

- StringTie transcript assembly and quantification (Galaxy Version 2.2.1+galaxy0)** - Includes a "Run Tool" button.
- GTF/GFF3 dataset to guide assembly ***: A dropdown menu shows "3: trichophyton_rubrum_cbs_118892_2_transcripts.gtf" with a yellow highlight and the text "Arquivo.gtf".
- Use Reference transcripts only?**: A toggle switch is set to "Yes", with a black arrow pointing to it and the text "Apenas transcritos conhecidos".
- Output files for differential expression?**: A dropdown menu shows "DESeq2/edgeR/limma-voom" with a yellow highlight and the text "Saída para o software DESeq2".
- Specify the average read length ***: A text input field contains "75".
- Cluster overlapping genes**: A toggle switch is set to "No".
- Prefix used for transcripts - optional**: An empty text input field.
- Prefix for clustering - optional**: An empty text input field.
- Output coverage file?**: A toggle switch is set to "No".
- Advanced Options**: A link to expand the configuration.

History Panel (Right): Titled "SRT-Aluno", it shows a list of job entries with details like "196 : RNA STAR on data 3, data 2, and others: mapped.bam".

Exemplo de Resultado StringTie (Gene counts)

Column 1	Column 2
gene_id	RNA_STAR_on_data_3_data_2_and_others_mapped_bam
TERG_00002	6030
TERG_00003	17457
TERG_00004	30772
TERG_00008	23418
TERG_00009	4228
TERG_00010	92
TERG_00011	1191
TERG_00012	12182
TERG_00013	1687
TERG_00014	176
TERG_00015	8428
TERG_00016	1406
TERG_00017	8658
TERG_00018	2903
TERG_00019	5389
TERG_00020	13537
TERG_00021	9995

Identificador do Gene

Quantidade de reads
mapeadas no Gene