

Contagem das leituras mapeadas no genoma de referência

Pablo Rodrigo Sanches

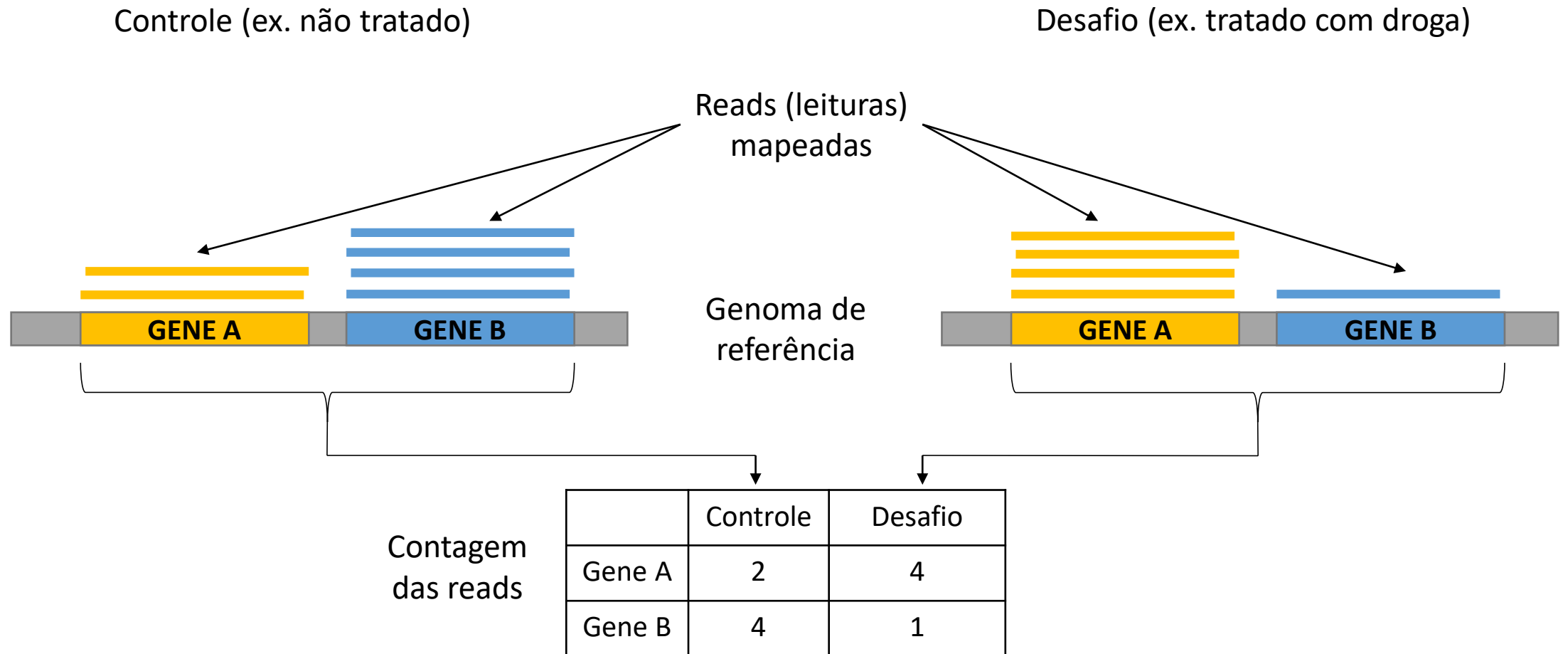
Departamento de Genética – FMRP/USP

psanches@usp.br

Roteiro de análise

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

Introdução



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 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.

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

The screenshot shows the Galaxy web interface at usegalaxy.org/datasets/edit. The main content area is titled "Edit dataset attributes" and contains a form for editing a dataset. The "Name" field is highlighted with a red box and contains the text "12h III RNA STAR on data 106, data 105, and others: mapped.bam". A black box with white text, containing the text "Oh II RNA STAR...; Oh III RNA STAR...; 3h II RNA STAR...; ...", is positioned over the "Name" field. A yellow arrow points to the "Save" button in the top right corner of the form. The "Info" section shows a log of events: "Apr 18 04:57:13 started STAR run", "Apr 18 04:57:13 ... starting to generate Genome files", "Apr 18 04:57:14 processing annotations GTF", and "Apr 18 04:57:14 ... starting to sort Suffix Array. This may take a long time...". The "Annotation" section is empty. The "Database/Build" dropdown menu is set to "----- Additional Species Are Below -----". The left sidebar shows a list of tools, including "GENERAL TEXT TOOLS", "GENOMIC FILE MANIPULATION", "COMMON GENOMICS TOOLS", and "GENOMICS ANALYSIS". The right sidebar shows the "History" section with a list of datasets, including "124: RNA STAR on data 106, data 105, and others: mapped.bam", "123: RNA STAR on data 106, data 105, and others: splice junctions.bed", "122: RNA STAR on data 106, data 105, and others: log", "121: RNA STAR on data 106, data 105, and others: mapped.bam", "120: RNA STAR on data 106, data 105, and others: splice junctions.bed", "119: RNA STAR on data 106, data 105, and others: log", "118: RNA STAR on data 106, data 105, and others: mapped.bam", "117: RNA STAR on data 106, data 105, and others: splice junctions.bed", and "116: RNA STAR on data 106, data 105, and others: log".

Arquivo .bam no Galaxy

The screenshot displays the Galaxy web interface. The main area shows a list of datasets, each with a unique identifier, a size, and a format. The datasets are organized into a table with columns for ID, size, and format. The history panel on the right shows a list of recent jobs, including RNA STAR and Trimmomatic runs, with their respective outputs and logs.

ID	Size	Format
HWI-ST697:173:D1N1UACXX:2:2109:2484:26168	345	Supercontig21
HWI-ST697:173:D1N1UACXX:2:2215:20629:9058	419	Supercontig21
HWI-ST697:173:D1N1UACXX:2:2305:19059:3686	355	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1211:1655:40186	419	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1108:8261:38044	163	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1310:13300:86402	419	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1311:5125:43902	419	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1311:12408:32050	419	Supercontig21
HWI-ST697:173:D1N1UACXX:2:2213:9496:34003	393	Supercontig21
HWI-ST697:173:D1N1UACXX:2:2206:16370:89706	419	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1314:11983:57787	419	Supercontig21
HWI-ST697:173:D1N1UACXX:2:2314:13926:9017	89	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1108:2208:24362	419	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1101:10540:19823	99	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1108:7798:9177	163	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1205:3211:48668	99	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1304:20379:29169	99	Supercontig21
HWI-ST697:173:D1N1UACXX:2:2103:1699:13193	99	Supercontig21
HWI-ST697:173:D1N1UACXX:2:2111:17386:80876	163	Supercontig21
HWI-ST697:173:D1N1UACXX:2:2215:2781:15226	163	Supercontig21
HWI-ST697:173:D1N1UACXX:2:2215:2881:65232	163	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1311:2286:53130	419	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1308:13748:31466	329	Supercontig21
HWI-ST697:173:D1N1UACXX:2:2206:16370:89706	419	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1104:9639:20843	345	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1311:12408:32050	419	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1211:3095:11645	419	Supercontig21
HWI-ST697:173:D1N1UACXX:2:2215:18513:69680	419	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1312:11366:84109	393	Supercontig21
HWI-ST697:173:D1N1UACXX:2:2106:3714:12029	99	Supercontig21
HWI-ST697:173:D1N1UACXX:2:2315:8406:67892	99	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1315:18038:38905	163	Supercontig21
HWI-ST697:173:D1N1UACXX:2:2106:3714:12029	147	Supercontig21
HWI-ST697:173:D1N1UACXX:2:2315:8406:67892	147	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1304:10338:46154	329	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1209:13620:8867	419	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1108:14361:36323	329	Supercontig21
HWI-ST697:173:D1N1UACXX:2:2315:4225:78082	163	Supercontig21
HWI-ST697:173:D1N1UACXX:2:2315:4225:78082	419	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1113:16281:38206	163	Supercontig21
HWI-ST697:173:D1N1UACXX:2:2315:17657:47026	163	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1113:8025:23752	419	Supercontig21
HWI-ST697:173:D1N1UACXX:2:1206:10901:47802	163	Supercontig21
HWI-ST697:173:D1N1UACXX:2:2206:16370:89706	419	Supercontig21
HWI-ST697:173:D1N1UACXX:2:2305:19059:3686	355	Supercontig21

History Panel:

- 111: 0h III RNA STAR on data 106, data 105, and others: splice junctions.bed
- 110: 0h III RNA STAR on data 106, data 105, and others: log
- 109: 0h II RNA STAR on data 106, data 105, and others: mapped.bam
- 108: 0h II RNA STAR on data 106, data 105, and others: splice junctions.bed
- 107: 0h II RNA STAR on data 106, data 105, and others: log
- 106: trichophyton_rubrum_cbs_118892_2_transcripts.gtf
- 105: trichophyton_rubrum_cbs_118892_2_supercontigs.fasta
- 104: Trimmomatic on data 44 and data 43 (log file)
- 103: Trimmomatic on data 44 and data 43 (trim log file)
- 102: Trimmomatic on 1 2h_III_R2.fastq.gz (R2 unpaired)
- 101: Trimmomatic on 1 2h_III_R2.fastq.gz (R2 unpaired)

StringTie – Via Galaxy

Galaxy

usegalaxy.org

Analyze Data Workflow Visualize Shared Data Help User

Using 30%

Tools

stringtie

Upload Data

Show Sections

StringTie transcript assembly and quantification

StringTie merge transcripts

Convert gffCompare annotated GTF to BED for StringTie results

RNA STAR Gapped-read mapper for RNA-seq data

DESeq2 Determines differentially expressed features from count tables

GffCompare compare assembled transcripts to a reference annotation

Remove Unwanted Variation from RNA-seq data

HISAT2 A fast and sensitive alignment program

WORKFLOWS

All workflows

StringTie transcript assembly and quantification (Galaxy Version 2.1.1)

Favorite Versions Options

Input mapped reads

109: 0h II RNA STAR on data 106, data 105, and others: mapped.bam

Arquivo mapped.bam

Input BAM/SAM file containing reads you want to assemble into transcripts

Input contains long reads?

No

Select if the input contains long error-prone reads, e.g. from Oxford Nanopore or Pacbio sequencing. (-L)

Specify strand information

Reverse (RF)

Fita "reverse"

Select 'Forward (FR)' if you are sequencing a stranded library, 'Reverse (RF)' if your reads are from a reverse-stranded library, or 'Unstranded' if your reads are not from a stranded library. See Help section below for more information. Default: Unstranded

Use a reference file to guide assembly?

Use reference GTF/GFF3

Use reference GTF

Use the reference annotation file to guide assembly (the output will include expressed reference transcripts as well as any novel transcripts that are assembled. This option is required by option -e (Use Reference transcripts only), see below. (-G)

Reference file

Use a file from history

Use a file from history

GTF/GFF3 dataset to guide assembly

106: trichophyton_rubrum_cbs_118892_2_transcripts.gtf

Arquivo.gtf

Use Reference transcripts only?

Yes

Apenas transcritos conhecidos

Limit the processing of read alignments to reference transcripts given with the -G option. With this option, read bundles with no reference transcripts (novel transcripts) will be entirely skipped, which may provide a speedup. Default: No (-e)

Output files for differential expression?

DESeq2/edgeR/limma-voom

Saída para o software DESeq2

Select to output additional files that can be used for differential expression analysis

Specify the average read length

75

Default: 75 (--length)

Cluster overlapping genes

History

search datasets

UDA

92 shown

77.35 GB

124: 12h III RNA STAR on data 106, data 105, and others: mapped.bam

123: RNA STAR on data 106, data 105, and others: splice junctions.bed

122: RNA STAR on data 106, data 105, and others: log

121: 12h II RNA STAR on data 106, data 105, and others: mapped.bam

120: RNA STAR on data 106, data 105, and others: splice junctions.bed

119: RNA STAR on data 106, data 105, and others: log

118: 3h III RNA STAR on data 106, data 105, and others: mapped.bam

117: RNA STAR on data 106, data 105, and others: splice junctions.bed

116: RNA STAR on data 106, data 105, and others: log

115: 3h II RNA STAR on data 106, data 105, and others: mapped.bam

114: RNA STAR on data 106, data 105, and others: splice junctions.bed

Exemplo de Resultado StringTie (Gene counts)

gene_id	0h_II_RNA_STAR_on_data_106_data_105_and_others_mapped_bam
TERG_00002	1248
TERG_00003	1406
TERG_00004	2301
TERG_00008	1189
TERG_00009	589
TERG_00010	58
TERG_00011	101
TERG_00012	1052
TERG_00013	161
TERG_00014	37
TERG_00015	806
TERG_00016	161
TERG_00017	1871
TERG_00018	333
TERG_00019	845
TERG_00020	1097

Identificador do Gene

Quantidade de reads mapeadas no Gene