

BS-Seq analysis using Galaxy

Konstantin Riege, Leibniz Institute on Aging (FLI)

RRBS-Seq

- Reduced Representation BS-Seq
 - Digestion of non-methylated DNA fragments
 - Bisulfite conversion & sequencing
 - ~1% DNA
 - Highly CpG enriched regions
 - Promoter and enhancer regions

Correlation with gene expression



RESEARCH ARTICLE

A complex association between DNA methylation and gene expression in human placenta at first and third trimesters

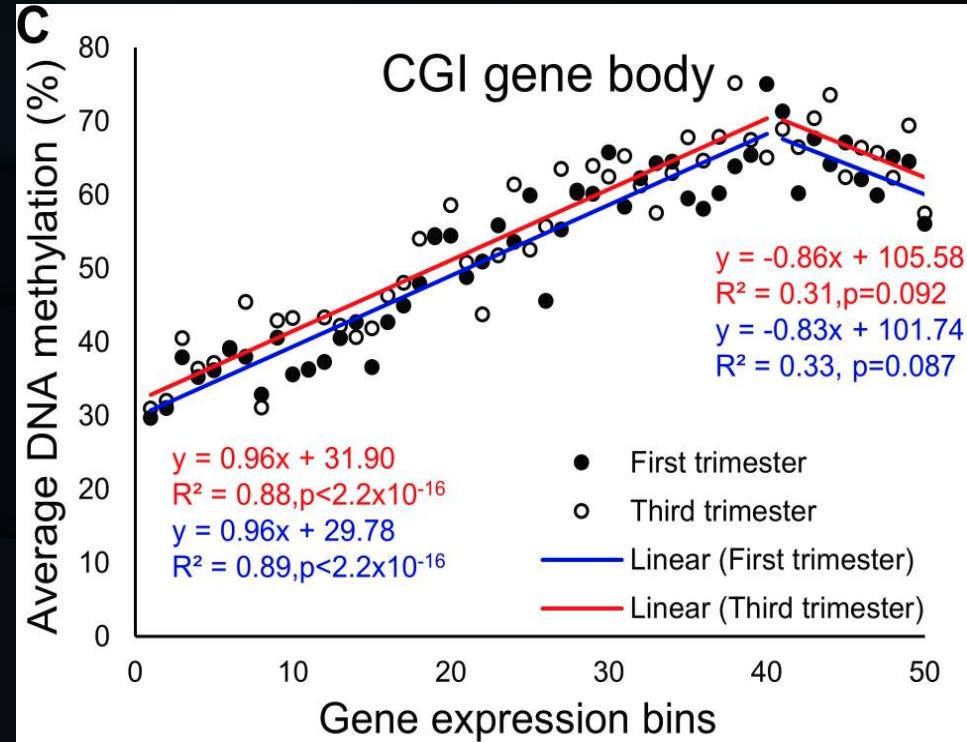
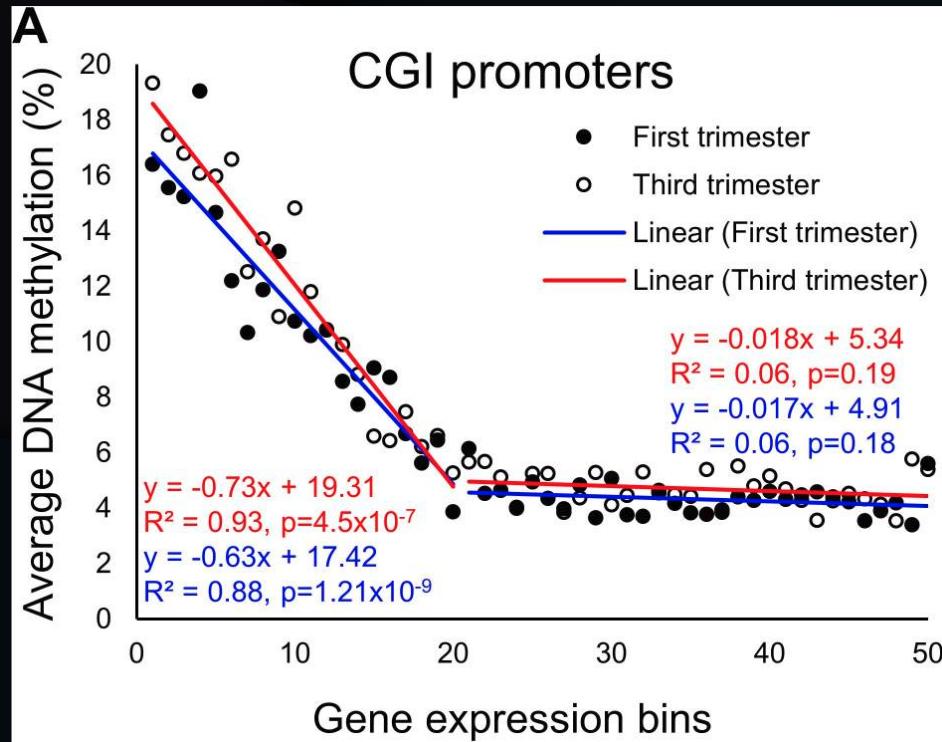
Yen Ching Lim¹, Jie Li¹, Yiyun Ni¹, Qi Liang¹, Junjiao Zhang¹, George S. H. Yeo², Jianxin Lyu^{1*}, Shengnan Jin^{1*}, Chunming Ding^{1*}

1 Key Laboratory of Laboratory Medicine, Ministry of Education of China, School of Laboratory Medicine and Life Science, Wenzhou Medical University, Wenzhou, Zhejiang, China, **2** KK Women's and Children's Hospital, Singapore, Singapore

Correlation with gene expression

- 1st trimester → 3rd trimester
 - ~200 diff. meth. Promoter regions
 - ~2200 diff. meth. genes
 - Hypermethylation for > 90%
 - 2400 diff. expressed genes
 - Downregulation for ~80%

Correlation with gene expression



25 DEG with DME

- Data preparation for 22 protein coding genes as listed in GENCODE – Ensembl v92
 - Minimalistic HG38 derived genome representation
 - Genes & 21 x 4kb intergenic regions: Fasta format
 - Minimalistic annotation
 - Genes: GTF format
 - Promoter - 1kb gene upstream regions: BED format

25 DEG with DME

- Data preparation for RRBS-Seq and RNA-Seq
 - Preprocessing & mapping
 - RRBS-Seq
 - RNA-Seq data
 - Subsampling to 5% & extraction from RAW data
 - Download:

<https://share.leibniz-fli.de/index.php/s/cQJAntRJJsrfMYb>

25 DEG with DME

- RRBS-Seq
 - 1T & 3T @ N=4, paired-end (PE)
- RNA-Seq
 - 1T & 3T @ N=4
 - 1T paired-end (PE)
 - **3T N1,N4 single (SE)**
 - **3T N2,N3 paired-end (PE)**

Galaxy – Upload

1

2

3

Galaxy / Europe

Tools

search tools

Get Data

Download from web or upload from disk

Analyze Data Workflow Visualize Shared Data Help User

Regular Composite Collection Rule-based

Please wait...30 out of 33 remaining.

Name	Size	Type	Genome	Settings	Status
hg38.goi.fa	1.7 MB	Auto-detect	Additional Speci...	100%	✓
hg38.goi.gtf	1.6 MB	Auto-detect	Additional Speci...	100%	✓
hg38.promoter.bed	816 b	Auto-detect	Additional Speci...	100%	✓
rbs_1T.N1.R1.fastq	14.4 MB	Auto-detect	Additional Speci...	Adding to history...	○
rbs_1T.N1.R2.fastq	14.4 MB	Auto-detect	Additional Speci...	0%	○
rbs_1T.N2.R1.fastq	20.9 MB	Auto-detect	Additional Speci...	0%	○
rbs_1T.N2.R2.fastq	20.9 MB	Auto-detect	Additional Speci...	0%	○

Type (set all): Auto-detect

Genome (set all): Additional Species Are B...

Choose local file Choose FTP file Paste/Fetch data Pause Reset Start Close

Learn how the NASA uses Galaxy!

Quality analysis: FastQC

Tools

fastqc

FASTA/FASTQ

Combine FASTA and QUAL into FASTQ

Manipulate FASTQ reads on various attributes

FastQC:Read QC reports using FastQC

fastp - fast all-in-one preprocessing for FASTQ files

FASTQ Quality Control

FastQC Read Quality reports

Mapping

Map with PerM for SOLiD and Illumina

FastQC Read QC reports (Galaxy Version 0.72)

Short read data from your current history

33: rna_3T.N4.fastq
32: rna_3T.N3.R2.fastq
31: rna_3T.N3.R1.fastq
30: rna_3T.N2.R2.fastq
29: rna_3T.N2.R1.fastq

This is a batch mode input

Contaminant list

Nothing selected

tab delimited file with 2 columns: name and sequence.

Submodule and Limit specifying file

Nothing selected

a file that specifies which submodules are to be executed

Execute

SE – Quality trimming: Trimmomatic

Tools

trimmomatic

FASTA/FASTQ
fastp - fast all-in-one processing for FASTQ files

FASTQ Quality Control
Trimmomatic flexible read trimming tool for Illumina NGS data

Assembly
Shovill Faster SPAdes assembly of Illumina reads

Workflows
▪ All workflows

Trimmomatic flexible read trimming tool for Illumina NGS data (Galaxy Version 0.36.5)

Single-end or paired-end reads?

Single-end

Input FASTQ file

33: rna_3TN4.fastq
32: rna_3TN3.R2.fastq
31: rna_3TN3.R1.fastq
30: rna_3TN2.R2.fastq
29: rna_3TN2.R1.fastq
28: rna_3TN1.fastq
27: rna_1TN4.R2.fastq

This is a batch mode input field. Separate jobs will be triggered for each file.

Perform initial ILLUMINACLIP step?

Yes No

Cut adapter and other illumina-specific sequences from the read

Trimmomatic Operation

1: Trimmomatic Operation

Select Trimmomatic operation to perform
Sliding window trimming (SLIDINGWINDOW)

Number of bases to average across
4

Average quality required
20

2: Trimmomatic Operation

Select Trimmomatic operation to perform
Drop reads below a specified length (MINLEN)

Minimum length of reads to be kept
20

PE – Quality trimming: Trimmomatic

Tools

trimmomatic

FASTA/FASTQ

fastp - fast all-in-one preprocessing for FASTQ files

FASTQ Quality Control

Trimmomatic flexible read trimming tool for Illumina NGS data

Assembly

Shovill Faster SPAdes assembly of Illumina reads

Workflows

- All workflows

Trimmomatic flexible read trimming tool Illumina NGS data (Galaxy Version 0.36.5)

Single-end or paired-end reads?

Paired-end (two separate input files)

Input FASTQ file (R1/first of pair)

23: rna_1T.N2.R2.fastq
22: rna_1T.N2.R1.fastq
21: rna_1T.N1.R2.fastq
20: rna_1T.N1.R1.fastq
19: rbs_3T.N4.R2.fastq
18: rbs_3T.N4.R1.fastq

This is a batch mode input field. Separate jobs will be triggered.

Input FASTQ file (R2/second of pair)

9: rbs_1T.N3.R2.fastq
8: rbs_1T.N3.R1.fastq
7: rbs_1T.N2.R2.fastq
6: rbs_1T.N2.R1.fastq
5: rbs_1T.N1.R2.fastq
4: rbs_1T.N1.R1.fastq

This is a batch mode input field. Separate jobs will be triggered.

Perform initial ILLUMINACLIP step?

Yes No

Cut adapter and other illumina-specific sequences from the read

RNA – SE – Mapping: Hisat2

The screenshot shows the Galaxy web interface with the Hisat2 tool selected for RNA-seq single-end mapping.

- Step 1: The search bar at the top left contains "hisat2".
- Step 2: The "Tools" dropdown menu is open, showing the selected "hisat2" entry.
- Step 3: The main configuration panel for the HISAT2 tool is displayed.
- Step 4: Under "Source for the reference genome", the option "Use a genome from history" is selected.
- Step 5: Under "Select the reference genome", the genome "hg38.gogi.fa" is chosen.
- Step 6: Under "Is this a single- or paired library?", the "Single-end" option is selected.
- Step 7: Under "FASTA/Q file", a batch mode input field lists several fastq files:
 - 96: Trimmomatic on rbs_1T.N1.R1.fastq (Paired)
 - 95: Trimmomatic on rna_3T.N4.fastq
 - 94: Trimmomatic on rna_3T.N1.fastq
 - 33: rna_3T.N4.fastq
 - 32: rna_3T.N3.R2.fastq
 - 21: rna_3T.N2.D1.fastqThe item "94: Trimmomatic on rna_3T.N1.fastq" is highlighted with a blue background.

Below the FASTA/Q file input field, a note states: "This is a batch mode input field. Separate jobs will be triggered for each file listed here." Another note specifies: "Must be of datatype 'fastqsanger' or 'fasta'".

Other sections visible include "Specify strand information" (set to "Unstranded") and "Summary Options" and "Advanced Options". A "Execute" button is at the bottom.

RNA – PE – Mapping: Hisat2

Tools hisat2

Mapping
HISAT2 A fast and sensitive alignment program

RNA Analysis
StringTie transcript assembly and quantification

Workflows
▪ All workflows

HISAT2 A fast and sensitive alignment program (Galaxy Version 2.1.0+galaxy4)

Source for the reference genome
Use a genome from history
Built-in references were created using default options
Select the reference genome
1 hg38.gi.fa

Is this a single-end or paired-end library
Paired-end

FASTA/Q file
4 This is a batch mode input file. Separate jobs will be triggered.
Must be of datatype "fastqsanger" or "fasta"

148: Trimmomatic on rna_3TN2.R1.fastq (R1 paired)
147: Trimmomatic on rna_3TN2.R2.fastq (R2 unpaired)
146: Trimmomatic on rna_3TN2.R1.fastq (R1 unpaired)
145: Trimmomatic on rna_3TN2.R2.fastq (R2 paired)
144: Trimmomatic on rna_3TN2.R1.fastq (R1 paired)

5 This is a batch mode input file. Separate jobs will be triggered.
Must be of datatype "fastqsanger" or "fasta"

FASTA/Q file
149: Trimmomatic on rna_3TN2.R2.fastq (R2 paired)
148: Trimmomatic on rna_3TN2.R1.fastq (R1 paired)
147: Trimmomatic on rna_3TN2.R2.fastq (R2 unpaired)
146: Trimmomatic on rna_3TN2.R1.fastq (R1 unpaired)
145: Trimmomatic on rna_3TN2.R2.fastq (R2 paired)
144: Trimmomatic on rna_3TN2.R1.fastq (R1 paired)

Specify strand information
Unstranded
'FR' means a read corresponds to a transcript. 'RF' means a read corresponds to a transcript. 'U' means a read is unused, every read alignment will have an XS attribute tag: '+' means a read belongs on '+' strand of genome. ('-rna-strandness')

Paired-end options
Specify paired-end parameters
See "Alignment Options" section of Help below for information

Select the upstream/downstream mate orientations for a valid paired-end alignment
5 --fr

RRBS – Mapping: Bismark

The screenshot shows the Galaxy web interface with the Bismark tool selected. Red numbered callouts point to specific fields:

- 1: The search bar at the top containing "bismark".
- 2: The "FASTA/FASTQ" section, which is currently empty.
- 3: The title of the Bismark Mapper Bisulfite reads mapper tool.
- 4: The question "Will you select a reference genome from your history or use a built-in index?"
- 5: The "Select the reference genome" dropdown menu, showing "hg38.goi.fa" as the selected option.
- 6: The "Paired End Pairs" section, specifically the "Mate pair 1" input field.
- 7: The "Mate pair 1" input field containing a list of FASTQ files:
 - 120: Trimmomatic on rbs_3T.N3.R1.fastq (R1 paired)
 - 119: Trimmomatic on rbs_3T.N2.R2.fastq (R2 unpaired)
 - 118: Trimmomatic on rbs_3T.N2.R1.fastq (R1 unpaired)
 - 117: Trimmomatic on rbs_3T.N2.R2.fastq (R2 paired)
 - 116: Trimmomatic on rbs_3T.N2.R1.fastq (R1 paired)
- 8: The "Mate pair 2" input field.
- 9: The "Mate pair 2" input field containing a list of FASTQ files:
 - 110: Trimmomatic on rbs_3T.N1.R1.fastq (R1 unpaired)
 - 117: Trimmomatic on rbs_3T.N2.R2.fastq (R2 paired)
 - 116: Trimmomatic on rbs_3T.N2.R1.fastq (R1 paired)
 - 115: Trimmomatic on rbs_3T.N1.R2.fastq (R2 unpaired)
 - 114: Trimmomatic on rbs_3T.N1.R1.fastq (R1 unpaired)
 - 113: Trimmomatic on rbs_3T.N1.R2.fastq (R2 paired)

Below the input fields are sections for "Minimum insert size for valid paired-end alignments" (set to 0) and "Maximum insert size for valid paired-end alignments" (set to 500).

Read filter: Samtools view

Tools

samtools view

SAM/BAM

- SAM-to-BAM convert SAM to BAM
- Slice BAM by genomic regions
- BAM-to-SAM convert BAM to SAM
- Filter SAM or BAM, output SAM or BAM files on FLAG MAPQ RG LN or by region
- Filter SAM or BAM files on FLAG MAPQ RG LN or by region

VCF/BCF

- bcftools call SNP/indel variant calling from VCF/BCF

Variant Calling

- bcftools call SNP/indel variant calling from VCF/BCF

Workflows

- All workflows

Filter SAM or BAM files on FLAG MAPQ RG LN or by region (Galaxy Version 1.8)

SAM or BAM file to filter

176: Bismark Mapper on data 114, data 124, and data 1: mapped reads (as bam)
174: Bismark Mapper on data 121, data 120, and data 1: mapped reads (as bam)
172: Bismark Mapper on data 117, data 116, and data 1: mapped reads (as bam)
170: Bismark Mapper on data 113, data 112, and data 1: mapped reads (as bam)
168: Bismark Mapper on data 109, data 108, and data 1: mapped reads (as bam)

This is a batch mode input field. Separate jobs will be triggered for each dataset selected.

Header in output

Include header

Minimum MAPQ quality score

1
(-q)

Filter on bitwise flag

yes

Only output alignments with all of these flag bits set

Select/Unselect all

- Read is paired
- Read is mapped in a proper pair
- The read is unmapped
- The mate is unmapped
- Read strand
- Mate strand
- Read is the first in a pair
- Read is the second in a pair
- The alignment or this read is not primary
- The read fails platform/vendor quality checks
- The read is a PCR or optical duplicate
- Supplementary alignment

Skip alignments with any of these flag bits set

Select/Unselect all

- Read is paired
- Read is mapped in a proper pair
- The read is unmapped
- The mate is unmapped
- Read strand
- Mate strand
- Read is the first in a pair
- Read is the second in a pair
- The alignment or this read is not primary
- The read fails platform/vendor quality checks
- The read is a PCR or optical duplicate
- Supplementary alignment

(-F)

Sort reads: Samtools sort

Tools

samtools sort

Fasta/FASTQ

[UMI-tools group](#) Extract UMI from fastq files

SAM/BAM

[Filter pileup](#) on coverage and SNPs

[Generate pileup](#) from BAM dataset

[SAM-to-BAM](#) convert SAM to BAM

[Samtools flagstat tabulate](#)
descriptive stats for BAM dataset

[samtools mpileup](#) multi-way pileup of variants

Samtools sort: order of storing aligned sequences (Galaxy Version 2.0.2)

BAM File

203: Filter SAM or BAM, output SAM or BAM on data 176: bam
202: Filter SAM or BAM, output SAM or BAM on data 174: bam
201: Filter SAM or BAM, output SAM or BAM on data 172: bam
200: Filter SAM or BAM, output SAM or BAM on data 170: bam
199: Filter SAM or BAM, output SAM or BAM on data 168: bam
198: Filter SAM or BAM, output SAM or BAM on data 166: bam
197: Filter SAM or BAM, output SAM or BAM on data 164: bam
196: Filter SAM or BAM, output SAM or BAM on data 162: bam
176: Bismark Mapper on data 125, data 124, and data 1: mapped

This is a batch mode input field. Separate jobs will be triggered

Primary sort key

coordinate

✓ Execute

Differential Gene expression



RNA – Quantification

- Requirements
 - Experiment strandness information
 - Prediction needs annotation in BED format

GTF to BED

The screenshot shows the Galaxy web interface for a 'Convert GTF to BED12' tool. The interface is divided into two main sections: a left sidebar and a right main panel.

Left Sidebar (Tools Panel):

- Header: Tools
- Search bar: gtf to bed
- Section: Convert Formats
 - [Convert gff3 to annotated GTF to BED for TilingTie results](#)
 - [Convert GTF to BED12](#)
 - [GTF-to-BEDGraph converter](#)
- Section: Workflows

Main Panel (Convert GTF to BED12):

- Section: Convert GTF to BED12 (Galaxy Version 77)
 - Header: GTF File to convert
 - File input: 2: hg38.goi.gtf
 - Section: Advanced options
 - Use default options
 - Advanced options for gtfToGenePred.
 - Execute button: A blue button with a white checkmark icon and the text "Execute".

Red numbered arrows point to specific elements:

- Red arrow pointing to the 'gtf to bed' search bar in the sidebar.
- Red arrow pointing to the 'Convert GTF to BED12' link in the sidebar.
- Red arrow pointing to the 'hg38.goi.gtf' file input field in the main panel.

RNA – SE – Infer experiment

Tools

1 infer_experiment 2

RNA Analysis

Infer Experiment speculates how RNA-seq were configured

FPKM Count calculates raw read count, FPM, and FPKM for each gene

BAM to Wiggle converts all types of RNA-seq data from .bam to .wig

RPKM Count calculates raw count and RPKM values for transcript at exon, intron, and mRNA level

Workflows

- All workflows

Infer Experiment speculates how RNA-seq were configured (Galaxy Version 2.6.4.1)

3 Input .bam file
4 153: HISAT2 on data 94 and data 1: aligned reads (BAM)
(--input-file)

Reference gene model
5 177: Convert GTF to BED12 on data 2
(--refgene)

Number of reads sampled from SAM/BAM file (default = 200000)
200000
(--sample-size)

6 Minimum mapping quality
0
Minimum mapping quality for an alignment to be considered as "uniquely mapped" (--mapq)

This is SingleEnd Data
Fraction of reads failed to determine: 0.0000
Fraction of reads explained by "++,--": 0.5033
Fraction of reads explained by "+-, -+": 0.4967

RNA – PE – Infer experiment

Tools 

infer_experiment 

RNA Analysis

Infer Experiment speculates how RNA-seq were configured

FPKM Count calculates raw read count, FPM, and FPKM for each gene

BAM to Wiggle converts all types of RNA-seq data from .bam to .wig

RPKM Count calculates raw count and RPKM values for transcript at exon, intron, and mRNA level

Workflows

- All workflows

Infer Experiment speculates how RNA-seq were configured (Galaxy Version 2.6.4.1)

Input BAM file   
155: HISAT2 on data 129, data 128, and data 1: aligned reads (BAM)
(--input-file)

Reference gene model   
177: Convert GTF to BED12 on data 2
(--refgene)

Number of reads sampled from SAM/BAM file (default = 200000)
200000
(--sample-size)

Minimum mapping quality
0
Minimum mapping quality for an alignment to be considered as "uniquely mapped" (--mapq)

This is PairEnd Data
Fraction of reads failed to determine: 0.0000
Fraction of reads explained by "1++,1--,2+-,2-+": 0.4941
Fraction of reads explained by "1+-,1--,2++,2--": 0.5059

RNA – SE – Quantify: featurecounts

Tools

featurecounts

Annotation

goseq tests for overrepresented gene categories

RNA Analysis

limma Perform differential expression with limma-voom or limma-trend

edgeR Perform differential expression of count data

DESeq2 Determines differentially expressed features from count tables

featureCounts Measure gene expression in RNA-Seq experiments from SAM or BAM files.

Remove Unwanted Variation from RNA-seq data

Workflows

All workflows

featureCounts Measure gene expression in RNA-Seq experiments from SAM or BAM files. (Galaxy Version 1.6.3+galaxy2)

Alignment file

157: HISAT2 on data 137, data 136, and data 1: aligned reads (BAM)
156: HISAT2 on data 133, data 132, and data 1: aligned reads (BAM)
155: HISAT2 on data 129, data 128, and data 1: aligned reads (BAM)
154: HISAT2 on data 95 and data 1: aligned reads
153: HISAT2 on data 94 and data 1: aligned reads

Count multi-mapping reads/fragments
Disabled; multi-mapping reads are excluded (default)
If specified, multi-mapping reads/fragments will be counted (ie. a multi-mapping tag can be used to find multi-mapping reads. (-M))

This is a batch mode input field. Separate jobs will be run for each alignment file. If you are using a Gene annotation file, the History, these files must have a Minimum mapping quality per read.

Specify strand information

Unstranded
Indicate if the data is stranded and strand-specific read counting should be performed on mapped BAM input(s) (-s)

Gene annotation file

in your history

Gene annotation file

2: hg38.gtf

The program assumes that the provided annotation file is in GTF format. It is used for the alignment.

Output format

Gene-ID "t" read-count (MultiQC/DESeq2/edgeR/limma-voom compatible)

The output format will be tabular, select the preferred columns here

Count multi-mapping reads/fragments
Disabled; multi-mapping reads are excluded (default)
If specified, multi-mapping reads/fragments will be counted (ie. a multi-mapping tag can be used to find multi-mapping reads. (-M))

This is a batch mode input field. Separate jobs will be run for each alignment file. If you are using a Gene annotation file, the History, these files must have a Minimum mapping quality per read.

Specify strand information

Unstranded
Indicate if the data is stranded and strand-specific read counting should be performed on mapped BAM input(s) (-s)

Gene annotation file

in your history

Gene annotation file

2: hg38.gtf

The program assumes that the provided annotation file is in GTF format. It is used for the alignment.

Long reads
Yes No
If specified, long reads such as Nanopore and PacBio reads will be counted.

If specified, long reads such as Nanopore and PacBio reads will be counted. (-L)

Count reads by read group
Yes No
If specified, reads are counted for each read group separately. The 'RG' tag is used for this.

If specified, reads are counted for each read group separately. The 'RG' tag is used for this.

Largest overlap
Yes No
If specified, the largest overlapping fragments will be assigned to the target that has the largest overlap.

If specified, the largest overlapping fragments will be assigned to the target that has the largest overlap.

Minimum bases of overlap
Specify the minimum required number of overlapping bases between a read and its mate. If extended from both ends. (-minOverlap)

RNA – PE – Quantify: featurecounts

Tools

featureCounts

Annotation

goseq tests for overrepresented gene categories

RNA Analysis

limma Perform differential expression with limma-voom or limma-trend

edgeR Perform differential expression of count data

DESeq2 Determines differentially expressed features from count tables

featureCounts Measure gene expression in RNA-Seq experiments from SAM or BAM files.

Remove Unwanted Variation from RNA-seq data

Workflows

- All workflows

featureCounts: Measure gene expression in RNA-Seq experiments from SAM or BAM files. (Galaxy Version 1.6.3+galaxy2)

Alignment file(s)

160: HISAT2 on data 149, data 148, and data 1: aligned reads (BAM)
159: HISAT2 on data 145, data 144, and data 1: aligned reads (BAM)
158: HISAT2 on data 141, data 140, and data 1: aligned reads (BAM)
157: HISAT2 on data 137, data 136, and data 1: aligned reads (BAM)
156: HISAT2 on data 133, data 132, and data 1: aligned reads (BAM)

This is a batch mode input field. Separate jobs will be triggered for each alignment file. If you are using a Gene annotation file in your history, these files must have the same name as the alignment files.

Count multi-mapping reads/fragments

Disabled; multi-mapping reads are excluded (default)
If specified, multi-mapping reads/fragments will be counted (ie. a multi-mapping program uses the -M tag to find multi-mapping reads. (-M))

Specify strand information

Unstranded
Indicate if the data is unstranded and no strand-specific read counting should be performed for unmapped BAM input(s) (-s)

Gene annotation file

in your history
2: hg38.gtf
The program assumes that the provided annotation file is in GTF format. Make sure the file is correctly formatted and contains all genes used for the alignment

Gene annotation file

Gene-ID "lt" read-count (MultiQC/DESeq2/edgeR/limma-voom compatible)
The output format will be tabular, select the preferred columns here

Create gene-length file

Yes No
Creates a tabular file that contains the effective (nucleotides used for counting reads)

Output format

Count reads by read group

Yes No
If specified, reads are counted for each read group separately. The 'RG' tag is required for each read.

Largest overlap

Yes No
If specified, fragments (or templates) will be assigned to the target that has the largest overlap.

Options for paired-end reads

Count fragments instead of reads

Enabled; fragments (or templates) will be counted instead of reads
If specified, fragments (or templates) will be counted instead of reads. (-p)

Minimum bases of overlap

10
Specify the minimum required number of overlapping bases between a read and its mate. The read is extended from both ends. (-minOverlap)

RNA – Diff. expression: DESeq2

The screenshot shows the Galaxy web interface for running the DESeq2 tool. A red triangle labeled '1' points to the 'deseq2' entry in the search bar. A red triangle labeled '2' points to the 'DESeq2' link under the 'RNA Analysis' section. A red triangle labeled '3' points to the 'Specify a factor name' field containing 'selects_drug_x or cancer_markers'. A red triangle labeled '4' points to the 'Specify a factor level' field containing '3t'. A red triangle labeled '5' points to the 'Counts file(s)' dropdown menu showing options like '188: featureCounts on data 2 and data 157: Counts'. A red triangle labeled '6' points to the 'Counts file(s)' dropdown menu showing options like '188: featureCounts on data 2 and data 157: Counts'. A red triangle labeled '7' points to the 'Specify a factor level' field containing '1t'. A red triangle labeled '8' points to the 'Counts file(s)' dropdown menu showing options like '182: featureCounts on data 2 and data 154: Counts'. A red triangle labeled '9' points to the 'Counts file(s)' dropdown menu showing options like '182: featureCounts on data 2 and data 154: Counts'.

Tools

deseq2

Annotation

Annotate DE(X)Seq results

RNA Analysis

edgeR Perform differential expression of count data

DESeq2 Determines differentially expressed features from count tables

Annotate DESeq2/DEXSeq output tables Append annotation from GTF to differential expression tool outputs

featureCounts Measure gene expression in RNA-Seq experiments from SAM or BAM files

Remove Unwanted Variation from RNA-seq data

StringTie transcript assembly and quantification

Peak Calling

PEAKachu Calls Peaks in CLIP data

DiffBind differential binding analysis of ChIP-Seq peak data

Workflows

All workflows

DESeq2 Determines differentially expressed features from count tables (Galaxy Version 2.11.40.6)

how

Select datasets per level

Factor

1: Factor

Specify a factor name, **selects_drug_x or cancer_markers**

placenta

Only letters, numbers and underscores will be retained in this field

Factor level

1: Factor level

Specify a factor level, **actual values could be 'tumor', 'normal', 'treated' or 'control'**

3t

Only letters, numbers and underscores will be retained in this field

Counts file(s)

188: featureCounts on data 2 and data 157: Counts
187: featureCounts on data 2 and data 156: Summary
186: featureCounts on data 2 and data 156: Counts
185: featureCounts on data 2 and data 155: Summary
184: featureCounts on data 2 and data 155: Counts

2: Factor level

Specify a factor level, **actual values could be 'tumor', 'normal', 'treated' or 'control'**

1t

Only letters, numbers and underscores will be retained in this field

Counts file(s)

182: featureCounts on data 2 and data 154: Counts
181: featureCounts on data 2 and data 153: Summary
180: featureCounts on data 2 and data 153: Counts
177: Convert GTF to BED12 on data 2
3: hg38.promoter.bed

+ Insert Factor level

+ Insert Factor

(Optional) provide a tabular file with additional batch factors to include in the model.

Cut columns: cut

The screenshot shows the Galaxy web interface with the 'Tools' panel on the left and the 'Cut columns from a table (cut)' tool configuration page on the right.

Tools Panel (Left):

- Search bar: 'cut' (marked with red triangle 1)
- Text Manipulation section:
 - Replace Text (entire line) (marked with red triangle 2)
 - Text reformatting with awk
 - Cut columns from a table (cut) (selected)
 - Cut columns from a table
- Filter and Sort section:
 - Filter sequences by ID from a tabular file
 - Select lines that match an expression
- FASTA/FASTQ section:
 - Filter by quality
 - fastp - fast all-in-one preprocessing for FASTQ files

Cut columns from a table (cut) Tool Configuration (Right):

- Title: Cut columns from a table (cut) (Galaxy Version 1.1.0) (marked with red triangle 3)
- File to cut: 247: Compute on data 246 (marked with red triangle 4)
- Operation: Keep
- Delimited by: Tab
- Cut by: fields (marked with red triangle 6)
- List of Fields:
 - Select/Unselect all
 - Column: 1 (marked with red triangle 5)
 - Column: 3
- Execute button (blue button at the bottom)

Tag results

Edit dataset attributes

Attributes Convert Datatypes Permissions

Edit attributes 2

Name: DGE.tsv

Info

Annotation

History

search datasets

Rostock

245 shown, 8 deleted, 8 hidden

868.02 MB

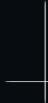
253: Cut on data 249 1

22 lines

format: tabular, database: ?

1	2
ENSG00000110934	2.23208401512721
ENSG00000136859	1.40905050600934
ENSG00000143850	-1.89543026151967
ENSG00000137868	-2.06915939568834
ENSG00000166444	-1.71738157515273

Difference in promoter methylation



Globally call meth. CpGs: MethylDackel

Tools

methyldackel

Epigenetics

MethylDackel A tool for processing bisulfite sequencing alignments

Workflows

- All workflows

MethylDackel tool for processing bisulfite sequencing alignments (Galaxy Version 0.3.0.1)

Load reference genome from

History

Use the following dataset as reference sequence

REFERENCE SEQUENCE; You can upload a FASTA sequence to the history and use it as reference

sorted_alignments.bam

This is a batch mode input field. Separate jobs will be triggered for each data item.

What do you want to do?

Extract methylation metrics from an alignment file in BAM/CRAM format

Merge per-Cytosine metrics from CpG and CHG contexts into per-CPG or per-CHG metrics

Yes No

(--mergeContext)

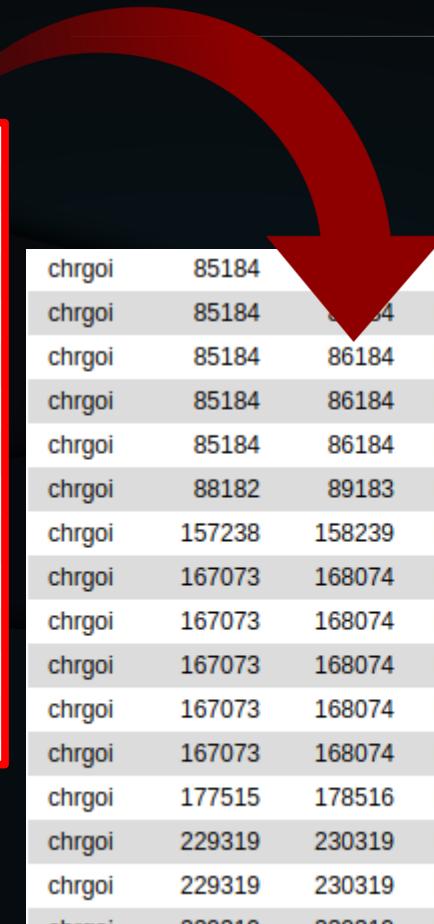
Fun with file formats

1	2	3	4	5	6
track 1					
chrroi	1	2	3	4	5
chrroi	track t	1	2	3	4
chrroi	chrroi	1	2	3	4
chrroi	chrroi	track type="bedGraph" description="output CpG merged methylation levels"	5	6	
chrroi	chrroi	1	2	3	4
chrroi	chrroi	2271	2273	50	2
chrroi	chrroi	2349	2351	25	1
chrroi	chrroi	9788	9790	3	4
chrroi	chrroi	9794	9796	29	31
chrroi	chrroi	9804	9806	34	37
chrroi	chrroi	9862	9864	50	48
chrroi	chrroi	9922	9924	75	15
chrroi	chrroi	9936	9938	100	5
chrroi	chrroi	9958	9960	20	1
chrroi	chrroi	12433	12435	85	6
chrroi	chrroi	12460	12462	100	7
chrroi	chrroi	12514	12516	14	1
chrroi	chrroi	35213	35215	13	4
chrroi	chrroi	35229	35231	48	14
chrroi	chrroi	35257	35259	41	12
chrroi	chrroi	35266	35268	13	4
chrroi	chrroi	35273	35275	37	11
chrroi	chrroi	35287	35289	9	3
chrroi	chrroi	37004	37006	0	0
chrroi	chrroi	37009	37011	50	2
chrroi	chrroi	37009	37011	50	2

chrroi	9788	9790	3	8	232
chrroi	9788	9790	5	20	325
chrroi	9788	9790	5	19	347
chrroi	9788	9790	4	12	229
chrroi	9794	9796	26	63	176
chrroi	9794	9796	37	126	206
chrroi	9794	9796	48	175	183
chrroi	9794	9796	41	99	139
chrroi	9804	9806	31	77	165
chrroi	9804	9806	chrroi	2271	2273
chrroi	9804	9806	chrroi	2349	2351
chrroi	9804	9806	chrroi	9788	9790
chrroi	9804	9806	chrroi	9794	9796
chrroi	9862	9864	chrroi	9862	9864
chrroi	9862	9864	chrroi	9862	9864
chrroi	9862	9864	chrroi	9862	9864
chrroi	9922	9924	chrroi	9922	9924
chrroi	9936	9938	chrroi	9936	9938
chrroi	9958	9960	chrroi	9958	9960
chrroi	12433	12435	chrroi	12433	12435
chrroi	12460	12462	chrroi	12460	12462
chrroi	12514	12516	chrroi	12514	12516
chrroi	35213	35215	chrroi	35213	35215
chrroi	35229	35231	chrroi	35229	35231
chrroi	35257	35259	chrroi	35257	35259
chrroi	35266	35268	chrroi	35266	35268
chrroi	35273	35275	chrroi	35273	35275
chrroi	35287	35289	chrroi	35287	35289
chrroi	37004	37006	chrroi	37004	37006
chrroi	37009	37011	chrroi	37009	37011

Fun with file formats

Chrom	Start	End	Name
chrroi	37553	38553	ENSG00000136859
chrroi	85184	86184	ENSG00000110934
chrroi	88182	chrroi	2271 2273 65
chrroi	157238	chrroi	2349 2351 47.5
chrroi	167073	chrroi	9788 9790 4.25
chrroi	177515	chrroi	9794 9796 38
chrroi	229319	chrroi	9804 9806 49
chrroi	239797	chrroi	9862 9864 46
chrroi	242795	chrroi	9922 9924 78.5
chrroi	250261	chrroi	9936 9938 81.33333333
chrroi	414709	chrroi	9958 9960 65.66666667
chrroi	420561	chrroi	12433 12435 24.33333333
chrroi	427763	chrroi	12460 12462 100
chrroi	619900	chrroi	12514 12516 71.66666667
chrroi	622898	chrroi	35213 35215 11.5
chrroi	805693	chrroi	35229 35231 56.5
chrroi	812937	chrroi	35257 35259 47
chrroi	813938	ENSG00000008853	
chrroi	849720	850721	ENSG00000137872
chrroi	1443843	1444844	ENSG00000139343
chrroi	1711347	1712347	ENSG00000166444
chrroi	1748149	1749149	ENSG00000137868
chrroi	1751147	1752148	ENSG00000171992



chrroi	85184	ENSG00000110934	chrroi	85189	85193	41.666666667
chrroi	85184	ENSG00000110934	chrroi	85201	85203	100
chrroi	85184	ENSG00000110934	chrroi	85229	85231	50
chrroi	85184	ENSG00000110934	chrroi	85234	85236	50
chrroi	85184	ENSG00000110934	chrroi	85243	85245	0
chrroi	88182	ENSG00000137948	.	-1	-1	.
chrroi	157238	ENSG00000121797	.	-1	-1	.
chrroi	167073	ENSG00000073067	chrroi	167297	167299	30.33333333
chrroi	167073	ENSG00000073067	chrroi	167301	167303	66
chrroi	167073	ENSG00000073067	chrroi	167364	167366	87.5
chrroi	167073	ENSG00000073067	chrroi	167373	167375	100
chrroi	167073	ENSG00000073067	chrroi	167387	167389	73
chrroi	177515	ENSG00000166801	.	-1	-1	.
chrroi	229319	ENSG00000269190	chrroi	229327	229331	0
chrroi	229319	ENSG00000269190	chrroi	229342	229344	0
chrroi	229319	ENSG00000269190	chrroi	229650	229652	0

Fun with file formats

chrgoi	85184	86184	ENSG00000110934	chrgoi	85189	85193	41.666666667
chrgoi	85184	86184	ENSG00000110934	chrgoi	85201	85203	100
chrgoi	85184	86184	ENSG00000110934	chrgoi	85229	85231	50
chrgoi	85184	86184	ENSG00000110934	chrgoi	85234	85236	50
chrgoi	85184	86184	ENSG00000110934	chrgoi	85243	85245	0
chrgoi	88182	89183	ENSG00000137948	.	-1	-1	.
chrgoi	157238	158239	ENSG00000121797	.	-1	-1	.
chrgoi	167073	168074	ENSG00000073067	chrgoi	167297	167299	30.333333333
chrgoi	167073	168074	ENSG00000073067	chrgoi	167301	167303	66
chrgoi	167073	168074	ENSG00000073067	chrgoi	167364	167366	87.5
chrgoi	167073	168074	ENSG00000073067	chrgoi	167373	167375	100
chrgoi	167073	168074	ENSG00000073067	chrgoi	167387	167389	73
chrgoi	177515	178516	ENSG00000166801	.	-1	-1	.
chrgoi	229319	230319	ENSG00000269190	chrgoi	229327	229331	0
chrgoi	229319	230319	ENSG00000269190	chrgoi	229342	229344	0
chrgoi	229319	230319	ENSG00000269190	chrgoi	229650	229652	0

chrgoi	37553	38553	ENSG00000136859	.
chrgoi	85184	86184	ENSG00000110934	40.27777778
chrgoi	88182	89183	ENSG00000137948	.
chrgoi	157238	158239	ENSG00000121797	.
chrgoi	167073	168074	ENSG00000073067	71.36666667
chrgoi	177515	178516	ENSG00000166801	.
chrgoi	229319	230319	ENSG00000269190	16.4
chrgoi	239797	240797	ENSG00000127951	.
chrgoi	242795	243796	ENSG00000189280	45
chrgoi	250261	251262	ENSG00000196208	46
chrgoi	414709	415710	ENSG00000170276	.
chrgoi	420561	421562	ENSG00000129009	.
chrgoi	427763	428764	ENSG00000172005	.
chrgoi	619900	620900	ENSG00000143850	2.135416667
chrgoi	623898	623899	ENSG00000132334	8.333333333
chrgoi	805693	806694	ENSG00000188060	33
chrgoi	812937	813938	ENSG00000008853	.
chrgoi	849720	850721	ENSG00000137872	0
chrgoi	1443843	1444844	ENSG00000139343	25.375
chrgoi	1711347	1712347	ENSG00000166444	16.29166667
chrgoi	1748149	1749149	ENSG00000137868	80
chrgoi	1751147	1752148	ENSG00000171992	36.66666667

Fun with file formats

chrgoi	37553	38553	ENSG00000136859	.	chrgoi	37553	38553	ENSG00000136859	.			
chrgoi	85184	86184	ENSG00000110934	34.57407407	chrgoi	85184	86184	ENSG00000110934	40.27777778			
chrgoi	88182	89183	ENSG00000137948	.	chrgoi	88182	89183	ENSG00000137948	.			
chrgoi	157238	158239	ENSG00000121797	.	chrgoi	157238	158239	ENSG00000121797	.			
chrgoi	167073	168074	ENSG00000073067	23.71428571	chrgoi	167073	168074	ENSG00000073067	71.36666667			
chrgoi	177515	178516	ENSG00000166801	.	chrgoi	177515	178516	ENSG00000166801	.			
chrgoi	229319	230319	ENSG00000269190	25.16666667	chrgoi	229319	230319	ENSG00000269190	16.4			
chrgoi	239797	240797	ENSG00000127951	.	chrgoi	239797	240797	ENSG00000127951	.			
chrgoi	242795	243796	ENSG00000189280	80	chrgoi	242795	243796	ENSG00000189280	45			
chrgoi	250261	251262	chrgoi	37553	38553	ENSG00000136859	0	chrgoi	37553	38553	ENSG00000136859	0
chrgoi	414709	415710	chrgoi	85184	86184	ENSG00000110934	34.57407407	chrgoi	85184	86184	ENSG00000110934	40.27777778
chrgoi	420561	421562	chrgoi	88182	89183	ENSG00000137948	0	chrgoi	88182	89183	ENSG00000137948	0
chrgoi	427763	428764	chrgoi	157238	158239	ENSG00000121797	0	chrgoi	157238	158239	ENSG00000121797	0
chrgoi	619900	620900	chrgoi	167073	168074	ENSG00000073067	23.71428571	chrgoi	167073	168074	ENSG00000073067	71.36666667
chrgoi	622898	623899	chrgoi	177515	178516	ENSG00000166801	0	chrgoi	177515	178516	ENSG00000166801	0
chrgoi	805693	806694	chrgoi	229319	230319	ENSG00000269190	25.16666667	chrgoi	229319	230319	ENSG00000269190	16.4
chrgoi	812937	813938	chrgoi	239797	240797	ENSG00000127951	0	chrgoi	239797	240797	ENSG00000127951	0
chrgoi	849720	850721	chrgoi	242795	243796	ENSG00000189280	80	chrgoi	242795	243796	ENSG00000189280	45
chrgoi	1443843	1444844	chrgoi	250261	251262	ENSG00000196208	29.53703704	chrgoi	250261	251262	ENSG00000196208	46
chrgoi	1711347	1712347	chrgoi	414709	415710	ENSG00000170276	15	chrgoi	414709	415710	ENSG00000170276	0
chrgoi	1748149	1749149	chrgoi	420561	421562	ENSG00000129009	0	chrgoi	420561	421562	ENSG00000129009	0
chrgoi	1751147	1752148	chrgoi	427763	428764	ENSG00000172005	17.83333333	chrgoi	427763	428764	ENSG00000172005	0
chrgoi	619900	620900	ENSG00000143850	5.416666667	chrgoi	619900	620900	ENSG00000143850	2.135416667			

Fun with file formats

ENSG00000136859	0.0
ENSG00000110934	5.70370371
ENSG00000137948	0.0
ENSG00000121797	0.0
ENSG00000073067	47.65238096
ENSG00000166801	0.0
ENSG00000269190	-8.76666667
ENSG00000127951	0.0
ENSG00000189280	-35.0
ENSG00000196208	16.46296296
ENSG00000170276	-15.0
ENSG00000129009	0.0
ENSG00000172005	-17.83333333
ENSG00000143850	-3.28125
ENSG00000132334	4.761904762
ENSG00000188060	9.66666667
ENSG0000008853	-19.0
ENSG00000137872	-3.111111111
ENSG00000139343	-0.45833333
ENSG00000166444	-42.64583333
ENSG00000137868	80.0
ENSG00000171992	36.66666667

Difference in promoter methylation

- 1) Compute mean CpG methylation for 1st & 3rd trimester replicates
 - Remove header line
 - Concatenate N=4 replicates
 - Sort files
 - Compute mean methylation reate of blocked (N=4) equal lines

Difference in promoter methylation

2) Intersect CpGs with promoter regions

- Promoter region intersection
- Compute mean promoter methylation rate

3) Compute difference between 1st & 3rd trimester

- Paste promoter methylation files horizontally
- Compute difference

1) Remove header line: tail

The screenshot shows the Galaxy web interface with the 'tail' tool selected. Red numbered arrows point to specific UI elements:

- 1: A red arrow points to the 'tail' tool icon in the 'Tools' sidebar.
- 2: A red arrow points to the 'Text Manipulation' section in the sidebar.
- 3: A red arrow points to the 'Text file' input field, which contains a list of dataset IDs.
- 4: A red arrow points to the 'Operation' dropdown menu, which is set to 'Keep everything from this line on'.
- 5: A red arrow points to the 'Number of lines' input field, which has the value '2'.
- 6: A red arrow points to the 'Execute' button at the bottom of the form.

The 'Text file' input field displays a batch mode input field with the following dataset IDs:

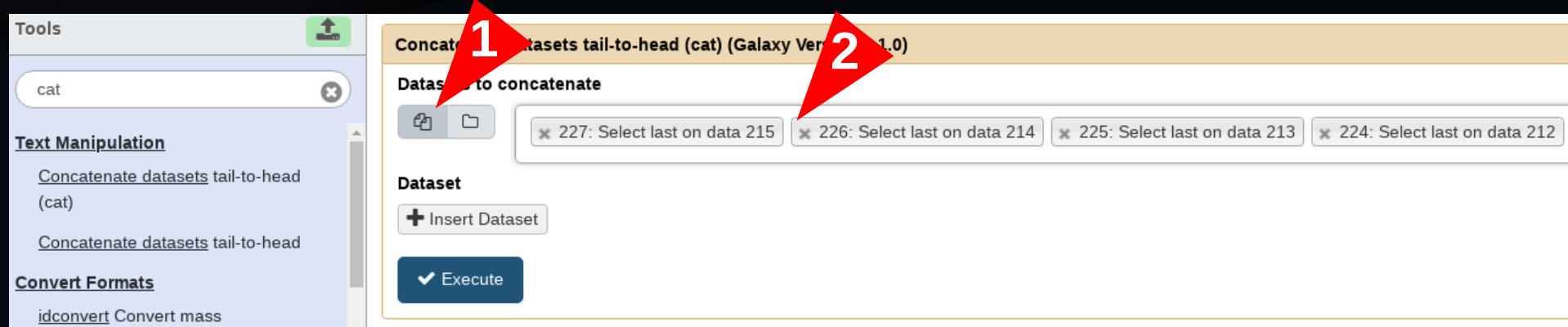
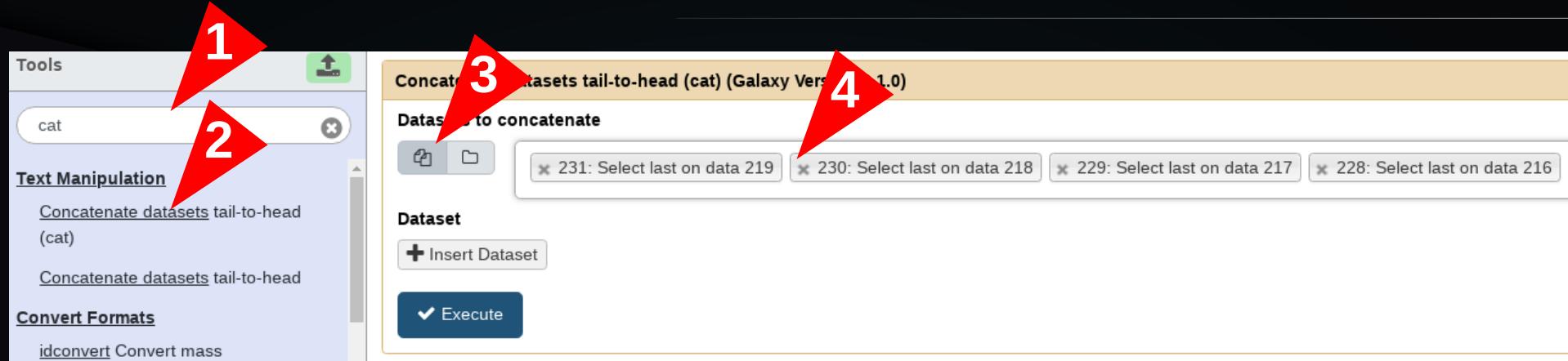
- 219: MethylDackel on data 211 and data 1
- 218: MethylDackel on data 210 and data 1
- 217: MethylDackel on data 209 and data 1
- 216: MethylDackel on data 208 and data 1
- 215: MethylDackel on data 207 and data 1
- 214: MethylDackel on data 206 and data 1
- 213: MethylDackel on data 205 and data 1
- 212: MethylDackel on data 204 and data 1

The 'Operation' dropdown menu contains the following options:

- Keep everything from this line on

The 'Number of lines' input field contains the value '2'. Below it, a note states: "These will be kept (depending on 'operation'). (--lines)".

1) Concatenate 1st & 3rd trimester: cat



Meta info on tracking down file origin

MethylDackel

Dataset Information

Number: 219
Name: MethylDackel on data 211 and data 1
Created: Mon Mar 4 19:58:00 2019 (UTC)
Filesize: 54.1 KB
Dbkey: ?
Format: bedgraph

Job Information

Galaxy Tool ID: toolshed.g2.bx.psu.edu/repos/bgruening/pileometh/pileometh/0.3.0.1
Galaxy Tool Version: 0.3.0.1
Tool Version: 0.3.0 (using HTSlib version 1.2.1)
Tool Standard Output: stdout
Tool Standard Error: stderr
Tool Exit Code: 0
History Content API ID: fdcb1fe8df798431
Job API ID: 3a0091bda5aac8da
History API ID: e3992a97a5c73694
UUID: 1d1157de-bfaa-4e3f-a8cd-08be2f60b9aa
Full Path: /data/dnb02/galaxy_db/files/008/514/dataset_8514149.dat

Tool Parameters

Input Parameter
Load reference genome from
Use the following dataset as the reference sequence
sorted_alignments.bam

	Value	Note for rerun
Load reference genome from	history	
Use the following dataset as the reference sequence	1: hg38.gzi.fa	
sorted_alignments.bam	211: Samtools sort on data 203	

The screenshot shows the Galaxy History panel with the following annotations:

- Red arrow 1 points to the dataset entry "219: MethylDackel on data 211 and data 1".
- Red arrow 2 points to the "eye" icon next to the dataset entry.
- Red arrow 3 points to the "display in IGB View" link.
- Red arrow 4 points to the "track type=bedGraph" entry in the tool parameters table.

1) Sort 1st & 3rd trimester: sort

The screenshot shows the Galaxy Sort tool interface with various configuration steps highlighted by red numbers:

- 1: Tools menu with "sort" selected.
- 2: "Text Manipulation" section with "Sort data in ascending or descending order" selected.
- 3: "Sort Query" button.
- 4: "Sort data in ascending or descending order" dropdown showing options like "234: Concatenate datasets on data 227, data 226, and others".
- 5: "Number of header lines" input field set to 0.
- 6: "Column selections" section for the first column.
- 7: "Flavor" section for the first column, with "Natural/Version sort (-V)" selected.
- 8: "Column selections" section for the second column.
- 9: "Flavor" section for the second column, with "Fast numeric sort (-n)" selected.
- 10: "Column selections" section for the third column.
- 11: "Flavor" section for the third column, with "Fast numeric sort (-n)" selected.

1) Get 1 & 3 t. mean: Bedtools merge

Tools

bedtools merge

BED

bedtools ClusterBed overlapping/nearby intervals

bedtools MergeBED combine overlapping/nearby intervals into a single interval

bedtools Merge BedGraph files combines coverage intervals from multiple BEDGRAPH files

Merge BedGraph files

Operate on Genomic Intervals

MergeBED combine overlapping/nearby intervals into a single interval

ClusterBed cluster overlapping/nearby intervals

Merge BedGraph files combines coverage intervals from multiple BEDGRAPH files

Workflows

- All workflows

bedtools Merge - combine overlapping/nearby intervals into a single interval (Galaxy Version 2.27.1)

Sort the following BAM/bed,bedgraph,gff,vcf file

236: Sort on data 234
235: Sort on data 233
234: Concatenate datasets on data 227, data 226, and others
233: Concatenate datasets on data 231, data 230, and others
231: Select last on data 219

This is a batch mode input field. Separate jobs will be triggered for each dataset selected.

Calculation based on strandedness?

Overlaps on either strand

Maximum distance between features allowed for features to be merged

0

That is, overlapping and/or book-ended features are merged. (-d)

Print the header from the A file prior to results

Yes No

(-header)

Applying operations to columns from merged intervals

1: Applying operations to columns from merged intervals

Specify the column(s) that should be summarized

Column: 4
(-c)

Specify the operation

Mean - numeric only

+ Insert Applying operations to columns from merged intervals

✓ Execute

2) Prom. intersection: bedtools intersect

Tools

bedtools intersect

BED

bedtools ClosestBed find the closest, potentially non-overlapping interval

bedtools ReDistBed calculate the distribution of relative distances

bedtools JaccardBed calculate the distribution of relative distances between two files

bedtools Multiple Intersect identifies common intervals among multiple interval files

bedtools WindowBed finds overlapping intervals within a window around an interval

bedtools Intersect intervals find overlapping intervals in various ways

bedtool 3: intersect intervals find overlapping intervals in various ways (Galaxy Version 2.27.1)

File A: intersect with B

3: hg38.promoter.bed

BAM/bed,bedgraph,gff,vcf format

Combined or separate output files

One output file per B' file

File(s) B to intersect with A

238: Merged Sort on data 234
237: Merged Sort on data 233
236: Sort on data 234
235: Sort on data 233
234: Concatenate datasets on data 227, data 226, and others

This is a batch mode input field. Separate jobs will be triggered for each dataset selection.

BAM/bed,bedgraph,gff,vcf format

Calculation based on strandedness?

Overlaps on either strand

What should be written to the output file?

Select/Unselect all

Perform a "left outer join". That is, for each feature in A report each overlap with B. If no overlaps are found, report a NULL feature for B (-loj)

2) Get prom mean: bedtools merge

Tools

bedtools merge

BED

- [bedtools LinksBed](#) create a HTML page of links to UCSC locations
- [bedtools BED12 to BED6 converter](#)
- [bedtools FlankBed](#) create new intervals from the flanks of existing intervals
- [bedtools ReldistBed](#) calculate the distribution of relative distances
- [bedtools GetFastaBed](#) get intervals to extract sequences from a FASTA file
- [bedtools BAM to BED converter](#)
- [bedtools MergeBED](#) combine overlapping/nearby intervals into a single interval
- [bedtools SortBED](#) order the intervals
- [bedtools MapBed](#) apply a function to a column for each overlapping interval
- [bedtools GroupByBed](#) group by common cols and summarize other cols
- [bedtools Convert from BAM to FastQ](#)
- [bedtools ExpandBed](#) replicate lines based on lists of values in columns
- [bedtools AnnotateBed](#) annotate coverage of features from multiple

bedtools MergeBED combine overlapping/nearby intervals into a single interval (Galaxy Version 2.27.1)

Sort the following BAM/bed,bedgraph,gff,vcf files

240: bedtools Intersect intervals on data 238 and data 3
239: bedtools Intersect intervals on data 237 and data 3
238: Merged Sort on data 234
237: Merged Sort on data 233
236: Sort on data 234

This is a batch mode input field. Separate jobs will be triggered for each dataset selected.

Calculation based on strandedness?

Overlaps on either strand

Maximum distance between features allowed for features to be merged

0

That is, overlapping and/or book-ended features are merged. (-d)

Print the header from the A file prior to results

Yes No

(-header)

Applying operations to columns from merged intervals

1: Applying operations to columns from merged intervals

Specify the column(s) that should be summarized

Column: 4
(-c)

Specify the operation

distinct (i.e., print a comma separated list) - numeric output

2: Applying operations to columns from merged intervals

Specify the column(s) that should be summarized

Column: 8
(-c)

Specify the operation

Mean - numeric only

+ Insert Applying operations to columns from merged intervals

3) Paste 1 & 3 t. prom.: Paste

Tools 1

paste 2

Text Manipulation

Paste two files side by side

Filter and Sort

Filter sequences by ID from a tabular file

Variant Calling

SNPEFF

SnpEff build: database from Genbank

Paste two files side by side (Galaxy Version 1.0.0) 4

Paste 3

and 4

Delimited 5

Tab 6

Execute

3) Replace '.' by 0: replace column

The screenshot shows the Biopython Text Processing tool interface. The main window is titled "Replace in a specific column" Version 1.1.3. The sidebar on the left lists various tools under categories like "Text Manipulation", "Filter and Sort", and "Convert Formats". Red numbered callouts point to specific elements:

- 1: The "Tools" button in the top-left corner of the sidebar.
- 2: The "replace" search bar in the top-left corner of the sidebar.
- 3: The title bar of the main window.
- 4: The "File to process" dropdown menu in the top-left corner of the main window.
- 5: The "Column: 5" input field in the "in column" section of the first replacement step.
- 6: The "Find pattern" input field containing the regular expression "\.".
- 7: The "Replace with" input field containing the value "0".
- 8: The "Column: 10" input field in the "in column" section of the second replacement step.
- 9: The "Find pattern" input field containing the regular expression "\.".
- 10: The "Replace with" input field containing the value "0".

The bottom right of the main window has a "Execute" button with a checkmark icon.

3) Subtract 1 & 3 prom. meth.: compute

The screenshot shows the Galaxy web interface with the 'Tools' panel open on the left and a specific tool configuration on the right.

Tools Panel:

- Search bar: 'compute' (highlighted by red arrow 1)
- Text Manipulation section:
 - [Compute an expression on every row](#)
- FASTA/FASTQ section:
 - [Compute sequence length](#)
 - [FLASH adjust length of short reads](#)
- BED section:
 - [bedtools Compute both the depth and breadth of coverage of features in file B on the features in file A \(bedtools coverage\)](#)
 - [bedtools OverlapBed computes the amount of overlap from two](#)

Tool Configuration Panel:

Compute an expression on every row (Galaxy Version 1.2.0)

Add expression: c10-c5 (highlighted by red arrow 3)

as a new column to: (highlighted by red arrow 4)

Dataset missing? See TIP below
246: Replace Text on data 243 (highlighted by red arrow 5)

Round result?: NO

Skip a header line?: no

characters are already considered as comments and kept

Execute button

Cut columns: cut

Tools

1

cut

2

Text Manipulation

- Replace Text in a line
- Text reformatting with awk
- Cut columns from a table (cut)
- Cut columns from a table

Filter and Sort

- Filter sequences by ID from a tabular file
- Select lines that match an expression

FASTA/FASTQ

- Filter by quality
- fastp - fast all-in-one preprocessing for FASTQ files

3 Cut columns from a table (cut) (Galaxy Version 1.1.0)

4 File to cut

247: Compute on data 246

5 Operation

Keep

Delimited by

Tab

Cut by

6 fields

List of Fields

Select/Unselect all

Column: 4 Column: 11
(-f)

Execute

Tag results

Edit dataset attributes

Attributes Convert Datatypes Permissions

Edit attribute **3**

Name: DME.tsv

Info

Annotation

Auto-detect Save

History

search datasets

Rostock

240 shown, 8 hidden
866 MB

248: Cut off data 247

22 lines
format: tabular, database: 2

Annotations

1	2
ENSG00000136859	0.0
ENSG00000110934	-5.70370371
ENSG00000137948	0.0
ENSG00000121797	0.0
ENSG00000073067	-47.65238096

Visualization



Join DGE & DME data: join

The screenshot shows the Galaxy 'join' tool interface. A sidebar on the left lists various tools and operations, with 'join' selected. The main panel is titled 'Join two files' (Galaxy Version 1.1.1). It has sections for '1st file' (DGE.tsv) and '2nd File' (DME.tsv), both with dropdown menus for selecting columns. There are also sections for 'Column to use from 1st file', 'Column to use from 2nd file', and 'Output lines appearing in'. Other settings include 'First line is a header' (No), 'Ignore case' (Yes), and 'Value to put in unpaired (empty) fields' (0). A large red arrow labeled '1' points to the 'join' tool in the sidebar. Red arrows labeled '2' through '8' point to the '1st file' section, '2nd File' section, 'Column to use from 1st file' dropdown, 'Column to use from 2nd file' dropdown, 'Output lines appearing in' dropdown, 'First line is a header' button, 'Ignore case' button, and the 'Value to put in unpaired (empty) fields' input field respectively.

Tools

join

Collection Operations

- Column Join on Collections

Text Manipulation

- Histogram of a numeric column
- Join two files on column allowing a small difference
- Subtract the intervals of two datasets
- Sort data in ascending or descending order
- Join two files
- Multi-Join (combine multiple files)
- Join the intervals of two datasets side-by-side

Filter and Sort

- Query Tabular using sqlite sql

Join, Subtract and Group

- Datamash (operations on tabular data)
- Reverse columns in a tabular file
- Transpose rows/columns in a

Join two files (Galaxy Version 1.1.1)

1st file: 253: DGE.tsv

Column to use from 1st file

2nd File: DME.tsv

Column to use from 2nd file

Output lines appearing in

Both 1st & 2nd file.

First line is a header line

Yes No

Ignore case

Yes No

Value to put in unpaired (empty) fields

0

Execute

Download table

The screenshot illustrates the process of downloading a dataset from the Galaxy platform.

Left Panel (Download Dialog):

- A modal dialog box is displayed, asking "Download 'Galaxy254-[Join_on_data_248_and_data_253].tabular'?"
- The file name is "Galaxy254-[Join_on_data_248_and_data_253].tabular".
- The file size is 930 bytes.
- An option "Always Save Files to Default Download Location" is checked.
- Buttons at the bottom include "Save" (highlighted with a red arrow labeled 3), "Save As...", "Open", and "Cancel".

Right Panel (History):

- A red arrow labeled 1 points to the dataset entry in the history list.
- A red arrow labeled 2 points to the "Edit" icon (pencil) next to the dataset entry.
- The dataset entry details:
 - 254: Join on data 248 and data 253**
 - 22 lines
 - format: tabular, database: ?
 - Icons: save, info, refresh, chart, help, edit, delete, comment
- The dataset table preview shows columns 1, 2, and 3 with the following data:

1	2	3
ENSG00000008853	0.542935631009544	19
ENSG00000073067	1.88003372546044	-4
ENSG00000110934	2.2320841512721	-5
ENSG00000121797	0.116862409130137	0.
ENSG00000127951	0.343302983265895	0.

Rstudio

<https://gen100.leibniz-fli.de/rstudio-demo/>

or <https://rstudio.cloud>

```
library('ggpubr');
library('ggExtra');
df=read.table('scatter.tsv', header = FALSE, sep = '\t');
colnames(df) = c('gene', 'DGE', 'DME');
p = ggscatter(df, x = 'DME', y = 'DGE', color = 'DGE', label = 'gene', repel = TRUE) +
  gradient_color(c('blue', 'white', 'red')) +
  geom_density_2d();
pdf('scatter.pdf');
ggMarginal(p, type = 'boxplot');
graphics.off();
```

Rstudio

The screenshot shows the RStudio interface with several numbered callouts:

- 1**: Points to the "Upload" button in the "Files" tab of the sidebar.
- 2**: Points to the code editor window showing R script content.
- 3**: Points to the "Run" button in the toolbar.
- 4**: Points to the "Cloud > project" section in the sidebar.

Code Editor Content:

```
library('ggpubr');
library('ggExtra');
df=read.table('scatter.tsv', header = FALSE, sep = '\t');
colnames(df) = c('gene', 'DGE', 'DME');
p = ggscatter(df, x = 'DME', y = 'DGE', color = 'DGE', label = 'gene', repel = TRUE) +
  gradient_color(c('blue', 'white', 'red')) +
  geom_density_2d();
pdf('scatter.pdf');
ggMarginal(p, type = 'boxplot');
graphics.off();
```

Console Output:

```
> library('ggpubr');
> library('ggExtra');
> df=read.table('scatter.tsv', header = FALSE, sep = '\t');
> colnames(df) = c('gene', 'DGE', 'DME');
> p = ggscatter(df, x = 'DME', y = 'DGE', color = 'DGE', label = 'gene', repel = TRUE) +
+   gradient_color(c('blue', 'white', 'red')) +
+   geom_density_2d();
> pdf('scatter.pdf');
> ggMarginal(p, type = 'boxplot');
> graphics.off();
>
```

Upload Files Dialog:

Target directory:
/cloud/project

File to upload:
Choose File No file chosen

TIP: To upload multiple files or a directory, create a zip file.
The zip file will be automatically expanded after upload.

OK Cancel

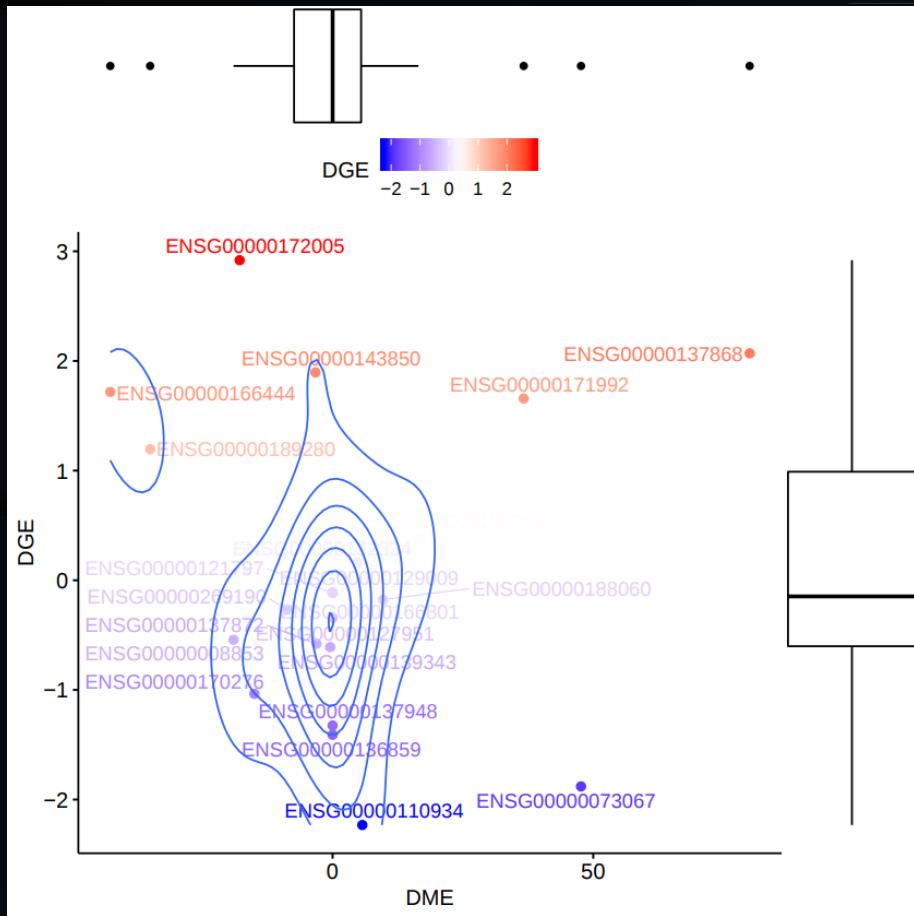
Environment Tab Data:

Object	Type	Description
df	22 obs. of 3 variables	Global Environment
p	List of 9	Global Environment

Files Tab Content:

- ..
- .Rhistory
- project.Rproj
- scatter.pdf
- scatter.tsv

Scatter plot



Possible impact of epigenetic modification on expression level of top DEGs:

UP (Q2)

MAL: T cell differentiation protein
ST5: Tumor suppressor
GJB5: Intercellular signalling

DOWN (Q4)

CYP2W1: hormone synthesis, vitamin D metabolism
BIN2: Onkogen
RAB42: Onkogen

UP (Q1)

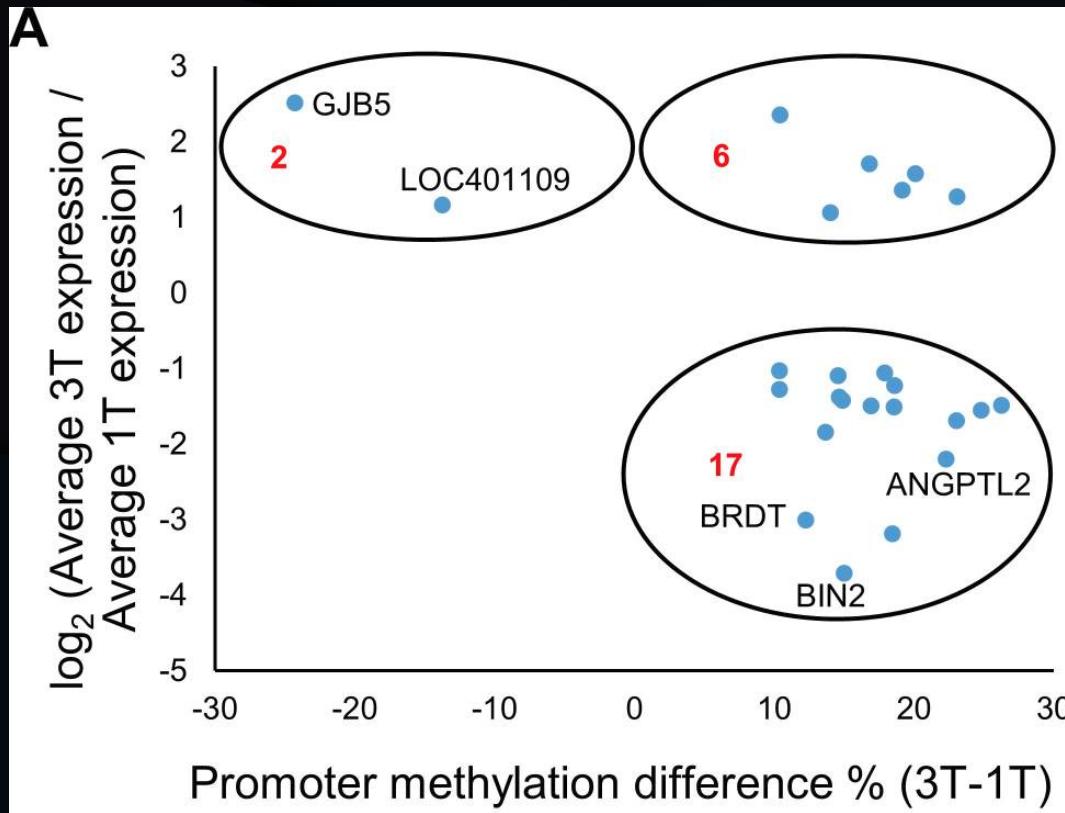
STRA6: Vitamin A metabolism
SYNPO: Cell motility

DOWN (Q3)

HSPB2: Heat shock protein
RHOBTB2: Tumor suppressor
SEMA6D: Secretome protein

- Enriched immune system processes due to fetus protective function of placenta
- Vitamins A and D known modulate pro-inflammatory immune response

Comparison with published plot



- 5% vs 100%
- DESeq2 vs RPKM + Mann-Whitney-U test
- Methyldackel vs ??
 - Similar extrema