Identification of somatic and germline variants from tumor and normal sample pairs

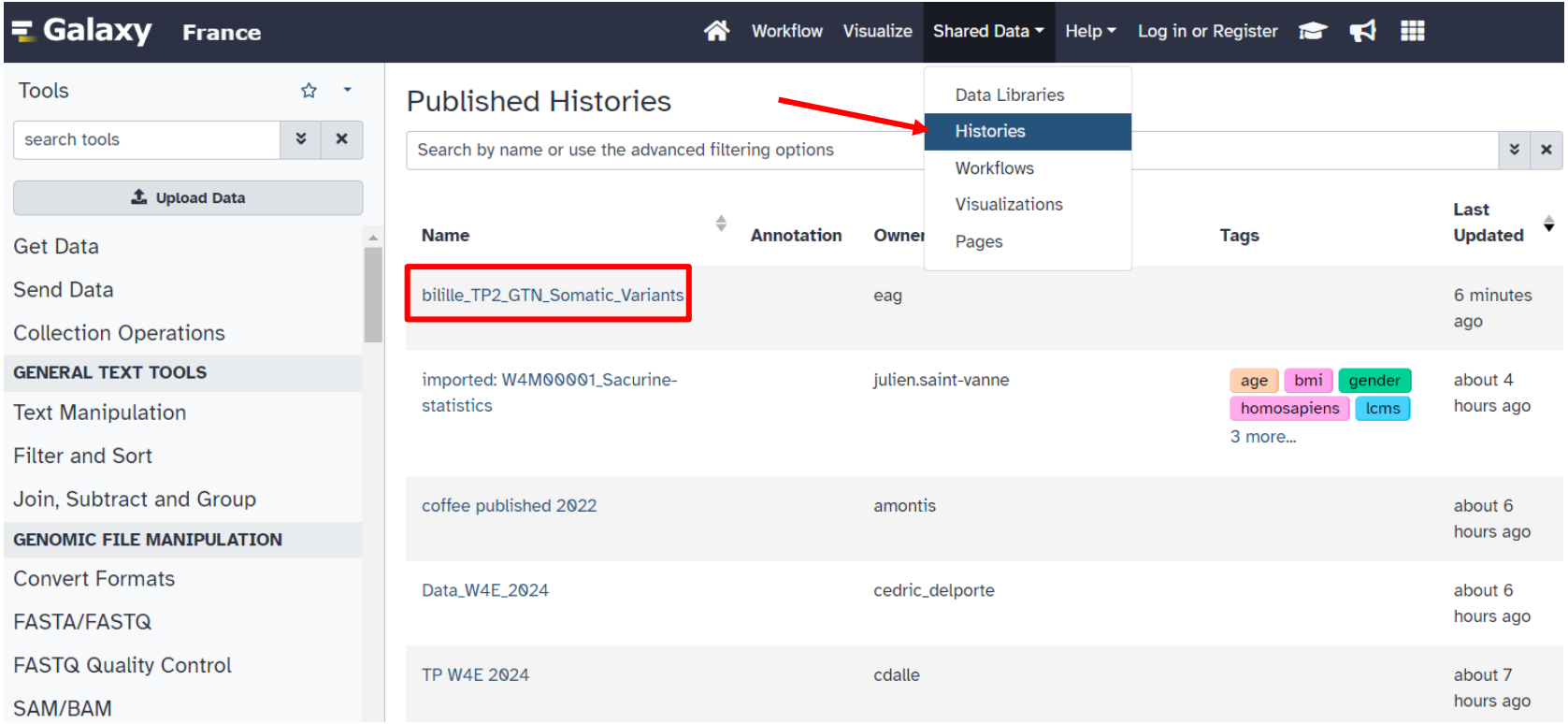
Somatic variants tutorial

Workflow

1. Mapped reads postprocessing
    a. Filtering on mapped reads properties
    b. Removing duplicate reads
    c. Left-align reads around indels
    d. Recalibrate read mapping qualities
    e. Refilter reads based on mapping quality
2. Variant calling and classification
3. Variant annotation and reporting
    a. Adding annotations to the called variants
    b. Reporting selected subsets of variants
    c. Generating reports of genes affected by variants
    d. Adding additional annotations to the gene-centered report

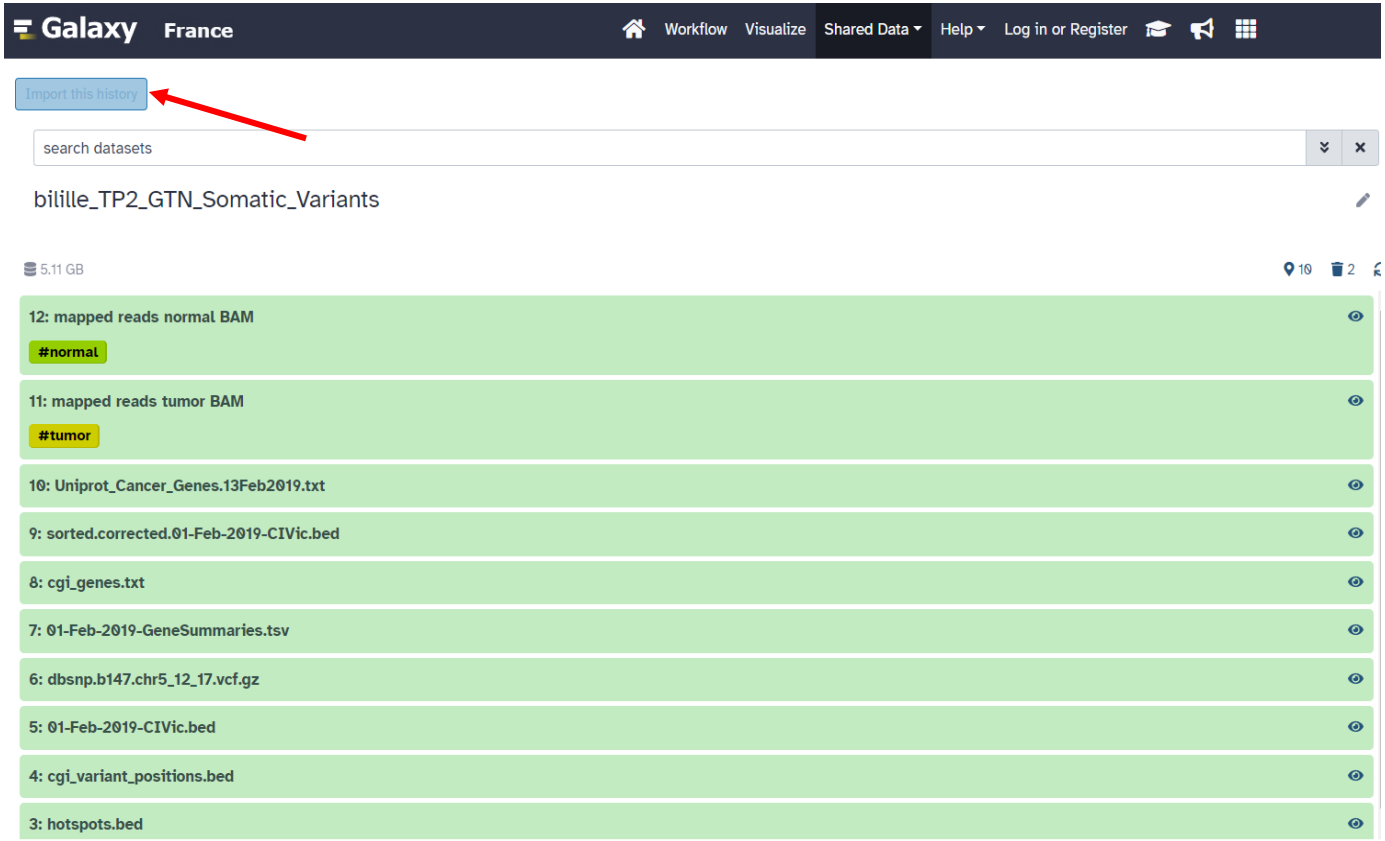
Starting from BAMs : Import Shared History

https://usegalaxy.fr/u/eag/h/somatic-tutorial



Starting from BAMs : Import Shared History

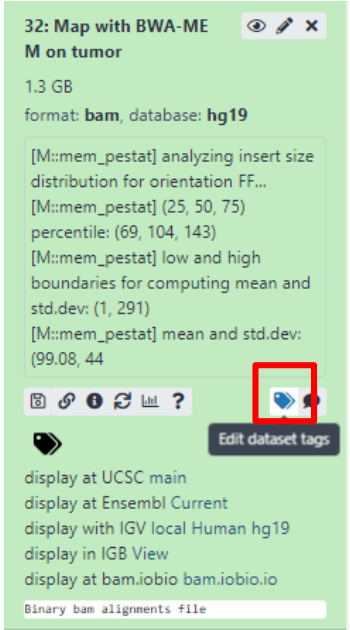
https://usegalaxy.fr/u/eag/h/somatic-tutorial

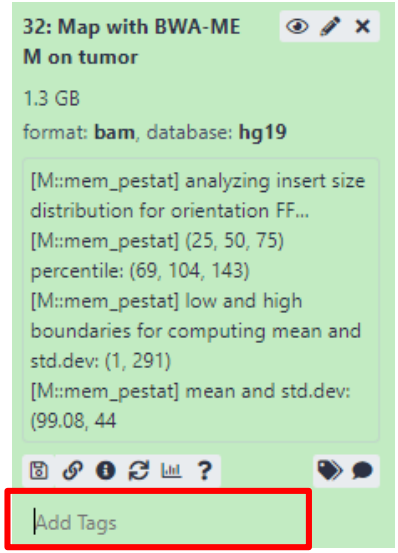


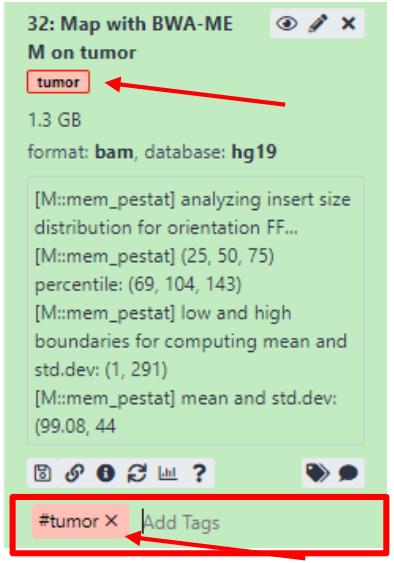
Prepare Data

Tip: Adding a tag

- Click on the dataset
- Click on **Edit dataset tags**
- Add a tag starting with `#`
  Tags starting with `#` will be automatically propagated to the outputs of tools using this dataset.
- Check that the tag is appearing below the dataset name







1. Mapped reads postprocessing

1. Mapped reads postprocessing

a. Filtering on mapped reads properties

Filtering for mapping status and quality

Galaxy Europe | Analyze Data | Workflow | Visualize | Shared Data | Help | User | Using 12%

Tools

Column arrange by header name

Upload Data

Hide Sections

Filter and Sort

- **Column arrange** by header name
- **Filter** data on any column using simple expressions
- **Filter GTF data by attribute values_list**
- **Filter sequences by ID** from a tabular file

Text Manipulation

- **Add input name as column** to an existing tabular file
- **Sort Column Order** by heading
- **Replace column** by values which are defined in a convert file
- **Replace Text** in a specific column
- **Replace chromosome names** in a tabular dataset using a mapping table
- **Histogram** of a numeric column
- **Add column** to an existing dataset
- **Add column** to an existing dataset
- **Join two files** on column allowing a small difference
- **Column Regex Find And Replace**
- **Add line to file** writes a line of text at the begining or end of a text file.

Filter BAM datasets on a variety of attributes (Galaxy Version 2.4.1) | Added | Versions | Options

BAM dataset(s) to filter

75: RmDup on data 73
74: RmDup on data 71
73: Filter on data 32: Filtered BAM
32: Map with BWA-MEM on tumor
31: Map with BWA-MEM on normal

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

Condition

1: Condition

Filter

1: Filter

Select BAM property to filter on

mapQuality

Filter on read mapping quality (phred scale)

>=1

You can use >, <, =, and ! (not) in your expression. E.g., to select reads with mapping quality of at least 30 use ">=30"

2: Filter ***"Insert Filter"***

Select BAM property to filter on

isMapped

Selected mapped reads

Yes

Checked = Mapped, Empty = NOT mapped

3: Filter

History

search datasets

Tumor Normal pair somatic pipeline TEST

73 shown, 41 deleted, hide hidden

31.3 GB

- tumor
- 76: BamLeftAlign on data 74 (alignments) — normal
- 75: RmDup on data 73 — tumor
- 74: RmDup on data 71 — normal
- 73: Filter on data 32: Filtered BAM — tumor
- 72: Filter on data 32: JSON filter rules — tumor
- 31: Filt
- 31: JS
- GeneS

In "3: Filter":

- ***"Select BAM property to filter on":*** `isMateMapped`
  - ***"Select reads with mapped mate":*** `Yes`

Filtering for mapping status and quality

There is not only one tool that can filter reads.

To Do : find another tool in Galaxy to perform the same operation

Filtering for mapping status and quality

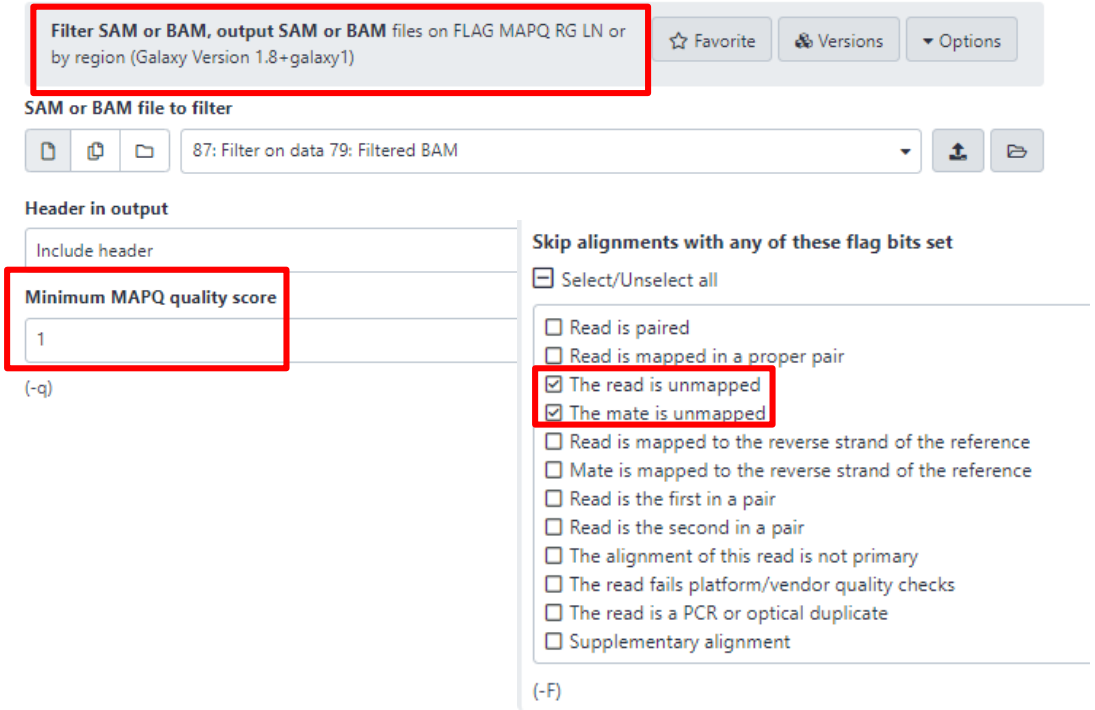
There is not only one tool that can filter reads.

To Do : find another tool in Galaxy to perform the same operation

Filter SAM or BAM, output SAM or BAM
based on samtools view

equivalent to

Filter BAM datasets on a variety of attributes
Based on bamtools filter



Mapped reads postprocessing

b. Removing duplicate reads

Remove duplicates with RmDup

Galaxy Europe | Analyze Data | Workflow | Visualize | Shared Data | Help | User | Using 12%

Tools

Column arrange by header name

Upload Data

Hide Sections

Filter and Sort

- **Column arrange** by header name
- **Filter** data on any column using simple expressions
- **Filter GTF data by attribute values_list**
- **Filter sequences by ID** from a tabular file

Text Manipulation

- **Add input name as column** to an existing tabular file
- **Sort Column Order** by heading
- **Replace column** by values which are defined in a convert file
- **Replace Text** in a specific column
- **Replace chromosome names** in a tabular dataset using a mapping table
- **Histogram** of a numeric column
- **Add column** to an existing dataset
- **Add column** to an existing dataset
- **Join two files** on column allowing a small difference
- **Column Regex Find And Replace**
- **Add line to file** writes a line of text at the begining or end of a text file.

RmDup remove PCR duplicates (Galaxy Version 2.0.1) | Favorite | Versions | Options

BAM File

75: RmDup on data 73
74: RmDup on data 71
73: Filter on data 32: Filtered BAM
71: Filter on data 31: Filtered BAM
31: Map with BWA-MEM on normal

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

Is this paired-end or single end data

BAM is paired-end

Treat as single-end

No

(-S)

Email notification

No

Send an email notification when the job completes.

Execute

What it does

Remove potential PCR duplicates: if multiple read pairs have identical external coordinates, only retain the pair with highest mapping quality. In the paired-end mode, this command ONLY works with FR orientation and requires ISIZE is correctly set. It does not work for unpaired reads (e.g. two ends mapped to different chromosomes or orphan reads).

Citations:

- *Definition of SAM/BAM format*. (n.d.). Retrieved from https://samtools.github.io/hts-specs/
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... and, R. D. (2009). The Sequence

History

search datasets

Tumor Normal pair somatic pipeline TEST

73 shown, 41 deleted, hide hidden

31.3 GB

tumor

76: BamLeftAlign on data 74 (alignments)
normal

75: RmDup on data 73
tumor

74: RmDup on data 71
normal

746.7 MB
format: **bam**, database: **hg19**

[bam_rmdup_core] processing reference chr10...
[bam_rmdup_core] 1 unmatched pairs
[bam_rmdup_core] processing reference chr11...
[bam_rmdup_core] 3 unmatched pairs
[bam_rmdup_core] processing reference chr12...
[bam_rmdup_core] inconsistent BAM file for

display at UCSC main
display at Ensembl Current
display with IGV local Human hg19

Mapped reads postprocessing

c. Left-align reads around indels

Left-align with BamLeftAlign

Galaxy Europe

Analyze Data | Workflow | Visualize | Shared Data | Help | User

Using 9%

Tools

BamLeftAlign

Upload Data

Hide Sections

VCF/BCF

BamLeftAlign indels in BAM datasets

Variant Calling

BamLeftAlign indels in BAM datasets

WORKFLOWS

All workflows

BamLeftAlign indels in BAM datasets (Galaxy Version 1.3.1)

Added | Versions | Options

Choose the source for the reference genome

Locally cached

Select alignment file in BAM format

75: RmDup on data 73

Using reference genome

Human (Homo sapiens): hg19

(--fasta-reference)

Maximum number of iterations

5

Iterate the left-realignment no more than this many times (--max-iterations)

Email notification

No

Send an email notification when the job completes.

Execute

When calling indels, it is important to homogenize the positional distribution of insertions and deletions in the input by using left realignment. Left realignment will place all indels in homopolymer and microsatellite repeats at the same position, provided that doing so does not introduce mismatches between the read and reference other than the indel. This method is computationally inexpensive and handles the most common classes of alignment inconsistency.

This is leftalign utility from FreeBayes package.

Citations:

- (N.d.). Retrieved from http://arxiv.org/abs/1207.3907

Requirements:

- freebayes (Version 1.3.1)
- samtools (Version 1.9)

History

search datasets

Tumor Normal pair somatic pipeline TEST

32 shown, 34 deleted, 13 hidden

22.87 GB

76: BamLeftAlign on data 74 (alignments)

746.7 MB

format: **bam**, database: **hg19**

display at UCSC main
display at Ensembl Current
display with IGV local Human hg19
display in IGB View
display at bam.iobio bam.iobio.io

Binary bam alignments file

75: RmDup on data 73

74: RmDup on data 71

73: Filter on data 32: Filtered BAM

72: Filter on data 32: JSON filter rules

71: Filter on data 31: Filtered BAM

normal

70: Filter on data 31: JSON filter rules

normal

32: Map with BWA-MEM on tumor

tumor

31: Map with BWA-MEM on normal

Mapped reads postprocessing

d. Recalibrate read mapping qualities

Recalibrate read quality scores with CalMD

Galaxy Europe

Analyze Data | Workflow | Visualize | Shared Data | Help | User | Using 13%

Tools

filter

Upload Data

Hide Sections

Sharpen

deepTools

estimateReadFiltering estimates the number of reads that would be filtered given certain criteria

alignmentsieve Filter BAM/CRAM files according to specified parameters

SAM/BAM

Filter BAM datasets on a variety of attributes

BAM filter Removes reads from a BAM file based on criteria

Samtools view - reformat, filter, or subsample SAM, BAM or CRAM

Filter SAM or BAM, output SAM or BAM files on FLAG MAPQ RG LN or by region

MiModD

MiModD VCF Filter extracts lines from a vcf variant file based on field-specific filters

Metagenomic Analysis

dada2: filterAndTrim Filter and trim short read data

sixgill filter a metapeptide database

khmer: Filter reads by minimal k-

CalMD recalculate MD/NM tags (Galaxy Version 2.0.2)

Added | Versions | Options

BAM file to recalculate

77: BamLeftAlign on data 75 (alignments)
76: BamLeftAlign on data 74 (alignments)
75: RmDup on data 73
74: RmDup on data 71
73: Filter on data 32: Filtered BAM
71: Filter on data 31: Filtered BAM
32: Map with BWA-MEM on tumor

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

Choose the source for the reference genome

Use a built-in genome

Using reference genome

Human (Homo sapiens): hg19

Do you also want BAQ (Base Alignment Quality) scores to be calculated?

No

(-r)

Additional options

Advanced options

Change identical bases to '='

No

Replace bases in read sequences that match the reference base at that position with an equal sign (-e)

Coefficient to cap mapping quality of poorly mapped reads

50

Higher values for this setting mean a stronger downgrade of the mapping quality of reads with excessive mismatches (50: recommended setting for reads aligned with BWA, 0: do not downgrade mapping qualities) (-C)

Email notification

No

Send an email notification when the job completes.

History

search datasets

Tumor Normal pair somatic Tutorial

40 shown, 45 deleted, 29 hidden

31.3 GB

84: Filter on data 78: JSON filter rules
normal

79: CalMD on data 77
tumor

78: CalMD on data 76
normal

754.4 MB

format: **bam**, database: **hg19**

[bam_fillmd1] different NM for read 'ST-K00265:137:HT33CBBXX:3:2115:18345:2
0 -> 1
[bam_fillmd1] different MD for read 'ST-K00265:137:HT33CBBXX:3:2115:18345:2
'98' -> '33n64'
[bam_fillmd1] different MD for read 'ST-K00265:137:HT33CBBXX:3:1127

display at UCSC main
display at Ensembl Current
display with IGV local Human hg19
display in IGB View
display at bam.iobio bam.iobio.io

Binary bam alignments file

Mapped reads postprocessing

e. Refilter reads based on mapping quality

Eliminating reads with undefined mapping quality

Galaxy Europe

Analyze Data Workflow Visualize Shared Data Help User

Tools

Filter BAM

Upload Data

Hide Sections

bedtools Genome Coverage compute the coverage over an entire genome

SAM/BAM

BedCov calculate read depth for a set of genomic intervals

BAM filter Removes reads from a BAM file based on criteria

Convert, Merge, Randomize BAM datasets and perform other transformations

BamHash Hash BAM and FASTQ files to verify data integrity

Filter BAM datasets on a variety of attributes

CalMD recalculate MD/NM tags

FASTA/FASTQ

Create binary barcodes from regular barcodes.

Extract barcodes according to pattern

Barcode Splitter

UMI-tools whitelist Extract cell barcodes from FASTQ files

Filter sequences by mapping from SAM/BAM file

Je-Demultiplex demultiplexes fastq files

Extract alignment ends from SAM or BAM

Proteomics

Filter BAM datasets on a variety of attributes (Galaxy Version 2.4.)

Added Versions Options

BAM dataset(s) to filter

79: CalMD on data 77
78: CalMD on data 76
77: BamLeftAlign on data 75 (alignments)
76: BamLeftAlign on data 74 (alignments)
75: RmDup on data 73
74: RmDup on data 71

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

Condition

1: Condition

Filter

1: Filter

Select BAM property to filter on

mapQuality

Filter on read mapping quality (phred scale)

<=254

You can use >, <, =, and ! (not) in your expression. E.g., to select reads with mapping quality of at least 30 use ">=30"

+ Insert Filter

+ Insert Condition

Would you like to set rules?

No

Allows complex logical constructs. See Example 4 below.

Email notification

No

Send an email notification when the job completes.

Execute

What is does

BAMTools filter is a very powerful utility to perform complex filtering of BAM files. It is based on BAMtools suite of tools by Derek Barnett (https://github.com/pezmaster31/bamtools).

How it works

2. Variant calling and classification

Variant calling with VarScan somatic

Galaxy Europe

Analyze Data | Workflow | Visualize | Shared Data | Help | User

Tools

VarScan

Upload Data

Hide Sections

VCF/BCF

VarScan somatic Call germline/somatic and LOH variants from tumor-normal sample pairs

VarScan mpileup for variant detection

VarScan copynumber Determine relative tumor copy number from tumor-normal pileups

VarScan for variant detection

Variant Calling

VarScan mpileup for variant detection

VarScan copynumber Determine relative tumor copy number from tumor-normal pileups

VarScan somatic Call germline/somatic and LOH variants from tumor-normal sample pairs

WORKFLOWS

All workflows

VarScan somatic Call germline/somatic and LOH variants from tumor-normal sample pairs (Galaxy Version 2.4.3.6)

Added | Versions | Options

Will you select a reference genome from your history or use a built-in genome?

Use a built-in genome

reference genome

Human (Homo sapiens): hg19

The fasta reference genome that variants should be called against.

aligned reads from normal sample

87: Filter on data 79: Filtered BAM

aligned reads from tumor sample

87: Filter on data 79: Filtered BAM

Estimated purity (non-tumor content) of normal sample

1

(--normal-purity)

Estimated purity (tumor content) of tumor sample

0.5

(--tumor-purity)

Generate separate output datasets for SNP and indel calls?

No

Settings for Variant Calling

Use default values

Settings for Posterior Variant Filtering

Use default values

Compatibility options for experts

Email notification

No

Send an email notification when the job completes.

Execute

VarScan Overview

Variant calling with VarScan somatic

Galaxy Europe

Analyze Data | Workflow | Visualize | Shared Data | Help | User

Tools

VarScan

Upload Data

Hide Sections

VCF/BCF

VarScan somatic Call germline/somatic and LOH variants from tumor-normal sample pairs

VarScan mpileup for variant detection

VarScan copynumber Determine relative tumor copy number from tumor-normal pileups

VarScan for variant detection

Variant Calling

VarScan mpileup for variant detection

VarScan copynumber Determine relative tumor copy number from tumor-normal pileups

VarScan somatic Call germline/somatic and LOH variants from tumor-normal sample pairs

WORKFLOWS

All workflows

VarScan somatic Call germline/somatic and LOH variants from tumor-normal sample pairs (Galaxy Version 2.4.3.6)

Added | Versions | Options

Will you select a reference genome from your history or use a built-in genome?

Use a built-in genome

reference genome

Human (Homo sapiens): hg19

The fasta reference genome that variants should be called against.

aligned reads from normal sample

87: Filter on data 79: Filtered BAM

aligned reads from tumor sample

87: Filter on data 79: Filtered BAM

Estimated purity (non-tumor content) of normal sample

1

(--normal-purity)

Estimated purity (tumor content) of tumor sample

0.5

(--tumor-purity)

Generate separate output datasets for SNP and indel calls?

No

Settings for Variant Calling

Settings for Posterior Variant Filtering

Use default values

Compatibility options for experts

Email notification

No

Send an email notification when the job completes.

Execute

Settings for Variant Calling

Customize settings

Read selection

Minimum base quality

28

The minimum base quality (default: 13) at a given position required to use a read for calling variants at that site (samtools mpileup -Q)

Minimum mapping quality

1

The minimum mapping quality (default: 0) required for a read to be considered in variant calling (samtools mpileup -q)

VarScan Overview

3. Variant annotation and reporting

Adding annotations to the called variants

a. Adding annotations to the called variants

a.1. Adding functional genomic annotations

Adding annotations with SnpEff

Galaxy Europe

Analyze Data | Workflow | Visualize | Shared Data | Help | User

Tools

SnpEff eff

Upload Data

Hide Sections

VCF/BCF

snippy Snippy finds SNPs between a haploid reference genome and your NGS sequence reads.

snippy-core Combine multiple Snippy outputs into a core SNP alignment

ococo consensus caller on SAM/BAM

Variant Calling

snippy-core Combine multiple Snippy outputs into a core SNP alignment

snippy-clean_full_aln Replace any non-standard sequence characters in snippy 'core.full.aln' file.

SnpEff eff: annotate variants for SARS-CoV-2

SnpEff eff: annotate variants

SnpEff databases: list available databases

SnpEff download: download a pre-built database

Variant Frequency Plot Generates a heatmap of allele frequencies grouped by variant type for SnpEff-annotated SARS-CoV-2 data

SnpEff chromosome-info: list chromosome names/lengths

SnpEff build: database from Genbank or GFF record

SnpEff to Peptide fasta to create a Search DB fasta for variant SAP peptides

SnpEff eff: annotate variants (Galaxy Version 4.3+T.galaxy1)

Favorite | Versions | Options

Sequence changes (SNPs, MNPs, InDels)

88: VarScan somatic on data 87 and data 85

Input format

VCF

Output format

VCF (only if input is VCF)

Create CSV report, useful for downstream analysis (-csvStats)

No

Genome source

Locally installed snpEff database

Genome

Homo sapiens : hg19

Regulation options

Upstream / Downstream length

5000 bases

(-ud)

Set size for splice sites (donor and acceptor) in bases

2 bases

(-ss)

spliceRegion Settings

Use Defaults

Produce Summary Stats

No

(-noStats)

Annotation options

Select/Unselect all

- Use 'EFF' field compatible with older versions (instead of 'ANN')
- Use Classic Effect names and amino acid variant annotations (NON_SYNONYMOUS_CODING vs missense_variant and G180R vs p.Gly180Arg/c.538G>C)
- Override classic and use Sequence Ontolgy terms for effects (missense_variant vs NON_SYNONYMOUS_CODING)
- Override classic and use HGVS annotations for amino acid annotations (p.Gly180Arg/c.538G>C vs G180R)
- Old notation style notation: E.g. 'c.G123T' instead of 'c.123G>T' and 'X' instead of '*'
- Use one letter Amino acid codes in HGVS notation. E.g. p.R47G instead of p.Arg47Gly
- Use transcript ID in HGVS notation. E.g. ENST00000252100:c.914C>G instead of c.914C>G

a. Adding annotations to the called variants

a.2. Adding genetic and clinical evidence-based annotations

Creating a GEMINI database from a variants dataset

Galaxy Europe

Analyze Data | Workflow | Visualize | Shared Data | Help | User

Using 8%

Tools

GEMINI load

Upload Data

Hide Sections

RNA Analysis

StringTie merge transcripts

Gene Body Coverage (Bigwig) Read coverage over gene body

StringTie transcript assembly and quantification

Gemini

GEMINI set_somatic Tag somatic mutations in a GEMINI database

GEMINI load Loading a VCF file into GEMINI

GEMINI fusions Identify somatic fusion genes from a GEMINI database

GEMINI amend Amend an already loaded GEMINI database.

GEMINI query Querying the GEMINI database

GEMINI annotate the variants in an existing GEMINI database with additional information

GEMINI database info Retrieve information about tables, columns and annotation data stored in a GEMINI database

GEMINI stats Compute useful variant statistics

GEMINI actionable_mutations Retrieve genes with actionable somatic mutations via COSMIC and DGIdb

GEMINI burden perform sample-wise gene-level burden calculations

GEMINI roh Identifying runs of

GEMINI load Loading a VCF file into GEMINI (Galaxy Version 0.20.1+galaxy2)

Favorite | Versions | Options

VCF dataset to be loaded in the GEMINI database

90: SnpEff eff: on data 88

Only build 37 (aka hg19) of the human genome is supported.

The variants in this input are

annotated with snpEff

GEMINI can parse and use annotations generated with either snpEff (both 'EFF'- and 'ANN'-style annotations are supported) or VEP. You can also load unannotated variants, but most of GEMINI's functionality will not be available or not be very useful without annotations. (-t)

This input comes with genotype calls for its samples

Yes

This is usually the case, but some published datasets, like some 1000G VCFs, are missing genotype information. (--no-genotypes)

Choose a gemini annotation source

GEMINI annotations w/ GERP & CADD (2019-01-12 snapshot)

Sample and family information in PED format

Nothing selected

The pedigree dataset is optional, but several GEMINI tools require the relationship between samples (i.e., the family structure) and/or the sample phenotype to be defined. The PED format is a simple tabular format (see the tool help below for details). If you choose to not provide sample information now, but later find that you need it for your analysis, you can also add it to an existing GEMINI database by using the GEMINI amend tool. (-p)

Load the following optional content into the database

Select/Unselect all

- [x] GERP scores
- [x] CADD scores (non-commercial use only; see licensing note below)
- [x] Gene tables
- [x] Sample genotypes
- [ ] Genotype likelihoods (sample PLs)
- [ ] only variants that passed all filters
- [x] variant INFO field

The preselected defaults should be ok for most use cases (feel free to enable CADD scores for non-commercial use). If you are not interested in certain annotations, you can speed up database creation and decrease the resulting database size slightly by not loading them into the database. Note: GERP and CADD scores are optional parts of the annotation source and can only be loaded if available.

Email notification

No

Send an email notification when the job completes.

Execute

History

search datasets

Tumor Normal pair somatic pipeline TEST

41 shown, 39 deleted, 13 hidden

21.27 GB

- 90: SnpEff eff: on data 88
- 88: VarScan somatic on data 87 and data 85
- 87: Filter on data 79: Filtered BAM
- 86: Filter on data 79: JSON filter rules
- 85: Filter on data 78: Filtered BAM
- 84: Filter on data 78: JSON filter rules
- 79: CalMD on data 77 (tumor)
- 78: CalMD on data 76 (normal)
- 77: BamLeftAlign on data 75 (alignments) (tumor)
- 76: BamLeftAlign on data 74 (alignments) (normal)
- 75: RmDup on data 73 (tumor)
- 74: RmDup on data 71 (normal)
- 73: Filter on data 32: Fil

Making variant call statistics accessible

Analyze Data Workflow Visualize Shared Data Help User Using 12%

GEMINI annotate the variants in an existing GEMINI database with additional information (Galaxy Version 0.20.1+galaxy2)

Favorite Versions Options

GEMINI database

91: GEMINI load on data 90

Only files with version 0.20.1 are accepted.

Dataset to use as the annotation source

88: VarScan somatic on data 87 and data 85

The tool can use the information from a BED or VCF dataset to annotate the database variants. (-f)

Strict variant-identity matching of database and annotation records (VCF format only)

Yes

The default is to consider VCF-formatted annotations only if a variant in the GEMINI database and a record in the annotation source describe the exact same nucleotide change at the same position in the genome. You can disable this option to make use of any annotation that overlaps with the position of a database variant. This setting is ignored for annotation sources in BED format, for which matching is always based on overlapping positions only. (--region-only)

Type of information to add to the database variants

Specific values extracted from matching records in the annotation source (extract)

(-a)

Annotation extraction recipe

1: Annotation extraction recipe

Elements to extract from the annotation source

History

search datasets

Tumor Normal pair somatic pipeline TEST

42 shown, 39 deleted, 13 hidden

30.2 GB

91: GEMINI load on dat a 90
normal tumor

90: SnpEff eff: on data 8 8
normal tumor

88: VarScan somatic on data 87 and data 85
normal tumor

87: Filter on data 79: Fil tered BAM
tumor

86: Filter on data 79: JS ON filter rules
tumor

Making variant call statistics accessible

Analyze Data | Workflow | Visualize | Shared Data | Help | User | Using 12%

1: Annotation extraction recipe

Elements to extract from the annotation source

SS

For an annotation source in BED format, specify the number of the column from which the annotations should be read. For a VCF source, name an INFO field element. (-e)

Database column name to use for recording annotations

somatic_status

A column with the name provided here will be added to the variants table of the GEMINI database to store the annotations (-c)

What type of data are you trying to extract?

- ○ Numbers with decimal precision
- ◉ Integer numbers
- ○ Text (text)

Your selection will determine the data type used to store the new annotations in the database. (-t)

If multiple annotations are found for the same variant, store ...

the first annotation found

Note: If indicated (in parentheses) an option is only applicable to annotations of a specific type. (-o)

2: Annotation extraction recipe

Elements to extract from the annotation source

GPV

For an annotation source in BED format, specify the number of the column from which the annotations should be read. For a VCF source, name an INFO field element. (-e)

Database column name to use for recording annotations

History

search datasets

Tumor Normal pair somatic pipeline TEST

42 shown, 39 deleted, 13 hidden

30.2 GB

91: GEMINI load on dat a 90
normal tumor

90: SnpEff eff: on data 8 8
normal tumor

88: VarScan somatic on data 87 and data 85
normal tumor

87: Filter on data 79: Fil tered BAM
tumor

86: Filter on data 79: JS ON filter rules
tumor

Making variant call statistics accessible

Analyze Data Workflow Visualize Shared Data Help User Using 12%

Note: If indicated (in parentheses) an option is only applicable to annotations of a specific type. (-o)

2: Annotation extraction recipe

Elements to extract from the annotation source

GPV

For an annotation source in BED format, specify the number of the column from which the annotations should be read. For a VCF source, name an INFO field element. (-e)

Database column name to use for recording annotations

germline_p

A column with the name provided here will be added to the variants table of the GEMINI database to store the annotations (-c)

What type of data are you trying to extract?

- (selected) Numbers with decimal precision
- Integer numbers
- Text (text)

Your selection will determine the data type used to store the new annotations in the database. (-t)

If multiple annotations are found for the same variant, store ...

the first annotation found

Note: If indicated (in parentheses) an option is only applicable to annotations of a specific type. (-o)

3: Annotation extraction recipe

Elements to extract from the annotation source

For an annotation source in BED format, specify the number of the column from which the annotations should be read. For a VCF source, name an INFO field element. (-e)

History

search datasets

Tumor Normal pair somatic pipeline TEST

42 shown, 39 deleted, 13 hidden

30.2 GB

91: GEMINI load on dat a 90
normal tumor

90: SnpEff eff: on data 8 8
normal tumor

88: VarScan somatic on data 87 and data 85
normal tumor

87: Filter on data 79: Fil tered BAM
tumor

86: Filter on data 79: JS ON filter rules
tumor

Making variant call statistics accessible

Analyze Data Workflow Visualize Shared Data Help User Using 12%

3: Annotation extraction recipe

Elements to extract from the annotation source

SPV

For an annotation source in BED format, specify the number of the column from which the annotations should be read. For a VCF source, name an INFO field element. (-e)

Database column name to use for recording annotations

somatic_p

A column with the name provided here will be added to the variants table of the GEMINI database to store the annotations (-c)

What type of data are you trying to extract?

- (●) Numbers with decimal precision
- ( ) Integer numbers
- ( ) Text (text)

Your selection will determine the data type used to store the new annotations in the database. (-t)

If multiple annotations are found for the same variant, store ...

the first annotation found

Note: If indicated (in parentheses) an option is only applicable to annotations of a specific type. (-o)

+ Insert Annotation extraction recipe

Email notification

No

Send an email notification when the job completes.

Execute

History

search datasets

Tumor Normal pair somatic pipeline TEST

42 shown, 39 deleted, 13 hidden

30.2 GB

91: GEMINI load on dat a 90
normal tumor

90: SnpEff eff: on data 8 8
normal tumor

88: VarScan somatic on data 87 and data 85
normal tumor

87: Filter on data 79: Fil tered BAM
tumor

86: Filter on data 79: JS ON filter rules
tumor

Adding further annotations from dbSNP

Analyze Data | Workflow | Visualize | Shared Data | Help | User | Using 12%

GEMINI annotate the variants in an existing GEMINI database with additional information (Galaxy Version 0.20.1 - galaxy2)

Favorite | Versions | Options

GEMINI database

92: GEMINI annotate on data 88 and data 91

Only files with version 0.20.1 are accepted.

Dataset to use as the annotation source

56: dbsnp.b147.chr5_12_17.vcf.gz

The tool can use the information from a BED or VCF dataset to annotate the database variants. (-f)

Strict variant-identity matching of database and annotation records (VCF format only)

Yes

The default is to consider VCF-formatted annotations only if a variant in the GEMINI database and a record in the annotation source describe the exact same nucleotide change at the same position in the genome. You can disable this option to make use of any annotation that overlaps with the position of a database variant. This setting is ignored for annotation sources in BED format, for which matching is always based on overlapping positions only. (--region-only)

Type of information to add to the database variants

Specific values extracted from matching records in the annotation source (extract)

(-a)

Annotation extraction recipe

1: Annotation extraction recipe

Elements to extract from the annotation source

History

search datasets

Tumor Normal pair somatic pipeline TEST

56 shown, 39 deleted, hide hidden

30.2 GB

es.13Feb2019.txt

56: dbsnp.b147.chr5_12_17.vcf.gz

55: 01-Feb-2019-ClVic.bed

54: cgi_variant_positions.bed

53: hotspots.bed

32: Map with BWA-MEM on tumor
tumor

31: Map with BWA-MEM on normal
normal

Adding further annotations from dbSNP

Analyze Data Workflow Visualize Shared Data Help User Using 12%

exact same nucleotide change at the same position in the genome. You can disable this option to make use of any annotation that overlaps with the position of a database variant. This setting is ignored for annotation sources in BED format, for which matching is always based on overlapping positions only. (--region-only)

Type of information to add to the database variants

Specific values extracted from matching records in the annotation source (extract)

(-a)

Annotation extraction recipe

1: Annotation extraction recipe

Elements to extract from the annotation source

SAO

For an annotation source in BED format, specify the number of the column from which the annotations should be read. For a VCF source, name an INFO field element. (-e)

Database column name to use for recording annotations

rs_ss

A column with the name provided here will be added to the variants table of the GEMINI database to store the annotations (-c)

What type of data are you trying to extract?

- (●) Numbers with decimal precision
- ( ) Integer numbers
- ( ) Text (text)

Your selection will determine the data type used to store the new annotations in the database. (-t)

If multiple annotations are found for the same variant, store ...

the first annotation found

Note: If indicated (in parentheses) an option is only applicable to annotations of a specific type. (-o)

History

search datasets

Tumor Normal pair somatic pipeline TEST

56 shown, 39 deleted, hide hidden

30.2 GB

es.13Feb2019.txt

56: dbsnp.b147.chr5_12_17.vcf.gz

55: 01-Feb-2019-ClVic.bed

54: cgi_variant_positions.bed

53: hotspots.bed

32: Map with BWA-MEM on tumor
tumor

31: Map with BWA-MEM on normal
normal

Adding further annotations from Cancer Hotspots v2

Analyze Data | Workflow | Visualize | Shared Data | Help | User | Using 12%

Dataset to use as the annotation source

53: cancerhotspots_v2.bed

The tool can use the information from a BED or VCF dataset to annotate the database variants. (-f)

Strict variant-identity matching of database and annotation records (VCF format only)

Yes

The default is to consider VCF-formatted annotations only if a variant in the GEMINI database and a record in the annotation source describe the exact same nucleotide change at the same position in the genome. You can disable this option to make use of any annotation that overlaps with the position of a database variant. This setting is ignored for annotation sources in BED format, for which matching is always based on overlapping positions only. (--region-only)

Type of information to add to the database variants

Specific values extracted from matching records in the annotation source (extract)

(-a)

Annotation extraction recipe

1: Annotation extraction recipe

Elements to extract from the annotation source

5

For an annotation source in BED format, specify the number of the column from which the annotations should be read. For a VCF source, name an INFO field element. (-e)

Database column name to use for recording annotations

hs_qvalue

A column with the name provided here will be added to the variants table of the GEMINI database to store the annotations (-c)

What type of data are you trying to extract?

- (selected) Numbers with decimal precision
- Integer numbers
- Text (text)

Your selection will determine the data type used to store the new annotations in the database. (-t)

If multiple annotations are found for the same variant, store ...

the smallest of the (numeric) values

History

search datasets

Tumor Normal pair somatic pipeline TEST

57 shown, 39 deleted, hide hidden

30.42 GB

93: GEMINI annotate on data 56 and data 92

92: GEMINI annotate on data 88 and data 91

normal | tumor

91: GEMINI load on data 90

normal | tumor

90: SnpEff eff: on data 88

normal | tumor

19,382 lines, 138 comments

format: **vcf**, database: **hg19**

display at UCSC main

display with IGV local

display at RViewer main

```
1.Chrom
##fileformat=VCFv4.2
##FILTER=<ID=PASS,Description="All filters
##reference=/data/db/reference_genomes/hg1
##source=varscan.py
```

Adding links to CIViC

Analyze Data | Workflow | Visualize | Shared Data | Help | User | Using 12%

Dataset to use as the annotation source

55: 01-Feb-2019-CIVic.bed

The tool can use the information from a BED or VCF dataset to annotate the database variants. (-f)

Strict variant-identity matching of database and annotation records (VCF format only)

Yes

The default is to consider VCF-formatted annotations only if a variant in the GEMINI database and a record in the annotation source describe the exact same nucleotide change at the same position in the genome. You can disable this option to make use of any annotation that overlaps with the position of a database variant. This setting is ignored for annotation sources in BED format, for which matching is always based on overlapping positions only. (--region-only)

Type of information to add to the database variants

Specific values extracted from matching records in the annotation source (extract)

(-a)

Annotation extraction recipe

1: Annotation extraction recipe

Elements to extract from the annotation source

4

For an annotation source in BED format, specify the number of the column from which the annotations should be read. For a VCF source, name an INFO field element. (-e)

Database column name to use for recording annotations

overlapping_civic_url

A column with the name provided here will be added to the variants table of the GEMINI database to store the annotations (-c)

What type of data are you trying to extract?

- Numbers with decimal precision
- Integer numbers
- Text (text)

Your selection will determine the data type used to store the new annotations in the database. (-t)

If multiple annotations are found for the same variant, store ...

a comma-separated list of non-redundant (text) values

History

search datasets

Tumor Normal pair somatic pipeline TEST

58 shown, 39 deleted, hide hidden

30.42 GB

94: GEMINI annotate on data 53 and data 93

93: GEMINI annotate on data 56 and data 92

92: GEMINI annotate on data 88 and data 91

normal | tumor

91: GEMINI load on data 90

normal | tumor

90: SnpEff eff: on data 88

normal | tumor

19,382 lines, 138 comments

format: vcf, database: hg19

display at UCSC main

display with IGV local

display at RViewer main

1.Chrom

##fileformat=VCFv4.2

Adding further annotations from Cancer Genome Interpreter (CGI)

Analyze Data | Workflow | Visualize | Shared Data | Help | User | Using 12%

GEMINI annotate the variants in an existing GEMINI database with additional information (Galaxy Version 0.20.1+galaxy2)

Favorite | Versions | Options

GEMINI database

95: GEMINI annotate on data 55 and data 94

Only files with version 0.20.1 are accepted.

Dataset to use as the annotation source

54: cgi_variant_positions.bed

The tool can use the information from a BED or VCF dataset to annotate the database variants. (-f)

Strict variant-identity matching of database and annotation records (VCF format only)

Yes

The default is to consider VCF-formatted annotations only if a variant in the GEMINI database and a record in the annotation source describe the exact same nucleotide change at the same position in the genome. You can disable this option to make use of any annotation that overlaps with the position of a database variant. This setting is ignored for annotation sources in BED format, for which matching is always based on overlapping positions only. (--region-only)

Type of information to add to the database variants

Binary indicator (1=found, 0=not found) of whether the variant had any match in the annotation source (boolean)

(-a)

Database column name to use for recording annotations

in_cgidb

A column with the name provided here will be added to the variants table of the GEMINI database to store the annotations (-c)

Email notification

No

Send an email notification when the job completes.

Execute

What it does

History

search datasets

Tumor Normal pair somatic pipeline TEST

59 shown, 39 deleted, hide hidden

30.42 GB

95: GEMINI annotate on data 55 and data 94

94: GEMINI annotate on data 53 and data 93

93: GEMINI annotate on data 56 and data 92

92: GEMINI annotate on data 88 and data 91

normal | tumor

91: GEMINI load on data 90

normal | tumor

90: SnpEff eff: on data 88

normal | tumor

19,382 lines, 138 comments

format: vcf, database: hg19

display at UCSC main

display with IGV local

display at RViewer main

b. Reporting selected subsets of variants

Querying the GEMINI database for somatic variants

Analyze Data | Workflow | Visualize | Shared Data | Help | User

GEMINI query Querying the GEMINI database (Galaxy Version 0.20.1+galaxy1)

Favorite | Versions | Options

GEMINI database

96: GEMINI annotate CGI infos on data 54 and data 95

Only files with version 0.20.1 are accepted.

Build GEMINI query using

Basic variant query constructor

Genotype filter expression

1: Genotype filter expression

Restrictions to apply to genotype values

gt_alt_freqs.NORMAL <= 0.05 AND gt_alt_freqs.TUMOR >= 0.10

(--gt-filter)

+ Insert Genotype filter expression

Sample filter expression

+ Insert Sample filter expression

Region Filter

+ Insert Region Filter

Filter variant sites by their position in the genome. If multiple Region Filters are specified, all variants that fall in ONE of the regions are reported.

Additional constraints expressed in SQL syntax

somatic_status = 2

Constraints defined here will become the WHERE clause of the SQL query issued to the GEMINI database. E.g. alt='G' or impact_severity = 'HIGH'.

Analyze Data | Workflow | Visualize | Shared Data | Help | User | Using 12%

Output format options

Type of report to generate

tabular (GEMINI default)

Add a header of column names to the output

Yes

(--header)

Set of columns to include in the variant report table

Custom (report user-specified columns)

Choose columns to include in the report

Select/Unselect all

- [ ] gene
- [x] chrom
- [x] start
- [ ] end
- [x] ref
- [x] alt
- [ ] impact
- [ ] impact_severity
- [ ] alternative allele frequency (max_aaf_all)

(--columns)

Additional columns (comma-separated)

gene, aa_change, rs_ids, hs_qvalue, cosmic_ids

Column must be specified by the exact name they have in the GEMINI database, e.g., is_exonic or num_hom_alt, but, for genotype columns, GEMINI wildcard syntax is supported. The order of columns in the list is maintained in the output.

Request drug-gene interaction info from DGIdb

No

(--dgidb)

Sort the output by the following column(s)

History

search datasets

Tumor Normal pair somatic pipeline TEST

61 shown, 39 deleted, hide hidden

31.3 GB

97: GEMINI query on data 96

normal | tumor

This job is waiting to run

96: GEMINI annotate CGI infos on data 54 and data 95

95: GEMINI annotate CIViC data on data 55 and data 94

normal | tumor

94: GEMINI annotate Cancer Hotspots V2 on data 53 and data 93

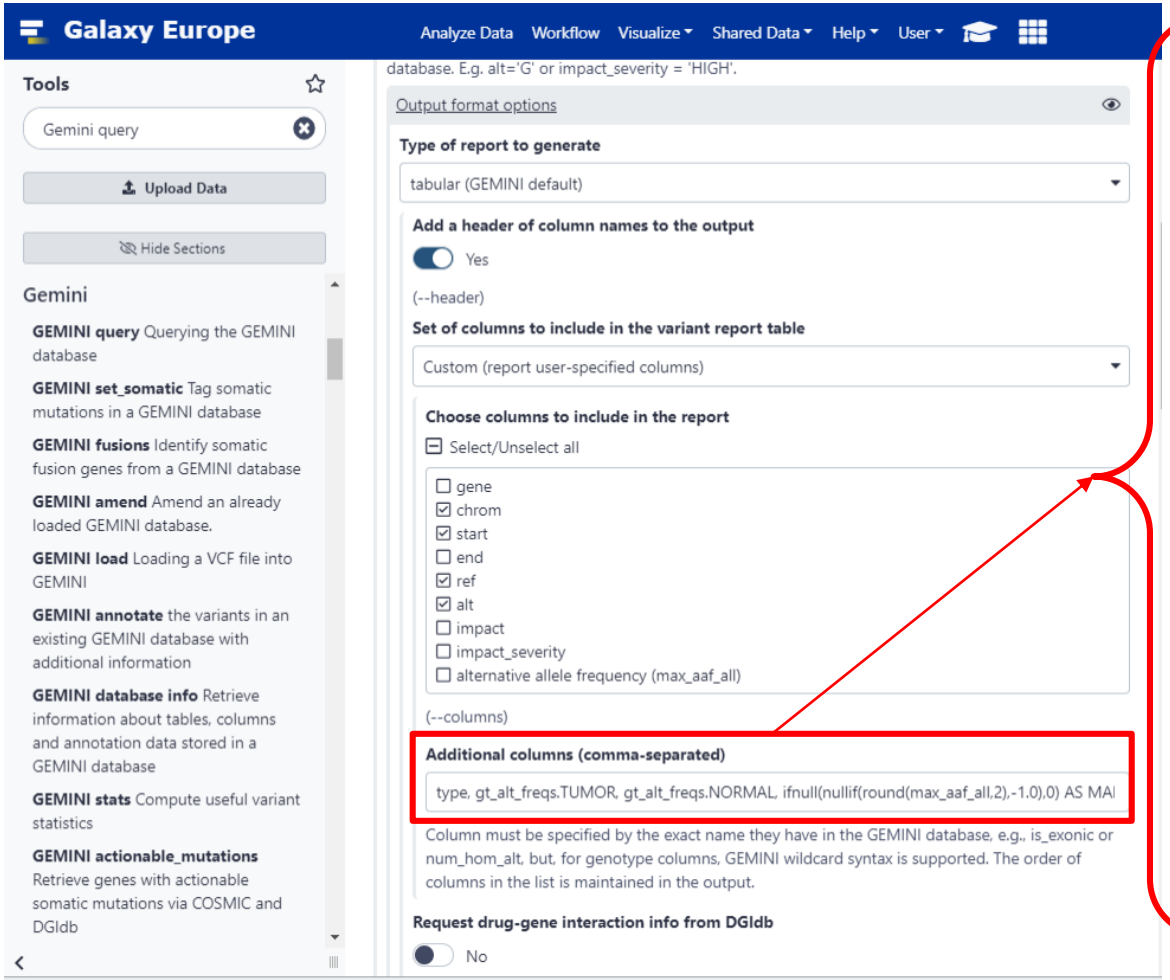
normal | tumor

93: GEMINI annotate dbSNP infos on data 56 and data 92

normal | tumor

92: GEMINI annotate VarScan Somatic infos on data 88 and data 91

GEMINI SQL-based output formatting




```
type,
gt_alt_freqs.TUMOR,
gt_alt_freqs.NORMAL,
ifnull(nullif(round(max_aaf_all,2),-1.0),0)
AS MAF,
gene,
impact_so,
aa_change,
ifnull(round(cadd_scaled,2),'.') AS
cadd_scaled,
round(gerp_bp_score,2) AS gerp_bp,
ifnull(round(gerp_element_pval,2),'.') AS
gerp_element_pval,
ifnull(round(hs_qvalue,2), '.') AS
hs_qvalue,
in_omim,
ifnull(clinvar_sig,'.') AS clinvar_sig,
ifnull(clinvar_disease_name,'.') AS
clinvar_disease_name,
ifnull(rs_ids,'.') AS dbsnp_ids,
rs_ss,
ifnull(cosmic_ids,'.') AS cosmic_ids,
ifnull(overlapping_civic_url,'.') AS
overlapping_civic_url,
in_cgidb
```

c. Generating reports of genes affected by variants

Turning query results into gene-centered reports

Analyze Data | Workflow | Visualize | Shared Data | Help | User | Using 12%

GEMINI query Querying the GEMINI database (Galaxy Version 0.20.1+galaxy1)

Favorite | Versions | Options

GEMINI database

96: GEMINI annotate CGI infos on data 54 and data 95

Only files with version 0.20.1 are accepted.

Build GEMINI query using

Advanced query constructor

The query to be issued to the database

SELECT v.gene, v.chrom, g.synonym, g.hgnc_id, g.entrez_id, g.rvis_pct, v.clinvar_gene_phenotype FROM variants v, gene_detailed g WHERE v.chrom = g.chrom AND v.gene = g.gene AND v.somatic_status = 2 AND v.somatic_p <= 0.05 AND v.filter IS NULL GROUP BY g.gene

Formulate your query using SQL syntax. (-q)

Genotype filter expression

1: Genotype filter expression

Restrictions to apply to genotype values

gt_alt_freqs.NORMAL <= 0.05 AND gt_alt_freqs.TUMOR >= 0.10

(--gt-filter)

+ Insert Genotype filter expression

Sample filter expression

+ Insert Sample filter expression

Output format options

History

search datasets

Tumor Normal pair somatic pipeline TEST

64 shown, 39 deleted, hide hidden

31.3 GB

100: GEMINI query on data 96

99: GEMINI query on data 96

98: GEMINI database info on data 96

97: GEMINI query on data 96

96: GEMINI annotate CGI infos on data 54 and data 95

95: GEMINI annotate CIViC data on data 55 and data 94 normal tumor

94: GEMINI annotate Cancer Hotspots V2 on data 53 and data 93 normal tumor

93: GEMINI annotate dbSNP infos on data 56 and data 92 normal tumor

```
SELECT v.gene, v.chrom,
g.synonym, g.hgnc_id,
g.entrez_id, g.rvis_pct,
v.clinvar_gene_phenotype

FROM variants v,
gene_detailed g

WHERE v.chrom = g.chrom AND
v.gene = g.gene AND
v.somatic_status = 2 AND
v.somatic_p <= 0.05 AND
v.filter IS NULL

GROUP BY g.gene
```

d. Adding additional annotations to the gene-centered report

Adding UniProt cancer genes information

Galaxy Europe

Analyze Data | Workflow | Visualize | Shared Data | Help | User

Using 12%

Tools

join two files

Upload Data

Show Sections

VCF-VCFintersect: Intersect two VCF datasets

Sub-sample sequences files e.g. to reduce coverage

idpQuery Creates text reports from idpDB files.

Join two files

fastq-join - Joins two paired-end reads on the overlapping ends

Join two files on column allowing a small difference

Join the intervals of two datasets side-by-side

Column Join on Collections

Column Join on Collections

QCMerger Merges two qcml files together.

FuzzyDiff Compares two files, tolerating numeric differences.

seqtk_mergefa merge two FASTA/Q files

Join +/- Ions Join positive and negative ionization-mode W4M

Join two files (Galaxy Version 1.1.2)

Favorite | Versions | Options

1st file

107: GEMINI query on data 96

Column to use from 1st file

Column: 1

2nd File

57: Uniprot_Cancer_Genes.13Feb2019.txt

Column to use from 2nd file

Column: 1

Output lines appearing in

Both 1st & 2nd file, plus unpairable lines from 1st file. (-a 1)

First line is a header line

Yes

Use if first line contains column headers. It will not be sorted.

Ignore case

No

Sort and Join key column values regardless of upper/lower case letters.

Value to put in unpaired (empty) fields

0

Email notification

No

Send an email notification when the job completes.

History

search datasets

Tumor Normal pair somatic pipeline TEST

37 shown, 45 deleted, 29 hidden

31.3 GB

108: Join on data 57 and data 107

This job is waiting to run

107: GEMINI query on data 96

43 lines

format: **tabular**, database: **hg19**

1	2	3	4
gene	chrom	synonym	hgnc_i
A2ML1	chr12	CPAMD9,FLJ25179	23336
ANKDD1B	chr5	None	32525
APC	chr5	DP3,PPP1R46,DP2,DP2.5	583
ARHGAP9	chr12	10C,MGC1295	14130

100: GEMINI query on data 96

normal | tumor

99: GEMINI query on da

Adding CGI biomarkers information

Galaxy Europe

Analyze Data | Workflow | Visualize | Shared Data | Help | User

Using 12%

Tools

join two files

Upload Data

Show Sections

VCF-VCFintersect: Intersect two VCF datasets

Sub-sample sequences files e.g. to reduce coverage

idpQuery Creates text reports from idpDB files.

Join two files

fastq-join - Joins two paired-end reads on the overlapping ends

Join two files on column allowing a small difference

Join the intervals of two datasets side-by-side

Column Join on Collections

Column Join on Collections

QCMerger Merges two qcml files together.

FuzzyDiff Compares two files, tolerating numeric differences.

seqtk_mergefa merge two FASTA/Q files

Join +/- Ions Join positive and negative ionization-mode W4M datasets for the same samples

Multi-Join (combine multiple files)

Join two files (Galaxy Version 1.1.2)

Favorite | Versions | Options

1st file

102: Join on data 57 and data 100

Column to use from 1st file

Column: 1

2nd File

58: cgi_genes.txt

Column to use from 2nd file

Column: 1

Output lines appearing in

Both 1st & 2nd file, plus unpairable lines from 1st file, (-a 1)

First line is a header line

Yes

Use if first line contains column headers. It will not be sorted.

Ignore case

No

Sort and Join key column values regardless of upper/lower case letters.

Value to put in unpaired (empty) fields

0

Email notification

No

Send an email notification when the job completes.

Execute

History

search datasets

Tumor Normal pair somatic pipeline TEST

66 shown, 39 deleted, hide hidden

31.3 GB

102: Join on data 57 and data 100

normal | tumor

70 lines

format: **tabular**, database: **hg19**

1	2	3	4	5	6	7
gene	chrom	start	ref	alt	type	MAF
A2ML1	chr12	8989045	A	C	snp	0
ANKDD1B	chr5	74930649	G	A	snp	0
APC	chr5	112175422	C	T	snp	0
ARHGAP9	chr12	57872993	G	A	snp	0.0

101: Join on data 57 and data 100

normal | tumor

70 lines

format: **tabular**, database: **hg19**

1	2	3	4	5	6	7
chrom	start	ref	alt	type	MAF	gene
chr12	113335665	G	A	snp	0.0	RPH3A
chr12	117582871	C	CT	indel	0.29	FBXO21
chr12	123286211	G	A	snp	0.0	CCDC62
chr12	126068542	A	G	snp	0	TMEM132

Adding gene information from CIViC

Galaxy Europe

Analyze Data | Workflow | Visualize | Shared Data | Help | User | Using 12%

Tools

join two files

Upload Data

Show Sections

VCF-VCFintersect: Intersect two VCF datasets

Sub-sample sequences files e.g. to reduce coverage

idpQuery Creates text reports from idpDB files.

Join two files

fastq-join - Joins two paired-end reads on the overlapping ends

Join two files on column allowing a small difference

Join the intervals of two datasets side-by-side

Column Join on Collections

Column Join on Collections

QCMerger Merges two qcml files together.

FuzzyDiff Compares two files, tolerating numeric differences.

seqtk_mergefa merge two FASTA/Q files

Join +/- Ions Join positive and negative ionization-mode W4M

Join two files (Galaxy Version 1.1.2)

Favorite | Versions | Options

1st file

109: Join on data 58 and data 108

Column to use from 1st file

Column: 1

2nd File

59: 01-Feb-2019-GeneSummaries.tsv

Column to use from 2nd file

Column: 3

Output lines appearing in

Both 1st & 2nd file, plus unpairable lines from 1st file. (-a 1)

First line is a header line

Yes

Use if first line contains column headers. It will not be sorted.

Ignore case

No

Sort and Join key column values regardless of upper/lower case letters.

Value to put in unpaired (empty) fields

.

Email notification

No

Send an email notification when the job completes.

History

search datasets

Tumor Normal pair somatic pipeline TEST

39 shown, 45 deleted, 29 hidden

31.3 GB

110: Join on data 59 and data 109

This job is waiting to run

109: Join on data 58 and data 108

108: Join on data 57 and data 107

107: GEMINI query on data 96

100: GEMINI query on data 96

normal | tumor

99: GEMINI query on data 96

normal | tumor

98: GEMINI database info on data 96

normal | tumor

Rearrange to get a fully annotated gene report

Galaxy Europe

Analyze Data Workflow Visualize Shared Data Help User

Tools

Column arrange

Upload Data

Show Sections

format into ranges of bases

Column arrange by header name

FASTQ Summary Statistics by column

BLAST parser Convert 12- or 24-column BLAST output into 3-column hcluster_sg input

Sort Column Order by heading

Split file according to the values of a column

Add input name as column to an existing tabular file

Replace Text in a specific column

Replace column by values which are defined in a convert file

FASTQ Trimmer by column

Column Join on Collections

Column Join on Collections

Histogram of a numeric column

Map peptides to a bed file for viewing in a genome browser

Manipulate loom object Add layers, or row/column attributes to a loom file

Summary Statistics for any numerical

Column arrange by header name (Galaxy Version 0.2)

Favorite Versions Options

file to rearrange

110: Join on data 59 and data 109

Specify the first few columns by name

1: Specify the first few columns by name

column

gene

2: Specify the first few columns by name

column

chrom

3: Specify the first few columns by name

column

synonym

4: Specify the first few columns by name

column

hgnc_id

5: Specify the first few columns by name

column

entrez_id

6: Specify the first few columns by name

column

Analyze Data Workflow Visualize Shared Data Help User

6: Specify the first few columns by name

column

rvis_pct

7: Specify the first few columns by name

column

is_OG

8: Specify the first few columns by name

column

is_TS

9: Specify the first few columns by name

column

in_cgi_biomarkers

10: Specify the first few columns by name

column

clinvar_gene_phenotype

11: Specify the first few columns by name

column

gene_civic_url

12: Specify the first few columns by name

column

description

+ Insert Specify the first few columns by name

Inspecting fully annotated gene report

gene	chrom	synonym	hgnc_id	entrez_id	rvis_pct	is_OG	is_TS	in_cgi_biom	clinvar_gen	gene_civic_	description	gene_id	entrez_id	last_review_date
A2ML1	chr12	CPAMD9,FLJ2517	23336	144568	98.52559566	0	0	0	nonsyndrom	.	.	.	.	.
ANKDD1B	chr5	None	32525	728780	None	0	0	0	None	.	.	.	.	.
APC	chr5	DP3,PPP1R46,DP2	583	324	0.902335456	0	1	1	apc-associa	https://civic	.	66	324	2017-02-09 21:58:08 UTC
ARHGAP9	chr12	10C,MGC1295	14130	64333	30.90351498	0	0	0	coronary_ar	.	.	.	.	.
C2CD5	chr12	CDP138,KIAA0528	29062	9847	None	0	0	0	None	.	.	.	.	.
CCDC62	chr12	ERAP75,CT109,FL	30723	84660	82.29535268	0	0	0	None	.	.	.	.	.
CDH12	chr5	Br-cadherin,CDH	1751	1010	29.48808681	0	0	0	None	.	.	.	.	.
CDH18	chr5	EY-CADHERIN,CD	1757	1016	5.602736494	0	0	0	None	.	.	.	.	.
CLEC4C	chr12	DLEC,CLECSF7,CD	13258	170482	86.47676339	0	0	0	None	.	.	.	.	.
CLEC6A	chr12	dectin-2,CLECSF1	14556	93978	89.95635763	0	0	0	None	.	.	.	.	.
COX7C	chr5	None	2292	1350	62.38499646	0	0	0	None	.	.	.	.	.
DDX51	chr12	None	20082	317781	64.96225525	0	0	0	None	.	.	.	.	.
ELAC2	chr17	FLJ10530,HPC2	14198	60528	10.12031139	1	0	0	combined_c	.	.	.	.	.
ERN1	chr17	IRE1P,IRE1	3449	2081	24.63434772	0	0	0	None	.	.	.	.	.
ESM1	chr5	None	3466	11082	56.64071715	0	0	0	None	.	.	.	.	.
FBXO21	chr12	FBX21,KIAA0875	13592	23014	12.77423921	0	0	0	None	.	.	.	.	.
HAPLN1	chr5	CRTL1	2380	1404	20.53550366	0	0	0	None	.	.	.	.	.
ITGB7	chr12	None	6162	3695	4.62962963	0	0	0	None	.	.	.	.	.
KRAS	chr12	KRAS1,KRAS2	6407	3845	42.87567823	1	0	1	acute_myel	https://civic	Mutations i	30	3845	2017-02-09 21:59:28 UTC
KRBA2	chr17	None	26989	124751	57.31304553	0	0	0	None	.	.	.	.	.
LINC01019	chr5	None	27742	285577	None	0	0	0	None	.	.	.	.	.
LYRM7	chr5	FLJ20796,C5orf31,	28072	90624	56.2514744	0	0	0	mitochondri	.	.	.	.	.
METTL2A	chr17	METTL2,FLJ12760	25755	339175	67.03231894	0	0	0	None	.	.	.	.	.
MROH2B	chr5	FLJ40243,DKFZp7	26857	133558	None	0	0	0	None	.	.	.	.	.
PCDHB9	chr5	PCDH-BETA9,PCD	8694	None	None	0	0	0	None	.	.	.	.	.
PCDHGB1	chr5	PCDH-GAMMA-B1	8708	56104	10.92238736	0	0	0	None	.	.	.	.	.
PCDHGB7	chr5	PCDH-GAMMA-B7	8714	56099	48.90894079	0	0	0	None	.	.	.	.	.
PDE3A	chr12	CGI-PDE	8778	5139	6.894314697	0	0	0	brachydacty	.	.	.	.	.
RACK1	chr5	GNB2L1,H12.3,Gn	None	10399	46.20193442	0	0	0	None	.	.	.	.	.
RNF213	chr17	NET57,KIAA1618,	14539	57674	97.65274829	1	0	0	moyamoya_	.	.	.	.	.
SLC16A5	chr17	MCT5,MCT6	10926	9121	36.22906346	0	0	0	None	.	.	.	.	.
SMARCC2	chr12	Rsc8,CRACC2,BAF	11105	6601	2.707006369	0	0	0	malignant_t	.	.	.	.	.
SOX5	chr12	L-SOX5,MGC3515	11201	6660	9.088228356	0	0	0	aplasia/hyp	.	.	.	.	.
SPEF2	chr5	KPL2,FLJ23577,CT	26293	79925	99.06817646	0	0	0	None	.	.	.	.	.
SYNPO	chr5	KIAA1029	30672	11346	40.67586695	0	0	0	None	.	.	.	.	.
TENM2	chr5	Ten-M2,KIAA1127	29943	57451	None	0	0	0	None	.	.	.	.	.
TMEM132B	chr12	KIAA1906,KIAA17	29397	114795	2.677518283	0	0	0	None	.	.	.	.	.
TP53	chr17	LFS1,p53	11998	7157	35.98726115	0	1	1	acute_mega	https://civic	TP53 mutati	45	7157	2018-03-30 15:05:39 UTC
TTC23L	chr5	FLJ25439	26355	153657	96.55579146	0	0	0	None	.	.	.	.	.
TTC37	chr5	KIAA0372	23639	9652	58.00896438	0	0	0	malignant_t	.	.	.	.	.
USP22	chr17	KIAA1063,USP3L	12621	23326	27.41802312	0	0	0	None	.	.	.	.	.
VCAN	chr5	CSPG2,PG-M	2464	1462	19.95753715	0	0	0	malignant_t	.	.	.	.	.