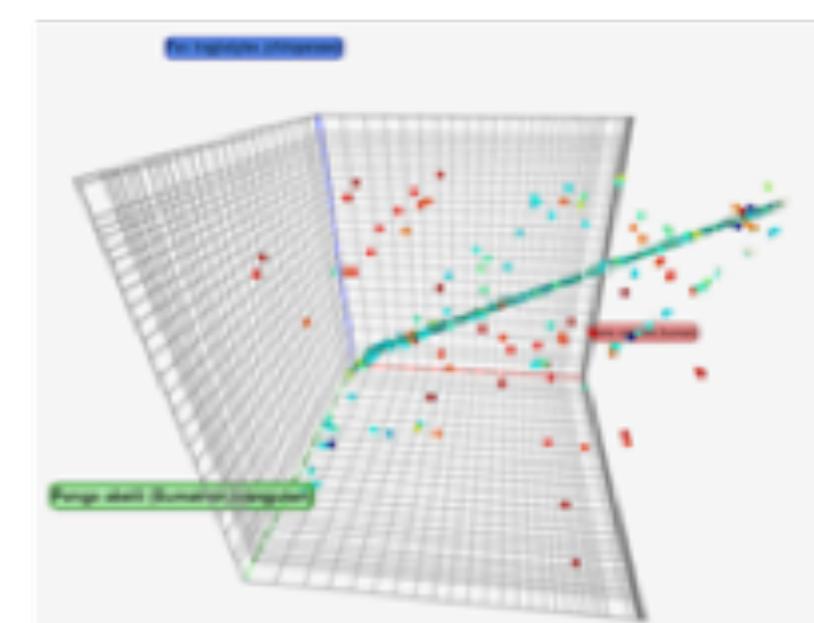
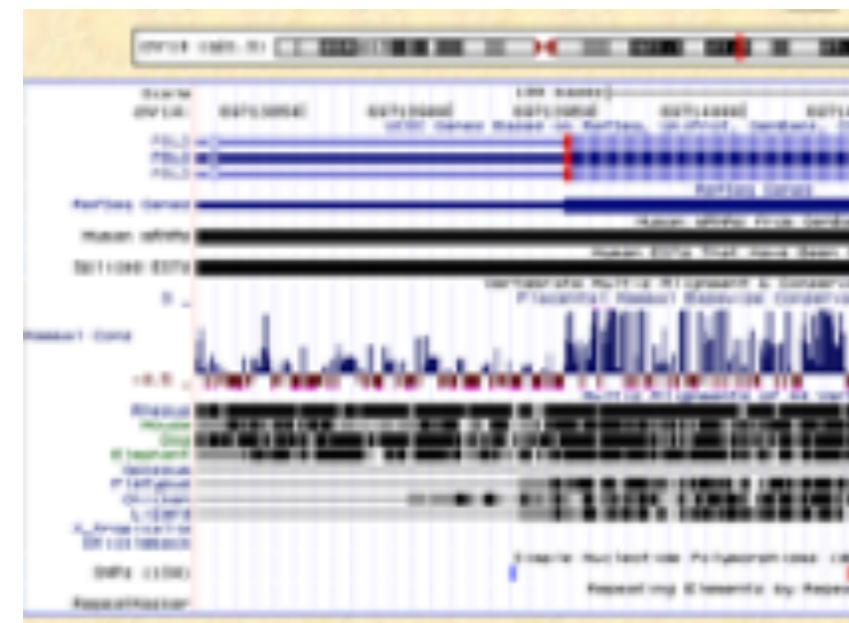


Computational Genomics

Computational Arithmetics



Computational Arithmetics

For Human Chromosome 22

- **How many transcripts are annotated?**
- **What is the size of the largest transcript?**
- **What is the size of the smallest transcript?**
- **What is the median size of all transcript?**

Computational Arithmetics

Tools

search tools x

Upload Data

HPRC

Get Data

Send Data

Collection Operations

Lift-Over

Text Manipulation

Convert Formats

Filter and Sort

Join, Subtract and Group

Fetch Alignments/Sequences

Operate on Genomic Intervals

Statistics

Graph/Display Data

Phenotype Association

BED

Annotation

Multiple Alignments

NCBI BLAST+

Mapping

SAM/BAM

Assembly

FASTQ Quality Control

FASTA/FASTQ

RNA-seq

CD-HIT

Best Practices for Kaiser Galaxy

- Kaiser Galaxy (docs, slides) is configured for teaching purposes so all users have a file quota of 1TB. How to [permanently delete nonessential files](#).
- Only certain tools that support multi-core processing have the Job Resources Parameters option which allow you to select cores, memory and time.
- The default job resource parameters for all tools is 1 core with 7GB memory for 24 hours (24 SUs).
 - Configuring a job to use 48 cores for 1 hour requires 48 SUs. (48 cores for 168 hours = 7872 SUs).
 - Configuring a job to use 360GB memory for 1 hour requires 48 SUs. (360GB memory for 168 hours = 7872 SUs).
 - If a tool you used failed because it needs the Job Resource Parameters option added, [contact the HPRC helpdesk](#).



Take an interactive tour: [Galaxy UI](#) [History](#) [Window Manager](#) [Deferred Datasets](#)

Galaxy is an open platform for supporting data intensive research. Galaxy is developed by The Galaxy Team with the support of many contributors.

The Galaxy Project is supported in part by NHGRI, NSF, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Johns Hopkins University.

History

search datasets x

L11-A

6 shown

12.56 MB checkbox eye edit trash

6: Hsapiens_SNPs_Chromosome 22 eye edit trash

5: Hsapiens_Repeats_Chromosome 22 eye edit trash

4: Hsapiens_CExons_Chromosome 22 eye edit trash

3: Hsapiens_Introns_Chromosome 22 eye edit trash

2: Hsapiens_Exons_Chromosome 22 eye edit trash

1: Hsapiens_Genes_Chromosome 22 eye edit trash

Computational Arithmetics

Tools

compute

Compute on rows

bedtools Compute both the depth and breadth of coverage of features in file B on the features in file A (bedtools coverage)

Compute quality statistics

Compute sequence length

bedtools OverlapBed computes the amount of overlap from two intervals

CollectWgsMetrics compute metrics for evaluating of whole genome sequencing experiments

CollectHsMetrics compute metrics about datasets generated through hybrid-selection (e.g. exome)

Upload File from your computer

bedtools Genome Coverage compute the coverage over an entire genome

Aggregate datapoints Appends the average, min, max of datapoints per interval

Mummer Align two or more sequences

Summary Statistics for any numerical column

FLASH adjust length of short reads

Sort data in ascending or descending order

Pear Paired-End read merger

Convert SAM to interval

FilterSamReads include or exclude aligned and unaligned reads and read lists

Compute on rows (Galaxy Version 2.0)

Input file: 1: Hsapiens_Genes_Ch22

Dataset missing? See TIP below

Input has a header line with column names?: No

Select Yes to be able to specify names for new columns and have them added to the header line. If you select No, the first line will be treated as a regular line: If it is empty or starts with a # character it will be skipped, otherwise the tool will attempt to compute the specified expression on it.

Expressions

1: Expressions

Add expression: c3-c2

Mode of the operation: Append

+ Insert Expressions

Avoid scientific notation in any newly computed columns: No

If yes, use fully expanded decimal representation when writing new columns with floating point values. To prevent scientific notation in just specific new columns, you can use numpy's format_float_positional function in the corresponding expression.

Error handling

Job Resource Parameters

Specify job resource parameters

Memory (GB): 7

Maximum Job Memory

Time (hours): 24

Maximum job time

History

search datasets

L11-A

6 shown

12.56 MB

6: Hsapiens_SNPs_Ch22	<input type="button" value=""/> <input type="button" value=""/> <input type="button" value="X"/>
5: Hsapiens_Repeats_Ch22	<input type="button" value=""/> <input type="button" value=""/> <input type="button" value="X"/>
4: Hsapiens_CExons_Ch22	<input type="button" value=""/> <input type="button" value=""/> <input type="button" value="X"/>
3: Hsapiens_Introns_Ch22	<input type="button" value=""/> <input type="button" value=""/> <input type="button" value="X"/>
2: Hsapiens_Exons_Ch22	<input type="button" value=""/> <input type="button" value=""/> <input type="button" value="X"/>
1: Hsapiens_Genes_Ch22	<input type="button" value=""/> <input type="button" value=""/> <input type="button" value="X"/>

2,197 regions
format: bed, database: hg38

?

display in IGB View
display with IGV local Human hg38
display at UCSC main test

1.Chrom	2.Start	3.End	4	5	6	7
chr22	24952715	25197448	ENST00000358431.8	0	+	2502
chr22	24952729	25194457	ENST00000406486.8	0	+	2502
chr22	25069968	25197448	ENST00000401395.1	0	+	2507
chr22	25111841	25197399	ENST00000637069.1	0	+	2511
chr22	33162225	33761476	ENST00000608642.6	0	-	3316

Computational Arithmetics

Tools ★ ⌂

search tools ×

Upload Data

HPRC

Get Data

Send Data

Collection Operations

Lift-Over

Text Manipulation

Convert Formats

Filter and Sort

Join, Subtract and Group

Fetch Alignments/Sequences

Operate on Genomic Intervals

Statistics

Graph/Display Data

Phenotype Association

BED

Annotation

Multiple Alignments

NCBI BLAST+

Mapping

SAM/BAM

Assembly

FASTQ Quality Control

FASTA/FASTQ

RNA-seq

CD-HIT

	244733
	241728
	127480
	85558
	599251
	599251
	599251
	647968
	299177
	384140
	488482
	488482
	588535
	607596
	647338
	647345
	649520
	649753
	649765
	648377
	648177
	7952
	7957
	8416
	8388
	7558
	236913
	20865
	57037
	62662
	66282
	66297
	66299
	39625
	47547
	16253
	16196
	14089

History ⟳ + ⌂ ⚙

search datasets ? ×

L11-A

7 shown

12.87 MB ✓ 🗑️ 💬

7: Compute on data 1 ✖

2,197 regions
format: **bed**, database: **hg38**

Computing 1 new columns with instructions
['c3-c2;;']
Computed new column values for 100.00% of 2197 lines written.

✖ display in IGB View display with IGV local Human hg38 display at UCSC main test

1.Chrom	2.Start	3.End	4	5	6	7
chr22	24952715	25197448	ENST00000358431.8	0	+	250
chr22	24952729	25194457	ENST00000406486.8	0	+	250
chr22	25069968	25197448	ENST00000401395.1	0	+	250
chr22	25111841	25197399	ENST00000637069.1	0	+	251
chr22	33162225	33761476	ENST00000608642.6	0	-	331

6: Hsapiens_SNPs_Ch22 ✖

5: Hsapiens_Repeats_Ch22 ✖

4: Hsapiens_CExons_Ch22 ✖

3: Hsapiens_Introns_Ch22 ✖

2: Hsapiens_Exons_Ch22 ✖

1: Hsapiens_Genes_Ch22 ✖

Computational Arithmetics

Tools ★ ⌂

search tools ×

Upload Data

HPRC

Get Data

Send Data

Collection Operations

Lift-Over

Text Manipulation

Convert Formats

Filter and Sort

Join, Subtract and Group Join_Subtract_and_Group

Fetch Alignments/Sequences

Operate on Genomic Intervals

Statistics

Graph/Display Data

Phenotype Association

BED

Annotation

Multiple Alignments

NCBI BLAST+

Mapping

SAM/BAM

Assembly

FASTQ Quality Control

FASTA/FASTQ

RNA-seq

CD-HIT

Best Practices for Kaiser Galaxy

- Kaiser Galaxy (docs, slides) is configured for teaching purposes so all users have a file quota of 1TB. How to permanently delete nonessential files.
- Only certain tools that support multi-core processing have the Job Resources Parameters option which allow you to select cores, memory and time.
 - The default job resource parameters for all tools is 1 core with 7GB memory for 24 hours (24 SUs). Configuring a job to use 48 cores for 1 hour requires 48 SUs. (48 cores for 168 hours = 7872 SUs).
 - Configuring a job to use 360GB memory for 1 hour requires 48 SUs. (360GB memory for 168 hours = 7872 SUs).
 - If a tool you used failed because it needs the Job Resource Parameters option added, contact the HPRC helpdesk.



Take an interactive tour: [Galaxy UI](#) [History](#) [Window Manager](#) [Deferred Datasets](#)

Galaxy is an open platform for supporting data intensive research. Galaxy is developed by The Galaxy Team with the support of many contributors.

The Galaxy Project is supported in part by NHGRI, NSF, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Johns Hopkins University.

History ⟳ + ⌂ ⚙

search datasets ? ×

L11-A

7 shown

12.87 MB 🔍 💾

7: Compute on data 1	👁️ 🖊️ ✎
6: Hsapiens_SNPs_Chromosome 22	👁️ 🖊️ ✎
5: Hsapiens_Repeats_Chromosome 22	👁️ 🖊️ ✎
4: Hsapiens_CExons_Chromosome 22	👁️ 🖊️ ✎
3: Hsapiens_Introns_Chromosome 22	👁️ 🖊️ ✎
2: Hsapiens_Exons_Chromosome 22	👁️ 🖊️ ✎
1: Hsapiens_Genes_Chromosome 22	👁️ 🖊️ ✎

Computational Arithmetics

Join, Subtract and Group

About GROUPING

For the following input:

chr22	1000	1003	TTT
chr22	2000	2003	aaa
chr10	2200	2203	TTT
chr10	1200	1203	ttt
chr22	1600	1603	AAA

- Grouping on column 4 while ignoring case, and performing operation Count on column 1 will return:

AAA	2
TTT	3

- Grouping on column 4 while not ignoring case, and performing operation Count on column 1 will return:

aaa	1
AAA	1
ttt	1
TTT	2

Computational Arithmetics

Tools

search tools x

Upload Data +

HPRC

Get Data

Send Data

Collection Operations

Lift-Over

Text Manipulation

Convert Formats

Filter and Sort

Join, Subtract and Group

Join two Datasets side by side on a specified field

Compare two Datasets to find common or distinct rows

Group data by a column and perform aggregate operation on other columns.

Fetch Alignments/Sequences

Operate on Genomic Intervals

Statistics

Graph/Display Data

Phenotype Association

BED

Annotation

Multiple Alignments

NCBI BLAST+

Mapping

SAM/BAM

Best Practices for Kaiser Galaxy

- Kaiser Galaxy (docs, slides) is configured for teaching purposes so all users have a file quota of 1TB. How to permanently delete nonessential files.
- Only certain tools that support multi-core processing have the Job Resources Parameters option which allow you to select cores, memory and time.
 - The default job resource parameters for all tools is 1 core with 7GB memory for 24 hours (24 SUs). Configuring a job to use 48 cores for 1 hour requires 48 SUs. (48 cores for 168 hours = 7872 SUs).
 - Configuring a job to use 360GB memory for 1 hour requires 48 SUs. (360GB memory for 168 hours = 7872 SUs).
 - If a tool you used failed because it needs the Job Resource Parameters option added, contact the HPRC helpdesk.



Group

Take an interactive tour: [Galaxy UI](#) [History](#) [Window Manager](#) [Deferred Datasets](#)

Galaxy is an open platform for supporting data intensive research. Galaxy is developed by The Galaxy Team with the support of many contributors.

The Galaxy Project is supported in part by NHGRI, NSF, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Johns Hopkins University.

History

search datasets ? x

L11-A

7 shown

12.87 MB ✓ dog comment

7: Compute on data 1	eye pencil x
6: Hsapiens_SNPs_Chromosome 22	eye pencil x
5: Hsapiens_Repeats_Chromosome 22	eye pencil x
4: Hsapiens_CExons_Chromosome 22	eye pencil x
3: Hsapiens_Introns_Chromosome 22	eye pencil x
2: Hsapiens_Exons_Chromosome 22	eye pencil x
1: Hsapiens_Genes_Chromosome 22	eye pencil x

Computational Arithmetics

Tools

search tools x

Upload Data

HPRC

Get Data

Send Data

Collection Operations

Lift-Over

Text Manipulation

Convert Formats

Filter and Sort

Join, Subtract and Group

Join two Datasets side by side on a specified field

Compare two Datasets to find common or distinct rows

Group data by a column and perform aggregate operation on other columns.

Fetch Alignments/Sequences

Operate on Genomic Intervals

Statistics

Graph/Display Data

Phenotype Association

BED

Annotation

Multiple Alignments

NCBI BLAST+

Mapping

SAM/BAM

Group data by a column and perform aggregate operation on other columns. (Galaxy Version 2.1.4)

Select data 7: Compute on data 1

Dataset missing? See TIP below.

Group by column

Column: 1

Ignore case while grouping? No

Ignore lines beginning with these characters

Select/Unselect all

>
 @
 +
 <
 *
 -
 =
 |
 ?
 \$
 .
 :
 &
 %
 ^
 #

lines beginning with these are not grouped

Operation

Job Resource Parameters

Specify job resource parameters

Memory (GB)

Maximum Job Memory

Time (hours)

History

search datasets ? x

L11-A

7 shown

12.87 MB

7: Compute on data 1

6: Hsapiens_SNPs_Chromosome 22

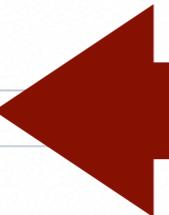
5: Hsapiens_Repeats_Chromosome 22

4: Hsapiens_CExons_Chromosome 22

3: Hsapiens_Introns_Chromosome 22

2: Hsapiens_Exons_Chromosome 22

1: Hsapiens_Genes_Chromosome 22



Group by Column 01 (C1)

Computational Arithmetics

Tools ★ ⌂

search tools ×

Upload Data

HPRC

Get Data

Send Data

Collection Operations

Lift-Over

Text Manipulation

Convert Formats

Filter and Sort

Join, Subtract and Group

Join two Datasets side by side on a specified field

Compare two Datasets to find common or distinct rows

Group data by a column and perform aggregate operation on other columns.

Fetch Alignments/Sequences

Operate on Genomic Intervals

Statistics

Graph/Display Data

Phenotype Association

BED

Annotation

Multiple Alignments

NCBI BLAST+

Mapping

SAM/BAM

Operations

%
^
#

lines beginning with these are not grouped

Operation

1: Operation

Type: Count

On column: Column: 1

Round result to nearest integer?: NO

Replace non numeric data: leave empty for no replacements. Will replace, e.g., empty cells and text cells.

+ Insert Operation

Job Resource Parameters

Specify job resource parameters

Memory (GB): 7

Maximum Job Memory

Time (hours): 24

Maximum job time

Execute

Count by C01 (to determine the number of transcripts)

TIP: If your data is not TAB delimited, use *Text Manipulation->Convert*

Syntax

This tool allows you to group the input dataset by a particular column and perform aggregate functions: Mean, Median, Mode, Sum, Max, Min, Count, Concatenate, and Randomly pick on any column(s).

The Concatenate function will take, for each group, each item in the specified column and build a comma delimited list. Concatenate Unique will do

History

search datasets ? X

L11-A

7 shown

12.87 MB

6: Hsapiens_SNPs_Ch22

5: Hsapiens_Repeats_Ch22

4: Hsapiens_CExons_Ch22

3: Hsapiens_Introns_Ch22

2: Hsapiens_Exons_Ch22

1: Hsapiens_Genes_Ch22

Computational Arithmetics

Tools

search tools x

Upload Data +

HPRC

Get Data

Send Data

Collection Operations

Lift-Over

Text Manipulation

Convert Formats

Filter and Sort

Join, Subtract and Group

Fetch Alignments/Sequences

Operate on Genomic Intervals

Statistics

Graph/Display Data

Phenotype Association

BED

Annotation

Multiple Alignments

NCBI BLAST+

Mapping

SAM/BAM

Assembly

FASTQ Quality Control

FASTA/FASTQ

RNA-seq

CD-HIT

2: Operation

Type: Maximum

On column: Column: 13

Round result to nearest integer? NO

Replace non numeric data: leave empty for no replacements. Will replace, e.g., empty cells and text cells.

+ Insert Operation

Job Resource Parameters

Specify job resource parameters

Memory (GB): 7

Maximum Job Memory

Time (hours): 24

Maximum job time

Execute

Maximum on C13 (contains lengths of transcripts)

History

search datasets

7: Compute on data 1

6: Hsapiens_SNPs_Ch22

5: Hsapiens_Repeats_Ch22

4: Hsapiens_CExons_Ch22

3: Hsapiens_Introns_Ch22

2: Hsapiens_Exons_Ch22

1: Hsapiens_Genes_Ch22

12.87 MB

Computational Arithmetics

Tools  

search tools 

Upload Data 

HPRC

Get Data

Send Data

Collection Operations

Lift-Over

Text Manipulation

Convert Formats

Filter and Sort

Join, Subtract and Group

Join two Datasets side by side on a specified field

Compare two Datasets to find common or distinct rows

Group data by a column and perform aggregate operation on other columns.

Fetch Alignments/Sequences

Operate on Genomic Intervals

Statistics

Graph/Display Data

Phenotype Association

BED

Annotation

Multiple Alignments

NCBI BLAST+

Mapping

SAM/BAM

3: Operation

Type: Minimum

On column: Column: 13

Round result to nearest integer? NO

Replace non numeric data: leave empty for no replacements. Will replace, e.g., empty cells and text cells.

+ Insert Operation

Job Resource Parameters

Specify job resource parameters

Memory (GB): 7

Maximum Job Memory

Time (hours): 24

Maximum job time

Execute

 TIP: If your data is not TAB delimited, use Text Manipulation->Convert

Syntax

This tool allows you to group the input dataset by a particular column and perform aggregate functions: Mean, Median, Mode, Sum, Max, Min, Count, Concatenate, and Randomly pick on any column(s).

The Concatenate function will take, for each group, each item in the specified column and build a comma delimited list. Concatenate Unique will do the same but will build a list of unique items with no repetition.

Count and Count Unique are equivalent to Concatenate and Concatenate Unique, but will only count the number of items and will return an integer.

- If multiple modes are present, all are reported.

History    

search datasets  

12.87 MB   

7: Compute on data 1   

6: Hsapiens_SNPs_Chromosome 22   

5: Hsapiens_Repeats_Chromosome 22   

4: Hsapiens_CExons_Chromosome 22   

3: Hsapiens_Introns_Chromosome 22   

2: Hsapiens_Exons_Chromosome 22   

1: Hsapiens_Genes_Chromosome 22   

Computational Arithmetics

Tools

search tools x

Upload Data +

HPRC

Get Data

Send Data

Collection Operations

Lift-Over

Text Manipulation

Convert Formats

Filter and Sort

Join, Subtract and Group

Join two Datasets side by side on a specified field

Compare two Datasets to find common or distinct rows

Group data by a column and perform aggregate operation on other columns.

Fetch Alignments/Sequences

Operate on Genomic Intervals

Statistics

Graph/Display Data

Phenotype Association

BED

Annotation

Multiple Alignments

NCBI BLAST+

Mapping

SAM/BAM

4: Operation

Type: Mean

On column: Column: 13

Round result to nearest integer?: NO

Replace non numeric data: leave empty for no replacements. Will replace, e.g., empty cells and text cells.

+ Insert Operation

Job Resource Parameters

Specify job resource parameters

Memory (GB): 7

Maximum Job Memory

Time (hours): 24

Maximum job time

Execute

Mean on C07 (contains lengths of transcripts)

History

search datasets

12.87 MB

7: Compute on data 1

6: Hsapiens_SNPs_Chromosome 22

5: Hsapiens_Repeats_Chromosome 22

4: Hsapiens_CExons_Chromosome 22

3: Hsapiens_Introns_Chromosome 22

2: Hsapiens_Exons_Chromosome 22

1: Hsapiens_Genes_Chromosome 22

Computational Arithmetics

Tools ☆ ⚙

search tools ×

Upload Data

HPRC

Get Data

Send Data

Collection Operations

Lift-Over

Text Manipulation

Convert Formats

Filter and Sort

Join, Subtract and Group

Fetch Alignments/Sequences

Operate on Genomic Intervals

Statistics

Graph/Display Data

Phenotype Association

BED

Annotation

Multiple Alignments

NCBI BLAST+

Mapping

SAM/BAM

Assembly

FASTQ Quality Control

FASTA/FASTQ

RNA-seq

CD-HIT

chr22 2197 701827 33 35456.4

Final Result

Final Result

Final Result

History ⟳ + ⚙

search datasets ? ×

L11-A

8 shown

12.87 MB ✓ 🌐 💬

8: Group on data 7	eye edit x
7: Compute on data 1	eye edit x
6: Hsapiens_SNPs_Chromosome 22	eye edit x
5: Hsapiens_Repeats_Chromosome 22	eye edit x
4: Hsapiens_CExons_Chromosome 22	eye edit x
3: Hsapiens_Introns_Chromosome 22	eye edit x
2: Hsapiens_Exons_Chromosome 22	eye edit x
1: Hsapiens_Genes_Chromosome 22	eye edit x