

1) Complete [this tutorial](#). (Variant Analysis)

a) Show your steps.

- Very briefly explain what you are doing with each step (What is your data format? What does it contain? What does the tool you are using do?)
- Make sure your assignment contains FreeBayes, SNPEff, Gemini outputs.

b) Answer the questions below the “Querying genotypes” title, using screenshots.

c) Share your Galaxy history link.

2) Visit [this tutorial](#). (IGV)

- a) Visualize the mapped reads from question 1 on IGV (Make sure you have selected Human hg19 on the top-left corner of the IGV window first). Share screenshots of some of the reads you find. Comment on your visuals (consider the colors, the directions of the reads, etc.).
- b) Do the “Encode RNA-Seq data” part from the tutorial.
  - i) Share a screenshot containing the aligned reads. Compared to what you saw in ‘part a’, what are the differences you observe in the aligned reads? What do the lines connecting the reads represent? What is the reason for this difference?
  - ii) Share a screenshot of sashimi plots. What do the red lines and the numbers represent?
- c) Download the FreeBayes output from question 1 and load it into IGV (again, make sure Human hg19 is selected).
  - i) Share a screenshot from the IGV window, showing some variants.
  - ii) Select a variant. Click on it to see the details for the mother, the father, and the son. Compared to the reference genome, who carries which variant?



# Calling variants in diploid systems

Authors: Anton Nekrutenko

Nicola Soranzo

Add Contributions!

## Overview

### Questions:

- How to find variable sites in diploid genomes?



### Objectives:

- Identification of genetic variations using variant calling

### Requirements:

- Introduction to Galaxy Analyses
- Sequence analysis
  - Quality Control: slides - hands-on
  - Mapping: slides - hands-on

Time estimation: 3 hours

Supporting Materials:

Topic Overview slides Datasets Workflows Tours Available on these Galaxies

OPEN CHAT

Using 1%

## History



search datasets



### variant calling exercise

(empty)



This history is empty. You can **load your own data** or **get data from an external source**

[https://usegalaxy.eu/u/lemannurnehri  
/h/variant-calling-exercise](https://usegalaxy.eu/u/lemannurnehri/h/variant-calling-exercise)

opening a new history

zenodo.org/search?page=1&size=20&q=GIAB-Ashkenazim-Trio.txt&file\_type=bam

MPN navigator yön... Makale draft leman... GeoGebra - 100 mil... Middle East Technic... Probabilistic Graphi... Nizamettin Aydin Diger yer işaretleri Okuma listesi

# zenodo

GIAB-Ashkenazim-Trio.txt

Upload Communities

All versions

Access Right

Open (12)  Restricted (1)

File Type

Gz (6)  Pdf (5)  Txt (3)  Bam (1)  Bed (1)  Md (1)  Rmd (1)

Found 1 result.

August 19, 2016 (v1) Dataset Open Access

## Training material for exome sequencing

Grünig, Björn; Erxleben, Anika; Houwaart, Torsten; Batut, Bérénice;

Exome sequencing means that all protein-coding genes in a genome are sequenced. In Humans, there are ~180,000 exons that makes up 1% of the human genome which contain ~30 million base pairs. Mutations in the exome have usually a higher impact and more severe consequences, than in the rema

Uploaded on September 8, 2016

Sort by: Best match asc.

Log in Sign up

downloaded  
the desired  
data: Zenodo is  
a tool for  
developing and  
organising the  
Big Data

< 1 >

< 1 >

11°C Kışman günəşli 14:07

## selecting desired data from zenodo

Files (102.0 MB)

### Name

### Size

dbSNP\_138.hg19.vcf

2.1 MB

 Download

md5:1bb54779b6e564062398ca593738d8f2 

father.bam

31.8 MB

 Download

md5:32b6da238924e0e8c702092891d32ede 

GIAB-Ashkenazim-Trio-hg19.gz

52.4 kB

 Download

md5:e7e4d5774877fb325335c2d4b0a1c015 

GIAB-Ashkenazim-Trio.txt

231 Bytes

 Download

md5:384ecad45c4f603d1f40baec5f2a0b79 

mother.bam

33.5 MB

 Download

md5:2463b4df4634b99b5ba49bb055e0c446 

patient.bam

34.4 MB

 Download

md5:2a856f42d30fd90efab48f51ebe1298b 

August 19, 2016

### DOI:

 DOI 10.5281/zenodo.60520

### Communities:

Galaxy Training Network

Zenodo

### License (for files):

 Creative Commons Attribution 4.0 International

### Share



### Cite as

Grüning, B., Erxleben, A., Houwaart, T., & Batut, B. (2016). Training material for exome sequencing [Data set]. Zenodo.

<https://doi.org/10.5281/zenodo.60520>

Start typing a citation style...



Workflow

Visualize

Shared Data

Help

User



## Download from web or upload from disk

Regular

Composite

Collection

Rule-based

uploading the desired data to history in galaxy

Please wait...2 out of 2 remaining.

Name	Size	Type	Genome	Settings	Status
GIAB-Ashkenazim-Tr	231 b	Auto-det...	unspecified (?)		Adding to history...
GIAB-Ashkenazim-Tr	141.9 KB	Auto-det...	unspecified (?)		0%

Download from web or upload from disk

Type (set all):

Auto-detect



Genome (set all):

unspecified (?)

Choose local files

Choose remote files

Paste/Fetch data

Start

Pause

Reset

Close

changing the database operations

Add an annotation or notes to a dataset; annotations are available when a history is viewed.

## Database/Build

Human Feb. 2009 (GRCh37/hg19) (hg19)

 Save

 Auto-detect

### Choose the source for the reference genome

Locally cached

changing the desired parameters to search on hg19

### Run in batch mode?

- Run individually  
 Merge output VCFs

Selecting individual mode will generate one VCF dataset for each input BAM dataset. Selecting the merge option will produce one VCF dataset for all input BAM datasets

### BAM dataset



2: GIAB-Ashkenazim-Trio-hg19



### Using reference genome

Human (Homo sapiens): hg19

### Limit variant calling to a set of regions?

Do not limit

Sets --targets or --region options

### Choose parameter selection level

5. Full list of options

changing the algorithmic parameters to calculate marginal prob.

(--read-dependence-factor)

**Calculate the marginal probability of genotypes and report as GQ in each sample field in the VCF output**



Yes

(--genotype-qualities)

**Email notification**

executing: freebayes:a  
bayesian genetic variation  
detector working on our data



Executed **FreeBayes** and successfully added 1 job to the queue.

The tool uses this input:

- **2: GIAB-Ashkenazim-Trio-hg19**

It produces this output:

- **3: FreeBayes on data 2 (variants)**

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

**Retain MNPs as separate events**

No

`--use-mnps` option

VcfAllelicPrimitives--changing the desired parameters to produce a dataset in VFC format

**Tag records which are split apart of a complex allele with this flag.**

Split primitives

`--tag-parsed` option**Do not manipulate records in which either the ALT or REF is longer than (bp)**

200

`--max-length` option**Maintain site and allele-level annotations when decomposing**

Yes

Note that in many cases, such as multisample VCFs, these won't be valid post-decomposition. For biallelic loci in single-sample VCFs, they should be usable with caution. (`--keep-info`)

**Maintain genotype-level annotations when decomposing**

Yes

~~Similar caution should be used for this as for --keep-info. (`--keep-geno`)~~

**Email notification**

executing



Executed **VcfAllelicPrimitives**: and successfully added 1 job to the queue.

The tool uses this input:

- **3: FreeBayes on data 2 (variants)**

It produces this output:

- **4: VcfAllelicPrimitives: on data 3**

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

## Sequence changes (SNPs, MNPs, InDels)



6: VcfAllelicPrimitives: on data 5



## Input format

VCF

changing the desired options in snpeff to  
annotate variants and estimate their effects on  
the genes

## 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



Executed **SnpEff eff:** and successfully added 1 job to the queue.

The tool uses this input:

- **6: VcfAllelicPrimitives: on data 5**

executing

It produces 2 outputs:

- **7: SnpEff eff: on data 6**
- **8: SnpEff eff: on data 6 - HTML stats**

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When a job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems have been encountered.

We need your support ...

### VCF dataset to be loaded in the GEMINI database



7: SnpEff eff: on data 6



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

#### The variants in this input are

annotated with snpEff

changing the options

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



1: GIAB-Ashkenazim-Trio.txt



The pedigree dataset is optional, but several GEMINI tools require the relationship between samples (i.e., the family structure).

If the sample relationships are not defined, the PED format is ignored, and variants from the VCF file are loaded as if they were all independent samples.

execution

GEMINI (GEnome MINIng):explores the genetic variations among genome annotations



Executed **GEMINI query** and successfully added 1 job to the queue.

The tool uses this input:

- **9: GEMINI load on data 1 and data 7**

It produces this output:

- **12: GEMINI query on data 9**

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

## GEMINI database



15: GEMINI load on data 1 and data 10

querying with SQL to  
search on the db



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 rs_ids, aaf, esp, ea, impact, clinvar_disease_name, clinvar_sig FROM variants WHERE filter is NULL and gene = 'POLRMT'
```

Formulate your query using SQL syntax. (-q)

### Genotype filter expression

 Insert Genotype filter expression

### Sample filter expression



Executed **GEMINI database info** and successfully added 1 job to the queue.

The tool uses this input:

execution

- **15: GEMINI load on data 1 and data 10**

It produces this output:

- **17: GEMINI database info on data 15**

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

## Dataset Information

Number	20
Name	GEMINI query on data 15
Created	Wednesday Dec 8th 6:14:34 2021 UTC
Filesize	<b>1.6 KB</b>
Dbkey	hg19
Format	tabular
File contents	contents
History Content API ID	11ac94870d0bb33aa6b7ae52e8aa8aea
History API ID	ec4f7f8d611e8bf5
UUID	4cc4584c-d881-40d7-bab8-423e964a8e14
Full Path	/data/dnb05/galaxy_db/files/4/c/c/dataset_4cc4584c-d881-40d7-bab8-423e964a8e14.dat

## Tool Parameters

Input Parameter	Value
-----------------	-------

## Job Dependencies

Dependency	Dependency Type	Version
\$gemini	\$conda	\$0.20.1

## Dataset Peek:

1	2	3	4	5
rs_ids	aaf_esp_ea	impact	clinvar_disease_name	clinvar_sig
rs41551212	0.169651162791	synonymous_variant	None	None
rs144281668	0.000116306117702	synonymous_variant	None	None
None	-1.0	intron_variant	None	None
rs11672829	-1.0	intron_variant	None	None

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 * from variants
```

Formulate your query using SQL syntax. (-q)

querying with  
gemini query ,  
for question 1  
which is  
selecting among  
variants with  
restrictions

### Genotype filter expression

1: Genotype filter expression



### Restrictions to apply to genotype values

```
gt_types.HG002_NA24385_son <> HOM_REF
```

|

(--gt-filter)

## Dataset Information

Number	21
Name	GEMINI query on data 15
Created	Wednesday Dec 8th 6:16:13 2021 UTC
Filesize	<b>14.4 KB</b>
Dbkey	hg19
Format	tabular
File contents	contents
History Content API ID	11ac94870d0bb33abe0e42c5ea759a84
History API ID	ec4f7f8d611e8bf5
UUID	7e34e324-10fd-4f7f-b957-b1df451d6026
Full Path	/data/dnb05/galaxy_db/files/7/e/3/dataset_7e34e324-10fd-4f7f-b957-b1df451d6026.dat

## Tool Parameters

This job requested 1 cores and 4 Gb. Given this, the smallest EC2 machine we could find is t2.medium (4 GB / 2 vCPUs / Intel Xeon Family). That instance is priced at 0.0536 USD/hour.

Please note, that those numbers are only estimates, all jobs are always free of charge for all users.

## Job Dependencies

Dependency	Dependency Type	Version
\$gemini	\$conda	\$0.20.1

## Dataset Peek:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
chrom	start	end	vcf_id	variant_id	anno_id	ref	alt	qual	filter	type	sub_type	call_rate	max_aaf_all	in_dbsnp
chr19	617803	617804	None	2	3	G	A	0.309944987297	None	snp	ts	1.0	0.00295186194369	1
chr19	618158	618159	None	4	2	A	G	193.703994751	None	snp	ts	1.0	0.7832	1
chr19	619020	619021	None	9	1	G	C	352.244995117	None	snp	tv	1.0	0.7822	1
chr19	619138	619139	None	10	2	G	A	137.481994629	None	snp	ts	1.0	0.0537	1

## GEMINI database



15: GEMINI load on data 1 and data 10



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 * from variants
```

querying with  
gemini query ,  
for question 2  
which is  
selecting among  
variants with  
restrictions

Formulate your query using SQL syntax. (-q)

## Genotype filter expression

1: Genotype filter expression



## Restrictions to apply to genotype values

```
(gt_types.HG002_NA24385_son <> HOM_REF AND gt_types.HG003_NA24149_father <> HOM_REF)
```



Executed **GEMINI query** and successfully added 1 job to the queue.

The tool uses this input:

- 15: **GEMINI load on data 1 and data 10**

execution

It produces this output:

- 22: **GEMINI query on data 15**

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

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 gts.HG002_NA24385_son, gts.HG003_NA24149_father from variants
```

querying with gemini query , for question 3 which is selecting among variants with restrictions

Formulate your query using SQL syntax. (-q)

#### Genotype filter expression

1: Genotype filter expression



#### Restrictions to apply to genotype values

```
(gt_types.HG002_NA24385_son <> HOM_REF AND gt_types.HG003_NA24149_father <> HOM_REF)
```

(--gt-filter)

Insert Genotype filter expression



Executed **GEMINI query** and successfully added 1 job to the queue.

The tool uses this input:

- **15: GEMINI load on data 1 and data 10**

execution

It produces this output:

- **23: GEMINI query on data 15**

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

## Dataset Information

Number	22
Name	GEMINI query on data 15
Created	Wednesday Dec 8th 6:17:10 2021 UTC
Filesize	<b>11.7 KB</b>
Dbkey	hg19
Format	tabular
File contents	contents
History Content API ID	11ac94870d0bb33a9e5dfc725abfd790
History API ID	ec4f7f8d611e8bf5
UUID	dc05269a-387c-4039-80a7-dcfecb29b2ea
Full Path	/data/dnb05/galaxy_db/files/d/c/0/dataset_dc05269a-387c-4039-80a7-dcfecb29b2ea.dat

## Tool Parameters

Input Parameter	Value

## Job Dependencies

Dependency	Dependency Type	Version
\$gemini	\$conda	\$0.20.1

## Dataset Peek:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
chrom	start	end	vcf_id	variant_id	anno_id	ref	alt	qual	filter	type	sub_type	call_rate	max_aaf_all	in_dbsnp	rs_ids
chr19	618158	618159	None	4	2	A	G	193.703994751	None	snp	ts	1.0	0.7832	1	rs1167
chr19	619020	619021	None	9	1	G	C	352.244995117	None	snp	tv	1.0	0.7822	1	rs1415
chr19	619138	619139	None	10	2	G	A	137.481994629	None	snp	ts	1.0	0.0537	1	rs1166
chr19	619407	619408	None	11	2	A	G	207.722000122	None	snp	ts	1.0	0.4581	1	rs1085

## GEMINI database



15: GEMINI load on data 1 and data 10



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 chrom, start, end, ref, alt, gene, impact, (gts).(*) FROM variants
```

the last question, selecting from variants table for chromosome, start and end points, ref. genome and gene and its impact

Formulate your query using SQL syntax. (-q)

## Genotype filter expression

1: Genotype filter expression



## Restrictions to apply to genotype values

(gt\_types).(\*).(=HET).(all)



Executed **GEMINI query** and successfully added 1 job to the queue.

The tool uses this input:

- 15: GEMINI load on data 1 and data 10

execution

It produces this output:

- 24: GEMINI query on data 15

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

## Dataset Information

Number	24
Name	GEMINI query on data 15
Created	Wednesday Dec 8th 6:19:42 2021 UTC
Filesize	<b>108</b> b
Dbkey	hg19
Format	tabular
File contents	contents
History Content API ID	11ac94870d0bb33aaf3d3aa2390deeab
History API ID	ec4f7f8d611e8bf5
UUID	eaf7b3ea-28b6-4a65-b6fd-7262d370aa25
Full Path	/data/dnb05/galaxy_db/files/e/a/f/dataset_eaf7b3ea-28b6-4a65-b6fd-7262d370aa25.dat

## Tool Parameters

## AWS estimate

**0.00 USD**

This job requested 1 cores and 4 Gb. Given this, the smallest EC2 machine we could find is t2.medium (4 GB / 2 vCPUs / Intel Xeon Family). That instance is priced at 0.0536 USD/hour.

Please note, that those numbers are only estimates, all jobs are always free of charge for all users.

## Job Dependencies

Dependency	Dependency Type	Version
\$gemini	\$conda	\$0.20.1

## Dataset Peek:

1    2    3    4    5    6    7    8

9

10

chrom start end ref alt gene impact gts.HG002\_NA24385\_son gts.HG003\_NA24149\_father gts.HG004\_NA24143\_mother

Genome ▾

Tracks ▾

Session ▾

Share

Bookmark

Save SVG

Help ▾

Local File ...

Dropbox  ...Google Drive  ...

URL ...

Human (GRCh38/hg38)

Human (hg38 1kg/GATK)

Human (GRCh37/hg19)

Human (hg18)

Mouse (GRCm39/mm39)

Mouse (GRCm38/mm10)

Mouse (NCBI37/mm9)

Rat (rn7)

Rat (RGCS 6.0/rn6)

Gorilla (Kamilah\_GGO\_v0/gorGor6)

Gorilla (gorGor4.1/gorGor4)

Chimp (panTro6) (panTro6)

Chimp (panTro5) (panTro5)

Chimp (SAC 2.1.4/panTro4)



5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19

Zoom in to see features

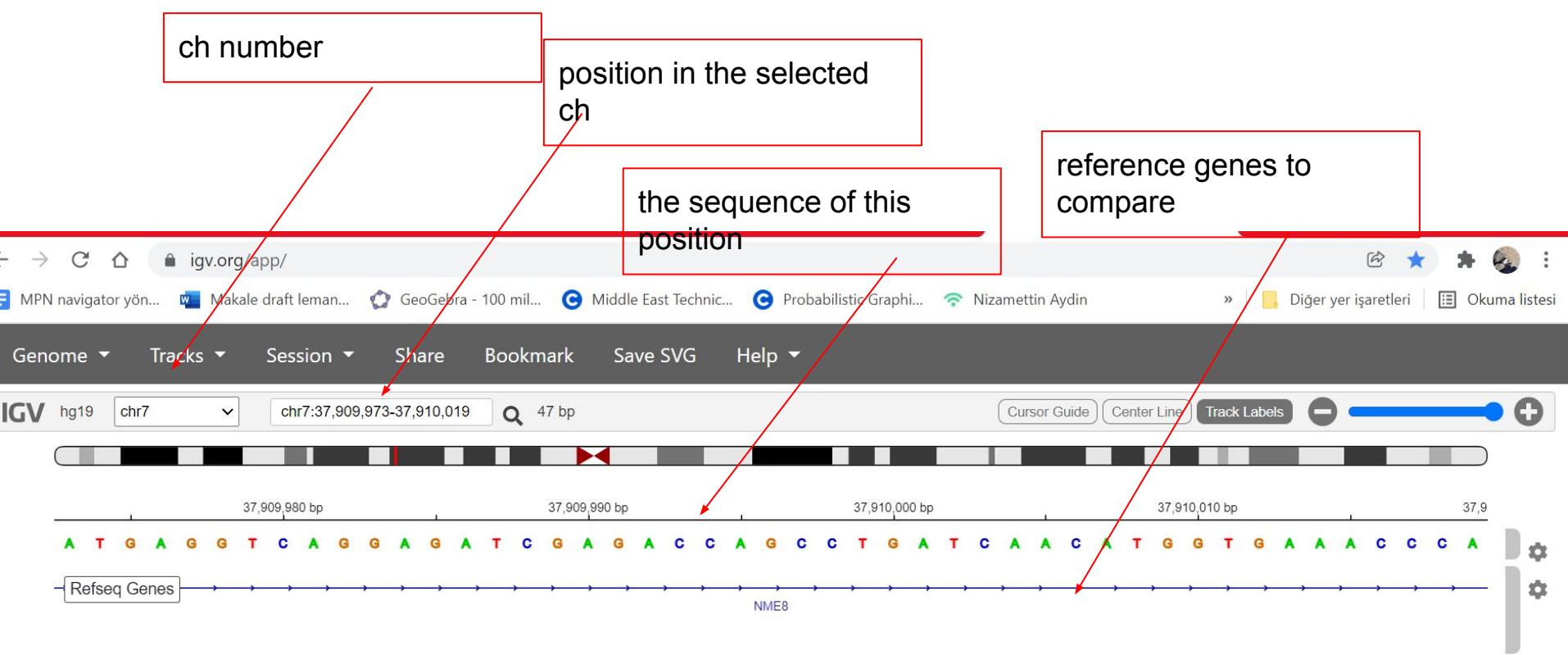
igv for the visualization of the alignment

ch number

position in the selected  
ch

reference genes to  
compare





M

Human (GRCh37/hg19)

All

## Encode Production Data

Filter:

19.954 rows

	cell	data Type	antibody	view	replicate	type	lab	hub
	8988T	DnaseSeq		Peaks		narrowPeak	Duke	Data
	AoSMC	DnaseSeq		Peaks		narrowPeak	Duke	Data
	Chorion	DnaseSeq		Peaks		narrowPeak	Duke	Data
	CLL	DnaseSeq		Peaks		narrowPeak	Duke	Data
	Fibrobl	DnaseSeq		Peaks		narrowPeak	Duke	Data
	FibroP	DnaseSeq		Peaks		narrowPeak	Duke	Data
	Gliobla	DnaseSeq		Peaks		narrowPeak	Duke	Data
	GM12891	DnaseSeq		Peaks		narrowPeak	Duke	Data
	GM12892	DnaseSeq		Peaks		narrowPeak	Duke	Data
	GM18507	DnaseSeq		Peaks		narrowPeak	Duke	Data
	GM19238	DnaseSeq		Peaks		narrowPeak	Duke	Data
	GM19239	DnaseSeq		Peaks		narrowPeak	Duke	Data
	GM19240	DnaseSeq		Peaks		narrowPeak	Duke	Data
	H9ES	DnaseSeq		Peaks		narrowPeak	Duke	Data
	HeLa-S3	DnaseSeq		Peaks		narrowPeak	Duke	Data
	Hepatocytes	DnaseSeq		Peaks		narrowPeak	Duke	Data
	HPDE6-E6E7	DnaseSeq		Peaks		narrowPeak	Duke	Data
	HSMM_emb	DnaseSeq		Peaks		narrowPeak	Duke	Data
	HTR8svn	DnaseSeq		Peaks		narrowPeak		Data
	Huh-7.5	DnaseSeq		Peaks		narrowPeak	Duke	Data
	Huh-7	DnaseSeq		Peaks		narrowPeak	Duke	Data
	iPS	DnaseSeq		Peaks		narrowPeak	Duke	Data
	Ishikawa	DnaseSeq		Peaks		narrowPeak	Duke	Data
	Ishikawa	DnaseSeq		Peaks		narrowPeak	Duke	Data
	LNCaP	DnaseSeq		Peaks		narrowPeak	Duke	Data
	MCF-7	DnaseSeq		Peaks		narrowPeak	Duke	Data
	Medullo	DnaseSeq		Peaks		narrowPeak	Duke	Data
	Melano	DnaseSeq		Peaks		narrowPeak	Duke	Data
	Mvometr	DnaseSeq		Peaks		narrowPeak	Duke	Data

Load

Cancel

## Encode Production Data



Filter: c2c

0 rows

^	cell	dataType	antibody	view	replicate	type	lab	hub
1								16

hocam there is  
no rnaseq for the  
C2C12 RNA-Seq  
BAM files , probably  
my version of igv  
(which is a desktop  
versian included  
java-2012) not  
include these  
sequences.

Load

Cancel

Human (GRCh37/hg19)

▼ All

## Encode Production Data

Filter: rnase

3.419 rows

	cell	dataType	antibody	view	replicate	type	lab	hub
	HSMM	RnaSeq		TranscriptG... 1		gtf	Caltech	Data
	HSMM	RnaSeq		TranscriptG... 2		gtf	Caltech	Data
	HSMM	RnaSeq		Signal 1		bigWig	Caltech	Data
	HSMM	RnaSeq		Signal 1		bigWig	Caltech	Data
	HSMM	RnaSeq		Signal 2		bigWig	Caltech	Data
	HSMM	RnaSeq		Signal 2		bigWig	Caltech	Data
✓	HUVEC	RnaSeq		Alignments 1		bam	Caltech	Data
✓	HUVEC	RnaSeq		Alignments 2		bam	Caltech	Data
	HUVEC	RnaSeq		MinusRawS... 1		bigWig	Caltech	Data
	HUVEC	RnaSeq		MinusRawS... 2		bigWig	Caltech	Data
	HUVEC	RnaSeq		PlusRawSig... 1		bigWig	Caltech	Data
	HUVEC	RnaSeq		PlusRawSig... 2		bigWig	Caltech	Data
	HUVEC	RnaSeq		Splices 1		bam	Caltech	Data
	HUVEC	RnaSeq		Splices 2		bam	Caltech	Data
	HUVEC	RnaSeq		MinusRawS... 1		bigWig	Caltech	Data
	HUVEC	RnaSeq		MinusRawS... 2		bigWig	Caltech	Data
	HUVEC	RnaSeq		PlusRawSig... 1		bigWig	Caltech	Data
	HUVEC	RnaSeq		PlusRawSig... 2		bigWig	Caltech	Data
	HUVEC	RnaSeq		MinusRawS... 1		bigWig	Caltech	Data
	HUVEC	RnaSeq		MinusRawS... 2		bigWig	Caltech	Data
	HUVEC	RnaSeq		PlusRawSig... 1		bigWig	Caltech	Data
	HUVEC	RnaSeq		PlusRawSig... 2		bigWig	Caltech	Data
	HUVEC	RnaSeq		Alignments 1		bam	Caltech	Data
	HUVEC	RnaSeq		Alignments 2		bam	Caltech	Data
	HUVEC	RnaSeq		ExonsGen... 1		gtf	Caltech	Data
	HUVEC	RnaSeq		ExonsGen... 2		gtf	Caltech	Data
	HUVEC	RnaSeq		GeneDeNovo 1		gtf	Caltech	Data
	HUVEC	RnaSeq		GeneDeNovo 2		gtf	Caltech	Data
	HUVEC	RnaSeq		GeneGenCV3c1		gtf	Caltech	Data

Load

Cancel

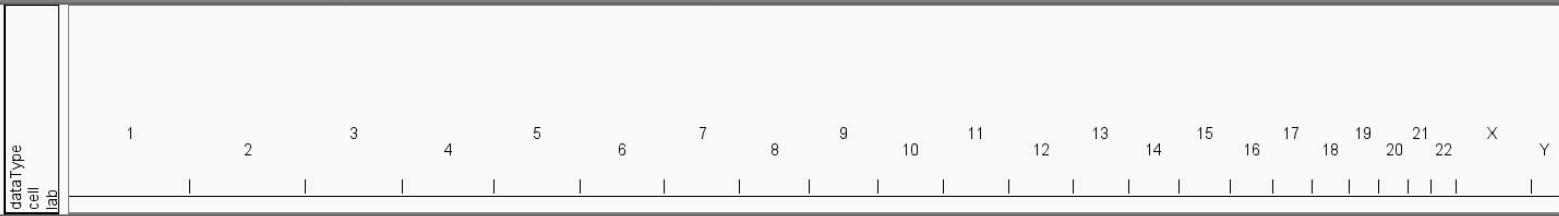
i selected HUVEC bam files to finish the tutorial

16 17 18 19 20 21 22 X

Human (GRCh37/hg19)

All

Go



HUVEC RnaSeq Alignments rep 1 C  
verage

Zoom in to see coverage.

output

HUVEC RnaSeq Alignments rep 1 J

HUVEC RnaSeq Alignments rep 1

Zoom in to see alignments.

HUVEC RnaSeq Alignments rep 2 C  
verage

Zoom in to see coverage.

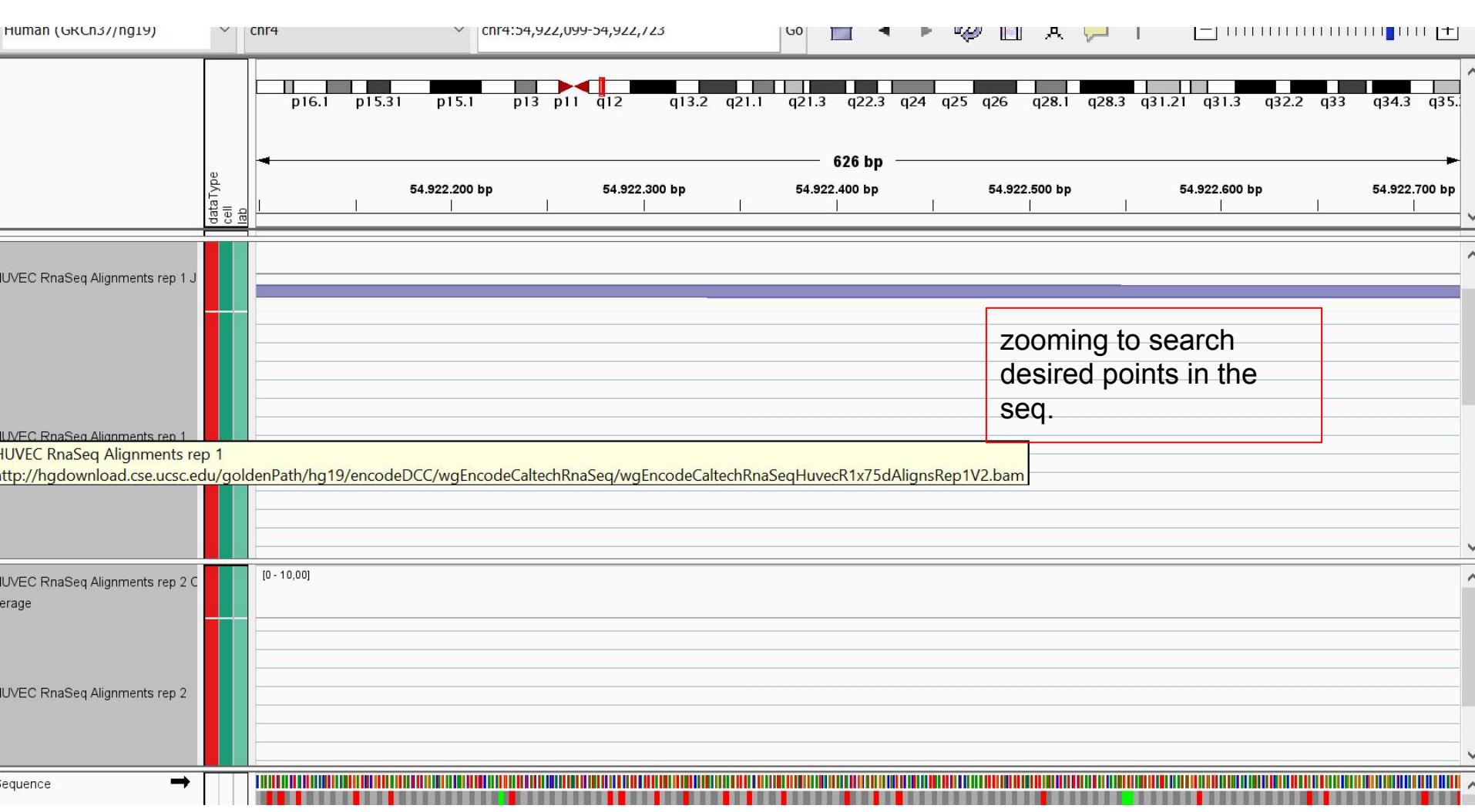
HUVEC RnaSeq Alignments rep 2

Zoom in to see alignments.

Refseq Genes

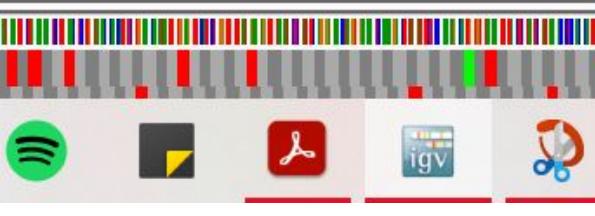
Zoom in to see features.

CHAT



visualization of splice  
junctions from aligned  
rna-seq data & gene  
annotation

[0 - 10,000]



Show all bases

Quick consensus mode

View as pairs

Go to mate

View mate region in split screen

Set insert size options ...

Collapsed

Expanded

Squished

Select by name...

Clear selections

Copy read sequence

BLAT read sequence

Copy consensus sequence

Sashimi Plot

Show Coverage Track

Show Splice Junction Track

9°C Çök



Ana

[File](#) [Genomes](#) [View](#) [Tracks](#) [Regions](#) [Tools](#) [Help](#)

Human (GRCh37/hg19)

chr4

chr4:54,922,099-54,922,723

Go



54.922.200 bp

54.922.300 bp

54.922.400 bp

54.922.500 bp

54.922.600 bp

 dataType  
 cell  
 lab

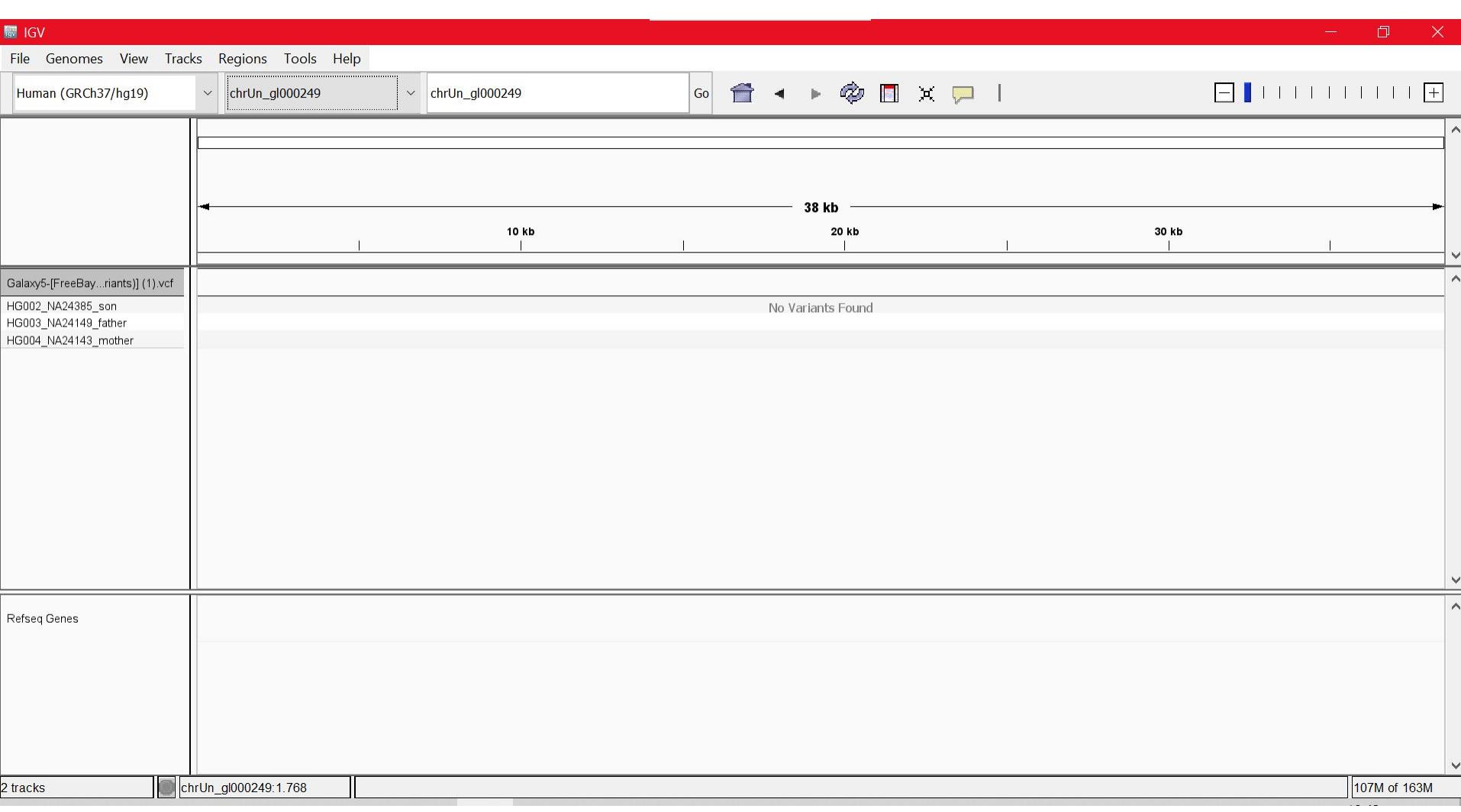
Galaxy5-[FreeBayesVariants] (1).vcf

HG002\_NA24385\_son

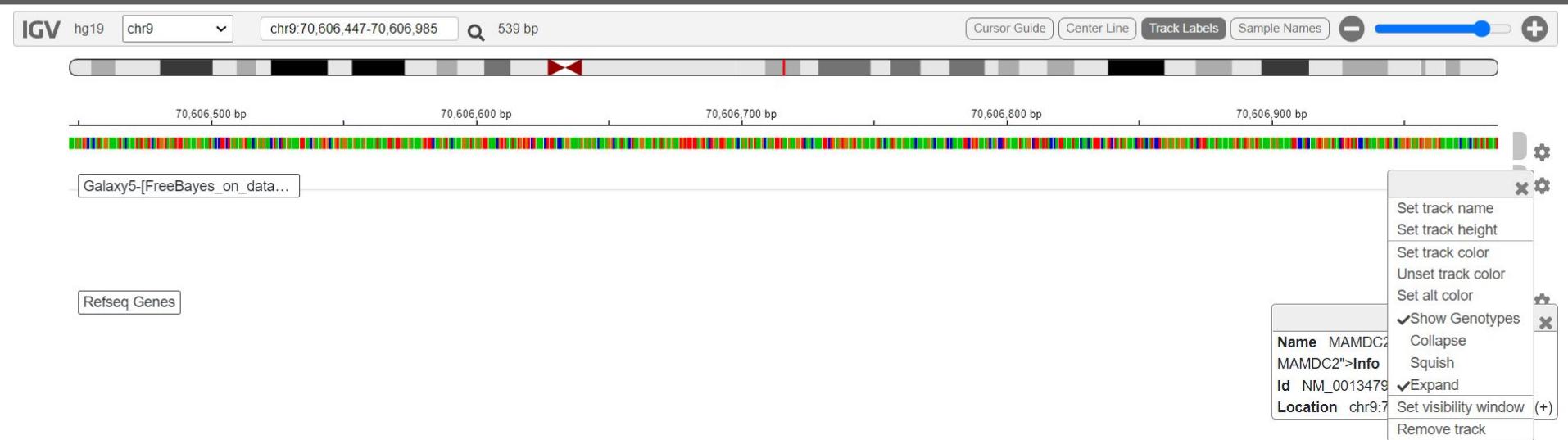
HG003\_NA24149\_father

HG004\_NA24143\_mother

No Variants Found



Genome Tracks Session Share Bookmark Save SVG Help



# References

Wernersson, R., & Pedersen, A. G. (2003). RevTrans: Multiple alignment of coding DNA from aligned amino acid sequences. Nucleic acids research, 31(13), 3537–3539. <https://doi.org/10.1093/nar/gkg609>

<https://www.ebi.ac.uk/seqdb/confluence/display/THD/Help+-+Clustal+Omega+FAQ#HelpClustalOmegaFAQ-What+isClustalOmega?>

<https://user.ceng.metu.edu.tr/~tcan/ceng465/Schedule/MSAComparison.pdf>