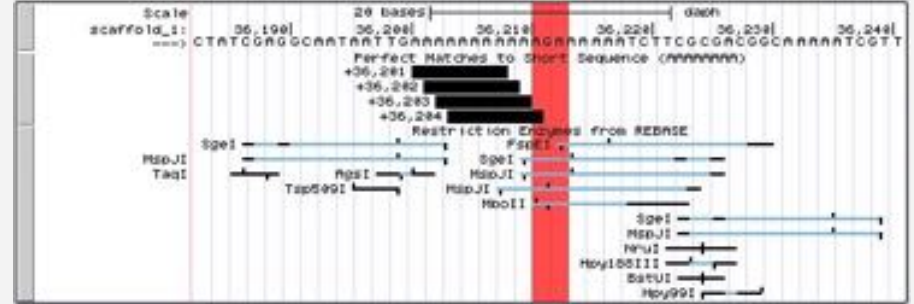


Benefits of Track Hubs For Visualizing Data

```
>scaffold_1
GTTGTAAATACTCTATTCTACAATAAAACCAA
TCATAGGTTGAATTGGCGTTGAAGTAAAAACAAA
...
>scaffold_2
AGTTATGACAAACTATAAAAAGTCGGTAGAGACAAAAG
TCGTTTCGTGGACGAAGCGACCAAAACTGAGCACAAGAT?
...
>scaffold_3
CATAAATTCATAAATCAATTCATGAAGAATAATT
TAGAAAATTTCCAGGAAGTTTGAAGTTGCTAT
...
```

```
hubDirectory
|__hub.txt
|__genomes.txt
|__hg19
|__trackDb.txt
```



via byte-range requests

1. What are Assembly Track Hubs?

2. Track Hub Resources

A 2bit file is a binary indexed version of a FASTA file
(stores sequence ACGT as 00 01 11 10)

faToTwoBit input.fasta output.2bit



```
>scaffold_1
GTTGTAAATACTCTATTCTACAATAAAACCAAAGATAACTCTTTATCAG
TCATAGGTTGAATTGGCGTTGAAGTAAAACAAAAATACTGCTCAAAAGG
...
>scaffold_2
AGTTATGACAAACTATAAAAAAGTCGGTAGAGACAAAAGCGAAAAGGATCT
TCGTTTCGTGGACGAAGCGACCAAACTGAGCACAAGATAAATCCCGAATA
...
>scaffold_3
CATAAATTCATAAATCAATTCATGAAGAATAATTTTAGAAAAATGGTTCA
TAGAAAATTTCCAGGAAGTTTGAAGTTGCTATAATGATTATTTCTCTTG
...
>scaffold_4
ATACTGCCCCATGCCTCATCTACTCCCTTCTTAACCTTTTCATCAACAACT
TGCTTAAGTTGGTCATCACTGACTGAGTGAGCCAAACTGGAATAGCTGT
...
```

The resulting indexed binary file
output.2bit can have data more easily
extracted and can also be viewed in the
UCSC Genome Browser.

Track Hubs

Track Hubs are text files (**hub.txt**, **genomes.txt**, **trackDb.txt**) that describe and point to the location of binary indexed data files (**twoBitPath**, **bigDataUrl**)

hubDirectory

|

|_hub.txt

|_genomes.txt

|_genomeDirectory1

|_trackDb.txt

Informs location of genome 2bit file(s): **twoBitPath**

Informs location of track data file(s): **bigDataUrl**

Track Hubs

Track Hubs are text files (**hub.txt**, **genomes.txt**, **trackDb.txt**) that describe and point to the location of binary indexed data files (**twoBitPath**, **bigDataUrl**)

hub.txt

hub *MyHubsNameWithoutSpaces*
shortLabel *My Hub's Name*
longLabel *Longer label about my hub.*
email *myEmail@address*
genomesFile *http://location_of/genomes.txt*



genomes.txt

genome *yourGenome*
trackDb *http://location_of/trackDb.txt*
twoBitPath *http://location_of/output.2bit*
...

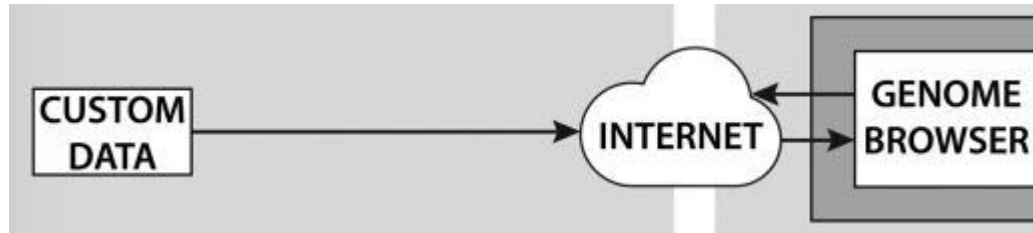
trackDb.txt

track *bam1*
type *bam*
shortLabel *BAM example*
longLabel *This BAM file is a RNA-seq example from the ENCODE project*
visibility *dense*
bigDataUrl *http://location_of/file.bam*
...
track *bam2*
...
track *bam3*
...

Each new track gets a new stanza

bigDataUrl lines point to location of binary indexed data

More trackDb parameters can be defined too (color, ect.)



twoBitPath http://location_of/output.2bit

genomes.txt shares where to find the 2bit (and what to call new genome)

genome *yourGenome*
trackDb http://location_of/trackDb.txt
scientificName *Your Genome*
description *Feb. 2017 Assembly*
organism *Your organism*
defaultPos *chr1:1000000-2000000*
groups *groups.txt*
htmlPath <http://yourGenome/description.html>
twoBitPath http://location_of/output.2bit

Genomic Window of Sequence viewed at UCSC

Three tracks are generated from the 2bit on the fly:

- Base Position
- Restriction Enzymes
- Short Match Track (30 bases)

Adding Annotations

simple as adding text-based custom tracks



track name=ex1 description="Blue example BED3 track" color=0,0,255

chr21 33030000 33040000

chr21 33050000 33060000

...

Adding tracks inside Track Hubs

create binary indexed versions of text files (`bedToBigBed`)

`bedToBigBed in.bed chrom.sizes out.bb`

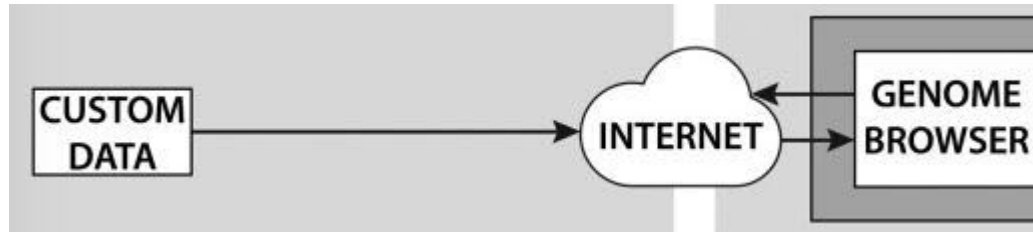
`chr1 33050000 33060000`
`chr1 33050000 33060000`
...
`chr21 33030000 33040000`
`chr21 33050000 33060000`
...
`chr22 33030000 33040000`
`chr22 33050000 33060000`
...

`chr1 248956422`
...
`chr21 46709983`
`chr22 50818468`
...

The resulting binary file **out.bb** (bigBed) can be hosted online:

[bigDataUrl=http://path.lab.edu/to/out.bb](http://path.lab.edu/to/out.bb)

The serving of files requires byte-range requests, which allows only portions of file to transfer. Some locations, like Dropbox, prevent such requests as people can then watch videos from their servers.



bigDataUrl <http://path.lab.edu/to/out.bb>

The **trackDb.txt** shares descriptions and where to find binary indexed tracks

View Data Tracks at UCSC

```

track ex1bb
type bigBed
shortLabel ex1
longLabel Blue example BED3 track
color 0,0,255
bigDataUrl http://path.lab.edu/to/out.bb

```

```

track track2
bigDataUrl http://path.lab.edu/to/out2.bb

```

...

A wide variety of track binary indexed formats exist:

- *BAM, CRAM, VCF*
- *bigGenePred, bigBarChart*
- *bigPsl, bigChain, bigMaf,*
- *bigNarrowPeak, halSnake*
- *bigBed, bigWig*

Additional Track Types In Hubs Example

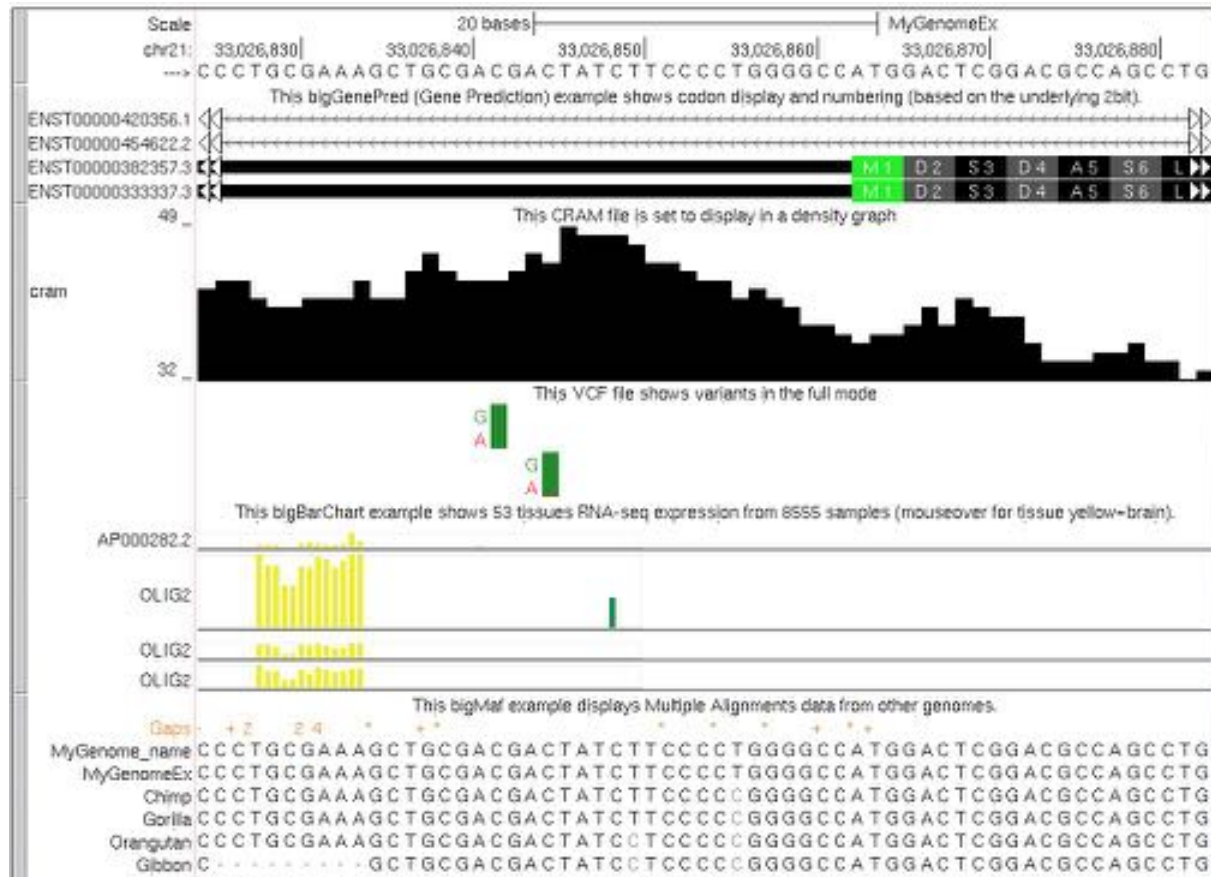
bigGenePred

CRAM

VCF

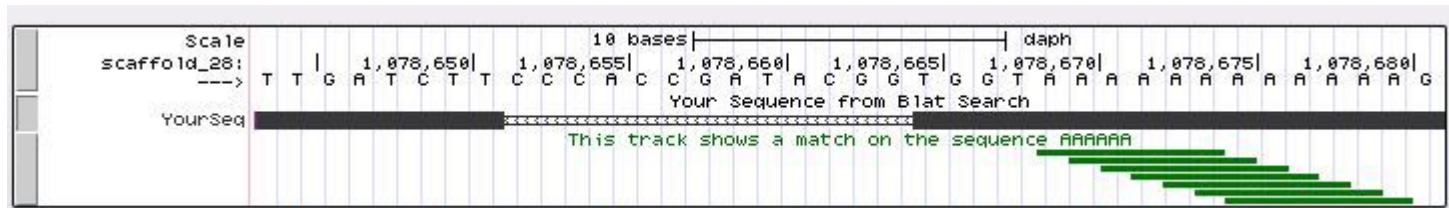
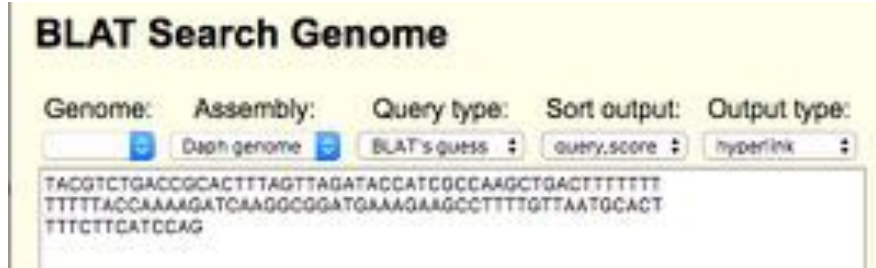
bigBarChart

bigMaf



Additional Tools Available on Assembly Hubs

- BLAT search sequence or amino-acid matched



Assembly Track Hub Summary:

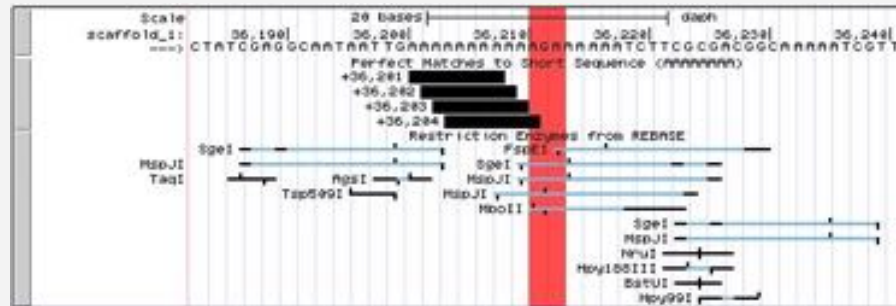
Remotely Hosted Data

Binary indexed files: **twoBitPath**,
bigDataUrl

```
>scaffold_1
GTTGTAATACTCTATTCTACAATAAAACCAA
TCATAGGTTGAATTGGCGTTGAAGTAAAAACAAA
...
>scaffold_2
AGTTATGACAAACTATAAAAAAGTCGGTAGAGACAAAAG
TCGTTTCGTGGACGAAGCGACCAAAACTGAGCACAAGAT
...
>scaffold_3
CATAAATTCATAAATCAATTCATGAAGAATAATT
TAGAAAATTTCCAGGAAGTTTGAAGTTGCTA
...
```

```
hubDirectory
|__hub.txt
|__genomes.txt
|__hg19
|__trackDb.txt
```

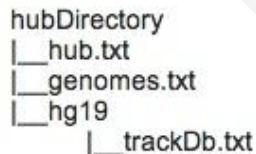
Visualizable at UCSC



Examples of Hubs and Resources

Genbank Assembly Hubs

<http://genome-test.soe.ucsc.edu/gbdb/hubs/genbank/>



```
hubDirectory
|__hub.txt
|__genomes.txt
|__hg19
|__trackDb.txt
```

ENCODE Project Dynamically Generated Hubs

<https://www.encodeproject.org/matrix/?type=Experiment>

Ensembl Track Hub Registry

<https://trackhubregistry.org/>

UCSC Public Hubs

<http://genome.ucsc.edu/cgi-bin/hgHubConnect>

CyVerse Discovery Environment byte-range supported data hosting

<https://de.cyverse.org/de/>

Existing Working Assembly Hub Templates

Genbank Assembly Hubs

<http://genome-test.soe.ucsc.edu/gbdb/hubs/genbank/>

Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

Genbank assembly hubs

Collection of hubs for Genbank genome assemblies.
These assemblies use the **UCSC naming patterns** on chromosomes.

Use these links to go to the index for that subset of assemblies:

| species subset ▲▼ | number of species ▲▼ | number of assemblies ▲▼ | total contig count ▲▼ | total nucleotide count ▲▼ | average contig size ▲▼ | average assembly size ▲▼ |
|--|----------------------|-------------------------|-----------------------|---------------------------|------------------------|--------------------------|
| other/synthetic assemblies | 3 | 3 | 94 | 2,835,004 | 30,159 | 945,001 |
| vertebrate other | 156 | 172 | 18,548,615 | 193,684,015,605 | 10,441 | 1,126,069,858 |
| vertebrate mammalian | 118 | 204 | 30,643,657 | 498,264,459,566 | 16,259 | 2,442,472,841 |
| plant | 190 | 269 | 34,577,423 | 145,341,422,954 | 4,203 | 540,302,687 |
| protozoa | 282 | 338 | 3,939,128 | 16,816,724,183 | 4,269 | 49,753,621 |
| invertebrate | 392 | 492 | 32,264,511 | 170,439,035,382 | 5,282 | 346,420,803 |
| fungi | 1,106 | 1,215 | 4,143,097 | 38,677,096,556 | 9,335 | 31,833,001 |
| archaea | 688 | 742 | 57,569 | 2,010,246,046 | 34,918 | 2,709,226 |
| bacteria | 34,009 | 58,661 | 8,397,216 | 234,147,691,500 | 27,883 | 3,991,539 |
| totals: | 36,944 | 62,096 | 132,571,310 | 1,299,383,526,796 | 9,801 | 20,925,398 |

Collection of hubs for Genbank genome assemblies.
These assemblies use the **NCBI accession naming patterns** on chromosomes.

Use these links to go to the index for that subset of assemblies:

| species subset ▲▼ | number of species ▲▼ | number of assemblies ▲▼ | total contig count ▲▼ | total nucleotide count ▲▼ | average contig size ▲▼ | average assembly size ▲▼ |
|--|----------------------|-------------------------|-----------------------|---------------------------|------------------------|--------------------------|
| other/synthetic assemblies | 3 | 3 | 94 | 2,835,004 | 30,159 | 945,001 |
| vertebrate other | 156 | 172 | 18,548,615 | 193,684,015,605 | 10,441 | 1,126,069,858 |
| vertebrate mammalian | 118 | 204 | 30,643,657 | 498,264,459,566 | 16,259 | 2,442,472,841 |

Existing Working Assembly Hub Templates

Genbank Assembly Hubs

<http://genome-test.soe.ucsc.edu/gbdb/hubs/genbank/>

Scroll Down to see assemblies... Click common name to load hub.

The "Taxon ID" link will go to the Entrez taxonomy for that ID.
The "common name" link will go to the UCSC genome browser for that assembly.
The "biosample" link will go to the Entrez biosample for that ID.
The "accession" link will go to the Entrez assembly for that ID.
The "assembly" link will go to the NCBI Genbank FTP source directory.

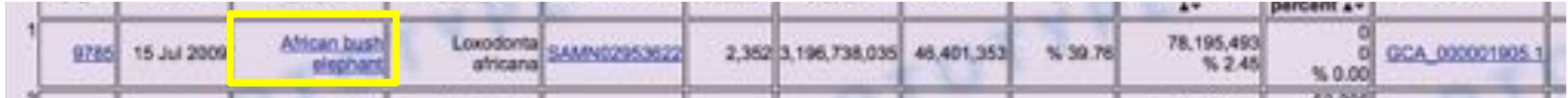
| Taxon ID ▲▼ | date ▲▼ | common name ▼ | scientific name ▲▼ | biosample ▲▼ | contig count ▲▼ | genome size ▲▼ | N50 size ▲▼ | GC percent ▲▼ | unknown bases count/percent ▲▼ | gene count bases percent ▲▼ | accession ▲▼ | |
|-------------|------------------------|---------------|--|--|------------------------------|----------------|---------------|---------------|--------------------------------|-----------------------------|----------------------------------|---------------------------------|
| 1 | 9785 | 15 Jul 2009 | African bush elephant | Loxodonta africana | SAMN02953522 | 2,352 | 3,196,738,035 | 46,401,353 | % 39.78 | 78,195,493 % 2.45 | 0 % 0.00 | GCA_000001905.1 |
| 2 | 9646 | 15 Dec 2009 | Alluropoda melanoleuca | Alluropoda melanoleuca | SAMN00008160 | 81,466 | 2,299,492,210 | 1,281,781 | % 40.62 | 54,198,184 % 2.36 | 52,206 473,410,926 % 20.59 | GCA_000004335.1 |
| 3 | 37293 | 16 Mar 2015 | Aotus nancymae | Aotus nancymae | SAMN03121886 | 29,222 | 2,926,565,220 | 8,280,397 | % 35.21 | 391,667,293 % 13.36 | 0 % 0.00 | GCA_000052055.1 |
| 4 | 9555 | 05 Jun 2012 | baboon | Papio anubis | SAMN02961400 | 63,250 | 2,948,380,710 | 139,646,187 | % 40.20 | 55,130,419 % 1.87 | 0 % 0.00 | GCA_000264685.1 |
| 5 | 118797 | 31 Jul 2013 | beji | Lipotes vexillifer | SAMN02953859 | 30,712 | 2,429,195,737 | 2,419,148 | % 40.75 | 32,195,939 % 1.33 | 0 % 0.00 | GCA_000442215.1 |

Existing Working Assembly Hub Templates

Genbank Assembly Hubs

<http://genome-test.soe.ucsc.edu/gbdb/hubs/genbank/>

1. Right-click and copy and paste a link to explore the hub's architecture.



| | | | | | | | | | percent | |
|---|----------------------|-------------|-----------------------|---------------------------|------------------------------|-------------------|------------|---------|----------------------|---|
| 1 | 9785 | 15 Jul 2009 | African bush elephant | <i>Loxodonta africana</i> | SAMN02953622 | 2,352,319,738,035 | 46,401,353 | % 39.76 | 78,195,493 % 2.45 | 0 0 GCA_000001905.1 |

2. Visit just the `hub.ncbi.txt` and then the `genomes.ncbi.txt` and then look at a specific `GCA...trackDb.ncbi.txt`

Existing Working Assembly Hub Templates

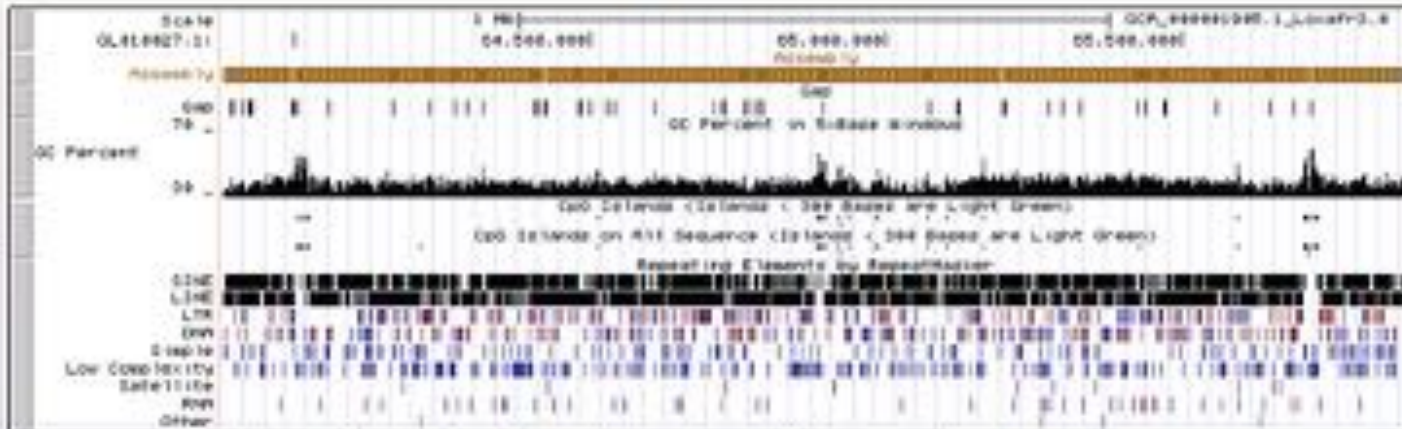
Genbank Assembly Hubs

<http://genome-test.soe.ucsc.edu/gbdb/hubs/genbank/>

Genome Browser on African bush elephant 15 Jul 2009/Loxafr3.0 Assembly

move <<< << < > >> >>> zoom in 1.5x 2x 10x base zoom out 1.5x 2x 10x 100x

GL010027.1:63,879,670-65,879,670 2,000,001 bp.



Example of Dynamically Built Hubs

ENCODE Project Hubs

<https://www.encodeproject.org/matrix/?type=Experiment>

BIO SAMPLE

14806 results

ASSAY

ChIP-seq
DNase-seq
PolyA RNA-seq
shRNA RNA-seq
total RNA-seq
eCLIP
RNA microarray
DNAme array
WGBS
RRBS
small RNA-seq
microRNA-seq
ATAC-seq
RAMPAGE
RNA Bind-n-Seq
genotyping array
CAGE
single cell RNA-seq
microRNA counts
sRNA RNA-seq
more

Immortalized cell line

| | | | | | | | | | | | | | | | | | |
|---------|-----|----|----|-----|----|-----|----|---|---|---|---|---|---|---|---|----|----|
| K562 | 627 | 7 | 19 | 276 | 12 | 190 | 12 | 3 | 1 | 1 | 8 | 1 | 1 | 2 | 9 | 1 | 50 |
| HepG2 | 334 | 3 | 11 | 257 | 6 | 161 | 7 | 3 | 2 | 2 | 3 | | | 2 | 8 | 1 | |
| A549 | 360 | 14 | 27 | | | | 2 | 2 | 1 | 1 | 9 | | | 2 | 3 | | |
| GM12878 | 242 | 2 | 11 | 6 | 6 | 3 | 2 | 2 | 6 | 1 | 1 | | | 2 | 8 | 13 | 1 |
| HEK293 | 250 | | | | | | 1 | 2 | | 2 | | | | 2 | | | |

+ See 179 more...

Tissue

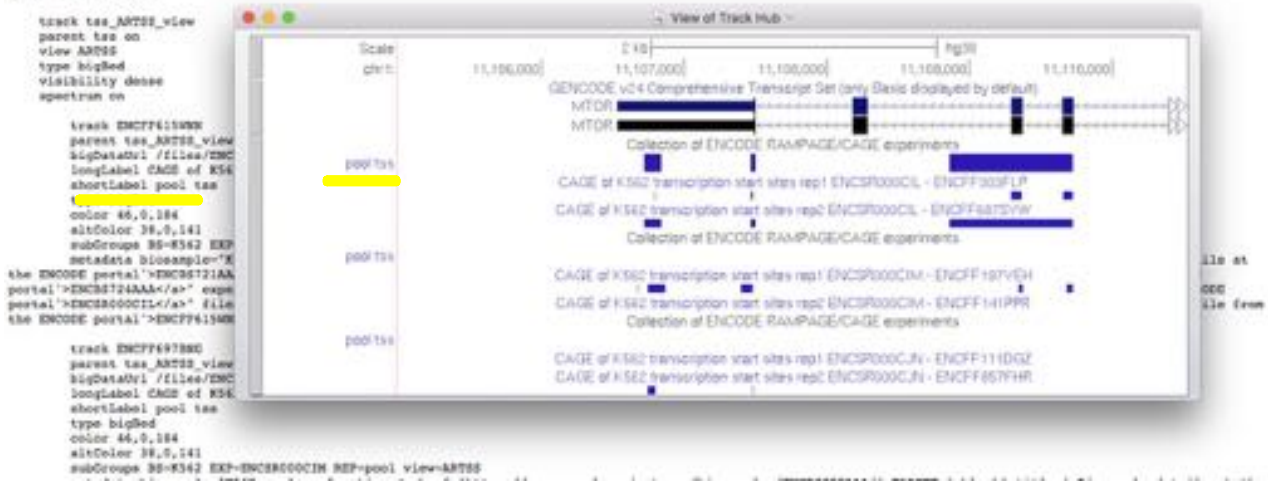
| | | | | | | | | | | | | | | | | |
|---------|-----|----|----|---|----|---|----|---|---|---|---|---|--|---|---|---|
| liver | 158 | 9 | 20 | 5 | | 1 | 10 | 1 | 1 | 7 | 7 | 2 | | 3 | 2 | 7 |
| heart | 100 | 23 | 13 | 6 | 10 | | 9 | 1 | 1 | 9 | 7 | 2 | | | 1 | 8 |
| stomach | 97 | 22 | 13 | 7 | | 3 | 8 | 1 | 4 | 4 | 6 | 5 | | 1 | | 4 |
| lung | 80 | 17 | 10 | 3 | 10 | 2 | 7 | 3 | 1 | 4 | 4 | 1 | | | | 4 |
| kidney | 69 | 18 | 11 | 2 | 2 | 2 | 4 | 4 | | 4 | 4 | | | | | 4 |

Example of Dynamically Built Hubs

ENCODE Project Hubs

<https://www.encodeproject.org/matrix/?type=Experiment>

```
track tsx
compositeTrack on
type bed 3
longLabel Collection of ENCODE RAMPAGE/CAGE experiments
shortLabel ENCODE RAMPAGE/CAGE
visibility full
permanentIcon https://www.encodeproject.org/static/img/persistent-encode.png https://www.encodeproject.org/ "This trackhub was automatically generated from the ENCODE files and metadata found at the ENCODE portal"
html TSS
subGroup1 view Views ARTS5-Replicated_TSSs ARTS5-TSSs SIGSL-signal_of_unique_reads SIGSM-signal_of_all_reads SIGLF-Five_signal_of_unique_reads SIGLS-Missed_signal_of_unique_reads SIGMT-Five_signal_of_all_reads SIGMS-Missed_signal_of_all_reads
subGroup2 BS Biosample BS42-K542
subGroup3 EXP Experiment ENCER000CL-ENCER000CL ENCER000CL-ENCER000CL ENCER000CL-ENCER000CL ENCER000CL-ENCER000CL ENCER000CL-ENCER000CL
subGroup4 REP Replicate pool-Pooled rep01-Replicate_1 rep02-Replicate_2
sortOrder BS-- EXP-- view-- REP--
dimension 21A-REP
```



Track Hub Resources

Ensembl Track Hub Registry

<https://trackhubregistry.org/>

The Track Hub Registry

A global centralised collection of publicly accessible track hubs

The goal of the Track Hub Registry is to allow third parties to advertise [track hubs](#), and to make it easier for researchers around the world to discover and use track hubs containing different types of genomic research data.



Track Hub Resources

UCSC Public Hubs *moderated collection of track hubs*

<http://genome.ucsc.edu/cgi-bin/hgHubConnect>

The screenshot shows the UCSC Public Hubs search interface. At the top, there are tabs for 'Public Hubs' and 'My Hubs'. Below the tabs, there is a search bar with the text 'mouse epigenomics' entered. A 'Search Public Hubs' button is next to the search bar. Below the search bar, there is a filter section with the text 'Filter hubs by assembly:' and a 'Search Public Hubs' button. The main content area displays a list of search results. The first result is for 'DNA Methylation' with a description 'Hundreds of analyzed methylomes from bisulfite sequencing data' and assemblies '[+] hg38, hg19, hg18, mm9, mm10, panTro2...'. The second result is for 'ENCODE Analysis Hub' with a description 'ENCODE Integrative Analysis Data Hub' and assembly 'hg19'. Each result has a 'Connect' button and a 'Search details ...' link.

Public Hubs My Hubs

Enter search terms to find in public track hub description pages:
mouse epigenomics

Filter hubs by assembly: Search Public Hubs

Displayed list **restricted by search terms:** mouse epigenomics Show All Hubs

When exploring the detailed search results for a hub, you may right-click on an assembly or track line to open it in a new window.
Clicking Connect redirects to the gateway page of the selected hub's default assembly.

| Display | Hub Name | Description | Assemblies |
|--|---------------------|--|---|
| <input type="button" value="Connect"/> | DNA Methylation | Hundreds of analyzed methylomes from bisulfite sequencing data | [+] hg38, hg19, hg18, mm9, mm10, panTro2... |
| Search details ... | | | |
| Hub Description: ... This session shows a collection of tracks from multiple blood cells. Mouse: Reprogramming: This session shows tracks from vari... | | | |
| 4 Matching Assemblies | | | |
| <input type="button" value="Connect"/> | ENCODE Analysis Hub | ENCODE Integrative Analysis Data Hub | hg19 |
| Search details ... | | | |
| Hub Description: ... d TALI genes. We tested a subset of these predicted enhancers in both Mouse and Fish transgenic models examples in Figure 6, | | | |

Track Hub Resources

CyVerse Discovery Environment Storage Solution

CyVerse `iCommands` (`rysnc` becomes `irsync`) allows transfer of 2-100GB files and recursive hub directory structures with commands like `irsync`:

```
$ irsync -r local_directory_name i:data_store_directory_name
```

hubDirectory

```
|  
|_hub.txt  
|_genomes.txt  
|_genomeDirectory1  
|_trackDb.txt
```

Move all your local files (BAMs, ect) with the single `irsync` command.

Then make the files publicly accessible with the `ichmod` command to view.

<https://de.cyverse.org/anon-files/>

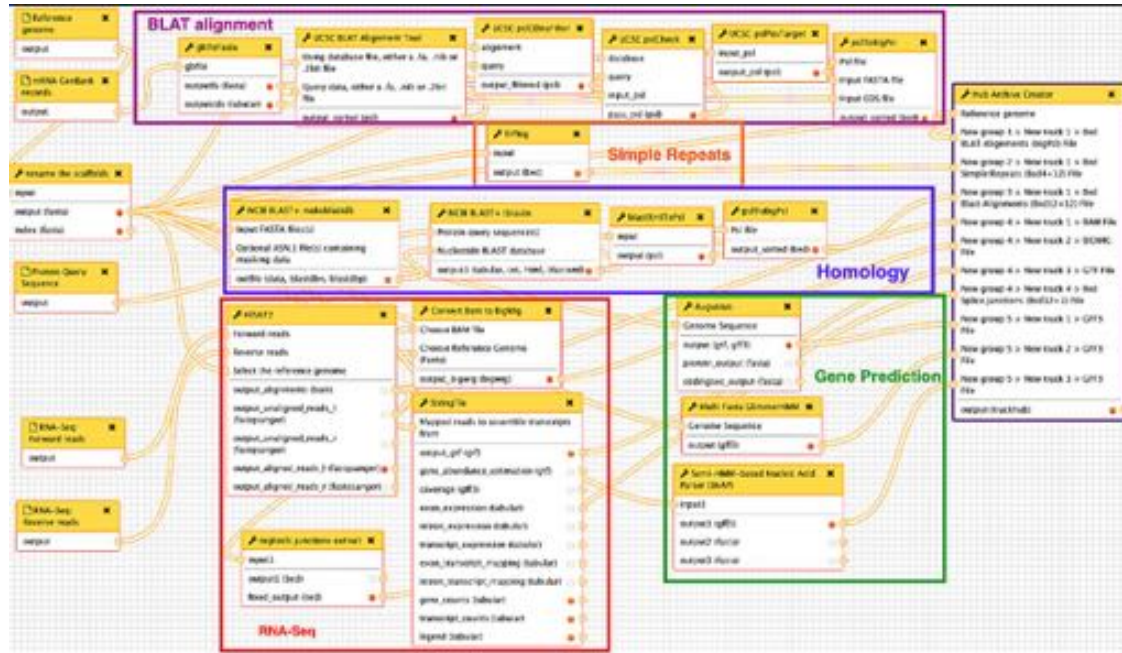
```
$ ichmod read anonymous data_store_directory_name
```

Track Hub Resources

G-OnRamp

Galaxy workflow turning data like RNA-Seq into Assembly Hubs

<http://gonramp.wustl.edu/>





Thank You!



UNIVERSITY OF CALIFORNIA
SANTA CRUZ | Genomics
Institute

