Biostars
BIOINFORMATICS EXPLAINED

🌐 Community   📶 Planet   ✏ New Post

🔓 Log In

Where Can I Download All Exons Of The Human Genome In Fasta Format (One Big File!) ?

👍

5

🔖

Hi to all,

Where can I download all exons of the human genome in FASTA format (one big file!)?

Thanks

12.2 years ago
mariusorion ▲ 70

`fasta` `human-genome` `exon` • 15k views

ADD COMMENT • link

updated 2.2 years ago by Ram ⚡ 45k • written 12.2 years ago by mariusorion ▲ 70

👍

0

🔖

Thank you very much to all. Thank you Malachi, this is what I am looking for.

ADD REPLY • link

12.2 years ago by mariusorion ▲ 70

👍

0

🔖

Not an answer... This should go as a comment. Thanks!

ADD REPLY • link

12.2 years ago by Josh Herr ⚡ 5.8k

👍

0

🔖

Here - ftp://ftp.ebi.ac.uk/pub/databases/astd/current_release/human/

ADD REPLY • link

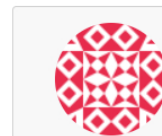updated 5.5 years ago by Ram ⚡ 45k • written 9.4 years ago by hamzakhanvit ▲ 10
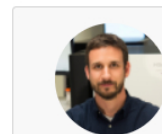

👍

9

🔖

Using Ensembl BioMART

1. Go to BioMART website.
2. Choose database: Ensembl Genes 70
3. Choose dataset: Homo sapiens
4. Click 'Attributes' then select the 'Sequences' option
5. Expand the sequences pane and select the 'Exon sequences' option
6. Expand the 'Header information' pane. Select the info you want to associate with each sequence record (e.g. Ensembl Gene ID, Ensembl Transcript ID, Ensembl Exon ID).
7. Click the 'Results' button.
8. Make sure the example output looks good, select the 'Compressed File' option and hit the 'Go' button.

The download took a while but eventually I got a file 'mart_export.txt.gz'. This file has output that looks like this

12.2 years ago

Malachi Griffith ⚡ 20k

Similar Posts

How big is the human exome? •
6.4 years ago by guido.leoni ▲ 10

Dear All According to your experience how big is the size of the human exome in megabases? Mining the web there are tons of sites tha…

Human genome full chromosomes build •
updated 2.8 years ago by Ram ⚡ 45k • written 9.6 years ago by priess1991 • 0

Hi! I need to use some data of the 1000 human genome project. On 1000genomes.org's ftp-server I found all data in form of 14gb big cram fi…

GTF/Reference genome adjustments? Gene-based differential expression analysis of genetically modified mouse line •
updated 20 months ago by Ram ⚡ 45k • written 20 months ago by alex • 0

Hello, I am trying to analyze my bulk RNAseq data set from hippocampal tissue extracted from our WT/KO mice. The knockout consists of a 10…

calculating the exome size from latest GTF file •
updated 3.7 years ago by Ram ⚡ 45k • written 11.1 years ago by Mingkun ▲ 40

I have downloaded the latest GTF file (Homo_sapiens.GRCh37.75.gtf) from Ensembl › I was trying to find all protein coding genes and calculte …

Extract mitogenome from seqences •
5.2 years ago by harmadikemil • 0

Dear All! I have a big database, full of different sequences, most of it in fastq.gz files. (archaic

```
>ENSG00000264715|ENST00000578347|ENSE00002715610
AGGAGTGACCAGAAGACAAGAGTGCGAGCCTTCTGTTATGCCCAGACAGGGCCACCAGAGGGCTCCTTGGTCTAGTGGTAACGCCA
>ENSG00000265161|ENST00000580394|ENSE00002716056
AGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAGGTGATCCATCCACC
>ENSG00000200917|ENST00000364047|ENSE00001438810
GCACATACATATACTAAAATTGGAACAATACAGAGAAGATTAGCATAGCCCCTGCGCAAGGATGACATGCAAATTCGTGAAGTGTTCCATATTAAA
. . .
```

ADD COMMENT • link   updated 5.5 years ago by Ram ⚡ 45k • written 12.2 years ago by Malachi Griffith ⚡ 20k

👍 Thank you Malachi, this is what I am looking for :)

0 ADD REPLY • link   12.2 years ago by mariusorion ▲ 70

👍 get the coordinates:

3 Exon coordinates of hg19 genome download

and fetch the sequences:

How to get the sequence of a genomic region from UCSC?

Batch fetching fasta sequences from bed file

ADD COMMENT • link   12.2 years ago by Pierre Lindenbaum ⚡ 166k

12.2 years ago
Pierre Lindenbaum ⚡
166k
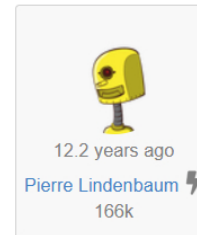
👍 Have a look at the UCSC Table Browser: position base query

2

12.2 years ago
aindap ▲ 120

ADD COMMENT • link   12.2 years ago by aindap ▲ 120

👍 The page does not work

0 ADD REPLY • link   12.2 years ago by mariusorion ▲ 70
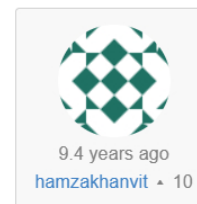

👍 Here - ftp://ftp.ebi.ac.uk/pub/databases/astd/current_release/human/

0

9.4 years ago
hamzakhanvit ▲ 10

ADD COMMENT • link   updated 5.5 years ago by Ram ⚡ 45k • written 9.4 years ago by hamzakhanvit ▲ 10

If anyone is looking for the same but for GRCh37 here is the link.

0

ADD COMMENT • link

8.8 years ago by Alejandro Jimenez Sanchez ▲ 180

→ Login before adding your answer.

know other people opinion about this, because I am mainly used to do RNA-Seq analysis, and I would apprec…

CDS FASTA file as Reference sequence for Ion Torrent •
updated 2.2 years ago by Ram
⚡ 45k • written 7.7 years ago by serenabivona ▲ 10

Hi! I'm sorry! My question will be stupid for somone else, but it is my first approach with ngs analysis... I will use PGM and Torrent Su…

How to align the sanger sequenced fragments against human reference gene? •
updated 7.0 years ago by WouterDeCoster ⚡ 48k • written 7.0 years ago by bioinforesearchquestions ▲ 370

Hi folks, I have been provided sequence file from Sanger sequencing platform. It has 12 exons, so the technician has captured 5'UTR+exon…

Extraction of first sequences from a big fasta file •
updated 3.1 years ago by Ram
⚡ 45k • written 10.3 years ago by vahapel ▲ 210

Dear All, I would like to ask a question regarding extraction of 100000 sequences in a big fasta file. In the forum, there is a bunch of s…

prepare a GFF file for MOCK fasta reference •
6.9 years ago by Sam ▲ 150

Dear Biostars I need your help to handle this problem, I'm working with GBS data and now I have a big list of SNP position in my output …

Identifying Exons and number of reads between exons from a BAM file •
updated 9.0 years ago by Alex Reynolds ⚡ 36k • written 9.0 years ago by startup_biostar ▲ 20

I have a big BAM file from which I need to identify the number of reads between Exon-1 and Exon-3 exon-21 and exon-24. What is the best too…

Extracting sequence and annotation of a particular gene •
updated 6.2 years ago by

ATpoint ⚡ 88k • written 6.2 years ago by Adrian Pelin ★ 2.7k

Hello, I have the human and mouse GENCODE annotations and the genome that goes with it. I am interested to create sequence and annotation f…

How To Reduce Fasta Output? • 13.2 years ago by Fabian ▲ 50

I'm using fasta's ggsearch and I'm wondering if I can get rid of all the verbose output because with the all vs. all aligments I'm doing, t…

Extract a BAM exome from the GTF of human genome • 2.8 years ago by Juliana • 0

Hi all! I am struggling for weeks with this problem and I hope someone can help. I would like to compare the target regions of human e…

Extracting exon neighbourhoods from the human reference genome • 3.6 years ago by Sergio Martínez Cuesta ▲ 230

Hi all, Starting from the GRCh38/hg38 human reference genome, I would like to scan through all the exons with an aim to extract all the ne…

Binning Exon And Plotting Read Densities (Bed Regions) With In And Around Binned Exons • updated 4.4 years ago by Biostar ⚡ 20 • written 13.8 years ago by Repineme ▲ 120

I want to map nucleosome (BED regions) around all hg18 exons. As you know human exons are of different sizes and they need to scaled. Is th…

NCBI vs Ensembl - which one to chose - for downloading protein fasta files • updated 8.0 years ago by Jean-Karim Heriche ⚡ 27k • written 8.0 years ago by Idit • 0

Hi [I'm a newbie in bioinformatics, my apologies for misusage of terms, if any..] I need to decide which resource to use, to download…

creating virtual library preparation of all possible exons in a gene in human • updated 4.3 years ago by Pierre

Lindenbaum ⚡ 166k • written
4.3 years ago by harry ▲ 40

Hi everyone, I want to make a virtual library fasta file for all the possible exons in a gene and also all possible exon-intron fasta file…

Strategy for generating a consensus sequence for 100 complete bacterial genomes? •
6.9 years ago by Alec Watanabe ▲ 60

Greetings, I am working with a 100 complete *M. tuberculosis* genomes in FASTA format. What I want to do is align all the sequences to s…

Content

Search

Users

Tags

Badges

Help

About

FAQ

Access

RSS

API

Stats

Use of this site constitutes acceptance of our User Agreement and Privacy Policy.