

Exercise

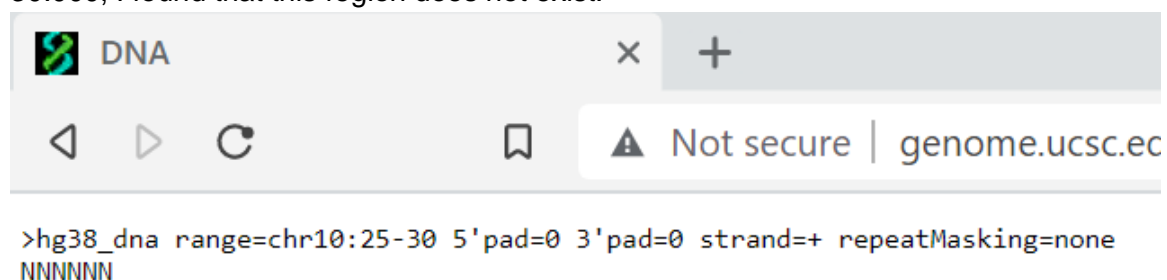
Using the Perl API on the latest human data for Ensembl release 104, convert coordinates on chromosome 10, from 25.000 to 30.000 to the same region in GRCh37.

SM Answer Exercise

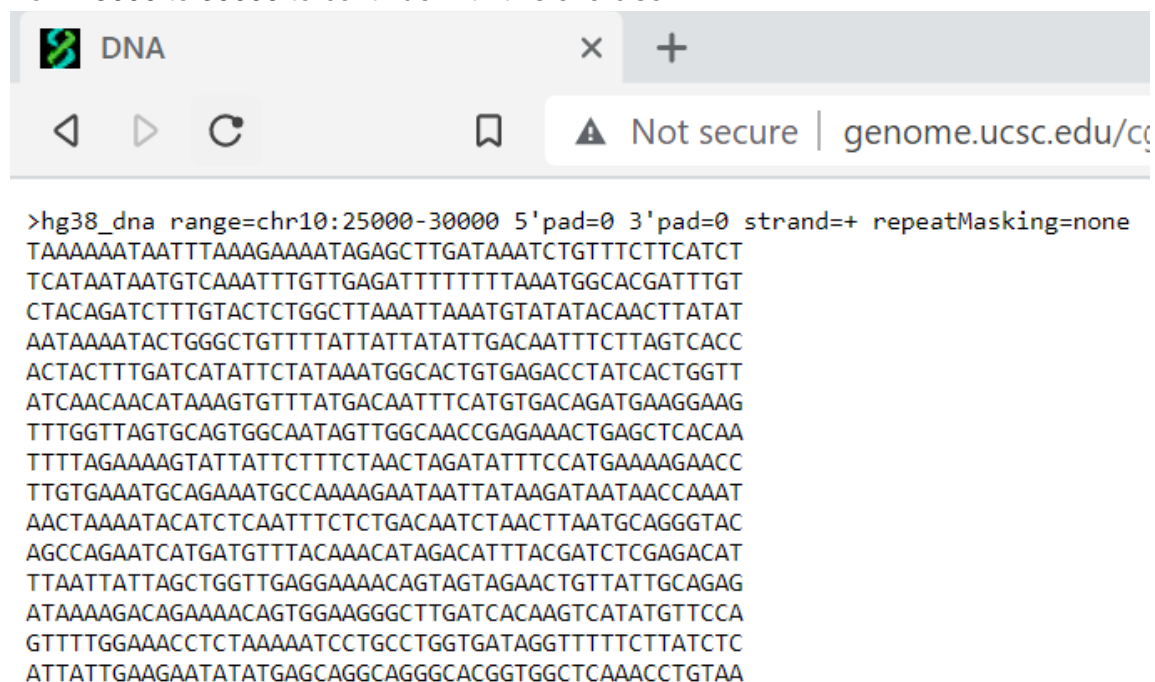
Git Link: https://github.com/smukher2/EMBL_git_ensembl_perl_API_Ex1.git

Point 1 of 4: Changed Input or Query Coordinates As Original One Provided In The Question Has No Sequence “No Data” or Gap region:

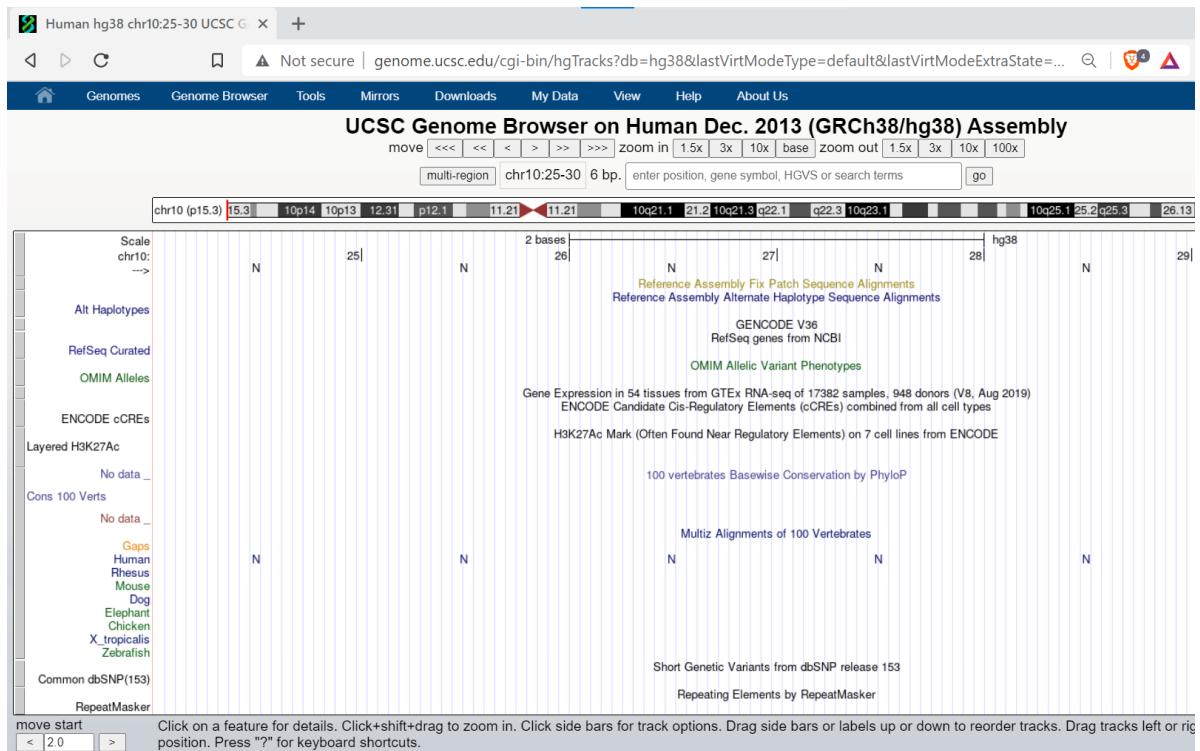
Upon checking UCSC genome browser for region GRCh38 chromosome 10, from 25.000 to 30.000, I found that this region does not exist.



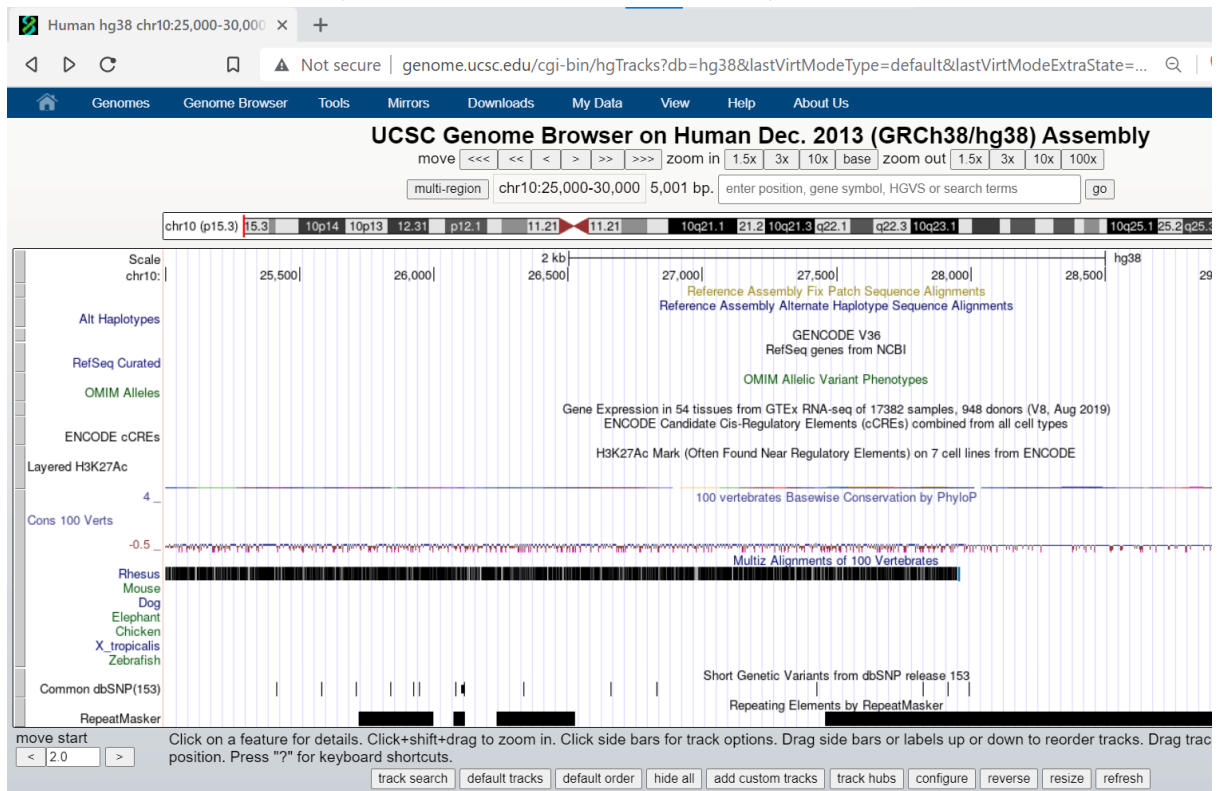
Therefore, for conversion to GRCh37, I used most likely region GRCh38 chromosome 10, from 25000 to 30000 to continue with this exercise.



Please see below screenshots of these genomic regions.
GRCh38 chromosome 10, from 25.000 to 30.000 does not exist.



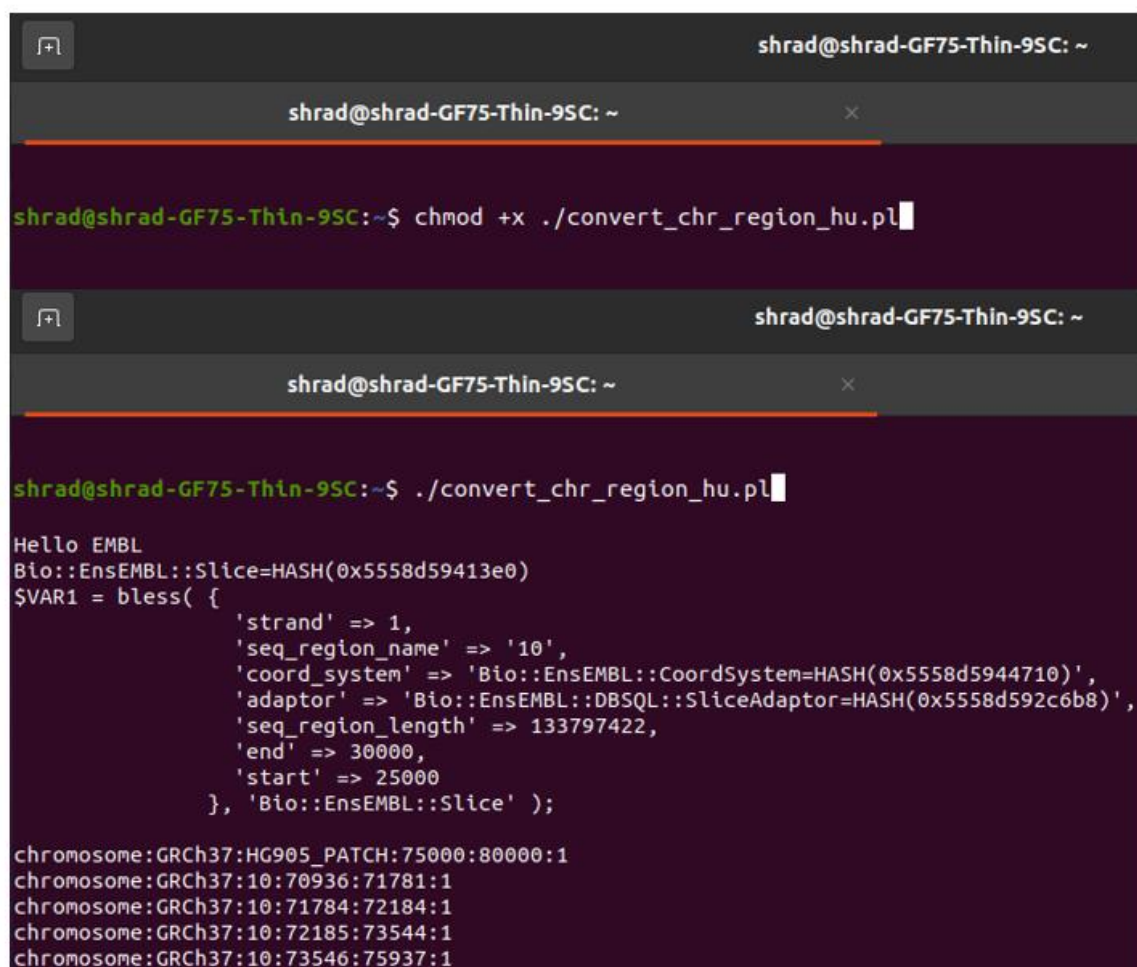
GRCh38 chromosome 10, from 25000 to 30000 does exist, and was used for this exercise.



Point 2 of 4: Conversion of Coordinates form GRCh38 to GRCh37: I used Ensemble Perl API to convert GRCh38 human genome region chr10: 25000-30000 to corresponding GRCh37. I used 'Atom' Editor to write the 'convert_chr_region_hu.pl' script and executed the script from Ubuntu Terminal using command (Ref: file permission <https://cets.seas.upenn.edu/answers/chmod.html> and running perl script <https://www.geeksforgeeks.org/hello-world-program-in-perl/>). I found helpful tutorials at <https://m.ensembl.org/info/website/tutorials/grch37.html> and https://rest.ensembl.org/documentation/info/assembly_map.

which perl #this gives location of perl used in first line of script for me its #!/usr/bin/perl
 chmod +x ./convert_chr_region_hu.pl #gives file permission to make script executable for all users

./convert_chr_region_hu.pl #run the script



```
shrad@shrad-GF75-Thin-9SC: ~
shrad@shrad-GF75-Thin-9SC: ~
shrad@shrad-GF75-Thin-9SC:~$ chmod +x ./convert_chr_region_hu.pl

shrad@shrad-GF75-Thin-9SC:~$ ./convert_chr_region_hu.pl

Hello EMBL
Bio::Ensembl::Slice=HASH(0x5558d59413e0)
$VAR1 = bless( {
    'strand' => 1,
    'seq_region_name' => '10',
    'coord_system' => 'Bio::Ensembl::CoordSystem=HASH(0x5558d5944710)',
    'adaptor' => 'Bio::Ensembl::DBSQL::SliceAdaptor=HASH(0x5558d592c6b8)',
    'seq_region_length' => 133797422,
    'end' => 30000,
    'start' => 25000
  }, 'Bio::Ensembl::Slice' );

chromosome:GRCh37:HG905_PATCH:75000:80000:1
chromosome:GRCh37:10:70936:71781:1
chromosome:GRCh37:10:71784:72184:1
chromosome:GRCh37:10:72185:73544:1
chromosome:GRCh37:10:73546:75937:1
```

Point 3 of 4: Results From Exercise: The results for this exercise, for conversion to GRCh37 is as follows,

chromosome:GRCh37:HG905_PATCH:75000:80000:1
 chromosome:GRCh37:10:70936:71781:1
 chromosome:GRCh37:10:71783:72184:1
 chromosome:GRCh37:10:72184:73544:1
 chromosome:GRCh37:10:73545:75937:1

Point 4 of 4: Pre-requisites for this exercise: I started with the assumption that anyone trying to run this exercise has internet connection and a computer (Mac, Windows or Linux), with no background in coding. Therefore, I uploaded detailed description of all setup/installation steps and a detailed tutorial in github repository **Git Link:**

https://github.com/smukher2/EMBL_git_ensembl_perl_API_Ex1.git.

Installation instruction files are:

2.1_Instal_Ubuntu_dual_boot_any_computer.pdf

2.2_Instal_Ubuntu_different_software.pdf

2.3_Install_Atom_Its_More_Than_Just_Text_Editor.pdf

2.4_Install_Perl_PerlModules_BioPerl_EnsemblPerlAPI.pdf

Tutorial file is:

4_Tutorial_Complete_Guide_Ubuntu_To_EnsemblPerlAPI.pdf

end of SM Answer Exercise

Alternatives

Describe at least one other way of retrieving the same information, along with its advantages and disadvantages.

SM Answer Alternatives

Git Link: https://github.com/smukher2/EMBL_git_ensembl_perl_API_Ex1.git

Point 1 of 5: Alternative Tool: As an alternative we can use a 'LiftOver' Ensembl tool <http://uswest.ensembl.org/info/docs/tools/index.html> that is web-browser based and does not require installation http://uswest.ensembl.org/Homo_sapiens/Tools/AssemblyConverter.

Point 2 of 5: Advantages Of Alternative Tool: The Ensembl "LiftOver" tool has the following advantaged:

1. Web-based ready to use tool that does not require setup/installation.
2. Usage of tool does not require any coding experience.

Point 3 of 5: Disadvantage Of Alternative Tool: The Ensembl "LiftOver" tool has the following disadvantaged:

1. Web results are deleted after 10 days. So user needs to manually download results in 10 days. In the coding exercise results are automatically saved/downloaded to the computer.
2. Its tedious to use when user has multiple coordinate files to convert, such as from samples of different experimental and control groups, and replicates. For repetitive tasks coding exercise method is preferable.

Point 4 of 5: Usage Of Alternative Tool: The steps to do conversion of GRCh38 chromosome 10, from 25000 to 30000 to GRCh37.

1. Visit 'LiftOver' tool
http://uswest.ensembl.org/Homo_sapiens/Tools/AssemblyConverter

- Input coordinates GRCh38 in bed format and click 'Run' to execute conversion, as shown in screenshot <https://m.ensembl.org/info/website/upload/bed.html>.
chr10 25000 30000

- Click 'Download' to download GRCh37 converted results as shown in screenshot below.

Point 5 of 5: Results From Alternative Tool: The results for this exercise, for conversion to GRCh37 is as follows,

```

HG905_PATCH      75000 80000
chr10 70936 71781
chr10 71783 72184
chr10 72184 73544
chr10 73545 75937

```

end of SM Answer Alternatives