

Please follow the instructions below and provide answers to be emailed back to recruitment@ebi.ac.uk within ten days. Your email should include links to your source code on GitHub. You should provide adequate information on how to run your script along with any other information we should be aware of.

Background

We provide gene annotation for the human genome
The latest assembly available is GRCh38 but some users still use the previous assembly
GRCh37 and want to convert coordinates between the two assemblies
You can find more information about genome assemblies here
<https://www.genome.gov/12011238/an-overview-of-the-human-genome-project/>
For this exercise, we use the latest Ensembl release

Ensembl Perl API

The Ensembl Perl API is used to generate annotation as well as access it programmatically
You can find more information about it here
http://www.ensembl.org/info/docs/api/core/core_tutorial.html

Exercise

Using the latest human data from Ensembl release and the Perl API to convert coordinates on chromosome (e.g chromosome 10 from 25000 to 30000) to the same region in GRCh37. Enable the script to be as generic as possible to be run as a command-line program.

Alternatives

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