

Homework 1 of Genomics 2019

Dehe Wang

September 17 2019

Homework

1. What is the Human Genome Project (HGP)? Please briefly explain and describe the history of HGP. (15')
2. Please briefly describe the contributions of HGP to the development of Life Sciences. (15')
3. Please search for the uc002rpz.3 transcript in the UCSC genome browser (hg19) and respond to following instructions or questions.
 - Extract this transcript as BED12 format (5')
 - Number of isoforms (from UCSC Genes annotation) (5')
 - Chromosome (5')
 - Strand (5')
 - Exon number (5')
 - mRNA length (5')
 - TSS position (5')
 - CDS end site position (5')
 - CDS length (hint: not including introns) (5')
 - Extract CDS protein sequence (5')
 - Extract the sequence of the exon 12 (5'), highlight this block (5') and take a screenshot as pdf format (5')
 - Get all transcript IDs within 100,000 bp upstream of this transcript TSS (UCSC Genes annotation) (5')

Please answer the first two questions and save them in the "StudentID_Answer.txt" file. Please write the answers for question 3 in the given "StudentID_Answer.tsv". Please save the screenshot as "StudentID_ScreenShot.pdf" and package the 3 files as "StudentID_HW1.zip". You should answer in the format as in the given example. And you should submit your homework to wangdehe@whu.edu.cn before Sept. 29, 2019.

Grading

The score S of this homework is calculated by the following formula:

$$S = (S_0 - P) * 0.97^d$$

where S_0 is the total score from all questions, P is a penalty and is obtained from the following table, and d is the late time (in day) calculated by the e-mail event stamp.

Table 1: Penalty list

Contents	Score
File naming error	-5' / file
Packaging error	-5'

Reference

1. **BED12 format:** <http://genome.ucsc.edu/FAQ/FAQformat.html#format1>
2. **Table browser:** <http://genome.ucsc.edu/goldenPath/help/hgTablesHelp.html>