

Search the UCSC genome browser at <https://genome.ucsc.edu/> and answer the following questions for the human gene TP53, using the hg38 human genome.

1. How many transcripts are reported by each annotation database of the NCBI RefSeq genes (curated subset) and the GENCODE V32?
 - NCBI RefSeq genes (curated subset): 15
 - GENCODE V32: 19
2. How many transcripts agree between the two databases?
 - 4
3. Choose a transcript from the NCBI RefSeq gene annotation: indicate the transcript ID and answer the following question about this transcript: How many exons are UTR and how many exons are coding?
 - Transcript ID: NM_001276698.2
 - UTR Exons: 3 (7668401,7670608,7673206)
 - Coding Exons: 5 (7673534, 7673700, 7674180, 7674858, 7675052)

Download the human gene annotation GTF file “Homo_sapiens.GRCh38.102.chr.gtf.gz” from Ensembl at <http://useast.ensembl.org/info/data/ftp/index.html>, analyze the file, and answer the questions below:

1. Which version of the human genome is the GTF file from (hg38, hg19, etc.)?
 - hg38
2. Which annotation version is this file?
 - 102
3. How many genes and transcripts in total are in this GTF file?
 - 292573
4. How many genes and transcripts are protein-coding?
 - 177062