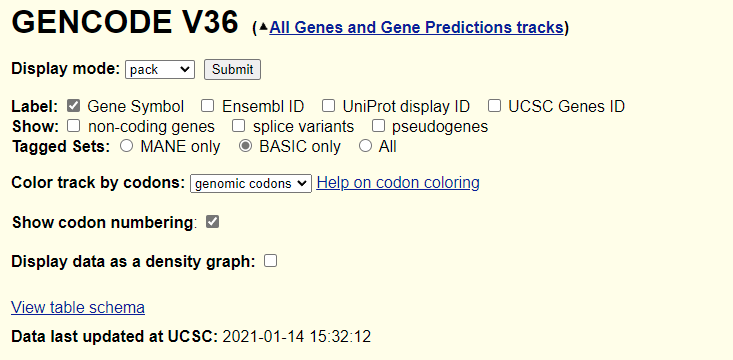
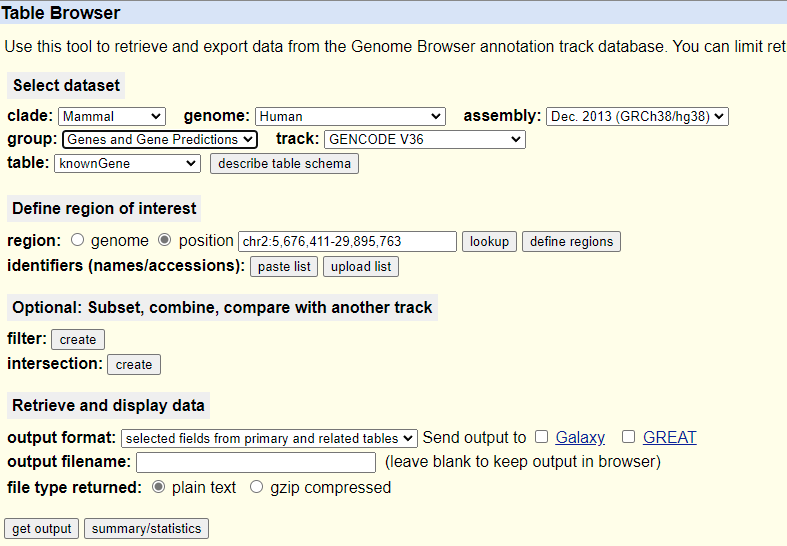
**2. Count the number of protein-coding genes on each chromosome.**

我选择以GENCODE V36为数据集，查询每个染色体上的蛋白质编码基因的数量。

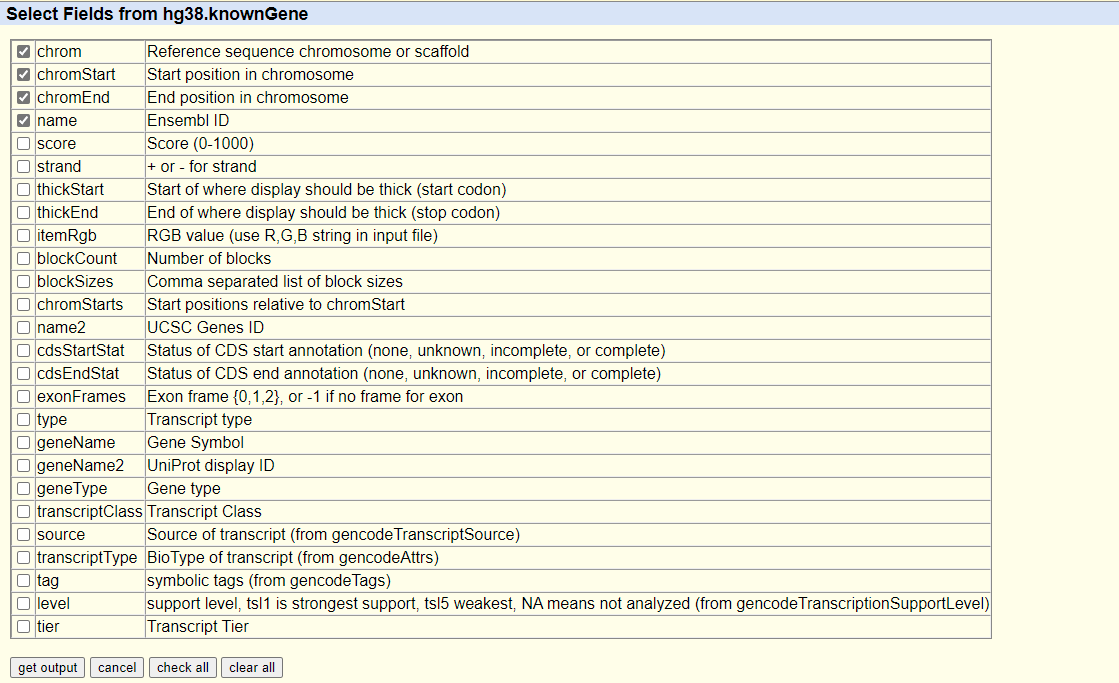
方法：点击进入如下链接：[Human hg38 chr2:5,676,411-29,895,763 UCSC Genome Browser v420](http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr2%3A5676411%2D29895763&hgsid=1174508739_uzGn5taxEZ6Htgrlq4e1v2JRnvLb)，输入某个染色体号（例如1号染色体就输入chr1），选择GENCODE V36，点击refresh，点击形成的图像框的左侧，在新弹出的页面中作如下选择：



接着点击Submit，在上访的Tools中点击选择Table Browser，在新弹出的页面中做出如下选择：



点击get output，在新弹出的页面中做出如下选择：

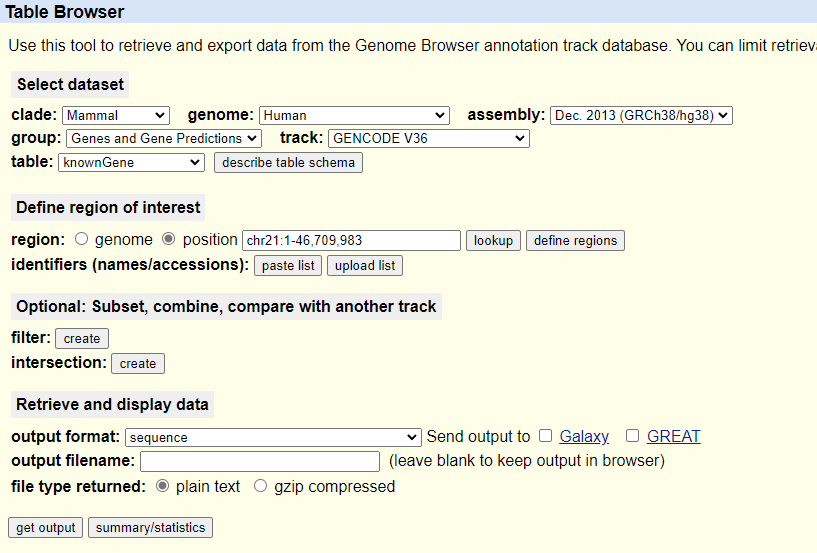


再次点击get output，就获得了1号染色体的所有蛋白质编码基因。以此类推，我们就可以获得每个染色体的蛋白质编码基因数量。

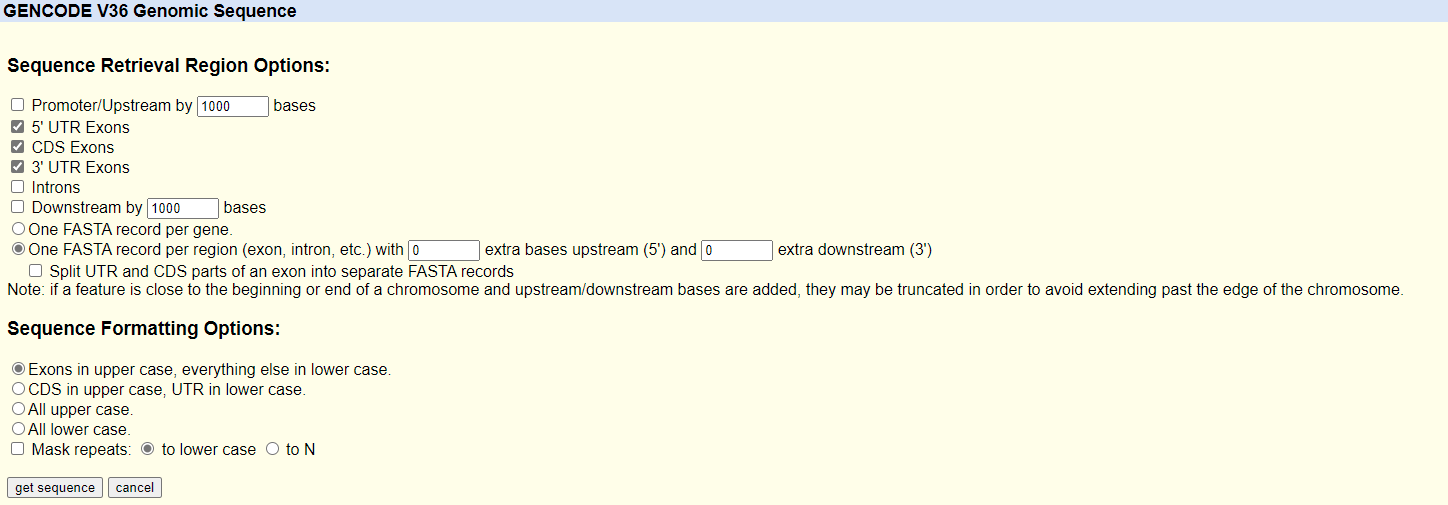
**3. Summarize length distributions of exons, introns and genes on each chromosome and in the whole genome using histograms, and also provide statistics like median, mean, max/min, etc..**

3.1计算外显子的分布

打开如下链接：[Table Browser (ucsc.edu)](http://genome.ucsc.edu/cgi-bin/hgTables)，在新弹出的页面中做出如下选择：



选定好对应的染色体区域，点击get output，在新弹出的页面中做出如下选择：



点击get sequence，就可以获得这个染色体上的所有外显子序列，调用我写的代码（第1次作业代码.ipynb）就可以获得分布。（需要将length\_distribution下的文件都解压后代码才可以正常运行！！！）

类似地，只需要修改一下选项，就可以获得内含子和基因的长度分布。