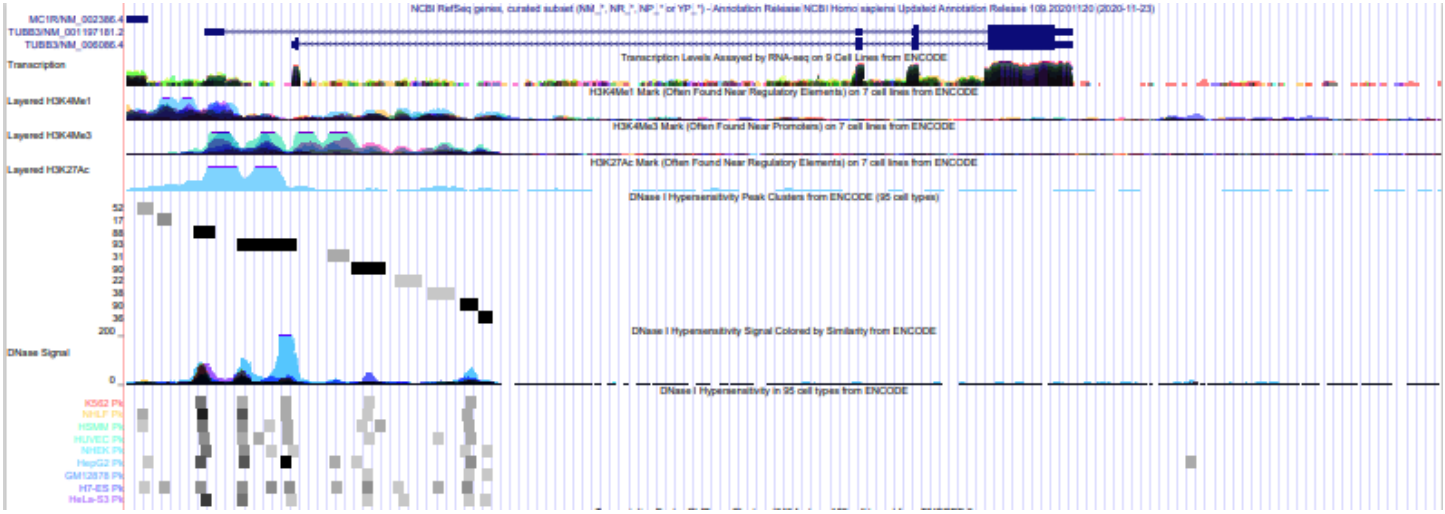
Find the human *TUBB3* gene using the UCSC Genome Browser (hg38). Turn on the Encode Regulation track (HINT: set display mode to “full” for these tracks) and NCBI RefSeq genes. In a few sentences, describe what you see at the *TUBB3* locus in terms of the Encode Regulation tracks. Include in your answer what histone modification(s) appear(s) near the transcription start site of the *TUBB3* gene. Submit a screenshot of this locus. (HINT: click View > PDF/EPS at the top of the browser page to export a PDF/EPS file.).

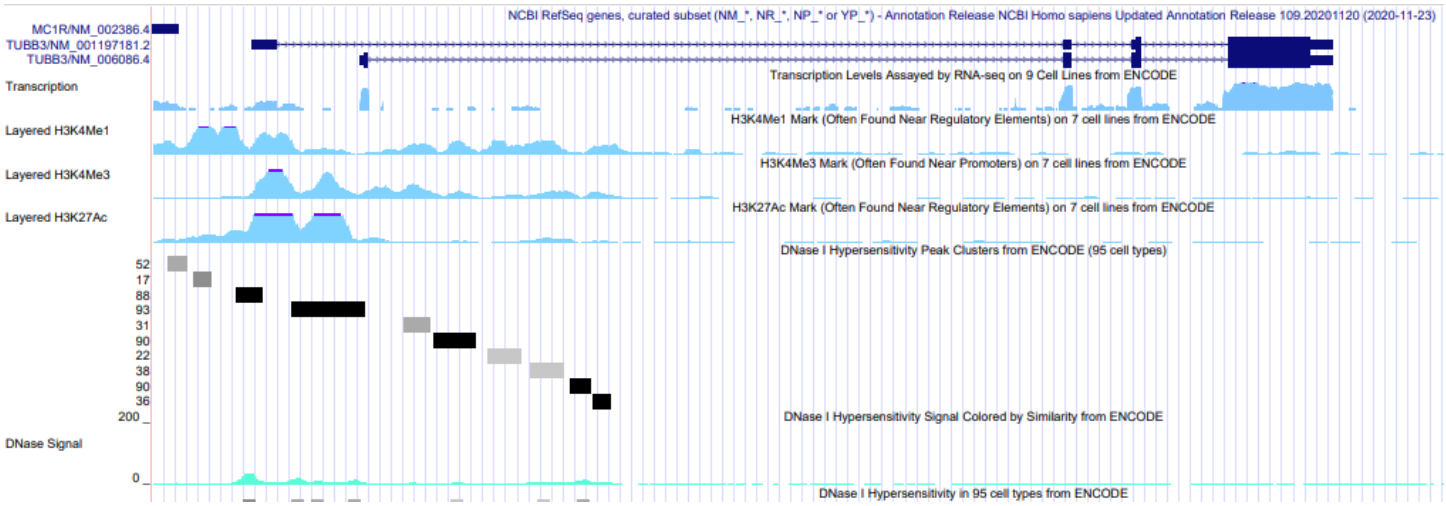
The ENCODE track at the *TUBB3* locus shows histone modifications near the TSS for H3K27Ac, H3K4Me1, and H3K4Me3. Near the TSS, there are also large DNase signal peaks. The transcription track shows high peaks near the end of the *TUBB3* locus. Lastly, there is also data for 340 transcription factors through the *TUBB3* locus.



Full screenshot of locus is also attached (Part2a…)

- At the same locus as Part 2.1, configure each Encode track to only show data for the HUVEC cell line. What is the HUVEC cell line? Based on the Encode tracks, do you think this gene is expressed in the HUVEC cell line? Why or why not? Submit a screenshot of this locus.

HUVEC stands for human umbilical vein endothelial cells, and the HUVEC cell line is the cell line derived from those cells. Based on the ENCODE tracks, I do believe that it is likely that TUBB3 is expressed in the HUVEC cell line. Given that there are peaks of histone modification for H3K4Me1, H3K4Me3, and H3K27Ac near the TSS; they could indicate increased levels of transcription.



Screenshot of locus with this view is also attached (Part2b…)

- Use Galaxy or any other tool (hg38) to find all UCSC flagged SNPs in (hg38/db147) on chromosome 1. Be sure to import the data as a BED file. Describe the result (number of lines, what's in columns, etc.). Also, is the imported BED file 0- or 1-indexed? How can you tell?

The BED file created from flagged SNPs resulted in 8,787 lines with six columns in BED format. In the columns are the chromosome number (1), the location of the flagged SNPs with beginning and ending location, the name of the SNP, the score, and orientation of the strand. The BED file appears to be 0 based. Since these locations represent SNPs and only represent a single nucleotide, if they were 0 based, the start and end location would have a different location, which is the case here for all the SNPs. If they were 1-based, then the start and end locations would be the same, which isn't the case in these.

***BED file is attached (Exam2_Part2C)**

- Use Biomart (web-based or R) to search among all SNPs (Ensembl Variation 99 or latest version) on human chromosome 1 (hg38). Use the Human Somatic Short Variants database and filter for the eye tumour phenotype. Output the following Attributes: Variant Name, Variant Source, Chromosome name, Chromosome position start, Chromosome position end, Variant start in translation, Variant end in translation, Variant Consequence, SIFT prediction, and PolyPhen prediction. Download unique results and submit a spreadsheet. How many total SNPs are reported? How many SNPs are in noncoding regions? SIFT and PolyPhen scores are optional to include.

There are 434 total results reported, of which 133 unique SNPs are represented. Of these unique SNPs, 48 were in noncoding regions (189 of total results, including non-unique SNPs).

***Excel sheet attached (Part2d…)**