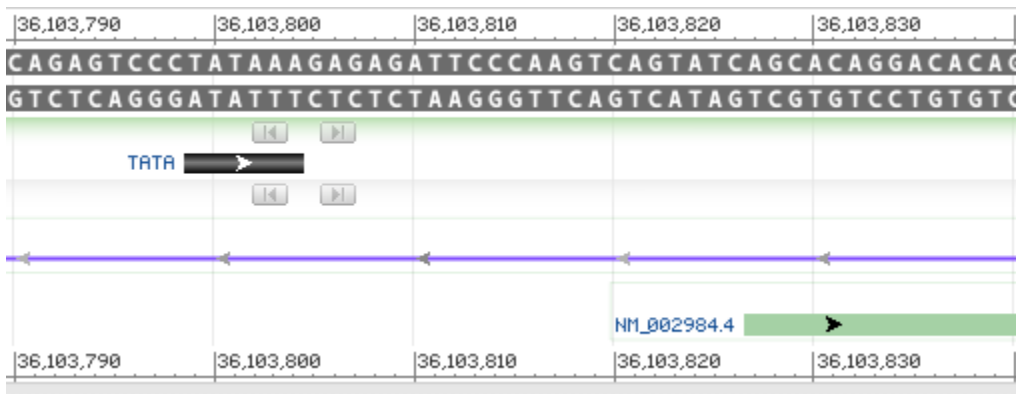


The human CCL4 gene has three exons. In 1999, the gene was called the Act-2 cytokine gene. In a paper in the Journal of Biological Chemistry from that year, Monica Napolitano et al. reported some gene features. There is a TATA box upstream from the transcription start site. The transcription start site is on chromosome 17 at position 36,103,827. There is also a palindromic sequence in exon 2. The authors reported an unknown significance for the palindrome.

4. Created and submitted a two-line 0-indexed BED file that includes the following locations for the CCL4 gene:
  - a. The TATA box, which begins 28 bases upstream from the transcription start site and ends 23 bases upstream from that same site.
  - b. The DNA location of the palindromic sequence, which runs from amino acid 40 through amino acid 45 in the protein.

```
chr17 36103798      36103804      TATA  1      +
chr17 36104568      36104586      Palin 1      +
```

5. The BED file displayed at the NCBI Variation Viewer for the CCL4 gene. Screenshot that shows both the TATA box (labeled) and the transcription start site.



6. Second screenshot that shows the palindromic sequence (labeled) in exon 2. Also zoomed in far enough for me to see the nucleotides in the sequence.



1. NCBI reports one transcript with three exons on its gene page for human CCL4. Ensembl lists three transcripts for the same gene. Briefly explained how the other two transcripts differ from the three-exon transcript.

**The one transcript with three exons that NCBI reports is 1,788 nucleotides long with a protein length of 92 amino acids . Ensembl's two additional transcripts include a retained intron at 1,774 base pairs with no protein, and a transcript with just two exons that produces a protein of 52 amino acids.**

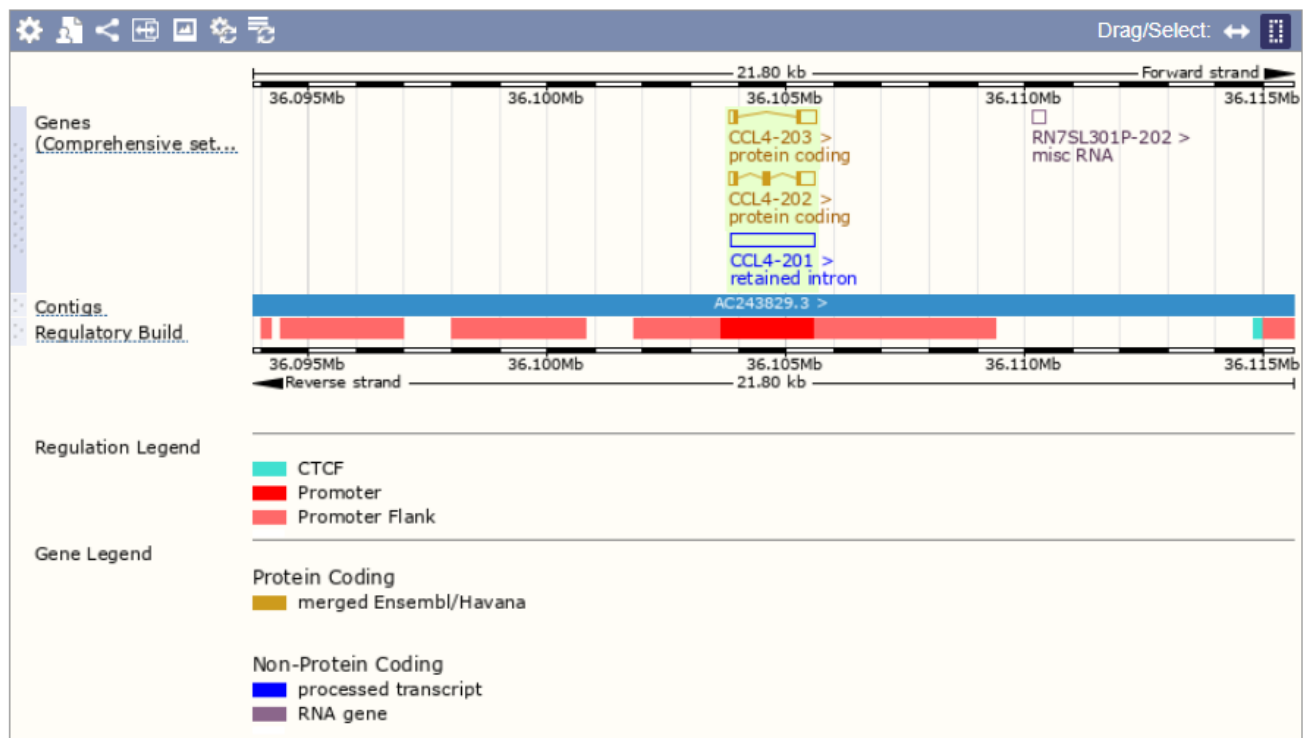
2. Retrieved an image of the Gene Summary image with the Genscan prediction track turned on.

## Summary ?

Name	<a href="#">CCL4</a> (HGNC Symbol)
CCDS	This gene is a member of the Human CCDS set: <a href="#">CCDS11308.1</a>
UniProtKB	This gene has proteins that correspond to the following UniProtKB identifiers: <a href="#">P13236</a>
RefSeq	This Ensembl/Gencode gene contains transcript(s) for which we have <a href="#">selected identical RefSeq transcript(s)</a> . If there are other RefSeq transcripts available they will be in the <a href="#">External references</a> table
Ensembl version	ENSG00000275302.2
Other assemblies	There is no ungapped mapping of this gene onto the GRCh37 assembly. Stable ID ENSG00000275302 not present in GRCh37.
Gene type	Protein coding
Annotation method	Annotation for this gene includes both automatic annotation from Ensembl and Havana manual curation, see <a href="#">article</a> .



Go to [Region in Detail](#) for more tracks and navigation options (e.g. zooming)



3. In Ensembl, found the amino acid location range for the Chemokine interleukin-8-like domain according to the Pfam database for CCL4-202 transcript.

**Amino acid location range for the Chemokine interleukin-8 like domain is: 27-89.**

1. Used Spleign to find the mRNA and CDS coordinates in the genomic DNA from mRNA and genomic files (znf214\_mrna.txt and znf214\_genomic.txt).

- a. mRNA locations: 614..896, 18114..18260, 19414..21651
- b. CDS locations: 18134..18260, 19414..21107

2. Created a BED6 file with 2 lines based on the paper from Takenaka et. al, 2015. Figure 3 from the paper shows the location of transcription factor DdIR binding to the promoter region of the ddIR-ddI operon in *Brevibacillus brevis*. The chromosomal location of the ddIR CDS is 2968133..2969623. The zero-based BED6 file should contain the location information of two genomic regions:

- a. Noted the region bound by the DdIR transcription factor (likely promoter). It is 170 bp in length, begins 140 nucleotides upstream from the start codon, and ends 29 nucleotides downstream from the start codon.
- b. The 5' UTR, noting that the transcription start site, as predicted by BPROM, begins 38 nucleotides upstream from the start codon. The 5' UTR is defined as the region from the transcription start site through the nucleotide that immediately precedes the start codon.

a.	chr1	2967992	2968162	promoter	0	+
b.	chr1	2968094	2968132	5'UTR	0	+

3. Screenshot of the BED6 from Problem 2. Using the NCBI Genome Browser for [Brevibacillus brevis NBRC 100599](#), loaded the BED6 file from the previous problem. Screenshot showing the entire promoter, 5' UTR region, and CDS of ddIR.

