Paper by Sun *et al* list three cow SNPs in the *CFL2* gene. One is intronic, one is synonymous, and the other is said to be nonsynonymous.
The SNPs are at the mRNA positions in the table (this is a minus strand gene). The DNA positions are based on the unspliced full-length mRNA. Are there issues with their listings?

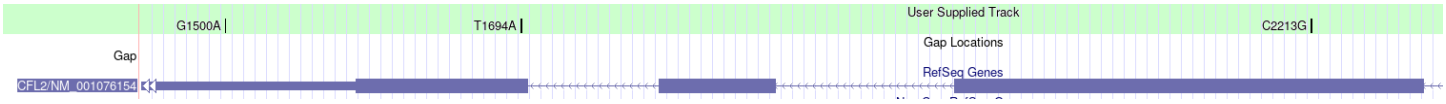
DNA change	Listed amino acid change	Listed location
C2213G	Proline-312-Alanine	exon 4
T1694A	Isoleucine-131-Isoleucine	exon 4
G1500A	noncoding	intron 2

- Create a zero-based BED file with the three SNP locations. Label each line with the DNA change. Submit the file. HINT: The SNP at position 1694 is located at position 45,888,611 on chr21 in cow (bosTau7 build). 45,888,611 is the actual position of the SNP nucleotide, not the zero-based position.

chr21 45889129 45889130 C2213G 0 -
chr21 45888610 45888611 T1694A 0 -
chr21 45888416 45888417 G1500A 0 -

***File also attached (4a…)**

- Load the BED file into the UCSC genome browser. Zoom into the CFL2 gene. Be sure the exon pattern and the custom track with the labeled SNPs are viewable. Submit the browser shot.



***Full browser shot file also included (Part4b…).**

- Based on the two SNPs with amino acid changes, what possible codon changes could cause those amino acids to change? Example: A C-to-A nucleotide change could make a His-to-Asn change by changing a CAU codon to an AAU codon.

Based on the amino acid changes, C2213G with a change from Proline to Alanine, there could be 4 possible codon changes. Codons CCT, CCC, CCA, and CCG could all change by the first "C" nucleotide to a "G" to GCT, GCC, GCA, and GCG, respectively. All those codon changes would result in the amino acid change from Pro-to-Ala.

T1694A, stays Ile-Ile, even with a nucleotide change from "T" to "A". In this scenario, only one codon would keep this result. Codon ATT could have the 3rd nucleotide change "T" to "A" to ATA and remain Ile.

- The CFL2 protein is 166 amino acids in length. That would make an amino acid change at position 312 a bit difficult. Looking at the exon sequence for CFL2 in **UCSC Genome Browser**, in which exon is the C2213G SNP? Specifically, what part of the exon: 5'UTR, CDS, or 3'UTR?

According to the UCSC Genome Browser, the C2213G SNP is actually located in exon 2 of 4 (negative strand). It is part of a CDS.

- The attached cow_cfl BED files contain all exons for the CFL2 gene and the coding exons for the CFL2 gene. Intersect your SNP BED file to each of those files using Galaxy or command-line BEDtools. How many of the three SNPs intersect with all exons? How many intersect with coding exons?

All three SNPs intersect with all exons, while 2 SNPs (C2213G and T1694A) intersect with coding exons.

- Submit either a screenshot of your output (Galaxy) or the commands you ran (command line) for the previous answer.

Intersections in coding exons:

chr21	45889129	45889130	C2213G	0	-
chr21	45888610	45888611	T1694A	0	-

Intersections in all exons:

chr21	45889129	45889130	C2213G	0	-
chr21	45888610	45888611	T1694A	0	-
chr21	45888416	45888417	G1500A	0	-