****

* **Identify the human gene that encodes a protein with the peptide sequence GPDGMPVIYHGHTLTTKIKFSDVLHTIKE**

 **and answer the following:**

* 1. **What is its chromosomal location?**
	2. **What neighboring genes are located in this region?**
	3. **How many splice variants (isoforms) have been reported for this gene?**
	4. **Pick an isoform and retrieve the genomic coordinates for its first three exons.**
	5. **Submit its DNA sequence highlighting all exons in blue and synonymous SNPs in red**
	6. **Submit its**

* + - **mRNA sequence(s) in FASTA format**
		- **Protein (amino acid) sequence(s) in FASTA format**
		- **1000 bp 5’ upstream potential promoter sequence**
* Find the genomic location of mouse BCL2 gene.
	+ What neighboring genes are located in that region?
	+ Have any SNPs and structural variations been reported?
	+ Find the following:
		- genomic sequence with all exons in blue and synonymous SNPs in red
		- mRNA sequence in FASTA format
		- amino acid sequence in FASTA format
		- 1000 bp 5’ upstream potential promoter sequence
		- presence of conserved transcription factor binding sites in the promoter region of its **human** homolog
* Determine the genomic location of human EGFR and its interaction partner, the peptide sequence :

**PSRQFVKDSIRLVKRCTKPDRKEFQKIAM**

* + Find the full protein sequence of this EGFR interaction partner
	+ Using the human protein sequence you have just retrieved, find the genomic location of its chicken (*Gallus gallus*) homolog
	+ After browsing through its neighboring genes, can you speculate on the transcription regulation of human EGFR and its above mentioned partner?

**NCBI Map Viewer:**

* During a positional cloning project aimed at finding a human disease gene, linkage data suggests that the gene of interest lies between two sequence-tagged site markers (D10S1676 and D10S1675).
	+ How can all known and predicted candidate genes in this interval be identified?
	+ What diseases are reported to be associated with this region?

**UCSC Table Browser**:

* Find chromosomal location overlap between the “known genes” dataset and the “simple repeat” dataset and then retrieve simple repeats of a copy number over ten.
* Retrieve all disease causing genes reported in OMIM database present in the region of human chromosome 7 between 55054219 to 55354319 bp (chr7:55054219-55354319).