**Subject: Ubiquitin**

**Ubiquitin is a challenging protein to explore because there are many sequences that make up the gene family. How many ubiquitins are in the human genome and what is the representative ubiquitin sequence? Please explain how you arrived at your answer.**

I took a stab at this. I found 4 human ubiquitin genes, UBA52, UBB, UBC, and RPS27A.

First, I typed **ubiquitin** into Entrez to see what it would come up with and I got 227,683 genes. Then I searched for **ubiquitin AND “homo sapiens”[Organism**] and that narrowed it down to 2669 genes, 2181 when I selected “Current”. I decided to go a different route through Taxonomy Browser, selected homo sapiens and then selected Gene. I narrowed this down by adding ubiquitin to the search bar so that the whole search looked like this: **txid9606[Organism:exp] ubiquitin**. I got 2181 results (the same as my search above). Using the Boolean operator AND gave the same results (AND is the default). I then selected “Genomic” and “Protein Coding”, and that narrowed it down to 1937, which was still too many. I adding phrases like **NOT ligase** and **NOT like**, but couldn’t get a definitive list.

Realizing that GeneBank contains redundant data and that this list also includes ubiquitin related genes and molecules, I tried a different route. I knew of one of the ubiquitin genes from another class, so I searched for ubiquitin C UBC and found one of the human ubiquitin protein coding genes. From there I went to Ensembl and found that this gene is a member of 5 protein families. On Ensembl I found 4 human ubiquitin genes, UBA52, UBB, UBC, and RPS27A.

The ubiquitin sequence I found by looking each of these up on GeneBank is MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG.

I’m sure there are other ways to do this, if anyone else has suggestions let me know.

References:

1. <https://www.ncbi.nlm.nih.gov/books/NBK49540/>
2. <https://www.ncbi.nlm.nih.gov/gene/>
3. <http://uswest.ensembl.org/Homo_sapiens/Gene/Family?db=core;g=ENSG00000150991;r=12:124911604-124917368>
4. <http://www.treefam.org/sequence/ENSG00000150991>
5. https://blast.ncbi.nlm.nih.gov/Blast.cgi