

# Science Summer School 2017



Researchers in the Centre for Chromosome Biology at NUI Galway work to understand the molecular mechanisms of genetics. This knowledge can allow us to diagnose and treat many diseases including diabetes and cancer. ([www.chromosome.ie](http://www.chromosome.ie))

This instruction sheet is available at [www.chromosome.ie/pomc](http://www.chromosome.ie/pomc)

## A - Human and dog chromosomes

*In the last decade, new technologies have enabled the complete genomes of many organisms to be sequenced. The most famous of these projects is the human genome. The European Bioinformatics Institute has a very powerful information service called Ensembl that provides access to genomic information freely for everyone.*

1. Open a new browser window and go to the website [www.ensembl.org](http://www.ensembl.org)
2. Under **Browse a Genome** on the left, click **Human** under Popular Genomes
3. After the page loads, click the **View Karyotype** link within the top left Genome Assembly panel. **Count how many chromosomes.**
4. Go back to the [www.ensembl.org](http://www.ensembl.org) homepage.
5. In the Browse a Genome panel, use the All Genomes pull-down menu (select a species) to look up **Dog** (*Canis lupus familiaris*).
6. Using View Karyotype again, **count how many chromosomes.**

**Q1. Is the human karyotype larger or smaller than the dog karyotype?**

**Genome:** The complete set of DNA encoding the genetic information of an organism

**Chromosome:** Microscopic structure that organises a part of the DNA genome

**Karyotype:** The number and appearance of chromosomes in a cell

## B - Classifying organisms

*Organisms are traditionally classified using their visible characteristics (phenotypes). Now we can also classify them using their genetic information (genotypes). This gives us an improved picture of evolutionary relationships.*

1. Open a browser window and go to the webpage [www.ncbi.nlm.nih.gov/taxonomy](http://www.ncbi.nlm.nih.gov/taxonomy) then follow the link for **Common Tree** under Taxonomy Tools
2. One at a time, type the scientific names of the organism below in the 'Enter name' box and click **Add**
  - *Homo sapiens* (human)
  - *Canis lupus* (dog)
  - *Gallus gallus* (chicken)
  - *Mus musculus* (mouse)
  - *Carassius auratus* (goldfish)

**Q2. Are dogs or are mice more closely related to humans in the phylogenetic tree?**

**Genotype:** The genetic makeup of an organism

**Phenotype:** The observable characteristics of an organism

**Taxonomy:** The science of classifying organisms

**Phylogenetic tree:** A diagram showing the evolutionary relationship between organisms

## C - Looking up gene and protein sequences

*Sequencing the entire genome of an organism allows us to identify all of its genes. The genetic code also allows us to translate these DNA sequences into the amino acid sequence of the protein they encode.*

*Databases like Ensembl and UniProt now hold all the identifiable gene and protein sequences for many organisms together with analyses of their properties. The databases are extensively inter-linked to make it easy for researchers to learn about the biochemistry they uncover.*

1. Open a new browser window and go to the website [www.ensembl.org](http://www.ensembl.org)
2. At the top of the homepage in the light blue Search panel, pull down the menu with "All Species" and select "Human"
3. In the box to the right of "for", type the gene name POMC then click Go
4. In the results listing, click on "**POMC (Human Gene)**"
5. The **Gene: POMC** page will load giving details about POMC

**Q3. What is the full name of POMC?** *Hint: Look at description at top of page.*

6. Click the Chromosome link given for the Location information.

**Q4. On which chromosome is the POMC gene located in the HUMAN genome?**

7. Go back to the Gene: POMC page
8. In the middle of the page is a list of possible transcripts for POMC. In the line for POMC-204, click on the link **P01189** in the UniProt column.
9. A new page will load from the UniProt database giving lots of information about the protein. Scroll down to the **Sequence** section and look to the right for the **Length**

**Q5. What is the length of the human POMC protein?** *Note: This is the number of amino acids in the protein.*

10. Repeat steps 1-9 above for POMC in the Dog genome. *Note that the UniProt ID code for Dog POMC is E2RQ39*

**Q6. What is the difference in length of human and dog POMC proteins?**

**Gene:** A region of DNA in the genome that codes for a protein

**Protein:** A long chain of amino acids that folds into a functional molecule in cell (eg enzyme, hormone)

**Sequence:** The linear order of nucleotides (DNA, RNA) or amino acids (protein) that make up a biomolecule

**Genetic code:** Set of rules by which DNA sequences can be translated into protein sequences by cells

**Nucleotide single letter code:** Letter for each DNA chemical building block A, C, G, T (eg A = adenine)

**Transcript:** A length of gene copied as a messenger RNA (mRNA) and then used in translation to protein

## D - Comparing protein sequences

One of the most useful things researchers do with DNA or protein sequences is to compare them to look for differences. This is called "multiple sequence alignment". A popular tool for alignments is called Clustal Omega.

1. Open a new browser window and go to the website [www.ebi.ac.uk/Tools/msa/clustalo/](http://www.ebi.ac.uk/Tools/msa/clustalo/)
2. Copy only the **Human POMC** and **Dog POMC** wild type protein sequences from below into the **Sequences** box under **Step 1** in Clustal Omega.  
*Note: Include the label lines starting >HumanPOMC and >DogPOMC*
3. Click the Submit button. It may take 10-20 seconds for the alignment to come back.

### Q7. How many short insertions/deletion locations are there between the human and dog POMC proteins?

4. Underneath the sequence alignment is a code showing the similarity at each amino acid position. Count the number of identical amino acids
  - \* (star/asterisk) is an identical match of amino acids
  - : (colon) are chemically very similar amino acids
  - . (period/full stop) are chemically somewhat similar amino acids
  - (space) means chemically different amino acids, or an insertion/deletion
5. Calculate the number of identical amino acids by count the differences (colons, periods and spaces) then subtract from the total protein length (from Q5 above).

### Q8. What is the percentage identity between human and dog POMC? Hint: Divide your answer in step 5 by the total protein length.

6. Go back to the Clustal Omega homepage [www.ebi.ac.uk/Tools/msa/clustalo/](http://www.ebi.ac.uk/Tools/msa/clustalo/)
7. Copy only the **Dog POMC** and **Labrador POMC mutant** protein sequences only from below into the **Sequences** box under **Step 1** in Clustal Omega. *Note: Include the label lines starting >DogPOMC and >LabradorPOMCmutant*
8. Click the Submit button. It may take 10-20 seconds for the alignment to come back

### Q9. At which amino acid position number does the Labrador POMC mutant protein sequence start to differ significantly from the wild type dog POMC proteins?

#### Single letter amino acid

**code:** Each of the 20 amino acids in proteins is represented by a letter (eg A for alanine, Y for Tyrosine)

**Mutation:** A change in the DNA sequence. If in a gene, this can then give a change in the protein sequence

**Insertion/deletion:** One sequence contains a stretch of sequence (insertion) that is missing in other sequences (deletion)

**Wild type:** Non-mutated form of a gene or protein

# POMC Protein Sequences

>HumanPOMC

MPRSCCSRSGALLLALLLQASMEVRGWCLESSQCQDLTTESNLLCIRACKPDLSAETPMFPGN  
GDEQPLTENPRKYVMGHFRWDRFGRRNSSSSGSSGAGQKREDVSAGEDCGPLPEGGPEPRSDGA  
KPGPREGKRYSYMEHFRWGKPVGKKRRPVKVYPNGAEDESAAEFPLEFKRELTGQRLREGDGPD  
GPADDGAGAQAADLEHSLLVAAEKKDEGPYRMEHFRWGSPPKDKRYGGFMTSEKSQTPLVTLFKN  
AIIKNAYKKGE

>DogPOMC

MPRSCCSRPGALLLALLLQASVEVSGWCLESSQCQDLTTESNLLACIRACKPDLSAETPVLPGN  
GDEQPLAENPRKYVMGHFRWDRFGRRNGSAGQKREEEVAAGGGRAPLPAGGPGPRGDGGELGL  
QEGKRYSYMEHFRWGKPVGKKRRPVKVYPNGAEDESAAEFVFEFKKELARQRLEPALGPGPAAG  
VAALADLEYGLVAEAGAAEKKDDGPYKMEHFRWGSPPKDKRYGGFMSSERSQTPLVTLFKNAII  
KNAHKKGQ

>LabradorPOMCmutant

MPRSCCSRPGALLLALLLQASVEVSGWCLESSQCQDLTTESNLLACIRACKPDLSAETPVLPGN  
GDEQPLAENPRKYVMGHFRWDRFGRRNGSAGQKREEEVAAGGGRAPLPAGGPGPRGDGGELGL  
QEGKRYSYMEHFRWGKPVGKKRRPVKVYPNGAEDESAAEFVFEFKKELGRGGAGRPGVRPGGGG  
RGGREEGRRALQD GALPLGQPAQGQALRRLHELGE E PDAPGDAVQKRHHQERPQEGPVRAQRGR  
RLLFPGS PRKAPLLPALLPSRPPPAPDVLLAQAVGGSPVVSLLKLPVVRK