



**Goal:** In this assignment, you have access to the human reference genome as well as the genome annotation. In addition, you have full genome sequence data from five individuals from a family at risk of carrying mutations related to the disease.

Your task is to write a Python program that will extract the CFTR gene, translate the gene sequence to its corresponding amino-acid sequence and based on the inferred amino-acid sequence determine whether any of the five given individuals is affected.

## » Fetch the appropriate files

The main task is divided in several steps. The first step is to fetch the sequence file (in `fasta` format) and the appropriate annotation file (in `GTF` format) from the [Ensembl database](#).

The CTFR gene is chromosome 7.

## » Warmup

1. What is the length of the chosen DNA sequence?
  - Tip
2. How many genes are annotated in the GTF file?
  - Note
3. What fraction of the chromosome is annotated as genes?

## » Architect a method

All the following tasks are now related to the CTFR gene.

In the annotation file (from the Ensembl database), that gene has the id `ENSG00000001626` on chromosome 7.

1. How many transcripts can this gene generate?