Thursday morning practical
The Genome Aggregation Database (gnomAD)

gnomAD has released data from 125K exomes and 15K genomes. It is possible to download the sites VCF (which has allele frequencies and annotations but not individual genotypes), but we will just take a look at the browser: http://gnomad.broadinstitute.org.

The database contains both common and very rare variation found in the ~140K individuals' exomes and genomes. You can enter any gene to see the variants detected in all these samples, as well as their frequencies.

Question 1: Sort the variants in your favorite gene (if you don't have one, look at PCSK9) by allele frequency. Are there lots of common variants?

Use the "LoF" button to show only the loss-of-function variants, which tend to be the most severe.

Question 2: Compare the loss-of-function variants observed in the genes *TCF4* and *PCSK9*. What do you observe? Why might this be?

Click on an individual variant and explore the information that is given on the next page.

Question 3: What is notable about the variant 7-117199644-ATCT-A in *CFTR* (otherwise known as *CFTR*-deltaF508)? If you found it in a patient of East Asian ancestry with a severe neurological disorder, would you think it a good candidate for being the pathogenic variant?

Question 4: You have sequenced an individual with a rare disease and discovered a G->A variant on chromosome 9 at position 127661645 (denoted 9:127661645). Why might you rule this out as being causal of her disease? What if you found a frameshift variant at 1:151638470?