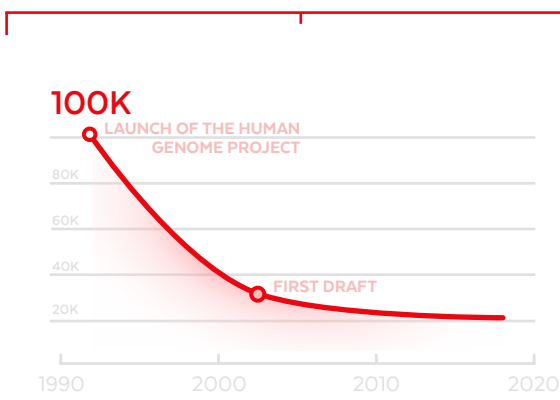


● GENE WHICH DEFICIENCY IS ASSOCIATED WITH DISORDERS ● PROTEIN CODING GENES ● NON-CODING GENES

# 60K Genes, 33% Protein-Coding

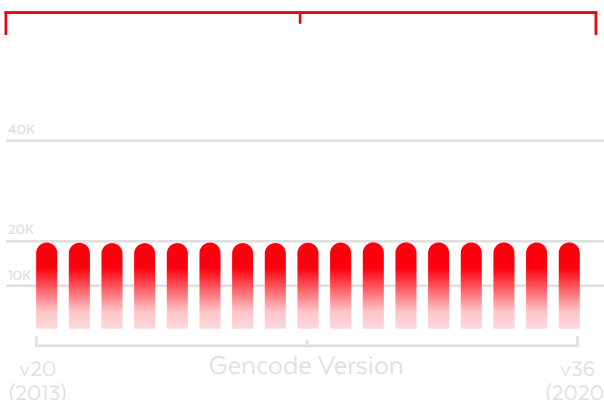
The estimative of genes in human genome varied greatly along the time. Following the genome size and gene content ratio of many procaryotes, one of the first estimate was about 3 million genes. Now, after the Human Genome project and massive improvement of sequencing technology, it's possible to narrow down that number to approximately 60 thousand genes. In fact, in Gencode - one of the main references for human genome annotation - there is 60,660 genes in our genome while only 1/3 of it (19,962) are protein coding gene. In OMIM, only 4,498 human genes, when disrupted by mutations are associated with some disorder. Now we only have to discovery what is the role of all those genes!

PROTEIN CODING ANNOTATION EVOLUTION



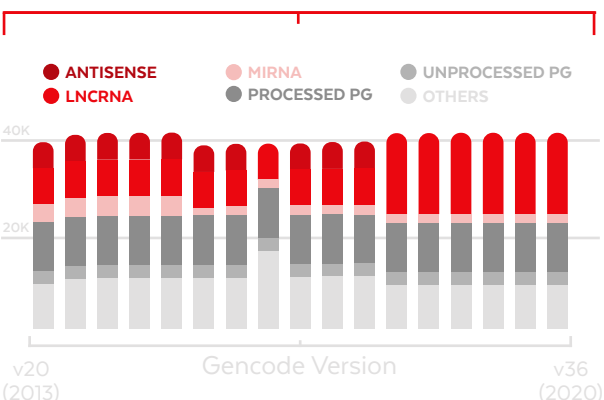
At the time of the launch of the Human Genome project, the protein-coding genes estimated number of in our genome was of about 100,000 genes. Then, since the first draft, in 2003, we got a lot closer to the actual number of protein-coding genes, about 20,000.

PROTEIN CODING IN GENCODE



The number of protein-coding gene in one of the main genome annotations , Gencode, varied less than 200 genes between the first version for the most recent human genome assembly in 2013, to the version 36 in 2020.

NON CODING IN GENCODE



While the protein-coding gene annotation didn't change that much, non-coding gene number varied no only in number (about 2 thousand genes) but in classification too (e.g. long intergenic non-coding RNA, lincRNA were merged in onlu long non-coding RNAs, about half of miRNA in the version 20 were removed or reclassified in the most recent version).

