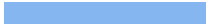
From proteins to base pairs: building the “no-mutation zone”

HUMAN PROTEOME

122,676

protein records



100 %

Ensembl GRCh38,
all isoforms

CONSERVED IN PRIMATES

1,919

proteins

1.6 % of the proteome

identical over their full length in all 10 primate genomes (2,383 if partial alignments are allowed)

WITH EXON COORDINATES

1,439

proteins, 7,071 exons

1.2 % of the proteome

the other 480 are absent from Ensembl's GRCh37 annotation

MERGED INTERVALS

889 kb

in 3,983 intervals (GRCh37)

0.03 % of the genome

overlapping exons of the 1,439 proteins merged

WHAT IS INSIDE

72 % UTR

28 % protein-coding



252 kb coding

637 kb UTR

whole exons were kept, including 5′ and 3′ UTR