Frequencies of codon changing SNPs in a human population

Anneke Brümmer

anneke.brummer@unil.ch

The genetic code is redundant

RNA codon table

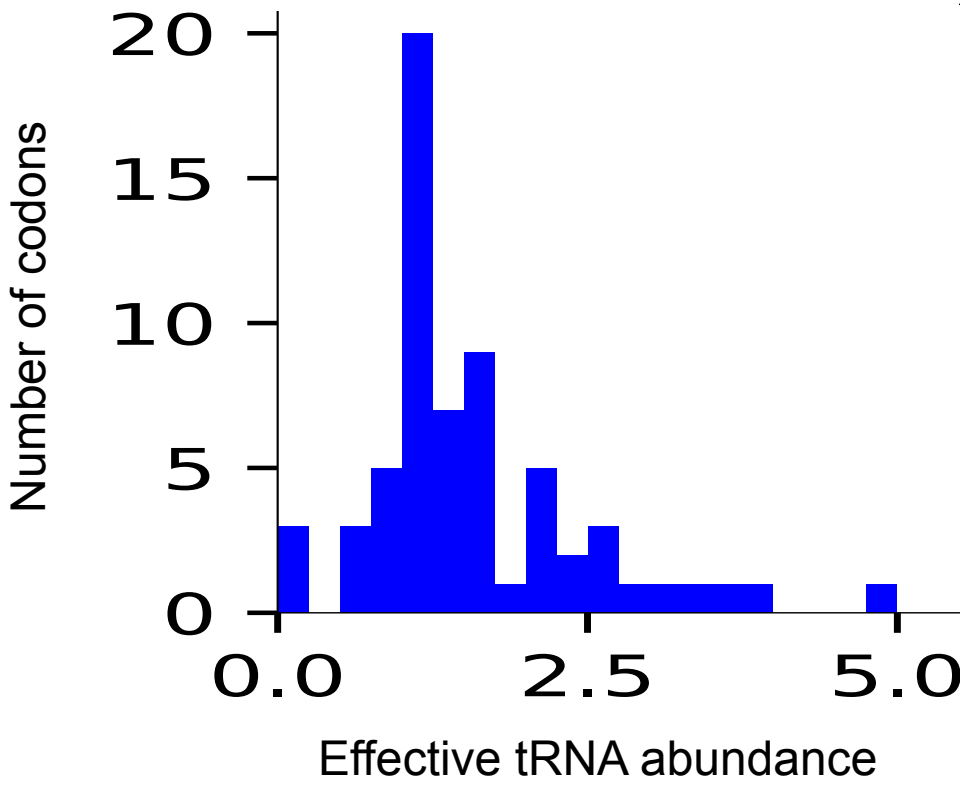
1st position	2nd position U	2nd position C	2nd position A	2nd position G	3rd position
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	stop	stop	A
	Leu	Ser	stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

SNPs can cause:

- non-synonymous (missense) codon change
- stop gain or loss
- synonymous codon change

tRNA abundances are not equal

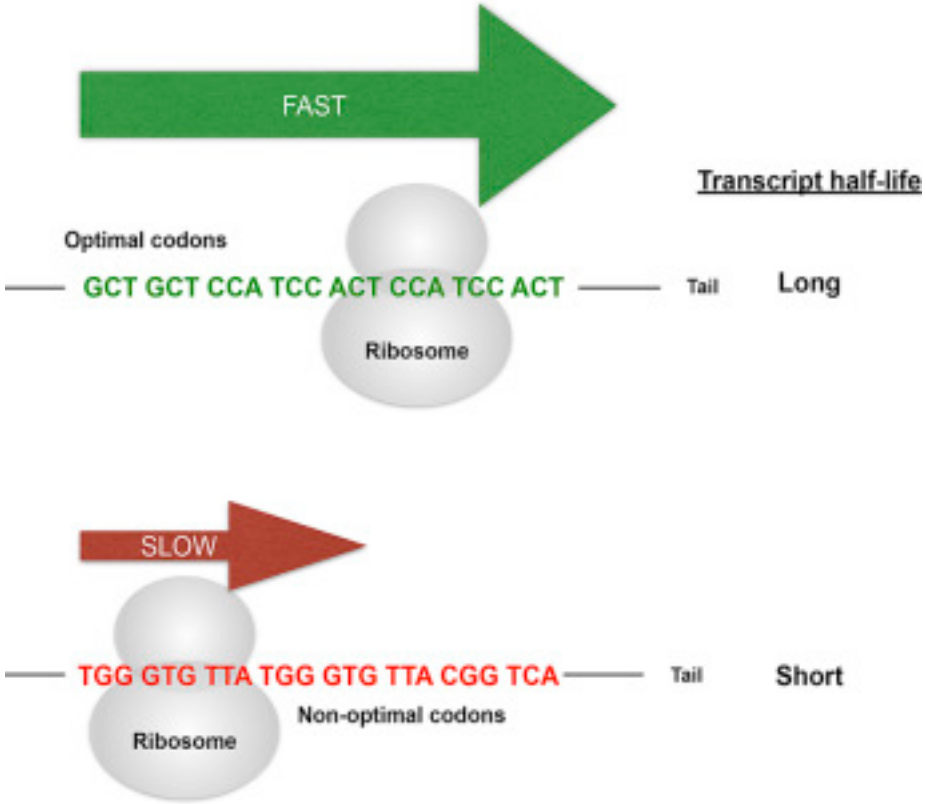
- unequal number of genes per tRNA
- differences in the expression of tRNA genes
- some codons can be recognized by multiple tRNAs (due to wobble base pairing and tRNA editing)



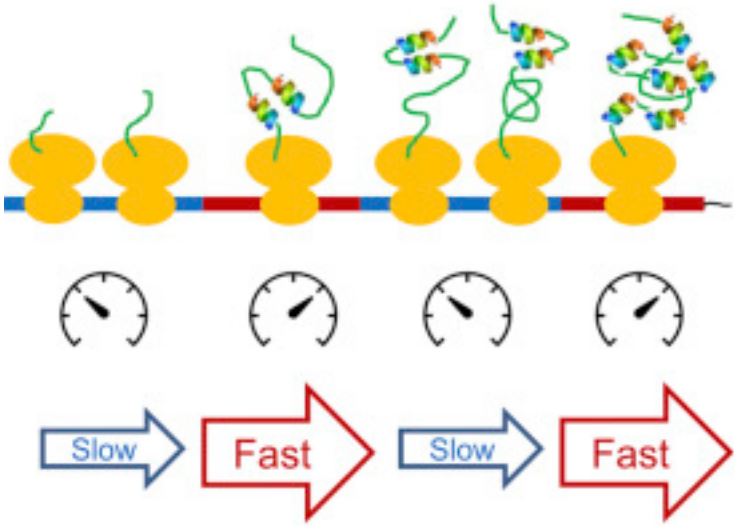
- codons corresponding to abundant tRNAs
  optimal codons
- codons corresponding to rare tRNAs
  non-optimal codons

Synonymous mutations can have a functional impact

- change in translation speed:
  -> mRNA stability
  -> protein abundance
  -> protein folding

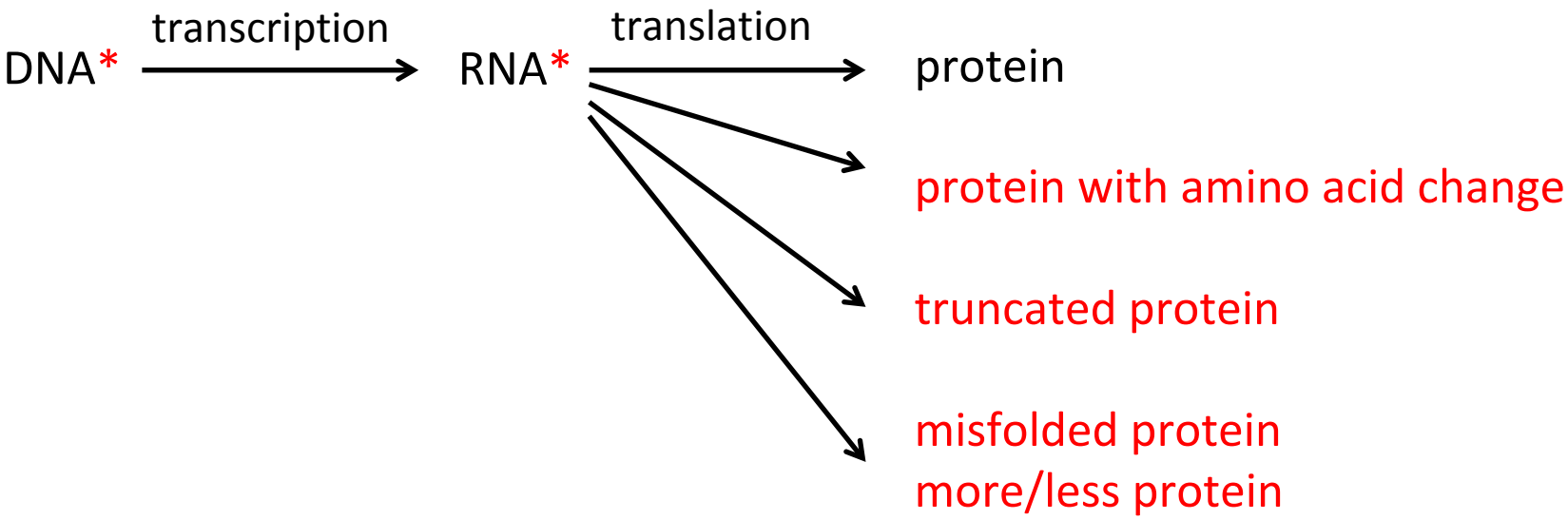


Codon Optimality Is a Major Determinant of mRNA Stability. Presnyak et al. Cell 2015.



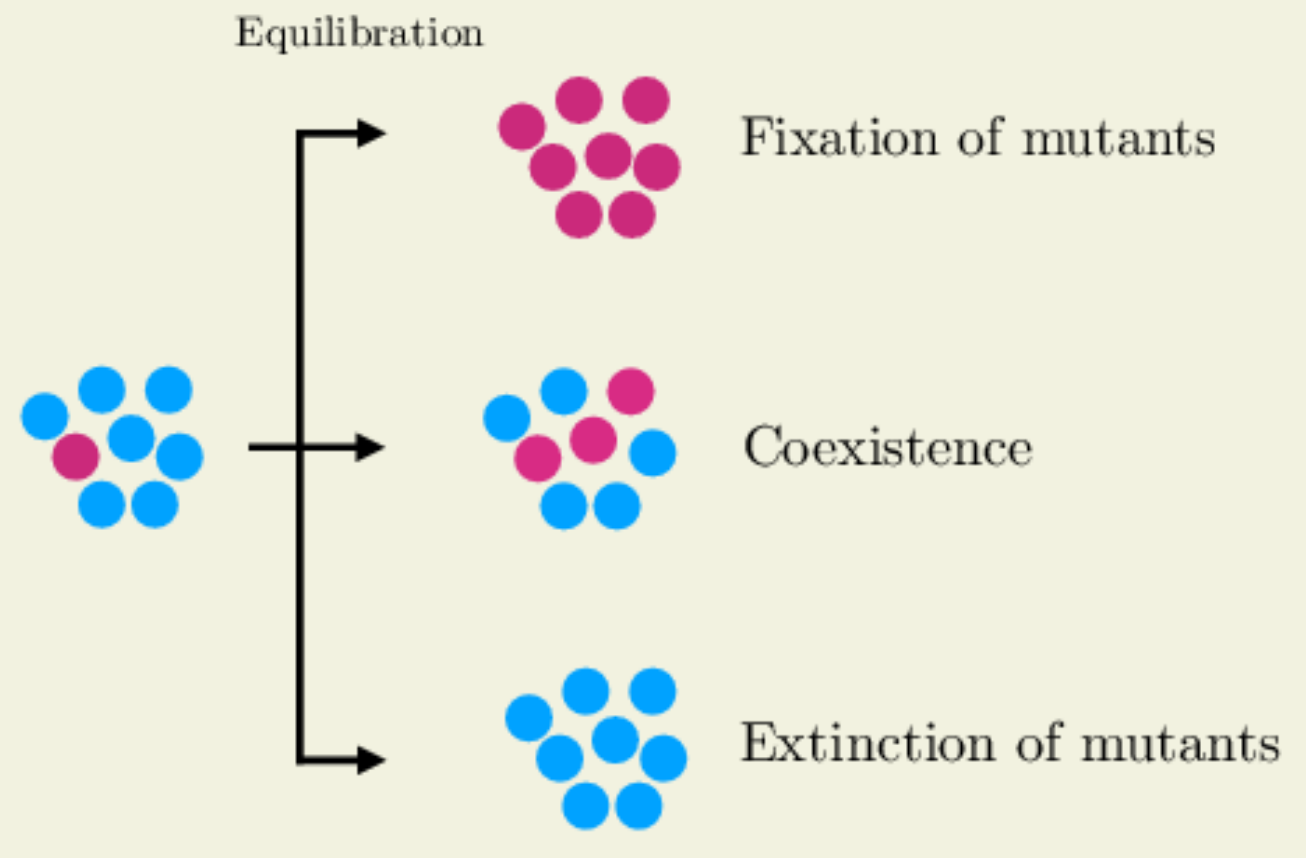
Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. Yu et al. Mol Cell 2016.

Possible consequences of mutations in coding regions

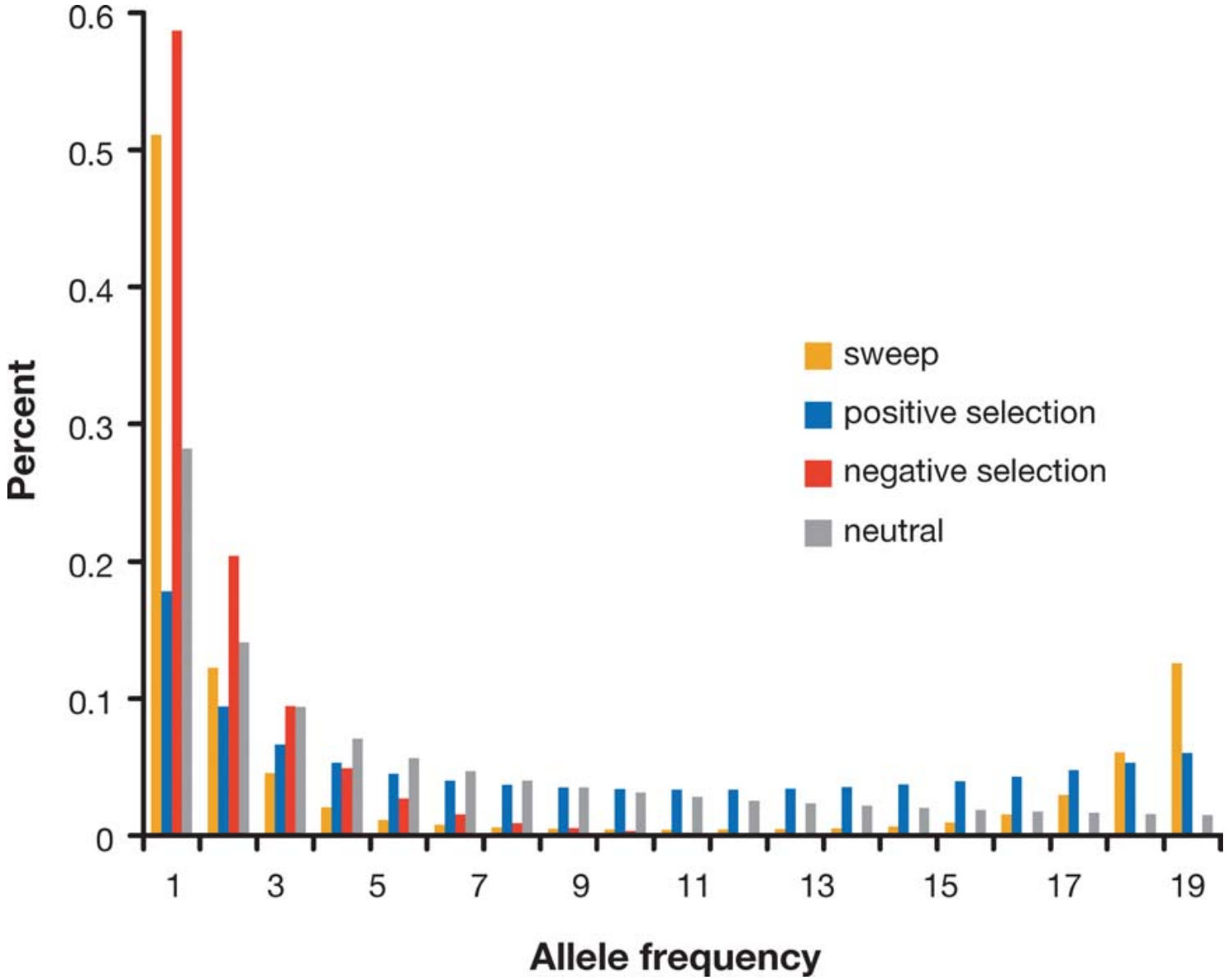


Which codon changes have a functional impact?

SNP frequencies in a population



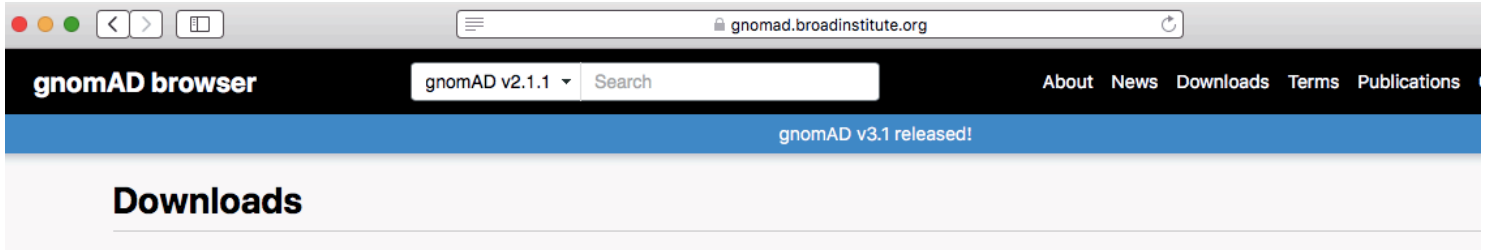
Allele frequency spectrum



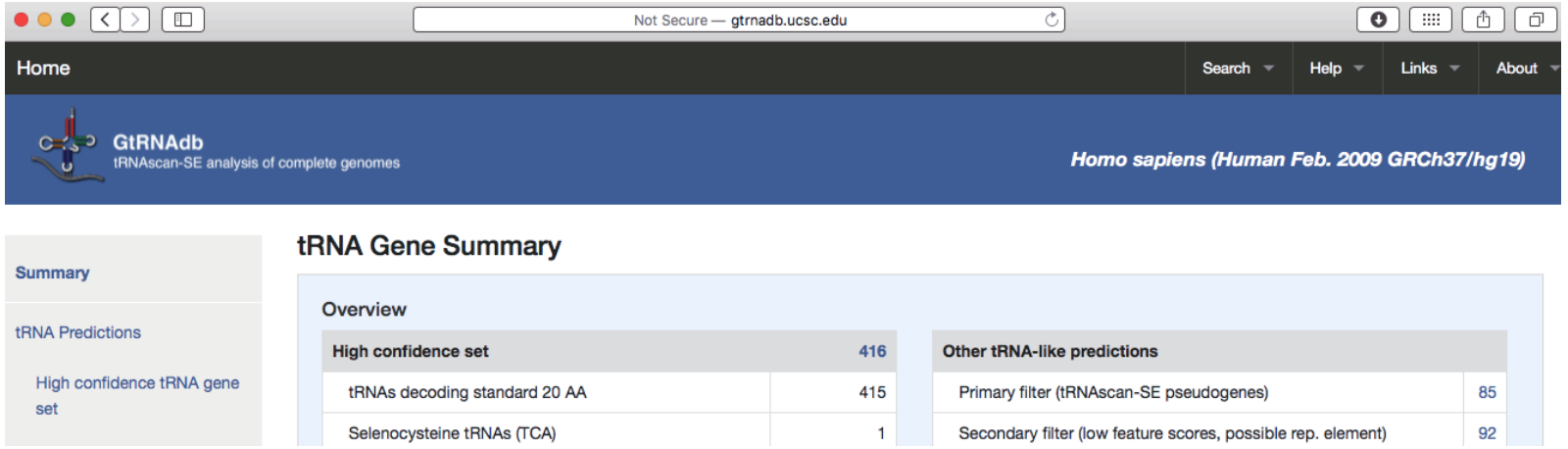
Nielsen R. Molecular signatures of Natural Selection. Annu Rev Genet, 2005.

Data:

- **SNP frequencies from gnomAD (gnomad.broadinstitute.org)**
  - >14 million SNPs from > 140000 exome sequencing data
  - minor allele frequencies down to ~10e-6



- **tRNA gene predictions from GtRNAdb (gtrnadb.ucsc.edu)**
  - Human tRNA gene copy numbers
  - calculate codon optimality measure



Questions:

- **How does the SNP frequency spectrum look for different codon changes?**
    - missense, nonsense, synonymous (slower, similar, faster)
    - different amino acid changes

- Which codon changes are better tolerated than others?

- Are there amino acid changes that can be tolerated?

- Are the frequencies different for SNPs in highly expressed genes?

- What genes have higher/lower frequency SNPs?

- Where within a gene are slower/faster mutations?

What you will learn:

- python
- reading and combining information from different data sets
- plotting
- biology

Questions?

anneke.brummer@unil.ch