

# **Koľko má človek génov a ako ich spočítat'?**

**Broňa Brejová, Katedra informatiky, FMFI UK**

## Gén v klasickej genetike

Gén = jednotka dedičnosti spôsobujúca určitý znak

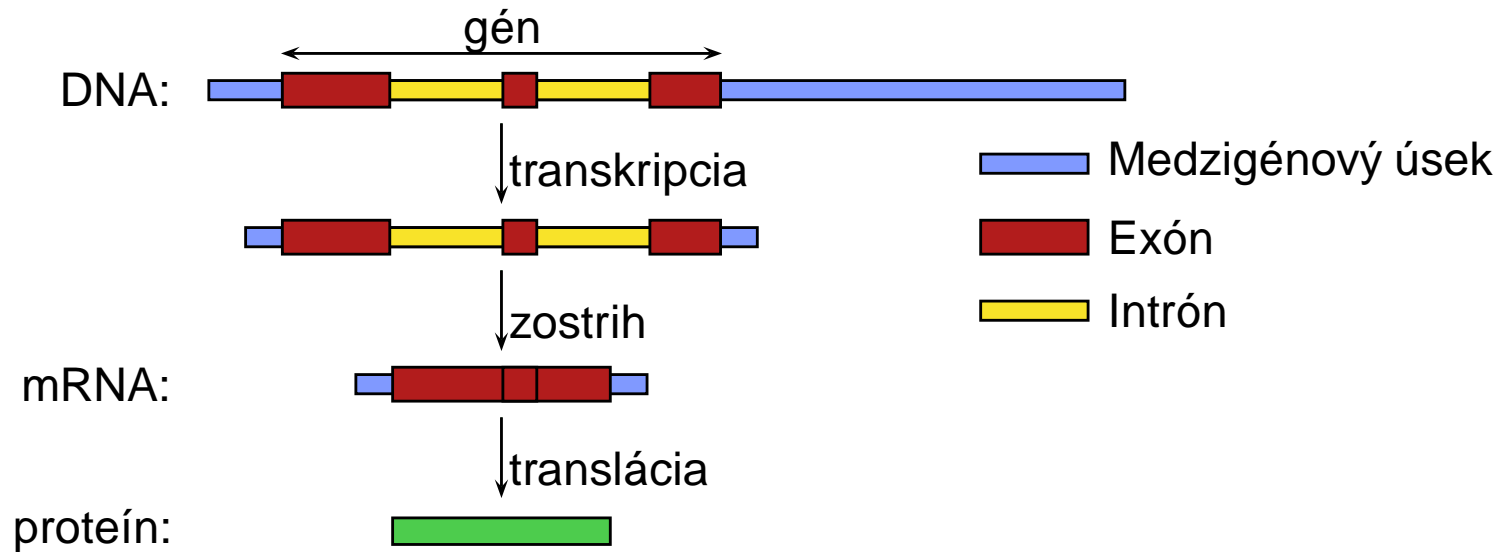
V organizme väčšinou dve kópie, recesívne a dominantné alely

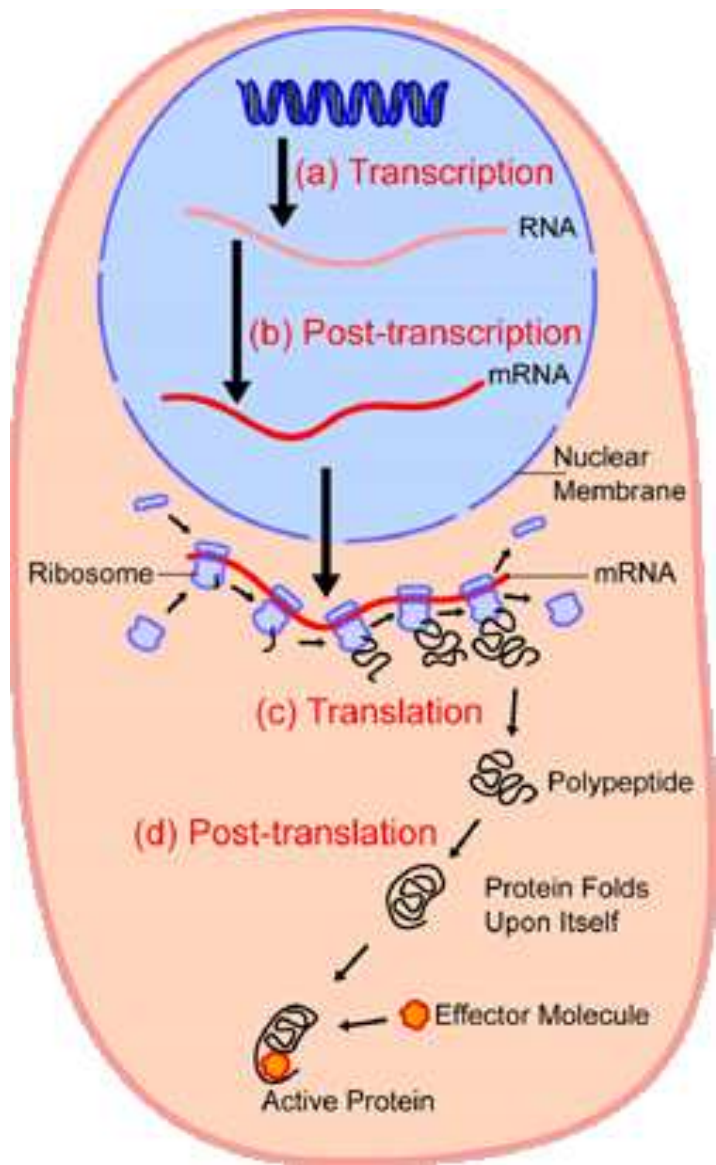
- 1866: Gregor Mendel formuluje zákony dedičnosti
- 1909: Wilhelm Johannsen zavádza slovo gén
- 1910: Thomas Morgan študuje segregáciu mutácií v drozofile, gény zoradené na chromozómoch



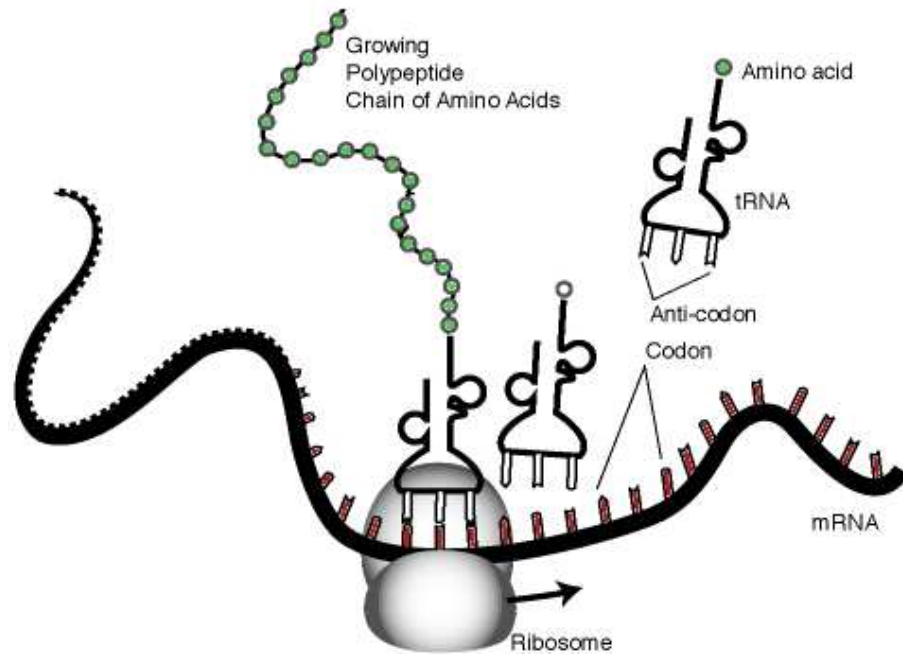
# Biochemická podstata génov 1940-1977

Gén ako úsek DNA kódujúci proteín

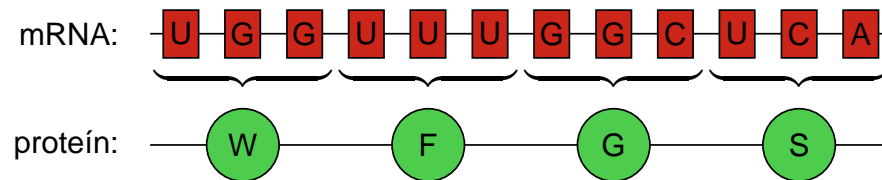




# Translácia



Kodón (trojica nukleotidov) určuje 1 aminokyselinu



## Sekvenovanie genómov

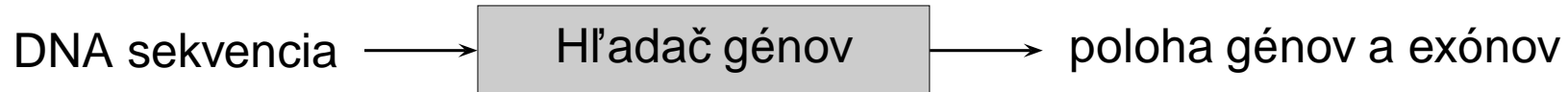
Určovanie poradia báz A,C,G,T

- 1976 vírus MS2 40 kB
- 1988 projekt sekvenovania ľudského genómu (15 rokov)
- 1995 baktéria *H. influenzae* 2 MB
- 1996 kvasinka *S. cerevisiae* 10 MB
- 1998 červ *C. elegans* 100 MB
- 1998 Celera: ľudský genóm do troch rokov!
- 2000 muška *D. melanogaster* 180 MB
- 2001 2x ľudský genóm 3 GB (NIH, Celera)
- po 2001 Myš, potkan, kura, šimpanz, pes, makak,...

Ľudský genóm má 3 miliardy báz, ktoré z nich tvoria gény?

## Výpočtové hľadanie génov

Čo by sme chceli naprogramovať:



**Problém:** nie je žiadne jednoduché pravidlo odlišujúce gény. . .

Len čiastková informácia:

Exóny sa vyznačujú iným zložením a signálmi na okrajoch.

Spájame štatistickým modelom.

Hľadače génov sú užitočné, ale robia dosť veľa chýb.

## Experimentálne hľadanie génov

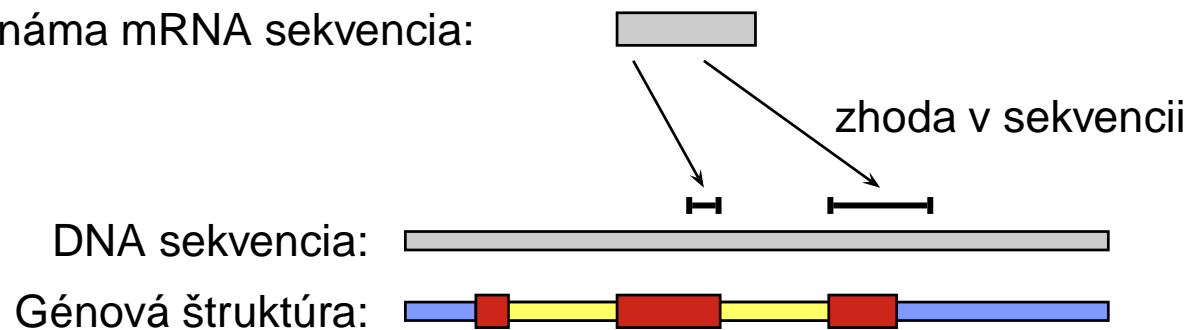
Sekvenujeme mRNA z bunky

Odkiaľ pochádza v genóme?

- problém hľadania rovnakých (podobných) reťazcov
- potrebujeme veľmi rýchle algoritmy

Zriedkavé mRNA je ťažké nájsť sekvenovaním.

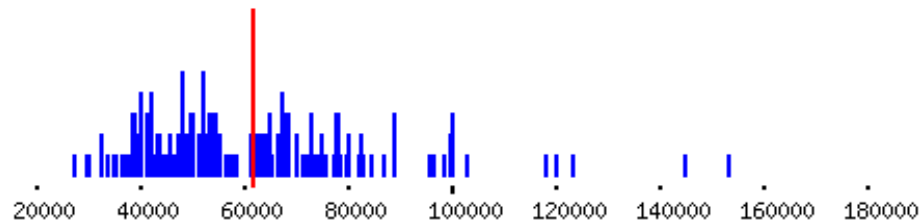
Známa mRNA sekvencia:





## Koľko génov má človek?

2000: Ewan Birney navrhuje stávkku na uhádnutie počtu ľudských génov  
165 účastníkov, cca \$1 200, odhady 28 000–140 000



2001: prvá verzia sekvencie ľudského genómu 30 000-40 000 génov

2004: “dokončená” sekvencia 20 000-25 000 génov

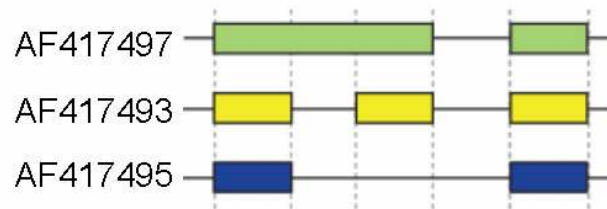
V súčasnosti sa predpokladá číslo bližšie 20 000

Plánované uzavretie stávky v 2003

Aj keď presný počet nebol známy, výhru rozdelili medzi najnižšie odhady

## Ďalšie komplikácie, čo je to vlastne gén?

- RNA gény, ktoré nepodliehajú translácii
- Alternatívny zostrih: z jedného génu rôzne proteíny



- Pseudogény: nefunkčné kópie génov
- Prekrývajúce sa gény
- Väčšina ľudského génu je transkribovaná (prepisovaná)

## Navrhovaná definícia génu, Mark Gerstein 2007

*The gene is a union of genomic sequences encoding a coherent set of potentially overlapping functional products.*

Ale nakoľko je ťažké overiť funkčnosť všetkých génov, môže trvať ešte dlho, kým budeme poznať ich presný počet.