HGVS-NOMENKLATURA

„Pravidla pro pojmenovávání sekvenčních variant“

HGVS-NOMENKLATURA

➢ **mezinárodní standard pro pojmenovávání sekvenčních variant nalezených v DNA, RNA či proteinech**

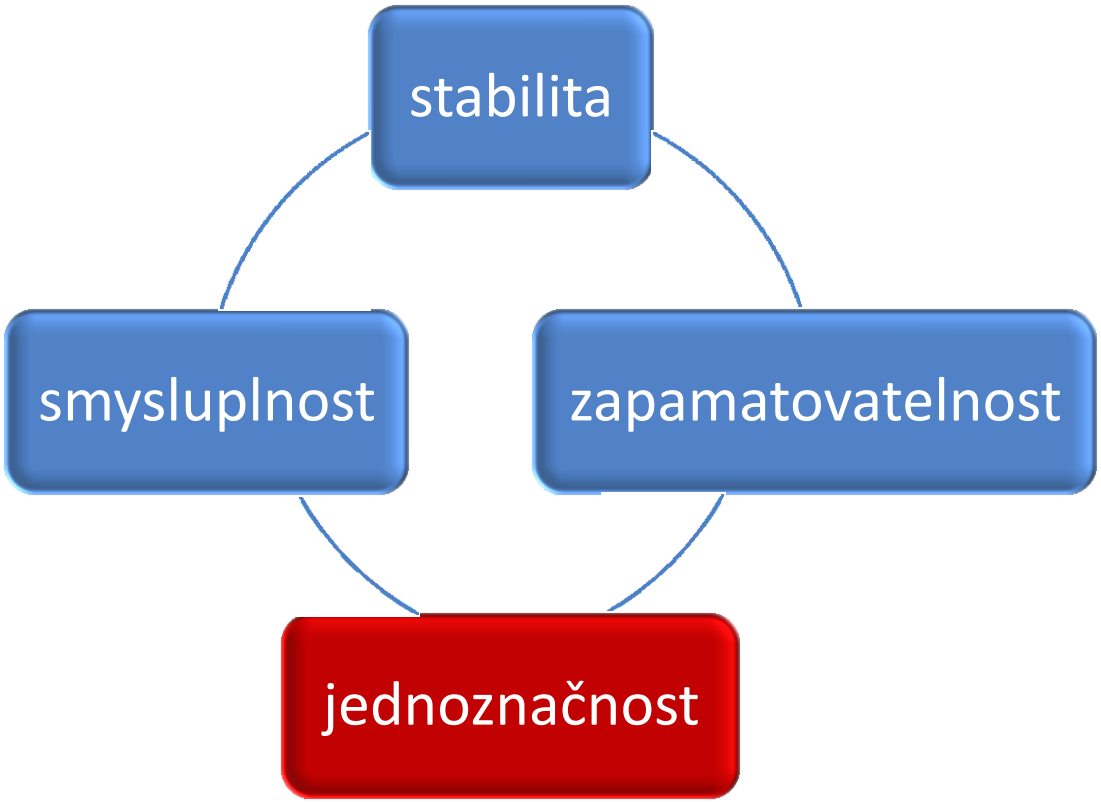
✓ autorizovaná Human Genome Variation Society (HGVS), the Human Variome Project (HVP), Human Genome Organization (HUGO)

publikace DEN DUNNEN, Johan T., Raymond DALGLEISH, Donna R. MAGLOTT, et al. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Human Mutation* [online]. 2016, **37**(6), 564-569 [cit. 2019-04-05]. DOI: 10.1002/humu.22981. ISSN 10597794. Dostupné z: http://doi.wiley.com/10.1002/humu.22981

online verze **http://www.HGVS.org/varnomen**

DŮVODY PRO ZAVEDENÍ

✓ **správné označení varianty zamezí nedorozuměním**



!vlastní pravidla vedou jen ke zmatkům a chybám!

HLAVNÍ PRAVIDLA

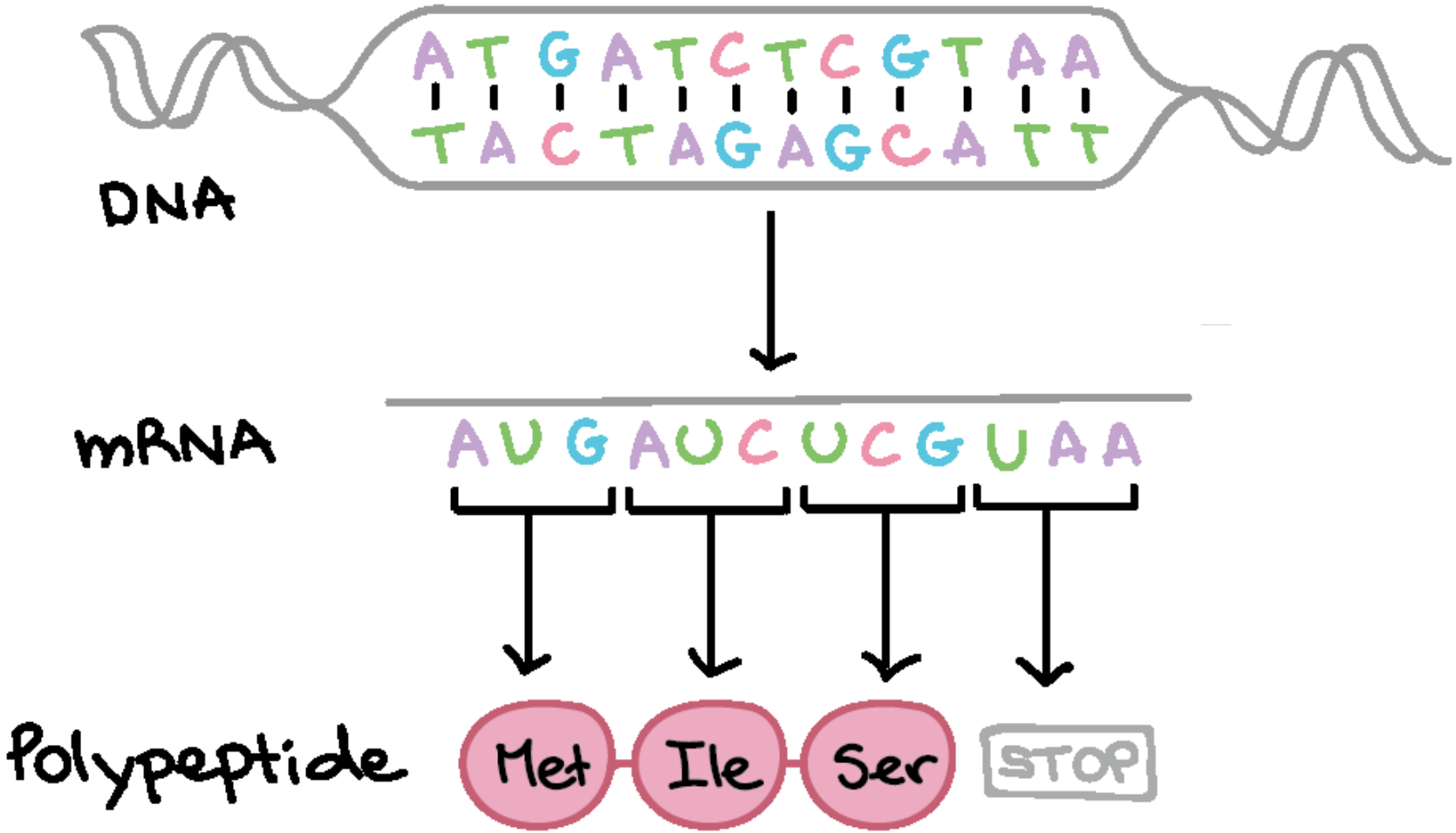
1) Pravidlo číslo jedna: užívej **oficiální symboly** (HGNC pro geny)!

2) Pravidlo číslo dvě: Vždy uváděj **referenční sekvenci**! (číslo i verzi)

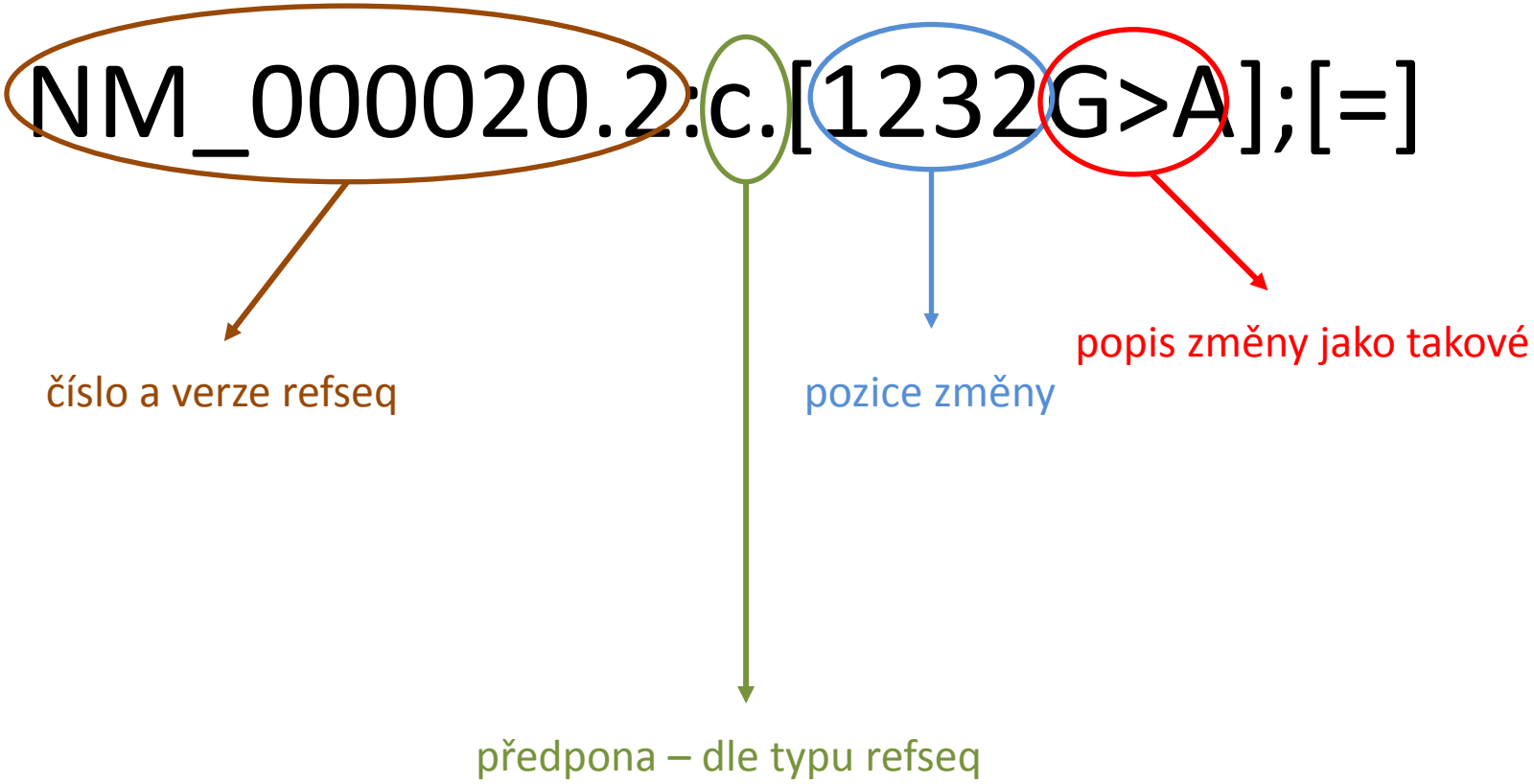
3) Pravidlo číslo tři: Vždy uveď **typ sekvence** na níž variantu pojmenováváš!

„Pravidlo číslo nula“:
Dodržuj doporučení a netvoř si vlastní pravidla!

KLÍČOVÉ MOLEKULY MOLEKULÁRNÍ GENETIKY



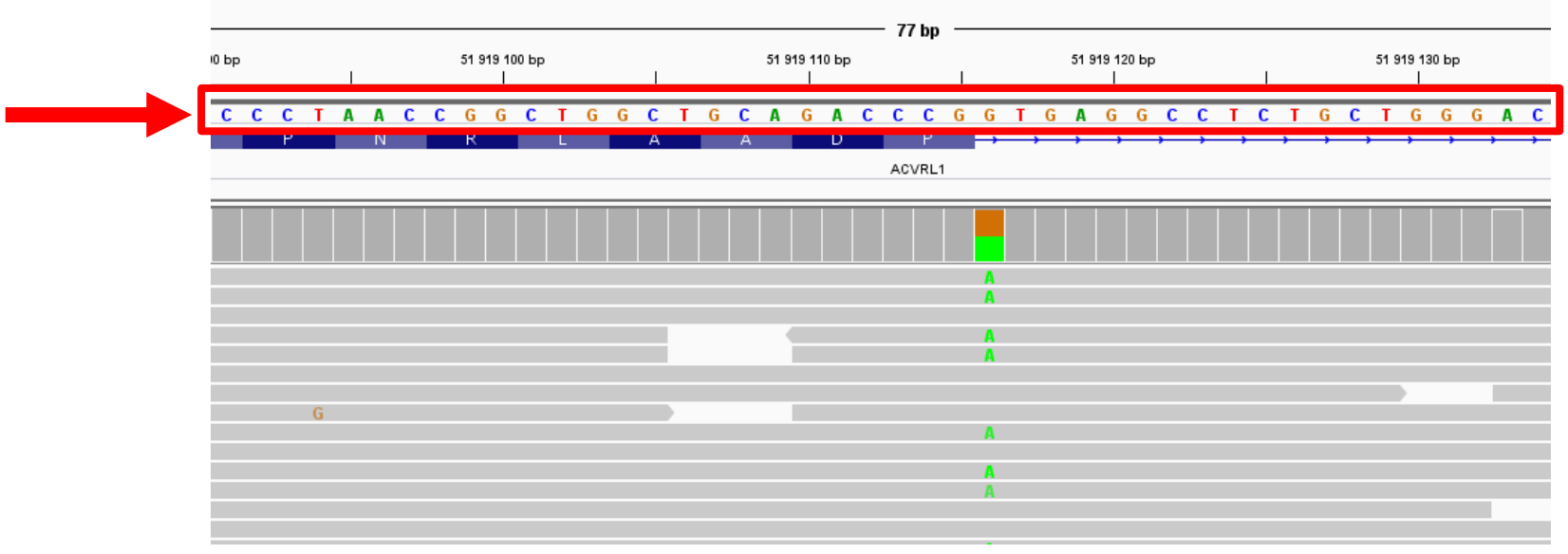
https://www.khanacademy.org/science/high-school-biology/hs-molecular-genetics/hs-rna-and-protein-synthesis/a/hs-rna-and-protein-synthesis-review

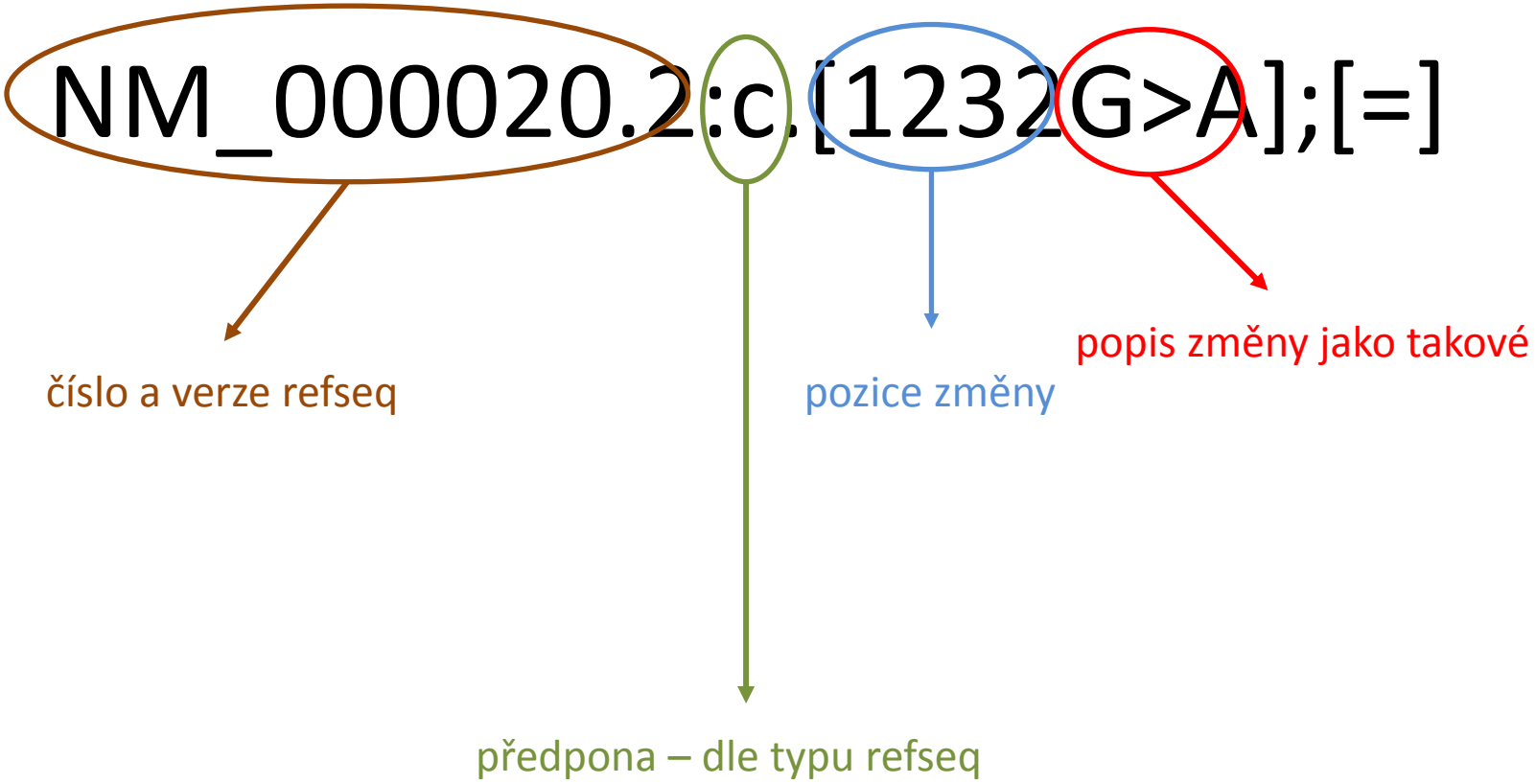

NM_000020.2:c.[1232G>A];[=]
číslo a verze refseq
předpona – dle typu refseq
pozice změny
popis změny jako takové

REFERENČNÍ SEKVENCE (RefSeq)

- veřejně dostupná, ověřená sekvence NK či proteinu
- představuje aktuálně nejspolehlivější zdroj sekvence považované za **standardní sekvenci** dané molekuly

➢ „ta sekvence, vůči které hledáme odchylky u vyšetřovaných vzorků“



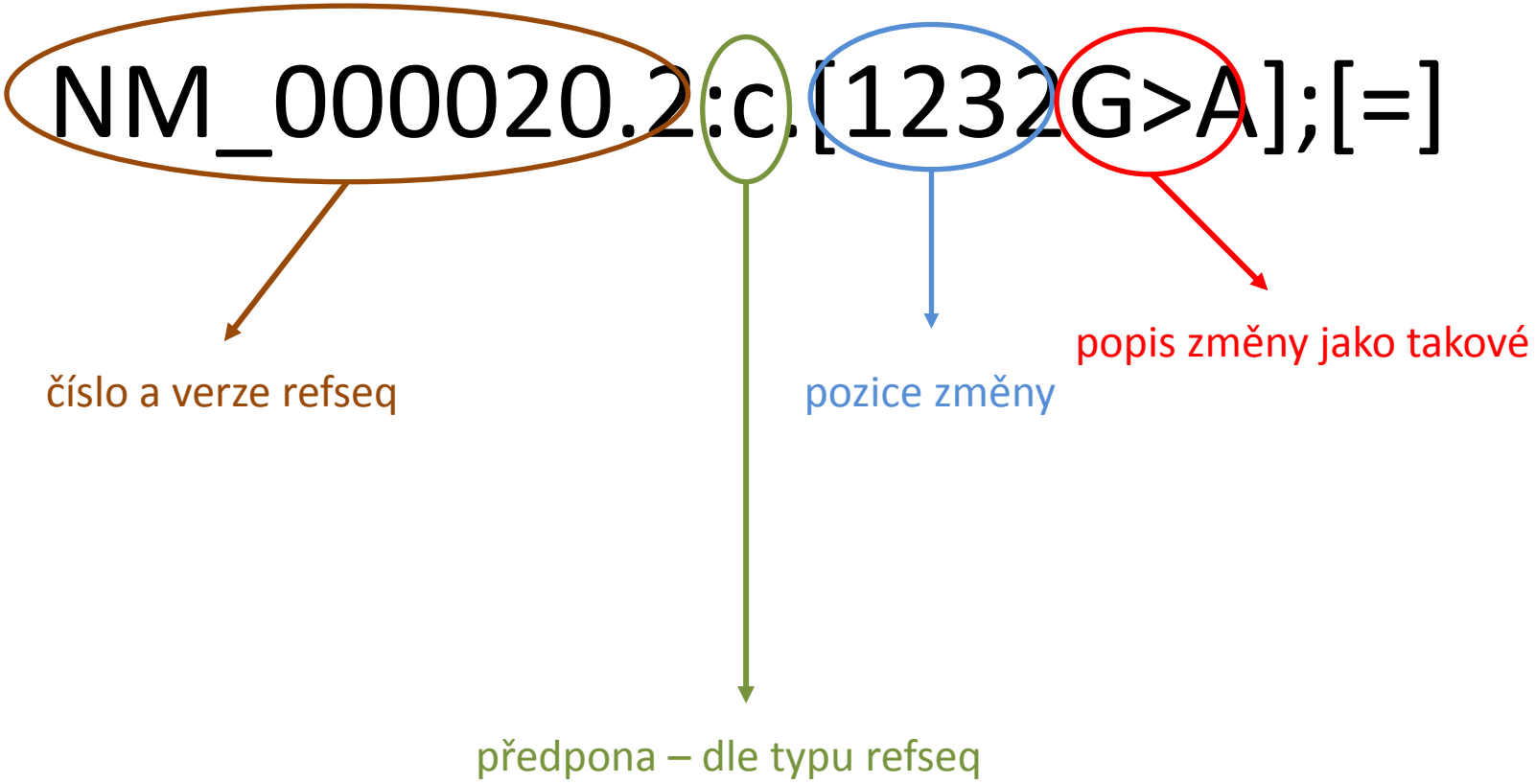

NM_000020.2:c.[1232G>A];[=]
číslo a verze refseq
předpona – dle typu refseq
pozice změny
popis změny jako takové

TYP REFERENČNÍ SEKVENCE

- **DNA**
  - A,T,C,G
  - **g. (genomická) vs c. (kódující)**
- RNA
  - a,u,c,g
  - r.
- PROTEIN (většinou odvozeno)
  - jednopísmenné/**třípísmenné** zkratky amk
  - p.

Amino acid	Three letter code	One letter code
Alanine	ala	A
Arginine	arg	R
Asparagine	asn	N
Aspartic acid	asp	D
Cysteine	cys	C
Glutamic acid	glu	E
Glutamine	gln	Q
Glycine	gly	G
Histidine*	his	H
Isoleucine*	ile	I
Leucine*	leu	L
Lysine*	lys	K
Methionine*	met	M
Phenylalanine*	phe	F
Proline	pro	P
Serine	ser	S
Threonine*	thr	T
Tryptophan*	trp	W
Tyrosine	tyr	Y
Valine*	val	V
Baronerocks.com	Baronerocks.com	Baronerocks.com

*Essential Amino Acids


NM_000020.2:c.[1232G>A];[=]
číslo a verze refseq
předpona – dle typu refseq
pozice změny
popis změny jako takové

ČÍSLOVÁNÍ REZIDUÍ

- **nukleotid 0 neexistuje!**

  -3, -2, -1, 1, 2, 3, 4

- **začátek 1**
  - u genomické sekvence 1 = první nukleotid daného souboru
  - u kódující DNA 1 = **A** start kodonu **A**TG

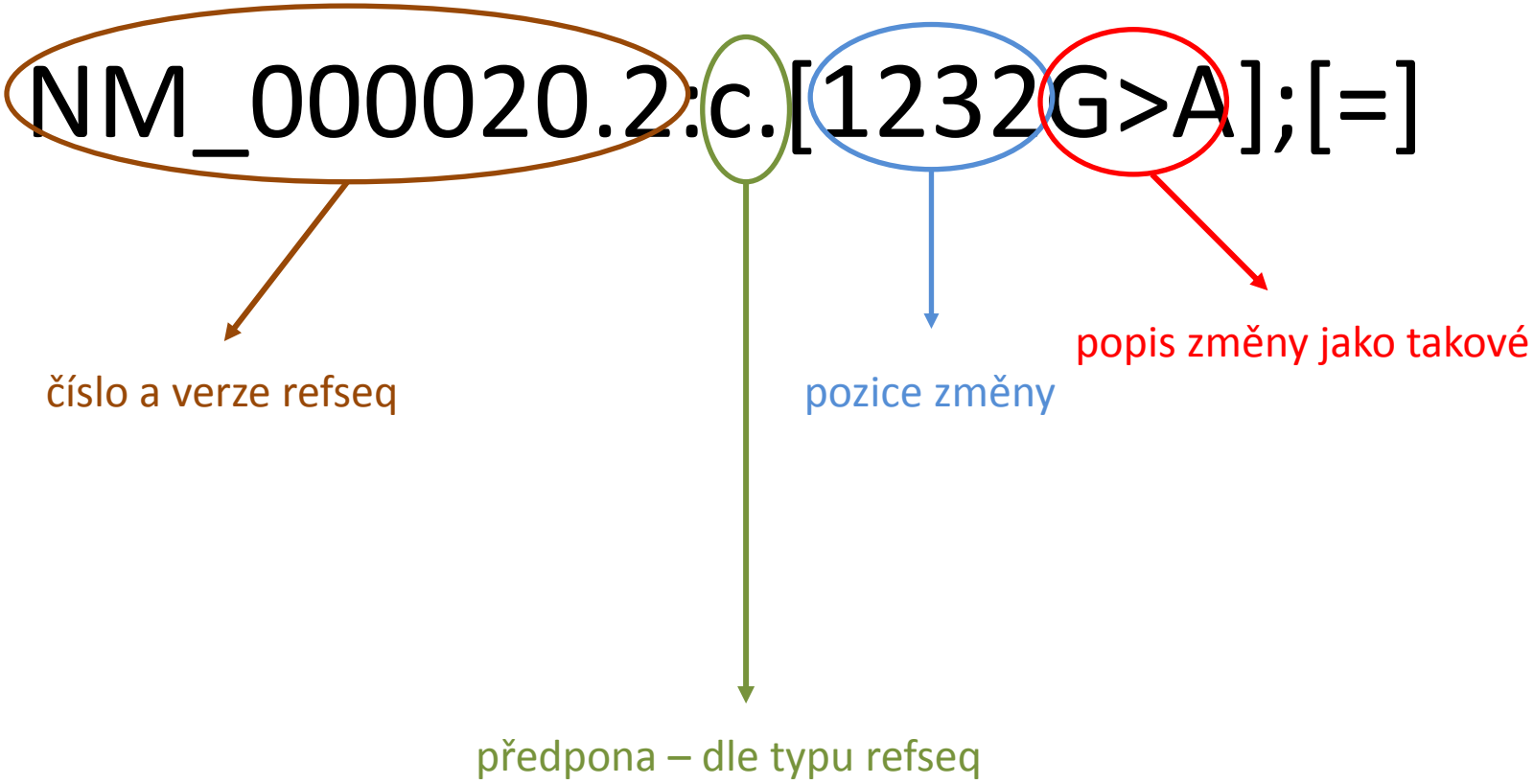
- **nepřekládané oblasti**
  - před start kodonem ATG přidáváme **mínus**
    -3, -2, -1, A, T, G
  - za terminačním kodonem přidáváme *****
    *STOP, *1, *2, *3*

ČÍSLOVÁNÍ REZIDUÍ

- **introny – spec. označení pouze u kódující DNA**
  - čísluje se ze strany, kde je exonu blíže
  - **plus/mínus** k poslednímu/prvnímu nukleotidu nejbližšího exonu

  → v polovině intronu se + láme na –

  c.589+15, c.590-5


NM_000020.2:c.[1232G>A];[=]
číslo a verze refseq
předpona – dle typu refseq
pozice změny
popis změny jako takové

UŽÍVANÉ ZNAČKY A ZKRATKY

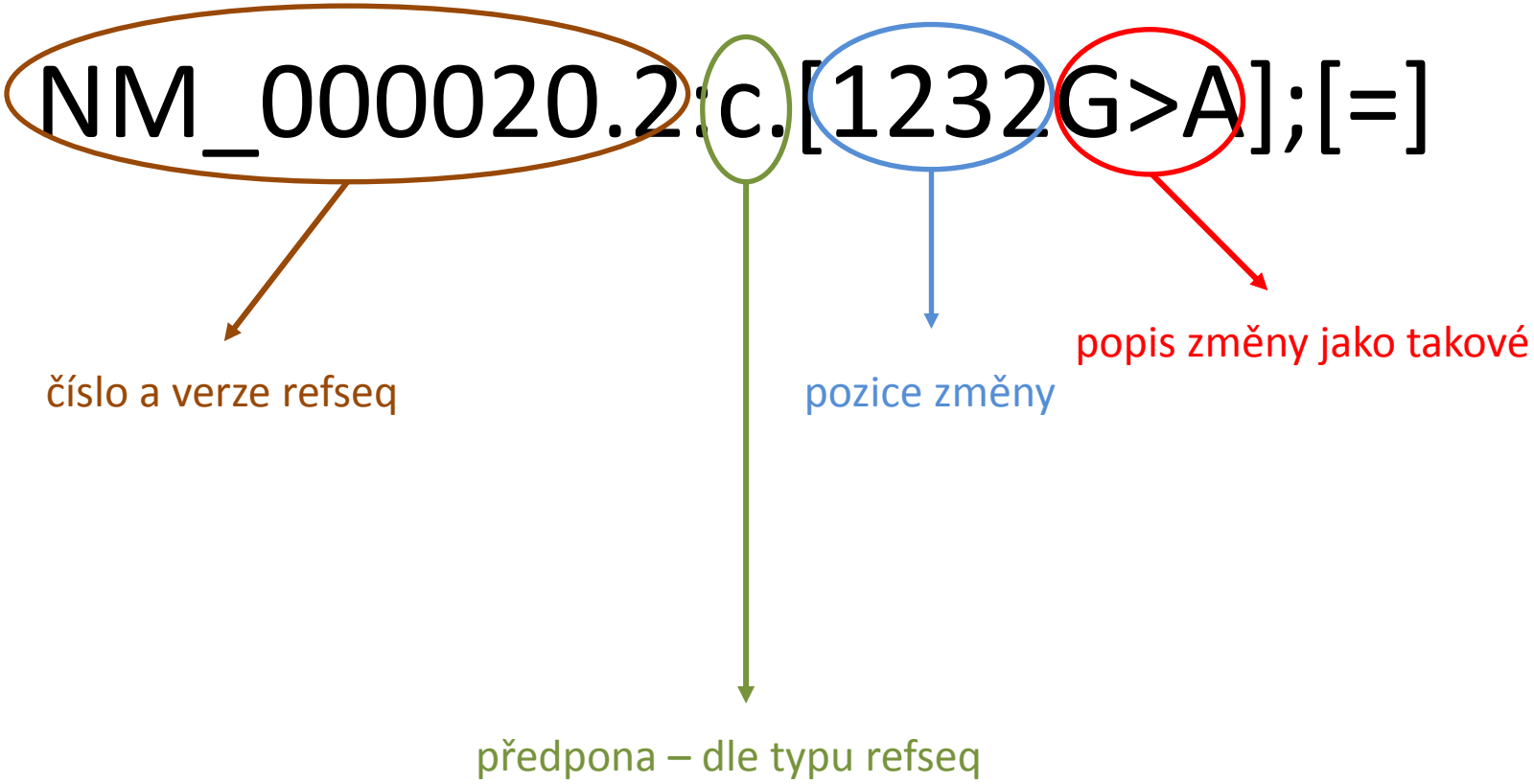
• **„+“**		u číslování nukleotidů
• **„-“**	**„mínus“**	u číslování nukleotidů
• **„*“**		u číslování nukleotidů
• **„_“**	**„až“**	označení rozpětí/rozsahu
• **„[ ]“**		pro alely
• **„ ; “**		odděluje různé varianty nacházející se na jedné alele nebo alely

UŽÍVANÉ ZNAČKY A ZKRATKY

- **„:“** odděluje ozn. refseq od zápisu varianty
- **„( )“** označení nejistoty nebo predikcí
- **„?“** označení neznámé pozice
- **„>“** označení substituce
- **„=„** označení testované pozice beze změny

UŽÍVANÉ ZNAČKY A ZKRATKY

- **„>“** substituce
- **„del“** delece
- **„dup“** duplikace
- **„ins“** inzerce
- **„inv“** inverze
- **„fs“** frame shift = posun čtecího rámce


NM_000020.2:c.[1232G>A];[=]
číslo a verze refseq
předpona – dle typu refseq
pozice změny
popis změny jako takové

VYBRANÉ NEJČASTĚJŠÍ TYPY ZMĚN

- **jednoduché**
  - substituce c.123A>G,
    „>“ neužívá se na úrovni proteinu p.(Gln182Leu)
  - delece c.123del (c.123delA), p.(Phe508del)
  - duplikace c.123dup (c.123dupA)
  - inzerce c.123_124insT – rozsah + sekvence!

- **komplexní**
  - indel c.123delinsGTAT

- **kombinace více variant**
  - změna na obou alelách c.[123A>G];[456delC]
  - více jak 1 varianta na 1 alele c.[123A>G;456delC];[=]

PŘÍKLADY: stop kodon -> protein

- **př. gen *CFTR* delece na DNA:** c.2250delT
  → posun čtecího rámce – změna amk + vznik **terminačního kodonu**
  → **p.(Arg751AlafsTer4)**

wt

```
2247 GCCTCGCATCAGCGTGATCAGCACTGGCCCCACGCTTCAGGCACGAAGGAGGCAGTCTGT   2306
      ProArgIleSerValIleSerIleGlyProThrLeuGlnAlaArgArgArgGlnSer
      750751752753754
```

c.2250delT

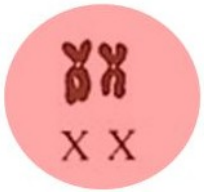
```
2247 GCCCGCATCAGCGTGATCAGCACTGGCCCCACGCTTCAGGCACGAAGGAGGCAGTCTGT   2306
      ProAlaSerAlaSTOP
      750751752753754
```

PŘÍKLADY: geny na chromozomu X

MUŽI

- muži – pouze 1 alela (1 chromozom X, 1 chromozom Y)
- c.[76A>G];[**0**] → **hemizygot**

ŽENY



- ženy – 2 alely (2 chromozomy X)
- c. [76A>G];[=]

PŘÍKLADY: více změn, cis/trans??

1) **detekovány 2 sekvenční varianty v poloze trans (na odlišných alelách)**

c.[2657+5G>A];[3528delC]

2) **detekovány 2 sekvenční varianty v poloze cis (na stejné alele), druhá alela je beze změny**

c.[12C>G;21A>G];[=]

PŘÍKLADY: více změn, cis/trans??

1) **detekovány 2 sekvenční varianty, zda jsou v poloze cis nebo trans nevíme**

c.2657+5G>A**(;)**3528delC

2) **detekovány 3 sekvenční varianty, u 2 víme, že jsou v poloze cis, u 3. nevíme**

c.**[**12C>G**;**21A>G**](;)**423dupG

PŘÍKLADY: MLPA

introny + rozsah + nejistota

1) **DELECE**

c.[(6438+1_6439-1)_(6614+1_6615-1)del];[0]

2) **DUPLIKACE**

c.[(649+1_650-1)_(2168+1_2169-1)dup];[0]

rada na závěr…

KDYŽ NEVÍŠ → HLEDEJ

http://varnomen.hgvs.org/