

# 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

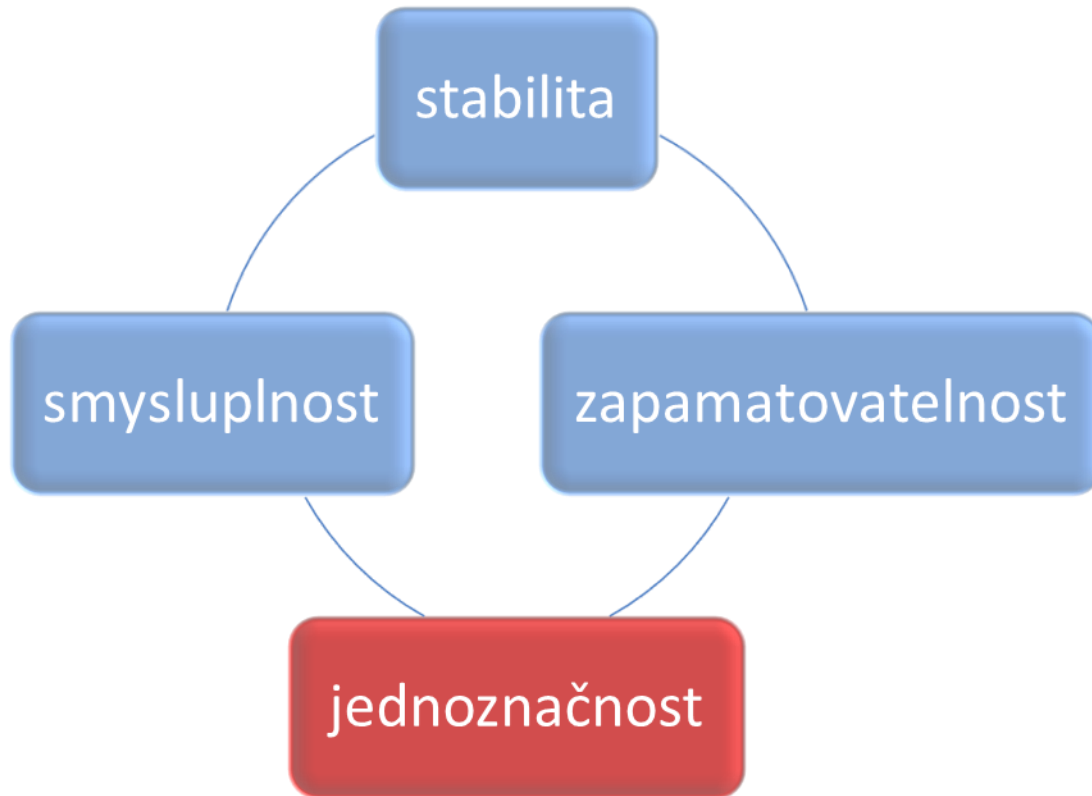
DEN DUNNEN, Johan T., Raymond DALGLEISH, Donna R. MAGLOTT, et al. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Human Mutation* [online]. 2016, **37**(6), 564-569 [cit. 2019-04-05]. DOI: 10.1002/humu.22981. ISSN 10597794. Dostupné z: <http://doi.wiley.com/10.1002/humu.22981>

online verze <http://www.HGVS.org/varnomen>



# DŮVODY PRO ZAVEDENÍ (a dodržování)

- ✓ **správné označení varianty zamezí nedorozuměním**



**! vlastní pravidla vedou ke zmatkům a chybám !**

# HLAVNÍ OBECNÁ PRAVIDLA

- 1) Užívat **oficiální symboly** (HGNC pro geny)
- 2) Uvádět **referenční sekvenci!** (číslo i verzi)
- 3) Uvádět **typ sekvence** na níž variantu pojmenováváme!

NM\_000020.2:c.[1232G>A];[=]

číslo a verze refseq

pozice změny

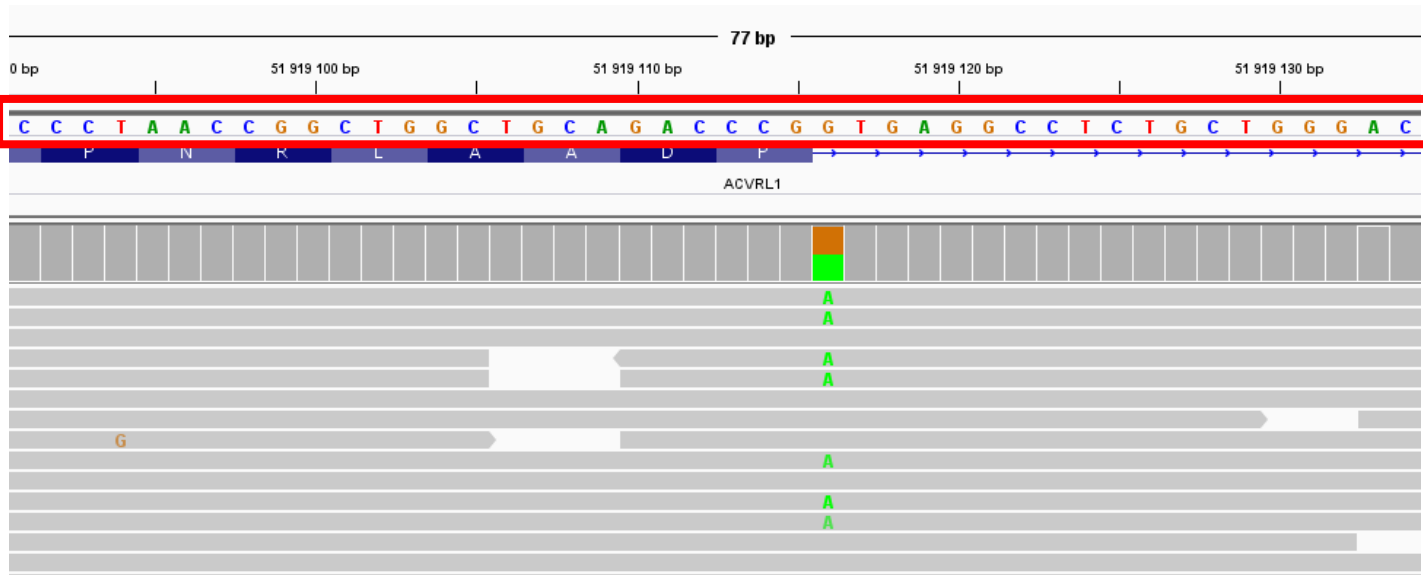
popis změny jako takové

předpona – dle typu refseq

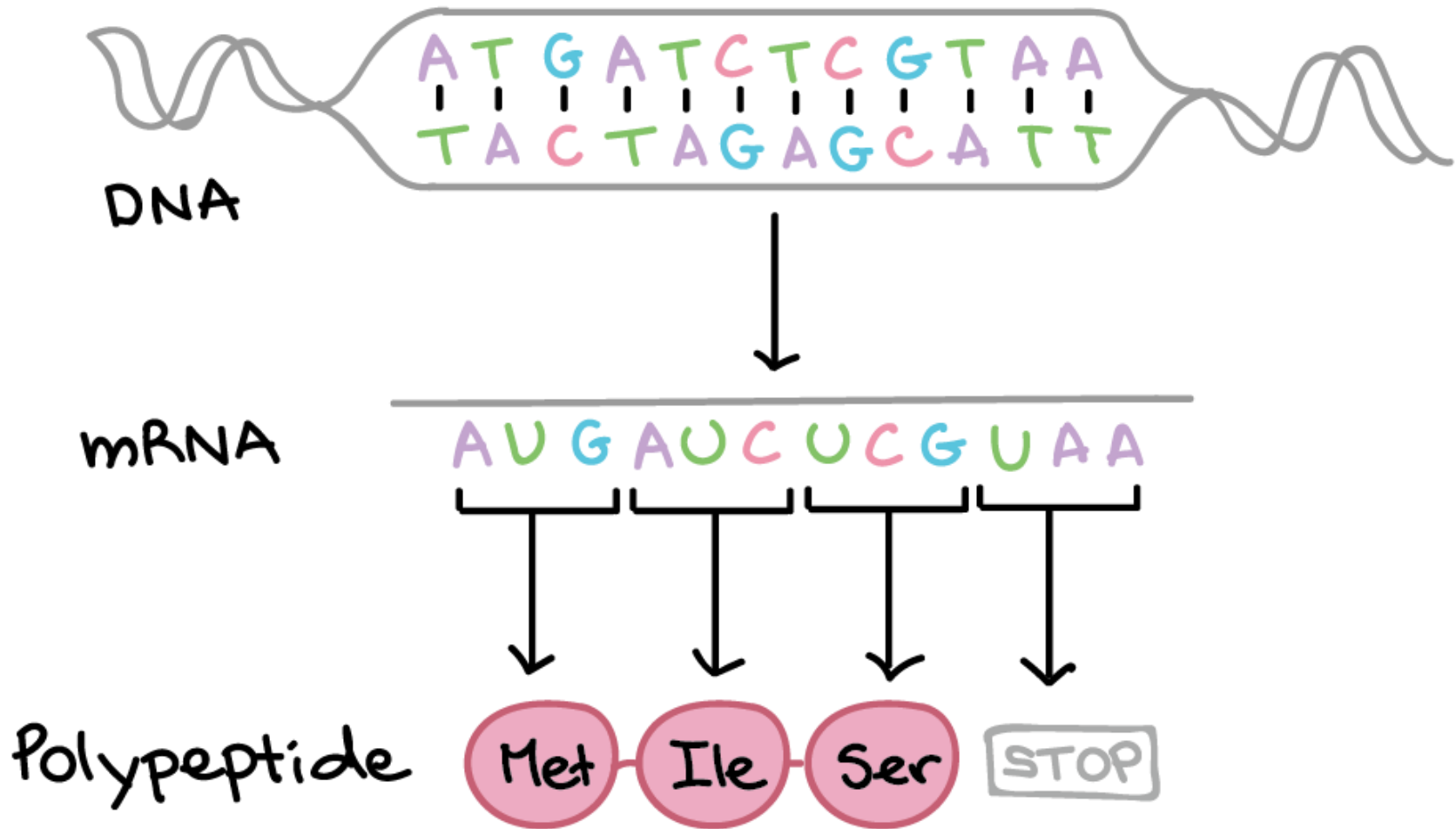
# REFERENČNÍ SEKVENCE (RefSeq)

- 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“



# KLÍČOVÉ MOLEKULY MOLEKULÁRNÍ GENETIKY



# TYP REFERENČNÍ SEKVENCE

- DNA
  - A,T,C,G
  - g. (genomická) vs c. (kódující)
- RNA
- PROTEIN (většinou odvozeno)

# GENOMICKÁ (g.)

# KÓDUJÍCÍ (c.)

>NM\_000020.2 Homo sapiens activin A receptor like type 1 (ACVRL1), transcript variant 1, mRNA

```
AGGAAACGGTTTATTAGGAGGGAGTGGTGGAGCTGGGCCAGGCAGGAAGACGCTGGAATAAGAAACATTT
TTGCTCCAGCCCCCATCCCAGTCCCAGGGAGGCTGCCGCGCCAGCTGCGCCGAGCGAGCCCCCTCCCCGGCT
CCAGCCCGGTCCGGGGCCCGCCCGGACCCAGCCCGCCGTCAGCGCTGGCGGTGCAACTGCGCCCGCG
CGGTGGAGGGGAGGTGGCCCCGGTCCGCCGAAGGCTAGCGCCCCCACCAGAGCGGGGCCCCAGAGGG
ACCATGACCTTGGGCTCCCCCAGGAAGGCCCTTCTGATGCTGCTGATGGCCCTGGTGACCCAGGGAGACC
CTGTGAAGCCGCTCTCGGGGCCCGCTGGTGACCTGCACGTGTGAGAGCCACATTGCAAGGGGCCTACCTG
CCGGGGGGCCTGGTGACAGTAGTGTGGTGCAGGGAGGGGAGGCACCCCCAGGAACATCGGGGCTGC
GGGAACCTGCACAGGGAGCTCTGCAGGGGGCGCCCCACCGAGTTCGTCAACCACTACTGCTGCGACAGCC
ACCTCTGCAACCACACGTGTCCCTGGTGTGGAGGCCACCCAACTCCTTCGGAGCAGCCGGGAACAGA
TGGCCAGCTGGCCCTGATCCTGGGCCCGCTGCTGGCCCTTGTGGCCCTGGTGGCCCTGGGTGTCTGGGC
CTGTGGCATGTCCGACGGAGGCAGGAGAAGCAGCGTGGCCCTGCACAGCGAGCTGGGAGAGTCCAAGTCTCA
TCCTGAAAGCATCTGAGCAGGGCGACAGCATGTTGGGGGACCTCCTGGACAGTGACTGCACCACAGGGAG
TAGGCTCAGGGCTCCCCCTTCTGGTGCAGAGGACAGTGGCACGGCAGGTTGCCCTGGTGGAGTGTGTGGGA
AAAGGCCGCTATGGCGAAGTGTGGCGGGGCTTGTGGCACGGTGAAGTGTGGCCGTCAAGATCTTCTCCT
CGAGGGATGAACAGTCTGGTTCGGGGAGACTGAGATCTATAACACAGTGTGCTCAGACACGACAACAT
CCTAGGCTTCATCGCCTCAGACATGACCTCCCGCAACTCGAGCACGCAGCTGTGGCTCATCAGCACTAC
CACGAGCACGGCTCCCTCTACGACTTTCTGCAGAGACAGACGCTGGAGCCCCATCTGGCTCTGAGGCTAG
CTGTGTCCGCGGCATGCGGCCCTGGCGCACCTGCACGTGGAGATCTTCGGTACACAGGGCACAACAGCCAT
TGCCACCAGCGACTTCAAGAGCCGCAATGTGCTGGTCAAGAGCAACCTGCAGTGTTCATCGCCGACCTG
GGCCTGGCTGTGATGCACTCAGAGGGCAGCGATTACCTGGACATCGGCAACAACCCGAGAGTGGGCACCA
AGCGGTACATGGCACCCGAGGTGTGGACGAGCAGATCCGCACGGACTGCTTTGAGTCTACAAGTGGAC
TGACATCTGGGCCTTTGGCCTGGTGTGTGGGAGATTGCCCGCCGGACCATCGTGAATGGCATCGTGGAG
GACTATAGACCACCCTTCTATGATGTGGTGCCTAATGACCCAGCTTTGAGGACATGAAGAAGGTGGTGT
GTGTGGATCAGCAGACCCCAACCTCCCTAACCGGCTGGCTGCAGACCCGGTCCCTCTCAGGCCTAGCTCA
GATGATGCGGGAGTGTGGTACCCAAACCCCTCTGCCGACTCACCGCGCTGCGGATCAAGAAGACACTA
CAAAAAATTAGCAACAGTCCAGAGAAGCCTAAAAGTATTCAATAGCCCAGGAGCACCTGATTCCTTTCTG
CCTGCAGGGGGCTGGGGGGGTGGGGGGCAGTGGATGGTGCCTATCTGGGTAGAGGTAGTGTGAGTGTGG
TGTGTGCTGGGGATGGGCAGCTGCGCCTGCCTGCTCGGCCCCCAGCCACCCAGCCAAAAATACAGCTGG
GCTGAAACCTGATCCCCTGCTGTCTGGCCTGCTCAAAGCGGCAGGCTCCCTGACGCCTGGCTCTCTCCCC
ACCCCTATGGCCAGCATGGTGCACCCCTTACCCTCCCGGGACAGGATGCAAAAAGAGGCTCCAGAGTCAG
AGTGCCAAGCCAGGGAATCCCAGTCCCAGACTCAGAGCCCGGGCCTGCACCTTGCCTCCCTGCCCTTGAATC
AACCCCACTGCCCCACCAGAGCTGCCAGGGTGGCACAGGGCCCTGTCCAGCCCTGGCACACACTTCCCT
GCCAGGCCTCAGCCTCTAGCATAAGCTCCAGAGAGCCAGGGCCCATCAGTTTCTCTCTGTGGATTTGTAT
CTCAGCTCCATGATGCCTTGGGCTTTCTGTCTCCTCAACAAGAGTGCAGCTTGTGAATGTCAGCTGCCT
GAGAGAGCTGGGGCCTGACTTACTAGGGCATTAAATCCTAAGAGGTCTACTGAGGTGTGGCAGGATCAC
AGGCCAGTGGAAAAAGGGCAGGTGAGATGGGCAAGGCCAGGACTTTCAGATTAAGTGGAGGATATCGA
GGCCAAAGCATGGCAGGGGGAAAGGTGAGTGGTGTCAAGAGACCCAGGCTGACCCCGGATGTTTGTCTCA
TGTGACAAAAGCAGGCCTGTCTCAGGACCTTTTCTTTCTTTTTCTTTTCTTTTGTGACACGGAG
TTTCGCTCTTGTGTCCAGGCTAGAGTGAATGGCATGATCCCAGCTCACCGCAACGTCTACCTCCCAGG
TTCAAATCATTCTTGCCTCAGACTCCCGAGTAGCTGGGATTACAGGCACATGCCACCATGCCTGGCTA
```



# 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

# 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 **ATG**

- nepřekládané oblasti

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

# GENOMICKÁ (g.)

# KÓDUJÍCÍ (c.)

>NM\_000020.2 Homo sapiens activin A receptor like type 1 (ACVRL1), transcript variant 1, mRNA

```
AGGAAACGGTTTATTAGGAGGGAGTGGTGGAGCTGGGCCAGGCAGGAAGACGCTGGAATAAGAAACATTT
TTGCTCCAGCCCCCATCCCAGTCCCAGGGAGGCTGCCGCGCCAGCTGCGCCGAGCGAGCCCCCTCCCCGGCT
CCAGCCCGGTCCGGGGCCCGCCCGGACCCAGCCCGCCGCTCCAGCGCTGGCGGTGCAACTGCGCCCGCG
CGGTGGAGGGGAGGTGGCCCCGGTCCGCCGAAGGCTAGCGCCCCCAGCCAGAGCGGGGCCCCAGAGGG
ACCATGACCTTGGGCTCCCCCAGGAAGGCCCTTCTGATGCTGCTGATGGCCCTGGTGACCCAGGGAGACC
CTGTGAAGCCGCTCTCGGGGCCCGCTGGTGACCTGCACGTGTGAGAGCCACATTGCAAGGGGCCTACCTG
CCGGGGGGCCTGGTGACAGTAGTGTGGTGCAGGGAGGGGAGGCACCCCCAGGAACATCGGGGCTGC
GGGAACCTGCACAGGGAGCTCTGCAGGGGGCGCCCCACCGAGTTCGTCAACCACTACTGCTGCGACAGCC
ACCTCTGCAACCACACGTGTCCCTGGTGTGGAGGCCACCCAACTCCTTCGGAGCAGCCGGGAACAGA
TGGCCAGCTGGCCCTGATCCTGGGCCCGCTGCTGGCCCTTGTGGCCCTGGTGGCCCTGGGTGTCTGGGC
CTGTGGCATGTCCGACGGAGGCAGGAGAAGCAGCGTGGCCCTGCACAGCGAGCTGGGAGAGTCCAAGTCTCA
TCCTGAAAGCATCTGAGCAGGGCGACAGCATGTTGGGGGACCTCCTGGACAGTGACTGCACCACAGGGAG
TAGGCTCAGGGCTCCCCTTCTGGTGCAGAGGACAGTGGCACGGCAGGTTGCCCTGGTGGAGTGTGTGGGA
AAAGGCCGCTATGGCGAAGTGTGGCGGGGCTTGTGGCACGGTGAAGTGTGGCCGTCAAGATCTTCTCCT
CGAGGGATGAACAGTCTGGTTCGGGGAGACTGAGATCTATAACACAGTGTGCTCAGACACGACAACAT
CCTAGGCTTCATCGCCTCAGACATGACCTCCCGCAACTCGAGCACGCAGCTGTGGCTCATCAGCACTAC
CACGAGCACGGCTCCCTCTACGACTTTCTGCAGAGACAGACGCTGGAGCCCCATCTGGCTCTGAGGCTAG
CTGTGTCCGCGGCATGCGGCCCTGGCGCACCTGCACGTGGAGATCTTCGGTACACAGGGCACAACAGCCAT
TGCCACCAGCGACTTCAAGAGCCGCAATGTGCTGGTCAAGAGCAACCTGCAGTGTTCATCGCCGACCTG
GGCCTGGCTGTGATGCACTCAGAGGGCAGCGATTACCTGGACATCGGCAACAACCCGAGAGTGGGCACCA
AGCGGTACATGGCACCCGAGGTGTGGACGAGCAGATCCGCACGGACTGCTTTGAGTCTCAAAGTGGAC
TGACATCTGGGCCTTTGGCCTGGTGTGTGGGAGATTGCCCGCCGGACCATCGTGAATGGCATCGTGGAG
GACTATAGACCACCCTTCTATGATGTGGTGCCTAATGACCCAGCTTTGAGGACATGAAGAAGGTGGTGT
GTGTGGATCAGCAGACCCCAACCTCCCTAACCCGGCTGGCTGCAGACCCGGTCCCTCTCAGGCCTAGCTCA
GATGATGCGGGAGTGTGGTACCCAAACCCCTCTGCCGACTCACCGCGCTGCGGATCAAGAAGACACTA
CAAAAAATTAGCAACAGTCCAGAGAAGCCTAAAAGTATTCAATAGCCCAGGAGCACCTGATTCTTTCTG
CCTGCAGGGGGCTGGGGGGGTGGGGGGCAGTGGATGGTGCCTATCTGGGTAGAGGTAGTGTGAGTGTGG
TGTGTGCTGGGGATGGGCAGCTGCGCCTGCCTGCTCGGCCCCCAGCCACCCAGCCAAAAATACAGCTGG
GCTGAAACCTGATCCCCTGCTGTCTGGCCTGCTCAAAGCGGCAGGCTCCCTGACGCCTGGCTCTCTCCCC
ACCCCTATGGCCAGCATGGTGCACCCCTTACCCTCCCGGGACAGGATGCAAAAAGAGGCTCCAGAGTCAG
AGTGCCAAGCCAGGGAATCCCAGTCCCAGACTCAGAGCCCGGGCCTGCACCTTGCCTCCCTGCCCTTGAATC
AACCCCACTGCCCCACCAGAGCTGCCAGGGTGGCACAGGGCCCTGTCCAGCCCTGGCACACACTTCCCT
GCCAGGCCTCAGCCTCTAGCATAAGCTCCAGAGAGCCAGGGCCCATCAGTTTCTCTCTGTGGATTTGTAT
CTCAGCTCCATGATGCCTTGGGCTTTCTGTCTCCTCAACAAGAGTGCAGCTTGTGAATGTCAGCTGCCT
GAGAGAGCTGGGGCCTGACTTACTAGGGCATTAAATCCTAAGAGGTCTACTGAGGTGTGGCAGGATCAC
AGGCCAGTGGAAAAAGGGCAGGTGAGATGGGCAAGGCCAGGACTTTCAGATTAAGTGGAGGATATCGA
GGCCAAAGCATGGCAGGGGGAAGGTGAGTGGTGTCAAGAGACCCAGGTCTGACCCCGGATGTTTGTCTCA
TGTGACAAAAGCAGGCCTGTCTCAGGACCTTTCTTTTCTTTTTCTTTTCTTTTGTGACACGGAG
TTTCGCTCTTGTGTCCAGGCTAGAGTGAATGGCATGATCCCAGCTCACCGCAACGTCTACCTCCCAGG
TTCAAATCATTCTTGCCTCAGACTCCCGAGTAGCTGGGATTACAGGCACATGCCACCATGCCTGGCTA
```

# 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*



# 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

# VYBRANÉ NEJČASTĚJŠÍ TYPY ZMĚN

- **jednoduché**

- substituce c.123A>G,  
*„>“ nežívá se na úrovni proteinu* p.(Gln182Leu)
- delece c.123del (c.123delA), p.(Phe508del)
- duplikace c.123dup (c.123dupA)
- inserce 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.(Arg751Alafs**Ter4**)

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

*rada na závěr...*

**POKUD NEVÍTE → HLEDEJTE**



**<http://varnomen.hgvs.org/>**