Mutation nomenclature

recommendations for the description of DNA changes

http://www.HGVS.org/mutnomen/
HUGO-MDI initiative

HGVS
HUMAN GENOME
VARIATION SOCIETY

© *JT den Dunnen*

Definitions

- **prevent confusion**
  - ***mutation***
    - ***- change***
    - ***- disease-causing change***
  - ***polymorphism***
    - ***- change in >1% population***
    - ***- not disease causing change***

- **better**
  - ***neutral terms***
  - ***sequence variant***
  - ***allelic variant***
  - ***alteration***
  - ***CNV*** ***(Copy Number Variant)***
  - ***SNV*** ***(not SNP)***

HGVS
HUMAN GENOME VARIATION SOCIETY

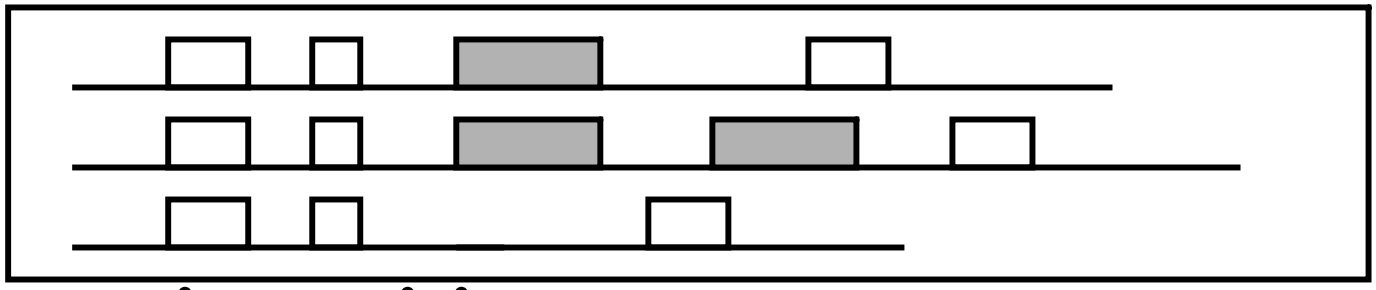
© JT den Dunnen

Possible variants[2]

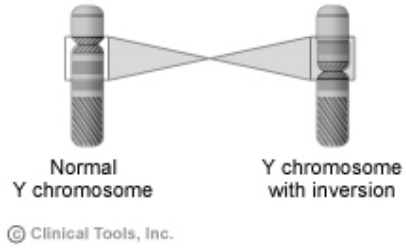
- **change in sequence**

```
ACATCAGGAGAAGATGTTC GAGACTTTGCCA
ACATCAGGAGAAGATGTTT GAGACTTTGCCA
ACATCAGGAGAAGATGTT  GAGACTTTGCCA
ACATCAGGAGAAGATGTTCCGAGACTTTGCCA
```

- **change in amount**



- **change in position**

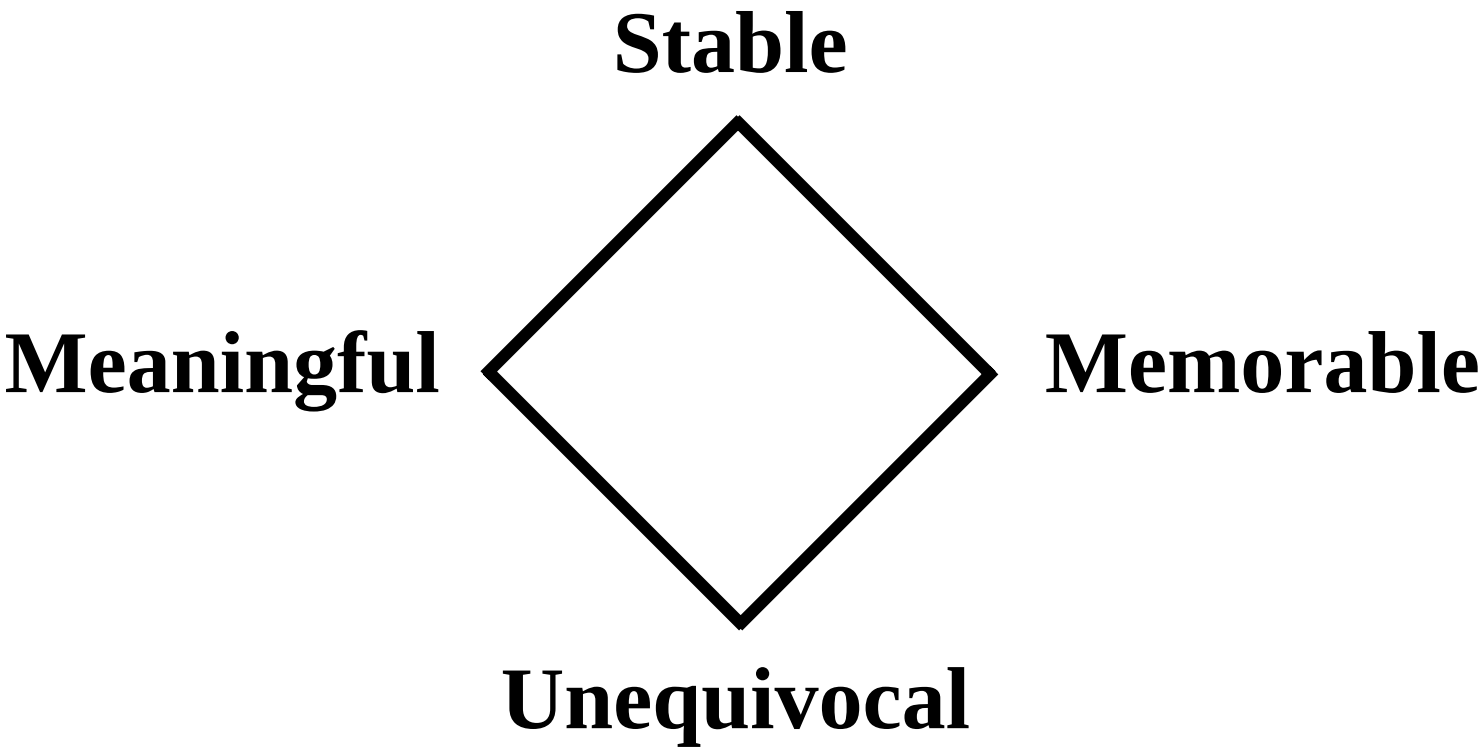


HGVS
HUMAN GENOME
VARIATION SOCIETY

© JT den Dunnen

Nomenclature

(describing DNA variants)



HGVS
HUMAN GENOME VARIATION SOCIETY

© *JT den Dunnen*

Mutation nomenclature 2

on behalf of HUGO MDI / HGVS

HUMAN MUTATION 15:7–12 (2000)

MDI SPECIAL ARTICLE

Mutation Nomenclature Extensions and Suggestions to Describe Complex Mutations: A Discussion

Johan T. den Dunnen[1*] and Stylianos E. Antonarakis[2*]
[1]*MGC-Department of Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands*
[2]*Division of Medical Genetics, University of Geneva Medical School, Geneva, Switzerland*

Consistent gene mutation nomenclature is essential for efficient and accurate reporting, testing, and curation of the growing number of disease mutations and useful polymorphisms being discovered in the human genome. While a codified mutation nomenclature system for simple DNA lesions has now been adopted broadly by the medical genetics community, it is inherently difficult to represent complex mutations in a unified manner. In this article, suggestions are presented for reporting just such complex mutations. Hum Mutat 15:7–12, 2000. © 2000 Wiley-Liss, Inc.

KEY WORDS: complex mutation; mutation detection; mutation database; nomenclature; MDI

HGVS
HUMAN GENOME VARIATION SOCIETY

© *JT den Dunnen*

Nomenclature for the description of sequence variations

(last modified February 12, 2007)

Prepared by Johan den Dunnen

Contents

Recent additions - ***(opinions please)***

- NEW uncertainties in descriptions
  (incl. arrayCGH, SNP-array, Southern blot data)

Current recommendations

- Introduction
- General recommendations
- Specific recommendations
  - DNA-level
  - RNA-level
  - Protein-level

Checklist - ***(online help when writing publications)***

Example descriptions

- DNA
- RNA
- Protein
- Quick Reference *(simple examples)*
- Characters used
- Codons and amino acids

Discussions

- general
- reference sequence
- nucleotide numbering

FAQ (frequently asked questions)

HGVS
HUMAN GENOME VARIATION SOCIETY

http://www.HGVS.org/

Follow the recommendations

when you disagree, ***start a debate*** – do not use private rules, this only causes confusion

DNA, RNA, protein

- **unique descriptions**
  prevent confusion
- **DNA**
  A, G, C, T
  c.957A>T
- **RNA** *(deduced mostly)*
  a, g, c, u
  r.957a>u
- **protein** *(deduced only)*
  three / one letter amino acid code
  X = stop codon
  p.Glu78Gln

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen

Reference Sequence

- **use official HGNC gene symbols**
- **set the residues and numbering**
  NM_012654.3 : *c.957A>T*
- **provide database reference**
  covering complete sequence
    largest transcript
  accession.version number
    e.g. NM_012654.3
  RefSeq database (curated seqs)
- **indicate type of Reference Sequence**
  DNA
    coding DNA c.
    genomic g.
    mitochondrial m.
  RNA r.
  protein p.

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen

coding DNA or genomic ?

- **human genome sequence**
  - ***complete***
  - ***covers all transcripts***
    - ***different promoters, splice variants,***
    - ***different polyA-sites, …***
  - ***but***
    - ***g.21,895,321_21,895,325del***
      - ***NT_035218.23 is 70 Mb file***
    - ***new builds follow each other regularly***
- **coding DNA**
  - ***does not cover all variants***
  - ***gives a clue towards position***

HGVS
HUMAN GENOME VARIATION SOCIETY

© *JT den Dunnen*

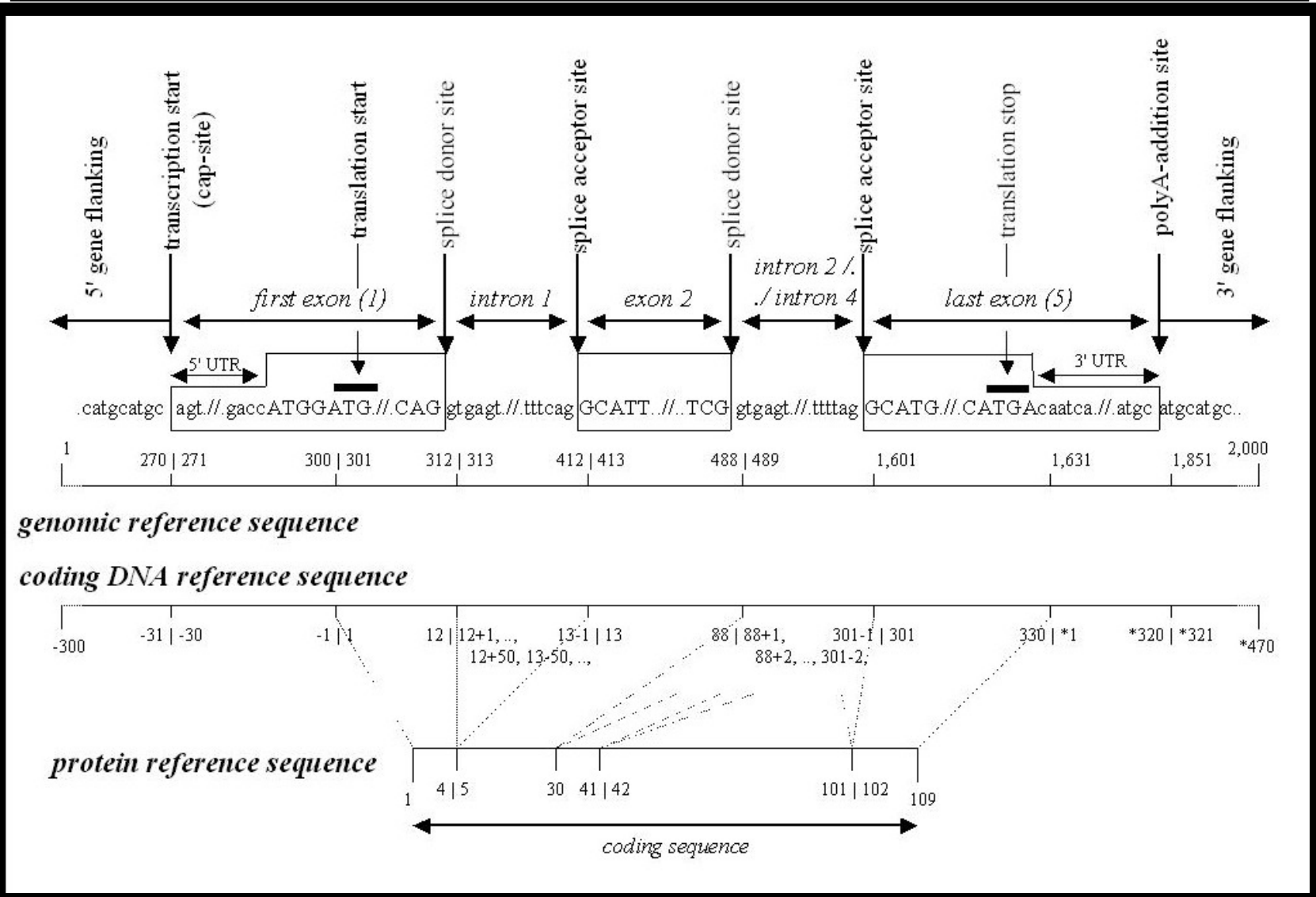
Residue numbering[1]

- **genomic reference sequence**
  - ***from first to last nucleotide***
    - ***1 to 13,562***
    - ***not +, - or other signs***
- **coding DNA**
  - ***from first of ATG to last of stop***
    - ***1 to 963***
  - ***- (minus) when 5' of ATG***
    - ***incl. 5' of cap site***
  - **** (asterisk) when 3'of stop codon***
    - ***incl. 3' of polyA-addition site***
  - ***intron***
    - ***362+1, 362+2, … start***
    - ***…, 363-2, 363-1 end***

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen

Reference Sequence



Residue numbering[2]

- **RNA** *(deduced mostly)*

  like coding DNA

- **protein** *(deduced only)*

  from first to last amino acid

  1 to 321

HGVS
HUMAN GENOME VARIATION SOCIETY

© *JT den Dunnen*

Types of variation

- **simple**
  - ***substitution*** ***c.123A>G***
  - ***deletion*** ***c.123delA***
  - ***duplication*** ***c.123dupA***
  - ***insertion*** ***c.123_124insC***
  - ***other***
    - ***inversion, translocation, transposition***
- **complex**
- **combinations**
  - ***two alleles*** ***c.[123A>G]+[456C>T]***
  - ***>1 per allele*** ***c.[123A>G; 456C>T]***

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen

Substitution

- **substitution designated by ">"**
  not used on protein level

- **examples**

cDNA	***c.546A>T***
	(NM_012654.3 : c.546A>T)
genomic	***g.54786A>G***
protein	***p.Gln78His***

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen

Deletion

- **deletion**

  designated by "del"
  range indicated by "_"

- **examples**

  c.546delT
  c.546del

  c.586_591del
  c.586_591delTGGTCA or c.586_591del6

  c.781-?_1392+?del
  = exon 3 to 6 deletion, breakpoint not sequenced

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen

Duplication

- **duplication**
  designated by "dup"
  range indicated by "_"

- **examples**

  c.546dupT
  c.546dup

  c.586_591dup
  c.586_591dup6 or c.586_591dupTGGTCA
  do not describe as insertion

  c.781-?_1392+?dup
  = exon 3 to 6 duplication, breakpoint not sequenced

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen

Insertion

- **insertion**
  - ***designated by "ins"***
  - ***range indicated by "_"***
  - ***give inserted sequence***

- **examples**

  c.546_547insT
  - ***NOT c.546insT***

  c.1086_1087insGCGTGA

  c.1086_1087insAB567429.2:g.34_12,567
  - ***when large insert submit to database and***
  - ***give database accession.version number***

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen

Inversion

- **inversion**
  designated by "inv"
  range indicated by "_"

- **example**

  c.546_2031inv

HGVS
HUMAN GENOME VARIATION SOCIETY

© *JT den Dunnen*

Conversion

- **conversion**
  - ***designated by "con"***
  - ***range indicated by "_"***

- **examples**

  c.546_657con917_1028

  c.546_2031conNM_023541.2:c.549_2034

HGVS
HUMAN GENOME VARIATION SOCIETY

© *JT den Dunnen*

Translocation

- **translocation**
  designated by "t"
  range indicated by "_"

- **examples**

  t(X;4) (p21.2;q35) (c.857+101_857+102)

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen

Repeated sequences

- **mono-nucleotide stretches**
  g.8932A(18_23)

  c.345+28T(18_23)
  alleles 345+28T[18]+[21]

- **di-nucleotide stretches**
  c.7TG(3_6)

- **larger**
  g.532_3886(20_45)
  3.3 Kb repeat

HGVS
HUMAN GENOME VARIATION SOCIETY

© *JT den Dunnen*

SNP's

- **SNP's**

  clear identifier for each SNP

  AC043217.2: g.78654 C>G

  rs2306220: A>G
  dbSNP entry

  DXS1219: g.117CA(18_26)
  alleles g.117CA[20]+[24]

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen

Specific codes

- **codes used**

+, -, *	
>	***substitution***
_	***range***
;	***more changes in one allele***
,	***more transcripts / mosaicism***
( )	***uncertain***
[ ]	***allele***
del	***deletion***
dup	***duplication***
ins	***insertion***
inv	***inversion***
con	***conversion***
ext	***extension***
X	***stop codon***
fsX	***frame shift***
o	***opposite strand***
t	***translocation***

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen

Changes in 2 alleles

- **recessive disease**
  report combination of changes
- **allele**
  indicated by "[ ]"
  separated by "+"
- **examples**

  c.[546C>T]+[1398delT]
  or [c.546C>T]+[c.1398delT]
  c.[546C>T]+[?]
  c.[546C>T]+[=]

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen

Alleles

- **recessive disease**
  - ***c.[546C>T]+[2398delT]***
  - ***c.[546C>T]+[?]***
- **more changes in 1 allele**
  - ***c.[546C>T; 2398delT]***
- **alleles unknown**
  - ***c.[246C>T(+)2398delT]***
    - ***parents not analysed***
- **more variants from 1 allele**
  - ***mosaicism - c.[=, 546C>T]***
  - ***two transcripts - r.[=, 512_636del]***

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen

Frame shifts

- **short form** *(sufficient)*

  p.Arg83fs

- **long from** *(more detail)*

  p.Arg83SerfsX15

  do not try to include changes at DNA level

  indicate

  first amino acid changed
  position
  first changed amino acid
  length shifted frame
  (from first changed to X incl.)
  do not describe del, dup, ins, etc.

HGVS
HUMAN GENOME VARIATION SOCIETY

© *JT den Dunnen*

Complex

- **deletion / insertions**

  "indel"

  c.1166_1177delinsAGT

- **descriptions nay become complex**

  only an expert understands the "code"
  consider database submission

  description: AC111747.1

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen

Publications

Published

patient#	protein	DNA	Remarks
QL43.2	Tyr151Asp	451T>C	parents unrelated
…	…	…	…
QL43.2	frame shift	976delA	parents unrelated

Correct

patient#	DNA	RNA	protein	Remarks
QL43.2	c.[451T>C(+) 976delA]	? or NA	? *(p.[Tyr151Asp (+)Phe326fs])*	parents unrelated

HGVS
HUMAN GENOME
VARIATION SOCIETY

© *JT den Dunnen*

Problematic[1]

- **coding DNA Reference Sequence**

no reference to genomic sequence

so, c.5477-137A>G ?

no reference to intron numbering

so, c.IVS20-1G>A ?

new exon identified ***(CFTR, SMN1)***

exons 12, 13, 13A, 14

exons –2, -1, 1, 2, 3, …

difference with genome browser

always from 1 to end

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen

Gene structure ?

- **how are exons numbered ?**
  - ***often confusing***
- **how are introns numbered ?**
  - ***SMA***
    - ***SMN1 exon 7***
  - ***exon 0 / intron 0***
    - ***where is exon 0 in a gene ?***
      - ***…, -2, -1, 0, 1, 2, … or***
      - ***…, -2, -1, 1, 2, …***

>> changes difficult to find

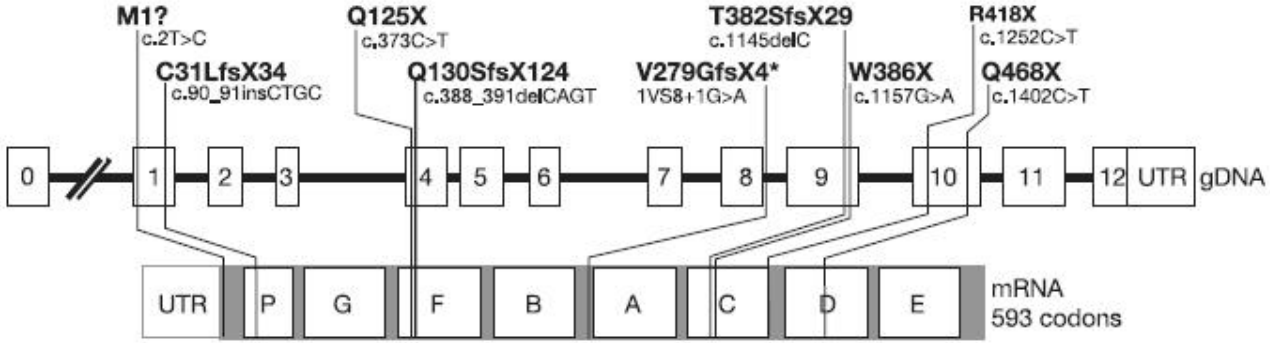
non-expert, student, …

HGVS
HUMAN GENOME VARIATION SOCIETY

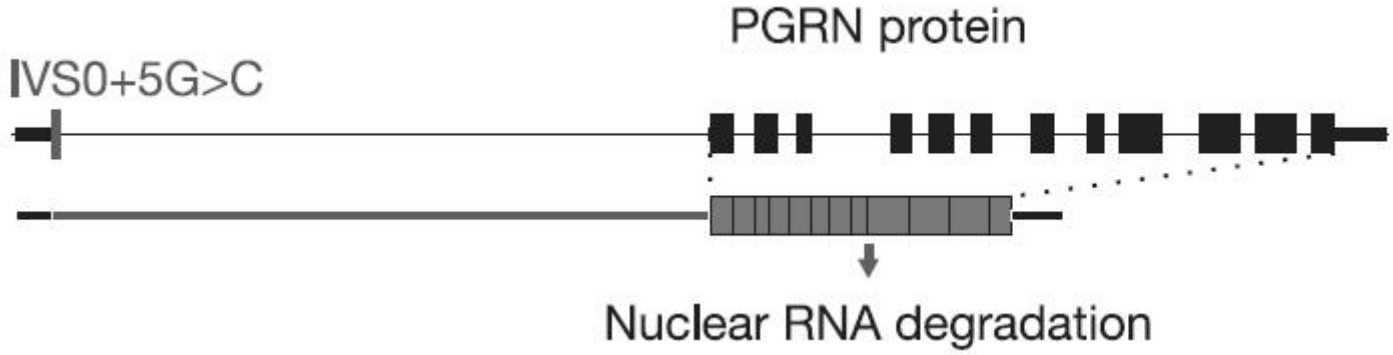
© *JT den Dunnen*

Exon 0

Progranulin
Aug.2006



Baker, Nature 442: 916



Cruts, Nature 442: 920

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen

Problematic 2

- **not described at DNA level**

  e.g. Abstract / Title / Results
  gives p.Tyr151Asp
  should give c.451T>C
  or c.451T>C (p.Tyr151Asp)
  at least on first appearance

- **range in description**

  c.635+12_14del seems clear
  (probably c.635+12_635+14del)
  but c.35+121_128del ?

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen

Problematic[3]

- **insertions**

  503insT is not clear

  ins at position 503 or after position 503 ?

  what about c.591-3insT ?

- **from name to one-letter AA-code**

  G ? - Glu, Gly, Gln
  A ? - Ala, Arg, Asn, Asp
  Phe - P ? (no F !)

- **assume most 3' residue is changed**

  ATGTC AAAAA TCGG
  c.10delA versus c.6delA

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen

Problematic[4]

- **recessive disease**
  - ***combination of alleles not listed***
- **AA numbering**
  - ***leader sequence not included***
- **reporting "polymorphisms"**
  - ***p.Arg123Arg***
    - ***useless and equivocal*** ***(no info, 5 possibilities DNA)***
  - ***36A/G***
    - ***unclear, c.36A>G or p.Ala36Gly ?***
- **experimental proof ?**
  - ***"change affects splicing"***
    - ***was RNA actually analysed ?***

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen

Problematic[4]

- **change detected**

  DNA c.2873G>T
  RNA r.2843_2873del
  protein Arg945fsX23

- **list this as ?**

  substitution,
  splice mutation or
  frame shift

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen

This meeting

(seen at this meeting)

- **XYZ_e01(-2599)**
- **-45A>C … -3I/D … 86C>G**
- **XYZ-267 … +10 … +80**
  (+10 positioned 5' of 5' UTR)
- **A196G**
- **IVS0+5G>C**

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen

One debate *, …of many*

reported: ***c.451T>C / p.Tyr151Asp***
text states: ***no RNA of this allele***

so correct: ***c.451T>C (r.? / p.0)***

author refuses to change,
editor accepts

what enters the database ?

is p.Tyr151Asp deleterious ?

pathogenic change may be elsewhere
other DNA change giving similar protein change

HGVS
HUMAN GENOME VARIATION SOCIETY

© *JT den Dunnen*

Why bother ?

Lydon, April 12, 2008—a XBG patient and his parents sued the department of clinical diagnosis in Lydon, the XBG mutation database, and the journal Human Mutation. *The complaint was that serious and culpable mistakes were made during the clinical diagnosis of the pregnancy in the XBG-family, that ultimately led to the birth of an affected child. A paper published in* Human Mutation *listed the sequence variant detected in the family as "nonpathogenic." Careful examination would have revealed that the change was clearly pathogenic (a nonsense mutation). However, the accused parties failed to verify the data of the original report and just copied it.*

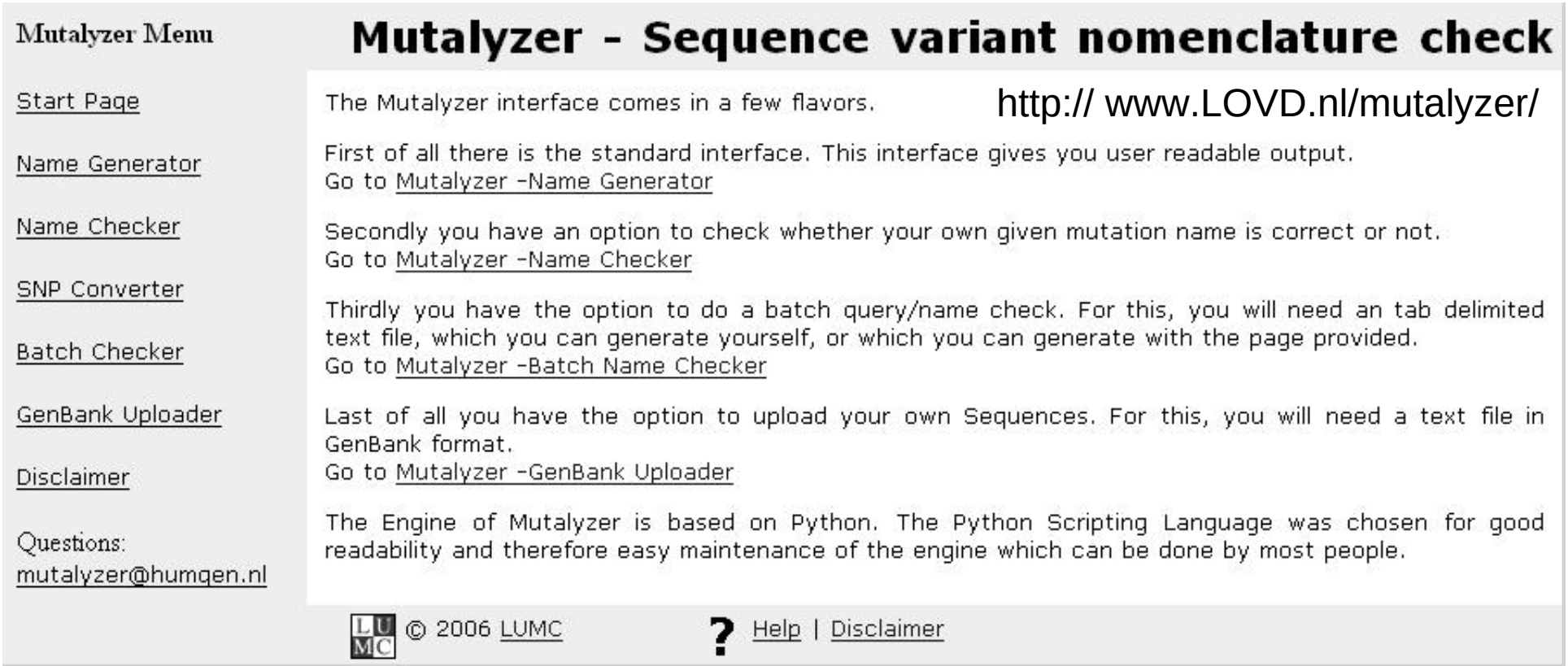
HUMAN MUTATION 22:181–182 (2003)

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen

Mutalyzer

- **input:** **reference sequence** ©Peter Taschner
  GenBank or private
- **output:** **correctly described change**
  DNA, protein, all annotated transcripts, RE-sites, …



Generates ref.seq.

Mutation database

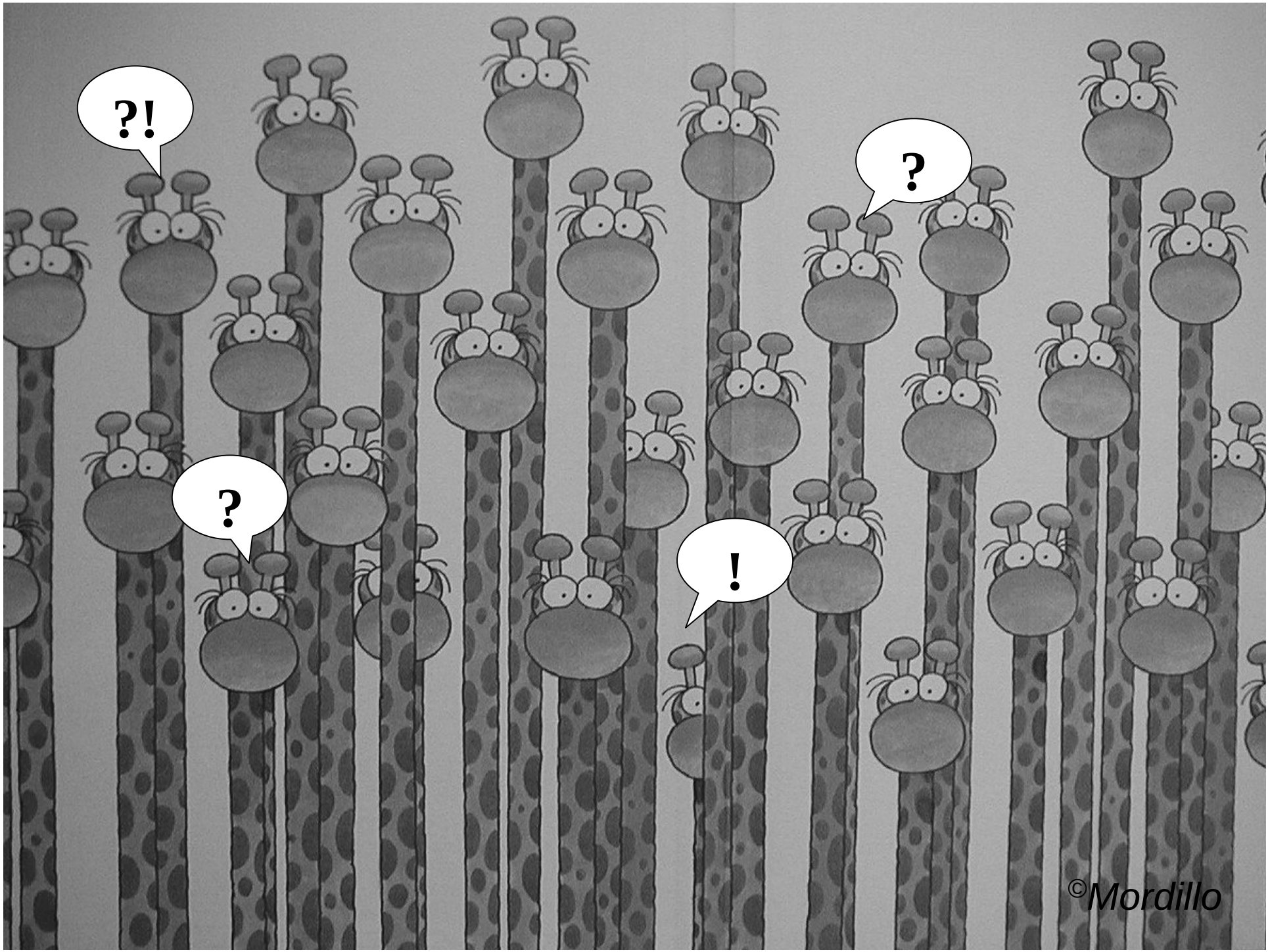
Submit <u>all</u> the changes

you have, NOW

(without errors)

HGVS
HUMAN GENOME
VARIATION SOCIETY

© *JT den Dunnen*


?!
?
?
!
©Mordillo

Recent suggestion

- **Copy Number Variants**

(last-present_first-deleted)_(last-deleted_first-present)del

BAC / PAC probe derived

(AC123322.10_AL109609.5)_(AL451144.5_AL050305.9)del
AL109609.5_AL451144.5del where non-del ?

g.(32,218,983_32,238,146)_(32,984,039_33,252,615)del
NCBI build 36.1

SNP-array

(rs2342234_rs3929856)_(rs10507342_rs947283)del

g.(32,218,983_32,238,146)_(32,984,039_33,252,615)del
NCBI build 36.1

HGVS
HUMAN GENOME VARIATION SOCIETY

© JT den Dunnen