Describing variants

"mutation nomenclature"

recommendations for the description of DNA changes

tinyurl.com/ VEP-HGVS

Johan den Dunnen

chair SVD-WG

http://varnomen.HGVS.org

Subjects

Reporting sequence variants

- *who decides*

- *where do I find the rules*
  www.HGVS/mutnomen

- *describing variants*
  brief, basics only

Discussion

problems
some tests
recommendations OK ?

Affiliations

THE HUMAN VARIOME PROJECT

get all variants/consequences shared

HGVS HUMAN GENOME VARIATION SOCIETY

standards for variant description and databases

ISCN

standards for cytogenetic variant descriptions

LOVD Leiden Open Variation Database

software for web-based gene databases

Standards

- **essential**
  to understand each other
  to exchange information

- **preferably ONE standard**
  used world-wide
  agreed by everybody

- **..but difficult**
  everybody agrees
  ...when their standard is used
  how to agree on changes ?
  which authority to decide ?

Celsius / Fahrenheit
kilometers / miles
liter / gallon

Still a problem?

Genetics inMedicine Official Journal of the American College of Medical Genetics

Home | Current Issue | Archive | Podcasts | For Authors & Referees | About the journal

Archive > Volume 17 > Issue 5 > Article

GENETICS IN MEDICINE | ACMG STANDARDS AND GUIDELINES

Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology

Sue Richards PhD, Nazneen Aziz PhD, Sherri Bale PhD, David Bick MD, Soma Das PhD, ... Gastier-Foster PhD, Wayne W. Grody MD, PhD, Madhuri Hegde PhD, Elaine Lyon PhD, Ek... Spector PhD, Karl Voelkerding MD & Heidi L. Rehm PhD ; on behalf of the ACMG Laborat... Quality Assurance Committee

ACMG: follow the HGVS recommendations !!

In addition, this ACMG recommendation supports three specific exceptions to the HGVS nomenclature rules: 1) "X" is still considered acceptable for use in reporting nonsense variants in addition to the current HGVS recommendation of "*" and "Ter"; 2) it is recommended

X to *

- **protein changes**

historically the X used for "stop codon"

> IUPAC amino acid codes
X = any amino acid

> NCBI amino acid codes
X = any amino acid, * = translation stop

>>> change X to * / Ter

p.Arg321*
p.Arg321Glufs*13
p.*535Glnext*17

The problem

alternative descriptions

- **share & retrieve**

  when alternative descriptions are accepted it becomes problematic to find previous reports

```
.          .          .          .                    g.73678
ttttgatgtgtgtgtgtgtgtgtgtgtttttttaacag                c.1210-1
```

NCBI

dbSNP
Short Genetic Variations

HGVS Names
NC_000007.13:g.117188660_117188661insTG
NC_000007.13:g.117188660_117188661insTGTG
NC_000007.14:g.117548606_117548607insTG
NC_000007.14:g.117548606_117548607insTGTG
NG_016465.4:g.87823_87824insTG
NG_016465.4:g.87823_87824insTGTG
NM_000492.3:c.1210-35_1210-34insTG
NM_000492.3:c.1210-35_1210-34insTGTG

HGVS Names
NC_000007.13:g.117188661_117188662delTG
NC_000007.14:g.117548607_117548608delTG
NG_016465.4:g.87824_87825delTG
NM_000492.3:c.1210-34_1210-33del
NM_000492.3:c.1210-34_1210-33delTG

HGVS Names
NC_000007.13:g.117188662_117188663insTG
NC_000007.13:g.117188662_117188663insTGTG
NC_000007.14:g.117548608_117548609insTG
NC_000007.14:g.117548608_117548609insTGTG
NG_016465.4:g.87825_87826insTG
NG_016465.4:g.87825_87826insTGTG
NM_000492.3:c.1210-33_1210-32insTG
NM_000492.3:c.1210-33_1210-32insTGTG

HGVS Names
NC_000007.13:g.117188684T>G
NC_000007.14:g.117548630T>G
NG_016465.4:g.87847T>G
NM_000492.3:c.1210-11T>G

HGVS Names
NC_000007.13:g.117188661_117188662TG[11][12]
NC_000007.14:g.117548607_117548608TG[11][12]
NG_016465.4:g.87824_87825TG[11][12]
NM_000492.3:c.1210-34_1210-33TG[11][12]

HGVS Names
NC_000007.13:g.117188662delG
NC_000007.14:g.117548628delG
NG_016465.4:g.87845delG
NM_000492.3:c.1210-13delG

HGVS Names
NC_000007.13:g.117188682_117188683insT
NC_000007.13:g.117188682_117188683insTGTT
NC_000007.13:g.117188682_117188683insTT

HGVS Names
NC_000007.13:g.117188684_117188685insG
NC_000007.14:g.117548630_117548631insG
NG_016465.4:g.87847_87848insG
NM_000492.3:c.1210-11_1210-10insG

HGVS Names
NC_000007.13:g.117188661_117188664delTGTG
NC_000007.14:g.117548607_117548610delTGTG
NG_016465.4:g.87824_87827delTGTG
NM_000492.3:c.1210-34_1210-31delTGTG

HGVS Names
NC_000007.13:g.117188682G>T
NC_000007.14:g.117548628G>T
NG_016465.4:g.87845G>T
NM_000492.3:c.1210-13G>T

HGVS Names
NC_000007.13:g.117188683delT
NC_000007.14:g.117548629delT
NG_016465.4:g.87846delT
NM_000492.3:c.1210-12delT

HGVS Names
NC_000007.13:g.117188688T[5][7][9]
NC_000007.14:g.117548634T[5][7][9]
NG_016465.4:g.87851T[5][7][9]

HGVS Names
NC_000007.13:g.117188661_117188666delTGTG
NC_000007.14:g.117548607_117548612delTGTG
NG_016465.4:g.87824_87829delTGTGTG
NM_000492.3:c.1210-34_1210-29delTGTGTG

HGVS Names
NC_000007.13:g.117188681_117188684delTGTT
NC_000007.14:g.117548627_117548630delTGTT
NG_016465.4:g.87844_87847delTGTT
NM_000492.3:c.1210-14_1210-11delTGTT

HGVS Names
NC_000007.13:g.117188683_117188684delTT
NC_000007.14:g.117548629_117548630delTT
NG_016465.4:g.87846_87847delTT
NM_000492.3:c.1210-12_1210-11delTT

HGVS Names
NC_000007.13:g.117188689_117188690insTT
NC_000007.14:g.117548635_117548636insTT
NG_016465.4:g.87852_87853insTT
NM_000492.3:c.1210-6_1210-5insTT

Variant description

the basis ***http:// www.HGVS.org / varnomen***

SPECIAL ARTICLE

Human Mutation

OFFICIAL JOURNAL HGVS HUMAN GENOME VARIATION SOCIETY www.hgvs.org

HGVS Recommendations for the Description of Sequence Variants: 2016 Update

Hum Mutat (2016) 37:564-569

Johan T. den Dunnen,[1]* Raymond Dalgleish,[2] Donna R. Maglott,[3] Reece K. Hart,[4] Marc S. Greenblatt,[5] Jean McGowan-Jordan,[6] Anne-Francoise Roux,[7] Timothy Smith,[8] Stylianos E. Antonarakis,[9] and Peter E.M. Taschner[10] on behalf of the Human Genome Variation Society (HGVS), the Human Variome Project (HVP), and the Human Genome Organisation (HUGO)

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, listing 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.

KEYWORDS: complex mutations; mutation detection; mutation database; nomenclature; MDI

THE HUMAN VARIOME PROJECT

HUGO

HGVS HUMAN GENOME VARIATION SOCIETY

HGVS / HVP / HUGO Sequence Variant Description working group

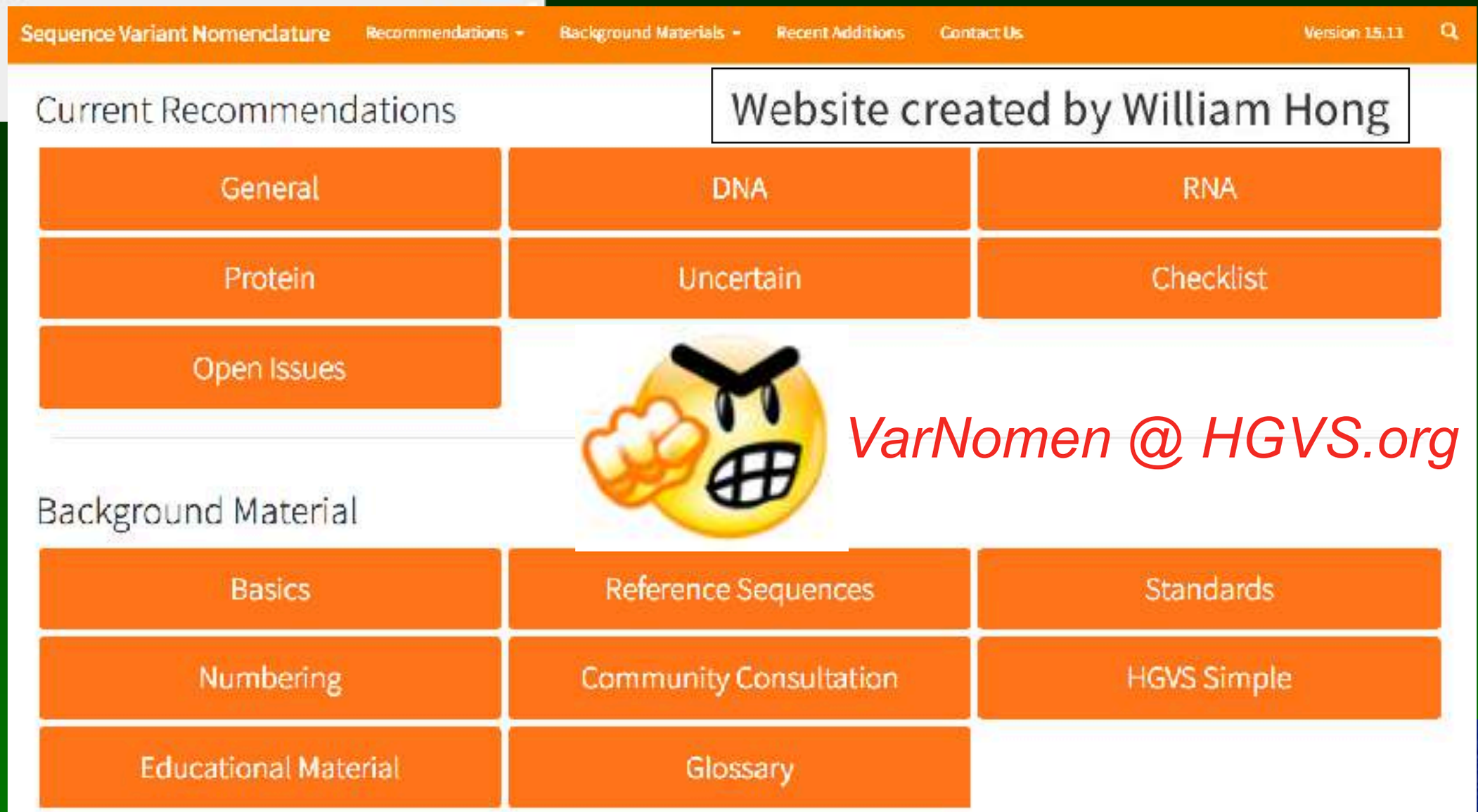
Working Group Members:

- Anne-Francoise Roux (EGT)
- Donna Maglott (NCBI/EBI)
- Jean McGowan-Jordan (ISCN)
- Peter Taschner (LSDBs)
- Raymond Dalgleish (LSDBs)
- Reece Hart (industry)
- Johan den Dunnen (chair)
- HGVS - Marc Greenblatt
- HUGO - Stylianos Antonarakis

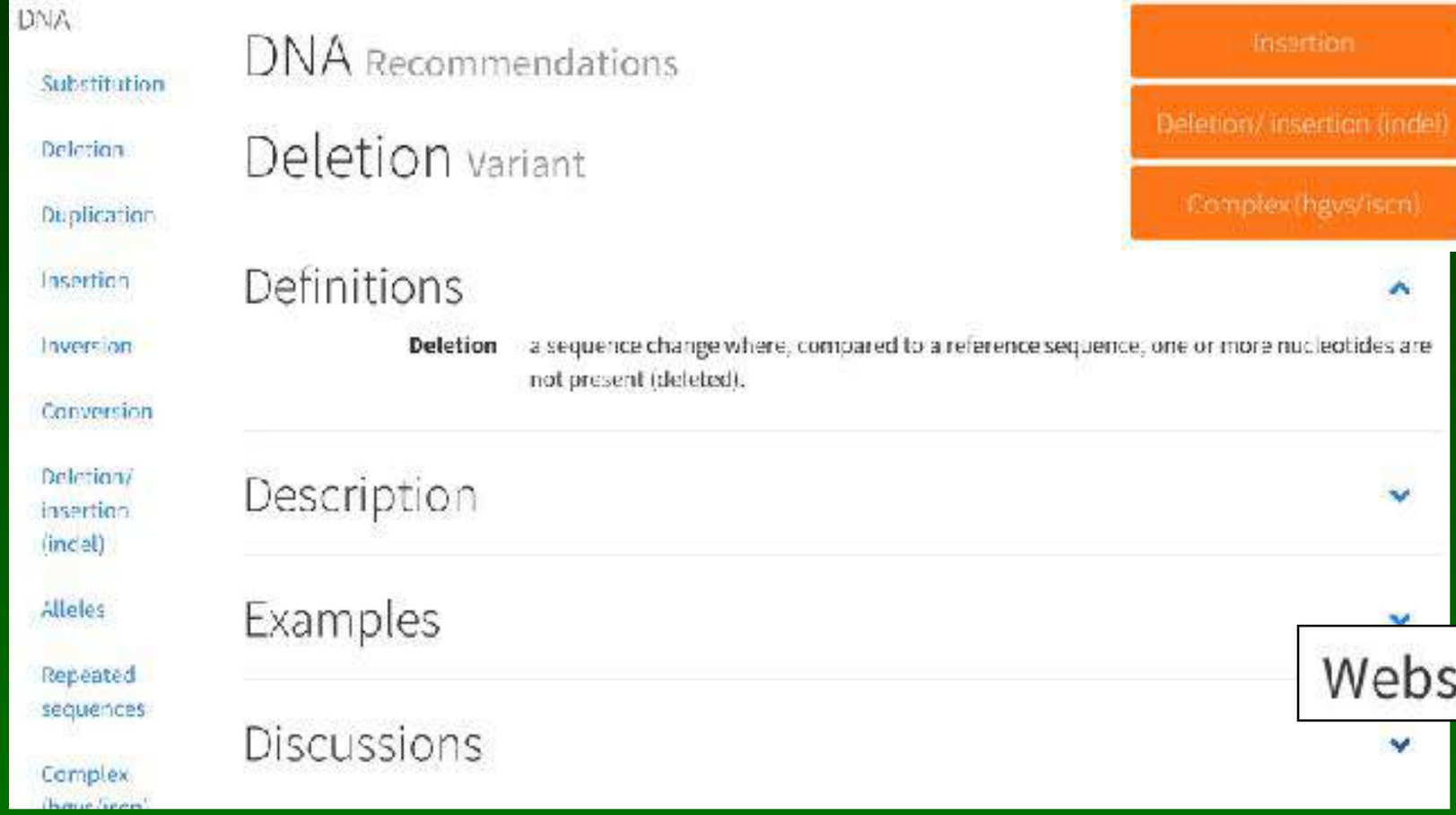
varnomen.HGVS.org

Follow the recommendations

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



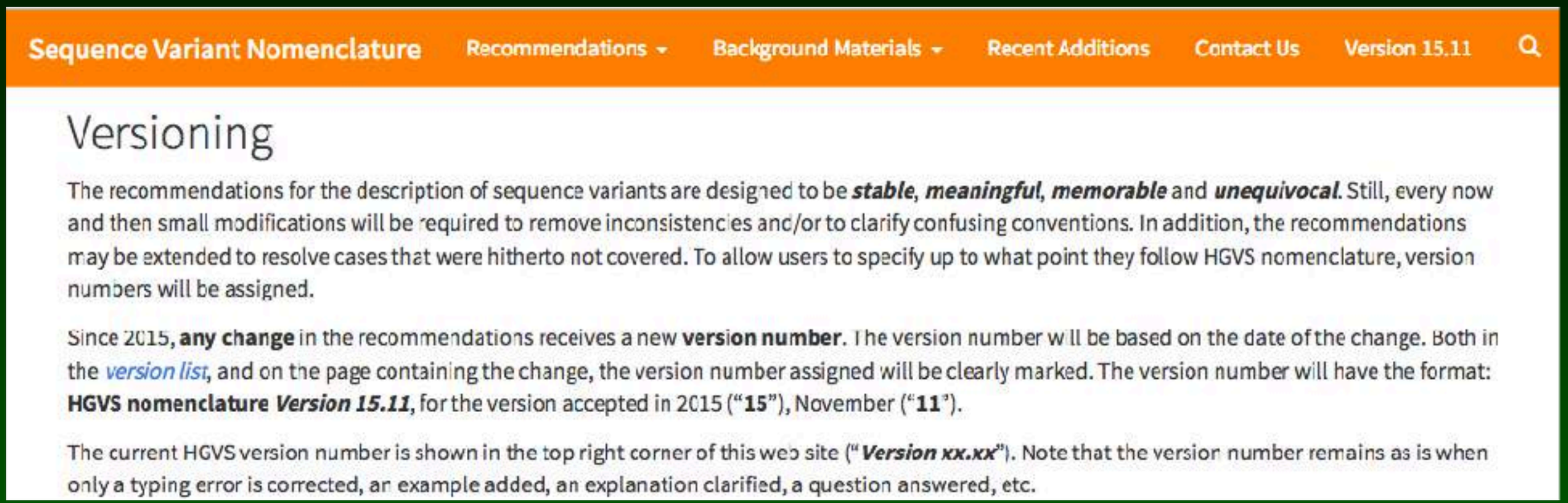
Per variant type



Website created by William Hong

Versioning

current version is 15.11 (Nov.2015)



Sequence Variant Nomenclature | Recommendations | Background Materials | Recent Additions | Contact Us | Version 15.11

Versioning

The recommendations for the description of sequence variants are designed to be ***stable***, ***meaningful***, ***memorable*** and ***unequivocal***. Still, every now and then small modifications will be required to remove inconsistencies and/or to clarify confusing conventions. In addition, the recommendations may be extended to resolve cases that were hitherto not covered. To allow users to specify up to what point they follow HGVS nomenclature, version numbers will be assigned.

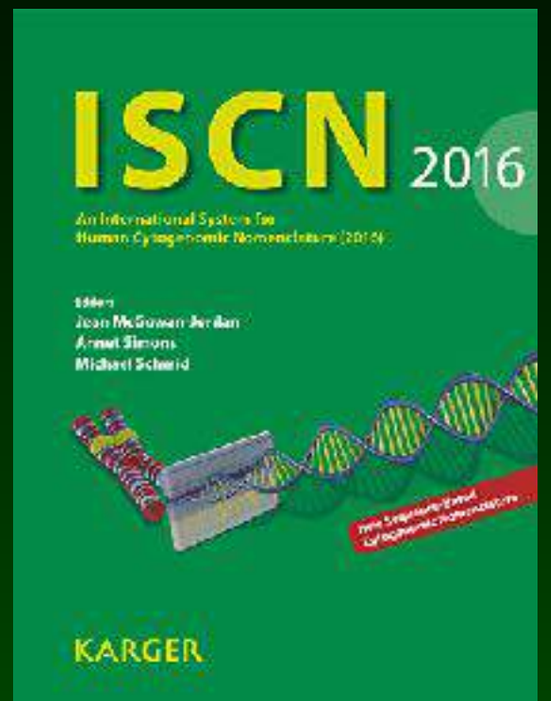
Since 2015, **any change** in the recommendations receives a new **version number**. The version number will be based on the date of the change. Both in the *version list*, and on the page containing the change, the version number assigned will be clearly marked. The version number will have the format: **HGVS nomenclature *Version 15.11***, for the version accepted in 2015 ("**15**"), November ("**11**").

The current HGVS version number is shown in the top right corner of this web site ("***Version xx.xx***"). Note that the version number remains as is when only a typing error is corrected, an example added, an explanation clarified, a question answered, etc.

Community Consultation

Community Consultation

HGVS nomenclature falls under the responsibility of the SVD-WG (*Sequence Va*
change or extend HGVS nomenclature operating according to a charter definir
Consultation step. Any proposal made by the SVD-WG will be **published on th**
2-month period (*register for email notification*). Everybody interested is asked
SVD-WG. Comments to proposals should be addressed to "Varnomen @ varior



- **SVD-WG004 (ISCN<>HGVS)**
  suggested to extend the recommendations to cover the description of structural variants, esp. translocations and chromothripsis.
  Status: under review. Closed Jan.15 (2016). Opened Nov.10 (2015).

facebook & twitter

Activities

- **daily**

  close link to gene variant databases

  website

  answer questions

  now 1-2 daily

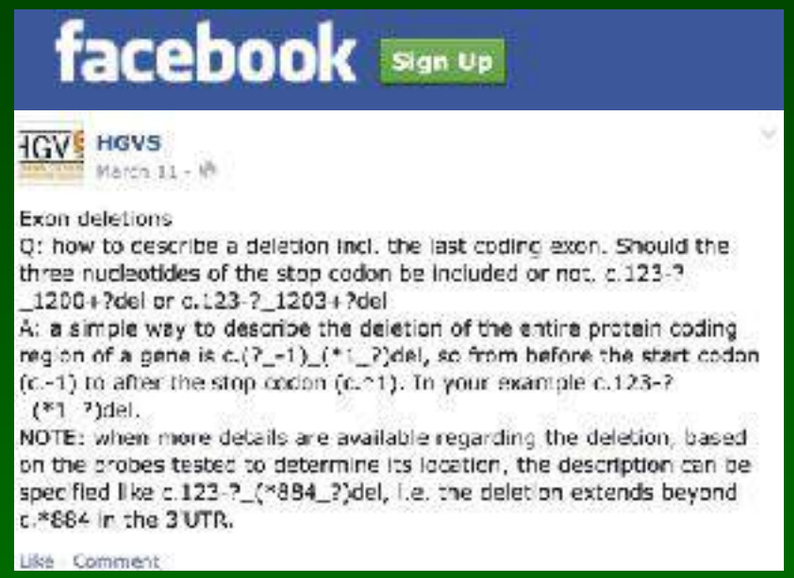
  started facebook/twitter

- **promote**

  lectures, posters, courses, write journals/authors/agencies, ...

  Socrative

  Questions/Answers



Nomenclature

(describing DNA variants)

Stable

Meaningful ***Memorable***

Unequivocal

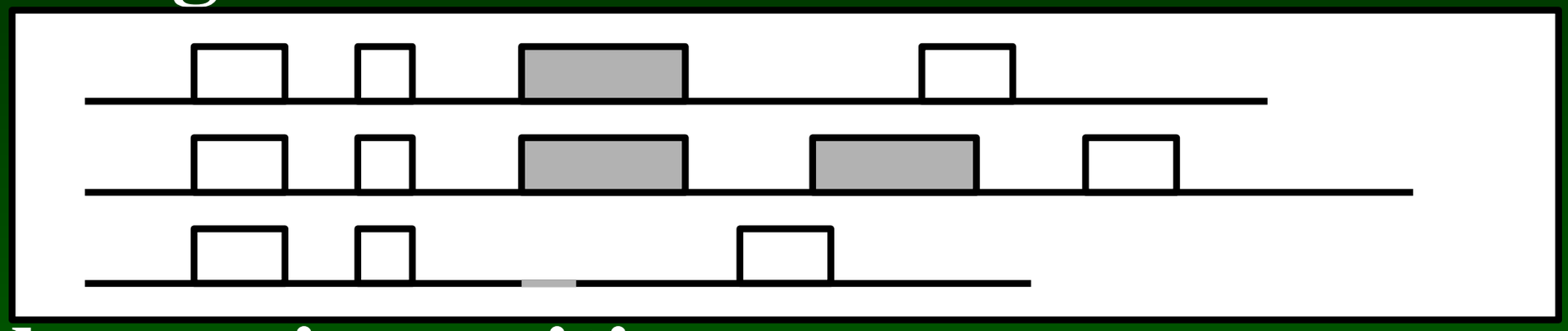
computer

Variant types

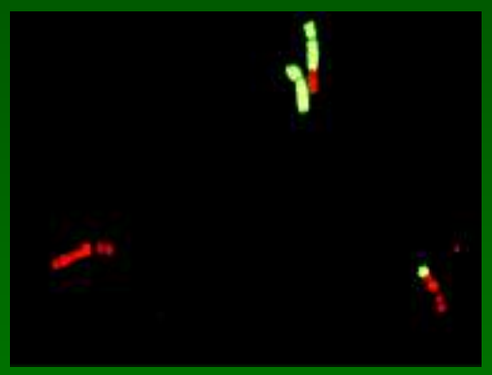
- **change in sequence**

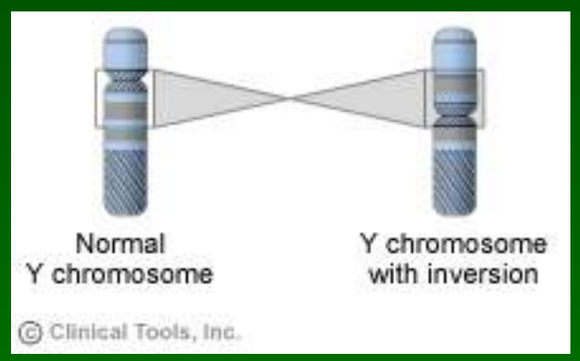
```
ACATCAGGAGAAGATGTTC GAGACTTTGCCA
ACATCAGGAGAAGATGTTT GAGACTTTGCCA
ACATCAGGAGAAGATGTT  GAGACTTTGCCA
ACATCAGGAGAAGATGTTCCGAGACTTTGCCA
```

- **change in amount** *(Copy Number Variation)*



- **change in position**





ISCN

Structural Variation (SV)

DNA, RNA, protein

- **unique descriptions**
  prevent confusion

- **DNA**
  A, G, C, T
  g.957A>T, c.63-3T>C

- **RNA**
  a, g, c, u
  r.957a>u, r.(?), r.spl?

- **protein** *(mostly deduced)*
  three / one letter amino acid code
  ** = stop codon*
  p.(His78Gln)

Basic rules

- **report what is detected**
  NOT what is predicted
  NOT p.Gly202Trp, but c.604G>T
  or c.604G>T (r.(?), p.(Gly202Trp))

- **give a reference sequence**
  accession.version number
  genomic (chromosomal) or LRG

- **use the 3' rule**
  shift change as far 3' as possible

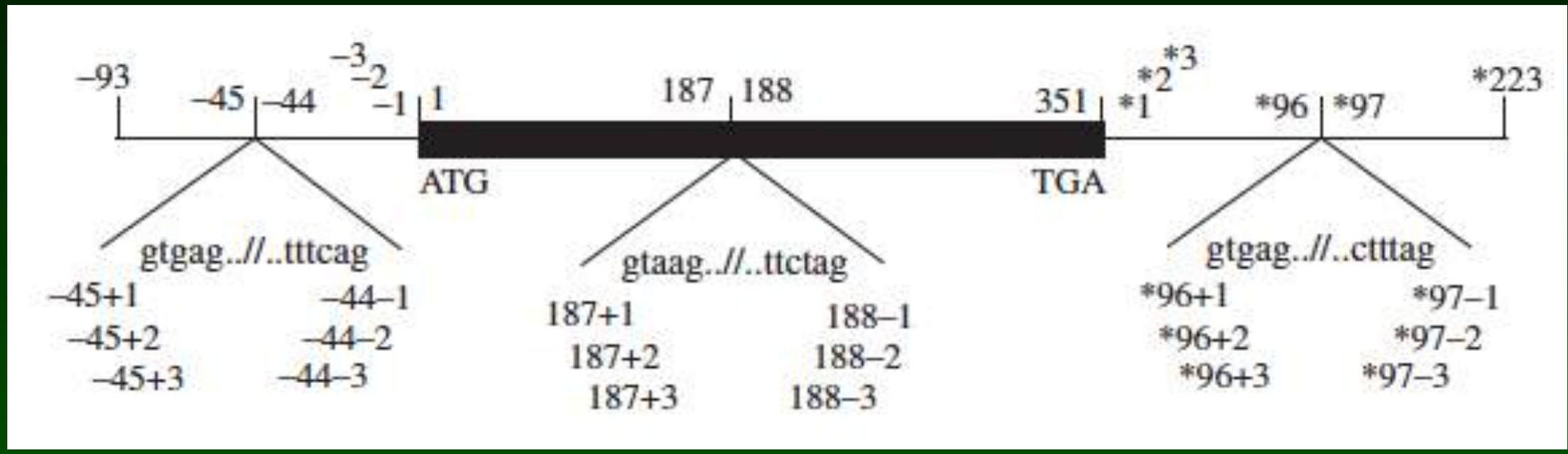
Numbering residues

- **start with 1**
  - *genomic* *1 is first nucleotide of file*
    - *no +, - or other signs*
  - *coding DNA* *1 is A of ATG*
    - *for introns refer to genomic Reference Sequence*

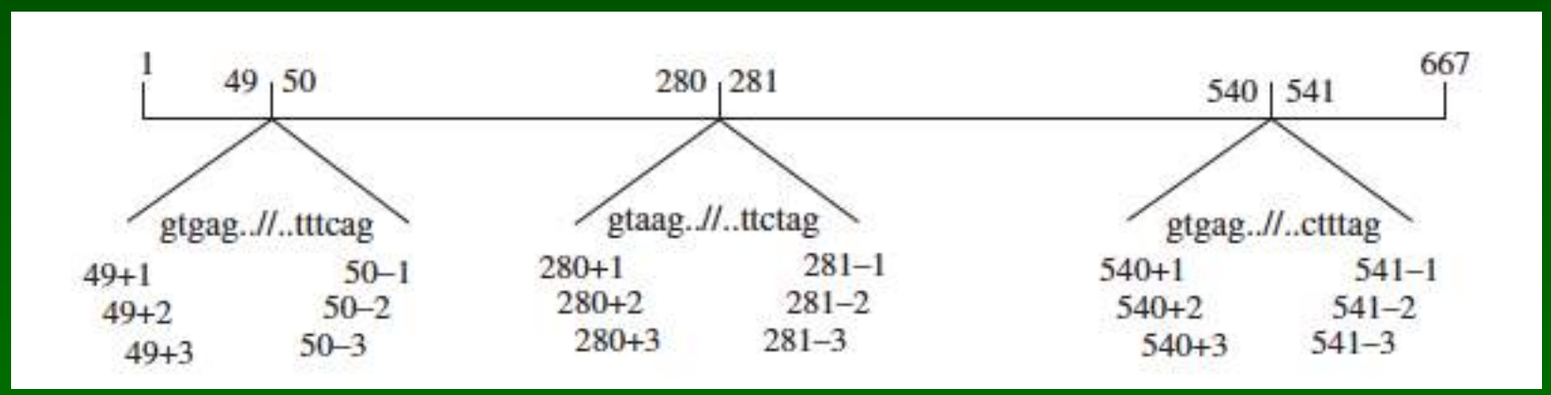
- **exception:** **coding DNA**
  - *5' of ATG* *…, -3, -2, -1, A, T, G, …*
    - *no nucleotide 0*
  - *3' of stop* **1, *2, *3, …*
    - *no nucleotide 0*
  - *intron*
    - *position between nt's 654 and 655*
    - *c.654+1, +2, +3, ……., -3, -2, c.655-1*
    - *change + to - in middle*

Reference Sequence

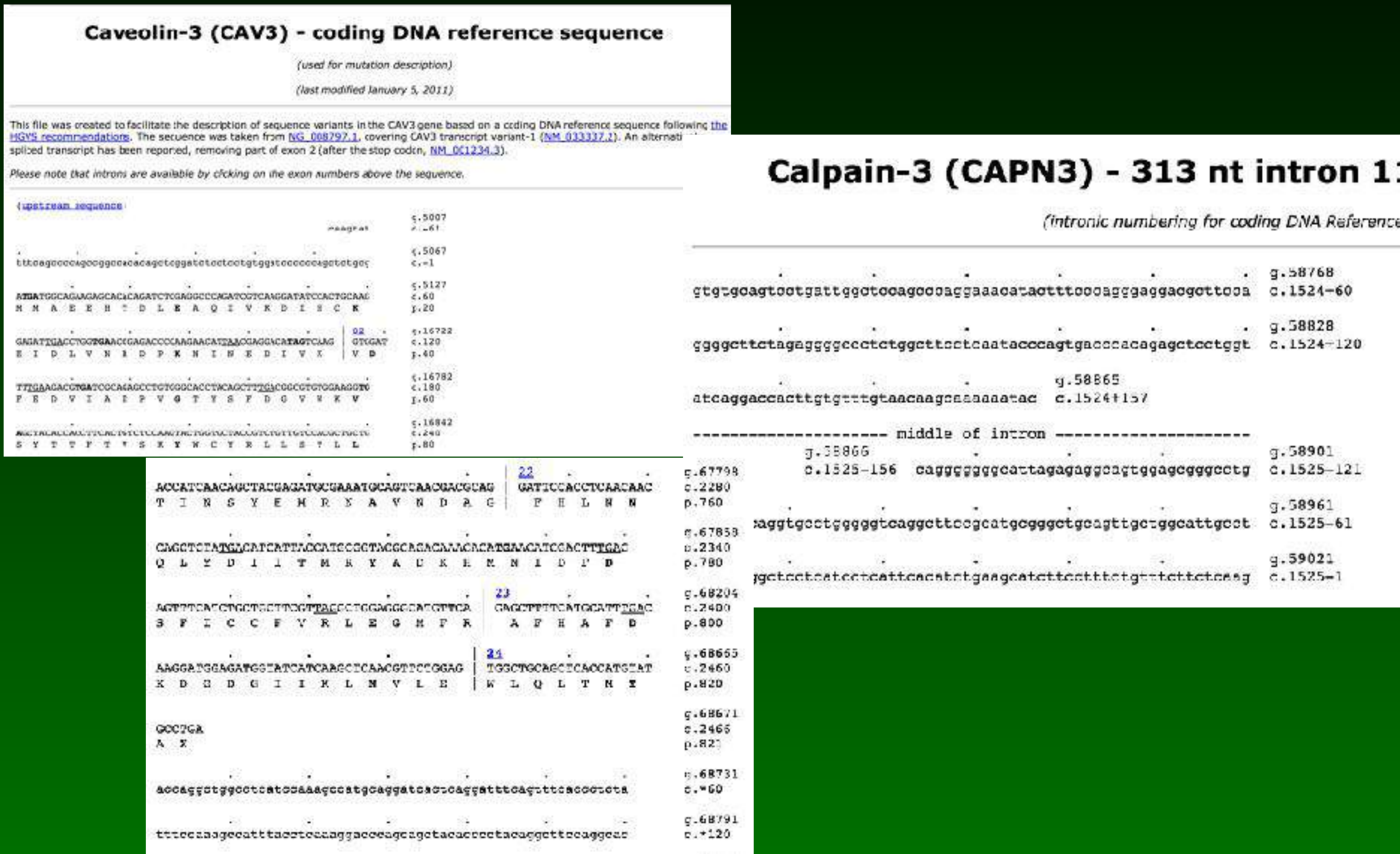
coding DNA reference sequence (c.)



non-coding DNA reference sequence (n.)



LSDB service



© JT den Dunnen

Computer prefered

- **g.12158663A>G**

hint chr.11 (hg19)

Computer prefered

- g.12158663A>G
- g.23669859>C
- g.89112396G>A

no relation to RNA & protein

- g.112775623C>G
- g.56569443A>T
- g.12741333T>G
- g.188153979G>C

Numbering - coding DNA

- **c.2396-6G>A**

in the 3' half of an intron,
6 nucleotides 5' of the splice acceptor site

splitting amino acid 799

Human prefered

- **c.1637A>G**
  protein coding region
- **c.859+12T>C**
  in intron (5' half)
- **c.2396-6G>A**
  in intron (3' half)

relation to RNA & protein

- **c.-23C>G**
  5' of protein coding region (5' of ATG)
- **c.*143A>T**
  3' of protein coding region (3' of stop)
- **c.-89-12T>G**
  intron in 5' UTR (5' of ATG)
- **c.-649+79G>C**
  intron in 3' UTR (3' of stop)

HGVS
HUMAN GENOME VARIATION SOCIETY

Types of variation

- **simple**
  - *substitution* — *c.123A>G*
  - *deletion* — *c.123delA*
  - *duplication* — *c.123dupA*
  - *insertion* — *c.123_124insC*
  - *other*
    conversion, inversion, translocation, transposition
- **complex**
  - *indel* — *c.123delinsGTAT*
- **combination of variants**
  - *two alleles* — *c.[123A>G];[456C>T]*
  - *>1 per allele* — *c.[123A>G;456C>T]*

Substitution

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

- **examples**

genomic	*g.54786A>T*
cDNA	*c.545A>T*
	(NM_012654.3 : c.546A>T)
RNA	*r.545a>u*
protein	*p.(Gln182Leu)*

Deletion

- **deletion**

  designated by "del"
  range indicated by "_"

- **examples**

  c.586_591del

  c.586_591delTGGTCA (not c.586_591del6)

  c.546del

  c.546delT

3' rule

```
Reference     ATAGCTTTCAGGA
Sample        ATAGCT TCAGGA
```

Describe as g.6del
g.7del
g.8del

?

By definition this is described as **g.8del**

HGVS 3’ rule

Applied correctly?

HGVS Nomenclature in Practice:
An Example from the United Kingdom National External Quality Assessment Scheme

OFFICIAL JOURNAL
HGVS
HUMAN GENOME VARIATION SOCIETY
www.hgvs.org

Zandra C. Deans,[1]* Jennifer A. Fairley,[1] Johan T den Dunnen,[2] and Caroline Clark[3]

[1]UK NEQAS for Molecular Genetics, NHS Lothian, Royal Infirmary of Edinburgh, Little France Crescent, Edinburgh, UK; [2]Clinical Genetics and Human Genetics, Leiden University Medical Center, Leiden, Nederland; [3]Department of Molecular Genetics, Medical Genetics, Polwarth Building, Aberdeen, UK

```
...ggccagcgtggaca          acCccc...
...ggccagcgtggacaGCGTGGACAacGccc...
```

(both variants on same chromosome)

c.[2303_2311dup;2314C>G] /
c.2312_2314delinsGCGTGGACAACG

or c.[2303_2311dup(;)2314C>G]

Applied correctly?

HGVS Nomenclature in Practice:
An Example from the United Kingdom National External Quality Assessment Scheme

OFFICIAL JOURNAL HGVS HUMAN GENOME VARIATION SOCIETY www.hgvs.org

Zandra C. Deans,[1]* Jennifer A. Fairley,[1] Johan T den Dunnen,[2] and Caroline Clark[3]

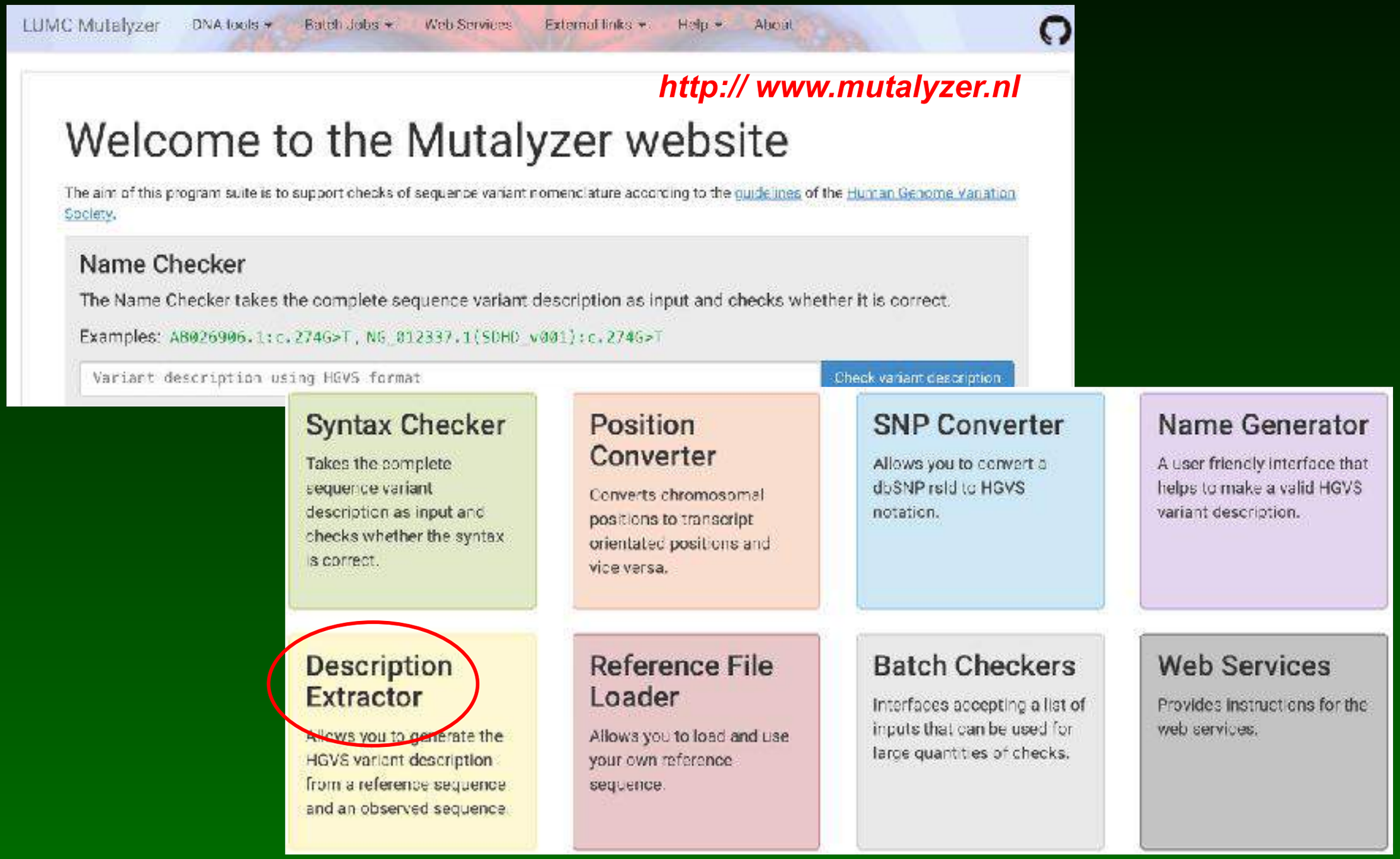
[1]UK NEQAS for Molecular Genetics, NHS Lothian, Royal Infirmary of Edinburgh, Little France Crescent, Edinburgh, UK; [2]Clinical Genetics and Human Genetics, Leiden University Medical Center, Leiden, Nederland; [3]Department of Molecular Genetics, Medical Genetics, Polwarth Building, Aberdeen, UK

- **26 participating labs**
- **21 different descriptions**
  (DNA & protein combined)
  5 shared + unique 21x

6 correct HGVS,
12 "correct" variant,
8 not correct

c.2303_2311dup
c.2312_2314delinsGCGTGGACAACG
c.2312_2314delACCinsGCGTGGACAACG
c.2311_2312insTGTCCACGC
c.2300_2301insCAGCGTGGA
c.2300_2301insCAGCGTGGA
c.2302_2310dup
c.2303_2311dup
c.2312_2320dupGCGTGGACA
c.2311ins/dupGCGTGGACA
c.2303_2311dup
c.2300_2301insCAGCGTGGA
c.2300_2301insCAGCGTGGA
c.2303_2311dup
c.2303_2311dupGCGTGGACA
c.2300_2301insCAGCGTGGA
c.2301_2302insCAGCGTGGA
c.2310_2311dupAGCGTGGAC
c.2301_2302ins9
c.2311_2312insGCGTGGACA
c.2311_2312ins9 and c.2314C>G

Support tools



Applied correctly?

Lab	*Mutalyzer*	
c.2303_2311dup	c.[2303_2311dup;2314C>G]	
c.2312_2314delinsGCGTGGACAACG	c.2312_2314delinsGCGTGGACAACG	
c.2312_2314delACCinsGCGTGGACAACG	c.2312_2314delinsGCGTGGACAACG	
c.2311_2312insTGTCCACGC	*c.2311_2312insTGTCCACGC*	*not corrected*
c.2300_2301insCAGCGTGGA	c.2303_2311dup	
c.2300_2301insCAGCGTGGA	c.2303_2311dup	
c.2302_2310dup	c.2303_2311dup	
c.2303_2311dup	c.2303_2311dup c.2314C>G	
c.2312_2320dupGCGTGGACA	c.2312_2320dup	*error Mutalyzer*
c.2311ins/dupGCGTGGACA	reports error	
c.2303_2311dup	c.2303_2311dup	
c.2300_2301insCAGCGTGGA	c.2303_2311dup	
c.2300_2301insCAGCGTGGA	c.2303_2311dup c.2314C>G	
c.2303_2311dup	c.2303_2311dup	
c.2303_2311dupGCGTGGACA	c.2303_2311dup c.2314C>G	
c.2300_2301insCAGCGTGGA	c.2303_2311dup	
c.2301_2302insCAGCGTGGA	*c.2301_2302insCAGCGTGGA*	*not corrected*
c.2310_2311dupAGCGTGGAC	c.2310_2311dup	*error Mutalyzer*
c.2301_2302ins9	Reports error c.2314C>G	
c.2311_2312insGCGTGGACA	*c.[2303_2311dup;2316C>G]*	*not corrected*
c.2311_2312ins9 and c.2314C>G	Reports error Reports error c.2314C>G	

Correct?

NOTE: reference sequences can be assumed to be known, spaces in descriptions used for clarity only.

- **NM_01234.3 : c.65+2T>A**

Not correct

NOTE: reference sequences can be assumed to be known, spaces in descriptions used for clarity only.

- **NM_01234.3 : c.65+2T>A**
  nucleotide not in reference
  correct NG_022335.1(NM_01234.3):c.65+2T>A
  NC_000005.10(NM_01234.3):c.65+2T>A

- **LRG_123:c.957G>T**
  reference transcript missing (LRG_123t1)

- **NG_01234.3:c.1A>G**
  reference transcript missing (NM_01234.3)

Correct?

NOTE: reference sequences can be assumed to be known, spaces in descriptions used for clarity only.

- **c.123-?_456+?del**

MLPA detected deletion

Correct?

NOTE: reference sequences can be assumed to be known, spaces in descriptions used for clarity only.

- **c.123-?_456+?del**

MLPA detected deletion

NO, it fails to describe the extent of the deletion it is something like

c.(122+1_123-1)_(456+1_457-1)del

or c.(?_123-1)_(456+1_?)del

Suggestions made

- **SVD-WG003 (exon del/dup)**
  suggested to describe exon deletions/duplications using the format c.(233+1_234-1)_(1234+1_1235-1)del.
  Status: Oct.6 (2015) new proposal to be made. Closed Jul.16 (2015). Opened May 14 (2015).

c.(122+1_123-1)_(456+1_457-1)del

c.(122+?_123-?)_(456+?_457-?)del

c.(122_?_123)_(456_?_457)del

c.(122_123)_(456_457)del

c.(122|123)_(456|457)del

c.(122^123)_(456^457)del

Correct?

NOTE: reference sequences can be assumed to be known, spaces in descriptions used for clarity only.

- **c.(122+1_123-1)_(456+1_457-1)dup**

MLPA detected

Duplication

NOTE: reference sequences can be assumed to be known, spaces in descriptions used for clarity only.

- **historically** ***MLPA detected***

  c.123-?_456+?dup

  c.123 and c.456 are flanking exon borders

- **...but**

  extend change must be indicated

  so c.(122+1_123-1)_(456+1_457-1)dup

- **...but is it a duplication?**

  you detected only an extra copy

  c.?_?ins(122+1_123-1)_(456+1_457-1)

  ISCN c.(122+1_123-1)_(345+1_346-1)x3

Correct?

NOTE: reference sequences can be assumed to be known, spaces in descriptions used for clarity only.

```
…TCT AGT TCT…   >   …TCT TCT…
…Ser Ser Ser…   >   …Ser Ser…
```

a. p.(Ser1)del
b. p.(Ser2)del
c. p.(Ser3)del

c. p.(Ser3)del

NOTE: reference sequences can be assumed to be known, spaces in descriptions used for clarity only.

- **report change at level described**

on protein level, forget DNA position

```
…TCT AGT TCT…   >   …TCT TCT…
…Ser Ser Ser…   >   …Ser Ser…
```

most 3' position possible is arbitrarily assigned

so p.(Ser3del) with c.4_6del

not p.(Gln2del)

request to change

not consistent with overall recommendations

Correct?

NOTE: reference sequences can be assumed to be known, spaces in descriptions used for clarity only.

g.1234567_1239870 I gom

Depends

NOTE: reference sequences can be assumed to be known, spaces in descriptions used for clarity only.

g.1234567_1239870 I gom

Community Consultation

- **SVD-WG005 (gom/lom)**
  suggests to extend the HGVS recommendations to allow description of changes in general methylation status
  Status: Open. Oct.20 (2016). Closes Dec.31 (2016).

gom = gain of methylation
lom = loss of methylation
I ('pipe') = change of state (not of sequence)

HGVS
HUMAN GENOME VARIATION SOCIETY

Emotions