

Sequence Variant Nomenclature

<http://varnomen.hgvs.org/recommendations/DNA/>

2017 年 6 月 27 日

Chapter 1

DNA

1.1 DNA Substitution

Definitions

Substitution: a sequence change where, compared to a reference sequence, one nucleotide is replaced by one other nucleotide

Description

Format: “prefix” “position_substituted” “reference_nucleotide” “>” new_nucleotide”, e.g. g.123A>G

“prefix” = reference sequence used = g.

“position_substituted” = position nucleotide substituted = 123

“reference_nucleotide” = nucleotide at reference position = A

“>” = type of change is a substitution = >

“new_nucleotide” = substituted nucleotide = G

Note

prefix reference sequences accepted are g., m., c. and n. (genomic, mitochondrial, coding DNA and non-coding DNA).

changes involving two or more consecutive nucleotides are described as deletion/insertions (indels) (see Deletion/insertion (indel)).

nucleotides that have been tested and found not changed are described as c.123A=, g.4567T= (see SVD-WG001 (no change)).

the description c.76_77delinsTT is preferred over c.[76A>T;77G>T].

NOTE: by definition this change can not be described as a substitution (like c.76_77AG>TT or c.76AG>TT)

it is not correct to describe “polymorphisms” as c.76A/G (see Discussions).

Examples

NC_000023.10:g.33038255C>A

a substitution of the C nucleotide at g.33038255 for an A

NG_012232.1(NM_004006.1):c.93+1G>T

a substitution of the G nucleotide at c.93+1 (coding DNA reference sequence) with a T

LRG_199t1:c.79_80delinsTT or c.[79G>T;80C>T]

the description c.79_80delinsTT is preferred over c.[79G>T;80C>T], unless either of the two variants (79G>T or c.80C>T) is known as a frequently occurring variant.

NOTE: based on the definition of a substitution, i.e. one nucleotide replaced by one other nucleotide, this change can not be described as a substitution like c.79_80GC>TT or c.79GC>TT

NM_004006.1:c.[145C>T;147C>G]

two substitutions replacing codon CGC (position c.145 to c.147) with TGG

NOTE: the variant can also be described as NM_004006.1:c.145_147delinsTGG, i.e. a deletion/insertion. The deletion/insertion format is preferred unless either of the two variants (c.145C>T or c.147C>G) is known as a frequently occurring variant.

LRG_199t1:c.54G>H

a substitution of the G nucleotide at c.54 (coding DNA reference sequence) with a A, C or T (IUPAC code “H” , see Standards)

NM_004006.1:c.123=

a screen was performed showing that nucleotide c.123 was a “C” as in the coding DNA reference sequence (the nucleotide was not changed). Alternative NM_004006.1:c.123C=.

LRG_199t1:c.85=/T>C

a mosaic case where at position 85 besides the normal sequence (a T, described as “=”) also chromosomes are found containing a C (c.85T>C)

NM_004006.1:c.85=/T>C

a chimeric case, i.e. the sample is a mix of cells containing c.85= and c.85T>C.

1.2 Deletion

Definitions

Deletion

a sequence change where, compared to a reference sequence, one or more nucleotides are not present (deleted).

Description

Format: “prefix” “position(s)_deleted” “del” , e.g. g.123_127del

“prefix” = reference sequence used = g.

“position(s)_deleted” = position nucleotide or range of nucleotides deleted = 123_127

“del” = type of change is a deletion = del

Note

prefix reference sequences accepted are g., m., c. and n. (genomic, mitochondrial, coding DNA and non-coding DNA).

“positions_deleted” should contain two different positions, e.g. 123_126 not 123_123.

the “position(s)_deleted” should be listed from 5’ to 3’ , e.g. 123_126 not 126_123.

for all descriptions the most 3’ position possible of the reference sequence is arbitrarily assigned to have been changed (3’ rule)

the 3’ rule also applies for changes in single residue stretches and tandem repeats (nucleotide or amino acid)

the 3’ rule applies to ALL descriptions (genome, gene, transcript and protein) of a given variant

exception

deletions around exon/intron and intron/exon borders when identical nucleotides flank these borders (see Numbering)

NG_012232.1(NM_004006.1):c.186+1del describes a deletion of a “G” at the exon/intron border ..CTGgta.. (positions c.186/c.186+1). When RNA analysis shows a G deletion (NM_004006.1:r.186del), so no effect on splicing, the change is described as NM_004006.1:c.186del.

NOTE: when in the above example the next exon would start with GGC.. the deletion is still described as NM_004006.1:c.186del (not c.188del).

Examples

NG_012232.1:g.19del (one nucleotide)

a deletion of the T at position g.19 in the sequence AGAATCACA to AGAA_CACA

NOTE: it is allowed to describe the variant as NG_012232.1:g.19delT

NG_012232.1:g.19_21del (several nucleotides)

a deletion of nucleotides g.19 to g.21 in the sequence AGAATCACA to AGAA____CA

NOTE: it is allowed to describe the variant as NG_012232.1:g.19_21delTGC

NG_012232.1(NM_004006.1):c.183_186+48del

a deletion of nucleotides c.183 to c.186+48 (coding DNA reference sequence), crossing an exon/intron border

NM_004006.1:c.1149del

when exon 10 ends with ..GAG and exon 11 starts with GGGT.. and the genomic DNA sequence shows that the last G-nucleotide of exon 10 is deleted (and not the G in exon 11), the deletion changing ..GAGGGGT.. to ..GAGGGT.. is described as c.1149del (not c.1152del, see exception in Numbering)

NOTE: it is allowed to describe the variant as NM_004006.1:c.1149delG

NM_004006.1:c.1152del

the deletion of the G nucleotide at the intron/exon border in the sequence CATGAGgt····/..agGGGTAC to CATGAGgt····/..agGG_TAC

NOTE: it is allowed to describe the variant as NM_004006.1:c.1152delG

NG_012232.1(NM_004006.1):c.1149+1del

the deletion of the G nucleotide at the exon/intron border in the sequence CATGAGgt.../..agGGGTAC to CATGAG_t.../..agGGGTAC (not c.1152del see Q&A)

NOTE: it is allowed to describe the variant as NG_012232.1(NM_004006.1):c.1149+1delG

NG_012232.1(NM_004006.1):c.4072-1234_5155-246del

a deletion of nucleotides c.4072-1234 to c.5155-246 removing exon 30 (starting at position c.4072) to exon 36 (ending at position c.5154) of the DMD-gene.

NOTE : c.4072-1234_5155-246delXXXXX, the size of the deletion (XXXXX) should not be described

NG_012232.1(NM_004006.1):c.(4071+1_4072-1)_(5154+1_5155-1)del

a deletion of exon 30 (starting at position c.4072) to exon 36 (ending at position c.5154) of the DMD-gene. The deletion break point has not been sequenced. Exons 29 (ending at c.4071) and 37 (starting at nucleotide c.5155) have been tested and shown to be not deleted. The deletion therefore starts in intron 29 (position c.4071+1 to c.4072-1) and ends in intron 36 (position c.5154+1 to c.5155-1).

NOTE : as mentioned (Uncertain) the description can also be probe-based. For a deletion of exons 30 to 36, detected using MLPA, the description would be NG_012232.1(NM_004006.1):c.(3996_4196)_(5090_5284)del, i.e. following the suggestion to use the central position (3' nucleotide) of the probe ligation site. E.g. the MLPA exon 29 probes hybridize from position c.3963 to c.4030 giving c.3996 as the position to use in the description.

NOTE : this description is part of proposal SVD-WG003 (undecided).

NOTE : previously, the suggestion was made to describe such deletions using the format NG_012232.1(NM_004006.1):c.4?_5154+?del. However, since c.4072-? indicates “to an unknown position 5' of c.4072” and c.5154+? “to an unknown position 3' of c.5154” this description is not correct when it is known that exons 29 and 37 are present.

NG_012232.1(NM_004006.1):c.(?-245)_(31+1_32-1)del

a deletion starting somewhere upstream of a gene, last position tested positive c.-244, and ending in the intron between nucleotides c.31+1 and c.32-1 (intron 1).

NG_012232.1(NM_004006.1):c.(?-1)_(+1_?)del

a deletion of the entire protein coding region of a gene based on a coding DNA reference sequence).

NOTE: when more details are available regarding the deletion, based on the probes tested to determine its location, the description can be specified like NG_012232.1(NM_004006.1):c.(?-189)_(+884_?)del, meaning the deletion starts 5' of c.-189 and extends 3' of c.+884.

NG_012232.1:g.19_21=/del

a mosaic case where from position g.19 to g.21 besides the normal sequence also chromosomes are found containing a deletion of this sequence

NG_012232.1:g.19_21del=//del

a chimeric case, i.e. the sample is a mix of cells containing g.19_g.21= and g.19_21del

1.3 Duplication

Definitions

Duplication

a sequence change where, compared to a reference sequence, a copy of one or more nucleotides are inserted directly 3' of the original copy of that sequence.

Description

Format: “prefix” “position(s)_duplicated” “dup” , e.g. g.123_345dup

“prefix” = reference sequence used = g.

“position(s)_duplicated” = position nucleotide or range of nucleotides duplicated = 123_345

“dup” = type of change is a duplication = dup

Note

prefix reference sequences accepted are g., m., c. and n. (genomic, mitochondrial, coding DNA and non-coding DNA).

“positions_duplicated” should contain two different positions, e.g. 123_126 not 123_123.

the “positions_duplicated” should be listed from 5' to 3' , e.g. 123_126 not 126_123.

by definition, duplication may only be used when the additional copy is directly 3' -flanking of the original copy (a “tandem duplication”).

when there is no evidence that the extra copy of a sequence detected is in tandem (directly 3' -flanking the original copy), the change can not be described as a duplication, it should be described as an insertion (see Insertion and proposal SVD-WG003).

inverted duplications are described as insertion (g.234_235ins123_234inv), not as a duplication (see Inversion)

when more then one additional copies are inserted directly 3' of the original copy the change is indicated using the format for Repeated sequences, like [3] (triplication), [4] (quadruplication), etc.

for all descriptions the most 3' position possible of the reference sequence is arbitrarily assigned to have been changed (3' rule)

the 3' rule also applies for changes in single residue stretches and tandem repeats (nucleotide or amino acid)

the 3' rule applies to ALL descriptions (genome, gene, transcript and protein) of a given variant

exception

deletions/duplications around exon/intron and intron/exon borders when identical nucleotides flank these borders (see Question below)

c.546+1dup describes a duplication of a “G” at the exon/intron border ..CAGgtg.. (positions c.546/c.546+1). When RNA analysis shows a G duplication (r.456dup), so no effect on splicing, the change is described as c.546dup.

NOTE: when in the above example the next exon starts with GGT.. the duplication is still described as c.546dup (not c.548dup) but based on a coding RNA sequence as r.548dup.

under discussion, see Proposal for complex variants

(curly braces) can be used to list any change in the duplicated sequence (“positions_duplicated”) which is different when compared to the source, e.g. g.123_345dup234A>G

Examples

g.7dup (one nucleotide)

the duplication of a T at position g.7 in the sequence ACTTACTGCC to ACTTACTTGCC

NOTE: it is allowed to describe the variant as g.7dupT

NOTE: it is not allowed to describe the variant as g.6_7insT (see prioritisation)

g.6_8dup (several nucleotides)

a duplication from position g.6 to g.8 in the sequence ACAATTGCC to ACAATTGCTGCC

NOTE: it is allowed to describe the variant as g.6_8dupTGC

c.120_123+48dup

a duplication of nucleotides c.120 to c.123+48 (coding DNA reference sequence), crossing an exon/intron border

c.123dup

based on the sequence of a genomic DNA sample, a duplication of the A nucleotide c.123 in the sequence CAAgt .../.agAAG to CAAAggt.../.agAAG, i.e. the duplication of the last nucleotide of an exon (see Question below)

NOTE: when RNA is sequenced and the variant does not alter splicing the description at the RNA level based on a coding RNA reference sequence is r.125dup (the 3' rule needs to be applied)

c.4072-1234_5146-246dup

a duplication of nucleotides c.4072-1234 to c.5146-246 duplicating exon 30 (starting at position c.4072) to exon 36 (ending at position c.5145) of the DMD-gene.

NOTE : c.4072-1234_5146-246dupXXXXX, the size of the duplication (XXXXX) should not be described

c.(4071+1_4072-1)_(5145+1_5146-1)dup

a duplication of exon 30 (starting at position c.4072) to exon 36 (ending at position c.5145) of the DMD-gene. The duplication break point has not been sequenced. Exons 29 (ending at c.4071) and 37 (starting at nucleotide c.5146) have been tested and shown to be not duplicated. The duplication therefore starts in intron 29 (position c.4071+1 to c.4072-1) and ends in intron 36 (position c.5145+1 to c.5156-1).

NOTE : this description is part of proposal SVD-WG003 (undecided).

NOTE : previously, the suggestion was made to describe such duplications using the format c.4072-?_5154+?dup. However, since c.4072-? indicates “to an unknown position 5' of c.4072” and c.5154+? “to an unknown position 3' of c.5154” this description is not correct when it is known that exons 29 and 37 are involved.

c.(4071+1_4072-1)_(5145+1_5146-1)[3]

a triplication of exon 30 (starting at position c.4072) to exon 36 (ending at position c.5145) of the DMD-gene (break points not sequenced).

NOTE : this description should only be used when the two additional copies are in tandem with the original copy. There is no specific recommendation yet how to describe such a change but following current recommendations

the format would be something like c.?ins(4071+1_4072-1)_(5145+1_5146-1)[2] ([2] since 2 additional copies have been inserted somewhere in the genome).

c.(?_-30)_(12+1_13-1)dup

a duplication starting somewhere upstream of a gene, last position tested duplicated c.-29, and ending in the intron between nucleotides c.12+1 and c.13-1 (intron 1).

c.(?_-1)_(1_?)dup

a duplication of the entire protein coding region of a gene based on a coding DNA reference sequence).

NOTE: when more details are available regarding the duplication, based on the probes tested to determine its location, the description can be specified like c.(?_-189)_(884_?)dup, meaning the duplication starts 5' of c.-189 and extends 3' of c.*884.

1.4 Insertion

Definitions

Insertion

a sequence change where, compared to the reference sequence, one or more nucleotides are inserted and where the insertion is not a copy of a sequence immediately 5'

Description

Format: “prefix” “positions_flanking” “ins” “inserted_sequence” , e.g. g.123_124insAGC

“prefix” = reference sequence used = g.

“positions_flanking” = position two nucleotides flanking insertion site = 123_124

“ins” = type of change is an insertion = ins

“inserted_sequence” = inserted sequence = AGC

Note

prefix reference sequences accepted are g., m., c. and n. (genomic, mitochondrial, coding DNA and non-coding DNA).

the “position” description should contain two flanking nucleotides, e.g. 123 and 124 but not 123 and 125.

an insertion can not be described using one nucleotide position, like g.123insG

for all descriptions the most 3' position possible of the reference sequence is arbitrarily assigned to have been changed (3' rule)

the 3' rule applies to ALL descriptions (genome, gene, transcript and protein) of a given variant

tandem duplications are described as a duplication (g.123_456dup), not an insertion (g.456_457ins123_456)

inverted duplications are described as insertion (g.234_235ins123_234inv), not as a duplication (see Inversion)

when the inserted sequence is very long it can best be submitted to a database (e.g. GenBank); the accession.version number obtained can then be used to describe the variant like g.123_124insL37425.1:23_361

under discussion, see Proposal for complex variants

(curly braces) can be used to list any change in the inserted sequence (“inserted_sequence”) which is different when compared to the source, e.g. g.123_124ins100_120111A>G

Examples

g.4426_4427insA

the insertion of an A nucleotide between nucleotides g.4426 and g.4427

g.5756_5757insAGG

the insertion of nucleotides AGG between nucleotides g.5756 and g.5757

g.123_124insL37425.1:23_361

the insertion of nucleotides 23 to 361 as described in GenBank file L37425.1 between nucleotides g.123 and g.124

insertion of inverted duplicated copies

g.122_123ins123_234inv

a copy of nucleotides g.123 to g.234 is inserted, in inverted orientation, 5' of the original sequence, between nucleotide g.122 and g.123

g.234_235ins123_234inv

a copy of nucleotides g.123 to g.234 is inserted, in inverted orientation, 3' of the original sequence, between nucleotide g.234 and g.235

g.122_123ins213_234invinsAins123_211inv

an inverted copy of nucleotides g.123 to g.234, with a G>A substitution of nucleotide g.212, is inserted directly 3' of the original sequence

g.122_123ins212_234inv123_199inv

an inverted copy of nucleotides g.123 to g.234, with a deletion from nucleotides g.200 to g.211, is inserted directly 3' of the original sequence

incomplete descriptions, preferably use exact descriptions only

c.(67_70)insG (p.Gly23fs)

the insertion of a G at an unknown position in the sequence encoding amino acid 23

g.549_550insN

the insertion of one not specified nucleotide (N) between position g.549 and g.550

g.15431_15432ins(5) (alternatively g.11_12insNNNNN)

the insertion of 5 not specified nucleotides (NNNNN) between position g.15431 and g.15432

g.1134_1135ins(100)

the insertion of 100 not specified nucleotides between position g.1134 and g.1135

g.?_?insNC_000023.10:(12345_23456)_(34567_45678)

the insertion of a sequence from the X-chromosome (NC_000023.10), maximally involving nucleotides 12345_45678 but certainly nucleotides 23456_34567, at an unknown position (g.?_?) in the genome (see Uncertain)

1.5 Inversion

Definitions

Inversion

a sequence change where, compared to a reference sequence, more than one nucleotide replacing the original sequence are the reverse complement of the original sequence.

Description

Format: “prefix” “positions_inverted” “inv” , e.g. g.123_345inv

“prefix” = reference sequence used = g.

“positions_inverted” = range of nucleotides inverted = 123_345

“inv” = type of change is an inversion = inv

Note

prefix reference sequences accepted are g., m., c. and n. (genomic, mitochondrial, coding DNA and non-coding DNA)

by definition, the region inverted (“positions_inverted”) contains more than one nucleotide. The description g.234inv is therefore not allowed; a one nucleotide inversion should be described as a substitution

for all descriptions the most 3’ position possible of the reference sequence is arbitrarily assigned to have been changed (3’ rule)

the 3’ rule applies to ALL descriptions (genome, gene, transcript and protein) of a given variant

inverted duplications are described as an insertion using the format g.234_235ins123_234inv, not as g.123_456dupinv (see Q&A)

under discussion, see Proposal for complex variants

(curly braces) can be used to list any change in the inverted sequence (“positions_inverted”) which is different when compared to the source, e.g. g.123_345inv233A>G

inversions are not used on Protein level

Examples

g.1077_1080inv

inversion of nucleotides g.1077 to g.1080, changing ..AGGCTGATT.. to ..AGGTCAGTT..

c.77_80inv

inversion of nucleotides c.77 to c.80 (coding DNA reference sequence), changing ..AGGCTGATT.. to ..AGGTCAGTT..

g.203_506inv

inversion of the 304 nucleotides from position g.203 to g.506

g.111754331_111966764inv

a large inversion (212,434 nucleotides) from position g.111754331 to g.111966764

g.122_123ins123_234inv

a copy of nucleotides g.123 to g.234 is inserted, in inverted orientation, 5' of the original sequence, between nucleotide g.122 and g.123

g.234_235ins123_234inv

a copy of nucleotides g.123 to g.234 is inserted, in inverted orientation, 3' of the original sequence, between nucleotide g.234 and g.235

g.122_123ins213_234invinsAins123_211inv

an inverted copy of nucleotides g.123 to g.234, with a G>A substitution of nucleotide g.212, is inserted directly 3' of the original sequence

g.122_123ins212_234inv123_199inv

an inverted copy of nucleotides g.123 to g.234, with a deletion from nucleotides g.200 to g.211, is inserted directly 3' of the original sequence

1.6 Conversion

Definitions

Conversion

a sequence change where, compared to a reference sequence, a range of nucleotides are replaced by a sequence from elsewhere in the genome

Description

Format: “prefix” “positions_converted” “con” “positions_replacing_sequence”, e.g. g.123_345con888_1110

“prefix” = reference sequence used = g.

“positions_converted” = range of nucleotides converted = g.123_345

“con” = type of change is a conversion = con

“positions_replacing_sequence” = description of replacing sequence = 888_1110

Note

prefix reference sequences accepted are g., m., c. and n. (genomic, mitochondrial, coding DNA and non-coding DNA)

the region converted (“positions_converted”) should start and end with a variant nucleotide

for all descriptions the most 3' position possible of the reference sequence is arbitrarily assigned to have been changed (3' rule)

the 3' rule applies to ALL descriptions (genome, gene, transcript and protein) of a given variant

under discussion, see Proposal for complex variants

(curly braces) can be used to list any change in the converted sequence (“positions_replacing_sequence”) which is different when compared to the source, e.g. g.123_345con888_1110999A>G

Examples

g.333_590con1844_2101

conversion replacing nucleotides g.333 to g.590 with nucleotides g.1844 to g.2101 from the same genomic reference sequence

g.415_1655conAC096506.5:g.409_1683

conversion replacing nucleotides g.414 to g.1655 with nucleotides g.409 to g.1683 as found in the genomic reference sequence AC096506.5

NC_000022.10:g.42522624_42522669con42536337_42536382

conversion in exon 9 of the CYP2D6 gene replacing exon 9 nucleotides g.42522624 to g.42522669 with those of the 3' flanking CYP2D7P1 gene, nucleotides g.42536337 to g.42536382

c.15_355conNM_004006.1:20_360

conversion replacing nucleotides c.15 to c.355 with nucleotides c.20 to c.360 as found in the coding DNA sequence file NM_004006.1

1.7 Deletion-insertion

Definitions

Deletion-insertion (delins)

a sequence change where, compared to a reference sequence, one or more nucleotides are replaced by one or more other nucleotides and which is not a substitution, inversion or conversion.

Description

Format: “prefix” “position(s)_deleted” “delins” “inserted_sequence” , e.g. g.123_127delinsAG

“prefix” = reference sequence used = g.

“position(s)_deleted” = position nucleotide or range of nucleotides deleted = 123_127

“delins” = type of change is a deletion-insertion (indel) = delins

“inserted_sequence” = description inserted sequence = AG

Note

prefix reference sequences accepted are g., m., c. and n. (genomic, mitochondrial, coding DNA and non-coding DNA).

by definition, when one nucleotide is replaced by one other nucleotide the change is a substitution.

two variants separated by one or more nucleotides should preferably be described individually and not as a “delins”

two variants separated by one nucleotide, together affecting one amino acid, can be described as a “delins” (e.g. c.142_144delinsTGG (p.Arg48Trp))

for all descriptions the most 3' position possible of the reference sequence is arbitrarily assigned to have been changed (3' rule)

Examples

g.6775delinsGA

a deletion of nucleotide g.6775 (a T, not described), replaced by nucleotides GA, changing ..AGGCTCATT.. to ..AGGCGACATT..

g.6775_6777delinsC

a deletion of nucleotides g.6775 to g.6777 (TCA, not described), replaced by nucleotides C, changing ..AGGCTCATT.. to ..AGGCCTT..

c.142_144delinsTGG (p.Arg48Trp)

a deletion replacing nucleotides c.142 to c.144 (CGA, not described) with TGG

NOTE: the variant can also be described as c.[142C>T;144A>G], i.e. two substitutions. This format is preferred when either of the two variants is known as a frequently occurring variant (“polymorphism”).

g.9002_9009delinsTTT

a deletion of nucleotides g.9002 to g.9009, replaced by nucleotides TTT

1.8 Alleles

Definitions

Allele

a series of variants on one chromosome.

Description

Format (one allele): “prefix” [“change1” ; “change2”], e.g. g.[123G>A;345del]

“prefix” = reference sequence used = g.

[= opening symbol for allele = [

“change1” = description first variant = 123G>A

; = separator symbol two changes = ;

“change2” = description second variant = 345del

] = closing symbol for allele =]

Format (two alleles): “prefix” [“change”]; [“change”], e.g. g.[123G>A];[345del]

“prefix” = reference sequence used = g.

[= opening symbol for allele-1 = [

“change” = description variant = 123G>A

];[= closing symbol for allele-1, separator symbol two alleles, opening symbol for allele-2 =];[

“change” = description variant = 345del

] = closing symbol for allele-2 =]

Note

humans are diploid organisms and have two alleles at each genetic locus, with one allele inherited from each parent

when two variants are identified in a gene that are on one chromosome (in cis) this should be described as “g.[variant1;variant2]” .

when two variants are identified in a gene that are on different chromosomes (in trans) this should be described as “g.[variant1];[variant2]” .

when two variants are identified in a gene, but when it is not known whether these are on one chromosome (in cis) or on different chromosomes (in trans), this should be described as “variant1(;variant)” , i.e. without using “[]” . NOTE: it is recommended to determine whether the changes are on the same chromosome or not.

descriptions combining variants based on different reference sequence types (e.g. c.[76A>C];g.[10091C>G]) should not be used.

Examples

LRG_199t1:c.[2376G>C;3103del]

one allele (chromosome) of a gene contains two different changes, g.2376G>C and c.3103del. The variants are found in cis.

LRG_199t1:c.[2376G>C];[3103del]

the two alleles (chromosomes) of a gene each contain a different change, c.2376G>C and c.3103del. A heterozygous case (compound heterozygote, e.g. in a recessive disease). The variants are found in trans.

LRG_199t1:c.2376[G>C];[G>C]

both alleles (chromosomes) of a gene contain the same variant, c.2376G>C. A homozygous case (e.g. in a recessive disease).

LRG_199t1:c.2376G>C(;3103del)

two variants in a gene, c.2376G>C and c.3103del, but it is not known whether they are on the same or on different alleles (chromosomes).

NOTE: when it is not known on which allele a variant is, allele brackets should not be used

LRG_199t1:c.2376[G>C];[(G>C)]

analysis detects one variant (c.2376G>C), suggesting both alleles (chromosomes) contain this variants but it can not be excluded the other allele is deleted.

LRG_199t1:c.2376[G>C];[=]

one allele (chromosome) of a gene contains a variant, c.2376G>C, the other allele (chromosome) contains the reference sequence, c.2376= (is wild-type).

NOTE: the description c.[2376G>C];[=], containing c.2376G>C and c.=, is different since it indicates the entire coding DNA reference sequence was analysed and the only variant identified was c.2376G>C (on one allele).

NOTE: for other variant types the format is c.2376[del];[=], c.2376_2399[dup];[=], c.2376_2377[insGT];[=], etc.

LRG_199t1:c.[2376G>C];[?]

one allele (chromosomes) of a gene contains a variant, c.2376G>C, while a variant for the other allele is expected but not yet identified (g.?) (e.g. in individuals affected by a recessive disease).

LRG_199t1:c.[296T>G;476C>T;1083A>C];[296T>G;1083A>C]

a sample contains variants c.296T>G and c.1083A>C on both alleles (chromosomes) and variant c.476C>T on only one allele.

NM_004006.2:c.[296T>G;476C>T];[476C>T](;)1083A>C

a sample contains a homozygous variant (c.476C>T) and two heterozygous variants (c.296T>G and c.1083G>C) for which it is not known on which allele (chromosome) they are.

LRG_199t1:c.[296T>G];[476C>T](;)1083G>C(;);1406del

a sample contains heterozygous variants on different alleles (c.296T>G and c.476C>T) and two additional heterozygous variants (c.1083G>C and c.1406del) for which it is not known on which allele (chromosome) they are.

NC_000023.10:g.[30683643A>G;33038273T>G]

one allele (X-chromosome) contains two different changes in two different genes, g.30683643A>G in the GK gene and g.33038273T>G in the DMD gene.

c.[NM_000167.5:94A>G;NM_004006.2:76A>C]

one allele contains two different changes in two different genes, NM_000167.5:c.94A>G in the GK gene and NM_004006.2:c.76A>C in the DMD gene.

1.9 Repeated sequences

Definitions

Repeated sequence

a sequence where, compared to a reference sequence, a segment of one or more nucleotides (the repeat unit) is present several times, one after the other.

Description

Format (repeat position): “prefix” “position_repeat_unit” “ ” [“ ” copy_number” ”] ” , e.g. g.123_125[36]

“prefix” = reference sequence used = g.

“position_repeat_unit” = position (range) first repeat copy = 123_125

[= opening symbol for allele = [

“copy_number” = number of repeat units = 36

] = closing symbol for allele =]

Format (sequence): “prefix” “position_repeat_start” “repeat_sequence” ” [“ ” copy_number” ”] ” , e.g. g.123GGC[36]

“prefix” = reference sequence used = g.

“position_repeat_start” = position first nucleotide repeat unit = 123

“repeat_sequence” = nucleotide sequence repeat copy = GGC

[= opening symbol for allele = [

“copy_number” = number of repeat units = 36

] = closing symbol for allele =]

Note

reference sequences accepted are g., m., c. and n. (genomic, mitochondrial, coding DNA and non-coding DNA). repeated sequences include both small (mono-, di-, tri-, etc., nucleotide) and larger (kilobase-sized) repeats.

the format based on repeat position is preferred, descriptions of the repeat sequence quickly become too lengthy.

NOTE: while g.123CAG[23] describes a repeat of 23 CAG units, g.123_125[23] describes a tri-nucleotide repeat of 23 units which could be interrupted with other units (e.g. a rare CAA). The description g.123CAG[23] can thus only be used when the repeat was sequenced.

the format g.123_124TG[4], should not be used; it contains redundant information (“123_124” and “TG”).

for composite repeats the basic format can be used, successively listing each different repeat unit; g.456_467[4]468_494[9]4

Examples

g.123_124[14] (when sequenced, alternatively g.123TG[14])

a repeated di-nucleotide sequence, with the first unit located from position g.123 to g.124, is present in 14 copies.

NOTE: when the repeat is variable in the population, sequenced, and the reference sequence has 15 units, the description g.123TG[14] is preferred over g.151_152del

NOTE: when the repeat is variable in the population, sequenced, and the reference sequence has 15 units, the description g.123TG[17] is preferred over g.149_152dup

g.123_124[14];[18] (when sequenced, alternatively g.123TG[14];[18])

a repeated di-nucleotide sequence, with the first unit located from position g.123 to g.124, is present in 14 copies on one allele and 18 copies on the other allele

FMR1 GGC-repeat

in literature the Fragile-X tri-nucleotide repeat is known as the CGG-repeat. However, based on a coding DNA reference sequence (GenBank NM_002024.5) and applying the 3’ rule, the repeat has to be described as a GGC-repeat see Recommendations.

c.-128_-126[79]

an extended repeat of exactly 79 units

NOTE : c.-128GGC[79] can only be used when the repeat has been sequenced, excluding it is interrupted by one or more GGA-triplets

c.-128_-126[(600_800)]

the repeated tri-nucleotide sequence, starting at position c.-128, has an estimated size of between 600 to 800 copies.

NOTE: the repeat can be pure or a mix of GGC and GGA triplets.

HD GCA-repeat

based on the HTT (huntingtin) coding DNA reference sequence (GenBank LRG_763t1 (NM_002111.8), applying the 3’ rule, the Huntington’ s Disease tri-nucleotide repeat is described as an GCA (not CAG) repeat.

c.54GCA[21]

NOTE: the coding DNA reference sequence (LRG_763t1 (NM_002111.8)) was determined and shown to contain an allele of 21 GCA repeats

NOTE: on protein level the reference allele contains 23 Gln's, described as p.Gln18[23] (alternatively p.Q18[23]). The difference derives from the fact that the GCA repeat is interrupted by ACA-triplet ("CAA" coding) at position 20.

c.54GCA[21]ACA[1]GCC[2]ACC[1]GCC[10]

the coding DNA reference sequence (LRG_763t1 (NM_002111.8)) was determined and shown to contain a tri-nucleotide allele of 21 GCA, 1 ACA, 2 GCC, 1 ACC and 10 GCC-repeats.

NOTE: when the sequence was not determined, but the repeat estimated based on PCR fragment size, the description is c.(54_56;117_119;120_122;126_128;129_131)[35]

g.456_457[4]466_468[9]490[12] (when sequenced, alternatively 456TG[4]TAA[9]T[12])

a complex repeated sequence has a first unit located from position g.456 to g.457, present in 4 copies, a second unit from position g.466 to g.468 present in 9 copies and a third unit (mono-nucleotide) starting at position position 490 present in 12 copies.

1.10 Complex

Definitions

Complex

a sequence change where, compared to a reference sequence, a range of changes occur that can not be described as one of the basic variant types (substitution, deletion, duplication, insertion, conversion, inversion, deletion-insertion, or repeated sequence).

Description

Sequence changes can be very complex, involving a range of changes at one specific location. The description of such changes can become rather complicated and at some point, although literally correct, effectively meaningless. Examples of complex changes, and suggestions how to describe them, have been published by Taschner and Den Dunnen. The topic is also discussed in SVD-WG004 (ISCN<>HGVS).

mosaicism

g.[17333296T=>A] describes a mosaic case where at position g.17333296 besides the normal sequence (a T, described as 'T=') also chromosomes are found containing an A (g.17333296T>A)

chimerism

g.[1323887G=//>C] describes a chimeric case where at position g.1323887 besides the normal sequence (a G, described as 'G=') also cells are found containing an C (c.1323887G>C)

translocation

translocations are described using the recommendations of the ISCN.