

Last modified June 23, 2015

Since references to WWW-sites are not yet acknowledged as citations, please mention [den Dunnen JT and Antonarakis SE \(2000\). Hum.Mutat. 15: 7-12](#) when referring to these pages.

Contents

- **Recommendations**

- [general recommendations](#)
 - [nucleotide numbering](#)
- **changes at DNA level**
 - [general](#)
 - [substitutions](#)
 - [deletions](#)
 - [duplications](#)
 - [insertions](#)
 - [insertion/deletions \(indels\)](#)
 - **NEW** [inversions](#)
 - **NEW** [conversions](#)
 - **NEW** [translocations](#)
 - **NEW** [two or more changes in one chromosome](#) (incl. mosaicism, chimerism)
 - **NEW** [repeated sequences \(incl. short sequence repeats\)](#)
 - **NEW** [uncertainties](#)
 - **NEW** [complex](#)
- [changes at RNA level](#)
- [changes at protein level](#)

- **Explanations / examples**
 - [Quick Reference](#)
 - [changes at DNA level](#)
 - [changes at RNA level](#)
 - [changes at protein level](#)
-

DNA level

(suggestions extending the [published](#) recommendations are **NEW** in italics)

- **nucleotides**

description of nucleotides at DNA level follows the recommendations of the [IUPAC-IUBMB](#). Nucleotides are designated by the bases, in upper case, A (adenine), C (cytosine), G (guanine), T (thymidine), including those for uncertain nucleotides like Y (pYrimidine) and R (puRine), ([see Standards](#)).
 - **nucleotide numbering** (for details and examples [see Reference Sequence discussions](#))
 - coding DNA reference sequence ([see Examples and Figure](#))
 - there is no nucleotide 0
 - nucleotide 1 is the A of the ATG-translation initiation codon
 - the nucleotide **5' of the ATG-translation initiation** codon is -1, the previous -2, etc.
 - **NEW** *the nucleotide 3' of the translation stop codon is *1, the next *2, etc.*
 - **intronic nucleotides** (coding DNA reference sequence only)
 - beginning of the intron: the number of the last nucleotide of the preceding exon, a plus sign and the position in the intron, like c.77+1G, c.77+2T, ...
 - end of the intron: the number of the first nucleotide of the following exon, a minus sign and the position upstream in the intron, like ..., c.78-2A, c.78-1G.
 - in the middle of the intron, numbering changes from "c.77+.." to "c.78-.."; for introns with an uneven number of nucleotides the central nucleotide is the last described with a "+" ([see Discussion](#))
 - **NOTE**: the format c.IVS1+1G and c.IVS1-2G should not be used ([see Discussion](#))
 - genomic reference sequence ([see Examples and Figure](#))
 - nucleotide numbering starts with 1 at the first nucleotide of the sequence
NOTE: the sequence should include all nucleotides covering the sequence (gene) of interest and should start well 5' of the promoter of a gene
 - no +, - or other signs are used
 - when the complete genomic sequence is not known, a coding DNA reference sequence should be used
 - for all descriptions the **most 3' position** possible is arbitrarily assigned to have been changed ([see Exception](#))
-

Substitutions

A nucleotide **substitution** is a sequence change where one nucleotide is replaced by one other nucleotide ([see Standards - Definition](#)). Nucleotide substitutions are described using a ">"-character (indicating "changes to").

NOTE: changes involving two or more consecutive nucleotides are described as deletion/insertions (indels, [see Deletion/insertions](#)).

- c.76A>C denotes that at nucleotide 76 an A is changed to a C
- c.-14G>C denotes a G to C substitution 14 nucleotides 5' of the ATG translation initiation codon
- c.88+1G>T denotes the G to T substitution at nucleotide +1 of an intron (in the coding DNA positioned between nucleotides 88 and 89)
- c.89-2A>C denotes the A to C substitution at nucleotide -2 of an intron (in the coding DNA positioned between nucleotides 88 and 89)
- c.*46T>A denotes a T to A substitution 46 nucleotides 3' of the translation termination codon
- the description c.76_77delinsTT is preferred over c.[76A>T; 77G>T]

NEW NOTE: based on the definition of a substitution ([see Standards - Definition](#); *one nucleotide replaced by one other nucleotide*) this change can not be described as a substitution (like c.76_77AG>TT or c.76AG>TT)

NOTE: it is not correct to describe "polymorphisms" as c.76A/G ([see Discussion](#)).

Deletions

A nucleotide deletion is a sequence change where one or more nucleotides are removed ([see Standards - Definition](#)). Deletions are described using "**del**" after an indication of the first and last nucleotide(s) deleted, separated by a "_" (*underscore*). For all descriptions the **most 3' position** possible is arbitrarily assigned to have been changed.

NOTE: to discriminate known variable sequences from other changes it is recommended to describe individual alleles differing from the reference sequence like g.210T[5] (preferred over g.210_211delTT) or g.121T[9] (preferred over g.210_211dupTT) ([see Repeated sequences](#)).

- c.76_78del (alternatively c.76_78delACT) denotes a ACT deletion from nucleotides 76 to 78
- deletions with **uncharacterised breakpoints** ([see Uncertainties](#))
 - c.(87+1_88-1)_(923+1_924-1)del denotes a deletion of exons 3 to 7 starting at an unknown position in intron 2 (between coding DNA nucleotides 87+1 and 88-1) and ending at an unknown position in intron 7 (between coding DNA nucleotides 923+1 and 924-1). The description indicates that exons 2 and 8 have been tested and shown not to be deleted
 - NOTE:** the description c.88-?_923+?del does not specify start/end of the deletion and is not correct when flanking sequences have been tested ([see Uncertainties](#))
 - c.(?-30)_(*220?)del denotes the deletion of the entire gene (coding DNA reference sequence running from -30 (cap site) to *220 (polyA-addition site))

- **NEW** c.88+101_oGJB2:c.355-1045del denotes a deletion which ends in the flanking GJB2 gene at position 355-1045 (in the intron between nucleotides 354 and 355) on the reverse strand (the genes are thus located and fused in opposite transcriptional directions, [see Discussion](#))
 - for all descriptions the **most 3' position** possible is arbitrarily assigned to have been changed ([see FAQ](#));
 - ACTTTGTGCC to ACTTGCC is described as c.5_7del (c.5_7delTGT, not as c.4_6delTTG)
 - ctttagGCATG to ctttagGCATG in an intron is described as c.301-3delT (not as c.301-5delT)
 - TCACTGTCTGCGGTAATC to TCACTG CGGTAATC is described as c.7_10del (c.7_10delTCTG) and not as c.4_7del (c.4_7delCTGT).
 - AAAGAAGAGGAG to AAAG GAG is described as c.5_9del (c.5_9delAAGAG) and not as c.3_7del (c.3_7delAGAAG)
 - **Exceptions**
 - using a *coding DNA reference sequence* there is **an exception** to the rule around exon/intron and exon/exon borders when identical nucleotides flank the exon/intron or exon/exon border;
 - when the exon 3/intron 3 border is ..CAGgtg.. and RNA analysis shows no effect on splicing but a deletion of a G the change ..CAGgtg.. to ..CAgtg.. is described as c.3delG **and not** c.3+1delG.
 - when exon 3 ends with ..CAA.. and exon 4 starts with ..ACG.. and the sequence of genomic DNA shows that the last A-nucleotide of exon 3 is deleted (and not the first A-nucleotide in exon 4), the deletion changing ..CAAACG.. to ..CAACG.. is described as c.3delA **and not** c.4delA
 - c.1210-12T(5_9) (*not c.1210-6T(5_9)*) describes the variable stretch of 5 to 9 T-residues in intron 9 of the CFTR gene. The most commonly used CFTR coding DNA reference sequence contains a stretch of 7 T's ([see Repeated sequences](#)).
-

Duplications

Duplications are designated by "**dup**" after an indication of the first and last nucleotide(s) duplicated. It should be noted that the description "**dup**" ([see Standards](#)) may **by definition** only be used when the sequence copy is directly 3'-flanking the original copy. For all descriptions the **most 3' position** possible is arbitrarily assigned to have been changed. For the addition of more than 1 copy (3, 4, 5, etc.) [see Repeated sequences](#) and [see Discussion](#).

NOTE: to discriminate known variable sequences from other changes it is recommended to describe individual alleles differing from the reference sequence like g.210T[5] (preferred over g.215_216del) or g.210T[9] (preferred over g.215_216dup) ([see Repeated sequences](#)).

- duplicating insertions should be described as duplications ([see Discussion](#))
 - g.5dupT (or g.5dup, *not g.5_6insT*) denotes a duplication ("*insertion*") of the T nucleotide at position 5 in the genomic reference sequence changing ACTCTGTGCC to ACTCT**T**GTGCC
 - g.7dupT (or g.7dup, *not g.5dupT*, *not g.7_8insT*) denotes a duplication ("*insertion*") of the T nucleotide at position 7 in the genomic reference sequence changing AGACTTTGTGCC to AGACTTT**T**GTGCC
 - g.7_8dup (or g.7_8dupTG, *not g.5_6dup*, *not g.8_9insTG*) denotes a TG duplication in the TG-tandem repeat sequence changing ACTTTGTGCC to ACTTTGTG**TG**CC
 - g.7_8[4] (or g.5_6[4], or g.5TG[4], *not g.7_10dup*) is the preferred description of the addition of two extra TG's to the variable

TG repeated sequence changing ACTTTGTGCC to ACTTTGTGTGTGCC ([see Repeated sequences](#))

- c.77_79dup (or c.77_79dupCTG) denotes that the three nucleotides 77 to 79 are duplicated (present twice)
 - duplications with **uncharacterised breakpoints** ([see Uncertainties](#))
 - c.(87+1_88-1)_(301+1_302-1)dup denotes a duplication of exons 3 to 4 starting at an unknown position in intron 2 (between coding DNA nucleotides 87+1 and 88-1) and ending at an unknown position in intron 5 (between coding DNA nucleotides 301+1 and 302-1). The description indicates that exons 2 and 5 have been tested and shown not to be duplicated
NOTE: the description c.88-?_301+?dup does not specify start/end of the duplication and is not correct when flanking sequences have been tested ([see Uncertainties](#))
NOTE: the description "**dup**" ([see Standards](#)) may **by definition** only be used when the additional copy is directly 3'-flanking of the original copy (tandem duplication). In many cases there will be **no experimental proof**, the additional copy may be anywhere in the genome (i.e. inserted / transposed). It has been suggested that, unless there is experimental evidence, exonic duplications should be described using the format, i.e. c.88-?_301+?(2) ([see Recommendations](#)).
 - **NEW** c.(1031+1_1032-1)_(1357+1_1358+1)(3) denotes the presence of two additional copies of an exon, starting at an unknown position in the flanking upstream intron (upstream of coding DNA nucleotide 1032) and ending at an unknown position in the flanking downstream intron (downstream of coding DNA nucleotide 1357). The description, using "(3)" indicates there is no proof that the three copies are in tandem ([see Repeated sequences](#), [see Uncertainties](#))
 - **NEW** c.(1031+1_1032-1)_(1357+1_1358+1)[3] denotes a **direct triplication** of an exon, starting at an unknown position in the flanking upstream intron (upstream of coding DNA nucleotide 1032) and ending at an unknown position in the flanking downstream intron (downstream of coding DNA nucleotide 1357) ([see Repeated sequences](#))
-

Insertions

Insertions are designated by "**ins**" after an indication of the nucleotides flanking the insertion site, followed by a description of the nucleotides inserted. Duplicating insertions should be described as duplications ([see Discussion](#)), not as insertion. For large insertions the number of inserted nucleotides should be mentioned, together with an accession.version number referring to a sequence database file containing the complete inserted sequence.

- c.76_77insT denotes that a T is inserted between nucleotides 76 and 77 of the coding DNA reference sequence
 - c.123+54_123+55insAB012345.2:g.76_420 denotes an intronic insertion (between nucleotides c.123+54 and 123+55) of 345 nucleotides (nucleotides 76 to 420 like in GenBank file AB012345 version 2)
NOTE: descriptions like c.123+54_123+55ins345 and c.123+54_123+55insAlu are not allowed: "ins345" and "insAlu" are not specified and the description can not be used to reconstruct the exact change described.
-

Deletion / insertions (indels)

Deletion/insertions of two or more consecutive nucleotides (indels) are described as a **deletion followed by an insertion** ([see Discussion](#)).

- c.112_117delinsTG (alternatively c.112_117delAGGTCAinsTG) denotes the replacement of nucleotides 112 to 117 (AGGTCA) by TG
 - c.113delinsTACTAGC (alternatively c.113delGinsTACTAGC) denotes the replacement of nucleotide 113 by 7 new nucleotides (TACTACG)
 - c.114_115delinsA (alternative c.[114G>A; 115delT])
-

NEW Inversions

Inversions are designated by "**inv**" after an indication of the first and last nucleotides affected by the inversion.

- *c.203_506inv (or 203_506inv304) denotes that the 304 nucleotides from position 203 to 506 have been inverted*
-

NEW Conversions

Conversions are designated by "**con**" after an indication of the first and last nucleotides affected by the conversion, followed by a description of the origin of the new nucleotides ([see Discussion](#)).

- *g.123_678conNG_012232.1:g.9456_10011 describes a gene conversion replacing nucleotides 123 to 678 of the reference genomic sequence with nucleotides 9456 to 10011 from the sequence as present in GenBank file NG_012232.1*
-

NEW Translocations

Translocations are described at the molecular level using the format "t(X;4)(p21.2;q34)", followed by the usual numbering, indicating the position translocation breakpoint. The sequences of the translocation breakpoints need to be submitted to a sequence database (Genbank, EMBL, DDJB) and the accession.version numbers should be given ([see Discussion](#)).

- *t(X;4)(p21.2;q35)(c.857+101_857+102) denotes a translocation breakpoint in the intron between coding DNA nucleotides 857+101 and 857+102, joining chromosome bands Xp21.2 and 4q34*
-

More changes in one individual

Two or more changes in a gene are described by combining the changes, per chromosome (maternal and paternal), between square brackets ("**[;];[;]**") and using a semicolon ("**;**") as separator: *[first change maternal; second change maternal]; [first change paternal; second*

change paternal]" ([see Discussion](#)). When changes are in different genes on different chromosomes a space (" ") is used to separate the different chromosomes ("[:] [:]").

NOTE: mixed descriptions like *c.[76A>C];g.[91C>G]* should not be used.

- **two changes in one gene on one chromosome**

c.[76A>C; 83G>C] describes two changes found in a gene on one chromosome; A to C change at nucleotide 76 and a G to C change at nucleotide 83

- **two changes in one gene on both chromosomes** (e.g. in *recessive diseases*)

c.[76A>C];[83G>C] describes two changes found in a gene on each chromosome (one paternal, one maternal); A to C change at nucleotide 76 on one chromosome and a G to C change at nucleotide 83 on the other chromosome

Examples

- *c.[76A>C];[76A>C]* denotes a homozygous A to C change at nucleotide 76
- *c.[76A>C];[(76A>C)]* denotes a homozygous A to C change at nucleotide 76, not confirmed by analysis of both parents, leaving the possibility of non-amplification of the sequences analysed on the other chromosome (e.g. due to a primer mismatch or a deletion)
- *c.[76A>C];[?]* denotes a A to C change at nucleotide 76 in a gene on one chromosome and an expected not yet detected change on the other chromosome
- *c.[76A>C];[=]* denotes a A to C change at nucleotide 76 in a gene on one chromosome and a normal coding DNA Reference Sequence of the other chromosome ([see FAQ](#))
- **NEW** *c.[76A>C];[0]* denotes a A to C change at nucleotide 76 in a gene on one chromosome and the absence of the entire coding DNA Reference Sequence on the other chromosome

NOTE: the description *c.0* should **preferably not be used**, it does not specify the extent (begin / end) of the deletion.

- *c.[350G>A(;)1210-12T[7];[9](;)1521_1523del]* describes a case where variants *c.350G>A*, *c.1210-12T[7]*, *c.1210-12T[9]* and *c.1521_1523del* were detected but without information on which variants are found together on one chromosome


- **NEW two changes in one gene with chromosomes unknown** are described as "[change1 (;) change2]" ([see FAQ](#))

c.[76A>C(;)283G>C] denotes that two changes were identified in one individual (an A to C change at nucleotide 76 and a G to C change at nucleotide 283), but it is not known whether these changes are on the same chromosome (in cis) or on different chromosomes (in trans)

- **changes in different genes on one chromosome** are described as "[change1;change2]"

When a coding DNA reference sequence is used the description should clearly indicate based on which reference sequence each variant is described. This can be done using either the **accession.version number** or the **Gene Symbol** (i.e. when elsewhere the reference sequence connected to each gene is specified).

Examples

- *hg19 chrX:g.[30683643A>G;33038273T>G]*
describes a A to G change at nucleotide g.30683643 (GK gene) and a T to G change at nucleotide g.33038273 (DMD gene) on one X-chromosome based on the genomic reference sequence of genome build hg19.
 - *c.[NM_000167.5:94A>G;NM_004006.2:76A>C]*
describes a A to G change at nucleotide c.94 (GK gene, based on coding DNA reference sequence NM_000167.5) and a A to C change at nucleotide c.76 (DMD gene, based on coding DNA reference sequence NM_004006.2) on one X-chromosome.
 - *c.[GK:94A>G;DMD:76A>C]*
describes a A to G change at nucleotide c.94 (GK gene) and a A to C change at nucleotide c.76 (DMD gene) on one X-chromosome. Elsewhere the coding DNA reference sequences are specified as NM_000167.5 for GK and NM_004006.2 for DMD.
-  **changes in different genes on different chromosome** are described as "[change1] [change2]"
*When a coding DNA reference sequence is used the description should clearly indicate based on which reference sequence each variant is described. This can be done using either the **accession.version number** or the **Gene Symbol** (i.e. when elsewhere the reference sequence connected to each gene is specified).*

Examples

- *hg19 chr1:g.[35227587C>G] chr13:g.[20763083A>T]*
describes a C to G change at nucleotide g.35227587 (GJB4 gene) on chromosome-1 and an A to T change at nucleotide g.35250398 (GJB2 gene) on chromosome-13 based on the genomic reference sequence of genome build hg19.
- *NM_153212.2:c.[732C>G] NM_004004.5:c.[638T>A]*
describes a C to G change at nucleotide c.732 (GJB4 gene, chromosome 1, based on coding DNA reference sequence NM_153212.2) and an A to T change at nucleotide c.638 (GJB2 gene, chromosome 13, based on coding DNA reference sequence NM_004004.5).
- *GJB4:c.[732C>G] GJB2:c.[638T>A]*
describes a C to G change at nucleotide c.732 (GJB4 gene, chromosome 1) and deletion of a G at nucleotide c.35 (GJB3 gene, chromosome 13). Elsewhere the coding DNA reference sequences are specified as NM_153212.2 for GJB4 and NM_004004.5 for GJB2.

Mosaicism

NEW *Mosaicism - two different nucleotides in one position caused by somatic mosaicism* are described as "[=/nucleotide 2]" ([see FAQ](#)).

- *c.[83G=/83G>C]* describes a mosaic case where at position 83 besides the normal sequence (a G, described as '=') also

chromosomes are found containing a C (c.83G>C)

Chimerism

NEW *Chimerism - two different nucleotides in one position caused by chimerism are described as "[=//nucleotide 2]"*


- *c.[=//83G>C] describes a chimeric case where at position 83 besides the normal sequence (a G, described as '=') also cells are found containing another chromosome containing a C at this position (c.83G>C)*

NEW Repeated sequences

A frequently occurring sequence change is the variability of repeated sequences. Within this category we discriminate both small sequences (mono-, di-, tri-, etc nucleotide repeats) as well as the much larger ones. Such changes are described using the format "***position-first-repeat-unit_[number]***" (e.g. g.123_124[4]) where ***position-first-repeat-unit*** gives the location of the first unit of the variable sequence repeat and ***[number]*** the number of units present in the allele described.

- the first unit of the repeat is preferably described based on position, like. g.123_124. *For short/simple repeats it is acceptable to include the content of the repeated unit, using the format "position-first-nucleotide-repeat_content" like g.123TG[4]. Do not use a mix of these descriptions like g.123_124TG[4]. This contains redundant information (123_124 and TG) with the danger of being in conflict.*

NOTE: *including the content of the sequence involved quickly gives descriptions which become too lengthy.*

- uncertainties regarding the number of repeated copies are given between brackets, like c.-128GGC[(600_800)].
-  the repeated sequences may have complex structures, consisting of a mix of several repeated sequences directly following each other. When sequenced, one should describe it including the individual repeated elements, like g.456TG[4]TA[9]TG[3] (or g.456_465[4]466_489[9]490_499[3]). When not sequenced, but ***based on fragment size***, it should be described like g.456_465[16].
- the same format can be used to describe the presence of multiple copies (triplication, quadruplication, etc.) of larger sequences, e.g. exons. Square brackets ("[]") should only be used when there is experimental evidence that the additional copies are in tandem on the same chromosome.

NOTE: *the description "dup" ([see Standards](#)) may by definition only be used when the additional copy is directly 3'-flanking of the original copy (tandem duplication). In many cases there will be **no experimental proof**, the additional copy may be anywhere in the genome (i.e. inserted / transposed). It has been suggested that, unless there is experimental evidence, exonic duplications should be described using the format, i.e. c.88-?_301+?(2).*

Examples

- g.123_124[4] (alternatively g.123TG[4]) describes a sequence variable in the number of TG repeats where the first unit is present at

position g.123_124 in the genomic reference sequence and the allele described contains 4 units.

- c.1210-12T[5_9] describes a famous variable stretch of 5 to 9 T-residues in intron 9 of the CFTR gene. The most commonly used CFTR coding DNA reference sequence contains a stretch of 7 T's. It is recommended to describe individual alleles differing from this reference sequence as c.1210-12T[5] (*not c.1210-7_1210-6delTT*) or c.1210-12T[9] (*not c.1210-7_1210-6dupTT*).

NOTE: the repeat should not be described as c.1210-6T[5_9]

- c.123+74TG[3_6] (*alternatively c.123+74_123+75[3_6]*) indicates that a TG di-nucleotide repeat is present, starting at nucleotide 74 in the intron following cDNA nucleotide c.123, which is found varying from 3 to 6 copies in the population
 - c.123+74TG[4];[5] denotes that a person carries a TG di-nucleotide repeat of length 4 on one chromosome and of length 5 on the other chromosome
- in literature the Fragile-X tri-nucleotide repeat is known as the CGG-repeat, but based on the coding DNA reference sequence (GenBank NM_002024.4) and the rule that for all descriptions the **most 3' position possible** should be arbitrarily assigned the repeat has to be described as a GGC-repeat. In addition the repeat is interrupted by GGA triplets (see e.g. [Eichler 1995](#)) making it a complex repeat which can not be accurately described based on sizing only. The sequence represented by the FMR1 coding DNA Reference Sequence (GenBank [NM_002024.5](#)) is c.-128GGC[9]GGA[1]GGC[10].

NOTE: the sequence represented by coding DNA reference sequence [NM_002024.3](#) is c.GGC[9]GGA[1]GGC[9]GGA[1]GGC[10]. Based on this reference sequence descriptions of variants will be significantly different. To prevent such differences the recommendation is to use a LRG (Locus Reference Genomic sequence, [Dalglish et al. 2010](#)) for variant descriptions.

- c.-128_-126[79] describes the presence of an allele with an extended GGC-repeat of **exactly 79 units**.
□ **NOTE:** the description c.-128GGC[79] can not be used since the repeat is probably interrupted by one or more GGA-triplets
- c.-128_-126[(600_800)] describes the presence of an extended GGC-repeat with an estimated length between 600 and 800 copies
□ **NOTE:** brackets are used to indicate uncertainties ([see Uncertainties](#)), the description c.-128GGC[(600_800)] can not be used since the repeat is probably interrupted by one or more GGA-triplets
- c.-128_-126[(1000)] describes the presence of an extended GGC-repeat of about 1000 units.
□ **NOTE:** brackets are used to indicate uncertainties ([see Uncertainties](#)), the description c.-128GGC[(1000)] can not be used since the repeat is probably interrupted by one or more GGA-triplets
- **NEW** c.1032-?_1357+?(3) denotes the presence of two additional copies of an exon, starting at an unknown position in the flanking upstream intron (upstream of coding DNA nucleotide 1032) and ending at an unknown position in the flanking downstream intron (downstream of coding DNA nucleotide 1357). The description, using "(3)" indicates there is no proof that the three copies are in tandem ([see Uncertainties](#))
- **NEW** c.1032-?_1357+?[3] denotes a **direct triplication** of an exon, starting at an unknown position in the flanking upstream intron (upstream of coding DNA nucleotide 1032) and ending at an unknown position in the flanking downstream intron (downstream of coding DNA nucleotide 1357)
NOTE: the use of tri (for triplication), qua (for quadruplication), etc. is not recommended ([see Discussion](#))
- g.1209_4523[12_45] denotes that a 3.3 kb repeat sequence of which the first copy is present in the genomic reference sequence from nucleotides 1209 to 4523 can be found repeated 12 to 45 times in the population
 - g.1209_4523[14];[23] denotes that a person carries a 3.3 kb repeat with 14 copies on one chromosome and of 23 copies on the

NEW Complex changes

Sequence changes can be very complex, involving several changes at a specific location. The description of such changes using the recommendations given above can become rather complicated and at some point, although literally correct, effectively meaningless. In such cases the recommendation is to submit the sequence that has been determined to GenBank and to use the accession.version number in the description.

- *c.123_678conNM_004006.1:c.123_678 describes a gene conversion replacing nucleotides c.123 to c.4567 of the coding DNA sequence of the transcript of interest with nucleotides c.123 to c.678 from a transcript sequence as present in GenBank file NM_004006 (version 1)*
- *c.88+101_oGJB2:c.355-1045del denotes a deletion which ends in the flanking GJB2 gene at position 355-1045 (in the intron between nucleotides 354 and 355) on the reverse strand (the genes are thus located and fused in opposite transcriptional directions, [see Discussion](#))*
- *c.123+54_123+55insAB012345.2:g.76_420 denotes an intronic insertion (between nucleotides c.123+54 and 123+55) of 345 nucleotides (nucleotides 76 to 420 like in GenBank file AB012345 version 2)*

| [Top of page](#) | [MutNomen homepage](#) | [Check-list](#) | [Symbols, codons, etc.](#) |
| **Recommendations:** [general](#), [DNA](#), [RNA](#), [protein](#), [uncertain](#) |
| [Discussions](#) | [FAQ's](#) | [Symbols, codons, etc.](#) | [History](#) |
| **Example descriptions:** [QuickRef](#), [DNA](#), [RNA](#), [protein](#) |