

Glossary

In preparation

Please note that this Glossary is **work in progress**. Do you encounter missing terms or want to suggest definitions, please let us know.

- **3'rule**
for all descriptions the most 3' position possible of the reference sequence is arbitrarily assigned to have been changed. When ATTTG changes to ATTG HGVS describes this as a change of the T at position 4 (not the T at position 2 or 3)
- **allele**
variant forms of the same gene (MESH (<http://www.ncbi.nlm.nih.gov/mesh?term=allele>))
HGVS: a series of variants on one chromosome.
descriptions *see Recommendations DNA* (/recommendations/DNA/variant/alleles/), *RNA* (/recommendations/RNA/variant/alleles/) or *protein* (/recommendations/protein/variant/alleles/).
- **amino acid**
a letter from the protein code (*see Standards* (/bg-material/standards/)).
- **cap site**
first nucleotide of a transcript (5' end) to which a specially altered nucleotide is added.
- **break point**
the site where two sequences which are in different positions in the reference sequence are joined as a consequence of genomic rearrangement (Structural Variant)
- **cDNA**
cDNA, "copy DNA" or "complementary DNA", is the DNA copy of a single stranded RNA molecule synthesized using the enzyme reverse transcriptase (Wikipedia (https://en.wikipedia.org/wiki/Complementary_DNA), MESH (<https://www.ncbi.nlm.nih.gov/mesh/68018076>)).
NOTE: cDNA is not the same as "coding DNA" (see below).
- **CDS**
coding DNA sequence, a sequence translated in to an amino acid sequence (protein).
- **chimerism**

the occurrence in one individual of two or more cell populations, derived from different zygotes, with different sequences (based on MESH (<http://www.ncbi.nlm.nih.gov/mesh?term=chimerism>)). Opposite of mosaicism.

descriptions *see General/Characters used* (/recommendations/general#characters).

- **cis**

two variants are “**in cis**” when they are on the same allele (DNA molecule, chromosome).

- **CNV**

copy number variant (CNV), a variant in a genome where the number of copies of a large stretch of DNA differs from that in the reference genome; a copy can be missing (deleted) or be present more than once (duplicated, triplicated, ..., or amplified).

NOTE: a “large stretch” is not defined precisely but usually covers at least an exon of a gene or 1,000 nucleotides or more.

alias CNP (copy number polymorphism)

- **coding DNA**

the segments of a genome or segment of a transcript (RNA molecule) which codes for a protein.

- **coding DNA reference sequence**

a DNA reference sequence (*see Reference Sequence* (/bg-material/refseq/#DNAC)), based on a protein-coding transcript of a gene, which can be used for nucleotide numbering using the “c.” prefix. Such a reference sequence includes the coding DNA sequence (CDS) and the 5’ and 3’ UTR regions.

NOTE: a coding DNA reference sequence is **not** a cDNA sequence (see above)

- **complex**

HGVS: 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).

- **compound heterozygote**

used in cases of *autosomal recessive* disease where the disease-causing variants on both alleles at a given locus are **not identical** (opposite of *homozygous*)

- **conversion**

HGVS-DNA: a sequence change where, compared to a reference sequence, a range of nucleotides are replaced by a sequence from elsewhere in the genome.

NOTE: conversion variants are described as a Deletion-Insertion (*see DNA* (/recommendations/DNA/variant/delins/) or *RNA* (/recommendations/RNA/variant/delins/)).

- **Crick strand**

see plus (+) strand.

- **deletion**

- one or more letters of the DNA code are missing (deleted). A deletion is indicated using a “**del**”
- **HGVS-DNA: a sequence change where, compared to a reference sequence, one or more nucleotides are not present (deleted).**

descriptions *see Recommendations DNA* (/recommendations/DNA/variant/deletion/), *RNA* (/recommendations/RNA/variant/deletion/) or *protein*

(/recommendations/protein/variant/deletion/).

- **deletion-insertion (delins)**
 - one or more letters in the DNA code are missing and replaced by several new letters
 - **HGVS-DNA: 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..**
descriptions *see Recommendations DNA* (/recommendations/DNA/variant/indel/), *RNA* (/recommendations/RNA/variant/indel/) or *protein* (/recommendations/protein/variant/indel/).
- **der**
see derivative chromosome
- **derivative chromosome**
a structurally rearranged chromosome carrying the intact centromere of the chromosome indicated (der#), generated by either more than one rearrangement within a single chromosome or a rearrangement involving two or more chromosomes
- **duplication**
 - one or more letters of the DNA code are present twice (doubled, duplicated)
 - **HGVS-DNA: 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.**
NOTE: diagnostic assays (like MLPA) usually detect **an additional copy** of a specific sequence. Whether the additional copy is a duplication or an insertion remains to be determined.
descriptions *see Recommendations DNA* (/recommendations/DNA/variant/duplication/), *RNA* (/recommendations/RNA/variant/duplication/) or *protein* (/recommendations/protein/variant/duplication/).
- **exon**
any nucleotide sequence within a gene which, during maturation of the RNA transcript, is not removed by a process called RNA splicing (Wikipedia (<https://en.wikipedia.org/wiki/Exon>), MESH (<https://www.ncbi.nlm.nih.gov/mesh/68005091>)). Every exon, except the first and last exon, is flanked by two **introns**.
- **extension**
a sequence change extending the reference amino acid sequence at the N- or C-terminal end with one or more amino acids (*protein* (/recommendations/protein/variant/extension/)).
- **frame (reading frame)**
frame 1 is the normal reading frame, using the first nucleotide of each coding triplet of the annotated amino acid reference sequence for translation, starting at the A of the ATG translation initiation codon (nucleotide c.1)
frame 2 is the reading frame using the second nucleotide of the annotated amino acid reference sequence as first nucleotide of a coding triplet for translation in the shifted reading frame
frame 3 is the reading frame using the third nucleotide of the annotated amino acid reference sequence as first nucleotide of a coding triplet for translation in the shifted reading frame.
- **frame shift**

a sequence change between the translation initiation (start) and termination (stop) codon where, compared to a reference sequence, translation shifts to another reading frame (*protein* (/recommendations/protein/variant/frameshift/))

- **fusion transcript**

a transcript (RNA molecule) which consist of parts of transcripts from two or more genes, resulting from a translocation, deletion, or inversion.

- **haplotype**

contiguous set of genetic variants that are co-located on one chromosome (molecule) and are inherited from the same parent

- **hemizygous**

an individual having **only one allele** at a given locus, either because the allele is absent (X and Y chromosome in males) or lost (deleted) (based on MESH (<http://www.ncbi.nlm.nih.gov/mesh?term=hemizygous>)).

- **heterozygous**

an individual in which both alleles at a given locus are **not identical** (based on MESH (<http://www.ncbi.nlm.nih.gov/mesh?term=heterozygous>)).

- **homozygous**

an individual in which both alleles at a given locus are **identical** (MESH (<http://www.ncbi.nlm.nih.gov/mesh?term=homozygous>)).

- **hypermorphic variant**

a variant causing the gene product to possess an increased level of activity, or in which the normal (wild-type) gene product is expressed at an increased level

- **hypomorphic variant**

a variant causing the gene product to possess a reduced level of activity, or in which the normal (wild-type) gene product is expressed at a reduced level

- **indel**

HGVS: confusing term, not used

sometimes: a sequence change where, compared to a reference sequence, one or more nucleotides are replaced by one or more other nucleotides

sometimes: a variant which is a deletion or an insertion.

sometimes: (evolutionary biology) a type of variant in which a specific nucleotide sequence is present (insertion) or absent (deletion).

MESH (<http://www.ncbi.nlm.nih.gov/mesh?term=indel>): a length difference between two alleles where it is unknowable if the difference was originally caused by a sequence insertion or a sequence deletion

- **insertion**

- one or more letters in the DNA, RNA or amino acid code are new (have been inserted)
- **HGVS-DNA: a sequence change where, compared to the reference sequence, one or more residues are inserted and where the insertion is not a copy of a sequence immediately upstream.**

descriptions *see Recommendations DNA* (/recommendations/DNA/variant/insertion/), *RNA* (/recommendations/RNA/variant/insertion/) or *protein* (/recommendations/protein/variant/insertion/).

- **intron**

any nucleotide sequence within a gene which, during maturation of the RNA transcript, is removed by a process called RNA splicing (Wikipedia (<https://en.wikipedia.org/wiki/Intron>), MESH (<https://www.ncbi.nlm.nih.gov/mesh/68007438>)). Every intron is flanked by two **exons**.

- **inversion**

HGVS-DNA: 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.

descriptions *see Recommendations DNA* (/recommendations/DNA/variant/inversion/) or *RNA* (/recommendations/RNA/variant/inversion/).

- **ISCN**

International System for Cytogenetic Nomenclature (*see ISCN*) (/bg-material/consultation/ISCN), covering the description of numerical and structural chromosomal changes detected using microscopic and cytogenetic techniques.

descriptions *see Recommendations DNA - Complex (HGVS<>ISCN)* (/recommendations/DNA/variant/complex/).

- **Kozak sequence**

a consensus sequence, including the ATG translation initiation codon, playing a role in the initiation of translation

- **LOH**

Loss of Heterozygosity (LOH) is a term originally derived from the analysis of tumor samples where, as a consequence of a somatic change, a cell that had originally two different alleles **loses one allele**. The LOH can be caused by different molecular mechanism, including the deletion of the allele, a gene conversion or uniparental disomy.

NOTE: the definition given by MESH (<http://www.ncbi.nlm.nih.gov/mesh/?term=loss+of+heterozygosity>), i.e. the loss of one allele at a specific locus caused by a deletion, is

therefore not correct

NOTE: the term LOH should thus **not** be used to indicate a **homozygous** region, i.e. a region where both chromosomes have the same sequence.

- **loss of heterozygosity**

see LOH.

- **minus (-) strand**

the bottom strand of the reference genome. Alias negative strand, Watson strand.

- **missense**

- a variant in which a codon is changed to one directing the incorporation of a different amino acid (based on MESH (<https://www.ncbi.nlm.nih.gov/mesh/?term=missense>)).

- **HGVS:** a variant in a protein sequence where compared to the reference sequence one amino acid is replaced by another amino acid.

- **mosaicism**

the occurrence in one individual of two or more cell populations, derived from a single zygote, with different sequences (based on MESH (<http://www.ncbi.nlm.nih.gov/mesh?term=mosaicism>)). Opposite of chimerism.

descriptions *see General/Characters used* (/recommendations/general#characters).

- **mutation**

NOTE: please do not use this term, *see Terminology* (/bg-material/basics/).

- **HGVS:** confusing term, do not use, use **variant** (*see Basics* (/bg-material/basics/))
- **biology:** a change in the sequence
- **medicine:** a sequence variant **associated** with a disease phenotype.

- **negative (-) strand**

see minus (-) strand.

- **nonsense**

- a variant that changed an amino acid-specifying codon to a stop codon (termination codon, based on MESH (<https://www.ncbi.nlm.nih.gov/mesh/?term=nonsense>)).
- **HGVS:** a variant in a protein sequence where compared to the reference sequence an amino acid is replaced by a translational stop codon (termination codon).

- **nucleotide**

a letter from the DNA code, e.g. A, C, G, or T (*see Standards* (/bg-material/standards/)).

- **plus (+) strand**

the top strand of the reference genome. Alias positive strand, Crick strand.

- **polyA addition site**

the 3' end of a precursor messenger RNA (pre-mRNA) transcript that is cleaved and to which subsequently a tail of A nucleotides is added (the polyA-tail)

- **polyA signal**

a sequence in the 3' UTR of a transcript signalling the downstream cleavage and addition of a polyA tail

- **polymorphism**

NOTE: please do not use this term, *see Terminology* (bg-material/basics/).

- **HGVS:** confusing term, do not use, use **variant** (*see Basics* (/bg-material/basics/))
- **biology:** a sequence variant present in the population at a frequency of 1% or higher
- **medicine:** a sequence variant **not associated** with a disease phenotype

- **positive (+) strand**

see plus (+) strand.

- **quadruplication**

a sequence change where, compared to a reference sequence, three copies of a sequence are inserted directly 3' of the original copy of that sequence (*see Recommendations DNA* (/recommendations/DNA/variant/duplication/)).

- **quintuplication**

a sequence change where, compared to a reference sequence, four copies of a sequence are inserted directly 3' of the original copy of that sequence (*see Recommendations DNA* (/recommendations/DNA/variant/duplication/)).

- **reading frame**
: one of three possible ways to translate a nucleotide sequence in to an amino acid sequence (a protein)
see also **frame**
- **readthrough transcript**
a transcript (RNA molecule) formed via the splicing of exons from more than one distinct gene. The two (or more) genes involved are found on the same chromosomal region, on the same strand, and are typically adjacent to one another.
- **repeated sequence**
HGVS: 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.
- **silent**
 - a variant in a DNA sequence that does not change the amino acid sequence of the encoded protein (based on MESH (<https://www.ncbi.nlm.nih.gov/mesh/?term=silent>)).
 - **HGVS**: an amino acid residue in a protein sequence where compared to the reference sequence the DNA sequence changed but not the encoded amino acid.
- **SNP**
Single Nucleotide Polymorphism (SNP). The preferred term is SNV (Single Nucleotide Variant), *see polymorphism*.
- **SNV**
Single Nucleotide Variant (SNV), a variant involving one nucleotide (e.g. A>C, A>T, A>G, delA, dupA, insA).
- **splice acceptor site (SA)**
the 3' splice site, at the end of the intron/start of the exon
- **splice donor site (SD)**
the 5' splice site, at the end of the exon/start of the intron
- **splice site**
the site in a precursor messenger RNA (pre-mRNA) transcript that is cleaved to remove the intron.
- **splicing**
the process removing specific segments (the introns) of a precursor messenger RNA (pre-mRNA) transcript. When an intron is removed the flanking RNA segments (the exons) are joined together (ligated)
- **strand**
one of the two strands of a DNA molecule (double stranded).
- **Structural Variant (SV)**
a variant in a genome where compared to the reference sequence the structure of a large stretch of DNA is changed. SVs include deletions/duplications (CNVs), inversions, insertions, deletion-insertions, conversions, transpositions, translocations, etc.
NOTE: a “large stretch” is not defined precisely but usually covers at least an exon of a gene or 1,000 nucleotides or more.

- **substitution**
 - one letter of the DNA, RNA or amino acid code is replaced (substituted) by one other letter
 - **HGVS-DNA: a sequence change where, compared to a reference sequence, one residue is replaced by one other residue.**
descriptions *see Recommendations DNA* (/recommendations/DNA/variant/substitution/), *RNA* (/recommendations/RNA/variant/substitution/) or *protein* (/recommendations/protein/variant/substitution/).
- **SV**
see Structural Variant.
- **trans**
two variants are “**in trans**” when they are on different alleles (DNA molecules, chromosomes).
- **transition**
a nucleotide variant changing a purine nucleotide to another purine nucleotide (A < > G), or a pyrimidine nucleotide to another pyrimidine nucleotide (C < > T).
- **translocation**
 - a chromosome abnormality characterized by chromosome breakage and transfer of the broken-off portion to a non-homologous chromosome (based on MESH (<https://www.ncbi.nlm.nih.gov/mesh/?term=translocation>))
 - **HGVS: a sequence change where, compared to a reference sequence, from a specific nucleotide position (the break point) all nucleotides upstream derive from another chromosome then those down stream**
NOTE: a translocation occurs when two chromosomes break and the fragments rejoin with the non-homologous chromosome. A full description of a (reciprocal) translocation consists of 2 parts, one describing the first junction, the second describing the other junction (e.g. the chromosome 4;X junction and the chromosome X;4 junction)
 - **translocation, balanced**
a translocation with an even exchange of DNA sequences and no segments deleted or duplicated
 - **translocation, unbalanced**
a translocation with an uneven exchange of DNA sequences and segments being deleted or duplicated
- **transposition**
a sequence change where, compared to a reference sequence, a large stretch of DNA moves from one position in the genome to another position, i.e. a deletion at one position combined with the insertion of the deleted sequence at another position. The variant is described as a deletion at the original location and an insertion at the new location.
- **transversion**
a nucleotide variant changing a purine nucleotide to a pyrimidine nucleotide (A or G > T or C), or a pyrimidine nucleotide to a purine nucleotide (C or T > A or G)
- **triplication**

a sequence change where, compared to a reference sequence, two copies of a sequence are inserted directly 3' of the original copy of that sequence (*see Recommendations DNA* (/recommendations/DNA/variant/duplication/)).

- **trisomy**

the presence of a third chromosome of any one type in an otherwise diploid cell (MESH (<http://www.ncbi.nlm.nih.gov/mesh/?term=trisomy>)).

- **UTR**

UnTranslated Region (UTR), the segments of of a protein coding RNA molecule that is not translated.

5'UTR = UTR 5' of the translation initiation codon (ATG start codon)

3'UTR = UTR 3' of the translation termination codon

- **variant**

a difference between a reference sequence and a sample sequence

- **VNTR**

Variable Number of Tandem Repeats, a nucleotide sequence consisting of units of a specific short sequence which is repeated in tandem copies and where the number of units is variable in the population.

- **Watson strand**

see minus (-) strand.