



SNP Class Definitions

Created: July 7, 2005; Updated: February 18, 2014.

Functional Class

I do not understand the meaning of the "Function" category or any of its components (e.g. synonymous, contig reference, nonsynonymous)?

The NCBI handbook has [documentation](#) that should address your questions. Below is an excerpt from that document that should address your question:

"...[fxn-class] defines variation functional classes. We base class on the relationship between a variation and any local gene features. When a variation is near a transcript or in a transcript interval but not in the coding region, then we define the functional class by the position of the variation relative to the structure of the aligned transcript. In other words, a variation may be near a gene (locus region), in a UTR (mrna-utr), in an intron (intron), or in a splice site (splice site). If the variation is in a coding region, then the functional class of the variation depends on how each allele may affect the translated peptide sequence.

Typically, one allele of a variation will be the same as the contig (contig reference), and the other allele will be either a synonymous change or a nonsynonymous change. In some cases, one allele will be a synonymous change, and the other allele will be a nonsynonymous change. If any allele is a nonsynonymous change, then the variation is classified as a nonsynonymous variation. Otherwise, the variation is classified as a synonymous variation.

The Four Basic Outcomes When a Variation Is in Coding Sequence

- The allele is the same as the contig (contig reference) and hence causes no change to the translated sequence.
- The allele, when substituted for the reference sequence, yields a new codon that encodes the same amino acid. This is termed a synonymous substitution.
- The allele, when substituted for the reference sequence, yields a new codon that encodes a different amino acid. This is termed a nonsynonymous substitution.
- A problem with the annotated coding region feature prohibits conceptual translation. In this case, we note the variation class as coding, based solely on position.

Because functional classification is defined by positional and sequence parameters, two facts emerge: (a) if a gene has multiple transcripts because of alternative splicing, then a variation may have several different functional relationships to the gene; and (b) if multiple genes are densely packed in a contig region, then a variation at a single location in the genome may have multiple, potentially different, relationships to its local gene neighbors." (3/14/05)

What does it mean when an SNP is labeled "synonymous" or "non-synonymous"?

The terms "synonymous" and "non-synonymous" are used for SNPs that are in predicted protein coding regions (i.e., exons of genes). Synonymous SNPs are those SNPs that have different alleles that encode for the

same amino acid. Non-synonymous SNPs are SNPs that have different alleles that encode different amino acids. For further details, I recommend querying for “genetic code” or “protein translation” at the [NCBI books website](#).

I noticed in the documentation where "coding nonsynonymous" is mentioned as a SNP function class, yet, I don't think I've ever seen it used. Instead I see the term "missense".

In ftp reports for older builds, we did use the term “nonsynonymous” to as a function class. Then about a year ago, the functional class code became more sophisticated so that it could describe the type of nonsynonymous change the SNP caused: “missense”, “nonsense” or “frameshift”.

So “missense”, “nonsense”, and “frameshift” are types of nonsynonymous change a variation can cause. (07/15/08)

Where are descriptions of your functional classifications located?

They are located in the [asn docsum](#). The portion of the docsum which answers your question is below:

```
NSE-FxnSet ::= SEQUENCE {

locusid INTEGER,                locus-id of gene as aligned to contig
                                SNPContigLocusId.locus_id
symbol VisibleString OPTIONAL,  symbol (official if present in LocusLink) of gene
                                SNPContigLocusId.locus_symbol
mrna-acc VisibleString OPTIONAL, mRNA accession if variation in transcript
                                SNPContigLocusId.mrna_acc
prot-acc VisibleString OPTIONAL, protein accession if variation in coding region
                                interval SNPContigLocusId.protein_acc

fxn-class-contig ENUMERATED {
locus-region (1),              variation in region of gene, but not in transcript
                                SNPContigLocusId.fxn_class
coding (2),                    variation in coding region of gene, assigned if the
                                allele-specific class is unknown
coding-synon (3)               no change in peptide for allele with respect to
                                contig sequence **allele specific class**
coding-nonsynon (4),          change in peptide with respect to contig sequence
                                **allele-specific class**
mrna-utr (5),                 variation in transcript, but not in coding region interval
intron (6),                   variation in intron, but not in first 2 or last 2 bases
                                of intron
splice-site (7),              variation in first 2 or last 2 bases of intron
reference (8),                 allele observed in reference contig sequence
                                **allele-specific class**
exception (9)                 variation in coding region with exception raised on
                                alignment. This occurs when protein with gap in sequence is
                                aligned back to contig sequence. variations 3' of the gap
                                have undefined functional inference.
```

Please clarify the definition for functional class code 8 (“reference”).

The word “reference” is used in two different contexts in dbSNP:

The first context for “reference” refers to the NCBI reference genome assembly, as opposed to alternative assemblies such as Celera.

The second context for “reference” is the “cds-reference” which is used as a functional class code (explained below).

I will use rs268 as an example of "cds-reference". rs268 has an A/G variation, which maps to reference contig NT_030737 at position 7658457. The contig at this position has the allele "A". This SNP can have either an "A" or a "G" allele, and when the allele happens to be "G", it will cause a non-synonymous coding change. So in the SNP functional class context, "cds-reference" just means the "allele on the contig" since synonymous and non-synonymous alleles are relative to the contig. Please note: rs268 also maps to Celera contig NW_923907, which also has the "A" allele at this position. (12/28/07)

What is the definition for the “Locus region” functional class?

Please see [Table 5](#) (refSNP Function Code Table) in the dbSNP section of the online NCBI handbook. (5/18/05)

Variation Class

The alleles for rs1611430 are reported A/G/T, yet the variation class is SNP. I thought this classification was reserved for biallelic SNPs, and that tri- and multi-alleles were classified as MNPs.

MNPs in dbSNP are variations that have alleles with multiple nucleotides like AT/GT.(6/8/07)

“snpclass1”, is a term I found in the SNP report header. What is its definition?

The snpclass values indicate the type of SNP using a classification scheme that is also [described in the spec file](#).

Type of variation	SubSNP.subsnp_class
Snp (1)	True single nucleotide polymorphism
in-del (2)	Insertion deletion polymorphism; deletions represented by '-' in allele string
het (3)	Variation has unknown sequence composition but is observed to be heterozygous
microsat (4)	Microsatellite/simple sequence repeat
named (5)	Allele sequences defined by name tag instead of raw sequence, e.g., (Alu)/-
no-variation (6)	Submission reports invariant region in surveyed sequence
mixed (7)	Mixed class
mnp (8)	Multiple nucleotide polymorphism (all alleles same length, where length >1)