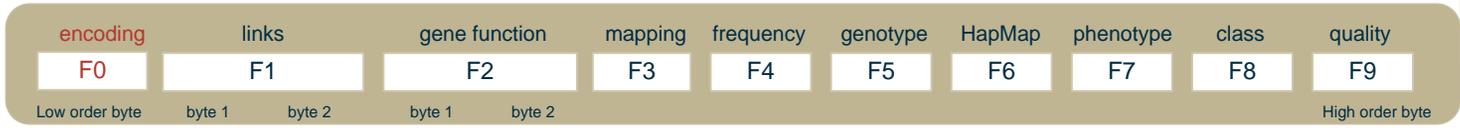




NCBI dbSNP bitfield encoding schema

Revisions:
 Ver 5.3, Jul 16, 2009; Ver 5.2 - May 12, 2009; ver 5.1, Rev. Nov.5, 2008; ver 5, Rev. May 21 2008
 ver 4, Rev. May 5 2008

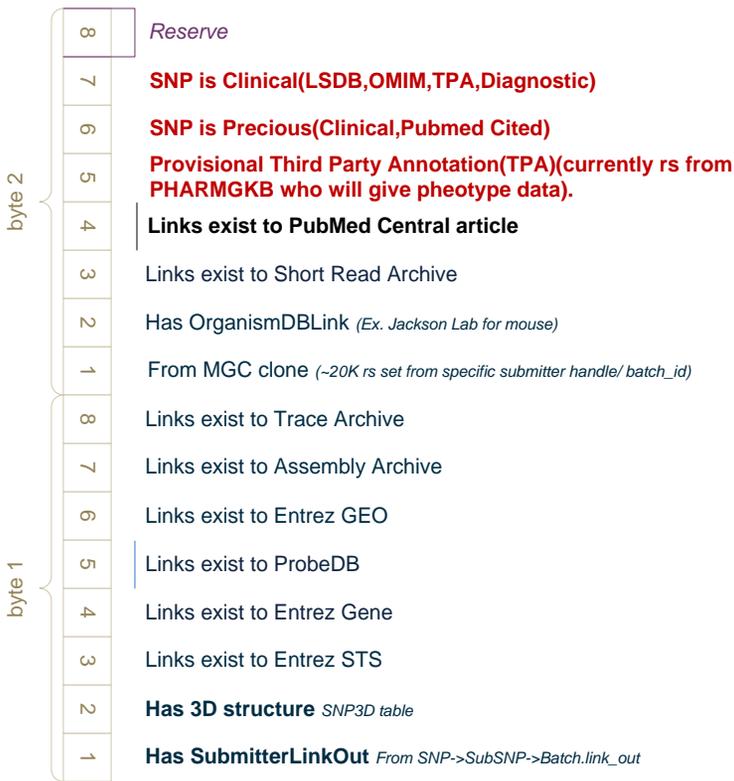
12 byte structure



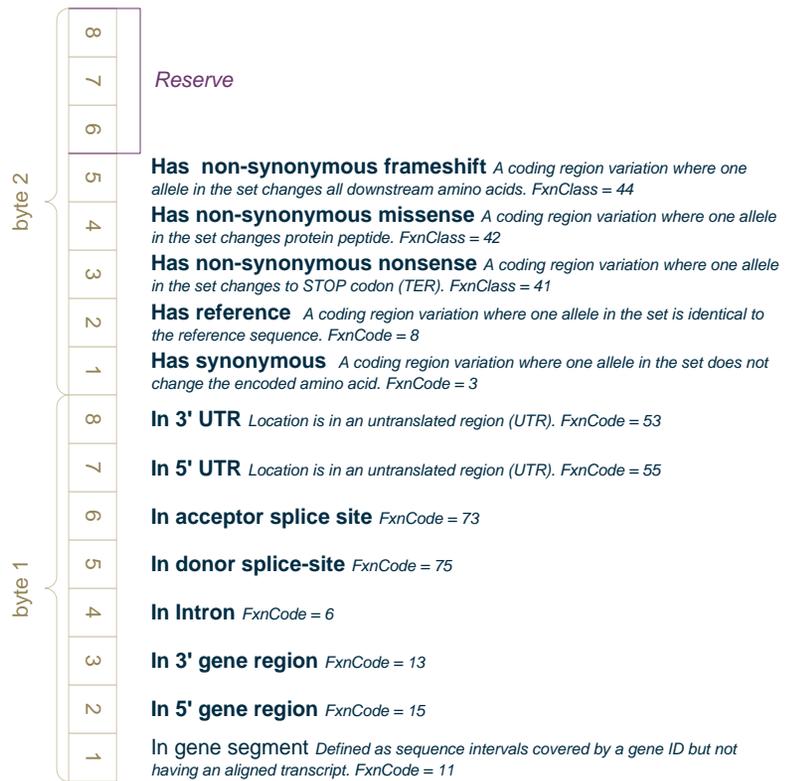
The dbSNP bitfield structure is a 12-byte object that defines variation properties in 9 areas.

- F1 and F2 are 2-byte structures
- F0 (lowest order byte) is the version of the encoding schema used for the data, see page 2 for definition.
- Bits labeled in **bold** are currently populated by dbSNP. Red color indicates newly implemented.

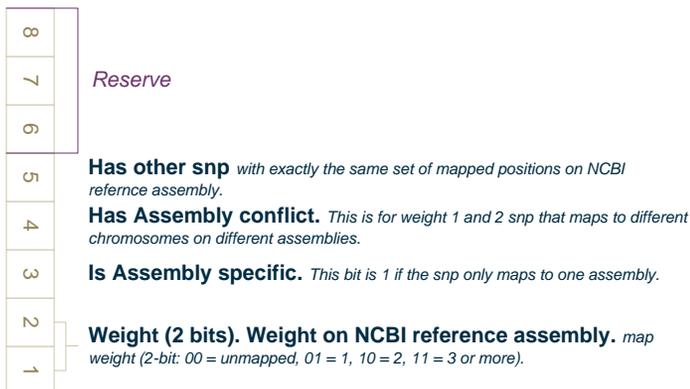
F1 – resource link properties



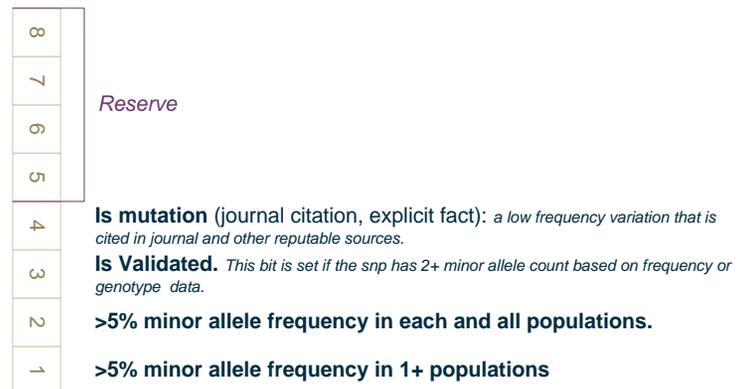
F2 – gene function properties



F3 – mapping properties



F4 – allele frequency properties





NCBI dbSNP bitfield encoding schema

Ver 5.3, Jul 16, 2009

F5 – genotype properties

| |
|---|
| 8 |
| 7 |
| 6 |
| 5 |
| 4 |
| 3 |
| 2 |
| 1 |

Reserve

Marker is on high density genotyping kit (50K density or greater).
The snp may have phenotype associations present in dbGaP.

In Haplotype tagging set

Genotypes available. *The snp has individual genotype (in SubInd table).*

F6 – HapMap properties

| |
|---|
| 8 |
| 7 |
| 6 |
| 5 |
| 4 |
| 3 |
| 2 |
| 1 |

Reserve

Phase 3 genotyped: filtered, non-redundant.

Phase 2 genotyped: filtered, non-redundant.

Phase 1 genotyped: filtered, non-redundant.

F7 – phenotype properties

| |
|---|
| 8 |
| 7 |
| 6 |
| 5 |
| 4 |
| 3 |
| 2 |
| 1 |

Has MeSH is linked to a disease.

Variation is interrogated in a clinical diagnostic assay

Has transcription factor

Submitted from a locus-specific database.

Has p-value $\leq 1 \times 10^{-3}$ in a dbGaP study association test

Has LOD score ≥ 2.0 in a dbGaP study genome scan

Microattribution/third-party annotation(TPA:GWAS,PAGE)

Has OMIM/OMIA

F8 – variation class

| |
|---|
| 8 |
| 7 |
| 6 |
| 5 |
| 4 |
| 3 |
| 2 |
| 1 |

Reserve

- 0001 = single base polymorphism SNP
- 0010 = dips (deletion/insertion)
- 0011 = HETEROZYGOUS
- 0100 = Microsatellite
- 0101 = Named variation, e.g. (Alu)
- 0110 = NOVARIATION
- 0111 = mixed class
- 1000 = multi-base polymorphism

F9 – quality check

| |
|---|
| 8 |
| 7 |
| 6 |
| 5 |
| 4 |
| 3 |
| 2 |
| 1 |

Reserve

Variation is somatic, not germline. *The variation was detected in a Somatic tissue (e.g. cancer tumor). The variation is not known to exist in heritable DNA.*

Contig allele not present in SNP allele list. *The reference sequence allele at the mapped position is not present in the SNP allele list, adjusted for orientation.*

Is Withdrawn by submitter *If one member ss is withdrawn by submitter, then this bit is set. If all member ss' are withdrawn, then the rs is deleted to SNPHistory.*

Rs cluster has non-overlapping allele sets. *True when rs set has more than 2 alleles from different submissions and these sets share no alleles in common.*

Is a strain-specific fixed difference

Has Genotype Conflict *Same (rs, ind), different genotype. N/N is not included.*

F0 – Version encoding

| |
|---|
| 8 |
| 7 |
| 6 |
| 5 |
| 4 |
| 3 |
| 2 |
| 1 |

Reserve

Bitmap schema version. *Versions increment as integer value (current is version 2, version 1.2 is encoded as version 1)*