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

### F1 – resource link properties

- **Reserve**
- **Links exist to Short Read Archive**
- **Has OrganismDBLink** (Ex. Jackson Lab for mouse)
- **From MGC clone** (~20K ns set from specific submitter handle/batch_id)
- **Links exist to Trace Archive**
- **Links exist to Assembly Archive**
- **Links exist to Entrez GEO**
- **Links exist to ProbeDB**
- **Links exist to Entrez Gene**
- **Links exist to Entrez STS**
- **Has 3D structure** SNP3D table
- **Has SubmitterLinkOut** From SNP::SubSNP::Batch::link_out

### F2 – gene function properties

- **Reserve**
- **Has non-synonymous frameshift** A coding region variation where one allele in the set changes all downstream amino acids. FxnClass = 44
- **Has non-synonymous missense** A coding region variation where one allele in the set changes protein peptide. FxnClass = 42
- **Has non-synonymous nonsense** A coding region variation where one allele in the set changes to STOP codon (TER). FxnClass = 41
- **Has reference** A coding region variation where one allele in the set is identical to the reference sequence. FxnCode = 8
- **Has synonymous** A coding region variation where one allele in the set does not change the encoded amino acid. FxnCode = 3
- **In 3’ UTR** Location is in an untranslated region (UTR). FxnCode = 53
- **In 5’ UTR** Location is in an untranslated region (UTR). FxnCode = 55
- **In acceptor splice site** FxnCode = 73
- **In donor splice-site** FxnCode = 75
- **In Intron** FxnCode = 6
- **In 3’ gene region** FxnCode = 13
- **In 5’ gene region** FxnCode = 15
- **In gene segment** Defined as sequence intervals covered by a gene ID but not having an aligned transcript. FxnCode = 11

### F3 – mapping properties

- **Reserve**
- **Has other snp** with exactly the same set of mapped positions on NCBI reference assembly.
- **Has Assembly conflict**. This is for weight 1 and 2 snp that maps to different chromosomes on different assemblies.
- **Is Assembly specific**. This bit is 1 if the snp only maps to one assembly.
- **Weight (2 bits)**. Weight on NCBI reference assembly. map weight (2-bit: 00 = unmapped, 01 = 1, 10 = 2, 11 = 3 or more).

### F4 – allele frequency properties

- **Reserve**
- **Is mutation** (journal citation, explicit fact): a low frequency variation that is cited in journal and other reputable sources.
- **Is Validated**. This bit is set if the snp has 2+ minor allele count based on frequency or genotype data.
- >5% minor allele frequency in each and all populations.
- >5% minor allele frequency in 1+ populations
F5 – genotype properties

- Markers are 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

- Phase 3 genotyped: filtered, non-redundant.
- Phase 2 genotyped: filtered, non-redundant.
- Phase 1 genotyped: filtered, non-redundant.

F7 – phenotype properties

- 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 significant association in dbGaP study.
- Has LOD score.
- Has SnpRIF.
- Has OMIM/OMIA.

F8 – variation class

- 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

- 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 is withdrawn by submitter, then this bit is set for all member sets. If all member sets are withdrawn, then the SNP is deleted to SNPHistory.
- Rs cluster has non-overlapping allele sets. True when rs set has more than two 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

- Bitmap schema version. Versions increment as integer value (NOTE: version 1.2 is encoded as version 1).