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

- **F0** 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 rs 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 doubleHit. This bit is set if the rs# is in Jim Mullikin’s double hit submission which has been only on human snp.
- >5% minor allele frequency in each and all populations.
- >5% minor allele frequency in 1+ populations.
NCBI dbSNP bitfield encoding schema version 3, Rev. October 2007

### F5 – genotype properties

- **8** Reserve
- **7** Marker is on high density genotyping kit (50K density or greater).
- **6** The snp may have phenotype associations present in dbGaP.
- **5** In Haplotype tagging set
- **4** Genotypes available. The snp has individual genotype (in SubInd table).

### F6 – HapMap properties

- **8** Reserve
- **7** Phase 3 genotyped: filtered, non-redundant.
- **6** Phase 2 genotyped: filtered, non-redundant.
- **5** Phase 1 genotyped: filtered, non-redundant.

### F7 – phenotype properties

- **8** Has MeSH is linked to a disease.
- **7** Variation is interrogated in a clinical diagnostic assay
- **6** Has transcription factor
- **5** Submitted from a locus-specific database.
- **4** Has significant association in dbGaP study
- **3** Has LOD score
- **2** Has SnpRIF
- **1** Has OMIM/OMIA

### F8 – variation class

- **8** Reserve
- **7** 0001 = single base polymorphism SNP
- **6** 0010 = dips (deletion/insertion)
- **5** 0011 = HETEROZYGOUS
- **4** 0100 = Microsatellite
- **3** 0101 = Named variation, e.g. (Alu)
- **2** 0110 = NOVARIATION
- **1** 0111 = mixed class
- **0** 1000 = multi-base polymorphism

### F9 – quality check

- **8** Reserve
- **7** 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.
- **6** 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.
- **5** 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.
- **4** Is a strain-specific fixed difference
- **3** Has Genotype Conflict. Same (rs, ind), different genotype. N/N is not included.

### F0 – Version encoding

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