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**

- **Reserve**
- **SNP is Clinical** (LSDB, OMIM, TPA, Diagnostic)
- **SNP is Precious** (Clinical, Pubmed Cited)
- **Provisional Third Party Annotation** (TPA) (currently rs from PHARMGKB who will give phenotype data).
- Links exist to PubMed Central article
- 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 STOP-Loss** A coding region variation where one allele in the set changes the encoded STOP CODON (TER). FxnClass = 43
- **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 STOP-Gain** 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 untranslatd 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**
- **GMAF>=0.01** (Global population Minor Allele Frequency. Population includes all samples in TGP(1000Genome project).
- **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

<table>
<thead>
<tr>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reserve</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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 – Validation by HapMap/TGP properties

<table>
<thead>
<tr>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reserve</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **TGP Phase 3:** val=32. Old: TGP_validated (for subset that passed positive second platform validation); old definition never implemented.
- **TGP Phase 1 (includes June Interim phase 1):** val=16
- **TGP pilot (1,2,3):** val=8
- **RS Cluster has TGP Submission as of June 2011 (include all current RS from TGP): VCF – KGPROD:** val=4
- **RS Cluster has none TGP Submission (set VCF OTHERKG):** val=2
- **HapMap Phase 3 genotyped: filtered, non-redundant. (VCF: PH3):** val=1

### F7 – phenotype properties

<table>
<thead>
<tr>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reserve</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **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 \( \geq 2.0 \) in a dbGaP study genome scan**
- **Microattribution/third-party annotation (TPA: GWAS, PAGE)**
- **Has OMIM/OMIA**

### F8 – variation class

<table>
<thead>
<tr>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reserve</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- 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

<table>
<thead>
<tr>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reserve</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Is suspect.** The variants are paralogous sequence differences. (added 01/19/11 ver 5.4) val=64
- **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

<table>
<thead>
<tr>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bitmap schema version. Versions increment as integer value (current is version 2, version 1.2 is encoded as version 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>