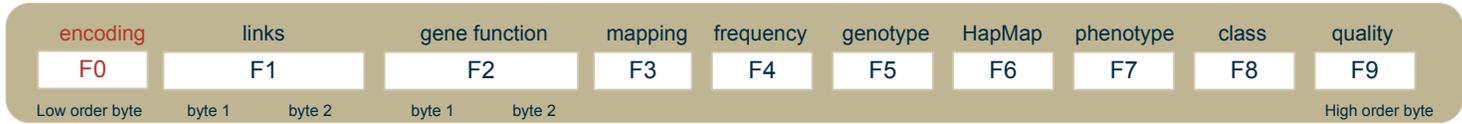




NCBI dbSNP bitfield encoding schema

Revisions:
 Ver 5.6 Aug, 2012; Ver.5.5 Jan, 2012; Ver 5.4 Jun 2, 2010; 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.

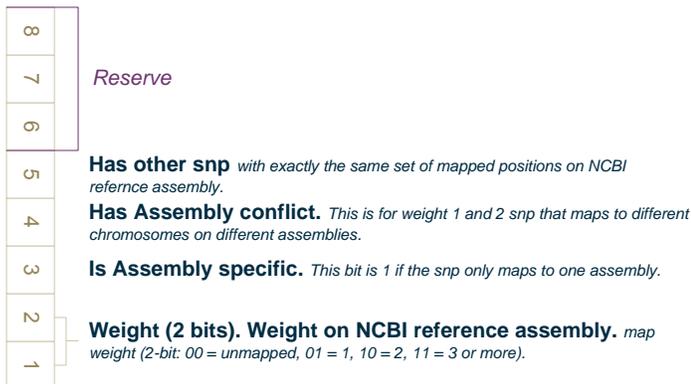
F1 – resource link properties



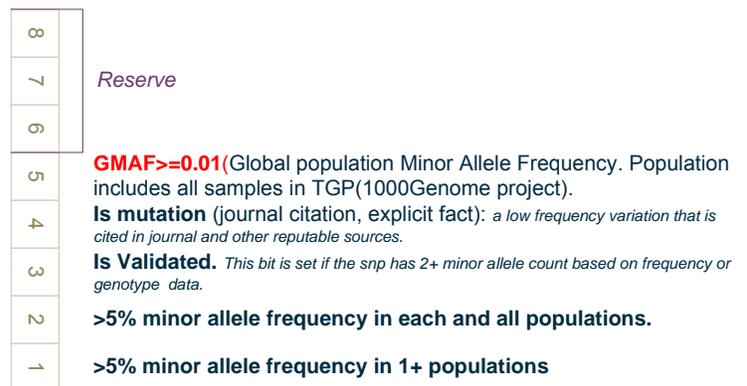
F2 – gene function properties



F3 – mapping properties



F4 – allele frequency properties





NCBI dbSNP bitfield encoding schema

Ver 5.3, Jul 16, 2009; Ver 5.4, Jan 19,2011; ver 5.5, Dec22,2011

F5 – genotype properties

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

F6 – Validation by HapMap/TGP properties

8	Reserve
7	
6	TGP Phase 3;val=32. Old: TGP_validated (for subset that passed positive second platform validation); old definition never implemented.
5	TGP Phase 1 (includes June Interim phase 1);val=16
4	TGP pilot (1,2,3);val=8
3	RS Cluster has TGP Submission as of June 2011(include all current RS from TGP): VCF – KGPROD; val=4
2	RS Cluster has none TGP Submission(set VCF OTHERKG); val=2
1	HapMap Phase 3 genotyped: filtered, non-redundant. (VCF: PH3); val=1

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 p-value $\leq 1 \times 10^{-3}$ in a dbGaP study association test
3	Has LOD score ≥ 2.0 in a dbGaP study genome scan
2	Microattribution/third-party annotation(TPA:GWAS,PAGE)
1	Has OMIM/OMIA

F8 – variation class

8	Reserve
7	
6	
5	
4	
3	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
2	
1	

F9 – quality check

8	Reserve
7	Is suspect. The variants are <i>paralogous sequence differences</i> . (added 01/19/11 ver 5.4) val=64
6	Variation is somatic, not germline. <i>The variation was detected in a Somatic tissue (e.g. cancer tumor). The variation is not known to exist in heritable DNA.</i>
5	Contig allele not present in SNP allele list. <i>The reference sequence allele at the mapped position is not present in the SNP allele list, adjusted for orientation.</i>
4	Is Withdrawn by submitter <i>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.</i>
3	Rs cluster has non-overlapping allele sets. <i>True when rs set has more than 2 alleles from different submissions and these sets share no alleles in common.</i>
2	Is a strain-specific fixed difference
1	Has Genotype Conflict <i>Same (rs, ind), different genotype. N/N is not included.</i>

F0 – Version encoding

8	Reserve
7	
6	
5	
4	
3	Bitmap schema version. Versions increment as integer value (current is version 2, version 1.2 is encoded as version 1)
2	
1	

