VCF/BCF format

```
##fileformat=VCFv4.2
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo
sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality”>
```

INFO

FILTER

FORMAT

```
#CHROM POS     ID        REF ALT    QUAL FILTER INFO                              FORMAT      NA00001 ...
20    14370   rs6054257 G    A       29   PASS   NS=3;DP=14;AF=0.5;               GT:GQ:DP:HQ 0|0:48:1:51,51
20    17330   .         T    A        3   q10    NS=3;DP=11;AF=0.017              GT:GQ:DP:HQ 0|0:49:3:58,50
20    1110696 rs6040355 A    G,T     67   PASS   NS=2;DP=10;AF=0.333,0.667;       GT:GQ:DP:HQ 1|2:21:6:23,27
20    1230237 .         T    .       47   PASS   NS=3;DP=13;AA=T                  GT:GQ:DP:HQ 0|0:54:7:56,6
20    1234567 microsat1 GTC  G,GTCT  50   PASS   NS=3;DP=9;AA=G                   GT:GQ:DP    0/1:35:4
```

VCF format

ID in dbSNP

Defined in ##INFO lines

```
#CHR  POS      ID        REF  ALT     QUAL  FILTER  INFO                              FORMAT      NA00001        ...
20    14370    rs6054257 G    A       29    PASS    NS=3;DP=14;AF=0.5;                GT:GQ:DP:HQ 0|0:48:1:51,51 ...
20    17330    .         T    A       3     q10     NS=3;DP=11;AF=0.017               GT:GQ:DP:HQ 0|0:49:3:58,50 ...
20    1110696  rs6040355 A    G,T     67    PASS    NS=2;DP=10;AF=0.333,0.667;        GT:GQ:DP:HQ 1|2:21:6:23,27 ...
20    1230237  .         T    .       47    PASS    NS=3;DP=13;AA=T                   GT:GQ:DP:HQ 0|0:54:7:56,6  ...
20    1234567  microsat1 GTC  G,GTCT  50    PASS    NS=3;DP=9;AA=G                    GT:GQ:DP    0/1:35:4       ...
```

DEL,INS

Defined in ##FILTER lines

Defined in ##FORMAT lines

Genotype; follows FORMAT specifications

https://samtools.github.io/hts-specs/VCFv4.2.pdf

VCF format: INFO

```
##INFO=<ID=NS,Number=1,Type=Integer,Description=”Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
```

Other standard fields:

1000G, AA (ancestral allele), CIGAR, MQ0 (# reads with MAPQ==0), etc.

Examples

```
NS=3;DP=14;AF=0.5;
NS=3;DP=11;AF=0.017;
NS=2;DP=10;AF=0.333,0.667;
NS=3;DP=13;AA=T
NS=3;DP=9;AA=G
```

VCF format: FORMAT

```
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype
Quality">
```

Other standard fields:

FT (filter pass), GL (genotype likelihood), MQ (mapping quality), etc.

Example

FORMAT	NA00001 ...
GT:GQ:DP:HQ	0|0:48:1:51,51
GT:GQ:DP:HQ	0|0:49:3:58,50
GT:GQ:DP:HQ	1|2:21:6:23,27
GT:GQ:DP:HQ	0|0:54:7:56,6
GT:GQ:DP	0/1:35:4

VCF format: Entries

1. SNP

```
Ref:  g c a G g t
Var:  g c a A g t
```

```
#CHROM  POS  ID  REF  ALT  QUAL  FILTER  INFO
14      4    .   G    A    .     PASS    DP=100
```

2. DEL

```
Ref:  g c a G g t
Var:  g c a - g t
```

```
14      4    .   AG   A    .     PASS    DP=100
```

3. Mixed

```
Ref:  g c a G g t
Var1: g c a - g t
Var2: g c a A g t
Var3: g c a Gtg t
```

```
14      4    .   AG   AG,A,AA,AGT A   .   PASS    DP=100
```

Decoding VCF entries