1st July 2008

HGVbaseG2P Genotype and Allele representations (where n represents the flanking sequence)

General rules for Alleles

1) Alleles are always presented in the context of 10 base upstream and 10 bases downstream, so that the DNA strand is clear

2) Alleles MUST ALWAYS be delimited with ()
e.g., nnnnnnnnnn(T)nnnnnnnnnn for the 'T' allele of a SNP

3) Allele sequences can be compressed down with the appropriate nomenclature e.g ATTTTTTAAAAAAAAA becomes (A(T)6(A)9)

4) Text-based rather than sequence-based alleles must exactly match the representation given in dbSNP (e.g LARGEDELETION) but with the text string additionally enclosed in Double Quotes "" and ()
eg. nnnnnnnnnn("LARGEDELETION")nnnnnnnnnn

General rules for Genotypes

1) Genotypes are always presented in the context of 10 base upstream and 10 bases downstream, so that the DNA strand is clear

2) Genotypes MUST ALWAYS be delimited with []
e.g., nnnnnnnnnn[(T)]nnnnnnnnnn for a 'T' allele homozygote

3) For genotypes of type 'QUALITATIVE' (i.e., where the utilized assay technology allows alleles to be merely seen or not seen), the detected alleles are listed with a '+' separator
e.g., nnnnnnnnnn[(G)+(T)]nnnnnnnnnn for a 'T' & 'G' allele heterozygote
Note: virtually all genotype data today are of type 'QUALITATIVE'

4) For genotypes of type 'RATIO' (i.e., where the utilized assay technology allows alleles to be quantified with respect to each other when more than one allele is detected), the detected alleles are listed with a '+' separator and a ratio number in front of the opening curved bracket for each allele
e.g., nnnnnnnnnn[2(P)+1(Q)]nnnnnnnnnn for a genotype where alleles P and Q are measured to be in the ratio of 2:1

5) For genotypes of type 'QUANTITATIVE' (i.e., where the utilized assay technology allows alleles to be quantified in absolute terms per cell), the detected alleles are listed with a '+' separator and a quantification number plus 'x' immediately after the opening curved bracket for each allele
e.g., nnnnnnnnnn[(1xP)+(0.5xQ)]nnnnnnnnnn for a genotype where alleles P is single copy per cell and allele Q is deleted from 50% of cells

See the Nomenclature tables at the end of this document for further examples of allele and genotype conventions

Single Nucleotide Polymorphism (SNP)

Definition: single base substitutions involving A, T, C, or G

Example allele representations:
dbSNP: G/T
HGVbaseG2P: nnnnnnnnnn(G)nnnnnnnnnn
nnnnnnnnnn(T)nnnnnnnnnn

Example qualitative genotype representations:
homozygote: nnnnnnnnnn[(G)]nnnnnnnnnn
homozygote: nnnnnnnnnn[(T)]nnnnnnnnnn
heterozygote: nnnnnnnnnn[(G)+(T)]nnnnnnnnnn

Multi-Nucleotide Polymorphism (MNP)

Definition: variations that are multi-base variations, with all alleles being the same length

Example allele representations:
dbSNP: ACG/TTC
HGVbaseG2P: nnnnnnnnnn(ACG)nnnnnnnnnn
nnnnnnnnnn(TTC)nnnnnnnnnn

Example qualitative genotype representations:
homozygote: nnnnnnnnnn[(ACG)]nnnnnnnnnn
homozygote: nnnnnnnnnn[(TTC)]nnnnnnnnnn
heterozygote: nnnnnnnnnn[(ACG)+(TTC)]nnnnnnnnnn

No-variation (none)

Definition: Segments of sequence that are assayed and determined to be invariant in a set of samples

dbSNP: NOVARIATION
HGVbaseG2P: nnnnnnnnnn(“NOVARIATION”)nnnnnnnnnn

Insertion/Deletion Polymorphism (in-del)

Definition: An insertion of one or more nucleotides in one version of a sequence relative to another. Since the molecular event that gave rise to this observation cannot be determined from the alleles alone (i.e. was it an insertion or a deletion), both events are incorporated into the name of this polymorphism type. In dbSNP, in-dels are designated using the full sequence of the insertion as one allele, and a "-" character to specify the deleted allele.

Note: In HGVbaseG2P the deleted allele is specified as "_" to avoid confusion that "-" is interpreted as a numeric operation. Also, "_" more appropriately reflects the alphabetical nature of the sequences.

Example allele representations:

dbSNP: -/T

HGVbaseG2P: nnnnnnnnnn(_)nnnnnnnnnn
nnnnnnnnnn(T)nnnnnnnnnn

Example qualitative genotype representations:

homozygote: nnnnnnnnnn[(_)]nnnnnnnnnn

homozygote: nnnnnnnnnn[(T)]nnnnnnnnnn

heterozygote: nnnnnnnnnn[(_)+(T)]nnnnnnnnnn

Microsatellite or short tandem repeat (STR)

Definition: Alleles consist of a repeated sequence motif and the number of tandem copies of this motif. Expansion of the motif into full-length sequence will be only an approximation of the true genomic sequence because microsatellite markers are typically not fully sequenced and are resolved as size variants only.

Example_1 allele representations:

dbSNP: (TAGATCATGCTGGAGCTTCTGGTGGG)28/41/49

HGVbaseG2P: nnnnnnnnnn((TAGATCATGCTGGAGCTTCTGGTGGG)28)nnnnnnnnnn
nnnnnnnnnn((TAGATCATGCTGGAGCTTCTGGTGGG)41)nnnnnnnnnn
nnnnnnnnnn((TAGATCATGCTGGAGCTTCTGGTGGG)49)nnnnnnnnnn
...note: the () enclosing brackets around each allele

Example_2 allele representations:

dbSNP: (A)1/10/11/T

HGVbaseG2P: nnnnnnnnnn(A)nnnnnnnnnn
nnnnnnnnnn((A)10)nnnnnnnnnn
nnnnnnnnnn((A)11)nnnnnnnnnn
nnnnnnnnnn(T)nnnnnnnnnn
...note: the () enclosing brackets around each allele

Example_3 allele representations:

dbSNP: (CA)11CGCACA(CG)6(CA)8/(CA)13CGCACA(CG)6(CA)8/(CA)14CGCACA(CG)6(CA)8/(CA)14CGCACA(CG)7(CA)8/(CA)15CGCACA(CG)6(CA)8/(CA)15CGCACA(CG)7(CA)8/(CA)17CGCA/(CA)18CGCACA(CG)7(CA)8/(CA)20CGCACA(CG)7(CA)8/(CA)20CGCACA(CG)7(CA)9/(CA)21CGCACA(CG)7(CA)8

HGVbaseG2P: nnnnnnnnnn((CA)11CGCACA(CG)6(CA)8)nnnnnnnnnn
nnnnnnnnnn((CA)13CGCACA(CG)6(CA)8)nnnnnnnnnn
nnnnnnnnnn((CA)14CGCACA(CG)6(CA)8)nnnnnnnnnn
nnnnnnnnnn((CA)15CGCACA(CG)6(CA)8)nnnnnnnnnn
nnnnnnnnnn((CA)15CGCACA(CG)7(CA)8)nnnnnnnnnn
nnnnnnnnnn((CA)17CGCA)nnnnnnnnnn
nnnnnnnnnn((CA)18CGCACA(CG)7(CA)8)nnnnnnnnnn
nnnnnnnnnn((CA)20CGCACA(CG)7(CA)8)nnnnnnnnnn
nnnnnnnnnn((CA)20CGCACA(CG)7(CA)9)nnnnnnnnnn
nnnnnnnnnn((CA)21CGCACA(CG)7(CA)8)nnnnnnnnnn
...*note: the () enclosing brackets around each allele*

Mixed variants

Definition: Markers that are comprised of alleles of different variation classes

Example_1 allele representations:

dbSNP: -/A/ATTTA/T

HGVbaseG2P: nnnnnnnnnn(_)nnnnnnnnnn
nnnnnnnnnn(A)nnnnnnnnnn
nnnnnnnnnn(ATTTA)nnnnnnnnnn
nnnnnnnnnn(T)nnnnnnnnnn
...*note: the* '-' allele is replaced by '_'

Example_2 allele representations:

dbSNP: -/G/T/TG/TTTTTTTG/TTTTTTTTTTTTTTG

HGVbaseG2P: nnnnnnn(_)nnnnnnnnnn
nnnnnnn(G)nnnnnnnnnn
nnnnnnn(T)nnnnnnnnnn
nnnnnnn(TG)nnnnnnnnnn
nnnnnnn((T)7G))nnnnnnnnnn
nnnnnnn((T)14G))nnnnnnnnnn
...*note: the* '-' allele is replaced by '_'

Named variants (Named)

Definition: insertion/deletion polymorphisms of longer sequence features, such as retroposons (presence or absence), Alus or LINEs. These variations frequently include a deletion "-" indicator for the absent allele.

Example_1 allele representations:

dbSNP: (LARGEDELETION)/-/G/T

HGVbaseG2P: nnnnnnnnnn(“LARGEDELETION”)nnnnnnnnnn
nnnnnnnnnn(_)nnnnnnnnnn
nnnnnnnnnn(G)nnnnnnnnnn
nnnnnnnnnn(T)nnnnnnnnnn
...note: quotation marks added to text-based allele
...note: the ‘-’ allele is replaced by ‘_’

Example_1 allele representations:

dbSNP: ([CT]+[CA]+[CT]STR, LENGTH 204)/([CT]+[CA]+[CT]STR, LENGTH194)/([CT]+[CA]+[CT]STR, LENGTHS 190-208)

HGVbaseG2P: nnnnnnnnnn(“[CT]+[CA]+[CT]STR, LENGTH 204”)nnnnnnnnnn
nnnnnnnnnn(“[CT]+[CA]+[CT]STR,LENGTH194”)nnnnnnnnnn
nnnnnnnnnn(“[CT]+[CA]+[CT]STR,LENGTHS190-208”)nnnnnnnnnn
...note: quotation marks added to text-based alleles