

TOP Format

At a given SNP, a strand is called TOP and the other BOT.

This definition depends on the alleles of the SNP or, when necessary, on the adjacent sequences.

Alleles	Strand	TOP A Allele	TOP B Allele
A-C	TOP	A	C
A-G	TOP	A	G
C-T	BOT	A	G
G-T	BOT	A	C
A-T	Undetermined, see below		
C-G	Undetermined, see below		

In case of undetermined, we look at the adjacent bases, on the right and on the left of the polymorphic base.

Alleles	Left base	Right base	Strand	A Allele	B Allele
A-T	A or T	C or G	TOP	A	T
A-T	C or G	A or T	BOT	T	A
A-T	A or T	A or T	Undetermined, see next base (s)		
A-T	C or G	C or G	Undetermined, see next base(s)		
C-G	A or T	C or G	TOP	C	G
C-G	C or G	A or T	BOT	G	C
C-G	A or T	A or T	Undetermined, see next base(s)		
C-G	C or G	C or G	Undetermined, see next base(s)		

Examples:

Sequence	Strand	A Allele (top or bot)	B Allele (top or bot)	TOP A Allele	TOP B Allele
tactc(G-A)gatcgt	TOP	A	G	A	G
cgagt(C-T)ttacct	BOT	T	C	A	G
agccc(T-G)gataga	BOT	T	G	A	C
tactcggtct(A-T)gcattcgatcgt	TOP	A	T	A	T
cgtagcatat(A-T)ttatcgatcgt	BOT	T	A	A	T
acgatcgataaa(A-T)atagctaacg	TOP	A	T	A	T
tactgctctat(G-C)gcccatcagatc	TOP	G	C	G	C

Note: No relationship between TOP/BOT and FOR/REV => 50% of the TOP strands are FOR. Same with REF/ALT.