

Ambiguous read assignment

Assignment of Mutant and Wild type reads:

While recalculating indel allele fraction using exonerate if a read sequence matches with both ALT and REF allele with identical score one can't confidently assign these reads to either of the alleles.

Example:

Tandem duplication at indel location:

Positions from VCF file

Chr	Pos	Ref	Alt
10	114911505	G	GGAAGAAGAAAA

Ref Sequence: [seq already in ref]

TGTTTATTGGGGGTTATGAAAAGGAAGAAGAAAAGGAAGAAGCCCCACATAAAGAAACCTCTTAATGCATTC

Alt Sequence: after insertion the region become tandem duplication.

TGTTTATTGGGGGTTATGAAAAGGAAGAAGAAAAGGAAGAAGAAAAGGAAGAAGCCCCACATAAAGAAACCTCTTAATGCATTC

New extended ref location calculation:

REF: TGTTTATTGGGGGTTATGAAAAGGAAGAAGAAAAGGAAGAAGCCCCACATAAAGAAACCTCTTAATGCATTC

ALT: TGTTTATTGGGGGTTATGAAAAGGAAGAAGAAAAGGAAGAAGAAAAGGAAGAAGCCCCACATAAAGAAACCTCTTAATGCATTC

207 218 226 229

Original indel position: 5' = 207, 3' = 218

New 5' ref = First mismatch position ==>226

New 3' ref = First mismatch position ==>226

New 3' alt = 229

Calculating 3' Alt:

insert length = 11 bp

insert modulus = ((226 - 207) % 11) = 8

uniquely inserted bases = insert length (11) - insert modulus (8) = 3

New alt 3'pos = 226 + 3 = 229

New indel position: 5' = 226, 3' = 229

Any read that overlaps above location will be calculated against respective allele.

Exonerate alignment for alt and ref sequence at indel location:

Indel position in below alignment is from 207-218, read completely overlaps with underlying indel sequence at same position in ref and alt.

Alignment

Query: HS33_12286:5:1205:12866:52259#13_30 [revcomp]

Target: alt

Model: ungapped:dna2dna

Raw score: 375

Query range: 75 -> 0

Target range: 149 -> 224

75 : TGTTTATTGGGTTGCGTCTGTTTTGTCTATCTCCAGAAAGCATCAGGACTCCAAAAAGGAAGAA :
12

|||||

150 : TGTTTATTGGGTTGCGTCTGTTTTGTCTATCTCCAGAAAGCATCAGGACTCCAAAAAGGAAGAA :
213

11 : GAAAAGAAGAA : 1

|||||

214 : GAAAAGAAGAA : 224

HS33_12286:5:1205:12866:52259#13_30 alt 75 + 149 224 -

C4 Alignment:

Query: HS33_12286:5:1205:12866:52259#13_30 [revcomp]

Target: ref

Model: ungapped:dna2dna

Raw score: 375

Query range: 75 -> 0

Target range: 149 -> 224

75 : TGTTTATTGGGTTGCGTCTGTTTTGTCTATCTCCAGAAAGCATCAGGACTCCAAAAAGGAAGAA :
12

|||||

150 : TGTTTATTGGGTTGCGTCTGTTTTGTCTATCTCCAGAAAGCATCAGGACTCCAAAAAGGAAGAA :
213

11 : GAAAAGAAGAA : 1

|||||

214 : GAAAAGAAGAA : 224