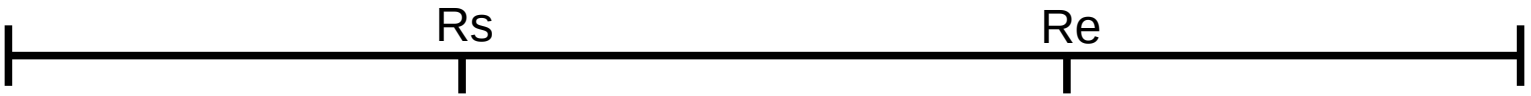
Overlap types

Defining feature-overlaps between sequences

Sequence-position definitions

- Reference Sequence
  - Rs is the **s**tarting location/position of a *feature* within the Reference
  - Re is the **e**nding location of a *feature* within the reference
  - For the purposes of this exercise (case where Rs-Re is an intron) Re is always greater in number than Rs and location number on the sequence increases from left to right



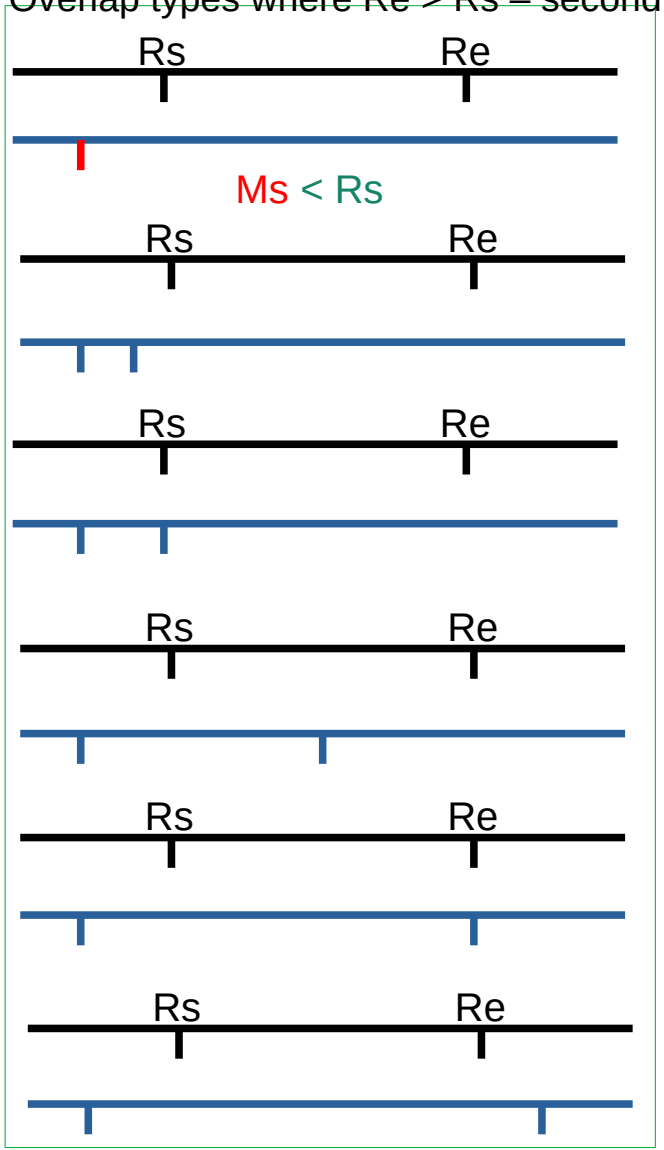
- Variant (mutant) sequence
  - Ms is the starting location of a *feature* within the Variant
  - Me is the ending location of a *feature* within the reference
  - When Ms has a higher value than Me it defines an *insertion* and Ms = Me +1(always)

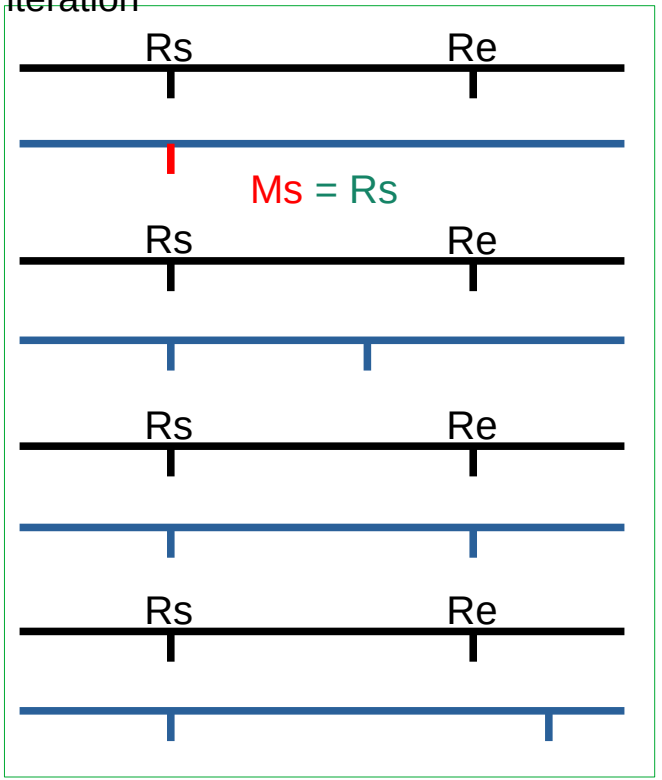
When Ms =Me, colour shown as purple

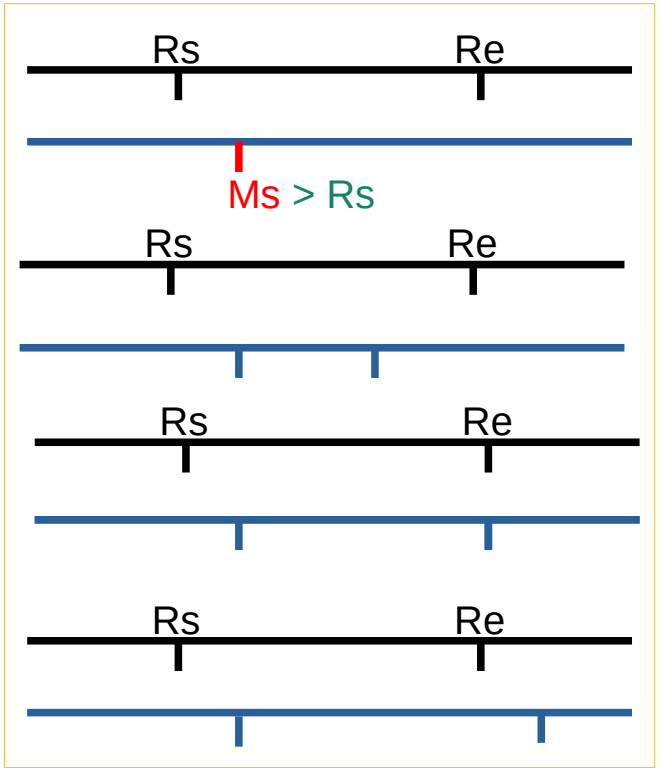
Overlap definitions

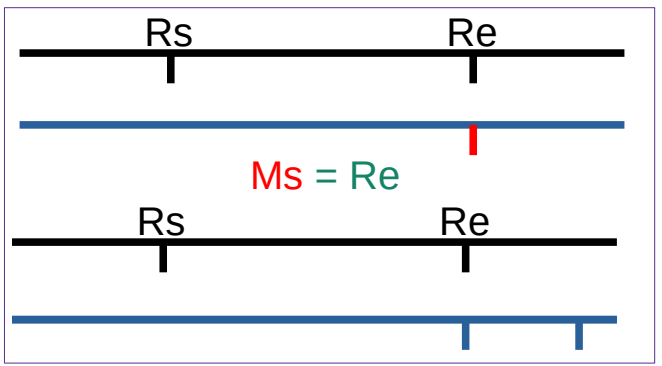
- The first iteration of this analysis (different number-set) was not systematic, missing cases where Ms = Me.
- Second iteration considered only definitions where Re > Rs
  - Simply because the main case I am interested in is where the range Re ... Rs is an intron range. This is called “skip” or “gap” in the code, to more generically describe the feature
  - The diagrams did not show a distinction in colour between Ms and Me
- Third iteration shows a distinction in colour between Ms & Me in diagrams
  - It thereby expands the number of cases to include where Ms = Me + 1
  - Specifically the second-iteration cases 2, 3, 13, 16, 18 have another type
  - The other second-iteration cases do not have another type because these are clearly separated by > 1 given their relationships to Re & Rs
  - Eg: if one of the points Re or Rs lie between Ms & Me, then it is impossible for Ms = Me + 1
- Re … Rs effectively becomes a mask on the variants that are to be included
- Where a deletion crosses the skip-boundary, it must be detected to retain the part of the deletion that lies outside the range.

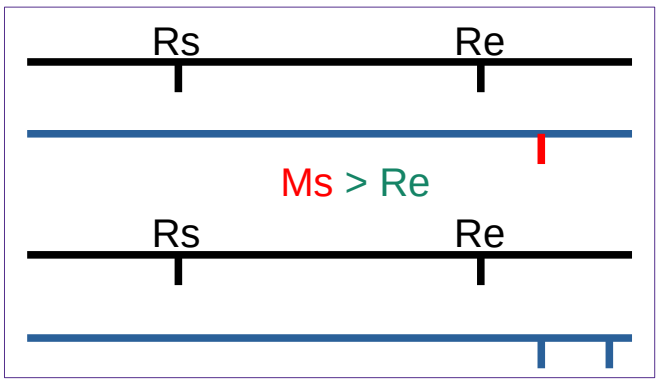
Overlap types where Re > Rs – second iteration



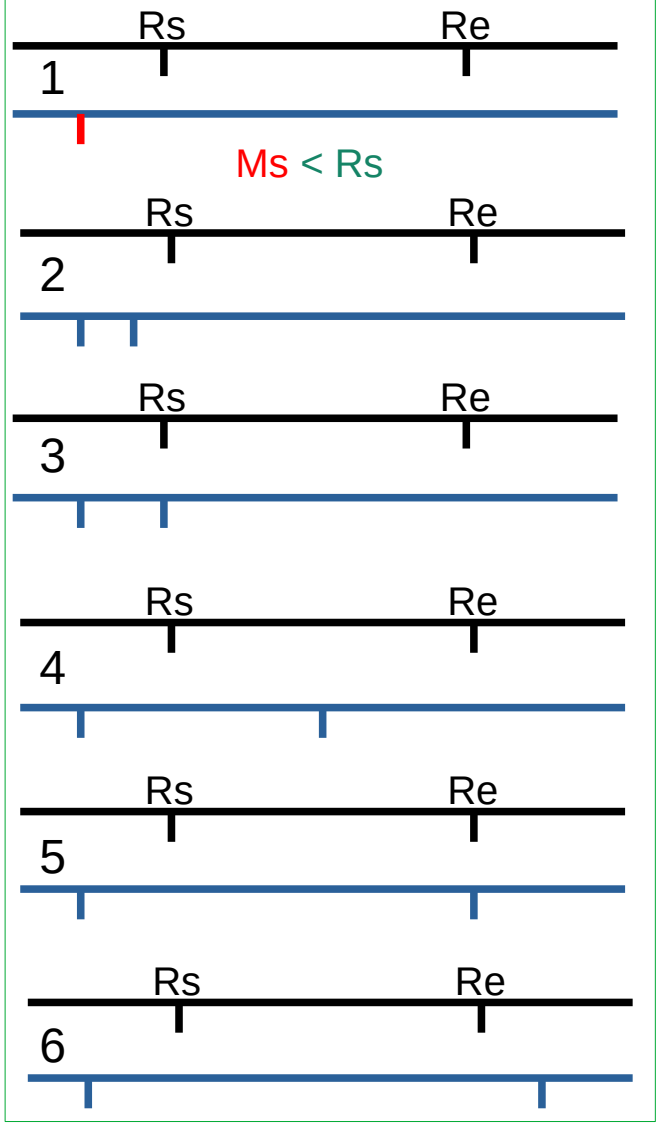


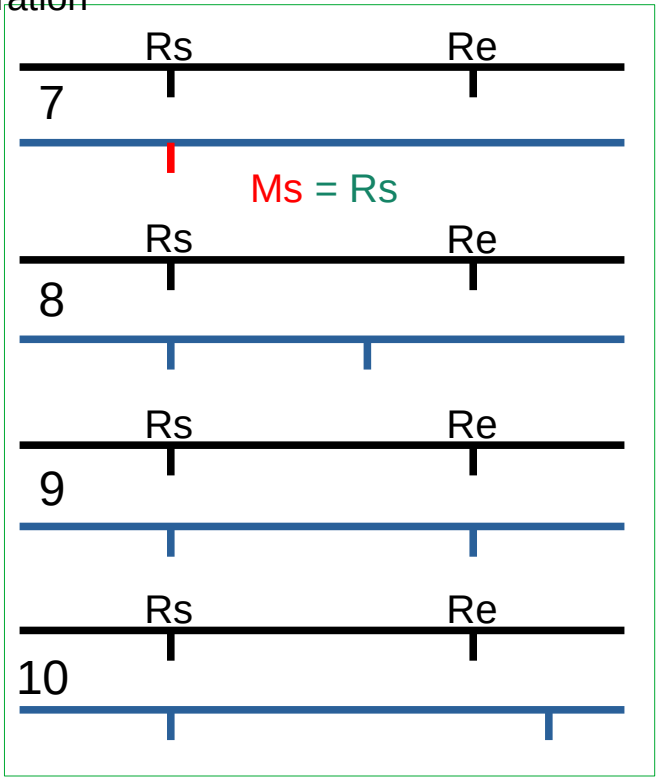


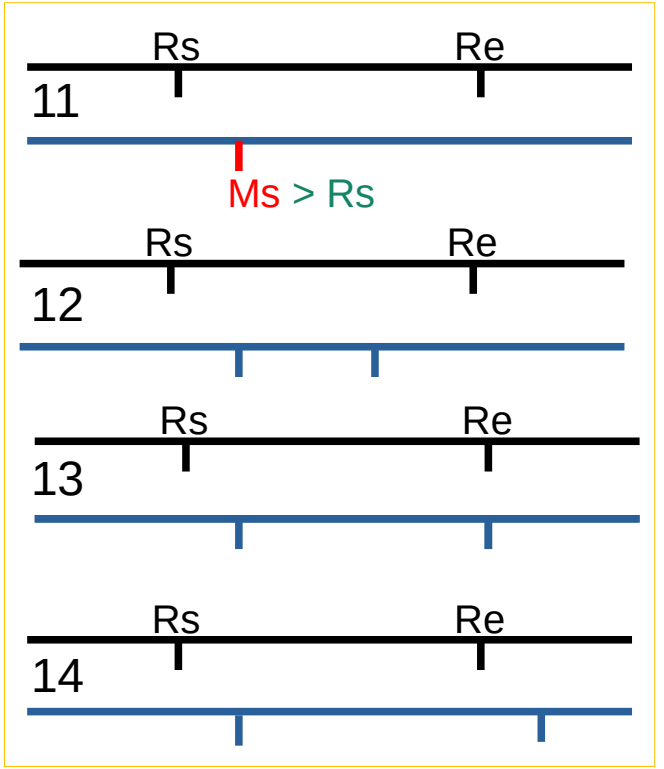


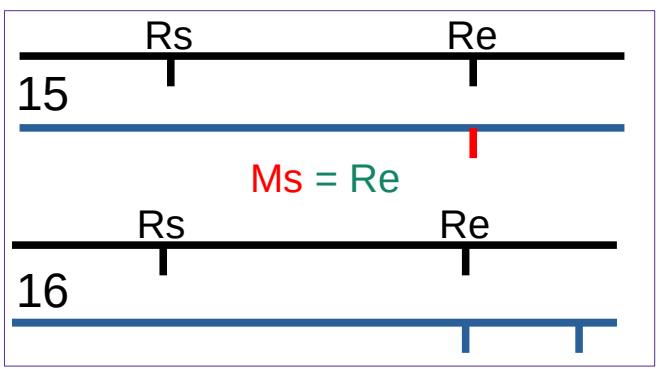


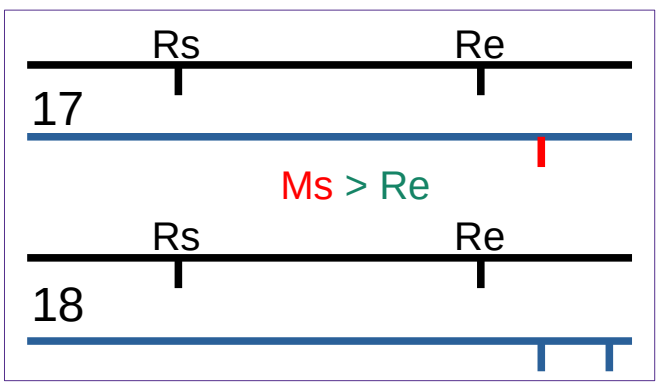
Numbering overlap types for second iteration





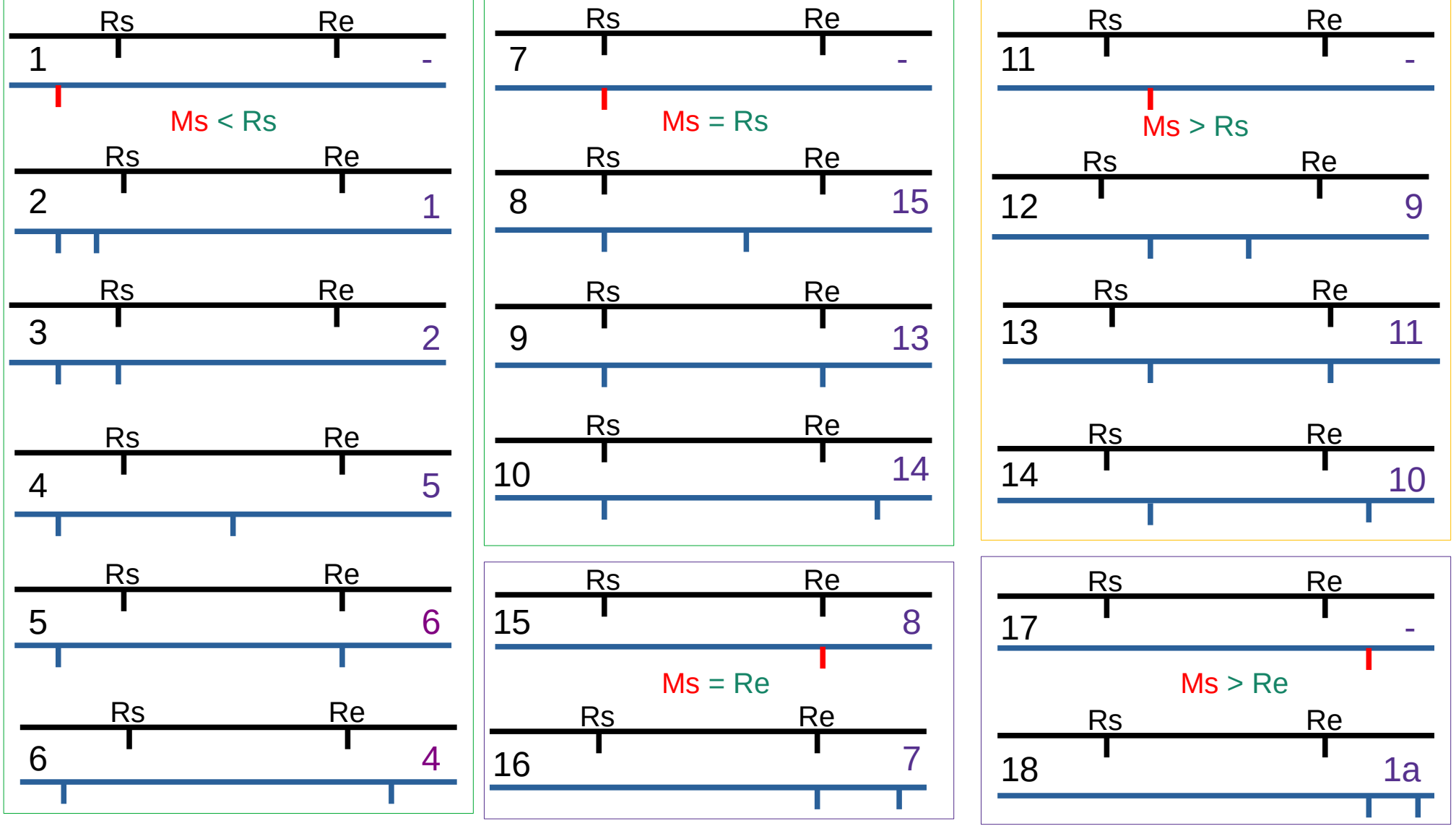




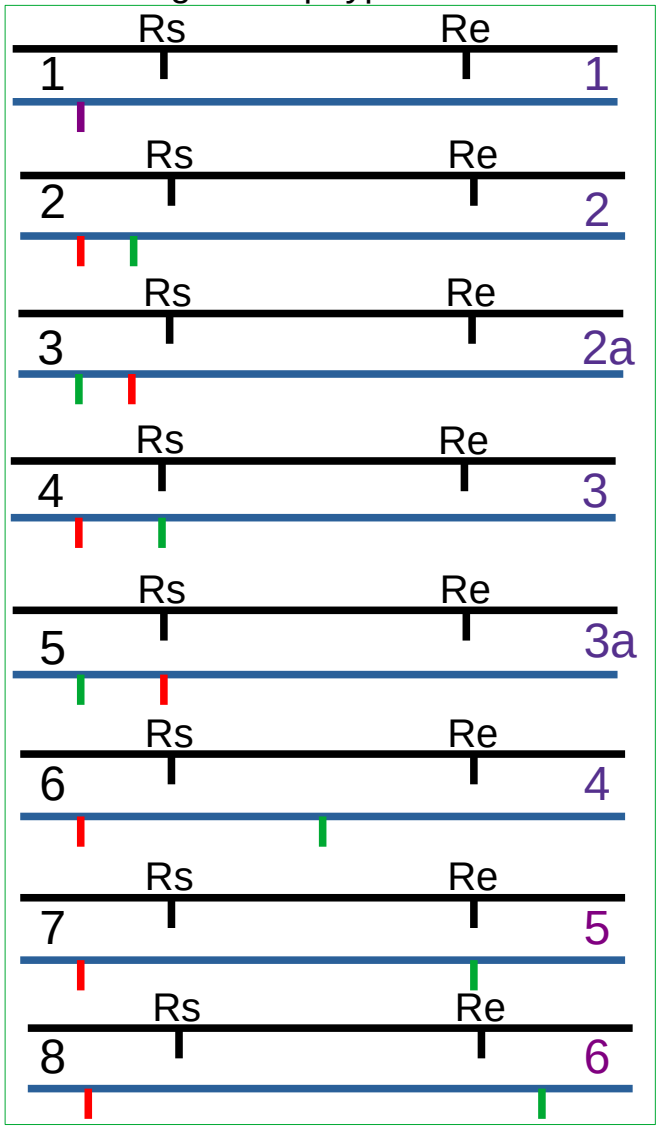


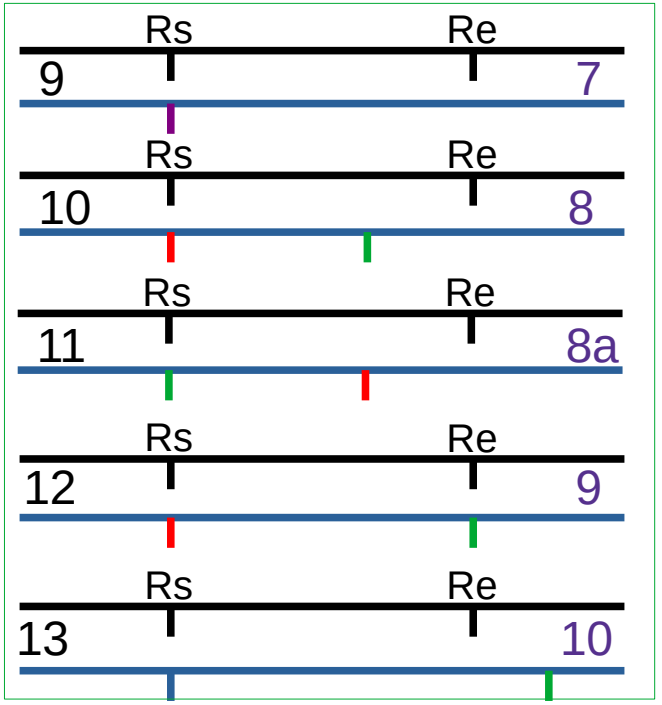
Numbering overlap types for second iteration

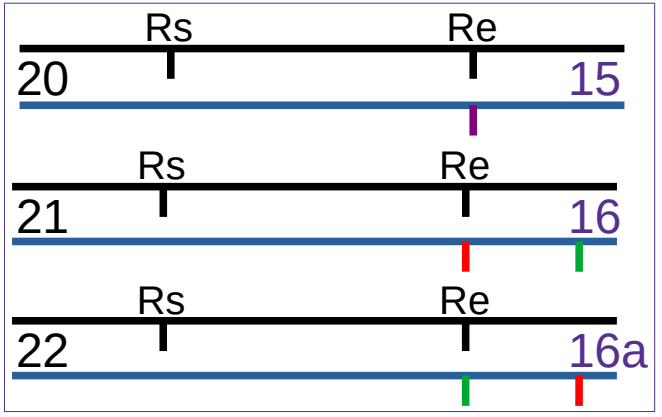
First iteration numbering



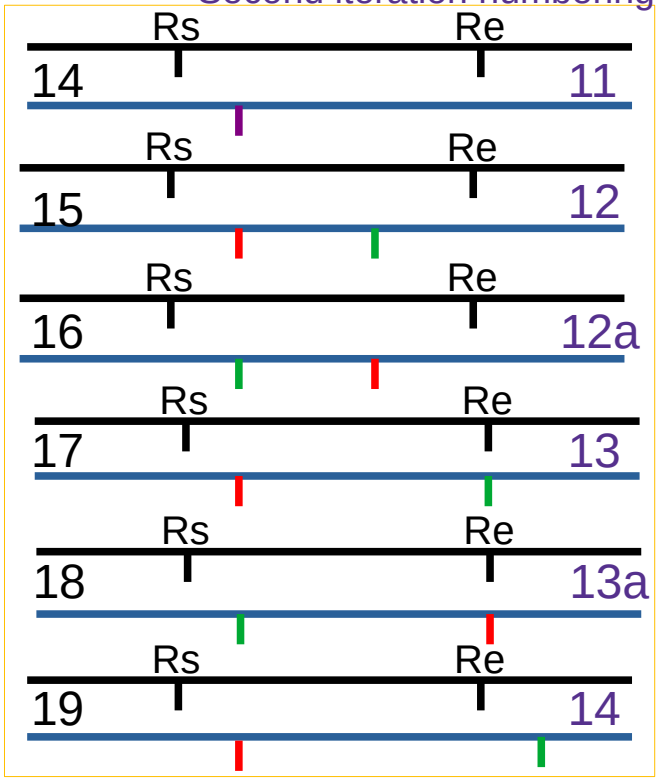
Numbering overlap types for third iteration

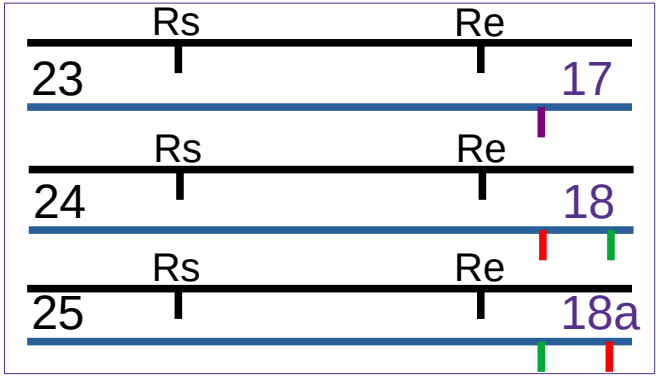






Second iteration numbering





FI: Overlap type 1
SI: OL 2; TI: ol2

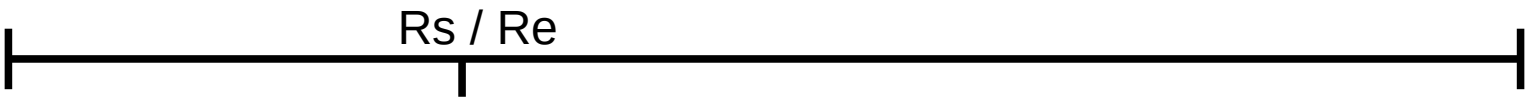
- First iteration (FI) overlap definitions

FI: Overlap type 2
SI: OL 3;TI: OL 4

- Distal concatenation (Ref distal to Var)
  - (Ms < Rs) AND (Me = Rs) AND (Re > Rs)

FI: Overlap type 3
SI: Not defined

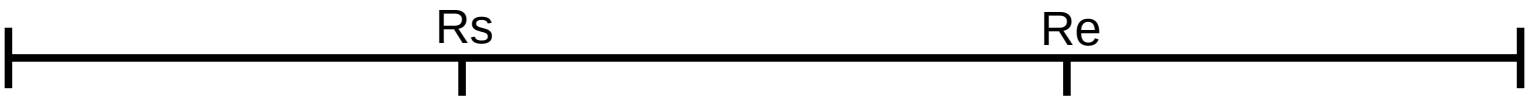
- Distal concatenation (Ref distal to Var)
  - (Ms < Rs) AND (Me = Rs) AND (Re = Rs)



  - Should not occur when Ref feature is a skip (Re > Rs)

FI: Overlap type 4
SI: OL 6; TI: OL 8

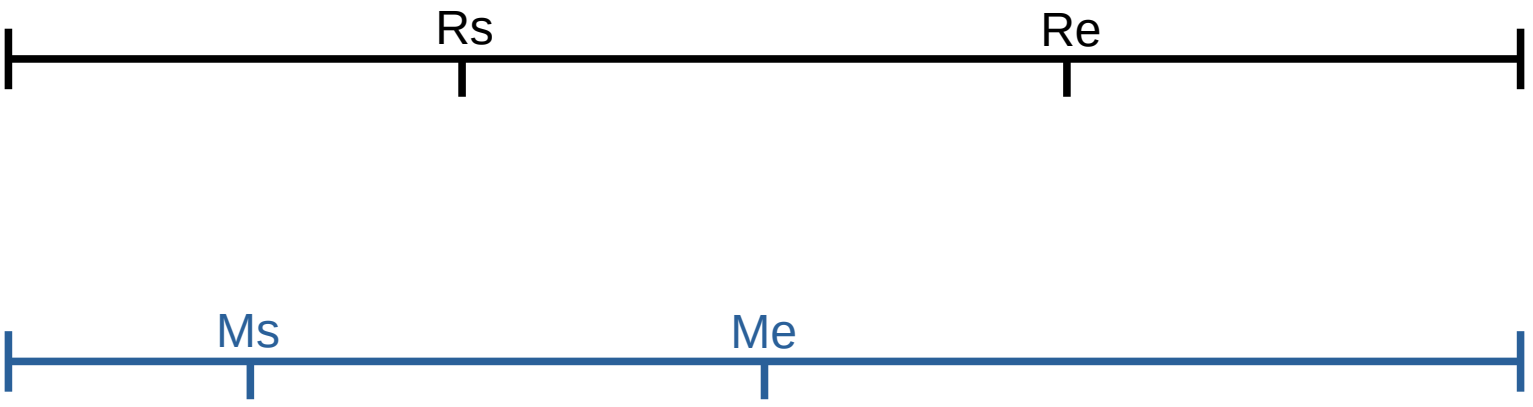
- Reference entirely within variant
  - (Ms < Rs) AND (Me > Re)



  - Re > Rs not tested, but should be true when Ref feature is a skip (Re > Rs)

FI: Overlap type 5
SI: OL 4; TI: OL 6

- Distal overlap (Ref distal to Var)
  - (Ms < Rs) AND (Me < Re)



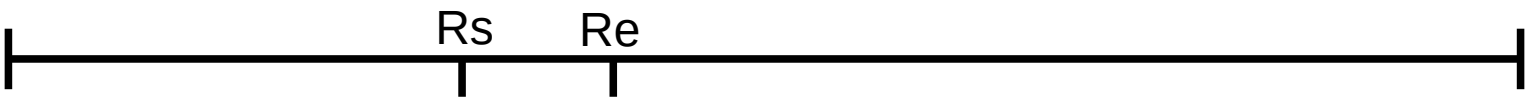
  - Re > Rs not tested, but should be true when Ref feature is a skip (Re > Rs)

FI: Overlap type 6
SI: OL 5; TI: OL 7

- Distal overlap (Ref distal to Var)
  - (Ms < Rs) AND (Me = Re) AND (Re > Rs)

FI: Overlap type 1A
SI: OL 18; TI: OL 24

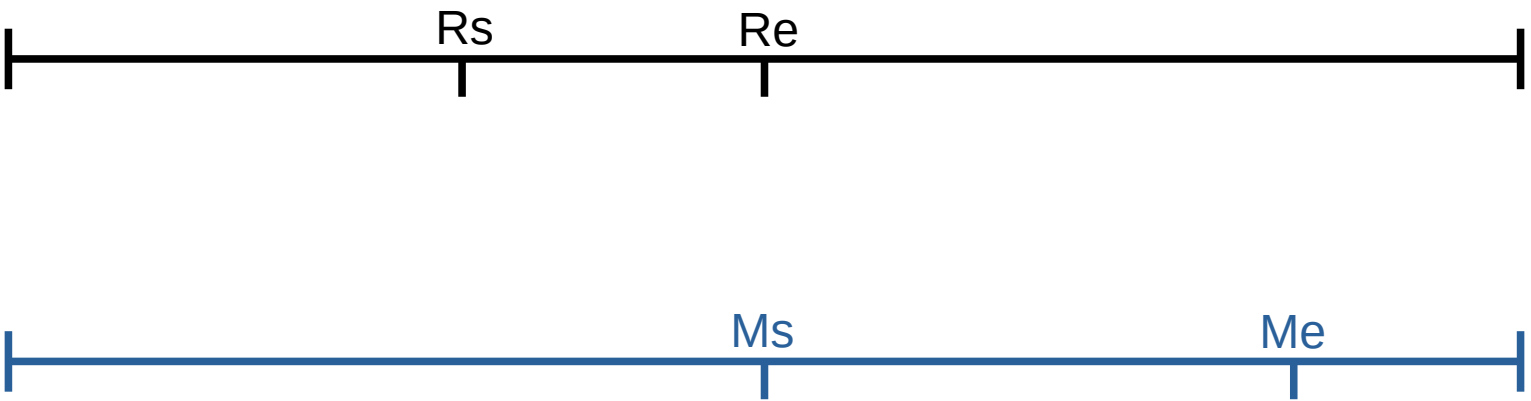
- No overlap
  - (Ms > Rs) AND (Re < Ms)



  - Re > Rs not tested, but should be true when Ref feature is a skip (Re > Rs)

FI: Overlap type 7
SI: OL 16; TI: OL 21

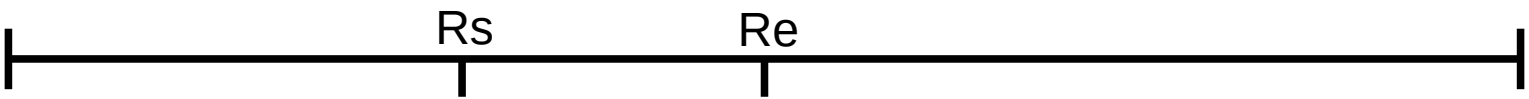
- Proximal concat
  - (Ms > Rs) AND (Re = Ms) AND (Me > Ms)



  - Re > Rs not tested, but should be true when Ref feature is a skip (Re > Rs)

FI: Overlap type 8
SI: OL 15; TI: OL 20

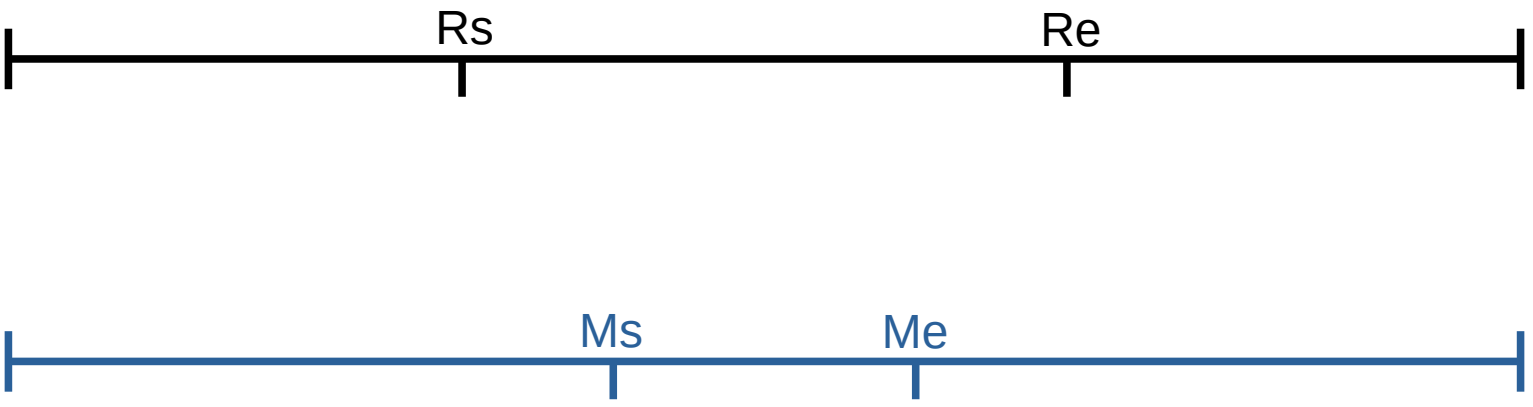
- Proximal point
  - (Ms > Rs) AND (Re = Ms) AND (Me = Ms)



  - Re > Rs not tested, but should be true when Ref feature is a skip (Re > Rs)

FI: Overlap type 9
SI: OL 12; TI: OL 15

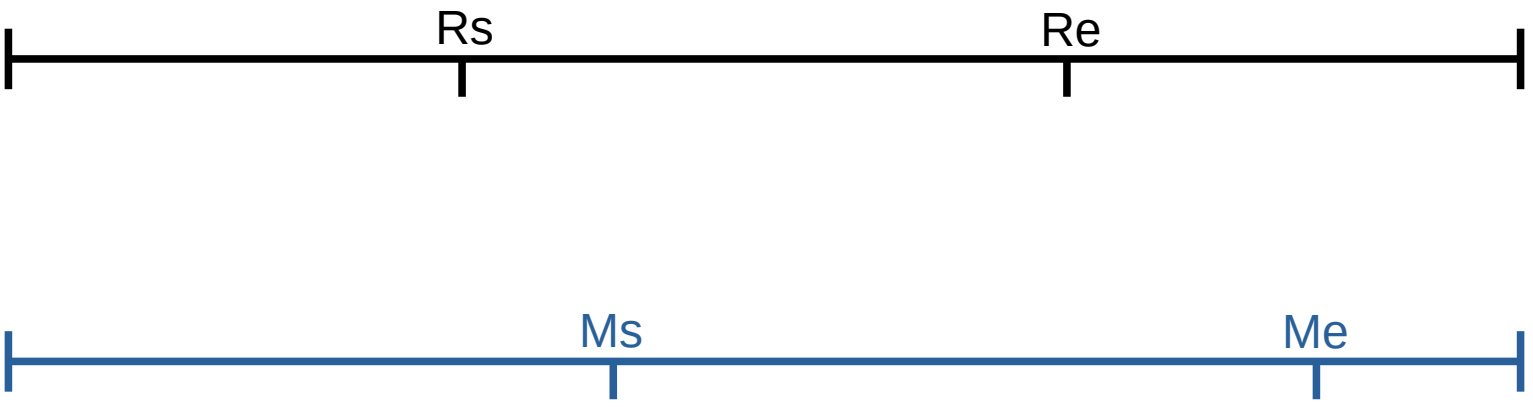
- The Variant range is entirely within the Reference region
  - (Re > Ms) AND (Me < Re)



  - Re > Rs not tested, but should be true when Ref feature is a skip (Re > Rs)

FI: Overlap type 10
SI: OL 14; TI: OL 19

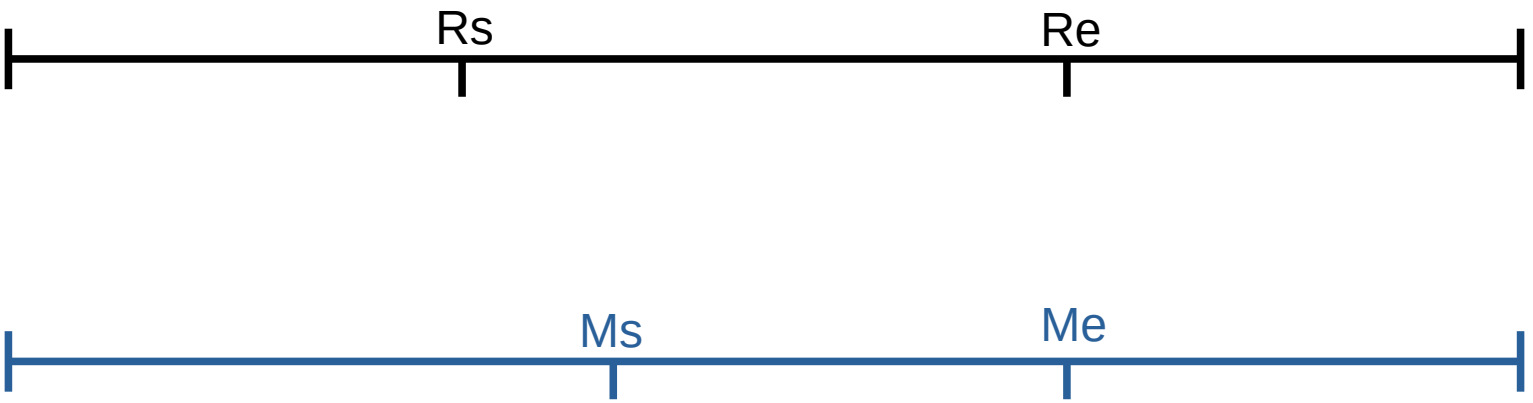
- Proximal overlap (Variant distal to ref)
  - (Re > Ms) AND (Me > Re)



  - Re > Rs not tested, but should be true when Ref feature is a skip (Re > Rs)

FI: Overlap type 11
SI: OL 13 ; TI: OL 17

- Variant within reference - distal
  - (Re > Ms) AND (Me = Re)



FI: Overlap type 12
SI: Not defined

- Coinciding points on both
  - (Rs = Ms) AND (Ms = Re) AND (Ms=Me)
  -

  - Should not occur when Ref feature is a skip (Re > Rs)

FI: Overlap type 13
SI: OL 9; TI: OL 12

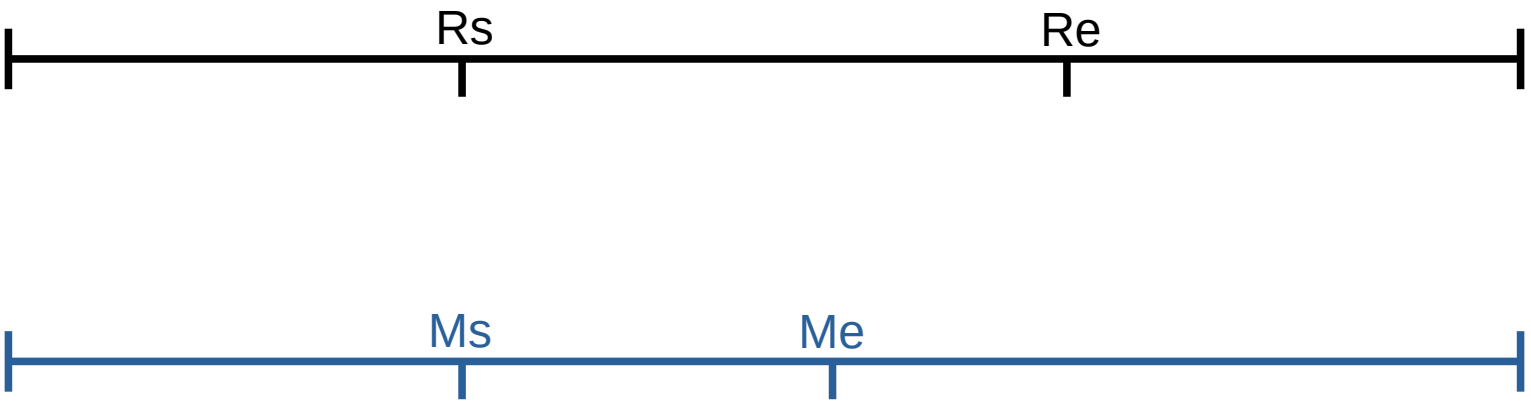
- Coincident feature positions
  - (Rs = Ms) AND (Re=Me)

FI: Overlap type 14
SI: OL 10 ; TI: OL 13

- Coincident start, variant longer than ref
  - (Rs = Ms) AND (Me > Re)

FI: Overlap type 15
SI: OL 8; TI: OL 10

- Coincident start, variant shorter than ref
  - (Rs = Ms) AND (Me < Re)



FI: Type 9
SI: OL 12; TI: OL 15

- Type 9: The Variant range is entirely within the Reference region