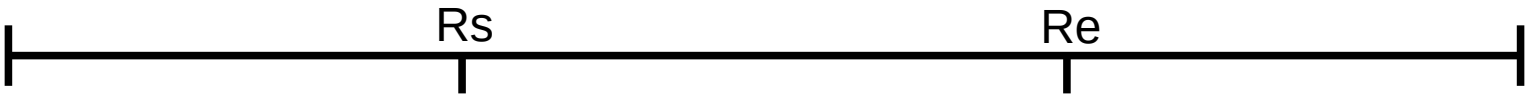
Overlap types

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



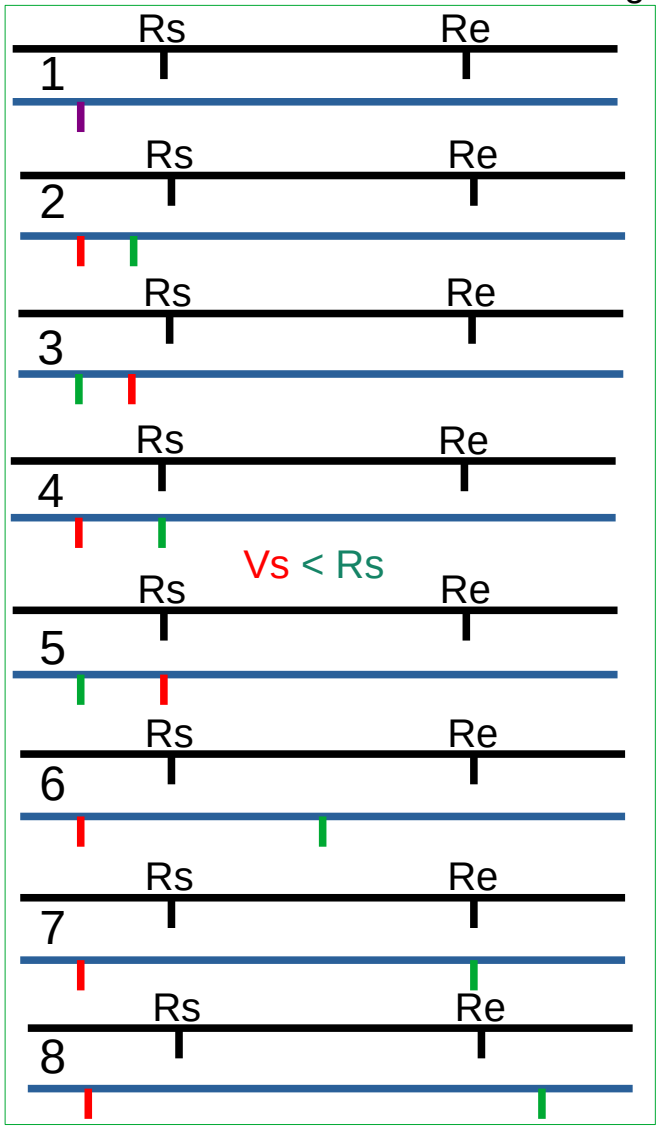
- Variant sequence
  - Vs is the starting location of a *feature* within the Variant
  - Ve is the ending location of a *feature* within the Variant
  - When Vs has a higher value than Ve it defines an *insertion* and Vs = Ve +1(always)

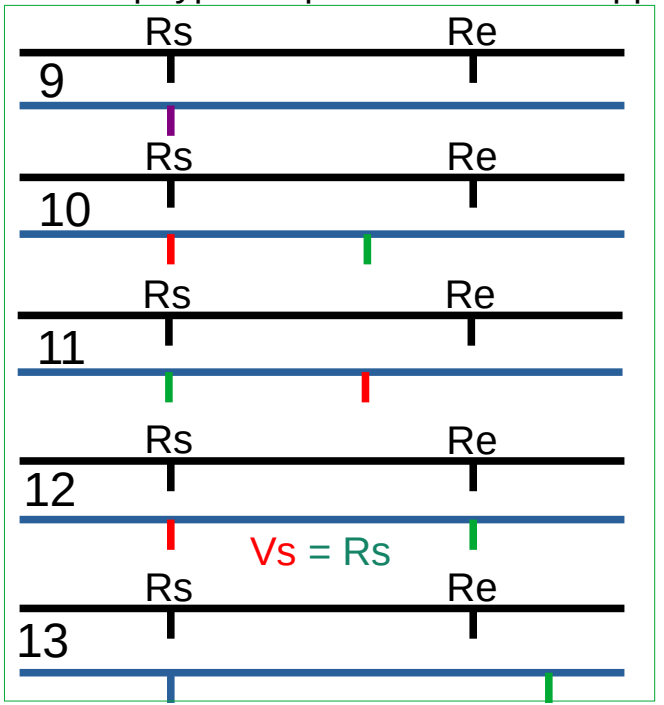
When Vs ==Ve, colour shown as purple

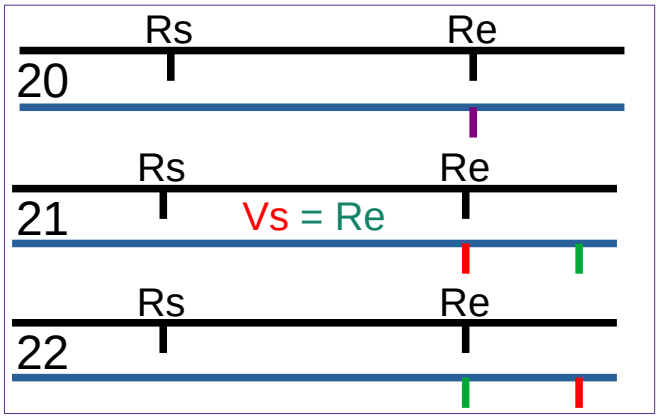
Overlap definitions

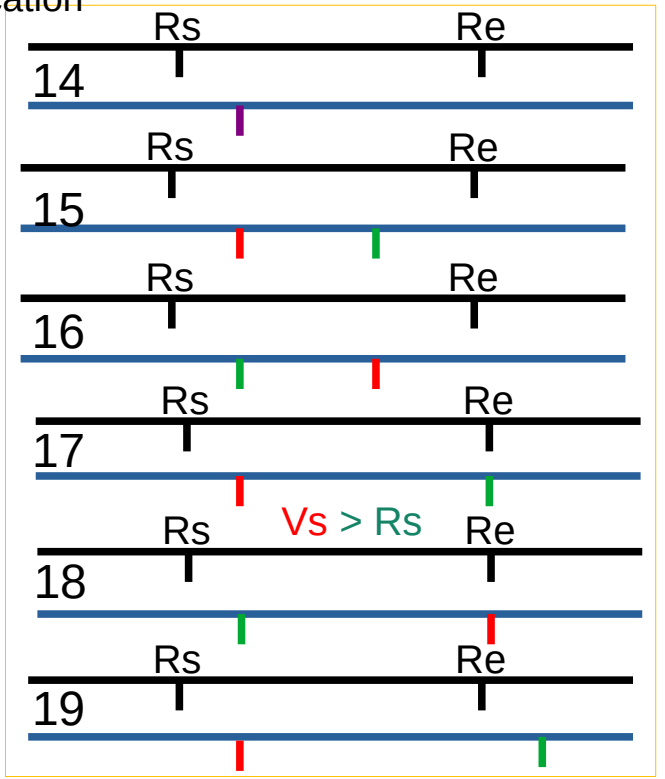
- The first iteration of this analysis (different number-set) was not systematic, missing cases where Vs = Ve.
- Second iteration considered only definitions where Re > Rs
  - Simply because the main case of interest is where the range Re ... Rs is an intron range.
  - This is called “skip” or “gap” in the code, to more generically describe the feature
  - The diagrams did not show a distinction in colour between Vs and Ve
- Third iteration shows a distinction in colour between Vs & Ve in diagrams
  - It thereby expands the number of cases to include where Vs = Ve + 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 Vs & Ve, then it is impossible for Vs = Ve + 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.

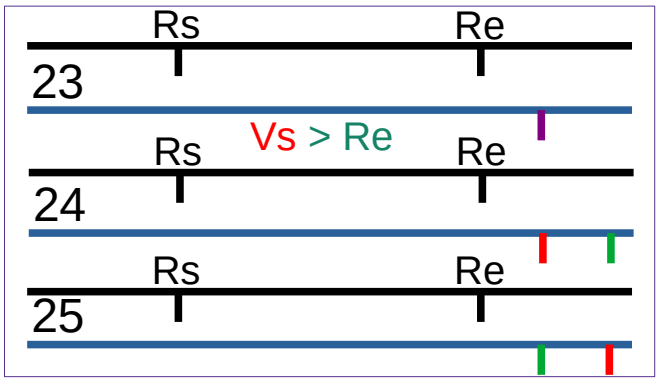
Numbering of overlap types implemented in the application







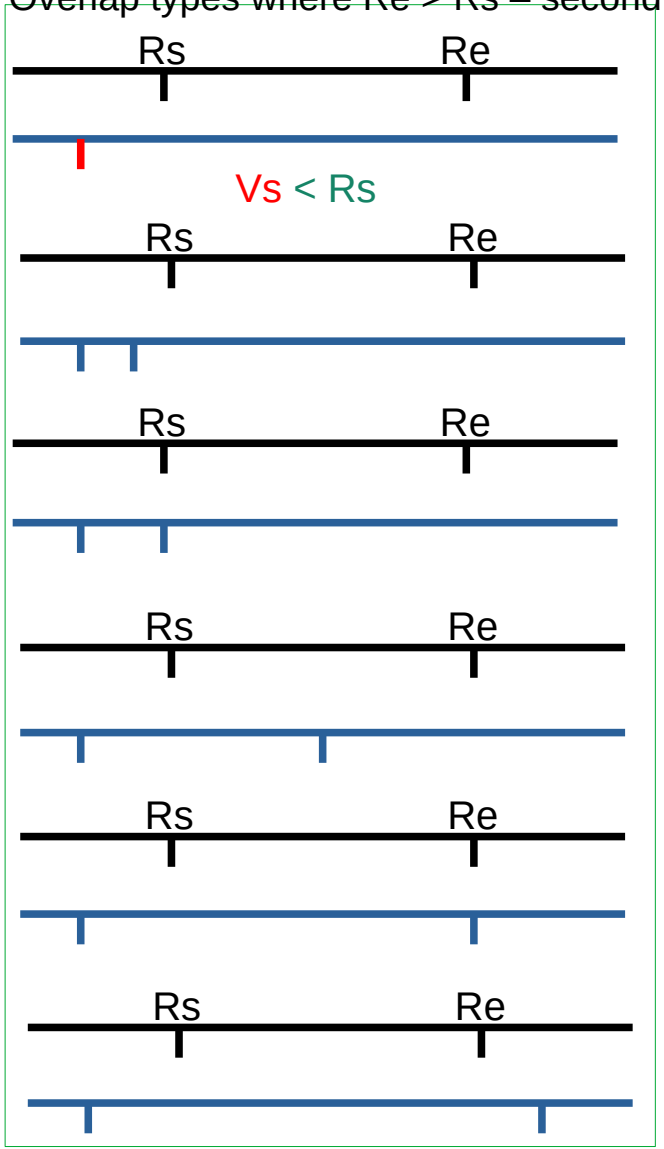


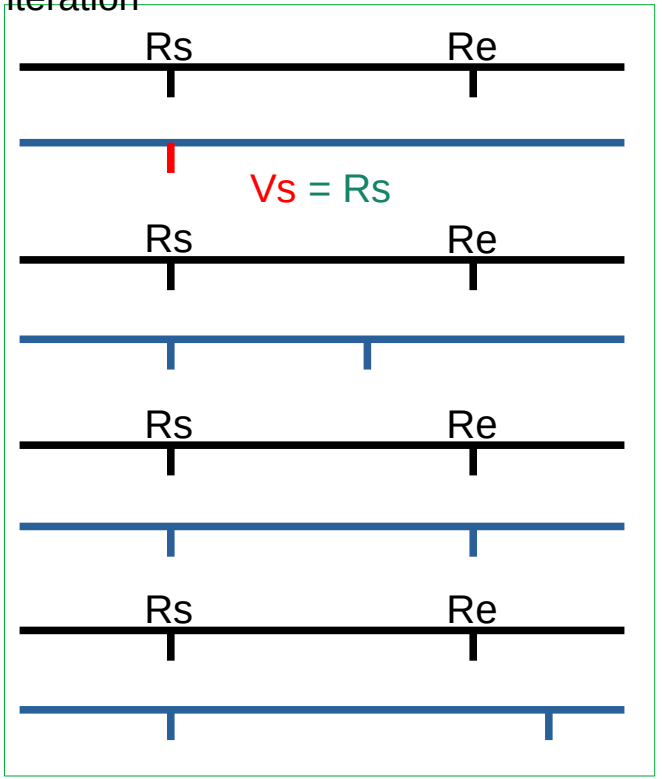


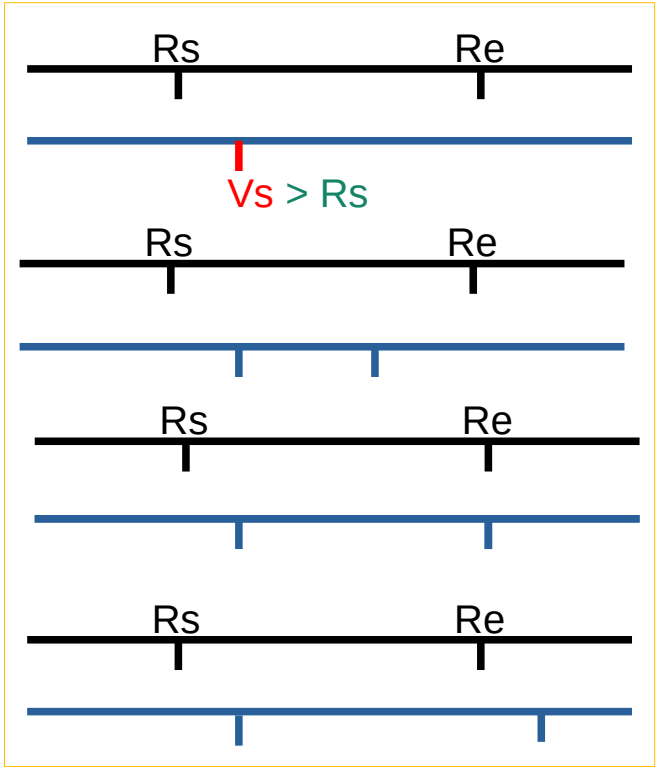
Older iterations documented

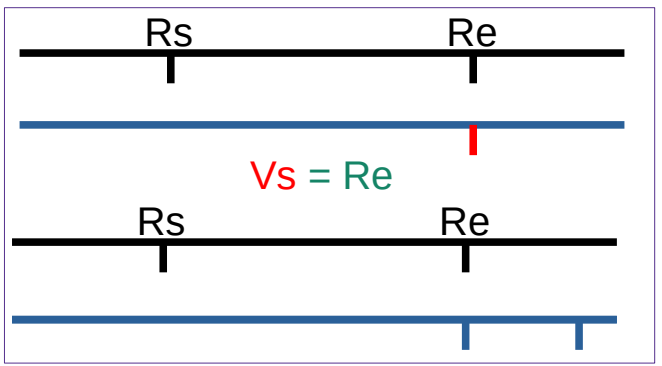
- Legacy documentation, tidied up into

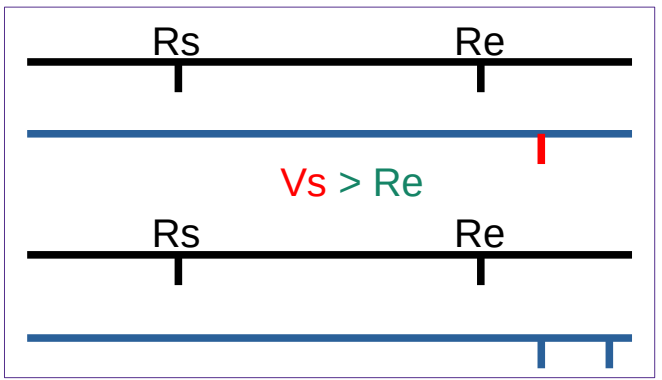
Overlap types where Re > Rs – second iteration



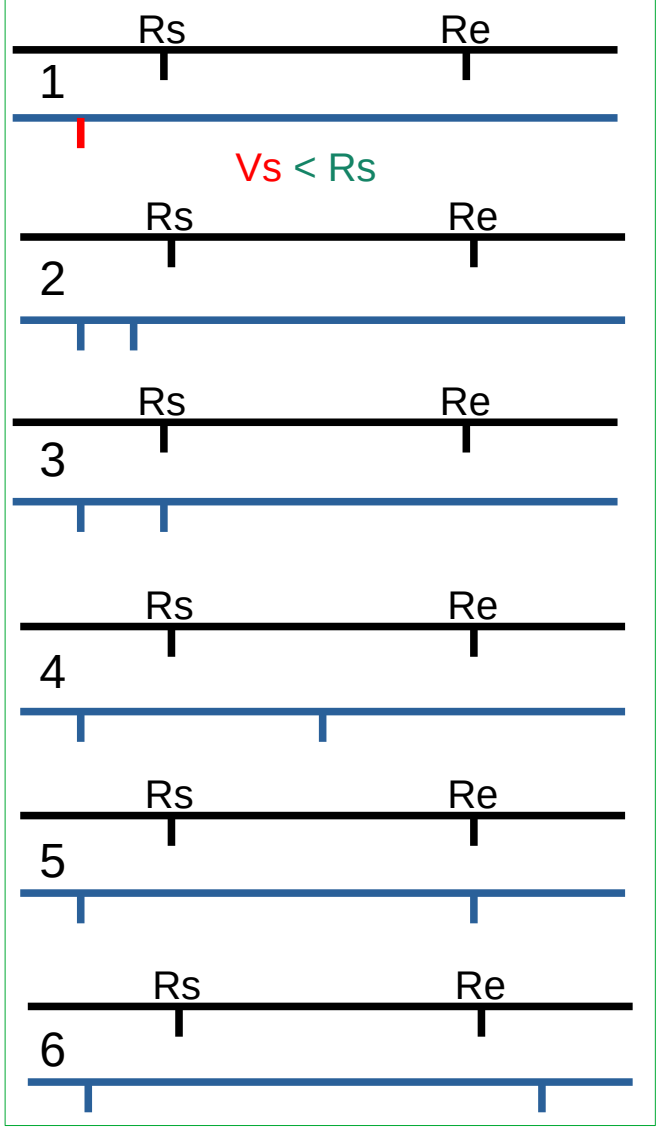


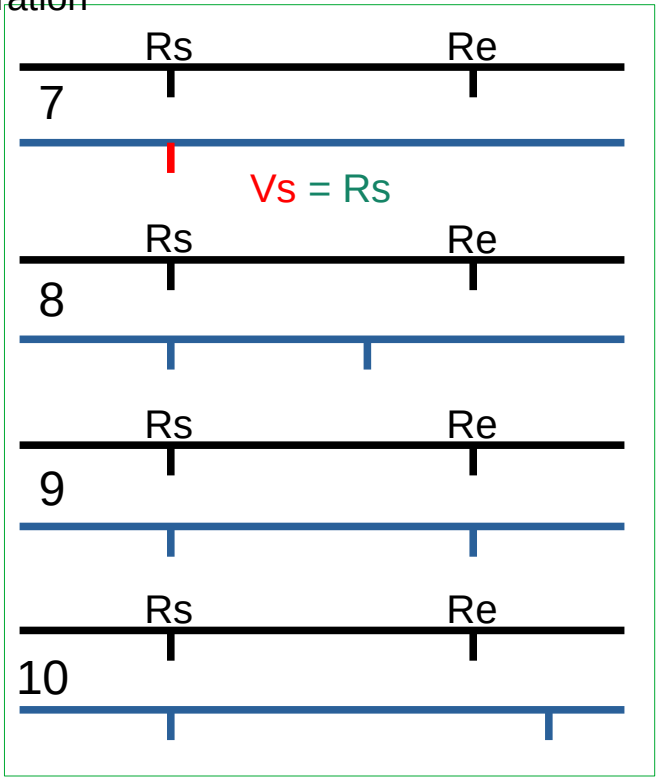


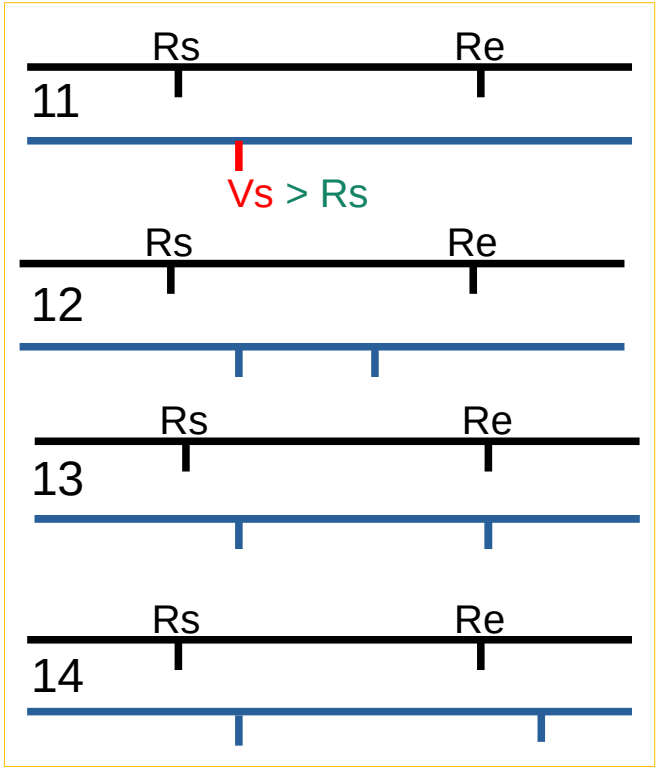


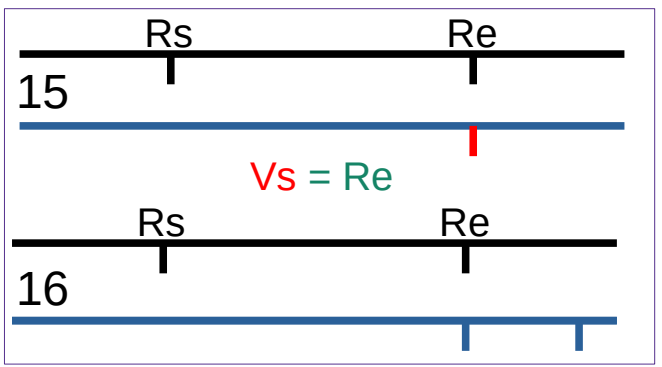


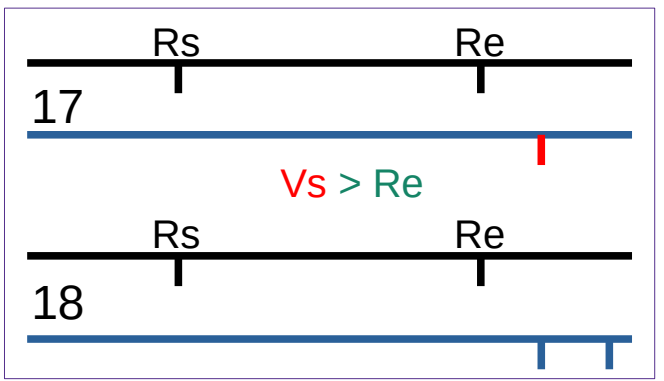
Numbering overlap types for second iteration





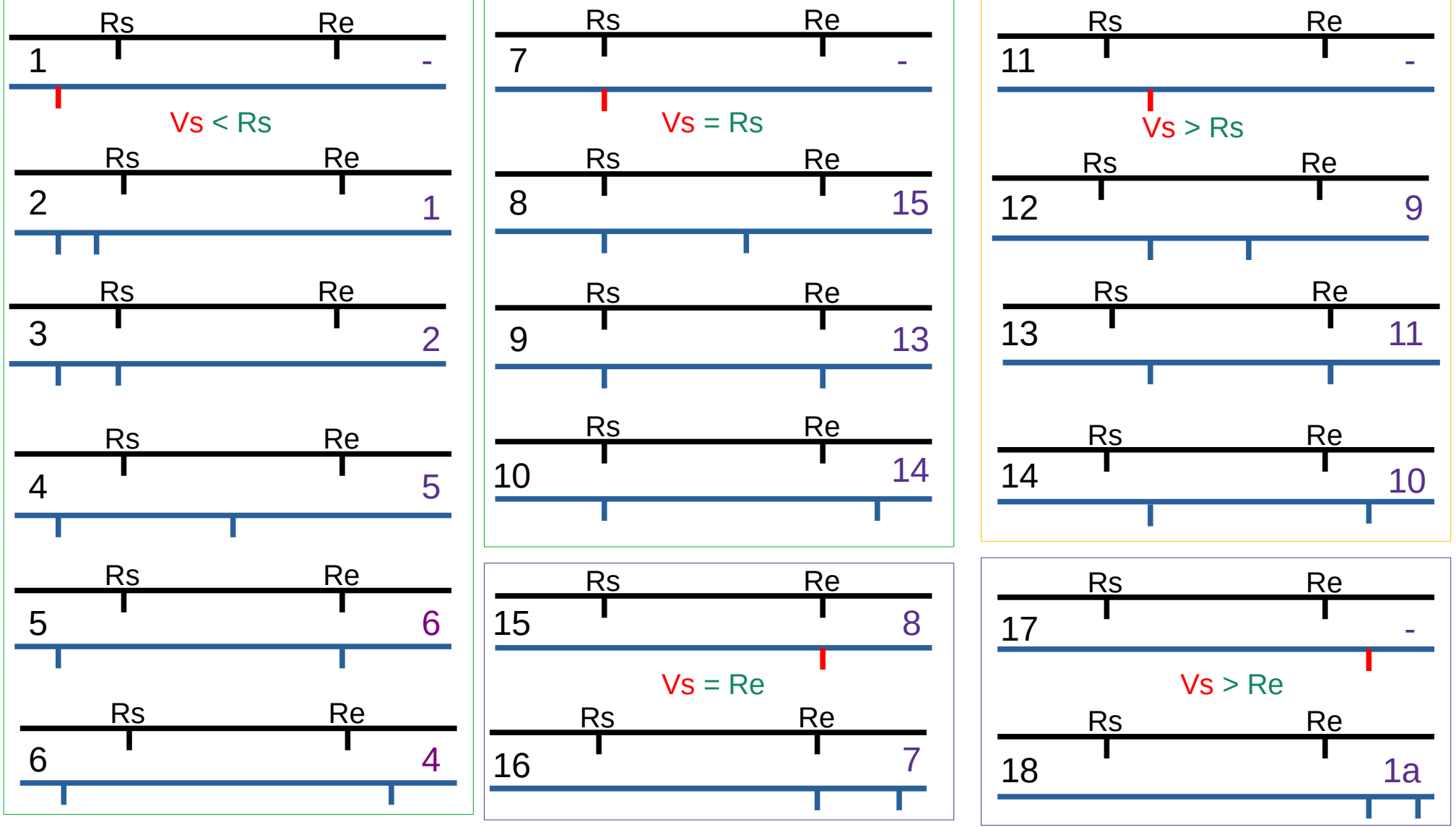


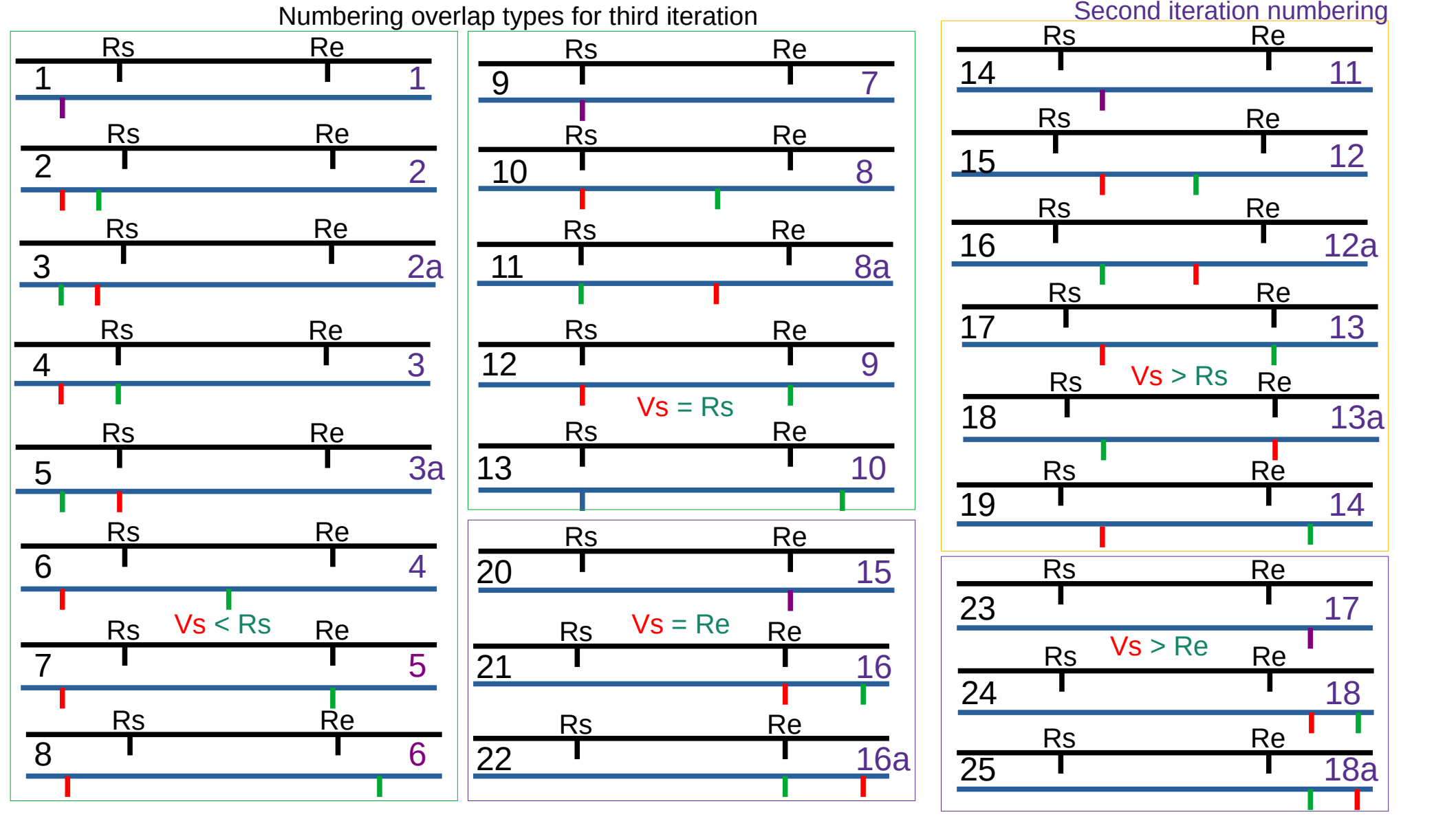




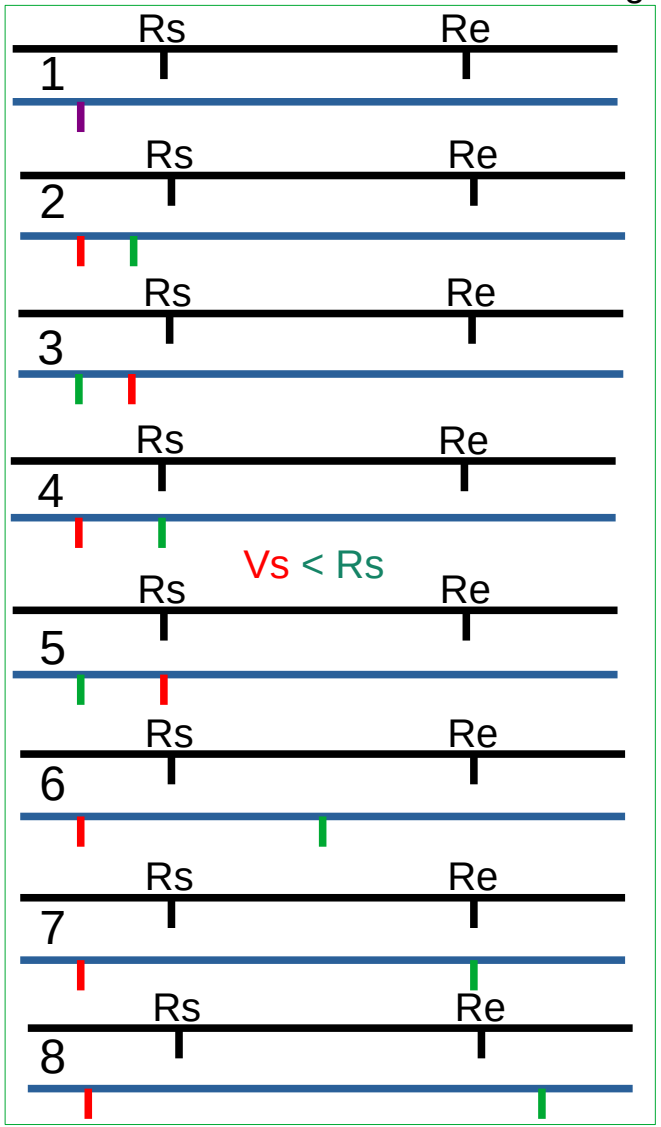
Numbering overlap types for second iteration

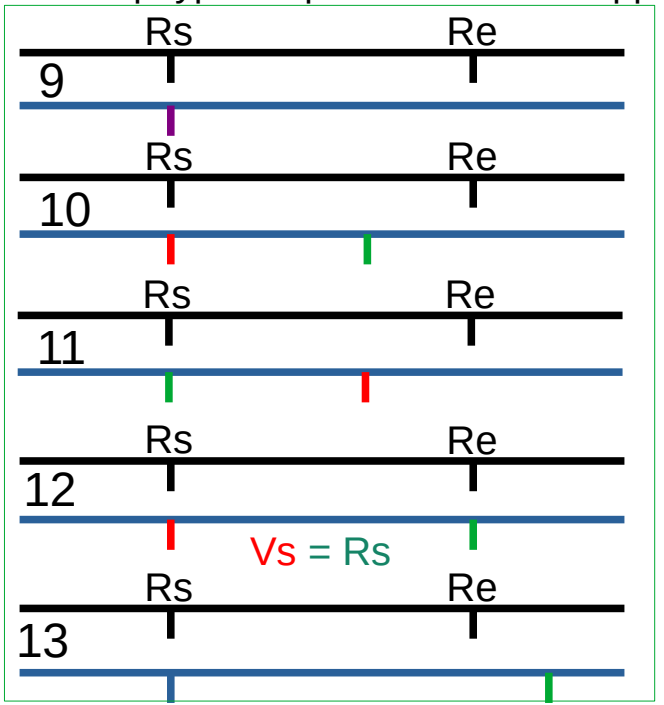
First iteration numbering

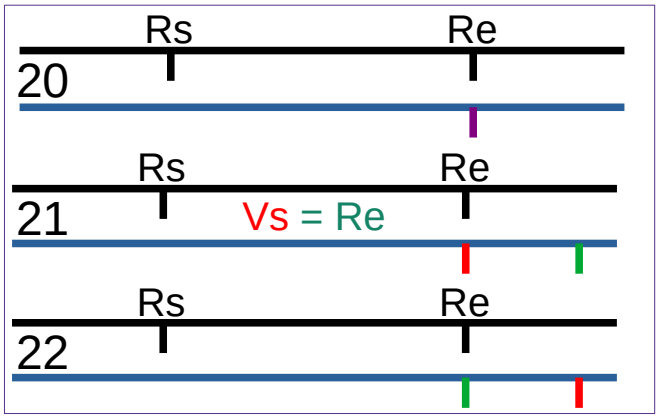


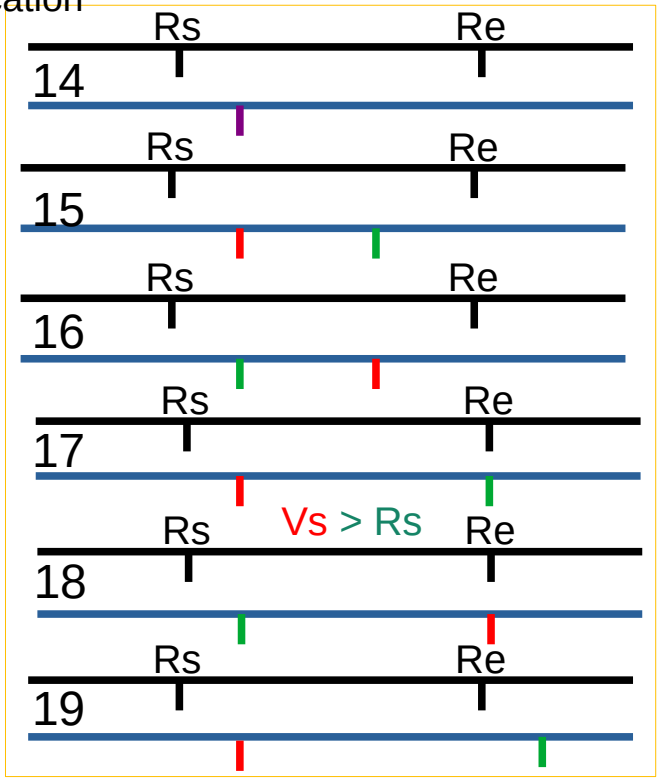

Numbering overlap types for third iteration
Second iteration numbering
Rs
Re
1
1
2
2
3
2a
4
3
5
3a
6
4
Vs < Rs
7
5
8
6
9
7
10
8
11
8a
12
9
Vs = Rs
13
10
20
15
Vs = Re
21
16
22
16a
14
11
15
12
16
12a
17
13
Vs > Rs
18
13a
19
14
23
17
Vs > Re
24
18
25
18a

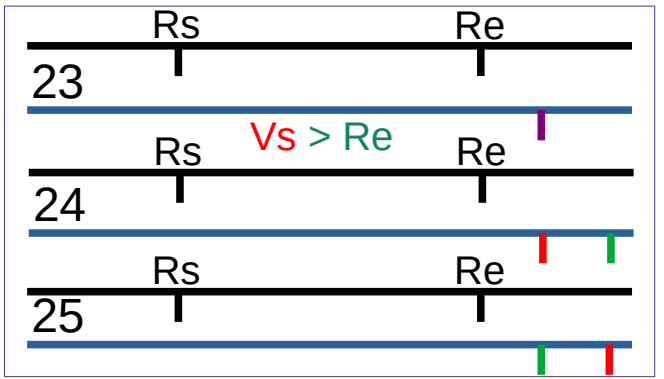
Numbering of overlap types implemented in the application











Annotated types

- The individual annotation of types is incomplete

Type 1

- Variant region is a point & precedes Reference
- No overlap

- FI: not defined ; SI: type 1 ; TI: type 1

Type 2

- Variant region precedes Reference
- No overlap

- FI: type 1; SI: type 2; TI: type 2

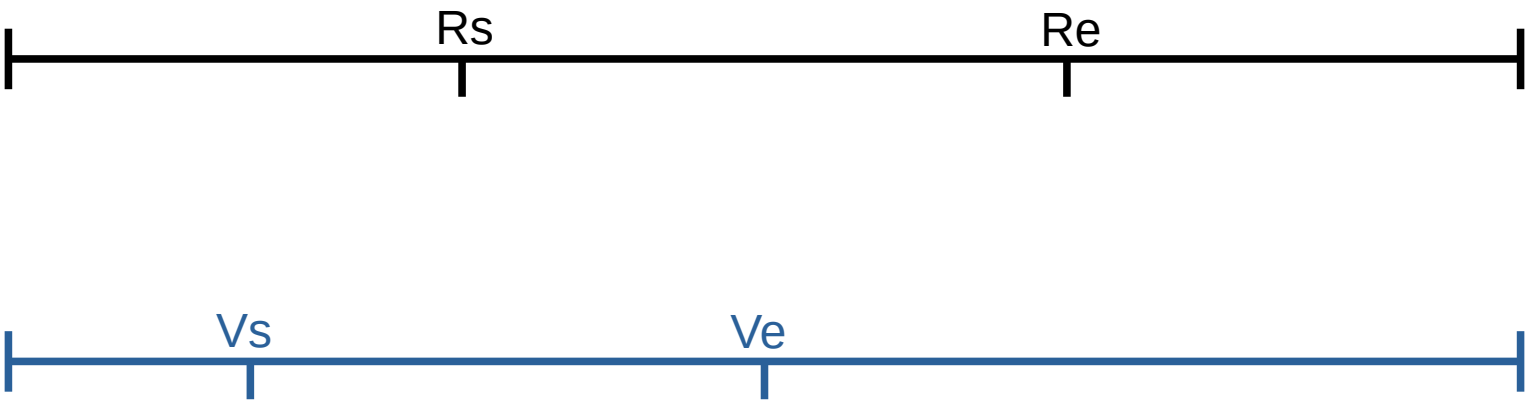
Type 4

- Variant concatenates start of Reference
  - (Vs < Rs) AND (Ve = Rs) AND (Re > Rs)

- FI: type 2; SI: type 3; TI: type 4

Type 6

- Variant begins before start of Reference, ends within Reference
  - (Vs < Rs) AND (Ve < Re)



  - Re > Rs not tested, but should be true when Ref feature is a skip (Re > Rs)
    - FI: type 5; SI: type 4; TI: type 6

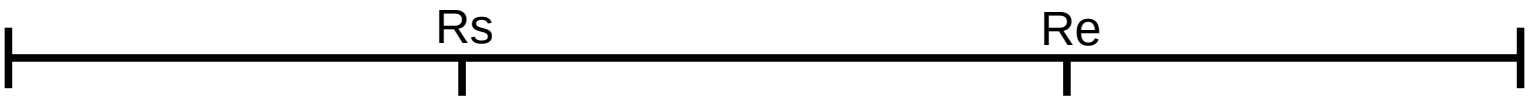
Type 7

- Variant begins before start of Reference
- Variant end coincides with Reference end
  - (Vs < Rs) AND (Ve = Re) AND (Re > Rs)

- FI: type 6; SI: type 5; TI: type 7

Type 8

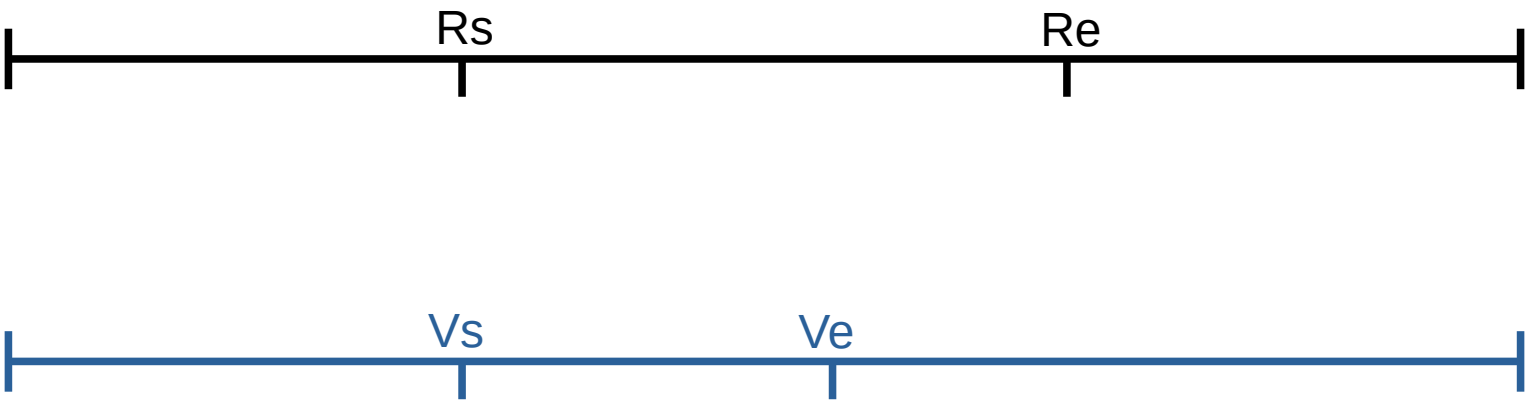
- Reference entirely contained within variant
  - (Vs < Rs) AND (Ve > Re)



  - Re > Rs not tested, but should be true when Ref feature is a skip (Re > Rs)
    - FI: type 4; SI: type 6; TI: type 8

Type 10

- Variant and Reference share start, ref longer than variant
  - (Rs = Vs) AND (Ve < Re)



- FI: type 15; SI: type 8; TI: type 10

Type 12

- Coincident feature positions
  - (Rs = Vs) AND (Re=Ve)

- FI: type 13; SI: type 9; TI: type 12

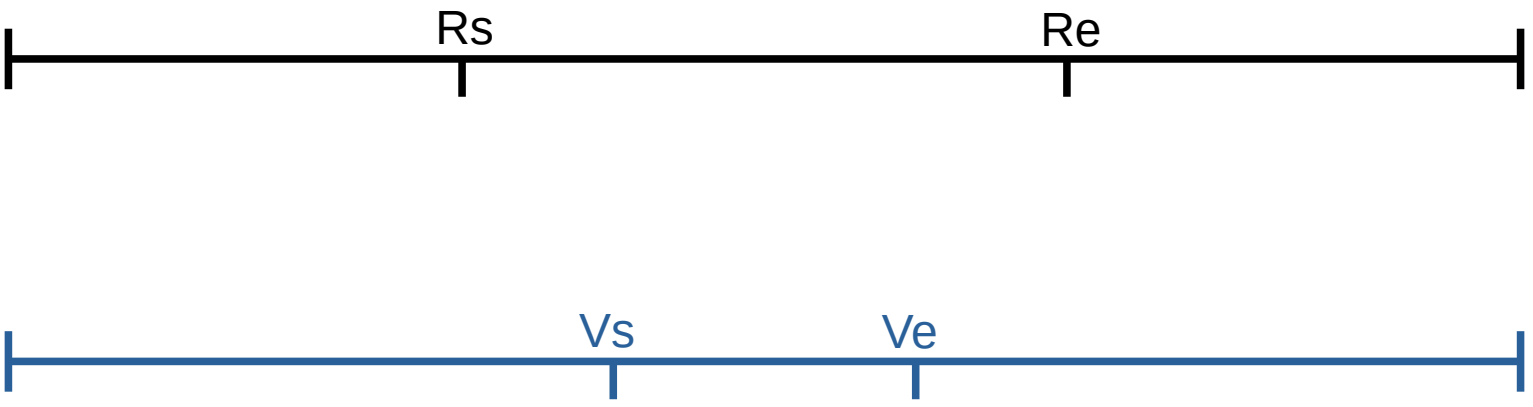
Type 13

- Variant and Reference share start, variant longer than ref
  - (Rs = Vs) AND (Ve > Re)

- FI: type 14; SI: type 10; TI: type 13

Type 15

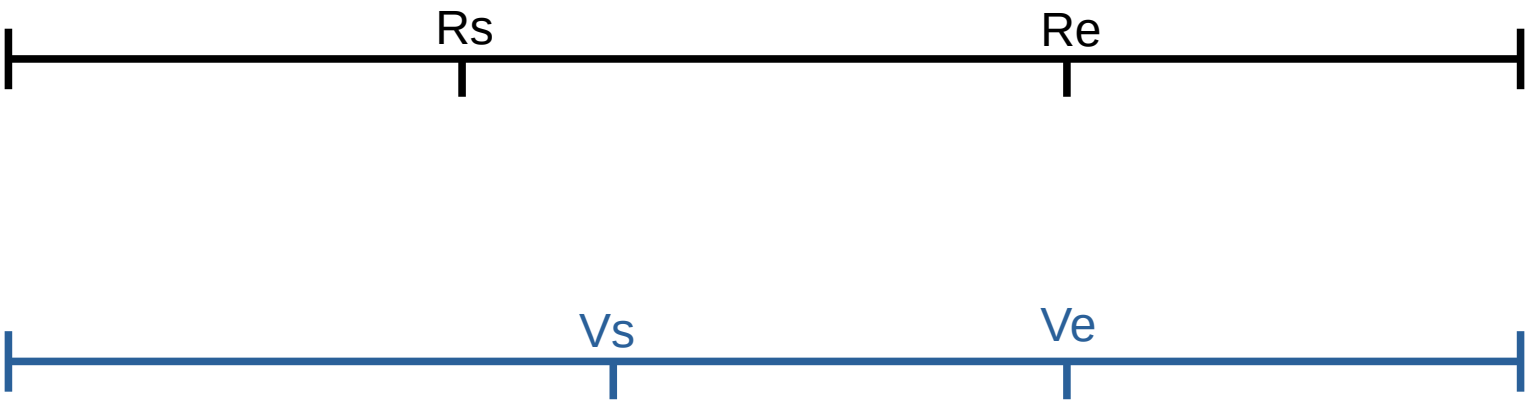
- The Variant range is entirely within the Reference range
  - (Re > Vs) AND (Ve < Re)



  - Re > Rs not tested, but should be true when Ref feature is a skip (Re > Rs)
    - FI: type 9; SI: type 12; TI: type 15

Type 17

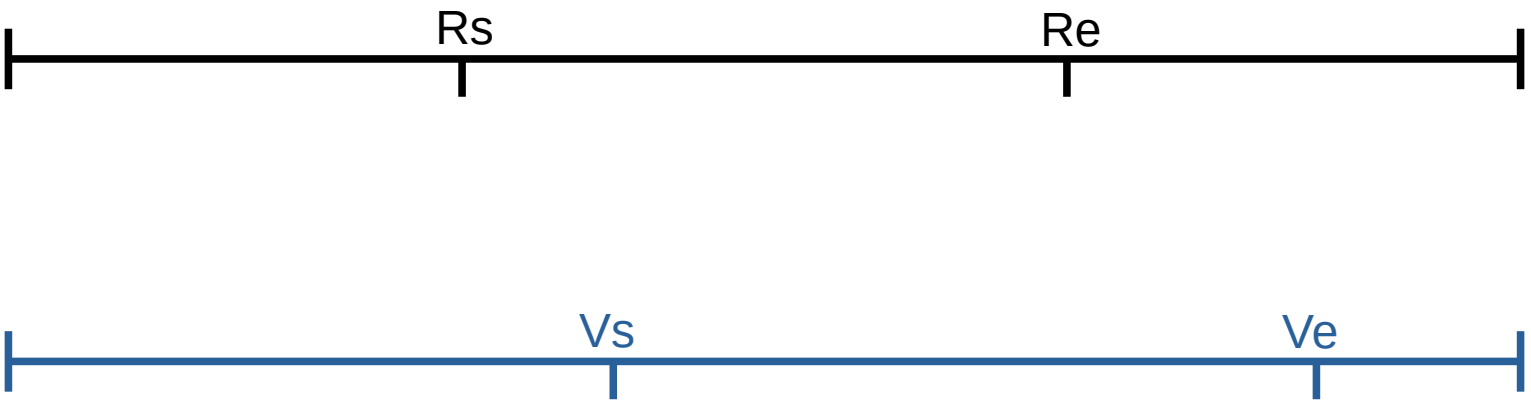
- Variant range entirely within reference, but ends coincide
  - (Re > Vs) AND (Ve = Re)



- FI: type 11; SI: type 13; TI: type 17

Type 19

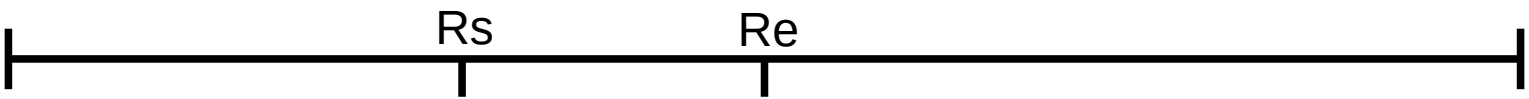
- Variant begins within Reference range but ends after Reference end
  - (Re > Vs) AND (Ve > Re)



  - Re > Rs not tested, but should be true when Ref feature is a skip (Re > Rs)
    - FI: type 10; SI: type 14; TI: type 19

Type 20

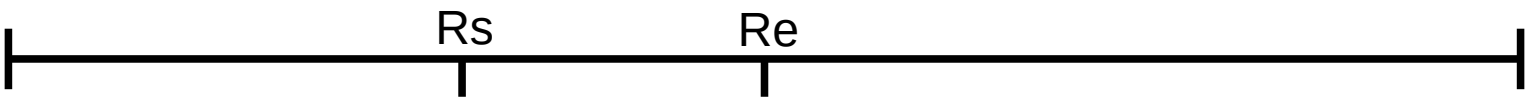
- Variant is a point coinciding with Reference end
  - (Vs > Rs) AND (Re = Vs) AND (Ve = Vs)



  - Re > Rs not tested, but should be true when Ref feature is a skip (Re > Rs)
    - FI: type 8; SI: type 15; TI: type 20

Type 21

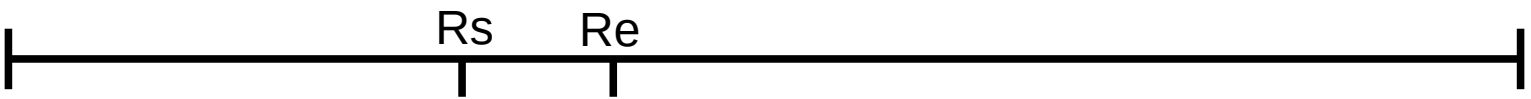
- Variant start coincides with Reference end
  - (Vs > Rs) AND (Re = Vs) AND (Ve > Vs)



  - Re > Rs not tested, but should be true when Ref feature is a skip (Re > Rs)
    - FI: type 7; SI: type 16; TI: type 21

Type 24

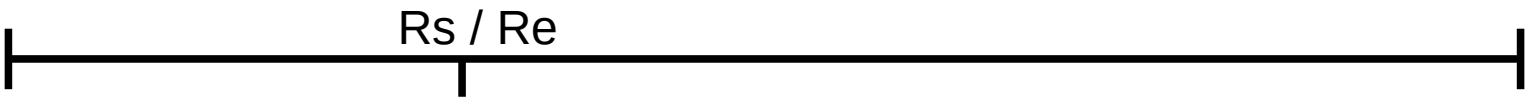
- Variant region after Reference
- No overlap
  - (Vs > Rs) AND (Re < Vs)



  - Re > Rs not tested, but should be true when Ref feature is a skip (Re > Rs)
    - FI: type 1A; SI: type 18; TI: type 24

Not defined (1)

- Variant concatenates start of Reference
- Reference is a point
  - (Vs < Rs) AND (Ve = Rs) AND (Re = Rs)



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

    - FI: type 3; SI and TI: not defined

Not defined (2)

- Coinciding points on both Variant and Reference
  - (Rs = Vs) AND (Vs = Re) AND (Vs=Ve)
  - 

  - Should not occur when Ref feature is a skip (Re > Rs)
    - FI: type 12; SI and TI: undefined