

1. „Variant Calling“

Eingabe: $n=4$ DNS-Sequenzen

Sequenz 0 als
Referenz

ref AAAACCCCGGGGTTTT

Seq0 AAAACCCCGGGGTTTT
Seq1 ATAACCCCGGGG
Seq2 ATAACCCCGGGGCG
Seq3 ATAACCCCGGGGCCTTTT

Einlesen der Sequenzen im
FASTA-Format

0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
ref AAAACCCCGGGGTTTT
Seq0 AAAACCCCGGGGTTTT

ref AAAACCCCGGGGTTTT
Seq1 ATAACCCCGGGG---- SNP, DEL

ref AAAACCCCGGGGTTTT
Seq2 ATAACCCCGGGGCG-- SNP, SV

ref AAAACCCCGGGG--TTTT
Seq3 ATAACCCCGGGGCCTTTT SNP, INS

Alignieren der Sequenzen
zur Referenz. Dabei werde
initiale seeds mithilfe eines q-
gram-Index gefunden und
anschließend durch *Sparse
Chaining* zu einem globalen
Alignment gekettet.

$e_{11} = (1, \text{SNP}, [0100], 'T', 1)$ $e_{12} = (12, \text{DEL}, [0100], '', 4)$
 $e_{21} = (1, \text{SNP}, [0010], 'T', 1)$ $e_{22} = (12, \text{SV}, [0010], 'CG', 4)$
 $e_{31} = (1, \text{SNP}, [0001], 'T', 1)$ $e_{32} = (12, \text{INS}, [0001], 'CC', 0)$

Generieren von *d-Events*
aus den Unterschieden zur
Referenzsequenz

2. Eventprocessing

$e_{11} = (1, \text{SNP}, [0111], 'T', 1)$ $e_{12} = (12, \text{DEL}, [0100], '', 4)$
 $e_{22} = (12, \text{SV}, [0010], 'CG', 4)$
 $e_{32} = (12, \text{INS}, [0001], 'CC', 0)$

Zusammenführen derselben
d-Events durch aktualisieren
des Sequenzvektors

Dependant region 1

Dependant region 2

$e_{11} = (1, \text{SNP}, [0111], 'T', 1)$ $e_{12} = (12, \text{DEL}, [0100], '', 4)$
 $e_{22} = (12, \text{SV}, [0010], 'CG', 4)$
 $e_{32} = (12, \text{INS}, [0001], 'CC', 0)$

Einteilen der d-Events in
abhängige Regionen
(dependent regions)

Dependant region 1

Dependant region 2

$e_{01} = (1, \text{SNP}, [1000], 'A', 1)$ $e_{02} = (12, \text{DEL}, [1000], '', 2)$
 $e_{11} = (12, \text{DEL}, [0100], '', 6)$
 $e_{21} = (12, \text{SV}, [0010], 'CG', 6)$

Änderung der Referenz je
nach berechnetem Score für
die einzelnen d-Events.
Der Referenzänderung
entsprechende Änderung der
d-Events.

0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
ref ATAACCCCGGGGCCTTTT
Seq0 AAACCCCGGGG--TTTT SNP, DEL

ref ATAACCCCGGGGCCTTTT
Seq1 ATAACCCCGGGG----- DEL

ref ATAACCCCGGGGCCTTTT
Seq2 ATAACCCCGGGGCG---- SV

ref ATAACCCCGGGGCCTTTT
Seq3 ATAACCCCGGGGCCTTTT