

Genome 540 Discussion

Feb. 17th 2026



HW6 Intro

Overview

1. Read in the mapped read count file.
2. Using the provided scoring file, scan the length of the segment, keeping track of wherever there are high scoring D-segments.
3. Look up the 3 highest scoring segments, and what they correspond to.

Input files:

Read count file:

chromosome (should stay const)

position

of reads

Score file:

scores, D and S

```
0 -0.3464
1 0.2488
2 0.8439
>=3 1.5337
D -33.219
S 33.219
```

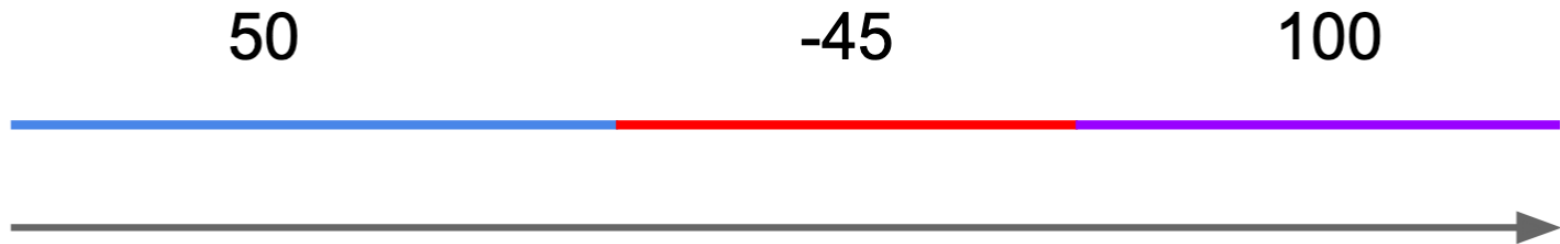
17	1	0
17	2	0
17	3	0
17	4	0
17	5	2
17	6	0
17	7	0
17	8	0
17	9	0
17	10	0
17	11	0
17	12	1
17	13	0
17	14	3
17	15	0
17	16	0
17	17	2
17	18	0
17	19	0
17	20	0
17	21	5
17	22	0
17	23	1
17	24	0
17	25	0
17	26	3

D-segments

Goal:

Find all high-scoring regions of the genome,
according to the given scoring scheme.

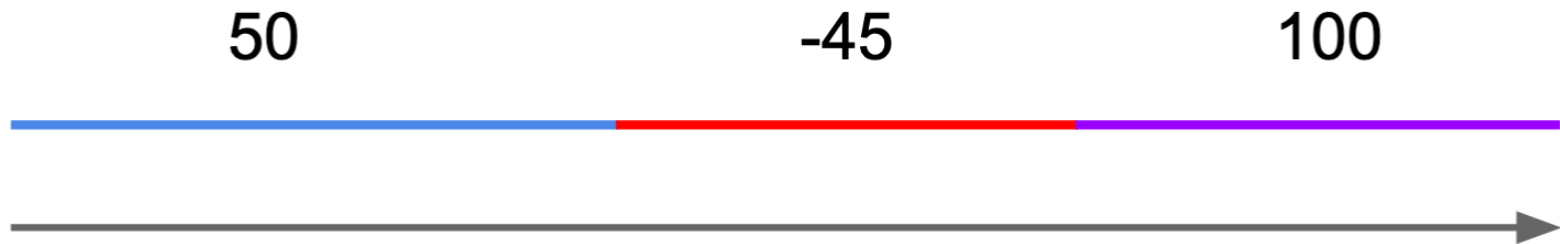
D-segments



Problem:

Maximally-scoring segments may have discontinuous segments of biological significance

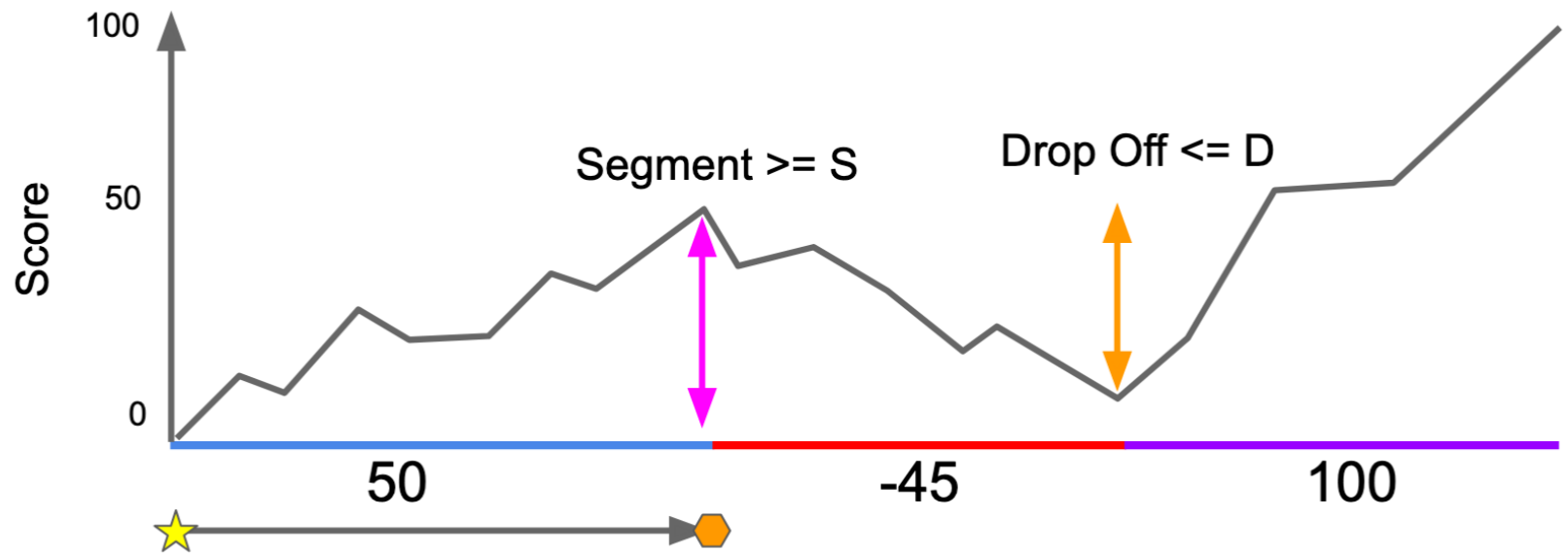
D-segments



Problem:

Maximally-scoring segments may have discontinuous segments of biological significance

D-segments



D-segments

$O(N)$ algorithm to find all maximal D-segs:

```
cumul = max = 0; start = 1;
for (i = 1; i ≤ N; i++) {
    cumul += s[i];
    if (cumul ≥ max)
        {max = cumul; end = i;}
    if (cumul ≤ 0 or cumul ≤ max + D or i == N) {
        if (max ≥ S)
            {print start, end, max; }
        max = cumul = 0; start = end = i + 1; /* NO BACKTRACKING
        NEEDED! */
    }
}
```

Example!

Position	1	2	3	4	5	6	7	8
Read Start Counts	0	1	2	3	0	2	0	0

Scoring Scheme:

0: -0.75

1: 0.5 D: -3

2: 1.0 S: 3

3: 1.25

$O(N)$ algorithm to find all maximal D-segs:

```
cumul = max = 0; start = 1;
for (i = 1; i ≤ N; i++) {
    cumul += s[i];
    if (cumul ≥ max)
        {max = cumul; end = i;}
    if (cumul ≤ 0 or cumul ≤ max + D or i == N) {
        if (max ≥ S)
            {print start, end, max; }
        max = cumul = 0; start = end = i + 1; /* NO BACKTRACKING
        NEEDED! */
    }
}
```