Genome 540 Discussion

February 8th, 2024

Clifford Rostomily

Genome Sciences
UNIVERSITY OF WASHINGTON

Assignment 5 Questions?

- Part 1
  - Build a weighted edit graph for 3 amino acid sequences of the insulin protein (human, frog, water buffalo) using the BLOSUM62 scoring matrix and save it as a text file
- Part 2:
  - Use your program from HW4 to find the max weight path through the edit graph

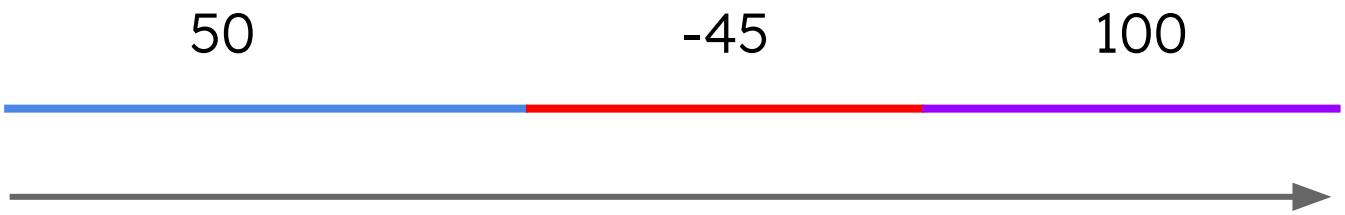
Assignment 6

Overview

- Write a program to identify regions of elevated copy-number using the D-segment algorithm
- Run the program on chromosome 16 from individual CHM13

D-segment motivation

D-segment motivation

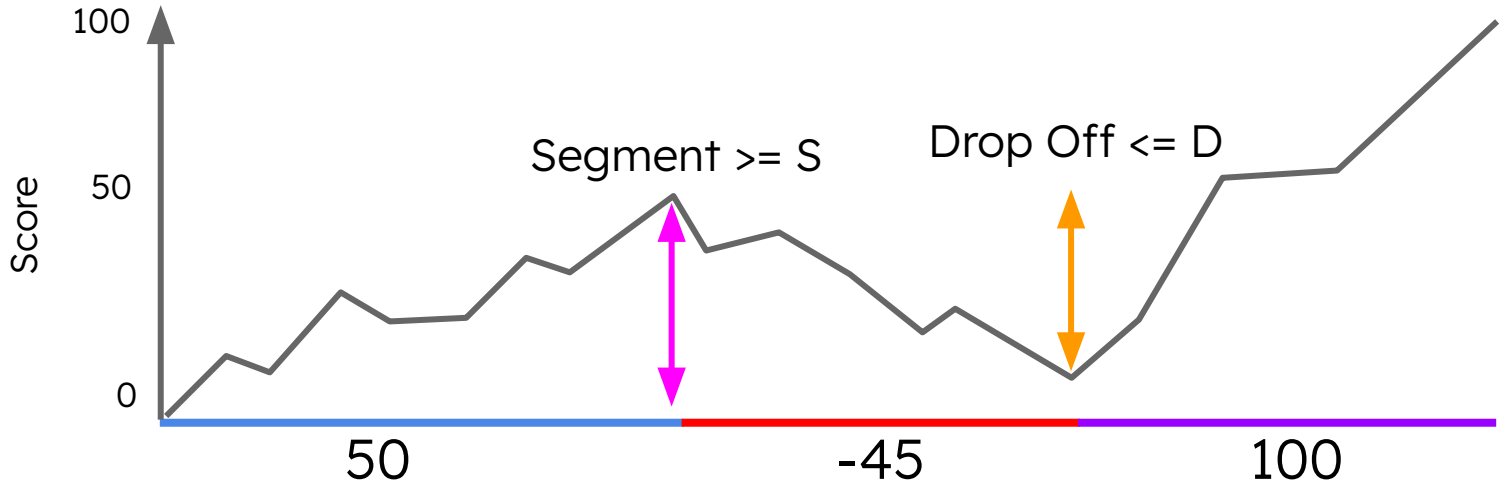


Whole region has a score of 105.

D-segment motivation

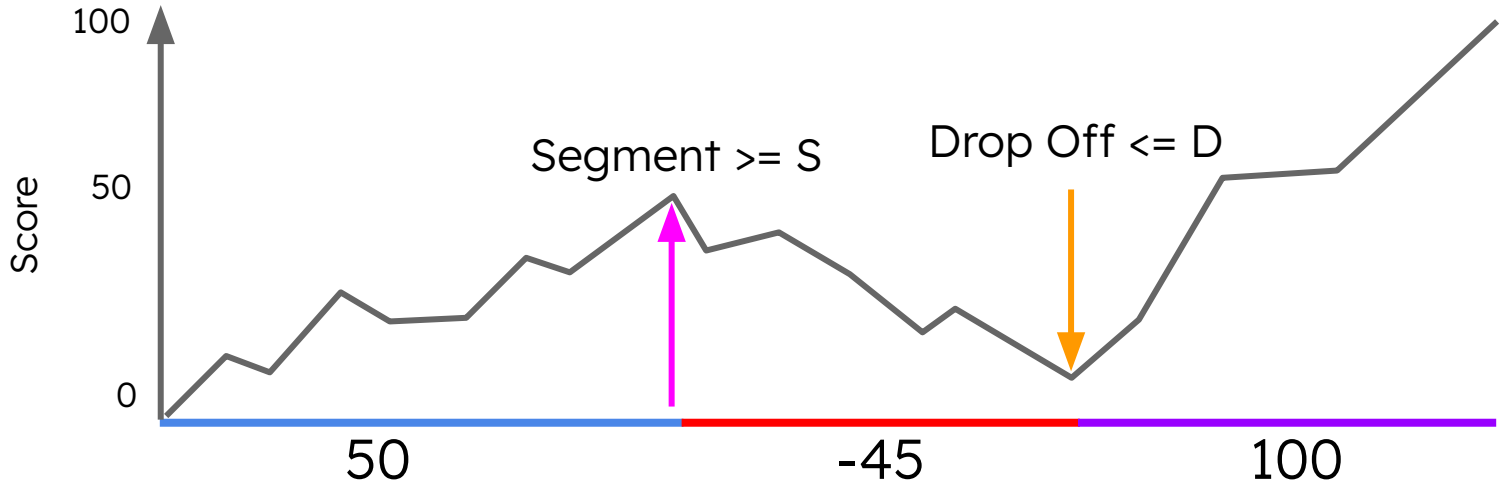
However, these two sub-segments may represent biologically distinct events...

D-segment algorithm



What values of S and D would separate these segments?

D-segment algorithm

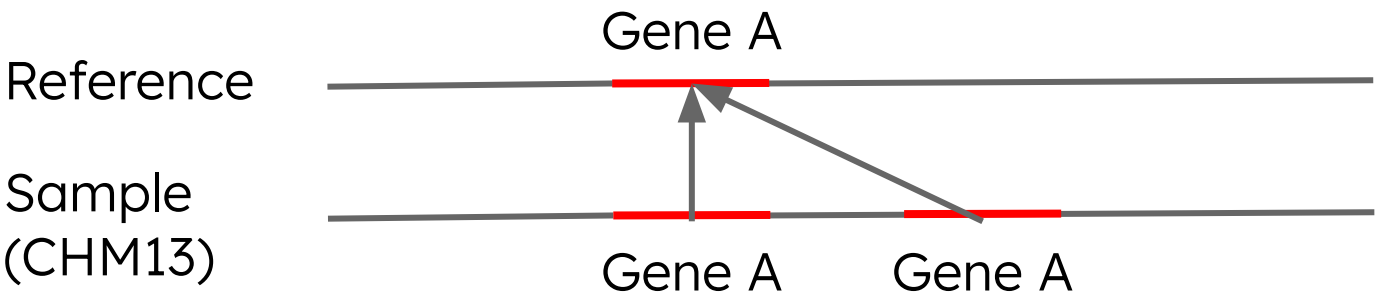


What values of S and D would separate these segments?

S <= 50 and D <= -45

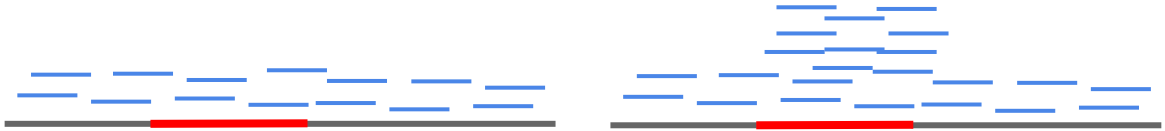
*** D would probably have to be less than -10 as well

Copy Number Variation



Both copies mapped to same reference allele

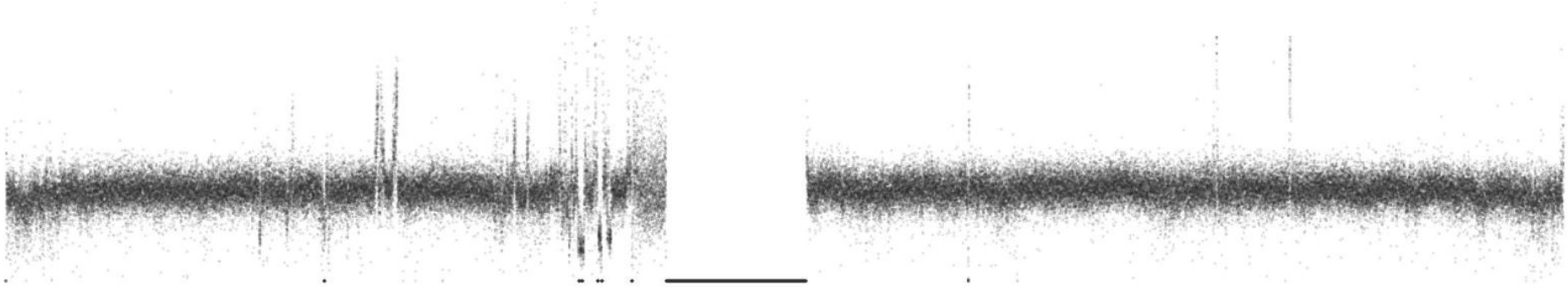
NGS read coverage will be higher for that gene when mapped to reference



Same as ref.

Duplication

Data - Read Start Counts



Position
(chr16)

Convert Counts to Scores

- Background:
  - m = mean(counts starts)
  - count = counts at a position
  - B ~ Poisson(m)
  - L(B|count) = P(count| B)
- Heterozygous duplication:
  - D ~ Poisson(1.5*m)
  - L(D|count) = P(count| D)
- Score
  - Score = log2(LR(L(D|count)/L(B|count)))

Pseudocode

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! */
    }
}
```

Reminders

- HW5 due this Sunday, 11:59pm
- Please have your name in the filename of your homework assignment and match the template