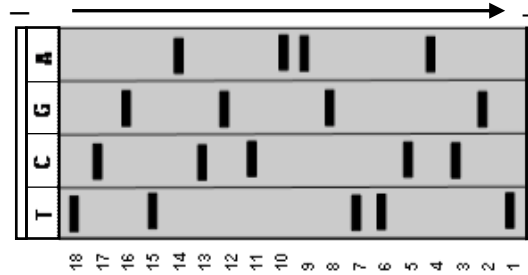


# Introduction to Computational and Systems Biology

## Exercise Set 1

Due: November 28, 2019

1. What is the DNA sequence corresponding to the Sanger plot below?



2. Given two sequences, which value is larger: their local similarity or their global similarity? Why? How does their semiglobal similarity compare with the other two values?
3. Find all best global alignments between sequences AAAC and AGC, where the scoring scheme is +1 for match, -1 for mismatch and -2 for an alignment with a gap.
4. Find all best global alignments between sequences ATAG and TTCG, where the scoring scheme is +1 for match, -1 for mismatch and -1 for an alignment with a gap.
5. Find all best local alignments between sequences ATACTGGG and TGACTGAG, using the same scoring scheme as in exercise 3.
6. What scoring schemes should you use to determine the longest common substring and the longest common subsequence for two given strings, using the algorithm for best global alignment?

**Note/definition:** Let  $u = a_1a_2 \dots a_n$  be a string of length  $n$ . We say that string  $v$  is a *substring* of  $u$  if  $v = a_i a_{i+1} \dots a_j$ , for some  $1 \leq i \leq j \leq n$ . We say that string  $w$  is a *subsequence* of  $u$  if  $w = a_{k_1} a_{k_2} \dots a_{k_l}$ , for some  $1 \leq k_1 \leq k_2 \leq \dots \leq k_l \leq n$ . In other words, a *substring* is a *consecutive sequence of letters* from the original string, whereas a *subsequence* is a *sequence of letters* from the original string, *not necessarily consecutive, but certainly in the order they occurred* in the original string.

7. Apply the scoring scheme you indicated in exercise 7 to find all longest common substrings for strings ATACTGGG and TGACTGGT.