

# COMP90016 – Group Assignment 1

## Naive Reads Alignment

Haonan Li, Xiaojing Li

March 19, 2018

### 1 Introduction

This task analysis simulated sequencing data from a short reference genome.

### 2 Experiments & Results

In this task, we use naive alignment method to align reads to the given reference sequence. The “naive” here means we only consider perfect matches to the either forward or reverse strand of the reference.

Table 1 shows the alignment statistics results. Specifically, how many reads were aligned to 0 positions and exactly N position (in absolute numbers and percent).

N-pos	0	1	2	3	4	5	6	7	8	9	10	11
Reads	23	91	52	24	16	7	1	1	2	0	0	1

Table 1: Alignment Statistics.

### 3 Discussion

Experiment results reflect there are 10.55% reads not aligning to the reference. Theoretically, Reads in highly polymorphic regions, repetitive regions, or reads with indels may not be aligned.

Reads length	5	6	7
Pieces num	6	204	8
Not Aligned	0	17	6

Table 2: Reads length related Results

But in particular work, we make Table 2 to analyze the not aligned reads related problem. In this Table, we find that 75% reads with length 7 were not aligned. Moreover, we also notice that the particular reads indicates in the last column of Table 1 whose length is 5. Based on these results, we think reads with long length tends is more likely fail to be aligned, these reads should be

created by insertion. At the same time, the short reads, which may be created by deletion, are less specific than the 6bp long fragments so they may have more alignment positions.