

## **BMEG3105 Scribing Lec 5**

20/9/2024

Zhai Xueer

1155174880

### **How to use the Scoring matrix to find the optimum score and optimum alignment:**

1. Fill in the numbers in each cell. Each cell has three possible paths to go to there, calculate the match, mismatch, and gap from each case and fill in the score that is the highest among them.
2. After filling in all the numbers, the score on the right bottom square should be the optimum alignment score. Trace back to see how the score is calculated, track the route, and match the letters.

### **Advantages of dynamic programming:**

Enumeration method:  $n=300:7*10^{88}$

Dynamic Programming Method:  $n=300:900$

Much less computation !!!!

### **There are tools that can help you do sequence alignment:**

Webserver and Biopython

### **Genome Sequencing:**

- only 0.001% of the human genome is different.
- We also need to know the gene expression (not just gene sequence) to account for the phenotype difference.

### **How to do Gene Expression Matrix:**

- Collect Gene Expression Data:
  - Acquire gene expression data, which can be obtained from various sources, such as:
    - RNA-sequencing (RNA-seq) experiments
    - Illumina Sequencing
- Preprocess the Data:
  - Perform necessary preprocessing steps, which may include:
    - Quality control and filtering of the raw data
    - Normalization to account for technical biases and variations
    - Transformation of the data (e.g., log-transformation, scaling)
    - Handling missing values
- Construct the Gene Expression Matrix:
  - Organize the preprocessed gene expression data into a matrix format, where:
    - Rows represent the genes (or transcripts/probes)
    - Columns represent the samples (e.g., different conditions, time points, or individuals)
    - The matrix elements represent the expression levels of the genes in each sample

