

# 바이오인포매틱스의 이해

실습 4

# 시작하기 전에

- 프로그램 언어는 (C/C++, Java, python) 사용 가능.
- 프로그램 작성하는 문제는 프로그램도 함께 제출하여야 함.
- 코드를 짤때 코드의 맨위에 자신의 학번 이름을 주석으로 적으시오.
- 코드 문제는 꼭 수업시간에 주어진 그대로 짤 필요 없습니다. 같은 알고리즘도 여러가지 방식으로, 때로는 더 효율적으로 구현도 가능하니 그 알고리즘내에서 자기가 원하는 방식으로 짜면 됩니다.
- 코드를 구현하는 문제는 반드시 결과화면을 첨부해야 합니다. 결과물을 제출할때 화면 캡처를 사용할 것. 이는 자신의 코드를 돌렸을때 나온 결과임을 보이기 위함으로 사용하는 언어나 에디터 등등에 따라 다를 수 있으므로 방법은 알아서 제출할 것. 어떤 방식이든 자신의 코드를 돌려서 나온 결과라는 것만 보여주면 됨. 예를들어, 자바 이클립스를 사용하면 이때 자신의 코드의 윗부분(학번 이름과 앞에 코드 5줄정도 포함)이 아래 실행 결과와 같이 캡처되도록 하시오.

# 시작하기 전에

## 1. 과제 제출 시 보고서를 반드시 제출해주셔야 합니다.

· 스크린샷, 각 문제에 대한 내용등을 보고서로 안내고 따로 제출하시는 분들이 있습니다. 반드시 보고서를 작성해서 제출해주세요. 미제출시 0점입니다.

## 2. 실행화면 제출시에 옆에 코드가 나오게 찍어주셔야 됩니다. 코드 가리시면 안됩니다.

· 실행화면 + 자신의 코드가 한 화면에 나오게 해주셔야합니다.

## 3. 모든 보고서에는 표지를 만들어주셔야 합니다.

· 인터넷에 동국대학교 레포트표지 라고 치면 많은 표지들이 나옵니다. 그 중 마음에 드는 것을 골라 자신의 레포트 표지로 이용해보세요.

## 4. 보고서에 실행화면을 넣어주실 때 어떤 문제에 대한 실행화면인지 간단한 설명 혹은 캡션을 달아주세요.

## 5. 서술형 문제가 나갈 경우 자세히 답해주셔야 합니다.

· 자신의 이해를 바탕으로 서술을 해야 하며 가능한 경우 예제를 사용해 설명해주세요. 서술형 문제에서 1~3줄사이의 간단한 답변은 오답으로 처리할 예정입니다.

## 6. 코드를 보고서에 캡처 혹은 복사해 넣으셔도 코드 소스파일은 반드시 제출하셔야 합니다.

· 미제출시 0점입니다.

## 7. 보고서 + 소스코드를 반드시 압축해서 제출해주세요.

## 8. 보고서에는 자신이 작성한 코드 설명이 들어가야 합니다.

· 주석으로 설명을 대체한 경우에는 주석을 캡처해서 보고서에 첨부해주세요.

# Short Read Matching

1. ATGC로 이루어진 길의 N의 무작위 염기서열을 생성하는 함수 makeRefDNA를 만드세요.
2. makeRefDNA에서 생성된 염기서열 RefDNA를 이용해서 myDNA를 만드는 함수 makeMyDNA를 만드세요. myDNA와 RefDNA는 10% 염기서열 차이를 보입니다. [차이를 보이는 서열들은 가급적 한곳에 연속적으로 위치하지 않고 균등하게 분포되도록 만들어 주셔야 합니다.]
3. 만들어진 myDNA를 이용해 길이 K의 Short Read를 P개를 만들어주는 함수 makeSRead를 작성하세요. [short read는 myDNA에서 순차적으로 자르는것이 아닌 매 위치를 랜덤으로 잡아서 길이 K만큼 잘라내는 방법으로 만들어야 합니다.]
4. M개의 MisMatch를 허용하면서 RefDNA와 ShortRead를 이용해 서열을 복원하고 복원된 서열과 myDNA가 얼마나 일치하는지를 보여주는 함수 MatchDNA를 작성하세요. [Match 방법은 trivial matching]
5.  $N, K, P = \{50\text{만}, 30, 2\text{만}\}, \{50\text{만}, 30, 4\text{만}\}, \{100\text{만}, 60, 2\text{만}\}, \{100\text{만}, 60, 3\text{만}\}$ , 등 추가로 여러가지 예제를 진행해보고 결과에 대한 분석을 보고서에 작성하세요. [어느 요인때문에 이러한 결과가 나왔는지에 대한 자신의 생각을 자유롭게 적어주시면 됩니다.]