

Central dogma와 전사 과정: https://www.youtube.com/watch?v=Ey_DNd1hEBQ

- DNA를 전사할 때, 특정 한 가닥만 전사하면 단백질 합성을 위한 정보를 일부 사용할 못 한다?

DNA는 상보적인 염기서열을 가지고 있기 때문에 한 가닥만 알아도 나머지도 알 수 있다.

- DNA를 전사하면, 항상 RNA는 mRNA만 생긴다?

DNA의 프로모터 부분에 따라 mRNA, tRNA 등 다양한 RNA가 생성된다.

- 생성된 mRNA의 염기서열은 DNA의 한 가닥에 있는 염기서열을 그대로 복제해온다?

ACGT → ACGU로 바뀌고 나머지는 그대로 복제

RNA-seq원리: <https://www.youtube.com/watch?v=-8BEOoMwNX8>

- 2:10 왜 U가 없는지?

서열을 읽을 때는 mRNA가 아니라 cDNA의 fragment들을 읽는 것이기 때문

- 2:10 1개의 DNA를 전사하면 1개의 mRNA가 나온다. RNA-seq에서 여러 RNA의 염기서열을 읽어낸다 했는데 이 RNA는 여러 DNA에서 전사된 mRNA인지 아니면 한 DNA에서 여러 번 프로모터를 바꿔가면서 전사한 mRNA인지?

둘 다 아님. 그냥 fragmentation에서 나온 여러 fragment들 각 각을 RNA로 명시해둔 것.

- 2:43 하나의 유전자는 alternative splicing(>)으로 인해 여러 전사체를 가질 수 있는데, 어떤 exon이 제외된 상태에서 전사되는가에 따라 1개의 유전자에서 수십개의 전사체가 나올 수 있다면, RNA-seq에서 추후 분석을 위해서 어떤 RNA의 염기서열을 읽어야 하는지?

일반적으로 전사 될 때는 DNA에 있는 Exon들 중 몇 개로만 mRNA가 구성되고 이것을 alternative splicing이라 한다. 항상 전사할 때마다 mRNA를 구성하는 exon의 종류가 다르기 때문에 여러 번 반복을 거쳐서 가장 자주 발현되는 유전자를 찾기 위해서 전사 전 DNA를 PCR이라는 방법을 통해서 증폭시킨다. 그래서 (유전자 단위인) DNA가 어떤 exon을 포함해서 (전사 체 단위인) mRNA를 만드는지에 따라 여러 전사체를 가질 수 있는 것이다.

즉, 증폭된 여러 DNA를 가지고 모두 전사시킨 뒤 cDNA를 만들고 fragmentation을 한 뒤 reference genome에 mapping 할 때 나온 모든 전사 체를 사용해서 일반적으로 어느 유전자 발현 량이 높은 지 결정하게 된다.

- 3:00 DNA가 전사될 때, 유전자의 시작과 끝을 의미하는 Exon만 남고 유전자의 정보가

들어있는 intron은 왜 전부 없어지는지?

Intron은 전사되지 않기 때문에 exon만 남는다.

- **3:20 역 전사했을 때, 생긴 cDNA의 두 가닥은 모두 Exon으로 구성되는지? 굳이 전사하고 다시 역 전사하는 이유?**

한 가닥은 exon으로 구성된 mRNA이고 나머지 가닥은 mRNA 염기서열과 상보적인 염기서열을 가지는 가닥이다.

- **4:40 역 전사된 cDNA를 잘라서 fragment를 만들었고, paired read를 했을 때 왜 mRNA에서 바로 fragmentation 하지 않고 cDNA를 만든 뒤 fragmentation 하는 이유?**

mRNA는 한 가닥으로 구성되어서 불안정하기 때문에 cDNA를 만들어서 진행한다.

- **4:55 Reference genome은 모두 Exon으로만 구성되어 있는지?**

RNA-seq 순서

■ download samples and reference genome files

Ensembl에서 다운 받은 자료는 read 데이터

■ Trimming (Trimmomatic)

Fragmentation된 fragment들을 paired end read한 데이터에서 양 끝 어댑터 혹은 퀄리티가 낮은 데이터를 제거하기 위한 단계

■ Quality control (FastQC)

Trim된 read의 퀄리티를 확인하기 위한 단계

■ Indexing

순서 상관없이 분해된 fragment = read 들을 mapping하기 전에 한 read와 동일한 염기서열을 reference genome에서 일일이 찾는 것보다 인덱싱해둔 reference genome을 참고했을 때 더 빠르게 찾을 수 있기 때문에 인덱싱을 만들어 두는 단계(사전처럼)

■ Alignment (HISAT2, STAR)

Reference genome에 각 reads 들을 매핑하는 단계

■ Sam to Bam to sorted Bam(samtools)

Mapping을 통해 얻은 sam 파일을 정렬한 뒤, 용량이 너무 크기 때문에 binary 형식의 bam파일로 변환하는 단계

■ Assembly (stringtie)

bam 파일에서 exon, intron 등의 gene 구조를 확인하기 위한 단계

■ Quantification (featureCounts)

Mapping된 bam파일에서 유전자 정보를 담은 GTF annotation files를 통해 각 유전자의 발현량을 수치화 하는 단계