

04

유전체 분석 기초 II

권기상 교수

원광보건대학교 임상병리과

05

NGS (Next Generation Sequencing) II

1. Whole genome shotgun sequencing
2. Next-Generation Sequencing



■ 05. NGS(Next Generation Sequencing) II

1 Whole genome shotgun sequencing

- ◆ 생어 염기서열 분석법
- ◆ 분석절차

- 1 DNA 단편화: 원하는 DNA분자를 일정한 길이로 단편화시킴
- 2 클로닝: 단편화된 DNA 조각을 plasmid 와 같은 vector에 삽입하고, 클로닝하여 library 만듦
- 3 염기서열분석: 만들어진 library 이용하여 염기서열 분석 – 생어법
- 4 Assembly: 알아낸 DNA 조각의 염기서열을 컴퓨터를 이용해 정렬
- 5 겹치는 부분 연결
- 6 반복하여 최대한 적은 단편수로 만듦

■ 05. NGS(Next Generation Sequencing) II

2 Next-Generation Sequencing(NGS; 차세대 염기서열 분석법)

- ◆ 대량으로 한꺼번에 유전체의 염기 서열 정보를 얻는 방법
(Massive parallel sequencing)
- ◆ 하나의 유전체를 작게 잘라 많은 조각으로 만든 뒤, 각 조각의 염기 서열을 읽은 데이터를 생성하여 이를 해독함

■ 05. NGS(Next Generation Sequencing) II

2 Next-Generation Sequencing(NGS; 차세대 염기서열 분석법)

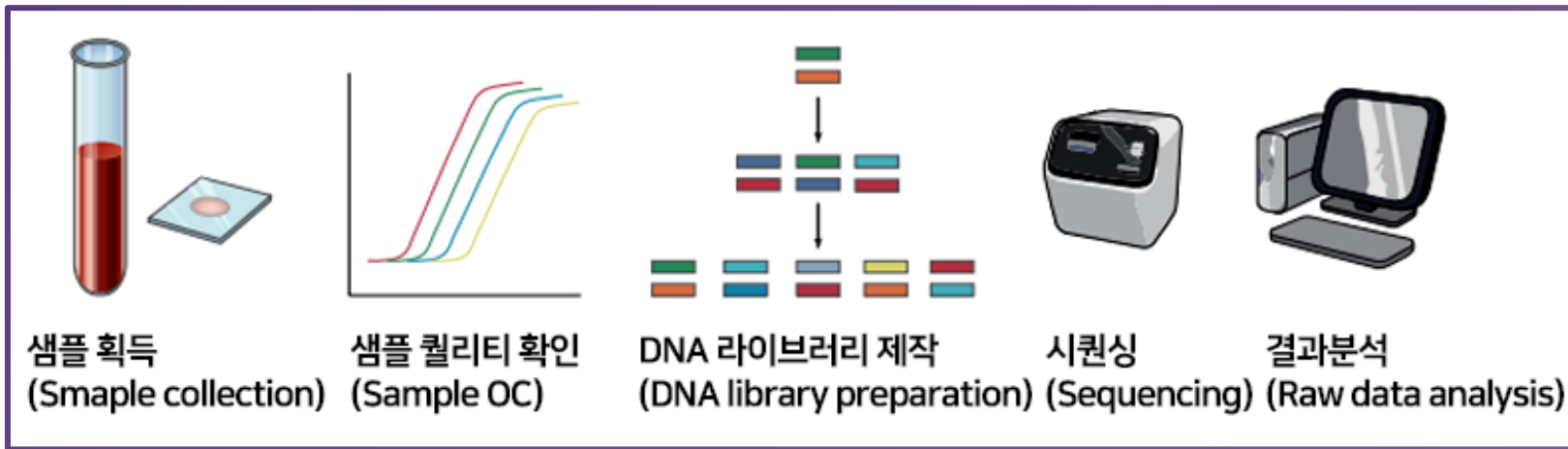
◆ NGS의 과정

- 샘플 준비: 조직, 혈액 등에서 genomic DNA 또는 RNA 추출
- 라이브러리 제작: DNA를 조각으로 자르고, 라이브러리를 제작한 뒤, 양과 질을 확인함
- 클러스터 제작: 주형을 증폭시키는
 - 오일과 물의 상질을 이용한 에멀전을 이용한 PCR법(emPCR)
 - 다리 형태로 증폭되는 PCR법(Bridge amplification)
- 시퀀싱 : 증폭된 주형의 염기 서열을 분석
 - probe의 연결에 의한 시퀀싱(sequencing-by-ligation)
 - 염기가 주형에 결합하면서 시퀀싱(sequencing-by-synthesis)
- 데이터 분석

05. NGS(Next Generation Sequencing) II

2 Next-Generation Sequencing(NGS; 차세대 염기서열 분석법)

◆ NGS의 과정



■ 05. NGS(Next Generation Sequencing) II

2 Next-Generation Sequencing(NGS; 차세대 염기서열 분석법)

- ◆ 라이브러리 제작
 - 분절화(fragmentation)
 - 말단수리(end-repair), 3'-말단 아데닌 추가(A-tailing)
 - 어댑터 부착(adapter ligation)
 - 라이브러리 증폭(library amplification)

■ 05. NGS(Next Generation Sequencing) II

2 Next-Generation Sequencing(NGS; 차세대 염기서열 분석법)

- ◆ 타겟 선별
 - 추가 작업 없이 NGS 분석을 하여
 - 모든 DNA의 데이터를 얻는 것을 whole genome sequencing(WGS), 모든 RNA의 정보를 분석하면 whole RNA sequencing
 - 타겟 패널 시퀀싱(targeted sequencing): 원하는 유전자 부위만 보고자 한다면 분석하고자 하는 부분의 DNA 혹은 RNA를 선별해야 하며 이것을 타겟 선별하여 분석
 - PCR 프라이머(primer)로 증폭을 하는 앰플리콘(amplicon)방법, 프로브(probe)를 이용하여 교합(hybridization)하는 캡처(capture) 방법

■ 05. NGS(Next Generation Sequencing) II

2 Next-Generation Sequencing(NGS; 차세대 염기서열 분석법)

◆ 데이터 정도관리

- 기술적 한계와 실험적 원인에 의한 다양한 오류(error)의 가능성이 있다. NGS 염기서열분석 결과의 원데이터(raw data)에서는 추정 오류 확률을 수치로 나타내며,
 - Phred 점수가 각 염기의 품질을 나타내는 지표로 활용
 - 시퀀싱 리드(read)의 염기서열과 Phred 점수를 같이 표시한 것을 FASTQ 파일
 - FASTQ 파일은 시퀀스 고유ID, 각 리드의 시퀀스, "+" 기호, 시퀀스의 각 베이스 별 Phred 점수로 된 네 개의 줄로 구성되어 있다.

■ 05. NGS(Next Generation Sequencing) II

2 Next-Generation Sequencing(NGS; 차세대 염기서열 분석법)

- ◆ 데이터 정도관리를 위한 프로그램
 - FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>)
 - FastQC 프로그램은 NGS 결과 생성된 FASTQ, SAM 또는 BAM 파일을 입력하여 시퀀싱 품질을 평가하는 기능을 제공하는 소프트웨어로 여러 품질 지표를 그래프로 보여줌
 - 일반적으로 NGS 기술의 한계로 시퀀싱을 계속하다 보면 시퀀싱 리드(read) 마지막 부분으로 갈수록 시퀀싱 품질 지표가 떨어짐.
 - 품질이 우수한 NGS 결과는 시퀀싱 리드(read)의 위치에 상관없이 A, T, G, C의 비율이 일정하게 유지되는 반면, 품질이 좋지 않은 NGS 결과는 시퀀싱 리드의 위치에 따라 A, T, G, C의 비율이 변함

■ 05. NGS(Next Generation Sequencing) II

2 Next-Generation Sequencing(NGS; 차세대 염기서열 분석법)

- ◆ 데이터 정도관리를 위한 프로그램
 - NGS QC Toolkit (<http://14.139.61.3:8080/ngsqctoolkit/>)
 - NGS 초기 데이터의 품질 정보
 - 다양한 시퀀싱 품질 지표를 그래프로 보여줌
 - 시퀀싱 리드 끝 부위의 품질이 떨어지는 것을 확인하게 되면 각 read의 끝부분을 일괄적으로 제거를 하여 품질이 우수한 부위만을 분석에 활용

■ 05. NGS(Next Generation Sequencing) II

2 Next-Generation Sequencing(NGS; 차세대 염기서열 분석법)

- ◆ 데이터 정도관리를 위한 프로그램
 - Trimmomatic (<https://github.com/timflutre/trimmomatic>)
 - FastQC를 통해 시퀀싱 리드의 끝 부위의 품질이 떨어지는 것을 확인
 - 각 시퀀싱 리드의 끝 부분을 일괄적으로 제거(trimming)하여 품질이 우수한 부위만을 분석에 활용
 - 이 때 사용할 수 있는 프로그램이 Trimmomatic이며, paired end 또는 single end 시퀀싱 결과인 FASTQ 파일을 입력으로 받게 됨

■ 05. NGS(Next Generation Sequencing) II

2 Next-Generation Sequencing(NGS; 차세대 염기서열 분석법)

- ◆ 매핑(Mapping) / 정렬(alignment)
 - 표준 유전체(reference genome)에서 위치를 찾아주는 작업
 - 인간의 표준 유전체는 Genome Reference Consortium(GRC)에서 공개한 유전체 염기서열이 널리 쓰이고 있음.
 - 인간 표준 유전체는 2009년 hg19(GRCh37) 버전이 공개된 후에 이의 약점을 보완하고 개선된 hg38(GRCh38) 버전이 2013년 공개
 - 그러나 지난 10년간 많은 유전체 연구들이 hg19 버전을 기반으로 결과를 발표하였기 때문에 아직은 hg19 버전이 표준으로 더 널리 쓰이고 있는 상황

■ 05. NGS(Next Generation Sequencing) II

2 Next-Generation Sequencing(NGS; 차세대 염기서열 분석법)

- ◆ 변이 검출(Variant calling)
 - 각 시퀀싱 리드를 분석하여 특정 위치에서 표준 유전체 서열과 다른 변이(variation)가 있는지 찾아내는 작업
 - 한 위치에 여러 개의 시퀀싱 리드를 종합하여 변이가 있는지 확률적으로 판단
 - 각 시퀀싱 리드에는 NGS 장비에서 분석할 때 발생할 수 있는 에러, 매핑 과정에서 발생할 수 있는 오류 등이 혼재되어 있기 때문에 이러한 것들을 배제하고 진양성(true positive) 변이를 걸러낼 수 있는 통계적 알고리즘들이 사용
 - 여기에 이용되는 프로그램의 종류, 프로그램에서 설정한 파라미터, 데이터 전처리, 결과 필터링(filtering) 과정에 따라 동일한 데이터에 대해서 다른 결과가 나올 수 있어 위양성(false positive) 또는 위음성(false negative)의 가능성에 대한 주의가 필요

■ 05. NGS(Next Generation Sequencing) II

2 Next-Generation Sequencing(NGS; 차세대 염기서열 분석법)

- ◆ 유전자 복제수 Copy number variation, CNV
- ◆ NGS 분석 서버 및 프로그램
- ◆ 유전체 데이터베이스database 및 주석annotation
- ◆ 컴퓨터 시뮬레이션 in silico 알고리즘
- ◆ 데이터 시각화visualization

■ 05. NGS(Next Generation Sequencing) II

2 Next-Generation Sequencing(NGS; 차세대 염기서열 분석법)

◆ NGS 이용

- 개인의 염기서열 데이터(sequencing reads)를 표준 염기서열(reference genome)과 비교하는 작업을 하게 됨(매핑; Mapping)
 - ⇒ 질병의 원인 유전자를 찾기 위해서는 기존의 유전자 염기서열로부터 어떤 변화가 일어났는지 조사해야하기 때문
- 매핑을 통해 개인과 표준 염기서열의 차이를 알아낸 후 이를 적당한 선택 기준을 정해 신뢰할 수 있는 염기서열 변이 정보만 추출하게 됨
 - ⇒ 이 변이 정보는 단일염기서열변이이거나 짧은 삽입/결실임
- 그런 다음 염기서열 변이 정보를 기존 데이터베이스와 비교하여 이미 밝혀진 변이인지 새롭게 발견된 변이인지 판단하게 됨
- 그리고 그 변이가 아미노산의 변화를 가져올 것인지 아닌지, 또한 단백질 구조에 있어서 어떤 영향을 줄 것인지 예측하게 됨

■ 05. NGS(Next Generation Sequencing) II

2 Next-Generation Sequencing(NGS; 차세대 염기서열 분석법)

- ◆ NGS 플랫폼의 종류
- ◆ 로슈(Roche)
 - 2007년 로슈의 454 장비는 인류 역사상 최초의 NGS 플랫폼
: 첫 번째 대상자는 DNA 구조를 발견해 노벨상을 받은 제임스 왓슨
- ◆ 써모피셔사이언티픽(Thermo Fisher Scientific 장비)
 - 미세한 비드(bead)에 DNA 라이브러리를 붙여 촘촘하게 짜여진 미세한 구멍(well)에 하나씩 들어가도록 하고 각 구멍 밑에는 전류를 측정할 수 있는 반도체 칩 회로가 설계되어 있어 상보적인 염기가 결합하는 합성 과정에서 방출되는 수소이온(H^+)으로 인하여 pH가 낮아지는 것을 반도체칩의 센서가 탐지하고 분석하는 방식

■ 05. NGS(Next Generation Sequencing) II

2 Next-Generation Sequencing(NGS; 차세대 염기서열 분석법)

◆ 일루미나(illumina)

- 브릿지 증폭(Bridge PCR)이라고 해서 슬라이드 위에 DNA 단편을 고정시킨 후에 최대 1000 분자까지 증폭시켜 같은 서열의 DNA 단편 집단을 형성
- 이 집단을 클러스터(cluster)라고 표현하며 이 때 염기서열의 양끝에 붙은 어댑터(adaptor)가 flow cell 표면에 부착해 이를 브릿지(Bridge) 방식
- 이 클러스터를 주형으로 네 종류 의 형광 표식 염기를 사용한 일 염기 합성반응인 SBS(Sequencing by synthesis)를 수행

◆ 옥스포드 나노(Oxford Nano)

- 15분 만에 시퀀싱이 되는 혁신적인 기술
- 나노포어 시퀀싱 : DNA 한 가닥을 생물학적 세공(biological pore) 속으로 통과시키면서 전기전도성의 차이를 측정해 다양한 염기를 판별하는 기술
- 미니온(MinION) : 스마트폰보다 작은 크기의 초소형 장치. 여기 끝에 USB가 달려 시퀀싱 결과가 15분 만에 바로 컴퓨터 화면에 보이는 놀라운 기술

■ 05. NGS(Next Generation Sequencing) II

2 Next-Generation Sequencing(NGS; 차세대 염기서열 분석법)

- ◆ NGS 검사의 임상적 응용분야
 - 유전자 패널
 - 엑솜Exome 및 전장유전체whole genome 시퀀싱
 - 단일염기다형성SNP 검출
 - 혈액기반 종양 검사Blood-based tumor diagnostics
 - 비침습적 산전검사Noninvasive prenatal testing, NIPT
 - 인간백혈구항원Human leukocyte antigen, HLA
 - 면역글로불린 유전자 재조합Immunoglobulin rearrangement
 - RNA 시퀀싱RNA-sequencing, RNA-Seq
 - DNA 메틸화methylation
 - .Chromatin immunoprecipitation(ChIP) 시퀀싱
 - 단일세포 시퀀싱Single cell sequencing
 - 미생물 분야에서 NGS 활용