

생명과학의 세계 · 7주차

유전체 정보는 어떻게 후천적으로 바뀔 수 있을까?

RNA에 의한 표현형 변화 · 후성 유전체-전사체 · 차세대 서열 분석 · 빅데이터 AI 시대 통합 정리

지성욱 교수 (고려대 생명과학부)

(1) RNA에 의한 표현형 변화

(2) 후성 유전체-전사체 변형

(3) 차세대 서열 분석

(4) 빅데이터 AI

강의 (1) RNA에 의한 표현형 변화

학습목표: RNA가 일으키는 표현형(phenotype) 변화를 이해함

1-1 연구의 틀 – in vivo · in vitro · in silico Molecular & Computational Genomics Lab

in vivo (생명과학·질병) → 시퀀싱 → in silico (생물정보학) → 데이터 기반 가설 → in vitro (분자생물학)

→ 시퀀싱 → 다시 in vivo

- **in vivo**: 살아 있는 몸 안. 환자 샘플, 마우스 동물 모델로 질병을 봄
- **in silico**: 컴퓨터 안. 시퀀싱 데이터를 생물정보학으로 분석해 **데이터에서 가설을 끌어냄(data-driven hypothesis)**
- **in vitro**: 시험관·배양접시 안. 세운 가설을 분자생물학 실험으로 확인
- 세 영역 한가운데에 **유전체학(특히 RNA)**이 있음

정리 ▪ 이번 주 네 강의는 모두 이 순환 구조로 읽으면 됨. 시퀀싱으로 데이터를 얻고(3강), 컴퓨터로 분석하고(4강), 실험으로 RNA의 기능을 확인(1·2강)함

1-2 DNA 유전체의 구조 DNA genome structure

- DNA 유전정보는 **핵**에 있고, 생명현상은 **세포** 전체에서 일어남 (DNA 유전정보(핵) > 생명현상(세포))
- DNA 이중나선의 염기쌍: **A=T, G≡C** (A-T는 수소결합 2개, G-C는 3개)
- 교안 표기: **유전체(genome) = 유전자(gene) + 염색체(chromosome)** → 보충 참고

세포분열과 염색체

- 세포분열 때 유전정보가 복제됨. **전기**에 염색질이 응축하기 시작하고, **중기**에 염색체가 가장 응축되어 가운데에 늘어섬
- 중기 염색체를 **김사(Giemsa)** 염색하면 줄무늬(G band)가 보임 → 염색체를 크기·모양·띠 순으로 나열한 것이 **핵형(karyotype)**

구조	설명
뉴클레오솜 (nucleosome)	DNA가 히스톤 단백질 8량체를 감은 기본 단위 (DNA + 히스톤)
염색질 (chromatin)	뉴클레오솜이 줄줄이 이어지고 다시 감긴 섬유. 평소 핵 안의 상태
염색체 (chromosome)	분열 때 염색질이 고도로 응축된 형태 (condensed chromosome)
텔로미어 (telomere)	염색체 양 끝. 반복 서열로 끝을 보호함
동원체 (centromere)	두 염색분체가 붙은 잘록한 부위. 방추사가 붙는 곳
팔 (arm)	동원체를 기준으로 짧은 팔(p)과 긴 팔(q)

+ 보충

"유전체 = 유전자 + 염색체"를 정확히 읽기

- 유전체(genome)는 gene + chromosome을 합친 **단어의 어원**을 보여 주는 표기임. 의미는 "한 생물이 가진 **DNA 전체 염기서열**"
- 유전자(gene)는 그중 RNA·단백질을 만드는 정보가 담긴 구간이고, 염색체는 그 DNA가 포장된 구조물임

교안 등식을 그대로 "더하기"로 이해하면 개념이 섞이기 쉬워 덧붙임.

1-3 인간 유전체와 서열 분석의 발전 Human genome · NGS

- 인간 유전체: 약 **30억 개 DNA 염기서열**, 교안 표기로 23종류의 염색체 (→ 보충)

시기	사건	걸린 시간·비용
1953	왓슨·크릭, DNA 이중나선 구조 발표 (Nature)	-
1990~2003	인간 유전체 프로젝트(HGP) - 30억 염기서열 결정	13년, 30억 달러
2008	차세대 서열 분석(NGS) 으로 한 개인(제임스 왓슨)의 전체 유전체 해독 (Wheeler 외, Nature)	2달, 백만 달러 미만
2017~	Illumina NovaSeq - 30억 염기서열을 약 10번 겹쳐 읽음(×10)	2일, 약 100달러(약 10만원)

정리 ▪ 같은 30억 염기를 읽는 데 13년·30억 달러 → 2일·100달러. 이 비용 하락이 개인 유전체·빅데이터 연구(3·4강)의 출발점임

+ 보충

"23종류의 염색체"와 "James Watson (1928-)"

- 사람의 체세포 염색체는 **23쌍(46개)**임. 종류로 세면 상염색체 22종 + X + Y = **24종**이 됨. 교안 "23종류"는 "23쌍"으로 이해하면 됨
- 교안 핵형 그림의 마지막 칸 "XX or XY"가 성염색체 한 쌍임 (여성 XX, 남성 XY)
- 제임스 왓슨은 2025년 11월 별세함(향년 97세). 교안의 "(1928-)" 표기는 제작 시점 기준임

1-4 중심이론과 선천적·후천적 형질 Central dogma · Nature vs. nurture

DNA (유전체, Genomics) → 전사 → RNA (전사체, Transcriptomics) → 번역 →

단백질 (단백질체, Proteomics) → 형질 발현

- **중심이론(central dogma)**: 유전정보는 DNA → RNA → 단백질 방향으로 흐름. 맨 아래 DNA가 "생명정보", 맨 위가 "형질 발현"
- 각 층을 통째로 연구하는 학문: 유전체학 · 전사체학 · 단백질체학 (-omics = 전체를 다룸)
- 영화 <엑스맨>의 **돌연변이(mutants)** = DNA 자체가 바뀌어 형질이 달라지는 경우의 비유

선천적 vs 후천적 형질 획득

선천적 형질

일란성 쌍둥이는 **DNA가 같음** → 타고난 형질은 같아야 함

후천적 환경 변화

그런데 쌍둥이도 한 명만 병에 걸리기도 함 → **유전정보의 "발현" 차이**로 설명

- 교안 그림: Nature 표지 "THE MS GENOME" – 다발성 경화증(MS)이 한 명에게만 생긴 일란성 쌍둥이의 유전체 연구 (2010)
- 결론: DNA 서열이 같아도 **어떤 유전자가 언제·얼마나 켜지는지**가 달라 형질이 달라질 수 있음 → 그 조절의 중심에 **RNA**가 있음

+ 보충

MS 쌍둥이 연구의 실제 결과

- 이 연구(Baranzini 외, Nature 2010)는 한 명만 MS인 일란성 쌍둥이의 유전체·전사체·DNA 메틸화를 비교했으나, 병을 설명할 만한 **뚜렷한 차이를 찾지 못함**
- 그래서 "DNA만으로는 설명이 안 되니 다른 층(발현·환경)을 봐야 한다"는 문제 제기의 상징으로 자주 인용됨

1-5 유전체의 80%가 RNA를 만든다 Noncoding RNA

- 단백질을 만드는 **코딩 유전자**는 유전체의 일부(교안 원그래프: 약 5%, 그중 실제 단백질 정보 약 1%)뿐임 → 나머지 "94%?"는 무엇을 하나?

- 신체 부위(머리·몸통·다리)별로 보면 각기 다른 부분에서 RNA가 만들어지고, 이를 합하면 **유전체의 약 80%가 RNA만 특이적으로 생성함**
- 이 RNA 대부분은 단백질을 만들지 않는 **논코딩 RNA(noncoding RNA)**임
- 질문: 단백질을 안 만드는 RNA를 왜 이렇게 많이 만들까? → RNA 자체가 **기능**을 가진다는 답으로 이어짐 (DNA → RNA → 단백질 → 기능)

+ 보충

"80%"라는 숫자를 볼 때 알아 둘 점

- 이 수치는 국제 공동연구 **ENCODE(2012)**가 "유전체의 대부분이 어떤 조직에서든 한 번은 전사된다"고 보고한 데서 나옴
- 단백질을 암호화하는 엑손만 따지면 유전체의 **약 1~2%** 수준임
- "전사된다 = 기능이 있다"로 볼 수 있는지는 학계에서 논쟁이 있었음. 교안 흐름에서는 "논코딩 RNA도 중요한 기능을 한다"는 메시지로 이해하면 됨

1-6 마이크로RNA의 유전자 조절 RNA-mediated regulation: microRNA

- 도입: 영화 <스파이더맨>(2002) – 거미에게 물린 뒤 유전체가 바뀌는 설정, 대사 "합성 전이 RNA를 이용해" → **"RNA로 유전체 정보를 바꿀 수 있을까?"**
- 실제 생물학의 답: RNA가 DNA를 고쳐 쓰지는 않지만, **특정 유전자의 발현을 억제**해 형질을 바꿀 수 있음

유전형: DNA → 전사 → mRNA × **miRNA가 억제** → 번역 × → 단백질 → 기능 = 표현형

마이크로RNA(miRNA)의 작동 방식

- 길이 약 **21 nt**의 짧은 RNA. 이중가닥으로 만들어진 뒤 한 가닥이 **아고너트(Argonaute, Ago)** 단백질에 실려 **RISC** 복합체가 됨
- miRNA의 앞쪽 **seed 부위(5' 끝 2~8번 염기)**가 표적 mRNA와 짝을 이룸. 나머지는 일부만 짝지음(**partial pairing**)
- 주로 표적 mRNA의 **3' UTR**에 붙어 → mRNA 분해 또는 번역 억제 (**target gene repression**)
- mRNA는 약 1,000 nt로 miRNA보다 훨씬 김. seed만 맞으면 되므로 miRNA 하나가 **수백 개 유전자**를 조절할 수 있음

표적 mRNA 5' 코딩 영역 3' UTR 3'

완전히 짝지를 필요 없음

seed가 핵심 → 부분 결합만으로 억제

결과: mRNA 분해 · 번역 억제



miRNA는 seed(2~8번)로 표적 mRNA의 3' UTR에 결합함 (교안 Partial pairing 그림을 단순화)

+ 보충

miRNA와 siRNA, RNAi

- 짧은 RNA로 유전자를 끄는 현상 전체를 **RNA 간섭(RNAi)**이라 함. 1998년 Fire와 Mello가 예쁜꼬마선충에서 발견하여 2006년 노벨 생리의학상을 받음
 - **miRNA**: 몸이 스스로 만드는 조절자. 부분 결합으로 여러 표적을 조금씩 억제
 - **siRNA**: 표적과 완전히 상보적. 한 표적 mRNA를 정확히 잘라냄 → 8주차 RNA 치료제(Onpattro 등)로 이어짐
- 교안 마지막 피라미드에 등장하는 "RNAi"의 뜻을 짚어 둬. 8주차 내용과 직접 연결됨.

1-7 AGO CLIP-Seq과 뇌의 마이크로RNA 표적지도 Ago HITS-CLIP

- 문제: miRNA가 실제로 어떤 mRNA에 붙는지 알기 어려움 (seed만 맞아도 붙으니 후보가 너무 많음)
- 해법: Ago-miRNA-표적 mRNA가 한데 붙은 **3중 복합체(Ago ternary complex)**를 통째로 잡아 NGS로 읽음 = **AGO CLIP-Seq**

조직(뇌)에 자외선 → RNA-단백질 교차결합 → Ago 항체로 면역침강(IP) → RNA 크기별 분리 →

cDNA → NGS

교안 실험 그림	읽는 법
Protein (IB: Ago)	대조군 IgG에는 없고 Ago로 잡은 칸에만 약 105 kD 띠 → Ago가 제대로 잡힘
RNA (³² P 표지)	Ago에 붙은 RNA가 90 / 110 / 130 kD 위치에 퍼져 나타남
cDNA	110 kD 부분 → 짧은 cDNA(miRNA), 130 kD 부분 → 더 긴 cDNA(miRNA + mRNA 조각)

대뇌피질 마이크로RNA 표적지도 결과

- **아고너트-miRNA**: 상위 20개 miRNA가 Ago에 실린 miRNA의 **약 90%** 차지 (1위 miR-30e, 이어서 miR-30d, miR-9 ... miR-124 등)
- **아고너트-mRNA**: Ago가 붙은 자리(Ago footprint)가 봉우리(peak)로 나타남 → 약 **10,000개** Ago-mRNA 클러스터

- 두 데이터로 **seed 서열**로 짝지어 → 어떤 miRNA가 어느 염색체의 어느 mRNA에 붙는지 그린 "**뇌의 마이크로RNA 표적지도**" 완성

+ 보충

이 연구의 출처

- Chi(지성욱) 외, "Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps", **Nature 2009**. 교안 사진의 두 연구자가 공동 연구자(Robert Darnell)와 강의 교수임
- HITS-CLIP = High-Throughput Sequencing of RNA isolated by CrossLinking ImmunoPrecipitation. 자외선 교차결합으로 "실제로 붙어 있던" 쌍만 골라낸다는 점이 핵심

1-8 마이크로RNA로 바뀌는 세포와 뇌 miRNA changes cell identity

① miR-124: 세포를 다른 종류로 바꿈

- miR-124 서열: **UAAGGCACGCGGUGAAUGCCAA** (뇌에 가장 많은 miRNA 중 하나)
- 세포에 miR-124를 넣으면 약 **86시간** 뒤 세포가 신경세포처럼 긴 돌기를 뻗음 → **miRNA 하나로 세포 정체성이 바뀜**

② 유인원 특이적 신경세포 마이크로RNA

- 생쥐 뇌 vs 인간 뇌: 인간 뇌가 훨씬 크고 주름이 많음 → 그 차이를 만드는 RNA가 있을까?
- 생쥐 신경세포주(N2a)에 **유인원에만 있는 miRNA**를 넣으면 신경돌기 모양이 달라짐
- 이 miRNA를 가진 생쥐는 (교안 사진에서) 대뇌 영역이 더 크고, **단일세포 RNA 분석(scRNA-Seq)**에서 **GABA성 억제 신경세포** 군집이 늘어남
- scRNA-Seq 지도 표지: GABA(억제성 신경세포), Glut1-Glut2(흥분성 신경세포), Oligo(희소돌기아교세포), MG(미세아교세포), Astro(별아교세포)

정리 ▪ DNA 몇 글자가 아니라 작은 RNA 하나가 들어오는 것만으로 세포 모양·뇌 구성이 바뀔 수 있음 = "RNA에 의한 표현형 변화"

1-9 RNA의 복잡성과 종간 차이 RNA complexity

	사람 (Human)	생쥐 (Mouse)	예쁜꼬마선충 (Worm)
코딩 유전자 (DNA)	21,550	22,670	20,389
전사체 (RNA)	126,019	77,177	29,873

- 코딩 유전자 수는 사람·생쥐·선충이 **거의 비슷함** → 유전자 개수로는 복잡성 차이를 설명 못함
- RNA 종류(전사체 수)는 사람이 선충의 4배 이상 → **RNA 수준의 다양성**이 종간 형질 차이와 맞물림

- 한 유전자에서 여러 RNA가 나오는 방법(RNA variants):
 - **선택적 스플라이싱(alternative splicing)**: 엑손을 다르게 이어 붙임 (교안 그림: 가운데 엑손을 넣거나 빼기)
 - **선택적 폴리아데닐화(alternative polyadenylation)**: RNA 꼬리(poly A)를 붙이는 위치를 달리해 3' UTR 길이가 달라짐 → miRNA가 붙을 자리 수도 달라짐

강의 (1) 학습정리

논코딩 RNA는 유전체의 발현을 조절하여 표현형을 바꿀 수 있음 (Noncoding RNAs can regulate the expression of genome thereby changing the phenotypes)

강의 (2) 후성 유전체-전사체 변형에 의한 표현형 변화

학습목표: 후성유전(epigenetic)·후성전사(epitranscriptional) 변화가 표현형을 바꾸는 방식을 이해함

2-1 후천적 변화의 세 갈래

층	무엇이 바뀌나	예
RNAi (강의 1)	작은 RNA가 mRNA 발현을 억제	miRNA, siRNA
후성유전체 (epigenome)	DNA 서열은 그대로, DNA·히스톤에 붙는 화학 표지 가 바뀜	H3K4me3, H3K27me3, DNA 메틸화
후성전사체 (epitranscriptome)	RNA 서열은 그대로, RNA 염기 에 화학 변형 이 생김	산화 구아닌(o ⁸ G)

- 도입: 영화 <어메이징 스파이더맨>(2012) – 다른 종의 유전자를 섞는 실험 설정 → "어떻게 **전체** DNA 유전체 정보를 바꿨을까?" → 서열이 아니라 **표지**를 바꾸는 방법

2-2 후성유전체 – 히스톤 변형과 DNA 메틸화

Epigenome: same DNA, different modification, different trait

- 포장 계층: 염색체 → 초나선 구조(supercoiled) → 솔레노이드 → **뉴클레오솜**(히스톤 H2A·H2B·H3·H4 각 2개 + 링커 H1) → DNA
- 히스톤 단백질 꼬리(histone tails)가 밖으로 빠져나와 있고, 여기 특정 아미노산(주로 **라이신 K**)에 화학기가 붙음

① 히스톤 단백질 변형

- **아세틸화**(파랑): 대체로 유전자 켜짐
- **허용 메틸화**(초록, permissive): 켜짐 예) H3K4
- **억제 메틸화**(빨강, repressive): 꺼짐 예) H3K9, H3K27

② DNA 염기 메틸화

- DNA의 **시토신(C)**에 메틸기가 붙음 (methylated cytosine)
- 주로 C 다음에 G가 오는 자리(CpG)에서 일어남
- 유전자 앞부분(프로모터)이 메틸화되면 대체로 **발현 억제**

표지	읽는 법	의미
H3K4me3	히스톤 H3 의 4번째 라이신(K)에 메틸기 3개 (me3)	활성 유전자 시작 부위 표지 → 켜짐
H3K27me3	히스톤 H3의 27번째 라이신에 메틸기 3개	억제 표지 → 꺼짐

정리 ▪ "같은 DNA, 다른 변형, 다른 형질". 같은 책이라도 어느 쪽에 클립을 끼워 두느냐에 따라 펼쳐 읽는 페이지가 달라지는 것과 같음. 일란성 쌍둥이의 차이, 세포마다 다른 모양을 설명하는 틀임

2-3 후성전사체 – RNA 염기 변형 Epitranscriptome

DNA = 생명의 책 (Book of life)

원본 책에 형광펜을 칠함 → 다른 DNA 염기 변형
= 후성유전체

RNA = 사본 (Copy script)

책에서 베껴 쓴 노트에 형광펜을 칠함 → 다른 RNA
염기 변형 = 후성전사체

- 두 경우 모두 글자(서열)는 그대로이고 표시만 달라짐
- 이런 변형 위치는 NGS 서열분석법을 응용해 유전체 전체에서 찾아냄

+ 보충

RNA 변형은 이미 백신 속에 있음

- RNA에는 100종이 넘는 화학 변형이 알려져 있음. 대표적인 것이 m⁶A(아데닌 메틸화), 슈도유리딘(Ψ) 등
- 코로나19 mRNA 백신은 우리딘(U)을 N¹-메틸슈도유리딘(m¹Ψ)으로 바꿔 넣어 면역 거부를 줄였음 → 8주차 (3) 강에서 자세히 다룸

2-4 산화된 마이크로RNA와 심비대증 Position-specific oxidation of miR-1

활성산소(ROS) → 산화스트레스 → miRNA의 G가 산화 → 심비대증 → 심근경색·심장마비 위험

산화 구아닌 o⁸G

- 정식 이름: 7,8-dihydro-8-oxoguanine (8-oxo-G, o⁸G). 구아닌의 8번 탄소에 산소(=O)가 붙은 산화변형 (Kasai & Nishimura, 1984)
- 보통 G는 C와 짝을 지음 → o⁸G는 모양이 바뀌어 A와 짝을 지음 (new base-pairing)
- 즉 서열은 G 그대로인데, 실제로는 "T/U처럼" 읽히는 글자가 됨

쥐 심근세포 실험

넣은 RNA	서열	결과
miR-1 (정상 마이크로RNA)	UGGAAUGUAAAGAAGUAUGUAU	심근세포 정상 (0 hr 모습 유지)
7o8G-miR-1 (산화 변형 마이크로RNA)	UGGAAUG ^{o8G} UAAAGAAGUAUGUAU 7번째 G가 o ⁸ G	72시간 뒤 심근세포가 크게 비대, 생쥐 심장도 커짐

- 결론: miRNA의 특정 위치 염기 하나가 산화되는 것만으로 심비대증이 유도됨 (Seok 외, Nature 2020; Nature Reviews Cardiology 2020.8.20 하이라이트 "Oxidation of miRNAs by ROS leads to cardiac hypertrophy")

+ 보충

왜 "7번째" 염기가 중요한가

- 7번 위치는 miRNA의 **seed(2~8번)** 안에 있음 (1-6 참고)
- seed의 G가 o⁸G가 되어 C 대신 A와 짝을 지으면, miR-1이 알아보는 표적 mRNA가 **원래 표적에서 새로운 표적으로 바뀜** → 심근세포를 키우는 방향의 유전자 조절이 일어남
- "DNA는 그대로인데 RNA 한 글자의 화학 상태가 바뀌어 질병 형질이 생긴다" = 후성전사체에 의한 표현형 변화의 대표 예

종합 - 후천적으로 생명정보를 바꾸는 방법

피라미드 층	후천적 변화	읽는 도구
DNA (유전체)	후성유전체: H3K27me3, H3K4me3, DNA 메틸화	NGS 서열분석
RNA (전사체)	RNAi(miRNA), 후성전사체: RNA 염기 변형(o ⁸ G)	NGS
형질 발현	질병, 노화	

강의 (2) 학습정리

후성유전·후성전사 변형은 유전체 정보를 바꾸어 표현형을 바꿀 수 있음 (Epigenetic and epitranscriptional modification can change information in the genome thereby changing the phenotypes)

강의 (3) 차세대 서열 분석을 통한 연구

학습목표: 차세대 서열 분석(NGS)과 그 응용을 이해함

3-1 인간 유전체 프로젝트에서 NGS까지 HGP → NGS

- **인간 유전체 프로젝트(1990-2003)**: 13년, 30억 달러로 30억 염기서열 결정
- 2000년 6월 백악관 발표("Decoding the Book of Life") → **2001년 초안**(Nature) → **2004년 완성본**(Nature) 발표
- **차세대 서열 분석(NGS, 2008~현재)**: **초고속 · 대용량 · 병렬화**
 - DNA를 잘게 조각내(DNA fragmentation) 수많은 조각을 **동시에** 읽은 뒤 컴퓨터로 이어 붙임
 - Illumina NovaSeq(2017): 30억 염기 ×10, 2일, 100달러
- 결과: 누구나 자기 유전체를 읽는 **개인 유전체학(Personal Genomics)**과 **개인맞춤 의학(Personalized Medicine)** 시대 (TIME 2012 표지 "Want to Know My Future?")

+ 보충

1세대(생어) vs 차세대(NGS) – 교안 "100개의 조각 서열분석/run" 그림 읽기

	1세대: 생어(Sanger) 서열분석	차세대: NGS
방식	모세관 전기영동, 모세관 1개당 1조각(1 read/capillary)	유리판 위 수억~수십억 조각을 동시에 읽음
한 번에 읽는 양	약 100개(96개 모세관) 조각	수십억 조각
쓰임	인간 유전체 프로젝트, 지금도 짧은 구간 확인용	전체 유전체·전사체·CLIP-Seq 등

교안 그림(International Neurourology Journal 2016)의 "Capillary array electrophoresis"는 사실 **1세대 방식**을 보여 줌. "한 번에 100개 정도밖에 못 읽던 시절"과 비교해 NGS의 병렬화를 이해하면 됨.

- 참고: HGP 완성본에도 읽지 못한 빈칸(약 8%)이 남아 있었고, 2022년 T2T 컨소시엄이 **빈틈 없는 완전한 인간 유전체**를 발표함

3-2 초소형·이동형 서열 분석 Portable, real-time sequencing

- **Oxford Nanopore MinION**: USB 크기 장치를 노트북에 꽂아 현장에서 실시간 분석 (portable, real-time)
- 원리: DNA 가닥이 막에 박힌 **단백질 구멍(나노포어)**을 지날 때 염기마다 전류가 다르게 변함 → 전류 파형을 염기서열로 바꿈
- 2016년 NASA/SpaceX 9번째 화물선으로 국제우주정거장에 실려 가 **우주에서 처음 DNA 서열 분석** (우주비행사 Kate Rubins)
- 더 작은 SmidgION은 스마트폰에 연결하는 형태

- **IoT(Internet of Living Things, 생물 사물인터넷)**: 농장·현장에서 얻은 서열을 클라우드(Metricor)로 올려 분석

미래의 유전체학 10가지 예측(2030): 미국 국립인간유전체연구소(NHGRI) 영상 - 교안은 제목만 제시함

3-3 개인 유전자 분석과 질병 변이 탐색 Personal genome service · SNP

- **23andMe**: 침 채취 키트(saliva collection kit)를 우편으로 보내면 유전 정보를 분석해 주는 개인 유전자 분석 서비스

유전체 서열분석을 통한 질병 유전자 변이 탐색

- 여러 사람의 DNA를 읽어 비교하면 특정 위치에서 **한 염기만 다른 곳이 있음** = **SNP**(단일염기다형성) 예) 어떤 사람은 G(SNP^G), 어떤 사람은 A(SNP^A)
- **질병 환자군과 정상인군**에서 SNP 비율을 비교
 - 정상인군: 일정 비율이 G, 나머지가 A
 - 환자군에서 G가 **예상보다 많으면** → SNP^G가 **질병과 관련** (또는 SNP^A가 보호 효과)

+ 보충

소비자 유전자 검사를 볼 때

- 23andMe 같은 서비스는 대부분 전체 유전체를 NGS로 읽기보다 미리 정한 수십만 개 SNP만 확인하는 **DNA 칩 (마이크로어레이)** 방식을 씀
- "관련(association)"은 위험도가 조금 높다는 통계일 뿐, 그 병에 걸린다는 진단이 아님
- 23andMe는 2025년 파산 절차를 밟으며 고객 유전 정보의 처리 문제가 큰 논란이 됨 → 유전 정보는 바꿀 수 없는 개인정보라는 점에서 4강의 윤리(ELSI) 기동과 연결됨

3-4 고인류 유전체와 단일세포 유전체학 Ancient genomes · Single-cell genomics

잃어버린 DNA: 고인류 유전체 분석

- **네안데르탈인**: 크로아티아 Vindija 동굴, 러시아 Mezmaiskaya 동굴 뼈에서 DNA 추출 (Science 2010 표지 "The Neandertal Genome")
- **데니소바인**: 시베리아 Denisova 동굴의 손가락뼈 조각 하나에서 DNA로 찾아낸 새로운 고인류
- **현생 인류(호모 사피엔스)** 여러 집단의 유전체와 NGS로 비교 → 고인류와 섞인 흔적 발견
- **스반테 페보(Svante Pääbo)**: "멸종한 고인류의 유전체와 인류 진화에 관한 발견"으로 **2022년 노벨 생리의학상** 단독 수상

+ 보충

우리 몸속의 고인류 DNA

- 아프리카 밖 현대인의 유전체에는 네안데르탈인 DNA가 약 1~2% 섞여 있음
 - 파푸아뉴기니 등 오세아니아 집단에는 데니소바인 DNA가 수 % 남아 있음
 - 티베트인이 고산지대 저산소에 잘 견디게 하는 *EPAS1* 유전자 변이는 데니소바인에게서 온 것으로 알려짐
- 교안의 노벨상 슬라이드 제목이 앞 슬라이드 제목("질병 유전자 변이 탐색")과 같게 들어가 있으나 내용은 고인류 유전체임.

단일세포 유전체학 (Single-cell genomics)

- 기존 분석은 수많은 세포를 갈아 **평균값**을 읽음 → 세포 하나하나의 차이를 놓침
- 미세유체 장치에서 세포를 **물방울 하나에 하나씩** 가둔 뒤 각각에 바코드를 붙여 NGS로 읽음 (교안 아래 영상 그림)
- 태아 그림처럼 몸을 이루는 세포 하나하나의 종류를 점으로 찍어 지도를 만들 (Science 2017 "Single-Cell Genomics" 특집). 1-8의 scRNA-Seq 결과가 이 기술임

3-5 유전형에서 표현형으로 – 금발 유전자 예 Genotype → Phenotype

DNA = 유전형(genotype) → 전사(transcription) → RNA → 번역(translation) → 단백질 →

기능 = 표현형(phenotype)

- *KITLG* 유전자에서 약 350 kb 떨어진 **모낭 인핸서(hair follicle enhancer)**에 한 염기 변이: CACTAAAG → CGCTAAAG
- 원래 서열(A)에는 전사인자 **LEF**가 잘 붙어 *KITLG*가 많이 발현 → **어두운 머리색**
- G 변이에서는 LEF 결합이 약해져 발현이 줄어들 (화살표가 가늘어짐) → **금발**
- 단백질 자체는 그대로이고 **발현량**만 달라져 형질이 바뀐 예 (Nature Genetics 2014)

강의 (3) 학습정리

차세대 서열 분석(NGS)이 개인 유전체학의 새 시대를 열 (Next-generation sequencing opens new era of personal genomics)

강의 (4) 빅데이터 AI 시대의 유전체 연구

학습목표(교안 표기): NGS와 그 응용 – 이번 강의는 NGS가 만든 빅데이터를 해석하는 RNA 생물학·AI·시스템 생물학으로 확장

4-1 RNA 세계와 RNA의 복잡성 RNA World

- 복습: 단백질 코딩유전자는 약 5%인데 유전체의 약 80%가 논코딩 RNA를 만들 (1-5) → 왜?

	현대 세계 (중심이론)	RNA 세계 (태고)
정보 저장	DNA	모두 RNA 가 담당 (스스로 복제)
정보 저장·전달	RNA	
기능	단백질	

- RNA는 한 가닥이라 스스로 접혀 **2차 구조**를 만들 수 있음 → **단백질처럼 기능(효소 작용)**도 가능
- 그래서 생명의 첫 유전물질은 RNA였다는 **RNA 세계 가설**이 나옴. 효소 기능을 하는 RNA(리보자임)를 발견한 **토머스 체크(Thomas Cech)**
- RNA의 복잡성: 사람 코딩 유전자 21,550개 → 전사체 126,019개 (1-9). 선택적 스플라이싱을 발견한 **필립 샤프(Phillip Sharp)**

+ 보충

교안 인물의 노벨상

- 토머스 체크: RNA의 촉매 기능 발견으로 **1989년 노벨 화학상**(시드니 올트먼과 공동)
- 필립 샤프: 유전자가 엑손·인트론으로 나뉘어 있음(split genes)을 발견해 **1993년 노벨 생리의학상**(리처드 로버츠와 공동)

4-2 RNA 결합 단백질(RBP)과 질병 RNA-binding protein

- RNA는 혼자 일하지 않고 **RNA 결합 단백질(RBP)**과 **RNA-단백질 복합체**를 이뤄 일함
- RBP가 하는 일: ① **전사체 변이**(선택적 스플라이싱으로 엑손 선택) ② **전사후 조절**(3' UTR에 붙어 안정성·번역 조절)
- 질문: "RNA-단백질 결합 복합체"가 생물학적 기능에서 매우 중요한가? → PND 사례로 답함

자가면역성 퇴행성 뇌질환 PND (Paraneoplastic Neurological Disorder)

몸 어딘가의 암(유방·난소·폐) → 암 항원에 면역반응 → 같은 단백질을 가진 신경세포까지 공격 →

자가면역 뇌 질환

- 환자 혈액(항체)으로 암 항원 유전자를 클로닝 → 유방/난소암 항원, 폐암 항원 확인

- 밝혀진 정체: 모두 **RNA 결합 단백질 (Nova, Hu)**. 원래 신경세포에만 있어야 할 RBP를 암세포가 만들어 면역계 표적이 된 것
- **Nova 표적 RNA 지도(HITS-CLIP)**: Nova가 붙는 RNA를 지도로 그리니 GABA·글리신 수용체, 이온 통로 등 **시냅스 단백질**이 대거 포함됨 → Nova의 생물학적 기능 = 시냅스 조절
- **Nova 결손(KO) 생쥐**: 운동 이상을 보이며 정상적으로 자라지 못함

RBP와 질병 네트워크

질병군	관련 RBP 예 (교안 그림)
암	SF2/ASF, eIF4E, Sam68, HuR, TLS/FUS-EWS(전좌로 육종)
신경질환	NOVA-1,2, elav/HuB·C·D(자가면역), ATXN2(실조증), FMRP(취약 X 증후군, FXS)
퇴행성 근육질환	SMN(척수성 근위축증, SMA), MBNL-CUGBP1(근긴장성 이영양증, DM), PABPN1(OPMD)

교안 그림 오른쪽: James Darnell, 『RNA: Life's Indispensable Molecule』

+ 보충

8주차와 이어지는 예 - 척수성 근위축증(SMA)

- SMA는 RBP의 일종인 SMN 단백질이 부족해 생기는 병임
- 8주차 RNA 치료제 표의 첫 줄 **Nusinersen(Spinraza)**이 바로 SMA 치료제로, ASO가 비슷한 유전자 **SMN2**의 **스플라이싱을 바꿔** 정상 SMN을 더 만들게 함
- "RBP-스플라이싱 이해(7주차) → RNA로 고치는 약(8주차)"의 흐름임

4-3 AlphaFold와 시스템 생물학 AI · Systems biology

- 여러 단계의 생명정보: 유전체·전사체는 NGS로 대량 측정 가능 → 단백질 단계는 **AI**가 돕기 시작

AlphaFold: AI로 단백질 3차원 구조 예측

- 구글 딥마인드(2016년 이세돌 9단과 대국한 알파고의 회사)가 개발
- 단백질은 아미노산 사슬 → 알파나선·베타병풍(pleated sheet) → 3차원 접힘 구조로 기능함
- 입력: 아미노산 서열(예: MALSVLLL...) → 유전 데이터베이스에서 진화적으로 비슷한 서열 정렬(MSA) + 구조 데이터베이스 주형 → **3차원 구조 출력**
- 단백질 구조 예측 대회(CASP) 점수(GDT): 2016년까지 40점 안팎 → AlphaFold(2018) 약 60 → **AlphaFold 2(2020) 90 근처**. 90점 이상이면 실험으로 구한 구조와 거의 같은 수준
- AlphaFold 단백질 구조 데이터베이스로 공개되어 누구나 검색 가능

+ 보충

2024년 노벨 화학상

- 데미스 허사비스·존 점퍼(AlphaFold)와 데이비드 베이커(단백질 설계)가 2024년 노벨 화학상을 받음
- AlphaFold 데이터베이스에는 2억 개가 넘는 단백질 예측 구조가 들어 있음

시스템 생물학: 네트워크

분자세포 생물학

신호 경로를 한 줄씩 따라감 (예: MAPK/ERK 경로
Ras→Raf→Mek→Erk → 핵 → 전사, cAMP
경로)

네트워크 생물학

수백 개 분자가 서로 얽힌 망 전체를 한꺼번에 봄
(세포막 → 세포질 → 핵 → 하위 유전자)

유전체 → 전사체 → 단백질체 → 대사체 → **형질체**

여러 층의 -omics 데이터를 쌓아 하나의 유전 요소가 어떻게 형질까지 이어지는지 통합적으로 분석함.

4-4 개인 유전체 맞춤 의학과 빅데이터 Big data in personalized medicine

- "1,000달러 유전체, 10만 달러 분석?" (The \$1,000 genome, the \$100,000 analysis?) – 읽는 비용은 싸졌지만 해석이 병목이 됨
- 교안 만화: 환자 침대에 서열이 적힌 종이가 잔뜩 쌓여 있지만 의사들은 어찌할 줄 모름
- 23andMe(데이터 생산), Google Genomics(클라우드 분석) 등 IT 기업이 뛰어듦
- NHGRI의 유전체 연구 미래 비전 그림: 인간 유전체 프로젝트를 토대로 "생물학 → 건강 → 사회" 층을 쌓고, 기둥으로 자원·기술개발·계산생물학(Computational biology)·교육훈련(Training)·윤리(ELSI)·교육이 받침 → 빨간 상자로 강조된 것이 계산생물학과 인력 양성
- 영화 <이노스페이스>처럼 세포 안 유전체를 "세포를 돌리는 마스터 프로그램"으로 보고, 컴퓨터 앞에서 분석하는 생명공학 시대

강의 (4) 학습정리

NGS가 가져온 빅데이터가 유전체학의 풀리지 않은 문제를 밝혀낼 것임 (Big data brought by NGS will elucidate unsolved problems in genomics)

용어 용어 정리와 교안 표기 정정

7주차 네 강의에 나온 핵심 용어

영문	국문	한 줄 정리
genome / gene / chromosome	유전체 / 유전자 / 염색체	DNA 전체 / 기능 단위 구간 / DNA가 포장된 구조
nucleosome · histone	뉴클레오솜 · 히스톤	DNA가 히스톤 8량체를 감은 기본 포장 단위
karyotype	핵형	중기 염색체를 크기·모양 순으로 나열한 그림 (김사 염색 G band)
central dogma	중심이론	DNA → (전사) → RNA → (번역) → 단백질
genomics · transcriptomics · proteomics	유전체학 · 전사체학 · 단백질체학	각 층을 통째로 측정·분석하는 학문
noncoding RNA	논코딩 RNA	단백질을 만들지 않고 RNA 자체로 기능하는 RNA
microRNA (miRNA)	마이크로RNA	약 21 nt, seed(2~8번)로 표적 mRNA 3' UTR에 붙어 발현 억제
Argonaute (Ago) · RISC	아고너트 · RNA 유도 침묵 복합체	miRNA를 싣고 표적을 찾아가는 단백질과 그 복합체
RNAi	RNA 간섭	짧은 RNA가 상보적 mRNA의 발현을 끄는 현상
CLIP-Seq (HITS-CLIP)	교차결합 면역침강 시퀀싱	자외선으로 RNA-단백질을 고정해 실제 결합 RNA를 NGS로 읽음
alternative splicing	선택적 스플라이싱	엑손 조합을 바꿔 한 유전자에서 여러 mRNA 생성
alternative polyadenylation	선택적 폴리아데닐화	poly(A) 붙는 위치를 바꿔 3' UTR 길이를 달리함
epigenome	후성유전체	서열 변화 없이 DNA·히스톤에 붙은 화학 표지 전체
epitranscriptome	후성전사체	서열 변화 없이 RNA 염기에 생긴 화학 변형 전체
H3K4me3 / H3K27me3	–	히스톤 H3 4번/27번 라이신 트라이메틸화 = 활성/억제 표지
8-oxoguanine (o ⁸ G)	8-옥소구아닌	산화된 G. C 대신 A와 짝지음
cardiac hypertrophy	심비대증	심근세포가 커져 심장이 비대해지는 상태
NGS	차세대 서열 분석	수십억 DNA 조각을 병렬로 읽는 초고속·대용량 기술
SNP	단일염기다형성	사람마다 한 염기가 다른 위치. 질병 관련성 탐색에 씀
single-cell RNA-Seq	단일세포 RNA 분석	세포 하나하나의 RNA를 따로 읽어 세포 지도를 만들
genotype / phenotype	유전형 / 표현형	가진 유전 정보 / 겉으로 드러난 형질

영문	국문	한 줄 정리
ribozyme · RNA world	리보자임 · RNA 세계	효소 기능 RNA · RNA가 정보와 기능을 모두 맡던 태고 가설
RNA-binding protein (RBP)	RNA 결합 단백질	RNA와 복합체를 이뤄 스플라이싱·안정성·번역을 조절
PND	부종양성 신경 질환	암 항원(신경 RBP)에 대한 면역이 신경세포를 공격하는 병
AlphaFold	-	아미노산 서열에서 단백질 3차원 구조를 예측하는 AI
systems biology	시스템 생물학	분자들의 네트워크 전체를 통합적으로 분석

+ 보충

교안 표기에서 헛갈릴 수 있는 부분

위치	교안 표기	바르게 읽기
(1) 인간 유전체	23종류의 염색체	23쌍 (종류로는 $22 + X + Y = 24$ 종)
(1)·(4) RNA 복잡성	Alternative poly adenylation (스프라이싱)	"스플라이싱"은 alternative splicing의 번역. poly adenylation은 "폴리아데닐화"
(1) NGS	James Watson (1928-)	2025년 11월 별세
(2) 심비대증	Shibatani S et al (1991)	Shibutani S 외 (1991), Nature
(3) 고인류	잃어버린 DNA	잃어버린 DNA
(3) 노벨상 슬라이드	제목 "질병 유전자 변이 탐색"	내용은 고인류 유전체 연구(페보, 2022)
(3) NGS 병렬화	100개의 조각 서열분석 / run	그림은 1세대 생어 방식. NGS는 수십억 조각
(4) 학습목표	NGS and its application	(3)강 목표가 그대로 들어감. 실제 내용은 RNA 생물학·AI·시스템 생물학

문제

4지선다 예상문제 10

정답은 각 문제 아래를 눌러 확인

1. 마이크로RNA가 표적 mRNA를 알아보는 데 가장 핵심이 되는 부위는?

1. miRNA의 3' 말단
2. miRNA 5' 쪽 2~8번째 염기(seed)
3. 표적 mRNA의 poly(A) 꼬리
4. 표적 mRNA의 5' cap

▼ 정답 · 해설

㉔ seed 부위만 맞으면 나머지는 부분 결합(partial pairing)만으로도 표적 mRNA의 3' UTR에 붙어 발현을 억제함.

2. AGO CLIP-Seq 실험에서 130 kD 부근에서 얻은 cDNA가 담고 있는 것으로 가장 알맞은 것은?

1. Ago 단백질만
2. miRNA만
3. miRNA와 표적 mRNA 조각
4. 대조군 IgG에 붙은 RNA

▼ 정답 · 해설

㉔ 110 kD는 Ago+miRNA, 130 kD는 Ago+miRNA+mRNA 3중 복합체로 더 긴 cDNA(mRNA 포함)가 나옴.

3. "H3K27me3"에 대한 설명으로 옳은 것은?

1. DNA의 27번째 시토신이 메틸화된 것
2. 히스톤 H3의 27번째 라이신에 메틸기 3개가 붙은 억제 표지
3. 히스톤 H3의 27번째 라이신이 아세틸화된 활성 표지
4. RNA의 27번째 구아닌이 산화된 것

▼ 정답 · 해설

㉔ H3 = 히스톤 H3, K27 = 27번 라이신, me3 = 트라이메틸. 유전자 꺼짐과 관련됨. H3K4me3는 반대로 활성 표지.

4. 산화된 구아닌(o⁸G)의 특징으로 옳은 것은?

1. 시토신과만 더 강하게 결합함
2. 아데닌과 새로운 염기쌍을 이룸
3. DNA에서만 생기고 RNA에서는 생기지 않음
4. 히스톤 단백질에 붙는 변형임

▼ 정답 · 해설

㉔ 8번 탄소가 산화된 G는 C 대신 A와 짝을 지음(new base-pairing). RNA(miRNA)에서도 생기며 후성전사체 변형의 예임.

5. 쥐 심근세포에 7o8G-miR-1(7번째 G가 산화된 miR-1)을 넣었을 때 나타난 결과는?

1. 심근세포가 신경세포로 바뀜
2. 변화 없음
3. 심근세포가 비대해지고 심장이 커짐
4. 심근세포가 모두 죽음

▼ 정답 · 해설

㉔ seed 안의 한 염기 산화로 표적이 바뀌어 심비대증이 유도됨 (Seok 외, Nature 2020).

6. 사람·생쥐·선충의 코딩 유전자 수는 비슷하지만 전사체 수는 사람이 훨씬 많다. 이를 설명하는 기전으로 가장 알맞은 것은?

1. 선택적 스플라이싱과 선택적 폴리아데닐화
2. 염색체 수의 차이
3. DNA 이중나선 구조의 차이
4. 리보솜 수의 차이

▼ 정답 · 해설

㉑ 한 유전자에서 엑손 조합·3' 끝 위치를 바꿔 여러 RNA 변이체를 만들기 때문임.

7. 차세대 서열 분석(NGS)에 대한 설명으로 옳지 않은 것은?

1. DNA를 조각내 수많은 조각을 동시에 읽음
2. 모세관 하나에 한 조각씩, 한 번에 약 100조각만 읽음
3. 사람 유전체 분석 비용을 100달러 수준까지 낮춤
4. 나노포어 방식은 휴대용·실시간 분석이 가능함

▼ 정답 · 해설

㉔ 모세관 전기영동으로 한 번에 약 100조각을 읽는 것은 1세대 생어 방식임.

8. 2022년 노벨 생리의학상을 받은 스펀테 페보의 업적은?

1. RNA 간섭 발견
2. 멸종한 고인류의 유전체와 인류 진화 연구
3. 인간 유전체 프로젝트 완성
4. 단백질 구조 예측 AI 개발

▼ 정답 · 해설

㉔ 네안데르탈인 유전체 해독, 데니소바인 발견 등.

9. 부종양성 신경 질환(PND) 환자의 혈액으로 찾아낸 암 항원 Nova, Hu의 정체는?

1. DNA 메틸화 효소
2. RNA 결합 단백질(RBP)
3. 히스톤 단백질
4. 마이크로RNA

▼ 정답 · 해설

㉔ 신경세포의 RBP를 암세포가 만들고, 그에 대한 면역반응이 신경세포까지 공격함.

10. AlphaFold가 입력받아 예측하는 것으로 옳은 것은?

1. DNA 서열 → 유전자 발현량
2. 아미노산 서열 → 단백질 3차원 구조
3. RNA 서열 → miRNA 표적
4. SNP → 질병 발생 여부

▼ 정답 · 해설

㉔ CASP 대회에서 AlphaFold 2(2020)는 실험 구조에 가까운 정확도(GDT 90 안팎)를 보임.

자료 기준: 「생명과학의 세계」(고려대학교) 7주차 지성욱 교수 강의 교안 - 유전체 정보는 어떻게 후천적으로 바뀔 수 있을까? (1) RNA에 의한 표현형 변화 (2) 후성 유전체-전사체 변형에 의한 표현형 변화 (3) 차세대 서열 분석을 통한 연구 (4) 빅데이터 AI 시대의 유전체 연구.

교안 그림 출처(Nature, Science, Wikimedia, Getty Images Bank 등)는 각 강의 교안의 "출처 및 참고문헌"을 따름. 이 문서의 그림은 교안 그림을 단순화해 새로 그린 것임.

'보충' 상자와 예상문제는 교안에 없는 참고 내용으로, 시험 범위 여부는 강의 기준을 따를 것.