KSBi-BIML 2024

Bioinformatics & Machine Learning(BIML) Workshop for Life and Medical Scientists

생명정보학 & 머신러닝 워크샵 (온라인)

Introduction to single cell multiomics

황병진 _ 연세대학교

KSBI KOREAN SOCIETY FOR BIOINFORMATICS | 한국생명정보학회

본 강의 자료는 한국생명정보학회가 주관하는 BIML 2024 워크샵 온라인 수업을 목적으로 제작된 것으로 해당 목적 이외의 다른 용도로 사용할 수 없음을 분명하게 알립니다.

이를 다른 사람과 공유하거나 복제, 배포, 전송할 수 없으며 만약 이러한 사항을 위반할 경우 발생하는 **모든 법적 책임은 전적으로 불법 행위자 본인에게 있음을 경고**합니다.

KSBi-BIML 2024

Bioinformatics & Machine Learning(BIML)
Workshop for Life and Medical Scientists

안녕하십니까?

한국생명정보학회가 개최하는 동계 교육 워크샵인 BIML-2024에 여러분을 초대합니다. 생명정보학 분야의 연구자들에게 최신 동향의 데이터 분석기술을 이론과 실습을 겸비해 전달하고자 도입한 전문 교육 프로그램인 BIML 워크샵은 2015년에 시작하여 올해로 벌써 10년 차를 맞이하게 되었습니다. BIML 워크샵은 국내 생명정보학 분야의 최초이자 최고 수준의 교육프로그램으로 크게 인공지능과 생명정보분석 두 개의 분야로 구성되어 있습니다. 올해 인공지능 분야에서는 최근 생명정보 분석에서도 응용이 확대되고 있는 다양한 인공지능 기반 자료모델링 기법들에 대한 현장 강의가 진행될 예정이며, 관련하여 심층학습을 이용한 단백질구조예측, 유전체분석, 신약개발에 대한 이론과 실습 강의가 함께 제공될 예정입니다. 또한 단일세포오믹스, 공간오믹스, 메타오믹스, 그리고 롱리드염기서열 자료 분석에 대한 현장 강의는 많은 연구자의 연구 수월성 확보에 큰 도움을 줄 것으로 기대하고 있습니다.

올해 BIML의 가장 큰 변화는 최근 연구 수요가 급증하고 있는 의료정보자료 분석에 대한 현장 강의를 추가하였다는 것입니다. 특히 의료정보자료 분석을 많이 수행하시는 의과학자 및 의료정보 연구자들께서 본 강좌를 통해 많은 도움을 받으실 수 있기를 기대하고 있습니다. 또한 다양한 생명정보학 분야에 대한 온라인 강좌 프로그램도 점차 증가하고 있는 생명정보 분석기술의 다양화에 발맞추기 위해 작년과 비교해 5강좌 이상을 신규로 추가했습니다. 올해는 무료 강좌 5개를 포함하여 35개 이상의 온라인 강좌가 개설되어 제공되며, 연구 주제에 따른 연관된 강좌 추천 및 강연료 할인 프로그램도 제공되며, 온라인을 통한 Q&A 세션도 마련될 예정입니다. BIML-2024는 국내 주요 연구중심 대학의 전임 교원이자 각 분야 최고 전문가들의 강의로 구성되었기에 해당 분야의 기초부터 최신 연구 동향까지 포함하는 수준 높은 내용의 강의가 될 것이라 확신합니다.

BIML-2024을 준비하기까지 너무나 많은 수고를 해주신 운영위원회의 정성원, 우현구, 백대현, 김태민, 김준일, 김상우, 장혜식, 박종은 교수님과 KOBIC 이병욱 박사님께 커다란 감사를 드립니다. 마지막으로 부족한 시간에도 불구하고 강의 부탁을 흔쾌히 허락하시고 훌륭한 현장 강의와 온라인 강의를 준비하시는데 노고를 아끼지 않으신 모든 강사분들께 깊은 감사를 드립니다.

2024년 2월

한국생명정보학회장 이 인 석

강의개요

Introduction to single cell multiomics

단일 세포 기술의 발달로 유전체, 전사체, 단백체, 그리고 후성 유전제를 분석할 수 있는 기술들이 빠른 속도로 개발되고 있다. 하지만 실제 biological한 의미를 가진 세포를 정확하게 정의하기 위해서는 여러 표현형을 동시에 측정하는 멀티오믹스 기술이 요구된다. 이런 방법론들의 예로, 한 세포에서 RNA와 표면 단백질 abundance를 측정하는 방법들 (CITE-seq, REAP-seq), 염색질과 전사체를 동시에 측정하는 기술 (10x multiome, sci- CAR) 이 대표적인 멀티오믹스 (multiomics) 기술들이 대표적이다.

본 강의에서는 최신 단일세포 멀티오믹스 데이터 종류들에 대해서 배우고, 이들이 어떻게 만들어지는지 기술적인 원리와 개념을 배우는 것을 목표로 한다. 또한 이런 기술들을 적용하여 실제 논문에서 분석된 예제들을 살펴본다.

강의는 다음의 내용을 포함한다:

- 단일세포 기술의 역사
- 단일세포 멀티오믹스 기술 I
- 단일세포 멀티오믹스 기술 II
- 단일세포 멀티오믹스 분석 방법론의 적용 예

* 교육생준비물: 노트북 (이론강의로 파워포인트나 PDF가 문제없이 열리면 됨)

* 강의 난이도: 초급

* 강의: 황병진교수 (연세대학교 의과대학 의생명과학부)

Curriculum Vitae

Speaker Name: Byungjin Hwang, Ph.D.

▶ **Personal Info**

Name	Byungjin Hwang
Title	Assistant Professor
Affiliation	Yonsei University, Department of Biomedical Sciences

▶ **Contact Information**

Address	502 Avison Biomedical Research Center (ABMRC), 50-1 Yonsei-ro, Seodaemun-gu, Seoul 120-752, Korea
Email	bjhwang113@yuhs.ac
Phone Number	+82-2-2228-0877

Research Interest

Single cell multi-omics, CRISPR engineering and Cancer-immunology

Educational Experience

2012	B.S. in Chemistry, Yonsei University, Korea
2018	Ph.D. in Genome engineering and Bioinformatics, Yonsei University, Korea

Professional Experience

2018.12-2022.8	Post-doc research fellow, Institute for Human Genetics, UCSF, USA
2022.7-2022.8	Visiting Scholar, University of Michigan, USA
2022.9-	Assistant Professor, Yonsei University, Severance Biomedical Science Institute, Korea

Selected Publications (5 maximum)

1. Connor A. Tsuchida, Nadav Brandes, Raymund Bueno, Marena Trinidad, Thomas Mazumder, Bingfei Yu, **Byungjin Hwang**, Christopher Chang, Jamin Liu, Yang Sun, Caitlin R. Hopkins, Kevin R. Parker, Yanyan Qi, Ansuman T. Satpathy, Edward A. Stadtmauer, Jamie H.D. Cate, Justin Eyquem, Joseph A. Fraietta, Carl H. June, Howard Y. Chang, Chun Jimmie Ye, Jennifer A. Doudna, ***Cell***, 2023, "Mitigation of chromosome loss in clinical CRISPR-Cas9-engineered T cells" (**Engineered main plasmid vector system for this CRISPR screen**)
2. **Byungjin Hwang***, David S. Lee*, Whitney Tamaki, Yang Sun, Anton Ogorodnikov, George Hartoularos, Aidan Winters, Bertrand Yeung, Kristopher L. Nazor, Yun S. Song, Eric D. Chow, Matthew H. Spitzer, Chun Jimmie Ye, ***Nature Methods***, 2021, doi: 10.1038/s41592-021-01222-3, "SCITO-seq: single-cell combinatorial indexed cytometry sequencing".
3. **Byungjin Hwang***, Wookjae Lee*, Soo-Young Yum*, Yujin Jeon, Namjin Cho, Goo Jang, Duhee Bang; ***Nature Communications***, 2019, doi:10.1038/s41467-019-09203-z, "Lineage tracing using a Cas9-deaminase barcoding system targeting endogenous L1 elements".
4. Namjin Cho*, **Byungjin Hwang***, Jung-Ki Yoon*, Sangun Park*, Joongoo Lee*, Han Na Seo, Jeewon Lee, Sunghoon Huh, Jinsoo Chung, and Duhee Bang, ***Nature Communications***, 2015, DOI:10.1038/ncomms9351, "De novo assembly and next-generation sequencing to analyze full-length gene variants from codon-barcoded libraries".

KSBI | 한국생명정보학회

KSBi-BIML 2024

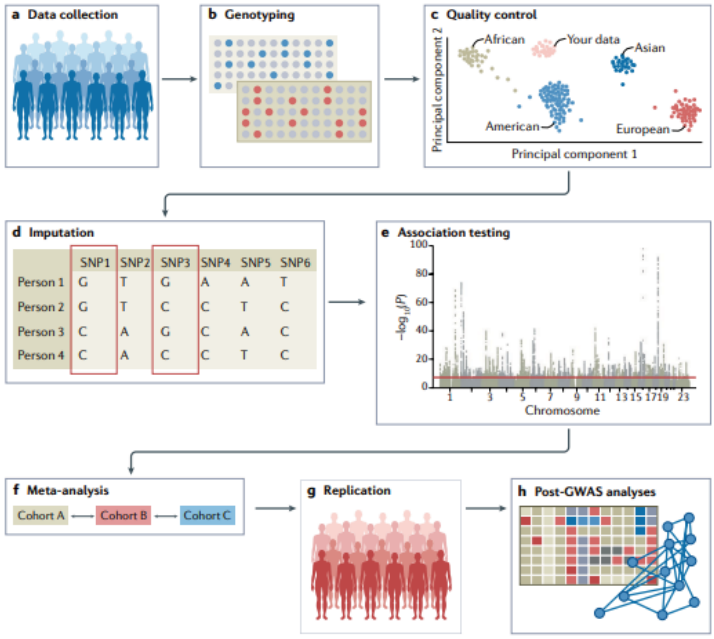
Introduction to single cell multiomics

연세대학교 의과대학 황병진

Contents

What we learned from the GWAS

GWAS (Genome-wide association studies)

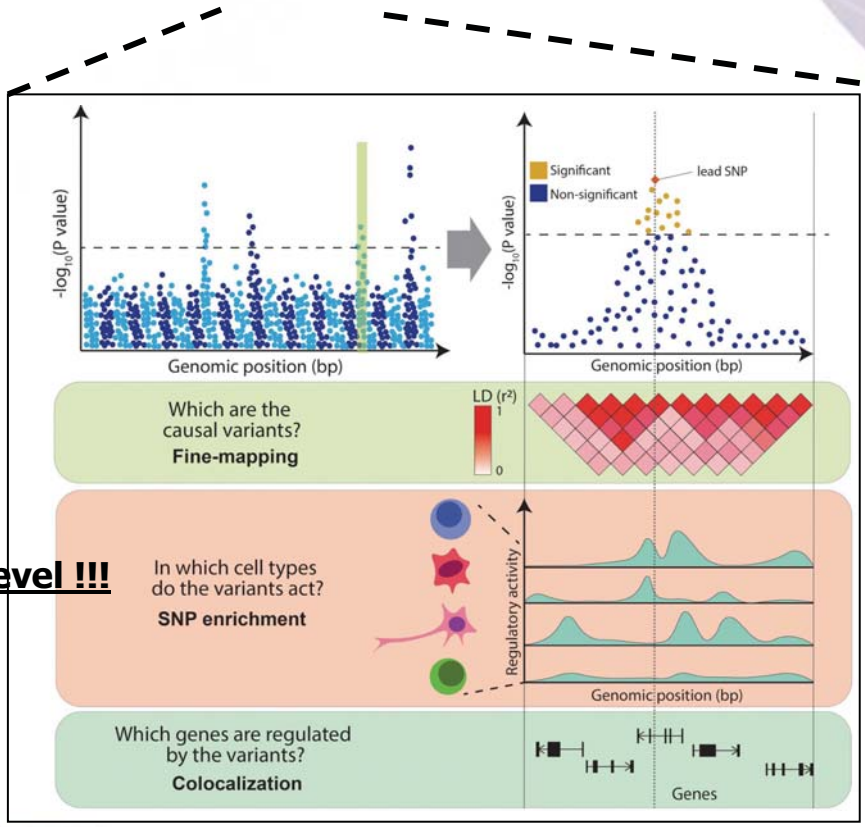


Fine mapping
SNP-to-Gene map
Gene-to-Function map
Pathway analysis
Gene-gene correlation

SNP (단일염기다형성) :
1,000개의 염기마다
하나씩 존재,
4,5백만/사람

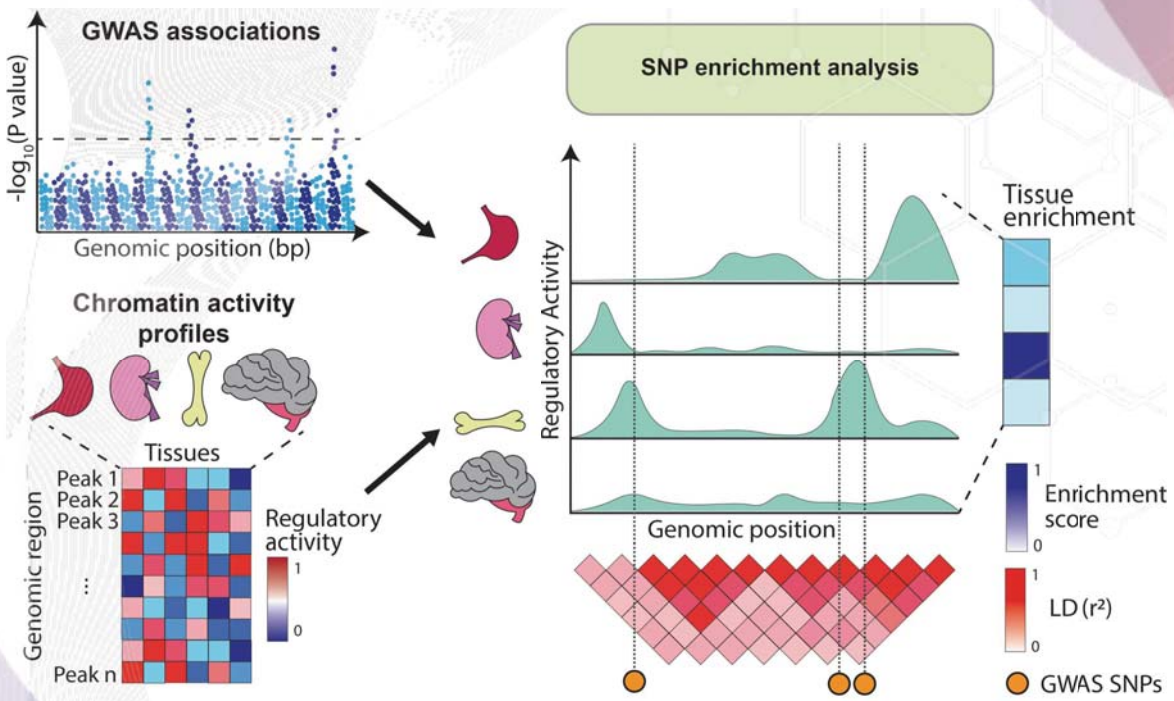
From GWAS to Function

DNA -> RNA -> Protein -> **Phenotype**



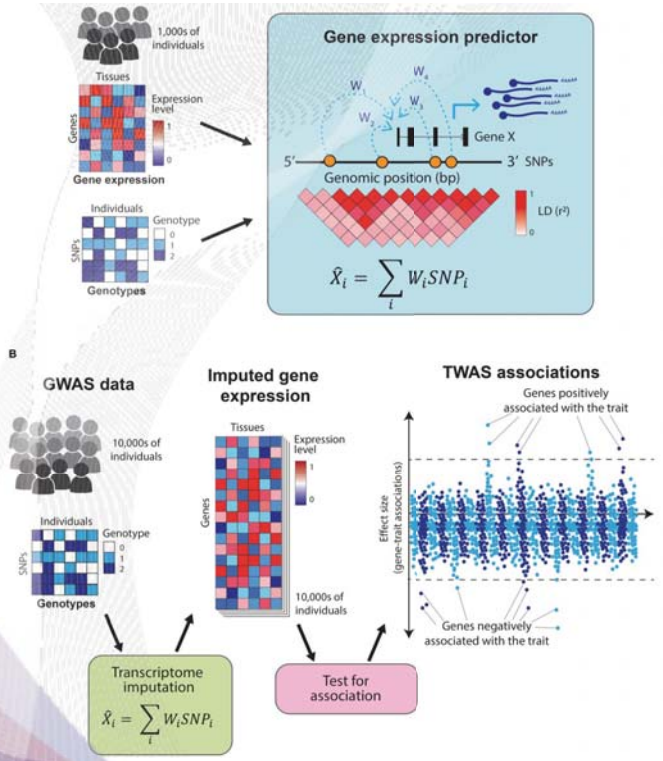
Singe cell level !!!

SNP enrichment and chromatin annotation



Chromatin activity : 염색체의 풀림정도를 측정

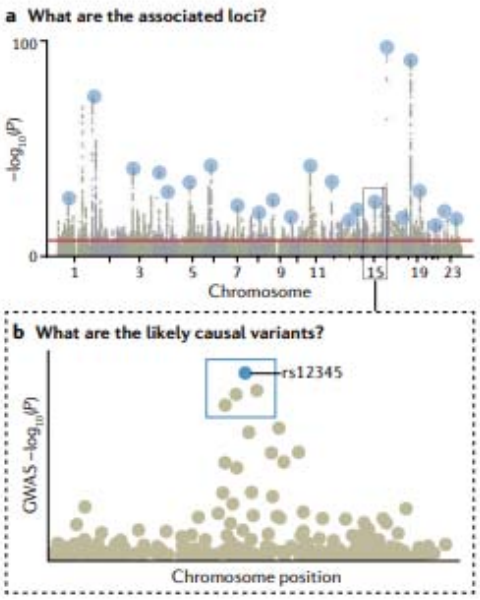
Overview of transcriptome-wide association studies (TWAS)



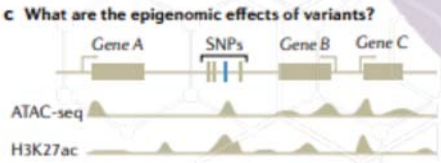
Gene level association
Vs
SNP level association

Less burden for testing size
(각 loci가 통계적으로 의미가 있는지)
3.3Gbp -> 20,000 genes

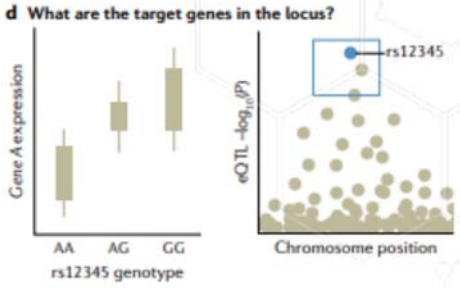
Functional follow-up of GWAS



ATAC
Methylation

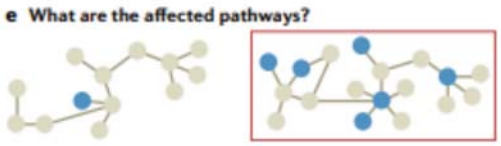


eQTL



3C, 4C, HiC

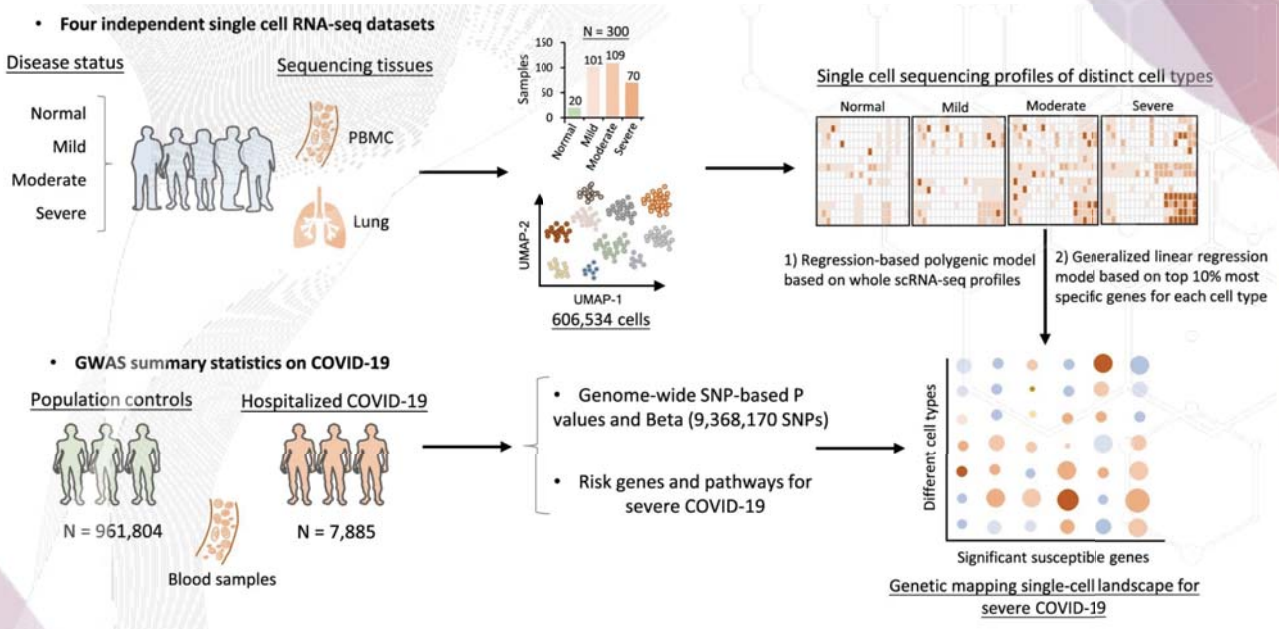
Pathway



Ex) IBD (inflammatory bowel disease) -> **12%** of risk loci as causal variant

(future **CRISPR** data?)

Integration of GWAS to scRNA-seq datasets



Cell type aware association is feasible to map DNA-RNA (genotype-phenotype) relationship

Moving the paradigm to phenotype cells better

?

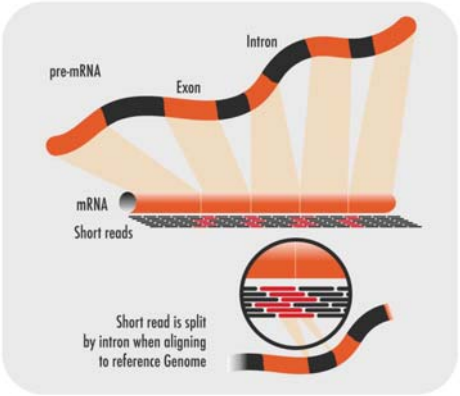
- **Genotype -> Phenotype**

- ~15 years of GWAS (DNA) was not sufficient (explained variance <20% for complex diseases)

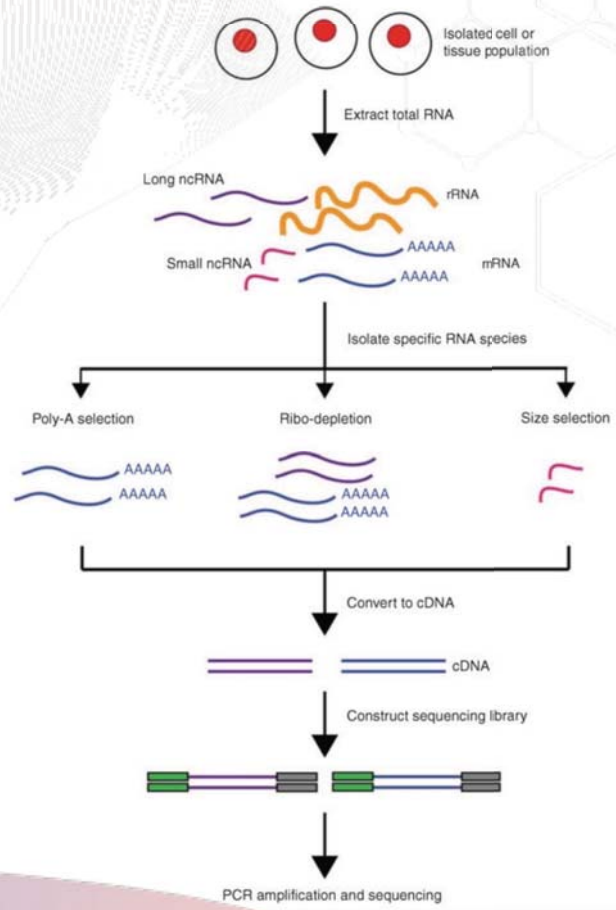
RNA-seq: a revolutionary tool for transcriptomics

- **For functional annotation, we measure 'Gene expression'**
- **Transcriptome** : The complete set of transcripts in a cell, and their quantity for a specific developmental stage or physiological condition

- Catalogue (mRNA, non-coding RNA and small RNAs), Structure (5' and 3' ands splicing patterns etc)

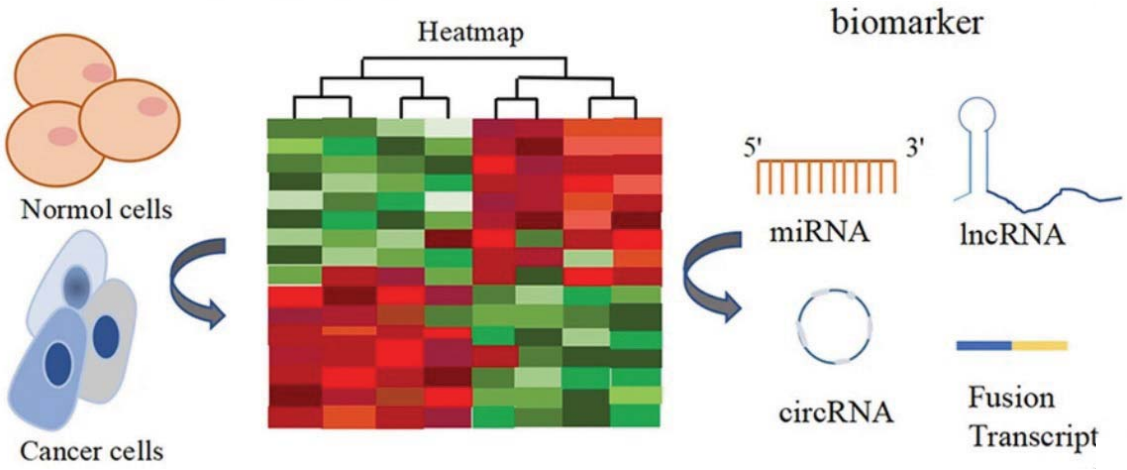


Overview of bulk RNA-seq



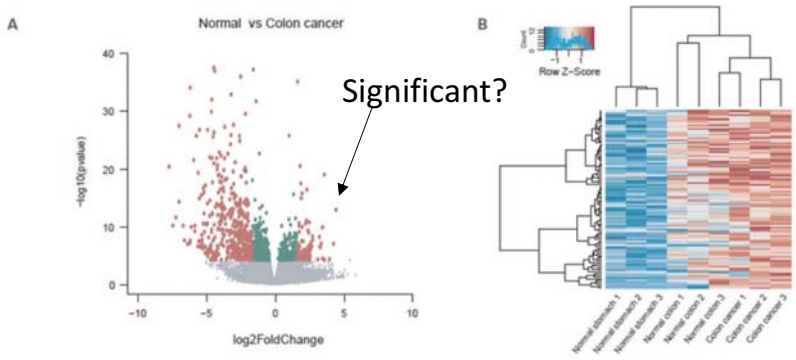
Various applications of bulk RNA-seq

1) Expression Profiling and **Differential expression (DE)** analysis



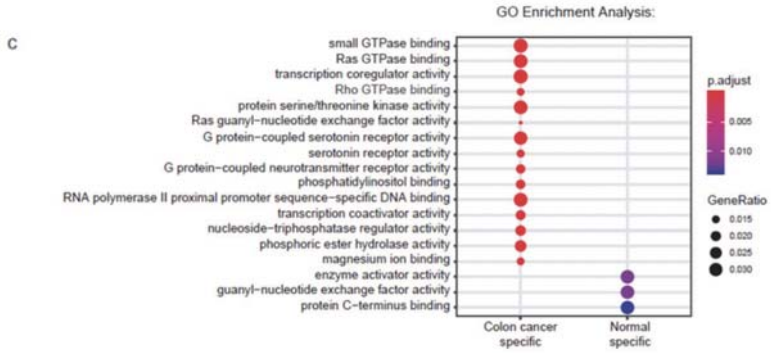
***(Differentially Expressed Gene, **DEG**) 차등 발현 유전자란 **두 실험 조건 하에서 샘플 집합의 유전자 발현량이 많이 차이나는 유전자 → 궁극적 질병 유전체 스터디의 목표**

DE analysis of cancer vs normal



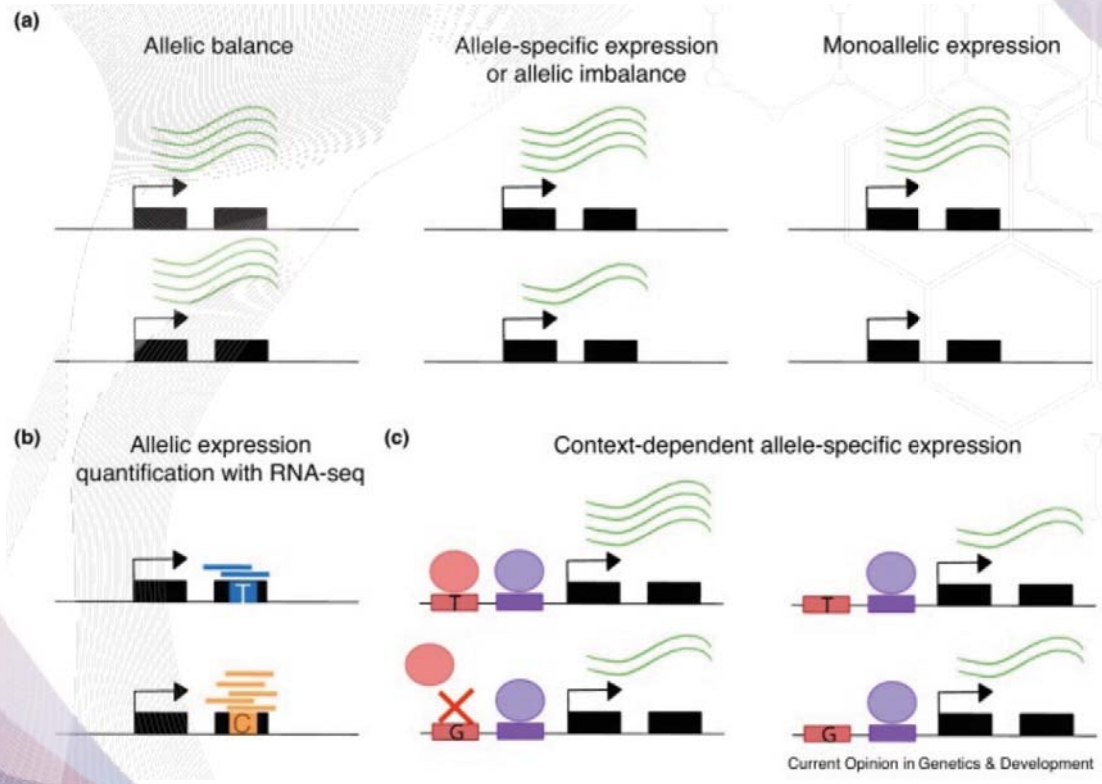
Fold-change volcano plot

Clustering heatmap

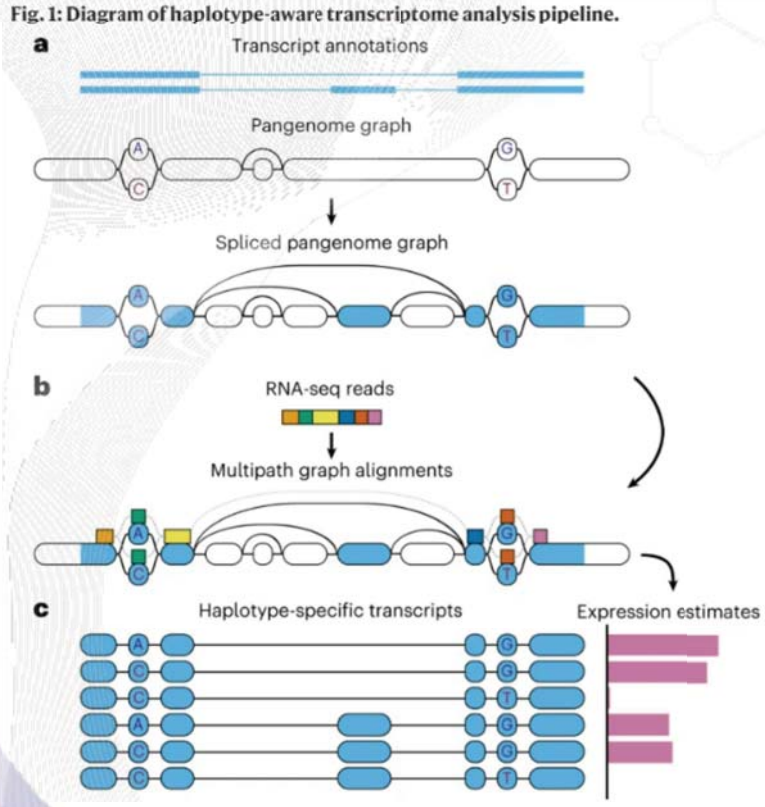


Gene set enrichment analysis

Allele-specific expression (ASE)



Haplotype-aware pantranscriptome



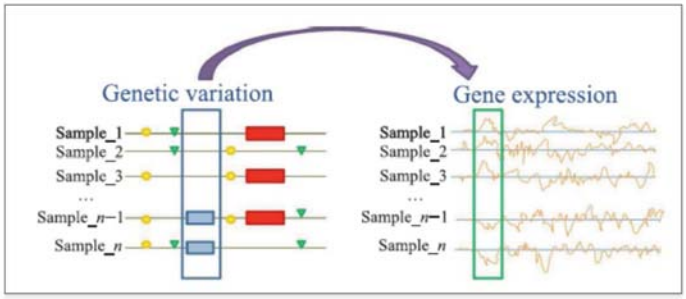
Pangenome -> 인류의 인종별 특징을 모두 취합한 reference genome임.

현재 phased variant (알려진 **haplotype block**) 에 대해서 분석됨

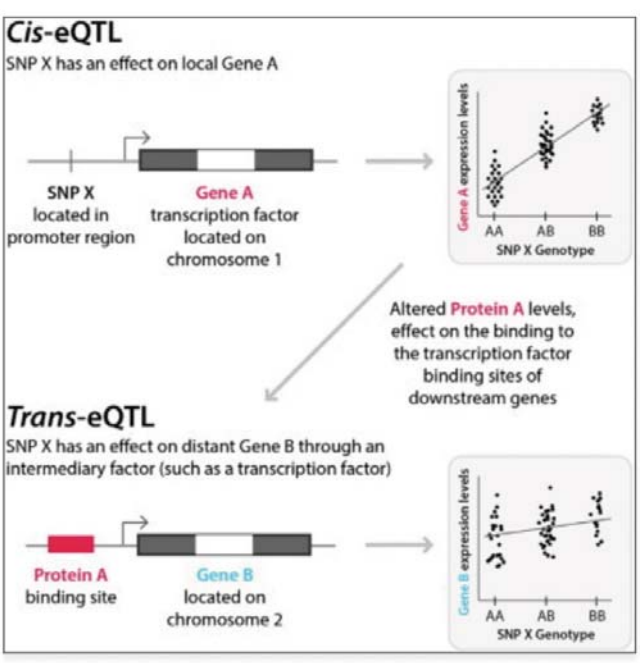
Long-read 시퀀싱 기술이 발전할수록 resolution증가

DEG (differentially expressed gene) 분석의 패러다임도 알려진**transcriptome(전사체) 뿐만 아니라 새로운 것들에 대한 재정의 필요해질 것임.**

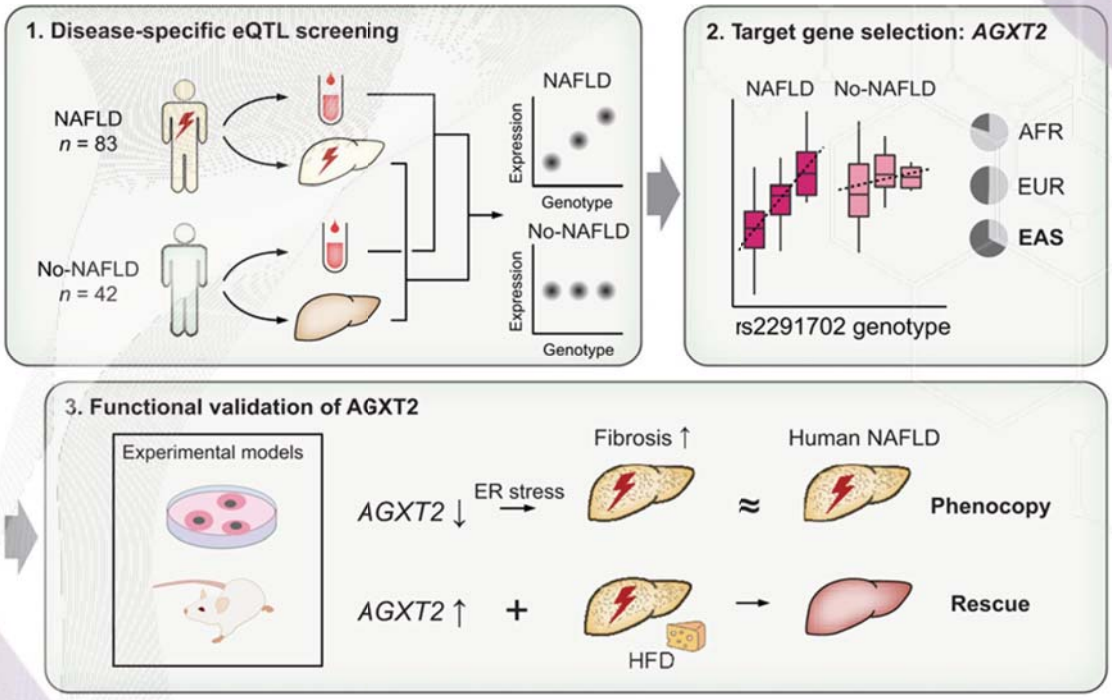
Expression quantitative trait loci (eQTL)



그동안의 많은 **GWAS** (genome-wide association study, DNA 염기서열의 변이 (SNP) 와 질병 유/무의 상관성 분석) Hit들이 **non-coding SNP (코딩 영역x 해석 어려움) -> 이 중 일부가 유전자 발현에 영향을 미칠 수 있다 (not protein), 새로운 질병/형질 연구의 가능성**



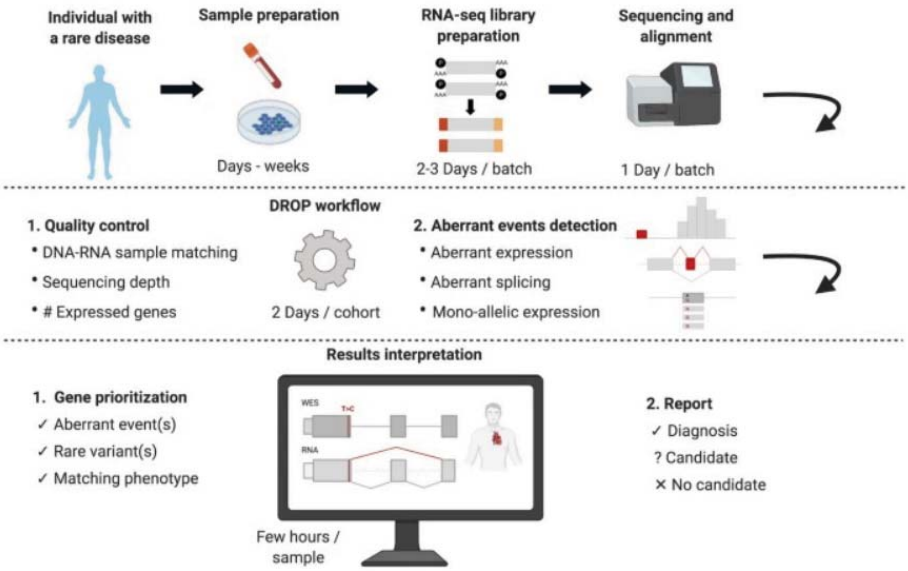
eQTL in liver disease



NAFLD (Non Alcoholic Fatty Liver Disease, 비알코올성지방간)

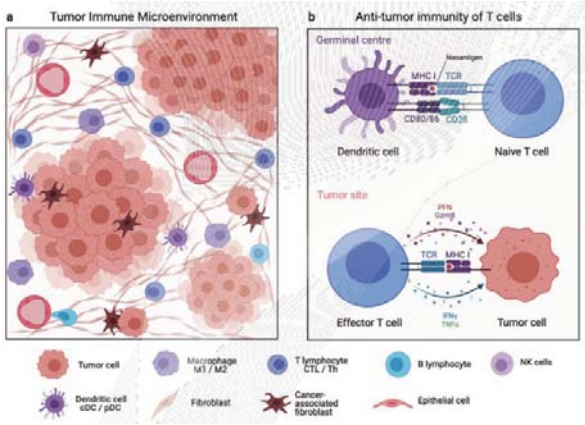
Clinical diagnostics of Mendalian diseases using RNA-seq

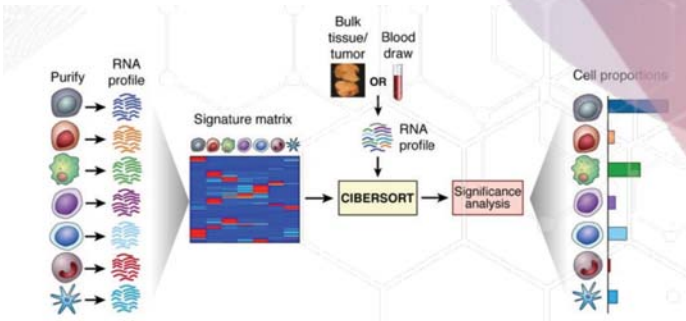
WES/RNA-seq of 303 people with Mendalian disease
(rare: 3~5% population, **80% of them are driven by genetic cause**)

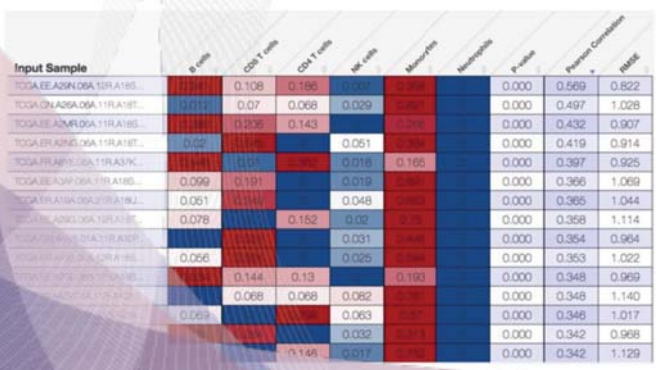


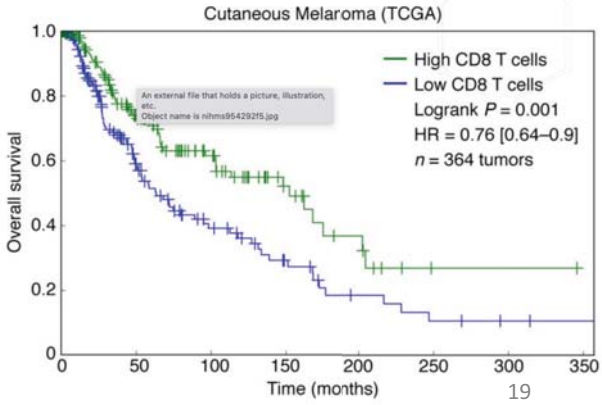
Able to genetically diagnosed **16% of inconclusive case from WES**

Tumor-immune-composition analysis

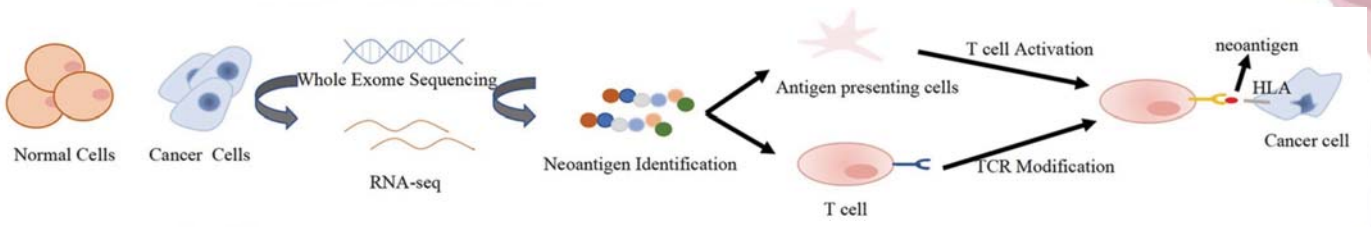








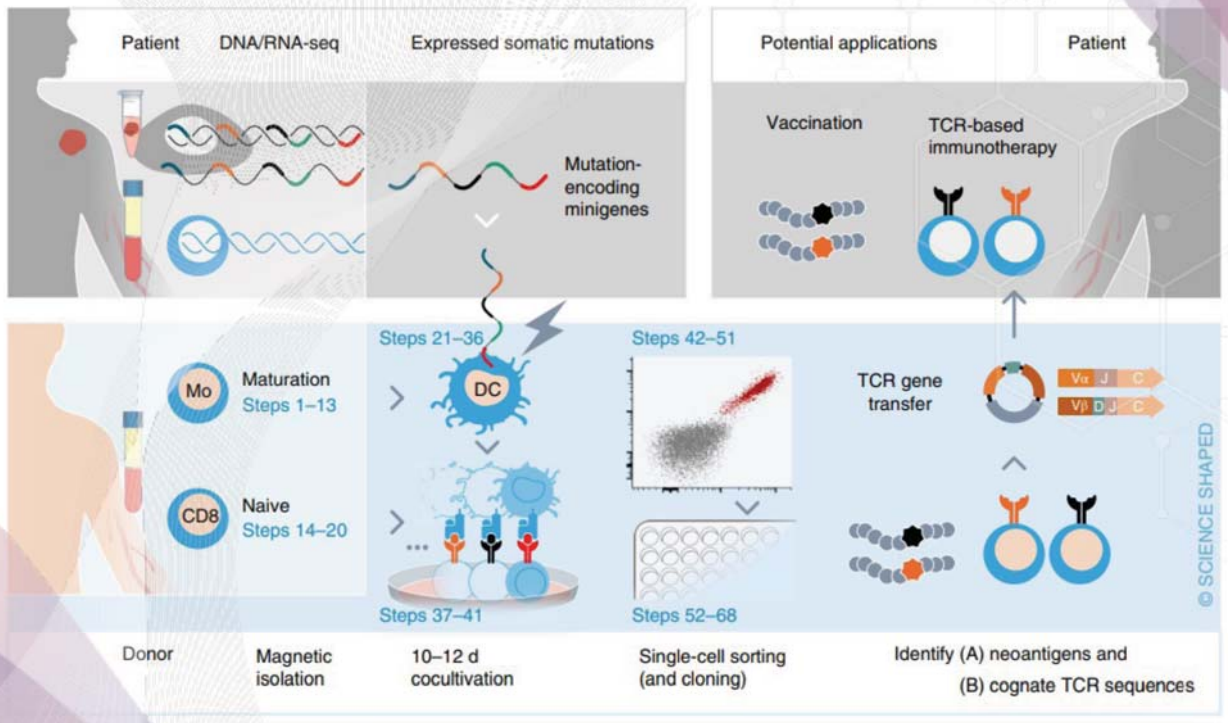
Neoantigen profiling by RNA-seq and TCR modification targeted neoantigens



1.**Whole exome sequencing to identify the mutations** by using different computational and mutation calling tools

2.**RNA-seq analysis to focus specifically on the expressed mutations** and identification of neoepitopes (신항원) in silico with computational algorithms for MHC class I and class II binding

Identifying neoantigen reactive T cells



유전자 편집 기술(ex CRISPR)의 발달로 allogenic donor에서 CAR-T, TCR-T같은 세포치료제가 획기적으로 발달하고 있다.

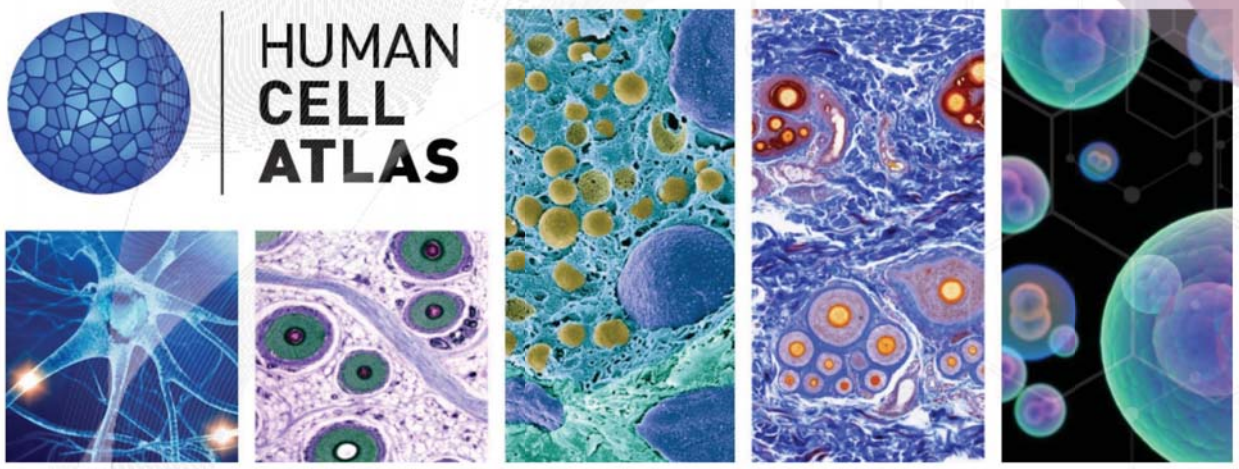
We need better resolution for phenotyping cells

?

- Genotype -> **RNA -> Phenotype**
- Bullk RNA-seq is routinely done in clinical labs along with GWAS (SNP) information
- Cancer heterogeneity not solved due to mixed signal (normal and tumor cells)

단일세포 오믹스의 필요성!!!

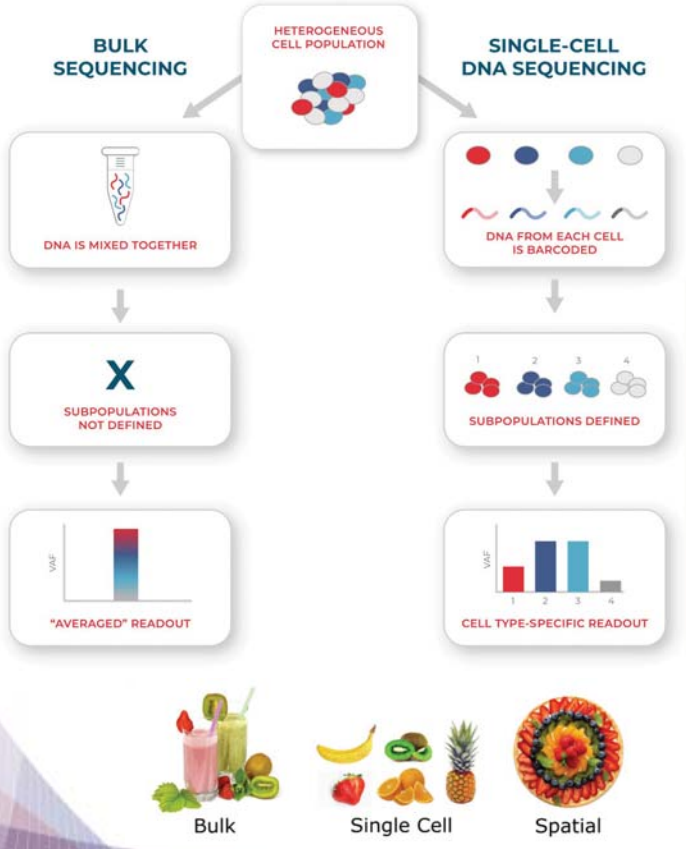
Human Cell Atlas (HCA) project



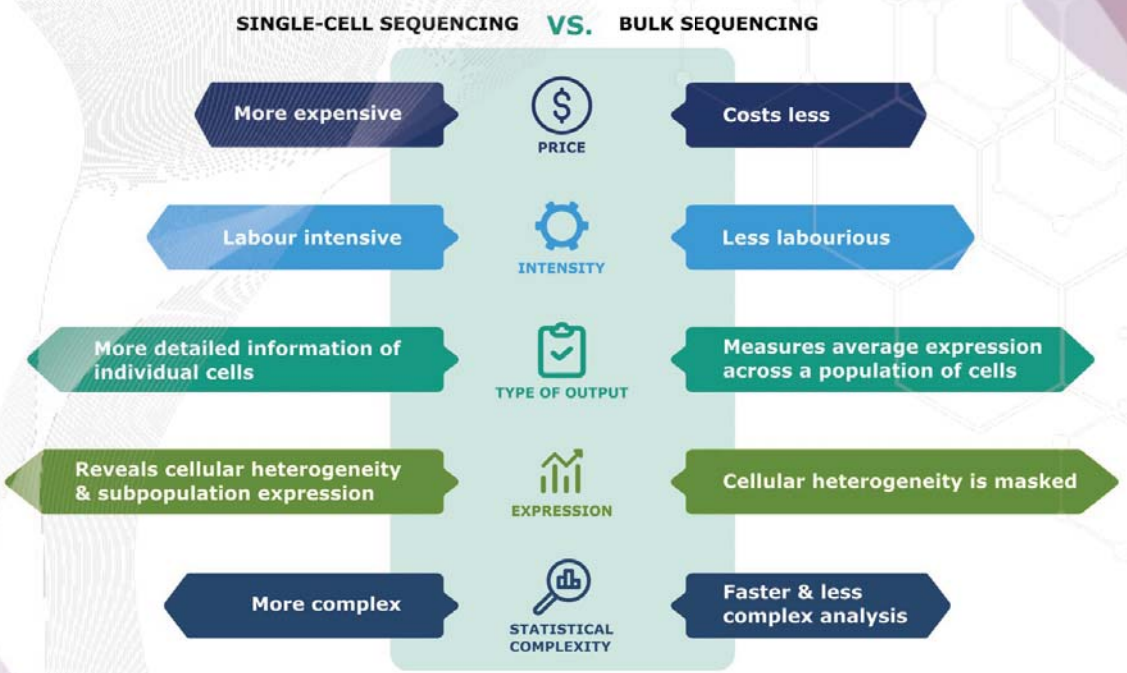
Bulk studies -> Single Cell Genomics (Catalogue of all cell types in the body from healthy and diseased individuals)

2016 결성, 초기목표: 모든 **단일세포의 전사체 (Transcriptome) 지도**

Why we care about single cell?

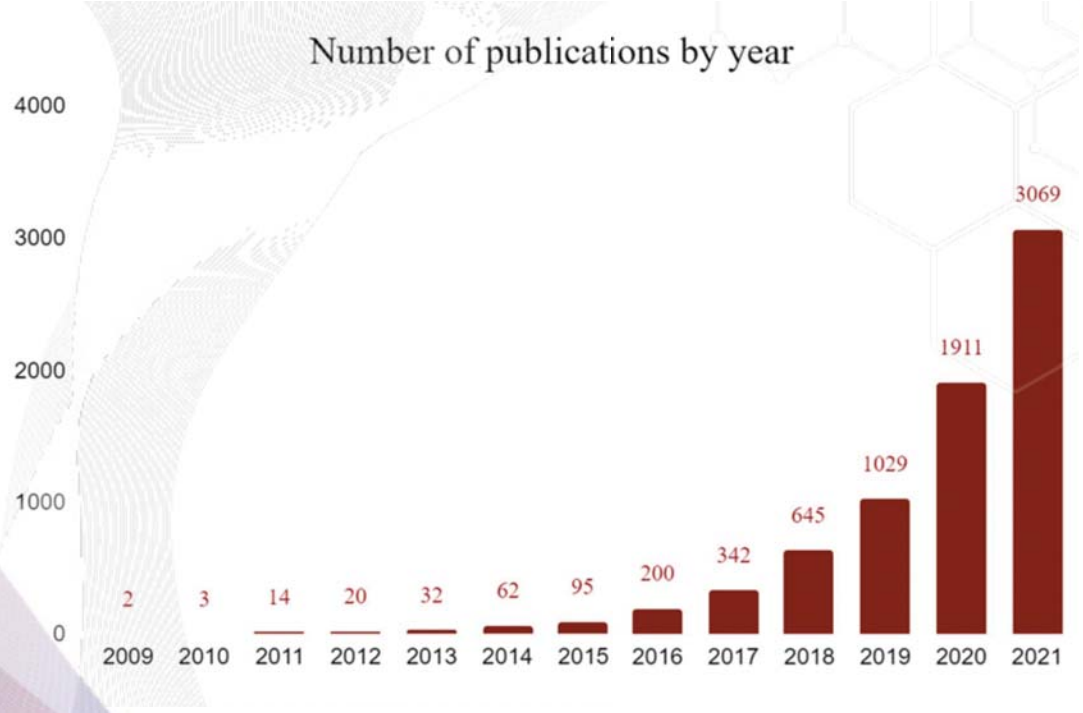


Bulk vs Single cell sequencing

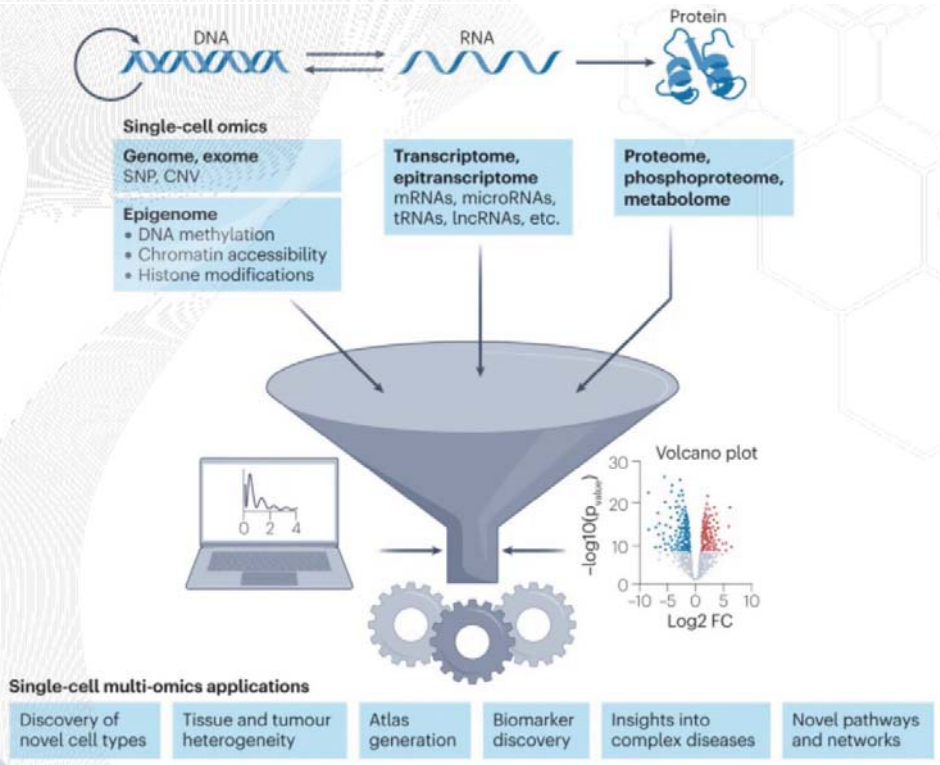


Heterogeneity : 이질성, 단일세포에서 중요한 개념임

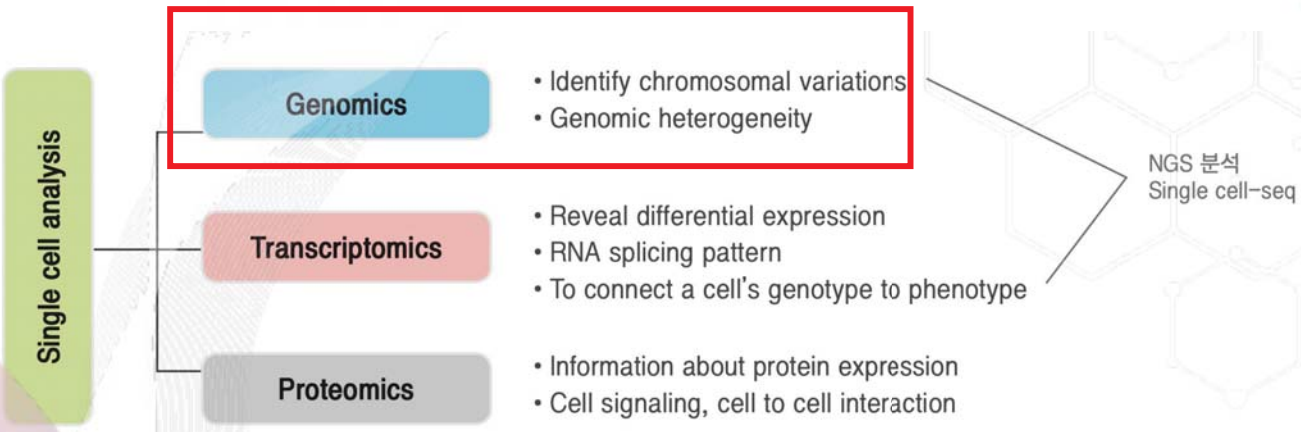
Increasing popularity of single cell RNA sequencing (scRNA-seq)



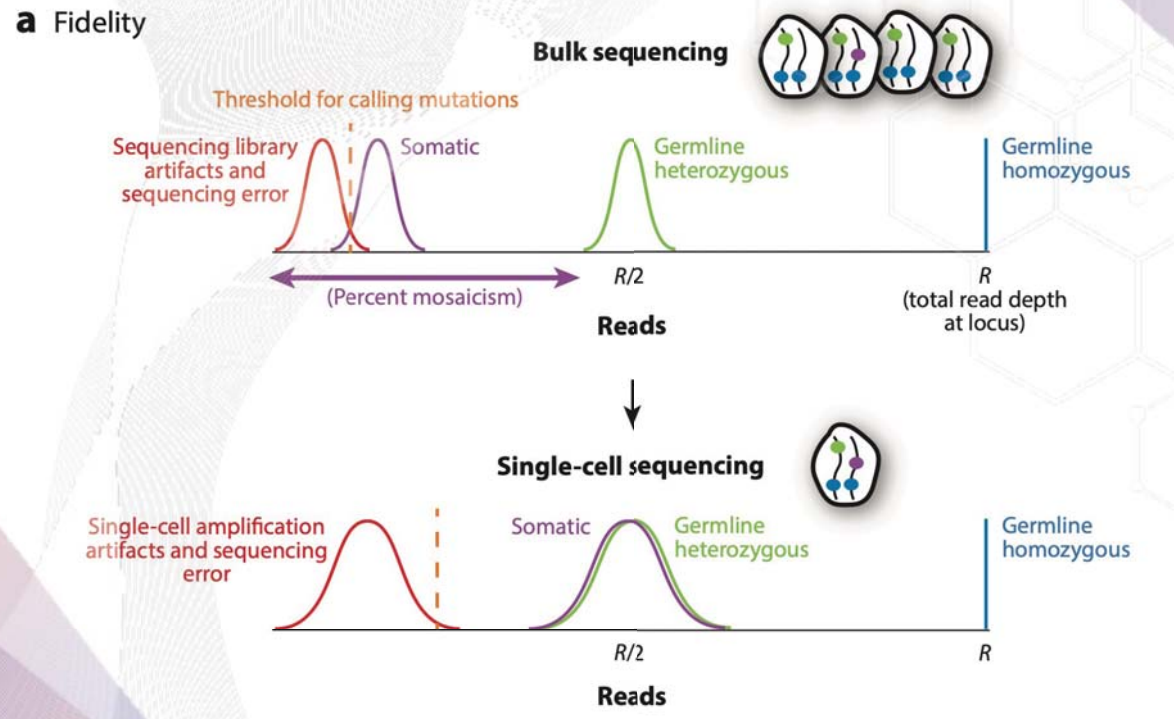
Single-omics to multi-omics



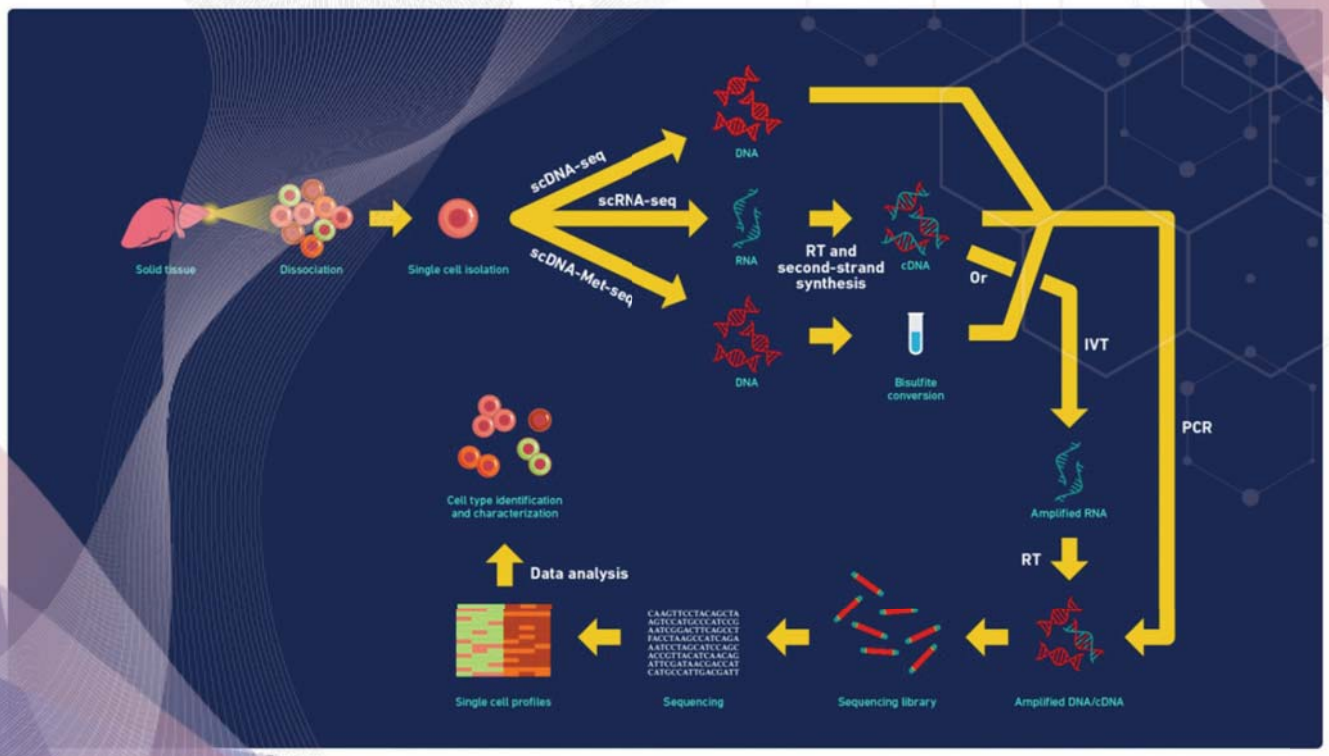
Single-cell analysis platforms



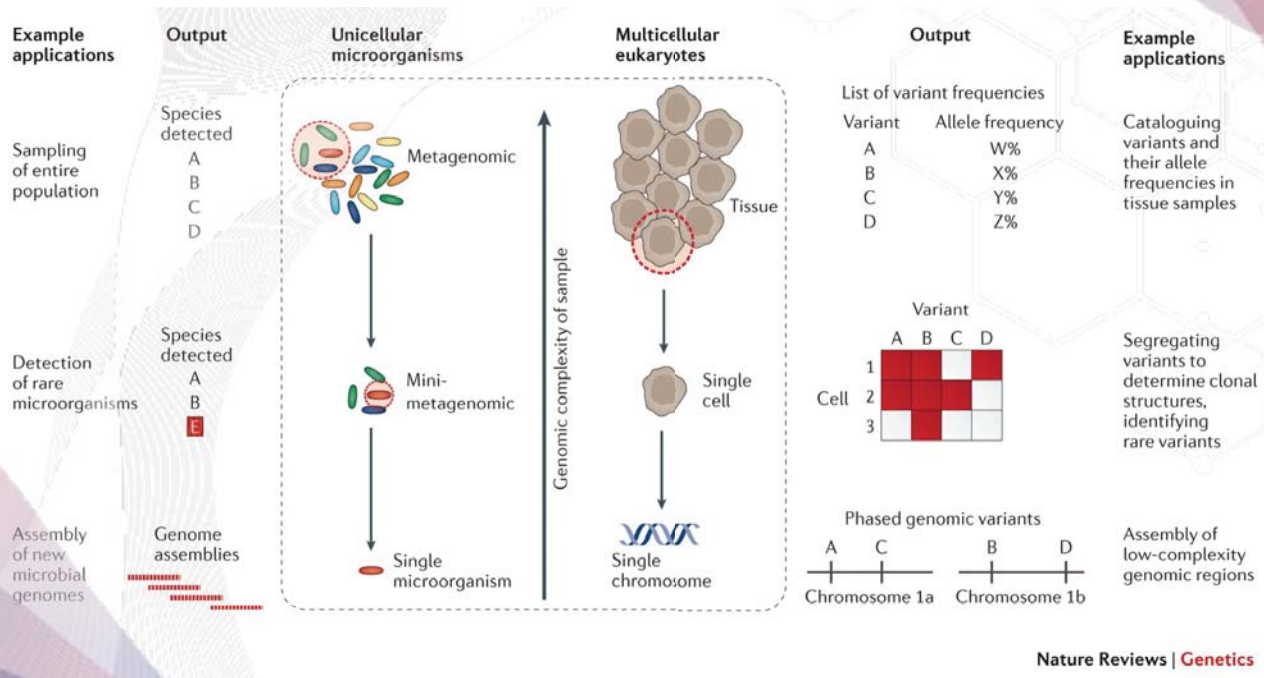
Fidelity of single cell genomics



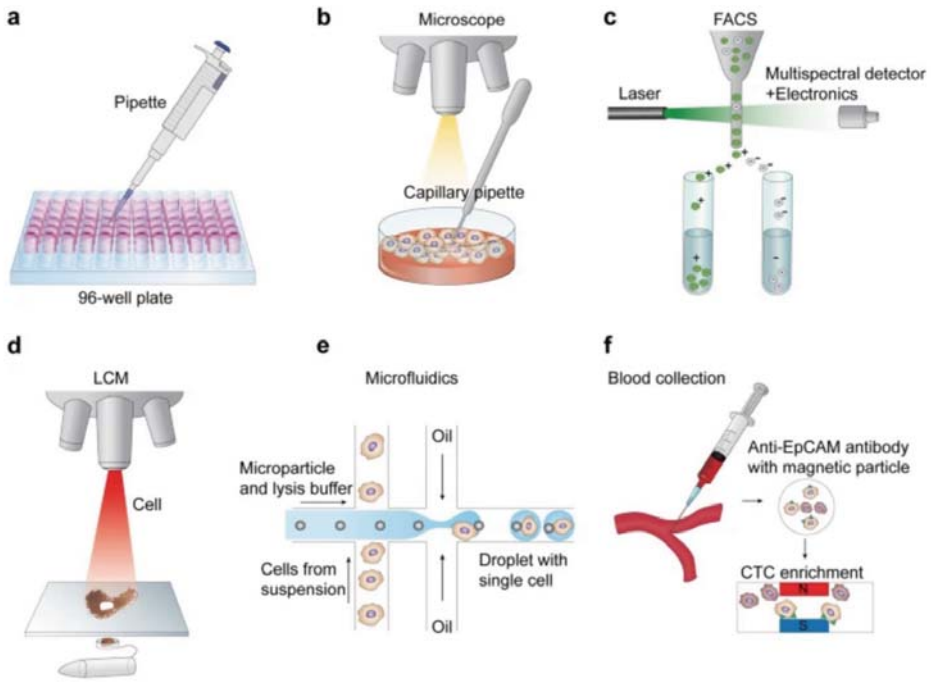
Single cell sequencing workflow



Opportunities enabled by single-cell genome sequencing



Single-cell isolation methods

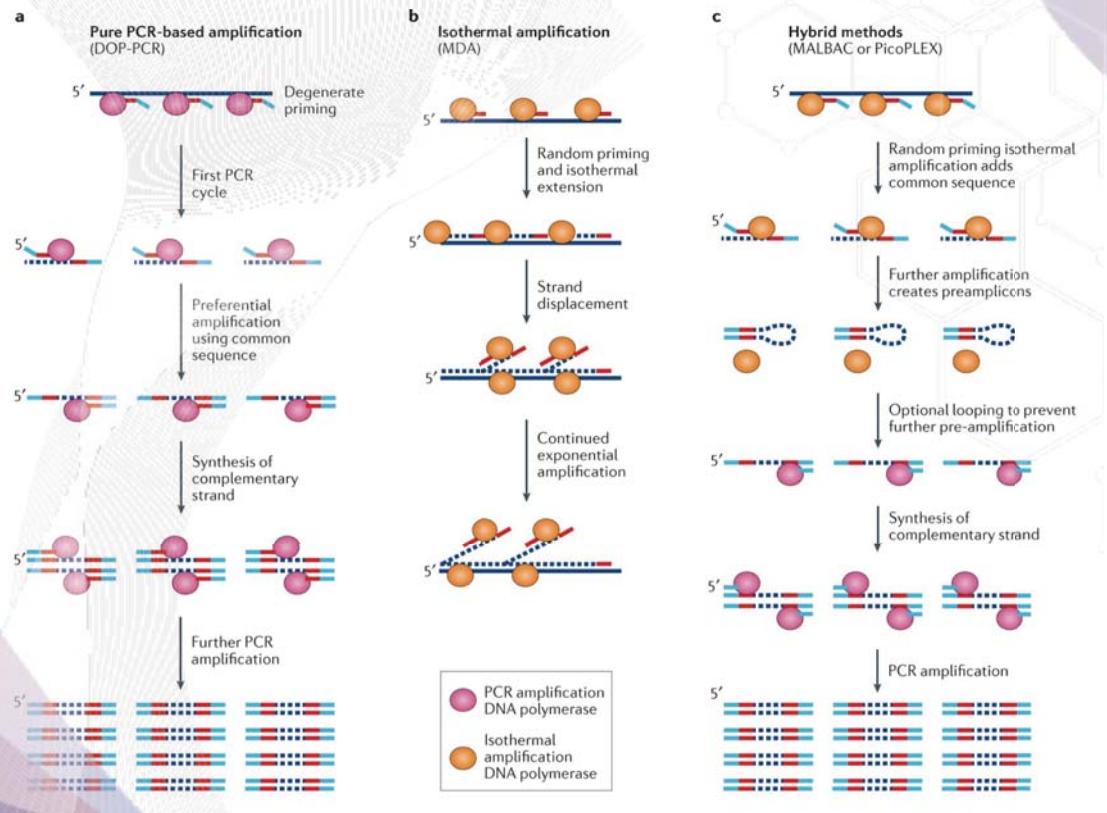


a. Limiting dilution b. tweezers c. FACS d. LCM
e. Microfluidics f. Bead based capture

How to amplify single-cell genomes?

- 10 yrs of progress for **WGA** (whole genome amplification) to reduce artefacts, amplification bias
- Merely 6pg in diploid DNA
- Needs to be amplified >100 times to generate sequencing library and analysis
- Cover as much of the genome (3 Billion) as possible without bias

Three main WGA methods



Pros and Cons

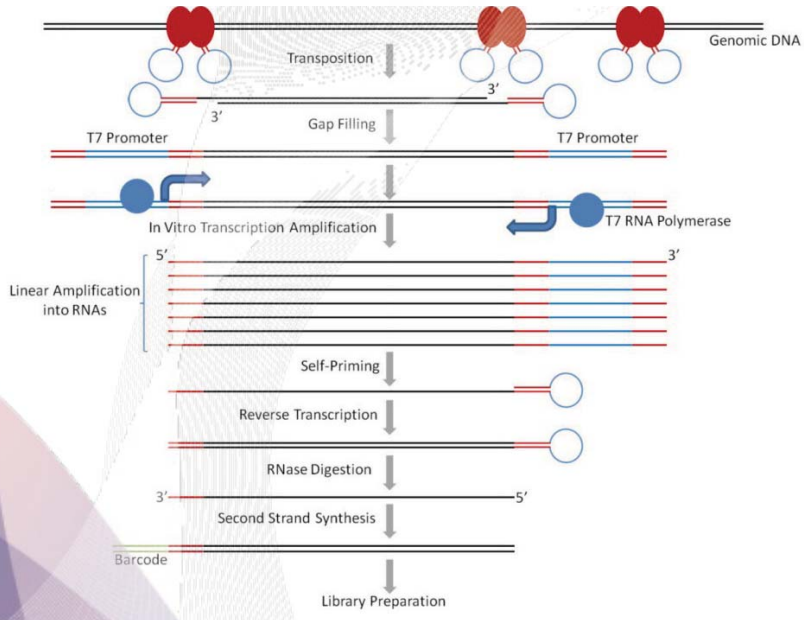
	PCR-based (DOP-PCR)	Isothermal (MDA)	Hybrid (MALBAC or PicoPLEX)
False-negative rate (coverage and allelic dropout)	High	Low	Intermediate
Non-uniformity	Low	High	Low
False-positive rate (amplification error rate)	High	Low	Intermediate

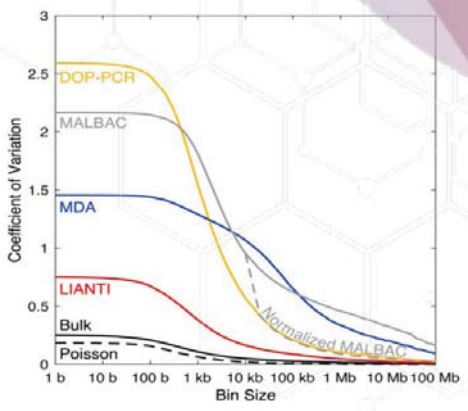
DOP-PCR: Degenerate oligonucleotide Primed

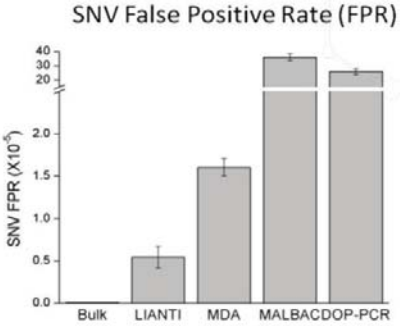
MALBAC : Multiple annealing and looping-based amplification cycles

MDA : Multiple displacement amplification

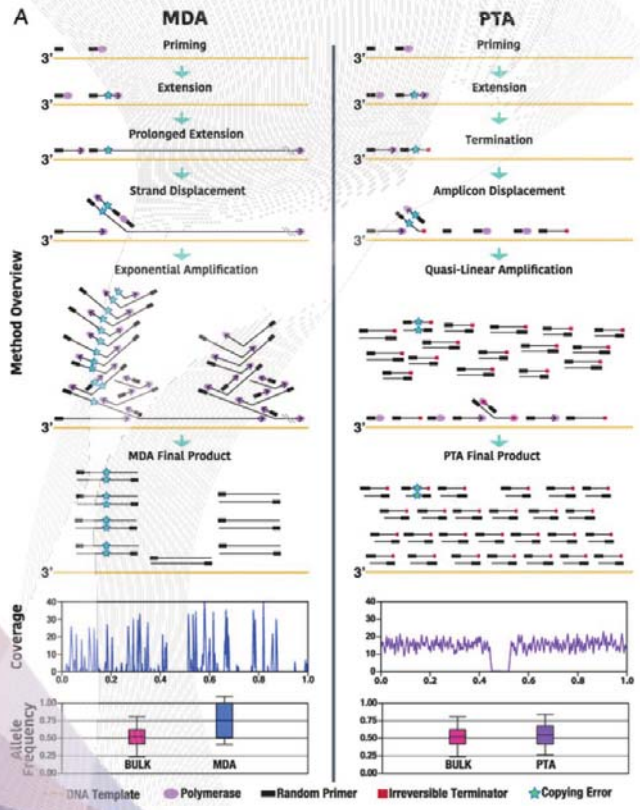
LIANTI (LInear AmplicatioN via Transposon Insertion)

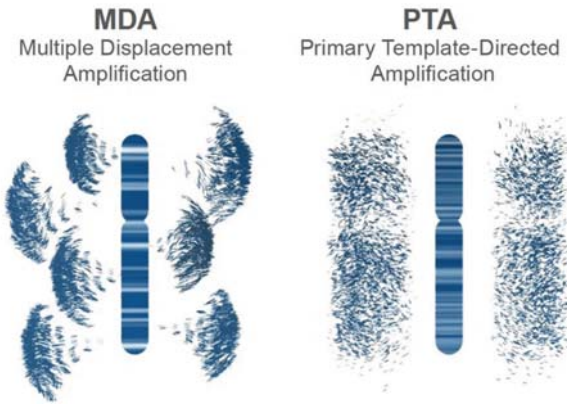






PTA (Primary Template directed Amplification)

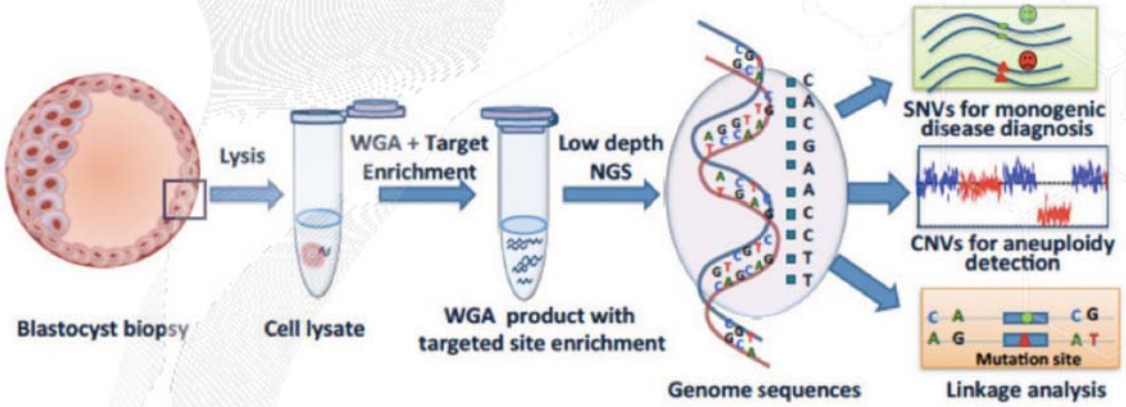




Future directions

- Long-read sequencing, CNV+SV detection
  - Brain diseases + others
- More efficient droplet-based WGA needed
  - Cost prohibitive (~only few hundred cells)

MALBAC Babies



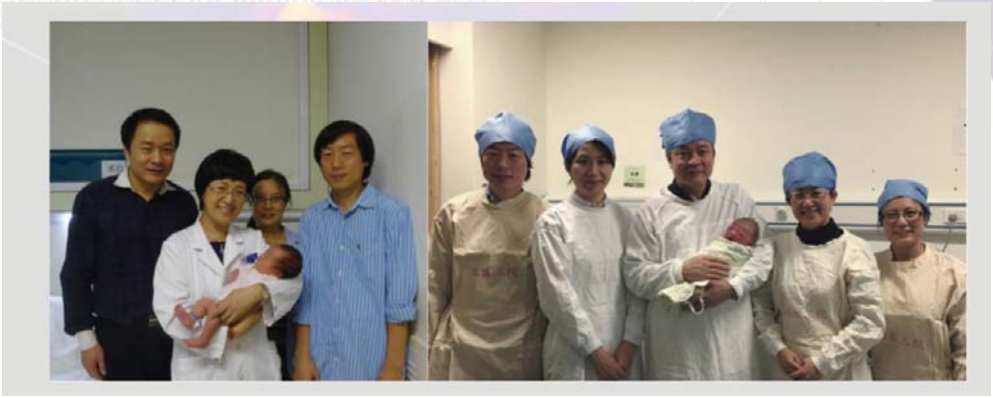
In vitro fertilization (IVF)

preimplantation genetic screening (PGS)

preimplantation genetic diagnosis (PGD)

외배엽생검을 통해 WGA 분석 후 착상

New opportunity for are genetic disorders



The first IVF baby from Sunny Xie's (Peking University)

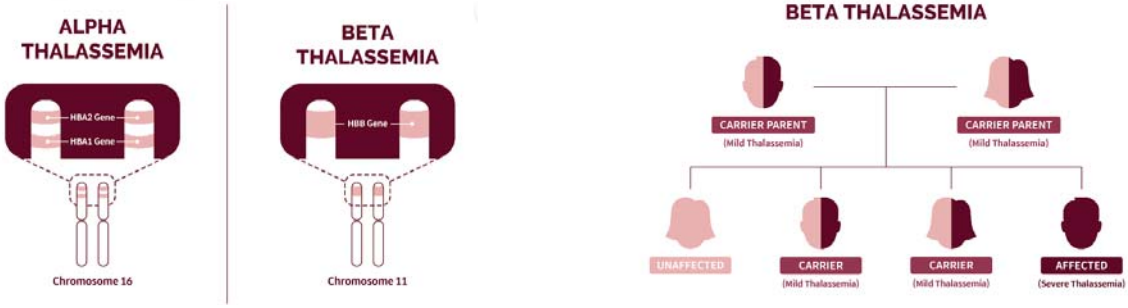
Case 1. Monogenic disease (husband, autosomal dominant disorder,hereditary multiple exostoses (HME, 유전적 다발성 외골증), c.233delC (frameshift point mutation in EXT2 gene)

Case 2. Monogenic disease (wife, X-linked disorder, hypohidrotic ectodermal dysplasia (HED, 외배엽이형성증), c.T1085G at EDA1 gene)

→ No mutation or no copy number variation cells were selected for transfer

지중해빈혈 (Thalassemia)

- 헤모글로빈 폴리펩타이드 사슬 합성저하로 산소운반 혈색소감소
- 이형접합체는 경증/정상인과 유사, 동형접합이라도 아이는 출생시 정상 소견 (베타사슬이 없는 태아, Hbf)
- 2500명당 1명꼴

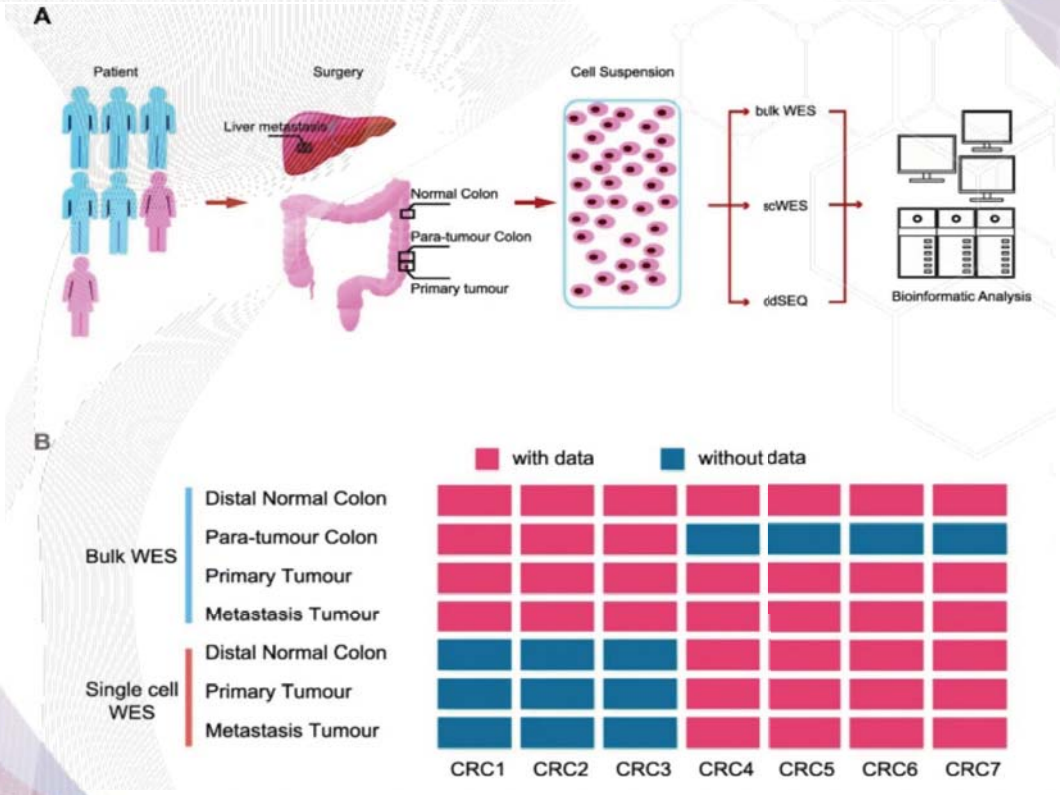


- **복합성 유전질환**이기에 MALBAC으로 진단은 가능하나 아직 교정 후 정상인이 태어나는 보고는 없음.
- iPSC와 같은 줄기세포 이용 조혈모세포 분화 방법이 현재 진행중

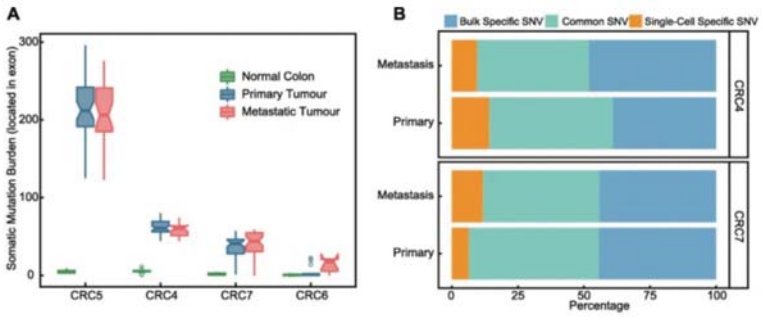
Single cell Whole genome vs exome

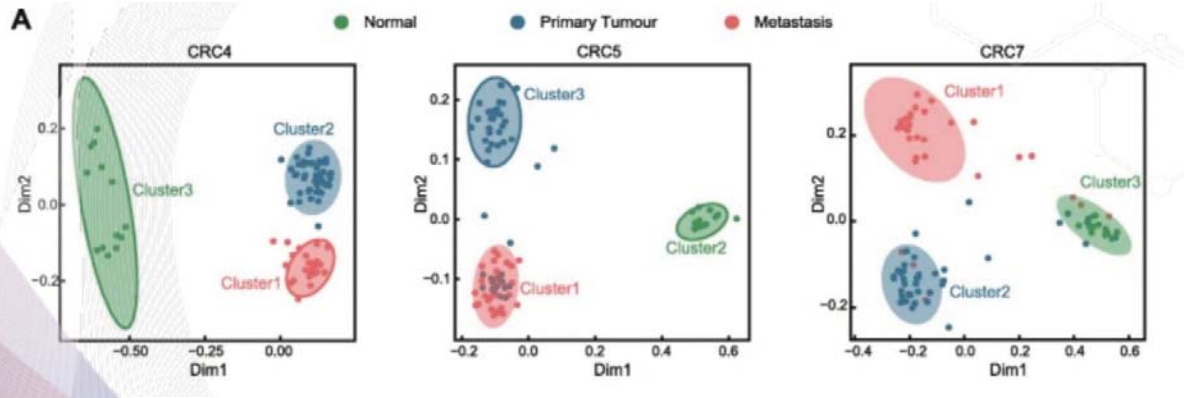
- **Whole-genome** :
  - More uniform amplification, suitable for detecting **SNV** (단일염기변이, single nucleotide variant), **CNV** (염색체 수 이상, copy number variation), **SV**(구조이상변이)
  - 30-fold more expensive than exome (only ~2% of the genome)
- **아직까지 WGS 가격이 hurdle임**
- **Droplet기반 기술은 아직 unstable**

Single cell Whole-exome seq for cancer

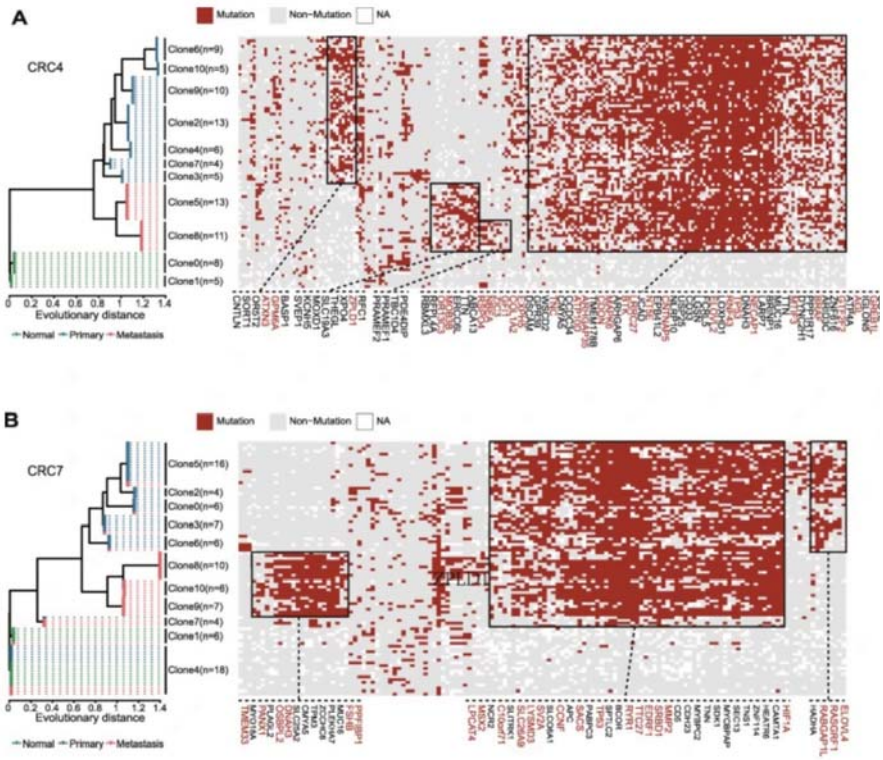


Mutation burden detection and clustering

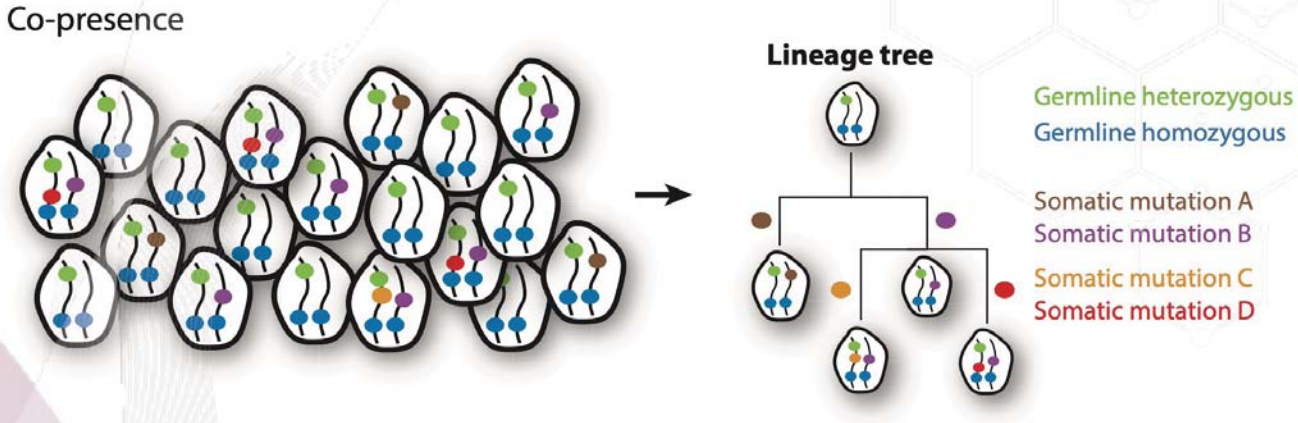




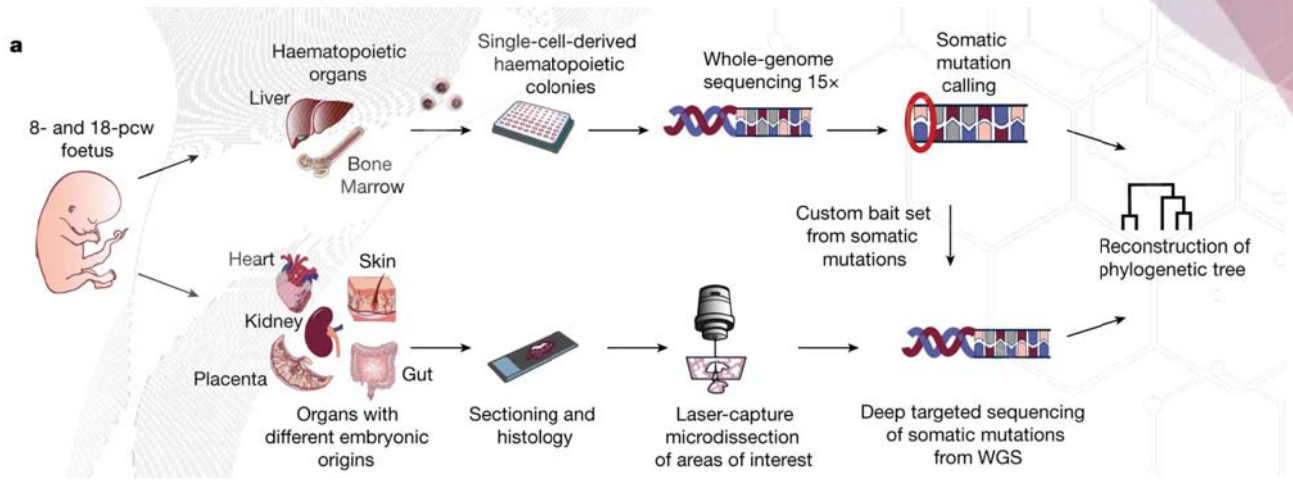
Sub-clonal analysis using single cell SNVs



Lineage tracing of human development through somatic mutations



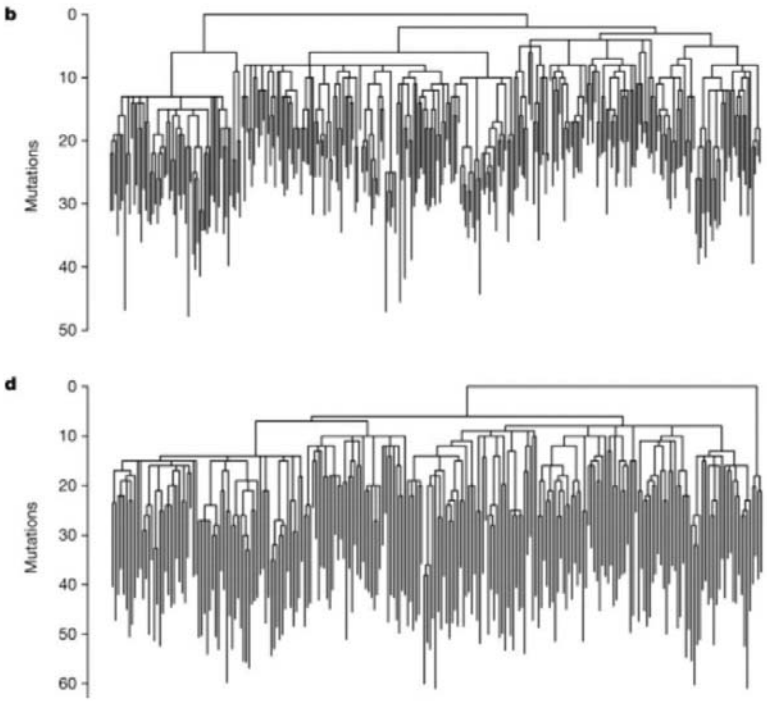
Lineage tracing experimental workflow



원시 조혈모세포의 유래?

최초의 조혈모세포 AGM에서 유래 -> 간에서 크게 증식 -> 출생 전 골수에 정착

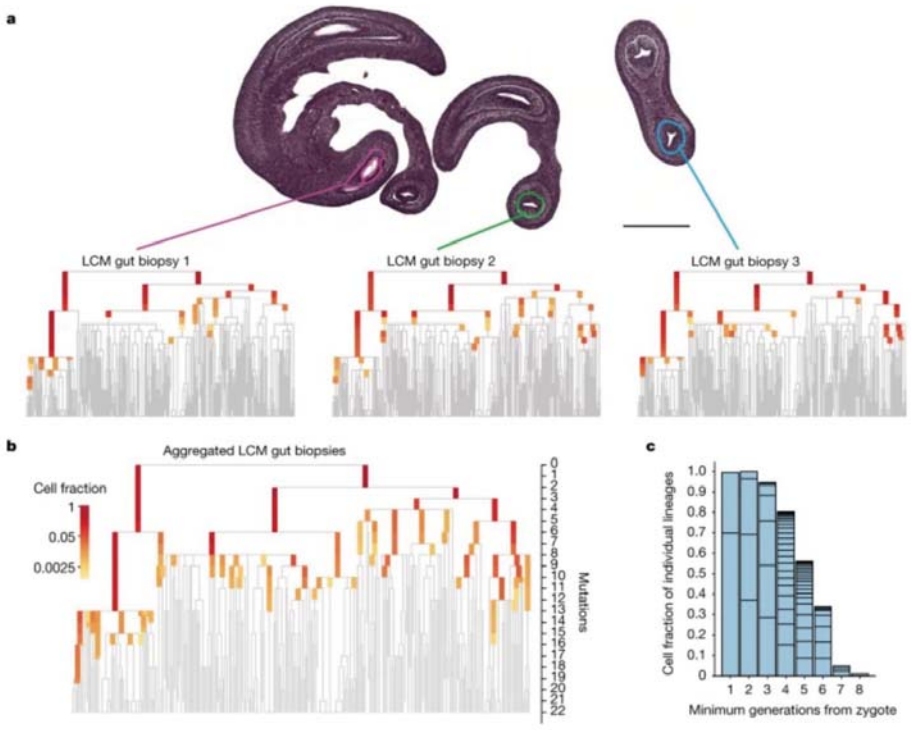
Phylogeny of 277 single cell 8-pcs liver HSPC



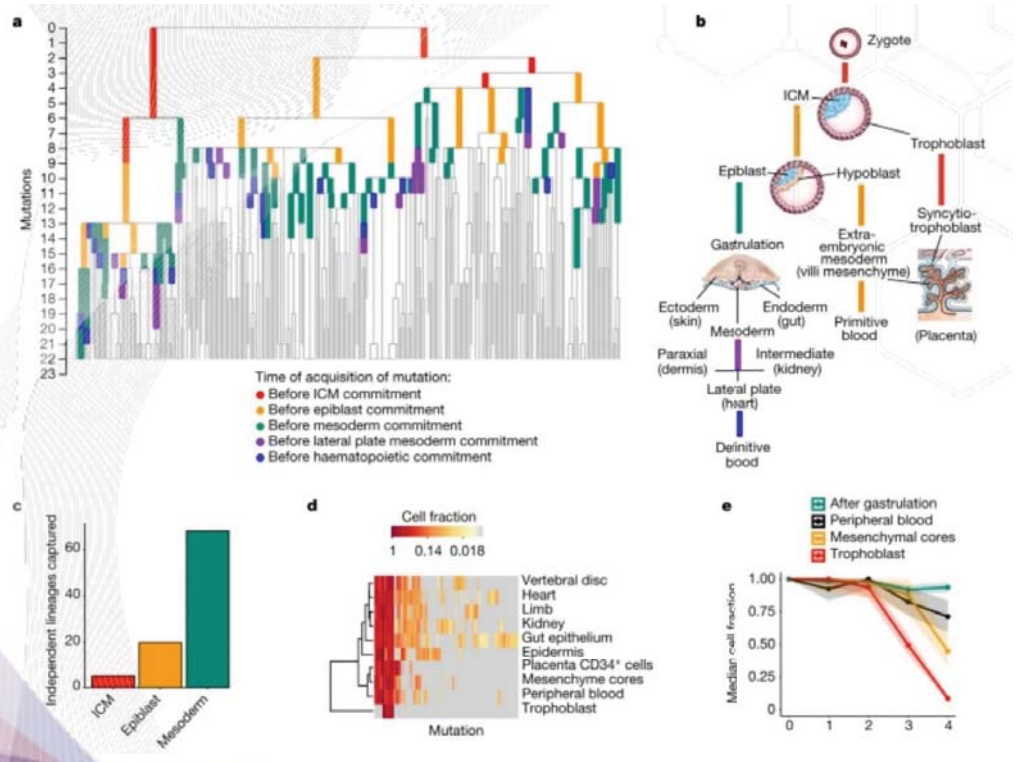
Pcw: post conception week

HSPC : haematopoetic stem and progenitor cell (조혈 줄기/전구 세포)

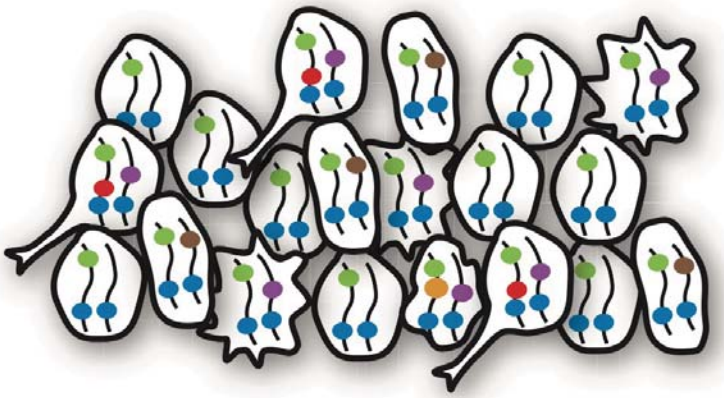
Reconstructing lineage divergence



Timing of divergence of lineages during development



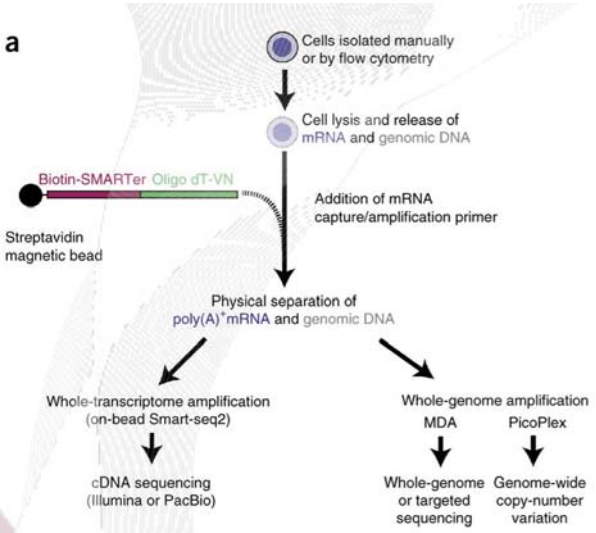
Phenotype association

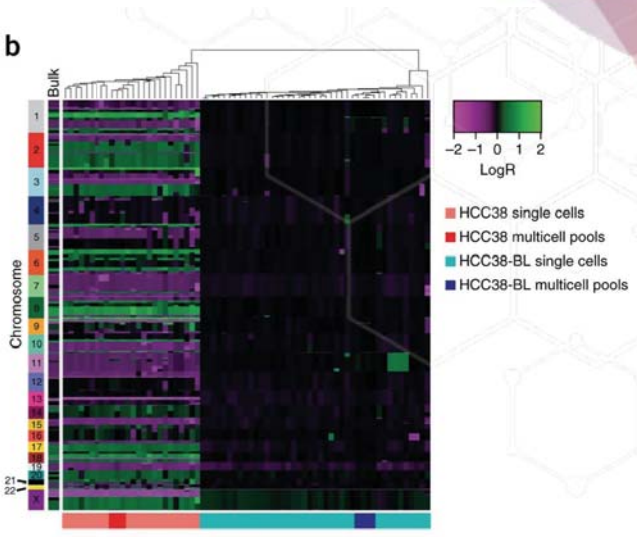


Somatic mutations	Cell type
●	Cell type 1
●	Progenitor of cell types 2 and 3
● + ●	Cell type 2
● + ●	Cell type 3

How do we associate DNA & RNA (phenotype) information?

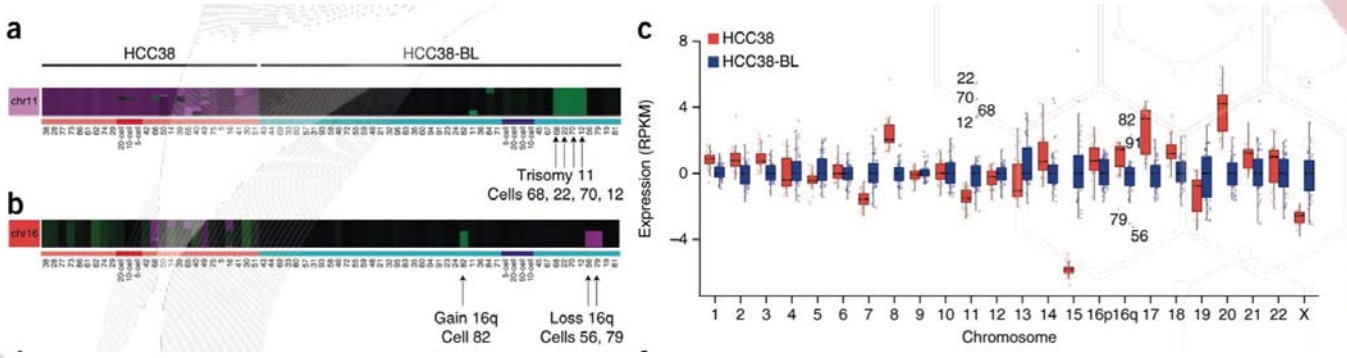
Single cell genome and transcriptome sequencing (G&T-seq)





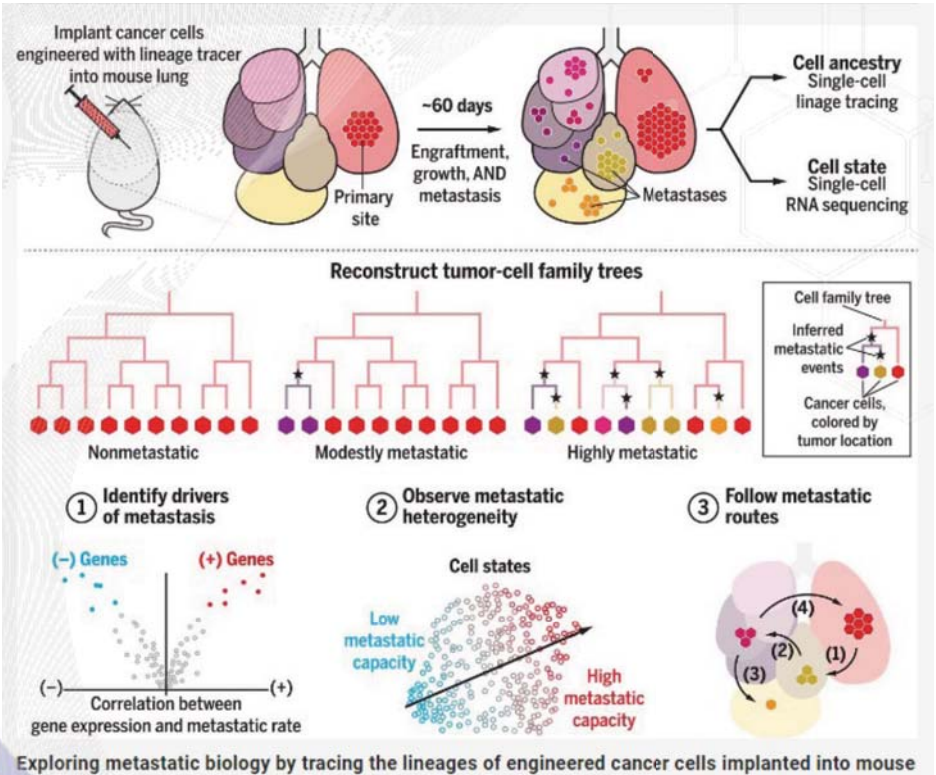
Physical isolation of DNA & RNA within a same **'single-cell'**

Simultaneous detection of chromosomal aneuploidy and gene expression



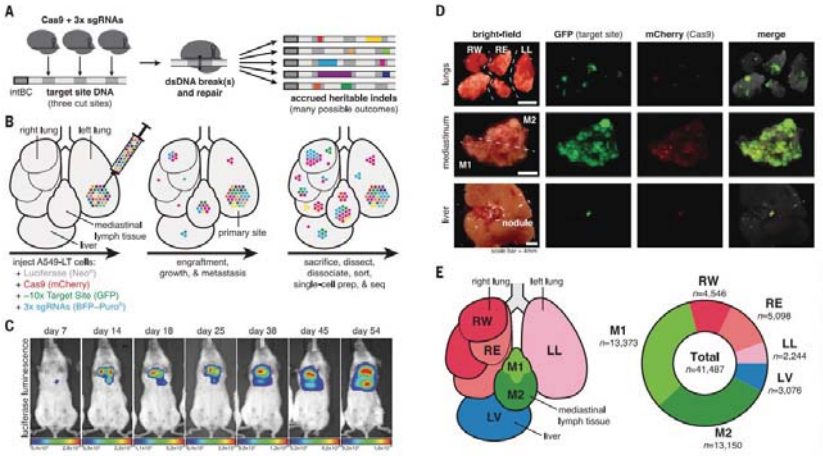
These data show that (sub)chromosomal copy number in a single cell is mostly positively correlated with gene expression in that cell.

Lineage tracing in engineered cancer cells



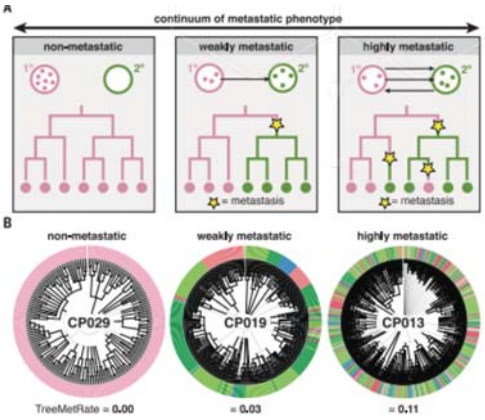
Exploring metastatic biology by tracing the lineages of engineered cancer cells implanted into mouse lungs.

CRISPR edited lineage tracing



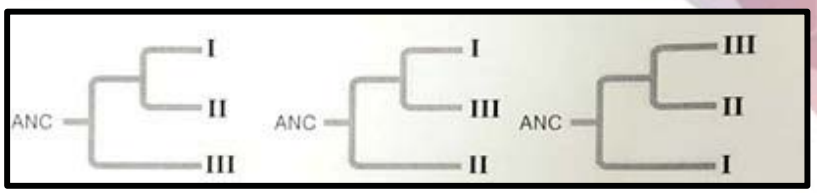
A549 cell line engineered to express :CRISPR, guide, target site

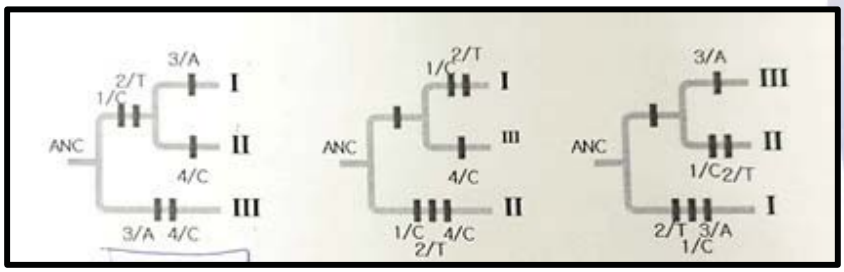
Barcode+transcript (DNA + RNA)
information to construct lineage tree



Maximum parsimony

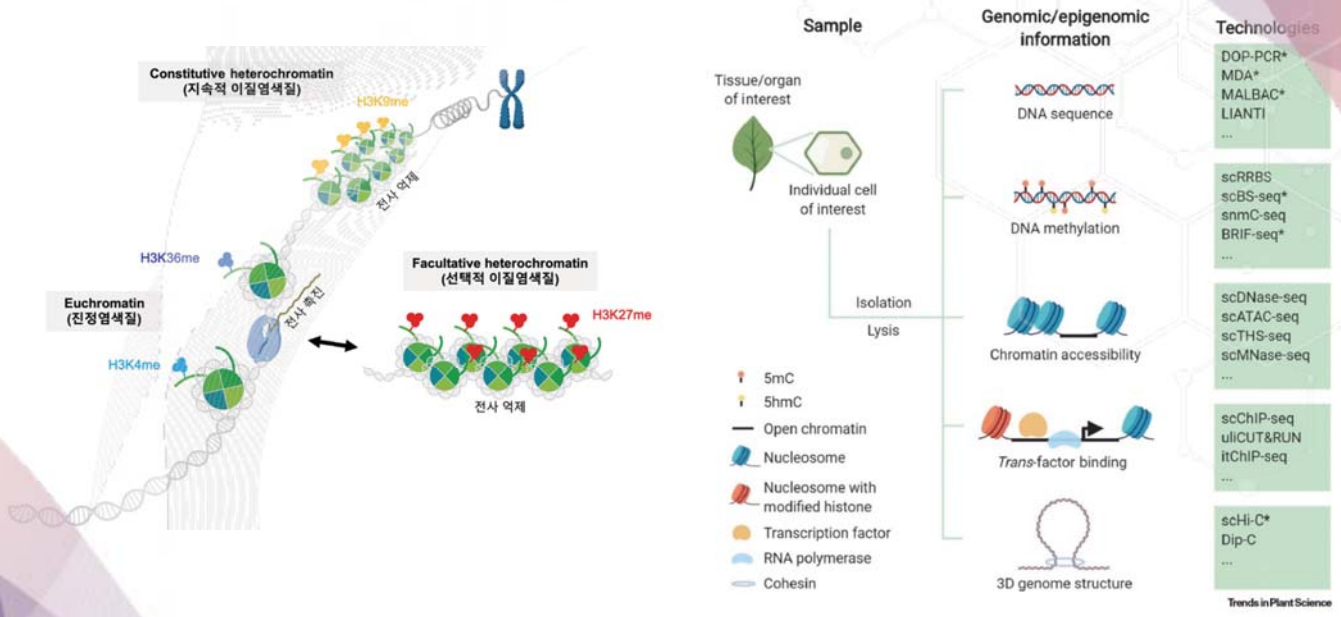
	Site 1	2	3	4
Species I	C	T	A	T
Species II	C	T	T	C
Species III	A	G	A	C
Ancestral state	A	G	T	T





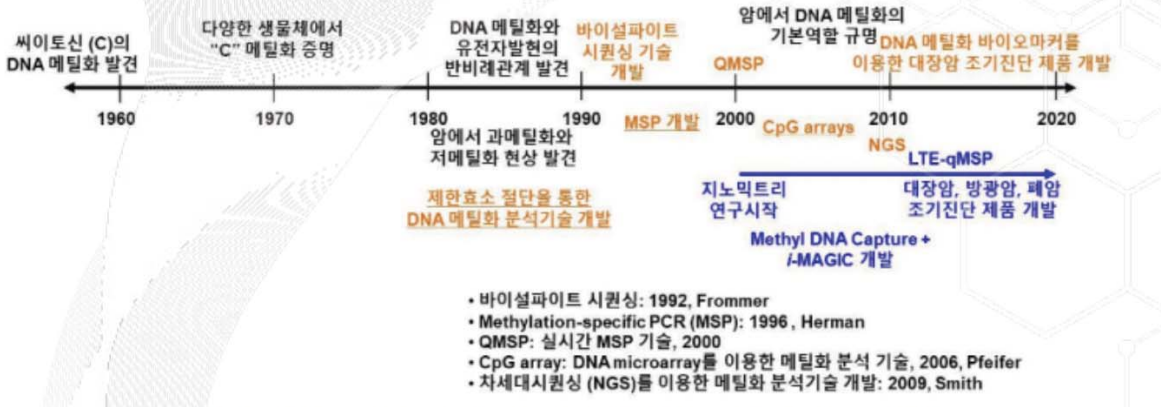
진화가 항상 변화 단계의 수를 최소화하는 방향으로 일어난다는 가정 하에 수행
계통 분류학에서 많이 쓰이며 evolutionary tree를 적용하는 모든 케이스에 사용

Epigenetics in single-cell genomics



후성유전(Epigenetics)은 **DNA 염기서열의 변화가 아닌** DNA의 메틸화, RNA의 메틸화 그리고 히스톤 단백질의 번역 후 변형(Post-translational modification; PTM) 에 의한 유전자 발현의 변화를 의미

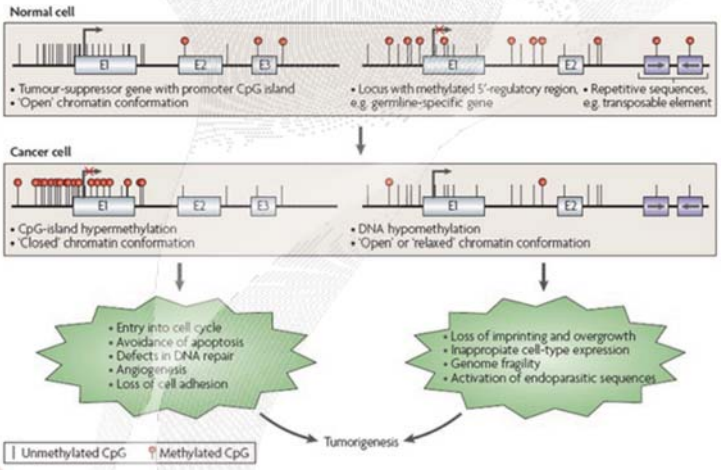
DNA methylation and cancer diagnosis

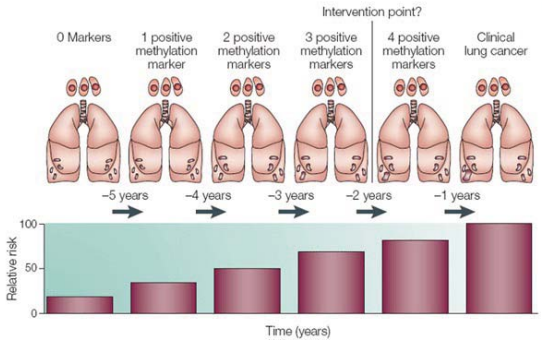


[DNA 메틸화 기술 개발 및 암 조기진단 기술 개발의 역사]

1979 robin holiday의해 메틸화가 암 연관있다는 것을 처음 증명
암후성유전체에서 가장 많이 연구된 것이 "메틸화" 임

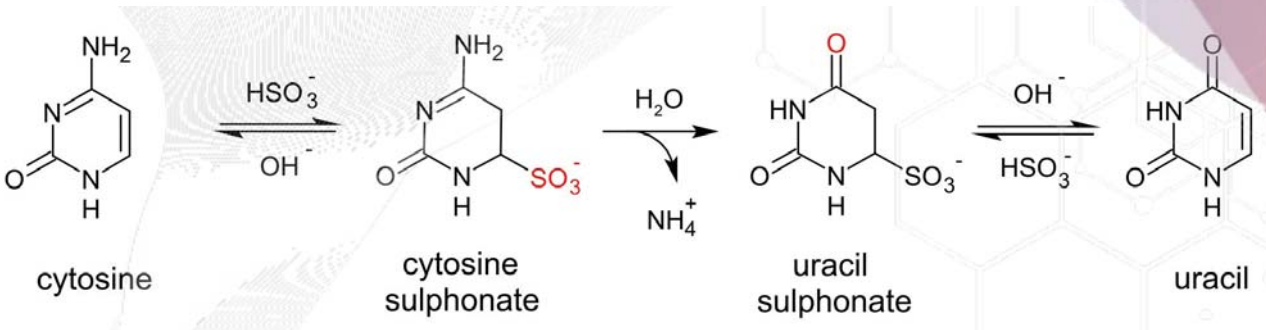
DNA methylation and cancer

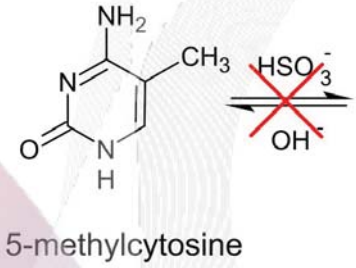




대부분 포유류는 CpG island에서 시토신 5번째 탄소가 메틸화됨
Tumor-suppressr유전자 de novo메틸화 문제
암세포 전반 저메틸화 – 염색체이상, translocation문제야기
이벤트는 **'초기 ' 에 일어나는것으로** 알려져 있어 진단이 시급함

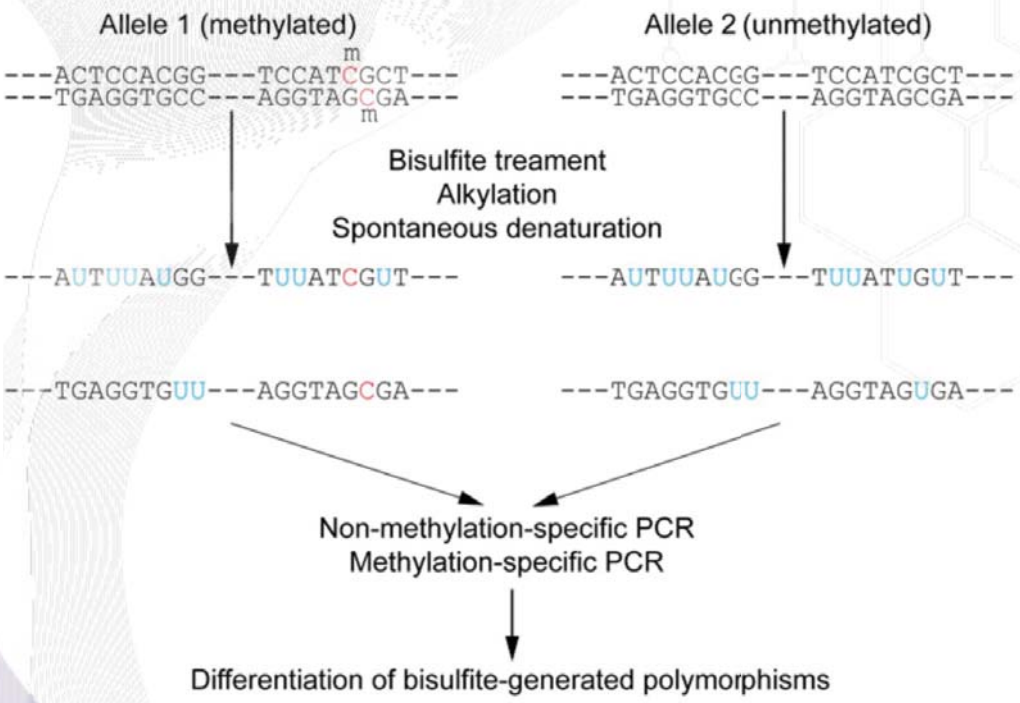
Bisulfite chemistry



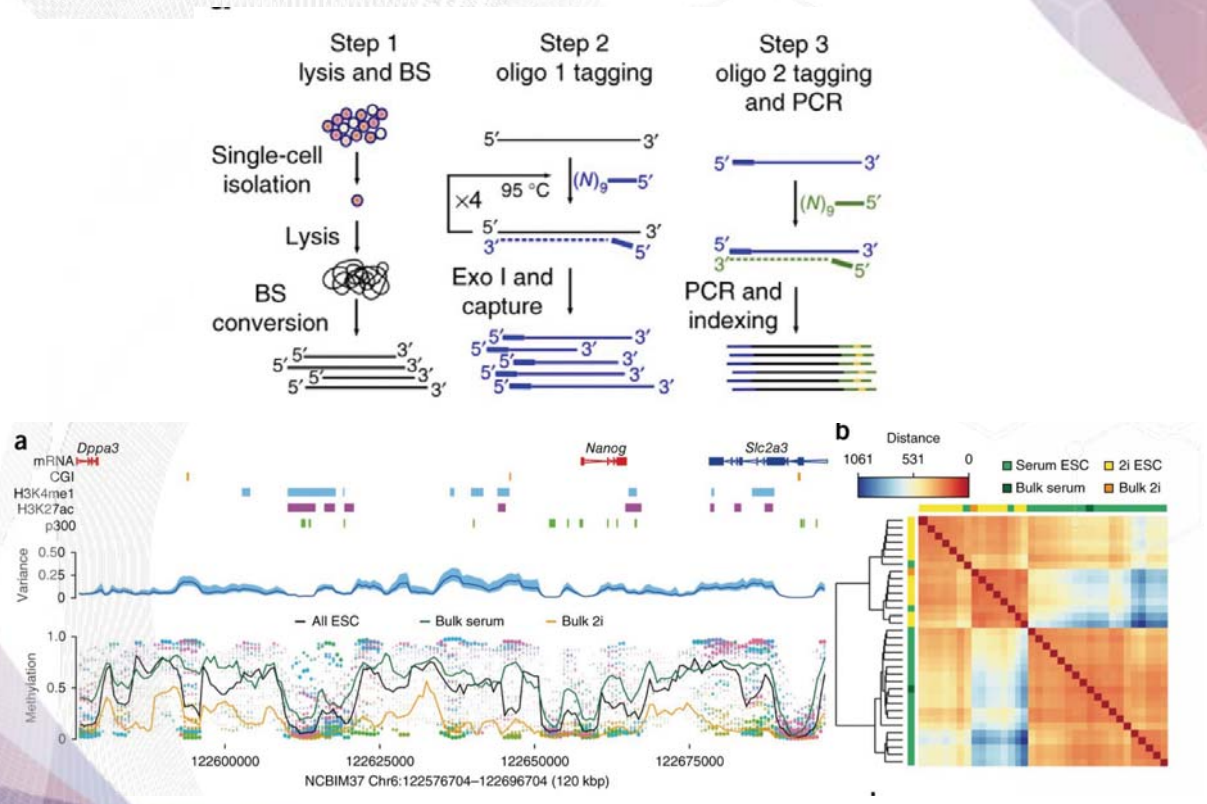


In animals it predominantly involves the addition of a methyl group to the carbon-5 position of cytosine residues of the dinucleotide CpG, and is implicated in repression of transcriptional activity.

How bisulfite conversion works

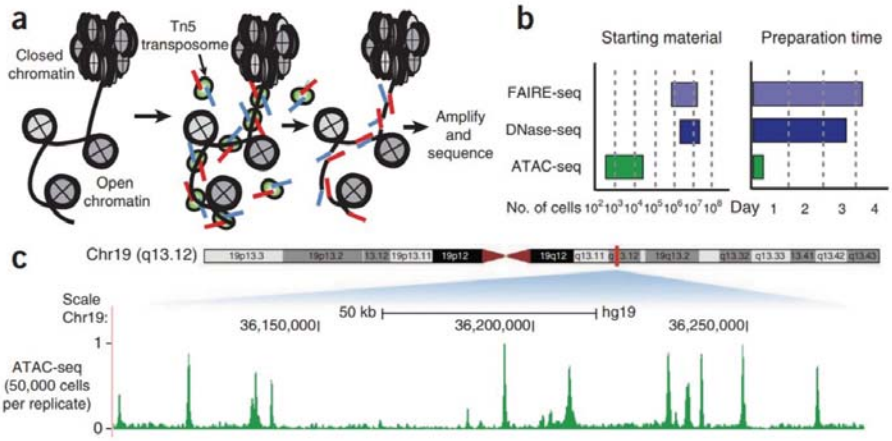


scBS-seq (single cell bisulfite sequencing)



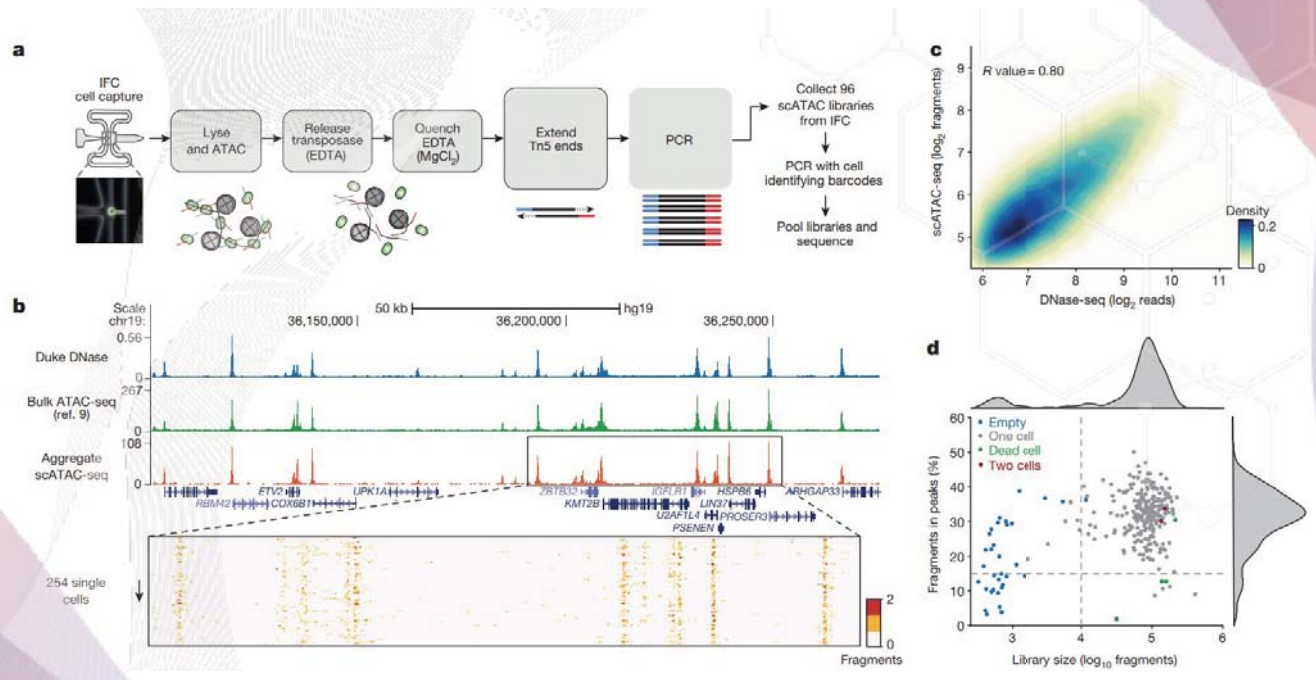
ATAC-seq
(Assays for Transposase Accessible Chromatin)

Chromatin accessibility (염색체 접근도)

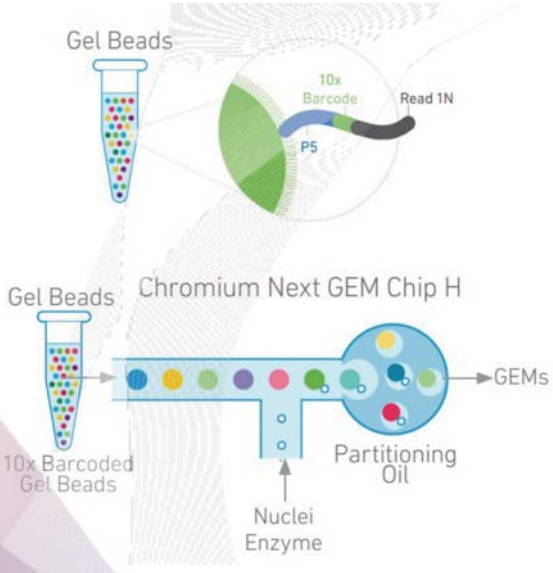


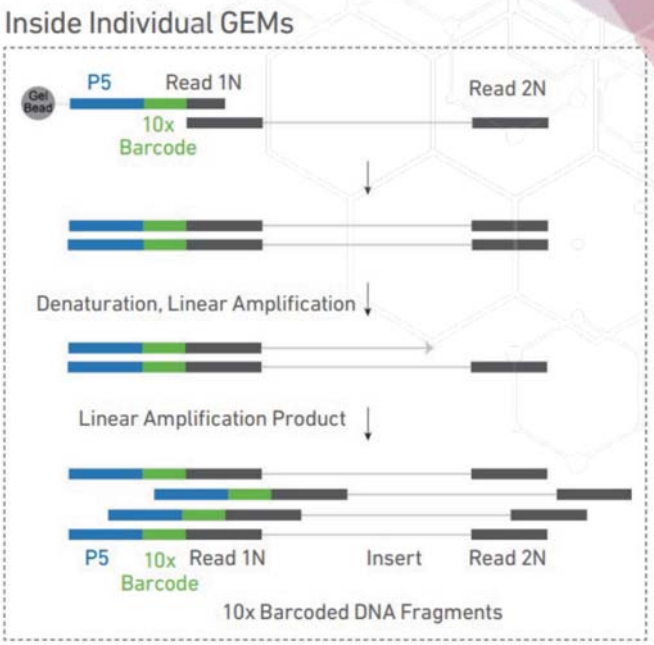
적은 시료 (적은 세포) 로 짧은 시간에 정확한 분석이 가능해 standard 가 되어감.

scATAC-seq

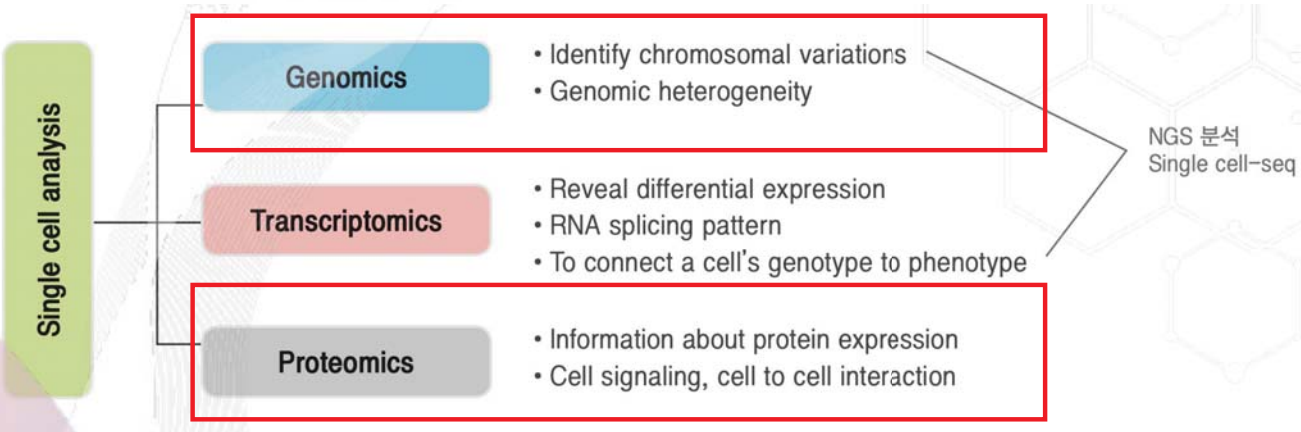


Commercialized scATAC-seq

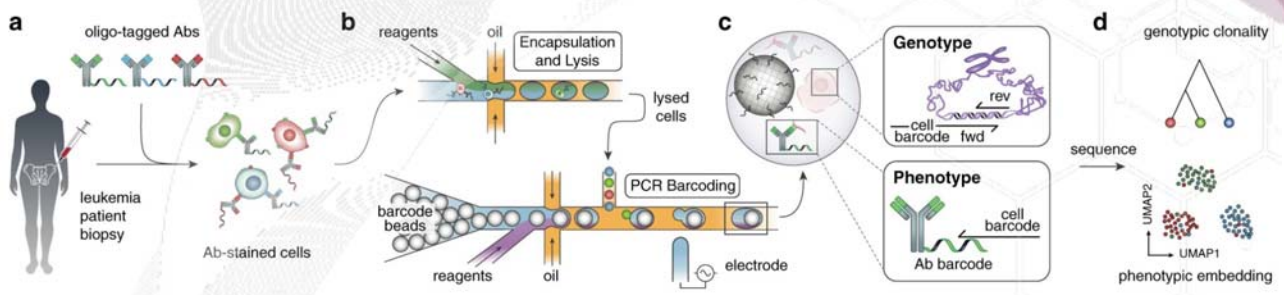




Single-cell analysis platforms



Joint profiling of DNA+Protein

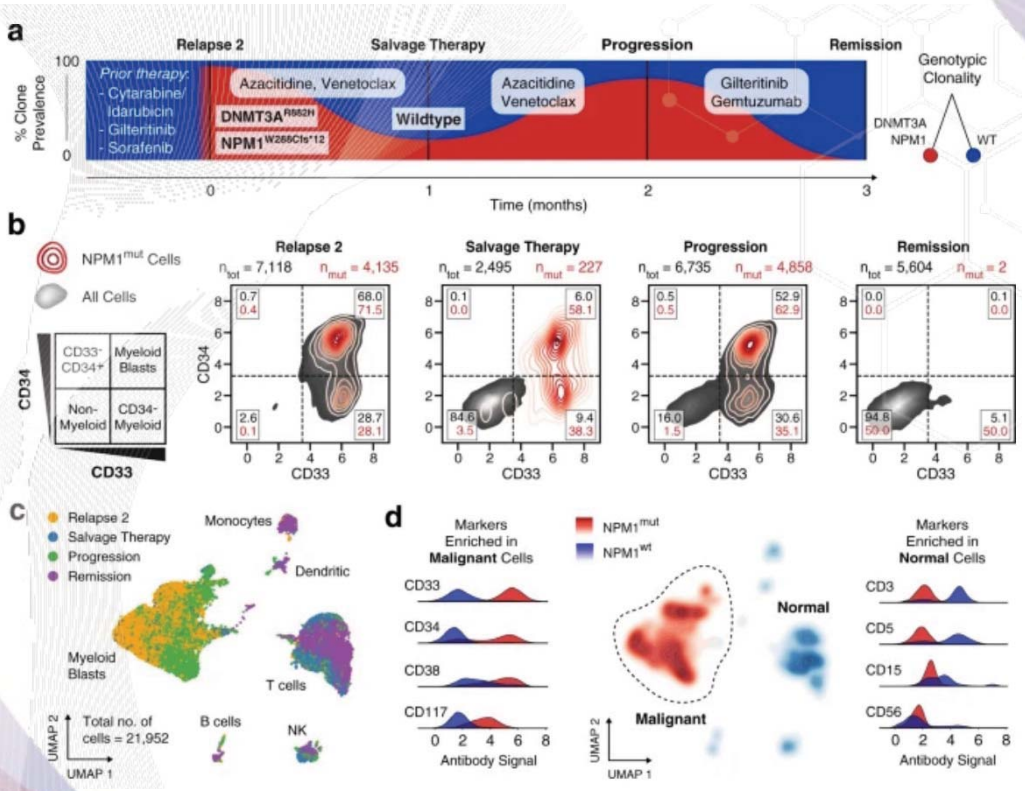


Dab-seq workflow

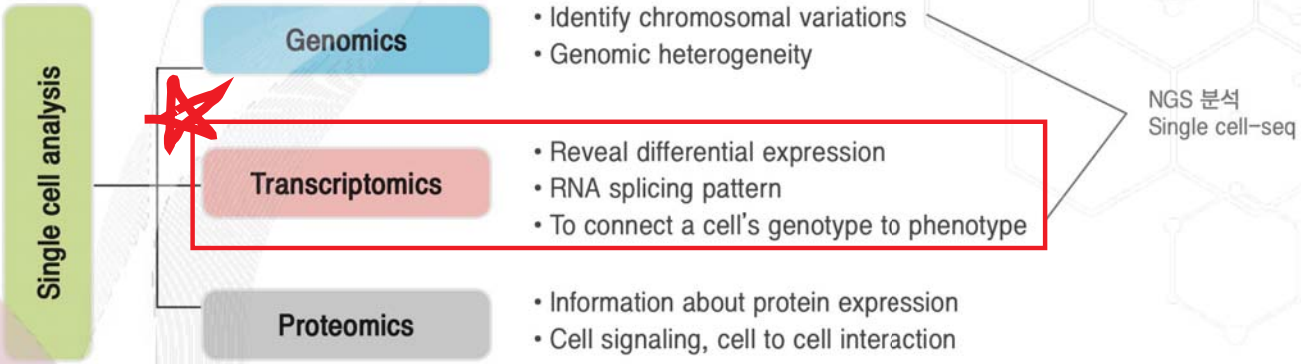
혈액암(AML)의 진단은 주로 Flow cytometry나 DNA mutation분석을 통해하지만 동시에 한세포에서 진단하는 방법론은 없었음.

1.Mission Bio's Tapestri platform을 modify하였음
2.Oligo conjugated Antibody를 활용하여 custom하게 실험이 가능함.
3.기존 Abseq그룹 (BD Rhapsody에서 상용화) 이 개발함.

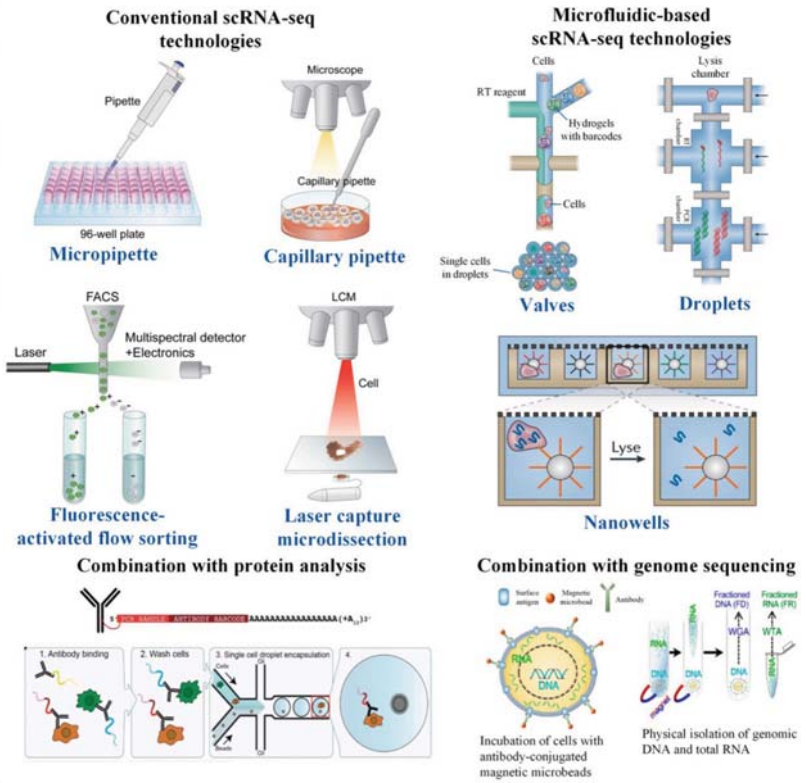
Multiomics profiling of patient dynamics



Single-cell analysis platforms



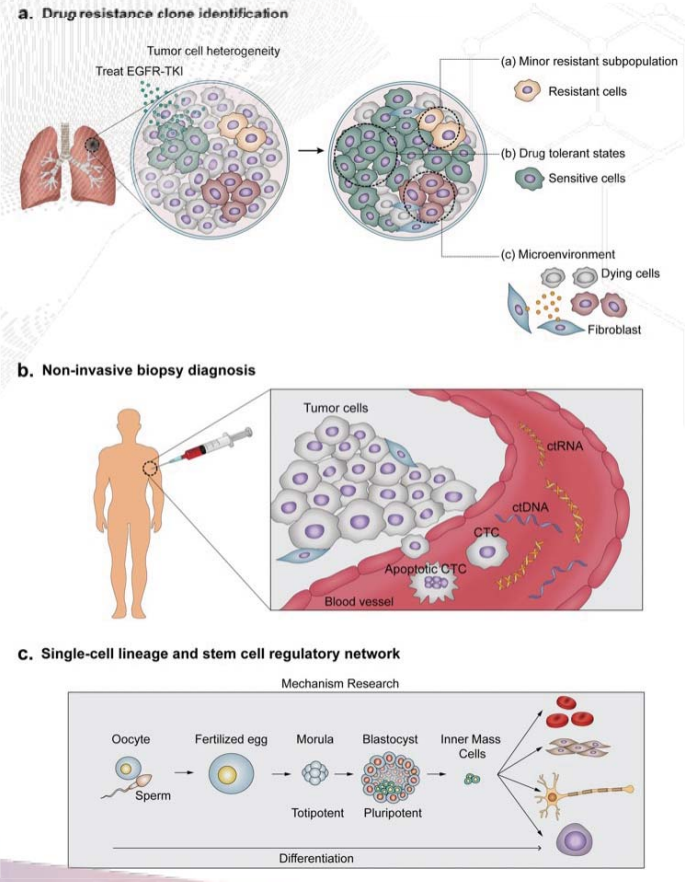
Single-cell RNA sequencing (scRNA-seq)



RNA+protein

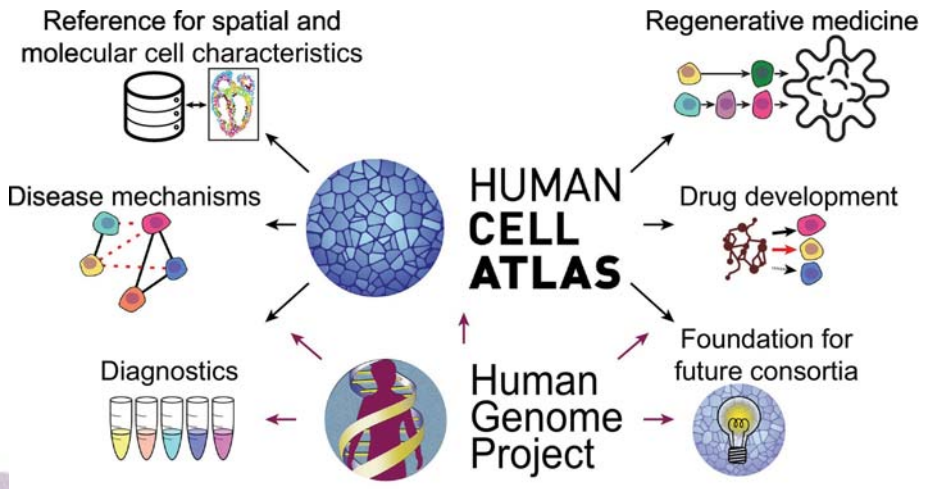
DNA+RNA

Applications of scRNA-seq



Towards a Human Cell Atlas (HCA)

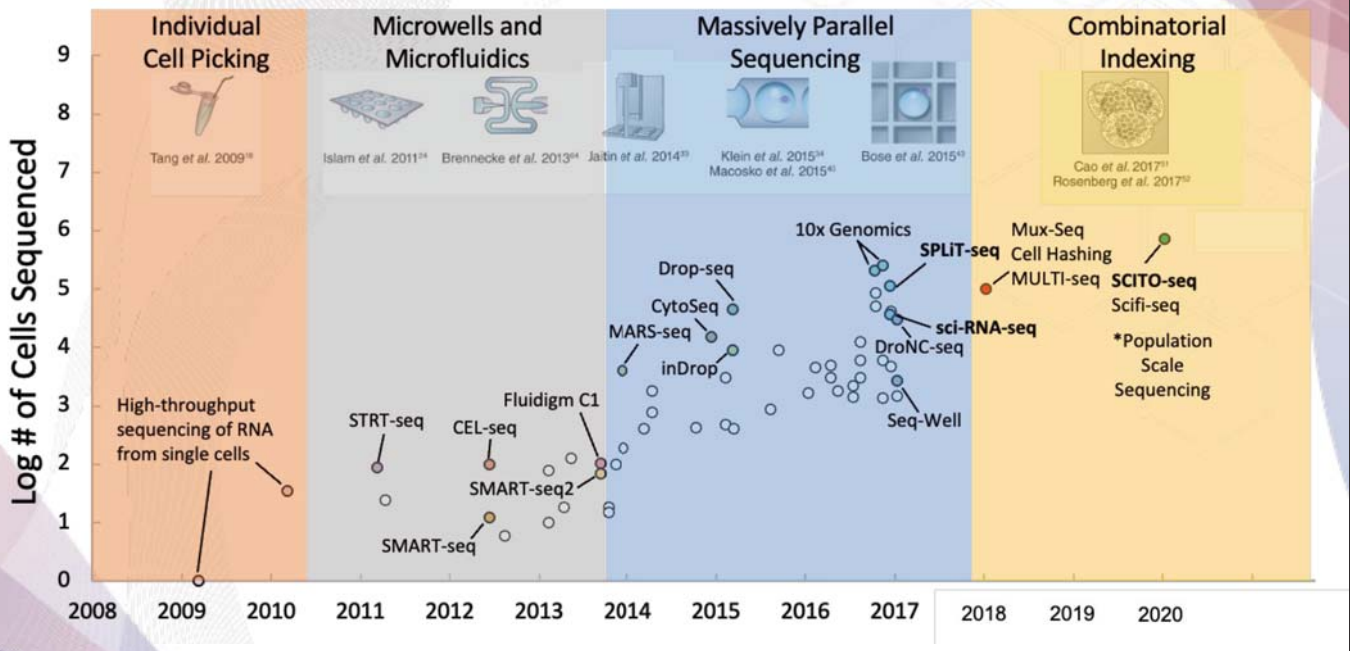
- Inspirations from HGP (human genome project) as a collaborative project
  - Impact is illustrated by world-wide collaboration w/COVID-19
- 39million cells from 15 organs
- Healthy people의 단일세포 전사체 맵 (scRNA-seq) → disease



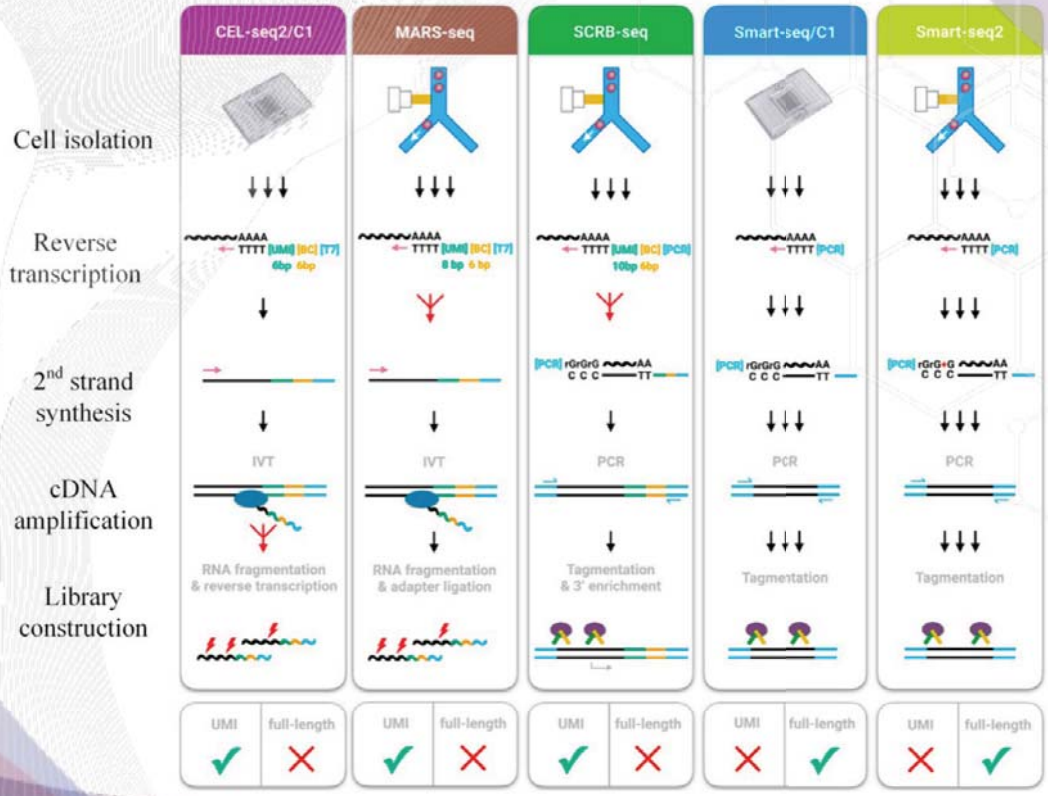
Google Maps of human cells is a milestone

- HCA Portal Site: https://data.humancellatlas.org/
- 하버드/MIT 브로드 연구소: https://singlecell.broadinstitute.org/single_cell
- 유럽연합 생물정보 연구소: https://www.ebi.ac.uk/gxa/sc/home

Exponential increase in scRNA-seq throughput

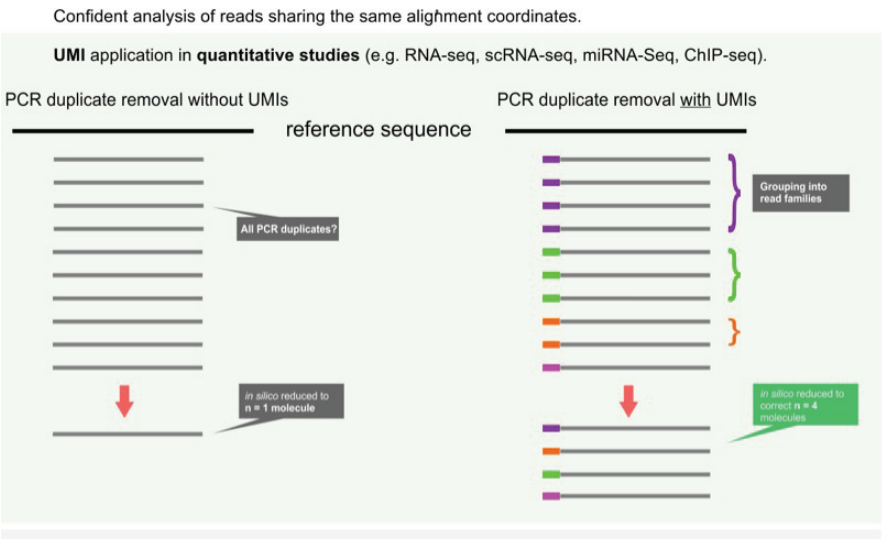


Initial phase of scRNA-seq technologies

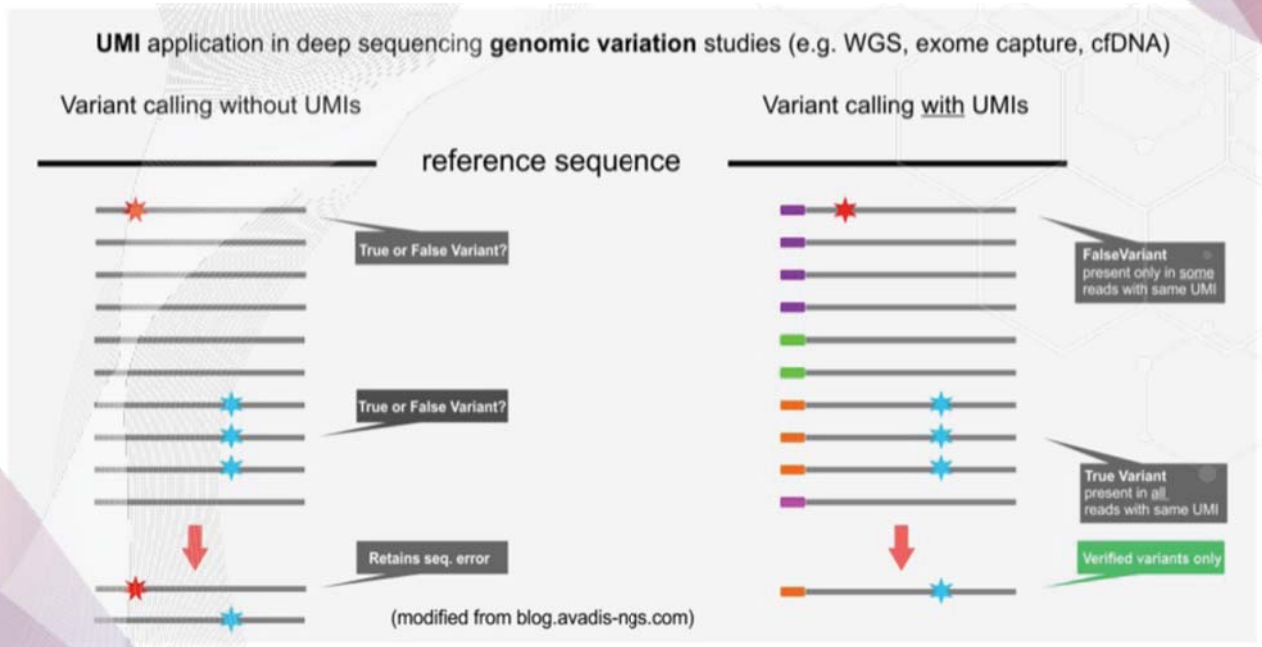


Unique Molecular Identifier (UMI) - Quantification

- Known as Molecular Barcodes (random 'N'염기서열)
- Complex DNA sequences added to reduce PCR amplification bias

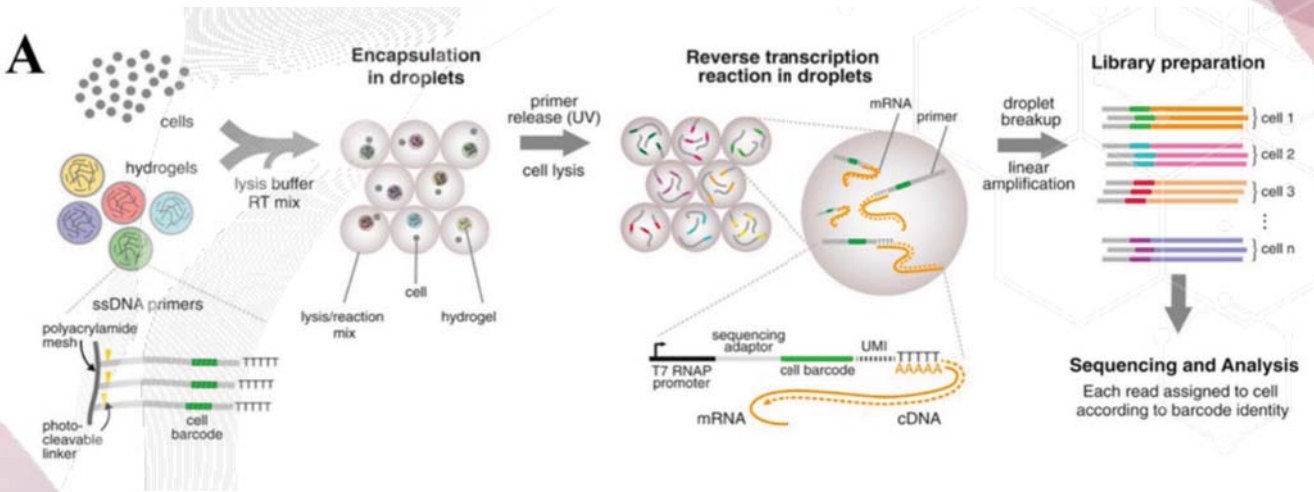


Unique Molecular Identifier (UMI) – Variant detection



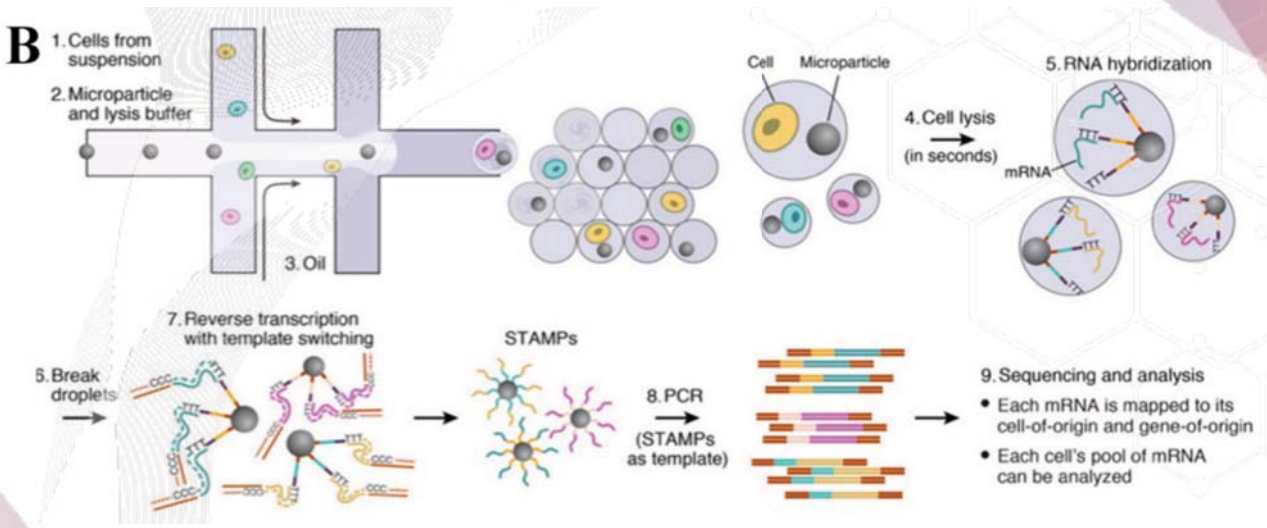
More high-throughput methods?

- Droplet-based scRNA-seq



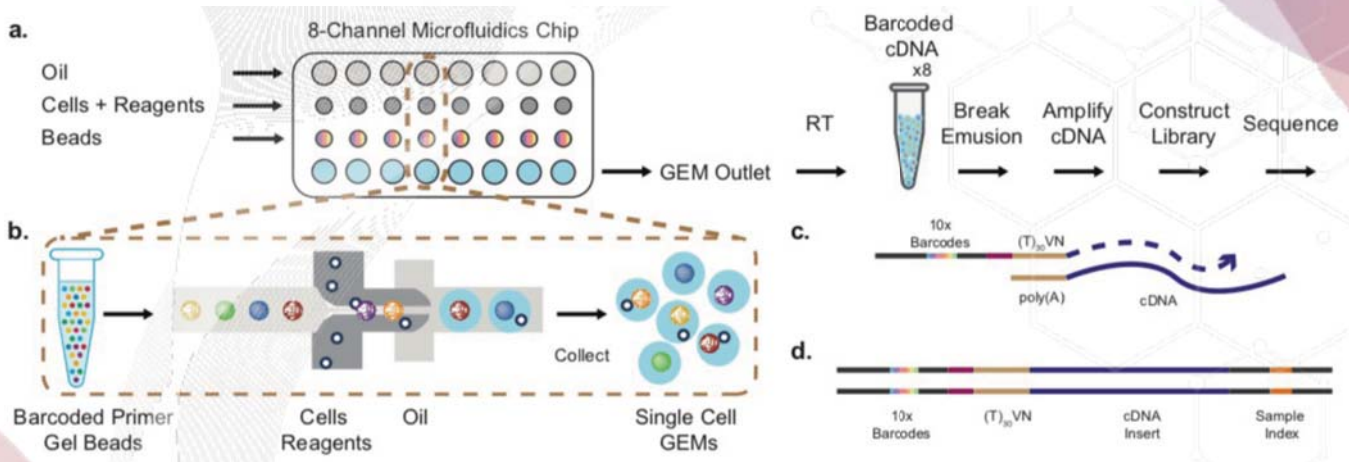
1. 초기 InDrop technology: low cell capture (~7%)
2. 20-50 copies/cell transcripts captured only

Drop-seq



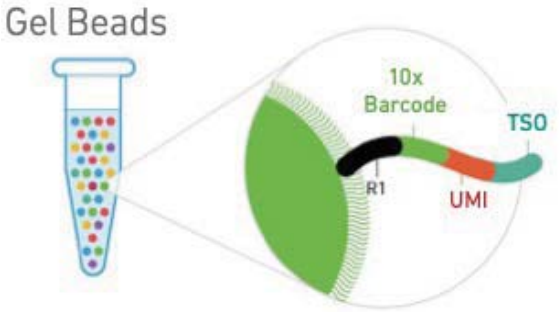
1.Use Barcoded Beads instead of hydrogel (InDrop)
2.Cell capture efficiency (~12.8%)
3.Captures 3' terminal fragments similar to In-Drop

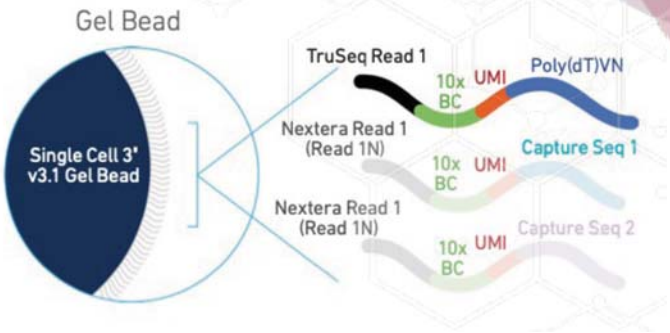
10x Genomics (commercial)



Uses Gel bead emulsion (GEM)
~50% Cell capture efficiency (Currently dominating the market!)

This can capture both 5’ and 3’ side of RNA



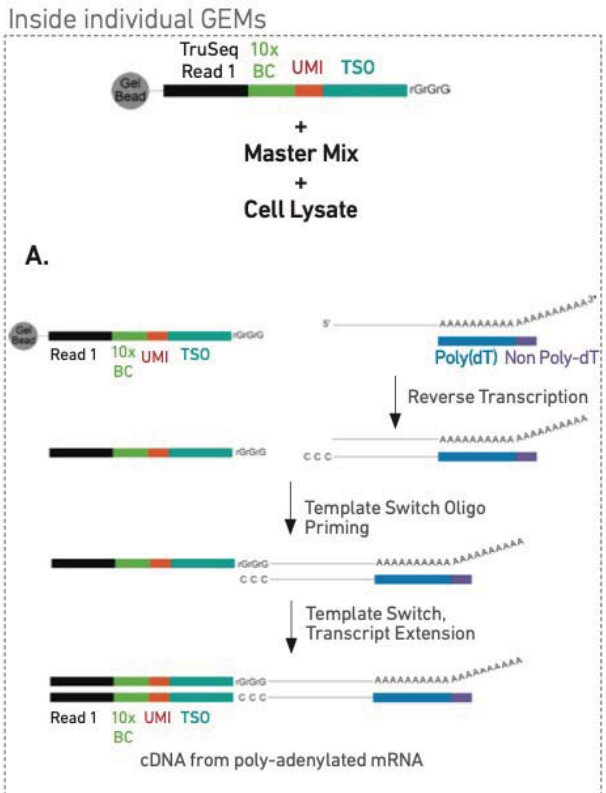


5’ GEM structure

3’ GEM structure

- They capture different parts of the transcript and show similar efficiency of capture
- Maybe limited to discovering alternative spliced transcripts (isoforms)
- 5’ technology can capture TCR (T-cell receptor) and BCR

How do you capture 5’ side?



TSO : template switching oligonucleotide

BC : barcode (random ‘N’ 염기 서열)

UMI : unique molecular identifier

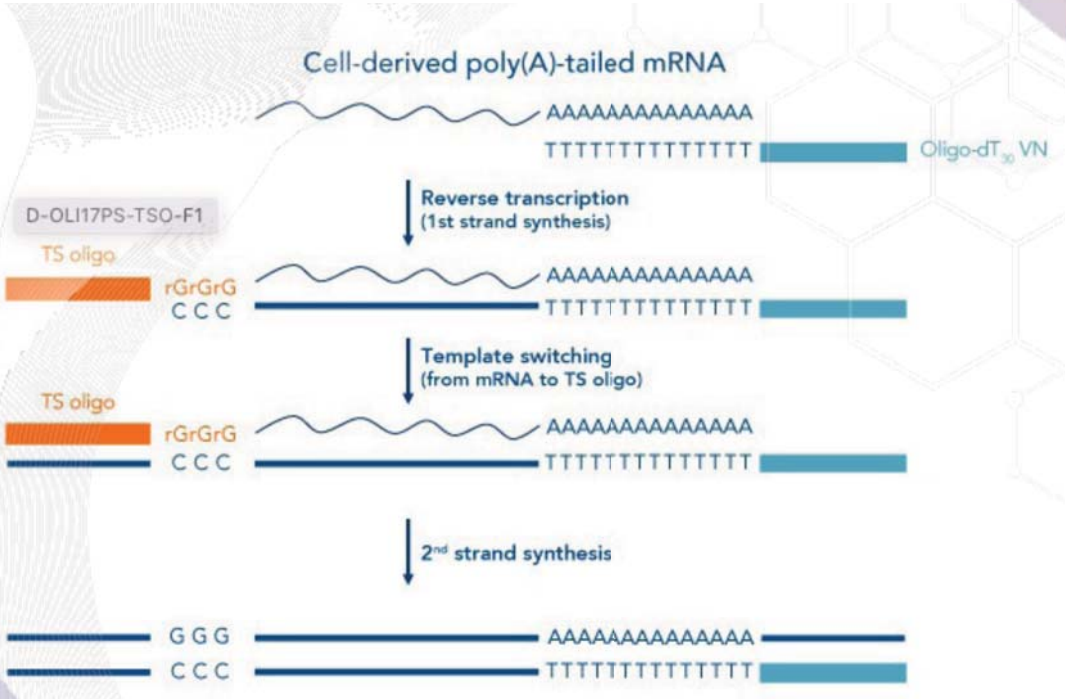
Template switching mechanism (RNA->cDNA)

- 1st cDNA 합성과정에서 cDNA 양 말단에 특정 서열을 삽입하는 획기적인 기술
  - 주로 Moloney Murine Leukemia Virus에서 유래한 MMLV RTase 라는 역전사 효소를 사용

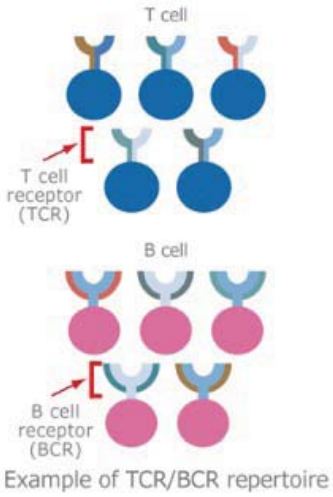
- 말단전이활성(Terminal transferase activity)
  - 역전사효소를 사용 mRNA 5'말단에 이르렀을때 일부 염기를 부가하는 능력

- 주형전환활성(Template switching activity)
  - 새로운 주형(template)으로 바꾸어 DNA를 합성하는 활성

Template switching schematic



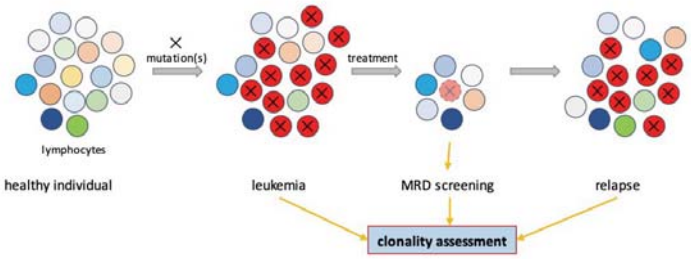
Why TCR and BCR sequencing is important



Main types of lymphocytes (T and B cells)
~10^12 diversity in DNA sequences

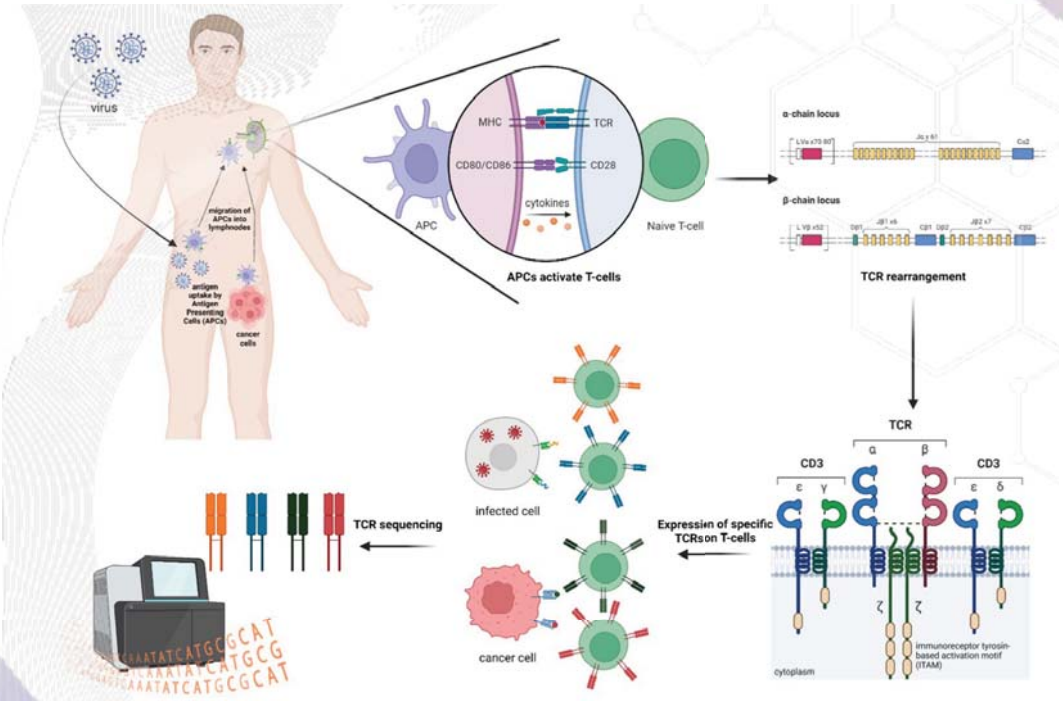
They recognize antigens

Malignant clones ~ 0.001%
→ **Need to sample many cells!**

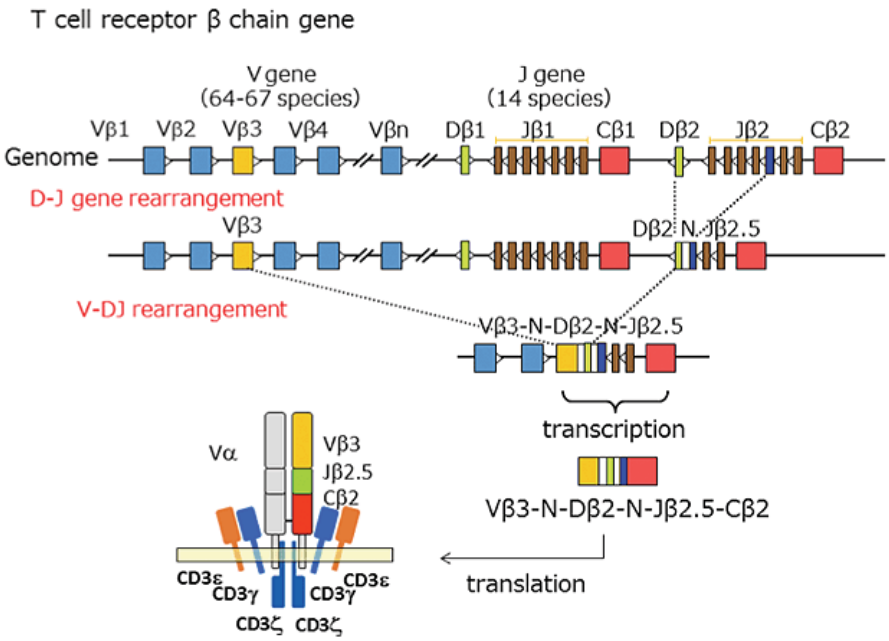


Detection of minimal residual disease (MRD)

TCR and HLA

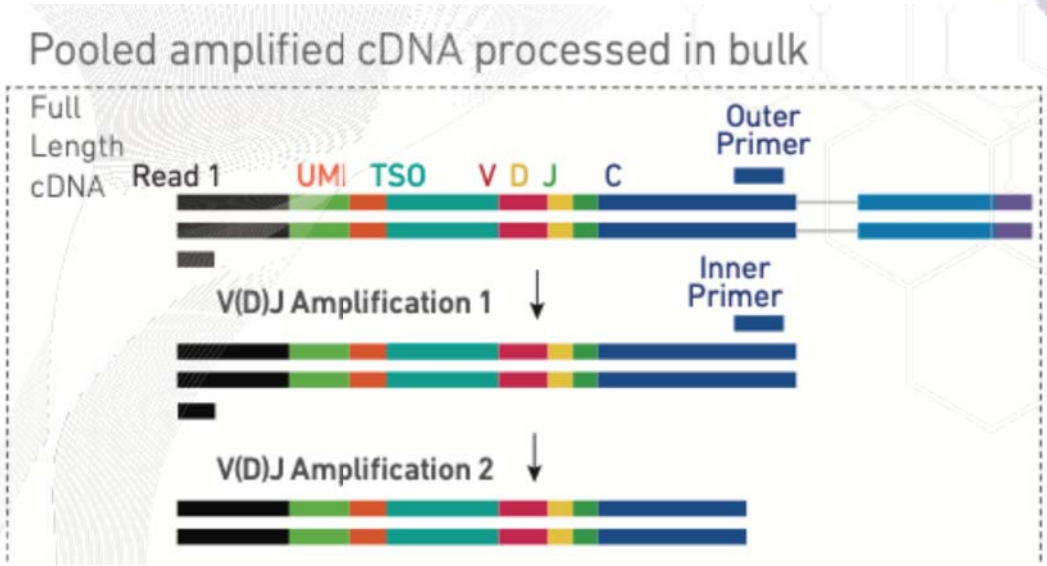


Gene arrangement in the T cell receptor beta chain gene



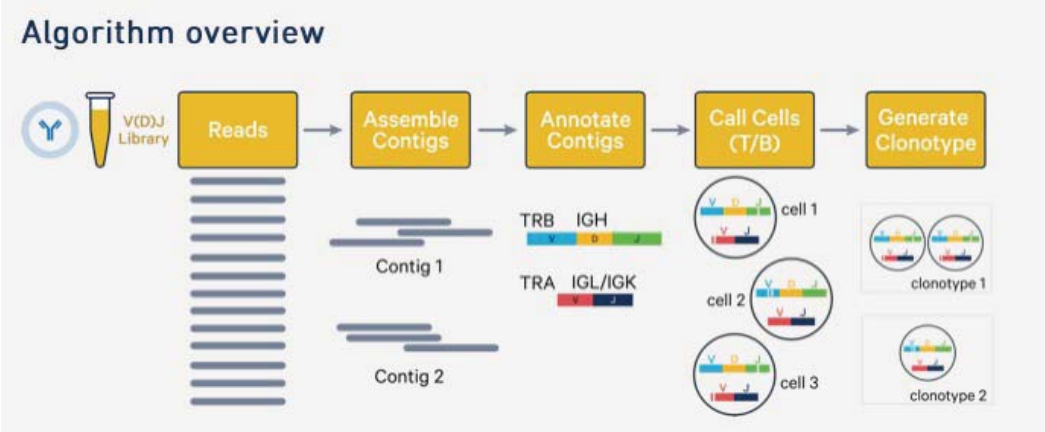
CDR3 is important for binding and determines 'Clonotype'

VDJ amplification from cDNA



TCR/BCR gene은 Constant region에 specific 한 primer로 증폭이 쉽게 가능
Outer + Inner primer 두 step의 Nested PCR이라는 방법으로 specificity높임

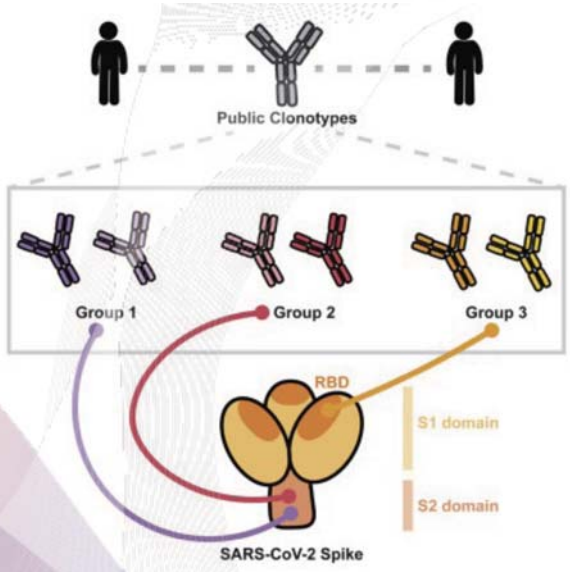
V(D)J sequence from assembly

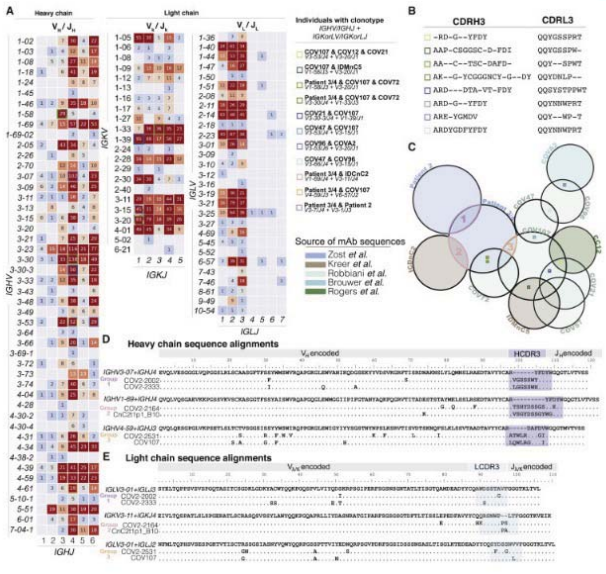


시퀀싱 서열분석을 한 결과는 ~150bp로 짧아 전체 TCR의 reconstruct (~800bp) 하기 위해 조각들을 이어붙이는 assembly를 진행한다.

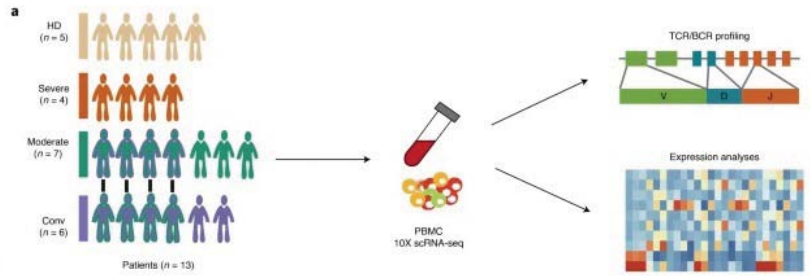
Clonotype (클론형) : 특정 항원에 반응하는 TCR/BCR의 염기서열 조합

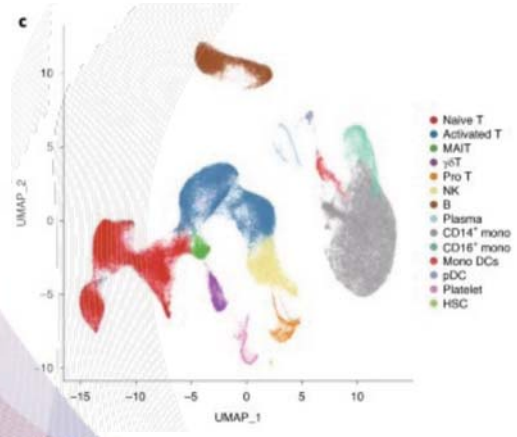
Convergent antibody response to the SARS-CoV-2 spike protein in convalescent and vaccinated individuals

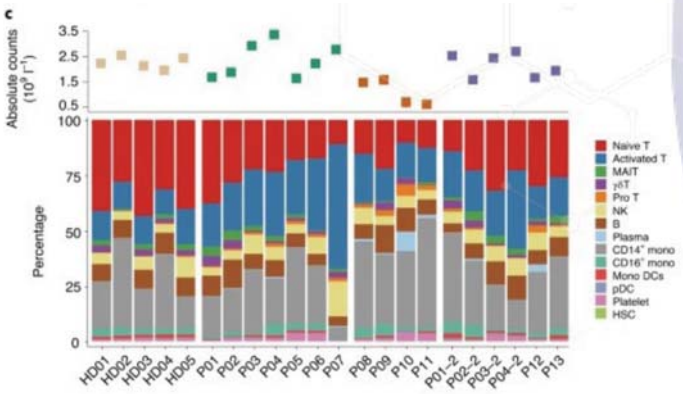




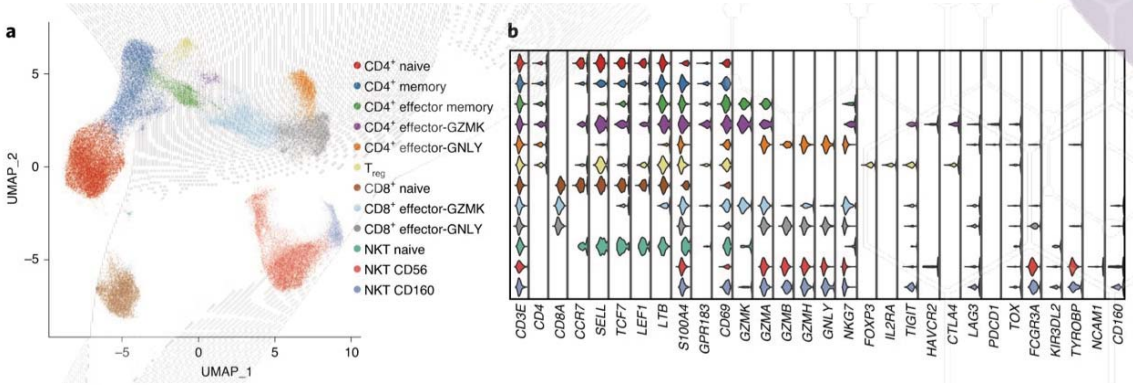
Case study with COVID-19

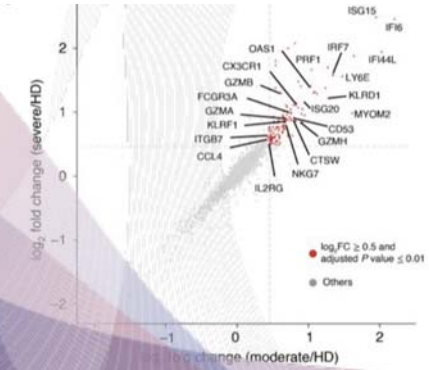


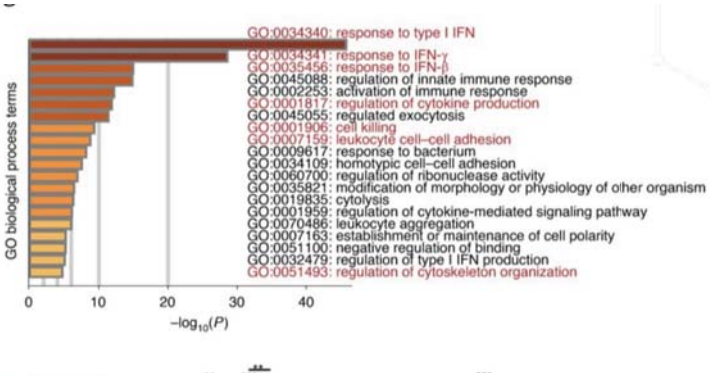




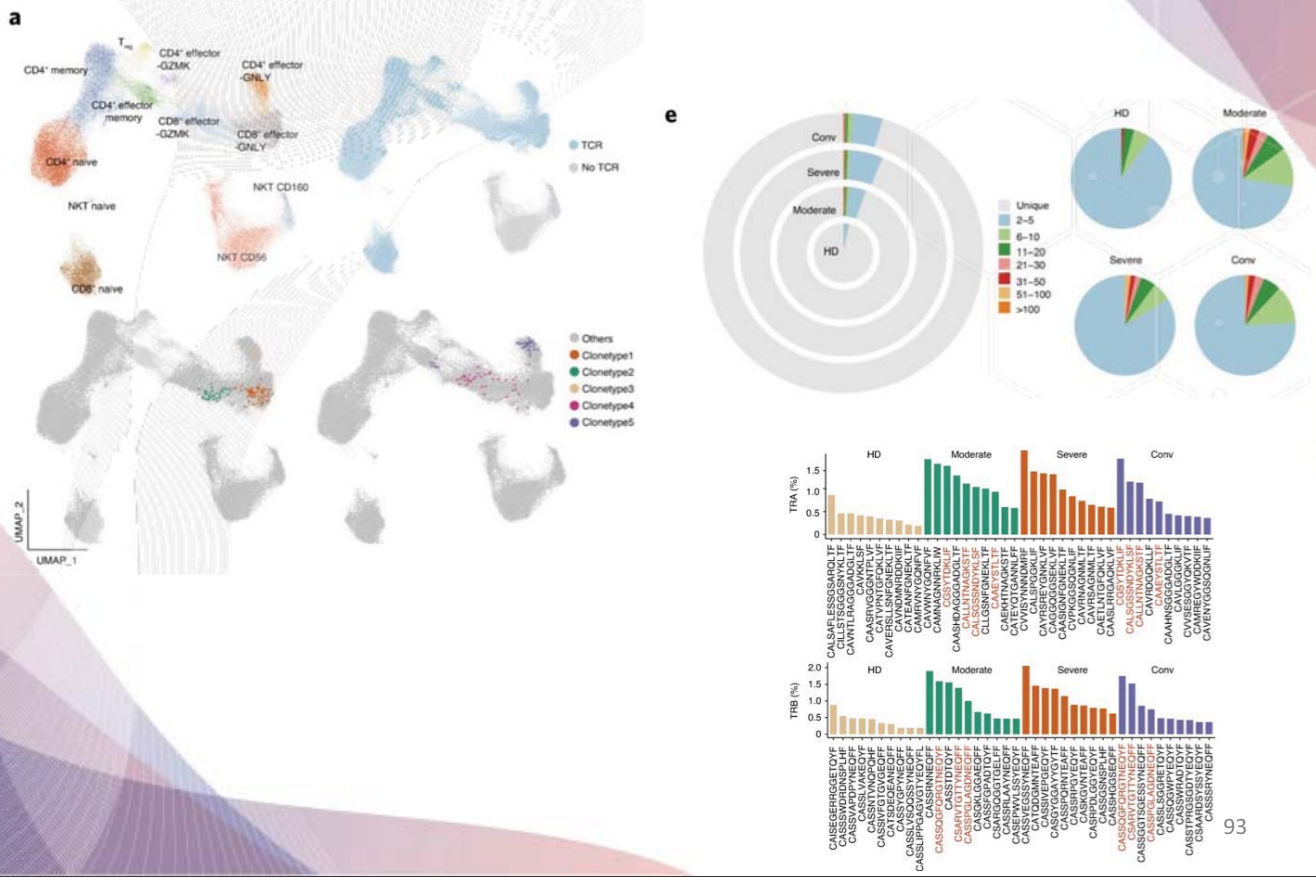
Immunological feature of T cells



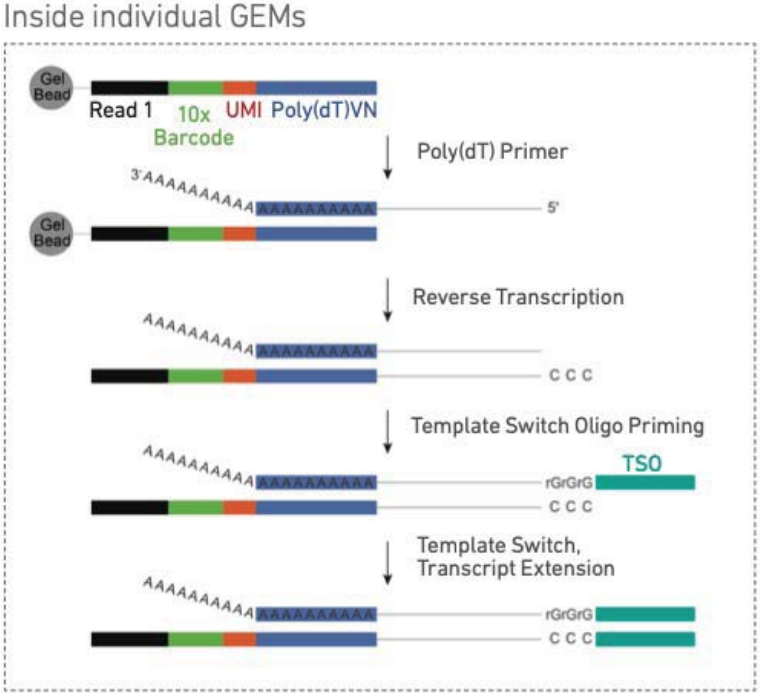




Expanded TCR clones and selective usage of V(D)J genes



How do you capture 3' side of RNA?

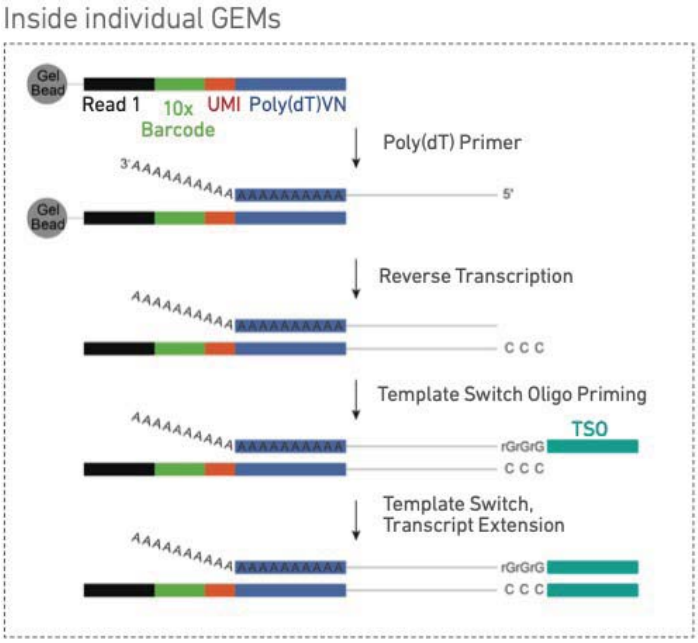


TSO : template switching oligonucleotide

BC : barcode
(random 'N' 염기 서열)

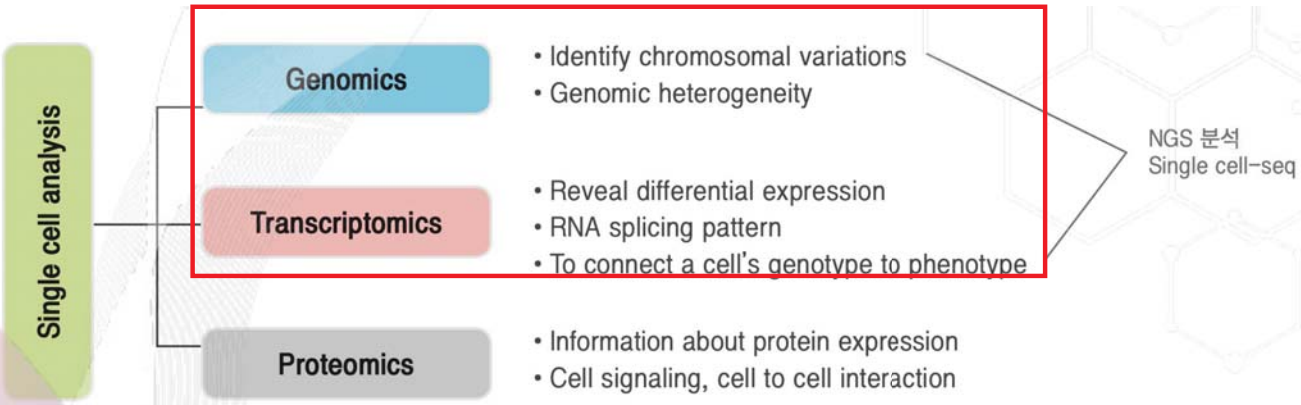
UMI : unique molecular identifier

Challenge: why 5' capture strategy is better to see V(D)J genes?

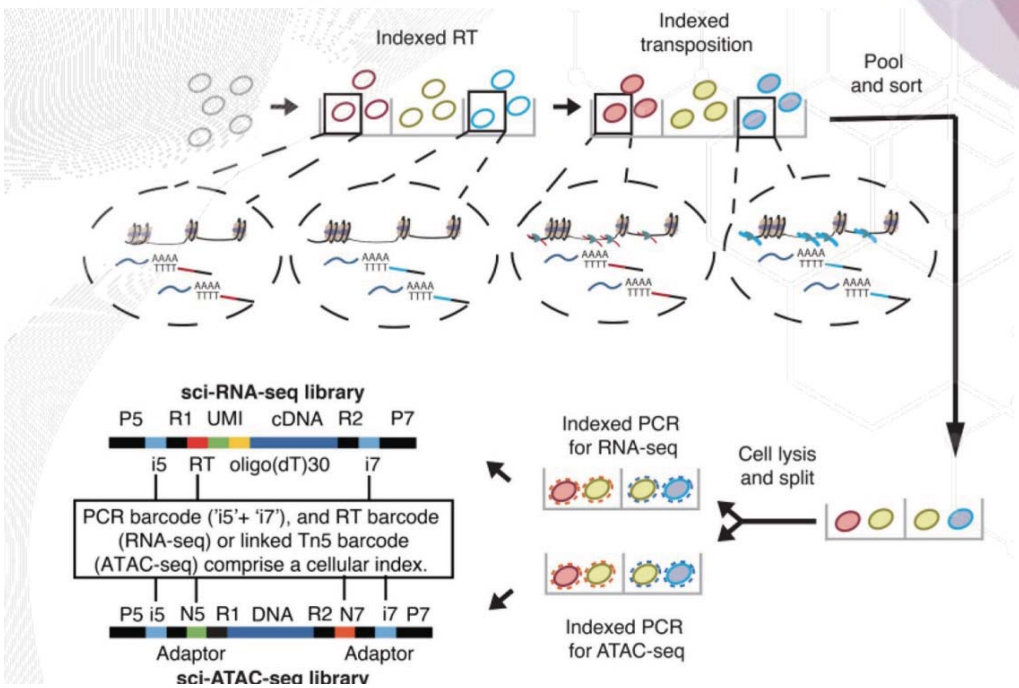


You will need many primers to target all Variable genes for 3' side!
Compare to 5' capture where you need 1 or 2 constant primers

Single-cell analysis platforms

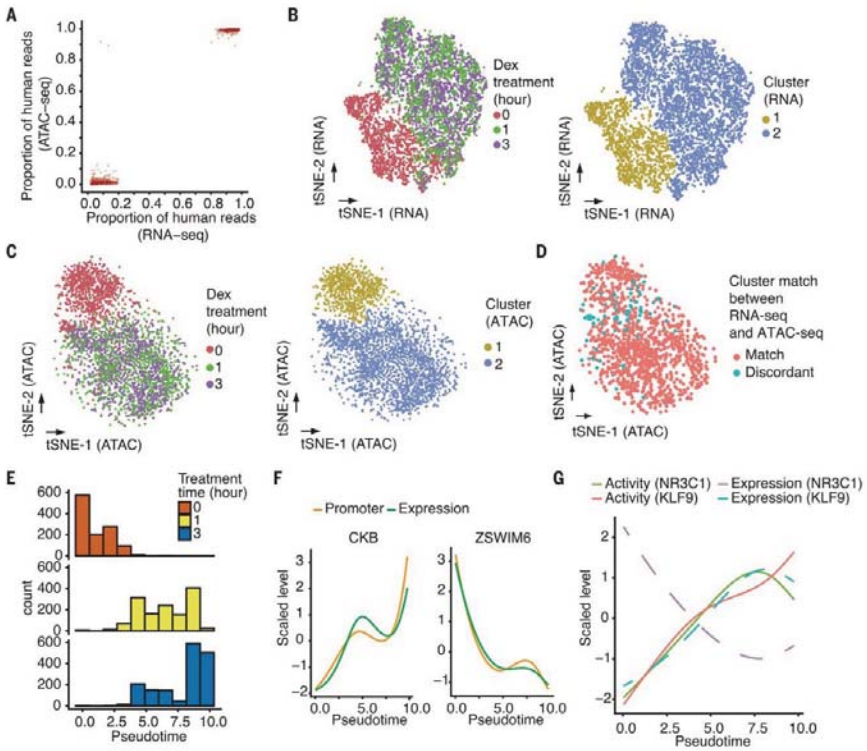


Joint profiling of RNA+ATAC (sci-CAR)

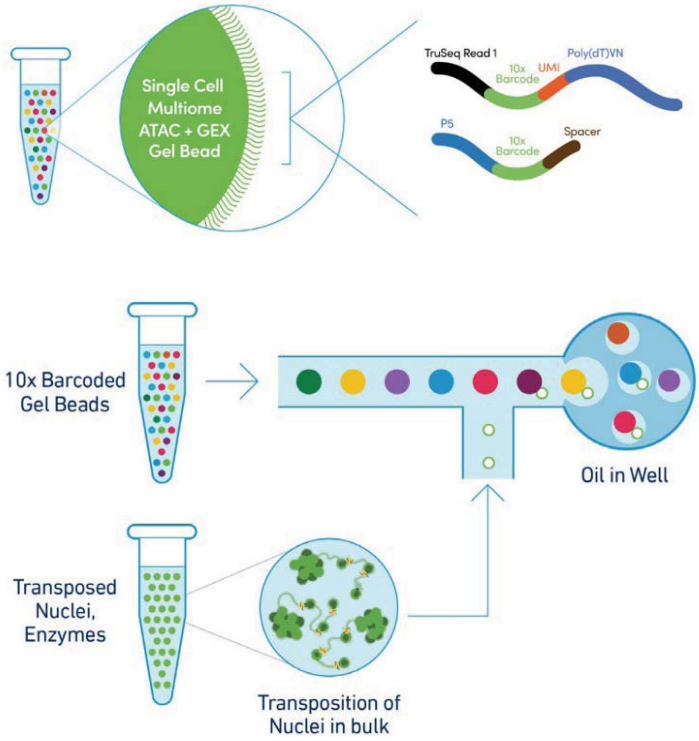


Different primer combinations for different modality amplification

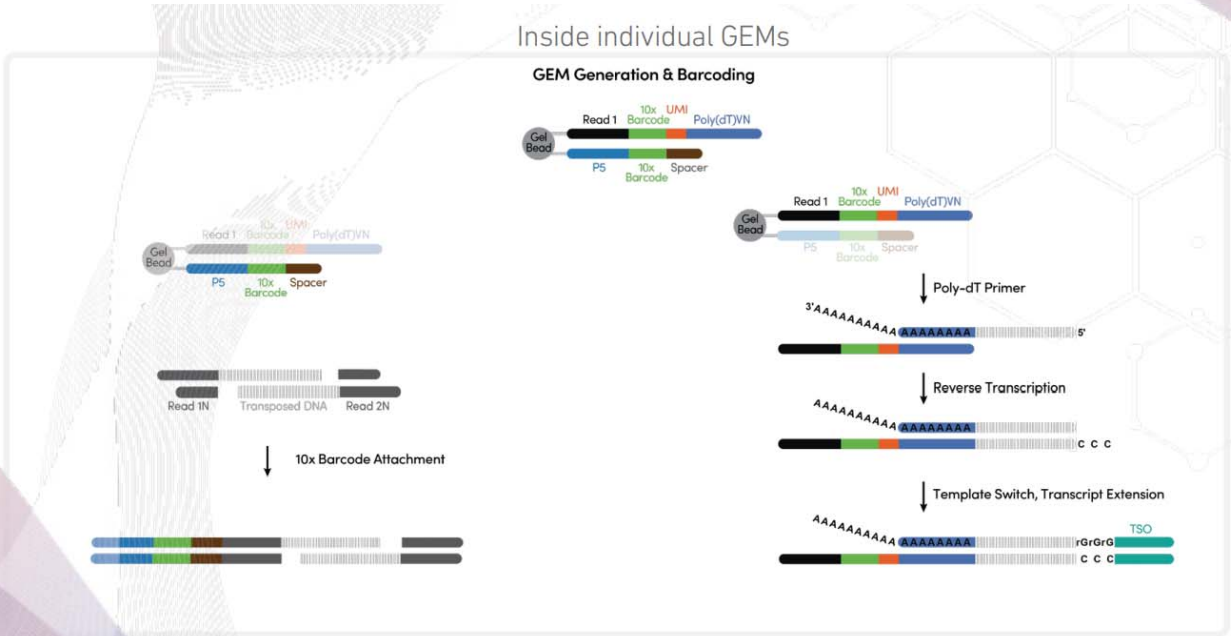
Gene regulation dynamics (open chromatin + RNA)



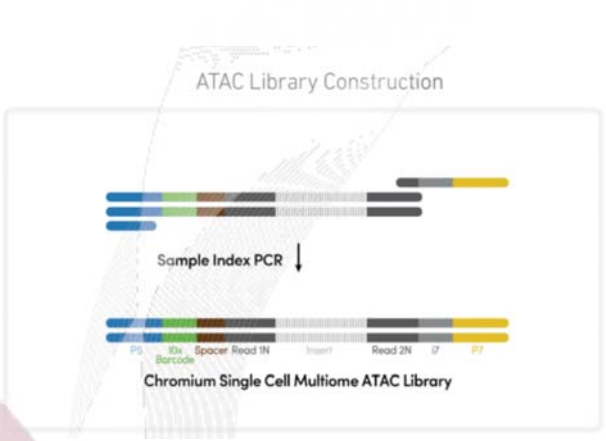
Multiome RNA+ATAC (Commercial)

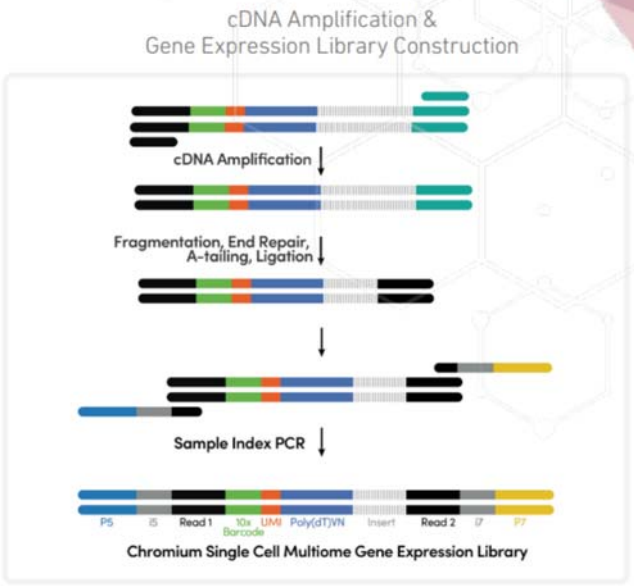


Split DNA and RNA reaction

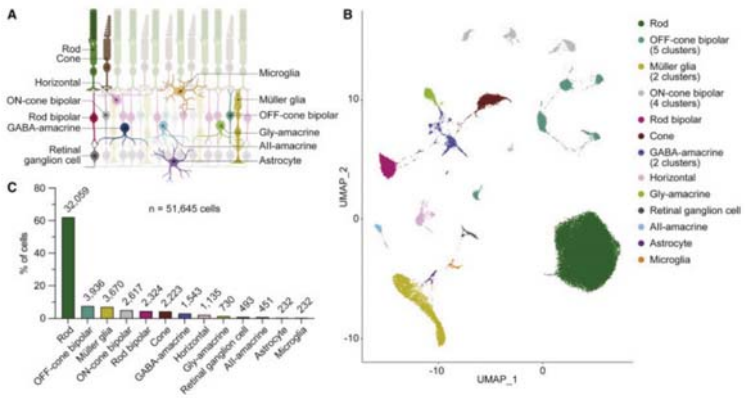


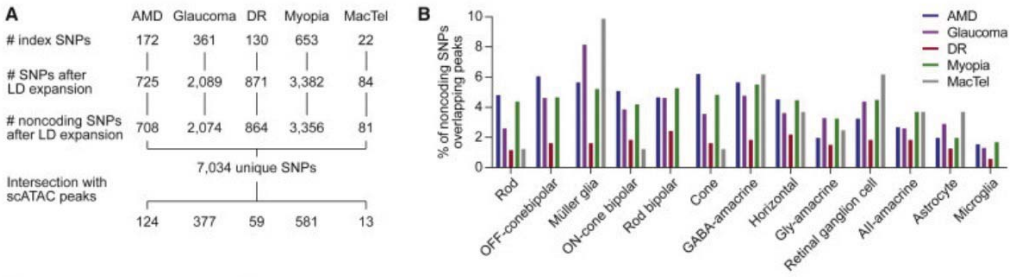
ATAC / RNA library preparation





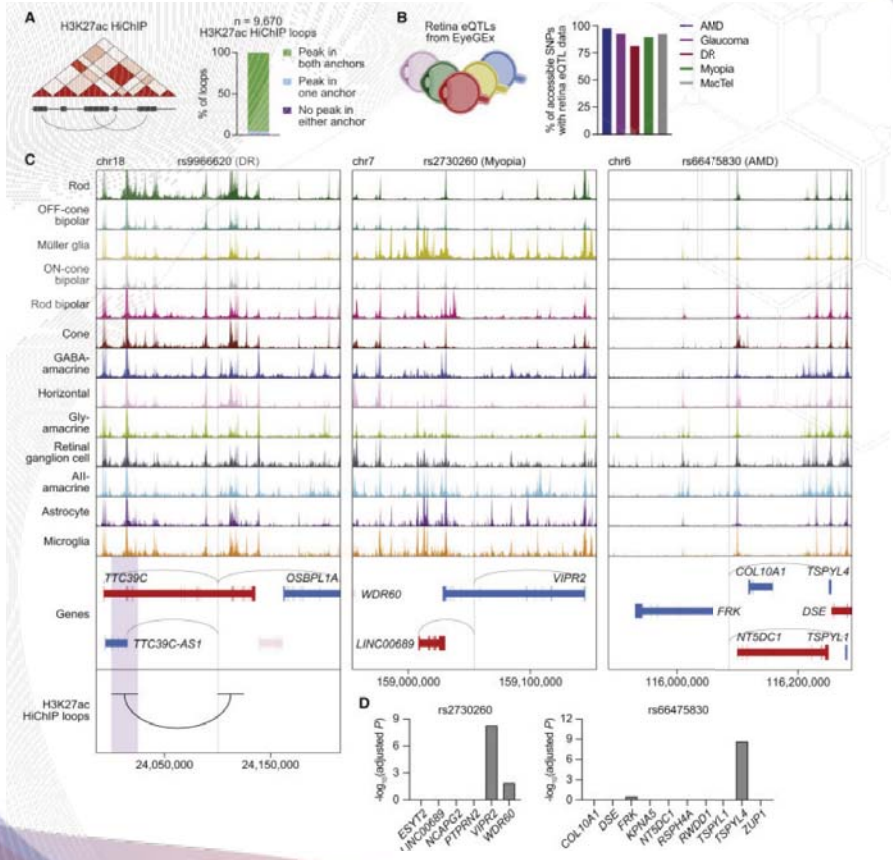
Multiome in retina cells



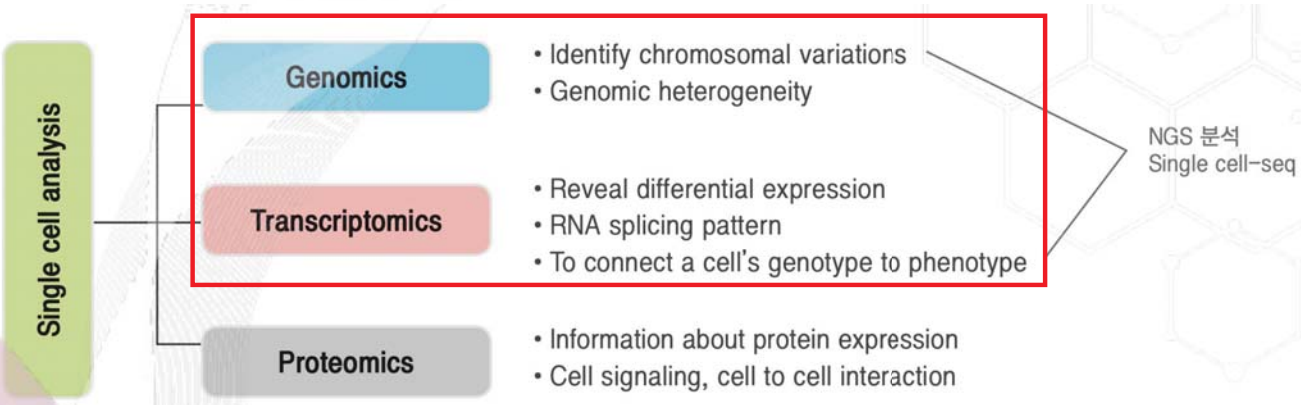


Prioritize cell-type specific ATAC peaks from GWAS datasets

Multiome + public data integration

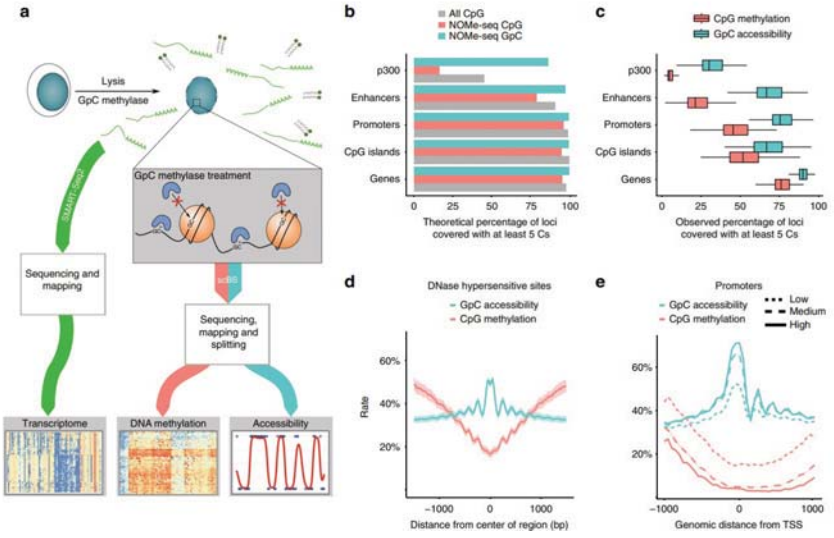


Single-cell analysis platforms

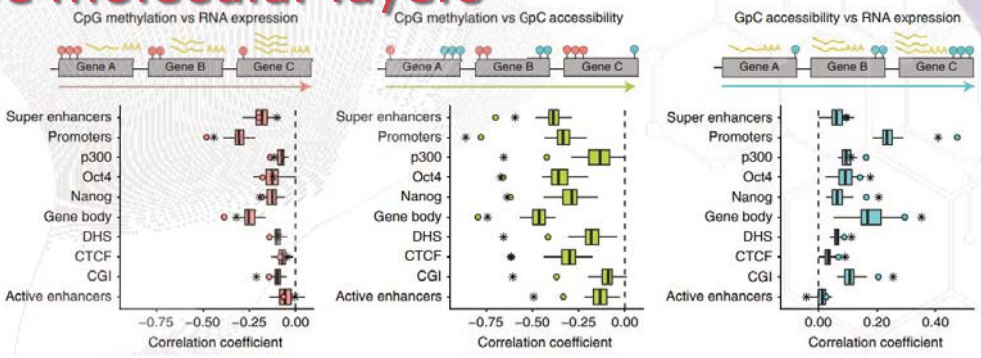


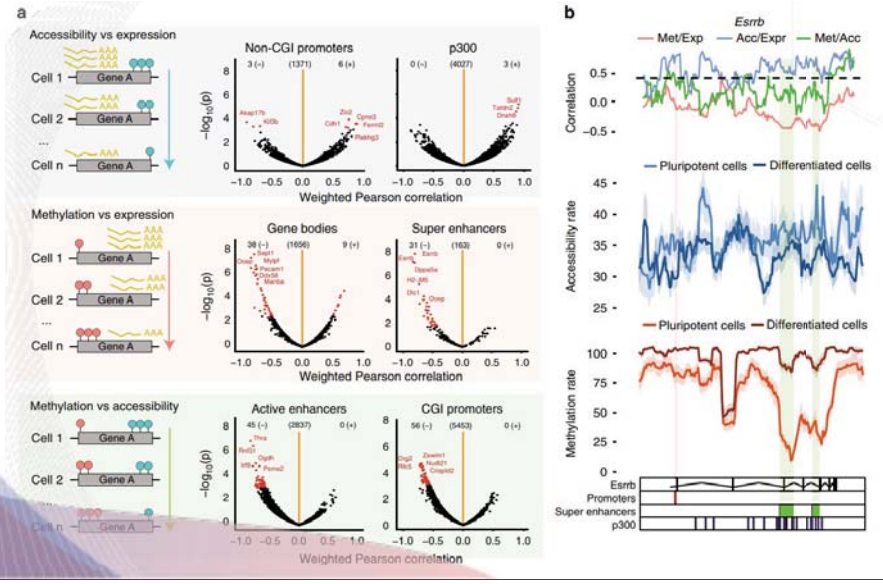
Joint profiling of methyl+chromatin+RNA

- scM&Tseq (methylation+RNA)
- NOMe-seq (nucleosome occupancy and methylation)
  - **Methyltransferase** (advantage over count based ATAC, DNase-seq methods)
  - **Frequency estimates** of CpG methylation doesn't suffer technical variation

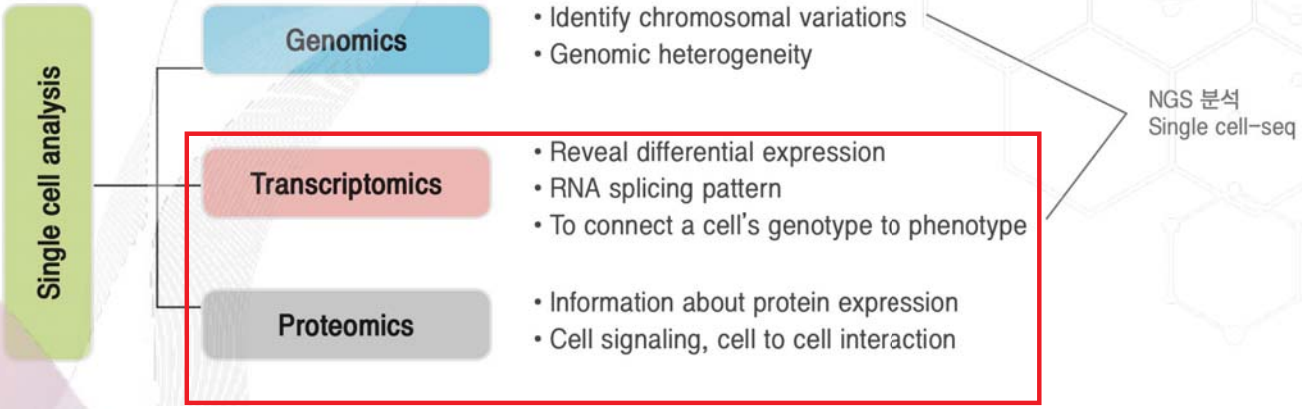


Known/novel association between three molecular layers





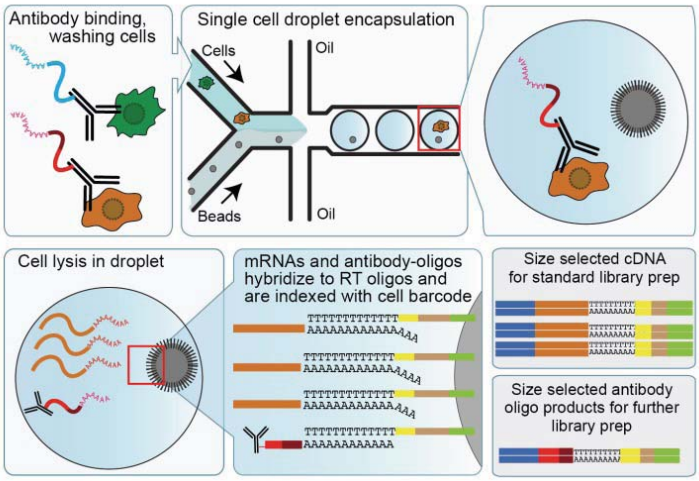
Single-cell analysis platforms



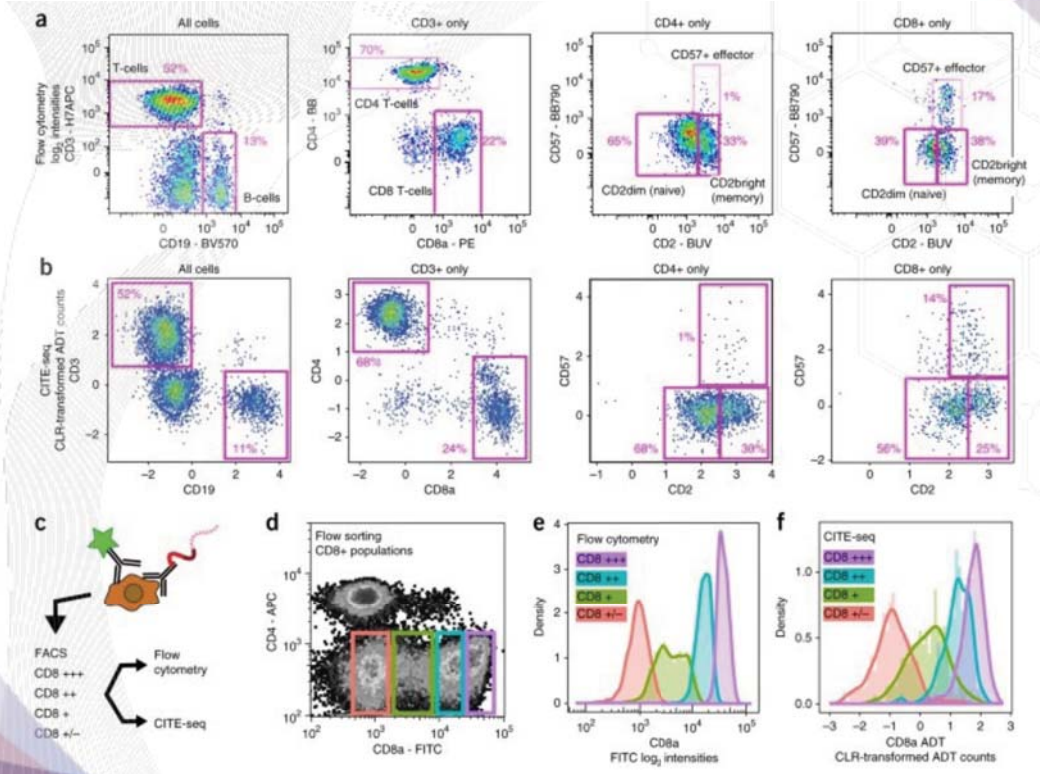
scRNA-seq + Surface protein

- **CITE-seq** (Cellular Indexing of Transcriptome and Epitopes by Sequencing)

CITE-seq uses DNA-barcoded antibodies to convert detection of proteins into a quantitative, sequenceable readout. Antibody-bound oligos act as synthetic transcripts that are captured during most large-scale oligodT-based scRNA-seq library preparation protocols (e.g. 10x Genomics, Drop-seq, ddSeq).



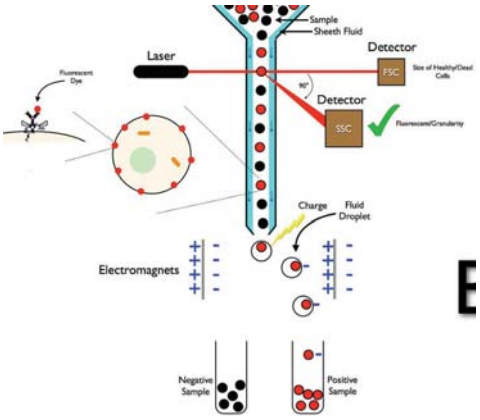
Comparison to FACS (fluorescence activated cell sorting)



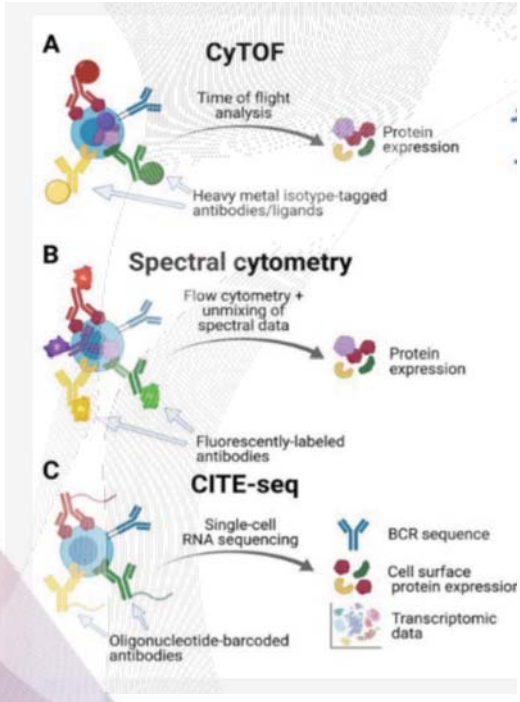
FACS (유세포 형광 분석기)

- 세포표현형을 분석하는 golden standard
- 세포 **표면마커**에 기반함.

FACS는 유세포 분석기의 분화한 타입이다.
이것은 **이질적으로 혼합된 세포들을 각각의 특정한 광산란과 형광 특징들에 기반하여 분류하는 방법**



What is the advantage of CITE-seq?

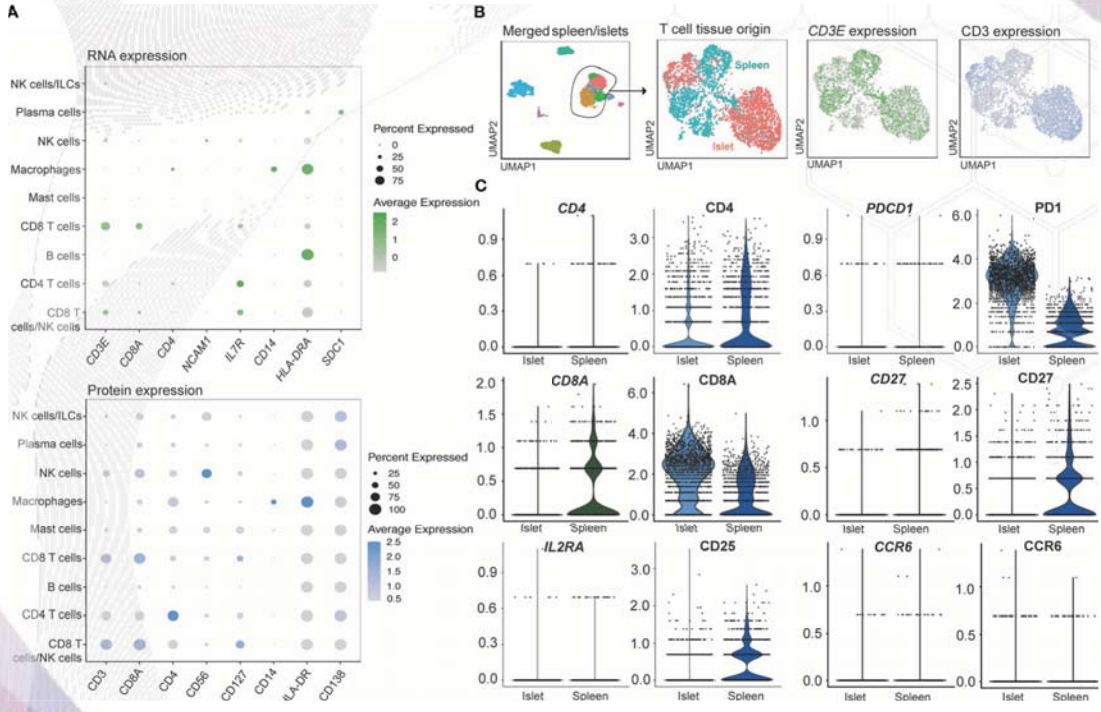


자연계에 존재하는 heavy metal동위원소 개수의 한계 존재 (~50)

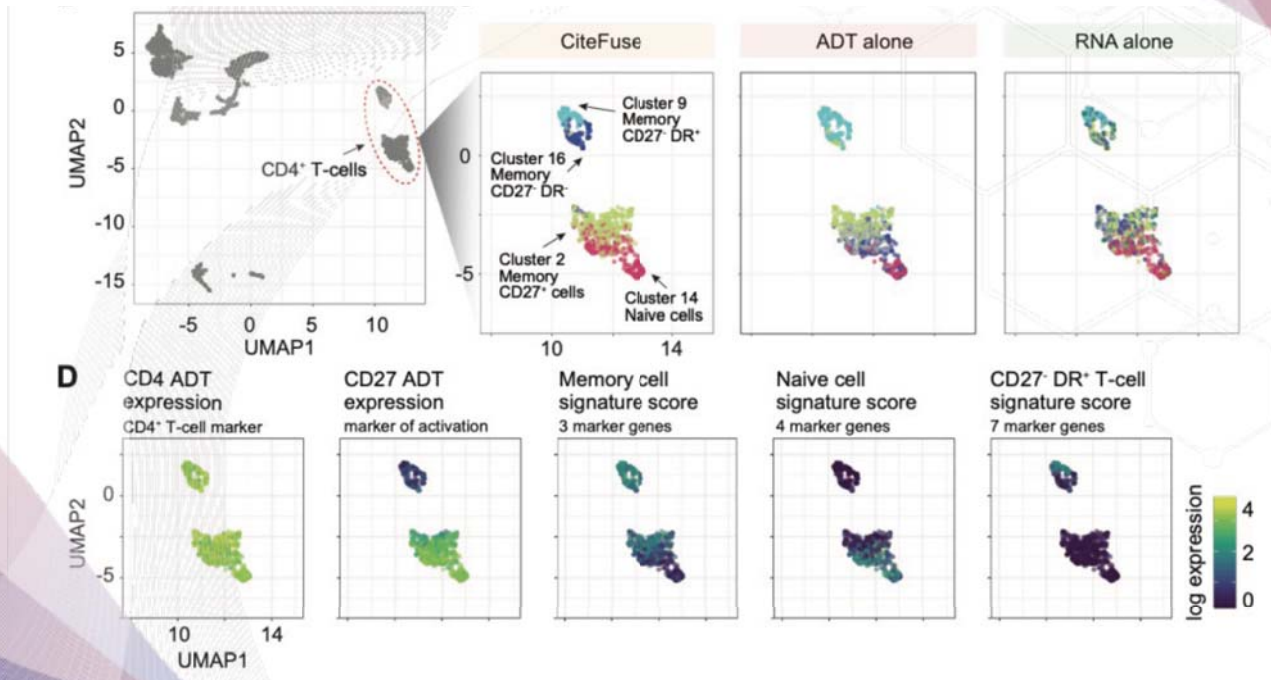
Spectral overlap해결이 어려움

Use of oligo sequence as a readout Is unlimited !!!

Both modalities are necessary to define clusters



CITE-seq enables novel cell type discovery



Future

- Ultra high-throughput multiomics technologies are coming along (ex: SCITO-seq, scifi-RNA-seq…)
- Trimodalities (ex: RNA+protein+ATAC..)
- Integration with public dataset (batch effect removal) + interpretation will be the key!

Closing

Thank you~!

Further readings:
Single-cell overview reading: Single-cell RNA sequencing technologies and bioinformatics pipelines | Experimental & Molecular Medicine (nature.com)
Single-cell multiomics:
https://www.nature.com/articles/s41580-023-00615-w

If you have any questions or inquiry about collaboration opportunity:
bjhwang113@yuhs.ac