KSBi-BIML 2026

Bioinformatics & Machine Learning(BIML) Workshop for Life Scientists

생명정보학 & 머신러닝 워크샵 (온라인)

Best practice for cancer immunity analysis

김권일 _ 경희대학교

KSBi-BIML 2026

Bioinformatics & Machine Learning (BIML) Workshop for Life Scientists

한국생명정보학회가 주최하는 BIML-2026 동계 Bioinformatics &Machine Learning 교육 워크숍에 여러분을 초대합니다.

BIML 워크숍은 생명정보학 연구자들이 최신 AI바이오 분야의 인공지능 기반 분석 기술과 바이오 데이터 분석 기법을 이론과 실습을 통해 체계적으로 배울 수 있는 전문 교육 프로그램입니다. 2015년에 시작된 BIML 워크숍은 올해로 12년 차를 맞이하며, 국내 생명정보학 분야의 최초이자 최고 수준의 교육 프로그램으로 자리 잡았습니다. 이번 워크숍은 크게 인공지능바이오(AI바이오) 분야와 디지털바이오 분야, 두 분야로 구성됩니다.

AI바이오 분야에서는 생명정보 분석에 폭넓게 응용되고 있는 다양한 인공지능 기반 자료 모델링 기법을 다룰 예정입니다. 특히, 인공지능 심층학습을 활용한 단백질 구조 예측, 유전체 분석, 신약 개발에 대한 이론 및 실습 강의가 진행됩니다.

또한 디지털바이오 분야에서는 단일세포오믹스, 공간오믹스, 멀티오믹스, 메타오믹스에 대한 강의도 마련되어 있어, 연구자들의 분석 역량 강화에 실질적인 도움을 줄 것으로 기대됩니다.

또한 2024년부터 추가된 의료정보 자료 분석을 다루는 강의를 올해도 지속해서 운영하고자 합니다. 이는 최근 의료정보 자료 분석에 관한 연구 수요 증가를 반영한 것으로, 관련 연구를 수행하는 의과학자 및 의료정보 연구자들에게 유용한 지침을 제공할 것입니다.

또한, 올해도 생명정보학 기술의 다양화에 발맞춰 온라인 강좌를 대폭 확대했습니다. 올해는 무료 강좌 10개를 포함한 총 40개 이상의 강좌가 개설되며, 연구 주제에 맞는 강좌 추천과 강연료 할인 혜택도 제공합니다.

BIML-2026는 국내 주요 연구 중심 대학의 전임 교수 및 각 분야 최고 전문가들의 강의로 구성되어 있으며, 기초 이론부터 최신 연구 동향까지 아우르는 심도 있는 교육의 장이 될 것으로 확신합니다.

여러분의 많은 관심과 참여를 기대합니다!

2026년 2월

한국생명정보학회장 류 성 호

Best practice for cancer immunity analysis

면역관문억제제를 포함한 면역항암치료는 현대 암 연구와 치료의 패러다임을 혁신적으로 변화시키고 있지만, 기존 치료와 마찬가지로 모든 환자에게 동일한 효과를 보이지는 않는다. 이에 따라 면역 저항성을 예측하고 극복하는 것이 중요한 도전 과제로 부각되고 있다. 이를 해결하기 위해서는 암 조직의 복잡성을 깊이 이해할 필요가 있다. 기존 연구가 암 세포의 유전적 이질성에 초점을 맞추어 진행되었다면, 최근에는 암 조직 및 그 주변에 분포하는 다양한 면역세포의 기능과 역할에 중점을 둔 연구가 활발히 진행되고 있다. 특히, 암과 면역계의 복잡한 상호작용을 이해하는 것은 암 면역(cancer immunity)을 이해하는 데 핵심적이며, 이를 위해 단일세포 오믹스(single cell omics)와 공간 오믹스(spatial omics) 분석 기술이 빠르게 발전하고 있다.

본 강의에서는 암 면역 분야에서 암세포와 면역세포 간의 상호작용을 분석한 최신 연구 사례를 방법론과 결과 해석에 중점을 두어 소개하고자 한다. 강의는 암 세포 면역성의 핵심인 신생항원 예측, 종양 침윤 면역세포의 특성화, 그리고 암세포와의 상호작용을 다룬 연구 내용을 포함한다. 해당 연구들은 exome-seq 및 RNA-seq 등 암 조직의 bulk sequencing data에서부터 scRNA-seq 등의 단일세포 오믹스(single cell omics) 데이터와 spatial transcriptomics 등의 공간 오믹스(spatial omics) 데이터를 포괄적으로 다루고 있다. 최종적으로는 암 조직에서 나타나는 cellular ecosystem을 특정하고, 이를 암 연구와 치료에 활용할 수 있는 분석 기술을 학습하는 것을 목표로 한다.

강의는 다음의 내용을 포함한다:

- Cancer immunity 주요 요소 및 기전
- Exome-seq 데이터를 활용한 HLA typing과 HLA-peptide binding prediction을 통한 신생항원 예측
- RNA-seq 데이터를 기반으로 한 digital cytometry 추론을 통한 면역미세환경 분석
- Single cell 및 spatial omics 데이터를 기반으로 한 세포 상태 궤적(cell status trajectory dynamics) 및 세포 간 상호작용(cell-cell interaction) 분석을 통한 세포 생태계(cellular ecosystem) 해석

* 강의 난이도: 초급

* 강의: 김권일 교수 (경희대학교 생물학과)

Curriculum Vitae

Speaker Name: Kwoneel Kim, Ph.D.

▶ **Personal Info**

Name	Kwoneel Kim
Title	Assistant Professor
Affiliation	Kyung Hee University

▶ **Contact Information**

Address	24, Kyungheedae-ro, Dongdaemun-gu, Seoul
Email	kwoneelkim@khu.ac.kr

Research Interest

Translational bioinformatics, Machine learning and computational genomics, Cancer genomics

Educational Experience

2009	B.S. Dept. Applied Bioscience, Konkuk University, Korea
2011	M.S. Dept. Functional Genomics, UST, Korea
2015	Ph.D. Dept. Bio and Brain Engineering, KAIST, Korea

Professional Experience

2015-2017	Post-Doctoral Researcher, Dept. Bio and Brain Engineering, KAIST
2017-2018	Senior Research Scientist, Asan Institute for Life Sciences, Asan Medical Center
2018-	Assistant Professor, Department of Biology, Kyung Hee University

Selected Publications (5 maximum)

1. Kim, H. B., Lee, S. H., Yang, D. Y., Lee, S. H., Kim, J. H., Kim, H. C., ... **Kim, K.*** & Hong, S. J.* (2024). PM exposure during pregnancy affects childhood asthma via placental epigenetic changes: neuronal differentiation and proliferation and Notch signaling pathways. *Environmental Pollution*, 125471. *Co-corresponding

2. Song, K. J*., Choi, S.*, **Kim, K.***, Hwang, H. S.*, Chang, E., Park, J. S., Shim, S. B., Choi, S., Heo, Y. J., An, W. J., Yang, D. Y., Cho, K. C., Ji, W., Choi, C. M., Lee, J. C., Kim, H. R., Yoo, J., Ahn, H. S., Lee, G. H., Hwa, C., ... Kim, K. P. (2024). Proteogenomic analysis reveals non-small cell lung cancer subtypes predicting chromosome instability, and tumor microenvironment. *Nature communications*, 15(1), 10164. *Co-first

3. Bang, H., Park, J. S., Kim, J. Y., Sung, C., An, J., Cho, D. Y., ... & **Kim, K.** Gene essentiality for tumour growth influences neoantigen-directed immunoediting. *Clinical and Translational Medicine*, 12(1). (2022)

4. Kim J-H*, **Kim K***, Yeom J*, Lee E, Kang M-J, Lee S-H, Kim K, Lee S-Y, Hong S-B, Oh DK, Lee K, Choi, S-J, Yang M-J, Kim J, Hong S-J. Integrative multi-omics approach for mechanism of humidifier disinfectant-associated lung injury. *Clinical and Translational Medicine*, 11, e562 (2021) *Co-first

5. **Kim K**, Kim HS, Jeong YK, Jung H, Sun J-M, Ahn JS, Ahn M-J, Park K, Lee S-H, Choi JK. Predicting clinical benefit of immunotherapy by antigenic or functional mutations affecting tumour immunogenicity. *Nature Communications*, 11. 951 (2020).

KSBi-BIML 2025

Best practice for cancer immunity analysis

Department of Biology, Kyung Hee University

Kwoneel Kim, PhD

강의개요

Best practice for cancer immunity analysis

면역관문억제제를 포함한 면역항암치료는 현대 암 연구와 치료의 패러다임을 혁신적으로 변화시키고 있지만, 기존 치료와 마찬가지로 모든 환자에게 동일한 효과를 보이지는 않는다. 이에 따라 면역 저항성을 예측하고 극복하는 것이 중요한 도전 과제로 부각되고 있다. 이를 해결하기 위해서는 암 조직의 복잡성을 깊이 이해할 필요가 있다. 기존 연구가 암 세포의 유전적 이질성에 초점을 맞추어 진행되었다면, 최근에는 암 조직 및 그 주변에 분포하는 다양한 면역세포의 기능과 역할에 중점을 둔 연구가 활발히 진행되고 있다. 특히, 암과 면역계의 복잡한 상호작용을 이해하는 것은 암 면역(cancer immunity)을 이해하는 데 핵심적이며, 이를 위해 단일세포 오믹스(single cell omics)와 공간 오믹스(spatial omics) 분석 기술이 빠르게 발전하고 있다.

본 강의에서는 암 면역 분야에서 암세포와 면역세포 간의 상호작용을 분석한 최신 연구 사례를 방법론과 결과 해석에 중점을 두어 소개하고자 한다. 강의는 암 세포 면역성의 핵심인 신생항원 예측, 종양 침윤 면역세포의 특성화, 그리고 암세포와의 상호작용을 다룬 연구 내용을 포함한다. 해당 연구들은 exome-seq 및 RNA-seq 등 암 조직의 bulk sequencing data에서부터 scRNA-seq 등의 단일세포 오믹스(single cell omics) 데이터와 spatial transcriptomics 등의 공간 오믹스(spatial omics) 데이터를 포괄적으로 다루고 있다. 최종적으로는 암 조직에서 나타나는 cellular ecosystem을 특정하고, 이를 암 연구와 치료에 활용할 수 있는 분석 기술을 학습하는 것을 목표로 한다.

강의는 다음의 내용을 포함한다:

- Cancer immunity 주요 요소 및 기전
- Exome-seq 데이터를 활용한 HLA typing과 HLA-peptide binding prediction을 통한 신생항원 예측
- RNA-seq 데이터를 기반으로 한 digital cytometry 추론을 통한 면역미세환경 분석
- Single cell 및 spatial omics 데이터를 기반으로 한 세포 상태 궤적(cell status trajectory dynamics) 및 세포 간 상호작용(cell-cell interaction) 분석을 통한 세포 생태계(cellular ecosystem) 해석

* 강의 난이도: 초급

* 강의: 김권일 교수 (경희대학교 생물학과)

2

Cancer immunity 주요 요소 및 기전: Cancer-Immunity Cycle

암-면역 주기는 암 세포와 면역 세포 간의 상호 작용을 설명하는 일련의 단계로 이루어짐

1. 신생항원 (neoantigen) 방출: 암세포가 사멸하면서 신생항원 방출

2. 항원 제시(antigen presentation): Dendritic cells (DC) 와 같은 항원 제시 세포가 신생항원을 포획하여 T cell 에 제시

3. T cell 활성화 및 증식 (T cell priming and activation): Lymph node 에서 DC 에 의해 활성화된 T cell 은 cancer cell 을 공격할 수 있는 effector T cell 로 분화하고 증식

4. 종양으로의 T cell 이동 및 종양 침투 (Tumor infiltration): 활성화된 T cell 은 혈류를 통해 tumor microenvironment (TME) 로 이동, 종양 조직으로 침투하여 cancer cell 을 인식하고 공격

5. Cancer cell 사멸: T cell 은 cancer cell 을 직접적으로 사멸시키거나 면역 반응을 통해 암세포를 제거
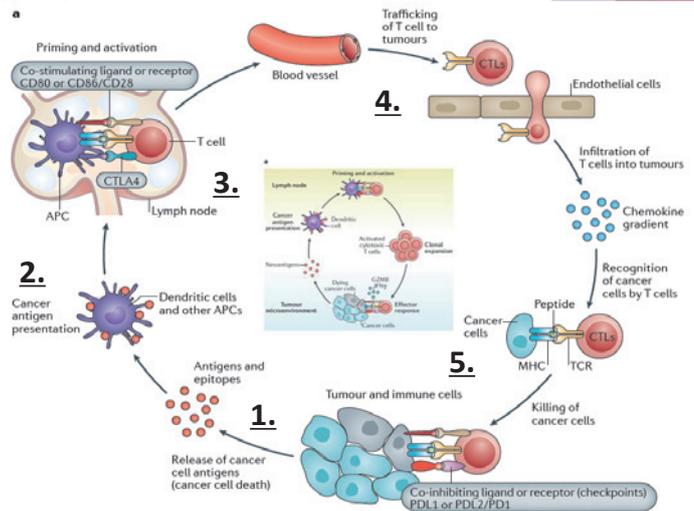
Nature Rev Genet. 17, 441-458 (2016)

3

Cancer immunity 주요 요소 및 기전: Tumor Microenvironment, TME

면역미세환경 (Tumor Microenvironment, TME) 주요 요소는 아래와 같음

- Dendritic cells, DCs: Cancer neoantigen 을 T cell에 제시
- Cytotoxic T lymphocytes, CTLs: 암세포를 직접적으로 사멸시키는 주요 immune cell
- Natural killer cells, NK cells: Cancer cell 을 인식하고 사멸시키는 선천 면역 담당 역할
- Regulatory T cells, Tregs: 면역 반응을 억제하는 역할을 하는 T cell
- Macrophages: TME 에서 다양한 역할을 수행하며, cancer cell 성장을 촉진하거나 억제

Cancer cell 은 TME 와 아래와 같은 방식으로 소통하여 면역 반응을 회피할 수 있음

- 면역 억제 (immunosuppression): immune cell 과의 crosstalk 을 통하여 면역 억제 환경을 조성하여 면역 세포의 활성을 억제하고 면역 회피를 촉진
- 혈관신생 (angiogenesis): 성장에 필요한 영양분과 산소를 공급받기 위해 혈관신생을 유도
- 세포 외 기질 (extracellular matrix, ECM): 세포 외 기질은 암세포의 성장, 침윤 및 전이를 촉진할 수 있음
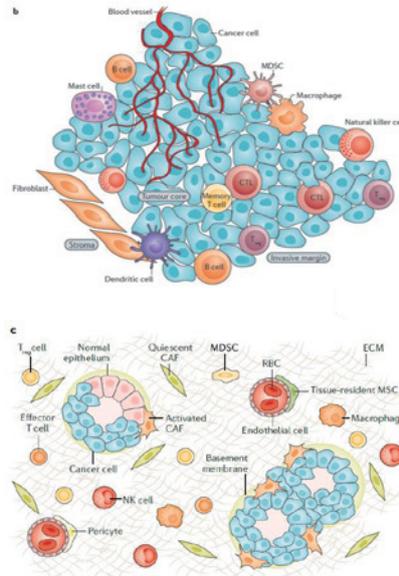
Nature Rev Genet. 17, 441-458 (2016)

4

Cancer immunity 주요 요소 및 기전: Neoantigen and immune checkpoint inhibitor (ICI) therapy

Immune checkpoint (IC) 는 면역 반응을 조절하는 분자로, 면역 항암 치료의 주요 표적
- **CTLA-4 (Cytotoxic T-lymphocyte-associated protein 4)**: T cell 활성화를 억제하는 IC
- **PD-1 (Programmed cell death protein 1)**: T cell 의 활성과 기능을 억제하는 IC
- **PD-L1 (Programmed death-ligand 1)**: PD-1에 결합하여 T cell 의 활성을 억제하는 IC ligand

신생항원 (neoantigen) 은 cancer cell 에서 발생하는 mutation 으로 인해 생성되는 새로운 항원
- **종양 특이적**: 정상 세포에는 존재하지 않으므로 면역 시스템에 의해 외부 물질로 인식되어 면역 반응을 유발
- **면역 치료 표적**: 면역 치료, 특히 cancer vaccine 및 T cell therapy 의 핵심 표적
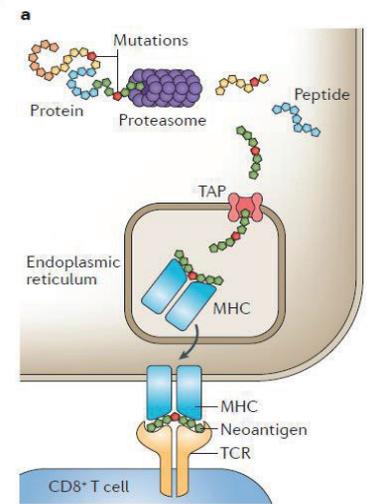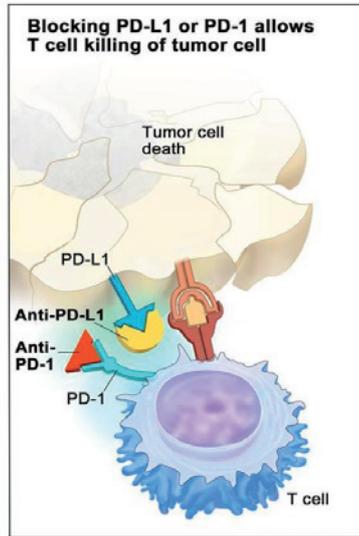
Nature Rev Genet. 17, 441-458 (2016)

5

Cancer immunity 주요 요소 및 기전: TME and ICI therapy

#Immune contexture = cellular ecosystem(niche)
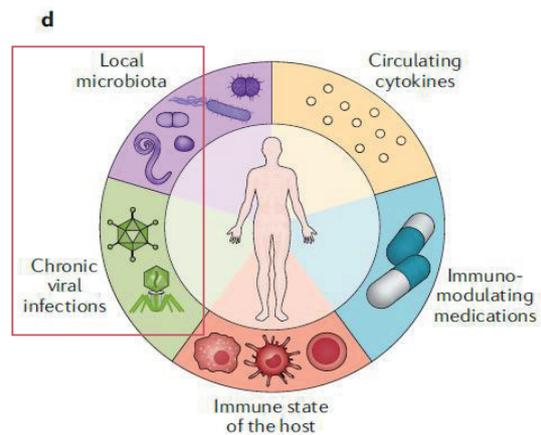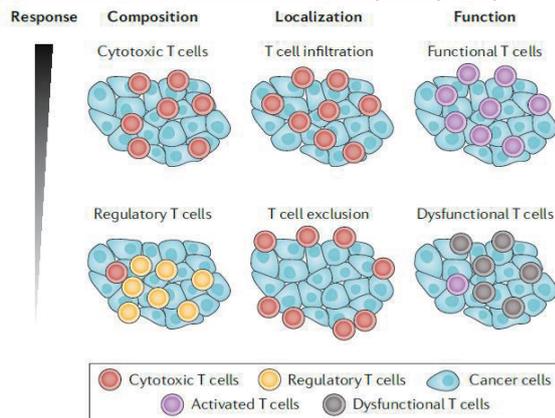
Nature Rev Genet. 20, 724-746 (2019)

6

Cancer immunity 주요 요소 및 기전 연구를 위한 분석 pipelines
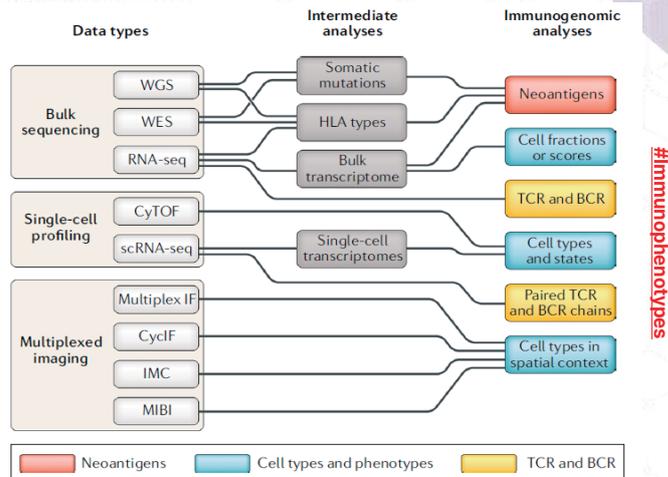
Fig. 2 | **Overview of technologies and analyses for interrogating cancer immunity.** Technologies for bulk sequencing allow the prediction of candidate neoantigens and deconvolution of cell fractions or computation of abundance scores using marker-gene-based approaches. Technologies for single-cell profiling allow the characterization of cell types and states from mass cytometry by time of flight (CyTOF) or single-cell RNA sequencing (scRNA-seq) data, but also the reconstruction of B cell receptors (BCRs) and T cell receptors (TCRs) of the same cells. Recent multiplexed imaging techniques can interrogate several cellular markers, enabling the phenotyping of distinct cell types and the reconstruction of the spatial architecture of the tumour microenvironment. CycIF, cyclic immunofluorescence; IF, immunofluorescence; IMC, imaging mass cytometry; MIBI, multiplexed ion beam imaging; RNA-seq, RNA sequencing; WES, whole-exome sequencing; WGS, whole-genome sequencing.

Nature Rev Genet. 20, 724-746 (2019)

7

Cancer immunity 주요 요소 및 기전 연구를 위한 분석 pipelines
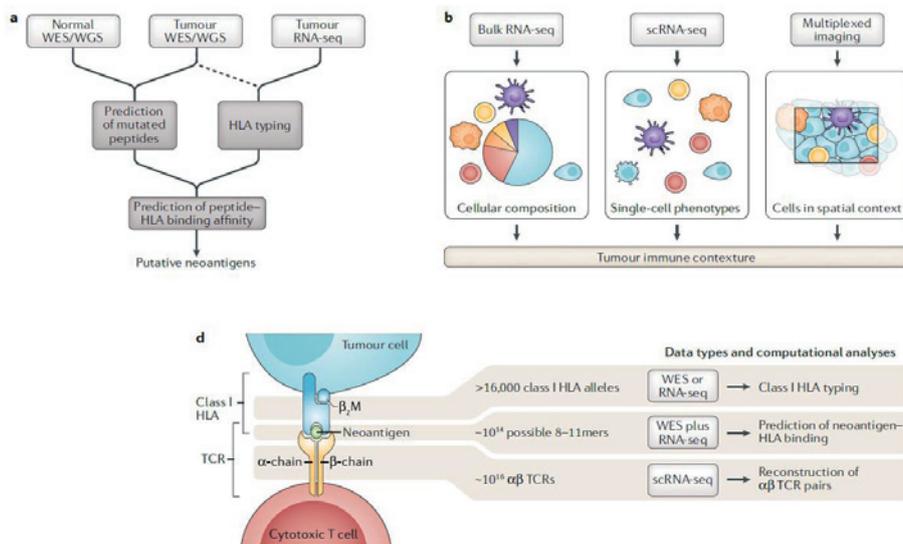
Nature Rev Genet. 20, 724-746 (2019)

8

Cancer immunity 주요 요소 및 기전 연구를 위한 data

- **TCGA (The Cancer Genome Atlas)**: 30가지 이상의 암 유형에 대한 임상 정보, 유전체 특성 데이터, 고수준 시퀀스 분석을 제공하는 가장 포괄적인 암 유전체 데이터베이스
- **COSMIC (Catalogue of Somatic Mutations in Cancer)**: 인간 암에서 발생하는 체세포 돌연변이에 대한 정보를 제공
- **cBioPortal**: 대규모 암 유전체 데이터 세트를 탐색, 분석, 다운로드할 수 있는 웹 서버. cBioPortal은 사용자 친화적인 인터페이스를 통해 복잡한 암 유전체 데이터를 쉽게 분석하고 시각화할 수 있도록 지원
- **IEDB (Immune Epitope Database and Analysis Resource)**: 실험적으로 검증된 면역원성 epitope 예측 및 분석을 위한 도구들을 제공하는 데이터베이스
- **ImmGen (Immunological Genome Project)**: 면역 세포의 유전자 발현 및 조절 네트워크에 대한 포괄적인 정보를 제공하는 프로젝트

Nature Rev Genet. 17, 441-458 (2016)

9

Cancer immunity 주요 요소 및 기전 연구를 위한 data

- **TCIA (The Cancer Immunome Atlas):** 20가지 고형암에 대한 포괄적인 면역유전체 분석 결과를 제공하며, 면역항암제 치료를 받은 환자 데이터도 포함. TCIA는 유전자 발현 데이터, HLA 유형 (제한적 접근), neoantigen, 면역 세포 비율, GSEA 및 디콘볼루션 점수, 유전체 및 임상 특징 데이터를 제공. 또한, 데이터 다운로드, 동적 시각화 및 분석 (예: 생존 분석) 기능도 제공
- **iAtlas**: 33가지 암 유형에 걸쳐 10,000개 이상의 TCGA 종양 샘플에 대한 면역유전체 데이터에 접근할 수 있는 웹 포털. iAtlas는 유전체, 면역학적 및 임상적 특징에 대한 접근을 제공하며, 데이터 분석 및 시각화 도구에 대한 간편한 접근을 제공

Nature Rev Genet. 20, 724-746 (2019)

10

Cancer immunity 주요 요소 및 기전 연구를 위한 analysis tool

Nature Rev Genet. 17, 441-458 (2016)
Rev Genet. 20, 724-746 (2019)

Cancer immunity 주요 요소 및 기전 연구를 위한 analysis tool

Net 20, 724-746 (2019)

12

Cancer immunity 주요 요소 및 기전 연구를 위한 analysis tool

Cancer immunity 주요 요소 및 기전 연구를 위한 analysis tool

Slide 15

OptiType Procedure:
1. Input: FASTQ raw data (single-end, paired-end both available)
2. Output: 4-digit HLA genotype for input sample + confidence score
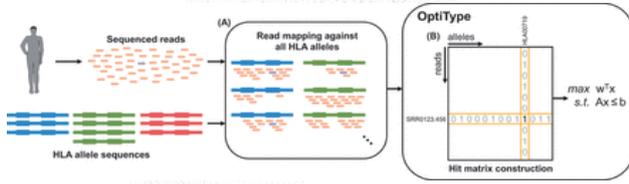
Usage

Optional step zero: you might want to filter your sequencing data for HLA reads. Should you have to re-run OptiType multiple times on the same sample (different settings, etc.) it could save you time if instead of giving OptiType the full, multi-gigabyte sequencing data multiple times, you would rather give it the relevant reads only, on the order of megabytes.

You can use any read mapper to do this step, although we suggest you use RazerS3. Its only drawback is that due to way RazerS3 was designed, it loads all reads into memory, which could be a problem on older, low-memory computing nodes.

Make sure to filter your files correctly depending on whether you have DNA (exome, WGS) or RNA-Seq data. The reference fasta files are `data/hla_reference_dna.fasta` and `data/hla_reference_rna.fasta` respectively. Below is an example for DNA sequencing data:

```
>razers3 -i 95 -m 1 -dr 0 -o fished_1.bam /path/to/OptiType/data/hla_reference_dna.fasta sample_1.fastq
>samtools bam2fq fished_1.bam > sample_1_fished.fastq
>rm fished_1.bam
```

If you have paired-end data, repeat this with the second ends' fastq file as well. Note: it's important that you filter the two ends individually. Don't use the read mapper's paired-end capabilities.

After the optional filtering, OptiType can be called as follows:

```
>python /path/to/OptiTypePipeline.py -i sample_fished_1.fastq [sample_fished_2.fastq]
                    (--rna | --dna) [--beta BETA] [--enumerate N]
                    [-c CONFIG] [--verbose] --outdir /path/to/out_dir/
```

This will produce a time-stamped directory inside the specified output directory containing a CSV file with the predicted optimal (and if enumerated, sub-optimal) HLA genotype, and a pdf file containing a coverage plot of the predicted alleles for diagnostic purposes.

15

Slide 16
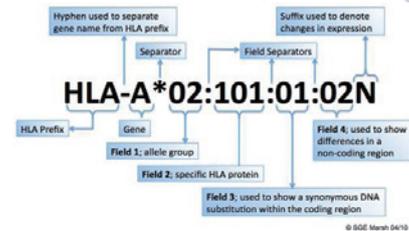
Field 3 (6-digit) 부터는 단백질 서열에는 영향을 주지 않는 유전자 level의 변이를 나타내는 (synonymous) 것이기 때문에 4-digit 까지가 가장 일반적으로 HLA 차이를 나타내는 방법입니다.

16

HLA typing과 HLA-peptide binding prediction을 통한 신생항원 예측: mutation calling

돌연변이 호출(Mutation Calling)Mutect2 실행: GATK의 Mutect2 도구를 사용하여 종양과 정상 샘플 간의 체세포 변이(somatic mutation)를 식별합니다.

돌연변이 단백질 동정주석(annotation): 필터링된 VCF 파일을 Ensembl VEP 또는 ANNOVAR와 같은 도구를 사용하여 주석을 추가하고, 변이가 단백질 수준에서 어떤 영향을 미치는지 확인합니다.

단백질 서열 예측: 변이로 인해 생성된 변이 단백질의 아미노산 서열을 예측합니다.

```bash
                                    📋 코드 복사
gatk Mutect2 \
  -R reference.fa \
  -I tumor.bam \
  -I normal.bam \
  --normal-sample normal_sample_name \
  --germline-resource af-only-gnomad.vcf.gz \
  --panel-of-normals pon.vcf.gz \
  -O somatic.vcf.gz
```

```bash
                                    📋 코드 복사
transvar ganno --vcf input.vcf --refseq --fasta reference.fa --out output.fasta
```


Current prediction of MHC class I binding neoantigens (NetMHC)

~60,000 data in The Immune Epitope Data Base (IEBD), Nucleic Acids Res. 38, D854-62 (2010)

MHC sequence (34 polymorphic residues)
ACAMYRNNM..

Peptide sequence (9 mers)
AEMKTDAA

Hidden layer [Artificial Neural Network]

Output = 29766.651099
(Non binding)

HLA typing과 HLA-peptide binding prediction을 통한 신생항원 예측: Neoantigen load and immunotherapy response

Cohort name	Tumor type	Cohort size	Target checkpoint	Reference
SMC	Lung cancer	122	PD-1/PD-L1	This work
Hellmann	Lung cancer	75	PD-1 & CTLA-4	Ref 25
Rizvi	Lung cancer	34	PD-1	Ref 24
Van Allen	Melanoma	110	CTLA-4	Ref 21
Snyder	Melanoma	64	CTLA-4	Ref 22
Roh	Melanoma	56	PD-1 & CTLA-4	Ref 10
Riaz	Melanoma	68	PD-1	Ref 12

Nature Comm. 11, 951 (2020)

Slide 19

HLA typing과 HLA-peptide binding prediction을 통한 신생항원 예측:
Therapeutic potential of neoantigen
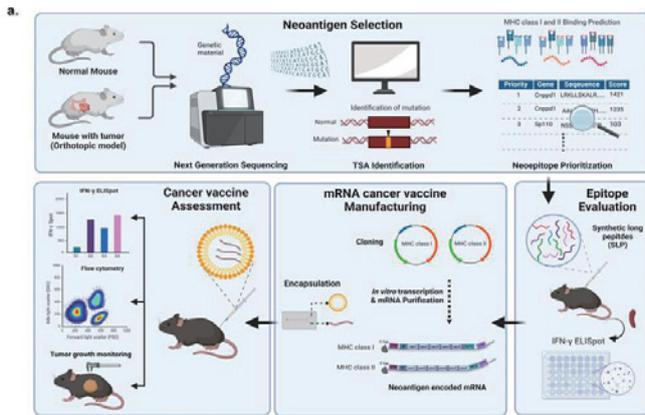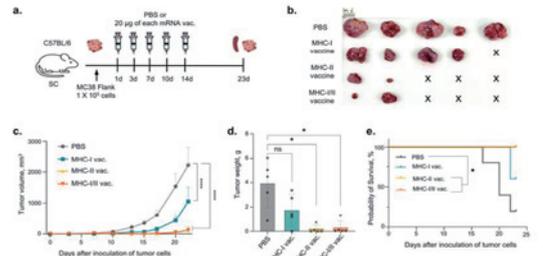
Advanced Science. In revision

Slide 20

Digital cytometry 추론을 통한 면역미세환경 분석
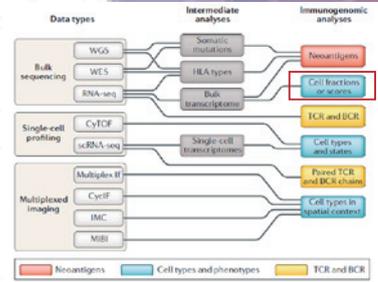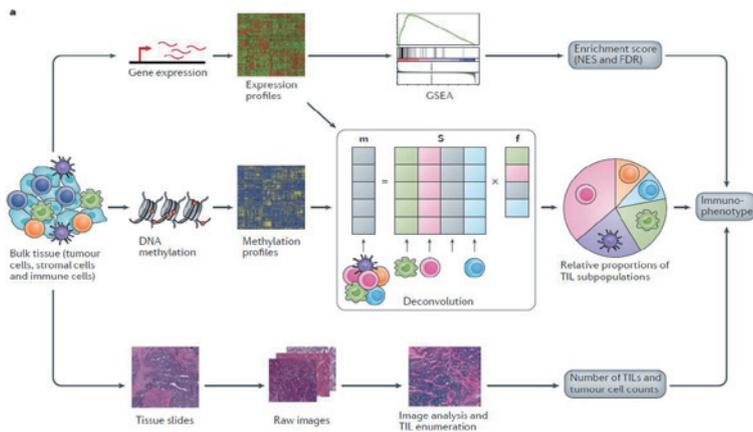
개별 세포 집단의 expression profile matrix를 대상으로..

1. GSEA 기반: 특정 면역 세포와 관련된 gene expression signature 를 통한 enrichment test 수행

2. Deconvolution 기반: 세포 유형별 gene expression reference 를 통한 각 세포 유형의 비율을 추론

Digital cytometry 추론을 통한 면역미세환경 분석: CybersortX

Cibersortx pipeline

Step 1 Signature matrix

IF each column of 'single cell reference matrix' is a phenotype label, proceed to step 2

```
<From cibersortx website>
The reference sample file is an input file required for custom signature genes file generation by CIBER
```

ELSE take the following two files and convert them using 'convert_reference_file.py' single cell reference matrix class matrix

```
python convert_reference_file.py MCA_liver_cell_expression.tsv MCA_liver_cell_class.tsv
```

Step 2 Run Cibersortx to create signature matrix

⭐ parameters (pro tem)
Single cell input options:
Min. Expression 0.5
▌Download output signature matrix

Step 3 Filter Mixure Matrix

Filter out columns with no expressions, using the signature matrix from step 2

```
python filter_by_signature_genes.py signature_gene_file.txt mixture_file.tsv
```

This creates a new mixture file named {mixture_file}_filtered.tsv

Step 4 Run Cibersortx to impute cell fractions

Upload the Mixture matrix from step 3 to the Cibersortx website, then run fraction imputation
⭐ parameters
default

📝 Memo
For the mouse liver data set, 19133+ samples exceeded the 'Allowed memory size of 536870912 bytes'
Use something like the following command to divide data set

```
# first 10000 samples
cat mixture_file_filtered.tsv | cut -f 1-10001 > mixture_file_filtered_1.tsv
# the rest of the samples (if <10000); first column == gene names (index)
cat mixture_file_filtered.tsv | cut -f 1,10002- > mixture_file_filtered_2.tsv
```

And to check the number of columns per file

```
cat file.tsv | awk '{print NF}'| sort -nu | tail -n 1
```

GeneSymbol	M1	Malignant	Oligodendrocyte
A1BG	0	0	0
A1BG-AS1	0	0.24001	0
A1CF	0	0	0

Gene	TCGA_02_0047_01A	TCGA_02_0055_01A	TCGA_02_2483_01A
A1BG	125.0069	391.8038	271.8522
A1BG-AS1	105.3013	162.1976	109.7288
A1CF	0	0	0

Mixture	M1	Malignant	Oligodendrocyte
TCGA_02_0047_01A	0.101649077	0.353447998	0.072751936
TCGA_02_0055_01A	0.17932788	0.261214401	0.026945961
TCGA_02_2483_01A	0.094886977	0.600738156	0.016771461

21

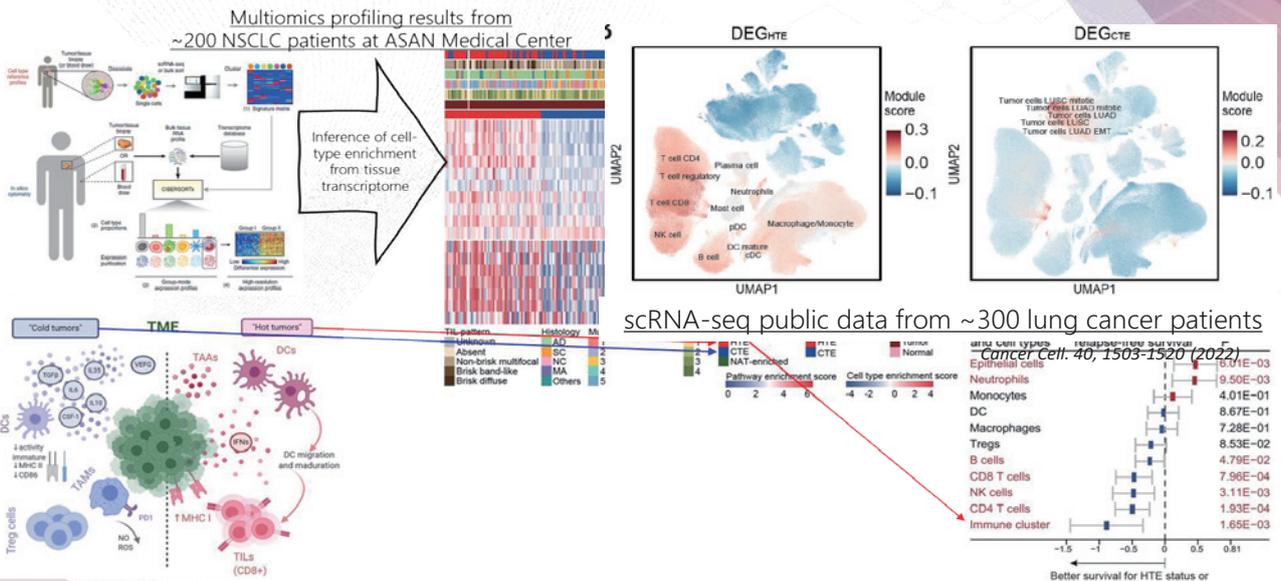## Digital cytometry 추론을 통한 면역미세환경 분석: Immune landscape associated with the load of neoantigens
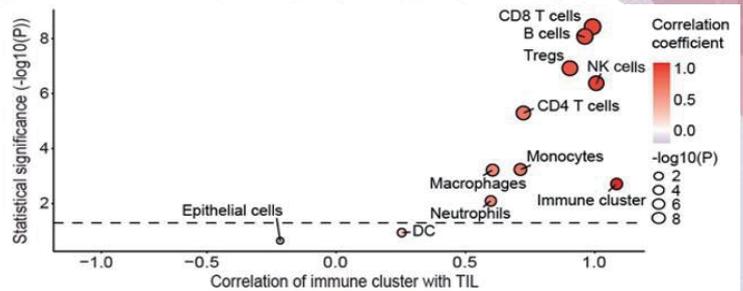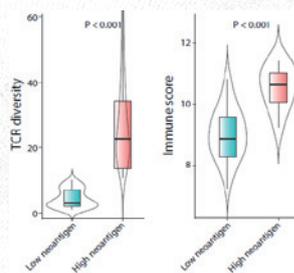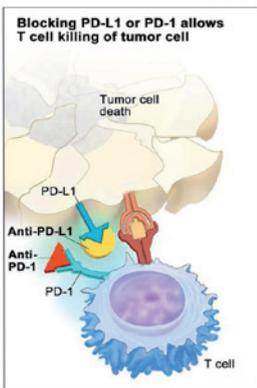

Nature Comm. 11, 951 (2020)
Comm. 15, 10164 (2024)

22

- 11 -

Digital cytometry 추론을 통한 면역미세환경 분석:
hot tumor enriched (HTE) and cold tumor enriched (CTE)

Multiomics profiling results from ~200 NSCLC patients at ASAN Medical Center

scRNA-seq public data from ~300 lung cancer patients

Cancer Cell. 40, 1503-1520 (2022)

Nature Comm. 15, 10164 (2024)

Digital cytometry 추론을 통한 면역미세환경 분석:
Patient prognosis was stratified by incorporating TME criteria into neoantigen
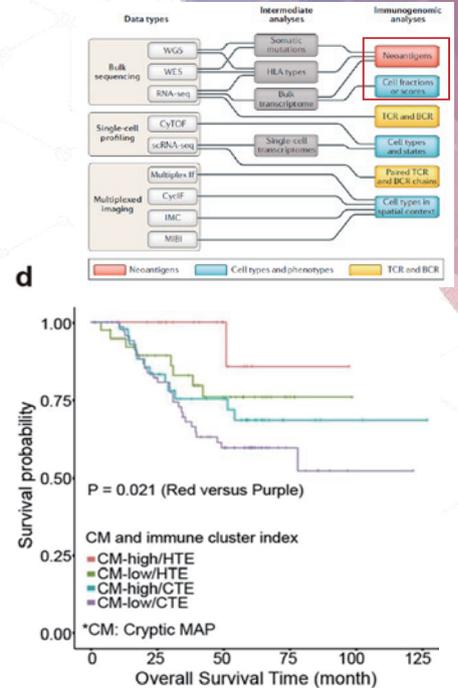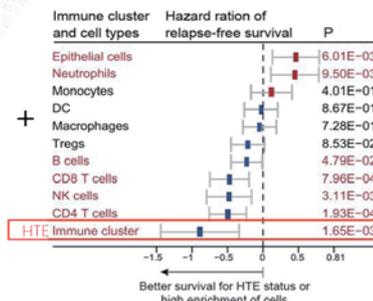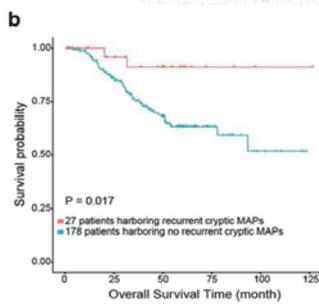
Nature Comm. 15, 10164 (2024)

From the limitations of bulk RNA-seq

> We can only estimate the average expression level for each gene across a population of cells, without regard for the heterogeneity in gene expression across individual cells of that sample.
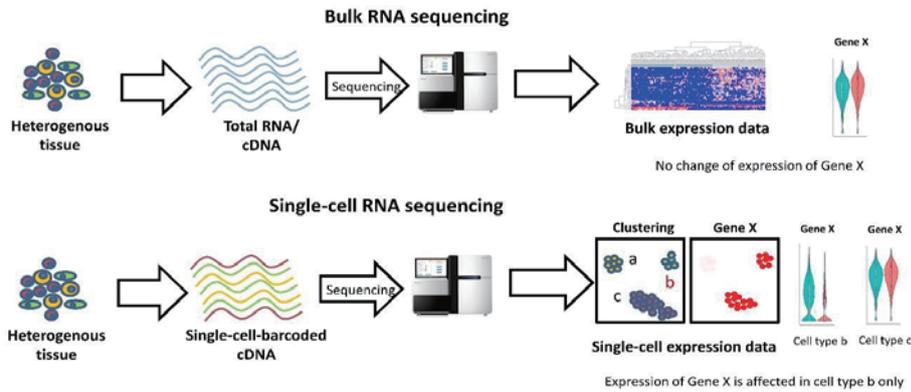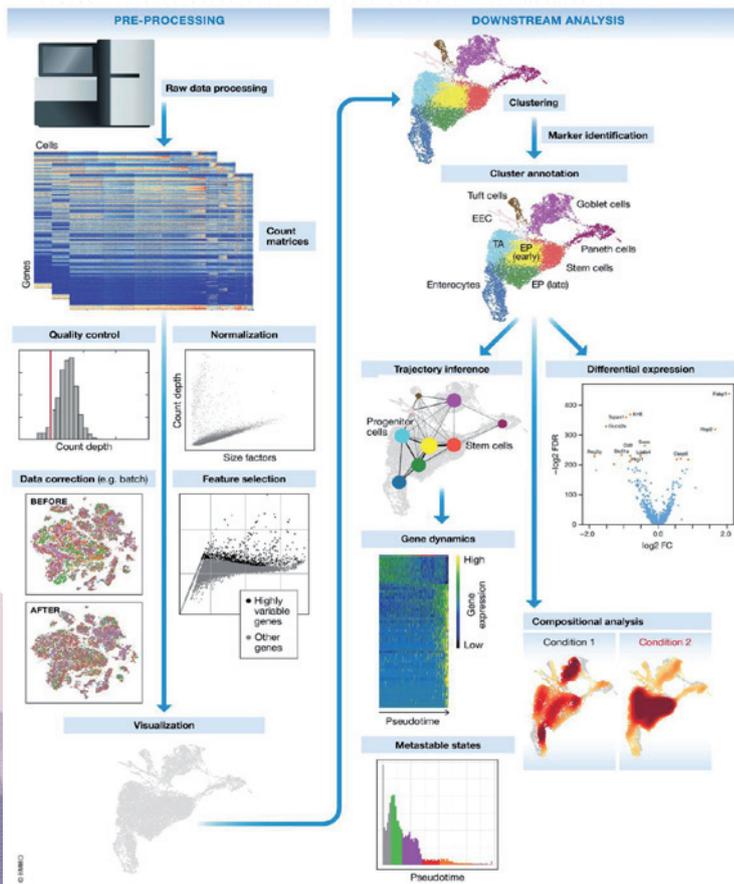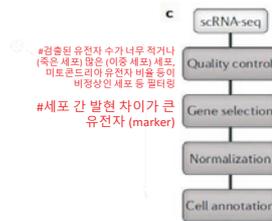
> Therefore, it is insufficient for studying heterogeneous systems, e.g. early development studies or complex tissues such as the brain.
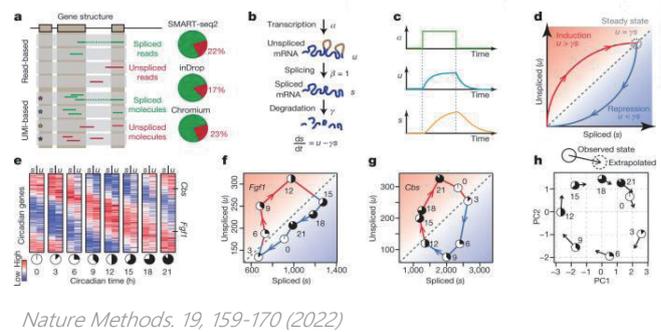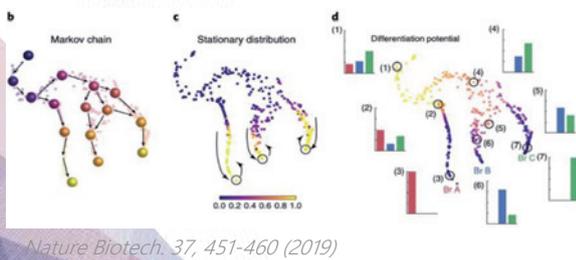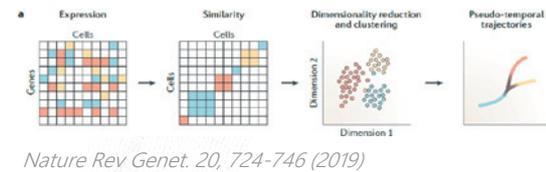
scRNA-seq analysis pipeline

Characterization of cell fate probabilities in single-cell data

> Expression similarity 기반으로 세포 (node) 간 network 구축

> Differentiation potential (DP): define a cell's differentiation potential (DP) to be the entropy over the branch probabilities, providing a quantitative metric for cell plasticity. 이 때, start cell은 prior-knowledge 기반으로 user 가 선택 (from cell annotation)

> Spliced RNA transcripts의 비율을 통해 세포의 future gene expression을 예측하는 RNA velocity 중심으로 여러가지 feature 를 통합

Nature Rev Genet. 20, 724-746 (2019)

Nature Methods. 19, 159-170 (2022)

Nature Biotech. 37, 451-460 (2019)

# Inferring cell-cell interactions

**Fig. 1: Overview of CellChat.**
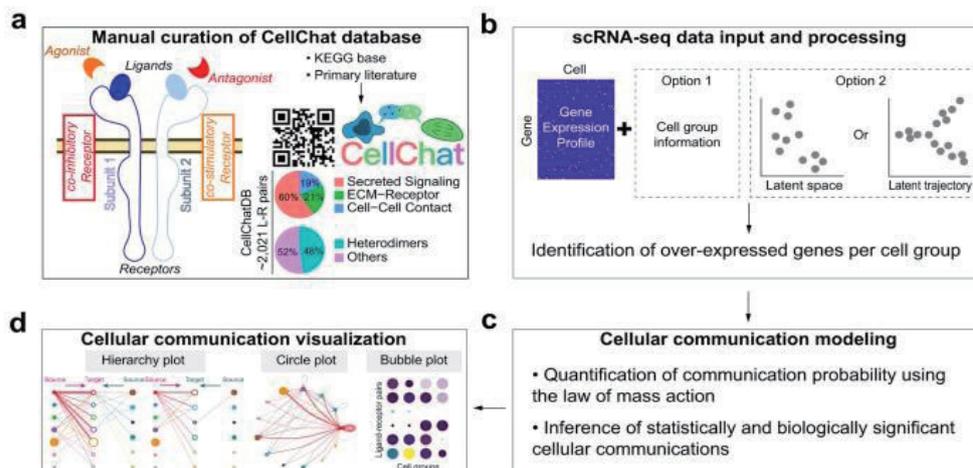
Nature Comm. 12, 1088 (2021)

More localized: spatially contextualizing cells

> Cell location affects intercellular interactions and associated gene expression.

> Molecules mediating CCIs form concentration gradients as they **diffuse** from their producing cells and trigger different signalling programmes in receiver cells.

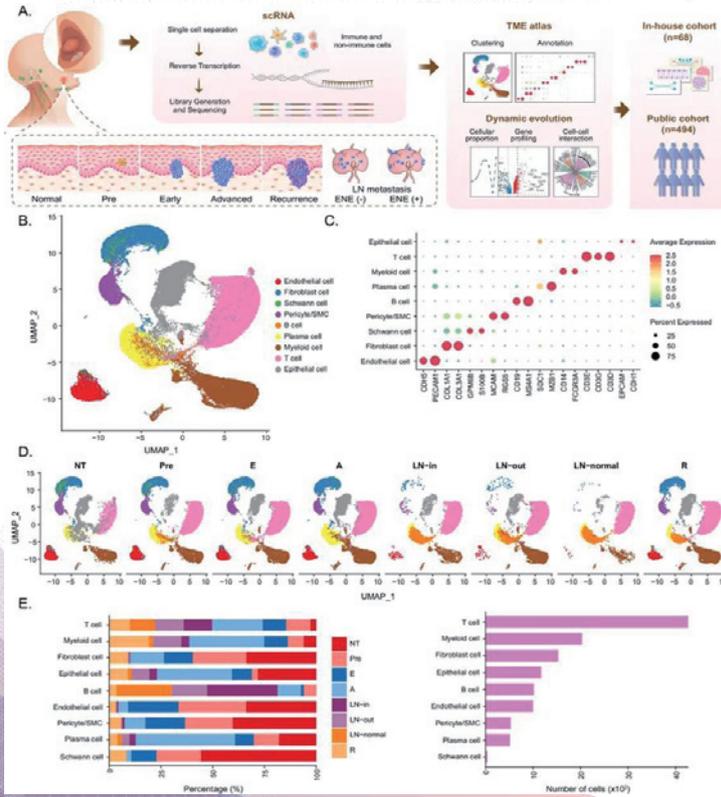> It is important to account for the **spatial context of cells** to understand how tissues function.

Spatial constrain reduces false positive interactions!
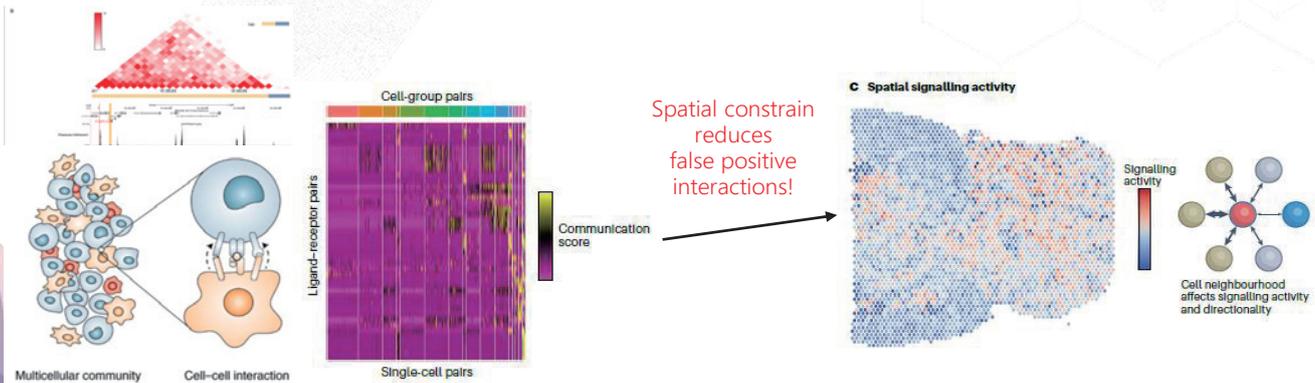
> After creating a representation of multicellular neighbourhoods, an important goal is to identify **recurrent cell state communities** across spatial domains, samples, or individuals. We refer to recurrent multicellular neighbourhood phenotypes, characterized by correlated cell states, as **'spatial ecotypes' (= "cellular niche").**

> The definitions of transcriptional ecotypes and spatial ecotypes also coincide with **'multicellular programmes'**, which are sets of transcriptional states across various cell types **that are co-associated in multiple samples or spatial regions.**
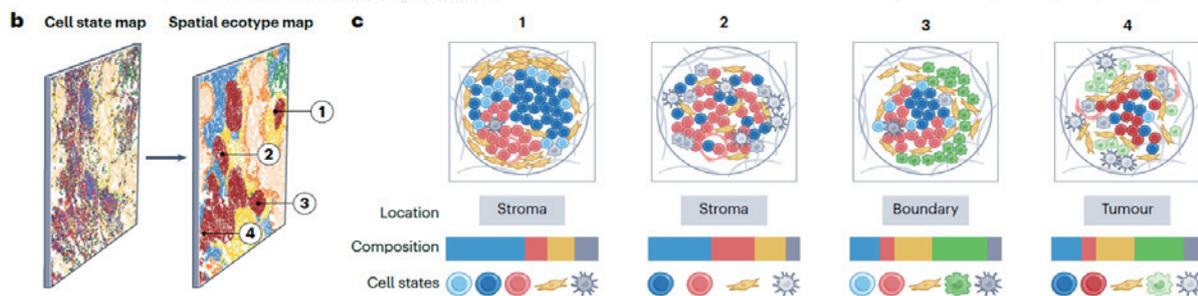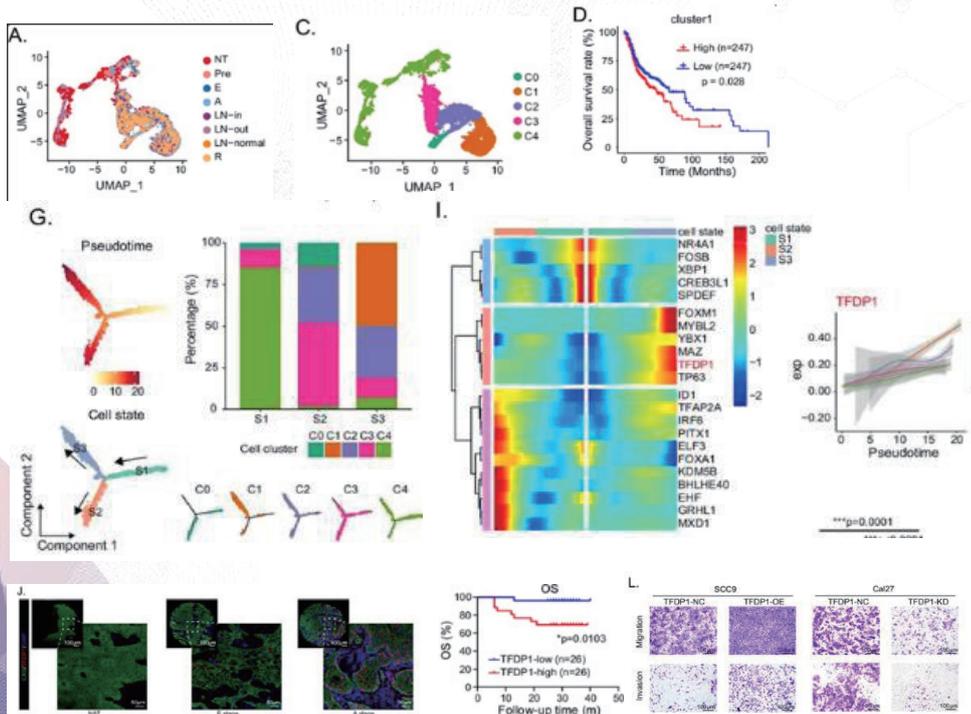
Case study: Single cell deciphering of progression trajectories of the tumor ecosystem in head and neck cancer
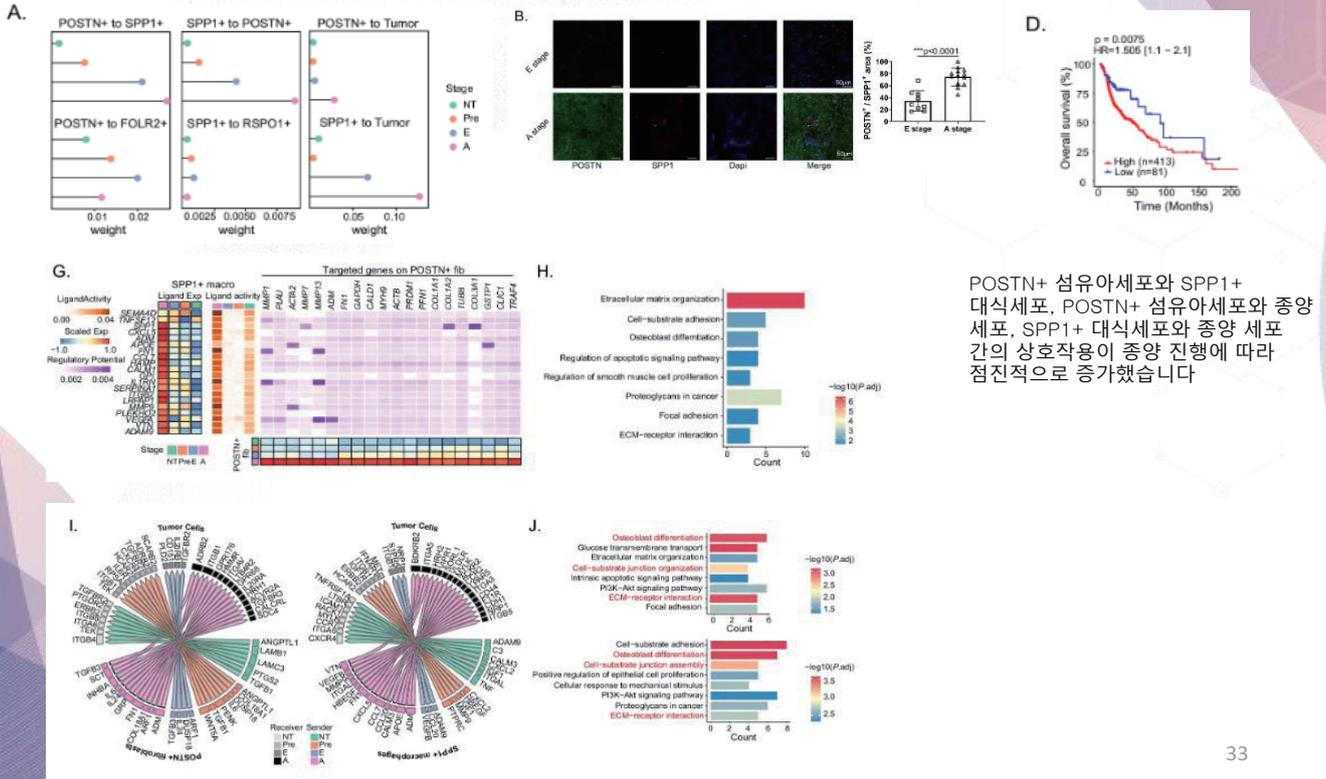
악성 상피 세포 아형의 특성

POSTN+ 섬유아세포, SPP1+ 대식세포 및 악성 세포의 세포 간 상호작용

POSTN+ 섬유아세포와 SPP1+ 대식세포, POSTN+ 섬유아세포와 종양 세포, SPP1+ 대식세포와 종양 세포 간의 상호작용이 종양 진행에 따라 점진적으로 증가했습니다

33

Summary

암-면역 상호작용은 암세포와 면역세포 간 신생항원 방출, 항원 제시, T 세포 활성화, 종양 침투, 암세포 사멸 단계로 구성됨.

면역 미세환경(TME)에서는 세포 간 상호작용과 신호 전달이 면역 억제, 혈관 신생, 세포외기질 형성으로 암 생존과 전이를 지원함.

최신 분석 도구들은 단일세포 RNA 및 공간 전사체 데이터를 통해 세포의 상호작용 네트워크와 질병 특이적 신호 경로를 밝혀내는 데 활용되고 있음.

34