COPD school 2026

Multi-omics Analysis in COPD



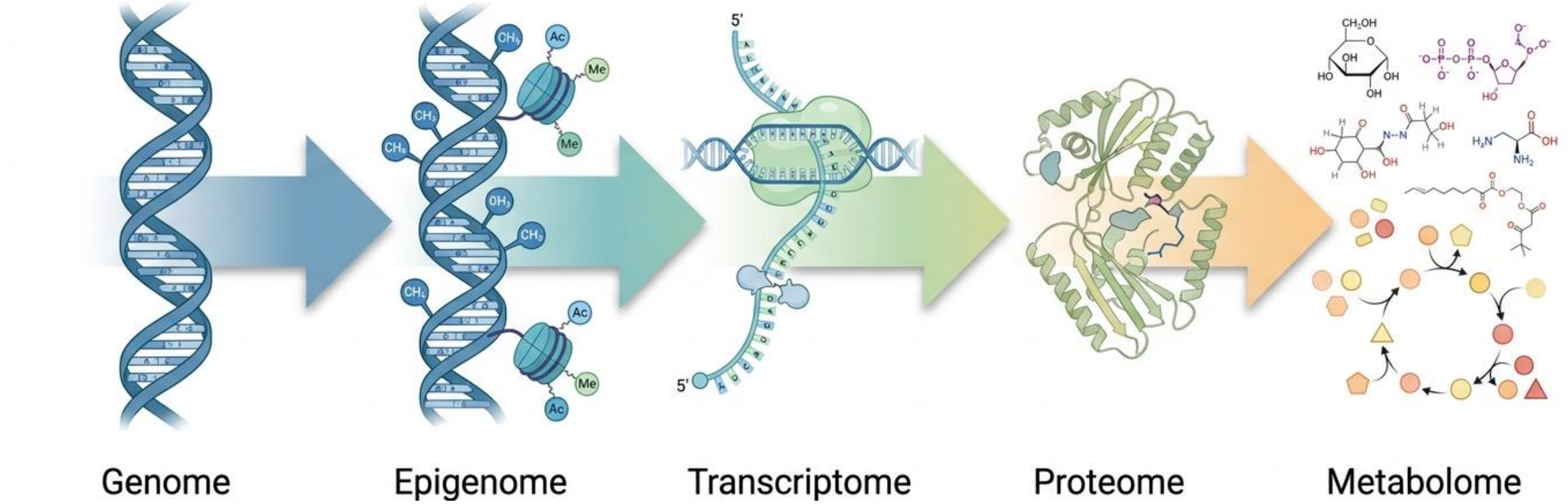

서울대학교 보건대학원 | 원성호 교수 | won1@snu.ac.kr

목차

WHY MULTI-OMICS ANALYSIS IN COPD?

Omics란 ?

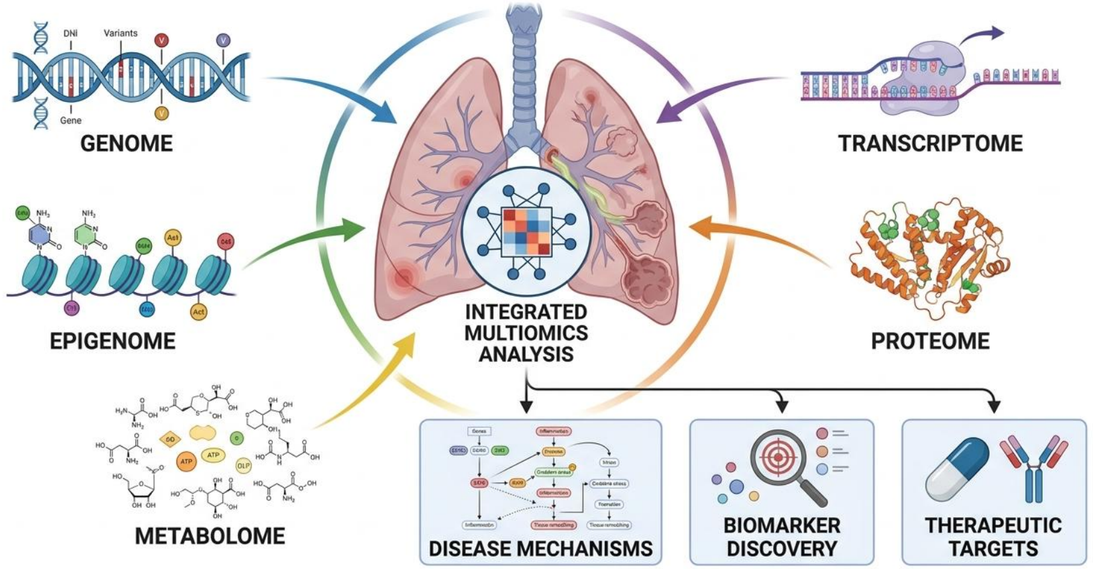
생체 내 특정 분자 집합 (genes, transcripts, proteins 등)을 측정하여 생명현상과 질병의 근본 원인을 규명하는 학문



Omics	Target	Platform / Design	얻을 수 있는 정보
Genome (유전체)	DNA	SNP array / WGS	SNP / Structure variation
Epigenome (후성 유전체)	DNA methylation, histone modification	EPIC array / WGBS / ATAC	DNA methylation / Chromatin accessibility
Transcriptome (전사체)	RNA	Bulk / Single-cell / Spatial	Tissue-, Cell type- , Spatial- specific gene expression level
Proteome (단백체)	Proteins	MS-based/ SOMAscan/ Olink	Proteins expression level
Metabolome (대사체)	Metabolites	LC-MS / GC-MS / NMR	Metabolites level

❖ 여러가지 오믹스 종류 및 특징

Why Multi-Omics in COPD ?



❖ COPD 관련 멀티오믹스 분석

- 각 오믹스는 COPD 병태 생리의 다른 층위를 반영함
- 단일 오믹스만으로는 질병의 메커니즘을 이해하기 어려움
- 멀티 오믹스 분석을 통해 극복할 수 있음
  - 질병 기전 규명
  - 바이오마커 발굴
  - 신규 치료 표적 발굴

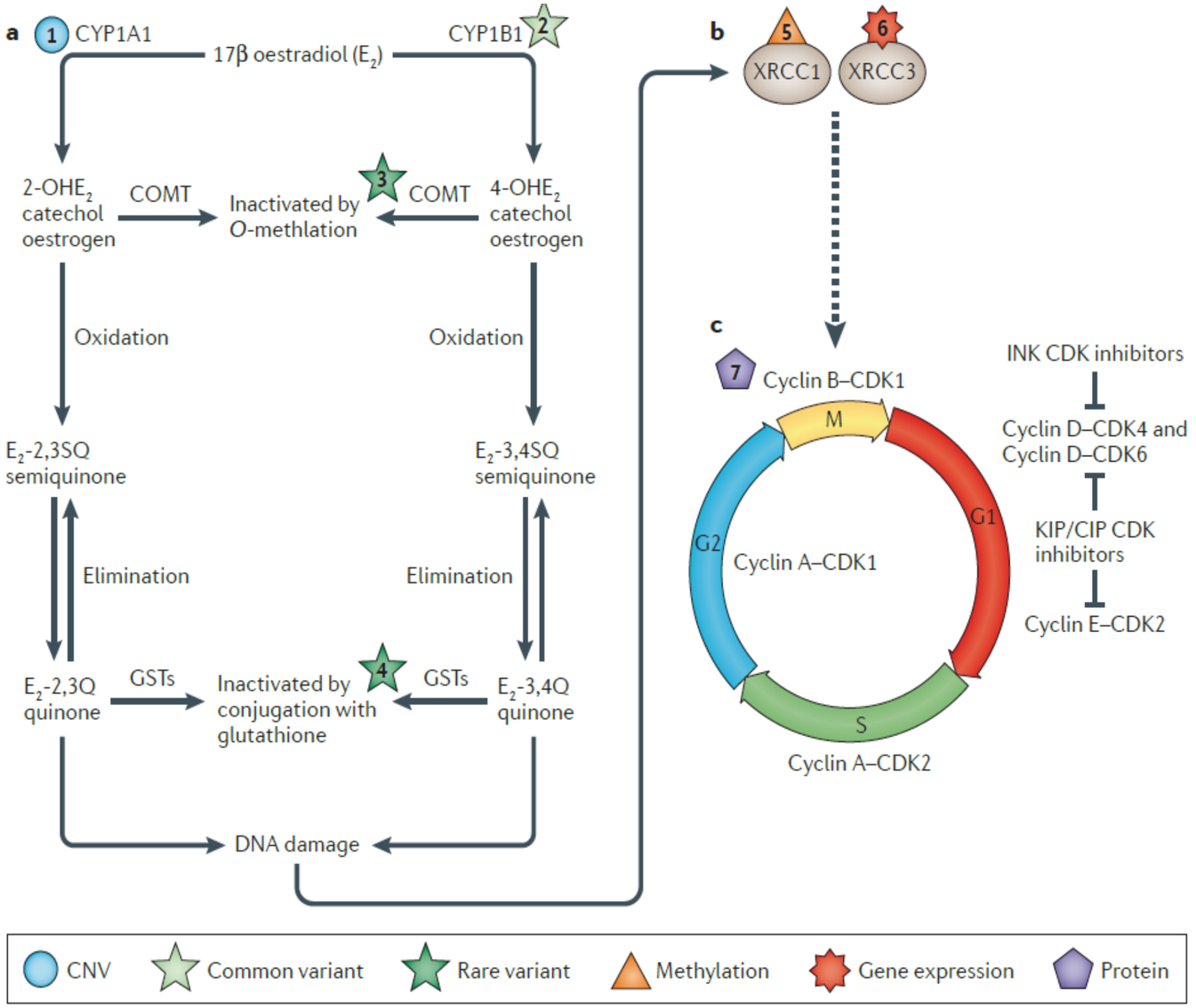
Omics	COPD에서의 의미	대표예시	한계
Genome (유전체)	COPD의 유전적 감수성을 설명 및 폐기능 저하 관련 유전적 요인 규명	- SERPINA1 (Alpha-1 antitrypsin deficiency) (Silverman & Sandhaus, NEJM, 2009) - FAM13A (COPD susceptibility locus) (Cho et al., Lancet Respir Med, 2014) - HHIP (Lung development, FEV1/FVC) (Sakornsakolpat et al., Nat Genet, 2019)	- 유전적 위험은 설명 가능하지만 현재 질병 상태를 반영하지 못함 - 환경 노출(흡연, 대기오염)의 영향을 직접 설명하기 어려움
Epigenome (후성 유전체)	흡연·환경 노출이 유전자 조절에 남기는 흔적	- AHRR hypomethylation (Smoking exposure marker) (Lee et al., Clin Epigenetics, 2016) - COPD-associated DMPs (BAL cells) (Eriksson Ström et al., AJRCMB, 2022) - Nrf2 pathway methylation (Oxidative stress) (Vucic et al., AJRCMB, 2014)	- DNA methylation 변화가 실제 기능 변화로 이어지는지 알기 어려움 - 조직 및 세포 특이성이 매우 강함
Transcriptome (전사체)	COPD에서 활성화되는 유전자 및 세포 특이적 반응 규명	- MUC5AC upregulation (Mucus hypersecretion) (Chan et al., Nat Commun, 2022) - AT2 cell dysfunction (Alveolar repair impairment) (Sauler et al., Nat Commun, 2022) - CXCL chemokine signaling (Inflammatory endothelial cells) (Sauler et al., Nat Commun, 2022)	- RNA 발현이 반드시 단백질 변화로 이어지지 않음 - 시점에 따라 변동성이 큼
Proteome (단백체)	COPD 관련 실제 기능 단백질을 설명	- IL-6 (Systemic inflammation) (Dickens et al., Respir Res, 2011) - Fibrinogen (COPD exacerbation biomarker) (Dickens et al., Respir Res, 2011) - MMP-2/MMP-9 (Emphysema progression) (Mahor et al., BMC Pulm Med, 2020)	- 단백질 변화의 원인(유전 vs 환경)을 구분하기 어려움 - 낮은 발현 단백질 검출에 한계
Metabolome (대사체)	COPD의 최종 대사 변화 및 생리적 상태 반영	- Ceramide (FEV1 decline) (Kim et al., Int J COPD, 2022) - Sphingomyelin (COPD-associated lipid metabolism) (Kim et al., Int J COPD, 2022) - Branched-chain amino acids ↓ (Disease severity) (Labaki et al., Sci Rep, 2019)	- 식이, 약물, 생활습관의 영향을 크게 받음 - 상류 분자기전(유전자·단백질)을 설명하기 어려움

❖ COPD 관련 단일 오믹스 분석 특징 및 관련 문헌

Multi-omics의 핵심 질문: “ 어떤 오믹스의 변화가 어떤 오믹스를 거쳐 COPD phenotype으로 나타나는가?”

Example how multiple data types may interact

Example involving 3 well-studied pathways for breast cancer: oestrogen metabolism, DNA damage repair and the cell cycle



Multi-omics data repositories

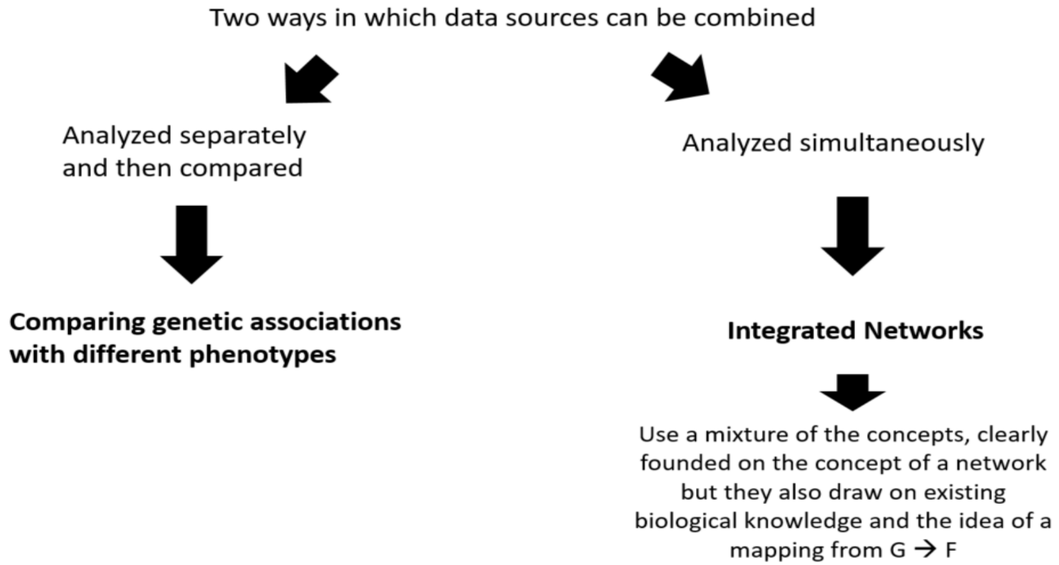
Table 1. List of multi-omics data repositories.

DATA REPOSITORY	WEB LINK	DISEASE	TYPES OF MULTI-OMICS DATA AVAILABLE
The Cancer Genome Atlas (TCGA)	https://cancergenome.nih.gov/	Cancer	RNA-Seq, DNA-Seq, miRNA-Seq, SNV, CNV, DNA methylation, and RPPA
Clinical Proteomic Tumor Analysis Consortium (CPTAC)	https://cptac-data-portal.georgetown.edu/cptacPublic/	Cancer	Proteomics data corresponding to TCGA cohorts
International Cancer Genomics Consortium (ICGC)	https://icgc.org/	Cancer	Whole genome sequencing, genomic variations data (somatic and germline mutation)
Cancer Cell Line Encyclopedia (CCLE)	https://portals.broadinstitute.org/ccle	Cancer cell line	Gene expression, copy number, and sequencing data; pharmacological profiles of 24 anticancer drugs
Molecular Taxonomy of Breast Cancer International Consortium (METABRIC)	http://molonc.bccrc.ca/aparicio-lab/research/metabric/	Breast cancer	Clinical traits, gene expression, SNP, and CNV
TARGET	https://ocg.cancer.gov/programs/target	Pediatric cancers	Gene expression, miRNA expression, copy number, and sequencing data
Omics Discovery Index	https://www.omicsdi.org	Consolidated data sets from 11 repositories in a uniform framework	Genomics, transcriptomics, proteomics, and metabolomics

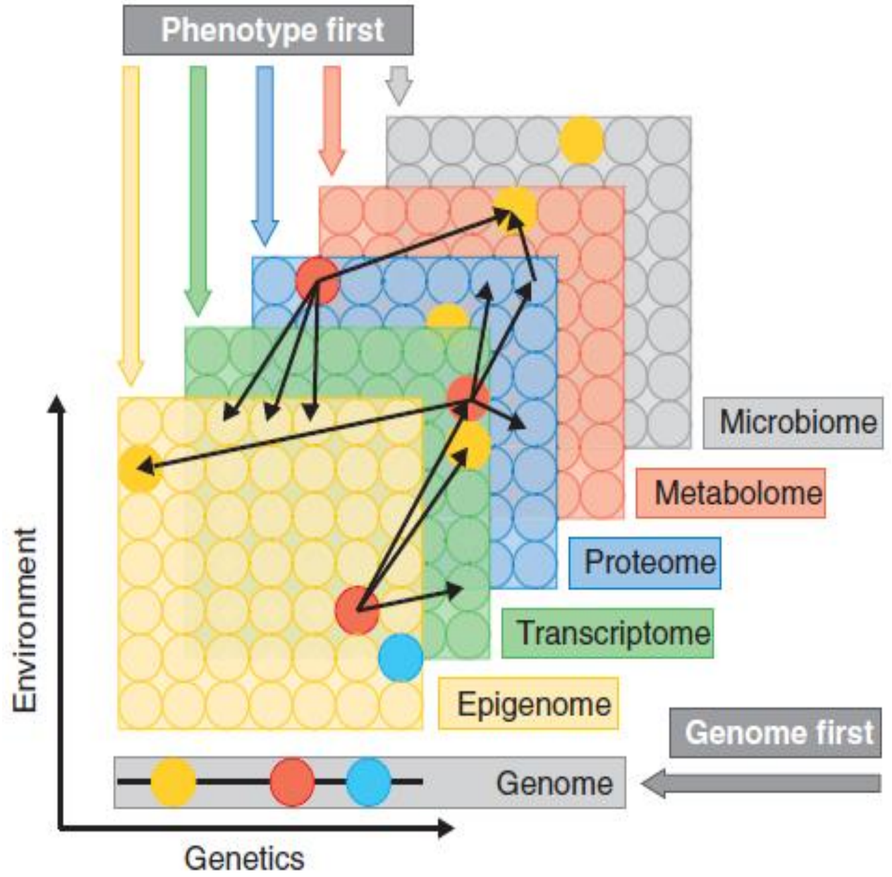
Subramanian et al, Bioinform Biol Insights, 2020

Methods of Multi-omics analysis

- Ritchie et al. classify multi-omics data integration methods into these 2 classes:
  ① Multi-staged approaches: consider different data types in a stepwise / linear / hierarchical manner.
  ② Multi-dimensional analysis: consider different data types simultaneously.



Multi-staged approaches



Hasin et al, Genome Biology, 2017

MULTI-STAGE ANALYSIS

Multi-stage analysis

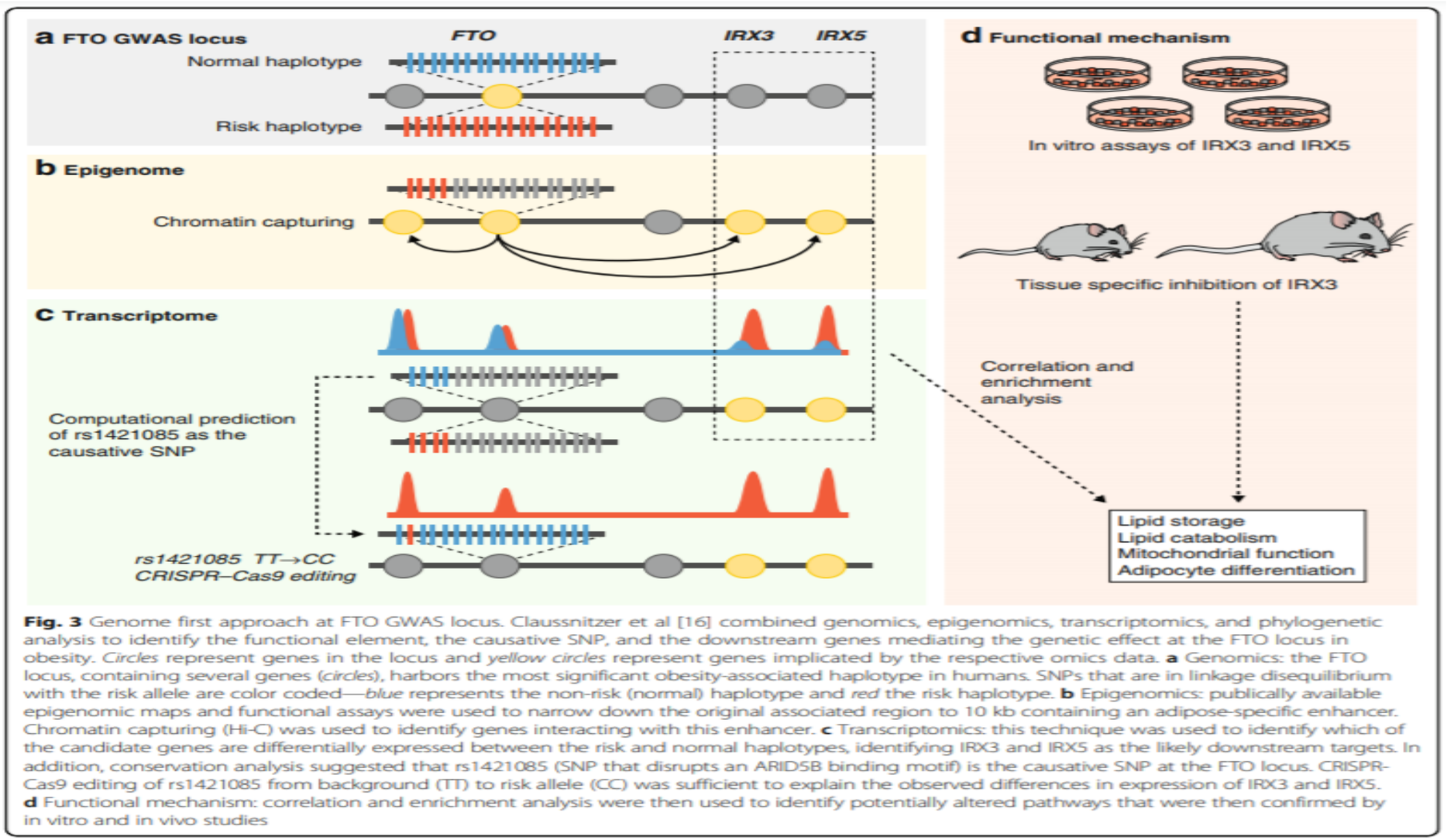


Fig. 3 Genome first approach at FTO GWAS locus. Claussnitzer et al [16] combined genomics, epigenomics, transcriptomics, and phylogenetic analysis to identify the functional element, the causative SNP, and the downstream genes mediating the genetic effect at the FTO locus in obesity. *Circles* represent genes in the locus and *yellow circles* represent genes implicated by the respective omics data. **a** Genomics: the FTO locus, containing several genes (*circles*), harbors the most significant obesity-associated haplotype in humans. SNPs that are in linkage disequilibrium with the risk allele are color coded—*blue* represents the non-risk (normal) haplotype and *red* the risk haplotype. **b** Epigenomics: publically available epigenomic maps and functional assays were used to narrow down the original associated region to 10 kb containing an adipose-specific enhancer. Chromatin capturing (Hi-C) was used to identify genes interacting with this enhancer. **c** Transcriptomics: this technique was used to identify which of the candidate genes are differentially expressed between the risk and normal haplotypes, identifying IRX3 and IRX5 as the likely downstream targets. In addition, conservation analysis suggested that rs1421085 (SNP that disrupts an ARID5B binding motif) is the causative SNP at the FTO locus. CRISPR-Cas9 editing of rs1421085 from background (TT) to risk allele (CC) was sufficient to explain the observed differences in expression of IRX3 and IRX5. **d** Functional mechanism: correlation and enrichment analysis were then used to identify potentially altered pathways that were then confirmed by in vitro and in vivo studies

Hasin et al, Genome Biology, 2017

Multi-stage analysis: Mol QTL

Steps:

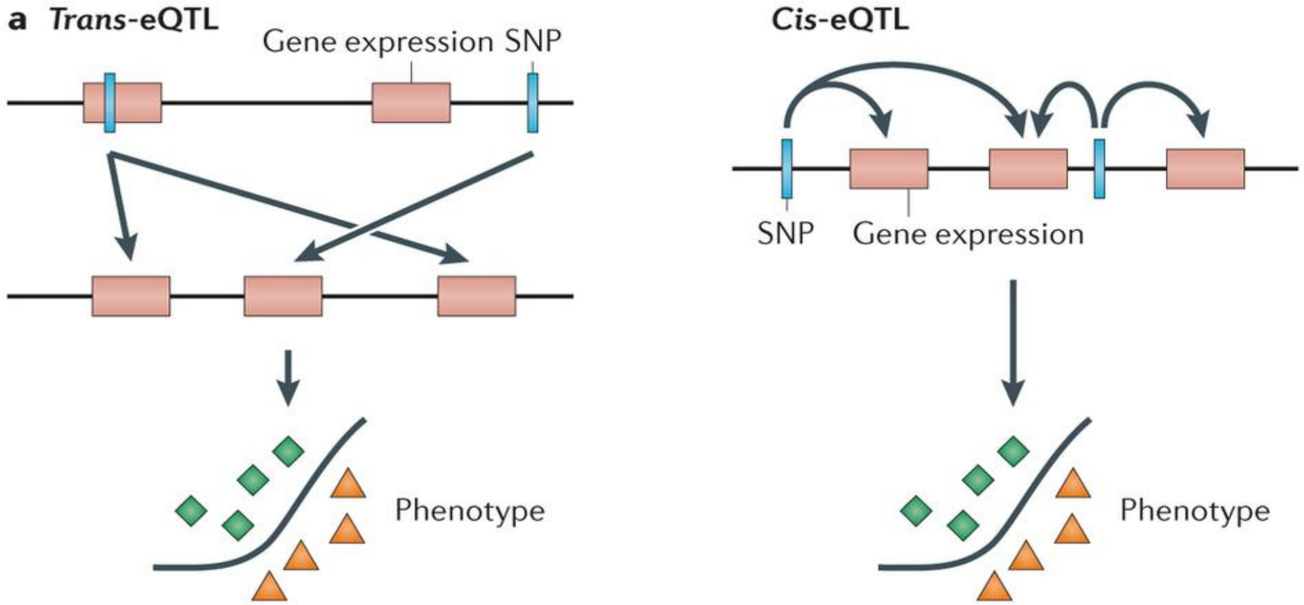
① associate SNPs with phenotype; filter by significance threshold

② Test the SNPs that are associated with phenotype with other omic data. E.g. check for the association with gene expression data -> eQTL (expression quantitative trait loci). Also: methylation QTLs, metabolite QTLs, protein QTLs …

③ (3) Test omic data used in step 2 for correlation with phenotype of interest.

Trans-eQTL: effect on remote gene

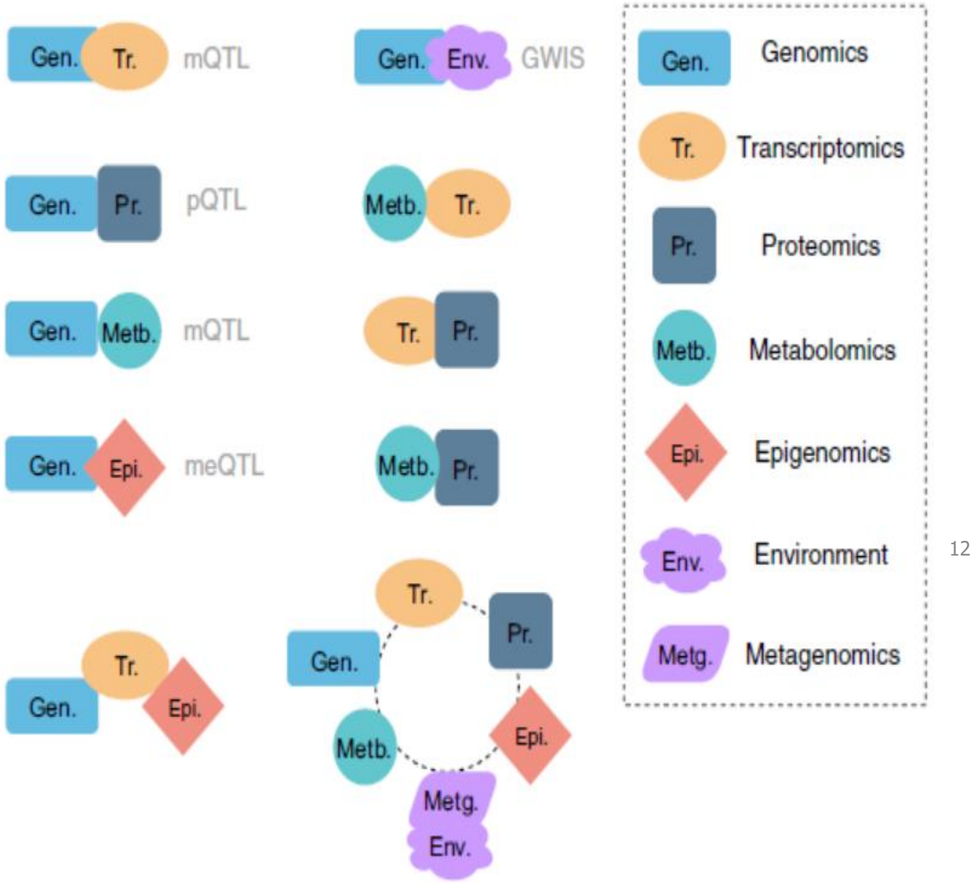
Cis-eQTL: effect on nearby gene



Ritchie et al. *Nature Rev Genet* **16**, 85 (2015)

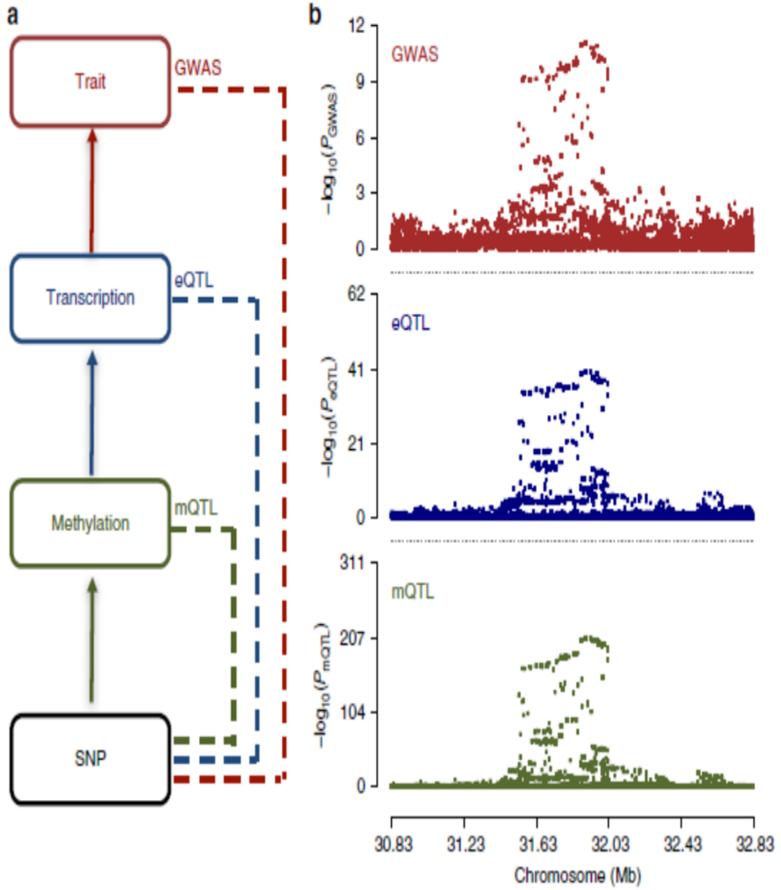
Definition of Mol-QTL

- Integrative analyses with multi-omics analysis



12

Possible integrations of multi-omics data



Molecular quantitative trait loci

Ivanova et al, *Allergy, 2019*

Definition of Mol-QTL

- Molecular quantitative trait loci (Mol QTL)
  - The most common type of human genetic variants, have important roles in shaping complex human traits and in causing diseases.
  - molQTL analysis is a statistical method to link genotyping and molecular phenotyping data to interpret the effects of genetic variants in complex traits.
  - Transcriptom, Proteoim, etc are sensitive to batch effect, reverse causation, etc and can be confounded by environmental factors.
  - Mol QTL allows us to identify causal effects of various omics data.
  - There are multiple open database: QTLbase (http://mulinlab.org/qtlbase)

Ye et al, Trends Genet, 2020

Definition of Mol-QTL

- Various types of Mol QTL

molQTL	Abbreviation name	Traits/diseases	Examples	Resources
Transcriptome				
Protein-coding gene	eQTL	Diverse populations [18]; cell lines/tissues (e.g., monocytes [19], human placentas [20]); diseases (e.g., schizophrenia [25]); external stimuli [27]	rs281437-*ICAM1* [19]	eQTLGen Consortium [31], PancanQTL [26], NephQTL [32], eQTL Catalog [33], ExSNP database [34], ImmunPop QTL browser [35]
Long noncoding RNA	lncR-eQTL	Lymphoblastoid cell lines [36], tissue [39], multiple sclerosis [38], cancer [40]	rs420259-*CTD-2196E14.9* [36]	ncRNA-eQTL [40]
miRNA	miR-eQTL	Blood [45], cancer [40]	rs7115089-*miR-125b* [45]	ncRNA-eQTL [40]
Circular RNA	Circ-eQTL	Lymphoblastoid cell lines [47], dorsolateral prefrontal cortex [48]	rs71023104-*circALOX5* [47]	
Post-transcriptional regulation				
Alternative splicing	sQTL	Population [56], diabetes [173], cancer [57]	rs56048322-*PTPN22* (skipping of exon 18) [173]	CancerSplicingQTL [57]
RNA editing	edQTL	Lymphoblastoid cell lines [67]	rs2028299; editing of *ARPIN* [66]	
APA	apaQTL	Tissues [72], cancer [74]	rs17497828-3′UTR of *MIER1* [53]	SNP2APA [74]

Ye et al, Trends Genet, 2020

Definition of Mol-QTL

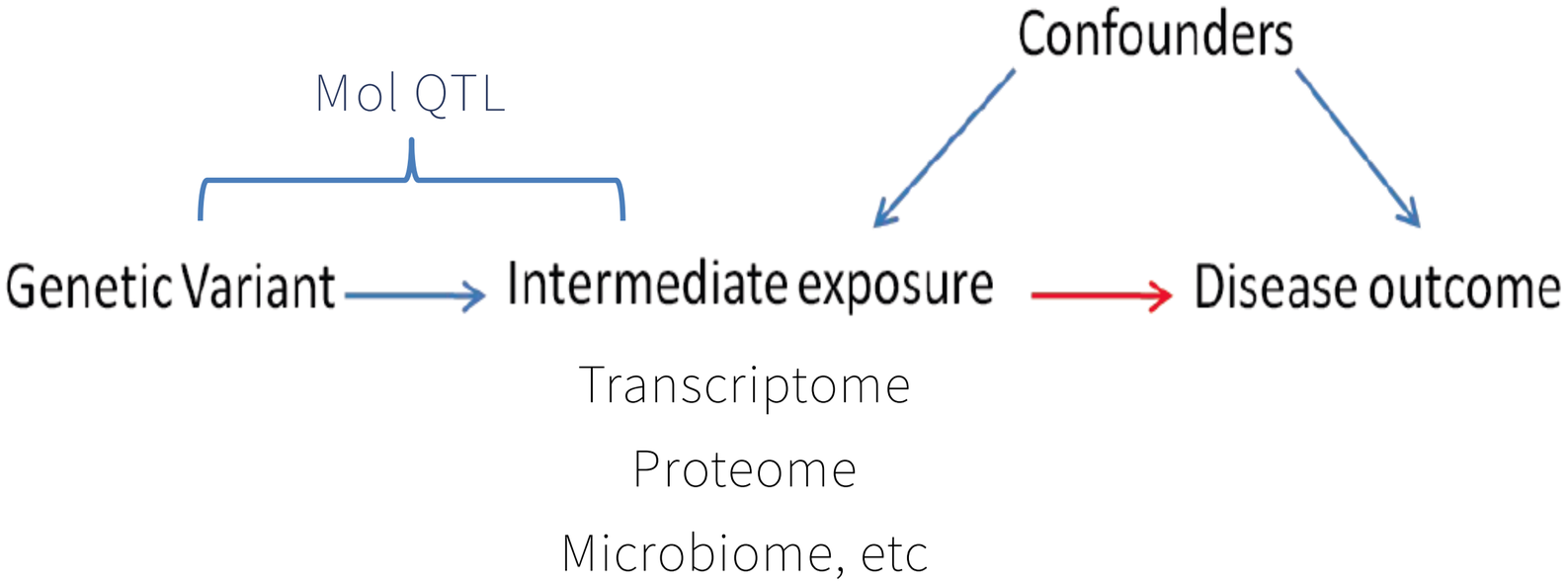
- Various types of Mol QTL

Epigenome				
DNA methylation	meQTL	PTSD [83], lung cancer [84]	rs2736100-CpG methylation in promoter of *TERT* [86]	GRASP [88], CDEG [89], Pancan-meQTL [90]
Histone modification	hQTL	Yoruba lymphoblastoid cell lines [92], autoimmune thyroid diseases [95]	rs8134436-H3K27ac in downstream of *ICOSLG* [94]	
Transcription factor binding	tfQTL	Immune, nervous, and metabolic diseases [96]	rs6537048-TFBS in promoter of *IL15* [96]	
Pol II binding	Pol II QTL	Yoruba lymphoblastoid individuals [92]	rs12723363-Pol II BS in promoter of *SNX7* [92]	
DNAse I hypersensitivity	dsQTL	Yoruba lymphoblastoid cells [97]	rs4953223-accessible region upstream of *NFKB* [97]	
Transposase accessibility	ATAC-QTL	Autoimmune diseases [98]	rs1217817-accessible region in promoter of *MAP1B* [98]	
Protein and protein post-translational modification				
Protein expression	pQTL	Body mass index (BMI) [105], height [140], Parkinson's disease [108], lung disease [109]	rs4129267-*IL6R* [106]	Obesity cohort [105]
Protein post-translational modifications	PTM-QTL	Huntington's disease [113]	rs118005095-*HTT* myristoylation [113]	AWESOME [115]
Metabolome				
Metabolites	mQTL	BMI [117], kidney disease [118], cardio-metabolic diseases [119]	rs4921914-formate [118]	UK Biobank [121]
Microbiome				
Microbiome	Microbiome QTL	BMI [122], inflammatory bowel disease [127], heart disease, meningitis [128]	rs11222579-*Ruminococcus* [128]	UK Twins [129]

Ye et al, Trends Genet, 2020

Mendelian Randomization

- The fundamental idea: If we cannot randomize the exposure, we can find a randomized instrumental variable to disentangle
  - Confounding
  - Reverse causation



Ye et al, Trends Genet, 2020

Mendelian Randomization

- Core assumptions

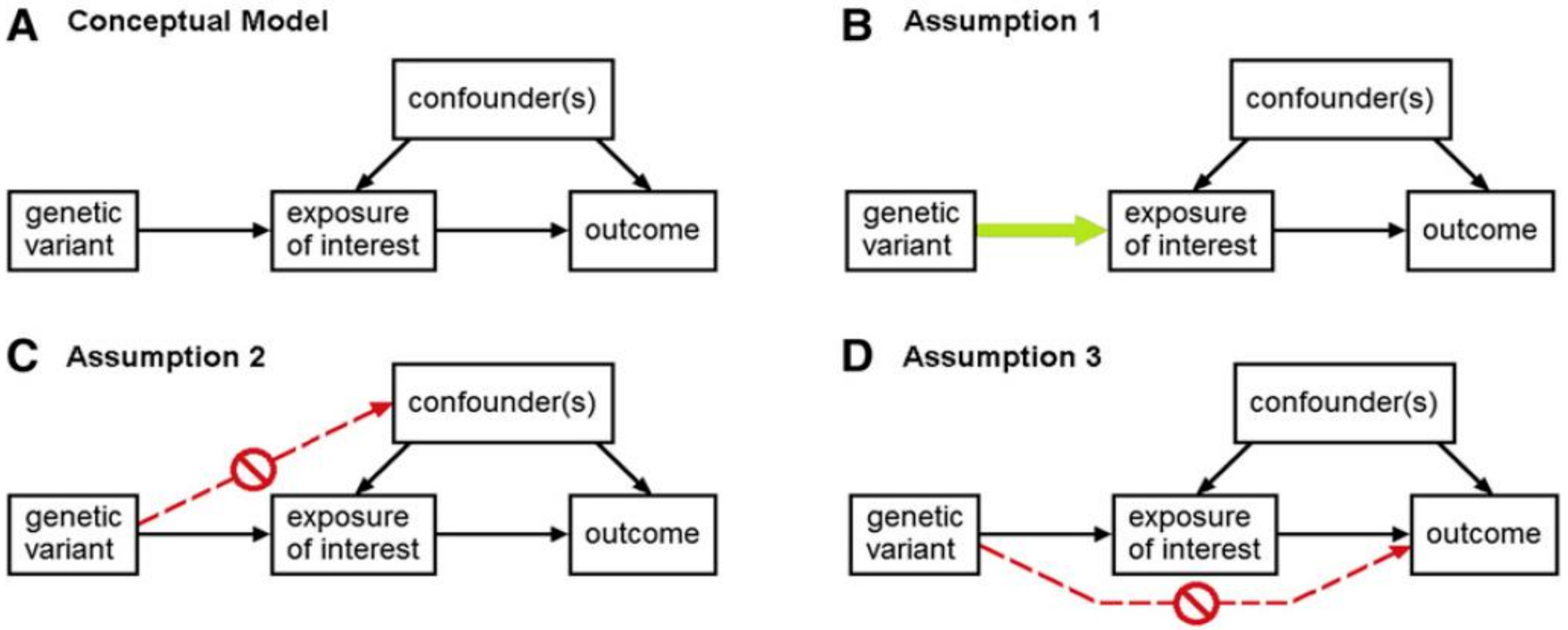


Figure 2. Conceptual illustration of the MR method and its three underlying core assumptions as directed acyclic graphs. (A) Conceptual model. (B) Assumption 1. (C) Assumption 2. (D) Assumption 3.

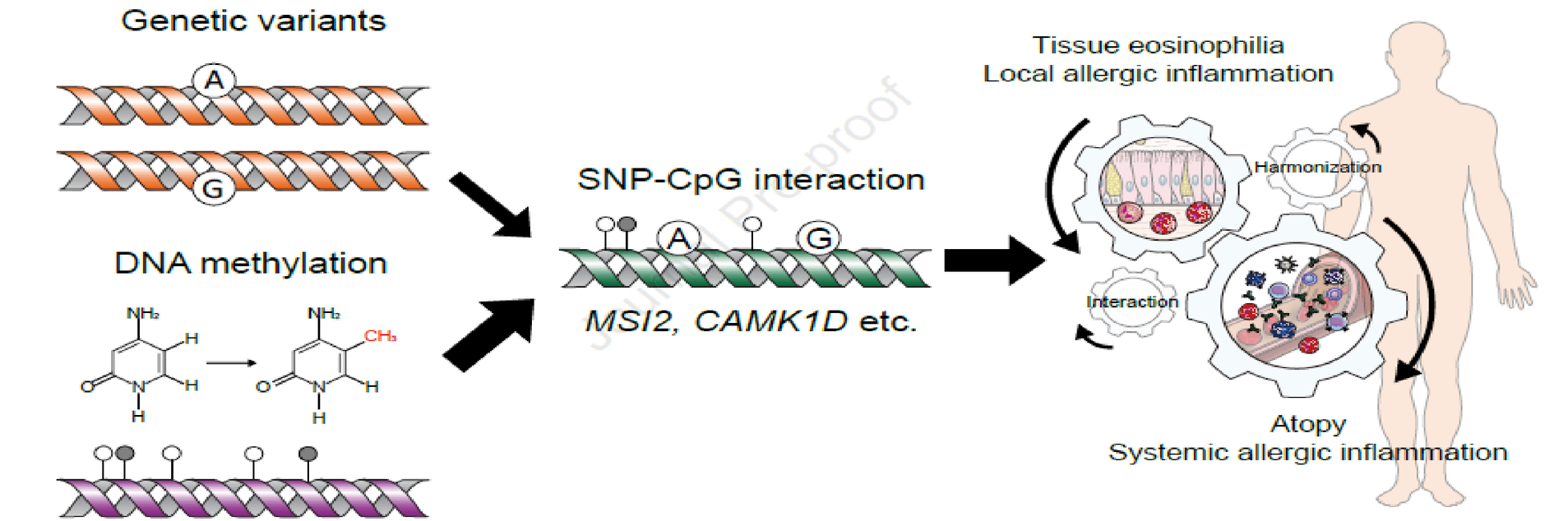
Ye et al, Trends Genet, 2020

Mol-QTL: allergic inflammation

Integrated genetic and epigenetic analyses uncover *MSI2* association with allergic inflammation

Kyung Won Kim, MD, PhD * • Sang-Cheol Park, PhD * • Hyung-Ju Cho, MD, PhD * • ...
Chang-Hoon Kim, MD, PhD • Sungho Won, PhD • Myung Hyun Sohn, MD, PhD *

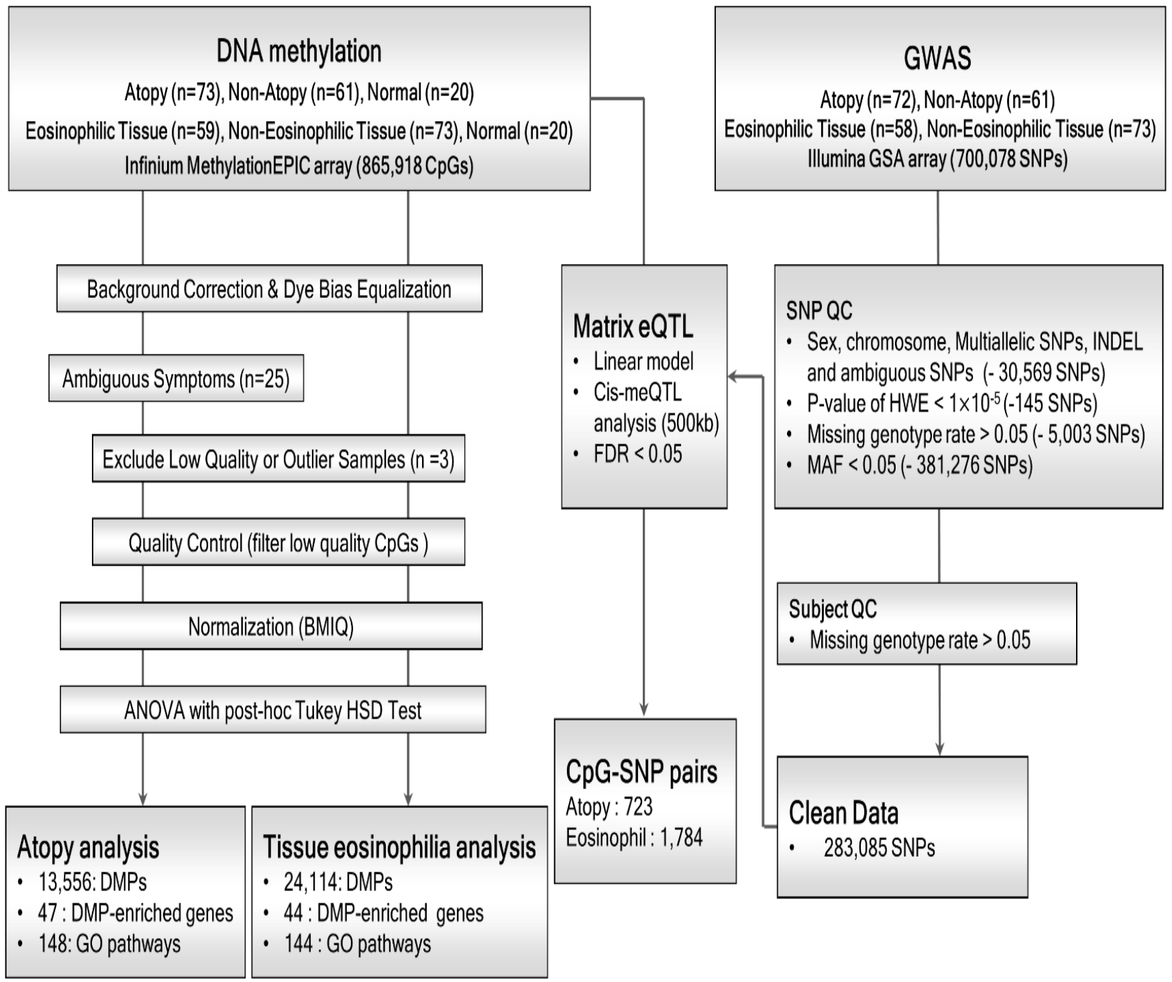
J Allergy Clin Immunol

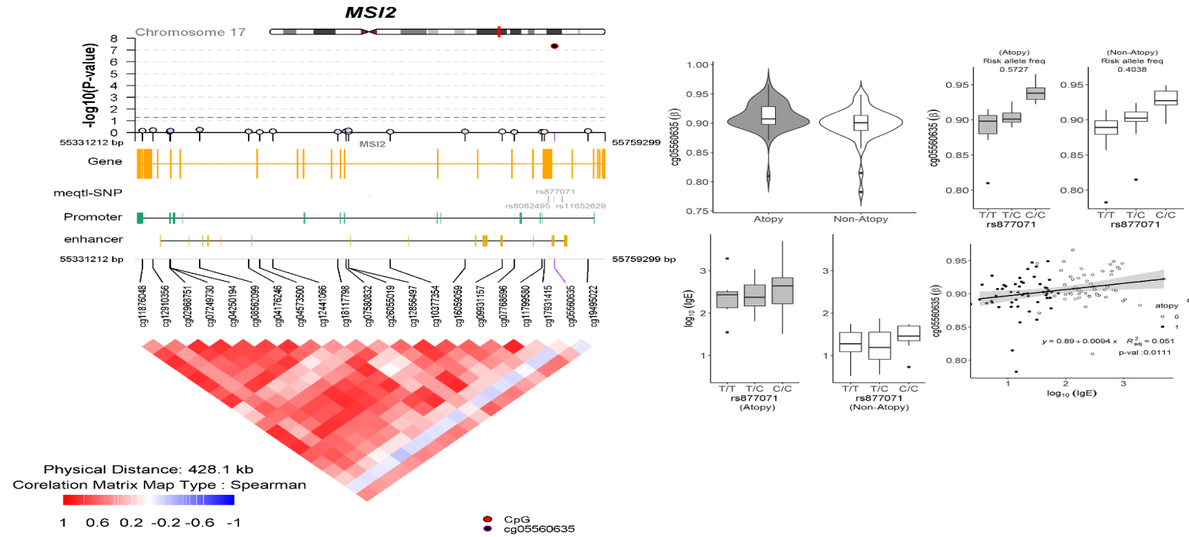


MSI2, Musashi RNA binding protein 2; CpG, 5'-C-phosphate-G-3'; *CAMK1D*, Calcium/calmodulin dependent protein kinase 1D

Kim et al, *J Allergy Clin Immunol*

Mol-QTL: allergic inflammation



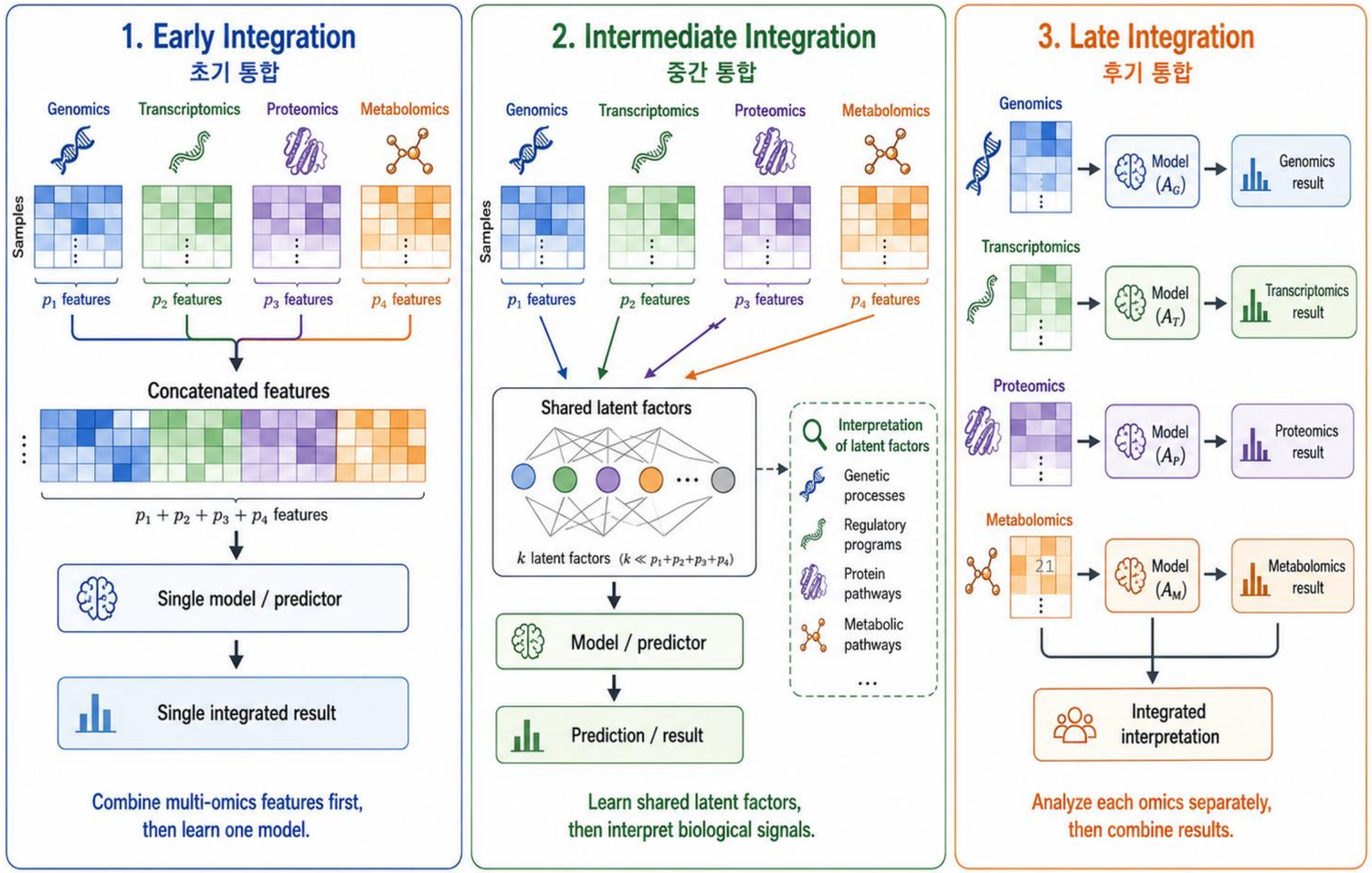


Model	Odds Ratios	95% CI	*P*value
Atopy ~			
Model 1 *			
CpG (cg05560635) +20 PC**			
CpG (cg05560635)	9.24×10^{18}	$3.17 \cdot 6.29 \times 10^{42}$	.072
Comp 1	4.46×10^{-03}	3.33×10^{-05} -1.17×10^{-01}	.007
Model 2			
Expression*** (*MSI2*)	0.13	2.27×10^{-2} - 5.46×10^{-1}	.009
Model 3 *			
1 CpGs + 20 PC**			
+ Expression			
CpG (cg05560635)	6.95×10^{30}	2.45×10^{-45} - 1.64×10^{18}	.046
Expression*** (*MSI2*)	6.45×10^{-3}	1.76×10^{-5} - 3.72×10^{-1}	.039
Comp 5	5.92×10^{12}	1.09×10^{3} - 3.67×10^{26}	.026

Kim et al, *J Allergy Clin Immunol*

MULTI-DIMENSION ANALYSIS

Multi-omics 통합 분석 방법



❖ 여러가지 Multi-omics 통합 분석 방법

Integration Strategy	개념/장점	단점
Early	모든 omics feature를 통합해서 사용	High dimensional, Overfitting
Intermediate	Omics 간 공통 signal을 latent factor로 요약할 수 있음	Latent factor의 해석이 명확하지 않을 수 있음
Late	각 omics간 특성을 보존하면서 결과 해석 가능	Omics간 직접적인 상호작용을 충분히 반영 어려움

❖ Multi-omics 통합 분석 방법 특징

- 3가지 Multi-omics 통합 분석 방법
  - Early Integration
  - Intermediate Integration
  - Late Integration
- 현재 선호되는 방법은 Late Integration

Early Integration 분석 사례: Atopic dermatitis [park et al, 2021, Allergy]

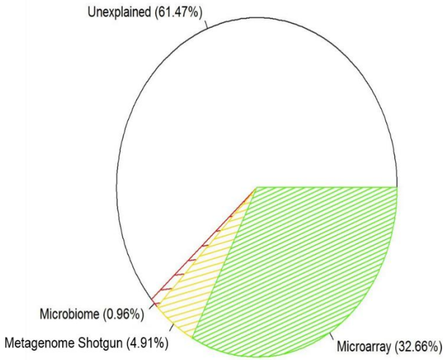
Allergy
EUROPEAN JOURNAL OF ALLERGY AND CLINICAL IMMUNOLOGY
EAACI

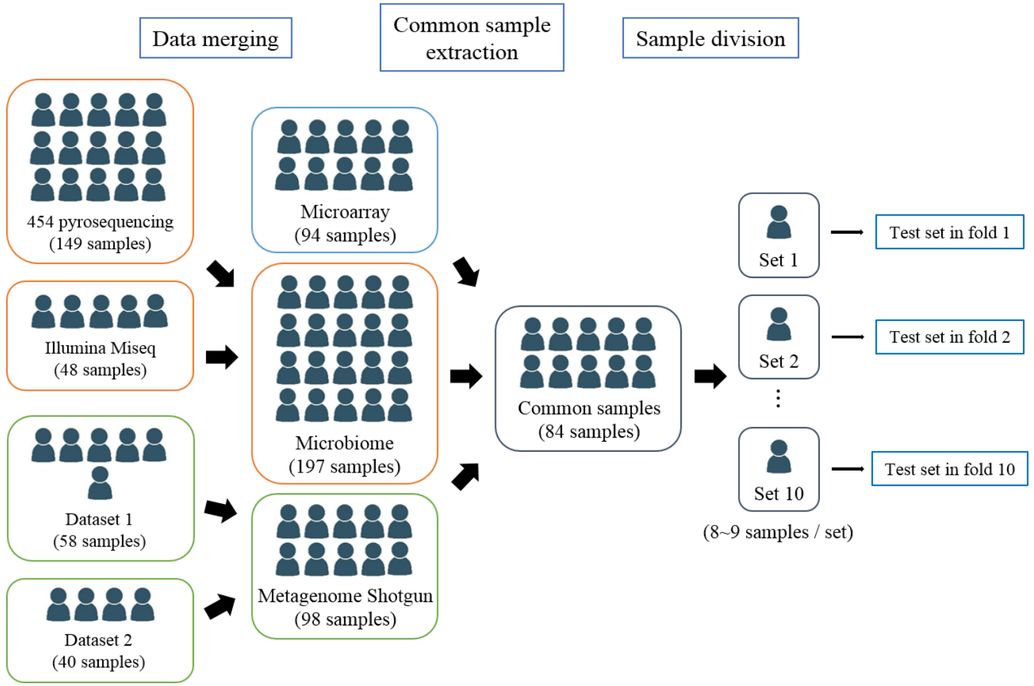
LETTER TO THE EDITOR

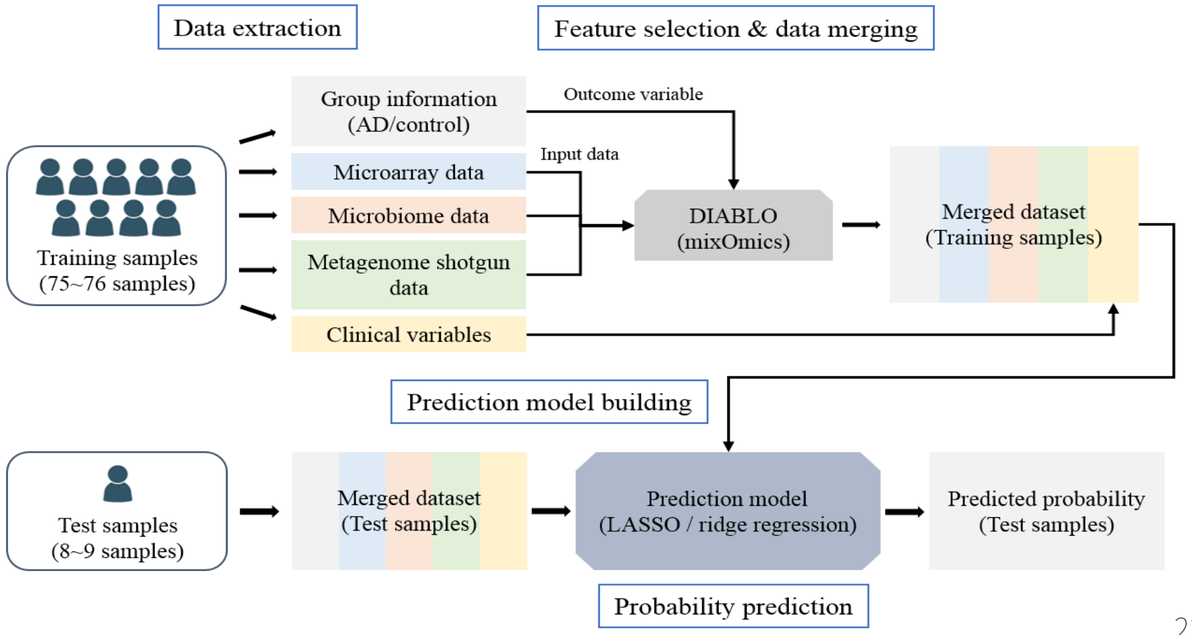
Multi-omics analyses implicate *EARS2* in the pathogenesis of atopic dermatitis

Jaehyun Park, Sharon M. Lutz, Seungil Choi, Sanghun Lee, Sang-Cheol Park, Kangjin Kim, Hosik Choi, Hansoo Park, So Yeon Lee, Scott T. Weiss, Soo-Jong Hong, Bong-Soo Kim ✉, Sungho Won ✉

Omics	Variable	Selection probability (%)	Average coefficient	OR
transcriptome	EARS2	100	7.061	1.166×10^3
transcriptome	SGOL1-AS1	80	2.810	16.610
transcriptome	LINC01036	60	3.474	32.266
transcriptome	C16orf72	60	4.589	98.396
metagenome	Carotenoid biosynthesis	60	-0.422	0.656







Early Integration 분석 사례: COPD multi-omics subtyping [Gillenwater, Lucas A., et al., 2021, Plos one]

PLOS ONE

RESEARCH ARTICLE

Multi-omics subtyping pipeline for chronic obstructive pulmonary disease

Lucas A. Gillenwater[1☯¤a], Shahab Helmi[2☯], Evan Stene[2☯], Katherine A. Pratte[1], Yonghua Zhuang[3], Ronald P. Schuyler[4¤b], Leslie Lange[5], Peter J. Castaldi[6], Craig P. Hersh[6], Farnoush Banaei-Kashani[2]*, Russell P. Bowler[1]*, Katerina J. Kechris[3]*

Dataset

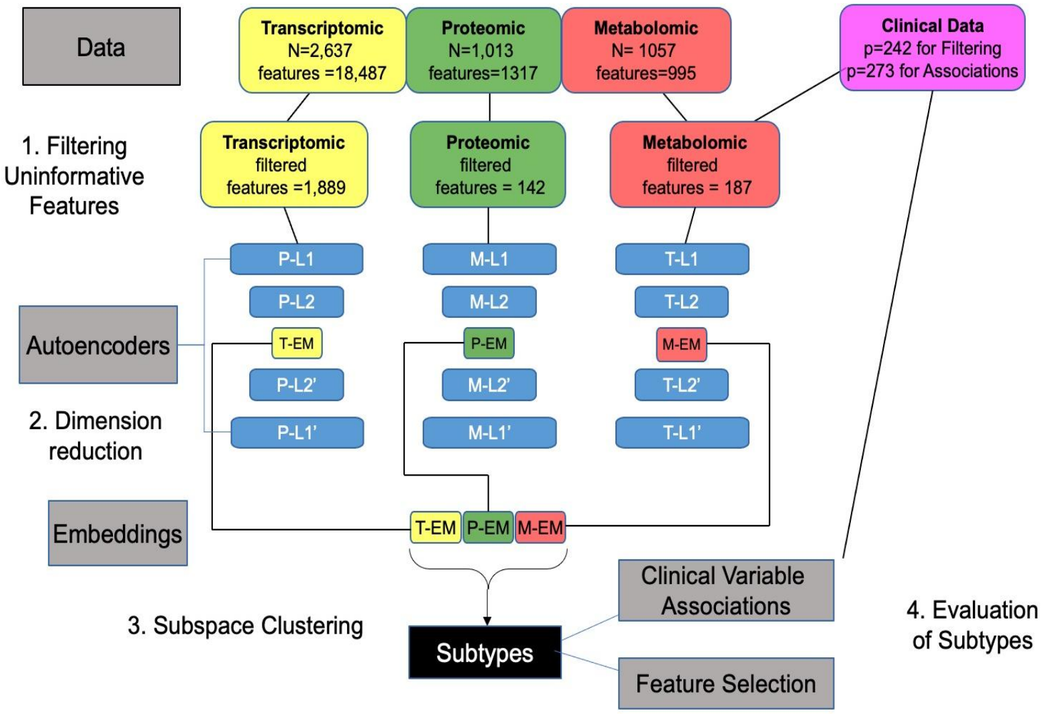
- COPD Gene cohort
- N = 489명 (COPD: 232명, non-COPD: 267명)

Omics

- Transcriptome (Bulk)
- Proteome
- Metabolome

Phenotypes

- COPD phenotypes
- 다른 여러가지 phenotypes



- Autoencoder: omics 별 autoencoder embedding 생성
- Early Integration: Transcriptome, Proteome, Metabolome embeddings concatenation
- Clustering: k-means, hierarchical clustering 등 다양한 clustering 기법 이용하여 4개의 subtype으로 구분
- ANOVA: 273개의 임상 변수들에 대하여 4개의 subtype에서 유의한 차이가 있는 지표를 선정 (유의한 임상 지표는 없었음)

Intermediate Integration 분석 사례: COPD multi-omics subtyping [Arda Halu., et al., 2026, preprint]

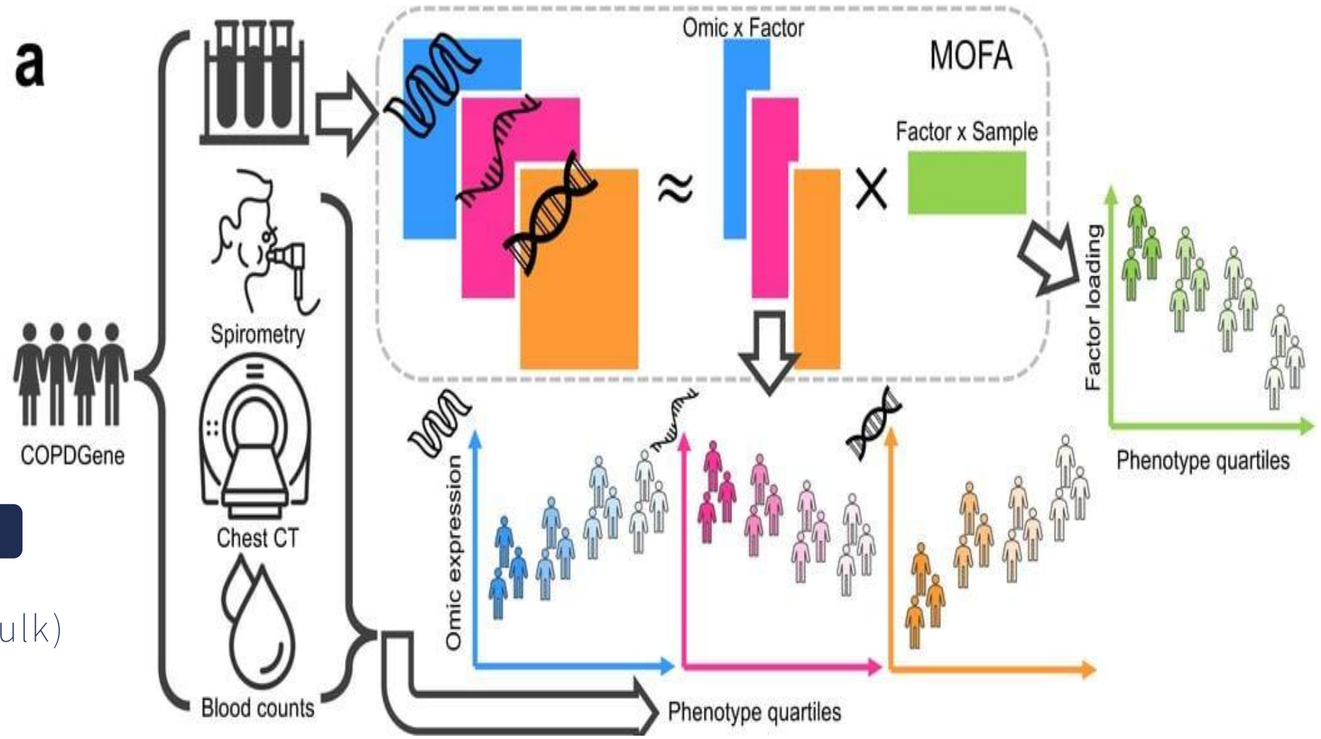
medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

Follow this preprint

Integrative multi-omic analyses identify major axes of heterogeneity in chronic obstructive pulmonary disease and uncover their molecular contributors

Arda Halu, Matthew Moll, Chengyue Zhang, Leonardo Martini, Per S. Bakke, Russell P. Bowler, Peter J. Castaldi, Michael H. Cho, Dawn L. DeMeo, Kimberly Glass, Craig P. Hersh, Brian D. Hobbs, Edwin K. Silverman

doi: https://doi.org/10.64898/2026.01.22.26344654

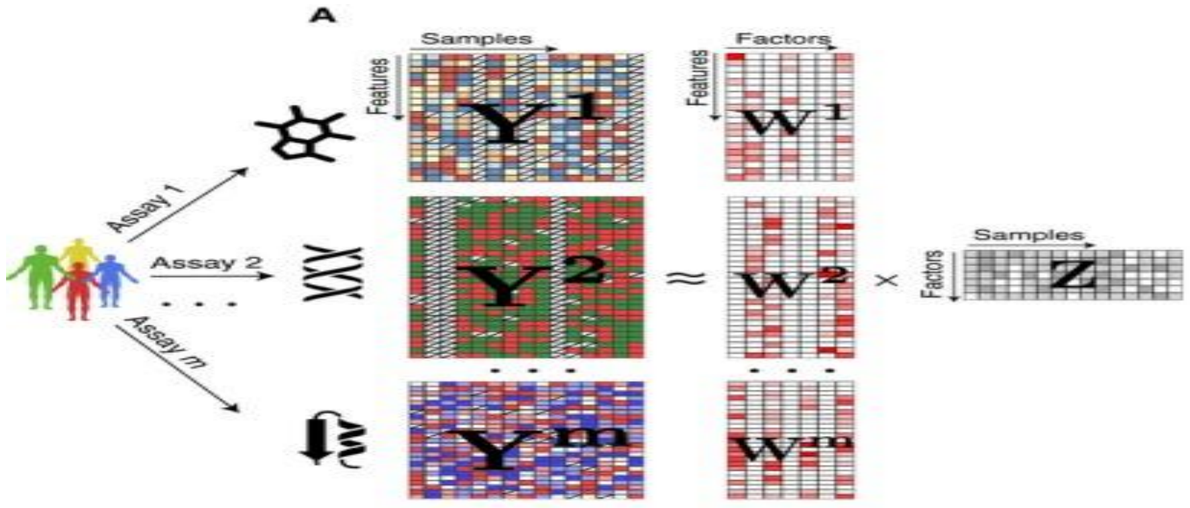


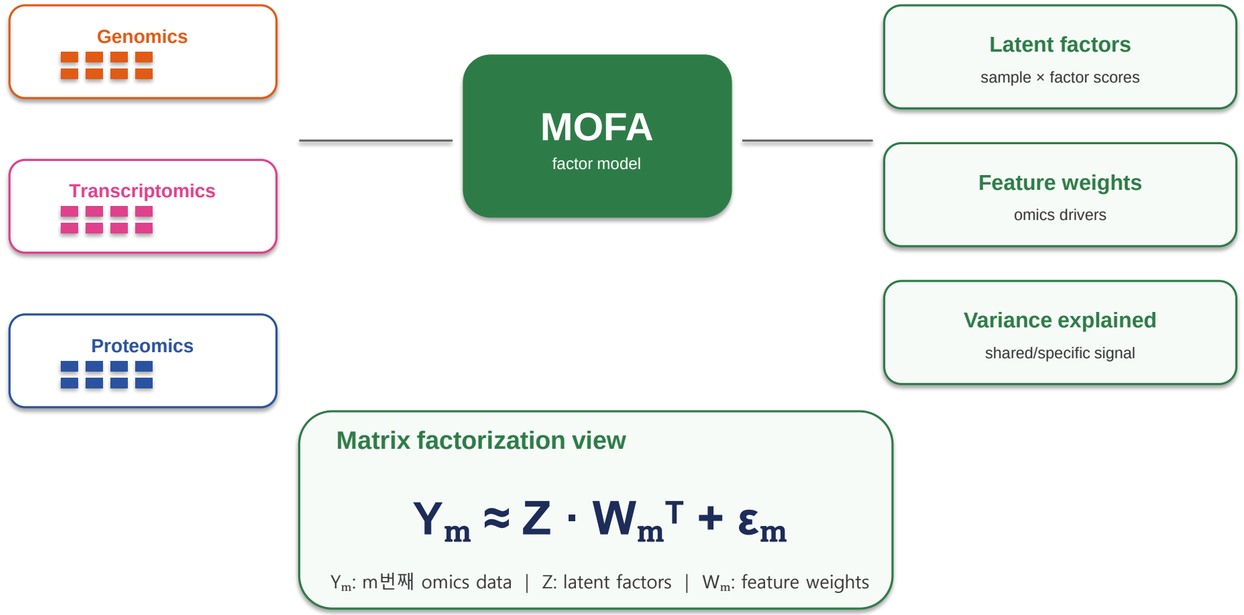
Dataset	Phenotypes	Omics
➢ COPD Gene cohort ➢ N = 1,872명 (COPD)	➢ 폐 기능 관련 지표 ➢ 흉부 CT 영상 수치 지표 ➢ 혈액 수치 ➢ 기타 임상 결과	➢ Genome ➢ Transcriptome (Bulk) ➢ Proteome

- MOFA(multi-omics factor analysis): 멀티 오믹스 데이터를 공통의 잠재 변수로 통합하여 COPD 환자 간의 이질성을 설명하는 주요 임상 변수 축과 핵심 분자를 식별함
- GFLASSO : 폐 기능이나 흉부 CT 영상 등 상호 상관관계가 있는 여러 임상 지표들 중에서, 이후 분석에 사용할 대표적인 표현형을 선택하는 데 사용함
- 일반화 선형 모델 (GLM), Cox 비례 위험 모델: MOFA를 통해 도출된 잠재 요인이 폐 기능, 흉부 CT 영상, 혈액 수치 및 전체 원인에 의한 사망률 등 환자의 실제 임상 결과와 어떠한 통계적 연관성이 있는지 평가

Intermediate Integration 분석 사례: COPD multi-omics subtyping [Arda Halu., et al., 2026, preprint]

◆MOFA (Multi-Omics Factor Analysis)란 ?



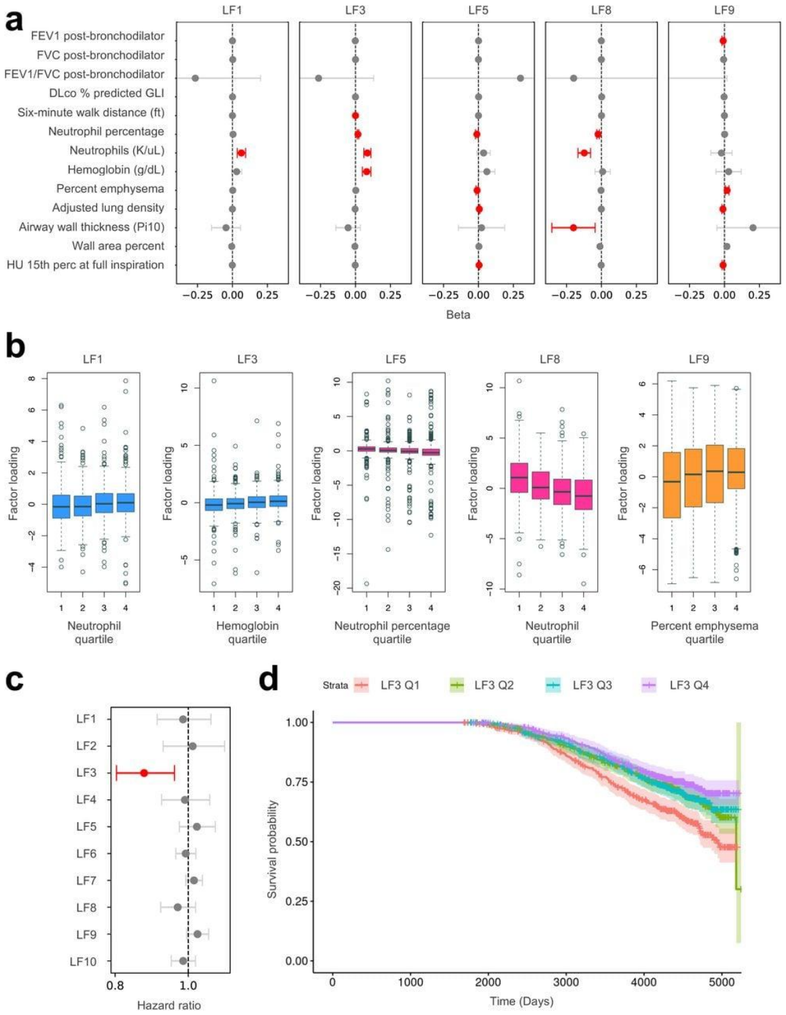


- Argelaguet, Ricard, et al., 2018, Molecular Systems Bio., 인용: 1,651회
- 여러 omics matrix를 동시에 입력으로 받는 unsupervised factor analysis
- 각 latent factor는 각 주요 변동 축을 나타냄
- Factor weight 는 해당 factor 를 설명하는 omics feature의 기여도를 의미함

MOFA는 omics 를 단순 결합 하지 않고, omics 간 공통 및 특이 변동을 latent factor 로 학습하는 intermediate integration 방법

Intermediate Integration 분석 사례: COPD multi-omics subtyping [Arda Halu., et al., 2026, preprint]

◆Latent factor와 COPD phenotype과의 연관성 확인



- Results
  - MOFA latent factors 중 일부가 lung function, chest CT imaging, blood count phenotype과 연관
  - LF3는 all-cause mortality와도 연관됨
  - factor loading을 phenotype quartile별로 비교하여 patient subgroup 차이를 해석
- Interpretation
  - 각 latent factor는 하나의 COPD subtype이라기보다, COPD heterogeneity를 설명하는 연속적인 biological axis로 해석 가능
  - 임상 phenotype과 factor의 연결을 통해 어떤 molecular axis가 어떤 COPD 특성과 관련되는지 파악

latent factor를 COPD phenotype 및 mortality와 연결하여 disease heterogeneity의 주요 축을 설명

Late Integration 분석 사례: Heavy metal exposure and its effect on APOC3 [Kim et al. 2025]

Journal of Hazardous Materials 482 (2025) 136574

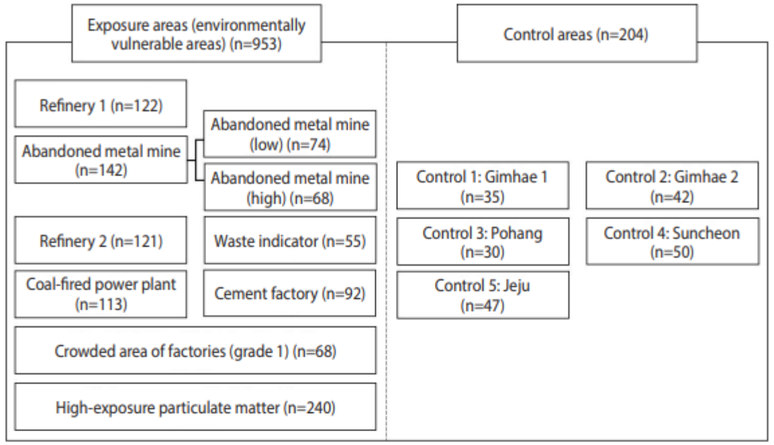
Contents lists available at ScienceDirect

Journal of Hazardous Materials

journal homepage: www.elsevier.com/locate/jhazmat

Heavy metal exposure and its effects on *APOC3*, *CFAI*, and *ZA2G*

Nam-Eun Kim [a,1], Min Heo [b], Hyeongyu Shin [b], Ah Ra Do [c], Jeeyoung Kim [d], Hee-Gyoo Kang [e], Sora Mun [e], Hyun Ju Yoo [f], Mi Jeong Kim [f], Jung-Woong Kim [g], Chul-Hong Kim [g], Young-Seoub Hong [h,i], Yong Min Cho [j,k], Heejin Jin [a], Kyungtaek Park [a], Woo Jin Kim [d,**,2], Sungho Won [a,b,c,l,*,2]



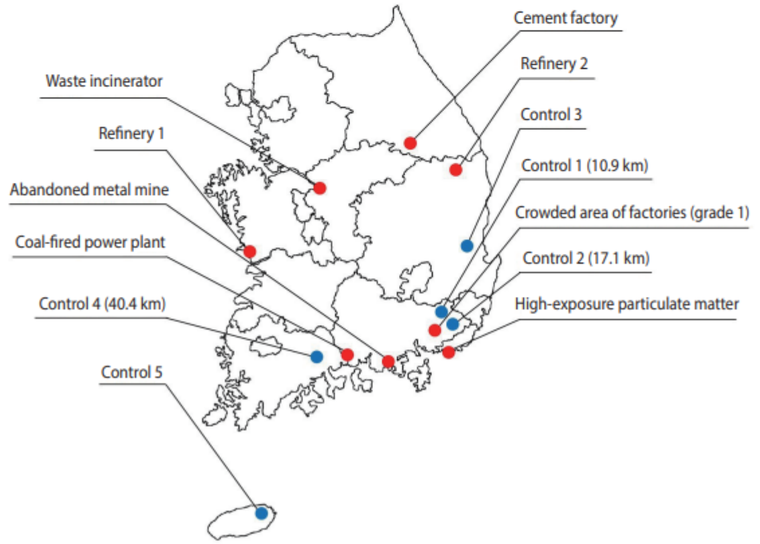
Volume: 46, Article ID: e2024062, 12 pages
https://doi.org/10.4178/epih.e2024062

DATA PROFILE

epiH
Epidemiology and Health

Introduction to the forensic research via omics markers in environmental health vulnerable areas (FROM) study

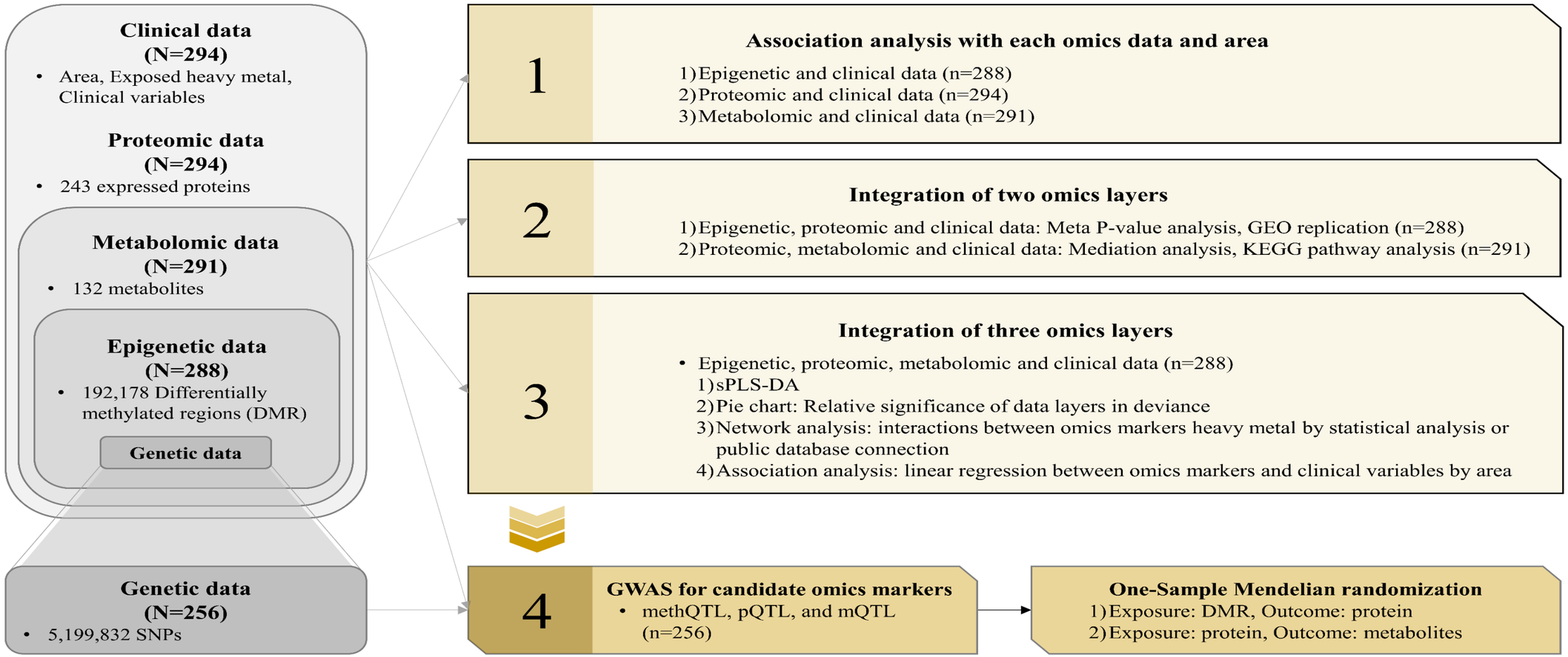
Jung-Yeon Kwon[1], Woo Jin Kim[2], Yong Min Cho[3], Byoung-Gwon Kim[1,4], Seungho Lee[1,4], Jee Hyun Rho[1], Sang-Yong Eom[5], Dahee Han[3], Kyung-Hwa Choi[6], Jang-Hee Lee[7], Jeeyoung Kim[2], Sungho Won[8], Hee-Gyoo Kang[9], Sora Mun[9], Hyun Ju Yoo[10], Jung-Woong Kim[11], Kwan Lee[12], Won-Ju Park[13], Seongchul Hong[14], Young-Seoub Hong[1,4]



Late Integration 분석 사례: Heavy metal exposure and its effect on APOC3 [Kim et al. 2025]

◆다중오믹스 샘플 및 연구의 흐름도

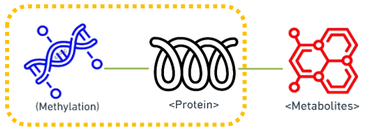
환경취약지역: 장항(제련소), 고성(폐광), 상촌(폐광)
대조지역: 김해



Late Integration 분석 사례: Heavy metal exposure and its effect on APOC3 [Kim et al. 2025]

◆후성유전체 – 단백체에서 공통적으로 유의한 유전자 추론

- Method
  - Meta-P 분석



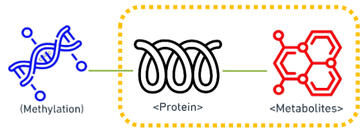
- Results
  - 메틸레이션과 단백체의 지역변수와의 유의성을 통합한 분석에서 5개의 유전자가 유의하게 나타남. APOC3 유전자의 경우 메틸레이션과 단백체 모두 오염지역에서 유의한 증가를 보임.

Gene name	Ensembl gene id	Methylation	Methylation meta t*	Methylation meta P	Protein t*	Protein P	Meta BH-P
APOH	ENSG00000091583	chr17: 66251975-66252185	0.55	0.58	-5.1	5.2E-07	**0.001**
APOC3	ENSG00000110245	chr11: 116827606-116827714	2.5	0.013	4.0	9.4E-05	**0.003**
A1BG	ENSG00000121410	chr19: 58351984-58352158	1.74	0.08	3.8	1.5E-04	**0.026**
APOC1	ENSG00000130208	chr19: 44918454-44918752	0.20	0.84	4.8	3.2E-06	**0.006**
CFAI	ENSG00000205403	chr4: 109792432-109792513	1.09	0.28	5.4	1.2E-07	**1.04E-04**

Late Integration 분석 사례: Heavy metal exposure and its effect on APOC3 [Kim et al. 2025]

◆중금속 – 단백체 – 대사체 간의 인과성 추론

- Method
  - Mediation analysis



- Results
  - 노출지역에서 bCd와 bHg가 증가할 때, 각각 APOC3와 ZA2G 단백체를 양의 방향과 음의 방향으로 중개하여 Inosine monophosphate(IMP), 3-phosphoglyceric acid(3PG), Serotonin 대사체가 증가하였음

Heavy metal (treatment)	Protein (mediator)	Metabolite (outcome)	Effect size (M)	P (M)	BH-P (M)	Effect size (D)	P (D)	BH-P (D)	Effect size (T)	P (T)	BH-P (T)
bCd	APOC3	IMP	0.0170	<0.001	<0.001	0.064	0.038	1.000	0.081	0.0160	0.592
		Serotonin	0.023	<0.001	<0.001	0.138	0.004	0.156	0.160	<0.001	<0.001
		Linoleic acid	0.0068	<0.001	<0.001	-0.033	0.016	0.592	-0.026	0.056	1.000
		C24 Cer	0.0079	<0.001	<0.001	-0.0195	0.188	1.000	-0.0116	0.428	1.000
		PE(16:0/18:2)	0.0114	<0.001	<0.001	0.0181	0.458	1.000	0.029	0.212	1.000
		3PG	0.0125	0.002	0.064	0.115	0.000	0.000	0.127	0.000	0.000
bHg	ZA2G	IMP	-0.0144	0.002	0.084	0.061	0.000	0.000	0.046	0.000	0.000
		3PG	-0.0099	0.002	0.084	0.068	0.000	0.000	0.058	0.000	0.000
		Serotonin	-0.0166	0.002	0.084	0.077	0.000	0.000	0.060	0.004	0.100
		Spermidine	-0.0088	0.002	0.084	0.051	0.000	0.000	0.042	0.000	0.000
		FBP	-0.0119	0.002	0.084	0.022	0.330	1.000	0.0103	0.630	1.000
		Putrescine	-0.0089	0.002	0.084	0.0053	0.696	1.000	-0.0036	0.834	1.000

Late Integration 분석 사례: Heavy metal exposure and its effect on APOC3 [Kim et al. 2025]

◆① 후성유전체 – 단백체, ② 단백체 – 대사체 간의 인과성 추론

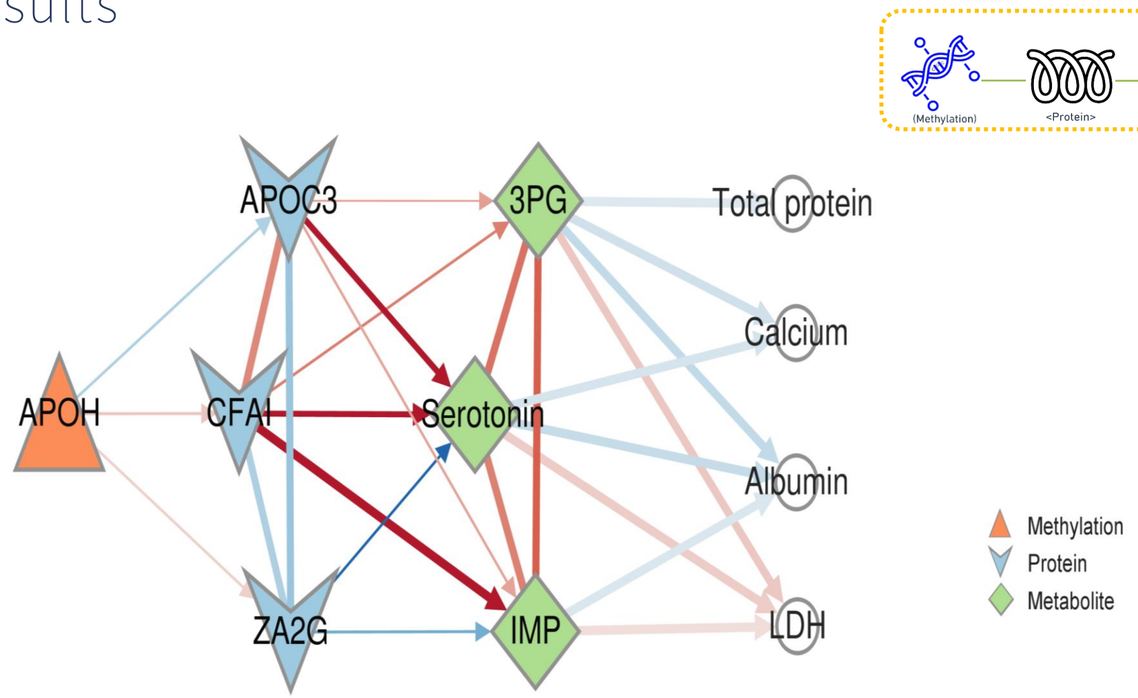
- Method
  - 맨델리안 랜덤화 기법 (MR; 2-stage least squares)
- Results
  - APOH 메틸레이션이 변화함에 따라 APOC3, CFAI, ZA2G 단백체가 변화하였으며, 오염 지역에서 APOC3, CFAI 단백체 증가, ZA2G 단백체 감소 시 IMP, 3PG, Serotonin 대사체가 증가하였음

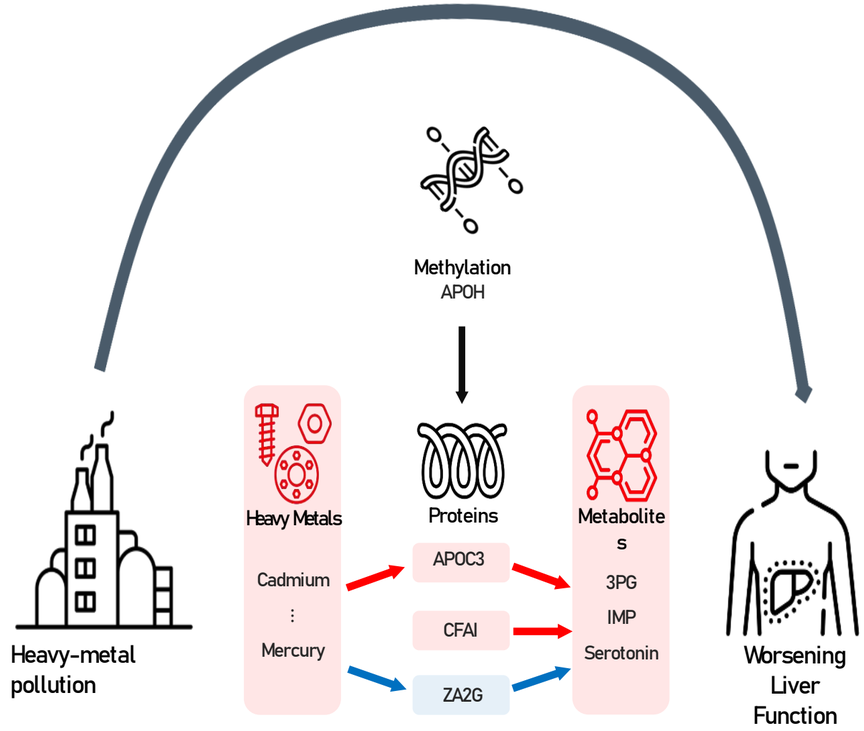
Exposure	Outcome	Weak Instrument P-value	DWH P-value	Sargan P-value	Estimate (β)	P-value
APOH (ME)	APOC3 (P)	3.7E-24	0.174	0.32	-0.22	0.017
APOH (ME)	CFAI (P)	9.3E-23	0.130	0.24	0.176	0.035
APOH (ME)	ZA2G (P)	8.5E-47	0.076	0.66	0.139	0.048
APOC3 (P)	IMP (M)	7.5E-18	0.53	0.95	0.30	0.023
APOC3 (P)	3PG (M)	7.5E-18	0.28	0.67	0.32	0.009
APOC3 (P)	Serotonin (M)	7.5E-18	0.26	0.83	0.56	0.003
CFAI (P)	IMP (M)	4.3E-21	0.099	0.88	0.86	<0.001
CFAI (P)	3PG (M)	4.3E-21	0.82	0.87	0.44	0.038
CFAI (P)	Serotonin (M)	4.3E-21	0.76	0.69	0.94	0.004
ZA2G (P)	IMP (M)	1.80E-24	0.72	0.25	-0.40	0.014
ZA2G (P)	Serotonin (M)	1.80E-24	0.64	0.44	-0.58	0.015

Late Integration 분석 사례: Heavy metal exposure and its effect on APOC3 [Kim et al. 2025]

◆ 다중 오믹스와 임상 변수 통합 네트워크 관계도

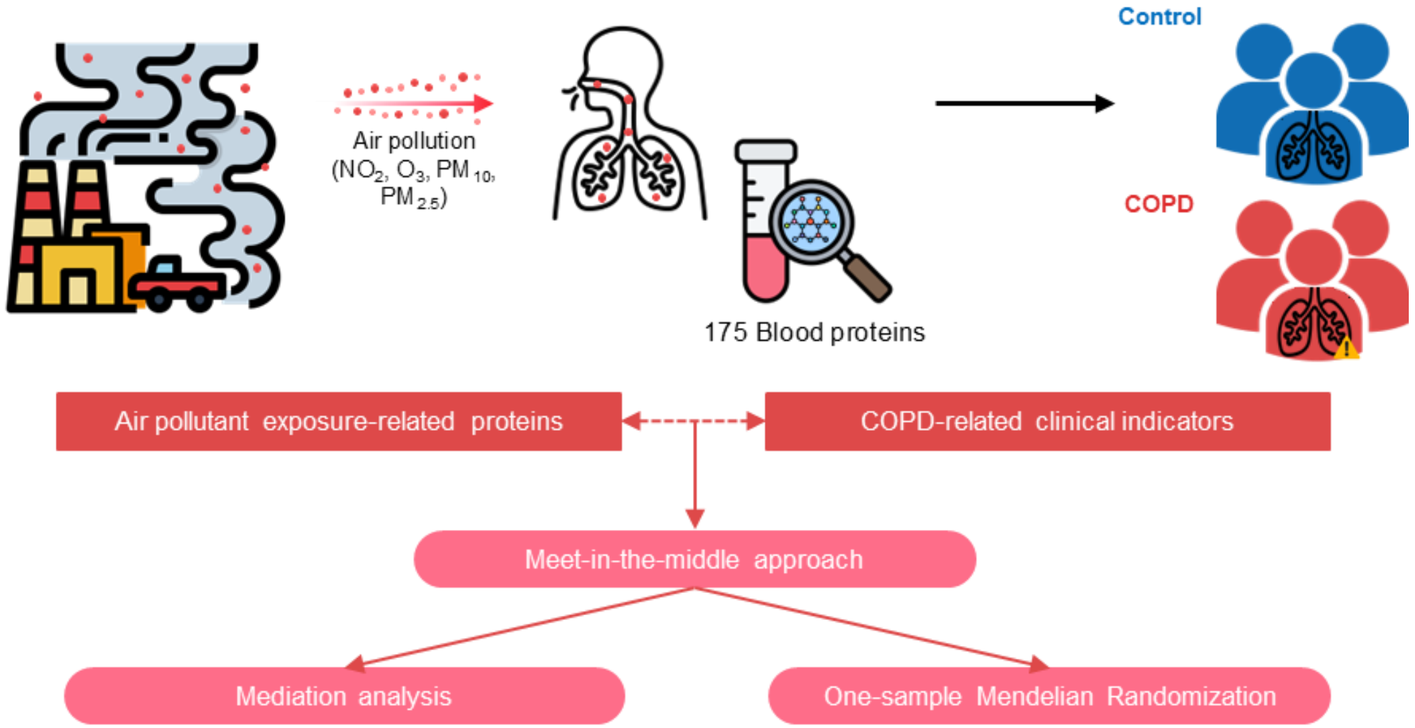
- Method
  - 다중회귀분석 (multiple linear regression) & Mediation, MR 결과 정리
- Results





Red lines signify positive correlations, whereas blue lines represent negative correlations, in association test, mediation analysis, and MR. Arrows at the end of lines indicate relationships identified through MR and mediation analyses, while lines without arrows represent results from association tests. The intensity of the color and thickness of the lines reflect the effect size and level of significance, respectively.

Late Integration 분석 사례: air pollution-related proteomic analysis in COPD [Unpublished data]



Dataset

- N = 486명 (COPD: 318명, non-COPD: 168명)
- Main analysis: male-specific

Omics

- Proteome
- Genome

Phenotypes

- Air pollutants
  - NO_2, O_3, PM_{10}, $PM_{2.5}$
- COPD phenotypes
  - COPD 진단
  - COPD phenotypes (FEV1, FVC, FEV1/FVC)

- Air pollutants exposure-related plasma proteins 식별
- Meet-in-the-middle analysis: Proteins - COPD-related clinical indicators associations
- Mediation & One-sample MR: air pollutants – proteins – COPD-related clinical indicator pathway

Discussion

- The goal of biosciences: Full understanding and predictive modelling of biological systems
- But the global genome-wide studies describe systems of a size that cannot be modelled to this level in the foreseeable future.
- Functional interpretation is attempted by integrative studies and systems biology but both of these techniques are still too high level to provide full functional explanations at a molecular or atomic level.
- This level of understanding will be the result of bottom-up approaches which provide a more detailed understanding of smaller systems or fewer genes.
- We are presently seeing the rise of high throughput studies.
  - The near future will probably see Mathematical Modelling being important to everyone.
  - and/or advances on Integrative Biology (top-down) & Systems Biology (bottom-up) and its relations

Acknowledgement

- 김남은 Ph.D., 서울대학교
- 허민 MS, 서울대학교
- 박재현 Ph.D., Vanderbilt Univ
- 김우진 MD/Ph.D., 강원대병원
- 김태범 MD/Ph.D., 서울아산병원
- 홍수종 MD/Ph.D., 국립중앙의료원
- 김봉수 Ph.D., 이화여대

Q & A

서울대학교 보건대학원 | 원성호 교수 | won1@snu.ac.kr