환경성질환 연구방법론

Sungho Won, Ph.D.

Department of Public Health Sciences

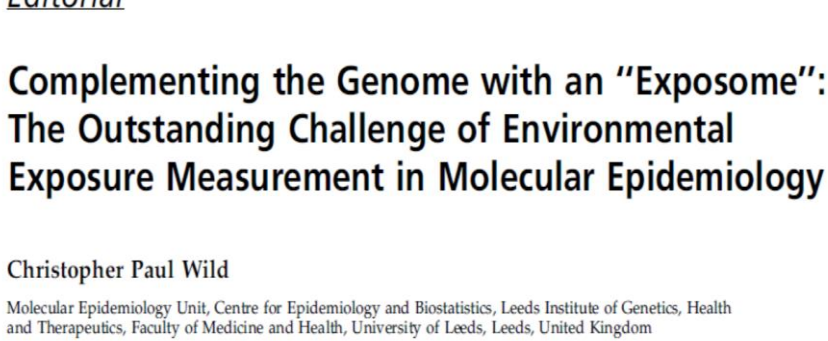
Seoul National University, Seoul, South Korea

Background

Environmental Diseases

- Definition
  - Environmental diseases (ENVDs) are diseases that can be directly attributed to environmental factors (as distinct from genetic factors or infection). *By wiki*
  - Example: asthma, allergy, atopic dermatitis, allergic rhinitis, etc

- Exposome
  - It was first proposed by Wild to encompass the totality of human environmental (meaning all non-genetic) exposures from conception onwards, complementing the genome.
  - This concept was developed to draw attention to the need for better and more complete environmental exposure data, in order to balance the investment, tools and knowledge in genetics.



Editorial

Complementing the Genome with an "Exposome": The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology

Christopher Paul Wild

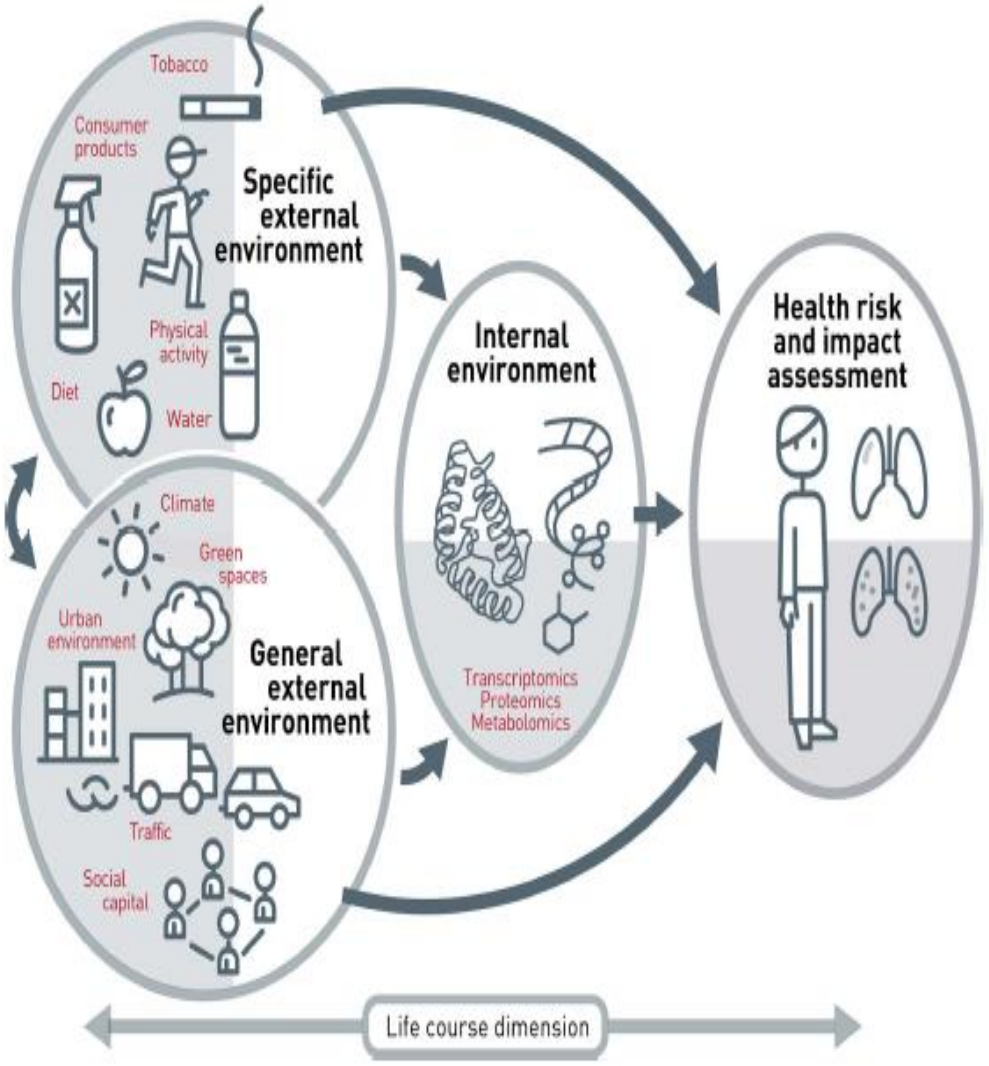
Molecular Epidemiology Unit, Centre for Epidemiology and Biostatistics, Leeds Institute of Genetics, Health and Therapeutics, Faculty of Medicine and Health, University of Leeds, Leeds, United Kingdom

Environmental Diseases

- Characteristics of exposome
  - Exposome is dynamic and vary on an hourly to yearly basis in the external and internal environments. This is opposed to the static genome.
  - Exposome do not have the same effect during the various developmental periods that are critical to health and disease.
  - Early-life may be important for defining the exposome because it is well recognized that the periods of organ development during prenatal life and infancy are especially vulnerable to the effects of environmental risk factors.

Environmental Diseases

- Domains of exposome
  - External exposome
    - General (community-level) external environment : urban environment, climate factors, social capital, stress, etc
    - Specific (individual-level) external environment : contaminants, diet, physical activity, tobacco, infections, etc
  - Internal exposome
    - Internal environment: metabolic factors, gut micro-flora, inflammation, oxidative stress



Vrijheid, Thorax, 2014

External Exposomes

- Individual-level exposures
  - Can be achieved through better modelling. For instance by using predictive exposure models that combine questionnaire information with biomarkers and monitors for validation (cotinine biomarkers & questionnaire)
  - Smartphone-linked diaries and imaging are also promising tools for more accurate and complete exposure assessment, for example of diet and use of consumer products
- Community-level exposures
  - Major improvements can be achieved by improving information on where people are, how they move through their environment, and in case of air pollution, how much air they inhale.
  - Smartphone applications that integrate global positioning systems location data with physical activity information and pollution measurements, are now being developed to better characterize the exposome.

Statistical Analyses

- Omics-based approach
  - The main contribution of omics techniques is to measure profiles of the biological response to a cumulative exposure experience.
  - Omics analysis: high-throughput molecular 'omics' techniques can analyze complete sets of biological molecules: smaller molecules (metabolomics), larger molecules (proteome), gene expression profiles (transcriptomics and epigenomics) and reactive electrophiles (adductomics).
  - Gene-environment interaction analysis: it is defined as "a different effect of an environmental exposure on disease risk in subjects with different genotypes" or "a different effect of a genotype on disease risk in subjects with different environmental exposures

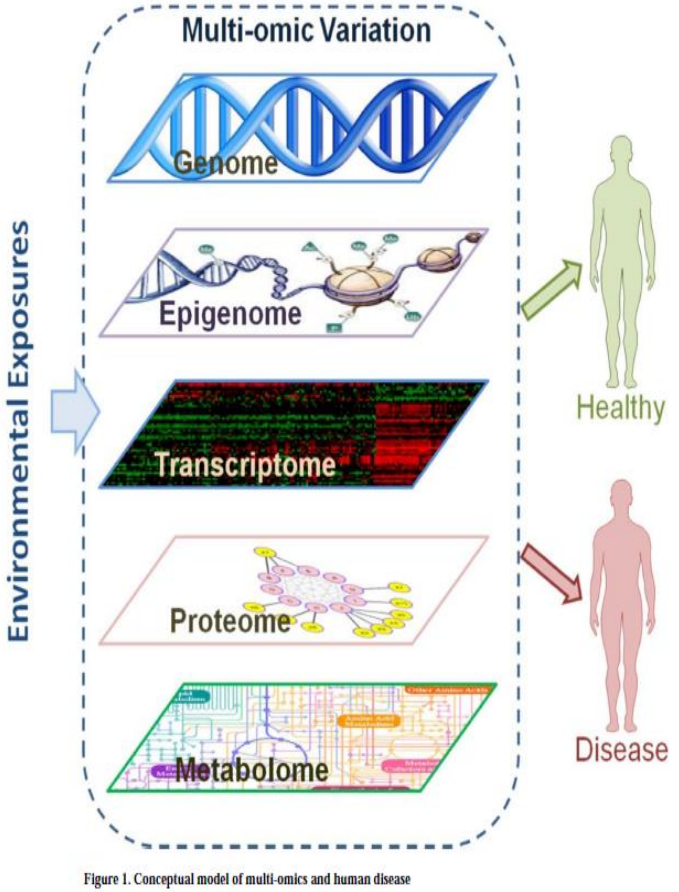


Figure 1. Conceptual model of multi-omics and human disease

Sun and Hu, Adv Genet, 2016

Environmental Diseases

- One-exposure-one-health-effect analysis
  - Traditional approaches
  - Regression models are often utilized
- Multiple and combined exposure analysis
  - Environment-wide association analysis: the evaluation of risk estimates for many single exposures in an agnostic, hypotheses generating manner.
  - Dimension reduction method : the risk estimates for combined exposures through data-driven dimension reduction methods are evaluated by using PCA, etc.
  - Group-based analysis: the risk estimates for groups of subjects sharing a similar exposome are evaluated by using a Bayesian profile regression analysis, etc.

Single-Omics Analysis

Single-Omics Analysis

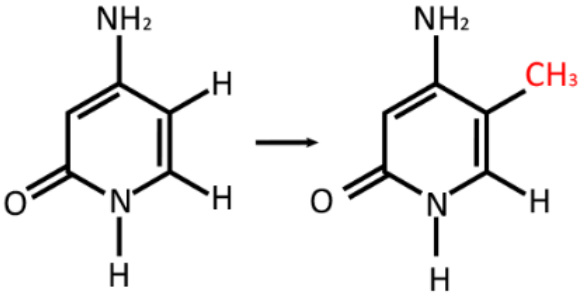
- Omics
  - Omics are used to form nouns meaning a study of the totality of something such as genomics, proteomics. *By wiki*
- Importance of Omics for exposome-health analysis
  - Biological processes, such as the development of human diseases, involve a highly dynamic and interactive system of molecular layers (e.g. genetics, epigenetics, mRNA transcripts, proteins and metabolites) and are influenced by many environmental factors.
  - Recent technological advancement has permitted high-throughput measurement of human genome, epigenome, metabolome, transcriptome and proteome at the population level

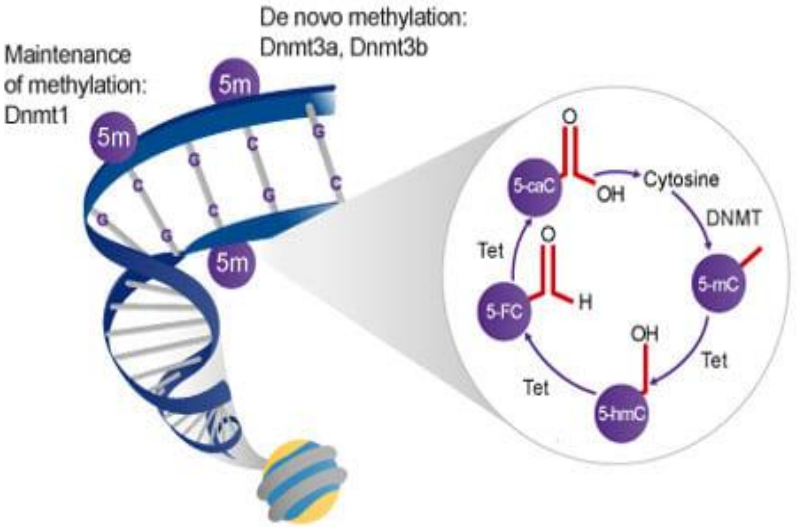
Single-Omics Analysis

- Types of Omics
  - Epigenomics
  - Genomics
  - Transcriptomics
  - Proteomics
  - Metabolomics
  - Microbiomics

DNA Methylation

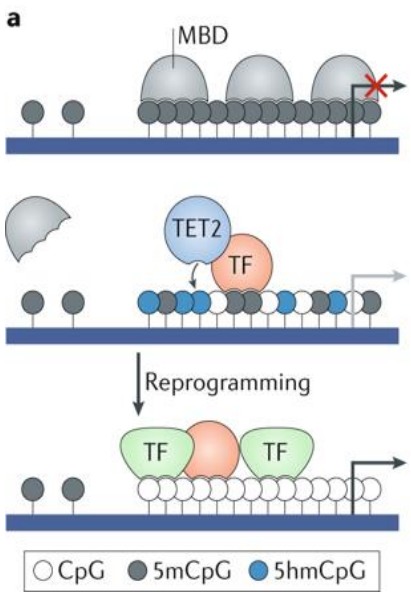
- Addition of a methyl group at the fifth carbon position of the cytosine base
- Numerous environmental such as organic pollutants, smoke, nutritional factors triggers global or site-specific DNA methylation alterations
- Transgenerational epigenetic inheritance
- Tissue specificity
  - Comprehensive epigenomic maps in multiple human tissues
  - Roadmap Epigenomics (www.roadmapepigenomics.org)
  - International Human Epigenome Consortium (ihec-epigenomics.org)
  - Epigenetic modification + genetic variants

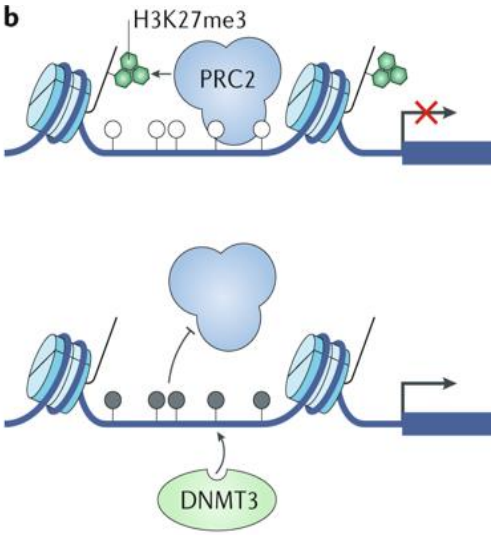


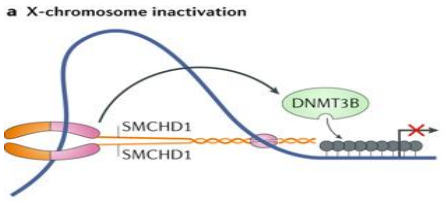


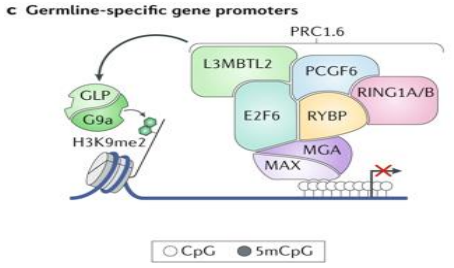
Role of DNA methylation

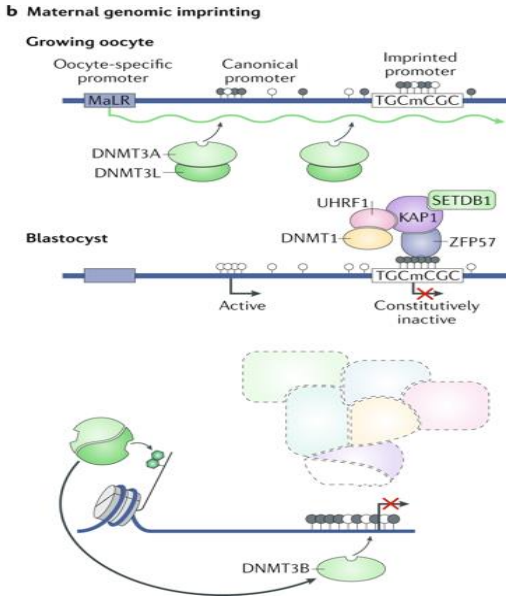
- Gene expression regulation
- Reprogramming : tissue differentiation & cell lineage
- Imprinting
- X-chromosome inactivation
- Genomic stability











Greenberg, M. V., & Bourc'his, D. (2019)

Epigenome-wide Association Studies

- Epigenome-wise association study (EWAS)
  - Discover correlation between phenotypes and epigenome data, such as DNA methylation
  - Use omics data made with microarray, bisulfite sequencing, ChIP-seq, …
  - Example: cord blood DNA methylation & prenatal lead exposure

ELSEVIER

Contents lists available at ScienceDirect

Environmental Research

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

Check for updates

Prenatal lead exposure and cord blood DNA methylation in the Korean Exposome Study

Jaehyun Park [a], Jeeyoung Kim [b], Esther Kim [b], Woo Jin Kim [b,*], Sungho Won [a,c,**]

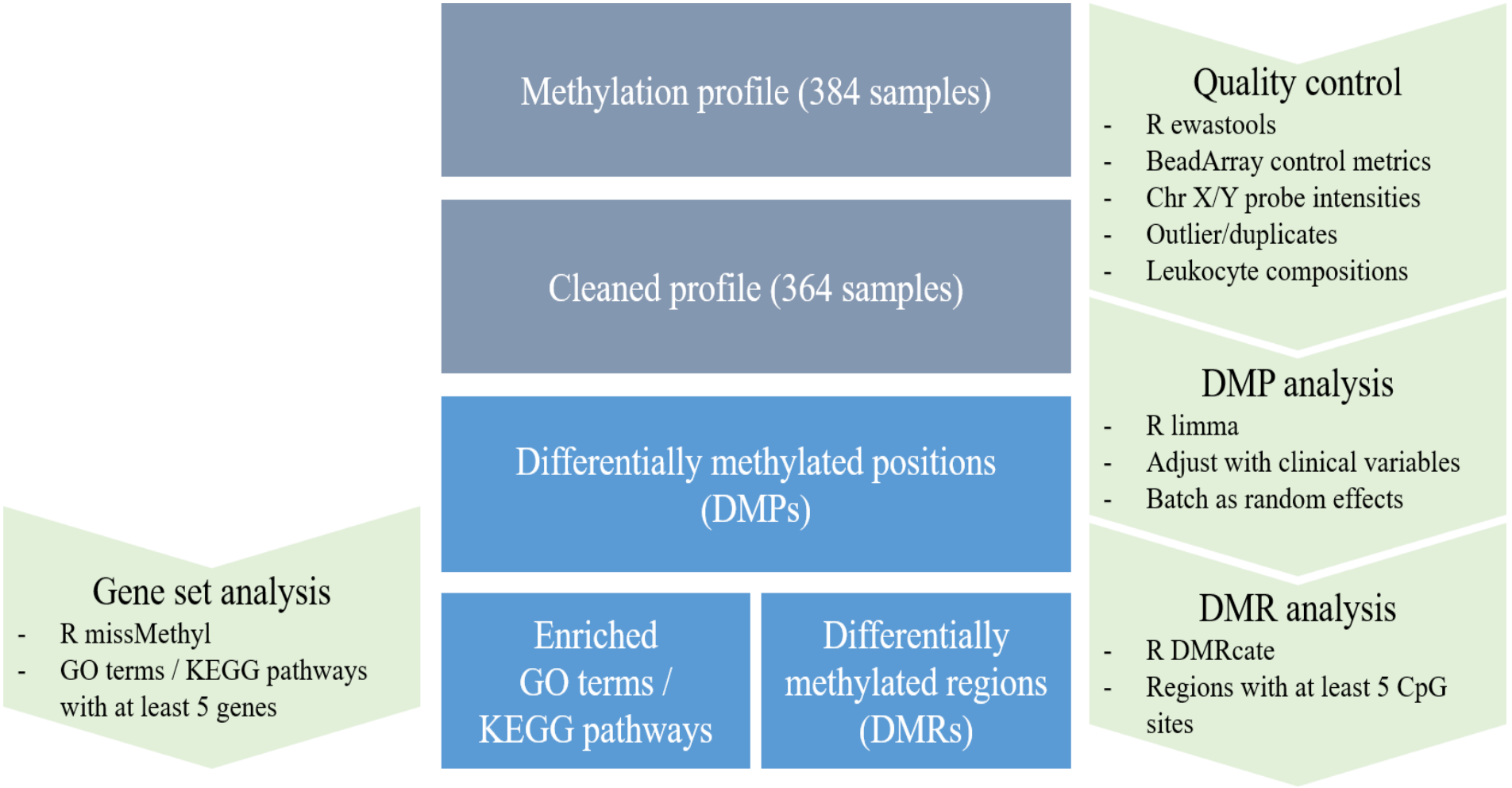
[a] Interdisciplinary Program of Bioinformatics, College of Natural Sciences, Seoul National University, Seoul, South Korea
[b] Department of Internal Medicine and Environmental Health Center, Kangwon National University, Chuncheon, South Korea
[c] Department of Public Health Sciences, Seoul National University, Seoul, South Korea

Epigenome-wide Association Studies

- Example: prenatal lead exposure (Park *et al*. 2021)
  - Association between DNA methylation from cord blood and prenatal lead exposure
  - Meta-analysis combining results from 3 centers
  - Linear model using limma, adjusting with infant sex, maternal pre-pregnancy BMI, maternal smoking status, estimated leukocyte compositions

Epigenome-wide Association Studies

- Example: prenatal lead exposure (Park *et al*. 2021)



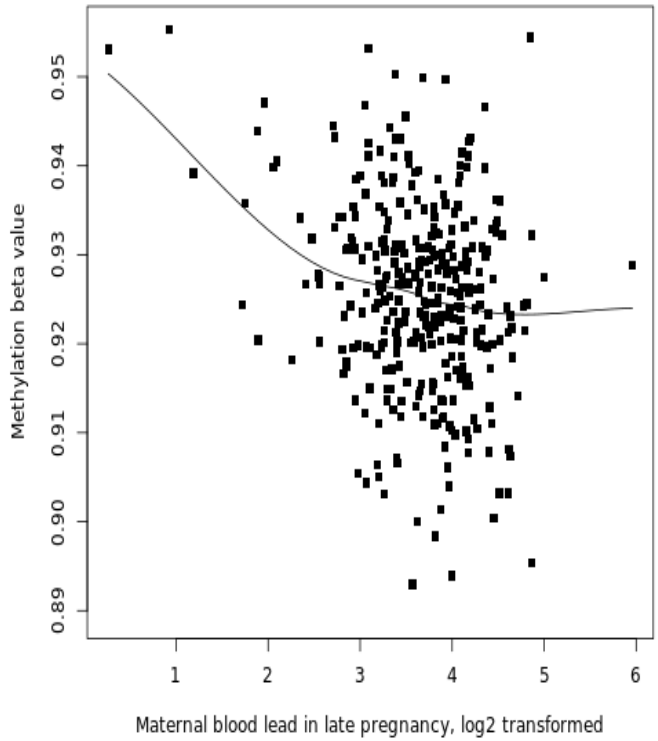
Epigenome-wide Association Studies

- Example: prenatal lead exposure (Park *et al*. 2021)
  - Meta-analysis combining results from 3 centers
  - 1 CpG site associated with maternal blood lead levels in late pregnancy

CpG ID	Chr	Position	Effect size (center 1)	P-value (center 1)	Effect size (center 2)	P-value (center 2)	Effect size (center 3)	P-value (center 3)
cg01912525	17	75,013,238	-6.71×10^{-3}	1.38×10^{-8}	-1.27×10^{-3}	0.222	-5.81×10^{-3}	0.095

P-value (combined)	Adjusted p-value*	Gene annotation	CpG island annotation
1.05×10^{-7}	0.0875	(Unknown)	(Unknown)

* P-values adjusted with Benjamini-Hochberg method



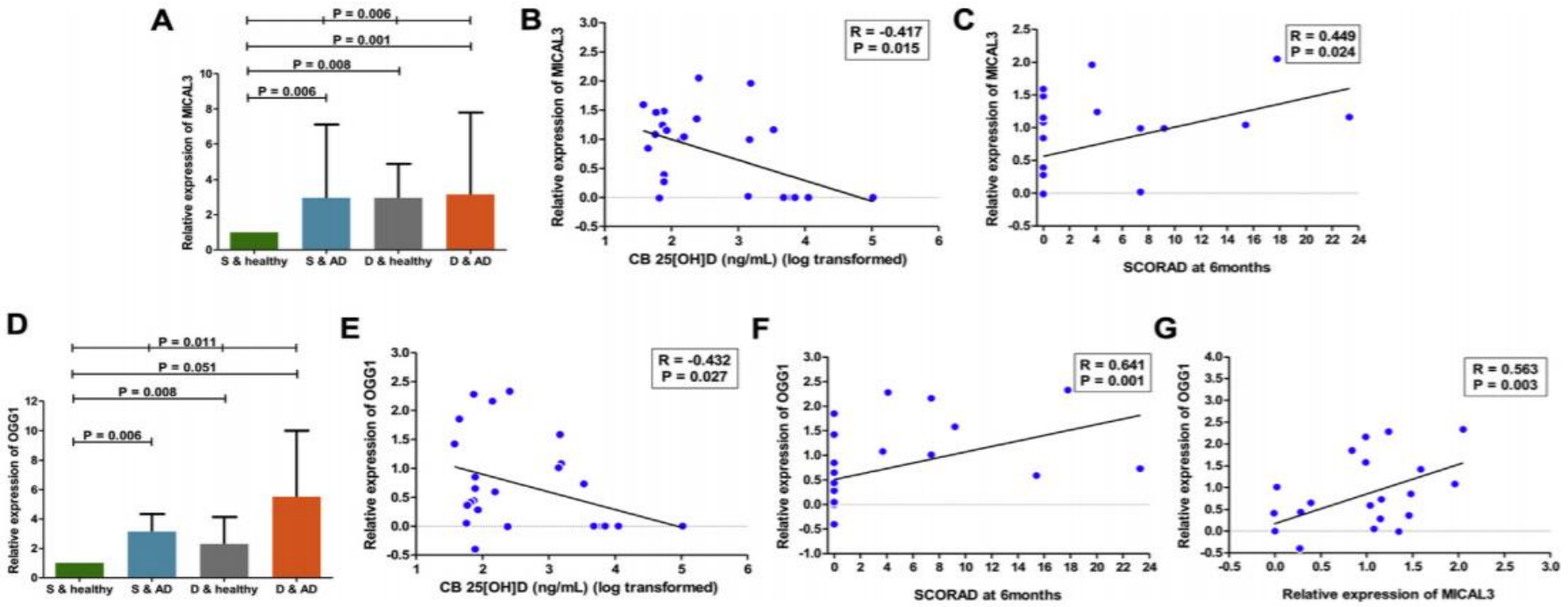
Park et al, Environmental Research, 2021

Epigenome-wide Association Studies

Prenatal 25-hydroxyvitamin D deficiency affects development of atopic dermatitis via DNA methylation

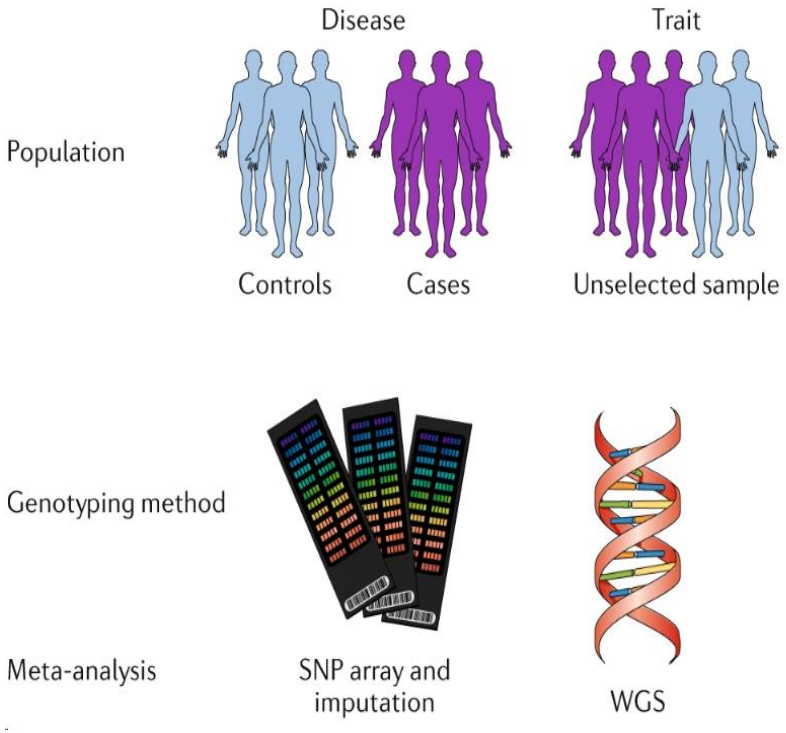
Hyun-Ju Cho, MD * • Youn Ho Sheen, MD, PhD * • Mi-Jin Kang, PhD • ... Sung-Ok Kwon, PhD • Se-Young Oh, PhD • Soo-Jong Hong, MD, PhD ✉ • Show all authors • Show footnotes

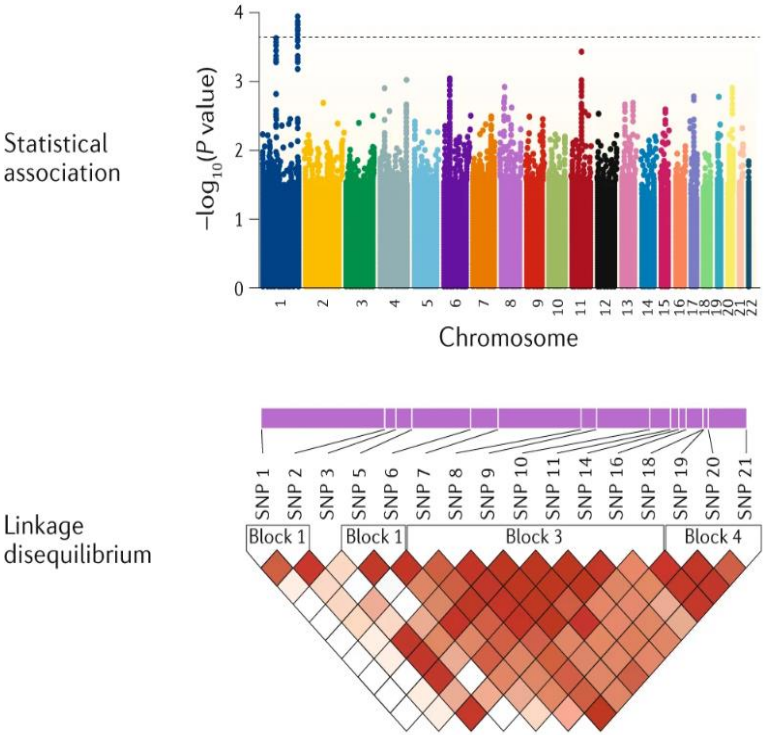
J Allergy Clin Immunol, 2019



Genome-wide Association Study

- Test genetic variants to identify genotype–phenotype association
  - Relatively stable within an individual
  - Successful in identifying novel variant–trait associations and biological mechanisms





Vivian Tam. (2019)

Genome-wide Association Studies

- Genotyping methods
  - PCR-based methods
  - Sequencing-based methods
  - Array-based hybridization

Platform	Total marker N	Annotated marker[a] N	Nonsyn marker[b] N (%)	ASN marker[c] N (%)
Affymetrix 5.0	500,568	489,457	2,179 (0.4)	769 (0.2)
Affymetrix 6.0	934,969	892,584	4,889 (0.5)	1,750 (0.2)
Illumina Omni 1 M	1,099,726	1,066,324	45,832 (4.3)	12,516 (1.2)
Illumina Exome array	242,901	241,923	217,775 (90.0)	39,480 (16.3)
Illumina GSA	700,078	688,062	87,759 (12.8)	21,371 (3.1)
Axiom Biobank	718,212	645,060	251,080 (38.9)	46,416 (7.2)
Axiom UK Biobank	845,487	823,336	104,058 (12.6)	19,487 (2.4)
Axiom PMRA	920,744	856,797	44,819 (5.2)	6,088 (0.7)
KoreanChip	**833,535**	**829,635**	**183,607 (22.1)**	**89,413 (10.8)**

Moon et al. Sci Rep, 2019

Genome-wide Association Studies

- Role of GWAS SNP Array

Analysis	Purpose	Discoveries
GWAS	detecting trait-SNP associations	~10,000 robust associations with diseases and disorders, quantitative traits, and genomic traits
Genome-wide CNV analysis	detecting trait-CNV associations	hundreds of associations with diseases and disorders
Genome-wide assessment of LD	quantifying genome architecture	large variation in LD in the genome
Estimation of SNP heritability[a]	genetic architecture	large proportion of genetic variation captured by common SNPs
Estimation of genetic correlation[a]	detecting and quantifying pleiotropy	pleiotropy is ubiquitous
Polygenic risk scores[a]	detecting pleiotropy; validating GWAS discoveries	out-of-sample prediction works as expected; detection of novel trait associations
Mendelian randomization[a]	testing causal relationships	replication of known causal relationships; empirical evidence of observational associations that are not causal
Population differences in allele frequencies	reconstructing human population history; detecting selection	genetic structure can mimic geographical structure; evidence of natural selection
Trait GWAS with -omics GWAS[a]	fine-mapping; detecting target genes; function	two-thirds of GWAS-associated loci implicate a gene that is not the nearest gene to the most associated SNP

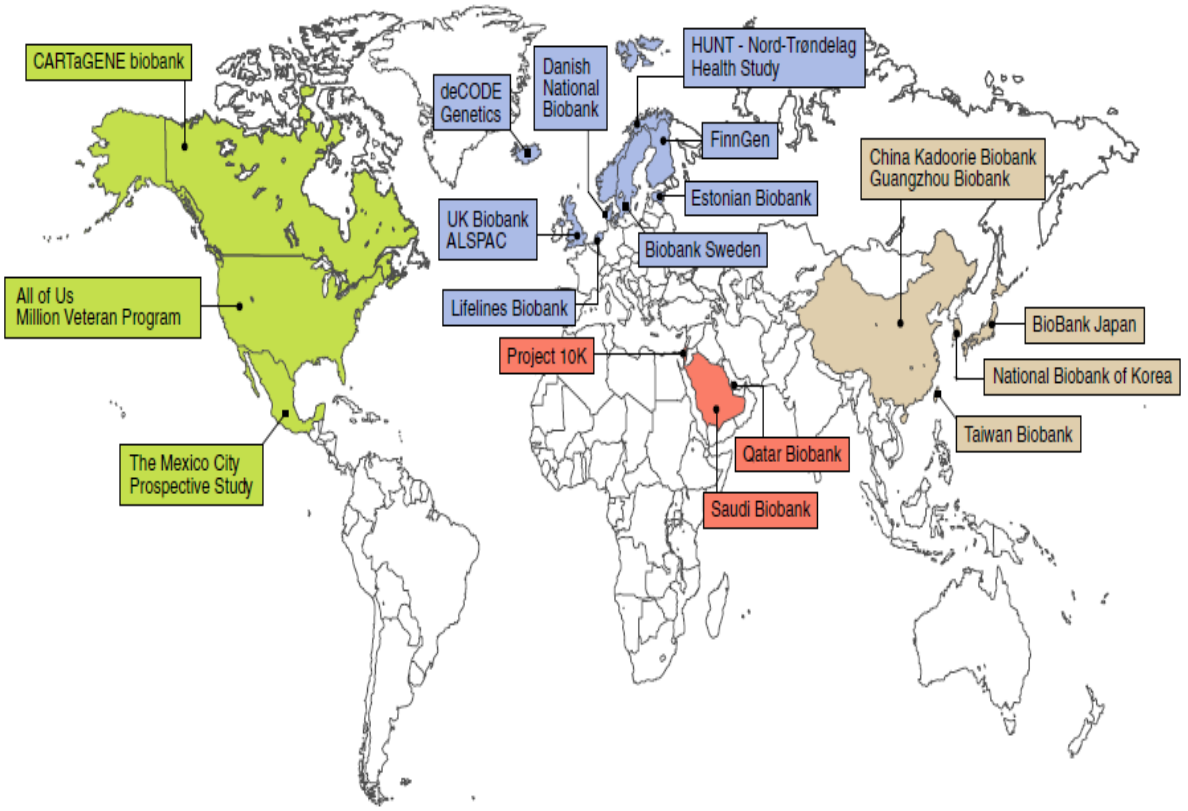
[a]These analyses can be performed with GWAS summary statistics.

Visscher et al. Am J Hum Genet, 2017

Genome-wide Association Studies

- Mega Biobank

Location	Biobank	N (goal)
Canada	CARTaGENE biobank[119]	43,000
USA	All of Us[33] Million Veteran Program[49]	1,000,000 > 600,000
Mexico	The Mexico City Prospective Study[52]	150,000
Iceland	deCODE Genetics	500,000
UK	UK Biobank[38] Avon Longitudinal Study of Parents and Children (ALSPAC)[20]	500,000 > 15,000
Netherlands	Lifelines Biobank[120]	> 167,000
Denmark	Danish National Biobank[121]	
Norway	HUNT - Nord-Trøndelag Health Study[122]	125,000
Sweden	Biobank Sweden	
Finland	FinnGen	500,000
Estonia	Estonian Biobank[123]	52,000
Israel	Project 10K	10,000
Saudi Arabia	Saudi Biobank	200,000
Qatar	Qatar Biobank[124]	60,000
China	China Kadoorie Biobank[51] Guangzhou Biobank[125]	> 500,000 30,000
Japan	BioBank Japan[126]	200,000
Korea	National Biobank of Korea[127]	500,000
Taiwan	Taiwan Biobank[128]	200,000



Shilo et al., Nat Med, 2020

UKBiobank Data

- UKBiobank data
  - Consists of 500K volunteers in the UK
  - Enrolled at ages from 40 to 69.
  - Initial enrollment took place during 2006 - 2010, and the volunteers will be followed for at least 30 years thereafter.
  - Affymetrix UK BiLEVE Axiom array: 50,000 + 450,000
  - Whole Exome data : 50K

KoGES Phenotype Dataset

▪ 지역사회 기반 (안산/안성) 코호트

✓ 2001년부터 2년 마다 기본정보, 일반정보, 생활습관 (음주력, 흡연력), 신체활동, 의료정보, 질병과거력, 치료력, 약물력, 가족력, 여성력, 호흡 순환기 질환, 관절질환, 수면력, 스트레스, Type A 진단, 식습관, 식품섭취 빈도조사, 임상검사, 신체계측, 폐기능, 심전도, 흉부 X-ray에 대한 설문 추적조사를 실시함

조사 구분	조사 연도	참여자 수 (명)	변수 개수
기반	'01 – '02년	10,030	1,938
1 차 추적	'03 – '04년	8,603	1,769
2 차 추적	'05 – '06년	7,515	2,482
3 차 추적	'07– '08년	6,688	2,086
4 차 추적	'09 – '10년	6,665	2,448
5 차 추적	'11 – '12년	6,238	2,600
6 차 추적	'13 – '14년	5,906	2,389
7 차 추적	'15 – '16년	6,318	1,854
8 차 추적	'17 – '18년	6,157	1,360

KoGES Phenotype Dataset

▪ 도시 기반 코호트

✓ 2004년부터 기본정보, 일반정보, 질병과거력, 수술력, 약물력, 가족력, 검진력, 생활습관 (체중변화, 흡연, 음주, 신체활동), 사회적심리스트레스, 여성력, 식습관, 식품섭취 빈도조사, 신체계측, 임상검사, 개방형 영양조사 (영양소) 에 대한 설문 추적조사를 실시함

조사구분	조사연도	참여자 수 (명)	변수 개수
기반	'04 – '13년	173,202	1,848
1 차 추적 (예비)	'07– '11년	4,606	418
1 차 추적 (CAPI)	'12 – '16년	65,611	1,065

CAPI: Computer Assisted Personal Interview

KoGES Phenotype Dataset

▪ 농촌 기반 코호트

✓ 2005년부터 기본정보, 일반정보, 질병과거력, 가족력, 생활습관 (음주력, 흡연력), 골절경험, 여성력, 우울증, 사회심리적 스트레스, 보충제, 식습관, 식품섭취 빈도조사, 신체계측, 임상검사 등에 대한 설문 추적조사를 실시함

조사구분	조사연도	참여자 수 (명)	변수 개수
기반	'05 – '11년	28,337	1,040
1 차 추적	'07– '14년	12,463	864
2 차 추적	'08 – '16년	11,399	864
3 차 추적	'11 – '16년	6,423	863
4 차 추적	'14 – '16년	1,449	205

KoGES Phenotype Dataset

▪ KoGES 통합 자료 (지역사회/도시/농촌 기반)

- ✓ 각각의 기반 자료에서 공통적으로 가지고 있는 변수들을 기준으로 함
- ✓ 일반정보, 생활습관 (음주력, 흡연력, 운동력), 약물력, 질병과거력, 치료력, 가족력, 여성력, 체성분검사, 심전도검사, 흉부 X-ray에 대한 변수 보유

KoGES 기반조사 통합자료	참여자 수 (명)	변수 개수
(지역사회, 도시, 농촌 코호트 기반조사 자료통합)	211,569	201

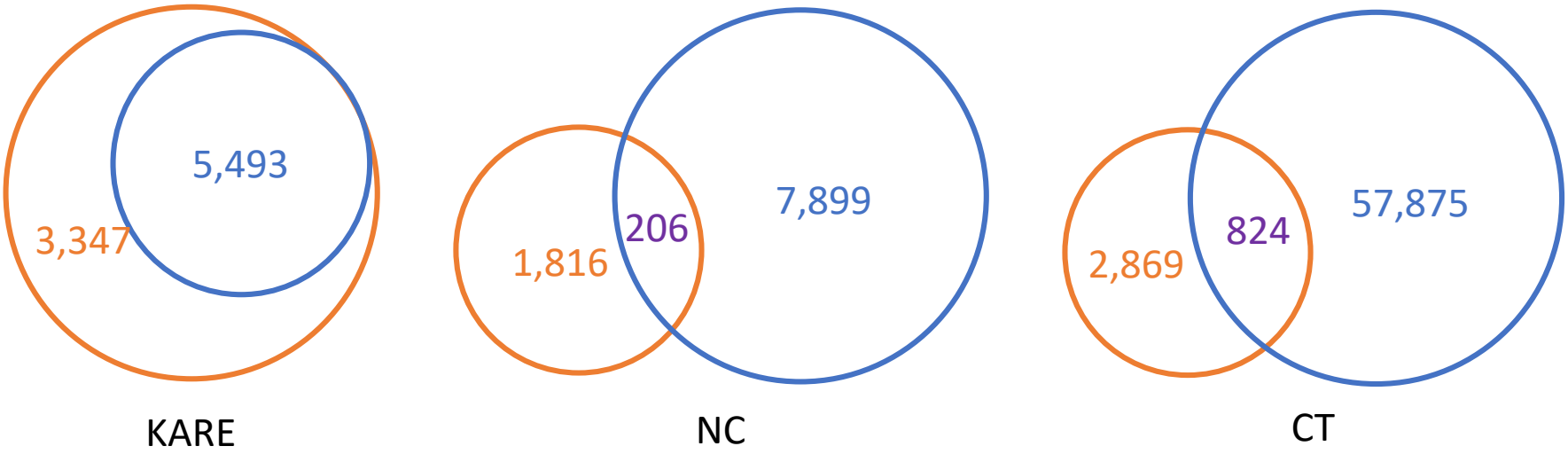
KoGES Genotype Dataset

- The number of samples by clinical data and array

Clinical data		KARE_AS		CT_HEXA		NC_CAVAS	
Baseline(1):		8,840		61,568		9,715	
Follow Up	1st (2)	7,586		52,666		5,920	
	2nd (3)	6,675		-		6,535	
	3rd (4)	5,918		-		4,620	
	4th (5)	5,907		-		1,100	
	5th (6)	5,529		-		-	
	6th (7)	5,240		-		-	
	7th (8)	5,589		-		-	
Genotype		Affy 5.0	Kchip	Affy 6.0	Kchip	Affy 6.0	Kchip
Sample		8,840	5,493	3,693	58,699	1,816	8,105
SNP		352,228	467,088	627,659	467,088	606,876	467,088

KOGES Genotype Dataset

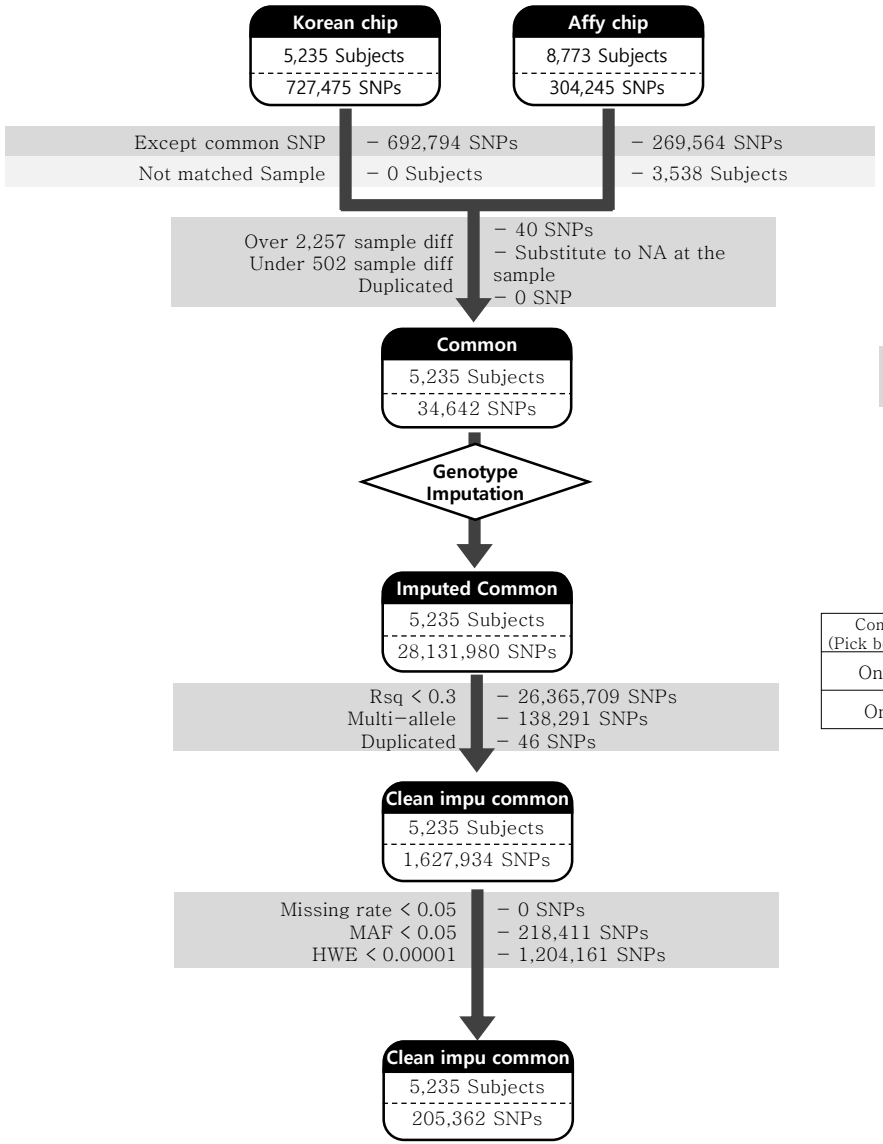
- The number of samples by array

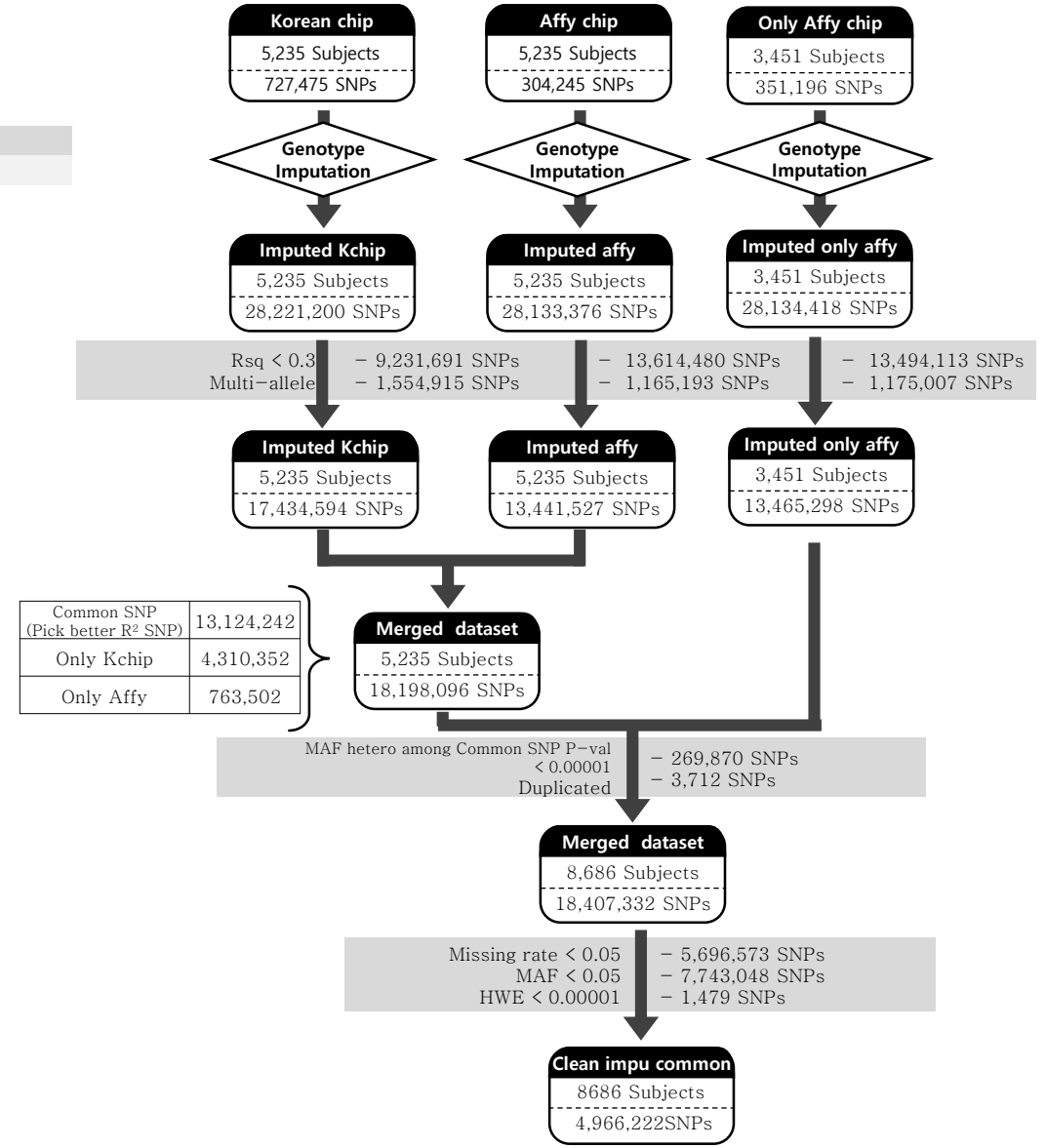


Total 80,239 samples

Korea biobank array
Affymetrix array

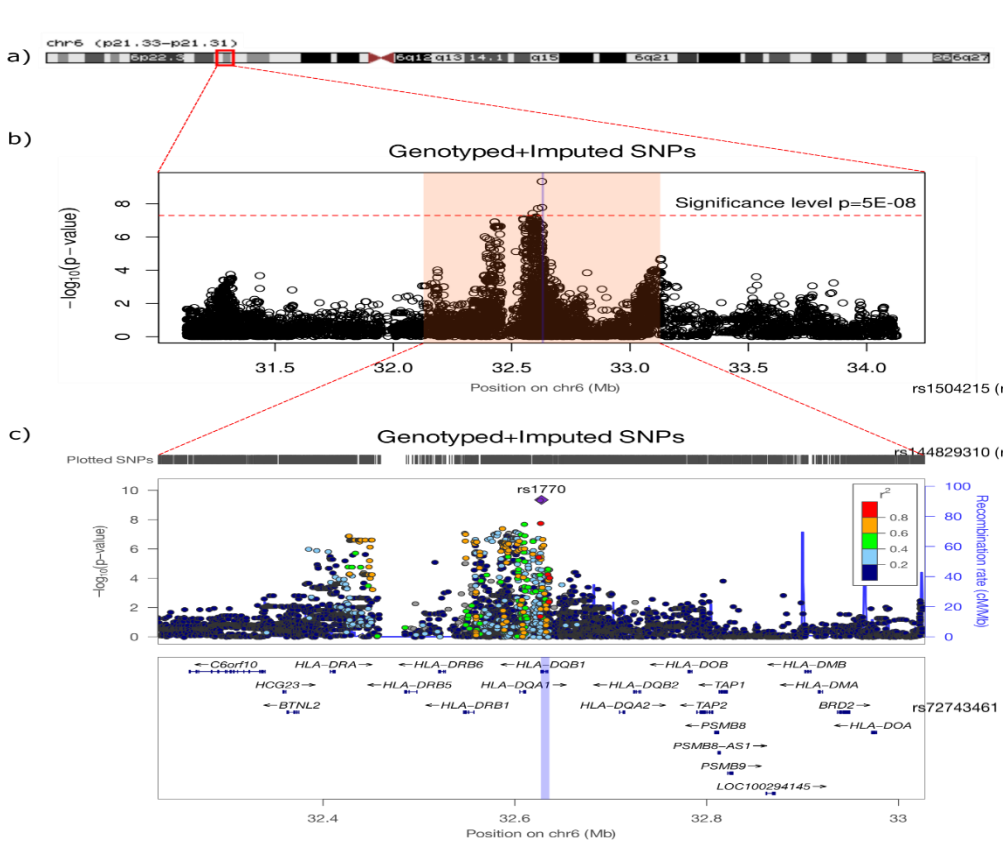
Genotype Imputation of Multiple Data

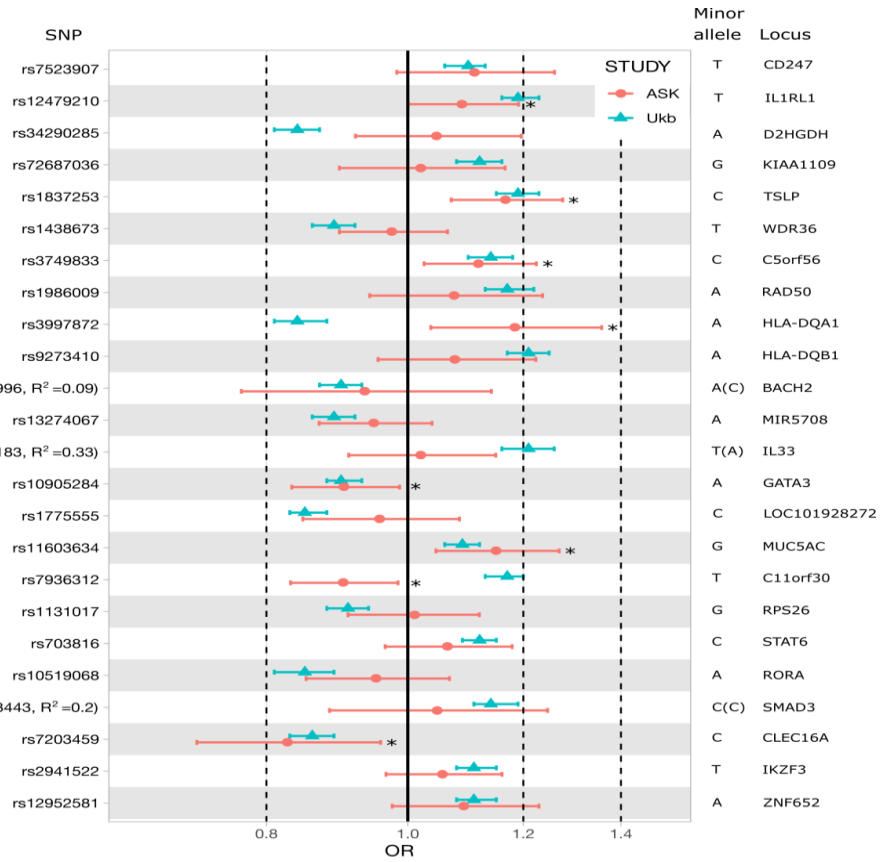




Genome-wide association study

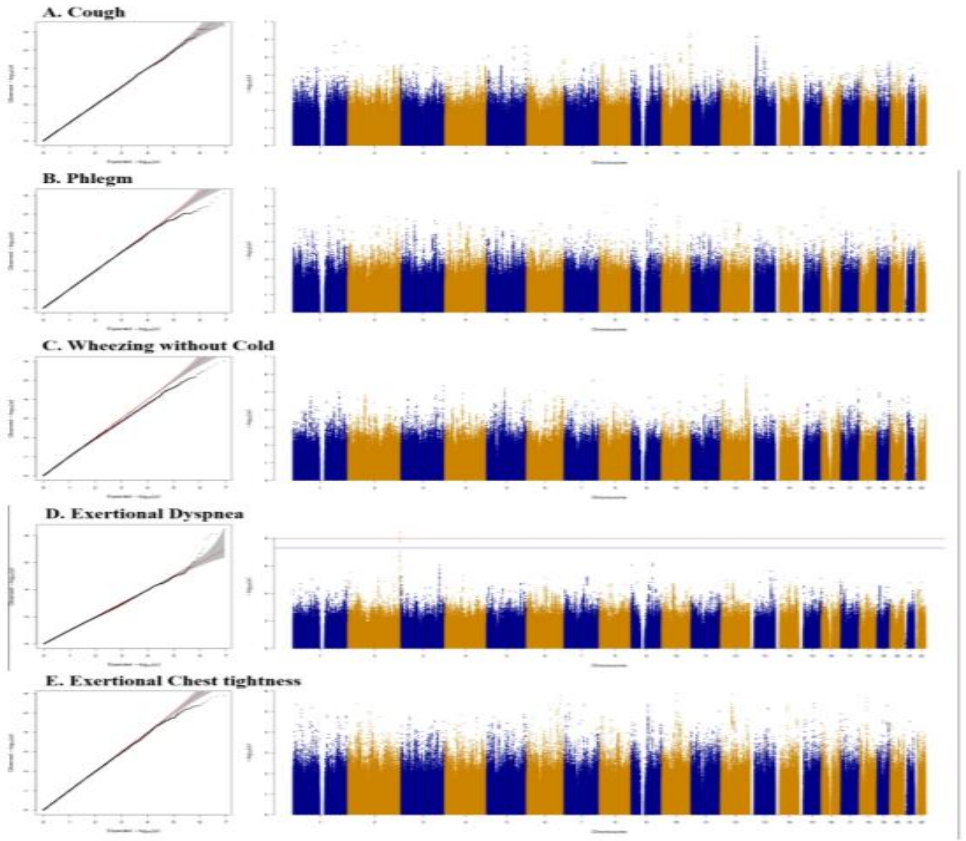
- Genome-wide Association Study of Korean Asthmatics: A Comparison with UK Asthmatics, An et al)

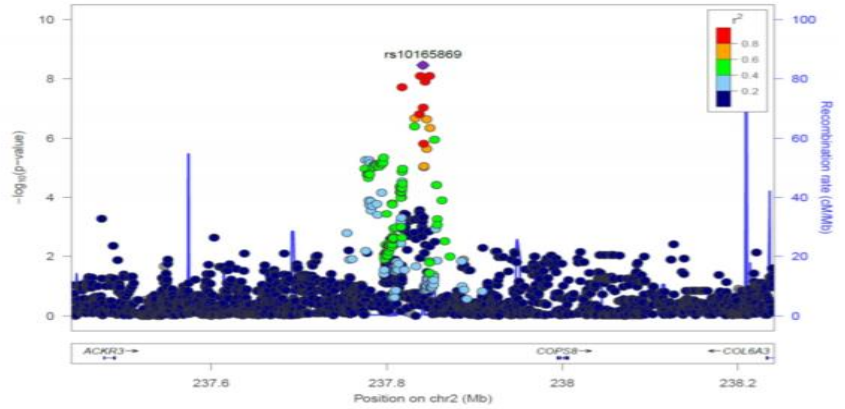




Genome-wide association study

- A Novel Locus for Exertional Dyspnea in Childhood Asthma, Lee et al (*Euro Respir J*)

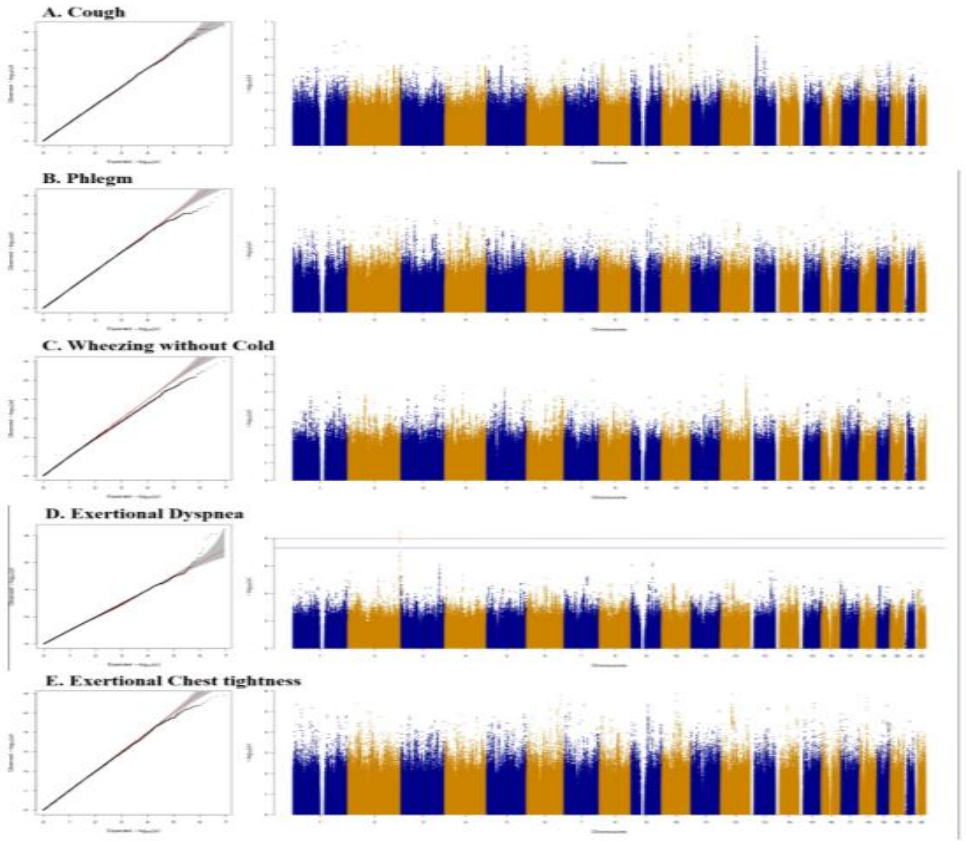


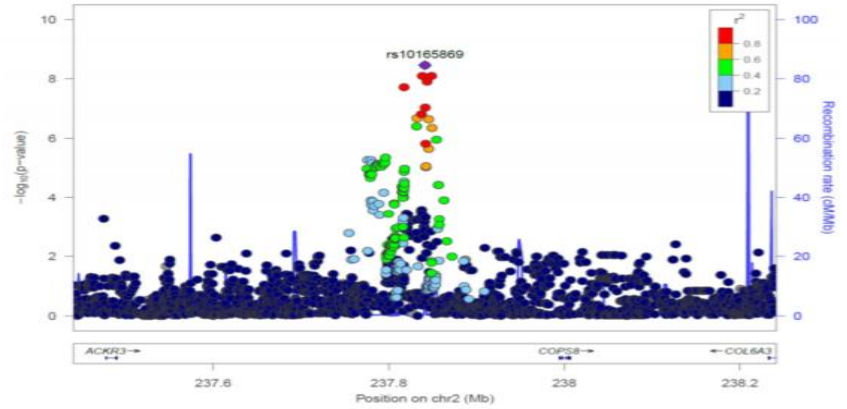


	Chr	Position	A1	A2	rsID	afreq	fam	S.E.S.	Var.S.	Z	p value
Costa Rican trios (N=894)	2	236932637	A	G	rs10165869	0.309	567	32.346	29.947	5.911	3.49E-09
	2	236940120	T	C	rs6725280	0.288	569	31.062	28.953	5.773	8.00E-09
	2	236929711	T	C	rs7607911	0.302	575	31.346	29.495	5.772	8.05E-09
	2	236935080	G	A	rs1865671	0.289	571	30.494	28.668	5.695	1.26E-08
	2	236908895	C	G	rs30102	0.304	591	30.523	29.528	5.617	1.94E-08
CAMP trios (N=286, dyspnea: 60.42%)	2	236932637	A	G	rs10165869	0.284	173	8.104	12.575	2.285	0.02229
	2	236940120	T	C	rs6725280	0.267	171	4.906	12.327	1.397	0.16229
	2	236929711	T	C	rs7607911	0.285	175	7.406	12.640	2.083	0.03724
	2	236935080	G	A	rs1865671	0.271	175	5.302	12.588	1.494	0.13507
	2	236908895	C	G	rs30102	0.275	171	6.313	12.549	1.782	0.07475

Genome-wide association study

- A Novel Locus for Exertional Dyspnea in Childhood Asthma, Lee et al (*Euro Respir J*)





	Chr	Position	A1	A2	rsID	afreq	fam	S.E.S.	Var.S.	Z	p value
Costa Rican trios (N=894)	2	236932637	A	G	rs10165869	0.309	567	32.346	29.947	5.911	3.49E-09
	2	236940120	T	C	rs6725280	0.288	569	31.062	28.953	5.773	8.00E-09
	2	236929711	T	C	rs7607911	0.302	575	31.346	29.495	5.772	8.05E-09
	2	236935080	G	A	rs1865671	0.289	571	30.494	28.668	5.695	1.26E-08
	2	236908895	C	G	rs30102	0.304	591	30.523	29.528	5.617	1.94E-08
CAMP trios (N=286, dyspnea: 60.42%)	2	236932637	A	G	rs10165869	0.284	173	8.104	12.575	2.285	0.02229
	2	236940120	T	C	rs6725280	0.267	171	4.906	12.327	1.397	0.16229
	2	236929711	T	C	rs7607911	0.285	175	7.406	12.640	2.083	0.03724
	2	236935080	G	A	rs1865671	0.271	175	5.302	12.588	1.494	0.13507
	2	236908895	C	G	rs30102	0.275	171	6.313	12.549	1.782	0.07475

Gene-environment-wide interaction studies (GEWIS)

- Occur when the risk of disease in exposed and susceptible individuals differs from that expected based on their individual effects
  - Expected effects can be additive or multiplicative

- Positive interaction
  - Synergistic

- Negative interaction
  - Antagonistic

GEWIS

- Statistical vs biological interaction
  - Statistical interaction: departure from additivity/multiplicity in a linear model on a selected scale of measurement
  - Biological interaction: the joint action of two or more factors, whether or not an additive statistical

- Quantitative vs qualitative interaction
  - Quantitative interaction: A form of statistical interaction in which the effects of one factor go in the same direction at different levels of the other, but differ in magnitude.
  - Qualitative interaction: Forms of statistical interaction in which: the effects go in opposite directions

Wang et al, *Hum Hered*, 2010

GEWIS

Data		Minor/ Major alleles	MAF	HWE	Main effects	Interaction (SNP – smoking status)			Interaction (SNP – pack years)	Overall effects
					β_{SNP} (P-value)	never vs former $\beta_{SNP\text{-}SM1}$ (P-value)	never vs current $\beta_{SNP\text{-}SM2}$ (P-value)	former vs current $\beta_{SNP\text{-}SM3}$ (P-value)	$\beta_{SNP\text{-}PY}$ (P-value)	
Discovery	KARE (Koreans)	C/T	0.384	0.604	**−0.025 (2×10^{-4})**	**−0.029 (0.043)**			0.0004 (0.185)	**2.70×10^{-7}**
Replication	GENIE (Koreans)	C/T	0.380	0.164	−0.004 (0.336*)	−0.018 (0.052*)	**−0.024 (0.049*)**		0.0003 (0.981*)	**0.0820**
	MESA-Lung (NHWs)	C/T	0.438	0.521	**−0.064 (0.008*)**	0.078 (0.941*)			**−0.0021 (0.014*)**	**0.0037**
	COPDGene (AAs)	C/T	0.177	0.377	0.042 (0.499)			−0.097 (0.082)	0.0005 (0.555)	0.2205
	COPDGene (NHWs)	C/T	0.459	0.433	**−0.066 (0.020)**			0.049 (0.054)	0.0006 (0.200)	0.0746

SCIENTIFIC REPORTS

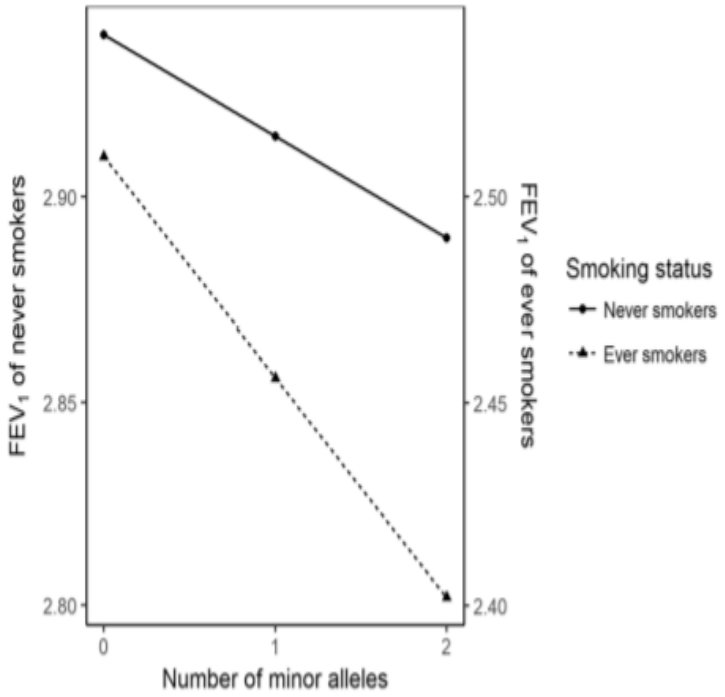
OPEN

Genome-wide assessment of gene-by-smoking interactions in COPD

Boram Park[1], So-My Koo[2,3], Jaehoon An[1], MoonGyu Lee[1], Hae Yeon Kang[4], Dandi Qiao[5], Michael H. Cho[5,6], Joohon Sung[1,7,8], Edwin K. Silverman[5,6], Hyeon-Jong Yang[3,9] & Sungho Won[1,7,8]

Received: 21 September 2017
Accepted: 30 May 2018
Published online: 18 June 2018

Cigarette smoke exposure is a major risk factor in chronic obstructive pulmonary disease (COPD) and its interactions with genetic variants could affect lung function. However, few gene-smoking interactions have been reported. In this report, we evaluated the effects of gene-smoking interactions on lung function using Korea Associated Resource (KARE) data with the spirometric variables—forced expiratory volume in 1 s (FEV_1). We found that variations in FEV_1 were different among smoking status. Thus, we considered a linear mixed model for association analysis under heteroscedasticity according to smoking status. We found a previously identified locus near *SOX9* on chromosome 17 to be the most significant based on a joint test of the main and interaction effects of smoking. Smoking interactions were replicated with Gene-Environment of Interaction and phenotype (GENIE), Multi-Ethnic Study of Atherosclerosis-Lung (MESA-Lung), and COPDGene studies. We found that individuals with minor alleles, rs17765644, rs17178251, rs11870732, and rs4793541, tended to have lower FEV_1 values, and lung function decreased much faster with age for smokers. There have been very few reports to replicate a common variant gene-smoking interaction, and our results revealed that statistical models for gene-smoking interaction analyses should be carefully selected.



Differentially Expressed Genes

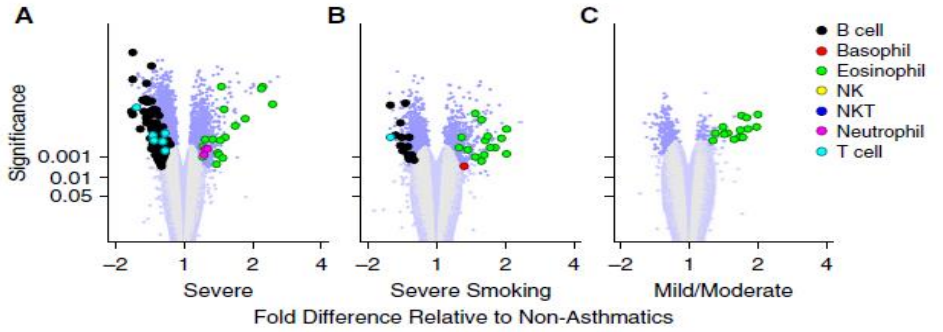
- Transciptomics
  - RNA levels are intermediate between DNA & proteins.
  - Large transcriptomic studies in the past decade has shown that while only ~3% of the genome encodes proteins, up to 80% of the genome is transcribed (Consortium EP, Nature 2012).
  - Thanks to RNA-seq technology, we found many isoforms of protein-coding transcriptome and non-coding RNAs such as long non-coding RNAs in mammalian cells (www.gencodegenes.org), short RNAs (miRNAs, piwi-interacting RNAs, snRNAs) and circular RNAs.
  - Associated technologies: probe-based arrays, RNA-seq
  - Statistical methods for identifying DEGs: DeSEQ2, Limma Voom, etc

Differentially Expressed Genes

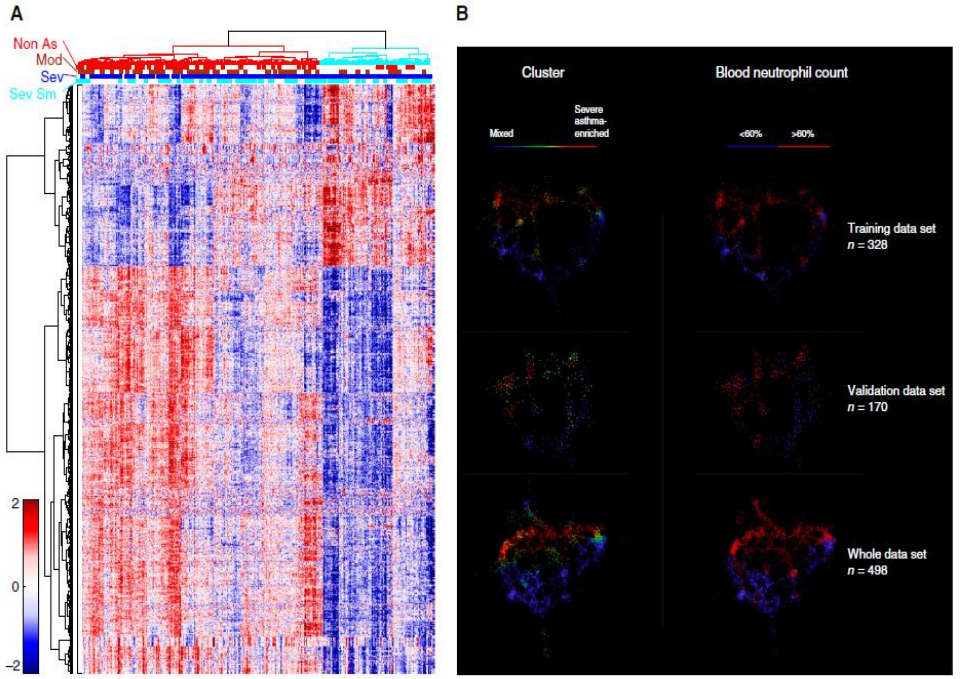
ORIGINAL ARTICLE

A Severe Asthma Disease Signature from Gene Expression Profiling of Peripheral Blood from U-BIOPRED Cohorts

Jeannette Bigler[1*‡], Michael Boedigheimer[2*], James P. R. Schofield[3], Paul J. Skipp[3], Julie Corfield[4,5], Anthony Rowe[6], Ana R. Sousa[7], Martin Timour[1], Lori Twehues[2], Xuguang Hu[8], Graham Roberts[9], Andrew A. Welcher[2], Wen Yu[1§], Diane Lefaudeux[10], Bertrand De Meulder[10], Charles Auffray[10], Kian F. Chung[11], Ian M. Adcock[11], Peter J. Sterk[12], and Ratko Djukanović[9]; on behalf of the U-BIOPRED Study Group‖



- Blood gene expression differences between clinically defined subgroups of patients with asthma and individuals without asthma, as well as subgroups of patients with severe asthma defined by transcript profiles, show the value of blood analysis in stratifying patients with asthma and identifying molecular pathways for further study.



Proteomics-based Analyses

- Proteomics
  - Proteomics is a research area that focusses on the large-scale study of proteins produced by cells, tissues and organisms.
  - Genomics and transcriptomics studies provide valuable contribution to asthma research, but a single gene or mRNA strand can generate several different protein isoforms as a result of alternative splicing of RNA and posttranslational modification of synthesized proteins.
  - In respiratory research, proteomics have previously been applied on different biological samples, such as serum, circulating cells, bronchoalveolar lavage fluid (BALF), nasal lavage fluid (NLF), induced sputum, exhaled breath condensate (EBC), epithelial lining fluid (ELF) and, albeit sporadically, biopsies
  - Associated technologies: immunoassays (Western blot, immunohistochemistry and ELISA) and mass spectrometry (MS). The former appeared earlier and is suitable for detection of small portion of proteins,125 whereas the latter aims to collect fractions of the whole proteome.

Proteomics-based Analyses

Analyses of asthma severity phenotypes and inflammatory proteins in subjects stratified by sputum granulocytes

Annette T. Hastie, PhD, Wendy C. Moore, MD, Deborah A. Meyers, PhD, Penny L. Vestal, MS, Huashi Li, MS, Stephen P. Peters, MD, PhD, Eugene R. Bleecker, MD, and the National Heart, Lung, and Blood Institute Severe Asthma Research Program *Winston-Salem, NC*

TABLE III. Sputum supernatant mediator levels stratified by sputum percentage eosionphils and percentage neutrophils

	<2% Eos +<40% Neu	<2% Eos + ≥40% Neu	≥2% Eos + <40% Neu	≥2% Eos + ≥40% Neu	*P* value for 4 groups*
BDNF pg/mL	9.5 (6-14.5)	18.4 (11.2-29)	15.4 (8-26.5)	20 (14-39)	**<.001**
BLC/CXCL13 pg/mL	110 ± 29	223 ± 48	94 ± 20	142 ± 49	.005
BMP-4 pg/mL	4.9 ± 1.5	2.8 ± 0.6	1.8 ± 0.3	3.7 ± 1.1	.33
EGF pg/mL	159 (108-205)	190 (116-263)	171 (90-212)	199 (145-256)	.13
Eotaxin 2/CCL24 pg/mL	0.88 (0.01-2.38)	1.13 (0.01-6.78)	2.39 (0.38-4.63)	8.90 (2.5-14.1)	.022
IFN-γ pg/mL	77 (37-149)	47 (2-128)	108 (49-175)	91 (23-125)	.11
IL-1β pg/mL	104 ± 14	224 ± 56	76 ± 12	228 ± 65	**<.001**
IL-8 ng/mL	1.5 ± 0.1	2.1 ± 0.2	1.6 ± 0.2	1.9 ± 0.2	.017
IL-13 pg/mL	91 ± 18	74 ± 9.9	64 ± 12	81 ± 18	.55
LIGHT/TNFSF14 pg/mL	36 (12-69)	92 (36-156)	33 (13-108)	58 (24-184)	.021
MIP-3α/CCL20 pg/mL	390 ± 41	781 ± 85	341 ± 52	668 ± 121	**<.001**
PARC/CCL18 pg/mL	6.7 (0.4-19)	12.7 (4.7-59)	8.3 (1.7-22)	11 (4.6-40)	.037
TNF-α pg/mL	0.3 (0.01-1.24)	1.1 (0.01-2.6)	0.39 (0.01-1.9)	0.7 (0.01-8.1)	.44

P values meeting Bonferroni correction are in boldface.
BMP, Bone morphogenic protein; *Eos*, eosinophils; *Neu*, neutrophils.
*ANOVA (mean ± SEM) or Kruskal-Wallis (median [25% to 75% quartiles]).

Metabolomics-based Analyses

- Metabolomics
  - The metabolic state is the result of both gene expression and environmental factors and can therefore be informative for disorders of multifactorial nature, such as asthma.
  - It simultaneously quantifies multiple small molecule types, such as amino acids, fatty acids, carbohydrates, or other products of cellular metabolic functions.
  - Metabolite levels and relative ratios reflect metabolic function, and out of normal range perturbations are often indicative of disease.
  - As inflammatory mediators usually have a short half-life, they are rapidly degraded to various metabolites. Potentially, the analysis of these metabolites allows determination of previous cellular responses involved in inflammatory processes.
  - Most metabolomic studies to date have focused on distinguishing asthma patients from healthy controls, and asthma severity. The main metabolites found in these studies are involved in tricarboxylic acid metabolism, hypermethylation, phospholipid regulation, hypoxia, oxidative stress and immune reactions (Reinke et al, Eur Respir J).
  - Associated technologies: MS-based methods

Metabolomics-based Analyses

Metabolomics analysis identifies different metabotypes of asthma severity

Euro Respir J, 2017

Stacey N. Reinke[1], Héctor Gallart-Ayala[1], Cristina Gómez[1,2], Antonio Checa[1,2], Alexander Fauland[1,2], Shama Naz[1], Muhammad Anas Kamleh[1], Ratko Djukanović[3,4], Timothy S.C. Hinks[3,4,5] and Craig E. Wheelock[1]

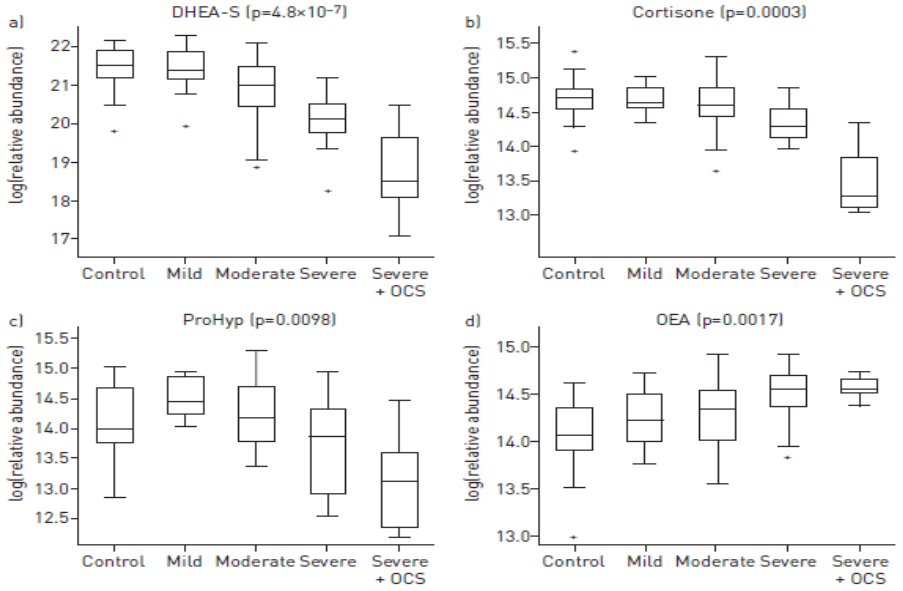


FIGURE 1 Association between oral corticosteroid (OCS) treatment and the relative abundance of selected metabolites. Severe asthma was classified according to treatment. Severe: treatment with inhaled corticosteroids only; Severe+OCS: treated also with oral corticosteroids. The line in the middle of each box equals the median value, the tops and bottoms of each box are the first and third quartile, respectively. The whiskers span from 1.5 times the interquartile range (IQR) above the third quartile to 1.5 times the IQR below the first quartile. Samples outside this range (crosses) are considered outliers. Kruskal–Wallis p-values are shown. DHEA-S: dehydroepiandrosterone sulfate; ProHyp: prolylhydroxyproline; OEA: oleoylethanolamide.

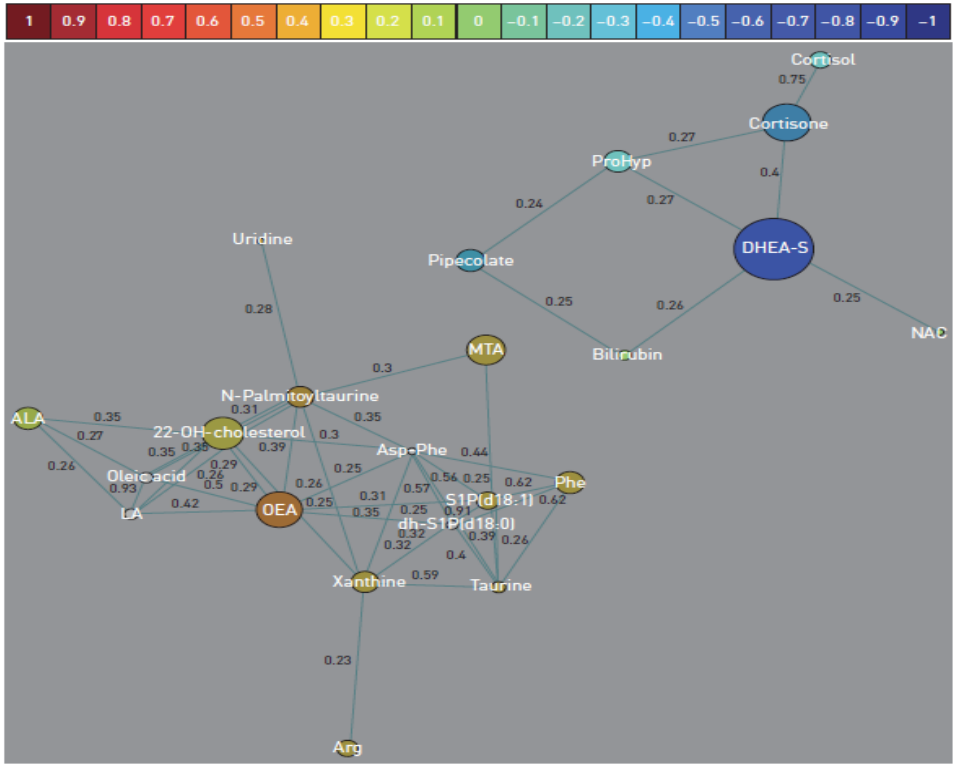


FIGURE 2 Spring-embedded plot illustrating the relationship between selected metabolites and disease severity. The size of the node is proportional to the significance of the relationship with disease severity (the larger the circle, the more significant the metabolite). The length of the line between the nodes (spring length) is proportional to the correlation strength (the shorter the length, the stronger the correlation with neighbouring metabolites). Node colour directly maps onto the correlation coefficient between the relative abundance of the metabolite and disease severity (see the colour bar above the figure: the intensity of the colours red and blue denote positive and negative correlations with disease severity, respectively) for the significant metabolites ($p<0.05$). Nodes were coloured grey if their corresponding p-value was 0.05–0.10. ALA: α-linolenic acid; Arg: arginine; Asp-Phe: aspartylphenylalanine; dh-S1P(d18:0): dihydrosphingosine-1-phosphate(d18:0); DHEA-S: dehydroepiandrosterone sulfate; LA, linoleic acid; MTA: methylthioadenosine; NAC: N-acetylcarnosine; OEA, oleoylethanolamide; Phe, phenylalanine; ProHyp, prolylhydroxyproline; S1P (d18:1), sphingosine-1-phosphate(d18:1).

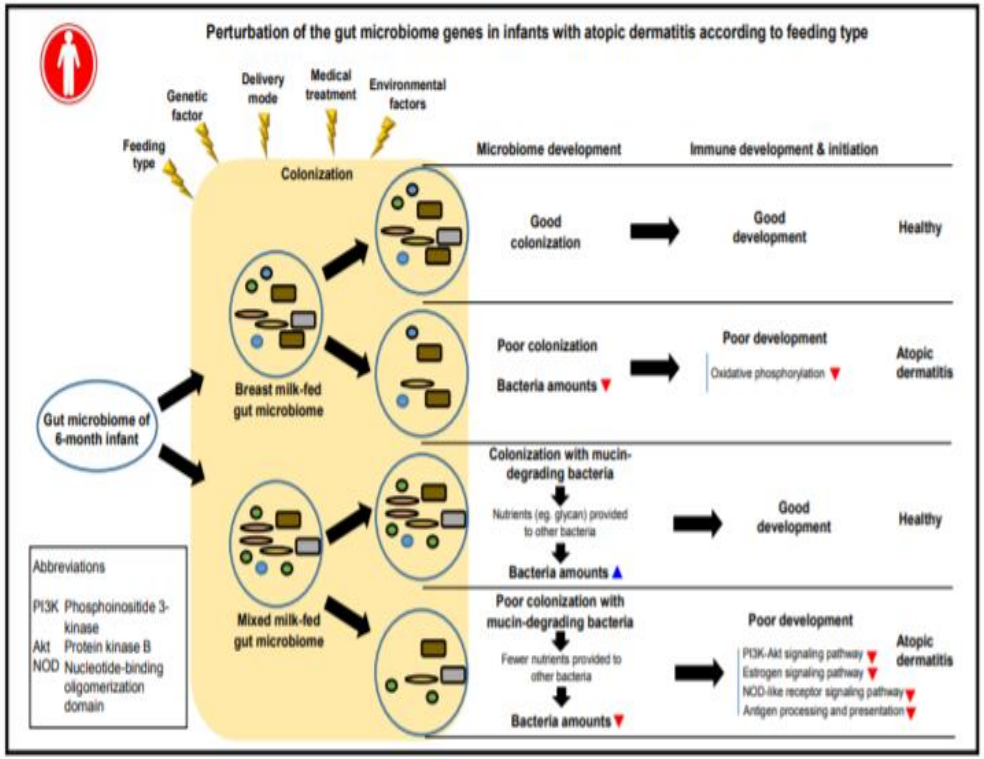
Microbiomics-based Analysis

- Microbiomics
  - Microbiome can be considered as exposure factor on epigenomic level as it is now believed to drastically influence a host organisms' health.
  - Huang et al demonstrated significant association between bacteria abundance and their diversity in airways with bronchial hyperresponsiveness (Huang et al, JACI, 2015).
  - Obesity sometimes co-occurs with asthma, and is connected to a different gut microbiome composition, which might lead to a systemic changes and aggravation of asthma symptoms (Castaner et al, Int J Endocrinol, 2018; Lvanova et al, Allergy, 2019)
  - Profiling: Amplification, 16s rRNA hypervariable region sequencing, Operational taxonomic unit clustering, Shotgun metagenomics sequencing (total DNA)

Microbiomics-based Analysis

Perturbations of gut microbiome genes in infants with atopic dermatitis according to feeding type

Min-Jung Lee, MS • Mi-Jin Kang, PhD • So-Yeon Lee, MD • ... Kangmo Ahn, MD • Bong-Soo Kim, PhD • Soo-Jong Hong, MD • Show all authors



J Allergy Clin Immunol, 2018

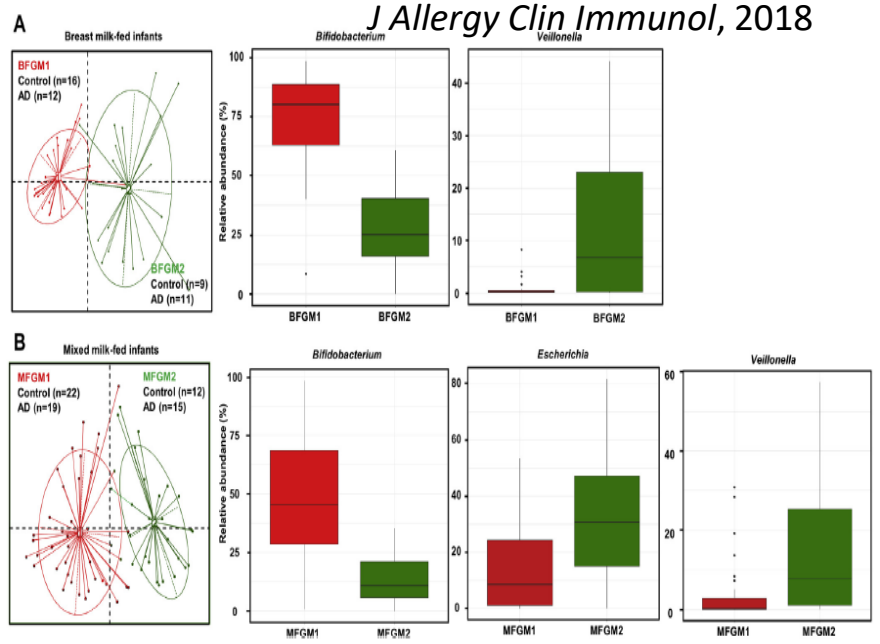


FIG 2. Enterotype clustering of gut microbiota from 6-month-old infants. **A**, Two enterotypes were detected in the gut microbiota of breast milk-fed infants. *Bifidobacterium* was the dominant genus in BFGM1, whereas *Veillonella* was dominant in BFGM2. **B**, Two enterotypes were detected in the gut microbiota of mixed milk-fed infants. *Bifidobacterium* was the dominant genus in MFGM1, whereas *Escherichia* and *Veillonella* were dominant in MFGM2.

Multi-Omics Analysis

Multi-Omics Integration

- Multi-Omics data
  - Single omics does not allow us to draw phenotypic traits accurately, in addition to the batch effects, large variability and lack of standardization are observed in omics studies.
  - A combination of information from several layers holds potential to compensate for this.

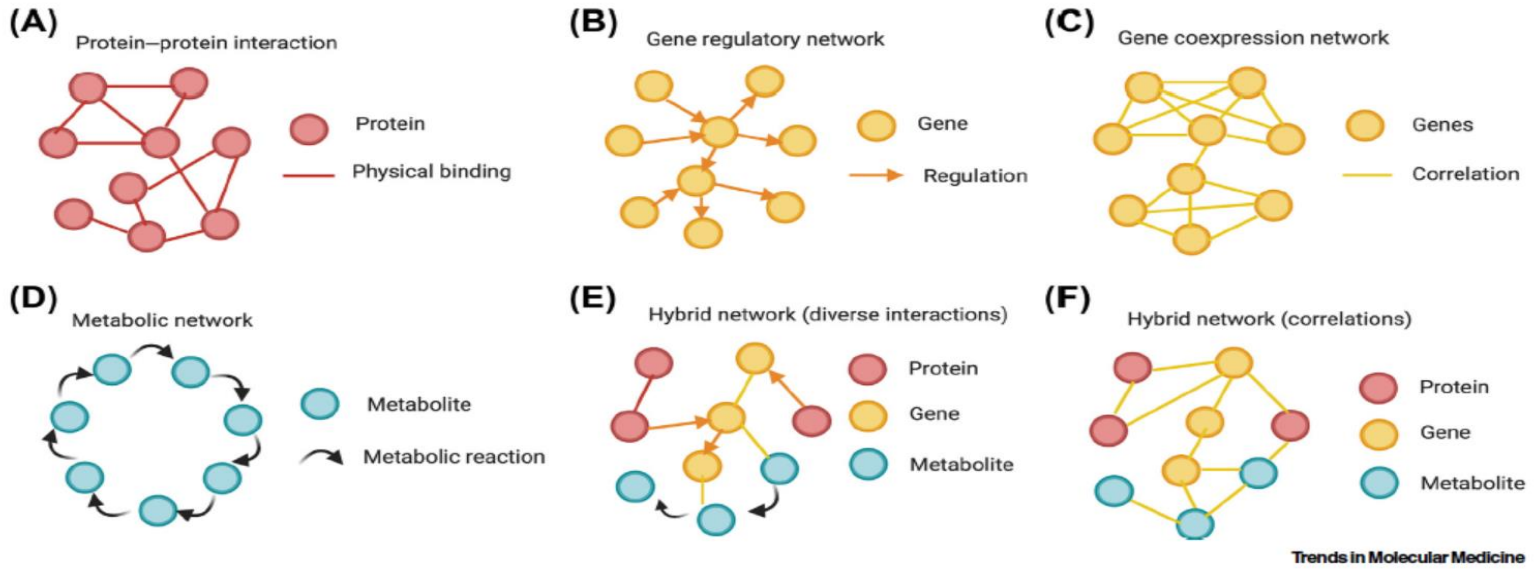


Figure 2. Main Types of Molecular Networks. (A) Protein–protein interaction networks. (B) Gene regulatory networks. (C) Gene coexpression networks. (D) Metabolic networks. (E) Hybrid networks based on various interaction types. (F) Hybrid networks based on correlations.

Yang, Trends in Molecular Medicine, 2020

Multi-Omics Integration

- Possible Research Goals for multi-Omics analyses
  - Biomarker discovery: studies aimed at detecting –omics characteristics indicating a disease state.
  - Subgroup identification: studies aimed at finding groups of patients that exhibit different therapeutic/prognostic outcomes.
  - Pathway analyses: studies aimed at discovering relation among -omics terms, such as genes or proteins in normal and asthma condition.
  - Drug repurposing/discovery: studies aimed at identifying new drugs to or existing effective drugs originally developed for other conditions.

Nicora et al, Frontiers in Oncology, 2020

Multi-Omics Analysis: Subgroup Identification

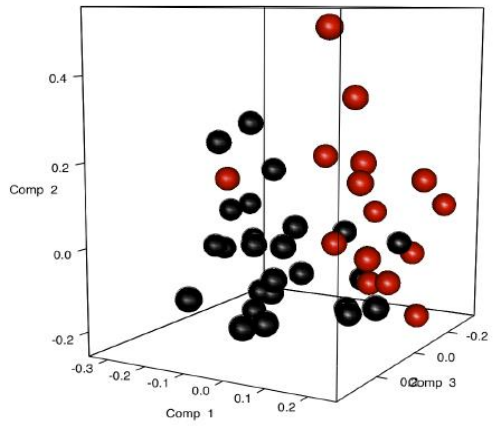
- Statistical Methods
  - Partial Least Squares (PLS) : It maximizes the covariance between each linear combination (components) associated to each omics data.
  - Sparse PLS can be used to conduct variable selection from two omics data sets.

Lê Cao et al., BMC Bioinformatics, 2009

Multi-Omics Analysis: Subgroup Identification

- Illustration of sparse PLS: samples' relationship
  - sPLS aims at selecting correlated variables (genes, proteins) across the same samples by performing a multivariate regression.
  - Regression mode: for instance, explain the protein abundance w.r.t the gene expression.
  - The latent variables (components) are determined based on the selected genes and proteins.

    => Give more insight into the samples similarities.
  - Unsupervised approach

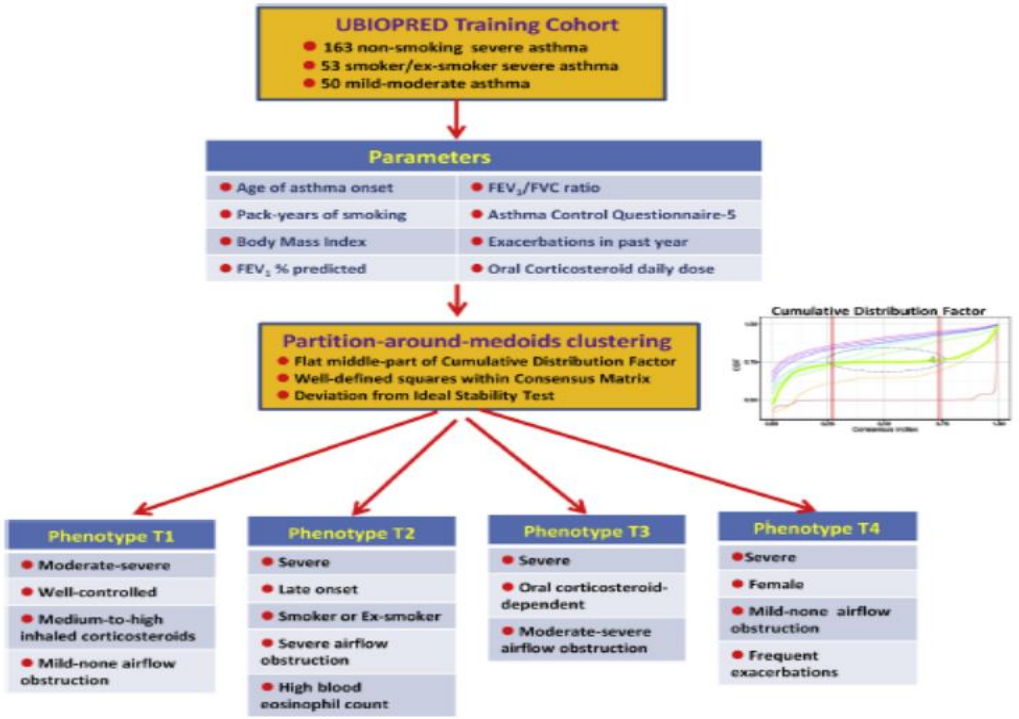


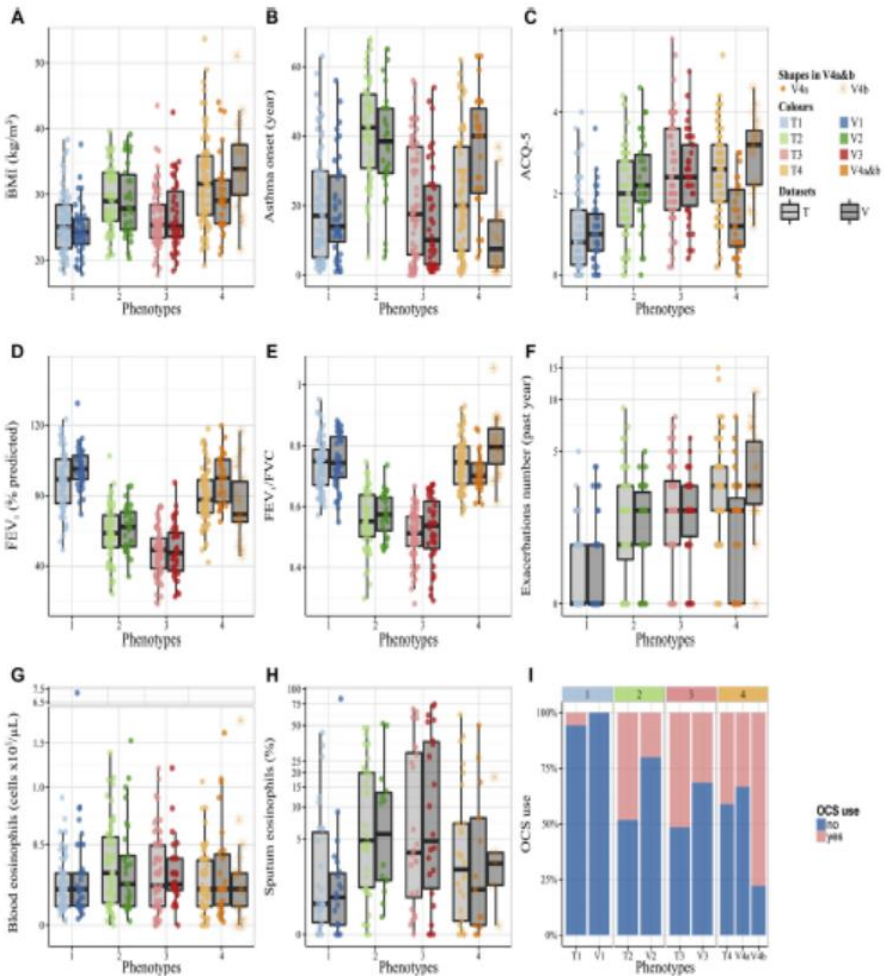
<Sample Plot>

Multi-Omics Analysis: Subgroup Identification

U-BIOPRED clinical adult asthma clusters linked to a subset of sputum omics

Diane Lefaudeux, MSc * • Bertrand De Meulder, PhD * • Matthew J. Loza, PhD • ... Charles Auffray, PhD •
Kian Fan Chung, MD and the U-BIOPRED Study Group ‡ • Show all authors • Show footnotes



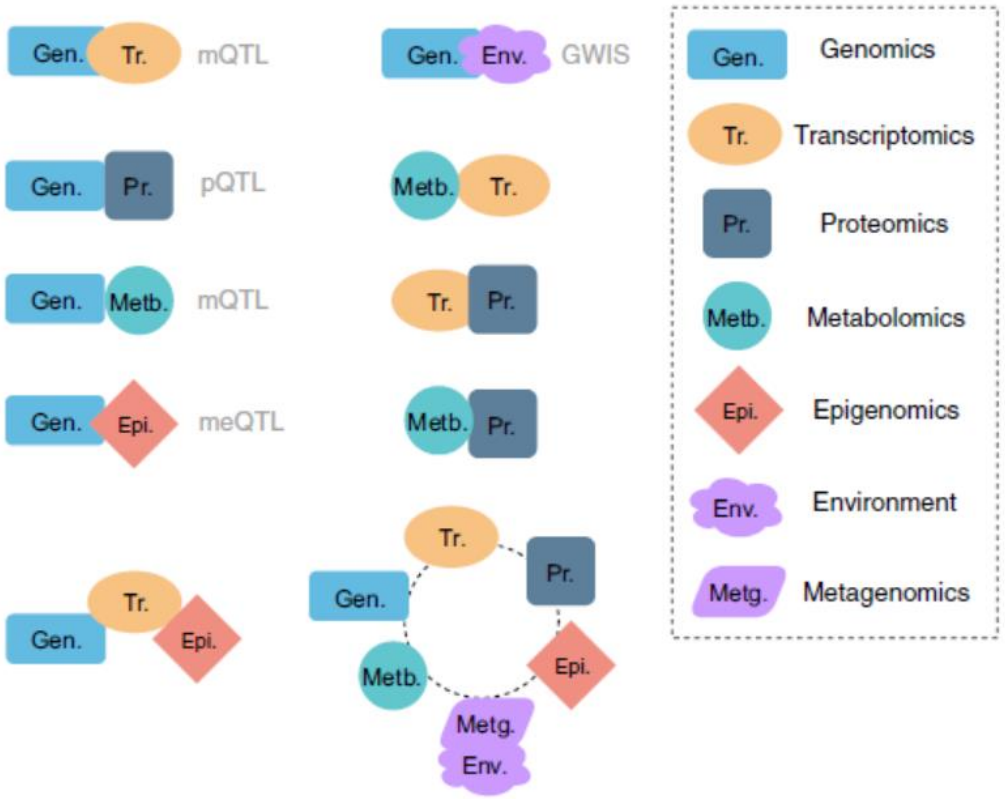


Multi-Omics Analysis: Subgroup Identification

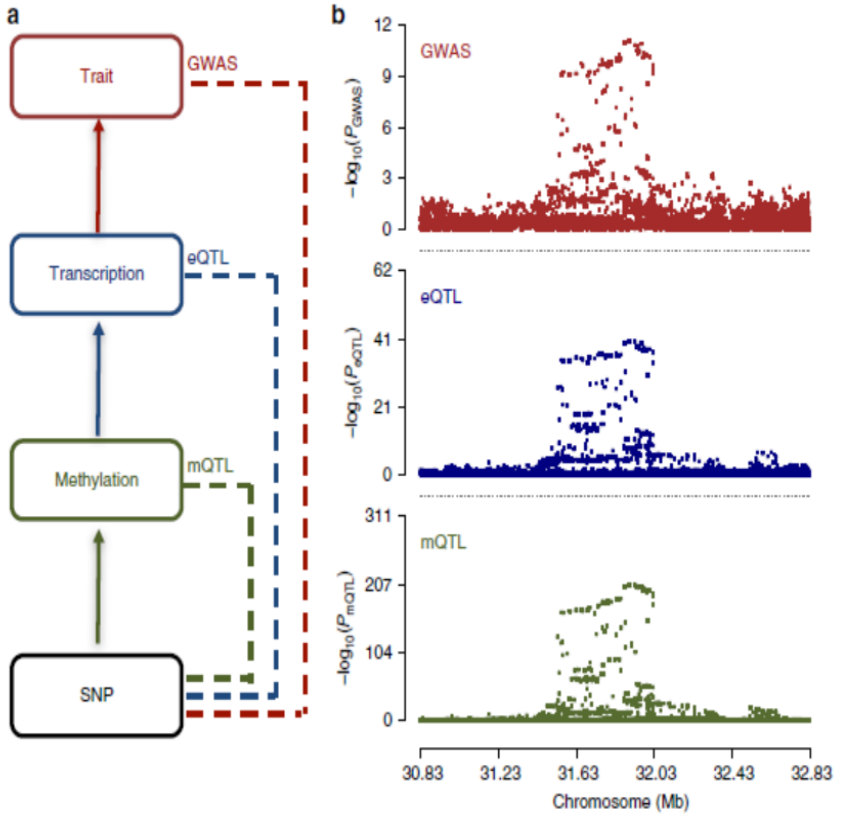
- Limitation
  - One of the main issues in omics analyses is “dimension hell”: the number of hypotheses (SNPs, genes, proteins, etc) being tested is large, in comparison with the number of patients.
  - To find statistically significant results, application of dimensionality reduction methods is required in a combination with large amounts of patients.
  - Otherwise it is almost impossible to find the pairs of multi-omics markers which jointly affect Asthma.

Multi-Omics Analysis: Biomarker Discovery

- Integrative analyses with multi-omics analysis



Possible integrations of multi-omics data



Molecular quantitative trait loci

Ivanova et al, *Allergy, 2019*

Multi-Omics Analysis: Biomarker Discovery

- 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

Multi-Omics Analysis: Biomarker Discovery

- 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

Multi-Omics Analysis: Biomarker Discovery

- 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

Multi-Omics Analysis: Biomarker Discovery

- Mendelian randomization
  - 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.

Ye et al, Trends Genet, 2020

Multi-Omics Analysis: Biomarker Discovery

- Mendelian randomization

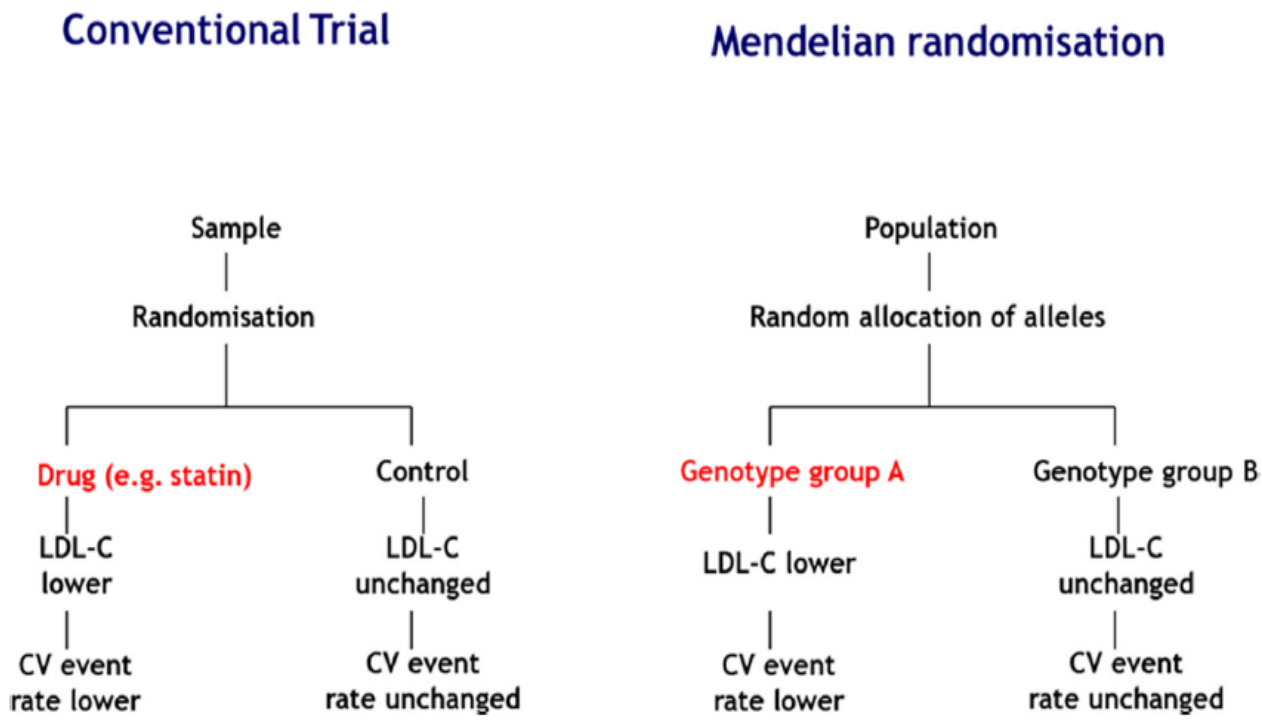


Figure 3 Comparison of a conventional trial with a Mendelian randomisation study. This illustrates the analogy between a conventional randomised controlled trial and a Mendelian randomisation study. CV, cardiovascular.

Bennet et al, Heart, 2017

Multi-Omics Analysis: Biomarker Discovery

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

Transcriptome

Proteome

Microbiome, etc

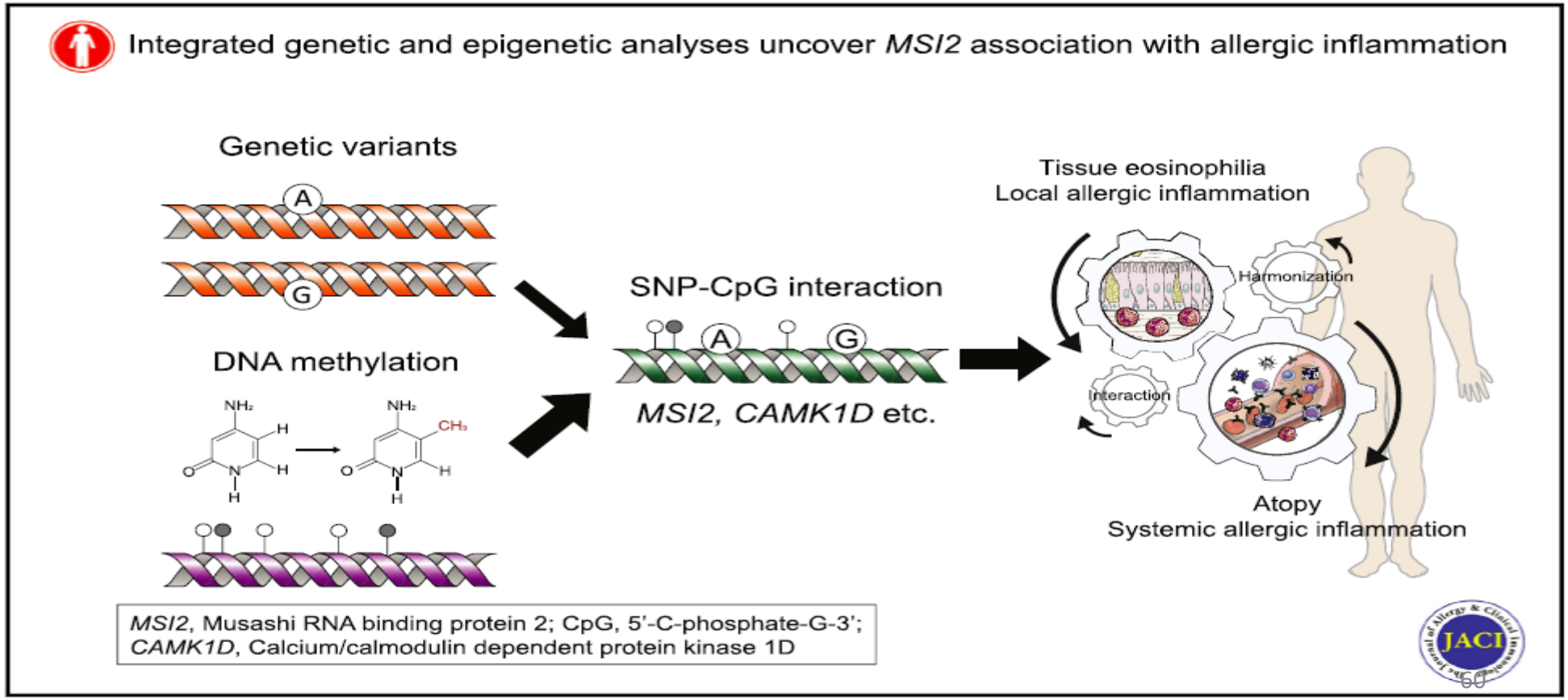
Multi-Omics Analysis: ATOPY

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

Kyung Won Kim, MD, PhD,[a]* Sang-Cheol Park, PhD,[b]* Hyung-Ju Cho, MD, PhD,[c]* Haerin Jang, BS,[a] Jaehyun Park, BS,[d] Hyo Sup Shim, MD, PhD,[e] Eun Gyul Kim, MS,[a] Mi Na Kim, PhD,[a] Jung Yeon Hong, PhD,[a] Yoon Hee Kim, MD, PhD,[f] Sanghun Lee, PhD,[g] Scott T. Weiss, MD, PhD,[h] Chang-Hoon Kim, MD, PhD,[c] Sungho Won, PhD,[b,d,i] and Myung Hyun Sohn, MD, PhD[a]

Seoul and Yongin, Korea, and Boston, Mass

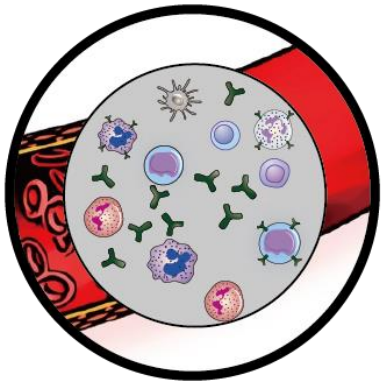
GRAPHICAL ABSTRACT



Allergic Inflammation

- The inflammation produced in sensitized subjects after exposure to a specific allergen
- Allergic inflammation can develop from persistent activation of type 2 immunity
- A hallmark of pathology of type 2 inflammatory disorders, such as asthma, atopic dermatitis, rhinosinusitis, and food allergy

Two Types of Allergic Reactions



- Atopy (allergic sensitization)
  - Allergic hypersensitivity reactions
  - Tendency to produce an exaggerated IgE immune response
  - Sneezing and Rhinorrhea
  - Predisposition to develop allergic diseases such as asthma, allergic rhinitis, and eczema
  - "Systemic allergic inflammation"

- Tissue eosinophilia
  - Increased number of eosinophils in the tissues
  - The level of eosinophils in bloodstream is likely normal
  - "Local allergic inflammation"

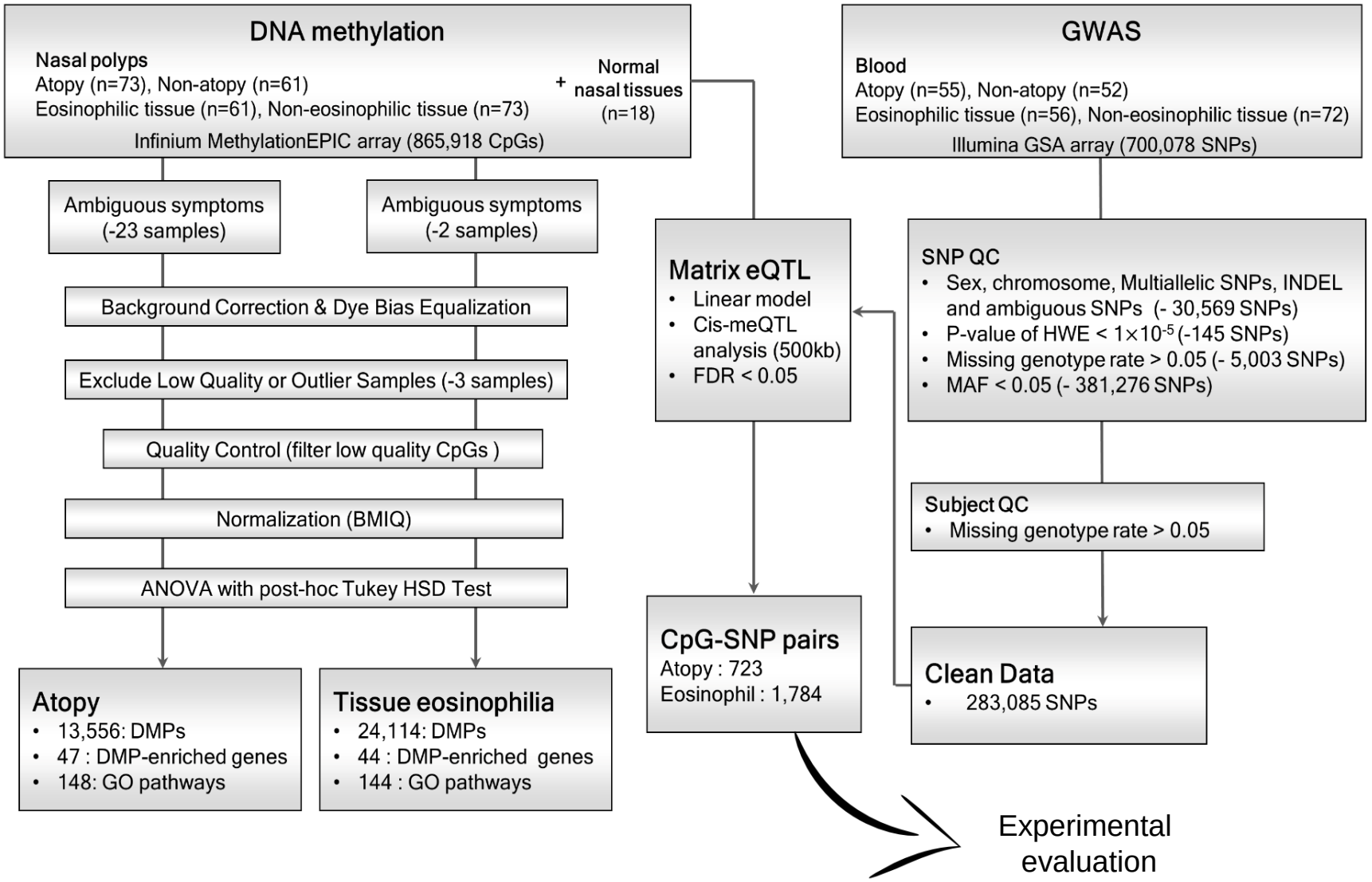
Sample Compositions

- Atopy was defined by
  - total serum IgE level of greater than 150 kU/L
  - Or specific IgE levels of greater than 0.7 kUA/L to at least 1 of the 6 common allergens
    - egg white, milk, *Dermatophagoides pteronyssinus, Dermatophagoides farinae, Alternaria species, or Blattella germanica.*
- Tissue eosinophilia was defined as
  - ≥ 70 eosinophils/HPF.1,2

	Atopy analysis (n=108)		Tissue eosinophilia analysis (n=129)		
	Atopy (n=56)	Non-atopy (n=52)	Eosinophilic tissue (n=57)	Non-eosinophilic Tissue (n=72)	Normal nasal tissue (n=18)
Male	44 (78.6)	31 (59.6)	45 (78.9)	47 (65.2)	6 (33.3)
Age, years	44.7 ± 13.1	44.9 ± 14.8	46.8 ± 12.6	44.1 ± 14.4	50.1 ± 17.9
Total serum IgE (kU/L)	446.2 ± 697.6	24.1 ± 18.0	196.3 ± 288.2	225.6 ± 620.5	32.0 ± 36.7

Data generated from nasal polyps +

Study Scheme

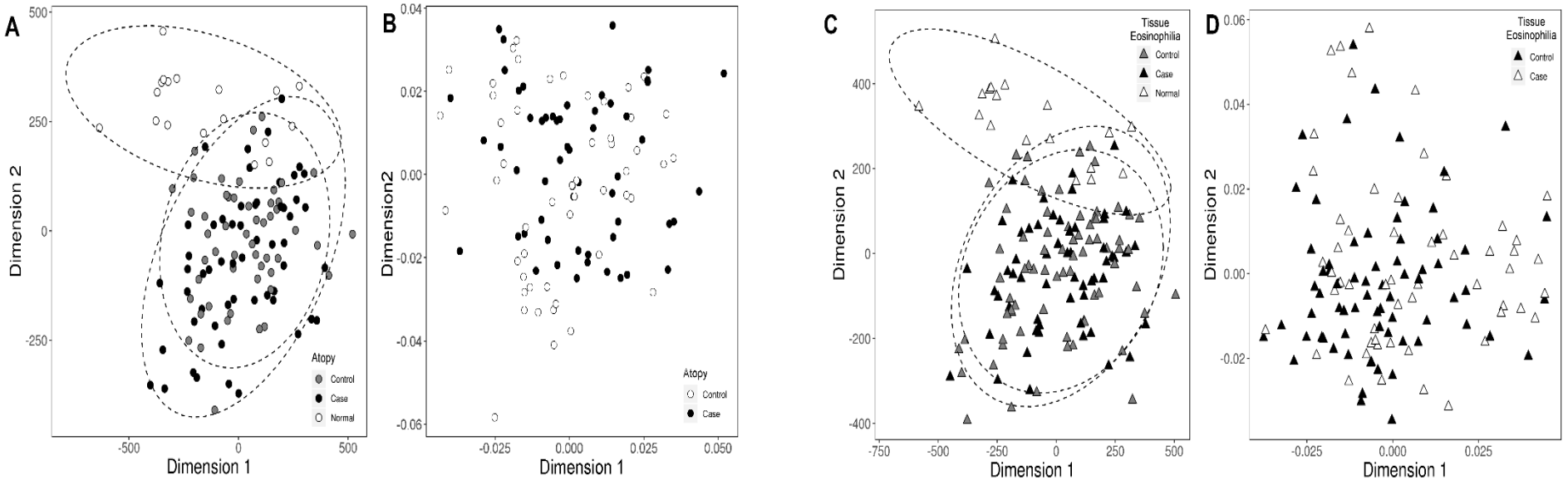


Overall Distribution by Group

- Clear cluster separation between polyps and normal tissues in methylation data
- But, no clear distinction between case and control group
  - Only subtle differences exist.

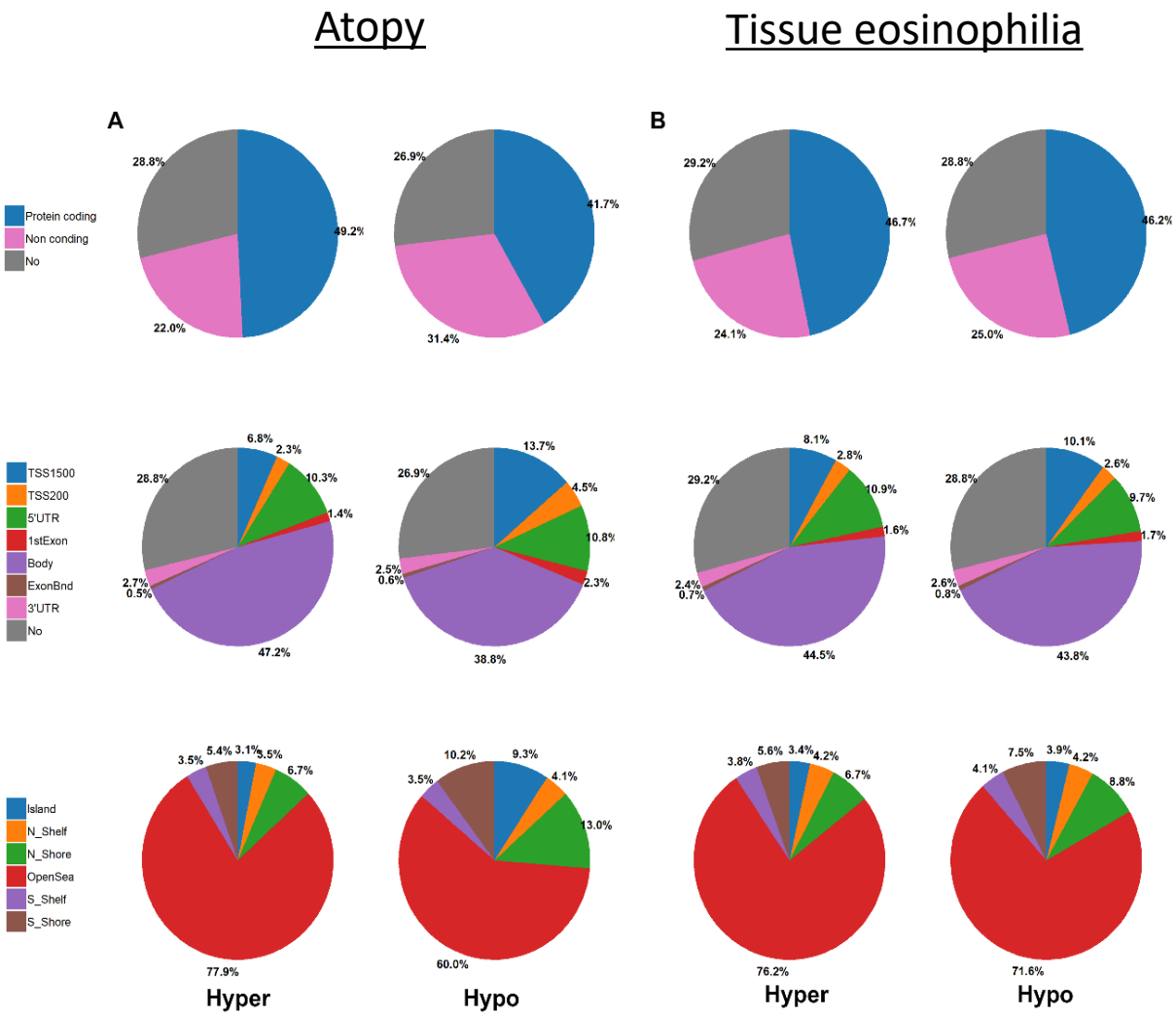
Atopy

Tissue eosinophilia



Multidimensional scaling (MDS)

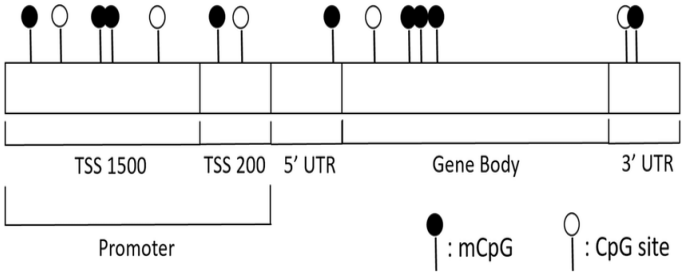
Summary of DMP Locus

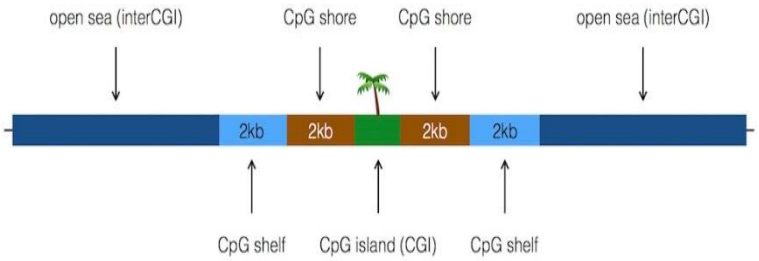


Atopy
- 13,556: DMPs
- 47 : DMP-enriched genes
- 148: GO pathways

Tissue eosinophilia
- 24,114: DMPs
- 44 : DMP-enriched genes
- 144 : GO pathways

DMP : differential methylation position

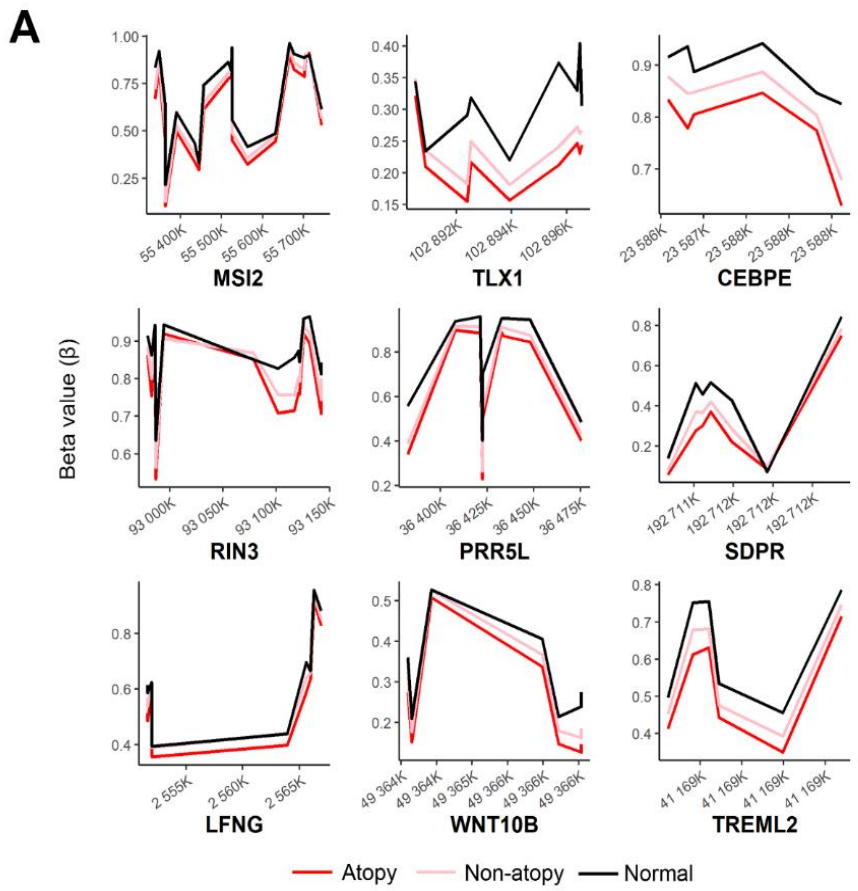


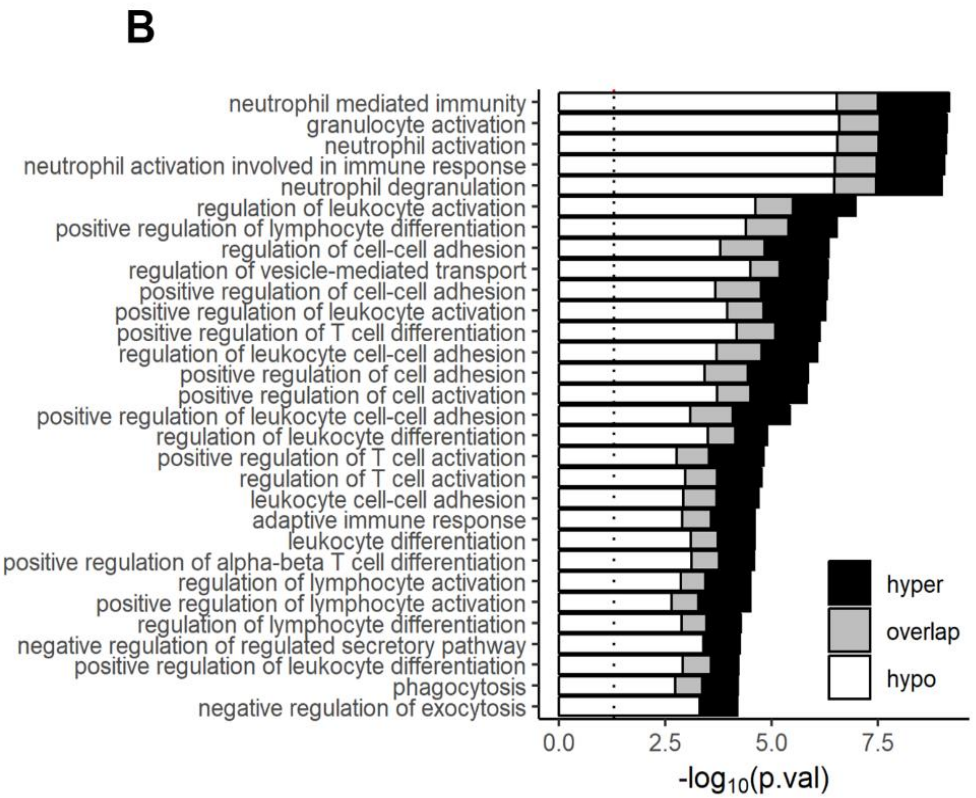


- A higher DMP concentration in gene bodies and open sea from either atopy or tissue eosinophilia analyses

Atopy-related Methylation Features

- Strongly associated with hypo-methylated CpGs

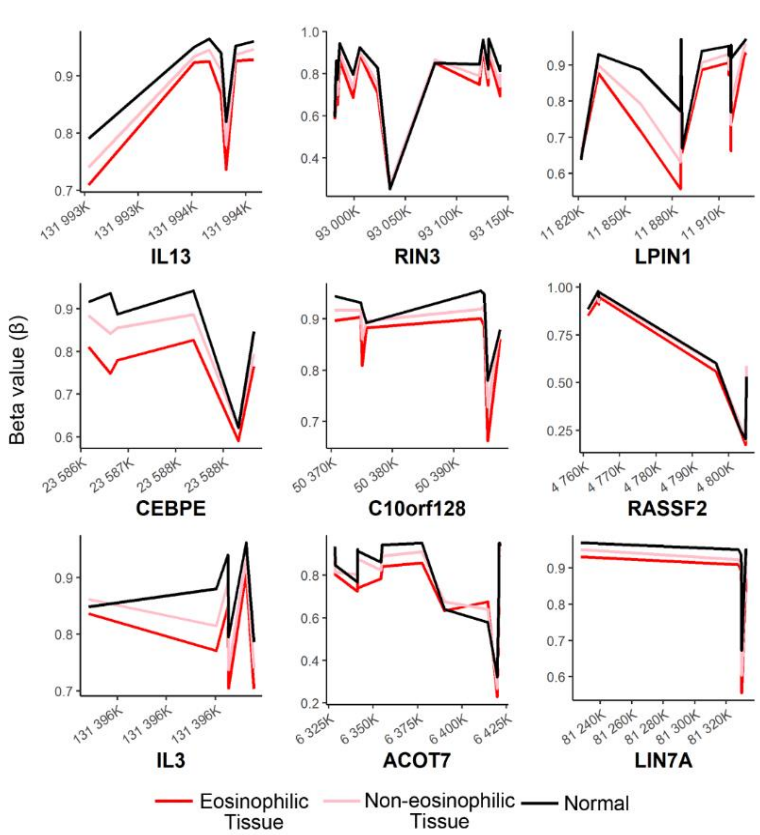


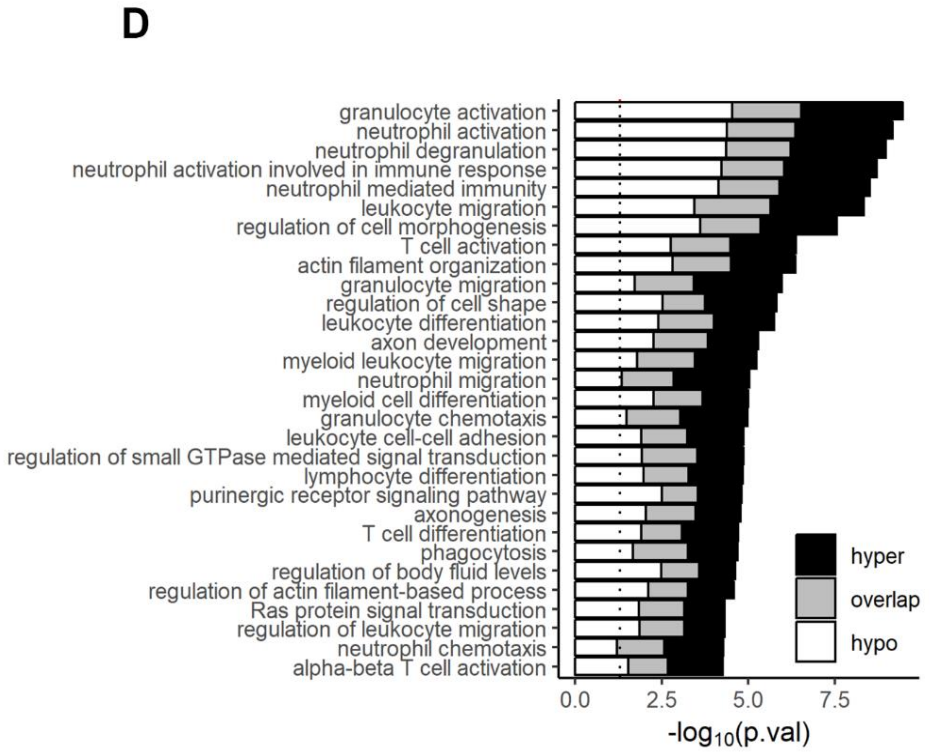


CpG : cytosines followed by guanine residues

Tissue Eosinophilia-related Features

- Strongly associated with hypo-methylated CpGs
- Similar to the result of atopy, but the different set of immune functions was more associated

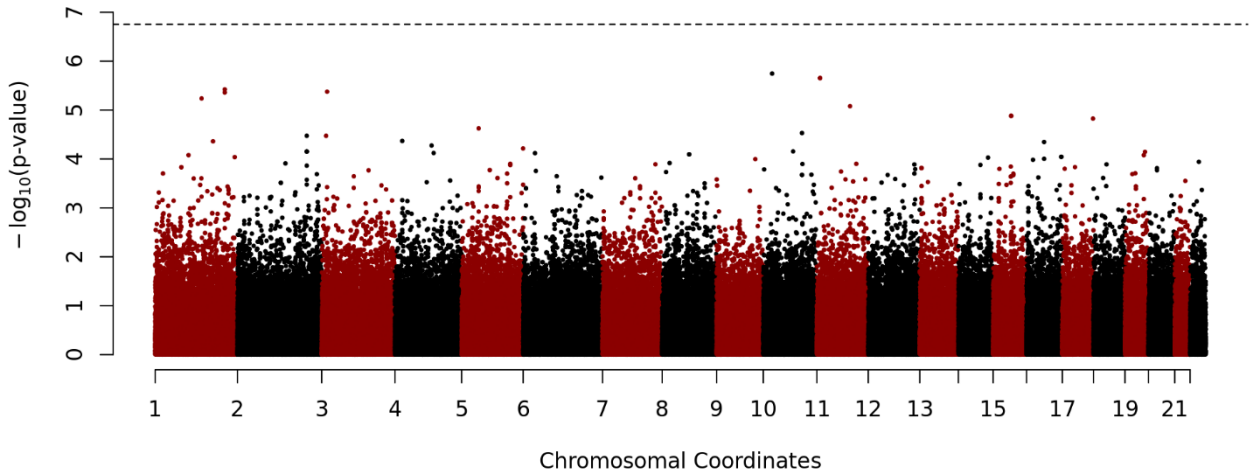


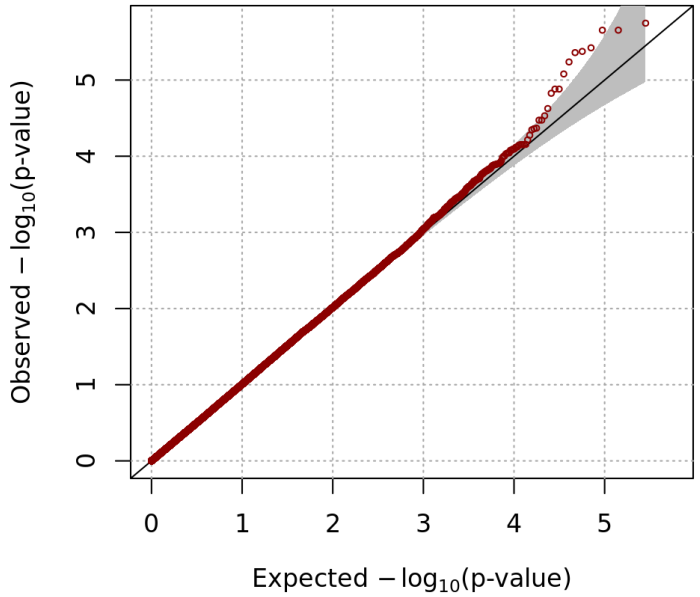


DMP : differential methylation position

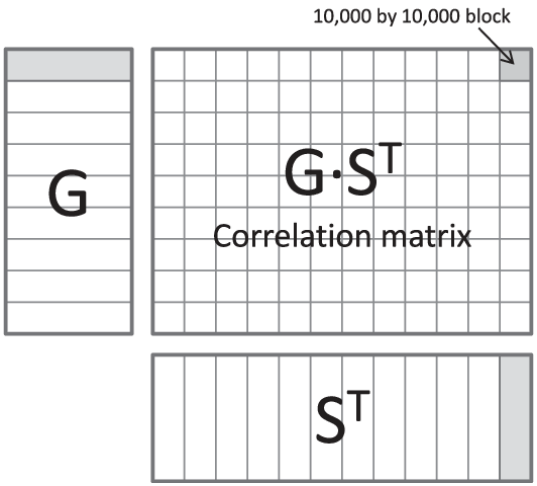
Limitation of Sample Size

- Not easy to find a statistically significant SNPs in general GWAS analysis with only 130 samples.





Interaction of Genetic & Epigenetic Feature



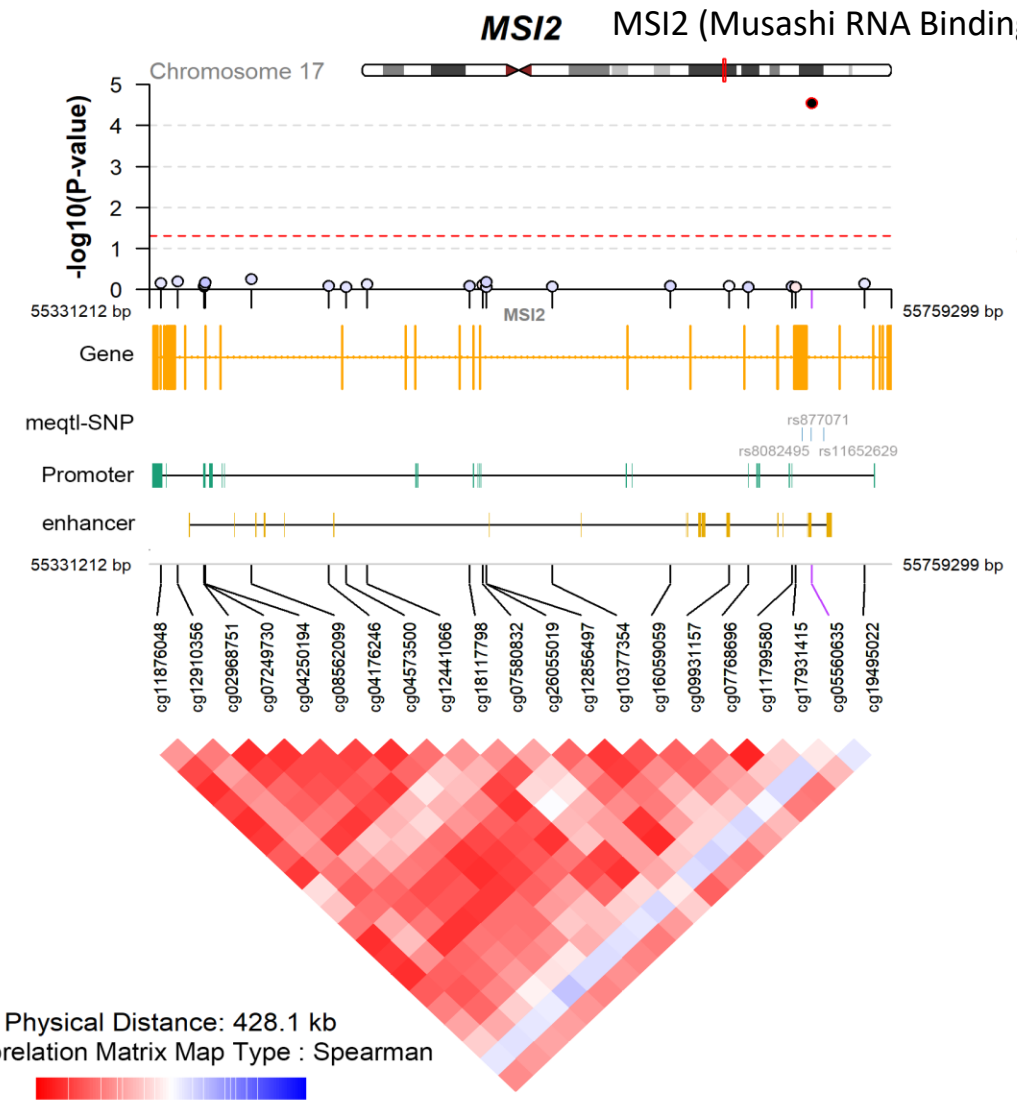
- Matrix eQTL
  - Technique to link epigenetics with genetic risk
  - Equation :

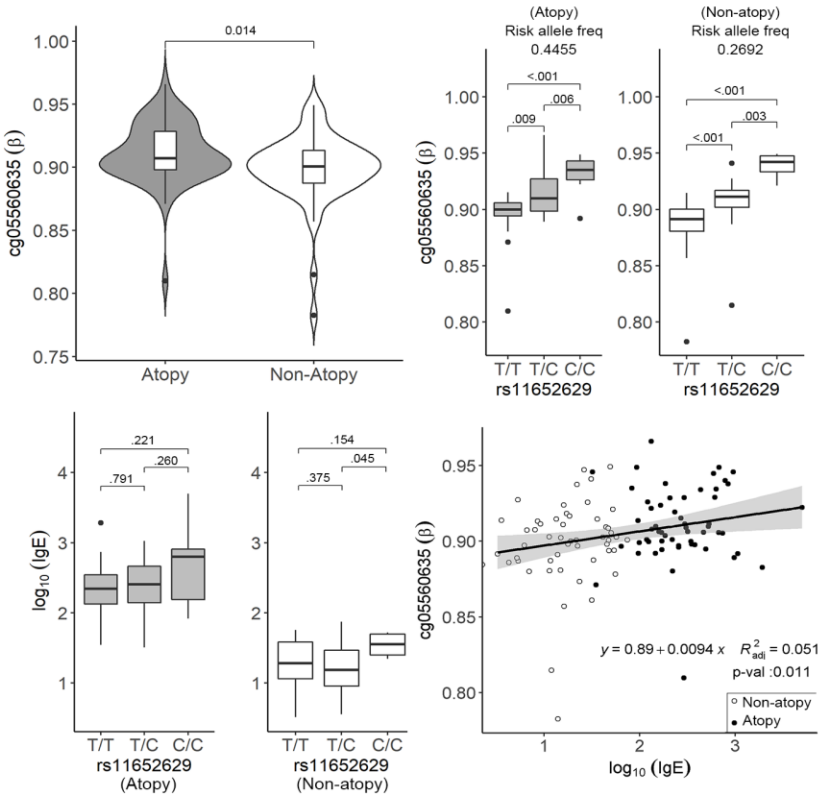
    $$\text{Methylation} = \alpha + \sum_k \beta_k \cdot \text{covariate}_k + \gamma \cdot \text{genotype\_additive}$$
  - Testing for significance of: γ
  - Test statistic: t-statistic
  - Covariate : Age, Sex
  - Only tested local pair within 500 kb

- 732 (atopy), 1314 (eosinophilia) SNP-CPG interaction pairs were founded
- Among these, genes with observed changes in DNA methylation mediated through genetic effects were selected
  - 4 genes (MSI2, JARID2, TRIO, and ACOT7)
  - 10 genes (RNF19A, CAMK1D, SLC45A4, ACOT7, PRDM16, NOTCH1, GJB2, ATAD3C, JAZF1, and TG)

eQTL : Expression quantitative trait loci

Noble Candidate Marker for Atopy





Significant SNP-CPG interaction pairs
(rs11652629 - cg05560635)

Replication in Costa Rican Cohort

- Whole genome sequencing data from 3770 Costa Rican subjects
- For three SNPs, atopy-related phenotypes were analyzed by GEMMA
  - serum total IgE
  - dust mite-specific IgE
  - peripheral blood eosinophil count

Total IgE

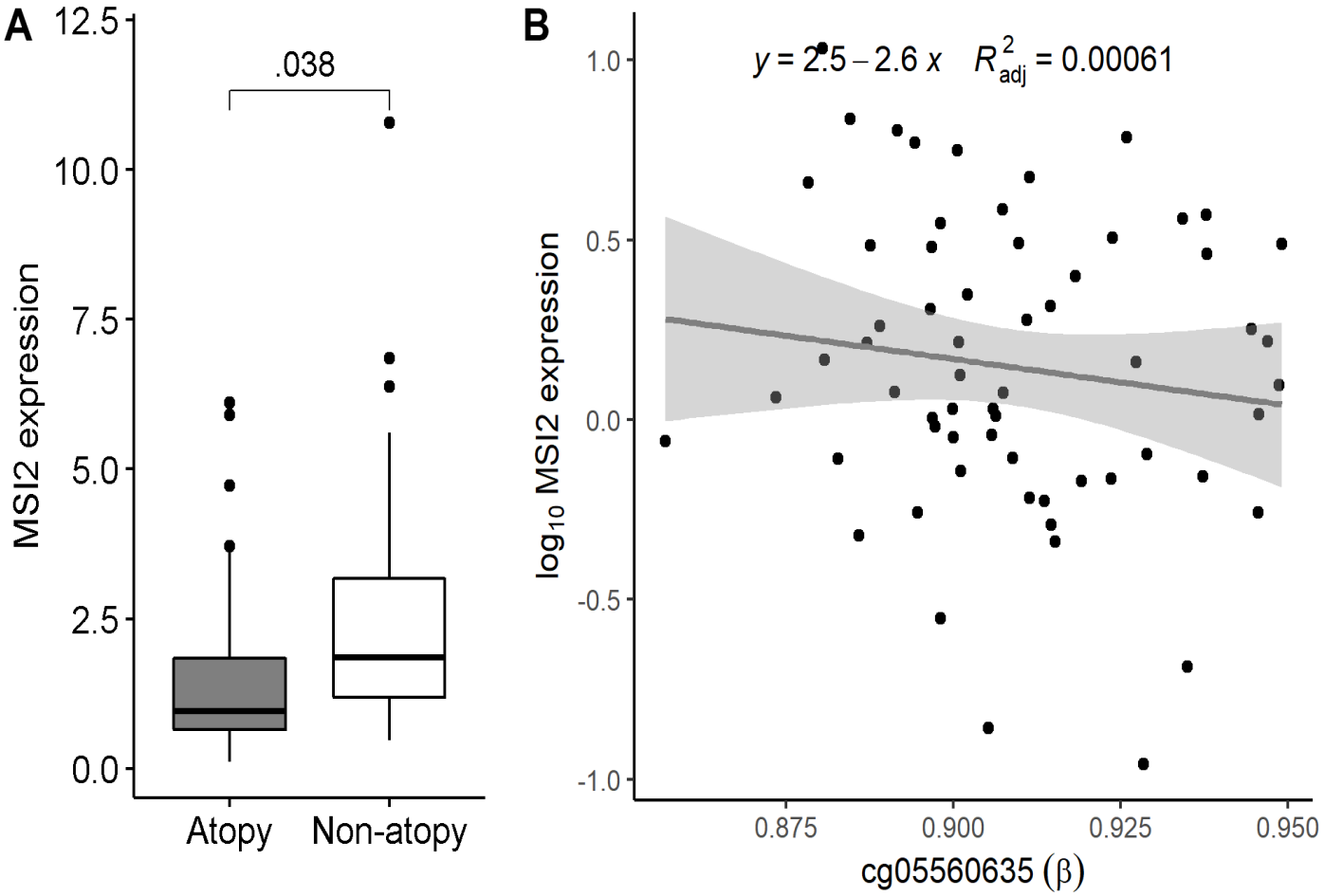
CHR	VARIANT	POS	ALT	Lambda	STAT_WALD	P_WALD	STAT_LRT	P_LRT	STAT_SCORE	P_SCORE
17	rs8082495	57630467	T	1.00E-05	9.31986	0.002302	9.30519	0.002285	9.27952	**0.002353**
17	rs877071	57635901	A	1.00E-05	0.881081	0.348041	0.881897	0.347683	0.881666	0.34788
17	**rs11652629**	57643001	C	1.00E-05	7.40083	0.006587	7.39338	0.006546	7.37717	**0.006673**

Log_{10} transformed Total IgE

17	rs8082495	57630467	T	1.00E-05	7.60816	0.005873	7.60003	0.005837	7.58291	**0.005956**
17	rs877071	57635901	A	1.00E-05	2.53348	0.111642	2.53459	0.111376	2.53268	0.111699
17	**rs11652629**	57643001	C	1.00E-05	5.61866	0.017882	5.61597	0.017798	5.60661	**0.018005**

MSI2 Expression Changed in Atopy

- MSI2 expression is significantly associated with the incidence of atopy ($P = 0.038$)
- The methylation levels of one CpG and expression were negative correlated



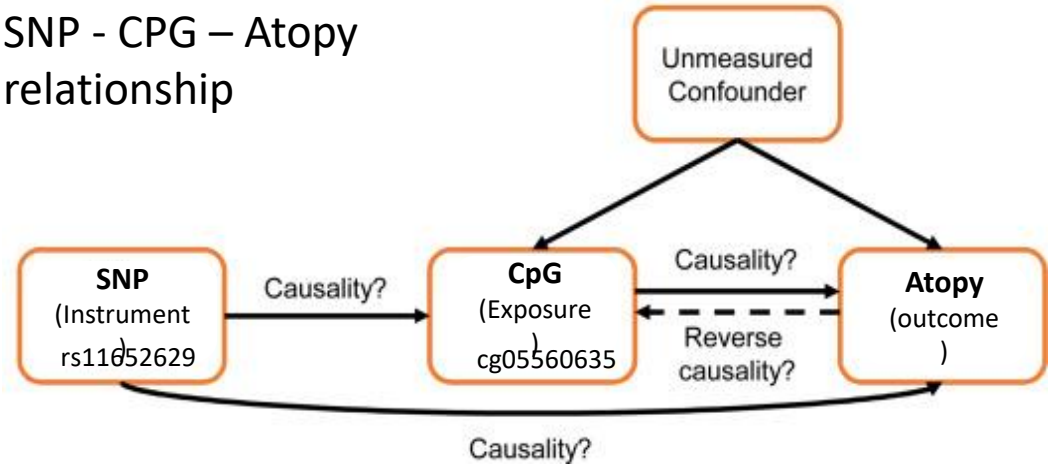
Casual Relationship

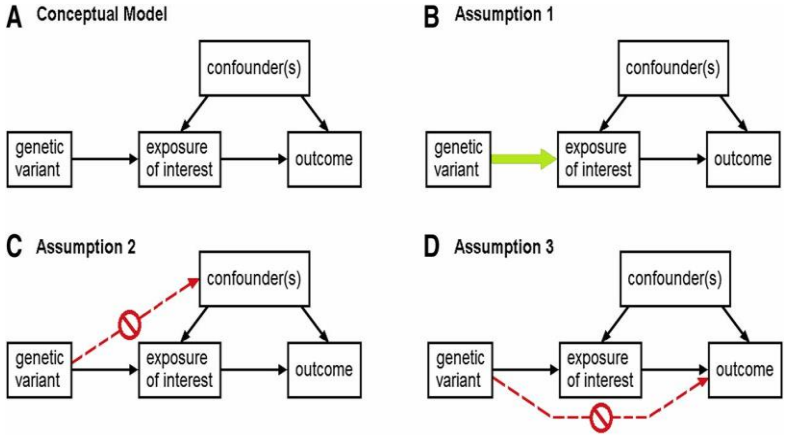
Initial hypothesis

- Genetic changes affect the methylation of adjacent CpG, which in turn leads to changes in the expression level in MSI, which is closely associated with atopy.

Mendelian Randomization and Mediation

SNP - CPG – Atopy relationship





Mendelian Randomization

Coef	ci95	t.stat	p.val
1.1921	(0.5934,1.7908)	1.9911	0.0494

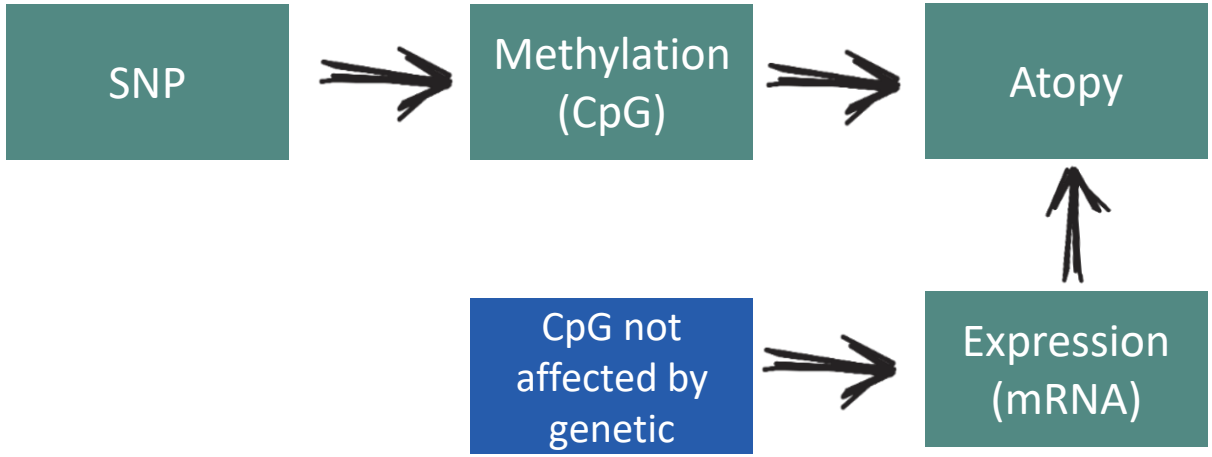
Mediation analysis

	Estimate	95% CI Lower	95% CI Upper	p-value
ACME (control)	0.19956	0.00896	0.39	0.044
ACME (treated)	0.19987	0.00973	0.39	0.044
ADE (control)	0.03737	-0.2375	0.33	0.812
ADE (treated)	0.03768	-0.22582	0.32	0.812
Total Effect	0.23724	-0.01789	0.46	0.07
Prop. Mediated (control)	0.80747	-2.44513	4.2	0.106
Prop. Mediated (treated)	0.81271	-2.4619	4.25	0.106
ACME (average)	0.19971	0.00967	0.39	0.044
ADE (average)	0.03753	-0.22902	0.32	0.812
Prop. Mediated (average)	0.81009	-2.53078	4.23	0.106

ACME : average causal mediation effect
ADE : average direct effect

MSI2 Expression Level Changes in Atopy

- Different hypothesis
  - Type types of CpG : CpG that is genetically affected and CpG that is not
  - The expression of MSI2 is likely to change due to changes in the methylation level of multiple CpGs rather than that of one CpG
  - we added other 20 CpGs associated with atopy on MSI2



MSI2 Expression Level Changes in Atopy

Linear regression analysis

MSI2 expression ~ PC scores from 20 CpGs (non-genetic)

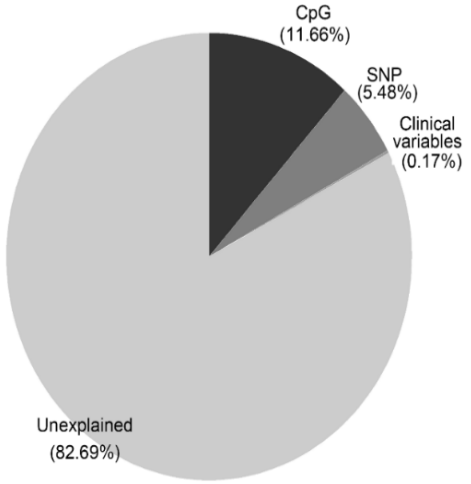
→ p-value: 0.0015, Adjusted R-squared: 0.3936

Model	Odds Ratios	95% CI	P value
Atopy ~			
Model 1			
CpG (cg05560635)	6.41×10^{15}	1.87×10^{4} - 1.45×10^{29}	.012
Model 2			
Expression* (MSI2)	0.13	2.27×10^{-2} - 5.46×10^{-1}	.009
Model 3 **			
1 CpGs + 20 PCs +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

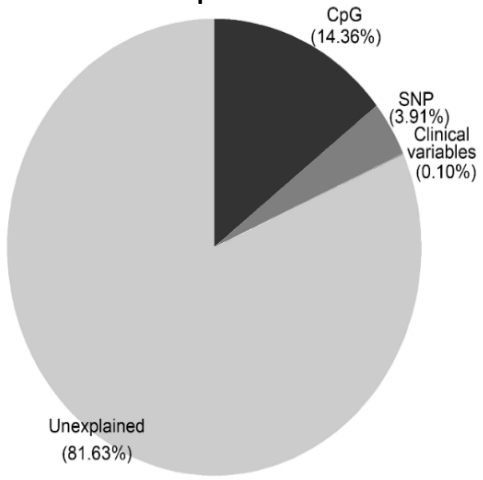
Genetically affected CpG and non-CpG are independently associated with atopy.

Global Effect on Phenotypes

Atopy



Tissue eosinophilia



A prediction model with SNPs, CpGs, and the clinical variables for atopy and tissue eosinophilia

Comparison of the McFadden's R2 improvement in accordance with each feature by 5-fold cross validation

$$\Delta = \sum_i \left\{ y_i \log \frac{y_i}{\widehat{y_i}} + (1 - y_i)\, log \frac{1 - y_i}{1 - \widehat{y_i}} \right\}$$

Although most proportions remained unexplained, Variations in CpGs yielded better predictions than SNPs or clinical variables for both phenotypes

Epigenetic changes due to environmental effects are better associated with atopy than with genetic effects

Question?