

인체 유래 환경유해물질 노출에 따른 멀티 오믹스 데이터 통합 분석 가시화 시스템

신가희[†], 홍지만^{**}, 박서우^{***}, 강병철^{****}, 이봉문^{*****}

Visualization for Integrated Analysis of Multi-Omics Data by Harmful Substances Exposed to Human

Ga-Hee Shin[†], Ji-Man Hong^{**}, Seo-Woo Park^{***},
Byeong-Chul Kang^{****}, Bong-Mun Lee^{*****}

ABSTRACT

Multi-omics data is difficult to interpret due to the heterogeneity of information by the volume of data, the complexity of characteristics of each data, and the diversity of omics platforms. There is not yet a system for interpreting to visualize research data on environmental diseases concerning environmental harmful substances. We provide MEE, a web-based visualization tool, to comprehensively explore the complexity of data due to the interconnected characteristics of high-dimensional data sets according to exposure to various environmental harmful substances. MEE visualizes omics data of correlation between omics data, subjects and samples by keyword searches of meta data, multi-omics data, and harmful substances. MEE has been demonstrated the versatility by two examples. We confirmed the correlation between smoking and asthma with RNA-seq and Methylation-Chip data, it was visualized that genes (*PHACTR3*, *PXDN*, *QZMB*, *SOCS3* etc.) significantly related to autoimmune or inflammatory diseases. To visualize the correlation between atopic dermatitis and heavy metals, we selected 32 genes related immune response by integrated analysis of multi-omics data. However, it did not show a significant correlation between mercury in blood and atopic dermatitis. In the future, should continuously collect an appropriate level of multi-omics data in MEE system, will obtain data to analyze environmental substances and diseases.

Key words: Multi-Omics, Exposure, Web-based Visualization, Environmental Harmful Substance, Integrated Analysis

1. 서 론

급격한 도시화와 산업화로 인해 환경오염은 가파

르게 증가하게 되었고, 환경유해물질 인자로 분류되고 있는 다수의 화학물질들은 생활용품 등의 형태로 일상생활에서 쉽게 접하게 되었다. 이는 특정 유전적

※ Corresponding Author : Ga-Hee Shin, Address: (16954) Heungdeok IT valley, 13, Heungdeok 1-ro, Yongin-si, Gyeonggi-do, Korea, TEL : +82-31-278-0061, FAX : +82-31-278-0062, E-mail : ghshin@insilicogen.com

Receipt date : Feb. 10, 2022, Approval date : Feb. 17, 2022

[†] Insilicogen Inc.

^{**} Insilicogen Inc.

(E-mail : jmhong@insilicogen.com)

^{***} Insilicogen Inc.

(E-mail : hspark@insilicogen.com)

^{****} DiF Inc.

(E-mail : bckang@d-if.kr)

^{*****} CKU center for health policy research., Catholic Kwandong University

(E-mail : leebm8407@gmail.com)

※ This research was supported by Korea Environment Industry & Technology Institute (KEITI) through “the Environmental Health Action Program”, funded by Korea Ministry of Environment (MOE) (2017001360005)

요인뿐만 아니라 환경 유해물질의 노출이 질병을 유발할 수 있다는 것을 의미한다[1-3]. 환경유해인자 노출에 의해 발병하는 질환을 ‘환경성질환’으로 정의하고 있고, 최근에 환경성 질환을 발생시키는 기전 중 증상을 악화시키는 원인으로 환경유해물질들이 알려지면서 환경관리에 대한 대중의 관심이 높다[4]. 임신 중 산모에게 노출된 환경유해인자가 태반을 통과하여 태아의 발달과 생후의 건강에 영향을 미친다는 주장과 함께 정산군(저노출 산모) 대비 임신결과(조산아, 저체중아, 출생 시 체중, 머리둘레 등), 성장과정 중 인지 발달, 아토피피부염, 알레르기 비염, 천식 등의 발병 비율 증가 등에 대한 연구 결과가 보고되고 있다[5-6]. 따라서 산모가 환경유해물질에 노출됨에 따라 태아에게 미치는 영향을 평가하기 위해서는 질병 발생과 같은 표현형 관찰과 더불어 태아 발달(differentiation) 과정부터 시작되는 독성영향에 대한 원론적 관점의 평가가 필요하다. 그러나 이는 생명윤리 상 인간을 대상으로 실시할 수 없는 연구 범위이므로, 이를 간접적으로 평가할 수 있는 시스템의 구축이 필요하다.

질병 발생에 대한 유전적 및 환경적 요인은 유전자-환경 상호작용에 관한 다양한 연구에서 밝혀진 바와 같이 질병이 독립적인 요인에 기인하기보다 상호 작용에 의해 발병하므로 환경 노출로 인한 발병 기전을 확인하고 질병과 환경 노출 간의 연관성을 이해하는 것이 중요하다[7-8]. 단일 오믹스 기술(예: 전사체학)은 한 유형의 생체 분자를 감지하므로 멀티 오믹스 데이터를 통해 포괄적인 분석이 필요하다. 그러나 멀티 오믹스 데이터는 여러 고차원 데이터 세트가 상호 연결된 특성으로 인해 데이터의 복잡성 때문에 분석하기가 어렵다. 이를 효과적으로 표현하기 위해서는 데이터 특성의 의미를 포함한 포괄적인 탐색이 가능하도록 사용하기 쉬운 시각화 도구를 이용하는 것이 필수적이다[9]. 멀티 오믹스 데이터의 시각화를 위한 생물정보학 도구로는 cBioPortal[10], UCSC Xena[11] 및 LinkedOmics[12]가 있으며, 이들은 유전자에 대한 쿼리 기반 탐색 및 시각화를 위한 웹 인터페이스를 제공한다. 이러한 방법은 소수의 선택 마커에 대한 정량적 값을 표시하거나 경로 또는 네트워크와 같은 추상적인 수준에서 전체 추세를 시각화하는 데 중점을 두고 있어 멀티 오믹스의 정량적 데이터를 직접적으로 시각화하는 도구는 거의 없다

[13-15].

우리는 다중 오믹스 데이터를 쉽게 탐색할 수 있도록 통합 가시화 도구인 MEE (Multi-Omics Explore Exposome)를 제안한다. MEE는 국내 외 환경 유해물질 노출에 따른 환경성 질환 연구 데이터들을 정규화하고 분석파이프라인을 통해 분석된 결과 및 시각화 정보를 제공하여 사용자의 가설에 따라 키워드 검색 기반으로 합리적인 규모로 좁혀진 정량적 다중 오믹스 데이터를 시각화하는 웹 인터페이스이다. 본 논문에서는 환경유해물질과 천식 또는 아토피와 같은 환경성 질환에 대한 멀티 오믹스 데이터 세트를 이용하여 시각화 기능을 보여주고자 한다. 첫 번째 데이터 세트는 RNA-Seq 데이터와 Methylation-Chip 데이터[16]로 천식 발병과의 상관관계를 밝히기 위한 데이터를 이용하여 가시화하였다. 두 번째 데이터 세트는 아토피 피부염과 혈중 중금속 농도에 따른 RNA-Seq 데이터(unpublished data, 삼성의 료원 그린코호트 2019~2021)의 상관관계를 가시화하였다. 이러한 예는 유전자 기반 키워드 검색 및 사용자 지정 옵션을 통해 멀티 오믹스 데이터의 주요 세그먼트를 시각화하기 위한 MEE의 다용성을 보여준다. MEE는 환경유해물질과 연관된 멀티 오믹스 데이터를 시각화하여 질병과 환경 노출간의 네트워크 정보를 확인하게 함으로써 분자적 수준에서의 다양한 연관성을 포괄적으로 이해할 수 있게 할 것이다.

2. 재료 및 방법

2.1 가시화 워크플로우

Fig. 1은 MEE에서의 일반적인 가시화 워크플로우를 나타낸 것이고, 웹기반의 시각화 인터페이스는 Fig. 2에 나타내었다. MEE는 유전자 기반 데이터 검색, 다양한 통계 차트 가시화로 유전자 단위의 상호 비교 분석을 지원하고, 연구대상자 및 검체 내 환경 유해물질의 연관분석을 지원하기 위한 유해물질의 교차 검색 및 PCA 분석 가시화 결과를 웹 기반으로 제공하는 시스템이다. 각 오믹스 데이터는 전처리 과정 및 정규화(normalization) 과정을 거쳐 가시화 시스템에서 활용할 수 있는 오믹스 데이터 파일 형식으로 변환한다. 예를 들어, 메틸화 프로파일 데이터 및 유전자 발현 프로파일 데이터는 유전자와 연구샘플의 행렬 파일로 변환하고, 유전체 변이 데이터는

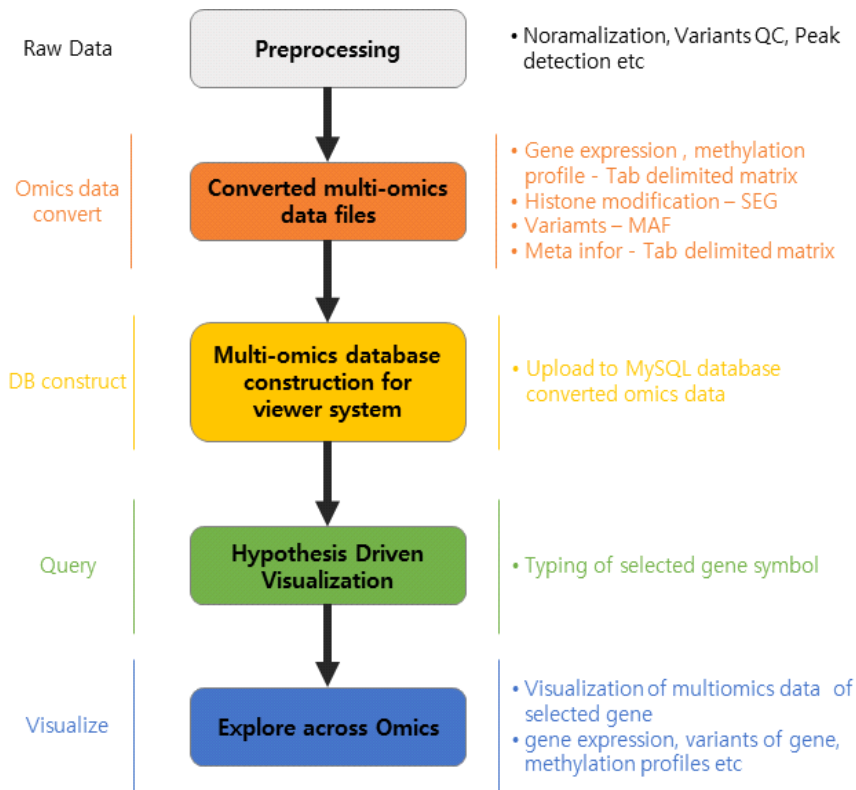


Fig. 1. Visualization work-flow of MEE.

MAF 형식의 파일로 변환하고, 연구/연구 개체 및 샘플등의 메타정보는 텍스트 파일 형태로 변환한다. 이후 메타 정보 및 오믹스 데이터 파일들은 가시화 및 분석을 위해 MySQL로 구성된 데이터베이스에 등록한다.

MEE는 등록된 연구 대상자 및 샘플의 메타 정보를 요약 하여 가시화하고, 각 연구대상자 또는 샘플 별 오믹스 데이터를 유전자 단위를 기준으로 조회하여 가시화할 수 있다. 즉, EnvPrint 메뉴를 통해 각 연구대상자 샘플의 유전자 별 발현 정보 및 변이 정보를 요약하여 가시화 하거나, 각 유전자 별 변이 정보 조회 및 간단한 통계 분석 기능 등을 사용할 수 있다(Fig. 2).

2.2 소프트웨어 아키텍처

환경유해인자 멀티오믹스 DB 가시화 시스템 MEE의 아키텍처 모식도는 Fig. 3에 나타내었다. 서버측은 2종의 웹서버, DB서버, Client 웹 브라우저를 포함한 4가지 구성 요소의 응용 프로그램이다. 사용자는 원하는 정보를 Client 웹 브라우저에서 검색,

조회, 가시화하고, 웹 인터페이스는 프론트엔드 프레임워크(AngularJS)를 적용하였다. 2종의 웹서버 중 WAR(Web application Archive)은 멀티 오믹스 DB 시스템을 구동하는 메인 어플리케이션으로 DB 서버와 통신하고, 사용자 웹 브라우저 및 PCA explorer과의 데이터를 통신한다. 또한 PCA explorer를 서비스 하기 위해 R 기반의 Shiny 전용 웹서버를 별도로 구축하였다. DB서버는 멀티 오믹스 DB 검색 시스템 내의 모든 데이터가 저장되어 있고, 유해물질 중심의 통합 검색 및 가시화를 위해 RDB(MySQL)와 유전자 단위의 프로파일링 검색이 가능하도록 구성하였다(Fig. 3).

2.3 데이터베이스

MEE는 다음과 같은 4가지의 DB가 관계형 DB로 구성되어 있다. 1) Exposome DB: 유해물질명, 클래스, 분류 등의 세부 정보가 포함된 환경유해물질 DB, 2) Cohort Pheotype DB: 환경유해성연구, 연구대상자, 연구대상자 히스토리, 연구 샘플 등의 연구를 지



Fig. 2. The main visualization interface for MEE. (a) Performing a query of a asthma multi-omics study, (b) Gene mutation frequencies and types in samples. The red bars indicate gene amplifications, blue bars are deep deletions, green bars are missense mutations, gray bars indicate truncating mutations, (c) Add clinical tracks such as disease, age, gender and smoking to sample set, and (d) Summary of gene profiling based on target genes for a set of samples.

원하기 위한 표현형 DB(삼성의료원 그린코호트 2019~2021), 3) Omics DB: 샘플별 SNP-genotyping chip, Exome-Seq, transcriptome, metabolite, epigenome, meDIP seq, hg37 ref 정보, 4) 변이체 관

련 유전자 검색 및 유전자 프로파일링 가시화 DB.

2.4 입력데이터

MEE는 간단하고 직관적인 인터페이스로 사용자

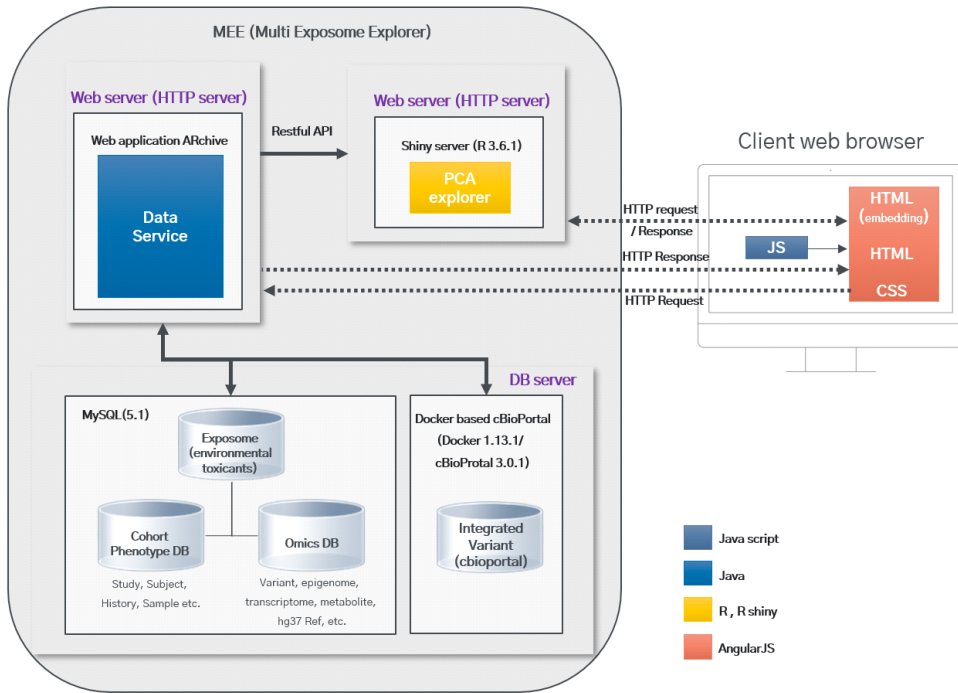


Fig. 3. System architecture for environmental harmful substance multi-omics DB.

의 데이터를 업로드하여 분석할 수 있도록 하였다. 전사체 데이터는 유전자에 대한 전체 샘플의 read count 데이터를 z-score값으로 정규화하고 후성유전체 데이터는 0과 1사이로 정규화하여 텍스트 파일로 구성하였다. 파일의 첫 번째 컬럼은 Hugo Gene Symbol 또는 Entrez Gene ID로 구성하였다. 유전체 데이터는 각 샘플 별 변이 정보에 대한 MAF 형식의 파일을 입력파일로 구성하였다. 각 오믹스별 데이터는 연구대상자 및 연구 샘플에 따라 고유한 아이디를 가지고 있어 오믹스 데이터와 샘플의 메타정보의 연관성을 가시화하여 확인할 수 있도록 하였다. MEE의 입력데이터는 원시 데이터를 전처리하고, 변환한 데이터 파일이며, 각 열에는 고유한 템플릿을 가지고 있다. 데이터 파일 외에도 샘플 속성(예: 임상 데이터 또는 표현형)이 포함된 별도의 샘플 정보 파일도 필요하고, 업로드 된 데이터 파일에서 시각화할 수 있도록 분자 하위 집합을 선택하는 기능을 제공한다.

연구대상자 정보 및 인체 샘플 정보 등 입력데이터는 원칙적으로는 사전에 동의를 구한 범위 내에서 활용하였다. 단, 추가정보의 사용 없이는 특정개인의 식별이 불가능하게 조치한 정보인 가명정보는

동계작성(상업적 목적 포함), 연구(산업적 목적 포함), 공익적 기록보존 등의 목적으로 사용하는 경우에는 동의 없이 활용 가능하지만, 가명정보와 개인정보(특정개인에 관한 정보, 개인 식별을 가능케 하는 정보)의 결합을 시도하거나, 개인정보를 안전하게 관리하지 못하는 경우에는 처벌 받을 수 있으므로 비식별화 된 입력데이터 관리를 철저히 하였다[17].

3. 결과 및 고찰

3.1 MEE 시스템 기능

3.1.1 멀티 오믹스 데이터 가시화

MEE 시스템은 환자의 메타 정보(질환 유무, 채취 시료 정보 등) 및 멀티 오믹스 정보(genome, transcriptome, epigenome 및 metabolome 등)와 환자별 유해물질노출 데이터를 통합하여 연구대상자 및 샘플의 메타 정보조회, 오믹스 데이터(transcriptome, epigenome, variants 등) 조회 그리고 오믹스 데이터 간 상관관계 (correlation), co-expression, enrichment 분석 등의 통계분석이 가능하도록 구현하였고, 이러한 기능은 2가지 방법으로 가시화 될 수 있다.

첫 번째는 특정 프로젝트에 등록된 환자 및 샘플에 대한 오믹스 정보를 조회(Explore selected studies)하여 가시화하는 방법이다. Explore selected studies 기능은 선택된 프로젝트에 대한 환자 및 샘플의 메타 정보를 가시화하고, 등록된 연구대상자나 샘플에 대하여 멀티 오믹스 데이터 중 유전자 변이, Histone modification (ChIP Seq 데이터) 요약 정보를 조회하여 전장 유전체에 대한 전체적인 Histone modification 수준을 시각화하였다. Fig. 4에는 현재 진행하고 있는 프로젝트의 일부 데이터를 MEE에 등록하여 가시화 하였다. 292명 연구대상자의 메타정보와 1,281개 샘플에 대한 오믹스 정보를 요약하였고, 이들 중 ChIP-Seq 데이터를 조회하여 712개 샘플의 Histon

modification 수준을 시각화한 결과이다. 두 번째는 유전자를 특정하여 해당 유전자에 대한 오믹스 정보를 조회(Query by Gene)하여 가시화될 수 있다. 두 번째 Query by Genes 기능은 사용자가 타겟하는 유전자를 기반으로 DB서버에 등록된 멀티 오믹스 데이터를 검색하고, 수치화된 멀티 오믹스 데이터를 heatmap으로 나타내어 발현의 정도로 상관관계를 시각화하는 것이다. 유전자 발현 수준에 대한 간단한 통계분석 결과와 mRNA 발현정도나 mutation 등에 대한 enrichment 분석, 이나 mRNA 발현 정보를 이용한 co-expression 정보를 제공하고, 그 외 유전자들에 대한 네트워크 분석도 가능하도록 구현되었다 (Fig. 5).



Fig. 4. Visualization of omics information on subjects and samples, (a) Summary of meta-data and (b) Histone modification data.

3.1.2 주성분 분석

환경유해물질의 화학정보와 멀티 오믹스 데이터를 검색하여 유해물질간의 상관관계를 분석할 수 있는 기능을 구현하였다. 환경유해물질은 PubChem DB와 연결되어 있고 연구대상자와 샘플은 각 유해물질의 노출 수치 값과 연결되어 R shiny 기반으로 주성분 분석이 가능하도록 구현하였다.

3.2 Case Study I

RNA-Seq 데이터와 Methylation-Chip 데이터로 천식 발병과의 상관관계를 통합 분석할 수 있도록

가시화하기 위해 NCBI에 등록된 데이터 셋(Accession No. GSE85567, GSE85568)을 활용하였다[16]. 총 83명의 기도상피세포(Airway Epithelial Cell, AEC)로부터 유전자 발현 프로파일(reads count) 데이터인 RNA-Seq 데이터를 확보하고, 이 중 28명은 정상군, 55명은 천식질환 군으로 분류하였다. FDR 5% 이하, log2(fold-change)값 1.5 이상 또는 log2(fold-change)값 -1.5이하의 조건으로 정상군과 천식질환 군 간 차등 발현 유전자 108개를 선별하였으며, Reads count 데이터는 In-house script를 통해 z-score로 변환하여 데이터를 등록하였다. Methylation

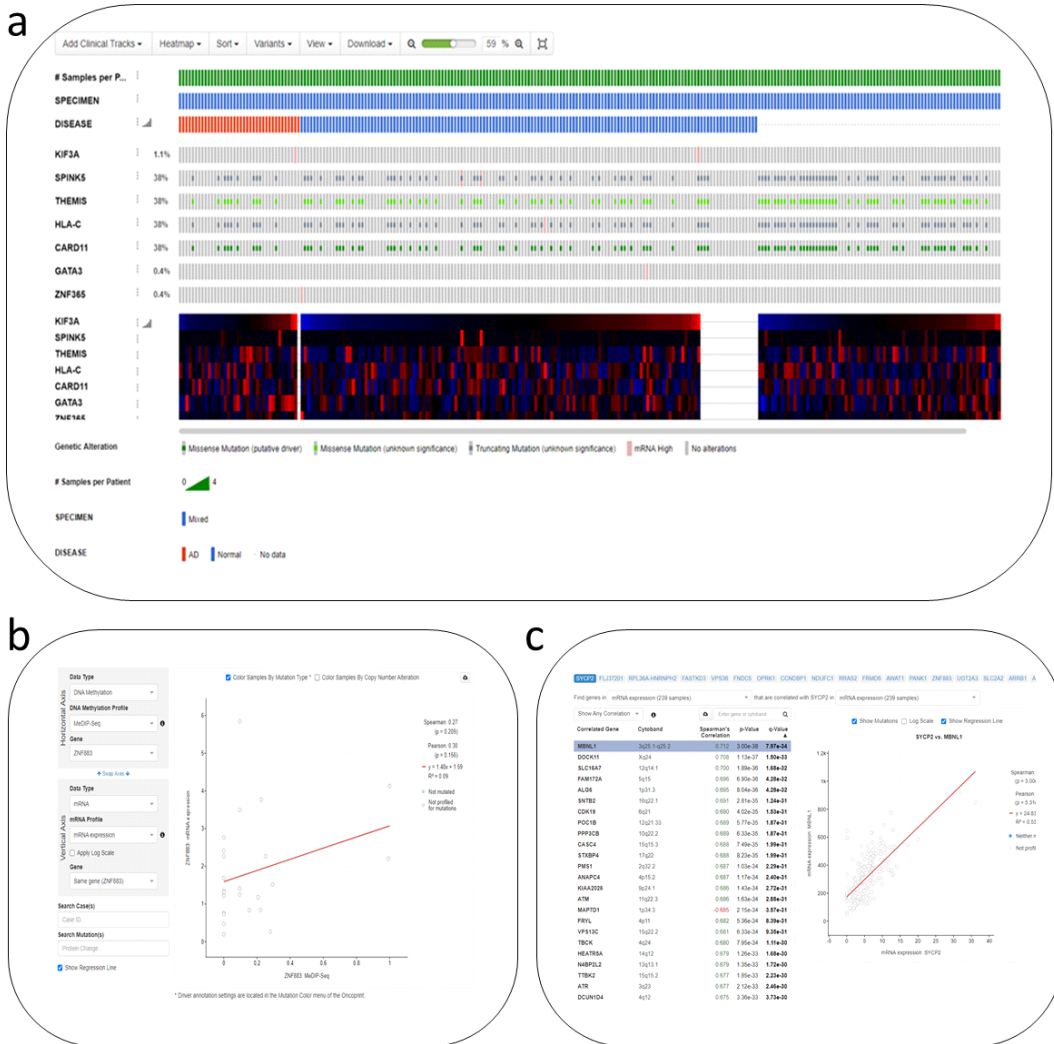


Fig. 5. Visualization of multi-omics data, (a) Meta-data, Gene profile of RNA-Seq, Medip-seq, Chip-seq, SNP, (b) Interaction of RNA-Seq and MeDIP-Seq, and (c) Co-expression.

ation Beads Array 데이터는 R package 인 minfi (v 1.32.0) 을 이용하여 전처리 하고, SWAN 알고리즘으로 메틸화 수준 데이터를 정규화한 후 각 프로브 별 유전자 주석을 분석하였다. 프로브 위치 정보를 기준으로 RNA-seq에서 확보한 108개 중 72개의 유전자를 선별하여 612개 프로브의 메틸화 데이터를 추출하고, 전사체 및 후성유전체 데이터를 통합 분석하기 위해 MEE 시스템에서 가시화하였다. Fig. 6은 천식 유무, 흡연 유무 등의 환자정보, 추출된 유전자 프로파일, 유전자 발현정보(RNA-Seq), 메틸화 수준 정보(Methylation Chip) 그리고 니코틴 주요대사 물질인 Cotinine 환경유해물질 정보를 통합하여 가시화한 결과이다. 그림에서 나타나는 바와 같이 흡연유무는 천식질환 군과는 연계성을 가지지 않고, 차등 발현된 일부 유전자(*PHACTR3*, *PXDN*, *QZMB*, *SOCS3*)중 *SOCS3* 의 3개 유전자에서 천식질환 군에서 발현이 높게 나타났으며, 유전자 별 메틸화 수준도 비교할 수 있었다. *PHACTR3*, *PXDN*, *QZMB* 은 자가면역 또는 면역관련 염증질환과 관련 있는

유전자들이다.

3.3 Case Study II

총 55명을 대상으로 아토피 피부염과 혈중 중금속 농도에 따른 RNA-Seq 데이터(unpublished, 삼성의료원 그린코호트 2019~2022)를 활용하여 상관관계를 가시화하였다. 확보된 연구대상자의 유전자 발현 데이터를 이용하여 아토피 피부염과 혈중 중금속 농도의 상관관계를 고려하여 Generalized linear model (GLM) 방법을 적용하여 차등발현 유전자를 탐색하였다. 수은, 망간, 카드뮴, 납에 대한 차등발현 유전자 분석을 수행하였으나, 수은 이외에는 혈중 중금속 농도에 대해 통계적으로 유의미한 유전자를 확인할 수 없었다. 혈중 수은 농도에 대하여 차등발현 유전자를 분석한 결과, FDR 5% 이하의 유의미한 유전자 32개가 탐색 되었으며, mast cell activation 등 면역과 연관된 유전자들이었다. 연구대상자등 메타정보, 중금속, RNA-seq데이터 등을 통합적으로 가시화하기 위해 MEE시스템에 반영하여 그 결과를 확인할 수는

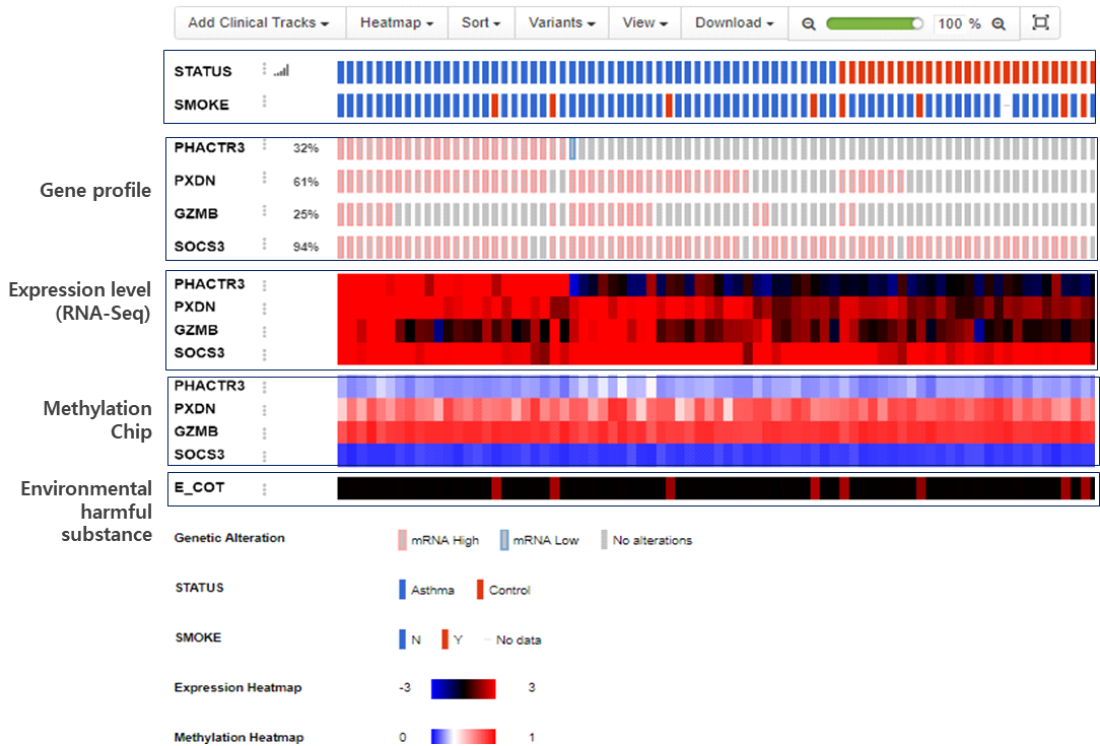


Fig. 6. Representation of the pattern and type of meta-data, gene profile of RNA-Seq and methylated level in AFC samples according to Cotinine.

있었으나, 유의한 상관성을 가지적으로 볼 수는 없었다. 이는 Case Study I과 같이 멀티 오믹스 데이터를 가시화하여 통합 분석하기 위해서는 적절한 수준의 데이터가 집적되어야 함을 알 수 있었다.

4. 결 론

본 논문에서는 저농도로 장기간 동안 환경유해물질이 인체에 노출될 경우, 인체 내 분자적 변화를 일으키므로 이러한 현상을 다양한 멀티 오믹스 데이터를 통해 환경성 질환과 환경유해물질과의 연관성을 통합적으로 해석할 수 있는 가시화 시스템을 제안하였다. 현재까지 MEE 시스템에는 총 1281개의 샘플이 등록되어 있으며, 웹을 통해 가시화에 활용할 수 있도록 하였다(<http://mee.insilicogen.com/>). 본 논문에서는 두 가지 사례를 통해 가시화 결과를 보고하였다. 먼저 흡연유무에 따른 천식 발병과의 상관관계를 통합 분석하여 천식 질환군에서 자가 면역 또는 염증 질환과 관련된 일부 유전자(PHACTR3, PXDN, QZMB, SOCS3)가 유의미한 발현을 확인하였다. 그리고 아토피 피부염과 혈중 수은농도와의 상관관계를 RNA-Seq 데이터를 활용하여 분석하여 면역과 관련된 유의미한 유전자 32개가 탐색하여 메타정보, 증금속, RNA-seq데이터 등을 통합적으로 가시화하였으나, 유의한 상관성을 가지적으로 확인할 수는 없었다. 본 연구를 통해 시스템에서 환경 물질과 질병을 분석하기 위해서는 적절한 수준의 멀티 오믹스 데이터가 지속적으로 집적되어야 세부적 분석을 수행할 수 있으므로 데이터의 정합성에 맞추어 다양한 멀티 오믹스 데이터를 확보하고자 한다. 따라서 멀티 오믹스 데이터와 메타정보가 지속적으로 쌓이게 되면, 환경과 질환 관련 연구를 수행하는 많은 연구자들에게 도움이 될 것이다.

REFERENCE

- [1] K.T. Cha, S.S. Oh, J.H. Yoon, K.H. Lee, S.K. Kim, B.S. Cha S.H. Kim et al., "Adverse Health Outcomes in Residents Exposed to Cement Dust," *Toxicology and Environmental Health Sciences*, Vol. 3. pp. 239-244, 2011.
- [2] F. Grun and B. Blumberg, "Environmental Obesogens: Organotins and Endocrine Disruption via Nuclear Receptor Signaling," *Endocrinology*, Vol. 147, pp. S50-S55, 2006.
- [3] R. Vermeulen, E.L. Schymanski, A.L. Barabási, and G.W. Miller, "The Exposome and Health: Where Chemistry Meets Biology," *Science*. Vol. 367, No. 6476, pp. 392-396, 2020.
- [4] S.B. Koh, "Environmental Diseases," *Journal of the Korean Medical Association*, Vol. 55, No. 3, pp. 212-213, 2012
- [5] O. Robinson and M. Vrijheid, "The Pregnancy Exposome," *Current Environmental Health Report*, Vol. 2, pp. 204-213, 2015.
- [6] K.A. Thayer, J.J. Heindel, J.R. Bucher, and M.A. Gallo, "Role of Environmental Chemicals in Diabetes and Obesity: a National Toxicology Program Workshop Review," *Environmental Health Perspectives*, Vol. 120, pp. 779-789, 2012.
- [7] M.M. Niedzwiecki, D.I. Walker, R. Vermeulen, M. Chadeau-Hyam, D.P. Jones, and G.W. Miller, "The Exposome: Molecules to Populations," *Annual Review of Pharmacology and Toxicology*, Vol. 59, pp. 107-127, 2019.
- [8] C.P. Wild, "Complementing the Genome with an "Exposome": The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology," *Cancer Epidemiology Biomarkers & Prevention*, Vol. 14, No. 8, pp. 1847-1850, 2005.
- [9] R. Hernández-de-Diego, S. Tarazona, C. Martínez-Mira, L. Balzano-Nogueira, P. Furió-Tarí, G.J. Pappas et al., "PaintOmics3: a Web Resource for the Pathway Analysis and Visualization of Multi-omics Data," *Nucleic Acids Research*. Vol. 46, pp. W503-W509, 2018.
- [10] J. Gao, B.A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S.O. Sumer et al., "Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal," *Science Signaling*. Vol. 6, p. pl1, 2013.
- [11] M.J. Goldman, B. Craft, M. Hastie, K. Repečka,

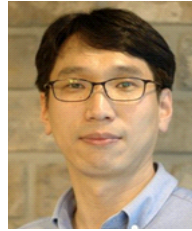
- F. McDade, A. Kamath et al., “Visualizing and Interpreting Cancer Genomics Data via the Xena Platform,” *Nature Biotechnology*, Vol. 38, pp. 675–678, 2020.
- [12] S.V. Vasaiakar, P. Straub, J. Wang, and B. Zhang, “LinkedOmics: Analyzing Multi-Omics Data within and across 32 Cancer Types,” *Nucleic Acids Research*, Vol. 46, pp. D956–D963, 2018.
- [13] R. Ihaka and R. Gentleman, “R: a Language for Data Analysis and Graphics,” *Journal of Computational and Graphical Statistics*, Vol. 5, No. 3, pp. 299–314, 1996.
- [14] E. Bonnet, L. Calzone, and T. Michoel, “Integrative Multi-Omics Module Network Inference with Lemon-Tree,” *PLoS Computational Biology*, Vol. 11, No. 2, pp. e1003983, 2015.
- [15] R. Argelaguet, B. Velten, D. Arnol, S. Dietrich, T. Zenz, J.C. Marioni et al., “Multi-Omics Factor Analysis—a Framework for Unsupervised Integration of Multi-Omics Data Sets,” *Molecular Systems Biology*, Vol. 14, No. 6, pp. 8124, 2018.
- [16] J. Nicodemus-Johnson, R.A. Myers, N.J. Sakabe, D.R. Sobreira, D.K. Hogarth, E.T. Naureckas et al., “DNA Methylation in Lung Cells is Associated with Asthma Endotypes and Genetic Risk,” *The Journal of Clinical Investigation Insight*, Vol. 1, No. 20, pp. 90151, 2016.
- [17] Y. Park and Y. Lee, “A Study on Countermeasure for Privacy in Mobile Office,” *Journal of Korea Multimedia Society*, Vol. 18, No. 2, pp. 178–188, 2015.



신 가 희

2008년 6월~현재 ㈜인실리코젠
수석연구원
2008년 10월 베를린공과대학교
생명공학과(공학박사)
2004년 4월 베를린공과대학교 생
명공학과(Dipl.Ing)

1997년 2월 동서대학교 식품공학과(공학사)
관심분야: 분자생물, Omics 정보, 생물공정



강 병 철

2006년 2월~현재 인실리코젠
R&D 연구소장
2004년 8월 부산대학교 생물정보
학과(공학박사)
1998년 2월 부경대학교 기계공학
과(공학석사)

1996년 2월 부경대학교 기계공학과(공학사)
관심분야: 생물정보, Omics 정보, 시맨틱 DB



홍 지 만

2013년 4월~현재 인실리코젠 책
임 연구원
2007년 2월 부산대학교 생물정보
학과(이학석사)
2005년 2월 청주대학교 유전공학
과(이학사)

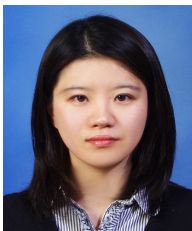
관심분야: 생물정보, Omics 정보, 머신러닝



이 봉 문

2021년 2월~현재 가톨릭관동대
학교 의학산업연구센터
센터장
2017년 5월~현재 가톨릭관동대
학교 기술지주회사 대표
2014년 9월~정년 가톨릭관동대
학교 의과대학 의학과 교수

2013년 2월 인하대학교 법과대학원(법학박사)
2000년 2월 연세대학교 법무대학원(법학석사)
1997년 2월 연세대학교 생물학과(이학사)
관심분야: 의학산업, 빅데이터, 의료산업, 개인정보보
호법



박 서 우

2017년 3월~현재 ㈜인실리코젠
BS실
2012년 3월~2013년 3월 ㈜인실
리코젠 Codes 사업부 SI
팀
2012년 2월 충남대학교 낙농자원
학 석사(농학석사)

2008년 2월 전남대학교 동물자원학부(농학사)
관심분야: 생물정보