



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

박희원 교수 지도
석사학위 청구논문

샘플의 특성값에 따른 분자 상호작용
추출을 위한 유전자 네트워크 분석

2025

성신여자대학교 대학원
통계학과
용혜민

샘플의 특성값에 따른 분자 상호작용 추출을 위한 유전자 네트워크 분석

박희원 교수 지도

이 논문을 석사학위논문으로 제출함

2024년 11월

성신여자대학교 대학원

통계학과

용혜민

인 준 서

용혜민의 석사학위 논문으로 인준함

2024년 12월

심사위원장

이성건



심사위원

최태화

심사위원

신준호

성신여자대학교 대학원

논문 개요

본 연구는 유전자 네트워크 분석에 기초하여 샘플의 특성값에 따른 유전자 상호작용을 추출하고 대량 대규모 유전자 네트워크를 해석하는 통계적 방법론을 제안하였다. Varing coefficient model을 이용하여 sample specific 유전자 네트워크를 추정하였다. 이후 제안한 통계량 Change of Regulatory Effect with Jaccard Index와 제안한 순열 검정법을 이용해 샘플의 특성값에 따라 유의미하게 변화하는 유전자 상호작용을 추출하였다. 새로운 통계량은 네트워크 구조의 유사성과 상호작용의 영향을 종합적으로 고려하였다. 새로운 순열 검정법은 균일 분포를 통해 생성된 순열 데이터를 토대로 순열 통계량을 계산한다.

제안한 통계적 방법론에 기초하여 급성 골수성 백혈병의 항암제 midostaurin에 대한 감수성의 변화에 따라 유의하게 달라지는 유전자 상호작용 추출을 시도하였다. 검정 대상으로 resistant와 sensitive cell line에서 추출된 subnetwork를 선정하였다.

분석 결과, resistant cell line과 sensitive cell line의 두 네트워크는 동일한 유전자 목록을 공유하고 있지만 네트워크 구조에서 큰 차이를 보였다. Resistant cell line의 네트워크에 존재한 hub gene (*CD53*, *MTIG*, *ANXA10* 등)이 sensitive cell line의 네트워크에서는 그의 상호작용이 손실되었음이 확인되었다. 반대로 sensitive cell line에서 존재한 hub gene(*BLNK*, *ANKRDI*, *NRN1* 등)이 resistant cell line의 네트워크에서 상호작용이 손실되었음이 확인되었다.

본 연구를 통해 항암제 감수성이 변화하는 환경에서 분자 상호작용을 이해하고 정밀 의료와 맞춤형 의료 분야의 새로운 치료타겟과 biomarker를 발굴할 가능성을 제시함으로써 의의를 가진다. 그러나 성능 평가가 실시되지 않았다는 것을 한계점으로 들어 향후 연구 과제로 Monte Carlo Simulation을 사용한 성능 평가와 개선이 필요하다.

핵심용어: 분자상호작용, Cell line specific 유전자 네트워크 분석, Midostaurin

목 차

논문 개요	i
목 차	iii
그림 목 차	iv
표 목 차	v
I. 서론	1
II. 유전자 네트워크 추정	4
1. 유전자 네트워크 추정	4
2. Sample specific 유전자 네트워크 추정	5
III. 추정된 유전자의 네트워크 해석	9
1. Regulatory Effect	10
2. Jaccard Index	12
3. Regulatory Effect with Jaccard Index	13
4. Change of Regulatory Effect with Jaccard Index	14
5. 샘플의 특성값에 따른 네트워크 변화의 유의성 검정	14
IV. 급성 골수성 백혈병 항암제 midostaurin 감수성을 특징 짓는 유전자 네트 워크 분석	16
1. 자료 및 평가 방법	16
2. 적용 결과	17
1) Midostaurin resistant specific 유전자 네트워크	18
2) Midostaurin sensitive specific 유전자 네트워크	19

3)	추출된 유전자 네트워크의 생물학적 경로 분석 결과	21
①	Midostaurin resistant specific 유전자 네트워크	21
②	Midostaurin sensitive specific 유전자 네트워크	22
V.	결론	23
	참고문헌	24
	ABSTRACT(영문초록)	34

그림 목 차

그림 1.	유전자 네트워크의 예시	1
그림 2.	대규모 유전자 네트워크의 예시	2
그림 3.	제안한 분석 프로세스의 overview	3
그림 4.	Gaussian kernel 함수를 통한 sample specific 네트워크 추정	6
그림 5.	대량 대규모 sample specific 유전자 네트워크의 예시	7
그림 6.	샘플의 특성에 따라 유전자 네트워크 추출의 예시	9
그림 7.	유전자 네트워크의 예시	10
그림 8.	Resistant cell line에서 추출한 네트워크 비교	18
그림 9.	Sensitive cell line에서 추출한 네트워크 비교	19
그림 10.	Resistant cell line specific 유전자 네트워크의 GO term 그래프	21
그림 11.	Sensitive cell linespecific 유전자 네트워크의 GO term 그래프	22

표 목 차

표 1.	Subnetwork 별 p-value 분포 결과	17
표 2.	Midostaurin resistant specific biomarker와 관련된 문헌	20
표 3.	Midostaurin sensitive specific biomarker와 관련된 문헌	20

I. 서론

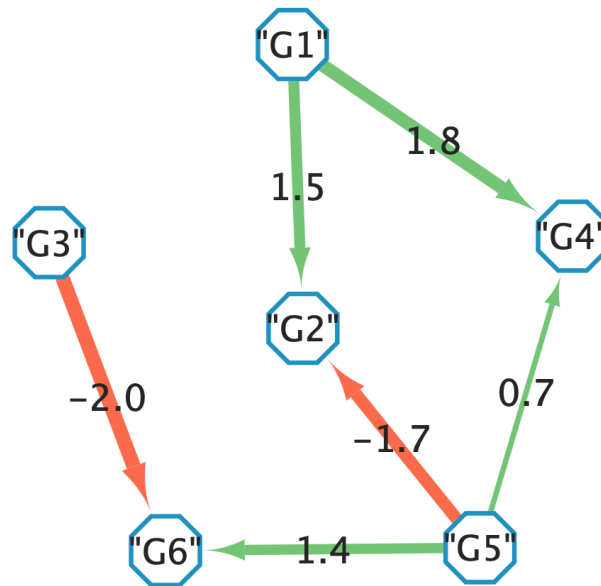


그림 1: 유전자 네트워크의 예시

유전자 네트워크는 유전자 또는 유전자 산물을 노드로 분자 간 상호작용을 예시로 표현한 네트워크이다. 그림 1과 같이 세포 내 다른 물질과 상호작용하여 mRNA와 단백질의 유전자 발현 수준을 조절하고 세포의 기능을 결정하는 분자 조절 물질의 집합체로 나타낼 수 있다. 유전자 네트워크의 상호작용에는 활성화와 억제가 있는데 그림 1에서 G1 → G2는 활성화를 나타내고 G3 → G6은 억제를 나타낸다. 이러한 방법으로 영향을 주는 조절 노드와 영향을 받는 타겟 노드의 관계를 설명하는 방향성 있는 네트워크로서 유전자 간의 복잡한 상호작용을 나타낸다.

유전자 간 상호작용은 질병의 유발 원인을 규명하는 데 도움을 줄 수 있다. 세포 내 유전자 수는 매우 많기 때문에 유전자들은 서로 복잡하게 얽혀 있다. 따라서 하나의 유전자의 변이 또는 상호작용에 기반하여 질병의 메커니즘을 이해하기에는 한계가

있다. 그렇기 때문에 유전자 집단 사이의 상호작용을 고려하여 생물학적 메커니즘을 이해하는 것이 필요하다. 유전자 간 상호작용은 세포의 기능과 다양한 생물학적 반응을 조절하는 중요한 메커니즘으로 특정 유전자 또는 유전자 집단의 조절이 암과 같은 복잡한 질병의 진행과 치료 반응에 영향을 미칠 수 있다 [1]. 유전자 네트워크 분석은 질병 메커니즘 이해의 핵심 도구로서 질병의 원인을 찾아내고 치료 타겟을 설정하는데에 중요한 역할을 한다.

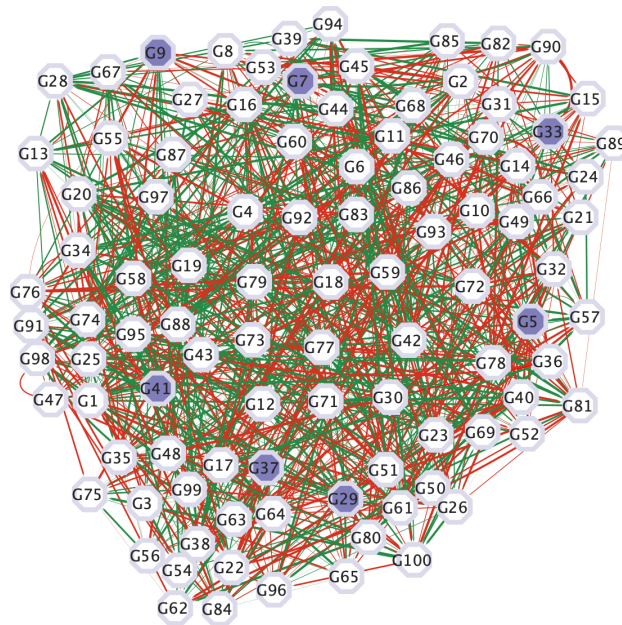


그림 2: 대규모 유전자 네트워크의 예시

그러나 기존 유전자 네트워크 연구에는 한계점이 있다. 기존 연구는 대규모 네트워크의 크기가 너무 크고 복잡하여 해석하기 어렵다는 단점이 있다. 그림 2과 같이 수천 개 이상의 유전자가 포함된 네트워크에서는 중요한 상호작용이나 핵심 유전자를 육안으로 찾기 어렵다. 그렇기 때문에 기존 연구에서는 주로 이전에 알려진 biomarker에만 의존하여 유전자 네트워크를 해석하였다. 이 방법은 중요한 생물학적 정보를 놓칠 가능성이 있고 전체적인 생물학적 메커니즘을 이해하는데 한계를 가진다. 따라

서 대규모 유전자 네트워크에서 유의미한 정보를 추출하고 해석하기 위한 효율적인 방법이 필요하다.

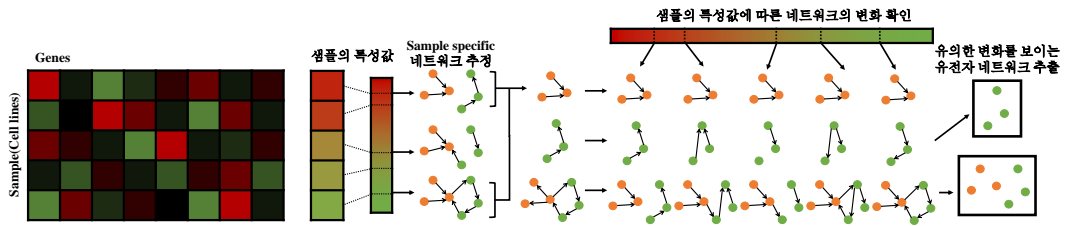


그림 3: 제안한 분석 프로세스의 overview

본 연구에서는 질병과 관련된 샘플의 특징에 따라 유의하게 변화하는 유전자 네트워크 추출하고 해석하기 위한 통계적 방법론을 제안한다. 제안한 통계적 방법론의 분석 프로세스는 그림 3와 같다. 유전자 발현량 데이터와 샘플의 특성값 데이터를 통해 sample specific 네트워크를 추정한다. 추정한 네트워크에서 검정 대상을 선정하여 샘플의 특성값에 따라 어떻게 변화하는지 확인한다. 검정 대상 중에서 유의한 변화를 보이는 유전자 네트워크를 추출하여 해석한다. 이러한 프로세스를 통해 추출된 네트워크는 질병과 관련된 샘플의 특성값을 설명하는데 중요한 분자 상호작용으로 고려된다.

II. 유전자 네트워크 추정

1. 유전자 네트워크 추정

R_1, \dots, R_q 를 k 번째 타겟 유전자(Target gene) T_k 의 전사를 조절하는 q 개의 조절 유전자들(Regulator genes) 이라고 가정한다. 본 연구에서는 타겟 유전자가 조절 유전자로부터 얼마나 영향을 받는지 추정하기 위해 선형 회귀 모형을 고려하였다. 다음은 타겟 유전자 T_k 를 반응변수로 두고 조절 유전자 R_j ($j = 1, \dots, q$)를 설명변수로 하는 선형 회귀 모형이다.

$$T_k = \sum_{j=1}^q \beta_{jk} \cdot R_j + \epsilon_k. \quad (2.1)$$

여기서 T_k 는 타겟 유전자 k 의 발현량을, R_j 는 조절 유전자 j 의 발현량을 의미한다. β_{jk} 는 R_j 가 T_k 에 미치는 영향력의 크기를 나타내는 회귀계수이다. ϵ_k 는 평균은 0이고 분산은 σ^2 인 무작위 오차이며 서로 독립이면서 동일한 분포를 가진다.

선형 회귀 모형의 β_{jk} 를 추정하기 위하여 다음과 같은 elastic net[2] 추정법을 고려하였다.

$$L_k(\beta_{1k}, \dots, \beta_{qk}) = \frac{1}{2} \sum_{i=1}^n \left\{ t_{ik} - \sum_{j=1}^q \beta_{jk} \cdot r_{ij} \right\}^2 + P_{\delta\lambda}(\beta_{jk}), \quad (2.2)$$
$$P_{\delta\lambda}(\beta_{jk}) = \lambda \sum_{j=1}^q \left\{ \frac{1}{2}(1 - \delta)\beta_{jk}^2 + \delta|\beta_{jk}| \right\}.$$

여기서 λ 는 모형의 복잡도를 제어하는 정규화 파라미터이며 패널티항의 영향력을 조절한다. δ 는 L_1 유형의 패널티(lasso[3])와 L_2 유형의 패널티(ridge[4])의 조합의 비율을 조절한다. δ 의 값이 커지면 L_1 유형의 패널티의 비중이 커지고 반대로 δ 의 값이

작아지면 L_2 유형의 패널티의 비중이 커진다. Elastic net은 L_1 유형의 패널티와 L_2 유형의 패널티로 구성된 패널티항을 적용한 모형 추정 방법으로서 에지 선택과 네트워크 추정을 동시에 수행할 수 있다. 샘플의 수보다 유전자의 수가 많은 경우에서 L_1 유형의 패널티와 L_2 유형의 패널티를 동시에 사용함으로써 L_1 유형의 패널티만 사용했을 때보다 더 많은 수의 에지를 선택할 수 있게 된다는 장점이 있다. 또한 elastic net의 grouping effect[5]를 통해 상관관계가 높은 유전자들이 비슷한 가중치를 가지도록 할 수 있다. 이러한 이유로 고차원 데이터인 유전자 발현량 데이터에 기초한 네트워크 추정의 유용한 방법으로 많이 사용되고 있다.

그러나 elastic net을 이용한 네트워크 추정 방법은 sample specific 유전자 네트워크를 추정하는 것은 불가능하다. Elastic net은 샘플마다 서로 다른 생물학적 메커니즘을 가지고 있는 경우에도 평균적인 네트워크 추정 결과만을 제공하기 때문이다.

2. Sample specific 유전자 네트워크 추정

본 연구에서는 sample specific 유전자 네트워크를 추정하는 방법으로 다음의 varying coefficient structural equation model[6]을 고려하였다.

$$T_k = \sum_{j=1}^q \beta_{jk}(m_\alpha) \cdot R_j + \epsilon_k. \quad (2.3)$$

m_α 는 α 번째 샘플의 특성값을 나타내는 모듈레이터 값(e.g., 항암제 감수성)을 의미한다. 식(2.3)은 α 번째 샘플의 varying coefficient structural equation model를 나타낸다. $\beta_{jk}(m_\alpha)$ 는 α 번째 샘플에서 R_j 가 T_k 에 미치는 영향력의 크기를 나타내는 회귀계수로 모듈레이터 m_α 에 따라 달라진다. 식(2.3)은 다음과 같이 나타낼 수 있다.

$$t_{ik} = \sum_{j=1}^q \beta_{jk\alpha} \cdot r_{ij} + \epsilon_{ik} \quad (i = 1, \dots, n). \quad (2.4)$$

여기서 $\beta_{jk\alpha} = \beta_{jk}(m_\alpha)$ 이며 이는 α 번째 샘플의 특성값에 따라 달라지는 회귀계수를 의미한다. m_α 에 상응하는 회귀계수를 구하기 위하여 다음과 같은 kernel based L_1 유형 패널티 정규화 기법을 고려하였다.

$$L_k(\beta_{1k\alpha}, \dots, \beta_{qk\alpha} | h_k) = \frac{1}{2} \sum_{i=1}^n \left\{ t_{ik} - \sum_{j=1}^q \beta_{jk\alpha} \cdot r_{ij} \right\}^2 K(m_\alpha - m_i | h_k) \quad (2.5)$$

$$+ \lambda_{k\alpha} \sum_{j=1}^q |\beta_{jk\alpha}| + \frac{\gamma_{k\alpha}}{2} \sum_{j=1}^q \beta_{jk\alpha}^2,$$

$$K(m_\alpha - m_i | h_k) = \exp \left\{ -\frac{1}{h_k} (m_i - m_\alpha)^2 \right\}.$$

m_i 는 i 번째 샘플의 특성값을 나타내는 모듈레이터 값이다. $\lambda_{k\alpha}$ 와 $\gamma_{k\alpha}$ 는 각각 L_1 유형 패널티와 L_2 유형 패널티를 조절하는 파라미터이다. α 번째 샘플에 대한 유전자 네트워크를 추정하기 위해 gaussian kernel 함수 $K(m_\alpha - m_i | h_k)$ 를 사용한다. Gaussian kernel 함수는 m_α 와 비슷한 샘플별 특성값을 가지는 샘플에는 큰 가중치를 주고 비슷하지 않은 샘플별 특성값을 가지는 샘플에는 작은 가중치를 주도록 조절하는 함수이다. 여기서 h_k 는 gaussian kernel 함수의 폭을 조절하는 bandwidth 파라미터이다.

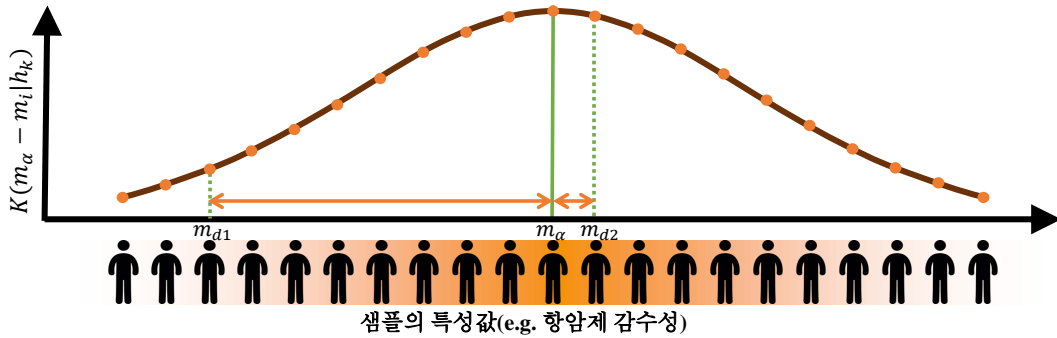


그림 4: Gaussian kernel 함수를 통한 sample specific 네트워크 추정

그림 4을 통해 α 번째 샘플과 근접한 샘플일수록 큰 가중치를 가지게 되는 것을

확인할 수 있다. m_{d2} 와 같이 α 번째 샘플과 거리가 가까운 샘플인 경우 m_α 와 m_i 의 값이 비슷하게 되어 큰 가중치를 가지게 된다. 이 때 α 번째 샘플과 거리가 가까운 샘플을 α 번째 샘플의 이웃 샘플이라고 표현할 수 있다. 반대로 m_{d1} 와 같이 α 번째 샘플과 거리가 먼 샘플의 경우에는 m_α 와 m_i 의 값의 차이가 많이 나게 되어 작은 가중치를 가지게 된다.

α 번째 샘플과 거리가 아주 멀게 되면 가중치의 값은 0에 수렴하게 된다. 이웃 샘플이 아닌 경우에는 가중치가 0이 되고 이웃 샘플의 경우에는 가중치가 0 이상의 값이 된다. 이웃 샘플의 범위를 결정하는 파라미터가 바로 h_k 이다. h_k 의 크기가 커지면 α 번째 샘플의 이웃의 수가 증가하게 되고 h_k 의 크기가 작아지면 이웃 샘플의 수가 감소하게 된다.

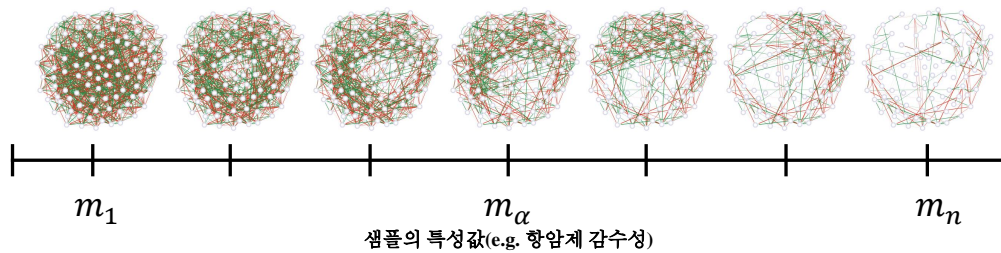


그림 5: 대량 대규모 sample specific 유전자 네트워크의 예시

그림 5에서와 같이 네트워크처럼 샘플의 특성값이 달라짐에 따라 큰 변화를 가지는 유전자 네트워크를 추출할 수 있다. 그림 5의 m_1 에 대한 네트워크는 가장 복잡한 네트워크를 보이고 있고, m_n 에 대한 네트워크는 가장 단순한 네트워크를 보이고 있다. 이는 해당 유전자들이 샘플의 특성값에 따라 서로 다른 분자 상호작용을 가지고 있다는 것을 알 수 있다.

그러나 이러한 sample specific 유전자 네트워크는 대량 대규모 네트워크 해석의 문제를 가지고 있다. 이 네트워크들은 규모가 크고 복잡한 네트워크들이기 때문에

육안으로는 중요한 유전자 간 상호작용을 찾기 어렵다. 특히, 대량 대규모 네트워크 해석에서는 샘플의 특성에 따른 유전자 네트워크를 서로 비교한 후 중요 상호작용을 찾기 어렵다는 문제가 존재한다. 이러한 문제를 해결하기 위하여 대규모 유전자 네트워크에서 유의미한 정보를 추출하고 해석하기 위한 효율적인 방법이 필요하다.

III. 추정된 유전자의 네트워크 해석

대량 대규모 네트워크의 해석을 위한 아이디어로 본 연구는 샘플의 특성에 따라 특징적 상호작용을 가지고 있어 유의미하게 변화하는 유전자 네트워크를 추출하는 통계적 방법론을 제안하였다. 특징적 상호작용을 가진 유전자 네트워크는 각각의 샘플의 특성값마다 서로 다른 네트워크의 특징을 가지고 있다.

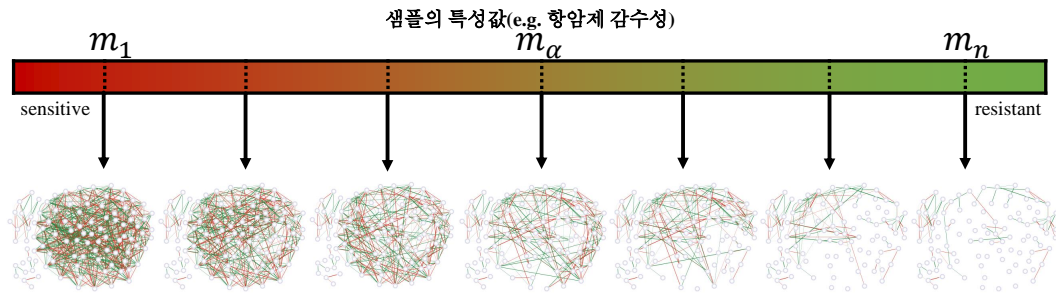


그림 6: 샘플의 특성에 따라 유전자 네트워크 추출의 예시

그림 6에서 m_1 , m_α , m_n 에 대하여 해당 샘플의 특성값에 따라 추출된 유전자 네트워크의 특징을 표현하는 통계량을 통해 샘플의 특성값에 따라 유의미하게 달라지는 유전자 네트워크를 추출할 수 있다.

본 연구에서는 추정된 대량의 대규모 sample specific 유전자 네트워크를 해석하기 위하여 샘플의 특성값(i.e., 모듈레이터 값)의 변화에 따라 유의한 차이를 보이는 네트워크 추출을 위한 통계적 방법론을 제안하였다. 네트워크의 특징을 표현하는 통계량으로는 Regulatory Effect와 Jaccard Index를 고려하였다. 종합적인 정보에 기초하여 네트워크의 특징을 표현하기 위하여 두 통계량을 결합한 Regulatory Effect with Jaccard Index를 제안하였다.

1. Regulatory Effect

Regulatory Effect[6]는 α 번째 샘플에서 조절 유전자 j 가 타겟 유전자 k 에 미치는 영향을 계산한 값으로 식으로 표현하면 다음과 같다.

$$RE_{jk\alpha} = \sum_{l \in \pi_{jk\alpha}} \hat{\beta}_l^{(j \rightarrow k)}(m_\alpha) \cdot r_{\alpha j}. \quad (3.1)$$

여기서 $\alpha (= 1, \dots, n)$ 는 샘플의 번호를 의미하고 $\pi_{jk\alpha}$ 는 j 번째 유전자로부터 k 번째 유전자에 도달할 수 있는 모든 경로를 말한다. l 은 $\pi_{jk\alpha}$ 에 속하는 경로의 인덱스이다. $\hat{\beta}_l^{(j \rightarrow k)}(m_\alpha)$ 는 α 번째 샘플에서의 k 번째 타겟 유전자에 대한 j 번째 조절 유전자의 회귀계수 추정량을 뜻한다. $r_{\alpha j}$ 는 α 번째 샘플의 j 번째 조절 유전자의 발현량을 의미한다.

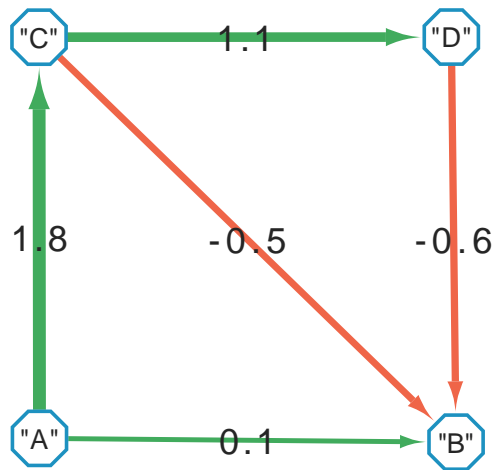


그림 7: 유전자 네트워크의 예시

그림 7에서 주어진 유전자 네트워크에서 조절 유전자 A 가 타겟 유전자 B 에 주는

영향은 다음과 같이 표현할 수 있다.

$$\begin{aligned}\pi_{AB\alpha} &= \{R_A \rightarrow T_B, R_A \rightarrow R_C \rightarrow T_B, R_A \rightarrow R_C \rightarrow R_D \rightarrow T_B\}, \\ RE_{AB\alpha} &= (\hat{\beta}_{AB\alpha} + \hat{\beta}_{AC\alpha} \cdot \hat{\beta}_{CB\alpha} + \hat{\beta}_{AC\alpha} \cdot \hat{\beta}_{CD\alpha} \cdot \hat{\beta}_{DB\alpha}) \cdot r_{\alpha A}, \\ RE_{AB\alpha} &= (0.1 + 1.8 \cdot -0.5 + 1.8 \cdot 1.1 \cdot -0.6) \cdot r_{\alpha A} = -1.988r_{\alpha A}.\end{aligned}$$

본 연구에서는 $\pi_{jk\alpha}$ 의 길이를 1로 단순화하여 계산하였다. 설정한 조건에 따르면 그림 7에서 $\pi_{AB\alpha}$ 에 포함되는 경로는 $\{R_A \rightarrow T_B\}$ 뿐이다. 즉 Regulatory Effect는 다음과 같이 표현된다.

$$RE_{jk\alpha} = \hat{\beta}_{jk\alpha} \cdot r_{\alpha j}. \quad (3.2)$$

전체 유전자 네트워크에서 하나로 연결된 네트워크를 subnetwork라고 한다. 각각의 subnetwork 안에는 조절-타겟 관계를 가지는 유전자들이 존재한다. Subnetwork 단위에서의 Regulatory Effect 값은 다음과 같이 표현된다.

$$sbRE_{S\alpha} = \sum_{j \in V_{S\alpha}} \sum_{k \in V_{S\alpha}} RE_{jk\alpha}. \quad (3.3)$$

S 는 subnetwork에 부여된 번호를 의미하고 $V_{S\alpha}$ 는 α 번째 샘플에서 subnetwork S 에 포함된 유전자 집합을 의미한다. $sbRE_{S\alpha}$ 는 subnetwork S 에 속한 모든 조절 유전자와 모든 타겟 유전자에 대한 Regulatory Effect이다. Subnetwork의 크기가 커지면 $sbRE_{S\alpha}$ 의 값이 커지고 반대의 경우에는 $sbRE_{S\alpha}$ 의 값이 작아진다. 이러한 부분을 보완하기 위하여 다음과 같은 scaled $sbRE_{S\alpha}$ 를 제안하였다.

$$(scale)sbRE_{S\alpha} = \frac{\sum_{j \in V_{S\alpha}} \sum_{k \in V_{S\alpha}} RE_{jk\alpha}}{|V_{S\alpha}|}. \quad (3.4)$$

여기서 $|V_{S\alpha}|$ 는 α 번째 샘플에서 subnetwork S 에 포함된 유전자의 수를 의미한다.

2. Jaccard Index

Jaccard Index[7]는 두 네트워크의 유사도를 나타내는 지표 중 하나이며 다음과 같이 나타낼 수 있다.

$$\text{sim}_{JI}(G_1, G_2) = \frac{|G_1 \cap G_2|}{|G_1 \cup G_2|}. \quad (3.5)$$

여기서 G_1 과 G_2 는 서로 다른 두 네트워크를 의미한다. Jaccard Index를 계산하기 위하여 노드나 에지 등의 네트워크 특징을 기준으로 합집합과 교집합을 구한다. 분모 $|G_1 \cup G_2|$ 는 네트워크 특징의 합집합의 원소 갯수를 의미하고 분자 $|G_1 \cap G_2|$ 는 네트워크 특징의 교집합의 원소 갯수를 의미한다. 본 연구에서는 Jaccard Index를 계산할 때 네트워크의 특징을 에지 목록을 기준으로 지정하여 에지들의 합집합과 교집합을 계산하였다. 이 때 에지의 방향성이 서로 다르면 다른 에지로 간주하고 목록을 작성한다. Jaccard Index는 두 네트워크의 에지 구조가 동일하다면 1, 완전히 다르다면 0으로 계산된다.

검정 대상이 항암제에 가장 민감한 샘플인 sensitive cell line인 경우 Jaccard Index를 다음과 같이 정의한다.

$$\text{sim}_{JI}(G_S^{\text{sen}}, G_S^{m_\alpha}) = \frac{|E_S^{\text{sen}} \cap E_S^{m_\alpha}|}{|E_S^{\text{sen}} \cup E_S^{m_\alpha}|}.$$

G_S^{sen} 는 sensitive cell line에서의 subnetwork S 이다. $G_S^{m_\alpha}$ 는 m_α 를 샘플의 특성값으로 가지는 샘플에서의 subnetwork S 이다. E_S^{sen} 는 sensitive cell line에서의 subnetwork S 의 에지 집합을 의미한다. $E_S^{m_\alpha}$ 는 m_α 를 샘플의 특성값으로 가지는 샘플에서의 subnetwork S 의 에지 집합을 나타낸다. 이를 통해 sensitive cell line과 m_α 를 샘플의 특성값으로 가지는 샘플을 비교하여 두 네트워크의 유사도를 정량적으로 평가할 수 있다. 샘플의 특성값 m_α 의 변화량에 따라 Jaccard Index의 값이 많이 변화할수록 샘플의 특성값에

따라 유의하게 변화하는 네트워크를 추출할 수 있다.

검정 대상이 항암제에 가장 둔한 샘플인 resistant cell line에서 추출된 subnetwork인 경우에는 다음과 같이 Jaccard Index를 계산한다.

$$sim_{JI}(G_S^{res}, G_S^{m\alpha}) = \frac{|E_S^{res} \cap E_S^{m\alpha}|}{|E_S^{res} \cup E_S^{m\alpha}|}.$$

G_S^{res} 는 resistant cell line에서의 subnetwork \mathcal{S} 이다. E_S^{res} 는 resistant cell line에서의 subnetwork \mathcal{S} 의 에지이다. 이와 같은 경우에는 G_S^{sen} 대신에 G_S^{res} 를 적용하여 동일한 절차를 진행한다.

3. Regulatory Effect with Jaccard Index

본 연구에서는 종합적인 정보에 기초해 유전자 네트워크의 특성을 표현하기 위하여 Regulatory Effect와 Jaccard Index를 결합한 Regulatory Effect with Jaccard Index를 제안하였다. 제안한 통계량 Regulatory Effect with Jaccard Index는 유전자 간의 상호작용과 네트워크 에지 구조의 변화까지 동시에 고려한다. 검정 대상이 sensitive cell line에서 추출된 subnetwork인 경우에는 다음과 같이 Regulatory Effect with Jaccard Index를 계산한다.

$$RE.JI_{S\alpha}^{sen} = sim_{JI}(G_S^{sen}, G_S^{m\alpha}) \cdot \frac{\sum_{j \in V_{S\alpha}} \sum_{k \in V_{S\alpha}} RE_{jk\alpha}}{|V_{S\alpha}|}. \quad (3.6)$$

검정 대상이 resistant cell line에서 추출된 subnetwork인 경우에는 G_S^{sen} 대신에 G_S^{res} 를 적용하여 동일한 절차를 진행한다.

4. Change of Regulatory Effect with Jaccard Index

각 subnetwork의 샘플의 특성값에 따라 유의하게 변화하는 네트워크를 추출하기 위하여 다음과 같은 Change of Regulatory Effect with Jaccard Index을 제안하였다. Change of Regulatory Effect with Jaccard Index는 주어진 샘플 중 최대 Regulatory Effect with Jaccard Index 값에서 최소 Regulatory Effect with Jaccard index 값의 차로 정의한다. 이는 네트워크의 특성을 나타내는 Regulatory Effect with Jaccard Index의 변화량을 나타낸다. 샘플의 특성값에 따라 sample-specific 네트워크의 에지 구조와 조절 유전자의 영향력이 얼마나 변화하는지 정량적으로 보여주는 통계량이다. 검정 대상이 sensitive cell line에서 추출된 subnetwork일 때 다음과 같이 Change of Regulatory Effect with Jaccard Index를 계산한다.

$$CRJ_S^{sen} = \max \{RE.JI_{S\alpha}^{sen}; \alpha = 1, \dots, n\} - \min \{RE.JI_{S\alpha}^{sen}; \alpha = 1, \dots, n\} \quad (3.7)$$

검정 대상이 resistant cell line에서 추출된 subnetwork인 경우에는 *sen* 대신에 *res*를 사용하여 Change of Regulatory Effect with Jaccard Index를 계산한다.

5. 샘플의 특성값에 따른 네트워크 변화의 유의성 검정

Change of Regulatory Effect with Jaccard Index에 기초하여 샘플의 특성값에 따라 유의하게 변화하는 유전자 네트워크 추출을 위해 제안된 통계량의 유의성 검정을 위한 순열 검정법을 제안하였다. 해당 순열 검정법은 샘플의 특성값에 따라 네트워크의 변화가 통계적으로 유의미한지 검정할 수 있다. 제안한 순열 검정법은 다음과 같이 세 가지 절차를 통해 이루어진다.

1. 원 데이터에 기초한 통계량 계산: 기존 데이터의 유전자 발현량과 모듈레이터 값을 사용하여 유전자 네트워크를 추정한다. 이후 CRJ_S^{sen} 와 CRJ_S^{res} 값을 계산

한다.

2. 순열 통계량 계산: 원 데이터의 모듈레이터 값 범위 내의 균일 분포를 통해 순열 모듈레이터 데이터를 생성한다. 순열 모듈레이터 데이터와 원 데이터의 유전자 발현량 데이터에 기초하여 순열 모듈레이터 값에 따른 네트워크를 추정한다. 여기서 추정된 네트워크는 균일 분포로부터 랜덤하게 생성된 순열 모듈레이터 값에 기초하여 추정되었다. 그렇기 때문에 샘플의 특성에 따른 정보를 포함하지 않으므로 샘플의 특성에 따른 분자 상호작용을 나타내지 못한다. 각 순열 모듈레이터 데이터에 대해 $CRJ_{S(v)}^{sen}$ 의 값과 $CRJ_{S(v)}^{res}$ 의 값을 계산한다. 이 과정을 총 $v = 1, \dots, \Upsilon$ 번 반복하여 충분히 많은 순열 데이터를 생성한다.
3. p-value 계산: 검정 대상이 sensitive cell line에서 추출된 subnetwork인 경우일 때 다음과 같이 p-value를 계산한다.

$$p_S^{sen} = \frac{\sum_{v=1}^{\Upsilon} I\left(|CRJ_{S(v)}^{sen}| \geq |CRJ_S^{sen}|\right)}{\Upsilon} \quad (3.8)$$

여기서 Υ 는 순열의 횟수를 나타내며 $I(\cdot)$ 는 조건이 참일 때 1, 거짓일 때 0을 반환하는 지시 함수이다. $CRJ_{S(v)}^{sen}$ 는 subnetwork S 에 대한 v 번째 순열 모듈레이터 데이터의 Change of Regulatory Effect with Jaccard Index 값이며 CRJ_S^{sen} 는 원 데이터의 Change of Regulatory Effect with Jaccard Index 값이다. Subnetwork S 가 resistant cell line에서 추출된 경우에는 위 과정에서 sen 대신 res 를 사용하여 동일한 절차를 수행한다.

본 연구에서는 p_S^{sen} 의 값이 0.01보다 작을 때 CRJ_S^{sen} 의 값이 유의하다고 판단하였다. 이는 일반적으로 많이 사용되는 유의수준 1%를 적용한 것으로 p-value의 값이 유의수준보다 작은 경우 해당 subnetwork의 변화가 통계적으로 유의미함을 나타낸다.

IV. 급성 골수성 백혈병 항암제 midostaurin 감수성을 특징 짓는 유전자 네트워크 분석

1. 자료 및 평가 방법

본 연구에서 사용한 유전자 네트워크 추정 방법과 제안하였던 추정된 유전자의 네트워크 해석 방법을 실제 데이터에 적용하였다. 샘플의 특징 중에서 항암제 감수성에 대한 값을 모듈레이터로 사용하였다. 분석 대상 데이터는 Genomics of Drug sensitivity in Cancer (GDSC) 데이터베이스(www.cancerrxgene.org)에서 제공하는 midostaurin에 대한 IC50 값과 유전자 발현량 데이터이다.

Midostaurin은 급성 골수성 백혈병(Acute Myeloid Leukemia, AML)의 항암제로 FLT3 유전자 변이가 있는 급성 골수성 백혈병 성인 환자 치료에 사용되는 다중 키나제 억제제(multitarget kinase inhibitor)이다. 급성 골수성 백혈병 외에도 공격성 전신 비만 세포증, 관련 혈액 종양을 동반한 전신 비만세포증, 비만 세포 백혈병 등의 치료에도 사용된다. IC50는 주어진 생물학적 과정 또는 생물학적 성분을 50% 억제하는데 필요한 특정 약물의 양을 나타내는 정량적 측정값이다. IC50 값이 클수록 약물에 둔감하여 저항성이 큰 샘플임을 의미하며 값이 작을수록 약물에 민감한 샘플임을 의미한다.

IC50 데이터는 970개의 cell line의 midostaurin에 대한 IC50 값을 가지고 있다. 유전자 발현량 데이터는 1018개의 cell line에 대한 17419개의 유전자 발현량 값을 가지고 있다. 17419개의 유전자 중 발현량의 분산 값이 높은 상위 1000개의 유전자와 IC50 데이터와 유전자 발현량 데이터에서 서로 중첩하는 831개의 cell line을 분석에 사용하였다.

본 연구에서는 midostaurin에 대한 IC50 값을 샘플의 특징을 나타내는 모듈레이터로 사용하여 IC50 값 변화에 따른 네트워크의 변화를 분석하였다. 분석을 수행하기

위하여 총 831 개의 cell line 중 IC50 값을 기준으로 백분위수를 적용하여 총 20 개의 cell line에 대한 샘플별 유전자 네트워크를 추정하였다. 추정된 네트워크의 에지에서 절대값 상위 1%인 에지로 구성된 네트워크를 분석 대상으로 사용하였다. 검정 대상은 모듈레이터 값이 가장 작은(항암제에 가장 민감한) cell line의 유전자 네트워크와 가장 큰(항암제에 가장 저항적인) cell line의 유전자 네트워크의 subnetwork들을 추출하였다. 이후 sensitive cell line과 resistant cell line에서 추출한 subnetwork에 대한 CRJ_S^{sen} 과 CRJ_S^{res} 값을 계산하였다.

순열 검정을 수행하기 위해 원 데이터의 모듈레이터 값의 범위에서 균일분포를 이용해 831 개의 순열 모듈레이터 값을 생성하였다. 앞서 사용한 방법과 동일하게 순열 모듈레이터 데이터에서 백분위수를 적용하여 20 개의 샘플을 선정하였다. 이후 샘플별 유전자 네트워크를 추정하였고 각 subnetwork에 대한 $CRJ_{S(\gamma)}^{sen}$ ($\gamma = 1, \dots, 200$) 과 $CRJ_{S(\gamma)}^{res}$ ($\gamma = 1, \dots, 200$) 값을 계산하였다.

2. 적용 결과

Subnetwork	p-value < 0.01	0.01 ≤ p-value
Total	46	17
Sensitive cell line에서 추출한 subnetwork	26	8
Resistant cell line에서 추출한 subnetwork	20	9

표 1: Subnetwork 별 p-value 분포 결과

검정 대상은 총 63 개의 subnetwork로 sensitive cell line에서는 34 개, resistant cell line에서는 29 개의 subnetwork를 추출하였다. 본 연구에서는 순열 검정법의 p-value의 값이 0.01보다 작을 때 해당 검정 대상의 변화가 유의하다고 판단했다. 전체 63 개의 subnetwork 중 46 개의 subnetwork가 추출되었다. 이 중 sensitive cell line의 네트워크는 26 개, resistant cell line의 네트워크는 20 개가 추출되었다.

1) Midostaurin resistant specific 유전자 네트워크

유의미한 네트워크 중에서 resistant cell line에서 추출한 subnetwork 중 Change of Regulatory Effect with Jaccard Index의 값이 가장 큰 subnetwork 5개를 선정하였다. 선정된 네트워크의 상호작용을 resistant cell line과 sensitive cell line에서 살펴보면 다음과 같은 결과를 나타낸다.

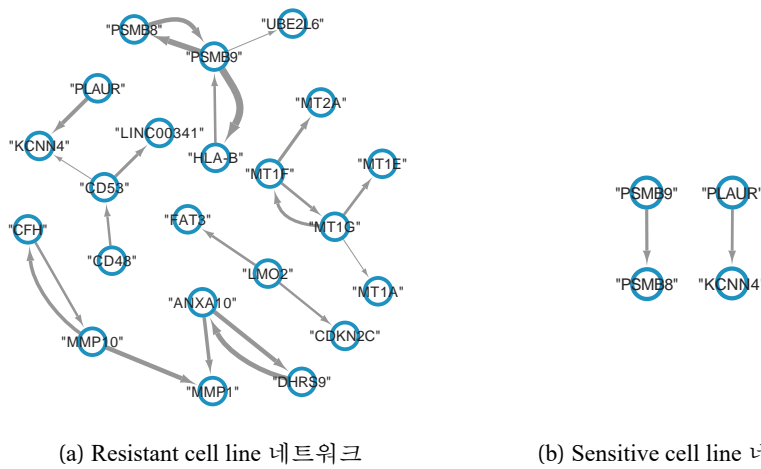


그림 8: Resistant cell line에서 추출한 네트워크 비교

Resistant cell line 네트워크와 sensitive cell line 네트워크를 비교한 결과, resistant cell line 네트워크에서는 존재하지만 sensitive cell line 네트워크에는 나타나지 않는 특정 유전자 간 상호작용을 확인하였다. Resistant cell line을 특징 짓는 유전자 간 상호작용으로는 $PSMB8 \rightarrow PSMB9$, $PSMB9 \rightarrow HLA-B$, $PSMB9 \rightarrow UBE2L6$, $HLA-B \rightarrow PSMB9$, $CD53 \rightarrow LINC00341$, $CD48 \rightarrow CD53$, $CFH \rightarrow MMP10$, $ANXA10 \rightarrow DHRS9$, $MMP10 \rightarrow MMP1$, $ANXA10 \rightarrow MMP1$ 등이 있다.

또한, resistant cell line 네트워크에서 관찰된 유전자 중 sensitive cell line 네트워크에는 존재하지 않는 유전자도 존재하였다. Resistant cell line specific biomarker로는 $MT1A$, $MT1G$, $MT1F$, $MT2A$, $CD53$, $MMP10$, $MMP1$, $FAT3$, $DHRS9$, $ANXA10$ 등이 확인

되었다. 이러한 상호작용과 유전자들은 midostaurin에 가장 저항적인 상태에서 나타나는 네트워크의 특징을 반영한다. 이는 본 연구를 통해 새롭게 밝혀진 resistant specific 분자 상호작용으로 이해될 수 있다.

2) Midostaurin sensitive specific 유전자 네트워크

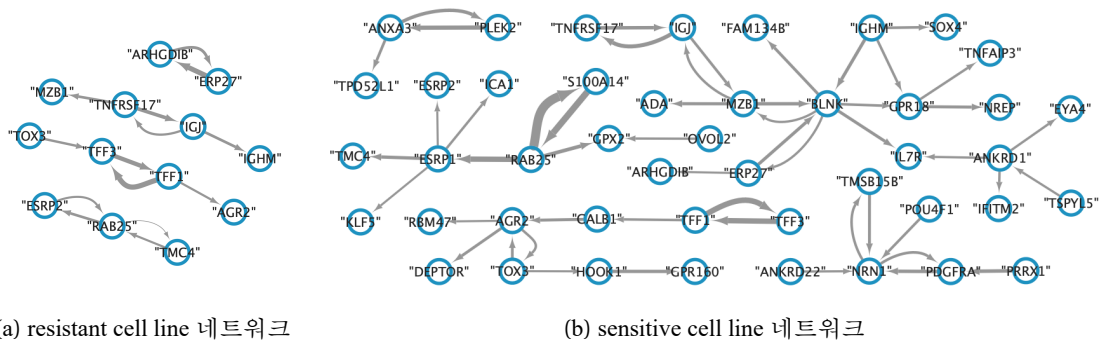


그림 9: Sensitive cell line에서 추출한 네트워크 비교

그림 9(b)는 유의한 sensitive cell line의 subnetworks 26개 중 Change of Regulatory Effect with Jaccard index의 값이 가장 큰 5개의 subnetwork를 나타낸다. 해당 subnetwork의 resistant cell line에서의 모습은 그림 9(a)에 나타나 있다.

Sensitive cell line 네트워크와 resistant cell line 네트워크를 비교한 결과, sensitive cell line을 특징 짓는 유전자 간 상호작용으로는 $ANXA3 \rightarrow PLEK2$, $ANXA3 \rightarrow TPD52L1$, $PLEK2 \rightarrow ANXA3$, $ESRP1 \rightarrow ICA1$, $ESRP1 \rightarrow ESRP2$, $ESRP1 \rightarrow TMC4$, $ESRP1 \rightarrow KLF5$, $S100A14 \rightarrow RAB25$, $OVOL2 \rightarrow GPX2$, $IGJ \rightarrow MZB1$ 등이 있다.

또한, sensitive cell line specific biomarker로는 $ANXA3$, $PLEK2$, $TPD52L1$, $ICA1$, $ESRP1$, $KLF5$, $RBM47$, $BLNK$, $ANKRD1$, $NRN1$ 등이 확인되었다. 이러한 상호작용과 유전자들은 midostaurin에 가장 민감한 상태에서 나타나는 네트워크의 특징으로 볼 수 있다. 이 또한 sensitive specific 분자 상호작용으로 이해될 수 있다.

Gene	AML	Midostaurin	Gene	AML	Midostaurin
<i>ANXA10</i>	[8]	-	<i>CD48</i>	-	[9]
<i>CD53</i>	[10]	-	<i>CDKN2C</i>	[11], [12]	-
<i>CFH</i>	[13]	-	<i>DHRS9</i>	[14]	[15]
<i>FAT3</i>	[16]	-	<i>HLA-B</i>	[17]	-
<i>LINC00341</i>	[18]	-	<i>LMO2</i>	[19]	-
<i>MMP1</i>	[20]	[21]	<i>MMP10</i>	[20]	[21]
<i>MT1A</i>	[22]	-	<i>MT1E</i>	[23]	-
<i>MT1F</i>	[22]	-	<i>MT1G</i>	[22]	-
<i>MT2A</i>	[23]	[24]	<i>UBE2L6</i>	[25]	[26]

표 2: Midostaurin resistant specific biomarker와 관련된 문헌

Gene	AML	Midostaurin	Gene	AML	Midostaurin
<i>ADA</i>	[27], [28]	-	<i>ANKRD1</i>	[29]	-
<i>ANKRD22</i>	[30]	-	<i>ANXA3</i>	[8], [31], [32]	-
<i>BLNK</i>	[33]	-	<i>CALB1</i>	[34]	-
<i>DEPTOR</i>	[35]	[36]	<i>EYA4</i>	[37]	-
<i>ESRP1</i>	[38]	-	<i>FAM134B</i>	[39]	-
<i>GPR160</i>	[40]	-	<i>GPR18</i>	[41]	-
<i>GPX2</i>	[42]	-	<i>HOOK1</i>	[43]	-
<i>ICAI</i>	-	[44]	<i>IFITM2</i>	[45]	-
<i>IL7R</i>	[46], [47]	[48]	<i>KLF5</i>	-	[49]
<i>NRN1</i>	[50]	-	<i>NREP</i>	-	[51]
<i>OVOL2</i>	[52]	-	<i>PDGFRA</i>	[53]	-
<i>PLEK2</i>	[54]	-	<i>POU4F1</i>	[55], [56]	[57]
<i>PRRX1</i>	[58]	[59]	<i>RBM47</i>	[60]	-
<i>S100A14</i>	[61]	-	<i>SOX4</i>	[62]	[63]
<i>TMSB15B</i>	[64]	-	<i>TNEAIP3</i>	[65]	[66]
<i>TPD52L1</i>	[67]	-	<i>TSPYL5</i>	[68]	-

표 3: Midostaurin sensitive specific biomarker와 관련된 문헌

표 2와 3는 각각 resistant cell line과 sensitive cell line에서 추출된 네트워크를 특징 짓는 유전자에 대한 급성 골수성 백혈병과 midostaurin에 대한 문헌 연구를 나타내고 있다. 해당 문헌 연구들은 본 연구에서 추출된 유전자들이 급성 골수성 백혈병과 midostaurin과 관련된 중요한 biomarker라는 사실을 강하게 뒷받침해주고 있다.

3) 추출된 유전자 네트워크의 생물학적 경로 분석 결과

본 연구에서 추출된 네트워크의 생물학적 의미를 파악하기 위해서 생물학적 경로 분석(Biological pathway analysis)을 실시하였다. 생물학적 경로 분석을 수행하기 위하여 DAVID (Database for Annotation, Visualization and Integrated Discovery) 데이터베이스에 기초하여 GO term (Gene Ontology term) 분석을 실시하였다. DAVID는 gene-enrichment analysis에 특화된 웹 기반 유전자 주석이 가능한 bioinformatics 도구로서 GO term, KEGG pathway 분석 등 array 데이터의 기능적 해석을 가능하게 한다. 이를 통하여 resistant와 sensitive cell line의 GO term 그래프를 통하여 생물학적 경로를 확인하였다.

① Midostaurin resistant specific 유전자 네트워크

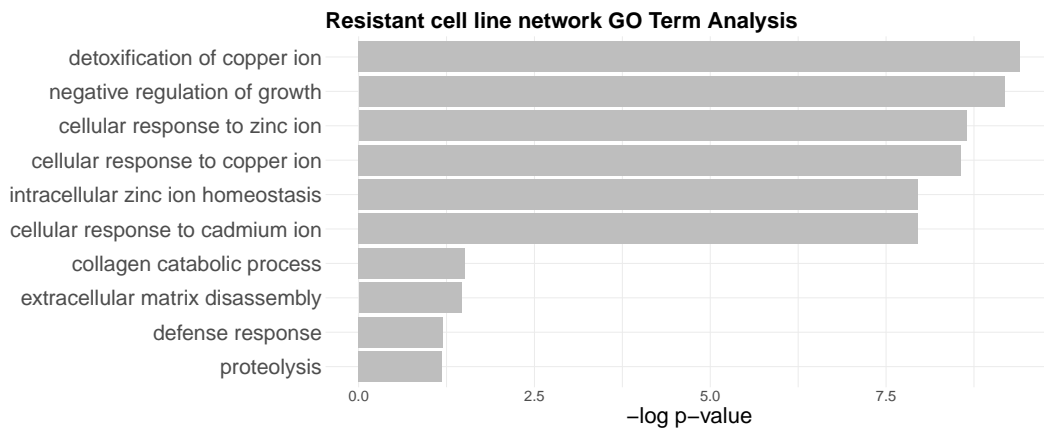


그림 10: Resistant cell line specific 유전자 네트워크의 GO term 그래프

Resistant cell line에서 DAVID GO Term 분석에서 구리 이온 해독(Detoxification of copper ion) 경로가 가장 높은 유의성을 보였다. 그 다음 순서로 유의한 생물학적 경로는 세포 성장 억제(Negative regulation of growth) 경로였다. 아연, 구리, 카드뮴 이온에

대한 세포 반응 (Cellular response to zinc, copper, cadmium ion) 경로도 유의한 생물학적 경로로 확인되었다.

② **Midostaurin sensitive specific** 유전자 네트워크

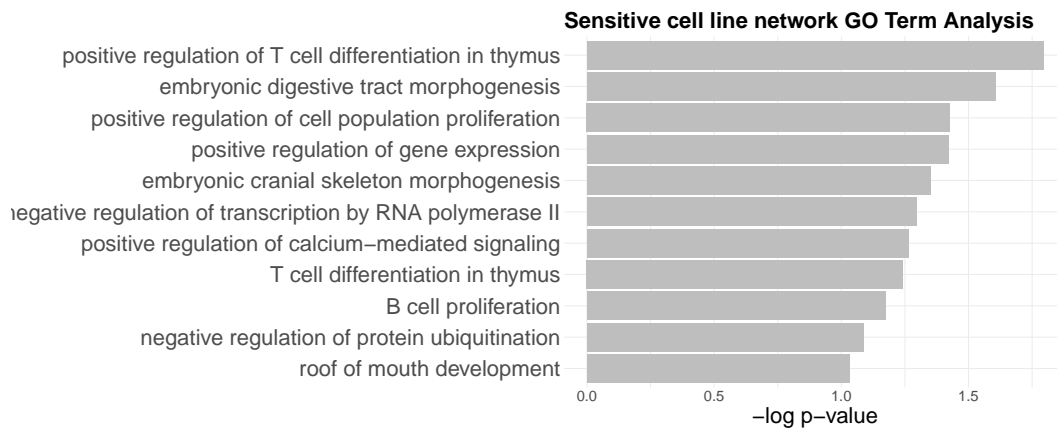


그림 11: Sensitive cell linespecific 유전자 네트워크의 GO term 그래프

Sensitive cell line에서 DAVID GO Term 분석 결과, 흉선 내 T 세포 분화의 양성 조절 (Positive regulation of T cell differentiation in thymus) 경로가 가장 높은 유의성을 나타냈다. 다음으로는 배아 소화관 형태형성 (Embryonic digestive tract morphogenesis) 경로가 유의한 생물학적 경로로 확인되었다. 세포 집단 증식의 활성화 (Positive regulation of cell population proliferation) 경로도 유의한 생물학적 경로임을 보였다.

V. 결론

본 학위 논문은 대량 대규모 네트워크 해석을 위한 샘플 특성을 특징 짓는 유전자 네트워크를 추출하는 것을 목표로 하였다. 이를 위하여 sample specific 네트워크를 추출하고 해석하기 위한 통계적 방법론을 제안하였다. 제안한 통계적 방법론은 특징적 상호작용을 보이는 네트워크를 찾기 위한 새로운 통계량 Change of Regulatory Effect with Jaccard Index와 대량 대규모 네트워크의 변화량을 검정하기 위한 새로운 순열 검정법으로 구성되어 있다. 제안한 통계적 방법론을 통해 샘플의 특성값에 따라 특징적인 분자 상호작용을 보이는 유전자 네트워크를 추출하는데에 기여하였다.

제안한 통계적 방법론은 실제 데이터를 활용한 분석에 적용되었다. 본 연구에서는 샘플의 특성을 항암제 midostaurin에 대한 감수성으로 설정하였고 resistant와 sensitive cell line을 특징 짓는 분자 상호작용 추출을 시도하였다. 데이터 분석 결과, 샘플의 특성값이 달라짐에 따라 동일한 유전자 목록을 공유하였으나 에지 구조에서 차이가 있다는 것을 확인하였다. 이를 통해 샘플의 특성값이 달라지면 변화하는 분자 상호작용들이 있다는 것을 알 수 있었다. 또한, 분석을 통해 추출된 분자 상호작용의 유전자들이 문헌 연구를 통해 급성 골수성 백혈병과 midostaurin과 관련이 있는 biomarker라는 것을 확인하였다. 이는 제안한 방법론이 생물학적으로 신뢰할 수 있는 결과를 제공한다는 것을 시사한다.

이러한 분석 과정을 거쳐 본 연구는 항암제 감수성이 변화하는 환경에서 분자 상호작용의 변화를 이해할 수 있다. 이를 통해 정밀 의료와 맞춤형 의료 분야에서 새로운 치료 타겟과 biomarker 발굴 가능성을 제시하며 의의를 가진다. 그러나 제안한 방법론의 성능 평가가 이루어지지 않았다는 점을 한계점으로 들 수 있다. 이러한 부분을 보완하기 위하여 향후 연구 과제로 Monte Carlo Simulation을 사용한 성능 평가와 개선이 필요하다.

참고문헌

- [1] Ballestar E, Esteller M. Epigenetic gene regulation in cancer. *Advances in genetics*. 2008;61:247–267.
- [2] Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 2005;67(2):301–320.
- [3] Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 1996;58(1):267–288.
- [4] Hoerl AE, Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*. 1970;12(1):55–67.
- [5] Zhou DX. On grouping effect of elastic net. *Statistics & Probability Letters*. 2013;83(9):2108–2112.
- [6] Shimamura T, Imoto S, Shimada Y, Hosono Y, Niida A, Nagasaki M, et al. A Novel Network Profiling Analysis Reveals System Changes in Epithelial-Mesenchymal Transition. *PLOS ONE*. 2011;6(6):1–17.
- [7] Jaccard P. Étude comparative de la distribution florale dans une portion des Alpes et du Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles*. 1901;37(142):547.
- [8] Niu Y, Yang X, Chen Y, Jin X, Xie Y, Tang Y, et al. Distinct prognostic values of Annexin family members expression in acute myeloid leukemia. *Clinical and Translational Oncology*. 2019;21:1186–1196.
- [9] Zhang T, Fang Q, Liu P, Wang P, Feng C, Wang J. Heme oxygenase 1 overexpres-

- sion induces immune evasion of acute myeloid leukemia against natural killer cells by inhibiting CD48. *Journal of Translational Medicine*. 2022;20(1):394.
- [10] Chen C, Chen Z, Chio CL, Zhao Y, Li Y, Liu Z, et al. Higher expression of WT1 with lower CD58 expression may be biomarkers for risk stratification of patients with cytogenetically normal acute myeloid leukemia. *Technology in Cancer Research & Treatment*. 2021;20:15330338211052152.
- [11] Lu-Qun W, Hao L, Xiang-Xin L, Fang-Lin L, Ling-Ling W, Xue-Liang C, et al. A case of simultaneous occurrence of acute myeloid leukemia and multiple myeloma. *BMC cancer*. 2015;15:1–6.
- [12] Schmalzbauer BS, Thondanpallil T, Heller G, Schirripa A, Sperl CM, Mayer IM, et al. CDK6 degradation is counteracted by p16INK4A and p18INK4C in AML. *Cancers*. 2022;14(6):1554.
- [13] Fratelli M, Bolis M, Kurosaki M, Dori M, Guarnaccia V, Spinelli O, et al. Association of CFHR1 homozygous deletion with acute myelogenous leukemia in the European population. *Leukemia & Lymphoma*. 2016;57(5):1234–1237.
- [14] Meester-Smoor MA, Janssen MJ, Grosveld GC, de Klein A, van IJcken WF, Douben H, et al. MN1 affects expression of genes involved in hematopoiesis and can enhance as well as inhibit RAR/RXR-induced gene expression. *Carcinogenesis*. 2008;29(10):2025–2034.
- [15] Kong D, Guo H. Construction and validation of a prognostic model for overall survival time of patients with ovarian cancer by metabolism-related genes. *Journal of Obstetrics and Gynaecology Research*. 2024;50(9):1622–1639.
- [16] Dehagi MO, Rostami S, Shamshiri A, Safari F, Hosseini RH, Thorne RF, et al. FAT1

- Gene Expression in Iranian Acute Lymphoid and Myeloid Leukemia Patients. *International Journal of Hematology-Oncology and Stem Cell Research*. 2023;17(2):81.
- [17] Hallner A, Bernson E, Hussein BA, Ewald Sander F, Brune M, Aurelius J, et al. The HLA-B- 21 dimorphism impacts on NK cell education and clinical outcome of immunotherapy in acute myeloid leukemia. *Blood, The Journal of the American Society of Hematology*. 2019;133(13):1479–1488.
- [18] Ou R, Huang J, Shen H, Liu Z, Zhu Y, Zhong Q, et al. Transcriptome analysis of CD34+ cells from myelodysplastic syndrome patients. *Leukemia Research*. 2017;62:40–50.
- [19] Lu L, Wang J, Fang F, Guo A, Jiang S, Tao Y, et al. LMO2 promotes the development of AML through interaction with transcription co-regulator LDB1. *Cell Death & Disease*. 2023;14(8):518.
- [20] J Hatfield K, Reikvam H, Bruserud O. The crosstalk between the matrix metalloprotease system and the chemokine network in acute myeloid leukemia. *Current medicinal chemistry*. 2010;17(36):4448–4461.
- [21] Richards CD. The enigmatic cytokine oncostatin m and roles in disease. *International Scholarly Research Notices*. 2013;2013(1):512103.
- [22] Xin X, Xu Z, Wei J, Zhang Y. MiR-376a-3p increases cell apoptosis in acute myeloid leukemia by targeting MT1X. *Cancer Biology & Therapy*. 2022;23(1):234–242.
- [23] Pan YQ, Niu M, Liu Sm, Bao YX, Yang K, Ma XB, et al. Effect of MT2A on apoptosis and proliferation in HL60 cells. *International Journal of Medical Sciences*. 2021;18(13):2910.
- [24] Grignano E, Cantero-Aguilar L, Tuerdi Z, Chabane T, Vazquez R, Johnson N, et al.

- Dihydroartemisinin-induced ferroptosis in acute myeloid leukemia: links to iron metabolism and metallothionein. *Cell Death Discovery*. 2023;9(1):97.
- [25] Orfali N, Shan-Krauer D, O' Donovan TR, Mongan NP, Gudas LJ, Cahill MR, et al. Inhibition of UBE2L6 attenuates ISGylation and impedes ATRA-induced differentiation of leukemic cells. *Molecular oncology*. 2020;14(6):1297–1309.
- [26] Cai J, Huang H, Hu X, Lang W, Fu W, Xu L, et al. Homoharringtonine Synergized with Gilteritinib Results in the Downregulation of Myeloid Cell Leukemia-1 by Upregulating UBE2L6 in FLT3-ITD-Mutant Acute Myeloid (Leukemia) Cell Lines. *Journal of Oncology*. 2021;2021(1):3766428.
- [27] Ratech H, Martiniuk F, Borer WZ, Rappaport H. Differential expression of adenosine deaminase isozymes in acute leukemia. *Blood*. 1988;72(5):1627–1632.
- [28] Morisaki T, Fujii H, Miwa S. Adenosine deaminase (ADA) in leukemia: clinical value of plasma ADA activity and characterization of leukemic cell ADA. *American journal of hematology*. 1985;19(1):37–45.
- [29] Chang F, Kong SJ, Wang L, Choi BK, Lee H, Kim C, et al. Targeting actomyosin contractility suppresses malignant phenotypes of acute myeloid leukemia cells. *International Journal of Molecular Sciences*. 2020;21(10):3460.
- [30] Li J, Chen N, Gong X. Prognostic implications of aberrantly expressed methylation-driven genes in hepatocellular carcinoma: A study based on The Cancer Genome Atlas. *Molecular Medicine Reports*. 2019;20(6):5304–5314.
- [31] Chen T, Zhang J, Wang Y, Zhou H. Identification of Survival-Related Genes in Acute Myeloid Leukemia (AML) Based on Cytogenetically Normal AML Sam-

- ples Using Weighted Gene Coexpression Network Analysis. *Disease Markers*. 2022;2022(1):5423694.
- [32] Handschuh L, Kaźmierczak M, Milewski MC, Góralski M, Łuczak M, Wojtaszewska M, et al. Gene expression profiling of acute myeloid leukemia samples from adult patients with AML-M1 and-M2 through boutique microarrays, real-time PCR and droplet digital PCR. *International journal of oncology*. 2018;52(3):656–678.
- [33] Silva FP, Swagemakers SM, Erpelinck-Verschueren C, Wouters BJ, Delwel R, Vrieling H, et al. Gene expression profiling of minimally differentiated acute myeloid leukemia: M0 is a distinct entity subdivided by RUNX1 mutation status. *Blood, The Journal of the American Society of Hematology*. 2009;114(14):3001–3007.
- [34] Ghannam G, Takeda A, Camarata T, Moore MA, Viale A, Yaseen NR. The oncogene Nup98-HOXA9 induces gene transcription in myeloid cells. *Journal of Biological Chemistry*. 2004;279(2):866–875.
- [35] Morales-Martinez M, Lichtenstein A, Vega MI. Function of Deptor and its roles in hematological malignancies. *Aging (Albany NY)*. 2021;13(1):1528.
- [36] Leifheit ME, Johnson G, Kuzel TM, Schneider JR, Barker E, Yun HD, et al. Enhancing Therapeutic Efficacy of FLT3 Inhibitors with Combination Therapy for Treatment of Acute Myeloid Leukemia. *International Journal of Molecular Sciences*. 2024;25(17):9448.
- [37] Huang S, Jiang MM, Chen GF, Qian K, Gao HH, Guan W, et al. Epigenetic silencing of eyes absent 4 gene by acute myeloid leukemia 1-eight-twenty-one oncoprotein contributes to leukemogenesis in t (8; 21) acute myeloid leukemia. *Chinese medical journal*. 2016;129(11):1355–1362.

- [38] He C, Chen Y, Zhang X, Feng H, Rao Y, Ji T, et al. Down-regulation of ESRP2 inhibits breast cancer cell proliferation via inhibiting cyclinD1. *Scientific Reports*. 2024;14(1):28475.
- [39] Chen W, Mao H, Chen L, Li L. The pivotal role of FAM134B in selective ER-phagy and diseases. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*. 2022;1869(8):119277.
- [40] Archer KJ, Fu H, Mrózek K, Nicolet D, Mims AS, Uy GL, et al. Improving risk stratification for 2022 European LeukemiaNet favorable-risk patients with acute myeloid leukemia. *The Innovation*. 2024;5(6).
- [41] Li J, Su H, Zhu Y, Cao Y, Ma X. ETS2 and microRNA-155 regulate the pathogenesis of heart failure through targeting and regulating GPR18 expression. *Experimental and therapeutic medicine*. 2020;19(6):3469–3478.
- [42] Zhang J, Peng Y, He Y, Xiao Y, Wang Q, Zhao Y, et al. GPX1-associated prognostic signature predicts poor survival in patients with acute myeloid leukemia and involves in immunosuppression. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*. 2022;1868(1):166268.
- [43] Xu Z, Guo C, Ye Q, Shi Y, Sun Y, Zhang J, et al. Endothelial deletion of SHP2 suppresses tumor angiogenesis and promotes vascular normalization. *Nature communications*. 2021;12(1):6310.
- [44] Park SM, Miyamoto DK, Han GY, Chan M, Curnutt NM, Tran NL, et al. Dual IKZF2 and CK1 α degrader targets acute myeloid leukemia cells. *Cancer Cell*. 2023;41(4):726–739.
- [45] Liu Y, Lu R, Cui W, Pang Y, Liu C, Cui L, et al. High IFITM3 expression predicts

- adverse prognosis in acute myeloid leukemia. *Cancer Gene Therapy*. 2020;27(1):38–44.
- [46] Tao Q, Wu Q, Xue Y, Chen C, Zhou Y, Shao R, et al. Prognostic impact of IL7R mutations on acute myeloid leukemia. *Therapeutic Advances in Hematology*. 2024;15:20406207241279533.
- [47] Xu N, Sun K, Wang YZ, Chen WM, Wang J, Li LD, et al. Low IL7R Expression at Diagnosis Predicted Relapse in Adult Acute Myeloid Leukemia Patients With t (8; 21). *Frontiers in Immunology*. 2022;13:909104.
- [48] Shimada A. Hematological malignancies and molecular targeting therapy. *European journal of pharmacology*. 2019;862:172641.
- [49] Wu J, Xiao Y, Sun J, Sun H, Chen H, Zhu Y, et al. A single-cell survey of cellular hierarchy in acute myeloid leukemia. *Journal of hematology & oncology*. 2020;13:1–19.
- [50] Liu H, Chen P, Yang YL, Zhu KW, Wang T, Tang L, et al. TBC1D16 predicts chemosensitivity and prognosis in adult acute myeloid leukemia (AML) patients. *European Journal of Pharmacology*. 2021;895:173894.
- [51] Zhang Z, Huang J, Zhang Z, Shen H, Tang X, Wu D, et al. Application of omics in the diagnosis, prognosis, and treatment of acute myeloid leukemia. *Biomarker Research*. 2024;12(1):60.
- [52] Hsu WY, Chiou SS, Lin PC, Liao YM, Yeh CY, Tseng YH. Prediction of miRNA-mRNA network regulating the migration ability of cytarabine-resistant HL60 cells. *Biomedical Reports*. 2023;20(2):20.

- [53] Metzgeroth G, Walz C, Score J, Siebert R, Schnittger S, Haferlach C, et al. Recurrent finding of the FIP1L1-PDGFR α fusion gene in eosinophilia-associated acute myeloid leukemia and lymphoblastic T-cell lymphoma. *Leukemia*. 2007;21(6):1183–1188.
- [54] Han X, Mei Y, Mishra RK, Bi H, Jain AD, Schiltz GE, et al. Targeting pleckstrin-2/Akt signaling reduces proliferation in myeloproliferative neoplasm models. *The Journal of Clinical Investigation*. 2023;133(6).
- [55] Fortier JM, Payton JE, Cahan P, Ley TJ, Walter MJ, Graubert TA. POU4F1 is associated with t (8; 21) acute myeloid leukemia and contributes directly to its unique transcriptional signature. *Leukemia*. 2010;24(5):950–957.
- [56] Dunne J, Gascoyne DM, Lister TA, Brady HJ, Heidenreich O, Young BD. AML1/ETO proteins control POU4F1/BRN3A expression and function in t (8; 21) acute myeloid leukemia. *Cancer research*. 2010;70(10):3985–3995.
- [57] Zhai Y, Singh P, Dolnik A, Brazda P, Atlasy N, Del Gaudio N, et al. Longitudinal single-cell transcriptomics reveals distinct patterns of recurrence in acute myeloid leukemia. *Molecular Cancer*. 2022;21(1):166.
- [58] Gervais C, Mauvieux L, Perrusson N, Helias C, Struski S, Leymarie V, et al. A new translocation t (9; 11)(q34; p15) fuses NUP98 to a novel homeobox partner gene, PRRX2, in a therapy-related acute myeloid leukemia. *Leukemia*. 2005;19(1):145–148.
- [59] Rasouli M, Troester S, Grebien F, Goemans BF, Zwaan CM, Heidenreich O. NUP98 oncofusions in myeloid malignancies: An update on molecular mechanisms and therapeutic opportunities. *HemaSphere*. 2024;8(9):e70013.
- [60] Zhang B, Yang L, Wang X, Fu D. Identification of survival-related alternative splicing signatures in acute myeloid leukemia. *Bioscience Reports*. 2021;41(7):BSR20204037.

- [61] Yang XY, Jin J, Huang J, Li P, Xue JW, Wu XJ, et al. Expression and clinical significance profile analysis of S100 family members in human acute myeloid leukemia. *European Review for Medical & Pharmacological Sciences*. 2020;24(13).
- [62] Lu J, Hsieh M, Hou H, Chen C, Tien H, Lin L. Overexpression of SOX4 correlates with poor prognosis of acute myeloid leukemia and is leukemogenic in zebrafish. *Blood Cancer Journal*. 2017;7(8):e593–e593.
- [63] Murray HC, Miller K, Dun MD, Verrills NM. Pharmaco-phosphoproteomic analysis of cancer-associated KIT mutations D816V and V560G. *Proteomics*. 2024;24(9):2300309.
- [64] Wang X, Björklund S, Wasik AM, Grandien A, Andersson P, Kimby E, et al. Gene expression profiling and chromatin immunoprecipitation identify DBN1, SETMAR and HIG2 as direct targets of SOX11 in mantle cell lymphoma. *PloS one*. 2010;5(11):e14085.
- [65] Lin Ms, Zhong HY, Yim RLH, Chen QY, Du Hl, He Hq, et al. Pan-cancer analysis of oncogenic TNFAIP2 identifying its prognostic value and immunological function in acute myeloid leukemia. *BMC cancer*. 2022;22(1):1068.
- [66] Reynolds C, Stein C, Atkinson T, Hurst A, Kimberlin D. P267 XMEN disease: an unexpected presentation of a rare primary immunodeficiency. *Annals of Allergy, Asthma & Immunology*. 2017;119(5):S67–S68.
- [67] Ha M, Han ME, Kim JY, Jeong DC, Oh SO, Kim YH. Prognostic role of TPD52 in acute myeloid leukemia: A retrospective multicohort analysis. *Journal of Cellular Biochemistry*. 2019;120(3):3672–3678.
- [68] Hernández-Sánchez JM, del Rey M, Benito R, Chakraborty S, Rodríguez-Vicente AE,

- González T, et al. Sequential Mutational and Gene Expression Profile shows marked changes in CD34 positive cells from MDS patients during 5-Azacytidine treatment. *Therapy-induced changes in the gene expression profile of patients with hematological diseases*. 2019; p. 87.
- [69] Jaccard P. THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE. *New Phytologist*. 1912;11(2):37–50.
- [70] Real R, Vargas JM. The Probabilistic Basis of Jaccard's Index of Similarity. *Systematic Biology*. 1996;45(3):380–385.
- [71] Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic acids research*. 2012;41(D1):D955–D961.

ABSTRACT

Gene network analysis to identify sample-specific molecular interactions

Haemin Yong

Department of Statistics

Graduate School of

Sungshin Women's University

This study aimed to identify differentially regulated gene networks under varying conditions of a characteristic of cell lines, especially anti-cancer drug resistance specific molecular interplays identification. We consider anti-cancer drug of Acute Myeloid Leukemia (i.e., midostaurin) and its sensitivity was used to a characteristic of cell lines. We then estimate the midostaurin sensitivity-specific gene networks and performed differential gene network analysis. To identify differentially regulated gene networks under varying conditions of the midostaurin sensitivity, we developed a novel statistic to describe the gene networks based on not only regulatory effects between genes, but also node structures. Thus, we can effectively describe the gene networks based on comprehensive information. We also developed a statistical test to access significance of the change of the gene regulatory system. In our strategy, the permuted cell line characteristic is generated from uniform distribution and then we estimate permuted cell line characteristic specific gene networks based on the randomly gen-

erated modulator. For the permutation gene networks, the developed statistic is computed. By comparing the permutation and ordinary statistics, we compute p-value to measure the significance of the change in networks. We applied our strategy to identify differentially regulated gene network under varying condition of midostaurin sensitivity. Our strategy revealed that the midostaurin sensitive and resistant gene networks show considerably difference gene regulatory systems. In particular, the sensitive specific gene network displayed more dense and complex network structures, while the molecular interplays become weaker and were disappeared in the resistant specific network. The identified midostaurin sensitive and resistant-specific molecular interplays can be considered as potential markers to understand mechanism of AML and of acquired midostaurin resistance. We expected that the proposed strategy will be a useful tool to uncover new biomarkers and therapeutic targets for overcoming cancer cell resistance.

Keywords: Molecular interactions, Cell line specific gene network analysis, Midostaurin

감사의 글(Acknowledgements)

2023년 3월, 통계학과 석사 과정을 시작하여 2024년 12월 석사학위 청구 논문을 완성하기까지 많은 분들의 도움과 격려가 있었습니다. 이 자리를 빌려 진심 어린 감사의 말씀을 전합니다.

부족한 저를 포기하지 않고 심혈을 기울여 지도해 주신 박희원 교수님께 깊이 감사드립니다. 제가 하고 싶은 연구 주제를 먼저 물어봐 주시고, 가지고 온 주제로 학위 논문을 쓸 수 있도록 이끌어 주신 덕분에 여기까지 올 수 있었습니다. 또한, 앞으로의 삶에서도 큰 지침이 될 귀중한 조언들로 저를 학문적 성장은 물론 인격적 성장의 길로도 인도해 주셨습니다. 교수님께서 주신 격려와 지지는 제가 더 나은 사람이 되는 데 큰 원동력이 되었습니다. 교수님의 가르침을 잊지 않고 마음 깊이 새기며, 앞으로 통계학 분야의 일원으로서 책임을 다하며 학문적 기여를 할 수 있는 사람이 되겠습니다.

학사 과정에서 석사 과정까지 통계학의 길을 함께해 주신 성신여자대학교 통계학과 교수님께 감사를 드립니다. 첫 통계 수업에서 만나 뵈었던 박만식 교수님께 감사의 마음을 표현하고 싶습니다. 통계 수학 및 연습 수업을 통해 많은 것을 배웠습니다. 특히, 홀로 공부하는 법뿐만 아니라 함께 공부하는 법을 배우며 성장할 수 있었습니다.

기초 통계학으로 처음 만나 뵈었던 정호현 교수님께 감사드립니다. 질문을 들고 연구실을 방문할 때마다 반갑게 맞이해 주시고, 함께 문제를 고민하며 눈높이에 맞는 가르침을 주셨기에 통계학의 기초를 다질 수 있었습니다.

회귀 분석 수업으로 처음 만나 뵈었던 이성건 교수님께 진심 어린 감사를 드립니다. 회귀 분석의 기초 개념을 배우며 통계학의 매력을 느꼈고, 이를 더 깊이 탐구하고자 하는 열망으로 석사 과정 진학을 결심하는 계기가 되었습니다.

인공지능 통계 분석 강의를 통해 처음 만나 뵈었던 김동하 교수님께 큰 감사를 드립니다. 교수님의 강의를 통해 통계와 인공지능의 접점을 이해할 수 있었으며, 이를

통해 새로운 통찰을 얻을 수 있었습니다.

조교 업무를 먼저 제안해 주셨던 박성오 교수님께도 깊은 감사를 드립니다. 프로젝트 참여와 다양한 연구를 접할 수 있는 기회를 주셔서 학업과 업무를 병행하며 성장할 수 있는 귀중한 경험을 쌓을 수 있었습니다.

선형 모형 수업으로 만나 뵈었던 박관영 교수님께도 감사의 마음을 전합니다. 교수님의 열정 넘치고 세심한 강의 덕분에 선형 모형을 공부하며 더 깊게 이해할 수 있었습니다.

생존 분석을 가르쳐 주셨던 최태화 교수님께도 마음 깊이 감사드립니다. 교수님의 강의를 통해 다른 분야의 의학 통계에 대한 지식을 접하고 생각을 넓힐 수 있었습니다.

학위 청구 논문 심사를 맡아 주신 신준호 교수님께 감사드립니다. 세심한 피드백 덕분에 논문을 한층 더 발전시켜 무사히 졸업의 문턱을 넘을 수 있었습니다.

대학원 생활 동안 함께해 주신 세리 언니, 지우 언니, 윤아, 지혜, 서연, 서영, 수지, 윤진, 나경, 주이, 진주, 경민, 세영, 민서 등 모든 대학원 식구들에게도 진심 어린 감사를 전합니다. 여러분 덕분에 힘든 순간도 웃으며 이겨낼 수 있었고, 소중한 추억을 많이 남길 수 있었습니다.

무엇보다도, 늘 곁에서 든든한 버팀목이 되어 주신 가족들께 깊이 감사드립니다. 아빠와 엄마는 항상 저를 믿어 주시고 학업과 연구에 집중할 수 있도록 격려해 주셨으며, 할머니와 오빠는 늘 따뜻한 응원으로 제게 큰 위로가 되었습니다.

또한, 김소연, 김소은, 김지혜, 김채림, 김하은, 김혜원, 박진희, 박현지, 박혜인, 배현아, 여소은, 우지혜, 이민주, 이서희, 이혜리, 정수아, 채수빈, 한지원을 포함한 모든 친구들께도 감사를 드립니다. 늘 곁에서 응원하고 격려해 주신 여러분 덕분에 어려운 시기도 잘 헤쳐 나갈 수 있었습니다.

이 모든 분들의 도움과 격려 덕분에 논문을 완성할 수 있었습니다. 보이지 않는 곳에서 도와주신 분들께 감사의 마음을 전합니다. 앞으로도 끊임없이 연구하고 발전하는 모습으로 보답하겠습니다. 다시 한 번 모든 분들께 깊은 감사의 뜻을 전합니다.