

다빈도 다표적 한약 성분의 Enrichment Test 분석의 영향: TCMSP를 활용한 염증 기전 사례 연구

장호 선임연구원

한국한의학연구원 한의약데이터부

Analyzing the impact of high-frequency multi-target herbal medicine compounds on enrichment test: a case study on inflammation response using TCMSP

Jang Ho

KM Data Division, Korea Institute of Oriental Medicine

Abstract

The increasing application of network pharmacological analysis offers a novel perspective for understanding the complex interactions between herbal compounds and their multiple protein targets. However, despite the importance of these analyses, the influence of high-frequency, multi-target compounds on network pharmacological analysis is not clear. This study investigates the impact of these prevalent multi-target compounds on network pharmacological analysis. To this end, we conducted a case study focusing on the analysis of the inflammation response, utilizing enrichment analysis results derived from the TCMSP database. We confirmed that compounds such as Kaempferol and Quercetin, which are part of various herbal medicines, may be highly relevant to inflammatory responses, as indicated by the enrichment test results. Our research suggests that the network pharmacological results of traditional medicine, utilizing public data, can be influenced by compounds with high-frequency, multi-target characteristics.

Correspondence: 장호(Jang Ho)

1672 Yuseong-daero, Yuseong-gu, Daejeon, 34054, Rep. of Korea

Tel: +82-10-8641-6311, E-mail: jh@kiom.re.kr

Received 2023-11-03, revised 2023-11-17, accepted 2023-11-23, available online 2023-11-27

doi:10.22674/KHMI-11-2-7



Keywords: Network pharmacology, Korean medicine, TCM, database, herb, formula

서론

한약 처방(Formulas), 약재(Medicinal plants), 성분(Compounds)의 분석을 위해 최근 네트워크 약리학(Network pharmacology)이 널리 활용되고 있다. 네트워크 약리학은 다성분(Multi compound) 다표적(Multi target) 특성을 띠는 한약과 질병과의 관계를 *in silico* 기반으로 탐색할 수 있는 수단을 제공한다. 2023년 11월 1일 기준으로 PubMed에서 'tcm AND network pharmacology'로 키워드 검색을 하면, 총 1,599개의 문헌이 검색되고, 출간되는 논문은 매해 증가하고 있다.

네트워크 약리학적 분석 대상이 되는 한약은 한의 처방, 약재이다. 때때로 분석 대상은 처방과 약재에 포함된 단일 성분이 되기도 한다. 만약 처방이나 약재가 대상일 경우 대상에 포함되는 구성성분이 무엇인지는 먼저 확인해야 한다. 연구자들은 LC/MS (Liquid chromatography - mass spectrometry)와 같은 실험을 통해 직접 성분 데이터를 생성하거나 인터넷의 공개 데이터베이스에 수록된 약재의 구성성분 정보를 탐색한다. 한약의 구성성분이 확인된 이후, 연구자는 각 성분이 표적(Target)하는 단백질이 무엇인지 확인한다. 이 성분의 표적 단백질들과 질병 관련 표적 단백질들과의 공통 표적(Co-target) 단백질들이 해석(Interpretation)에 활용된다. 연구자는 주로 네트워크 시각화를 통해서 약재, 구성성분, 표적 단백질 사이의 연결 관계를 파악하거나, Enrichment test와 같은 통계적인 기법을 이용해서 생체경로(Biological pathway)와 같은 Functional gene group을 파악하여 한약과 질병사이의 관계를 탐색한다.

처방 또는 약재 선정

↓ 성분 구성 (TCMSp)

Compound	DL	OB
beta-sitosterol	0.75	36.91
kaempferol	0.24	41.88
Stigmasterol	0.76	43.83
...

↓ 표적 구성 (TCMSp)

Compound	Target	score	Symbol
beta-sitosterol	Prostaglandin G/H synthase 1	0.90	PTGS1
kaempferol	Progesterone receptor	0.90	PGR
Stigmasterol	Beta-1 adrenergic receptor	0.83	ADRB1
...

↓ 표적 해석 (DAVID)

Category	Term	RT	Genes	Count	%	P-Value	Benjamini
GOTERM_BP_DIRECT	response to xenobiotic stimulus	RT	17	22.1	9.3E-15	1.6E-11	
GOTERM_BP_DIRECT	response to lipopolysaccharide	RT	11	14.3	3.2E-10	2.6E-7	
GOTERM_BP_DIRECT	response to hypoxia	RT	11	14.3	1.7E-9	9.5E-7	
GOTERM_BP_DIRECT	positive regulation of apoptotic process	RT	13	16.9	6.9E-9	2.9E-6	
	adenylate cyclase						

그림 1. 한약의 네트워크 약리학적 분석 예시

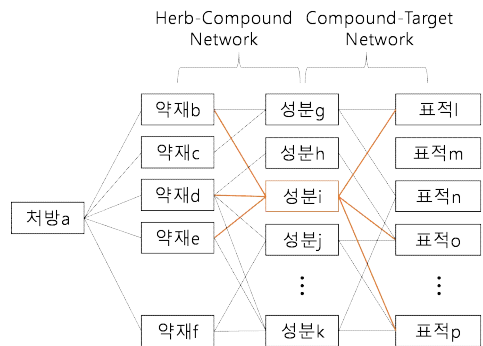


그림 2. 공개데이터의 약재-성분-표적의 결합

<그림 1>은 공개 데이터베이스를 활용한 한약의 네트워크 약리학적 분석의 예시이다. 지금까지



발표된 한약의 네트워크 약리학 연구는 이와같이 분석 대상이 될 한약과 질병의 선정, 한약의 구성 성분의 탐색, 성분이 표적하는 단백질의 탐색, 표적의 해석(Interpretation)의 단계로 진행이 된다.

많은 연구자들이 웹상에 공개된 데이터베이스로부터 구성성분 정보와 성분이 표적 하는 단백질 정보를 가져와서 연구하고 결과를 발표했다. 약재의 구성성분 데이터베이스로는 Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform (TCMSP)¹⁾, Traditional Chinese Medicine Information Database (TCM-ID)²⁾, Database of medicinal materials and chemical compounds in Northeast Asian traditional medicine (TM-MC)³⁾ 등이 있고 성분의 표적 정보를 제공하는 데이터베이스는 TCMSP, STITCH⁴⁾, BATMAN-TCM⁵⁾ 등이 있다. 이처럼 공개 데이터베이스에서 제공하는 구성성분, 표적 정보를 결합(Join)하여 다양한 약재와 다양한 질병의 관계를 탐색할 수 있다. <그림 2>의 예시처럼, 성분정보는 한약 네트워크 약리학에서 한약과 표적 사이를 연결해 주는 핵심이다. 성분들은 특정 약재에서만 등장할 수도 있고 다양한 약재에 등장할 수도 있다. 또한 성분들은 하나의 단백질만 표적 하거나 수많은 단백질을 표적 할 수도 있다. 성분의 이러한 특성으로 인해 한약의 네트워크 약리학적 분석 결과에 편향을 유발될 가능성이 있으나, 이러한 점을 다루고 있는 연구 결과는 보고 되지 않았다. 따라서 본 연구에서는 대표적인 공개 데이터베이스인 TCMSP를 사례로 여러 약재에서 등장하고 또한 다수의 단백질을 표적 하는 일부 성분들이 한약 네트워크 약리학적 분석 결과에 미칠 수 있는 영향을 확인하고자 했다. 구체적으로, TCMSP에 수록된 약재와 염증 반응(Inflammatory response)의 관계를 분석할 때, 몇몇 성분이 Enrichment test의 결과에 미치는 영향을 확인하였고, 문헌조사를 통해서 해당 성분들이 활성 성분으로서 빈번하게 등장함을 있음을 확인했다. 우리는 이를 통해 공개 데이터베이스를 활용한 한약의 네트워크 약리학적 분석에 있어서 일부 성분의 다빈도 대표적 특성이 분석 결과에 미치는 영향에 대해 주의할 필요가 있음을 보이고자 했다.

본 연구는 다음의 구성으로 이루어져 있다. '본론'의 1절에서 연구를 위한 절차 및 방법을 설명했고, 2절에서 5절까지는 연구 결과를 기술했고, 마지막 6절에서는 결과에 대한 저자의 의견을 기술했고, '결론'에서는 연구를 요약했다.

본론

1. 연구 단계 및 방법 소개

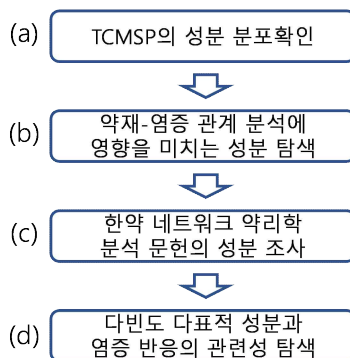


그림 3. 연구의 분석단계



〈그림 3〉은 다빈도 대표적 성분이 네트워크 약리학 분석 결과에 미치는 영향을 확인하기 위한 본연구의 단계를 나타낸다. 먼저 TCMSP의 기초데이터 분석을 수행했다. 구체적으로 성분과 관계 맺는 약제 개수 및 표적 단백질 개수의 분포를 확인했고, 여러 약제에 등장하며 다양한 단백질을 표적하는 성분을 선별했다(〈그림 3〉 (a)).

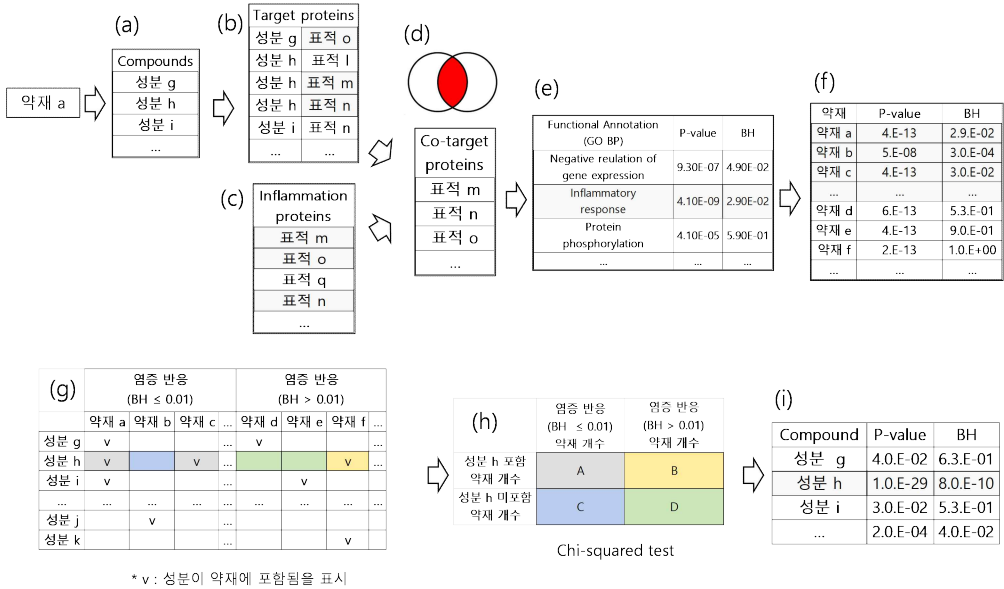


그림 4. 약제와 염증반응의 관련성에 대한 Enrichment test 결과에 영향을 끼친 성분을 탐색하기 위한 절차

이후 우리는 TCMSP의 모든 약제를 대상으로 염증과의 관련성에 대한 간이 네트워크 약리학 분석을 수행했다(〈그림 3〉 (b)). 〈그림 4〉 (a) ~ (f)는 이 분석의 단계를 보여준다. TCMSP의 각각의 약제에 대해, 약제와 연관된 성분을 찾고(〈그림 4〉 (a)), 성분과 연관된 표적 단백질을 찾는다(〈그림 4〉 (b)). 그리고 DisGeNET⁶⁾에서 찾은 염증 관련 단백질(〈그림 4〉 (c))과 약제의 표적 단백질 사이의 교집합인 Co-target을 구한다(〈그림 4〉 (d)). Co-target 단백질을 대상으로 Enrichment test를 수행한다(〈그림 4〉 (e)). 구체적으로 Gene Ontology (GO)⁷⁾의 다양한 Biological process (BP) 중 염증 반응(Inflammatory response)이 해당 약제와의 관계에서 통계적으로 유의하게 나오는지 Fisher's exact test를 수행하여 p-value를 계산하고, Benjamini-Hochberg (BH) procedure를 통해 Adjusted p-value (BH p-value)를 계산한다. 모든 약제에 대해서 이 과정을 통해 염증 반응과의 관련성을 확인했다(〈그림 4〉 (f)).

TCMSP의 수백 가지의 약제에 대해서 위의 과정을 수행해야 하기에 R프로그래밍을 이용했다. Enrichment test는 DAVID⁸⁾에서 제공하는 Enrichment test와 동일한 방법을 직접 구현하여서 결과를 산출했는데, 직접 구현한 Enrichment test 결과와 DAVID에서 제공하는 test 결과에 큰 차이가 없음을 확인했다⁹⁾.

다음으로 우리는 앞의 단계에서 수행한, 약제들과 염증반응 사이의 Enrichment test 결과에 어떤 성분이 주로 기여했는지 탐색했다. 〈그림 4〉 (g) ~ (i)가 해당 분석의 절차를 나타낸다. 각각의 성분



들에 대해서, 염증 반응에 대해 통계적으로 유의하게 나온 약재 그룹(BH p-value ≤ 0.01)과 통계적으로 유의하지 않게 나온 약재 그룹(BH p-value > 0.01)의 두 그룹으로 나누고(〈그림 4〉(g)), 그 성분을 포함한 약재 그룹과 포함하지 않은 약재 그룹으로 나눈 후 분할표(Contingency table)를 만들어서 염증 반응과 성분의 포함 여부 사이의 독립성 테스트(Chi-square test)를 수행한다(〈그림 4〉(h)). 모든 TCMSP성분에 대해서 독립성 테스트의 p-value를 계산하고 BH procedure를 통해 BH p-value를 계산해서 그 값이 0.05 이하인 성분들을 약재와 염증 반응의 관련성 분석 결과에 영향을 끼친 성분들로 보았다(〈그림 4〉(i)).

다음 단계에서는 앞에서 도출한 성분들에 대해 PubMed 문헌 검색을 수행했다(〈그림 3〉(c)). TCMSP 데이터를 활용한 염증 관련 네트워크 약리학 연구 논문들에서 이 성분들이 주요한 활성 성분으로 언급되는 논문들의 빈도수를 확인하였다. 이를 통해서 실제 출간된 연구들에서 다빈도 대표적 성분들이 얼마나 빈번하게 등장하는지 확인하였다.

마지막 단계에서는 다빈도 대표적 성분 각각을 대상으로 네트워크 약리학적 분석을 수행해서 염증 반응과의 관련성을 확인하였다(〈그림 3〉(d)). 〈그림 4〉(b) ~ (f)의 단계를 따라서 개별 성분들을 표적 단백질과 연계한 후 Enrichment test를 수행하여 염증 반응이 통계적으로 유의하게 나오는지 테스트 해 보았다. 이 분석을 통해서 단일한 대표적 성분만으로도 염증 반응과의 관련성이 높게 나올 수 있음을 보였다. 이후의 절에서는 다빈도 대표적 성분이 네트워크 약리학적 분석 결과에 미치는 영향을 앞서 설명한 단계별 결과를 통해 소개한다.

2. TCMSP의 약재-성분, 성분-표적 정보의 분포

TCMSP에는 총 502개의 약재에 대한 13,727개의 성분정보가 존재했다. 〈그림 5〉(a)는 성분별 관계를 맺는 약재 개수의 분포를 보여준다. 최소 3분위수(Quantile)까지의 성분들은 하나의 약재와만 관계를 맺으므로, TCMSP의 대부분 성분은 하나의 약재에만 속한다고 볼 수 있다. 그러나 Beta-sitosterol 같은 성분은 502개 중 245개(48.8%)의 약재에 포함되는 것으로 확인되었다. 〈표 1〉은 다양한 약재에 속하는 상위 10개의 성분을 보여준다. 〈그림 5〉(b)는 성분별 표적 하는 단백질 개수의 분포를 보여준다. 성분은 평균 5개의 단백질을 표적하고 있었다. 가장 많은 단백질을 표적하는 성분은 FMT (Formic acid)로 최대 552개의 단백질 표적과 관련이 있었다. 〈표 2〉는 가장 많은 단백질을 표적하고 있는 상위 10개의 성분을 보여준다. 〈그림 3〉은 TCMSP의 2차원 평면에서의 성분의 분포를 보여준다. 성분이 포함된 약재의 개수를 X축에, 성분이 표적하는 단백질의 개수를 Y축에 놓았을 때, 대부분 성분은 원점에 가깝게 분포해 있었다(〈그림 6〉의 검정색 점). 약재의 개수 및 표적의 개수가 모두 사분위수 0.99 이상인 성분(〈그림 6〉의 빨간색 점)들은 총 8개로 〈표 3〉과 같았다. Quercetin, Kaempferol, Oleic acid, Luteolin, Apigenin, Ursolic acid, Succinic acid, Eugenol 성분은 58개 이상의 약재에서 등장하고, 41개 이상의 단백질을 표적 하는 다빈도, 대표적 성분들이다.

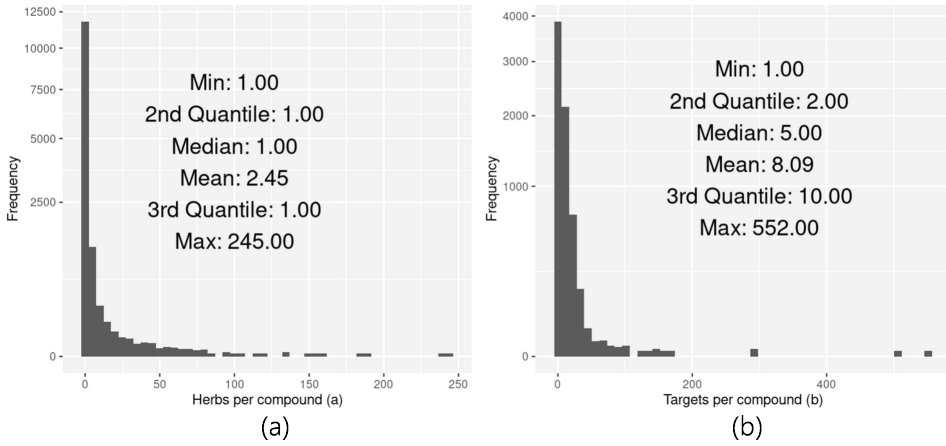


그림 5. TCMSPP 성분과 관계를 맺는 약재의 개수(a) 및 표적의 개수(b) 분포

표 1. 상위 10개의 다빈도 성분

Compound	#Herbs
Beta-sitosterol	245
Palmitic acid	239
Quercetin	188
Sitogluside	184
Sitosterol	160
EIC	154
Stearic acid	148
Kaempferol	133
Stigmasterol	132
Myristic acid	121

표 2. 상위 10개의 대표적 성분

Compound	#Targets
FMT	552
Acetic acid	504
IPA	294
Azeton	290
GLY	173
Quercetin	154
Resveratrol	151
Acetaldehyde	142
(-)-epigallocatechin-3-gallate	140
OXL	125

표 3. 다빈도 대표적 성분 8개

Compound	#Herbs	#Targets
Quercetin	188	154
Kaempferol	133	63
Oleic acid	117	48
Luteolin	93	57
Apigenin	81	80
Ursolic acid	81	55
Succinic acid	69	60
Eugenol	58	41

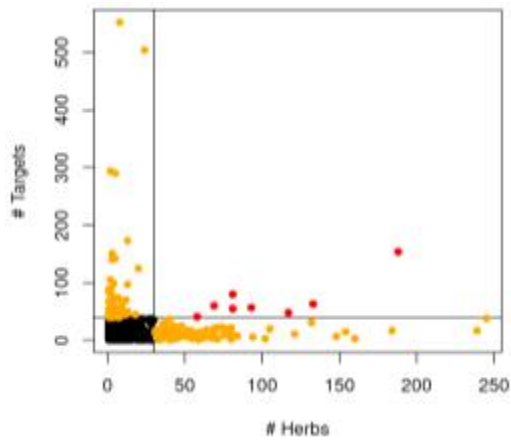


그림 6. 2차원 평면에서 성분의 분포

3. 약제의 염증 반응 관계 분석에서 Enrichment test의 결과에 영향을 미치는 다빈도 다표적 성분

앞에서 우리는 TCMSPP의 성분정보와 관계를 맺는 약제 개수 및 표적 개수의 통계를 구하고, 일부 다빈도 다표적 성분을 확인했다. 다음으로 약제와 염증 반응의 관계 판단을 위해 Enrichment test 적용 시, 그 결과에 다빈도 다표적 성분이 미치는 영향을 확인하였다. 이를 위해 ‘1. 연구 단계 및 방법 소개’에서 언급한 네트워크 약리학 분석단계를 수행했다. 실제 분석 상황에 가깝도록, 분석에 사용된 구성성분은 기존의 한약 네트워크 약리학 연구들에서 사용되는 성분 스크리닝 조건인 Oral bioavailability (OB)를 30 이상, Druglikeness (DL)를 0.18 이상인 성분들로만 한정하였다¹⁾. 성분들을 표적 단백질과 연계하고, 염증 관련 단백질과 공통된 Co-target 단백질까지 연계가 가능했던 약제는 총 502개 중 482개였다. 이 482개의 약제에 대해서 약제-성분-표적 정보를 연계하고 표적 정보를 쿼리 Gene set으로, Gene Ontology (GO) Biological process (BP)에 대한 Enrichment test를 수행한 후, GO BP의 Inflammatory response (GO:0006954) Gene set의 통계적 유의성 여부를 판단하였다. 482개 중 270개(56%)의 약제에서 염증 반응이 BH p-value ≤ 0.01로 통계적으로 유의한 것으로 나왔다. 나머지 212개(44%)의 약제들은 Enrichment test에서 염증 반응이 통계적으로 유의하지 않게 나왔거나 염증 반응과 연관된 유전자가 존재하지 않아서 테스트를 수행할 수 없는 경우였다.

염증 반응과 관련성이 통계적으로 유의하게 나온 270개의 약제 그룹과 그렇지 않은 약제 그룹에 등장하는 성분에 서로 차이가 있는지를 확인하기 위해 Chi-square test를 수행하였다. BH procedure로 정정했을 때 유의하게(BH p-value ≤ 0.05) 차이가 나는 성분들은 <표 4>과 같았다. 이 성분들 중 Kaempferol, Quercetin, Luteolin은 <표 3>에서 언급한 다빈도 다표적 성분들이다. 다음 절에서 이 세 가지 성분이 실제 출간된 논문들에서 활성 성분으로 언급된 빈도수를 확인해보았다.

표 4. Enrichment test에서 염증 반응과 관련 되어 있는 것으로 나온 약제들에 포함된 성분들

Compound	약제 개수 (염증반응 O 성분포함 O)	약제 개수 (염증반응 O 성분포함 X)	약제 개수 (염증반응 X 성분포함 O)	약제 개수 (염증반응 X 성분포함 X)	p-value	BH p-value
Quercetin	188	82	0	212	6.E-54	9.E-51
Kaempferol	133	137	0	212	1.E-32	8.E-30
Luteolin	93	177	0	212	6.E-21	3.E-18
Isorhamnetin	38	232	2	210	5.E-07	2.E-04
(+)-catechin	36	234	5	207	4.E-05	1.E-02
Acacetin	21	249	0	212	9.E-05	2.E-02
Beta-sitosterol	158	112	87	125	2.E-04	4.E-02

4. 염증관련 한약 네트워크 약리학 연구 문헌에서 다빈도 대표적 성분의 등장

앞에서 우리는 약재와 염증 반응의 관련성 판단을 위한 Enrichment test의 결과에 영향을 미친 것으로 보이는 일부의 성분을 확인하였다. 이번에는 발표된 한약과 염증의 관련성을 분석한 논문으로 발표된 네트워크 약리학 연구들에서 이 성분들이 활성 성분으로 얼마나 자주 언급되는지를 확인하였다. 구체적으로 TCMSp 데이터베이스를 활용한 한약 네트워크 약리학 문헌들에서, Kaempferol, Quercetin, Luteolin이 해당 한약에 대한 활성 성분으로 얼마나 많이 언급되는지를 확인하였다.

표 5. Kaempferol, Quercetin, Luteolin이 활성 성분으로 언급된 문헌들

Year	PMID	질병, 증상	한약 처방, 약재	종류	Kaempferol	Quercetin	Luteolin
2020	32501408	Respiratory tract infection	Shuanghuanglian	처방	1	1	1
2020	32854331	Influenza-associated inflammation	Fritillariae thunbergii	약재	0	0	0
2020	33000580	Airway neurogenic inflammation	Chan Qin (CQ) granules	처방	1	1	0
2020	33424592	Pelvic inflammatory disease	Sargentodoxa cuneata Patriniascabiosifolia	약재	1	0	0
2021	33519947	Inflammation	Shiyifang Vinum	약재	1	1	1
2021	34754318	Stroke	Dengzhan Shengmai	처방	1	1	0
2021	34966437	Asthma	Guizhi Decoction	처방	0	1	0
2022	35154344	Intervertebral disc degeneration	Bu-Shen-Huo-Xue Decoction	처방	1	1	1
2022	35368759	Patients with ankylosing spondylitis	Xinfeng Capsule	처방	1	1	0
2022	35799189	COVID-19	Lianhua Qingwen Capsules	처방	0	0	0
2022	35873643	Upper respiratory tract infection	Citrus changshan-Huyou Y. B. Chang	약재	0	0	1
2022	36158488	Inflammatory bowel disease	zedoary turmeric-trisomes	처방	0	0	0
2022	36212968	Inflammation	Pan Ji Sheng	처방	1	1	0
2022	36401196	Diabetic nephropathy	Chuanxiong	약재	0	0	0
2023	36624435	Inflammation	Huangqi Guizhi Wuwu decoction	처방	1	1	0
2023	37056910	Colitis and colitis-related carcinogenesis	Qingchang Wenzhong Decoction	처방	1	1	0
2023	37287697	Airway inflammations	Tuo-Min-Ding-Chuan Decoction	처방	1	1	1
2024	37454750	Acute lung inflammation	Jingfang granule	처방	1	1	1
2023	37657026	Inflammation	Radix Astragali	약재	1	1	0
2023	37670946	Gastrointestinal motility disorder	Wuda granule	처방	0	0	1
2023	37822694	Inflammation	Dodonea angustifolia	약재	0	0	1
총개수 (비율)					13 61.9%	13 61.9%	8 38.1%



PubMed에서 'TCMSP AND Inflamm*[TI]'를 키워드로 검색하면 2023년 10월 22일 기준 총 49개의 문헌 결과가 검색되었다. 이 중에서 Mass Spectrometry 등을 활용해서 직접 한약의 성분을 분석한 논문, 중국어 원문을 구할 수 없는 논문, 단일 성분(Single compound)을 대상으로 삼은 연구 등은 제외하였다. TCMSP를 주로 활용하여 한약의 항염증에 대한 네트워크 약리학을 적용한 연구 문헌의 개수는 21개였고, <표 5>는 이 문헌들에서 다빈도 대표적 성분들이 활성 성분으로 언급된 횟수를 보여준다. Kaempferol, Quercetin, Luteolin을 활성 성분으로 언급한 문헌은 각각 13개(61.9%), 13개(61.9%), 8개(38.1%)였다.

5. 다빈도 대표적 성분들에 대한 염증 반응 관련 Enrichment test 결과

앞에서 네트워크 약리학적 분석으로 염증 반응과 관련하여 통계적으로 유의하게 나온 많은 약제에서 주로 등장하는 다빈도 대표적 성분을 확인했고, 문헌 분석에서도 이 성분들이 활성 성분으로 빈번하게 언급됨을 확인하였다. 이번에는 TCMSP의 모든 성분 각각을 대상으로 Enrichment test를 수행하여 염증 반응과의 관련성을 확인해보았다. <표 6>는 통계적으로 유의하게 나온 8개 성분의 Enrichment test의 '염증 반응'의 순위와 BH p-value를 보여준다. 앞에서 언급한 Kaempferol, Quercetin, Luteolin은 모두 통계적으로 유의하게 나왔다.

표 6. Enrichment test에서 염증 반응과 관련된 것으로 나온 성분들

Compound	# Common genes	p-value	BH p-value
Quercetin	20	7.35.E-19	9.28.E-15
Triptolide	10	8.48.E-13	1.07.E-08
Fisetin	10	2.33.E-12	2.94.E-08
Wogonin	9	1.56.E-10	1.97.E-06
Kaempferol	8	4.26.E-08	5.38.E-04
(-)-epigallocatechin-3-gallate	8	6.87.E-06	1.93.E-03
Luteolin	7	3.58.E-06	3.23.E-03
Irisolidone	5	8.17.E-06	2.58.E-02

6. 고찰

2012년경부터 다양한 한약을 대상으로 네트워크 약리학 방법론을 적용한 여러 논문이 발표되었고, 해가 갈수록 그 수는 계속 늘고 있다. 한약의 성분을 LC/MS와 같은 실험적 방법을 통해 직접 규명한 후 확인된 성분 정보만을 활용하는 연구 결과들도 있지만, 공개 데이터베이스에서 얻은 정보를 활용하여 분석 연구를 수행한 논문들도 다수이다. 공개 데이터베이스는 약제와 그 구성성분 및 표적 정보까지 연구자들에게 제공하여, 다양한 한약 처방, 약제, 성분에 대한 접근을 쉽게 만들었다.

본 연구에서 우리는 TCMSP의 약제와 GO BP의 염증 반응의 관계 분석이라는 특정 사례를 연구했다. 모든 약제에 대해 Enrichment test를 수행했을 때 270개(56%)의 약제가 염증 반응과 관련된 것으로 나오는 것을 확인했다. 이러한 결과는 실제 수많은 약제가 염증 반응과 관련이 있음을



함의할 수도 있고 또는 결과에 위양성(False positive)이 포함되어 있을 가능성을 함의할 수도 있으나 이는 전적으로 *in silico* 분석 결과이므로 검증은 쉽지 않다. 한약 네트워크 약리학에서 '약재-성분-표적'의 결합은 '약재-성분' 관계와 '성분-표적' 관계를 테이블 결합(Join)하는 형태로 이루어지므로 특정 한약과 특정 표적의 관계 여부는 그 사이를 이어주는 성분의 유무에 달려있다. 우리는 Enrichment test 결과를 통해 염증 반응과 관련된 것으로 나온 약재들에서 Quercetin, Kaempferol 같은 일부 성분들이 자주 등장함을 확인했다.

특히 Quercetin, Kaempferol은 각각 188개(37.5%), 133개(26.5%)의 TCMSP의 약재들에 구성성분으로 포함되었다. 즉 다양한 약재들에 흔하게 등장하는 성분이다. 또한 이 성분들은 각각 154개와 63개의 단백질 표적을 가지고 이들 중 적지 않은 단백질들이 염증 반응과 관련이 있었기에 <표 6>에서 확인한 것처럼 개개의 성분만으로도 Enrichment test에서 염증 반응과 관련이 있는 것으로 나왔다. 이러한 다빈도 대표적 특성을 띠는 일부 성분의 존재로 인해 502개 중 270개에 달하는 TCMSP의 약재들이 Enrichment test의 결과로 염증 반응과 관련 있는 것으로 나오게 된 것으로 보인다. 또한 출간된 네트워크 약리학 연구 문헌들에서 이 성분들이 염증 관련 질병의 치료를 위한 활성 성분으로 자주 언급됨을 확인했다. <표 5>의 문헌들은 실험적 성분 분석을 하지 않고 주로 TCMSP의 약재 성분 정보만을 활용해서 네트워크 약리학 연구를 수행한 것들이다.

우리는 공개 데이터베이스를 활용한 네트워크 약리학 분석에서 다빈도 대표적 특성을 가지는 성분들이 분석 결과에 미칠 수 있는 영향을 살펴보았다. 데이터의 이러한 특성을 다룬 기존 연구는 없었기에 본 연구는 이전의 연구들과 차별성이 있다. 그러나 우리의 연구는 특정 성분이 분석 결과에 영향을 미칠 수 있음을 보였을 뿐이다. 실제 한약과 질병 또는 기전 사이의 네트워크 약리학 분석 결과에 포함될 수 있는 위양성을 통제하고 결과의 신뢰성을 높이기 위해서는, 실험적인 방법을 통한 성분의 규명을 통해 성분의 실제 존재 여부를 확인해야 하고, 세포 및 동물 실험을 통한 효능 검증이 따라야 할 것으로 보인다.

우리는 한약 네트워크 약리학에서 다빈도 대표적 성분으로 인한 영향이라는 특정한 사례를 다루었으나, 이 사례를 네트워크 분석에서 Hub node의 Degree가 미치는 영향이라는 문제로 확장 또는 일반화하여 문제를 바라볼 수 있다. 이종의 데이터를 통합하여 단백질과 질병의 관계를 규명하기 위한 네트워크 분석 연구¹⁰⁾에서 Hub node의 높은 Degree 값에 의해 이러한 관계 규명이 어려움을 보고하였다. 연구에서는 이러한 Hub node에 의한 영향을 낮추기 위해서 Down-weight path count와 같은 방법론을 제안하였다. 차후의 연구에서는 다빈도 대표적 성분을 Hub node로 보고 유사한 방법론을 우리의 문제에 적용하여 다빈도 대표적 성분의 중요도를 낮추는 방법을 고안함으로써 이런 다빈도 대표적 성분으로 인한 문제에 대한 더 나은 방법을 모색할 수도 있을 것이다.

결론

우리는 공개 데이터베이스에 기반한 한약의 네트워크 약리학적 분석에서, 다빈도 대표적 특성을 지닌 일부 성분으로 인해 분석 결과에 있어서 편향이 있을 수 있음을 살펴보았다. 구체적으로, TCMSP의 약재에 빈번하게 등장하는 일부 성분들로 인해 Enrichment test의 결과가 크게 영향을 받을 수 있음을, 직접 분석 및 기존의 문헌 사례를 통해 확인하였다. 네트워크 약리학은 *In silico* 기반의 연구이므로 분석 결과의 신뢰성을 위해서는 데이터에 대한 정확한 이해나 실험적 검증이 필요할 것으로 보인다.



감사의 글

이 논문은 한국한의학연구원 기본사업 “AI 한의사 개발을 위한 ICT 기반 한의 증점 질환 진단 예측 기술 개발(KSN1823130)”, “AI 한의사 개발을 위한 임상 빅데이터 수집 및 서비스 플랫폼 구축(KSN1922110)”, “폐암의 면역개선 통합 항암치료기술 개발(KSN2321240)” 지원을 받아 수행된 연구임.

참고문헌

1. Ru J, Li P, Wang J, Zhou W, Li B, Huang C, Li P, Guo Z, Tao W, Yang Y, Xu X, Li Y, Wang Y, Yang L. TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. *Journal of cheminformatics*. 2014;6:1-6.
2. Wang JF, Zhou H, Han LY, Chen X, Chen YZ, Cao ZW. Traditional Chinese medicine information database. *Clinical Pharmacology & Therapeutics*. 2005;78(1):92-3.
3. Kim SK, Nam SJ, Jang HC, Kim A, Lee JJ. TM-MC: a database of medicinal materials and chemical compounds in Northeast Asian traditional medicine. *BMC Complementary and Alternative Medicine* 2015;15(1):1-8.
4. Kuhn M, von Mering C, Campillos M, Jensen LJ, Bork P. STITCH: interaction networks of chemicals and proteins. *Nucleic acids research*. 2007;36:D684-8.
5. Liu Z., Guo F, Wang Y, Li C, Zhang X, Li H, Diao L, Gu J, Wang W, Li D, He F. BATMAN-TCM: a bioinformatics analysis tool for molecular mechanism of traditional Chinese medicine. *Scientific reports*, 2016;6(1):21146.
6. Pinero J, Bravo A, Queralt-Rosinach N, Gutierrez-Sacristan A, Deu-Pons J, Centeno E, Garcia-Garcia J, Sanz F, Furlong LI, DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic acids research*, 2017;45(D1):D833-9.
7. Aleksander, S.A., Balhoff, J, Carbon, S, Cherry JM, Drabkin HJ, Ebert D, Feuermann M, Gaudet P, Harris NL, Hill DP, The Gene Ontology knowledgebase in 2023, *Genetics*, 2023;224(1):iyad031, DOI:10.1093/genetics/iyad031
8. National Institutes of Health, DAVID Functional Annotation Bioinformatics Microarray Analysis, Version v2023q2, Available from <https://david.ncifcrf.gov> (accessed 2023-10-30)
9. Jang H, The Development of Herbal Medicine Network Analysis System, *Journal of The Korea Society of Computer and Information*, 2023;28(10):113-21.
10. Himmelstein DS, Baranzini SE, Heterogeneous network edge prediction: a data integration approach to prioritize disease-associated genes. *PLoS computational biology*, 2015;11(7):e1004259.

