

The Development of Herbal Medicine Network Analysis System

Ho Jang*

*Senior Researcher, KM Data Division, Korea Institute of Oriental Medicine, Daejeon, Korea

[Abstract]

Network pharmacology in traditional Korean and Chinese medicine studies the molecular and biological aspects of herbal medicine using computational methods. Despite variations in databases, techniques, and criteria, most studies follow similar steps: constructing herb-compound networks, compound-target networks, and target interpretation. To ensure efficient and consistent analysis in herbal medicine network pharmacology, we designed and implemented a common analysis pipeline. We showed its reliability with existing databases. The proposed system has a potential to facilitate network pharmacology analysis in traditional medicine, ensuring consistent analysis of various herbal medicines.

▶ **Key words:** Network Pharmacology, Korean Medicine, Traditional Chinese Medicine, Data Integration, Bioinformatics

[요 약]

한의학 및 중의학 분야에서 네트워크 약리학은 계산학적인 방법을 통해 한약의 분자생물학적인 기전을 연구하기 위해 널리 활용된다. 개별 연구를 위해 사용되는 데이터베이스, 분석기법, 분석 기준 등은 다양하나, 대부분의 한의학 네트워크 약리학 연구들은 약제-성분 네트워크 구축, 성분-표적 네트워크 구축, 표적의 해석이라는 유사한 단계로 수행된다. 효율적이고 일관성있는 한의학 네트워크 약리학 분석을 위해서, 우리는 일반적으로 적용할 수 있는 한의학 네트워크 분석 파이프라인 시스템을 설계하고 구현하였다. 우리는 이 시스템의 신뢰성을 약제의 네트워크 약리학 분석을 위해 널리 사용되는 데이터베이스를 활용해서 확인하였다. 제안된 시스템은 다양한 한약에 대한 네트워크 약리학적 분석을 용이하게 하고, 일관성 있는 분석을 도울 것이다.

▶ **주제어:** 네트워크 약리학, 한의학, 중의학, 데이터 통합, 생물정보학

I. Introduction

한약 및 한약재는 수천 년 동안 동아시아 곳곳에서 질병 및 증상을 치료하기 위해서 사용되어왔고 현대의학이 자리잡은 이후에도 꾸준히 활용되고 있다. 한국한의학진흥원의 ‘2022년 한방의료이용 실태조사 일반국민 보고서’에 따르면 만 19세 이상 국민의 71%는 평생동안 한방의료 이용 경험이 있다고 응답했고, 처방이나 한약제제를 이용한 적이 있는 응답자는 각각 26.7%와 28.5%였다[1]. 동 기관의 ‘2021년 한약 소비 실태조사’에 따르면 근골격계통의 치료에는 오직산이, 소화계통에는 평위산이, 호흡계통에는 소청룡탕이나 갈근탕이 가장 많이 사용되는 것으로 조사되었다[2]. 또한 한약 및 한약재는 현대에 새로 등장한 건강 문제에 대한 해법으로서 여전히 활발히 연구 및 사용되고 있다. 한약재 팔각회향에서 유래한 Shikimic acid을 원료로 하여 얻어지는 Oseltamivir가 인플루엔자 바이러스의 증식을 억제하는 치료제로 널리 사용되고 있는 것이 그 사례이다[3].

이처럼 한약 및 한약재는 현대에 이르러서도 그 활용 가치가 유효하고 현대에 등장한 다양한 질병에 대한 해법으로서도 그 활용성이 기대되나, 그 효능의 기전을 분자 수준에서 과학적으로 규명하는 것은 도전적인 문제로 여겨지고 있다. 주요한 이유 중 하나는 한약 및 약재가 단일 성분(Compound)이 아닌 다양한 성분으로 이루어져 있고 이 성분들이 인체의 여러 표적(Target)에 영향을 미치는 다성분-다표적(Multi Compound-Multi Target) 특성 때문에 약재와 표적의 해석이 복잡하고 어렵다는 점이다[4] (Fig. 1).

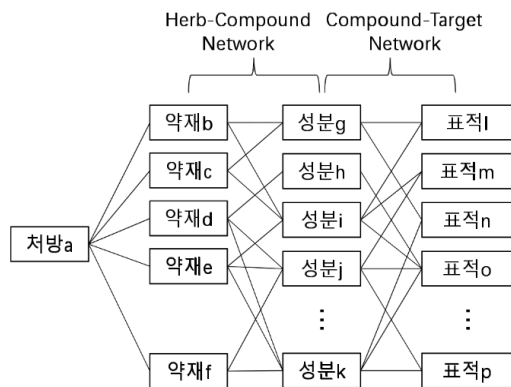


Fig. 1. Connections of formula, herbs, compounds, and targets in a herbal medicine networks

네트워크 약리학(Network Pharmacology)은 약리를 해석하기 위한 종래의 단일성분(Single Compound)-단일 표적(Single Target)의 틀에서 벗어나 성분이 영향을 미칠

수 있는 다중 표적(Multi Targets)의 상황을 분석하기 위한 분석 방법으로서 약물 발굴 연구를 위해서 적용되어왔다 [4]. 한의학에서는 약재 또는 처방을 이루는 여러 성분들이 조합되어 인체의 다양한 표적에 영향을 미쳐서 치료 효과를 보인다고 가정하고 있는데, 네트워크 약리학은 이런 가정을 바탕으로 약재나 처방에 대한 분자생물학적인 기전을 탐색하거나 활성 성분을 탐색하기 위한 좋은 틀을 제공해주고 있기에 2010년 이후 한약, 약재, 성분을 대상으로 한 다양한 네트워크 약리학 분석 연구가 발표되고 있다[5].

몇몇 사례로서, 천연물 유래 단일성분인 Cinnamic Aldehyde의 골관절염에 대한 항염증 효과를 분석한 연구 [6]에서는 Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform(TCMSP) 데이터베이스[7]를 이용해서 이 성분의 표적을 탐색하였고, 갈근금련탕의 궤양성 대장염에 대한 치료 기전 규명 연구[8]에서는 TCMSP를 이용해서 처방을 이루는 구성약재와 성분, 성분과 표적사이의 네트워크를 구성하고 Kyoto Encyclopedia of Genes and Genomes(KEGG)[9], Gene Ontology(GO)[10]와 같은 데이터베이스를 이용해서 생물학적인 기전을 규명했다. 이원영 외[5]의 문헌분석 연구에 의하면 대부분의 분석연구들은 약재(또는 처방) 선정, 약재의 구성성분의 식별, 성분-표적 관계 구축, 표적의 해석 단계로 진행이 된다. 즉 약재-성분 네트워크 구성, 성분-표적 네트워크 구성, 표적의 해석이라는 동일한 프로세스로 진행이 되나, 각 네트워크 약리학 연구마다 네트워크를 구성하기 위해 사용하는 데이터베이스, 성분과 표적의 필터링 방법, 표적의 해석방법 등이 각기 달라서 모든 연구에 통용될 수 있는 골드스탠다드는 없고, 이 때문인지 한의학 네트워크 약리학 분석을 위한 통합적 환경을 제공해주는 분석 소프트웨어 툴도 아직 보고되지 않았다.

비록 분석하려는 대상과 분석에 사용하려는 세부적인 데이터베이스는 연구마다 다를지라도, 네트워크 약리학의 공통된 단계들을 단일한 프레임워크 안으로 포괄하는 분석 파이프라인을 구축할 수 있다면, 복잡한 네트워크 약리학 분석작업에서의 실수를 줄이고 분석 효율을 높이고, 더 나아가 표준화된 분석 방법을 모색할 수 있을 것이다. 본 연구에서는 한의학 네트워크 약리학 분석을 위한 단일 프레임워크 분석시스템의 설계 및 구현에 대해서 논할 것이다.

본 논문의 구성은 다음과 같다. 2장에서는 시스템의 설계 및 구현을 위한 단계를 논의하고 3장에서는 2장의 방법을 적용한 결과를 논의한다. 4장에서는 개발된 시스템에 대한 한계점 및 향후 발전 방향을 논하고, 5장에서는 논문의 결론을 요약한다.

II. Research Method



Fig. 2. Steps for herbal medicine network analysis system development

1. Research steps

Fig. 2은 한의학 네트워크 약리학 분석 파이프라인 구축 및 시스템 개발의 단계를 보여준다. 네트워크 약리학 분석 방법을 주제로 한 리뷰 문헌을 수집하여 공통적인 분석 단계를 확정한다. 분석 과정을 구체화한 분석 파이프라인을 설계하고 이를 지원하기 위한 데이터베이스 테이블을 설계한다. 컴퓨터 언어 및 패키지를 통해서 실제 분석 시스템을 구현한다. 마지막으로 구현된 시스템의 신뢰성을 확인한다.

구체적으로, 네트워크 약리학 분석법을 대상으로 하는 리뷰 논문을 수집하고 정성분석을 통해서 다양한 네트워크 약리학 분석 조건을 포괄할 수 있는 단계를 정한다. 분석 파이프라인 설계에서는 약재, 성분, 표적의 필터링을 위한 방법을 추가하여 한의학 네트워크 약리학 분석기능을 구체화한다. 다음으로 분석 파이프라인의 지원을 위한 데이터베이스 테이블을 설계한다. 이후 웹서비스를 구현할 수 있는 언어 및 패키지 또는 라이브러리를 선정하고 실제 구현한다. 구현된 시스템의 성능은 본 분석 시스템을 이용하지 않고 직접 수행한 네트워크 약리학 분석 결과를 본 시스템에서 구현된 기능을 활용하여 얼마나 가깝게 재현할 수 있는지를 측정해서 확인하였다. 다음 장에서는 각 연구 단계별 수행 결과를 기술한다.

III. Result

1. Common Steps for Herbal Medicine Network Pharmacology

네트워크 약리학 연구를 리뷰한 논문을 PubMed에서 검색했다. 구체적으로 2018년 1월 1일부터 2022년 12월 31일 사이에 출간된 문헌들 중 제목에 'network

pharmacology'를 포함하고 PubMed의 문헌 종류 (Publication Type)는 Review로 지정된 논문들로 한정했다. 총 62개의 검색 문헌이 있었다. 이 문헌들을 대상으로, 제목 또는 초록 내용을 확인하여 특정 질병이나 특정 치료법으로 주제가 국한된 문헌은 제외하고 네트워크 약리학 자체를 논한 논문들로 한정했다. 또한 자료 접근성의 문제로 원문이 중국어인 논문들도 제외했다. 한의학 네트워크 약리학이 주제가 아닌, 네트워크 약리학 일반론을 다루는 연구들도 제외를 했다. 한의학 네트워크 약리학 분석 단계를 기술한 논문은 62개 중 1개가 있었고[11], Table 1.에 해당 논문에 기술된 분석 단계를 정리하였다. 약재의 구성 성분은 TCMSP 같은 외부 데이터베이스에서 얻어오거나 Liquid Chromatography and Mass Spectrometry (LC/MS) 등을 이용하여 실험적인 방법으로 직접 구성하였다. 이후에 TCMSP, Traditional Chinese Medicine Integrated Pharmacology Platform(TCMIP)[12], Bioinformatics Analysis Tool for Molecular Mechanism of TCM(BATMAN-TCM)[13]과 같은 데이터베이스에서 성분의 표적을 찾는다. 특정 질병과 관련된 표적들에 관심이 있다면 Online Mendelian Inheritance in Man(OMIM) [14], Genecards[15], Uniprot[16], Therapeutic Targets Database(TTD)[17]와 같은 유전자 정보 제공해주는 데이터베이스에서 그 질병과 연관된 표적 정보를 직접 얻어오거나, Gene Expression Omnibus(GEO)[18], The Cancer Genome Atlas(TCGA)[19]와 같은 다양한 생물체의 전사체 (Transcriptome) 데이터셋을 직접 분석하여 질병 관련 표적 정보를 구성하였다. 성분과 관련된 표적과 질병과 관련된 표적의 교집합을 이후의 단계에서 활용하는데, 여기에 더하여 Search Tool for the Retrieval of Interacting Genes(STRING)[19]과 같은 유전자간 인터랙션 정보를 제공해주는 데이터베이스를 이용해서 Protein-Protein Interaction 네트워크를 구성하여 표적에 대한 정보를 변형시키는 단계를 거치기도 한다. 마지막으로 KEGG나 GO와 같은 다양한 생물학적 기전과 관련한 유전자 셋(Gene set) 정보를 제공해주는 데이터베이스를 활용하여 생물학적 기전과의 관련성을 탐색한다. 서론에서 언급한 이원영 외[5]의 연구는 PubMed에 Review로 등록되지 않아서 본 문헌 조사에서 제외 되었으나, 그 연구는 2011년부터 2018년 사이에 출간된 총 147개의 네트워크 약리학 연구에 대해 분석하여 Table 2.와 같이 단계를 정리하였으므로, 우리는 두 연구의 분석 단계를 통합한 Table 3.과 같은 한의학 네트워크 약리학 분석 단계를 확

정하였다. Table 1.의 2~5 단계는 표적을 정하는 작업이므로 Table 3.의 2단계에 포함되어 Table 3.와 같이 3단계로 설정을 하였다. Table 3.의 ‘시스템 검증에 이용된 데이터’는 실제 시스템 구현 후, 성능의 검증을 위해서 사용한 데이터이다. TCMSP는 약재와 성분의 관계, 성분과 표적의 관계에 대한 정보를 제공하고, GO의 Biological Process(BP)는 유기체에서 다양한 생물학적 기능을 탐색하기 위해 널리 사용되는 유전자 셋 데이터이다.

Table 1. Analysis steps, techniques and databases for herbal medicine network pharmacology (Yuan et al.[11])

#	Step	Databases and techniques
1	Active Compound Search	Databases, LC/MS
2	Compound-Target Relation Search	TCMSP, TCMIP, BATMAN-TCM
3	Disease- related Target Search	OMIM, Genecards, Uniprot, TTD, GEO, TCGA
4	Target Intersection	Intersections between Drug Targets and Disease Targets
5	PPI Network Construction	STRING
6	Mechanism Study	KEGG, GO

Table 2. Analysis steps, techniques and databases for herbal medicine network pharmacology (Lee et al.[5])

#	Step	Databases and techniques
1	Herb-Compound Network Construction	TCMSP, TCM Database@taiwan, TCMID, TCMGeneDIT, HPLC, Literature mining, UPLC, BATMAN-TCM, TCM-PTD, Organchem, SciFinder, In-house database, NMR, Korean traditional knowledge portal database, LC-MS, Chinese Pharmacopoeia, BindingDB, TCM mesh
2	Compound-Target Network Construction	TCMSP, Molecular docking, PharmMapper, TCMGeneDIT, Microarray, CNKI, drugCIPHER, SEA, STITCH, ChemMapper, WES, TTD, Tanimoto coefficient, Medchem, AlzhCPI, Metadrag, TarNet, SwissTargetPrediction, ChemMapper
3	Target Interpretation	TTD, KEGG, Drugbank, GO, PharmGkb, RCSB PDB, Pubchem, Pubmed, PharmGkb, UTRI, GeneCards, Association database, FDA orange book, Wangfang Database, OMIM, IPA, CADGene, NCBI Gene database, MalaCards, PANTHER, OncoDB, DisGeNET, OncoDB

Table 3. Common steps for herbal medicine network pharmacology analysis

#	Step	The Databases used for the system validation
1	Herb-Compound Network Construction	Herb-compound relations from TCMSP
2	Compound-Target Network Construction	Compound-target relations from TCMSP
3	Target Interpretation	GO Biological Process (BP)

2. Analysis Pipeline

앞서 논한 분석 단계 설정은 약재-성분 관계 구축, 성분-표적 관계 구축, 표적 해석으로 이어지는 네트워크 약리학의 큰 틀에서의 분석 단계를 정하는 작업이었고, 분석 파이프라인은 단계별로 실제 분석을 위한 세부 사항을 구체화하는 작업이다. Fig 3. 는 앞선 네트워크 약리학 연구 문헌들로부터 추출된 분석 파이프라인을 보여준다. ‘분석 대상 선정’은 약재 조합(처방), 약재, 활성 성분 등 분석할 천연물을 선정하는 단계이다. ‘약재-성분 네트워크 구성’에서는 해당 대상에 포함되는 성분 정보를 검색하여 네트워크를 구성한다. 이때 성분을 필터링 할 수 있는 정보, 예를 들어 Drug likeness(DL) 나 Oral bioavailability(OB)와 같은 정보가 있다면 선별된 성분만 다음 단계인 성분-타겟 네트워크 구성에서 쓰이도록 추가 옵션을 넣을 수 있다. ‘성분-표적 네트워크 구성’은 성분이 표적으로 하는 유전자(단백질)와의 관계로 네트워크를 구성한다. 여기에서도 성분-표적을 필터링 할 수 있는 정보, 예를 들어 Search tool for interactions of chemicals(STITCH)[20]의 Confidence Score와 같은 정보가 있다면, 성분과 연결되는 표적의 개수를 줄일 수 있도록 추가 옵션을 넣을 수 있다. ‘질병-표적 필터링’에서는 표적을 미리 정해진 유전자 리스트, 예를 들어 특정 질병과 관련된 유전자의 목록을 가지고 있다면, 해당 유전자들과의 교집합인 표적들만을 대상으로 다음 단계인 ‘표적 해석’을 진행할 수 있다. ‘표적 해석’ 단계에서는 표적들과 관련된 것으로 보이는 생체경로나 기전을 탐색하거나 네트워크 시각화를 통해 요소들 사이의 관계를 시각적으로 파악하는 일이 이루어진다.

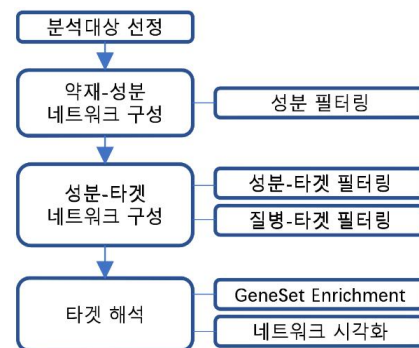


Fig. 3. Analysis pipeline for herbal medicine network pharmacology

3. Database tables

앞서 다루었던 분석 파이프라인은 시스템 이용자가 실제 분석에 사용하려는 다양한 약재 데이터, 성분 표적 데

이터, 표적 해석 기법을 모두 포괄할 수 있어야 한다. 이것을 시스템 내에서 구현하기 위해서는 다양성을 포괄할 수 있는 데이터베이스 테이블(Table)의 설계가 필요하다. 주요한 테이블 사이의 관계는 Fig 4.와 같고 각 테이블의 속성(Attributes)은 Table 4.와 같다. Fig 1.에서 보았던 것처럼 약재-성분-표적 관계 정보는 수많은 중복이 있으므로, 약재, 성분, 표적을 개별 엔티티 테이블로 선정하였고, 약재-성분, 성분-표적 관계 테이블로 분리하였다.

Table 4.의 테이블 HERB는 약재의 식별자, 명칭과 같은 기본 정보와 약재 데이터를 얻어온 출처 같은 정보를 포함한다. COMPOUND는 성분의 식별자와 이름, DL, OB와 같은 약리학적 특성 정보를 담고 있다. TARGET은 유전자 표적의 정보를 저장한다. HERB_COMPOUND를 통해서 약재와 성분 관계가 연계되고, COMPOUND_TARGET을 통해 성분과 표적이 연계된다. GENE_SET은 표적 해석 단계에서 사용되는 정보들이 저장되는데, 예를 들어 GO의 Biological Process나 KEGG Pathway와 같은 특정 생체 기전과 관련있는 유전자 집합의 정보를 저장하는 용도이다. 약재, 성분, 유전자는 각각 다양한 이명(Alias)을 가질 수 있으므로 약재, 성분, 유전자 검색의 용이성을 위해서 HERB_ALIAS, COMPOUND_ALIAS, TARGET_ALIAS에 이명을 저장하도록 했다.

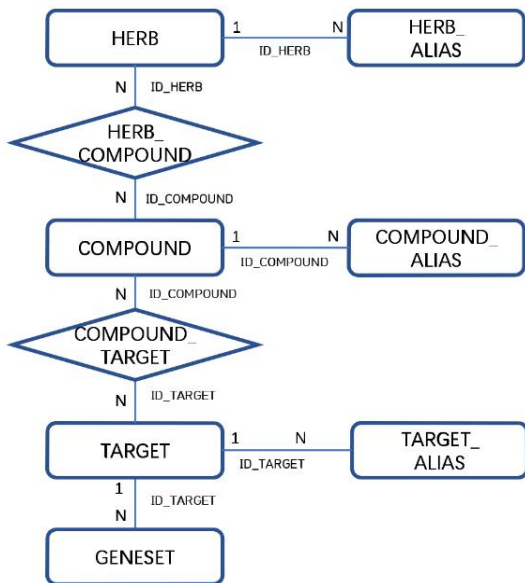


Fig. 4. The table relations of herbal medicine network pharmacology analysis system

Table 4. The database tables and attributes of herbal medicine network pharmacology analysis system

Table	Attribute	Description
HERB	ID_HERB	Herb Identifier
	KOR	Herb name (Korean)
	CHI	Herb name (Chinese)
	ENG	Herb name (English)
	SOURCE	Data source
COMPOUND	ID_COMPOUND	Compound identifier
	NAME	Compound name
	DL	Druglikeness
	OB	Oral Bioavailability
	SOURCE	Data source
TARGET	ID_TARGET	Target identifier
	NAME	Target name
	SOURCE	Data source
GENESET	ID_GENESET	Gene set identifier
	ID_TARGET	Target identifier
	DESC	Gene set description
	SOURCE	Data source
HERB_COMPOUND	ID_HERB	Herb identifier
	ID_COMPOUND	Compound identifier
	SCORE	Herb-compound relation score
	SOURCE	Data source
COMPOUND_TARGET	ID_COMPOUND	Compound identifier
	ID_TARGET	Target identifier
	SCORE	Compound-target relation score
	SOURCE	Data source
HERB_ALIAS	ID_HERB	Herb identifier
	ALIAS	Synonym
	LANG	Language
COMPOUND_ALIAS	ID_COMPOUND	Compound identifier
	ALIAS	Synonym
TARGET_ALIAS	ID_TARGET	Target identifier
	ALIAS	Synonym
	SOURCE	Data source

4. System Implementation

시스템의 구현을 위해 R Shiny 패키지[21]를 사용했고, DBMS는 SQLite를 사용했다. Fig 5.는 구현된 시스템 화면이고 Fig 3.에서 정한 단계를 따라 사용할 수 있도록 구현되었다. 사용자는 Fig 5.(a-e)에서 ‘약재-성분 네트워크 구성’ 단계의 작업을 수행 할 수 있다. 사용자는 먼저 등록된 약재를 검색하고(Fig 5.(a)), 분석할 약재를 확정한다(Fig 5.(b)). 다음으로 성분 정보를 가져올 특정 DB를 선택하거나, 성분을 필터링하기 위해 OB나 DL같은 약리학적 인 조건을 걸어서 검색될 성분의 수를 줄일 수 있다(Fig 5.(c)). 사용자는 조건에 맞는 성분들을 확인하게 된다(Fig 5.(d)). 또는 사용자는 이후 단계에서 사용될 성분들을 직접 입력 할 수 있다(Fig 5.(e)).



Fig. 5. The implemented system

사용자는 Fig 5.(f-i)에서 ‘성분-표적 네트워크 구성’ 단계의 작업을 수행할 수 있다. 사용자는 먼저 성분의 표적 정보를 가져올 DB를 선택할 수 있다. 또한 선택한 DB가 성분-표적 관계를 필터링 할 수 있는 추가적인 정보를 제공하는 경우, 필터링을 통해 선별된 성분-표적 관계만 분

석에 사용할 수 있다(Fig 5. (f)). 이 과정을 통해 사용자는 선별된 약재-성분-표적 관계를 열람할 수 있다(Fig 5. (g)). 여기에 더하여, 사용자는 표적 정보를 특정 유전자 목록에 포함되는 것들로 추가적으로 필터링 할 수 있다 (Fig 5. (h)). 예를 들어 만약 사용자가 염증과 관련된 유전자들의 목록을 넣는다면, 이후 단계에서는 염증 관련 유전자들과 공통된 표적들만 이후 단계에서 사용될 것이다. 사용자는 최종적으로 필터링 된 표적 정보를 테이블로 확인할 수 있다 (Fig 5. (i)).

사용자는 Fig 5.(j-l)에서 ‘표적의 해석’ 단계의 작업을 수행할 수 있다. 사용자는 먼저 검색된 표적을 해석하기 위한 데이터베이스를 선택한다(Fig 5. (j)). 또한 사용자는 분석할 표적을 직접 입력 할 수 있다(Fig 5. (k)). 입력의 결과로, 사용자는 검색된 표적들과 관련성이 높을 것으로 보이는 생물학적인 기전에 대한 정보를 열람하게 된다(Fig 5. (l)).

5. System Validation

본 연구는 기존의 한의학 네트워크 약리학적 분석 수행을 단일한 파이프라인으로 통합한 것이 주요한 기여점이므로, 기존 방식의 수행 결과와, 구현된 시스템을 통해서 수행한 네트워크 약리학 분석 결과의 일치도를 비교해서 시스템의 성능을 측정 해보았다.

Fig 6.는 우리의 시스템을 이용하지 않은 직접 분석의 예시이다. 약재-성분 네트워크와 성분-표적 네트워크 구성을 위해 TCMSP를 이용하고, 표적 해석을 위해서 GO BP를 DAVID(Database for annotation, visualization and integrated discovery) 서비스[22]에서 제공하는 Gene set enrichment test를 이용하는 가상의 예시이다. 선택한 약재에 대해서 약재-성분-표적 네트워크를 구성해서 도출된 표적들과 연관성이 큰 GO BP의 Biological process를 찾는 과정이고 이를 DAVID에서는 도출된 표적 리스트와 GO BP에서 제공하는 모든 Biological process사이에 Fisher exact test를 수행하여 p-value를 계산한다. GO BP의 Biological process들은 유전자 목록으로 구성되어 있다. DAVID gene set enrichment test에서는 다양한 Biological process에 대해서 Simultaneously 하게 통계 검정을 수행하므로 Multiple testing problem으로 인한 type 1 error를 컨트롤 하기 위해서 Benjamini-Hochberg 방법을 사용하여 False discovery rate을 컨트롤 한 후, 통계적으로 유의한 Biological process들만 해당 약재와 관련이 있다고 여긴다. 우리는 TCMSP의 500개의 약재에 각각에 대해서 통계적으로 유의한 Biological process들을 산출하였다.

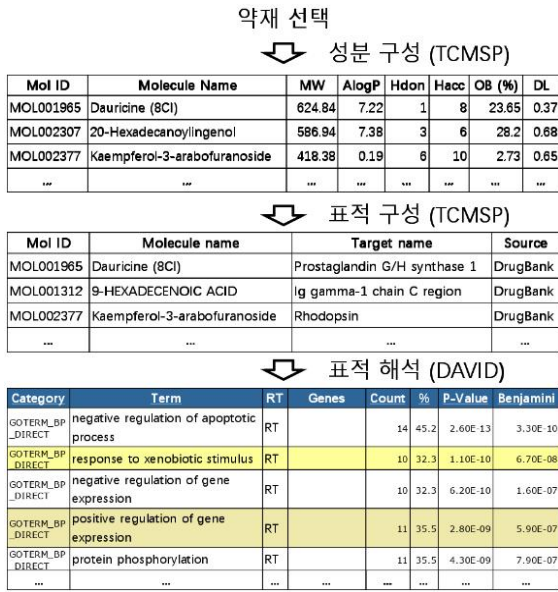


Fig. 6. Examples of manual network pharmacology analysis

한편 위의 작업을 우리의 시스템을 이용하여 수행해보았다. 먼저, TCMSP의 약재, 성분, 표적 데이터 및 DAVID의 GO BP 데이터를 시스템에 업로드 하였고, 위와 동일하게 표적 해석을 하기 위해서 Fisher exact test 및 Benjamini-Hochberg 분석 기능을 추가하였다. 그리고 시스템을 활용하여 각각의 약재에 대해서 통계적으로 유의한 Biological process를 산출하였다.

Fig 7.는 TCMSP의 500종의 약재에 대해서 두 방법으로 산출된, 통계적으로 유의한 Biological process의 일치도의 분포를 보여준다. 히스토그램에 의하면 대부분의 약재에 대해서 100%에 가깝게 재현이 되었고, 평균적으로 99.78%의 일치도를 보였다. 시스템으로 결과가 100% 완벽하게 재현되지 않는 이유는 데이터베이스의 버전, 데이터의 전처리 방식, 계산학적인 분석을 위해 사용한 프로그래밍 라이브러리의 차이 때문일 수 있다.

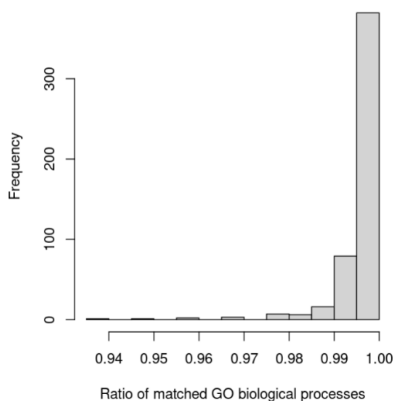


Fig. 7. Histogram of the ratio of matched GO biological processes for 500 TCMSP herbs

IV. Discussions

우리는 한의학 분야의 네트워크 약리학 방법론 체계적이고 일관성 있게 적용하기 위한 분석 시스템의 개발을 연구했다. 기존에 발표된 수많은 네트워크 약리학 분석 연구들에서, 데이터 통합 및 분석을 위해 채택한 데이터베이스 및 분석 방법론들은 서로 다르더라도, 다수의 네트워크 약리학 연구 문헌들에서 약재-성분-표적으로 이어지는 절차들에는 유사성이 있기에, 그 공통요소들을 추출하여 파이프라인을 구축하고, 단일한 플랫폼에서 분석이 이루어질 수 있도록 시스템을 설계하고 구현하였다. 또한 TCMSP 및 GO BP DIRECT Gene set의 활용이라는 특정한 사례를 바탕으로, 시스템의 Enrichment 분석 방법의 재현성 측면에서 구현된 시스템의 기존의 분석 방법을 잘 재현함을 보였다. 시스템은 체계적이고 일관성있는 네트워크 약리학 분석을 위해 제안되었고, 한의 처방, 약재, 성분을 활용하고자 하는 한의학 및 생의학 연구자들의 네트워크 약리학 분석 방법에 대한 접근성을 증대시킬 수 있다. 그러나 제안된 시스템은 다음과 같은 한계점 또는 개선이 필요한 부분이 있다.

첫째, 데이터 통합 기능의 개발이 필요하다. 기존에 발표된 네트워크 약리학 연구들은 약재-성분 네트워크 구축을 위해서 출처가 다양한 여러 성분 데이터를 합치거나, 성분-표적 네트워크 구축을 위해서 여러 출처의 성분-표적 데이터를 합쳐서 활용한다. 서로 다른 출처의 데이터를 통합하는 작업은 본격적인 분석의 전 단계인 전처리 작업에 해당하지만 최종 분석 결과에 매우 큰 영향을 끼치므로 많은 시간과 노동력이 들어가는 중요한 작업이다. 본 연구는 네트워크 약리학 분석 및 탐색의 편의를 높이기 위한 시스템의 개발에 맞춰져 있으므로 데이터의 통합에 대한 이슈는 논하지 않았으나, 이종의 데이터의 통합을 위한 기능을 제공한다면 본 시스템의 실제 활용성을 높일 수 있을 것으로 기대한다. 데이터베이스 통합과 관련한 이슈로는 성분 정보 및 표적 정보의 중복 및 모호성 해결 방법이 될 것으로 보인다. 개별 데이터는 각각 고유의 체계로 데이터를 배포 및 관리하기 때문이다. 성분 정보는 다양한 명칭으로 기술이 되므로 데이터 통합에 있어서, 모호성 해결 또는 중복 제거 작업은 최종 분석 결과의 정확성 및 신뢰성을 높이기 위해서 매우 중요하다. 마찬가지로 표적 정보를 분석함에 있어서 유전자 명칭도 데이터마다 HUGO Gene Nomenclature Committee(HGNC) 유전자 심볼, Ensembl genome database project의 Ensembl ID, National Center for Biotechnology Information(NCBI)

의 NCBI ID 등 다양한 방식으로 기술이 되므로 성분-표적 네트워크 구축 시 동일 표적 여부를 식별하여 중복을 제거하는 작업이 요구된다. 여러 출처에서 얻어진 성분-표적 관계를 통합하는 작업 역시 각 성분-표적 데이터마다 각각 고유한 방법으로 관계의 신뢰성(Confidence) 정보를 제안하므로 데이터 통합에서 이슈가 된다.

둘째, 약재, 성분, 표적 요소들간 관계를 보여주는 네트워크 시각화 기능이 필요하다. 발표된 대부분의 한의학 네트워크 약리학 분석 연구에서는 데이터 통합 및 분석 결과를 시각적으로 보여준다. 본 시스템에서는 수많은 약재, 성분, 표적 요소들 사이의 연결관계를 시각화 해주는 기능이 부재하고 자체 분석 결과를 내보내어(Export), 네트워크 시각화를 위해 널리 사용되는 사이토스케이프(Cytoscape)의 입력 데이터로 사용할 수 있는(Import) 기능만 포함하였다. 그러나 본 시스템을 활용하여 주요한 성분 및 표적을 탐색할 때, 각 요소 사이의 네트워크 연결 관계를 즉각적으로 볼 수 있다면, 시스템을 활용하는 연구자의 성분 및 표적 탐색작업의 효율성이 높아질 것으로 기대한다.

셋째, Molecular docking 및 Protein-Protein Interaction 의 기능 추가가 필요하다. 성분과 표적의 관계를 계산학적으로 보여주는 Molecular docking 기능 및 네트워크 분석에서 사용되는 Degree 및 Betweenness 의 필터링을 통한 연관 유전자의 확장은 많은 네트워크 약리학 분석 연구에서 활용되므로 해당 기능이 분석 파이프라인에 통합되면 분석결과의 신뢰성을 높혀 줄 것이다.

넷째, 전사체 분석 데이터 통합 기능이 필요하다. 다양한 처방, 약재, 성분에 대해, Intervention 전후의 전사체 데이터의 차이를 비교하여 표적유전자를 찾는 연구들이 실제로 실험현장에서 생성되고 있고, 기존에 발표되었던 연구의 전사체 데이터는 GEO와 같은 데이터베이스에서 다운로드가 가능하다. 이러한 전사체 데이터와 기존의 분석방법론을 통합하는 기법의 개발을 통해 성분-표적의 정보에 중요도를 부여하여 후보 표적을 좁히거나 관련 생체 경로를 특정하는데 도움이 될 수 있다.

V. Conclusions

본 연구에서 제안한 네트워크 약리학 분석 시스템의 초기 버전은 서로 다른 목적의 데이터베이스를 연계해야 하기에 여러 단계를 거쳐야 하고 성분, 표적 등을 선별해야 하는 복잡한 네트워크 약리학 분석 작업을 체계적으로 일관성 있게 수행하도록 설계 및 구현 되었다. 다양한 처방,

약재에 대해서 주요성분, 표적 및 생체경로를 탐색할 수 있는 편의를 제공하기 위해서 개발되었기에 데이터분석 및 전산학 비전공자의 네트워크 약리학 분석의 접근성을 향상시킬 수 있을 것으로 기대한다. 추후의 기능 추가를 통해 전문적인 기능을 추가하여 실제 현장에 활용될 수 있는 수준으로 높일 수 있을 것으로 기대한다.

ACKNOWLEDGEMENT

This research was funded by the Korea Institute of Oriental Medicine (No. KSN1823130 and No. KSN1922110).

REFERENCES

- [1] Korean Medicine Utilization Survey in 2022, National Institute for Korean Medicine Development
- [2] Herbal Medicine Consumption Survey in 2021, National Institute for Korean Medicine Development
- [3] S. Ghosh, Y. Chisti, and U. Banerjee, "Production of shikimic acid," *Biotechnology advances*, Vol. 30, No. 6, pp. 1425-1431, Nov, 2012. DOI: 10.1016/j.biotechadv.2012.03.001
- [4] L. Zhao, H. Zhang, N. Li, J. Chen, H. Xu, Y. Wang, and Q. Liang, "Network pharmacology, a promising approach to reveal the pharmacology mechanism of Chinese medicine formula," *Journal of Ethnopharmacology*, Vol. 309, Feb, 2023. DOI: 10.1016/j.jep.2023.116306
- [5] W. Lee, C. Lee, Y. Kim, and C. Kim, "The methodological trends of traditional herbal medicine employing network pharmacology," *Biomolecules*, Vol. 9, No. 8, pp. 362. Aug, 2019. DOI: 10.3390/biom9080362.
- [6] P. Chen, J. Zhou, A. Ruan, L. Zeng, J. Liu, and Q. Wang, "Cinnamic Aldehyde, the main monomer component of Cinnamon, exhibits anti-inflammatory property in OA synovial fibroblasts via TLR4/MyD88 pathway," *Journal of Cellular and Molecular Medicine*, Vol. 26, No. 3, pp. 913-924, Feb, 2022. DOI: 10.1111/jcmm.17148
- [7] J. Ru, P. Li, J. Wang, W. Zhou, B. Li, C. Huang, P. Li, Z. Guo, W. Tao, Y. Yang, X. Xu, Y. Li, Y. Wang, and X. Xu, "TCMSP: a database of systems pharmacology for drug discovery from herbal medicines," *Journal of cheminformatics*, Vol. 6, pp. 1-6, Dec, 2014. DOI: 10.1186/1758-2946-6-13.
- [8] M. Wei, H. Li, Q. Li, Y. Qiao, Q. Ma, R. Xie, R. Wang, Y. Liu, C. Wei, B. Li, and C. Zheng, "Based on network

pharmacology to explore the molecular targets and mechanisms of Gegen Qinlian decoction for the treatment of ulcerative colitis,” *BioMed Research International*, Nov, 2020. DOI: 10.1155/2020/5217405

- [9] M. Kanehisa, S. Goto, “KEGG: kyoto encyclopedia of genes and genomes,” *Nucleic acids research*, Vol. 28, No. 1, pp. 27-30, Jan, 2000. DOI: 10.1093/nar/28.1.27.
- [10] S. Aleksander, J. Balhoff, S. Carbon, J.M. Cherry, H.J. Drabkin, D. Ebert, M. Feuermann, P. Gaudet, N.L. Harris, and D.P. Hill, “The Gene Ontology knowledgebase in 2023,” *Genetics*, Vol. 224, No. 1, May, 2023. DOI: 10.1093/genetics/iyad031
- [11] Z. Yuan, Y. Pan, T. Leng, Y. Chu, H. Zhang, J. Ma, and X. Ma, “Progress and prospects of research ideas and methods in the network pharmacology of traditional Chinese medicine,” *Journal of Pharmacy & Pharmaceutical Sciences*, Vol. 25, pp. 218-226, Jun, 2022. DOI: 10.18433/jpps32911
- [12] M. Kan, J. Wang, S. Ming, X. Sui, X., Z. Zhang, Q. Yang, X. Liu, J. Lin, Y. Zhang, Q. Pang, and Y. Liu, “Investigating the mechanism of ShengmaiYin (codonopsis pilosula) in the treatment of heart failure based on network pharmacology,” *Combinatorial Chemistry & High Throughput Screening*, Vol. 25, No. 13, pp. 2191-2202, 2022. DOI: 10.2174/1386207325666220221093415
- [13] Z. Liu, F. Guo, Y. Wang, C. Li, X. Zhang, H. Li, L. Diao, J. Gu, W. Wang, D. Li, and F. He, “BATMAN-TCM: a bioinformatics analysis tool for molecular mechANism of traditional Chinese medicine,” *Scientific reports*, Vol. 6, No. 1, pp. 21146, Feb, 2016. DOI: 10.1038/srep21146
- [14] Online Mendelian Inheritance in Man, <https://omim.org>
- [15] GeneCards – the human gene database, <https://www.genecards.org>
- [16] Uniprot, <https://www.uniprot.org>
- [17] Therapeutic Target Database, <https://db.idrblab.net/ttd>
- [18] Gene Expression Omnibus, <https://www.ncbi.nlm.nih.gov/geo>
- [19] STRING: functional protein association networks, <https://string-db.org>
- [20] STITCH: interaction networks of chemicals and proteins, <http://stitch.embl.de>
- [21] Shiny: Web Application Framework for R, <https://github.com/rstudio/shiny>
- [22] DAVID Functional Annotation Bioinformatics Microarray Analysis, <https://david.ncifcrf.gov>

Authors



Ho Jang received his Ph.D. degree from School of Electrical Engineering and Computer Science at Gwangju Institute of Science and Technology in 2017. Dr. Jang has been working for Korea Institute of

Oriental Medicine (KIOM) since 2018. He is currently a senior researcher in KM Data Division at KIOM. His research interests include computational biology and machine learning.