Cluster Analysis of Medicinal Plants and Targets Based on Multipartite Network

Namgil Lee[1,2], Hojin Yoo[2], Heejung Yang[2,3,*]

[1]Department of Information Statistics, Kangwon National University, Gangwondaehak-gil 1, Chuncheon, Gangwon 24341, Republic of Korea

[2]Bionsight, Inc., Gangwondaehak-gil 1, Chuncheon, Gangwon 24341, Republic of Korea

[3]Laboratory of Natural Products Chemistry, College of Pharmacy, Kangwon National University, Gangwondaehak-gil 1, Chuncheon, Gangwon 24341, Republic of Korea

*** Correspondence:**
Heejung Yang
heejyang@kangwon.ac.kr

Keywords: medicinal plants, multi-chemicals, multi-targets, multipartite network, network analysis.

Word count: 3981

Number of figures: 4

Abstract

Although medicinal plants have played an important role in the development of therapeutics as natural compound sources, the synergistic effects of the multiple chemicals in the plants on multiple targets have not been studied widely. Recently, network-based methods for the analysis of drug-target interactions have gained attention and relies on the paradigm that a single drug can act on multiple targets rather than a single target. In this study, we have presented a novel approach to analyze the interactions between the chemicals in the medicinal plants and multiple targets based on the complex multipartite network of the medicinal plants, multi-chemicals, and multiple targets. The multipartite network was constructed via the conjunction of two relationships: chemicals in plants and the biological actions of those chemicals on the targets. By analyzing the structural characteristics of the network using the network analysis techniques, we investigated the similarities among the plants via their chemical profiles obtained from the hierarchical clustering analysis. In doing so, we introduced an index of the efficacy of chemicals in a plant on a protein target of interest, called target potency score (TPS). We showed that the analysis can identify specific chemical profiles from each group of plants, which can then be employed for discovering new alternative therapeutic agents. Furthermore, specific clusters of plants and chemicals acting on specific targets were retrieved using TPS that suggested potential drug candidates with high probability of clinical success.

We expect that this approach may open a way to predict the biological functions of multi-chemicals and multi-plants on the targets of interest and enable repositioning of the plants and chemicals.

1 Introduction

Plants have been used as therapeutic agents for a very long time in the human history (Petrovska, 2012). Single bioactive compounds in plants have been developed as well-known commercial drugs, such as paclitaxel, vinblastine, morphine, and artemisinin (Koehn and Carter, 2005; Harvey et al., 2015). Occasionally, single components from natural products have provided chemical scaffolds for developing more potent semi-synthetic drugs, such as rosuvastatin, tramadol, and eribulin (Rodrigues et al., 2016). Similar to these single compound drugs that act on a specific target, a mixture of components from the whole plants or their sub-fractions have been widely used as botanical drugs, such as Veregen® and Fulyzaq® (Galanakis, 2019). However, the synergistic effects of multi-chemicals in plants on multiple targets and the underlying complex biological pathways involved are less understood.

For the purpose of developing drugs composed of multi-chemicals, the identification of the interactions between the multi-chemicals and multi-targets is important to understand the therapeutic effects as well as the side effects, i.e., the on-target and off-target effects (Roth et al., 2004; Keiser et al., 2009; Besnard et al., 2012; Lounkine et al., 2012; Anighoro et al., 2014; Zhang et al., 2015, 2017; Ramsay et al., 2018). Identifying the synergistic drug combinations is also critical to develop effective therapies against cancers and to improve the therapeutic efficacy (Jia et al., 2009; NCI-DREAM Community et al., 2014; Sun et al., 2015; Han et al., 2017). However, the complexity of screening all possible candidate combinations of multi-chemicals and multi-targets prevents the use of the classical way of identifying the interactions via biological experiments, using *in vitro* and *in vivo* assays. Since a medicinal plant produces a large number of chemical compounds, the target identification of every single chemical is impossible for the development of botanical drugs, especially to acquire approval in the global market requiring well-controlled clinical studies (Ahn, 2017).

In recent decades, computational approaches for the systematic analysis of drug-target interactions have become available, improving efficiency and reducing costs of studying the interactions (Zheng et al., 2013; Lavecchia and Cerchia, 2016; Luo et al., 2017; Zhang et al., 2017; Ezzat et al., 2019; Forouzesh et al., 2019). Among the computational approaches, network-based approaches provide alternative methods to deal with the complex molecular interactions (Hopkins, 2008; 2019; Monteiro et al., 2020). Compared with other computational methods, network-based approaches do not require three-dimensional structures of the targets or negative control samples (Cheng et al., 2019). These approaches have also been proposed for botanical drug design in search of optimal combinations of medicinal plant components and prediction of multiple target activities (Leung et al., 2013). However, studies based on the integrated networks of medicinal plants, chemicals, and multi-targets are still rare. Most studies have focused only on drug-target networks (Yıldırım et al., 2007) or the molecular networks of traditional Chinese medicinal plants (Zhao et al., 2010).

In this study, we introduced an approach based on the integrated network of medicinal plants, chemical components, and multi-targets to analyze the complex biological roles of medicinal plants in human body. First, we constructed an integrated multipartite network of the biomedical entities including Korean medicinal plants, chemicals, and protein targets from heterogeneous databases. The databases have been curated for biomedical entities, such as genes, proteins, chemicals, and species and are publicly accessible. In particular, we employed the ChEMBL database, which is one of the

largest public databases of the activities of small molecules on biological targets (Koutsoukas et al., 2011; Martínez-Jiménez et al., 2013; Davies et al., 2015; Gaulton et al., 2017; Bosc et al., 2019). It also provides the built-in, ligand-based target prediction methods that we used to generate the links between chemicals with targets in the interaction network. We also used the database on compound ingredients extracted from Korean medicinal plants, which was released in 2007 by the Korean Traditional Knowledge Portal (http://www.koreantk.com) (Shin et al., 2010).

We then preprocessed the multipartite network and adapted the spectral co-clustering method to identify the subnetworks of closely related biomedical entities within their own structural properties, such as the average bipartite clustering coefficient (Dhillon, 2001). The clustering results could be used for the single or combination use of Korean medicinal plants in common target domains. The hierarchical clustering analysis was followed for further identification of the plant clusters based on the similarities in their chemical composition. In the analysis, we introduced a network-based mathematical index of interactions between the medicinal plants and targets, called the target potency score (TPS). The suggested index is used for ranking the plant clusters or protein targets, and for retrieving the best subnetworks of the medicinal plants, multi-chemicals, and multi-targets customized to specific medicinal purposes. Finally, we have demonstrated the utility of the simplified multipartite network with the hierarchical clusters.

2 Materials and Methods

2.1 Raw Data Sets

We employed several heterogeneous data sets to construct a multipartite network of medicinal plants, chemicals, and targets. The data set of the medicinal plants was obtained from the Korean Traditional Knowledge Portal (KTKP, http://www.koreantk.com). The database was built by the Korean Intellectual Property Office, Ministry of Trade, Industry and Energy, Republic of Korea (Shin, Jin-Sup et al., 2010). The KTKP data sets contain a list of 5,500 Korean medicinal plants and 10,056 chemicals, and their relations. The data set of chemicals and targets was obtained from the ChEMBL database (ChEMBL25, March 2019, http://ftp.ebi.ac.uk/pub/databases/chembl/target_predictions). The ChEMBL25 database contains 1,879,206 compounds, 12,482 targets, and 15,504,603 bioactivity entries from 72,271 publications. All the chemicals in the KTKP data sets were identified in the ChEMBL25 database, and the links between the chemicals and single-protein targets were generated using the prediction model reported previously (Koutsoukas et al., 2011; Martínez-Jiménez et al., 2013), and is available on the ChEMBL25 database. Specifically, the probability of the activation of a compound against a target was estimated by the 10-uM activity cut-off model in the ChEMBL25 database, and we determined the interaction between a pair of a chemical and a target at the probability of >0.9.

2.2 Preprocessing

We excluded non-plants from the raw dataset. For instance, minerals (non-living things) and animals (including corals) were removed, but the fungi were not. After excluding the non-plants, the raw data sets containing 5,500 instances of plants and non-plants were reduced to a set containing 3,614 plants. Then, the duplicated instances were removed based on their identical “scientific name,” “Latin name” and “use target”, which led to a remaining list of 2,886 plants in the data set.

The multipartite network of the medicinal plants, chemicals, and targets was constructed by combining the two bipartite networks: the plants-chemicals network and chemicals-targets network (**Figure 1A**). The nodes in the network represented the plants, chemicals, and targets, and the edges

in the network represented the relationships between the plants and chemicals, and between the chemicals and targets. From the multipartite network, we removed the nodes of degree zero, i.e., the plants, chemicals, or targets that were not linked to any other entities. After all, 1,138 plants, 10,043 chemicals, and 441 targets remained for the combined multipartite network. The basic statistics of the preprocessed network data are summarized in **Table 1**.

2.3 Methods

We analyzed the multipartite network using the spectral co-clustering algorithm (Dhillon, 2001). The algorithm clusters the medicinal plants, chemicals, and targets by solving a network partitioning problem. It partitions the set V of the connected nodes into k subsets, $V_1, V_2, \dots, V_k$ in such a way that the k subsets approximately minimize the normalized cut criterion:

$$Q(V_1, V_2, \dots, V_k) = \frac{1}{k}\sum_{i=1}^{k}\frac{cut(V_i, \bar{V}_i)}{vol(V_i)}$$

where $\bar{V}_i = V \backslash V_i$ is the complement of V_i, the $cut(M, N) = \sum_{m\in M, n\in N} w_{mn}$ of two disjoint sets M and N is defined by the sum of the edge weights, w_{mn}, between two nodes from each of M and N, and $vol(V_i) = \sum_{m\in V_i} d_m = \sum_{m\in V_i}\sum_n w_{mn}$ is the sum of degrees of all nodes in V_i. The normalized cut criterion represents the level of connectivity between the k subsets, where a smaller value indicates a weaker connectivity between them.

To determine the number of clusters, k, we computed the modularity, defined by

$$Modularity(k) = \frac{1}{2E}\sum_{i=1}^{k}\sum_{m\in V_i, n\in V_i}(w_{mn} - d_m d_n/2E)$$

where E is the number of edges and dm is the degree of a node m (Newman, 2004, 2006; Newman and Girvan, 2004). The modularity measures the strength of a partition with respect to the distribution of edges because it compares the number of edges, $\sum_{m,n} w_{mn}$, with the expected number of edges placed at random, $\sum_{m,n} d_m d_n/2E$, within the partition. An optimal number of clusters is determined at the maximum point of the modularity.

The hierarchical clustering analysis led to the identification of the clusters of medicinal plants based on their chemical profiles. The goal of the analysis was to provide insight into the biologically or chemically inter-related plants and search for their potentials medicinal effects in the integrated network. We use the Bray-Curtis distance for the hierarchical clustering, which is defined by, between two plants u and v,

$$d(u, v) = \frac{\sum_j |u_j - v_j|}{\sum_j |u_j| + |v_j|}$$

where u_j (or v_j) is equal to one if the plant u (or v)is connected to the jth chemical in the network, and zero otherwise. The Bray-Curtis distance takes a value between 0 and 1, and it has two favorable properties for the multipartite network analysis: (1) It is a scale-free distance metric, that is, it depends only on the fraction of the number of chemicals connected to the plants, rather than the total number of chemicals in the network. The scale-freeness is a crucial property for analyzing sparsely

connected networks, such as the multipartite network in this study where most of the u_j (or v_j) values are zero; see, for example, **Figure S1** in Supplemental Information for a visualization. (2) It is suitable for the unweighted networks because it only considers the number of edges while ignoring the edge weights.

We used the complete linkage method for hierarchical clustering, that is, the distance between two clusters is defined by the distance between two elements that are farthest from each other. The combined use of the Bray-Curtis distance and complete linkage yields a meaningful interpretation of clustering results. That is, if we combine two clusters one whenever they are at a distance of less than one, then every pair of elements will have a distance less than one. It implies that every pair of medicinal plants in a cluster will have at least one common chemical because $d(u, v) < 1$ implies that $u_j = v_j = 1$ for some chemical j.

Finally, we proposed an index called TPS, which is a similarity measure between a plant cluster and a target. TPS of a plant cluster P to a target T is defined by the fraction

$$TPS(P,T) = \frac{|C_P \cap C_T|}{|C_T|} = \frac{number\ of\ chemicals\ connected\ to\ both\ P\ and\ T}{number\ of\ chemicals\ connected\ to\ T}$$

where C_P is the set of all chemicals connected to at least one plant in the plant cluster P, C_T is the set of all chemicals connected to the target T, and $|\cdot|$ denotes the number of elements in a set. TPS can represent potential efficacy of a plant cluster for predicting a protein target. We used the TPS to search for the multipartite network for plant clusters of the highest efficacy or most active targets.

3 Results

3.1 Multipartite Network Analysis

We constructed the multipartite network with the nodes of medicinal plants, chemicals, and targets for the interpretation of the biological roles of these chemicals and plants containing them on multi-targets. The preprocessed network is shown in **Figure 1B**, where 10% of the nodes are randomly selected for visualization. Note that the plants and targets have connections only to the chemicals. **Figure 1C** shows the distribution of the node degrees for the plants, chemicals, or targets, where the vertical axes are on a log scale. The distribution of the power-law degree imply that the network is a scale-free network (Goh et al., 2002; Newman, 2005; Lima-Mendez and van Helden, 2009).

The structural properties of the multipartite network were further analyzed by the network analysis techniques. We divided the network into five subnetworks by applying the spectral co-clustering algorithm. Then, the structural properties of the multipartite network and the five subnetworks were computed and summarized in **Table S1** in Supplemental Information. The computed structural properties are a standard set of theoretic graphical measures, such as the diameter, average bipartite clustering coefficient, and degree of assortativity. **Figure S3** in Supplemental Information illustrates the multipartite network, modularity, and five subnetworks, where only 10% of the whole nodes were randomly selected for visualization.

3.2 Hierarchical Clustering Analysis of Plants

We performed the hierarchical clustering analysis of the medicinal plants in the multipartite network, and generated dendrograms to visually inspect the clusters of the medicinal plants (**Figure 2** and **Figure S2** in Supplemental Information). In the two largest clusters of the plants (**Figure 2A** and **2C**,

192 **Table S2** and **S3** in Supplemental Information), the same species with different usage parts have
193 been aggregated at the bottom level (i.e., at the height of zero) in the dendrogram, and different
194 species belonging to the same family are grouped together at low levels. However, some species
195 belonging to the same family are far apart or the ones belonging to a different family are closely
196 located. In the largest cluster, many different species sharing β-sitosterol, which is one of the most
197 common phytosterols and is distributed over a wide range in the plant kingdom, existed (Tao et al.,
198 2019); many plants belonging to different families share more chemicals than the ones belonging to
199 the same family (**Figure 2A**). The chemical diversity of the cluster shows that the specialized
200 distribution of the secondary metabolites in plants is not dependent on their family level. However, in
201 the second largest cluster, the chemical profiles of the plants depending on the family contributed to
202 the classification of the smaller clusters (**Figure 2C**). Additionally, the chemical diversity of the
203 plants in the second largest cluster (123 types of subclasses and 311 types of direct parent levels) was
204 bigger than the largest cluster (109 types of subclasses and 257 types of direct parent levels) (**Figures**
205 **2B** and **2D**). Among the smaller clusters, two clusters containing the specialized types of chemicals,
206 protoberberine alkaloids (**Figures 2E** and **2F**, and **Table S4** in Supplemental Information) and
207 vitamins (**Figure 2G** and **2H**, **Table S5** in Supplemental Information), were clearly categorized by
208 the family. We expect that the clustering information by the chemical profiles between the plants can
209 be utilized not only for the chemotaxonomic study of the medicinal plants, but also for alternative use
210 of the medicinal plants.

211 **3.3 Interaction Between Plants and Targets Based on TPS**

212 In the multipartite network, the interaction between the plants and targets is mediated by the
213 connectivity of the chemicals in the network. We used the TPS to quantitatively measure the
214 interaction that represents the potential efficacy of the plants for the targets. Because the plants in the
215 same cluster share chemicals in common, the TPS was calculated between a cluster of plants and
216 target.

217 The TPS was applied in two ways. First, the targets with the highest TPS values were selected for
218 each of the four clusters of plants that were mentioned in the previous section and were manually
219 investigated to validate our approaches. The subnetworks of the selected targets, chemicals connected
220 to the targets, and corresponding cluster of plants connected to the chemicals are illustrated in
221 **Figures 3A-D**, for the inspection of the effects of the mixture of chemicals or the plants in the
222 cluster. Specifically speaking, the targets with the TPS values greater than 0.05~0.5 were selected
223 while the number of chemicals connected to the targets was kept sufficiently large enough (at least
224 20) to produce reliable TPS values.

225 In each network, a single target had multiple edges with chemicals, but only a few chemicals
226 contributed to the connection between the plants and targets. In the largest cluster, six targets, signal
227 transducer and activator of transcription 3 (CHEMBL4026), phosphodiesterase 10A
228 (CHEMBL4409), anandamide amidohydrolase (CHEMBL3229), prostanoid EP2 receptor
229 (CHEMBL4909), sphingosine 1-phosphate receptor Edg-3 (CHEMBL3892), and pepsin A
230 (CHEMBL2714), are displayed (**Figure 3A**). For example, in case of phosphodiesterase 10A, which
231 has been studied as a therapeutic target for colon cancer and neurological symptoms (Li et al., 2015),
232 12 chemicals, mainly carboline-type alkaloids, are displayed with linkages to the target (**Table S6** in
233 Supplemental Information). The well-known targets, such as cannabinoid CB1 (CHEMBL218) and
234 CB2 (CHEMBL253) receptors, in the network from the second largest cluster (**Figures 2C** and **3B**)
235 shared the monoterpenoids and benzopyran-type chemicals including cannabinoids from the genus
236 *Cannabis* (Wu, 2019). In the third network, protoberberine-type alkaloids from three families,

237 Papaveraceae, Ranunculaceae, and Rutaceae, were linked with the targets, dopamine receptors
238 (CHEMBL217, CHEMBL265, CHEMBL1850, CHEMBL2056, and CHEMBL339), and alpha
239 adrenergic receptors (CHEMBL315 and CHEMBL1867), which are known as the main target of
240 these chemicals (Kawano et al., 2015; Fan et al., 2019). Tetraterpenoids, such as lycopene, lutein, and
241 β-carotene, which are the well-known chemicals targeting the retinoid signaling pathway, had
242 multiple linkages with retinoic acid receptor gamma (RAR-γ, CHEMBL2003) and retinoid X
243 receptor beta (RXR-β, CHEMBL4047) (Aydemir et al., 2012, 2013; Huang et al., 2016). These
244 results show that the target information predicted by the chemicals in the multipartite network is
245 reliable enough to investigate alternative targets of the chemicals, for which biological roles have not
246 been defined in previous studies or they existed as mixed forms with other chemicals in a plant.

247 Second, the proposed TPS was applied for searching the closest related plants, which are predictive
248 of a given target based on the highest TPS value. The results provide clues to understand the complex
249 biological roles of the combination of medicinal plants or the chemicals produced from the plants. In
250 this scenario, a cluster of plants is recommended from among the large-scale network as the
251 candidate medicinal plant for a single target of interest. For example, we used 'tyrosinase'
252 (CHEMBL3318) as the model target to validate the suggested method and to investigate the
253 medicinal effects of the retrieved cluster of plants (**Figure 4A**). In the multipartite network retrieved
254 for the target "tyrosinase," 165 chemicals from 31 different species were linked with the target node.
255 Among them, 50 chemicals from *Morus alba* were found in the middle of the network (**Table S7** in
256 Supplemental Information). Additionally, *Carthamus tinctorius*, *Cudrania tricuspidata*, *Glycyrrhiza*
257 *inflata,* and *Glycyrrhiza uralensis* had more than 10 chemicals, which were predicted to act on
258 tyrosinases (**Table S8** in Supplemental Information). These five plants have been known as the most
259 popular and successful sources for the whitening products used in the cosmetic industry (Smit et al.,
260 2009; Xin et al., 2017).

261 Next, we combined three subnetworks-networks from three targets, steroid 5α-reductase 1 (S5R1,
262 ChEMBL1787), steroid 5α-reductase 2 (S5R2, ChEMBL1856), and androgen receptor (AR,
263 ChEMBL1871), which are involved in the androgen metabolism, for the interaction of plants and
264 chemicals on multi-targets (**Figure 4B**). In total, 44 among 52 and 54 chemicals from S5R1 and
265 S5R2, respectively, shared with each other, but 10 of 52 from S5R1 and 7 of 54 from S5R2 only
266 shared those with 58 chemicals from AR (**Table S9-11** in Supplemental Information). The types of
267 chemicals between the pairs of targets showed significantly different patterns. The chemicals co-
268 targeting S5R1 and S5R2 were mainly the steroid derivatives, such as cholestanes, ergostanes, and
269 stigmastanes, which were synthesized via the triterpenoid pathways (**Figure 4C**); however,
270 sesquiterpenoids and diterpenoids were the main components sharing the pairs of S5R1 and AR, or
271 S5R2 and AR (**Figure 4D** and **4E**). We found that only 6 chemicals simultaneously acted on three
272 targets and plants containing them (**Figure 5** and **Table 2**).

273 4 Discussion

274 In this study, we tried to unravel the complex interaction between multi-chemicals and multiple
275 targets using the computational target prediction model and multipartite networks comprising the
276 relationship between the plants, chemicals, and targets. In hierarchical cluster analysis, plants
277 producing similar specialized metabolites are clustered only within a taxon. In particular, in
278 chemotaxonomy, secondary metabolites can be used as specialized markers to distinguish among
279 plants (Wink, 2003). However, not all secondary metabolites are produced in the same taxon. For
280 example, flavonoids are widely distributed in plants and even in fungi, fulfilling many functions,
281 such as ultraviolet filtration, symbiotic nitrogen fixation, and floral pigmentation (Kumar and

282 Pandey, 2013). The chemical profile of each plant has a minor effect on the classification of the
283 plant.

284 To unravel the clusters of multipartite networks from hierarchical clustering analysis, we introduced
285 an index of potential efficacy of plant clusters toward a target, called TPS. TPS indicated the
286 proximity of subnetworks of the metabolic relationships between plant clusters and targets. We
287 utilized it to search for and suggest candidate plants and chemicals acting on specific targets (**Figure**
288 **3**). It was validated that the plants with more than 10 chemicals had enough experimental evidence to
289 explain the biological roles on specific targets retrieved in the network. TPS was also used for the
290 extraction of a list of plants and chemicals acting on a specific target of interest (**Figure 4**).

291 We also obtained potent plant and chemical candidates for the specific target, "tyrosinase,"
292 confirming the validity of the approach. The node information of the plants in the network is
293 expected to guide toward new alternative sources developed by plant combinations.

294 The approach is based on the multipartite network of plants and chemicals for a single target using
295 the TPS and could be expanded for multi-targets that closely interact with each other. The plants and
296 chemicals, which are expected to act on multiple targets, including S5R1, S5R2, and AR, could be
297 suggested to control the levels of testosterone (T) and dihydrotestosterone (DHT), involved in the
298 androgen metabolism. It is known that the blockade of AR can decrease benign prostate hyperplasia
299 (BPH) symptoms and further relieve the lower urinary tract symptoms (Izumi et al., 2013). Since the
300 alterations in AR are associated with prostate cancer, the inhibition of AR by drugs is the prevention
301 target for prostate cancer growth (Tan et al., 2015). The 5α-reductase inhibitors, which suppress the
302 conversion of T into DHT, contribute to the reduction in BPH or preventing prostate cancer (Izumi et
303 al., 2013; Liss and Thompson, 2018). In the present study, the AR-targeted chemicals were
304 sesquiterpenoid and diterpenoid-type, while the steroid derivatives were suggested as the chemicals
305 targeting 5α-reductases (S5R1 and S5R2). We expected that the plants containing the chemicals
306 acting on all three targets, S5R1, S5R2, and AR, can be the new candidates for controlling the
307 symptoms or diseases involving androgen metabolism, such as BPH and prostate cancer. The well-
308 organized multipartite network is expected to explain the synergistic effects between the multi-
309 chemicals in single or multiple medicinal plants.

310 We have constructed a multipartite network by integrating the databases of the medicinal plants,
311 chemicals, and targets for the prediction of the biological roles of the plants and chemicals on
312 specific targets. We have presented a novel computational approach based on the multipartite
313 network with which the plant species can be distinguished based on their chemical profiles. In the
314 clustered bipartite network of plants and chemicals, the chemical profiles of the plants in identical or
315 similar species were clustered closely at the bottom level of the dendrogram. However, the plants in
316 the same family were found far apart in the dendrogram in many cases, which implies that the
317 chemical profiles of the plants do not significantly contribute to the classification of the species.

318 In conclusion, we have demonstrated that the network analysis helps to understand the biological
319 roles of multi-chemicals on multi-targets. With the target prediction models, it also provides a way to
320 discover alternative therapeutic agents from natural sources by extrapolating from the known
321 bioactivities of chemicals on targets. The predicted results about the biological roles of multi-
322 chemicals on multi-targets allow to explain the therapeutic activity of the medicinal plant extracts via
323 subnetworks of molecular mechanisms, and suggest the combinatory preparation of several medicinal
324 plants on multi-targets to be involved in the symptoms or diseases of interest. In our future study, we

will study the candidates, plants or chemicals, predicted from the multipartite network in *in vitro* and *in vivo* models to validate our approaches and further develop novel therapeutic botanical drugs.

5 References

Ahn, K. (2017). The worldwide trend of using botanical drugs and strategies for developing global drugs. BMB Rep. 50, 111-116.

Anighoro, A., Bajorath, J., and Rastelli, G. (2014). Polypharmacology: challenges and opportunities in drug discovery: miniperspective. J. Med. Chem. 57, 7874-7887.

Aydemir, G., Carlsen, H., Blomhoff, R., and Rühl, R. (2012). Lycopene induces retinoic acid receptor transcriptional activation in mice. Mol. Nutr. Food Res. 56, 702-712.

Aydemir, G., Kasiri, Y., Birta, E., Béke, G., Garcia, A. L., Bartók, E.-M., et al. (2013). Lycopene-derived bioactive retinoic acid receptors/retinoid-X receptors-activating metabolites may be relevant for lycopene's anti-cancer potential. Mol. Nutr. Food Res. 57, 739-747.

Besnard, J., Ruda, G. F., Setola, V., Abecassis, K., Rodriguiz, R. M., Huang, X. P., et al. (2012). Automated design of ligands to polypharmacological profiles. Nature 492, 215-220.

Bosc, N., Atkinson, F., Felix, E., Gaulton, A., Hersey, A., and Leach, A. R. (2019). Large scale comparison of QSAR and conformal prediction methods and their applications in drug discovery. J. Cheminformatics 11.

Cheng, F., Kovács, I. A., and Barabási, A.-L. (2019). Network-based prediction of drug combinations. Nat. Commun. 10, 1197. doi:10.1038/s41467-019-09186-x.

Davies, M., Nowotka, M., Papadatos, G., Dedman, N., Gaulton, A., Atkinson, F., et al. (2015). ChEMBL web services: streamlining access to drug discovery data and utilities. Nucleic Acids Res. 43, W612-W620.

Dhillon, I. S. (2001). Co-clustering documents and words using bipartite spectral graph partitioning. in Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '01 (San Francisco, California: ACM Press), 269-274.

Ezzat, A., Wu, M., Li, X., and Kwoh, C.-K. (2019). "Computational Prediction of Drug-Target Interactions via Ensemble Learning," in Computational Methods for Drug Repurposing, ed. Q. Vanhaelen (New York, NY: Springer New York), 239-254.

Fan, J., Zhang, K., Jin, Y., Li, B., Gao, S., Zhu, J., et al. (2019). Pharmacological effects of berberine on mood disorders. J. Cell. Mol. Med. 23, 21-28.

Forouzesh, A., Samadi Foroushani, S., Forouzesh, F., and Zand, E. (2019). Reliable target prediction of bioactive molecules based on chemical similarity without employing statistical methods. Front. Pharmacol. 10, 835.

Galanakis, C. M. (2019). Nutraceuticals and natural product pharmaceuticals.

Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., et al. (2017). The ChEMBL database in 2017. Nucleic Acids Res. 45, D945-D954.

361 Goh, K.-I., Oh, E., Jeong, H., Kahng, B., and Kim, D. (2002). Classification of scale-free networks.
362 Proc. Natl. Acad. Sci. 99, 12583-12588.

363 Han, K., Jeng, E. E., Hess, G. T., Morgens, D. W., Li, A., and Bassik, M. C. (2017). Synergistic drug
364 combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. Nat.
365 Biotechnol. 35, 463-474.

366 Harvey, A. L., Edrada-Ebel, R., and Quinn, R. J. (2015). The re-emergence of natural products for
367 drug discovery in the genomics era. Nat. Rev. Drug Discov. 14, 111-129.

368 Hopkins, A. L. (2008). Network pharmacology: the next paradigm in drug discovery. Nat. Chem.
369 Biol. 4, 682-690.

370 Huang, X., Gao, Y., Zhi, X., Ta, N., Jiang, H., and Zheng, J. (2016). Association between vitamin A,
371 retinol and carotenoid intake and pancreatic cancer risk: Evidence from epidemiologic studies.
372 Sci. Rep. 6, 38936.

373 Izumi, K., Mizokami, A., Lin, W.-J., Lai, K.-P., and Chang, C. (2013). Androgen receptor roles in
374 the development of benign prostate hyperplasia. Am. J. Pathol. 182, 1942-1949.

375 Jia, J., Zhu, F., Ma, X., Cao, Z. W., Li, Y. X., and Chen, Y. Z. (2009). Mechanisms of drug
376 combinations: interaction and network perspectives. Nat. Rev. Drug Discov. 8, 111-128.

377 Kawano, M., Takagi, R., Kaneko, A., and Matsushita, S. (2015). Berberine is a dopamine D1- and
378 D2-like receptor antagonist and ameliorates experimentally induced colitis by suppressing
379 innate and adaptive immune responses. J. Neuroimmunol. 289, 43-55.

380 Keiser, M. J., Setola, V., Irwin, J. J., Laggner, C., Abbas, A. I., Hufeisen, S. J., et al. (2009).
381 Predicting new molecular targets for known drugs. Nature 462, 175-181.

382 Koehn, F. E., and Carter, G. T. (2005). The evolving role of natural products in drug discovery. Nat.
383 Rev. Drug Discov. 4, 206-220.

384 Koutsoukas, A., Simms, B., Kirchmair, J., Bond, P. J., Whitmore, A. V., Zimmer, S., et al. (2011).
385 From in silico target prediction to multi-target drug design: Current databases, methods and
386 applications. J. Proteomics 74, 2554-2574.

387 Kumar, S., and Pandey, A. K. (2013). Chemistry and Biological Activities of Flavonoids: An
388 Overview. Sci. World J. 2013, 1-16.

389 Lavecchia, A., and Cerchia, C. (2016). In silico methods to address polypharmacology: current
390 status, applications and future perspectives. Drug Discov. Today 21, 288-298.

391 Leung, E. L., Cao, Z.-W., Jiang, Z.-H., Zhou, H., and Liu, L. (2013). Network-based drug discovery
392 by integrating systems biology and computational technologies. Brief. Bioinform. 14, 491-505.

393 Li, N., Lee, K., Xi, Y., Zhu, B., Gary, B. D., Ramírez-Alcántara, V., et al. (2015). Phosphodiesterase
394 10A: a novel target for selective inhibition of colon tumor cell growth and β-catenin-dependent
395 TCF transcriptional activity. Oncogene 34, 1499-1509.

396 Lima-Mendez, G., and van Helden, J. (2009). The powerful law of the power law and other myths in
397 network biology. Mol. Biosyst. 5, 1482.

398 Liss, M. A., and Thompson, I. M. (2018). Prostate cancer prevention with 5-alpha reductase
399 inhibitors: concepts and controversies. Curr. Opin. Urol. 28, 42-45.

400 Lounkine, E., Keiser, M. J., Whitebread, S., Mikhailov, D., Hamon, J., Jenkins, J. L., et al. (2012).
401 Large-scale prediction and testing of drug activity on side-effect targets. Nature 486, 361-367.

402 Luo, Y., Zhao, X., Zhou, J., Yang, J., Zhang, Y., Kuang, W., et al. (2017). A network integration
403 approach for drug-target interaction prediction and computational drug repositioning from
404 heterogeneous information. Nat. Commun. 8, 573.

405 Martínez-Jiménez, F., Papadatos, G., Yang, L., Wallace, I. M., Kumar, V., Pieper, U., et al. (2013).
406 Target Prediction for an Open Access Set of Compounds Active against Mycobacterium
407 tuberculosis. PLoS Comput. Biol. 9, e1003253.

408 Monteiro, N. R. C., Ribeiro, B., and Arrais, J. (2020). Drug-target interaction prediction: end-to-end
409 deep learning approach. IEEE/ACM Trans. Comput. Biol. Bioinform.

410 NCI-DREAM Community, Bansal, M., Yang, J., Karan, C., Menden, M. P., Costello, J. C., et al.
411 (2014). A community computational challenge to predict the activity of pairs of compounds.
412 Nat. Biotechnol. 32, 1213-1222.

413 Newman, M. (2005). Power laws, Pareto distributions and Zipf's law. Contemp. Phys. 46, 323-351.

414 Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. Phys. Rev.
415 E 69, 066133.

416 Newman, M. E. J. (2006). Modularity and community structure in networks. Proc. Natl. Acad. Sci.
417 103, 8577-8582.

418 Newman, M. E. J., and Girvan, M. (2004). Finding and evaluating community structure in networks.
419 Phys. Rev. E 69, 026113.

420 Petrovska, B. (2012). Historical review of medicinal plants′ usage. Pharmacogn. Rev. 6, 1.

421 Ramsay, R. R., Popovic-Nikolic, M. R., Nikolic, K., Uliassi, E., and Bolognesi, M. L. (2018). A
422 perspective on multi-target drug discovery and design for complex diseases. Clin. Transl. Med.
423 7, 3.

424 Rodrigues, T., Reker, D., Schneider, P., and Schneider, G. (2016). Counting on natural products for
425 drug design. Nat. Chem. 8, 531-541.

426 Roth, B. L., Sheffler, D. J., and Kroeze, W. K. (2004). Magic shotguns versus magic bullets:
427 selectively non-selective drugs for mood disorders and schizophrenia. Nat. Rev. Drug Discov.
428 3, 353-359.

429 Shin, Jin-Sup, Lee, Yu-Seon, and Lee, Myeong-Seon (2010). Protection and utilization of
430 traditional knowledge resources through Korean traditional knowledge portal(KTKP). J. Korea
431 Contents Assoc. 10, 422-426.

432 Smit, N., Vicanova, J., and Pavel, S. (2009). The hunt for natural skin whitening agents. Int. J. Mol.
433 Sci. 10, 5326-5349.

434 Sun, Y., Sheng, Z., Ma, C., Tang, K., Zhu, R., Wu, Z., et al. (2015). Combining genomic and
435 network characteristics for extended capability in predicting synergistic drugs for cancer. Nat.
436 Commun. 6, 8481.

437 Tan, M. E., Li, J., Xu, H. E., Melcher, K., and Yong, E. (2015). Androgen receptor: structure, role in
438 prostate cancer and drug discovery. Acta Pharmacol. Sin. 36, 3-23.

439 Wink, M. (2003). Evolution of secondary metabolites from an ecological and molecular phylogenetic
440 perspective. Phytochemistry 64, 3-19.

441 Wu, J. (2019). Cannabis, cannabinoid receptors, and endocannabinoid system: yesterday, today, and
442 tomorrow. Acta Pharmacol. Sin. 40, 297-299.

443 Xin, L.-T., Yue, S.-J., Fan, Y.-C., Wu, J.-S., Yan, D., Guan, H.-S., et al. (2017). Cudrania
444 tricuspidata: an updated review on ethnomedicine, phytochemistry and pharmacology. RSC
445 Adv. 7, 31807-31832.

446 Yıldırım, M. A., Goh, K.-I., Cusick, M. E., Barabási, A.-L., and Vidal, M. (2007). Drug—target
447 network. Nat. Biotechnol. 25, 1119-1126.

448 Zhang, C., Hong, H., Mendrick, D. L., Tang, Y., and Cheng, F. (2015). Biomarker-based drug safety
449 assessment in the age of systems pharmacology: from foundational to regulatory science.
450 Biomark. Med. 9, 1241-1252.

451 Zhang, W., Pei, J., and Lai, L. (2017). Computational multitarget drug design. J. Chem. Inf. Model.
452 57, 403-412.

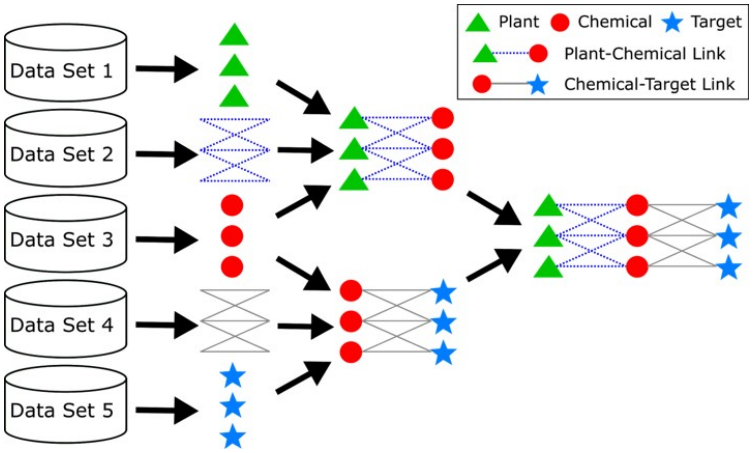
453 Zhao, J., Jiang, P., and Zhang, W. (2010). Molecular networks for the study of TCM Pharmacology.
454 Brief. Bioinform. 11, 417-430.

455 Zheng, M., Liu, X., Xu, Y., Li, H., Luo, C., and Jiang, H. (2013). Computational methods for drug
456 design and discovery: focus on China. Trends Pharmacol. Sci. 34, 549-559.

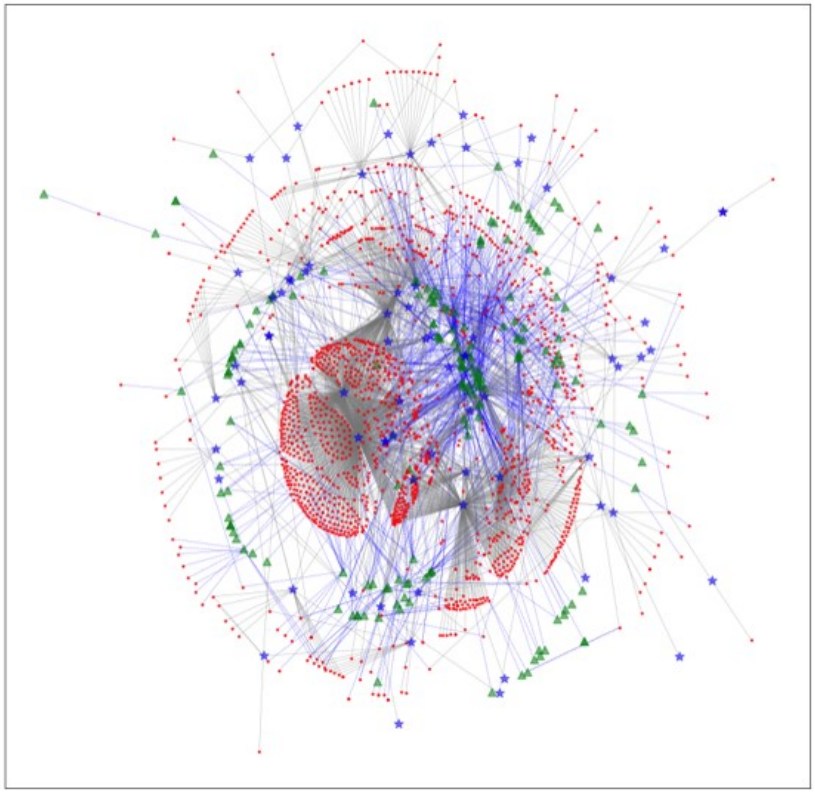
457

458

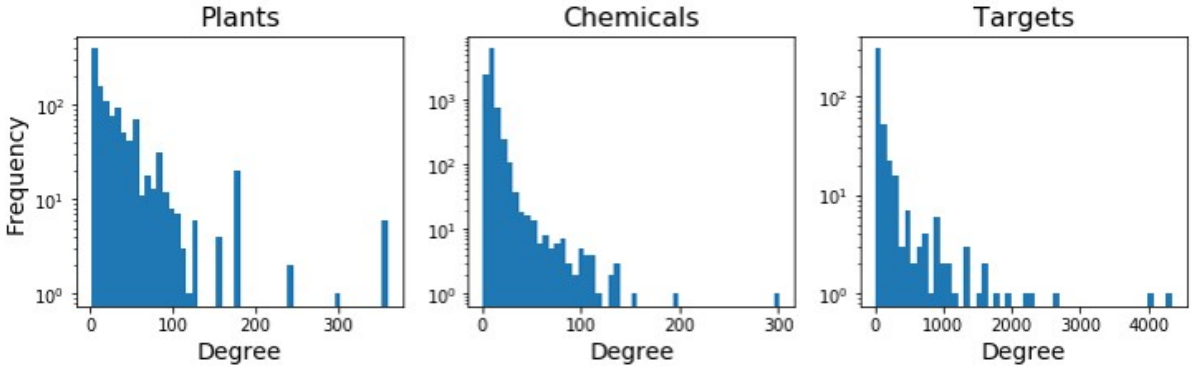
Figures and Tables



(A)



(B)



(C)

Figure 1. (A) The construction and visualization of the multipartite network database. **(B)** The representative multipartite network generated from the nodes which were randomly selected from 10% of the nodes 1,138 plants, 10,043 chemicals, and 441 targets in the whole network. **(C)** The distribution of the node degrees for the plants, chemicals, or targets.

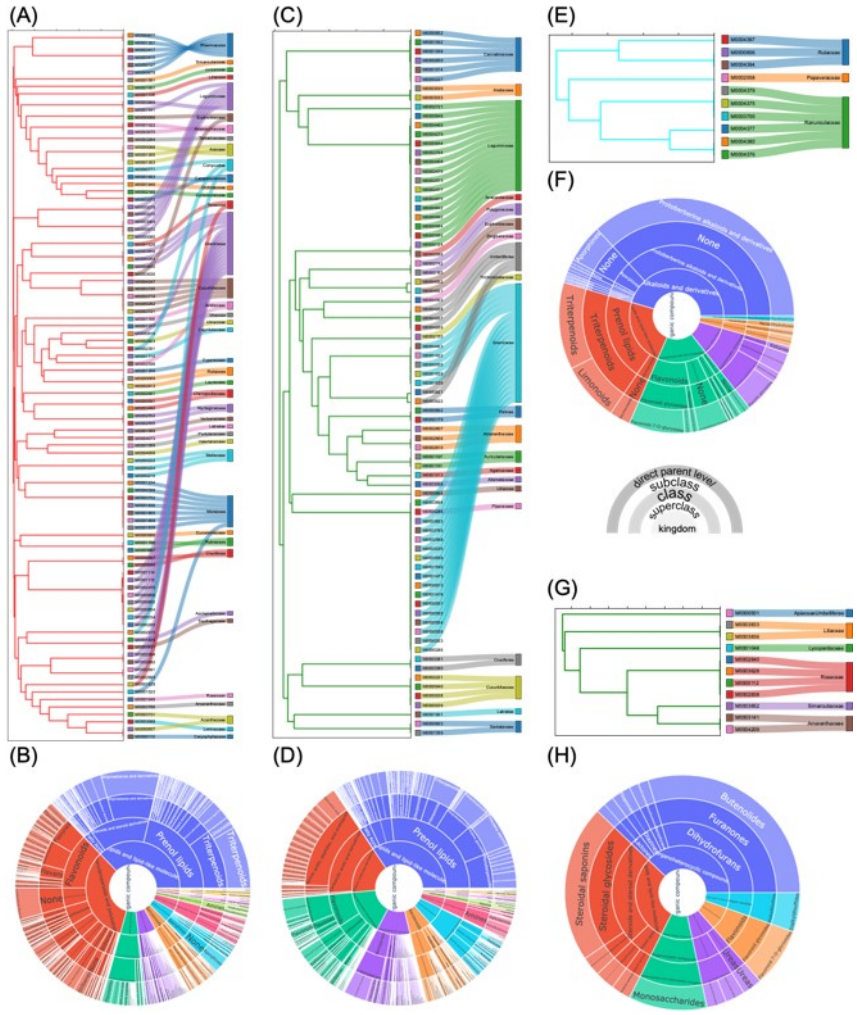


Figure 2. The four sub-clusters in the whole dendrogram (**Figure S2**). **(A)** The largest sub-cluster in the whole network. **(B)** The classification of chemicals in **(A)**. **(C)** The second largest sub-cluster in the whole network. **(D)** The classification of chemicals in **(C)**. **(E)** The sub-cluster of the nodes sharing protoberberine-type chemicals in the whole network. **(F)** The classification of chemicals in **(E)**. **(G)** The sub-cluster of the nodes sharing vitamins in the whole network. **(H)** The classification of chemicals in **(G)**.

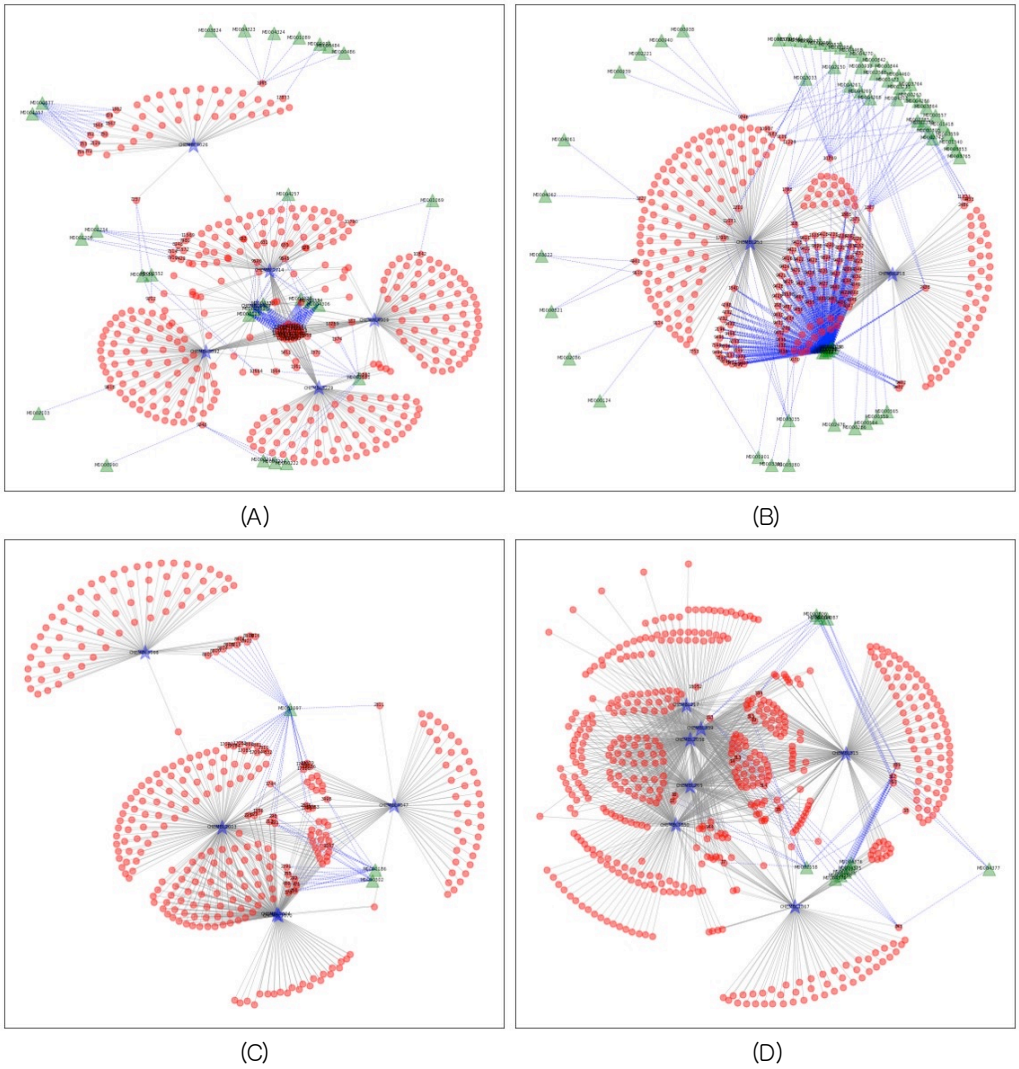
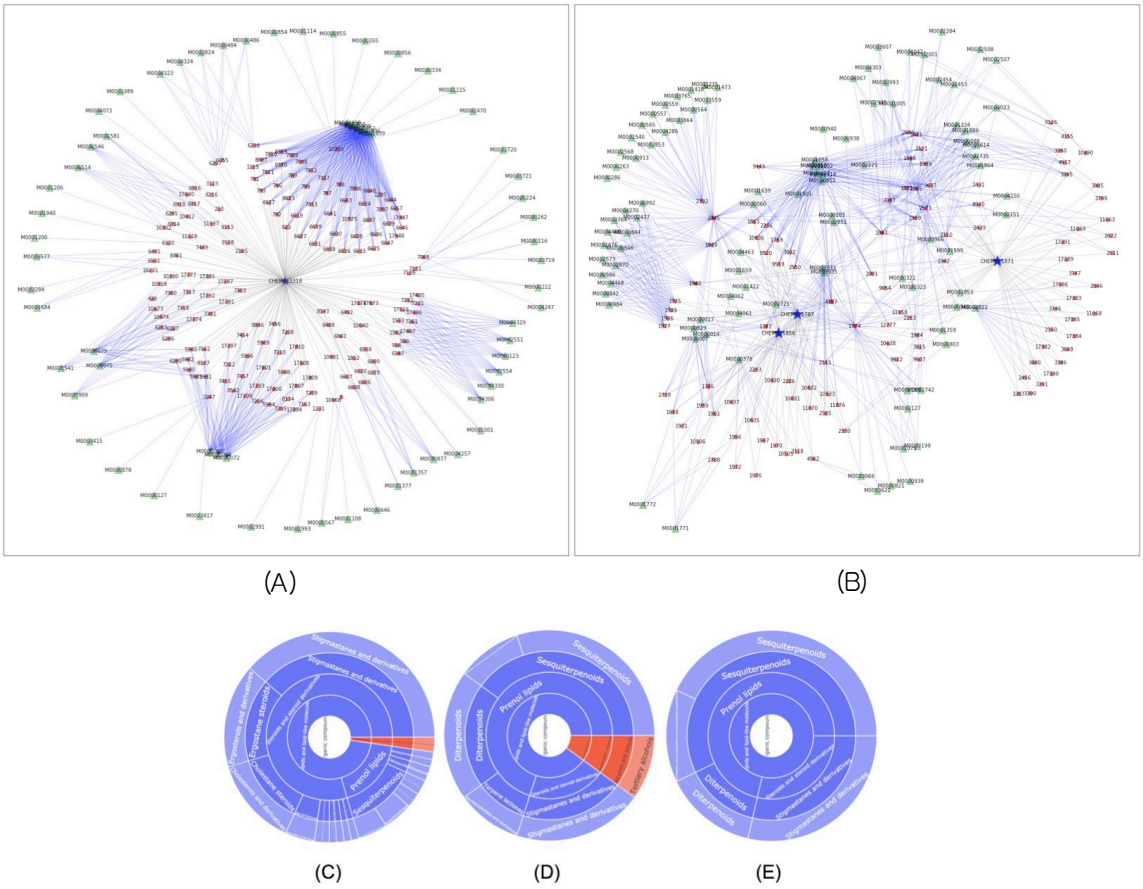


Figure 3. (A) The network diagram for the plants cluster in **Figure 2A** and the 6 targets retrieved based on the TPS index greater than or equal to 0.1. The chemicals effective to the targets are depicted in red circles. **(B)** The network diagram for the plants cluster in **Figure 2C** and the 2 targets retrieved based

481 on the TPS index greater than or equal to 0.5. The chemicals effective to the targets are depicted in red
482 circles. **(C)** The network diagram for the plants cluster in **Figure 2E** and the 7 targets retrieved based
483 on the TPS index greater than or equal to 0.05. The chemicals effective to the targets are depicted in
484 red circles. **(D)** The network diagram for the plants cluster in **Figure 2G** and the 5 targets retrieved
485 based on the TPS index greater than or equal to 0.05. The chemicals effective to the targets are depicted
486 in red circles.



487

488 **Figure 4. (A)** The network diagram for the plants cluster having the largest TPS index to the target
489 'Tyrosinase' (CHEMBL3318). **(B)** The combined network generated from three network diagrams for
490 the plants cluster having the largest TPS index to the three targets, 'steroid 5α-reductase 1' (S5R1,
491 ChEMBL1787), 'steroid 5α-reductase 2' (S5R2, ChEMBL1856) and 'androgen receptor' (AR,
492 ChEMBL1871). The chemical profiles sharing from chemicals from S5R1 and S5R2 **(C),** from S5R1
493 and AR **(D)** and from S5R2 and AR **(E)**.

Table 1. Basic statistics of the combined multipartite network data of plants, chemicals, and targets. The density of a multipartite network is computed by E/(Np*Nc+Nc*Nt).

	Number of Nodes				Number of Edges			Density
	Plant (Np)	Chemical (Nc)	Target (Nt)	Total (N)	Plant-Chemical	Chemical-Target	Total (E)	
Raw Data	5,500	10,056	1,224	16,780	58,068	100,290	158,358	0.00234
without Non-plants	3,614	10,056	1,224	14,894	54,585	100,290	154,875	0.00318
without Non-plants or Duplicates	2,886	10,056	1,224	14,166	34,549	100,290	134,839	0.00326
Only targets with the probability of above 0.9	2,886	10,056	441	13,383	34,549	73,112	107,661	0.00322
Pre-processed Data	1,138	10,043	441	11,622	34,549	73,112	107,661	0.00679

498 **Table 2**. List of chemicals simultaneously acting on three targets, S5R1, S5R2 and AR.

Chemical ID	Chemical Name	inchikey	Chemicals classification				
			Kingdom	Superclass	Class	Subclass	Direct_parent
9612	Kusunol	MQWIFDHBNGIVPO-UHFFFAOYSA-N	Organic compounds	Lipids and lipid-like molecules	Prenol lipids	Sesquiterpenoids	Eremophilane, 8,9-secoeremophilane and furoeremophilane sesquiterpenoids
10638	Stigmasta-4,6-dien-3-one	KEAZWUZFBSXOMV-UHFFFAOYSA-N	Organic compounds	Lipids and lipid-like molecules	Steroids and steroid derivatives	Stigmastanes and derivatives	Stigmastanes and derivatives
2001	Stigmastan-3-one (5-alpha)	BVVFRHKBULZQCQ-UHFFFAOYSA-N	Organic compounds	Lipids and lipid-like molecules	Steroids and steroid derivatives	Stigmastanes and derivatives	Stigmastanes and derivatives
11158	Dehydroabietinal	YCLCHPWRGSDZKL-UHFFFAOYSA-N	Organic compounds	Lipids and lipid-like molecules	Prenol lipids	Diterpenoids	Diterpenoids

1847	(-)-alpha-Copaene	VLXDPFLIRFYIME-UHFFFAOYSA-N	Organic compounds	Lipids and lipid-like molecules	Prenol lipids	Sesquiterpenoids	Sesquiterpenoids
9654	(-)-Ylangene	VLXDPFLIRFYIME-UHFFFAOYSA-N	Organic compounds	Lipids and lipid-like molecules	Prenol lipids	Sesquiterpenoids	Sesquiterpenoids

499