

质谱 + 网络药理学分析

2024-02-01

LiChuang Huang



@ 立效研究院

Contents

1 摘要	1
2 前言	1
3 材料和方法	1
3.1 材料	1
3.2 方法	1
4 分析结果	2
5 结论	2
6 附：分析流程	2
6.1 已被鉴定的成分	2
6.2 化合物信息	3
6.2.1 分类学	3
6.2.2 化合物靶点	4
6.3 网络药理学	5
6.3.1 靶点网络 (+ 化合物含量)	5
6.3.2 炎症	5
6.3.3 靶点网络 (+ 化合物含量) + 炎症	6
6.3.4 相关靶点与转录因子 FOXO 相关	7
6.3.5 富集分析	8
6.3.6 口服利用度筛选	9
6.3.7 最终筛选过的网络药理图	10
6.4 分子对接	11
6.4.1 对接的成分	11
6.4.2 对接结果	11
Reference	13

List of Figures

1 Classification hierarchy	3
2 Compounds classify	4
3 Network pharmacology	5
4 Overall targets number of datasets	6
5 Network pharmacology with disease	6
6 Targets intersect with targets of diseases	7
7 Ids KEGG enrichment	8
8 Ids GO enrichment	9

9	HOB 20 prediction	10
10	FINAL Network pharmacology	11
11	Overall combining Affinity	12

List of Tables

1	Identified compounds records in table CompoundDiscovery	2
2	Targets predicted by Super Pred	4
3	Transcription Factor binding sites	7
4	Overall combining Affinity rawData	12

1 摘要

需求:

- 质谱对复方内的五味药草匚、土茯苓、泽泻、川牛膝和生米仁进行预测，得到复方药物的具体的有效药物成分 XX1、XX2、XX3
- 分子对接检测 XX1、XX2、XX3 与 FOXO 信号的对接能量
- 通过 GO 功能富集分析和 KEGG 通路富集分析，对 FOXO 信号通路的下游靶点进行预测，进而得到所需关于抑制炎症的下游靶点 YY

结果:

- 以网络药理学的方式筛选成分:
 - 获取所有成分的靶点 (Tab. 2)
 - 获取炎症的靶点 (Fig. 4)
 - 炎症和疾病靶点交集 (Fig. 6)
 - 过滤，以转录因子 FOXO 可结合的靶点 (Tab. 3)
 - 将上述靶点富集分析，聚焦到 GO 显著结果：通路 regulation of inflammatory response (Fig. 8)
 - 获取对应富集到该通路的靶点的基因，以及靶向这些基因的成分
 - 对这些成分再次以 HOB 口服利用度过滤 (Fig. 9)
 - 上述步骤之后的网络图见 Fig. 10
 - 最后获取含量(峰面积)前 5 的成分，见 6.4.1
- 分子对接结果见 Fig. 11, Tab. 4
- 富集分析是以 FOXO 转录因子结合靶点的基础上富集的，可能的通路为 regulation of inflammatory response，对应的靶点见 6.3.5

2 前言

3 材料和方法

3.1 材料

3.2 方法

Mainly used method:

- Database PubChem used for querying information (e.g., InChIKey, CID) of chemical compounds; Tools of Classyfire used for get systematic classification of chemical compounds^{1,2}.
- R package ClusterProfiler used for gene enrichment analysis³.
- Databases of DisGeNet, GeneCards, PharmGKB used for collating disease related targets⁴⁻⁶.
- Python tool of HOB was used for prediction of human oral bioavailability⁷.
- Python tool doctr <https://github.com/mindee/doctr> used for interpreted character from images.
- R package PubChemR used for querying compounds information.

- Web tool of Super-PRED used for drug-targets relationship prediction⁸.
- The Transcription Factor Target Gene Database (<https://tfbsdb.systemsbiology.net/>) was used for discovering relationship between transcription factors and genes..⁹
- The CLI tools of AutoDock vina and ADFR software used for auto molecular docking¹⁰⁻¹⁴.
- Other R packages (eg., dplyr and ggplot2) used for statistic analysis or data visualization.

4 分析结果

5 结论

6 附：分析流程

6.1 已被鉴定的成分

Table 1 (下方表格) 为表格 Identified compounds records in table CompoundDiscovery 概览。

(对应文件为 [Figure+Table/Identified-compounds-records-in-table-CompoundDiscovery.xlsx](#))

注：表格共有 129 行 7 列，以下预览的表格可能省略部分数据；表格含有 129 个唯一 ‘en.name’。

1. en.name: 化合物英文名
2. cn.name: 化合物中文名
3. rt.min: 保留时间 (分钟)
4. formula: 化合物分子式
5. mw: 分子质量
6. peak_area: 峰面积

Table 1: Identified compounds records in table CompoundDiscovery

en.name	cn.name	rt.min	formula	mw	file_area	peak_area
Betaine	甜菜碱	1.53	C5H11NO2	117.0791	compound_d...	5.99e+09
Astilbin	落新妇苷	21.96	C21H22O11	450.1163	compound_d...	4.53e+09
Sucrose	蔗糖	1.58	C12H22O11	342.1162	compound_d...	2.3e+09
Citric acid	柠檬酸	1.67	C6H8O7	192.027	compound_d...	1.13e+09
Pseudoprot...	伪原薯蓣皂苷	24.66	C51H82O21	1030.5348	compound_d...	8.76e+08
Raffinose	棉籽糖	1.57	C18H32O16	252.0846	compound_d...	7.83e+08
Protodioscin	原薯蓣皂苷	24.66	C51H84O22	1048.5452	compound_d...	7.11e+08
Trigonell...	盐酸葫芦巴碱	1.61	C7H7NO2	137.0479	compound_d...	6.38e+08
Taxifolin	二氢槲皮素	22	C15H12O7	304.0584	compound_d...	6.34e+08
Engeletin	黄杞苷	22.9	C21H22O10	434.1214	compound_d...	5.57e+08
2-Pyrrolid...	L-脯氨酸	1.58	C5H9NO2	115.0635	compound_d...	4.06e+08

en.name	cn.name	rt.min	formula	mw	file_area	peak_area
Stachyose	水苏糖	1.6	C24H42O21	666.2221	compound_d...	3.91e+08
Cyasterone	杯苋甾酮	22.75	C29H44O8	566.3094	compound_d...	3.27e+08
5-Hydroxym...	5-羟甲基糠醛	1.53	C6H6O3	126.0319	compound_d...	2.07e+08
Maltopentaose	麦芽五糖	1.61	C30H52O26	828.2749	compound_d...	2.07e+08
...

6.2 化合物信息

6.2.1 分类学

Figure 1 (下方图) 为图 Classification hierarchy 概览。

(对应文件为 Figure+Table/Classification-hierarchy.pdf)

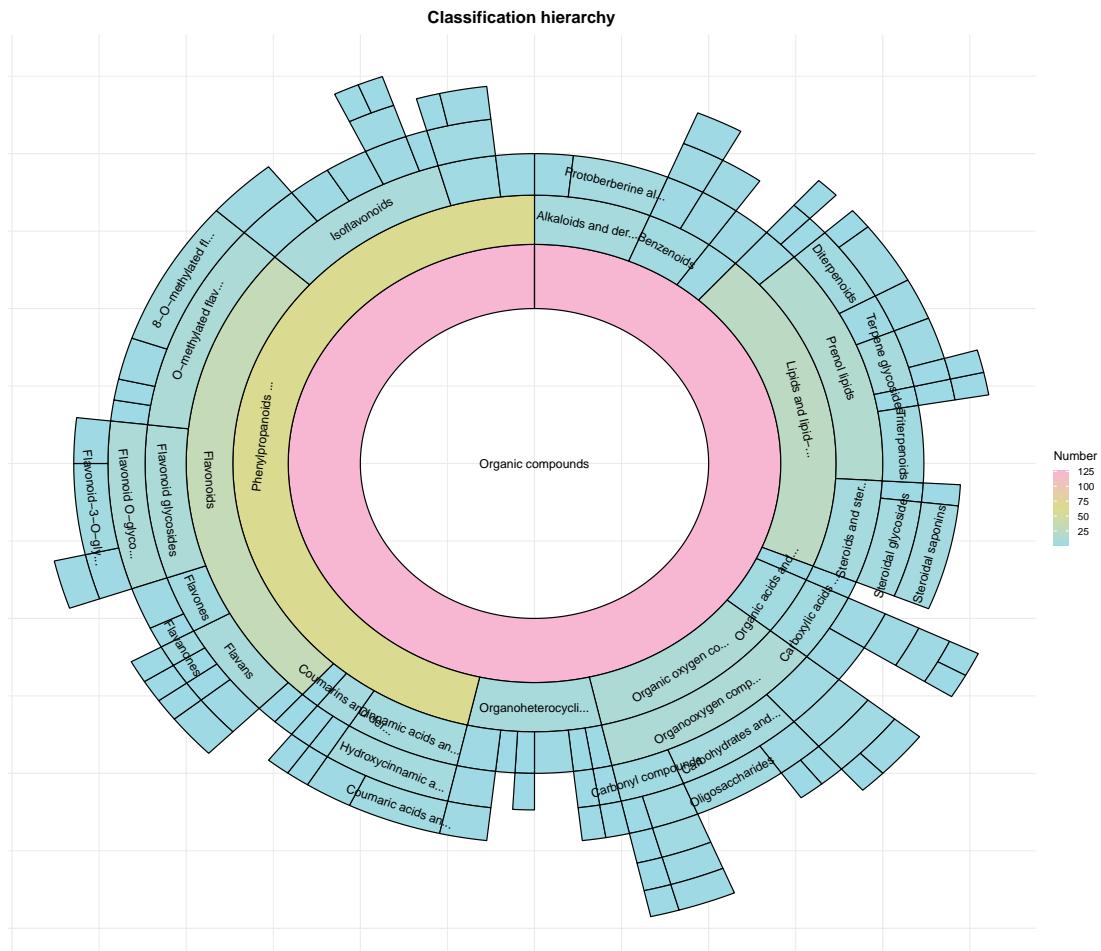


Figure 1: Classification hierarchy

Figure 2 (下方图) 为图 Compounds classify 概览。

(对应文件为 Figure+Table/Compounds-classify.pdf)

Compounds classify

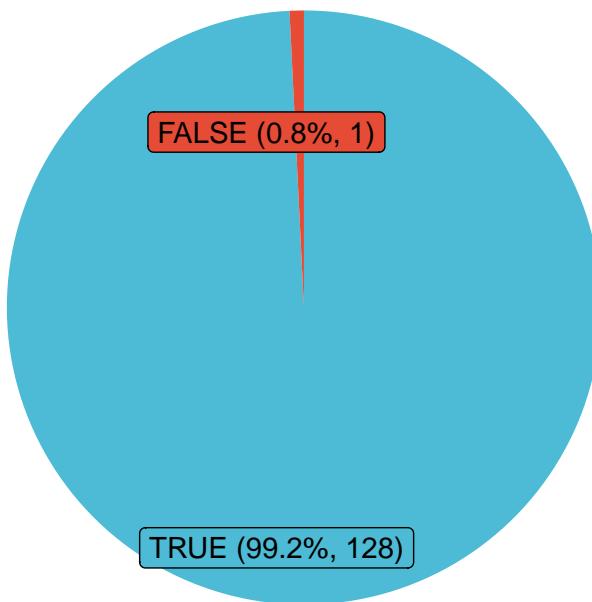


Figure 2: Compounds classify

6.2.2 化合物靶点

Table 2 (下方表格) 为表格 Targets predicted by Super Pred 概览。

(对应文件为 [Figure+Table/Targets-predicted-by-Super-Pred.xlsx](#))

注：表格共有 13670 行 9 列，以下预览的表格可能省略部分数据；表格含有 129 个唯一‘.id’。

Table 2: Targets predicted by Super Pred

.id	Target...	ChEMBL-ID	UniPro...	PDB Vi...	TTD ID	Probab...	Model ...	symbols
C1[C@H...]	Tyrosy...	CHEMBL...	Q9NUW8	6N0D	Not Av...	99.57%	71.22%	TDP1
C1[C@H...]	DNA-(a...	CHEMBL...	P27695	6BOW	T13348	97.3%	91.11%	APEX1
C1[C@H...]	Cathep...	CHEMBL...	P07339	4OD9	T67102	92.9%	98.95%	CTSD
C1[C@H...]	LSD1/C...	CHEMBL...	O60341	5L3D	Not Av...	91.71%	97.09%	KDM1A
C1[C@H...]	Endopl...	CHEMBL...	Q99714	2O23	Not Av...	89.3%	70.16%	HSD17B10
C1[C@H...]	Transc...	CHEMBL...	O15164	4YBM	Not Av...	89.13%	95.56%	TRIM24
C1[C@H...]	DNA to...	CHEMBL...	P11388	6ZY5	T17048	88.76%	89%	TOP2A
C1[C@H...]	Glycin...	CHEMBL...	P23415	4X5T	T50269	87.03%	90.71%	GLRA1
C1[C@H...]	Estrog...	CHEMBL242	Q92731	1QKM	T80896	86.62%	98.35%	ESR2
C1[C@H...]	Nuclea...	CHEMBL...	P19838	1SVC	Not Av...	85.02%	96.09%	NFKB1

.id	Target...	ChEMBL-ID	UniPro...	PDB Vi...	TTD ID	Probab...	Model ...	symbols
C1[C@H...]	Riboso...	CHEMBL...	Q15418	2Z7Q	Not Av...	84.42%	85.11%	RPS6KA1
C1[C@H...]	IgG re...	CHEMBL...	P55899	6FGB	Not Av...	84.41%	90.93%	FCGRT
C1[C@H...]	Egl ni...	CHEMBL...	Q9GZT9	4BQY	Not Av...	83.24%	93.4%	EGLN1
C1[C@H...]	Arachi...	CHEMBL...	P18054	3D3L	Not Av...	81.74%	75.57%	ALOX12
C1[C@H...]	Casein...	CHEMBL...	P67870	6TLS	T51565	81.25%	99.23%	CSNK2B
...

6.3 网络药理学

6.3.1 靶点网络 (+ 化合物含量)

Figure 3 (下方图) 为图 Network pharmacology 概览。

(对应文件为 Figure+Table/Network-pharmacology.pdf)

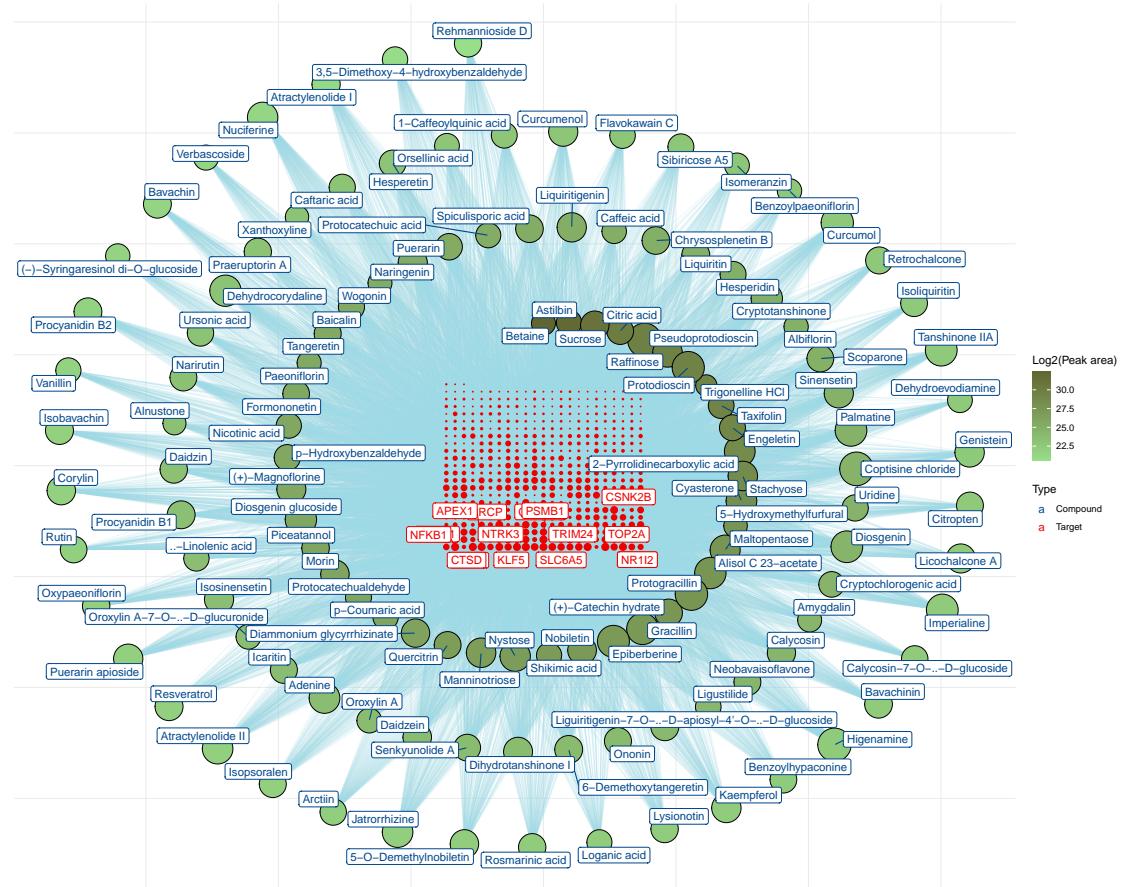


Figure 3: Network pharmacology

6.3.2 炎症

Figure 4 (下方图) 为图 Overall targets number of datasets 概览。

(对应文件为 Figure+Table/Overall-targets-number-of-datasets.pdf)

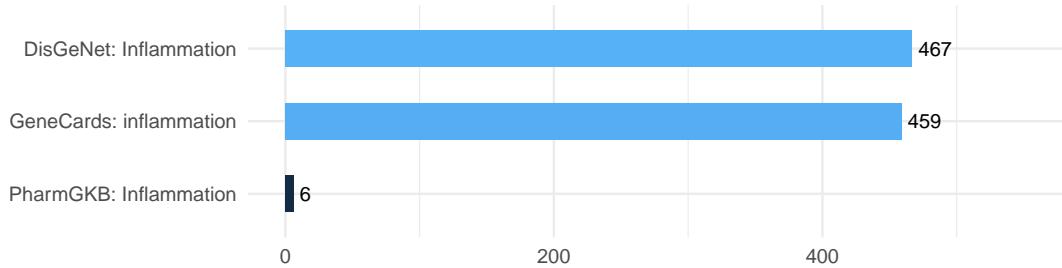


Figure 4: Overall targets number of datasets

6.3.3 靶点网络 (+ 化合物含量) + 炎症

Figure 5 (下方图) 为图 Network pharmacology with disease 概览。

(对应文件为 Figure+Table/Network-pharmacology-with-disease.pdf)

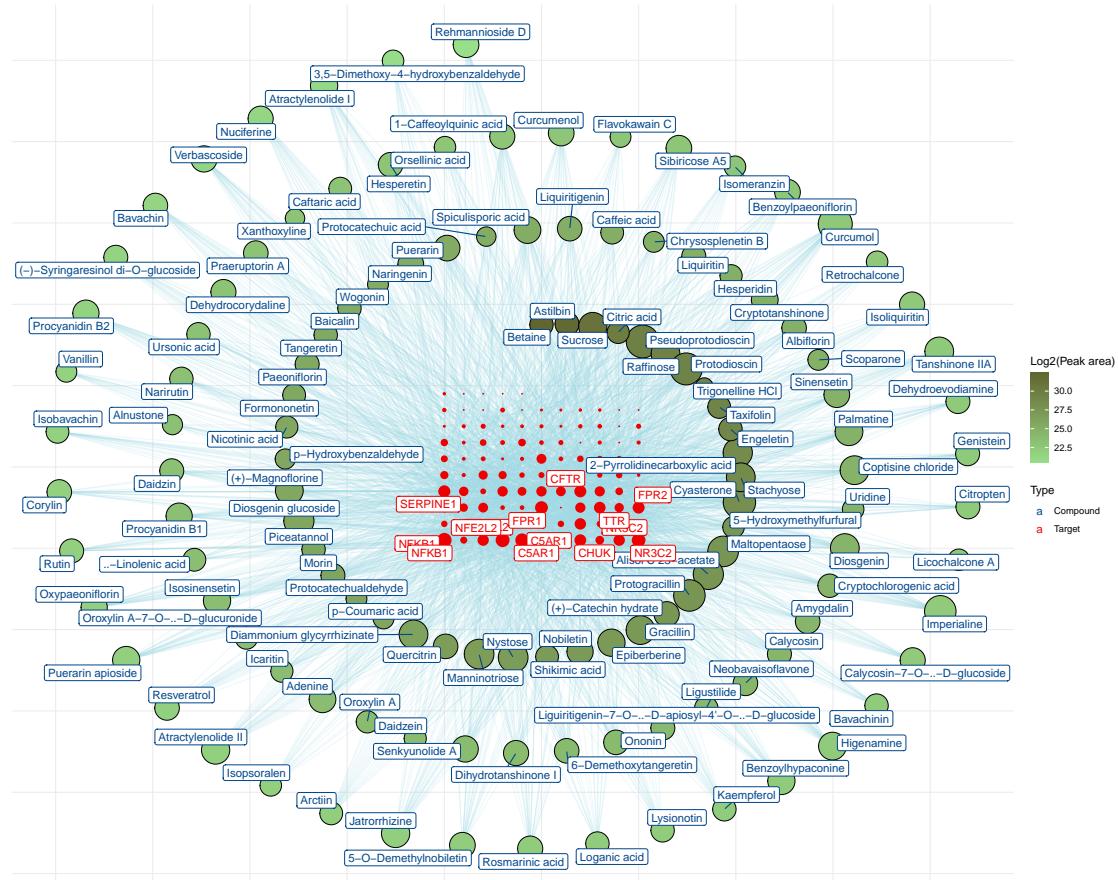


Figure 5: Network pharmacology with disease

Figure 6 (下方图) 为图 Targets intersect with targets of diseases 概览。

(对应文件为 Figure+Table/Targets-intersect-with-targets-of-diseases.pdf)

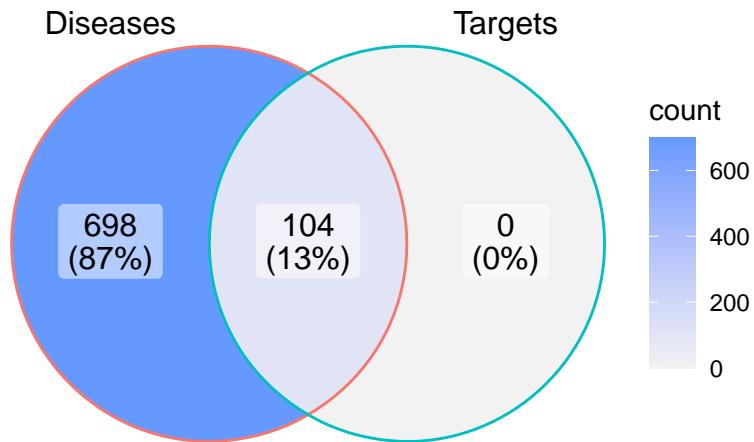


Figure 6: Targets intersect with targets of diseases

Intersection :

ACE, PTGS2, MIF, PPARG, TLR4, MMP9, PARP1, NOS2, MMP2, PPARA, F2R, IDO1, STAT3, RORC, CHRNB2, TRPA1, NR1H4, PTPN1, ABCB1, KLK1, CDK5, CCR5, TACR1, ALOX5, PDE5A, ADORA1, PIK3CG, MTOR, OPRM1, PROC, NFKB1, PLAU, NOS3, SERPINE1, CNR2, NFE2L2, MAPK14, APP, CFTR, NR3C2, CD38, C5AR1, TGM2, HIF1A, FPR2, ...

(上述信息框内容已保存至 [Figure+Table/Targets-intersect-with-targets-of-diseases-content](#))

6.3.4 相关靶点与转录因子 FOXO 相关

数据库 Transcription Factor Target Gene Database 检索：

Table 3 (下方表格) 为表格 Transcription Factor binding sites 概览。

(对应文件为 [Figure+Table/Transcription-Factor-binding-sites.csv](#))

注：表格共有 30090 行 10 列，以下预览的表格可能省略部分数据；表格含有 104 个唯一‘target’。

1. Start: 起始点

Table 3: Transcription Factor binding sites

target	TF_symbol	Motif	Source	Strand	Start	Stop	PValue	MatchS...	Overla...
ACE	CTCF	CTCF_M...	JASPAR	-	61556433	61556451	0.0E+00	GGACCA...	19
ACE	FOXJ2	FOXJ2_...	SELEX	-	61551959	61551971	4.0E-06	GTAAAA...	13
ACE	MESP1	MESP1_...	SELEX	-	61554628	61554637	9.0E-06	AGCACCC...	10

target	TF_symbol	Motif	Source	Strand	Start	Stop	PValue	MatchS...	Overla...
ACE	FOXJ2	FOXJ2_...	SELEX	-	61551854	61551861	5.0E-06	ATAAACAA	8
ACE	TBX1	TBX1_T...	SELEX	-	61552066	61552085	1.0E-06	TTCACA...	20
ACE	HOXD9	Hoxd9_...	SELEX	-	61551499	61551508	4.0E-06	ACAATT...	10
ACE	BARX1	BARX1_...	SELEX	-	61552008	61552024	4.0E-06	TCATTA...	17
ACE	target	SRY_HM...	SELEX	+	61551570	61551584	5.0E-06	TTCTAT...	15
ACE	KLF16	KLF16_...	SELEX	-	61554376	61554386	7.0E-06	GACACA...	11
ACE	KLF16	KLF16_...	SELEX	-	61554725	61554735	1.0E-05	GCCCCG...	11
ACE	FOGX1	FOGX1_...	SELEX	+	61551524	61551540	6.0E-06	AAAAAC...	17
ACE	FOGX1	FOGX1_...	SELEX	-	61551902	61551918	2.0E-06	ATAAAT...	17
ACE	MEF2D	MEF2D_...	SELEX	+	61551941	61551952	1.0E-06	CCTAAA...	12
ACE	FOXC2	FOXC2_...	SELEX	-	61551853	61551864	3.0E-06	TGTATA...	12
ACE	PDX1	PDX1_h...	SELEX	-	61551528	61551545	7.0E-06	ACAATT...	18
...

6.3.5 富集分析

Figure 7 (下方图) 为图 Ids KEGG enrichment 概览。

(对应文件为 Figure+Table/Ids-KEGG-enrichment.pdf)

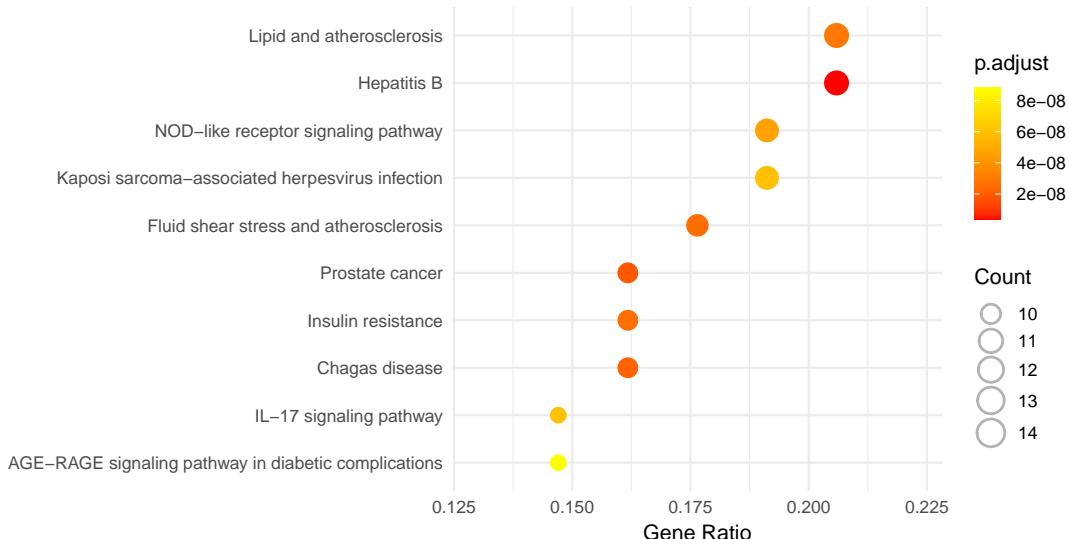


Figure 7: Ids KEGG enrichment

Figure 8 (下方图) 为图 Ids GO enrichment 概览。

(对应文件为 Figure+Table/Ids-GO-enrichment.pdf)

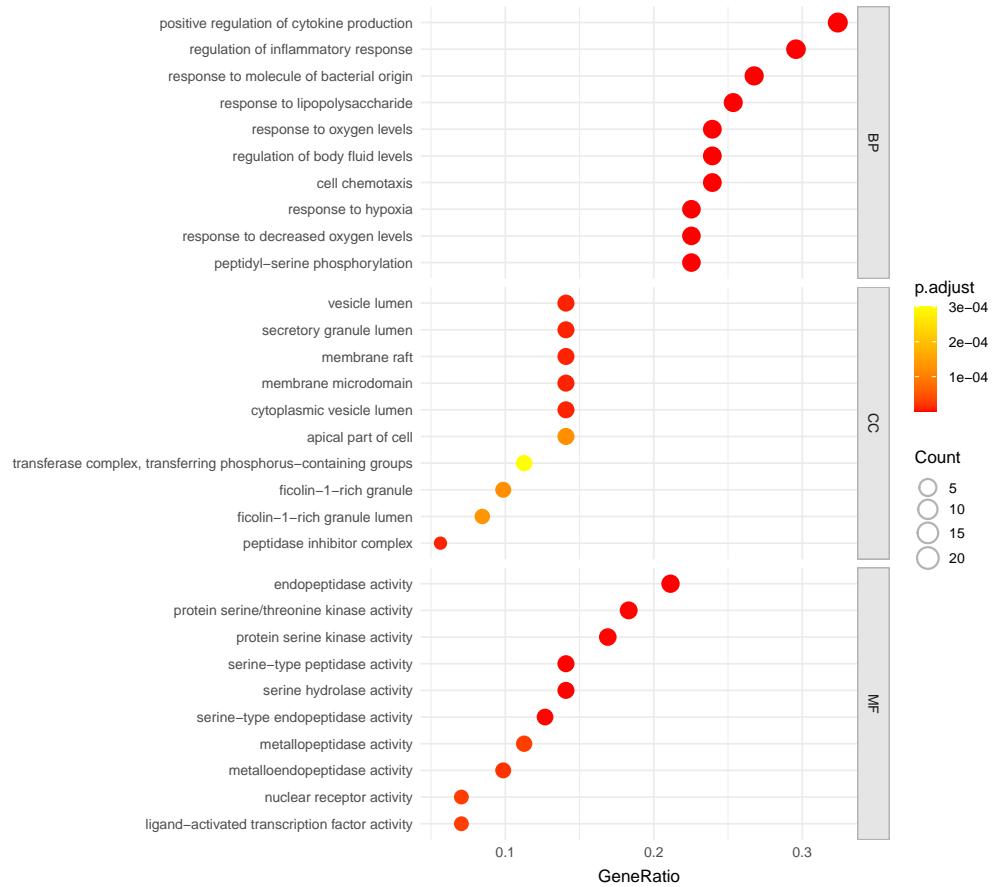


Figure 8: Ids GO enrichment

GO 中, regulation of inflammatory response 通路的基因:

GO :

ACE, ADORA1, ALOX5, APP, CCR2, CNR1, CNR2, CYP19A1, IDO1, MAPK14, MMP8, MMP9, NFKB1, NR1H3, PIK3CG, PLA2G2A, PRKCD, PROC, PTGS2, SERPINE1, TLR4

6.3.6 口服利用度筛选

Figure 9 (下方图) 为图 HOB 20 prediction 概览。

(对应文件为 Figure+Table/HOB-20-prediction.pdf)

HOB (20%) Prediction

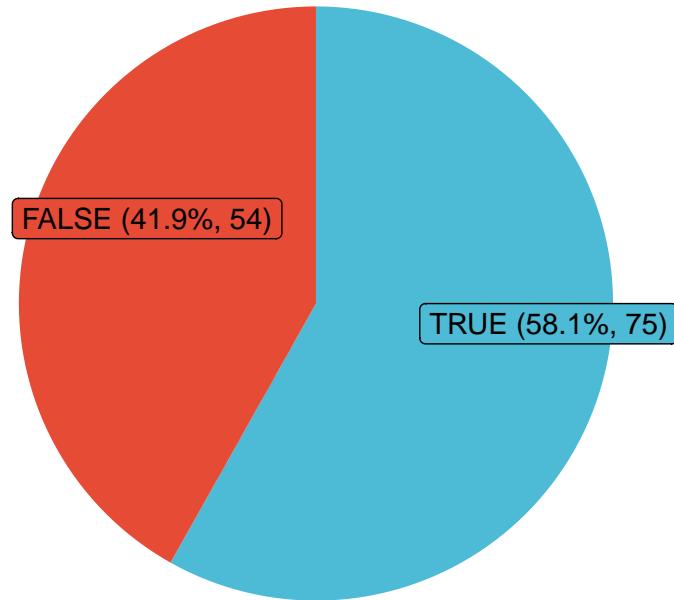


Figure 9: HOB 20 prediction

6.3.7 最终筛选过的网络药理图

Figure 10 (下方图) 为图 FINAL Network pharmacology 概览。

(对应文件为 Figure+Table/FINAL-Network-pharmacology.pdf)

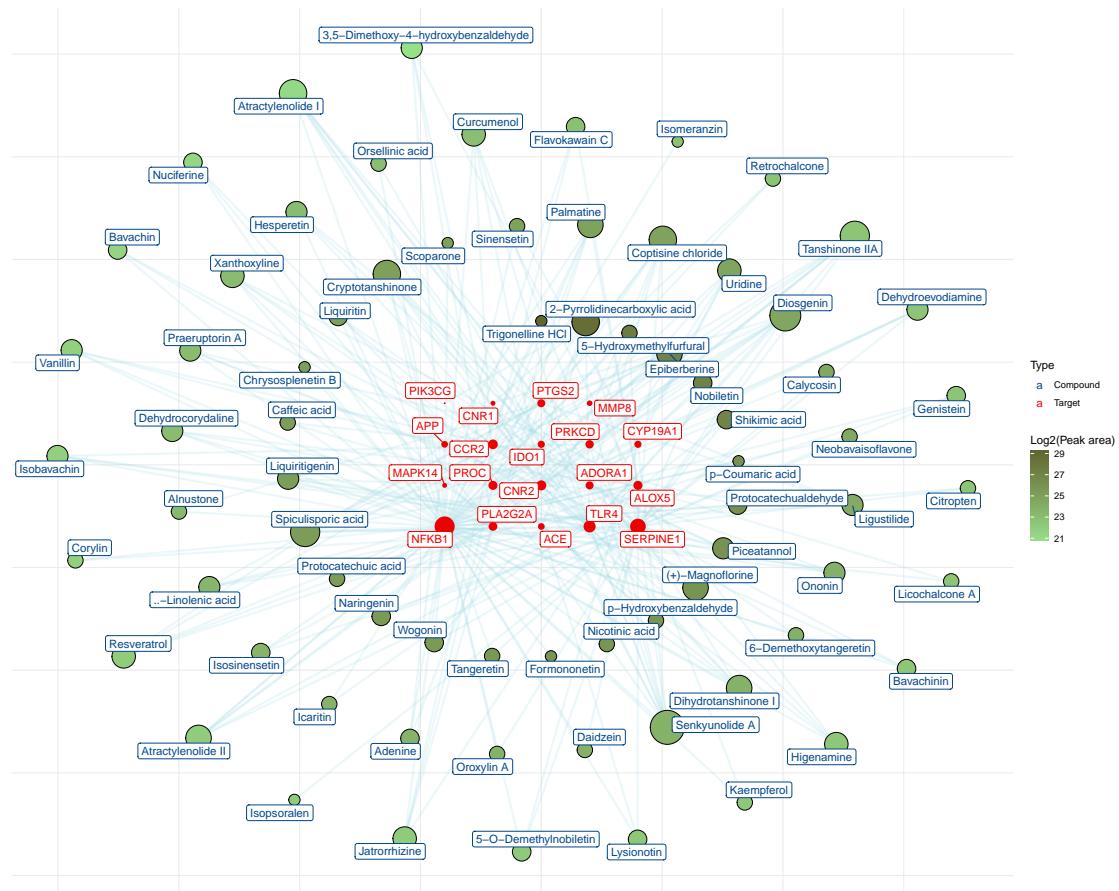


Figure 10: FINAL Network pharmacology

6.4 分子对接

6.4.1 对接的成分

选择了含量排名前 5 的成分对接 (从 Fig. 5)。

Compounds :

Trigonelline HCl, 2-Pyrrolidinecarboxylic acid, 5-Hydroxymethylfurfural, Epiberberine, Nobletin

6.4.2 对接结果

Figure 11 (下方图) 为图 Overall combining Affinity 概览。

(对应文件为 Figure+Table/Overall-combining-Affinity.pdf)

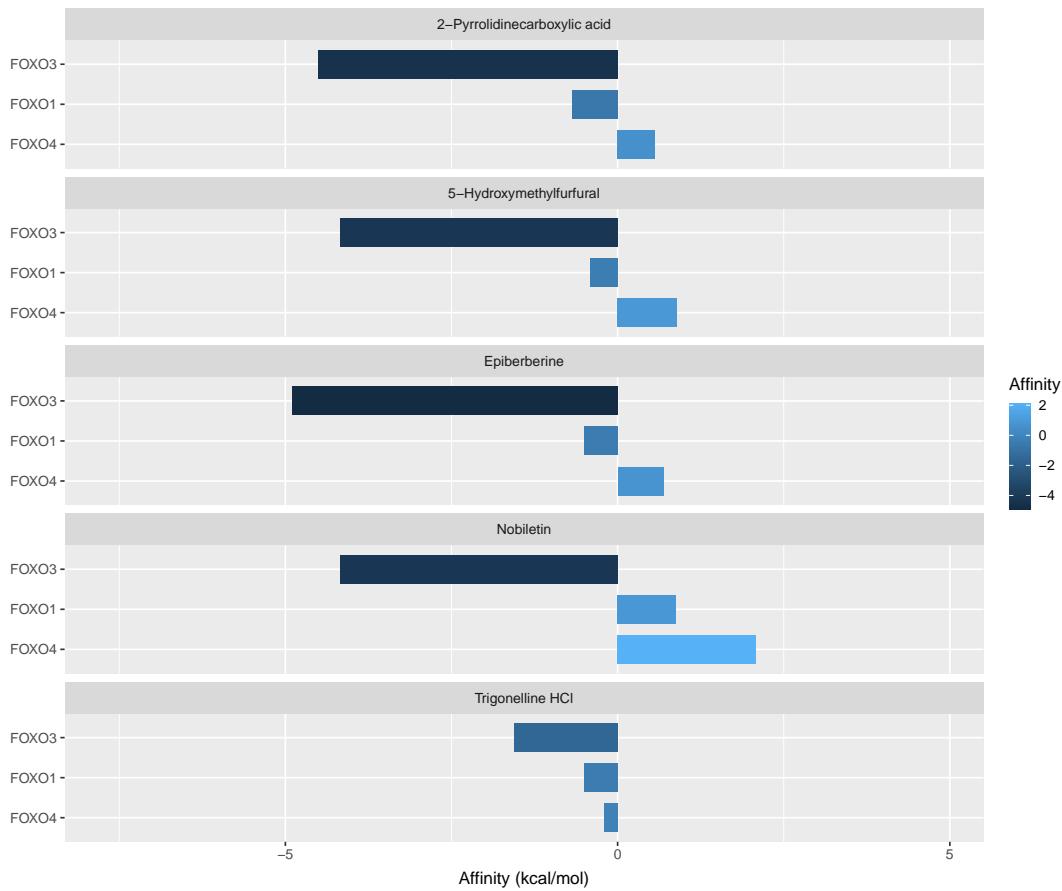


Figure 11: Overall combining Affinity

Table 4 (下方表格) 为表格 Overall combining Affinity rawData 概览。

(对应文件为 [Figure+Table/Overall-combining-Affinity rawData.csv](#))

注：表格共有 15 行 8 列，以下预览的表格可能省略部分数据；表格含有 5 个唯一‘PubChem_id’。

1. hgnc_symbol: 基因名 (Human)

Table 4: Overall combining Affinity rawData

PubChe...	PDB_ID	Affinity	dir	file	Combn	hgnc_s...	Ingred...
160876	2uzk	-4.9	vina_s...	vina_s...	160876...	FOXO3	Epiber...
145742	2uzk	-4.502	vina_s...	vina_s...	145742...	FOXO3	2-Pyrr...
237332	2uzk	-4.177	vina_s...	vina_s...	237332...	FOXO3	5-Hydr...
72344	2uzk	-4.173	vina_s...	vina_s...	72344_...	FOXO3	Nobiletin
134606	2uzk	-1.563	vina_s...	vina_s...	134606...	FOXO3	Trigon...
145742	6qvw	-0.687	vina_s...	vina_s...	145742...	FOXO1	2-Pyrr...
160876	6qvw	-0.508	vina_s...	vina_s...	160876...	FOXO1	Epiber...

PubChe...	PDB_ID	Affinity	dir	file	Combn	hgnc_s...	Ingred...
134606	6qvw	-0.506	vina_s...	vina_s...	134606...	FOXO1	Trigon...
237332	6qvw	-0.413	vina_s...	vina_s...	237332...	FOXO1	5-Hydr...
134606	3l2c	-0.203	vina_s...	vina_s...	134606...	FOXO4	Trigon...
145742	3l2c	0.559	vina_s...	vina_s...	145742...	FOXO4	2-Pyrr...
160876	3l2c	0.691	vina_s...	vina_s...	160876...	FOXO4	Epiber...
72344	6qvw	0.874	vina_s...	vina_s...	72344_...	FOXO1	Nobiletin
237332	3l2c	0.889	vina_s...	vina_s...	237332...	FOXO4	5-Hydr...
72344	3l2c	2.084	vina_s...	vina_s...	72344_...	FOXO4	Nobiletin

Reference

1. Kim, S. *et al.* PubChem substance and compound databases. *Nucleic Acids Research* (2015).
2. Djoumbou Feunang, Y. *et al.* ClassyFire: Automated chemical classification with a comprehensive, computable taxonomy. *Journal of Cheminformatics* **8**, 61 (2016).
3. Wu, T. *et al.* ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, (2021).
4. Piñero, J. *et al.* The disgenet knowledge platform for disease genomics: 2019 update. *Nucleic Acids Research* (2019) doi:10.1093/nar/gkz1021.
5. Stelzer, G. *et al.* The genecards suite: From gene data mining to disease genome sequence analyses. *Current protocols in bioinformatics* **54**, 1.30.1–1.30.33 (2016).
6. Barbarino, J. M., Whirl-Carrillo, M., Altman, R. B. & Klein, T. E. PharmGKB: A worldwide resource for pharmacogenomic information. *Wiley interdisciplinary reviews. Systems biology and medicine* **10**, (2018).
7. Wei, M. *et al.* HobPre: Accurate prediction of human oral bioavailability for small molecules. *Journal of Cheminformatics* **14**, (2022).
8. Nickel, J. *et al.* SuperPred: Update on drug classification and target prediction. *Nucleic acids research* **42**, W26–W31 (2014).
9. Plaisier, C. L. *et al.* Causal mechanistic regulatory network for glioblastoma deciphered using systems genetics network analysis. *Cell systems* **3**, 172–186 (2016).
10. Eberhardt, J., Santos-Martins, D., Tillack, A. F. & Forli, S. AutoDock vina 1.2.0: New docking methods, expanded force field, and python bindings. *Journal of Chemical Information and Modeling* **61**, 3891–3898 (2021).
11. Zhang, Y., Forli, S., Omelchenko, A. & Sanner, M. F. AutoGridFR: Improvements on autodock affinity maps and associated software tools. *Journal of computational chemistry* **40**, 2882–2886 (2019).
12. Zhang, Y. & Sanner, M. F. AutoDock crankpep: Combining folding and docking to predict protein-peptide complexes. *Bioinformatics (Oxford, England)* **35**, 5121–5127 (2019).

13. Ravindranath, P. A. & Sanner, M. F. AutoSite: An automated approach for pseudo-ligands prediction-from ligand-binding sites identification to predicting key ligand atoms. *Bioinformatics (Oxford, England)* **32**, 3142–3149 (2016).
14. Ravindranath, P. A., Forli, S., Goodsell, D. S., Olson, A. J. & Sanner, M. F. AutoDockFR: Advances in protein-ligand docking with explicitly specified binding site flexibility. *PLoS computational biology* **11**, (2015).