

Analysis

Huang LiChuang of Wie-Biotech

Contents

1 摘要	2
2 研究设计流程图	2
3 材料和方法	2
4 分析结果	2
4.1 网络药理学分析主要活性成分	2
4.1.1 从 HERB 网站获取中药和成分以及靶点数据	2
4.1.2 从 Genecards 获取疾病的靶点数据	5
4.1.3 数据透视	6
4.1.4 以 STRINGdb 构建 PPI 网络	9
4.1.5 筛选 HubGenes	10
4.1.6 通路富集分析	12
4.2 分析糖尿病肾病肠道差异菌群	13
4.2.1 选择公共数据库的 16s rRNA 数据	13
4.2.2 下载和预处理 SRA 数据	14
4.2.3 Qiime2	14
4.3 分析糖尿病肾病的肠道代谢组学差异	15
4.4 分析差异菌群与代谢物的相关性	15
5 结论	15
Reference	15

List of Figures

1	Intersect of target genes of TCMs	7
2	Intersect of compounds of TCMs	8
3	Intersect of targets of compounds and disease targets	9
4	MCC Top 30	11
5	GO enrichment of MCC top 30	12
6	KEGG enrichment of MCC top 30	13

List of Tables

1	TCM information	2
2	TCM compounds	3
3	Compounds targets	4
4	Compounds targets with annotation of biomaRt of ensembl dataset	4
5	Desease targets of diabetic nephropathy	5
6	Desease targets with annotation of biomaRt of ensembl dataset	6
7	ID mapped by STRINGdb	10
8	All MCC scores	11
9	Metadata of the sra data of PRJNA824185	13

1 摘要

2 研究设计流程图

3 材料和方法

4 分析结果

4.1 网络药理学分析主要活性成分

4.1.1 从 HERB 网站获取中药和成分以及靶点数据

HERB <http://herb.ac.cn/>

Table 1为表格 TCM information 概览。

(对应文件为 **Figure+Table/TCM-information.xlsx**)

注：表格共有 10 行 18 列，以下预览的表格可能省略部分数据；表格含有 10 个唯一 ‘Herb_’。

Table 1: TCM information

Herb_	Herb_...	Herb_...	Herb_...	Herb_...	Prope...	Merid...	UsePart	Function	Indic...	Toxicity
HERB0...	CHEN PI	陈皮	Dried...	Peric...	Warm;...	Lung;...	NA	To re...	Treat...	NA
HERB0...	DA HUANG	大黄	root ...	Radix...	Cold;...	Splee...	root ...	To ca...	Reple...	NA
HERB0...	DAN SHEN	丹参	root ...	Radix...	Minor...	Liver...	root	To re...	Angin...	NA
HERB0...	FU LING	茯苓	India...	Poria	Mild;...	Splee...	scler...	To ca...	Neura...	NA
HERB0...	GAN CAO	甘草	Root ...	Radix...	Mild;...	Lung;...	root ...	To re...	1. It...	NA
HERB0...	HONG HUA	红花	Saffl...	Flos ...	Warm;...	Liver...	flower	To ac...	Ameno...	NA
HERB0...	HUANG...	黄连	rhizo...	Rhizo...	Cold;...	Large...	rhizome	To re...	Febri...	NA
HERB0...	NIU XI	牛膝	root ...	Radix...	Mild;...	Liver...	root	To re...	Achin...	NA

Herb__	Herb__...	Herb__...	Herb__...	Herb__...	Prope...	Merid...	UsePart	Function	Indic...	Toxicity
HERB0...	TAI Z...	太子参	Heter...	Radix...	Mild;...	Lung;...	NA	To re...	Treat...	NA
HERB0...	ZHI B...	制半夏	Prepa...	Rhizo...	Warm;...	Lung;...	Pinel...	Treat...	Treat...	Extre...

Table 2为表格 TCM compounds 概览。

(对应文件为 **Figure+Table/TCM-compounds.xlsx**)

注：表格共有 1761 行 4 列，以下预览的表格可能省略部分数据；表格含有 10 个唯一 ‘herb_id’。

Table 2: TCM compounds

herb_id	Ingre.....2	Ingre.....3	Ingre.....4
HERB0...	HBIN0...	1-A-t...	NA
HERB0...	HBIN0...	(1S,5...	(1S,5...
HERB0...	HBIN0...	1-Und...	10268...
HERB0...	HBIN0...	22410...	2,6-O...
HERB0...	HBIN0...	2-(2-...	NA
HERB0...	HBIN0...	2,5,5...	2,5,5...
HERB0...	HBIN0...	2,6,1...	2,6,1...
HERB0...	HBIN0...	()-2-...	Bicyc...
HERB0...	HBIN0...	(2S)-...	NA
HERB0...	HBIN0...	(2S)-...	(2S)-...
HERB0...	HBIN0...	3-carene	(1S)-...
HERB0...	HBIN0...	3-dec...	3-dec...
HERB0...	HBIN0...	[(3R)...	butan...
HERB0...	HBIN0...	4-ACE...	17745...
HERB0...	HBIN0...	5,7-d...	(2R)-...
...

Table 3为表格 compounds targets 概览。

(对应文件为 **Figure+Table/compounds-targets.csv**)

注：表格共有 14888 行 6 列，以下预览的表格可能省略部分数据；表格含有 707 个唯一 ‘Ingredient_id’。

Table 3: Compounds targets

Ingre...	Targe.....2	Targe.....3	Datab...	Paper.id	PubMe...
HBIN0...	HBTAR...	PGR	NA	NA	NA
HBIN0...	HBTAR...	PTGS2	NA	NA	NA
HBIN0...	HBTAR...	RXRA	NA	NA	NA
HBIN0...	HBTAR...	DPP4	NA	NA	NA
HBIN0...	HBTAR...	AR	NA	NA	NA
HBIN0...	HBTAR...	NR3C1	NA	NA	NA
HBIN0...	HBTAR...	PRSS1	NA	NA	NA
HBIN0...	HBTAR...	CA2	NA	NA	NA
HBIN0...	HBTAR...	ESR1	NA	NA	NA
HBIN0...	HBTAR...	GABRA1	NA	NA	NA
HBIN0...	HBTAR...	NCOA2	NA	NA	NA
HBIN0...	HBTAR...	ACHE	NA	NA	NA
HBIN0...	HBTAR...	PRSS1	NA	NA	NA
HBIN0...	HBTAR...	PRSS1	NA	NA	NA
HBIN0...	HBTAR...	SLC6A2	NA	NA	NA
...

将 Tab. 3 的基因信息注释:

Table 4为表格 compounds targets with annotation of biomaRt of ensembl dataset 概览。

(对应文件为 **Figure+Table/compounds-targets-with-annotation-of-biomaRt-of-ensembl-dataset.xlsx**)

注: 表格共有 14888 行 13 列, 以下预览的表格可能省略部分数据; 表格含有 4230 个唯一 'Target.name'。

Table 4: Compounds targets with annotation of biomaRt of ensembl dataset

Targe.....1	Ingre...	Targe.....3	Datab...	Paper.id	PubMe...	ensem.....7	ensem.....8	entre...	...
6PGD	HBIN0...	HBTAR...	NA	HBREF...	27270429	NA	NA	NA	...
AAMP	HBIN0...	HBTAR...	NA	NA	NA	ENSG0...	ENST0...	14	...
AANAT	HBIN0...	HBTAR...	NA	NA	NA	ENSG0...	ENST0...	15	...
AANAT	HBIN0...	HBTAR...	NA	NA	NA	ENSG0...	ENST0...	15	...
AARS1	HBIN0...	HBTAR...	NA	NA	NA	ENSG0...	ENST0...	16	...
AARS1	HBIN0...	HBTAR...	NA	NA	NA	ENSG0...	ENST0...	16	...
AARS1	HBIN0...	HBTAR...	NA	NA	NA	ENSG0...	ENST0...	16	...
AARS1	HBIN0...	HBTAR...	NA	NA	NA	ENSG0...	ENST0...	16	...
AARS1	HBIN0...	HBTAR...	NA	NA	NA	ENSG0...	ENST0...	16	...
AARS2	HBIN0...	HBTAR...	NA	NA	NA	ENSG0...	ENST0...	57505	...

Targe.....1	Ingre...	Targe.....3	Datab...	Paper.id	PubMe...	ensem.....7	ensem.....8	entre...	...
AASS	HBIN0...	HBTAR...	NA	NA	NA	ENSG0...	ENST0...	10157	...
AASS	HBIN0...	HBTAR...	NA	NA	NA	ENSG0...	ENST0...	10157	...
ABAT	HBIN0...	HBTAR...	NA	NA	NA	ENSG0...	ENST0...	18	...
ABAT	HBIN0...	HBTAR...	NA	NA	NA	ENSG0...	ENST0...	18	...
ABAT	HBIN0...	HBTAR...	NA	NA	NA	ENSG0...	ENST0...	18	...
...

4.1.2 从 Genecards 获取疾病的靶点数据

Genecards <https://www.genecards.org/>

Table 5为表格 disease targets of diabetic nephropathy 概览。

(对应文件为 Figure+Table/disease-targets-of-diabetic-nephropathy.xlsx)

注：表格共有 1309 行 7 列，以下预览的表格可能省略部分数据；表格含有 1309 个唯一‘Symbol’。

Table 5: Disease targets of diabetic nephropathy

Symbol	Descr...	Category	UniPr...	GIFtS	GC_id	Score
ACE	Angio...	Prote...	P12821	55	GC17P...	84.68
HNF1B	HNF1 ...	Prote...	P35680	47	GC17M...	83.38
GCK	Gluko...	Prote...	P35557	53	GC07M...	83.04
KCNJ11	Potas...	Prote...	Q14654	50	GC11M...	80.35
HNF1A	HNF1 ...	Prote...	P20823	51	GC12P...	78.94
ABCC8	ATP B...	Prote...	Q09428	50	GC11M...	73.88
IL6	Inter...	Prote...	P05231	55	GC07P...	70.31
HNF4A	Hepat...	Prote...	P41235	53	GC20P...	67.73
PPARG	Perox...	Prote...	P37231	57	GC03P...	64.11
PDX1	Pancr...	Prote...	P52945	50	GC13P...	61.95
COL4A5	Colla...	Prote...	P29400	48	GC0XP...	58.60
WFS1	Wolfr...	Prote...	O76024	50	GC04P...	57.65
TCF7L2	Trans...	Prote...	Q9NQB0	51	GC10P...	56.48
VEGFA	Vascu...	Prote...	P15692	53	GC06P...	56.25
COL4A3	Colla...	Prote...	Q01955	50	GC02P...	55.86
...

Table 6为表格 disease targets with annotation of biomaRt of ensembl dataset 概览。

(对应文件为 Figure+Table/disease-targets-with-annotation-of-biomaRt-of-ensembl-dataset.xlsx)

注：表格共有 674 行 8 列，以下预览的表格可能省略部分数据；表格含有 674 个唯一 ‘hgnc_symbol’。

Table 6: Disease targets with annotation of biomaRt of ensembl dataset

hgnc_...	ensem.....2	ensem.....3	entre...	chrom...	start...	end_p...	descr...
ABCB1	ENSG0...	ENST0...	5243	7	87503017	87713323	ATP b...
ABCC8	ENSG0...	ENST0...	6833	11	17392498	17476894	ATP b...
ABCG1	ENSG0...	ENST0...	9619	21	42199689	42297244	ATP b...
ACE	ENSG0...	ENST0...	1636	17	63477061	63498380	angio...
ACHE	ENSG0...	ENST0...	43	7	10088...	10089...	acety...
ACSL1	ENSG0...	ENST0...	2180	4	18475...	18482...	acyl-...
ACTB	ENSG0...	ENST0...	60	7	5526409	5563902	actin...
ADA	ENSG0...	ENST0...	100	20	44584896	44652252	adeno...
ADA2	ENSG0...	ENST0...	51816	22	17178790	17258235	adeno...
ADAMTS13	ENSG0...	ENST0...	11093	HG203...	240452	285496	ADAM ...
ADAR	ENSG0...	ENST0...	103	1	15458...	15462...	adeno...
ADCY3	ENSG0...	ENST0...	109	2	24819169	24920237	adeny...
ADCY5	ENSG0...	ENST0...	111	3	12328...	12344...	adeny...
ADD1	ENSG0...	ENST0...	118	4	2843844	2930076	adduc...
ADORA1	ENSG0...	ENST0...	134	1	20309...	20316...	adeno...
...

4.1.3 数据透视

以下，以 UpSet 图展示各个数据集之间的交集。

Figure 1为图 intersect of target genes of TCMs 概览。

(对应文件为 Figure+Table/intersect-of-target-genes-of-TCMs.pdf)

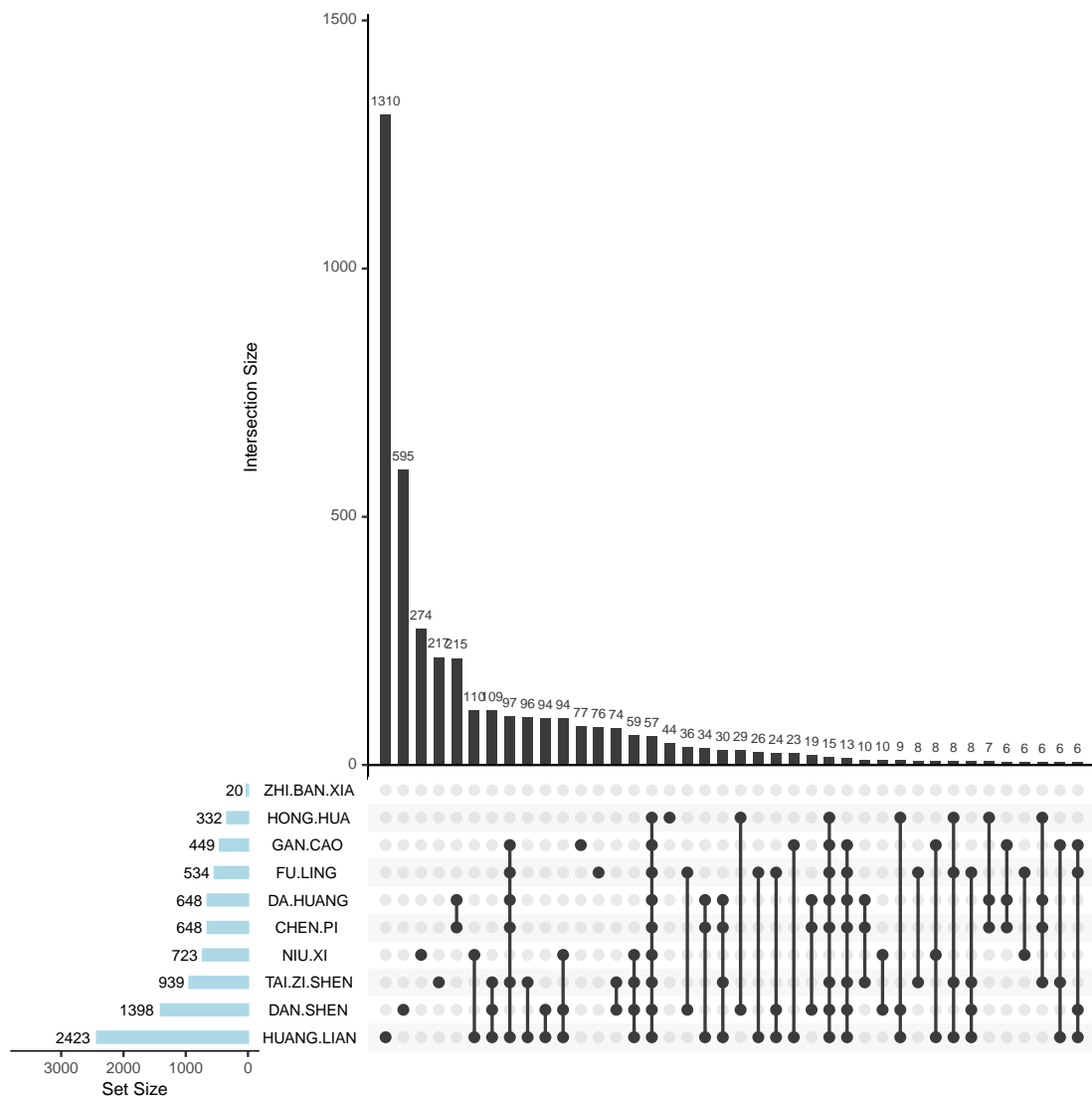


Figure 1: Intersect of target genes of TCMs

Figure 2为图 intersect of compounds of TCMs 概览。

(对应文件为 Figure+Table/intersect-of-compounds-of-TCMs.pdf)

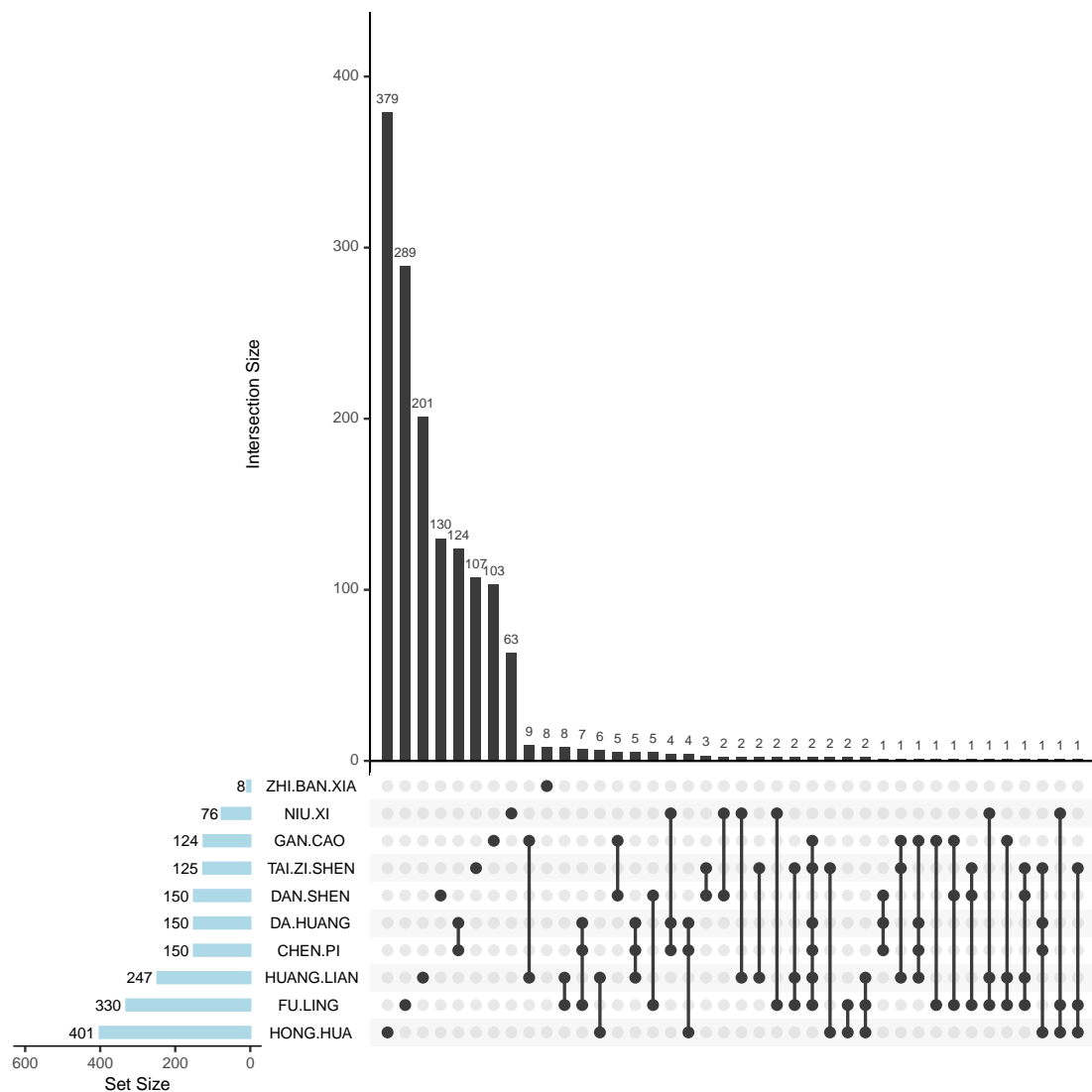


Figure 2: Intersect of compounds of TCMs

Figure 3为图 intersect of targets of compounds and disease targets 概览。

(对应文件为 Figure+Table/intersect-of-targets-of-compounds-and-disease-targets.pdf)

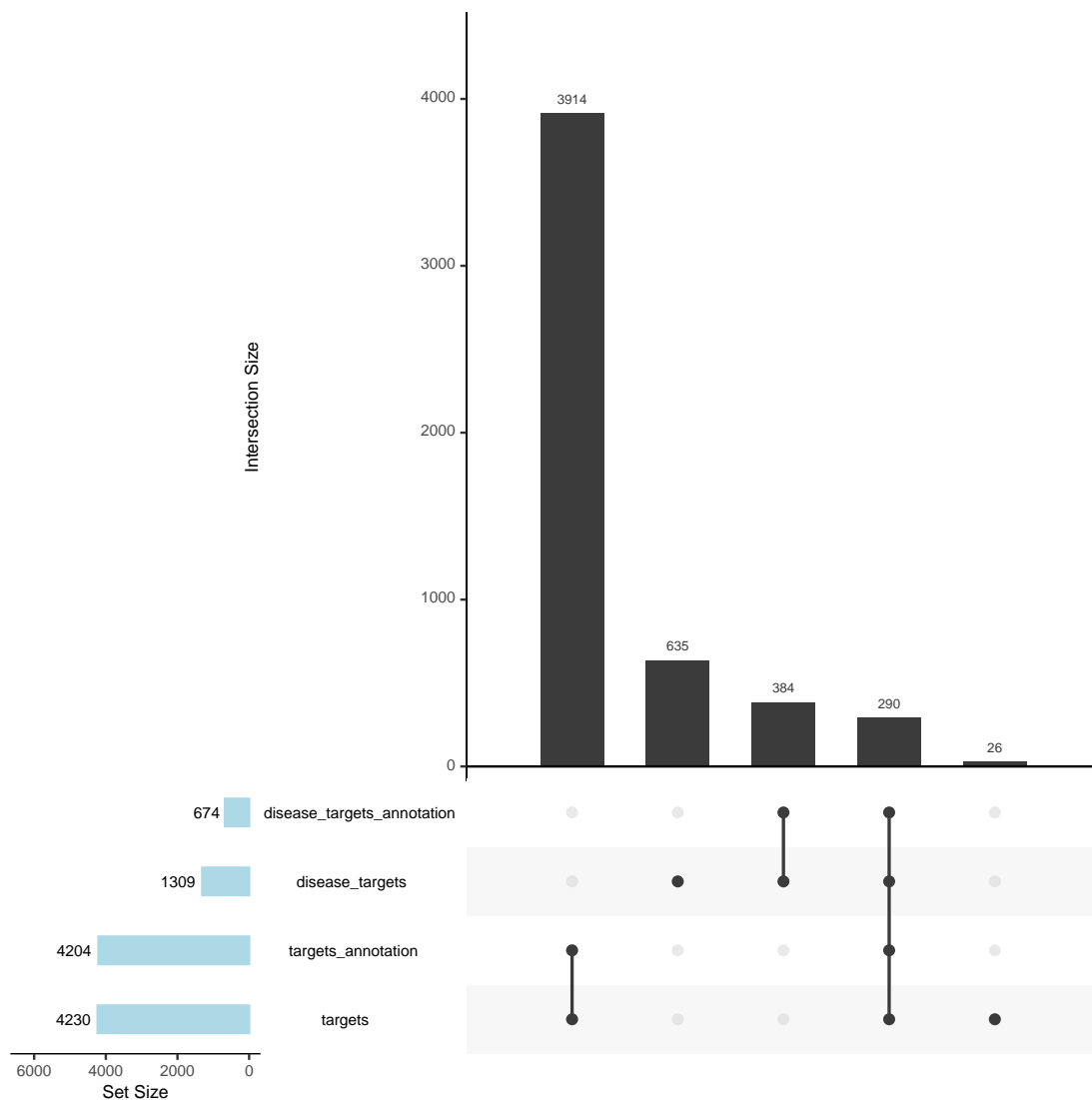


Figure 3: Intersect of targets of compounds and disease targets

4.1.4 以 STRINGdb 构建 PPI 网络

以 Fig. 3 中的四个数据集的交集，以 STRINGdb 创建 PPI 网络。

由于 PPI 网络包含过多节点，这里不展示 PPI 图（较为混乱）。

Table 7为表格 ID mapped by STRINGdb 概览。

(对应文件为 Figure+Table/ID-mapped-by-STRINGdb.xlsx)

注：表格共有 292 行 9 列，以下预览的表格可能省略部分数据；表格含有 289 个唯一 ‘hgnc_symbol’。

Table 7: ID mapped by STRINGdb

hgnc_...	STRIN...	ensem.....3	ensem.....4	entre...	chrom...	start...	end_p...	descr...
ABCB1	9606....	ENSG0...	ENST0...	5243	7	87503017	87713323	ATP b...
ACE	9606....	ENSG0...	ENST0...	1636	17	63477061	63498380	angio...
ACHE	9606....	ENSG0...	ENST0...	43	7	10088...	10089...	acety...
ACSL1	9606....	ENSG0...	ENST0...	2180	4	18475...	18482...	acyl-...
ACTB	9606....	ENSG0...	ENST0...	60	7	5526409	5563902	actin...
ADA	9606....	ENSG0...	ENST0...	100	20	44584896	44652252	adeno...
ADA2	9606....	ENSG0...	ENST0...	51816	22	17178790	17258235	adeno...
ADA2	9606....	ENSG0...	ENST0...	51816	22	17178790	17258235	adeno...
ADAR	9606....	ENSG0...	ENST0...	103	1	15458...	15462...	adeno...
ADD1	9606....	ENSG0...	ENST0...	118	4	2843844	2930076	adduc...
ADORA1	9606....	ENSG0...	ENST0...	134	1	20309...	20316...	adeno...
ADRB1	9606....	ENSG0...	ENST0...	153	10	11404...	11404...	adren...
AGXT	9606....	ENSG0...	ENST0...	189	2	24086...	24088...	alani...
AKR1B10	9606....	ENSG0...	ENST0...	57016	7	13452...	13454...	aldo-...
ALAD	9606....	ENSG0...	ENST0...	210	9	11338...	11340...	amino...
...

4.1.5 筛选 HubGenes

利用 Cytoscape 的插件 CytoHubba¹ 提供的 MCC 算法计算 Hub 基因得分（这里 MCC 算法被集成到 R 中，独立计算）。

Figure 4为图 MCC Top 30 概览。

(对应文件为 Figure+Table/MCC-Top-30.pdf)

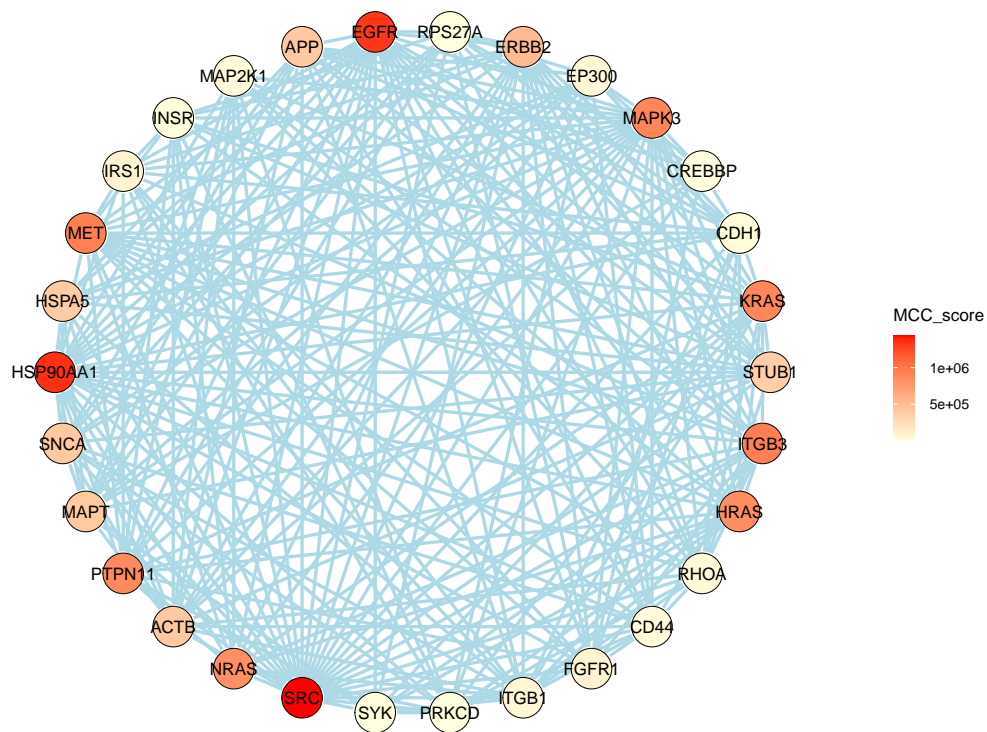


Figure 4: MCC Top 30

Table 8为表格 all MCC scores 概览。

(对应文件为 **Figure+Table/all-MCC-scores.xlsx**)

注：表格共有 292 行 10 列，以下预览的表格可能省略部分数据；表格含有 289 个唯一 ‘hgnc_symbol’。

Table 8: All MCC scores

hgnc_...	MCC_s...	STRIN...	ensem.....4	ensem.....5	entre...	chrom...	start...	end_p...	descr...	...
SRC	1441311	9606...	ENSG0...	ENST0...	6714	HG410...	254187	315552	SRC p...	...
HSP90AA1	1353642	9606...	ENSG0...	ENST0...	3320	14	10208...	10213...	heat
EGFR	1328596	9606...	ENSG0...	ENST0...	1956	7	55019017	55211628	epide...	...
ITGB3	941356	9606...	ENSG0...	ENST0...	3690	17	47253827	47313743	integ...	...
MET	937035	9606...	ENSG0...	ENST0...	4233	7	11667...	11679...	MET p...	...
MAPK3	894838	9606...	ENSG0...	ENST0...	5595	16	30114105	30123506	mitog...	...
KRAS	890536	9606...	ENSG0...	ENST0...	3845	12	25205246	25250936	KRAS
PTPN11	871709	9606...	ENSG0...	ENST0...	5781	12	11241...	11250...	prote...	...
HRAS	849299	9606...	ENSG0...	ENST0...	3265	HSCHR...	61883	66928	HRas

hgnc_...	MCC_s...	STRIN...	ensem.....4	ensem.....5	entre...	chrom...	start...	end_p...	descr...	...
NRAS	828404	9606...	ENSG0...	ENST0...	4893	1	11470...	11471...	NRAS
ERBB2	531274	9606...	ENSG0...	ENST0...	2064	17	39687914	39730426	erb-b...	...
ACTB	420124	9606...	ENSG0...	ENST0...	60	7	5526409	5563902	actin...	...
APP	417130	9606...	ENSG0...	ENST0...	351	21	25880535	26171128	amylo...	...
SNCA	416493	9606...	ENSG0...	ENST0...	6622	4	89700345	89838315	synuc...	...
MAPT	406245	9606...	ENSG0...	ENST0...	4137	HSCHR...	760287	893653	micro...	...
...

4.1.6 通路富集分析

使用 clusterProfiler 富集分析 top 30 的基因。

Figure 5为图 GO enrichment of MCC top 30 概览。

(对应文件为 Figure+Table/GO-enrichment-of-MCC-top-30.pdf)

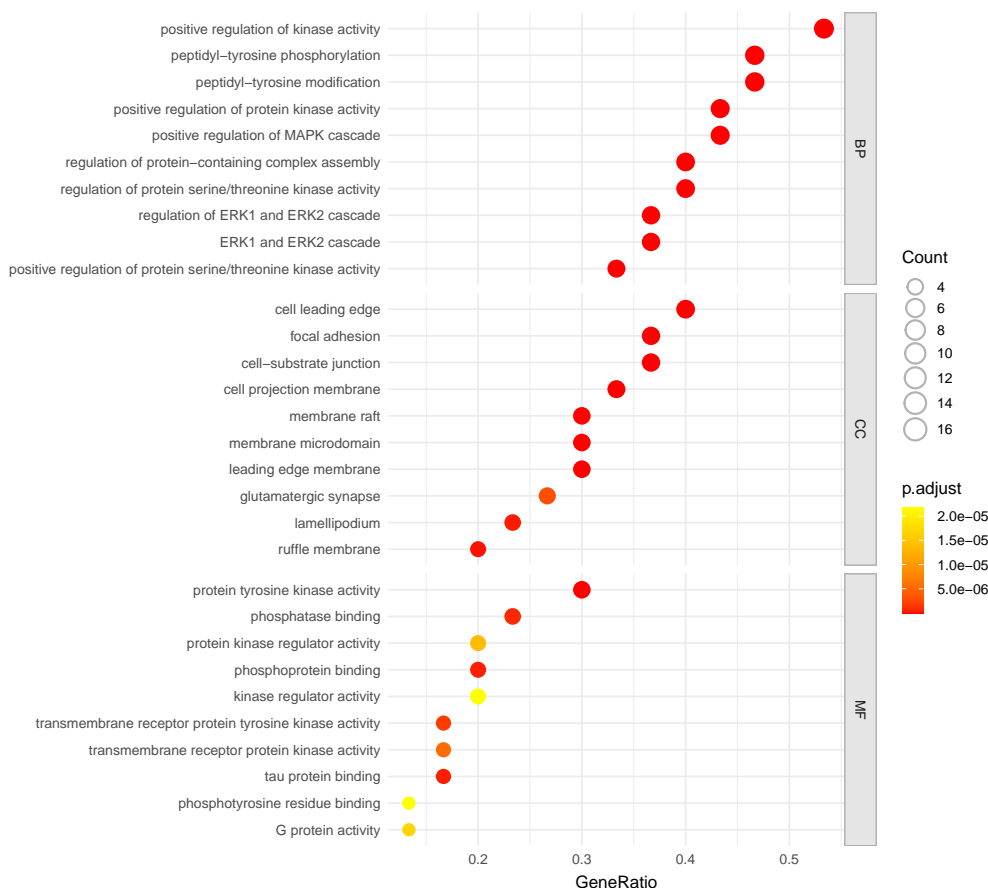


Figure 5: GO enrichment of MCC top 30

Figure 6为图 KEGG enrichment of MCC top 30 概览。

(对应文件为 Figure+Table/KEGG-enrichment-of-MCC-top-30.pdf)

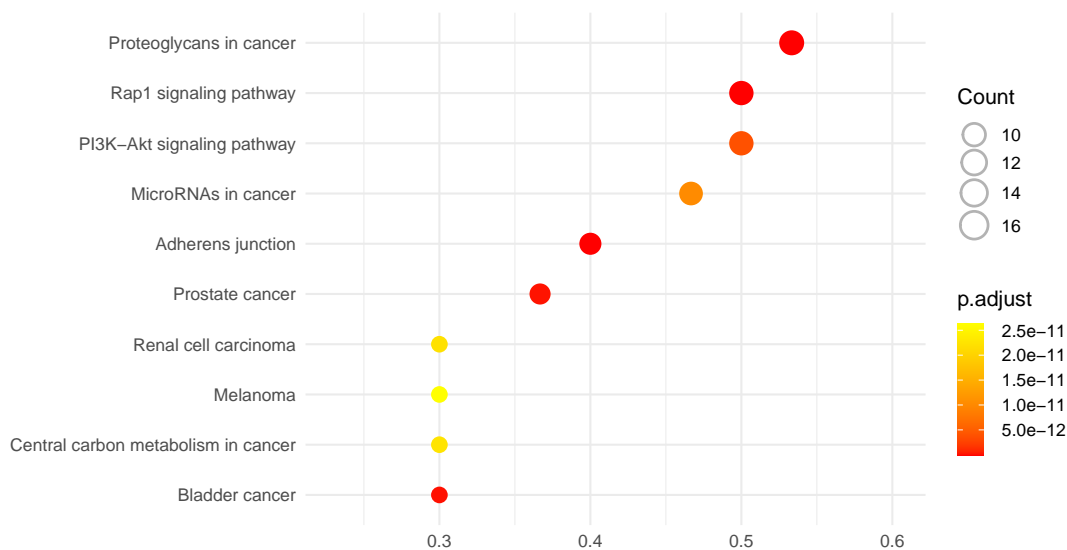


Figure 6: KEGG enrichment of MCC top 30

4.2 分析糖尿病肾病肠道差异菌群

4.2.1 选择公共数据库的 16s rRNA 数据

- 所选数据的文献来源²
 - Title: The Intestinal Microbiota Composition in Early and Late Stages of Diabetic Kidney Disease
 - PMID: 37341590
 - BioProject: PRJNA824185
- 相关信息
 - 引物: 805R (5'-GACTACHVGGGTATCTAATCC-3') and 341F (5'-CCTACGGGNGGCWGCAG-3')

Table 9为表格 metadata of the sra data of PRJNA824185 概览。

(对应文件为 **Figure+Table/metadata-of-the-sra-data-of-PRJNA824185.csv**)

注：表格共有 108 行 47 列，以下预览的表格可能省略部分数据；表格含有 108 个唯一 'Run'。

Table 9: Metadata of the sra data of PRJNA824185

Run	Relea...	LoadDate	spots	bases	spots...	avgLe...	size_MB	Assem...	downl...	Exper...	Libra.....	12 ..
SRR18..2023-	2022-	84114	42057000	500	13	NA	https...	SRX14..C_1	...			
...			
SRR18..2023-	2022-	85133	42566500	500	13	NA	https...	SRX14..C_10	...			
...			
SRR18..2023-	2022-	83213	41772920	502	12	NA	https...	SRX14..C_11	...			
...			

Run	Relea...	LoadDate	spots	bases	spots...	avgLe...	size_MB	Assem...	downl...	Exper...	Libra.....	12 ...
SRR18..2023-	...	2022-	80525	40423550	...	502	12	NA	https...	SRX14..C_12
...
SRR18..2023-	...	2022-	83961	41980500	...	500	13	NA	https...	SRX14..C_2
...
SRR18..2023-	...	2022-	81782	40891000	...	500	13	NA	https...	SRX14..C_3
...
SRR18..2023-	...	2022-	56959	28479500	...	500	9	NA	https...	SRX14..C_4
...
SRR18..2023-	...	2022-	85406	42703000	...	500	15	NA	https...	SRX14..C_5
...
SRR18..2023-	...	2022-	87838	43919000	...	500	15	NA	https...	SRX14..C_6
...
SRR18..2023-	...	2022-	87784	43892000	...	500	13	NA	https...	SRX14..C_7
...
SRR18..2023-	...	2022-	85552	42776000	...	500	13	NA	https...	SRX14..C_8
...
SRR18..2023-	...	2022-	82562	41281000	...	500	12	NA	https...	SRX14..C_9
...
SRR18..2023-	...	2022-	80309	40154500	...	500	12	NA	https...	SRX14..DM_10
...
SRR18..2023-	...	2022-	84125	42062500	...	500	13	NA	https...	SRX14..DM_11
...
SRR18..2023-	...	2022-	82632	41316000	...	500	13	NA	https...	SRX14..DM_12
...
...

4.2.2 下载和预处理 SRA 数据

使用 `sra-toolkit` 工具组的 `prefetch` 下载 SRA 数据，并用 `fastq-dump` 转化为 fastq 文件。

4.2.3 Qiime2

以下分析参照：<https://docs.qiime2.org/2023.7/tutorials/moving-pictures-usage/>

4.3 分析糖尿病肾病的肠道代谢组学差异

4.4 分析差异菌群与代谢物的相关性

5 结论

Reference

1. Chin, C.-H. *et al.* CytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology* **8**, S11 (2014).
2. Zhang, L. *et al.* The intestinal microbiota composition in early and late stages of diabetic kidney disease. *Microbiology Spectrum* **11**, (2023).