

Analysis

Huang LiChuang of Wie-Biotech

Contents

1 摘要	3
2 研究设计流程图	3
3 材料和方法	3
4 分析结果	4
4.1 网络药理学分析	4
4.1.1 从 HERB 网站获取中药和成分以及靶点数据	4
4.1.2 从 Genecards 获取疾病的靶点数据	6
4.1.3 数据透视	8
4.1.4 以 STRINGdb 构建 PPI 网络	11
4.1.5 筛选 HubGenes	12
4.1.6 通路富集分析	14
4.2 分析糖尿病肾病肠道差异菌群	15
4.2.1 选择公共数据库的 16s rRNA 数据	15
4.2.2 下载和预处理 SRA 数据	16
4.2.3 使用 Qiime2 作为上游分析	17
4.2.4 使用 MicrobiotaProcess 作为下游分析	17
4.2.5 Alpha 多样性	17
4.2.6 物种丰度	18
4.2.7 Beta 多样性	18
4.2.8 差异分析	21
4.3 分析糖尿病肾病的代谢组学差异	23
4.3.1 数据来源	23
4.3.2 代谢物筛选	25
4.3.3 富集分析	25
4.4 关联分析: 差异菌群与代谢物	27
4.4.1 使用关联数据	27
4.4.2 从差异菌群到差异代谢物	27
4.5 分析糖尿病肾病的转录组学差异	29
4.5.1 数据来源	29

4.5.2	数据标准化	30
4.5.3	差异分析	31
4.6	转录组学和网络药理学结合	33
4.7	代谢小分子靶点蛋白分析	33
4.7.1	分子对接	33
4.7.2	对接可视化	34
4.7.3	靶点蛋白的表达量以及相关通路	38
4.8	复方活性成分	39
4.8.1	作用于靶点蛋白的主要活性成分	39
5	结论	40
	Reference	40

List of Figures

1	Intersect of target genes of TCMs	9
2	Intersect of compounds of TCMs	10
3	Intersect of targets of compounds and disease targets	11
4	MCC Top 30	13
5	GO enrichment of MCC top 30	14
6	KEGG enrichment of MCC top 30	15
7	Alpha rarefaction curve	18
8	Evaluate alpha diversity in multiple methods	18
9	PCoA	19
10	Abundance in Family	20
11	Abundance in Genus	21
12	Significant Microbiota boxplot	22
13	Selected metabolites with multiple selection methods	25
14	Use MetaboAnalyst for metabolites enrichment	26
15	Use FELLA for metabolites enrichment	27
16	Integration of datasets for metabolites and microbiota selection	28
17	Metabolite Trimethylamine levels	29
18	RNA filtered genes	30
19	RNA nomalization	31
20	RNA seq DEG	32
21	Intersection of target genes as receptor	33
22	Docking affinity	34
23	Docking with FOXO3	35
24	Docking with IRAK1	36
25	Docking with CDKN1A	37
26	Compound structure 2D	37

27	Compound structure 3D	38
28	Pathway enrichment of receptors	39

List of Tables

1	TCM information	4
2	TCM compounds	5
3	Compounds targets	5
4	Compounds targets with annotation of biomaRt of ensembl dataset	6
5	Desease targets of diabetic nephropathy	7
6	Desease targets with annotation of biomaRt of ensembl dataset	7
7	ID mapped by STRINGdb	12
8	All MCC scores	13
9	Metadata of the sra data of PRJNA824185	15
10	Metadata of used 16s rRNA data	17
11	FDR significant Microbiota	22
12	Metadata of serum	23
13	Source data of serum	24
14	RNA DEG top table	32
15	Main active compounds to receptors	40

1 摘要

结合网络药理学、肠道菌微生物组 (16s rRNA)、代谢组、转录组、分子对接等技术, 筛选复方 (肾衰宁) 对糖尿病肾病 (DN) 治疗的主要活性成分和作用靶点。

16s rRNA 和代谢组结合, 筛选到差异菌 Lachnospiraceae 和差异代谢物 Trimethylamine。DN 患者 Lachnospiraceae 丰度下降, 而血清 Trimethylamine 水平升高, 提示 DN 患者由于 Lachnospiraceae 菌失调, 降低 Trimethylamine 的代谢能力。进一步结合网络药理学和转录组 (肾组织), 筛选到靶点蛋白 FOXO3, IRAK1, CDKN1A。DN 患者的 FOXO3, IRAK1, CDKN1A 表达量下降。分子对接应用于 Trimethylamine 和 FOXO3, IRAK1, CDKN1A 蛋白, 结果显示, Trimethylamine 对三者都有结合可能, 且对 FOXO3 有最高的亲和力。作用于 FOXO3, IRAK1, CDKN1A 等蛋白的复方的活性成分共有 12 个, 请参考 Tab. 15。

2 研究设计流程图

3 材料和方法

- Qiime2
- MetaboAnalystR
- gutMDisorder
- MicrobiotaProcess
- STRINGdb

- clusterProfiler
- FELLA
- limma
- Autodock Vina
- pymol
- R
- ...

4 分析结果

4.1 网络药理学分析

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
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

Ingre...	Targe.....2	Targe.....3	Datab...	Paper.id	PubMe...
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	...
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	Gluco...	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...

hgnc_...	ensem.....2	ensem.....3	entre...	chrom...	start...	end_p...	descr...
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**)

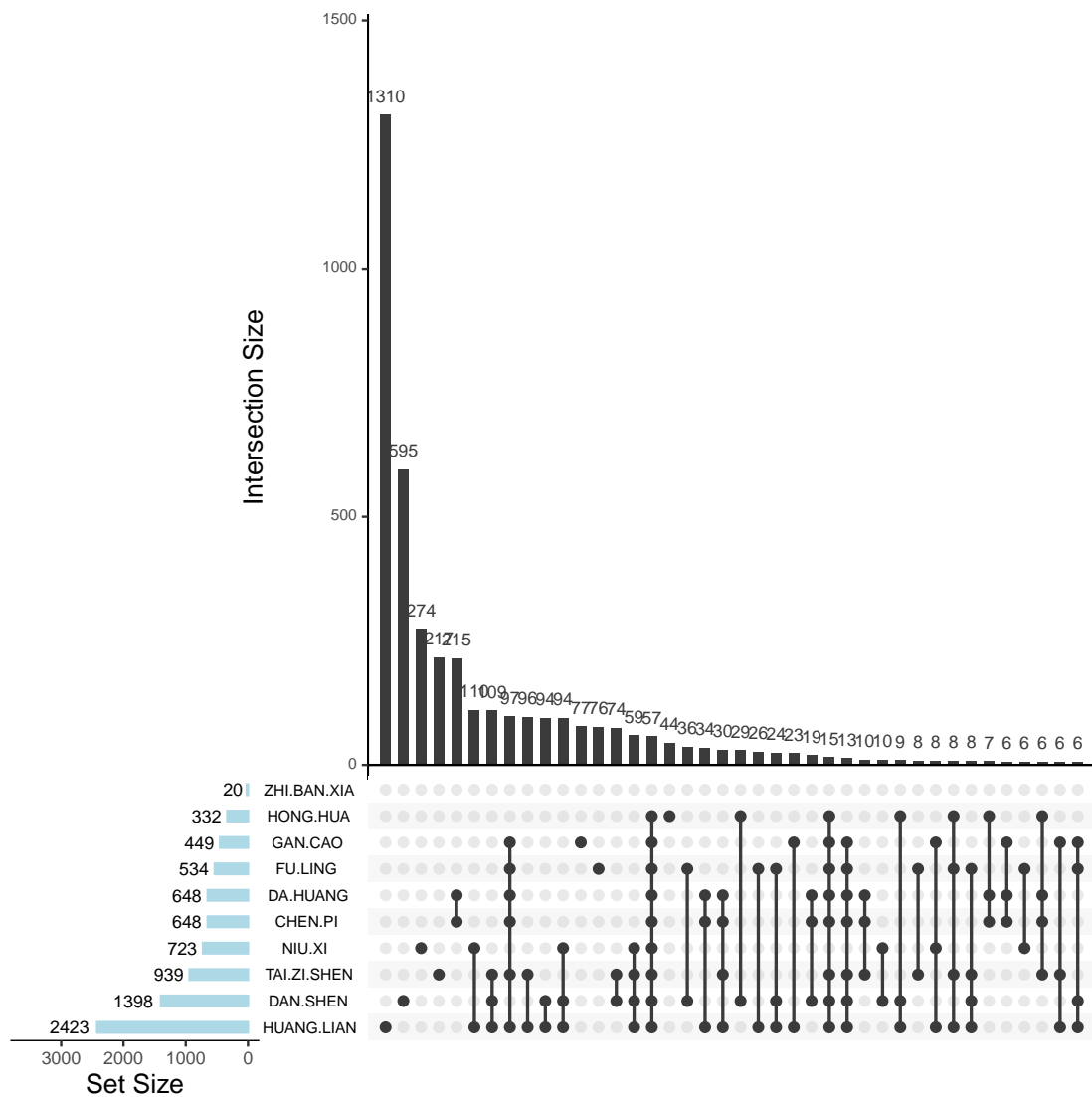


Figure 1: Intersect of target genes of TCMs

Figure 2为图 intersect of compounds of TCMs 概览。

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

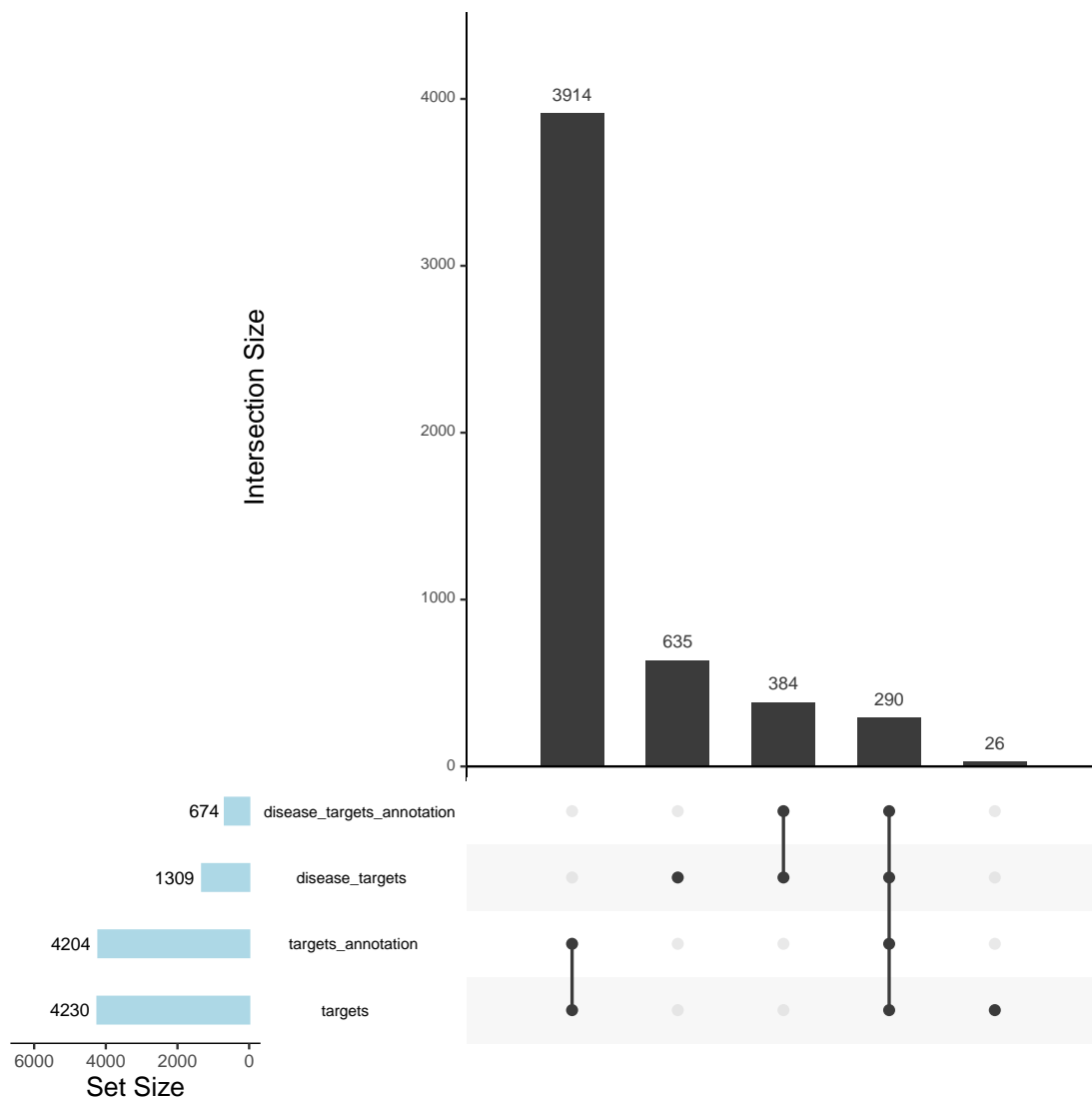


Figure 3: Intersect of targets of compounds and disease targets

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)

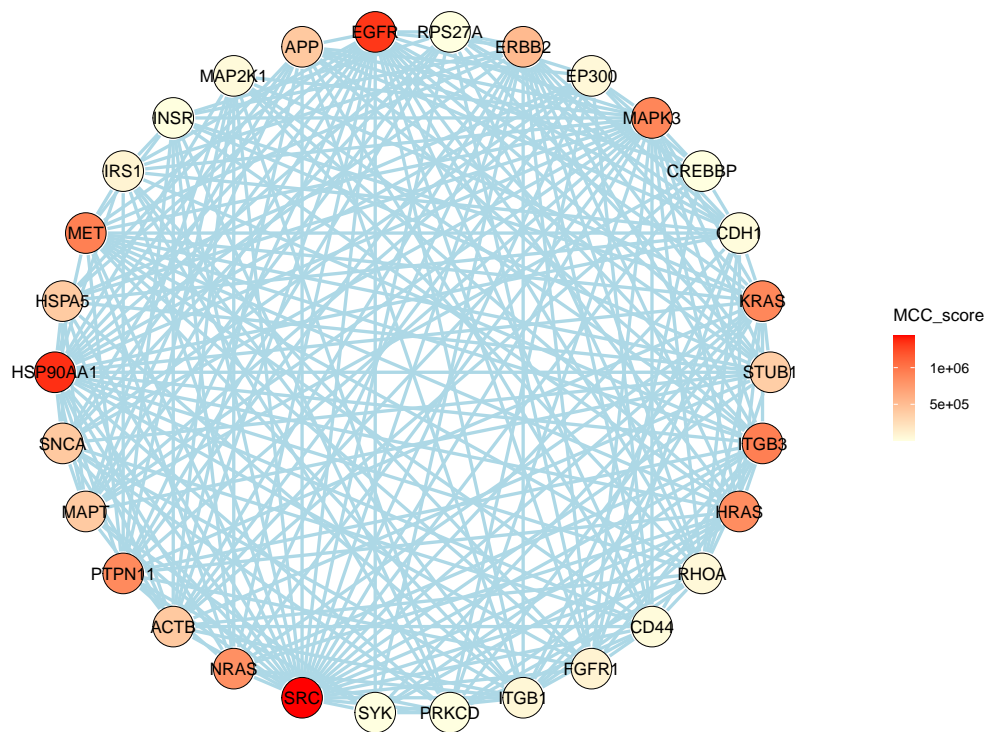


Figure 4: MCC Top 30

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)

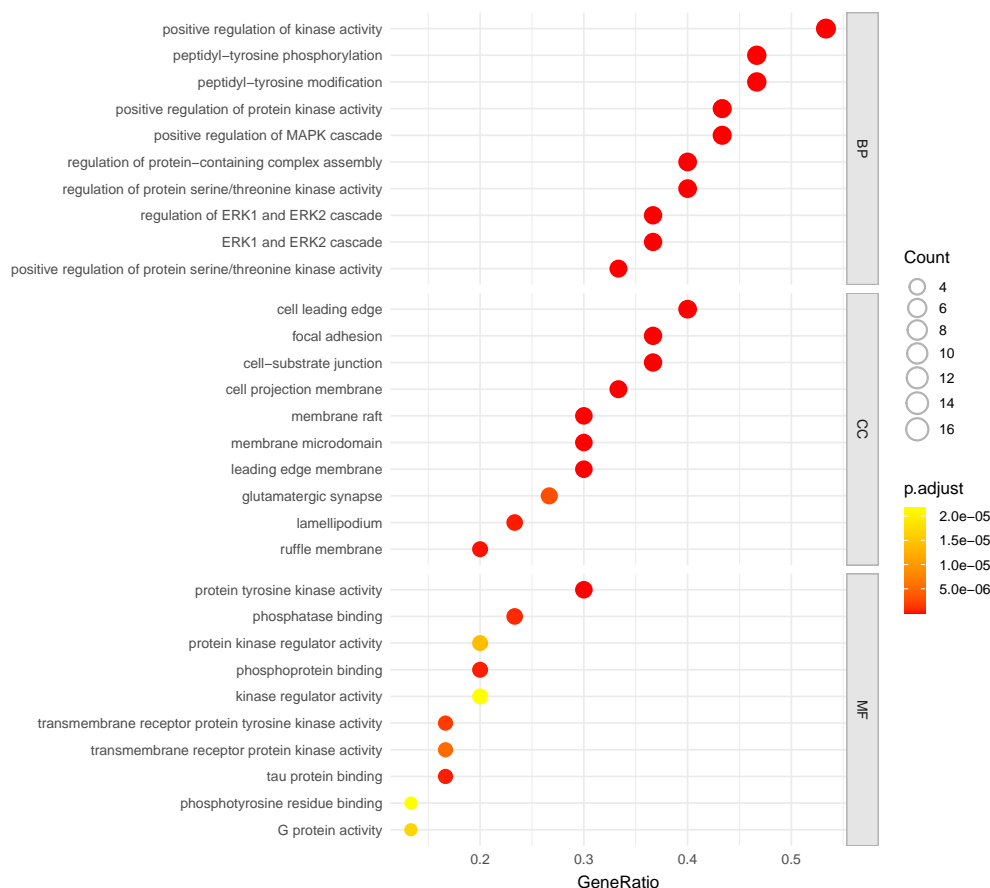


Figure 5: GO enrichment of MCC top 30

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)

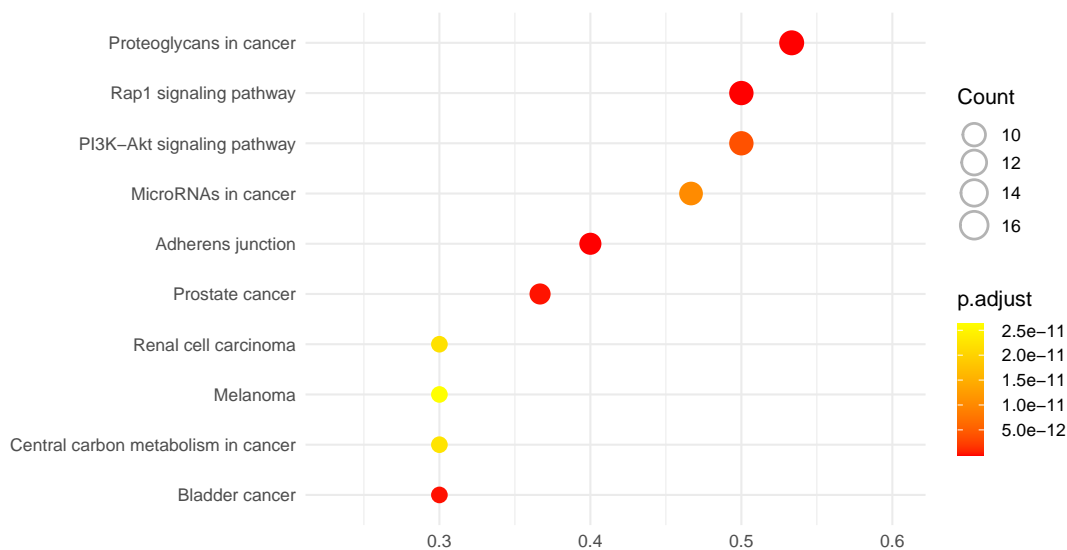


Figure 6: KEGG enrichment of MCC top 30

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。

实际使用的数据为：

- Control 组
- Diabetic Nephropathy 组

Table 10为表格 metadata of used 16s rRNA data 概览。

(对应文件为 `Figure+Table/metadata-of-used-16s-rRNA-data.csv`)

注：表格共有 78 行 5 列，以下预览的表格可能省略部分数据；表格含有 2 个唯一 ‘group’。

Table 10: Metadata of used 16s rRNA data

sampl...	Run	forwa...	rever...	group
C.1	SRR18...	/home...	/home...	C
C.10	SRR18...	/home...	/home...	C
C.11	SRR18...	/home...	/home...	C
C.12	SRR18...	/home...	/home...	C
C.2	SRR18...	/home...	/home...	C
C.3	SRR18...	/home...	/home...	C
C.4	SRR18...	/home...	/home...	C
C.5	SRR18...	/home...	/home...	C
C.6	SRR18...	/home...	/home...	C
C.7	SRR18...	/home...	/home...	C
C.8	SRR18...	/home...	/home...	C
C.9	SRR18...	/home...	/home...	C
DN.16	SRR18...	/home...	/home...	DN
DN.17	SRR18...	/home...	/home...	DN
DN.18	SRR18...	/home...	/home...	DN
...

4.2.3 使用 Qiime2 作为上游分析

以下分析参照³⁻⁷: <https://docs.qiime2.org/2023.7/tutorials/moving-pictures-usage/>

- importing data
- Demultiplexing sequences
- Sequence quality control and feature table construction
 - DADA2
- FeatureTable and FeatureData summaries
- Generate a tree for phylogenetic diversity analyses
- Alpha and beta diversity analysis
- Alpha rarefaction plotting
- Taxonomic analysis
- Differential abundance testing with ANCOM

4.2.4 使用 MicrobiotaProcess 作为下游分析

结果见以下部分。

4.2.5 Alpha 多样性

Control 组和 Diabetic Nephropathy 组的 Alpha 多样性没有显著差异 (Fig. 8)。

Figure 7为图 alpha rarefaction curve 概览。

(对应文件为 Figure+Table/alpha-rarefaction-curve.pdf)

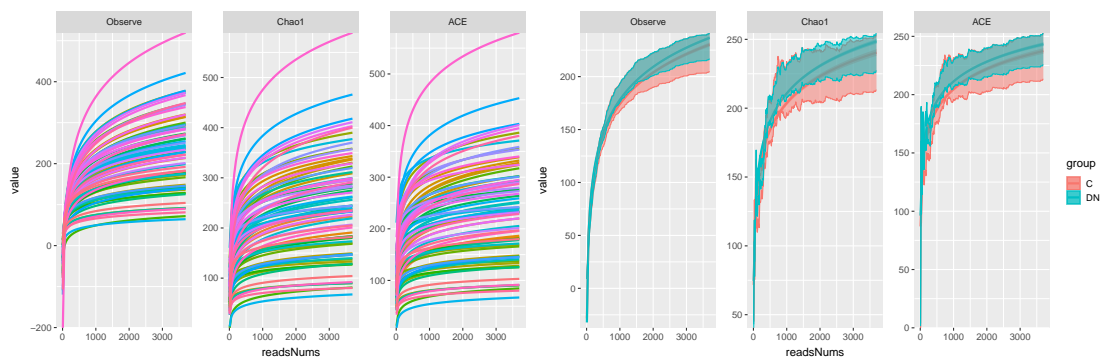


Figure 7: Alpha rarefaction curve

Figure 8为图 evaluate alpha diversity in multiple methods 概览。

(对应文件为 Figure+Table/evaluate-alpha-diversity-in-multiple-methods.pdf)

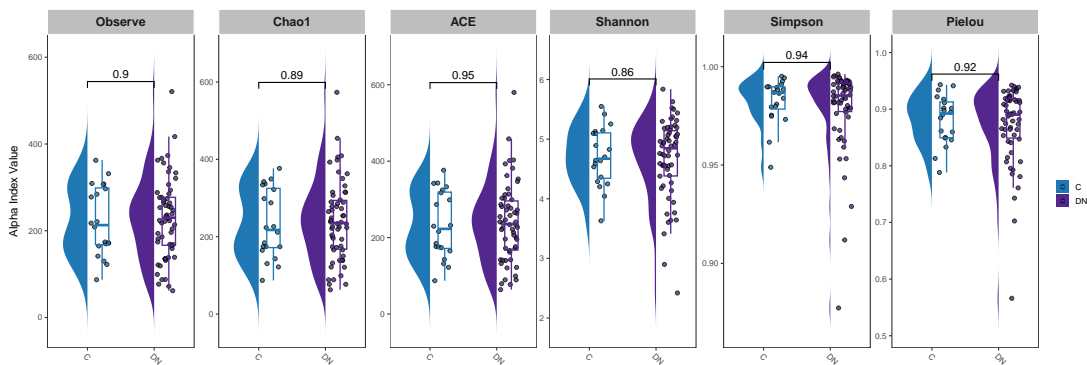


Figure 8: Evaluate alpha diversity in multiple methods

4.2.6 物种丰度

门、纲、目、科、属、种对应为：

taxonomy :

Phylum, Class, Order, Family, Genus, Species

4.2.7 Beta 多样性

在 Beta 分析中，C 组和 DN 组未发生完全区分的聚类。

Figure 9为图 PCoA 概览。

(对应文件为 Figure+Table/PCoA.pdf)

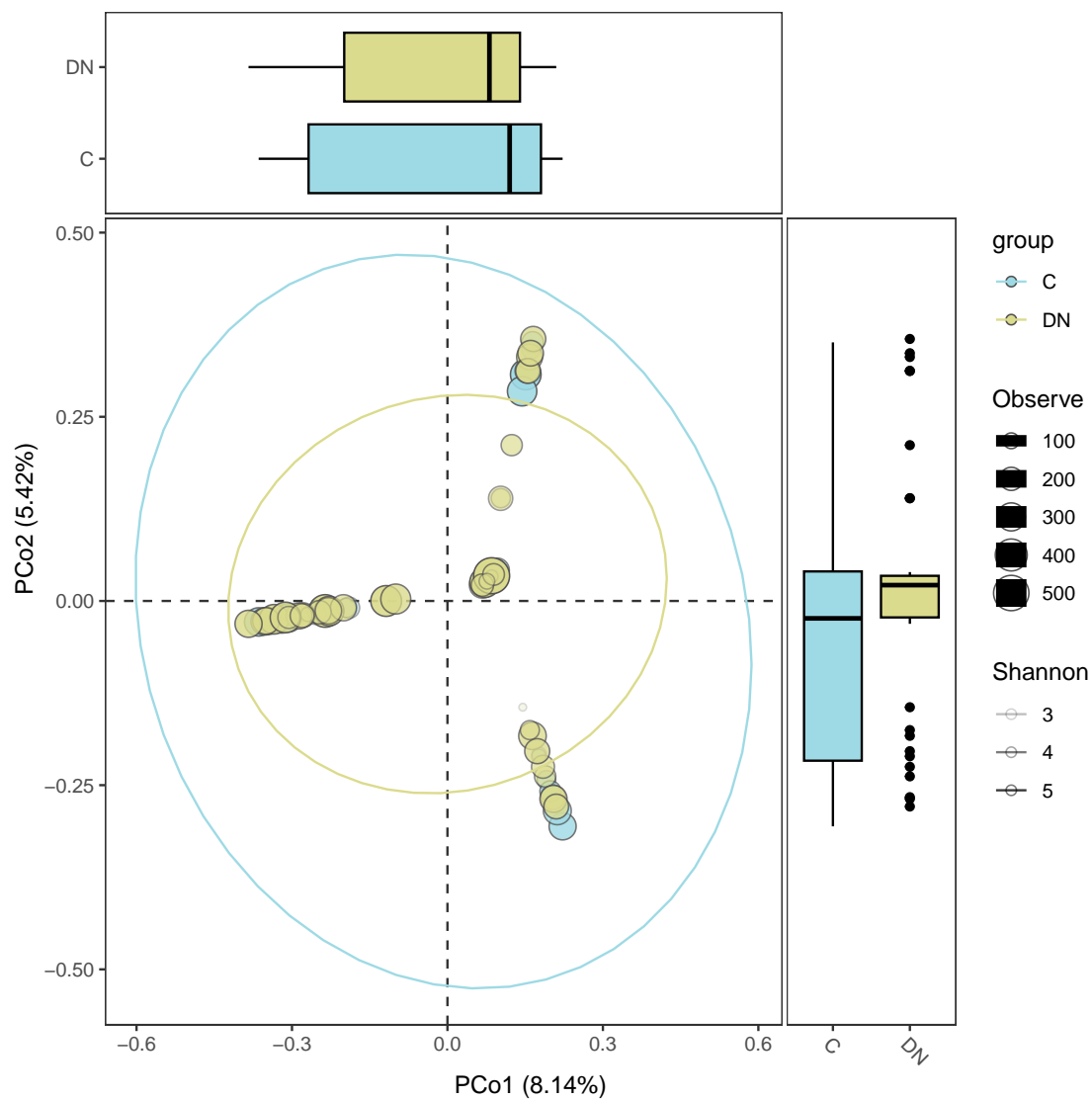


Figure 9: PCoA

Figure 10为图 abundance in Family 概览。

(对应文件为 Figure+Table/abundance-in-Family.pdf)

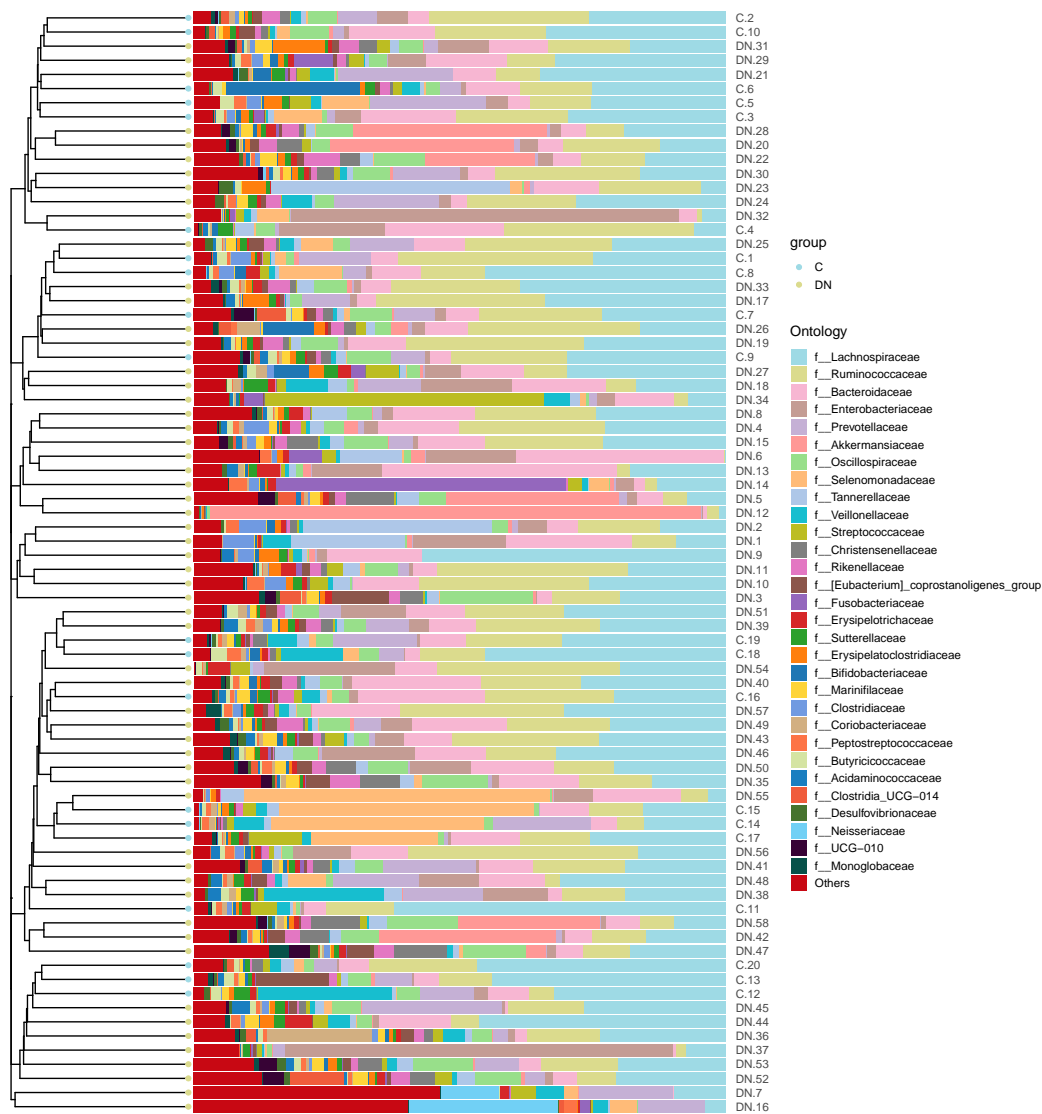


Figure 10: Abundance in Family

Figure 11为图 abundance in Genus 概览。

(对应文件为 Figure+Table/abundance-in-Genus.pdf)

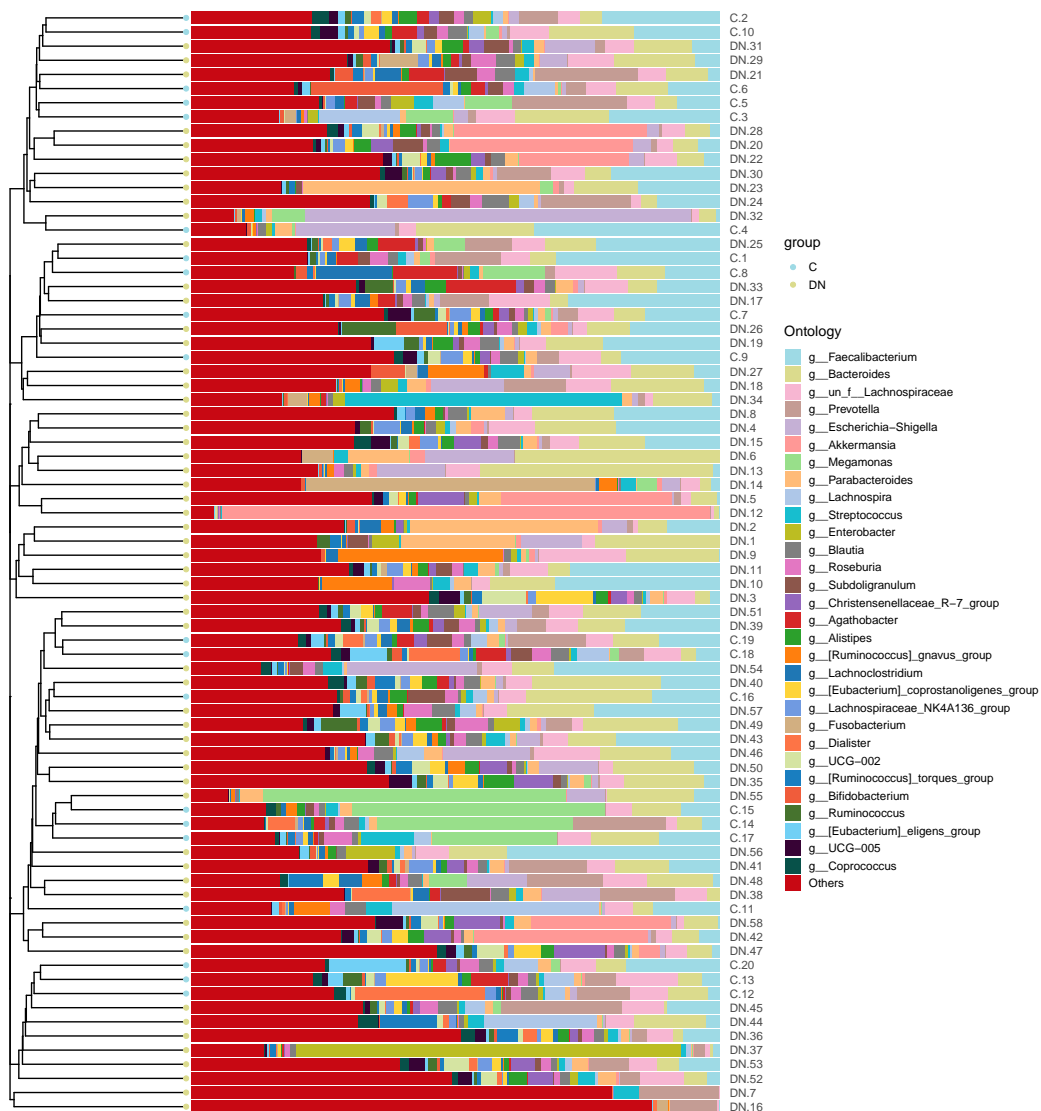


Figure 11: Abundance in Genus

4.2.8 差异分析

相比于 C 组，DN 组的部分菌丰度显著降低 (Tab. 11)。

Figure 12为图 significant Microbiota boxplot 概览。

(对应文件为 [Figure+Table/significant-Microbiota-boxplot.pdf](#))

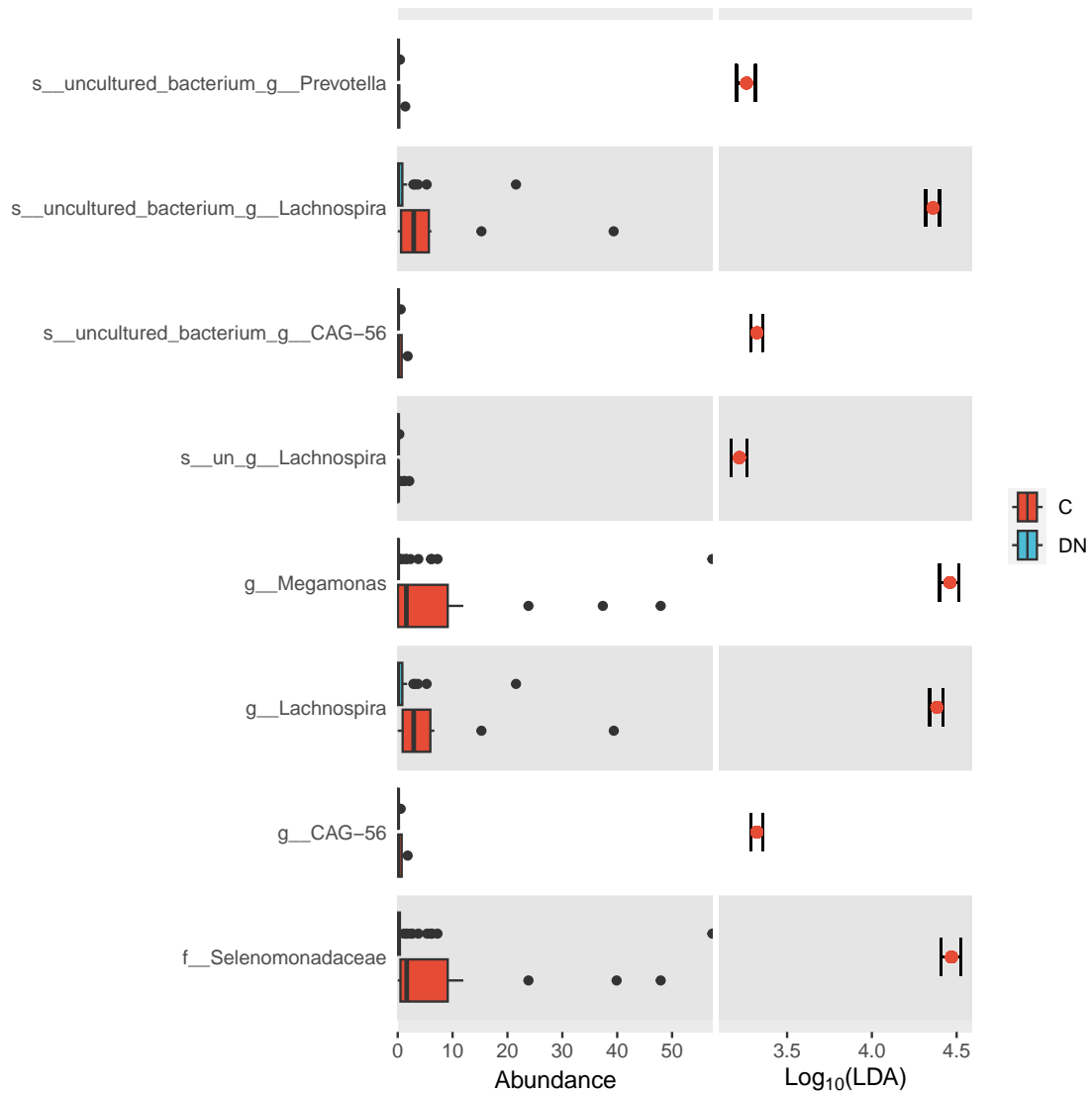


Figure 12: Significant Microbiota boxplot

Table 11为表格 FDR significant Microbiota 概览。

(对应文件为 **Figure+Table/FDR-significant-Microbiota.csv**)

注：表格共有 10 行 8 列，以下预览的表格可能省略部分数据；表格含有 10 个唯一 ‘label’。

Table 11: FDR significant Microbiota

label	nodeC...	pvalue	fdr	LDAupper	LDAmean	LDAlower	Sign_...
g__CA...	Genus	6.558...	0.004...	3.355...	3.321...	3.285...	C
s__un...	Species	6.558...	0.004...	3.355...	3.321...	3.285...	C
g__La...	Genus	6.577...	0.032...	4.421...	4.383...	4.341...	C
7924a...	OTU	1.637...	0.037...	3.598...	3.539...	3.469...	C

label	nodeC...	pvalue	fdr	LDAupper	LDAmean	LDAlower	Sign_...
d581e...	OTU	1.637...	0.037...	3.465...	3.418...	3.365...	C
f__Se...	Family	1.512...	0.037...	4.523...	4.469...	4.408...	C
s__un...	Species	1.842...	0.037...	3.313...	3.260...	3.201...	C
s__un...	Species	2.010...	0.037...	4.400...	4.360...	4.317...	C
g__Me...	Genus	2.279...	0.037...	4.514...	4.460...	4.398...	C
s__un...	Species	2.963...	0.043...	3.263...	3.218...	3.168...	C

关于 LDA⁸。

4.3 分析糖尿病肾病的代谢组学差异

4.3.1 数据来源

该数据为血清代谢组。血清代谢组可以反映肠道菌丰度，请参考⁹。

数据来源：

- Title: Serum Metabolomics Benefits Discrimination Kidney Disease Development in Type 2 Diabetes Patients
- PMID: 35615098

Table 12为表格 metadata of serum 概览。

(对应文件为 `Figure+Table/metadata-of-serum.csv`)

注：表格共有 89 行 2 列，以下预览的表格可能省略部分数据；表格含有 3 个唯一 ‘group’。

Table 12: Metadata of serum

group	sample
Control	C301
Control	C317
Control	C320
Control	C321
Control	C322
Control	C328
Control	C329
Control	C331
Control	C334
Control	C337
Control	C340
Control	C302
Control	C343

group	sample
Control	C345
Control	C347
...	...

Table 13为表格 source data of serum 概览。

(对应文件为 **Figure+Table/source-data-of-serum.xlsx**)

注：表格共有 1042 行 90 列，以下预览的表格可能省略部分数据；表格含有 1042 个唯一‘rownames’。

Table 13: Source data of serum

rownames	C301	C317	C320	C321	C322	C328	C329	C331	C334	C337	C340	C302	C343	C345	C347	C348	C349	C353	C355	...
Acryl...	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	...
	0.11...	0.08...	0.08...	0.11...	0.09...	0.08...	0.14...	0.19...	0.12...	0.02...	0.12...	0.19...	0.10...	0.18...	0.11...	0.14...	0.13...	0.13...	0.06...	...
Butanal	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	...
	0.27...	0.24...	0.35...	0.30...	0.28...	0.35...	0.30...	0.29...	0.33...	0.31...	0.25...	0.29...	0.32...	0.30...	0.30...	0.32...	0.34...	0.26...	0.20...	...
Acety...	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	...
	0.18...	0.16...	0.22...	0.21...	0.21...	0.22...	0.20...	0.19...	0.24...	0.21...	0.19...	0.20...	0.23...	0.23...	0.20...	0.20...	0.22...	0.17...	0.15...	...
Time...	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	...
	0.10...	0.05...	0.00...	0.12...	0.09...	0.09...	0.08...	0.08...	0.11...	0.03...	0.15...	0.12...	0.17...	0.18...	0.13...	0.10...	0.06...	0.05...	0.12...	...
N-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	...
oxide	0.27...	0.24...	0.35...	0.30...	0.28...	0.35...	0.30...	0.29...	0.33...	0.31...	0.25...	0.29...	0.32...	0.17...	0.30...	0.32...	0.34...	0.26...	0.20...	...
HydroL	153.0	484.1	108.0	266.0	188.1	310.0	474.2	068.5		0.546	0.050	0.606	1.468	0.820	0.768	0.207	2.574	0.437	0.087	...
																				0.00...
Cyste...	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	...
	0.09...	0.15...	0.22...	0.20...	0.17...	0.21...	0.18...	0.10...	0.33...	0.20...	0.25...	0.19...	0.10...	0.15...	0.19...	0.32...	0.12...	0.17...	0.20...	...
Benzene	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	...
	0.27...	0.24...	0.35...	0.30...	0.28...	0.35...	0.30...	0.29...	0.33...	0.31...	0.25...	0.29...	0.32...	0.30...	0.30...	0.32...	0.34...	0.26...	0.20...	...
2-	0.852	0.719	0.893	0.610	0.425	0.704	0.467	0.650	0.843	0.590	0.506	0.710	0.969	0.939	0.900	0.636	0.732	0.713	0.475	...
Chl...																				...
Sulfite	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	...
	0.07...	0.11...	0.14...	0.09...	0.13...	0.14...	0.14...	0.11...	0.09...	0.14...	0.11...	0.15...	0.14...	0.12...	0.14...	0.11...	0.15...	0.12...	0.08...	...
5-	0.419	0.336	0.639	0.379	0.454	0.622	0.419	0.447	0.528	0.446	0.374	0.410	0.520	0.418	0.455	0.596	0.583	0.414	0.297	...
Ami...																				...
Imida...	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	...
	0.09...	0.05...	0.14...	0.18...	0.06...	0.01...	0.06...	0.14...	0.13...	0.15...	0.14...	0.04...	0.09...	0.19...	0.03...	0.05...	0.05...	0.03...	0.07...	...
Penta...	0.001																			...
	0.16...		0.06...	0.16...	0.14...	0.15...	0.17...	0.12...	0.18...	0.07...	0.15...	0.13...	0.14...	0.16...	0.15...	0.14...	0.12...	0.05...	0.13...	...

rowname	C301	C317	C320	C321	C322	C328	C329	C331	C334	C337	C340	C302	C343	C345	C347	C348	C349	C353	C355	...
2-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	...
Ami...	0.15...	0.12...	0.18...	0.20...	0.18...	0.18...	0.13...	0.16...	0.15...	0.12...	0.15...	0.14...	0.19...	0.18...	0.12...	0.19...	0.18...	0.17...	0.07...	...
Pyruv...	0.181...	0.120...	0.252...	0.152...	0.000...	0.287...	0.164...	0.220...		0.140...		0.183...	0.144...	0.022...	0.177...	0.294...	0.268...	0.174...	0.004...	...
								0.03...		0.00...										...
...

4.3.2 代谢物筛选

Figure 13为图 selected metabolites with multiple selection methods 概览。

(对应文件为 Figure+Table/selected-metabolites-with-multiple-selection-methods.pdf)

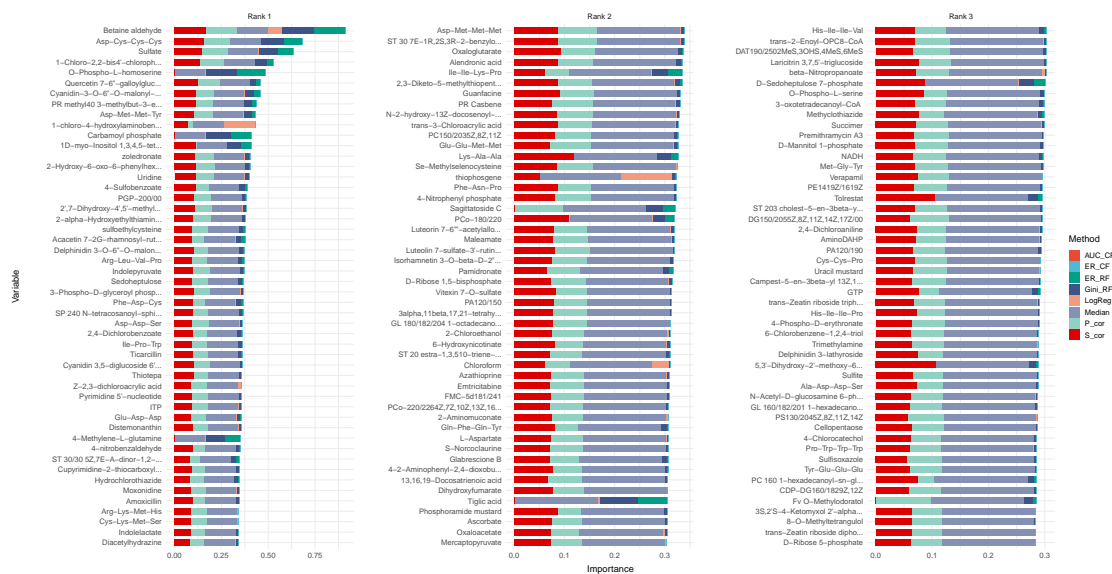


Figure 13: Selected metabolites with multiple selection methods

4.3.3 富集分析

取 Fig. 13 的代谢物做富集分析。

以下内容仅供参考，无直接结论。

4.3.3.1 MetaboAnalyst Figure 14为图 use MetaboAnalyst for metabolites enrichment 概览。

(对应文件为 Figure+Table/metabolites_ORA_dot_kegg_pathway_dpi72.pdf)

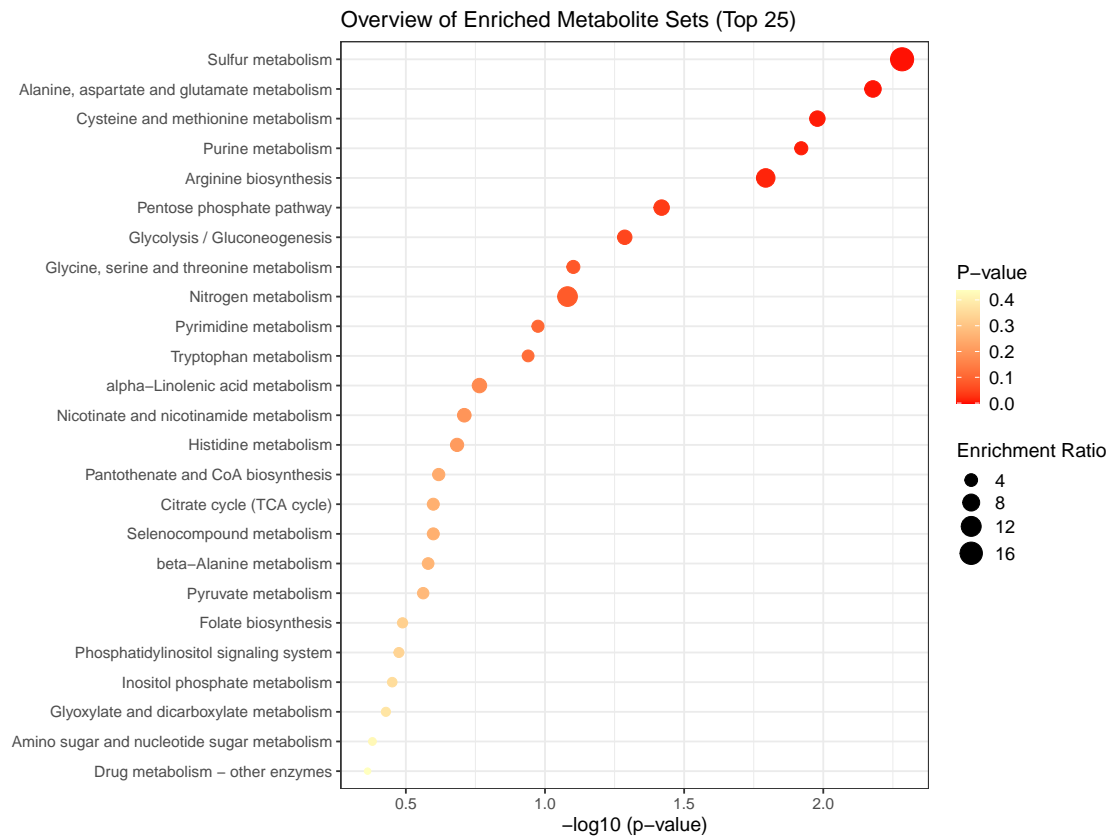


Figure 14: Use MetaboAnalyst for metabolites enrichment

4.3.3.2 FELLA Figure 15为图 use FELLA for metabolites enrichment 概览。

(对应文件为 `Figure+Table/use-FELLA-for-metabolites-enrichment.pdf`)

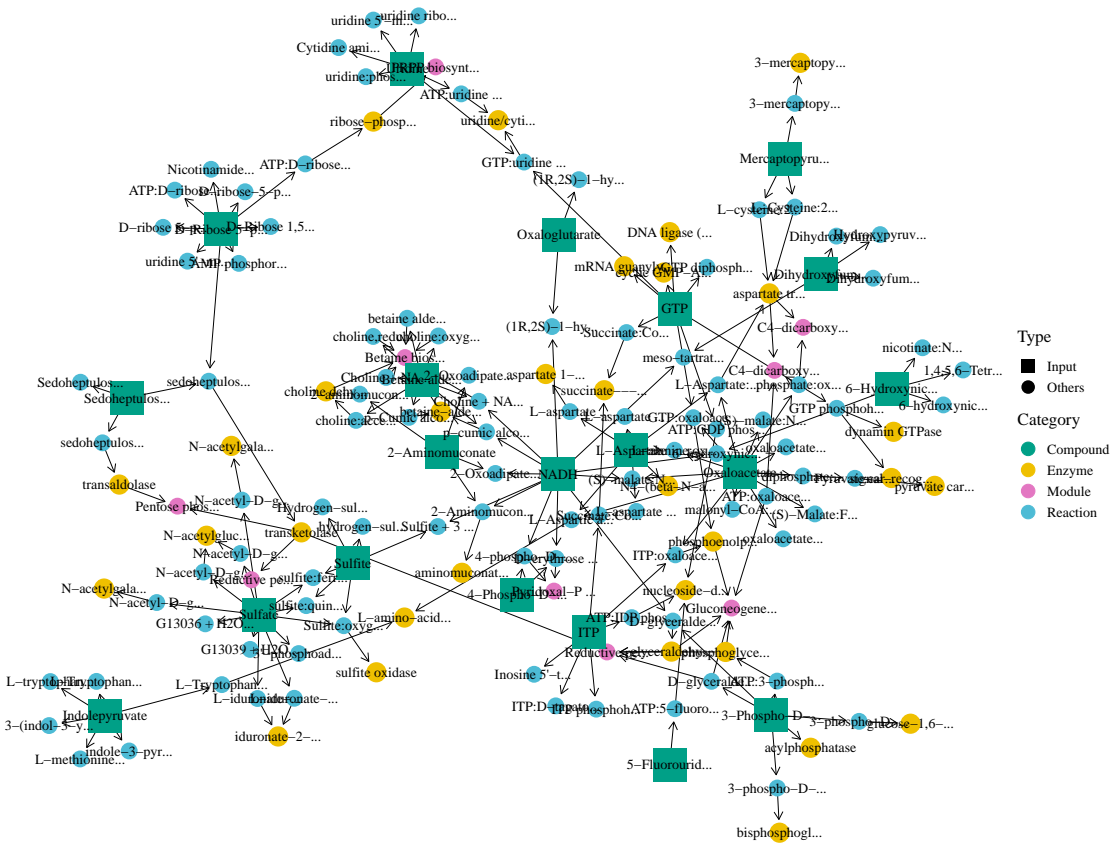


Figure 15: Use FELLA for metabolites enrichment

4.4 关联分析：差异菌群与代谢物

4.4.1 使用关联数据

血清代谢物和肠道菌群的相关性¹⁰

- Title: gutMDisorder: a comprehensive database for dysbiosis of the gut microbiota in disorders and interventions

该数据集反映了肠道菌与代谢物之间的关联。

4.4.2 从差异菌群到差异代谢物

将 Fig. 13 的代谢物和 Fig. 12 的差异菌，结合 gutMDisorder 数据集，进一步筛选差异代谢物和差异菌。

Figure 16为图 Integration of datasets for metabolites and microbiota selection 概览。

(对应文件为 [Figure+Table/Integration-of-datasets-for-metabolites-and-microbiota-selection.pdf](#))

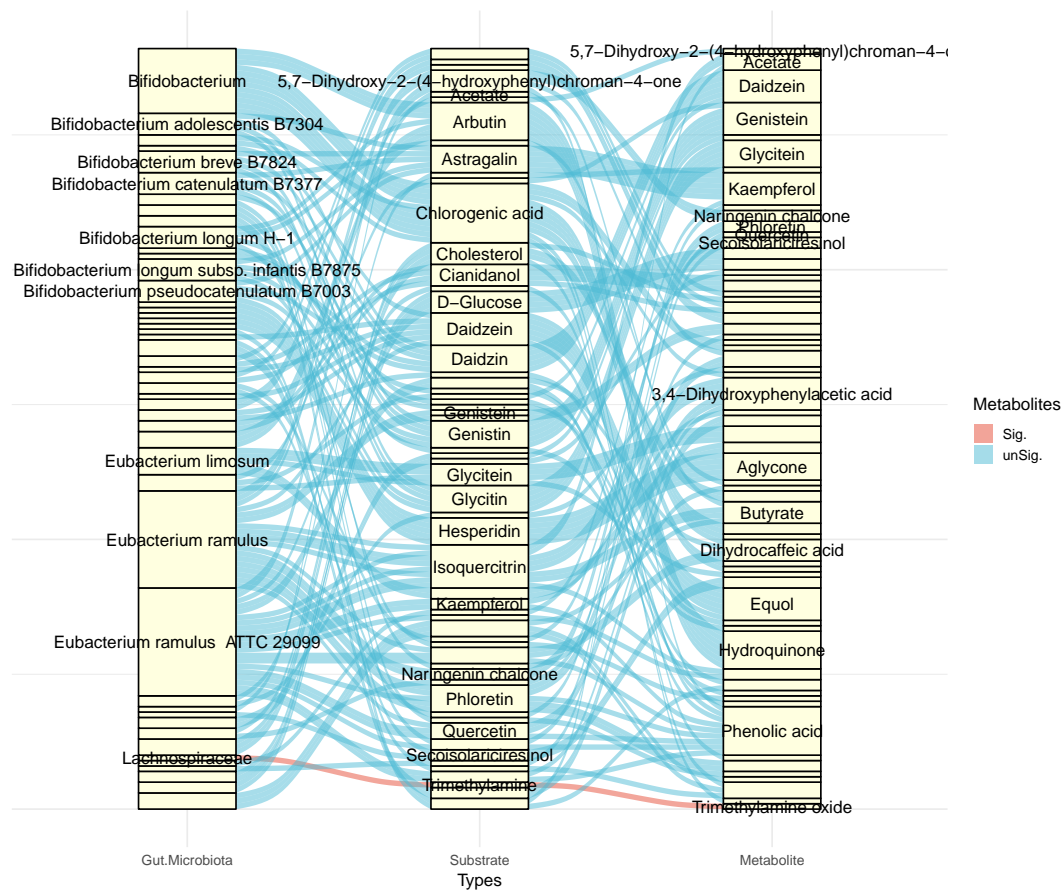


Figure 16: Integration of datasets for metabolites and microbiota selection

Fig. 16 筛选到:

- Lachnospiraceae¹¹
- Trimethylamine¹²

Figure 17为图 Metabolite Trimethylamine levels 概览。

(对应文件为 Figure+Table/Metabolite-Trimethylamine-levels.pdf)

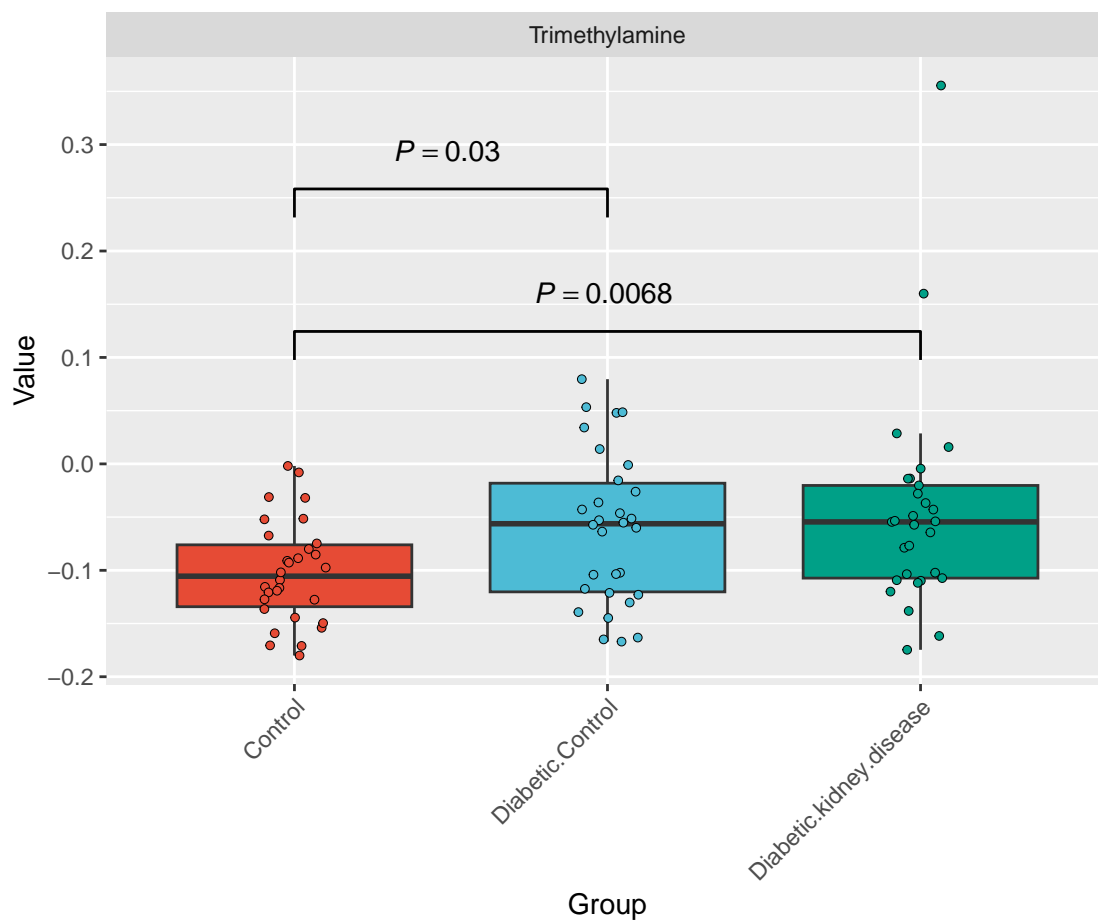


Figure 17: Metabolite Trimethylamine levels

4.5 分析糖尿病肾病的转录组学差异

4.5.1 数据来源

- GSE199838

data_processing :

Illumina Bcl2FastQ software used for basecalling.

data_processing.1 :

Sequenced reads were trimmed for adaptor sequence, and masked for low-complexity or low-quality sequence. The remaining reads were filtered against the rRNA database to remove possible ribosomal RNA contamination, and then mapped to the hg19 whole genome using Hisat2 v2.1.0 with default parameters.

data_processing.2 :

HTSeq v0.11.2 was subsequently employed to convert aligned short reads into read counts for each gene model.

data_processing.3 :

Assembly: hg19

data_processing.4 :

Supplementary files format and content: tab-delimited text files include RPKM for each Sample

4.5.2 数据标准化

Figure 18为图 RNA filtered genes 概览。

(对应文件为 `Figure+Table/RNA-filtered-genes.pdf`)

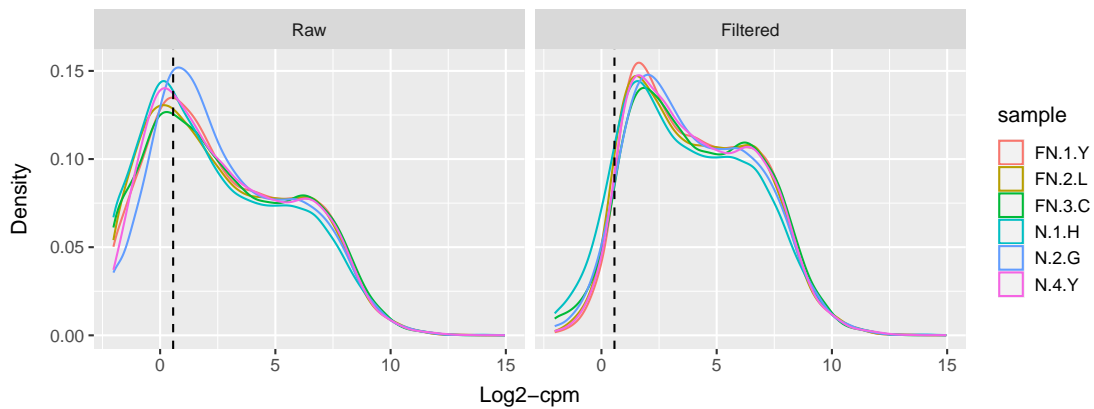


Figure 18: RNA filtered genes

Figure 19为图 RNA nomalization 概览。

(对应文件为 Figure+Table/RNA-normalization.pdf)

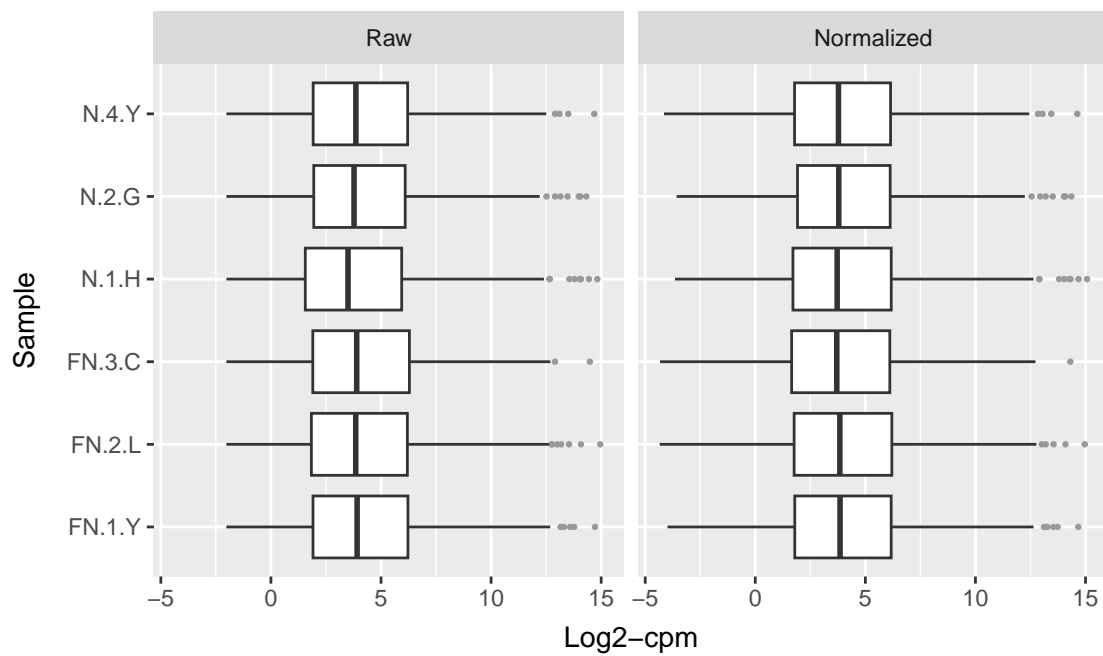


Figure 19: RNA normalization

4.5.3 差异分析

Figure 20为图 RNA seq DEG 概览。

(对应文件为 Figure+Table/RNA-seq-DEG.pdf)

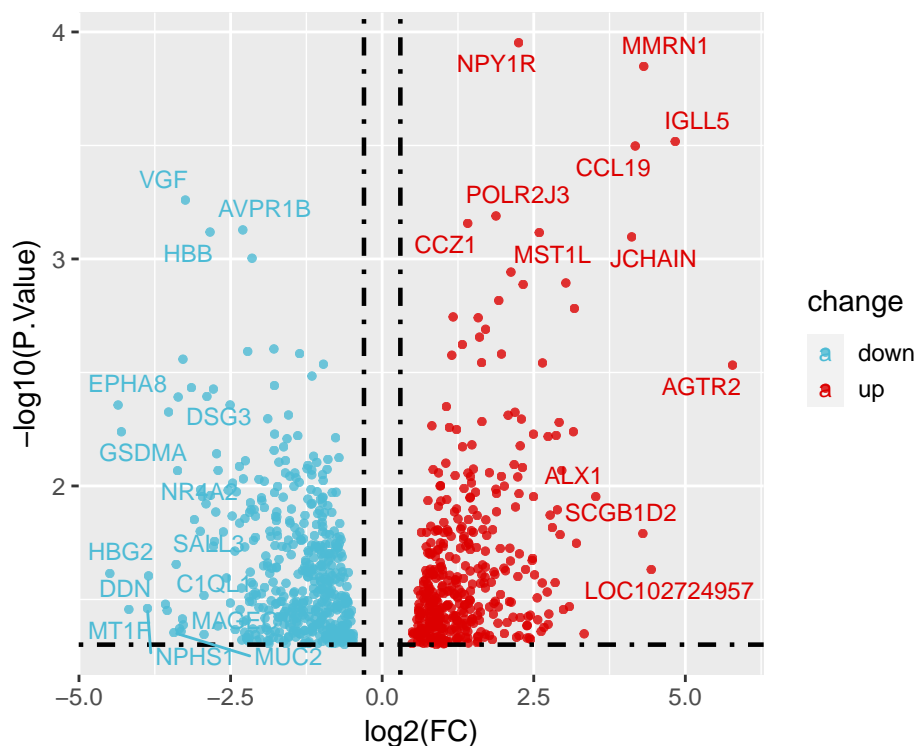


Figure 20: RNA seq DEG

Table 14为表格 RNA DEG top table 概览。

(对应文件为 Figure+Table/RNA-DEG-top-table.csv)

注：表格共有 865 行 7 列，以下预览的表格可能省略部分数据；表格含有 865 个唯一‘hgnc_symbol’。

Table 14: RNA DEG top table

hgnc_...	logFC	AveExpr	t	P.Value	adj.P...	B
NPY1R	2.251...	6.749...	7.572...	0.000...	0.618...	-3.90...
HBB	-2.83...	8.479...	-5.57...	0.000...	0.618...	-3.91...
COL14A1	2.124...	7.990...	5.201...	0.001...	0.618...	-3.94...
ASAH1	1.170...	7.424...	4.806...	0.001...	0.618...	-3.97...
VAR5	-0.96...	7.518...	-4.40...	0.002...	0.618...	-3.99...
RHOB	-1.15...	9.861...	-4.31...	0.003...	0.618...	-4.01...
MT1E	-3.14...	8.226...	-4.21...	0.003...	0.618...	-4.02...
MT2A	-2.89...	8.428...	-4.14...	0.004...	0.618...	-4.02...
FGL2	1.058...	8.156...	4.069...	0.004...	0.618...	-4.03...
C4B	-1.88...	7.510...	-3.97...	0.005...	0.618...	-4.04...
SLC27A4	-0.76...	7.183...	-3.82...	0.006...	0.618...	-4.06...
VASN	-1.13...	7.761...	-3.68...	0.007...	0.618...	-4.07...

hgnc_...	logFC	AveExpr	t	P.Value	adj.P...	B
PABPC1	-0.71...	9.322...	-3.67...	0.007...	0.618...	-4.08...
TTC28	-0.90...	8.207...	-3.58...	0.008...	0.618...	-4.08...
TRABD2B	-1.78...	7.019...	-3.73...	0.006...	0.618...	-4.08...
...

4.6 转录组学和网络药理学结合

取 4.1.5 中 top 100 的基因，与 Tab. 14 中满足条件 ($P.Value < .05$, $|\log FC| > 1$) 的基因。

Figure 21为图 intersection of target genes as receptor 概览。

(对应文件为 Figure+Table/intersection-of-target-genes-as-receptor.pdf)

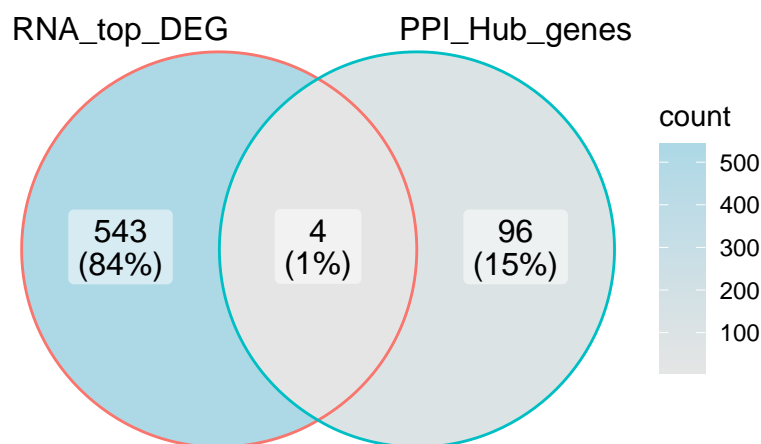


Figure 21: Intersection of target genes as receptor

所有的候选受体（用于分子对接 4.7

receptors :
DCN, IRAK1, CDKN1A, FOXO3

4.7 代谢小分子靶点蛋白分析

4.7.1 分子对接

使用 Autodock vina 对接¹³。

流程请参考文献¹⁴ 或者 https://autodock-vina.readthedocs.io/en/latest/docking_basic.html。

Figure 22为图 docking affinity 概览。

(对应文件为 Figure+Table/docking-affinity.pdf)

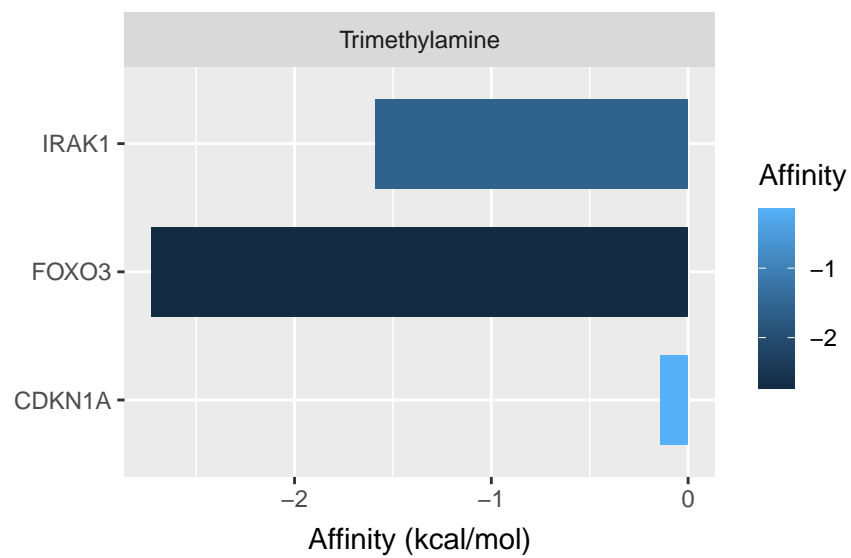


Figure 22: Docking affinity

4.7.2 对接可视化

使用 PyMol 工具将结果可视化¹⁵。

Figure 23为图 docking with FOXO3 概览。

(对应文件为 Figure+Table/1146_into_2lqh.png)

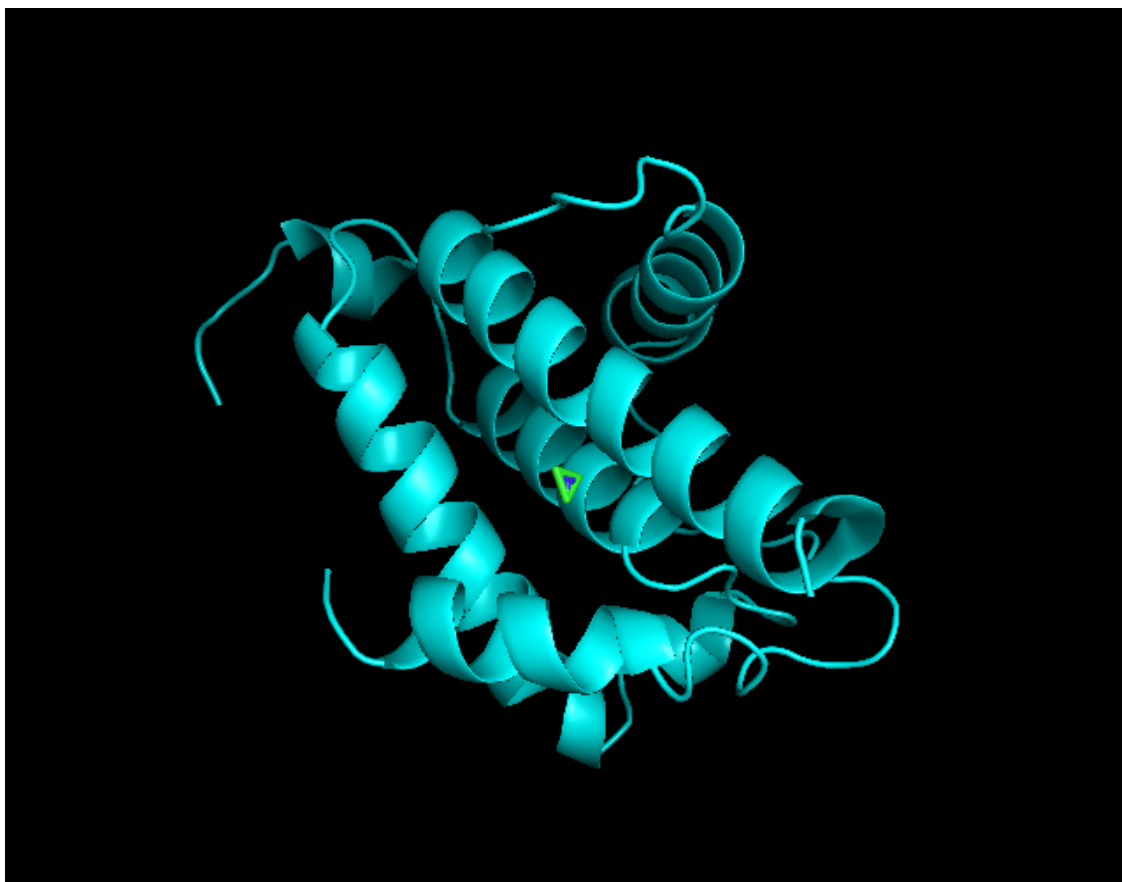


Figure 23: Docking with FOXO3

Figure 24为图 docking with IRAK1 概览。

(对应文件为 Figure+Table/1146_into_6bfn.png)

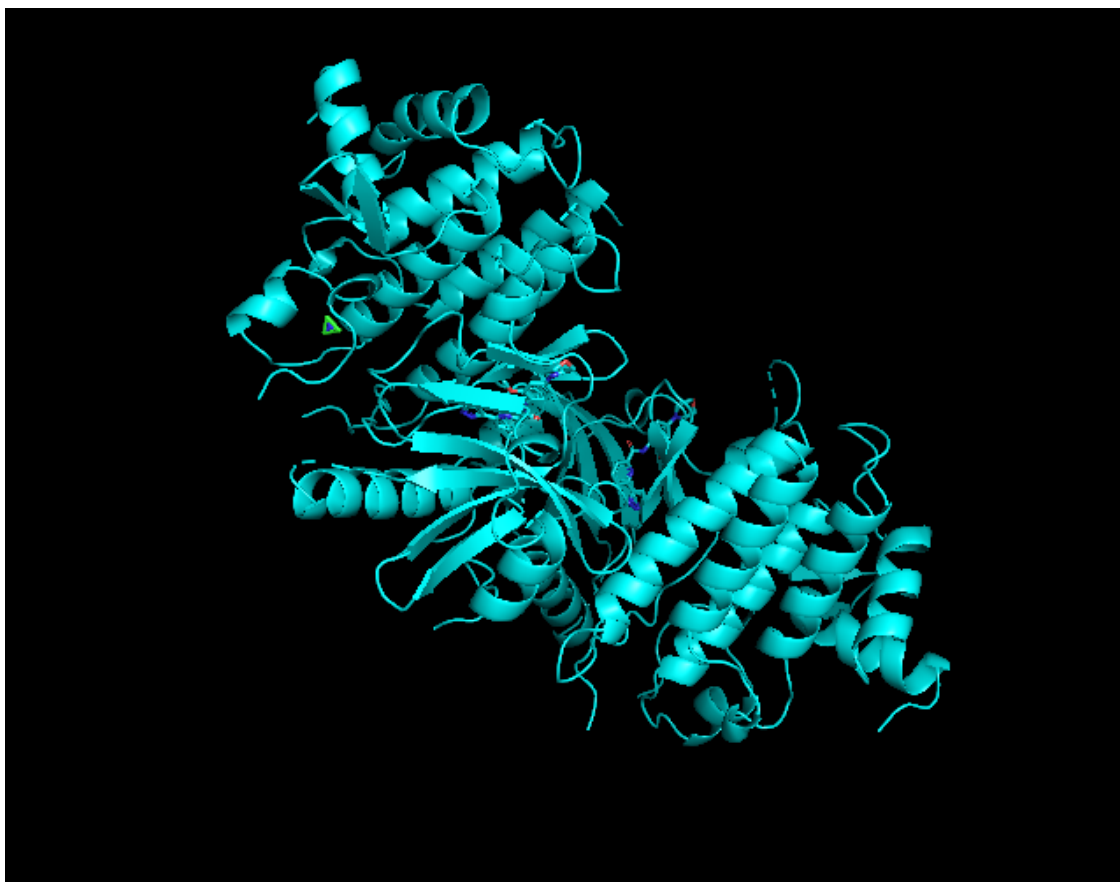


Figure 24: Docking with IRAK1

Figure 25为图 docking with CDKN1A 概览。

(对应文件为 Figure+Table/1146_into_6p8h.png)

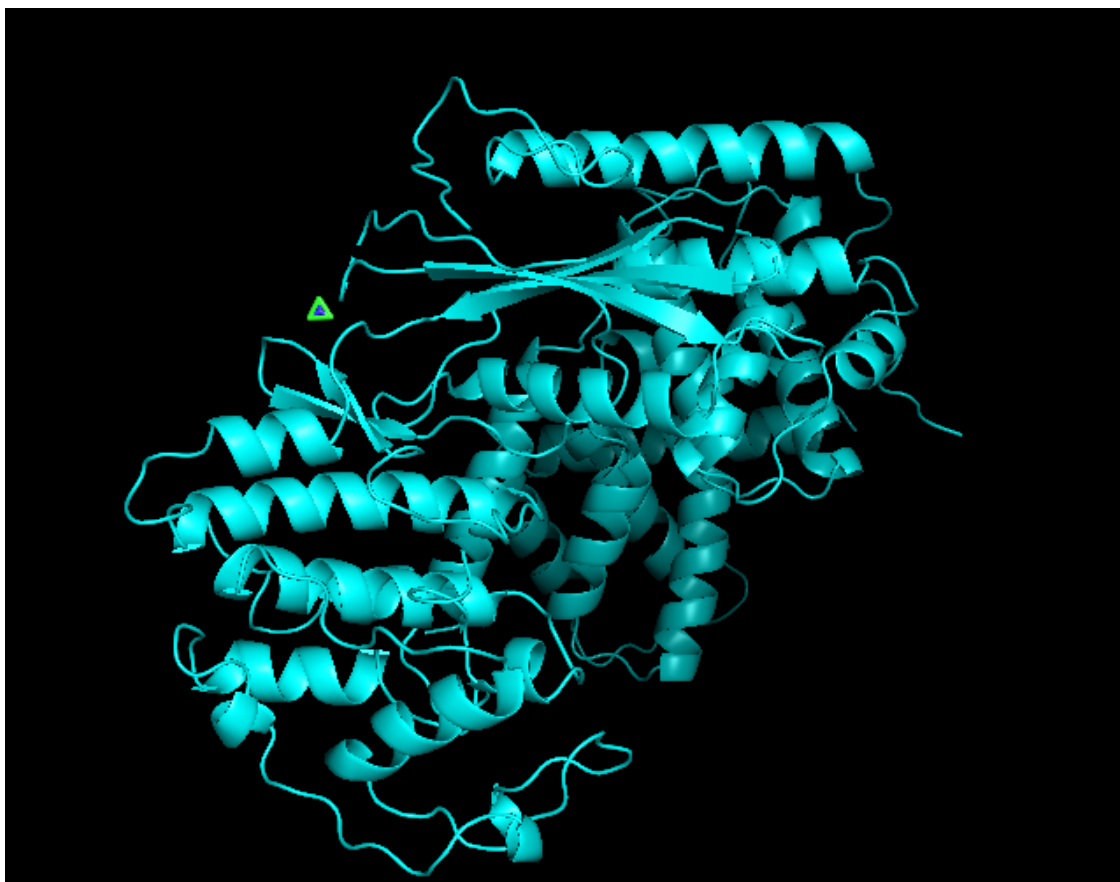


Figure 25: Docking with CDKN1A

对接的化合物的结构为：

Figure 26为图 compound structure 2D 概览。

(对应文件为 Figure+Table/1146_2D.pdf)

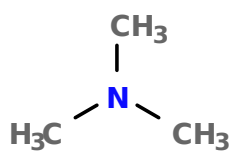


Figure 26: Compound structure 2D

Figure 27为图 compound structure 3D 概览。

(对应文件为 Figure+Table/1146.png)

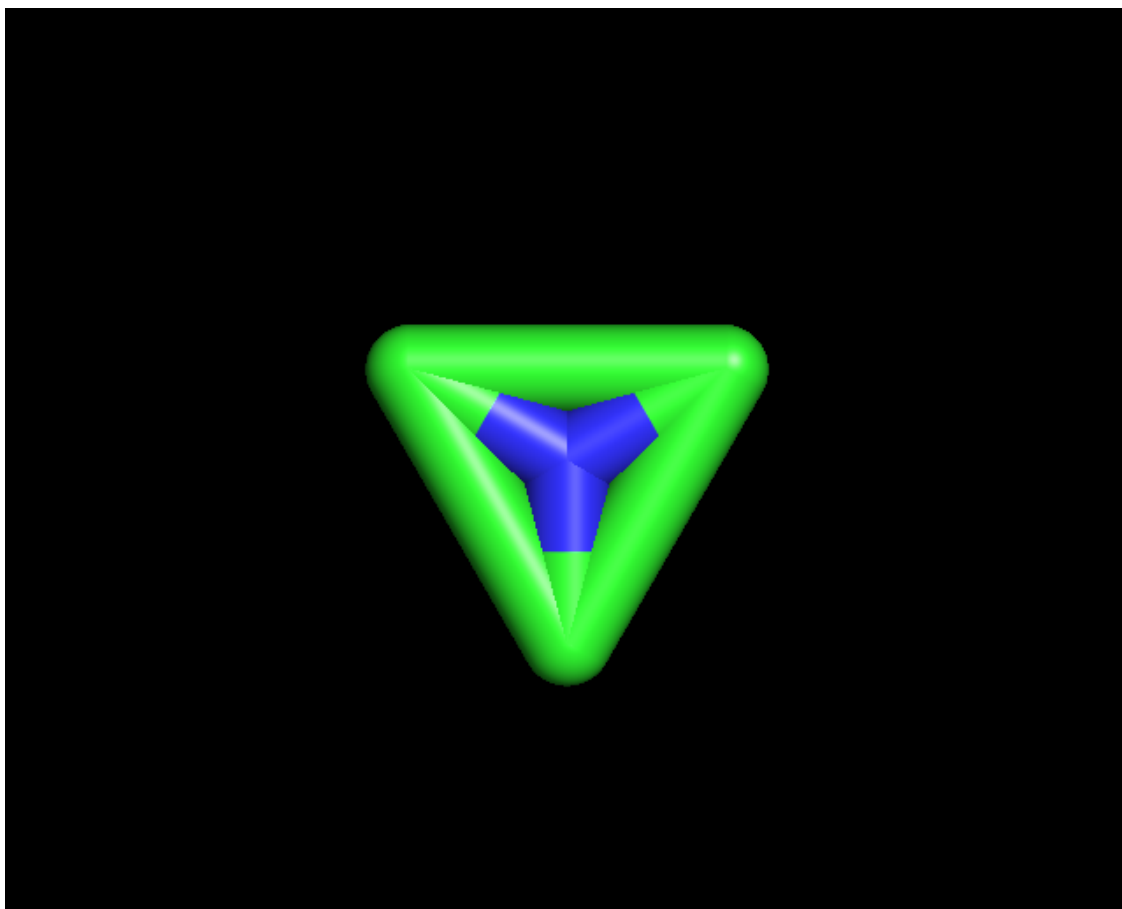


Figure 27: Compound structure 3D

4.7.3 靶点蛋白的表达量以及相关通路

Figure 28为图 pathway enrichment of receptors 概览。

(对应文件为 `Figure+Table/pathway-enrichment-of-receptors.pdf`)

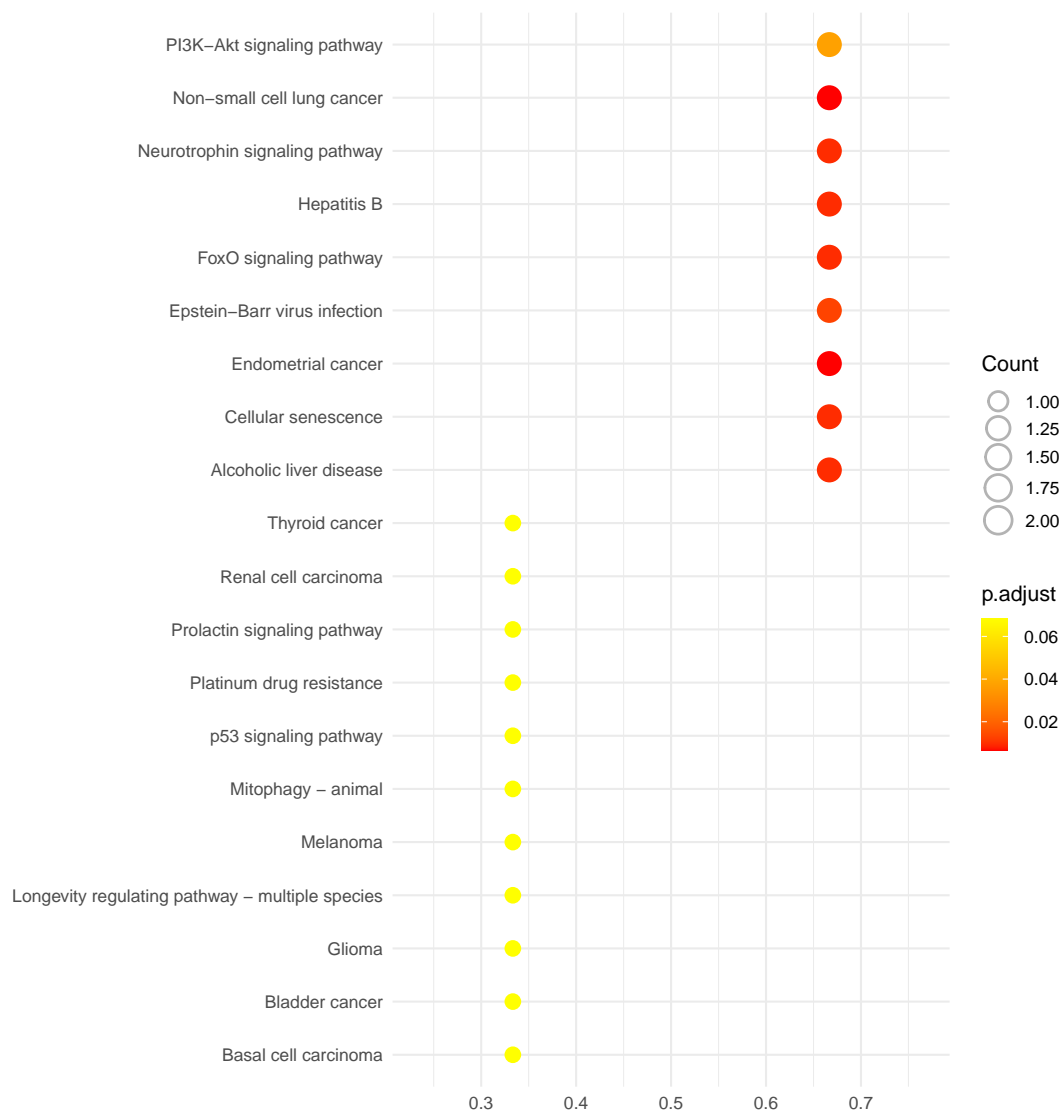


Figure 28: Pathway enrichment of receptors

4.8 复方活性成分

4.8.1 作用于靶点蛋白的主要活性成分

Table 15为表格 main active compounds to receptors 概览。

(对应文件为 **Figure+Table/main-active-compounds-to-receptors.xlsx**)

注：表格共有 16 行 8 列，以下预览的表格可能省略部分数据；表格含有 12 个唯一 ‘Ingredient.name’。

Table 15: Main active compounds to receptors

Herb_...	Ingre.....2	Ingre.....3	Targe.....4	Targe.....5	Datab...	Paper.id	PubMe...
DAN SHEN	aloe...	AC1Q6...	HBTAR...	CDKN1A	NA	NA	NA
FU LING	aloe...	AC1Q6...	HBTAR...	CDKN1A	NA	NA	NA
GAN CAO	beta-...	MolPo...	HBTAR...	CDKN1A	NA	NA	NA
GAN CAO	capsa...	capsa...	HBTAR...	CDKN1A	NA	NA	NA
NIU XI	d-man...	1630-...	HBTAR...	IRAK1	NA	NA	NA
CHEN PI	naringin	(2S)-...	HBTAR...	CDKN1A	NA	NA	NA
DA HUANG	naringin	(2S)-...	HBTAR...	CDKN1A	NA	NA	NA
HONG HUA	naringin	(2S)-...	HBTAR...	CDKN1A	NA	NA	NA
CHEN PI	nobil...	Spect...	HBTAR...	CDKN1A	NA	NA	NA
DA HUANG	nobil...	Spect...	HBTAR...	CDKN1A	NA	NA	NA
GAN CAO	pachy...	pachy...	HBTAR...	CDKN1A	NA	NA	NA
FU LING	salvi...	4CN-1...	HBTAR...	FOXO3	NA	HBREF...	31193821
FU LING	tansh...	I14-1...	HBTAR...	CDKN1A	NA	NA	NA
FU LING	tansh...	SCHEM...	HBTAR...	CDKN1A	NA	NA	NA
NIU XI	tetra...	NCI60...	HBTAR...	CDKN1A	NA	HBREF...	29605511
...

5 结论

见 1

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).
3. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using qiime 2. *Nature Biotechnology* **37**, 852–857 (2019).
4. McDonald, D. *et al.* The biological observation matrix (biom) format or: How i learned to stop worrying and love the ome-ome. *GigaScience* **1**, 7 (2012).
5. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from illumina amplicon data. *Nature methods* **13**, 581 (2016).
6. Hamday, M., Walker J., J., Harris, J. K., Gold J., N. & Knight, R. Error-correcting barcoded primers allow hundreds of samples to be pyrosequenced in multiplex. *Nature Methods* **5**, 235–237 (2008).

7. Hamday, M. & Knight, R. Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Research* **19**, 1141–1152 (2009).
8. Rai, S. N. *et al.* Microbiome data analysis with applications to pre-clinical studies using qiime2: Statistical considerations. *Genes & Diseases* **8**, (2021).
9. Wilmanski, T. *et al.* Blood metabolome predicts gut microbiome α -diversity in humans. *Nature Biotechnology* **37**, (2019).
10. Cheng, L., Qi, C., Zhuang, H., Fu, T. & Zhang, X. GutMDisorder: A comprehensive database for dysbiosis of the gut microbiota in disorders and interventions. *Nucleic Acids Research* **48**, (2019).
11. Vacca, M. *et al.* The controversial role of human gut lachnospiraceae. *Microorganisms* **8**, (2020).
12. Praveenraj, S. S. *et al.* The role of a gut microbial-derived metabolite, trimethylamine n-oxide (tmao), in neurological disorders. *Molecular neurobiology* **59**, 6684–6700 (2022).
13. 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).
14. Forli, S. *et al.* Computational proteinligand docking and virtual drug screening with the autodock suite. *Nature Protocols* **11**, (2016).
15. Seeliger, D. & Groot, B. L. de. Ligand docking and binding site analysis with pymol and autodock/vina. *Journal of Computer-Aided Molecular Design* **24**, (2010).