

肠道菌群宏基因组群落分析联合 RNA-seq

2024-01-29

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 宏基因组群落分析	2
6.1.1 数据质控	2
6.1.2 群落鉴定和丰度定量	2
6.1.3 群落分析	3
6.1.3.1 Alpha 多样性	3
6.1.3.2 Beta 多样性	3
6.1.3.3 差异分析	7
6.2 肠道菌群关联代谢物分析	9
6.3 代谢物关联蛋白质 (基因) 分析	11
6.4 蛋白质关联到 RNA-seq 的 DEG	12
6.4.1 Liver	12
6.4.2 Ileum	16
6.5 进一步验证代谢物的存在	20
6.5.1 Liver	21
6.5.1.1 富集分析	22
6.5.2 Ileum	23
6.5.2.1 富集分析	24
Reference	26

List of Figures

1 Alpha diversity	3
2 PCoA	4
3 Sample distance	5
4 Beta diversity group test	6
5 Species hierarchy	7
6 The abundance and LDA from Phylum to Class	9

7	MICRO alluvium plot of Matched data in gutMDisorder	10
8	Liver Intersection of Microbiota associated Genes with DEGs	13
9	Liver Top 1000 relationship network	15
10	Ileum Intersection of Microbiota associated Genes with DEGs	17
11	Ileum Top 1000 relationship network	19
12	LIVER ids KEGG enrichment	22
13	LIVER ids GO enrichment	23
14	ILEUM ids GO enrichment	25
15	ILEUM ids KEGG enrichment	25

List of Tables

1	Merged abundance table	2
2	Statistic of all difference microbiota	8
3	MICRO Matched data in gutMDisorder	10
4	MICRO Discover relationship between Microbiota with Host genes by matching metabolites .	11
5	Liver DEGs	12
6	Liver Microbiota associated Genes filtered by DEGs	14
7	Liver Top 1000 relationship data	15
8	Ileum DEGs	16
9	Ileum Microbiota associated Genes filtered by DEGs	18
10	Ileum Top 1000 relationship data	19
11	Unnamed chunk 38	20
12	Liver results filtered by validation	21
13	Ileum results filtered by validation	23

1 摘要

需求：以客户提供的数据 (RNA-seq + 肠道菌宏基因组数据)，筛选 DEGs、代谢物、肠道菌群关系链。

结果：

- 肠道菌群群落分析，以 MetaPhlAn 注释和定量肠道菌，MicrobiotaProcess 下游分析：
 - Alpha 和 Beta 分析均表明，对照组和模型组有显著差异 (6.1.3.1, 6.1.3.2)
 - 所有差异肠道菌见 6.1.3.3
- 建立肠道菌和代谢物联系，以 gutMDisorder 数据库：
 - 将上述差异肠道菌在数据库筛选相关代谢物 (6.2)
- 建立代谢物和蛋白质 (基因) 的联系，以发表的文献数据¹
 - 关联性筛选结果见 Tab. 4
- 建立上述筛选的蛋白质 (基因) 与此前的 RNA-seq 筛选的 DEGs 之间的关联：
 - (Liver 和 Ileum 的 DEGs 是此前已分析的，这里不再重新分析。见表格 Tab. 5 和 Tab. 8)
 - 筛选的基因与 Liver 和 Ileum DEGs 的交集见 Fig. 8 和 Tab. ??
 - 按代谢物与蛋白质 (基因) 关联强弱 (关联系数 rho)，将排名前 1000 的差异肠道菌、代谢物、DEGs 之间的关系呈现，Liver 和 Ileum 分别见 Fig. 9, Fig. 11
- 上述，肠道菌、差异基因 DEGs 均有数据支撑，而中间环节代谢物尚缺少验证；因此，这里选择已发表的胆结石研究² 中的数据以进一步验证：
 - 根据该文献报道的胆结石小鼠模型差异代谢物 (肝脏) 进行验证筛选，经过滤后的结果见 Tab. 12, Tab. 13
- 最后，试着对上一步的最终基因筛选结果做进一步富集分析：
 - 结果见 6.5.1.1, 6.5.2.1

2 前言

3 材料和方法

3.1 材料

Other data obtained from published article (e.g., supplementary tables):

- Supplementary file from article refer to².
- Supplementary file from article refer to¹.

3.2 方法

Mainly used method:

- The `biomart` was used for mapping genes between organism (e.g., `mgc_symbol` to `hgnc_symbol`)³.
- R package `ClusterProfiler` used for gene enrichment analysis⁴.
- `Fastp` used for Fastq data preprocessing⁵.
- Database `gutMDisorder` used for finding associations between gut microbiota and metabolites⁶.
- R package `Limma` and `edgeR` used for differential expression analysis^{7,8}.

- R package `MicrobiotaProcess` used for microbiome data visualization⁹.
- `MetaPhlAn` used for profiling the composition of microbial communities from metagenomic data¹⁰.
- 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 宏基因组群落分析

6.1.1 数据质控

‘Fastp QC’ 数据已全部提供。

(对应文件为 `./fastp_local/`)

注：文件夹 `./fastp_local/` 共包含 18 个文件。

1. `L1EGG121102-Chow1.html`
2. `L1EGG121103-Chow2.html`
3. `L1EGG121104_L1EGG121104-Chow3.html`
4. `L1EGG121105_L1EGG121105-Chow4.html`
5. `L1EGG121106_L1EGG121106-Chow5.html`
6. ...

6.1.2 群落鉴定和丰度定量

从这里开始，仅选择对照组和模型组进行分析。

Table 1 (下方表格) 为表格 Merged abundance table 概览。

(对应文件为 `Figure+Table/Merged-abundance-table.csv`)

注：表格共有 1590 行 13 列，以下预览的表格可能省略部分数据；表格含有 1590 个唯一 ‘clade_name’。

Table 1: Merged abundance table

clade_...	L1EGG1.....2	L1EGG1.....3	L1EGG1.....4	L1EGG1.....5	L1EGG1.....6	L1EGG1.....7	...
k__Bac...	99.99835	100	100	100	100	100	...
k__Arc...	0.00165	0	0	0	0	0	...
k__Bac...	43.04207	26.21962	35.11351	38.43636	41.23258	37.83933	...
k__Bac...	27.14994	0.70409	41.94514	22.4391	10.30688	24.7957	...

clade__...	L1EGG1.....2	L1EGG1.....3	L1EGG1.....4	L1EGG1.....5	L1EGG1.....6	L1EGG1.....7	...
k__Bac...	22.28786	1.84668	11.54645	3.45731	0.37113	2.64639	...
k__Bac...	3.94504	66.67222	7.0249	26.20566	19.89094	18.6702	...
k__Bac...	1.61135	4.04429	1.66165	7.25594	12.58176	14.00675	...
k__Bac...	1.21374	0	0.25707	2.02386	2.48308	1.89131	...
k__Bac...	0.4301	0.51236	2.40227	0.15918	13.13363	0.15032	...
k__Bac...	0.31824	0.00074	0.04902	0.0226	0	0	...
k__Arc...	0.00165	0	0	0	0	0	...
k__Bac...	20.71366	0.86067	7.24419	2.1026	0.13696	1.07217	...
k__Bac...	15.78352	0.3191	27.43063	16.53408	7.19353	19.10865	...
k__Bac...	11.64828	0.2736	10.04731	0.41265	0.67944	0.81364	...
k__Bac...	10.10471	15.68872	14.30816	17.52233	19.43315	18.17767	...
...

6.1.3 群落分析

6.1.3.1 Alpha 多样性

对照组和模型组有显著差异。

Figure 1 (下方图) 为图 Alpha diversity 概览。

(对应文件为 Figure+Table/Alpha-diversity.pdf)

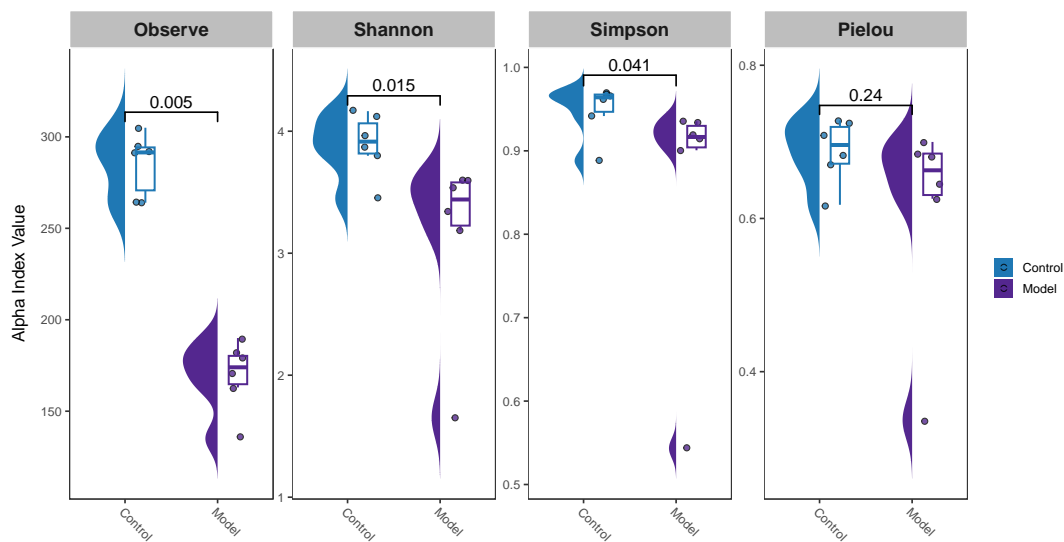


Figure 1: Alpha diversity

6.1.3.2 Beta 多样性

Figure 2 (下方图) 为图 PCoA 概览。

(对应文件为 Figure+Table/PCoA.pdf)

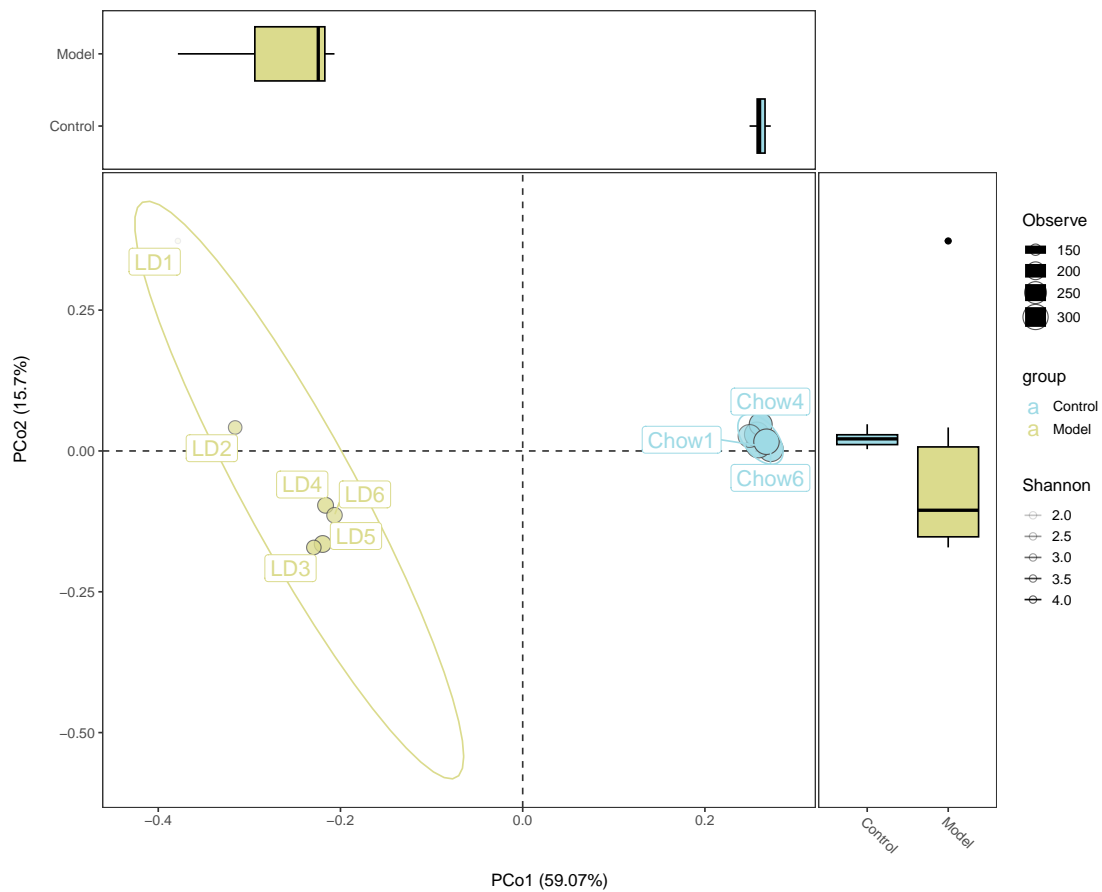


Figure 2: PCoA

Figure 3 (下方图) 为图 Sample distance 概览。

(对应文件为 Figure+Table/Sample-distance.pdf)

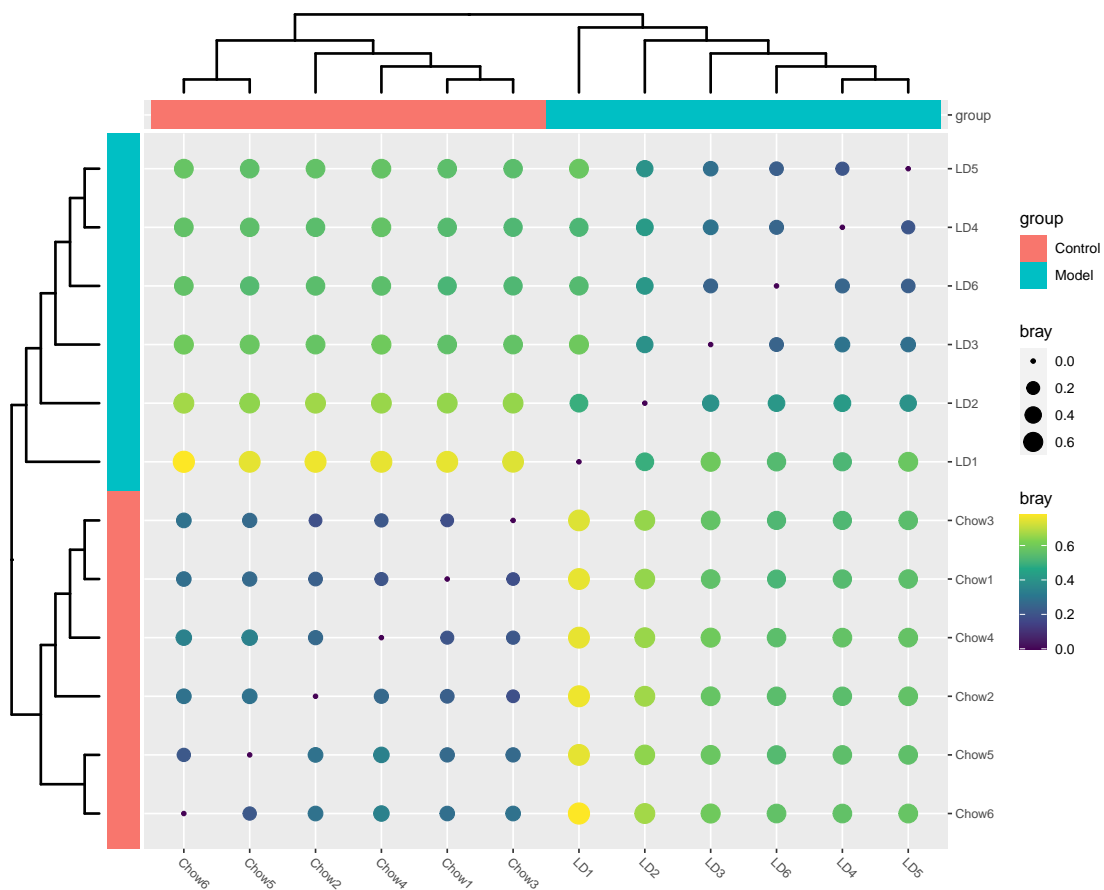


Figure 3: Sample distance

Figure 4 (下方图) 为图 Beta diversity group test 概览。

(对应文件为 [Figure+Table/Beta-diversity-group-test.pdf](#))

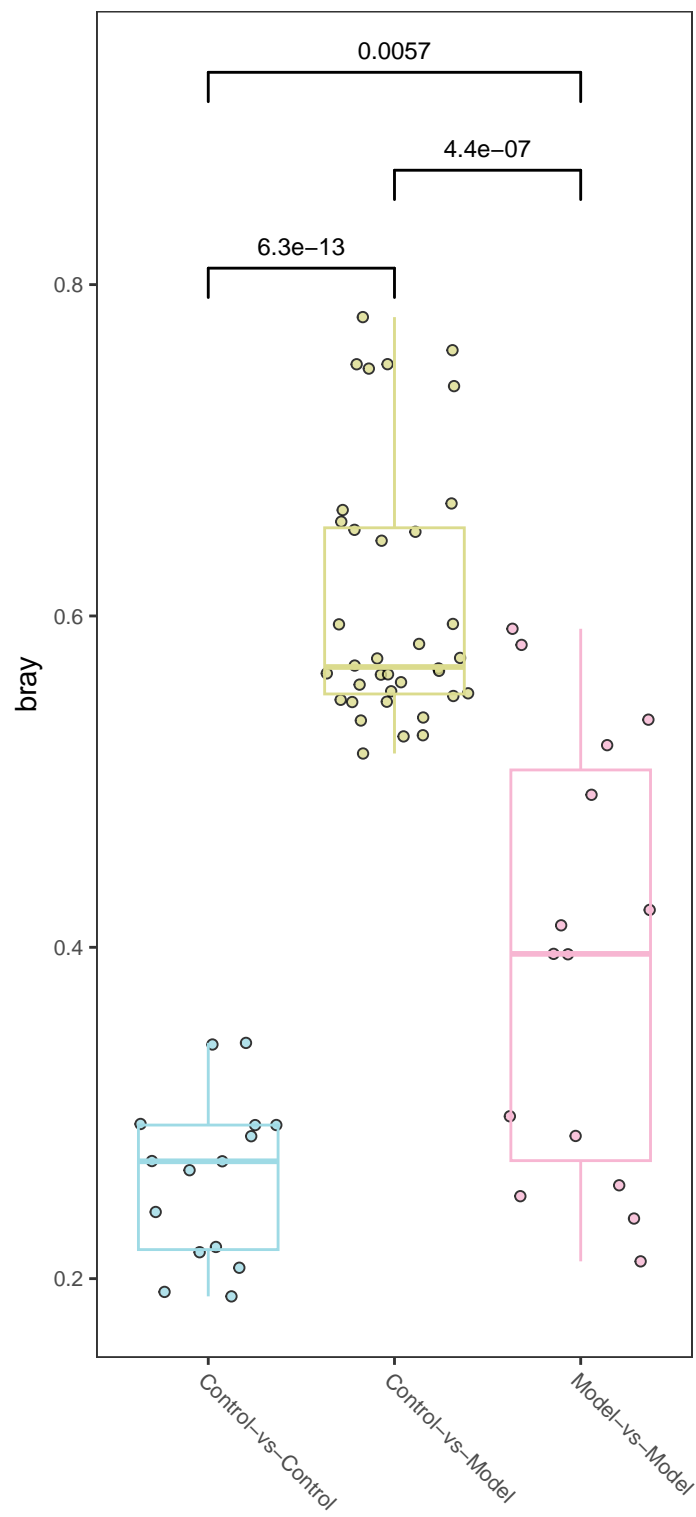


Figure 4: Beta diversity group test

‘All hierarchy data’ 数据已全部提供。

(对应文件为 Figure+Table/All-hierarchy-data)

注：文件夹 Figure+Table/All-hierarchy-data 共包含 6 个文件。

1. 1_Phylum.pdf
2. 2_Class.pdf
3. 3_Order.pdf
4. 4_Family.pdf
5. 5_Genus.pdf
6. ...

Figure 5 (下方图) 为图 Species hierarchy 概览。

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

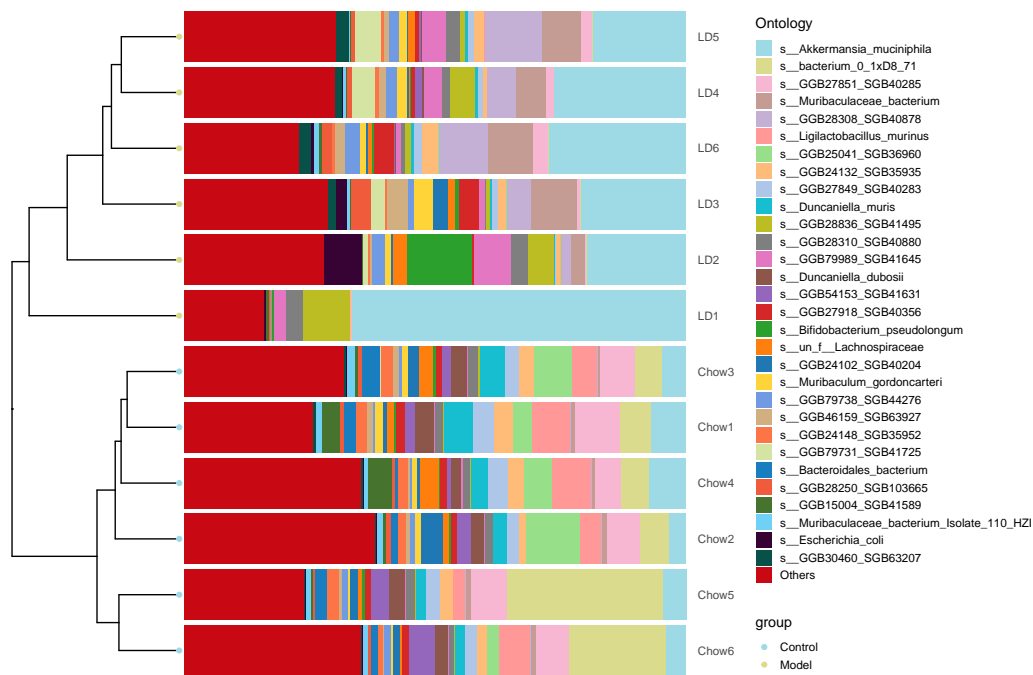


Figure 5: Species hierarchy

6.1.3.3 差异分析

Table 2 (下方表格) 为表格 Statistic of all difference microbiota 概览。

(对应文件为 Figure+Table/Statistic-of-all-difference-microbiota.csv)

注：表格共有 1693 行 8 列，以下预览的表格可能省略部分数据；表格含有 1693 个唯一 'label'。

1. pvalue: 显著性 P。

Table 2: Statistic of all difference microbiota

label	nodeClass	pvalue	fdr	LDAupper	LDAmean	LDAlower	Sign_g...
t__SGB...	OTU	0.0020...	0.0160...	3.7811...	3.7217...	3.6527...	Control
t__SGB...	OTU	0.0020...	0.0160...	3.3656...	3.3470...	3.3275...	Control
t__SGB...	OTU	0.0020...	0.0160...	2.0906...	2.0529...	2.0118...	Control
t__SGB...	OTU	0.0020...	0.0160...	2.3222...	2.2279...	2.1074...	Control
t__SGB...	OTU	0.0020...	0.0160...	2.6348...	2.6044...	2.5717...	Control
t__SGB...	OTU	0.0020...	0.0160...	2.7349...	2.7128...	2.6896...	Control
t__SGB...	OTU	0.0020...	0.0160...	3.2904...	3.2657...	3.2395...	Control
t__SGB...	OTU	0.0020...	0.0160...	2.4401...	2.4141...	2.3864...	Control
t__SGB...	OTU	0.0020...	0.0160...	2.5684...	2.5329...	2.4943...	Control
t__SGB...	OTU	0.0020...	0.0160...	2.1686...	2.1377...	2.1045...	Control
t__SGB...	OTU	0.0020...	0.0160...	3.7672...	3.7140...	3.6534...	Control
t__SGB...	OTU	0.0020...	0.0160...	2.6050...	2.5764...	2.5458...	Control
t__SGB...	OTU	0.0020...	0.0160...	2.4451...	2.4190...	2.3911...	Control
t__SGB...	OTU	0.0020...	0.0160...	3.4353...	3.3707...	3.2948...	Control
t__SGB...	OTU	0.0020...	0.0160...	NA	NA	NA	NA
...

Figure 6 (下方图) 为图 The abundance and LDA from Phylum to Class 概览。

(对应文件为 Figure+Table/The-abundance-and-LDA-from-Phylum-to-Class.pdf)

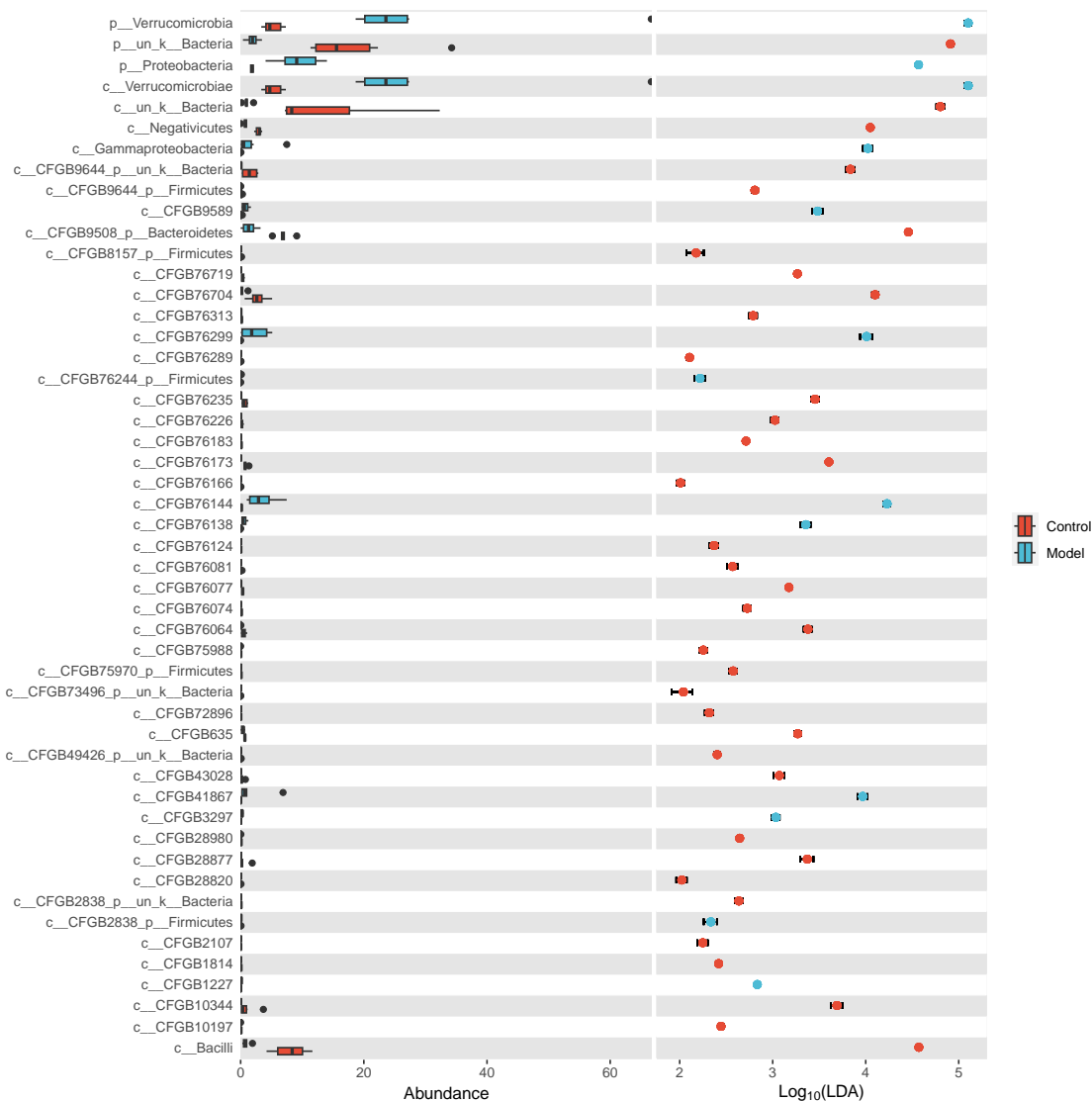


Figure 6: The abundance and LDA from Phylum to Class

6.2 肠道菌群关联代谢物分析

Figure 7 (下方图) 为图 MICRO alluvium plot of Matched data in gutMDisorder 概览。

(对应文件为 [Figure+Table/MICRO-alluvium-plot-of-Matched-data-in-gutMDisorder.pdf](#))

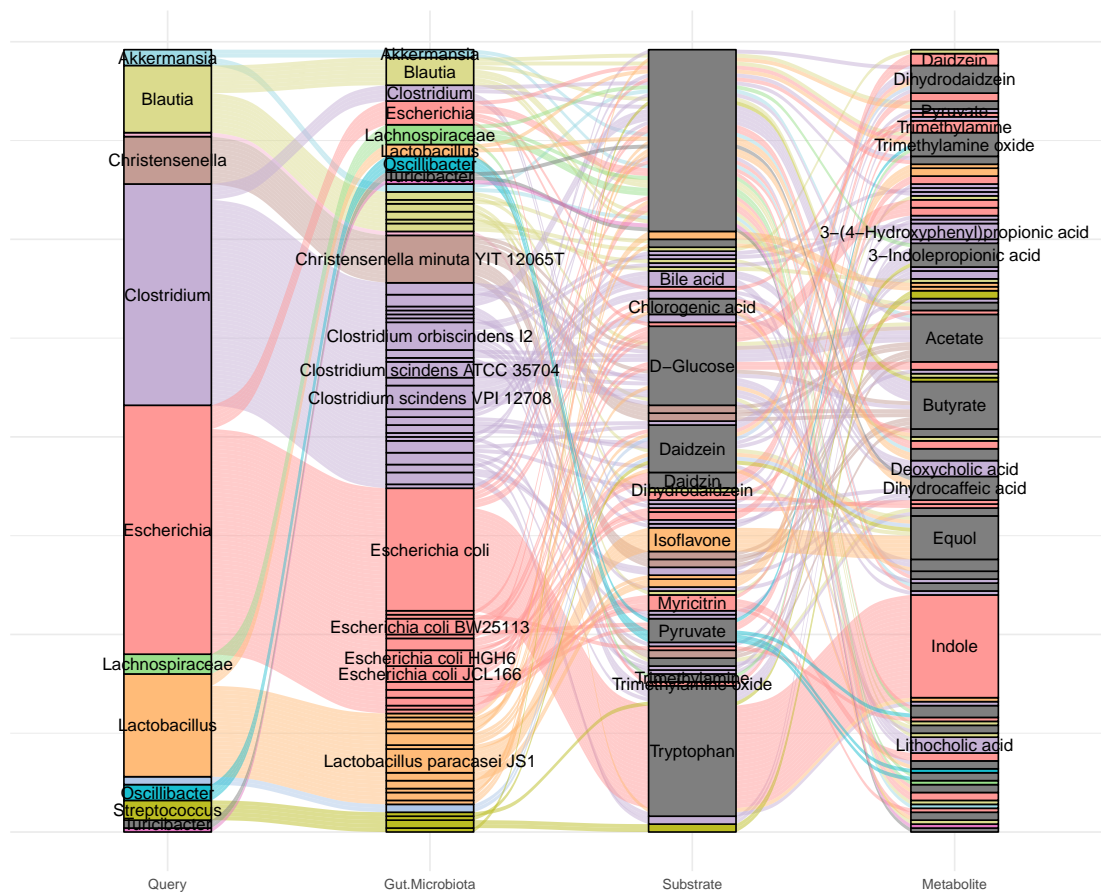


Figure 7: MICRO alluvium plot of Matched data in gutMDisorder

Table 3 (下方表格) 为表格 MICRO Matched data in gutMDisorder 概览。

(对应文件为 **Figure+Table/MICRO-Matched-data-in-gutMDisorder.xlsx**)

注：表格共有 198 行 13 列，以下预览的表格可能省略部分数据；表格含有 13 个唯一‘Query’。

Table 3: MICRO Matched data in gutMDisorder

Query	Gut.Mi.....2	Gut.Mi.....3	Gut.Mi.....4	Classi...	Substrate	Substr.....7	Substr.....8	...
Christ...	Christ...	NA	gm0883	strain	D-Glucose	5793	HMDB00...	...
Christ...	Christ...	NA	gm0883	strain	Salicin	439503	HMDB00...	...
Christ...	Christ...	NA	gm0883	strain	D-Xylose	135191	HMDB00...	...
Christ...	Christ...	NA	gm0883	strain	L-Arab...	439195	HMDB00...	...
Christ...	Christ...	NA	gm0883	strain	L-Rham...	25310	HMDB00...	...
Christ...	Christ...	NA	gm0883	strain	D-Mannose	18950	HMDB00...	...
Christ...	Christ...	NA	gm0883	strain	D-Glucose	5793	HMDB00...	...
Christ...	Christ...	NA	gm0883	strain	Salicin	439503	HMDB00...	...
Christ...	Christ...	NA	gm0883	strain	D-Xylose	135191	HMDB00...	...

Query	Gut.Mi.....2	Gut.Mi.....3	Gut.Mi.....4	Classi...	Substrate	Substr.....7	Substr.....8	...
Christ...	Christ...	NA	gm0883	strain	L-Arab...	439195	HMDB00...	...
Christ...	Christ...	NA	gm0883	strain	L-Rham...	25310	HMDB00...	...
Christ...	Christ...	NA	gm0883	strain	D-Mannose	18950	HMDB00...	...
Clostr...	Clostr...	29347	gm0885	strain	Bile acid	439520		...
Clostr...	Clostr...	29347	gm0885	strain	Cholic...	221493	HMDB00...	...
Clostr...	Clostr...	29347	gm0885	strain	Chenod...	10133	HMDB00...	...
...

6.3 代谢物关联蛋白质 (基因) 分析

<https://github.com/aeisman/protein-metabolite> <https://mbenson.shinyapps.io/protein-metabolite/>¹

Table 4 (下方表格) 为表格 MICRO Discover relationship between Microbiota with Host genes by matching metabolites 概览。

(对应文件为 **Figure+Table/MICRO-Discover-relationship-between-Microbiota-with-Host-genes-by-matching-metabolites**)

注：表格共有 52210 行 10 列，以下预览的表格可能省略部分数据；表格含有 22 个唯一 ‘id’。

1. META_Rho: 关联分析结果的关联系数，绝对值越大，说明关联性越强 (源自文献的分析)
2. META_Q: 关联分析结果 P 的校正值 (源自文献的分析)
3. META_P: 关联分析结果 P 的值 (源自文献的分析)

Table 4: MICRO Discover relationship between Microbiota with Host genes by matching metabolites

.id	.id_from	Substrate	Metabo.....4	Gut.Mi...	Target...	Metabo.....7	META_Rho	META_Q	META_P...
586	Metabo...	Creatine	Akkerm...	NEGR1	creatine	-0.217...	9.9886...	2.5777...	...
586	Metabo...	Creatine	Akkerm...	RGMA	creatine	-0.204...	3.3272...	6.2440...	...
586	Metabo...	Creatine	Akkerm...	RGMB	creatine	-0.200...	1.6028...	9.8838...	...
586	Metabo...	Creatine	Akkerm...	CD55	creatine	-0.197...	4.2927...	2.1866...	...
586	Metabo...	Creatine	Akkerm...	MB	creatine	-0.196...	1.1517...	3.3965...	...
586	Metabo...	Creatine	Akkerm...	RELT	creatine	-0.194...	1.3574...	1.7104...	...
586	Metabo...	Creatine	Akkerm...	IGFBP2	creatine	-0.193...	5.9415...	6.8494...	...
586	Metabo...	Creatine	Akkerm...	IGFBP6	creatine	-0.187...	5.2454...	6.1876...	...
586	Metabo...	Creatine	Akkerm...	CD59	creatine	-0.187...	7.7980...	8.7806...	...
586	Metabo...	Creatine	Akkerm...	NPPB	creatine	-0.189...	1.7690...	9.4854...	...
586	Metabo...	Creatine	Akkerm...	CDNF	creatine	-0.181...	1.3228...	7.0930...	...
586	Metabo...	Creatine	Akkerm...	EFNB2	creatine	-0.180...	1.8860...	1.4601...	...
586	Metabo...	Creatine	Akkerm...	UNC5D	creatine	-0.172...	3.7962...	3.8674...	...

.id	.id_from	Substrate	Metabo.....4	Gut.Mi...	Target...	Metabo.....7	META_RM	META_Q	META_P...
586	Metabo...		Creatine	Akkerm...	CST3	creatine	-0.169...	1.5765...	3.3814... ..
586	Metabo...		Creatine	Akkerm...	FSTL3	creatine	-0.169...	1.6812...	2.9297... ..
...

6.4 蛋白质关联到 RNA-seq 的 DEG

6.4.1 Liver

Table 5 (下方表格) 为表格 Liver DEGs 概览。

(对应文件为 **Figure+Table/Liver-DEGs.xlsx**)

注：表格共有 2998 行 11 列，以下预览的表格可能省略部分数据；表格含有 2998 个唯一 ‘hgnc_symbol’。

1. hgnc_symbol: 基因名 (Human)
2. mgi_symbol: 基因名 (Mice)
3. logFC: estimate of the log2-fold-change corresponding to the effect or contrast (for ‘topTableF’ there may be several columns of log-fold-changes)
4. AveExpr: average log2-expression for the probe over all arrays and channels, same as ‘Amean’ in the ‘MarrayLM’ object
5. t: moderated t-statistic (omitted for ‘topTableF’)
6. P.Value: raw p-value
7. B: log-odds that the gene is differentially expressed (omitted for ‘topTreat’)

Table 5: Liver DEGs

hgnc_s...	mgi_sy...	ensembl...	entrez...	descri...	logFC	AveExpr	t	P.Value	adj.P.Val
ENHO	Enho	ENSMUS...	69638	energy...	-4.453...	2.2957...	-17.04...	2.0685...	0.0002...
CES2	Ces2a	ENSMUS...	102022	carbox...	1.6614...	8.8361...	15.281...	5.6258...	0.0004...
HSD17B6	Hsd17b6	ENSMUS...	27400	hydrox...	2.8011...	8.6106...	14.201...	1.0950...	0.0006...
FMO5	Fmo5	ENSMUS...	14263	flavin...	1.2618...	8.1280...	13.790...	1.4285...	0.0007...
ABCB11	Abcb11	ENSMUS...	27413	ATP-bi...	1.2515...	7.7893...	11.121...	9.7706...	0.0033...
GNAT1	Gnat1	ENSMUS...	14685	G prot...	-2.232...	2.9799...	-10.64...	1.4365...	0.0044...
NNMT	Nnmt	ENSMUS...	18113	nicoti...	-3.804...	5.1567...	-9.928...	2.6415...	0.0065...
CSAD	Csad	ENSMUS...	246277	cystei...	-1.560...	7.2232...	-9.535...	3.7482...	0.0083...
ABCB1	Abcb1a	ENSMUS...	18671	ATP-bi...	3.2131...	3.2740...	9.4563...	4.0267...	0.0084...
FGFR2	Fgfr2	ENSMUS...	14183	fibrob...	2.9129...	4.2716...	9.2494...	4.8706...	0.0087...
DDAH1	Ddah1	ENSMUS...	69219	dimeth...	1.3322...	6.8518...	9.1694...	5.2476...	0.0087...
ABCG5	Abcg5	ENSMUS...	27409	ATP bi...	1.5044...	7.3228...	9.0073...	6.1127...	0.0089...
SLC1A2	Slc1a2	ENSMUS...	20511	solute...	-1.900...	5.0951...	-8.951...	6.4435...	0.0089...

hgnc_s...	mgc_sy...	ensembl...	entrez...	descri...	logFC	AveExpr	t	P.Value	adj.P.Val
TTC39C	Ttc39c	ENSMUS...	72747	tetrat...	-1.946...	6.2930...	-8.431...	1.0707...	0.0106...
WNK4	Wnk4	ENSMUS...	69847	WNK ly...	2.3302...	2.6719...	8.3971...	1.1085...	0.0106...
...

Figure 8 (下方图) 为图 Liver Intersection of Microbiota associated Genes with DEGs 概览。

(对应文件为 [Figure+Table/Liver-Intersection-of-Microbiota-associated-Genes-with-DEGs.pdf](#))

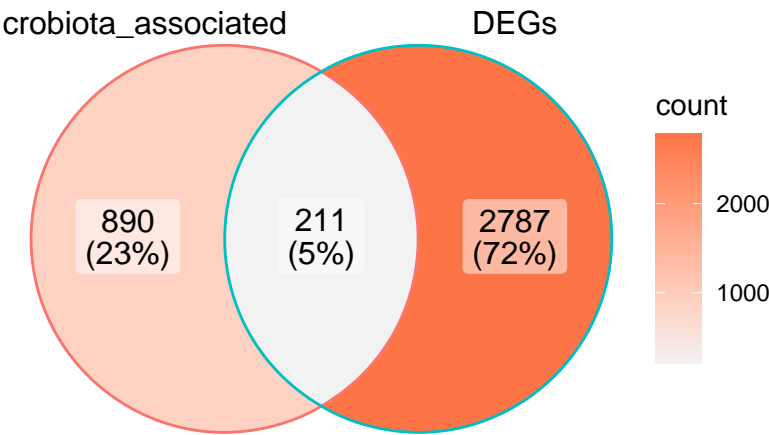


Figure 8: Liver Intersection of Microbiota associated Genes with DEGs

Intersection :

CD59, FGFR1, TGFBR3, ACY1, GHR, CXCL12, CD93, POSTN, BMP1, SMOC1, SERPINF2, OMD, APOL1, DCTPP1, APOA1, MAP2K2, CCL23, INHBA, CFB, CAPG, SPP1, GFRA1, SERPINA3, IL15RA, CADM1, IL1R1, ALCAM, CYP3A4, GDI2, IL18BP, GPT, IGFBP1, F11, COL18A1, SERPINA1, HNRNPAB, AFM, IGFBP5, RBP4, TKT, DDR2, APOM, KYNU,...

(上述信息框内容已保存至 [Figure+Table/Liver-Intersection-of-Microbiota-associated-Genes-with-DEGs-content](#))

Table 6 (下方表格) 为表格 Liver Microbiota associated Genes filtered by DEGs 概览。

(对应文件为 [Figure+Table/Liver-Microbiota-associated-Genes-filtered-by-DEGs.xlsx](#))

注：表格共有 10454 行 10 列，以下预览的表格可能省略部分数据；表格含有 22 个唯一 'id'。

1. META_Rho: 关联分析结果的关联系数，绝对值越大，说明关联性越强 (源自文献的分析)
2. META_Q: 关联分析结果 P 的校正值 (源自文献的分析)
3. META_P: 关联分析结果 P 的值 (源自文献的分析)

Table 6: Liver Microbiota associated Genes filtered by DEGs

.id	.id_from	Substrate	Metabo.....4	Gut.Mi...	Target...	Metabo.....7	META_RM	META_Q	META_P ...
588	Metabo...		2-Imin...	Clostr...	CD59	creati...	0.4459...	6.9021...	1.8504... ..
750	Substrate	Glycine	Acetyl...	Clostr...	GHR	glycine	-0.425...	4.5068...	2.4165... ..
750	Metabo...		Glycine	Blautia	GHR	glycine	-0.425...	4.5068...	2.4165... ..
750	Metabo...		Glycine	Lactob...	GHR	glycine	-0.425...	4.5068...	2.4165... ..
5793	Substrate	D-	Acetate	Christ...	PLXNB2	glucose	0.3849...	1.3255...	3.5538... ..
		Glucose							
5793	Substrate	D-	Butyrate	Christ...	PLXNB2	glucose	0.3849...	1.3255...	3.5538... ..
		Glucose							
5793	Substrate	D-	2,3-Bu...	Escher...	PLXNB2	glucose	0.3849...	1.3255...	3.5538... ..
		Glucose							
5793	Substrate	D-	Acetoin	Escher...	PLXNB2	glucose	0.3849...	1.3255...	3.5538... ..
		Glucose							
5793	Substrate	D-	2,3-Bu...	Escher...	PLXNB2	glucose	0.3849...	1.3255...	3.5538... ..
		Glucose							
5793	Substrate	D-	2,3-Bu...	Escher...	PLXNB2	glucose	0.3849...	1.3255...	3.5538... ..
		Glucose							
5793	Substrate	D-	Acetoin	Escher...	PLXNB2	glucose	0.3849...	1.3255...	3.5538... ..
		Glucose							
5793	Substrate	D-	2,3-Bu...	Escher...	PLXNB2	glucose	0.3849...	1.3255...	3.5538... ..
		Glucose							
5793	Substrate	D-	Ethanol	Lactob...	PLXNB2	glucose	0.3849...	1.3255...	3.5538... ..
		Glucose							
5793	Substrate	D-	Acetate	Clostr...	PLXNB2	glucose	0.3849...	1.3255...	3.5538... ..
		Glucose							
5793	Substrate	D-	Butyrate	Clostr...	PLXNB2	glucose	0.3849...	1.3255...	3.5538... ..
		Glucose							
...

Figure 9 (下方图) 为图 Liver Top 1000 relationship network 概览。

(对应文件为 **Figure+Table/Liver-Top-1000-relationship-network.pdf**)

Gut.Microbiota	Metabolite	Target_Gene
Escherichia coli JCL166	2,3-Butanediol	PLXNB2
Escherichia coli JCL166	Acetoin	PLXNB2
Escherichia coli JCL166	2,3-Butanedione	PLXNB2
Lactobacillus fermentum	Ethanol	PLXNB2
Clostridium pasteurianum	Acetate	PLXNB2
Clostridium pasteurianum	Butyrate	PLXNB2
...

6.4.2 Ileum

Table 8 (下方表格) 为表格 Ileum DEGs 概览。

(对应文件为 **Figure+Table/Ileum-DEGs.xlsx**)

注：表格共有 2554 行 11 列，以下预览的表格可能省略部分数据；表格含有 2554 个唯一 ‘hgnc_symbol’。

1. hgnc_symbol: 基因名 (Human)
2. mgi_symbol: 基因名 (Mice)
3. logFC: estimate of the log2-fold-change corresponding to the effect or contrast (for ‘topTableF’ there may be several columns of log-fold-changes)
4. AveExpr: average log2-expression for the probe over all arrays and channels, same as ‘Amean’ in the ‘MarrayLM’ object
5. t: moderated t-statistic (omitted for ‘topTableF’)
6. P.Value: raw p-value
7. B: log-odds that the gene is differentially expressed (omitted for ‘topTreat’)

Table 8: Ileum DEGs

hgnc_s...	mgi_sy...	ensem...	entrez...	descri...	logFC	AveExpr	t	P.Value	adj.P.Val
HMGCS1	Hmgcs1	ENSMUS...	208715	3-hydr...	-1.476...	7.1006...	-10.01...	3.4647...	0.0446...
LYPD8	Lypd8	ENSMUS...	70163	LY6/PL...	-1.422...	6.7523...	-9.413...	5.7988...	0.0446...
ACAA1	Acaa1b	ENSMUS...	235674	acetyl...	3.2729...	4.4478...	9.3290...	6.2488...	0.0446...
FDFT1	Fdft1	ENSMUS...	14137	farnes...	-1.885...	4.1659...	-9.280...	6.5253...	0.0446...
COL1A1	Col1a1	ENSMUS...	12842	collag...	0.9944...	6.5221...	9.2532...	6.6835...	0.0446...
SQLE	Sqle	ENSMUS...	20775	squale...	-1.933...	5.8928...	-9.198...	7.0178...	0.0446...
TCF23	Tcf23	ENSMUS...	69852	transc...	1.8100...	3.0507...	9.0283...	8.1816...	0.0446...
CCL23	Ccl6	ENSMUS...	20305	chemok...	-1.331...	7.5855...	-8.558...	1.2650...	0.0511...
CCL15	Ccl6	ENSMUS...	20305	chemok...	-1.331...	7.5855...	-8.558...	1.2650...	0.0511...
CCL15-...	Ccl6	ENSMUS...	20305	chemok...	-1.331...	7.5855...	-8.558...	1.2650...	0.0511...

hgnc_s...	mgi_sy...	ensem...	entrez...	descri...	logFC	AveExpr	t	P.Value	adj.P.Val
PIGR	Pigr	ENSMUS...	18703	polyme...	-1.150...	10.832...	-8.463...	1.3856...	0.0511...
GZMA	Gzma	ENSMUS...	14938	granz...	-3.299...	4.3784...	-8.447...	1.4064...	0.0511...
CEACAM21	Ceacam10	ENSMUS...	26366	CEA ce...	-3.021...	2.5056...	-7.995...	2.1919...	0.0638...
MSMO1	Msmo1	ENSMUS...	66234	methy...	-1.455...	5.8017...	-7.623...	3.2019...	0.0777...
INSIG1	Insig1	ENSMUS...	231070	insuli...	-1.306...	5.1683...	-7.562...	3.4134...	0.0784...
...

Figure 10 (下方图) 为图 Ileum Intersection of Microbiota associated Genes with DEGs 概览。

(对应文件为 [Figure+Table/Ileum-Intersection-of-Microbiota-associated-Genes-with-DEGs.pdf](#))

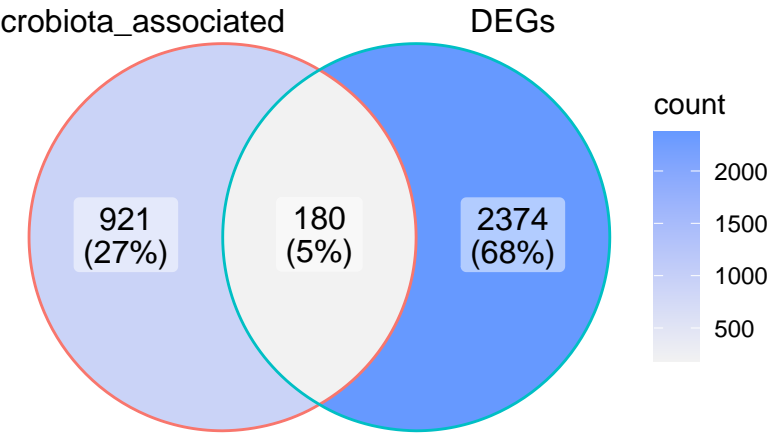


Figure 10: Ileum Intersection of Microbiota associated Genes with DEGs

Intersection :

RGMB, GSN, POSTN, TNFRSF21, BMP1, WFIKKN2, B2M, SMOC1, SLITRK5, JAM2, DSC2, PDE11A, CST6, APOL1, CCDC80, TNFRSF19, CCL23, NOG, IL15RA, NRXN1, ALCAM, MDH1, GDI2, LCN2, IL18BP, GZMA, SLPI, PTN, GPT, LGALS3BP, COL18A1, SERPINA1, PDE2A, SOD3, CD109, EFNA3, CXCL10, SGTA, RET, POR, GAPDH, CHL1, RAC1, A...

(上述信息框内容已保存至 [Figure+Table/Ileum-Intersection-of-Microbiota-associated-Genes-with-DEGs-content](#))

Table 9 (下方表格) 为表格 Ileum Microbiota associated Genes filtered by DEGs 概览。

(对应文件为 [Figure+Table/Ileum-Microbiota-associated-Genes-filtered-by-DEGs.xlsx](#))

注：表格共有 9208 行 10 列，以下预览的表格可能省略部分数据；表格含有 22 个唯一 ‘id’。

1. META_Rho: 关联分析结果的关联系数，绝对值越大，说明关联性越强 (源自文献的分析)
2. META_Q: 关联分析结果 P 的校正值 (源自文献的分析)
3. META_P: 关联分析结果 P 的值 (源自文献的分析)

Table 9: Ileum Microbiota associated Genes filtered by DEGs

.id	.id_from	Substrate	Metabo.....4	Gut.Mi...	Target...	Metabo.....7	META_RM	META_Q	META_P ...
588	Metabo...		2-Imin...	Clostr...	B2M	creati...	0.5130...	0	...
588	Metabo...		2-Imin...	Clostr...	DSC2	creati...	0.5128...	0	...
588	Metabo...		2-Imin...	Clostr...	RGMB	creati...	0.4166...	5.3138...	1.4246... ..
750	Substrate	Glycine	Acetyl...	Clostr...	RET	glycine	-0.407...	9.5712...	5.1320... ..
750	Metabo...		Glycine	Blautia	RET	glycine	-0.407...	9.5712...	5.1320... ..
750	Metabo...		Glycine	Lactob...	RET	glycine	-0.407...	9.5712...	5.1320... ..
588	Metabo...		2-Imin...	Clostr...	JAM2	creati...	0.4070...	2.8618...	7.6726... ..
588	Metabo...		2-Imin...	Clostr...	CST6	creati...	0.3307...	2.1455...	5.7522... ..
588	Metabo...		2-Imin...	Clostr...	SPOCK2	creati...	-0.321...	1.3977...	1.1241... ..
588	Metabo...		2-Imin...	Clostr...	LCN2	creati...	0.3110...	2.0900...	1.1206... ..
588	Metabo...		2-Imin...	Clostr...	TNFRSF24	creati...	0.3098...	7.2094...	7.7313... ..
588	Metabo...		2-Imin...	Clostr...	SMOC1	creati...	0.3071...	2.8839...	7.7317... ..
588	Metabo...		2-Imin...	Clostr...	TNFRSF10	creati...	0.2890...	4.7330...	1.2689... ..
588	Metabo...		2-Imin...	Clostr...	COL18A1	creati...	0.2883...	3.0855...	3.3089... ..
750	Substrate	Glycine	Acetyl...	Clostr...	SLITRK5	glycine	0.2801...	1.7545...	4.7038... ..
...

Figure 11 (下方图) 为图 Ileum Top 1000 relationship network 概览。

(对应文件为 **Figure+Table/Ileum-Top-1000-relationship-network.pdf**)

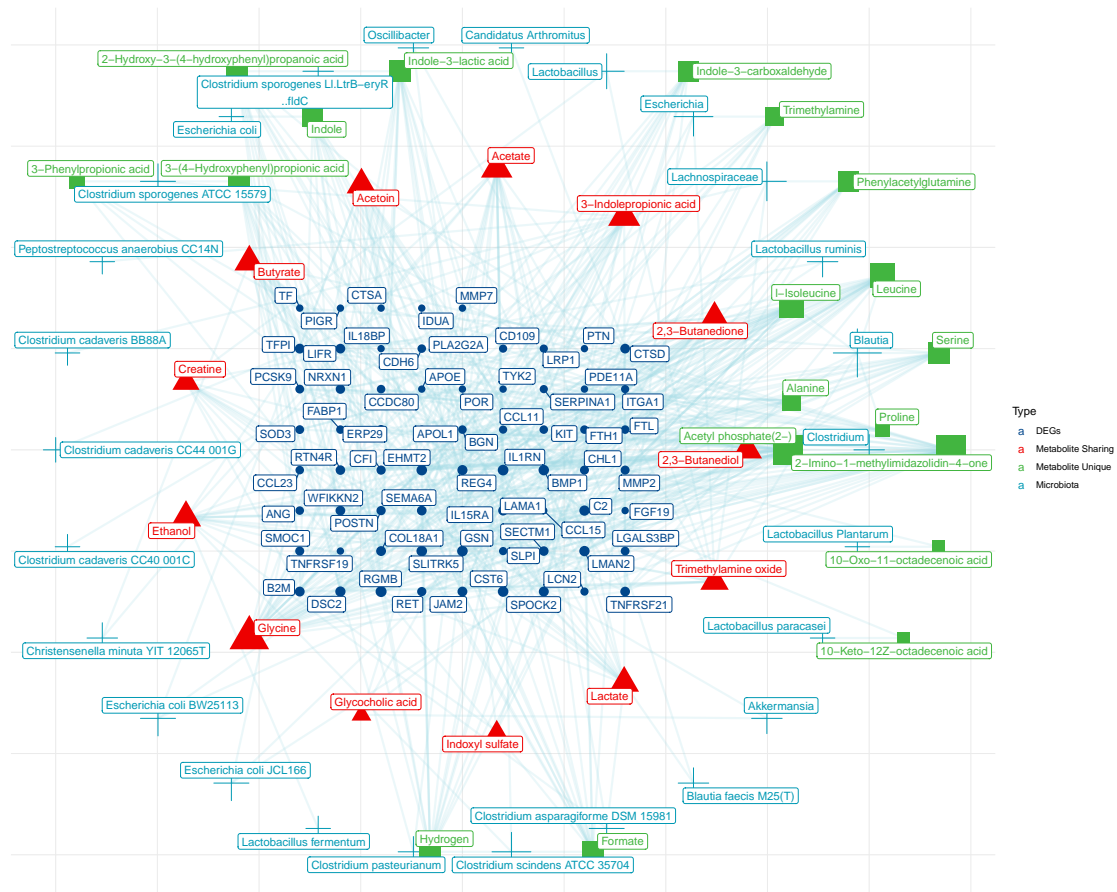


Figure 11: Ileum Top 1000 relationship network

Table 10 (下方表格) 为表格 Ileum Top 1000 relationship data 概览。

(对应文件为 **Figure+Table/Ileum-Top-1000-relationship-data.xlsx**)

注：表格共有 1000 行 3 列，以下预览的表格可能省略部分数据；表格含有 26 个唯一 ‘Gut.Microbiota’。

Table 10: Ileum Top 1000 relationship data

Gut.Microbiota	Metabolite	Target_Gene
Clostridium	2-Imino-1-methylimidazolidi...	B2M
Clostridium	2-Imino-1-methylimidazolidi...	DSC2
Clostridium	2-Imino-1-methylimidazolidi...	RGMB
Clostridium	Acetyl phosphate(2-)	RET
Blautia	Glycine	RET
Lactobacillus ruminis	Glycine	RET
Clostridium	2-Imino-1-methylimidazolidi...	JAM2
Clostridium	2-Imino-1-methylimidazolidi...	CST6
Clostridium	2-Imino-1-methylimidazolidi...	SPOCK2

Gut.Microbiota	Metabolite	Target_Gene
Clostridium	2-Imino-1-methylimidazolidi...	LCN2
Clostridium	2-Imino-1-methylimidazolidi...	TNFRSF21
Clostridium	2-Imino-1-methylimidazolidi...	SMOC1
Clostridium	2-Imino-1-methylimidazolidi...	TNFRSF19
Clostridium	2-Imino-1-methylimidazolidi...	COL18A1
Clostridium	Acetyl phosphate(2-)	SLITRK5
...

6.5 进一步验证代谢物的存在

以来自于文献² 胆结石小鼠模型研究的差异代谢物 (肝脏) 验证

以下是来源数据：

Table 11 (下方表格) 为表格 unnamed chunk 38 概览。

(对应文件为 **Figure+Table/unnamed-chunk-38.xlsx**)

注：表格共有 3104 行 9 列，以下预览的表格可能省略部分数据；表格含有 100 个唯一 ‘metabolite’。

1. cor: 皮尔逊关联系数，正关联或负关联。
2. pvalue: 显著性 P。
3. -log2(P.value): P 的对数转化。
4. significant: 显著性。
5. sign: 人为赋予的符号，参考 significant。

Table 11: Unnamed chunk 38

metabo...	microb...	cor	pvalue	AdjPvalue	-log2(...	signif...	sign	cid
PE(16:...	Prevot...	0.6120...	0.0049...	0.0159...	7.6581...	< 0.05	*	NA
PE(16:...	Alloba...	-0.559...	0.0115...	0.0218...	6.4339...	< 0.05	*	NA
PE(16:...	[Eubac...	-0.461...	0.0419...	0.0636...	4.5738...	< 0.05	*	NA
PE(16:...	A2	-0.514...	0.0218...	0.0428...	5.5171...	< 0.05	*	NA
PE(16:...	Trepon...	0.5303...	0.0161...	0.0471...	5.9517...	< 0.05	*	NA
PE(16:...	Anaero...	0.5185...	0.0191...	0.0383...	5.7051...	< 0.05	*	NA
PE(16:...	Bifido...	-0.670...	0.0016...	0.0160...	9.2801...	< 0.05	*	NA
PE(16:...	Entero...	-0.475...	0.0357...	0.0567...	4.8046...	< 0.05	*	NA
PE(16:...	Turici...	-0.524...	0.0176...	0.0299...	5.8208...	< 0.05	*	NA
PE(16:...	Tyzzzer...	-0.568...	0.0100...	0.0197...	6.6310...	< 0.05	*	NA
PE(16:...	[Eubac...	-0.478...	0.0345...	0.0931...	4.8570...	< 0.05	*	NA
PE(16:...	GCA-90...	-0.498...	0.0252...	0.0406...	5.3097...	< 0.05	*	NA

metabo...	microb...	cor	pvalue	AdjPvalue	-log2(...	signif...	sign	cid
PE(16:...	Rumino...	-0.466...	0.0382...	0.0868...	4.7099...	< 0.05	*	NA
PE(16:...	Tyzzzer...	0.6169...	0.0037...	0.0096...	8.0559...	< 0.05	*	NA
PE(16:...	[Rumin...	-0.472...	0.0370...	0.0699...	4.7527...	< 0.05	*	NA
...

6.5.1 Liver

Table 12 (下方表格) 为表格 Liver results filtered by validation 概览。

(对应文件为 **Figure+Table/Liver-results-filtered-by-validation.csv**)

注：表格共有 148 行 10 列，以下预览的表格可能省略部分数据；表格含有 1 个唯一 ‘.id’。

1. META_Rho: 关联分析结果的关联系数，绝对值越大，说明关联性越强 (源自文献的分析)
2. META_Q: 关联分析结果 P 的校正值 (源自文献的分析)
3. META_P: 关联分析结果 P 的值 (源自文献的分析)

Table 12: Liver results filtered by validation

.id	.id_from	Substrate	Metabo.....4	Gut.Mi...	Target...	Metabo.....7	META_RM	META_Q	META_P ...
222528	Metabo...	Bile acid	Deoxyc...	Clostr...	PCSK9	deoxyc...	0.0946...	6.9429...	1.4891... ..
222528	Metabo...	Cholic...	Deoxyc...	Clostr...	PCSK9	deoxyc...	0.0946...	6.9429...	1.4891... ..
222528	Metabo...	Chenod...	Deoxyc...	Clostr...	PCSK9	deoxyc...	0.0946...	6.9429...	1.4891... ..
222528	Metabo...	Bile acid	Deoxyc...	Clostr...	PCSK9	deoxyc...	0.0946...	6.9429...	1.4891... ..
222528	Metabo...	Bile acid	Deoxyc...	Clostr...	CD59	deoxyc...	-0.070...	4.5859...	2.0409... ..
222528	Metabo...	Cholic...	Deoxyc...	Clostr...	CD59	deoxyc...	-0.070...	4.5859...	2.0409... ..
222528	Metabo...	Chenod...	Deoxyc...	Clostr...	CD59	deoxyc...	-0.070...	4.5859...	2.0409... ..
222528	Metabo...	Bile acid	Deoxyc...	Clostr...	CD59	deoxyc...	-0.070...	4.5859...	2.0409... ..
222528	Metabo...	Bile acid	Deoxyc...	Clostr...	AHSG	deoxyc...	0.0688...	0.0001...	3.0112... ..
222528	Metabo...	Cholic...	Deoxyc...	Clostr...	AHSG	deoxyc...	0.0688...	0.0001...	3.0112... ..
222528	Metabo...	Chenod...	Deoxyc...	Clostr...	AHSG	deoxyc...	0.0688...	0.0001...	3.0112... ..
222528	Metabo...	Bile acid	Deoxyc...	Clostr...	AHSG	deoxyc...	0.0688...	0.0001...	3.0112... ..

.id	.id_from	Substrate	Metabo.....4	Gut.Mi...	Target...	Metabo.....7	META_RM	META_Q	META_P...
222528	Metabo...	Bile acid	Deoxyc...	Clostr...	CCL23	deoxyc...	-0.065...	0.0001...	6.9947...
222528	Metabo...	Cholic...	Deoxyc...	Clostr...	CCL23	deoxyc...	-0.065...	0.0001...	6.9947...
222528	Metabo...	Chenod...	Deoxyc...	Clostr...	CCL23	deoxyc...	-0.065...	0.0001...	6.9947...
...

6.5.1.1 富集分析

Figure 12 (下方图) 为图 LIVER ids KEGG enrichment 概览。

(对应文件为 [Figure+Table/LIVER-ids-KEGG-enrichment.pdf](#))

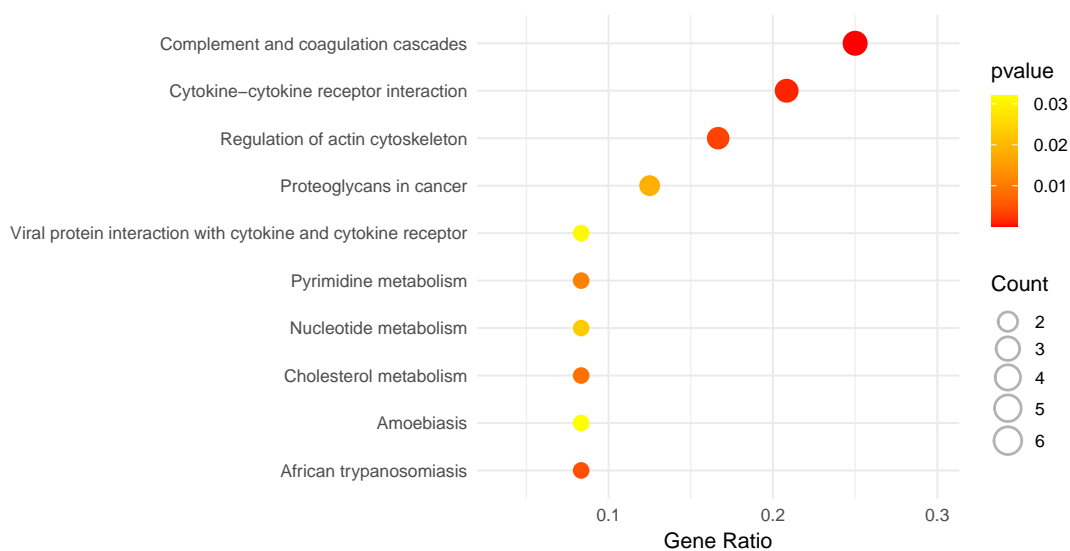


Figure 12: LIVER ids KEGG enrichment

Figure 13 (下方图) 为图 LIVER ids GO enrichment 概览。

(对应文件为 [Figure+Table/LIVER-ids-GO-enrichment.pdf](#))

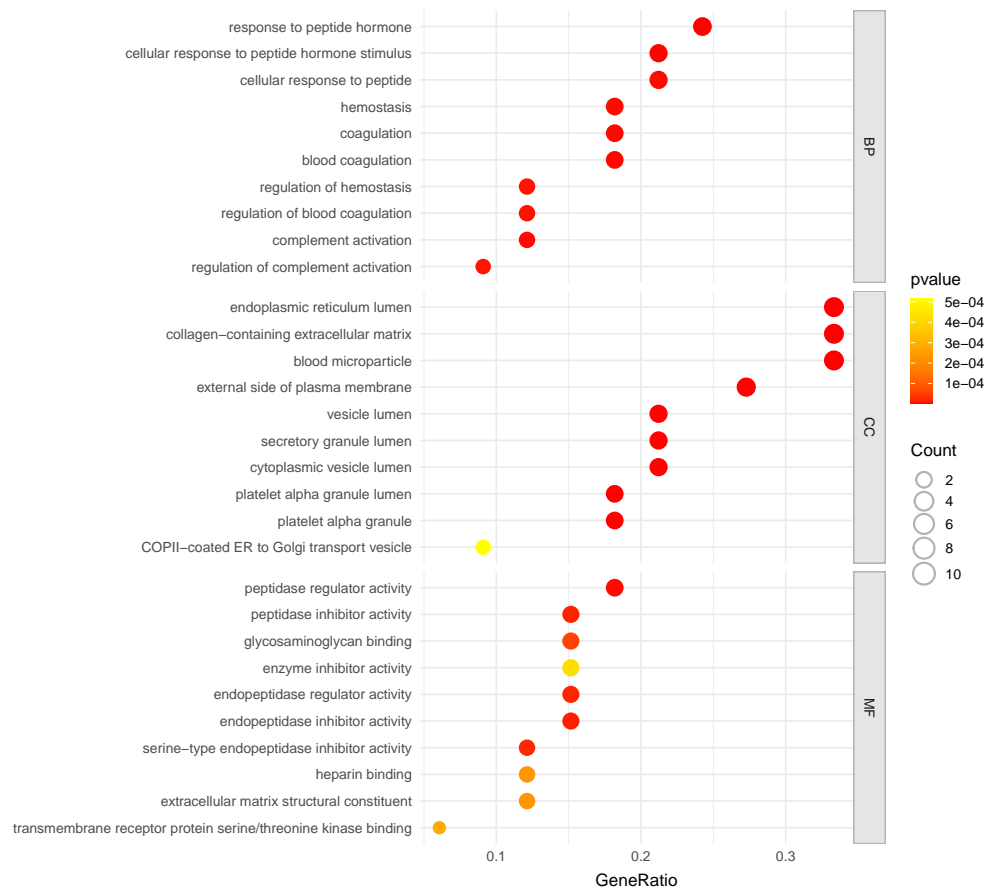


Figure 13: LIVER ids GO enrichment

6.5.2 Ileum

Table 13 (下方表格) 为表格 Ileum results filtered by validation 概览。

(对应文件为 **Figure+Table/Ileum-results-filtered-by-validation.csv**)

注：表格共有 104 行 10 列，以下预览的表格可能省略部分数据；表格含有 1 个唯一 ‘id’。

1. META_Rho: 关联分析结果的关联系数，绝对值越大，说明关联性越强 (源自文献的分析)
2. META_Q: 关联分析结果 P 的校正值 (源自文献的分析)
3. META_P: 关联分析结果 P 的值 (源自文献的分析)

Table 13: Ileum results filtered by validation

.id	.id_from	Substrate	Metabo.....4	Gut.Mi...	Target...	Metabo.....7	META_Rho	META_Q	META_P...
222528	Metabo...	Bile acid	Deoxyc...	Clostr...	FGF19	deoxyc...	0.0981...	8.8819...	4.5243...
222528	Metabo...	Cholic...	Deoxyc...	Clostr...	FGF19	deoxyc...	0.0981...	8.8819...	4.5243...

.id	.id_from	Substrate	Metabo.....4	Gut.Mi...	Target...	Metabo.....7	META_RM	META_Q	META_P ...
222528	Metabo...	Chenod...	Deoxyc...	Clostr...	FGF19	deoxyc...	0.0981...	8.8819...	4.5243... ..
222528	Metabo...	Bile acid	Deoxyc...	Clostr...	FGF19	deoxyc...	0.0981...	8.8819...	4.5243... ..
222528	Metabo...	Bile acid	Deoxyc...	Clostr...	PCSK9	deoxyc...	0.0946...	6.9429...	1.4891... ..
222528	Metabo...	Cholic...	Deoxyc...	Clostr...	PCSK9	deoxyc...	0.0946...	6.9429...	1.4891... ..
222528	Metabo...	Chenod...	Deoxyc...	Clostr...	PCSK9	deoxyc...	0.0946...	6.9429...	1.4891... ..
222528	Metabo...	Bile acid	Deoxyc...	Clostr...	PCSK9	deoxyc...	0.0946...	6.9429...	1.4891... ..
222528	Metabo...	Bile acid	Deoxyc...	Clostr...	CCL23	deoxyc...	-0.065...	0.0001...	6.9947... ..
222528	Metabo...	Cholic...	Deoxyc...	Clostr...	CCL23	deoxyc...	-0.065...	0.0001...	6.9947... ..
222528	Metabo...	Chenod...	Deoxyc...	Clostr...	CCL23	deoxyc...	-0.065...	0.0001...	6.9947... ..
222528	Metabo...	Bile acid	Deoxyc...	Clostr...	CCL23	deoxyc...	-0.065...	0.0001...	6.9947... ..
222528	Metabo...	Bile acid	Deoxyc...	Clostr...	CCL23	deoxyc...	-0.060...	0.0005...	0.0002... ..
222528	Metabo...	Cholic...	Deoxyc...	Clostr...	CCL23	deoxyc...	-0.060...	0.0005...	0.0002... ..
222528	Metabo...	Chenod...	Deoxyc...	Clostr...	CCL23	deoxyc...	-0.060...	0.0005...	0.0002... ..
...

6.5.2.1 富集分析

Figure 14 (下方图) 为图 ILEUM ids GO enrichment 概览。

(对应文件为 **Figure+Table/ILEUM-ids-GO-enrichment.pdf**)

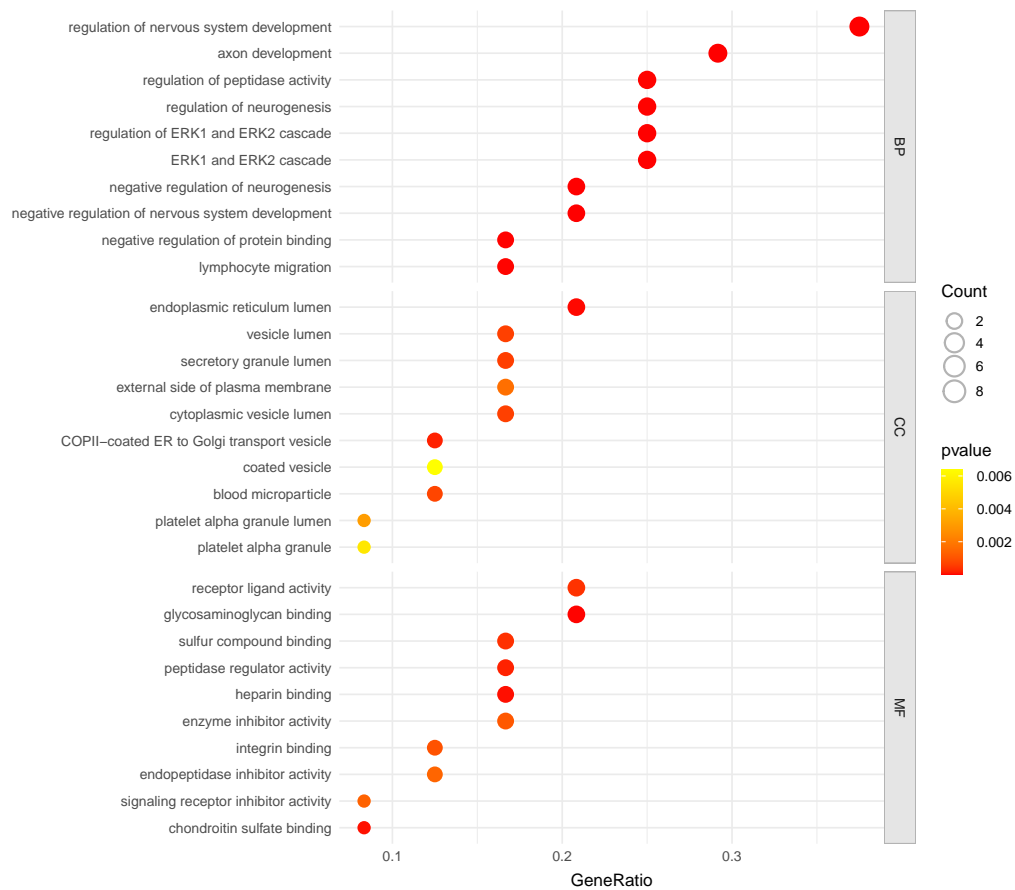


Figure 14: ILEUM ids GO enrichment

Figure 15 (下方图) 为图 ILEUM ids KEGG enrichment 概览。

(对应文件为 [Figure+Table/ILEUM-ids-KEGG-enrichment.pdf](#))

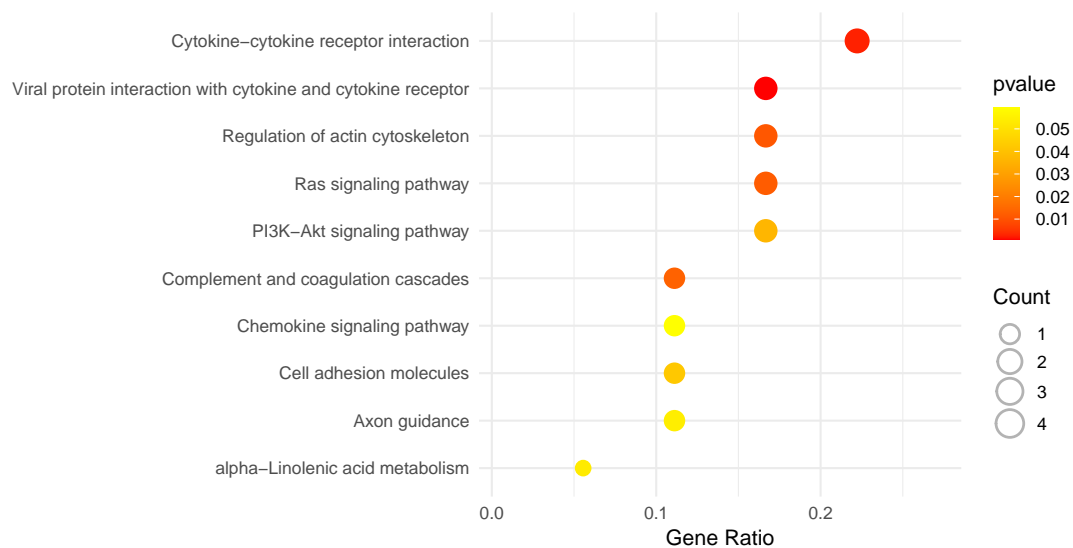


Figure 15: ILEUM ids KEGG enrichment

Reference

1. Benson, M. D. *et al.* Protein-metabolite association studies identify novel proteomic determinants of metabolite levels in human plasma. *Cell Metabolism* **35**, 1646–1660.e3 (2023).
2. Chen, Y. *et al.* Changes and correlations of the intestinal flora and liver metabolite profiles in mice with gallstones. *Frontiers in physiology* **12**, (2021).
3. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the r/bioconductor package biomaRt. *Nature protocols* **4**, 1184–1191 (2009).
4. Wu, T. *et al.* ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, (2021).
5. Chen, S. Ultrafast one-pass fastq data preprocessing, quality control, and deduplication using fastp. *iMeta* **2**, (2023).
6. 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).
7. Ritchie, M. E. *et al.* Limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Research* **43**, e47 (2015).
8. Chen, Y., McCarthy, D., Ritchie, M., Robinson, M. & Smyth, G. EdgeR: Differential analysis of sequence read count data user’s guide. 119.
9. Xu, S. *et al.* MicrobiotaProcess: A comprehensive r package for deep mining microbiome. *The Innovation* **4**, 100388 (2023).
10. Blanco-Míguez, A. *et al.* Extending and improving metagenomic taxonomic profiling with uncharacterized species using metaphlan 4. *Nature Biotechnology* **41**, 1633–1644 (2023).