胆结石 RNA-seq 结合肠道菌、代谢物筛选关键差异表达基因

2023-12-25

LiChuang Huang

@ 立效研究院

Contents

List of Figures

List of Tables

1 摘要

需求：

根据客户提供的 RNA-seq，结合肠道菌、代谢物筛选关键差异表达基因，基因不要是 FXR 及其相关信号通路（CYP7A1 等），要与胆固醇代谢、胆固醇摄取、胆固醇合成、胆固醇重吸收和胆汁酸代谢相关；同时结合肠道菌群大数据库，结合菌群代谢产物。注：客户研究的疾病是胆固醇胆结石（cholesterol gallstones），如果没有使用胆结石也可。

结果：见 4。

2 前言

客户拥有的数据类型仅为 RNA-seq，反映的是组织 mRNA 水平。当前公共数据缺少同时结合胆结石 (gallstones, G) 疾病的 RNA-seq、肠道菌、代谢组的分析类型。因此，为了将客户的 RNA-seq 分析结果与肠道菌和代谢物建立联系，设计思路为：

- DEGs -> eQTL -> SNP -> GWAS -> Metabolites and Microbiota

eQTL 分析的本质是以全部的 DNA 变异位点为自变量，轮流以每种 mRNA 表达量为因变量，用大量的个体数据做样本进行线性回归，得到每一个 SNP 位点和每一个 mRNA 表达量间的关系 (https://www.nature.com/scitable/topicpage/quantitative-trait-locus-qtl-analysis-53904/)。

本次分析，通过寻找 mRNA 和 SNP 之间的关联，让 RNA-seq 筛选的 DEGs 联系到已有的关于代谢物或微生物的 GWAS 大数据研究 (3.2 这些数据反映了 SNP 与代谢物或微生物之间的关联性) (即 SNP 作为桥梁) 筛选出关键 DEGs 和对应的肠道微生物和代谢物，最后再联系已有的胆结石 (gallstones, G) 的肠道菌或代谢物的研究进行验证。

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 to[1].
- Supplementary file from article refer to[2].

3.2 方法

Mainly used method:

- R package `biomaRt` used for gene annotation.[3]
- The `biomart` was used for mapping genes between organism (e.g., mgi_symbol to hgnc_symbol).[3]
- R package `ClusterProfiler` used for gene enrichment analysis.[4]
- The QTL data were abtained from GTEx database.[5]
- R package `ClusterProfiler` used for GSEA enrichment.[4]
- Database `gutMDisorder` used for finding associations between gut microbiota and metabolites.[6]

- R package `Limma` and `edgeR` used for differential expression analysis.[7,8]
- Other R packages (eg., `dplyr` and `ggplot2`) used for statistic analysis or data visualization.

4 分析结果

4.1 Liver:

- 根据 Model vs Control 初步筛选 DEGs (Tab. 1)
- DEGs 从 Mouce 到 Human 映射 (Tab. 2)
- 对上述映射后的基因进行 KEGG 的 GSEA 富集，结果发现 'Steroid biosynthesis' 为首要富集结果 (Fig. 2
- 为了找到 DEGs 可能对应的 SNP，使用 eQTL 数据集，并筛选该数据集 (Fig. 4)

- 上述数据建立了：DEGs -> SNP 之间的关联，随后需要建立 SNP -> metablite 或者 microbiota 的关联，因此这里使用了相关的 GWAS 数据，并做了筛选 (Tab. 5、Tab. 6) 。这样，SNP -> metablite 或者 microbiota 的关联就确立了。往上对应到 DEGs (Human)，它们是：ITGB3, C9orf152。
- 随后，为了发现更多的与上述筛选的 metabolite 或者 microbiota 相关的 metabolite 或者 microbiota，使用了 gutMDisorder 数据库，挖掘到的信息见 Tab. 7
- 为了验证上述的发现，使用了[1] 的数据 (这是一批研究胆结石 (gallstones, G) 的代谢物和肠道微生物的 mice 的数据) (Fig. 7) 。筛选后发现，Ruminococcus 的确在胆结石 (gallstones, G) 中属于差异微生物。这样，串联上述线索，发现了关系链：
  - Microbiota:Ruminococcus -> Metabolite:Leucine -> SNP:`chr17_47247224_A_G_b38` -> DEG:ITGB3
- 这里，进一步将 ITGB3, C9orf152 与 Steroid biosynthesis 通路的其它基因做了关联分析，发现它们主要成显著的负关联 (Fig. 8)。
- 这些基因在 human 或者 mice 中的基因名的对应关系见 Tab. 10
- 建议以 ITGB3 或上述其它基因 (Steroid biosynthesis 通路) 作为目标基因进一步分析。

5 结论

6 附：分析流程 (Liver)

6.1 差异表达基因

6.1.1 Model vs Control

Table 1 (下方表格) 为表格 Liver raw DEGs Model vs control 概览。

(对应文件为 `Figure+Table/Liver-raw-DEGs-Model-vs-control.xlsx`)

> 注：表格共有 3908 行 11 列，以下预览的表格可能省略部分数据；表格含有 3908 个唯一 'ensembl_transcript_id'。

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 1: Liver raw DEGs Model vs control

ensemb…	mgi_sy…	entrez…	hgnc_s…	descri…	logFC	AveExpr	t	P.Value	adj.P.Val
ENSMUS…	Cyp2c70	226105	NA	cytoch…	4.1662…	5.9781…	23.094…	1.2533…	3.9223…
ENSMUS…	Scd1	20249	NA	stearo…	2.8187…	11.748…	19.213…	6.8879…	0.0001…
ENSMUS…	Ces2a	102022		carbox…	1.6614…	8.8361…	15.281…	5.6258…	0.0004…
ENSMUS…	Hsd17b6	27400		hydrox…	2.8011…	8.6106…	14.201…	1.0950…	0.0006…
ENSMUS…	Fmo5	14263		flavin…	1.2618…	8.1280…	13.790…	1.4285…	0.0007…
ENSMUS…	Hsd17b6	27400		hydrox…	3.0271…	4.8530…	13.490…	1.7415…	0.0007…
ENSMUS…	Enho	69638	NA	energy…	-4.453…	2.2957…	-17.04…	2.0685…	0.0002…
ENSMUS…	Abcb11	27413		ATP-bi…	1.2515…	7.7893…	11.121…	9.7706…	0.0033…
ENSMUS…	Hsd17b6	27400		hydrox…	3.6061…	4.4081…	11.158…	9.4863…	0.0033…
ENSMUS…	Gsta4	14860	NA	glutat…	1.9687…	6.1777…	10.257…	1.9887…	0.0056…
ENSMUS…	Gnat1	14685	NA	G prot…	-2.232…	2.9799…	-10.64…	1.4365…	0.0044…
ENSMUS…	Nnmt	18113	NA	nicoti…	-3.804…	5.1567…	-9.928…	2.6415…	0.0065…
ENSMUS…	Csad	246277	NA	cystei…	-1.560…	7.2232…	-9.535…	3.7482…	0.0083…
ENSMUS…	Hsd17b6	27400		hydrox…	2.9137…	3.3218…	9.8923…	2.7263…	0.0065…
ENSMUS…	Mup7	100041658	NA	major …	-10.24…	7.5916…	-9.232…	4.9474…	0.0087…
…	…	…	…	…	…	…	…	…	…

Figure 1 (下方图) 为图 Liver plot DEGs Model vs control 概览。

(对应文件为 `Figure+Table/Liver-plot-DEGs-Model-vs-control.pdf`)

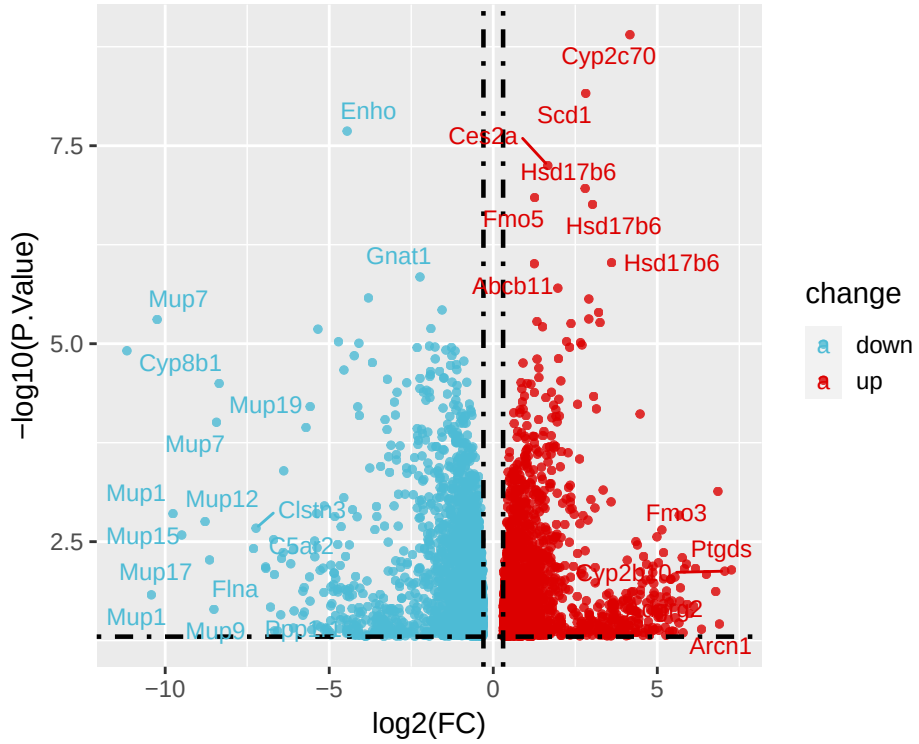

Figure 1: Liver plot DEGs Model vs control

6.2 DEGs 从 Mouce 到 Human 映射

6.2.1 Biomart mapping

客户的数据为 Mouce 的数据，这里将 Mouce 的基因映射为 Human 的基因 (因为后续的数据来源主要为 Human)。

Table 2 (下方表格) 为表格 Liver DEGs mapping from Mice to Human 概览。

(对应文件为 `Figure+Table/Liver-DEGs-mapping-from-Mice-to-Human.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 2: Liver DEGs mapping from Mice to Human

hgnc_s…	mgi_sy…	ensemb…	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…
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…
…	…	…	…	…	…	…	…	…	…

6.3 GSEA 富集 (Human)

6.3.1 pathways

对映射完毕的 DEGs (Tab. 2) 进行富集分析，首要富集结果为 'Steroid biosynthesis'。

Figure 2 (下方图) 为图 LIVER KEGG enrichment with enriched genes 概览。

(对应文件为 `Figure+Table/LIVER-KEGG-enrichment-with-enriched-genes.pdf`)

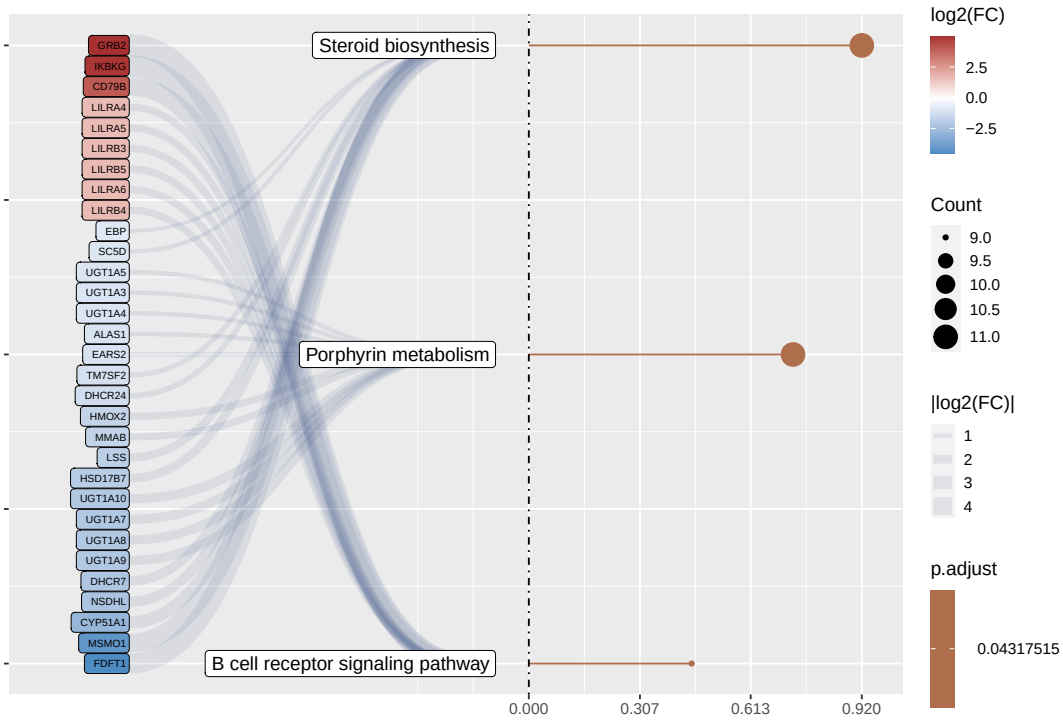

Figure 2: LIVER KEGG enrichment with enriched genes

Figure 3 (下方图) 为图 LIVER GSEA plot of the pathways 概览。

(对应文件为 `Figure+Table/LIVER-GSEA-plot-of-the-pathways.pdf`)

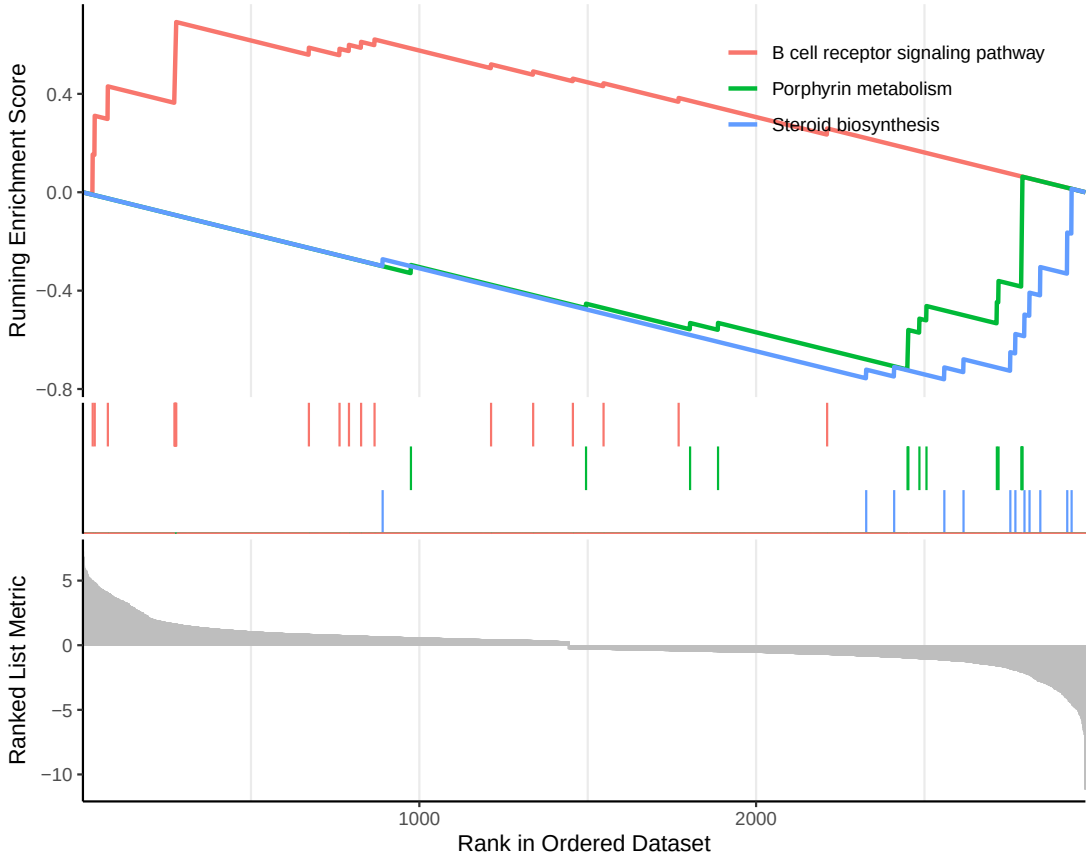

Figure 3: LIVER GSEA plot of the pathways

6.4 eQTL 数据：寻找基因表达变化（DEGs）和突变（SNP）的关联

6.4.1 eQTL 数据

使用的 eQTL 数据集（经过注释的，来源见 3.2 QTL 说明）：

Table 3（下方表格）为表格 LIVER all used eQTL data 概览。

（对应文件为 `Figure+Table/LIVER-all-used-eQTL-data.csv`）

> 注：表格共有 341233 行 13 列，以下预览的表格可能省略部分数据；表格含有 221715 个唯一 'variant_id'。

1. hgnc_symbol: 基因名 (Human)
2. gene_id: GENCODE/Ensembl gene ID
3. variant_id: variant ID in the format {chr}_{pos_first_ref_base}_{ref_seq}_{alt_seq}_b38
4. tss_distance: distance between variant and transcription start site (TSS). Positive when variant is downstream of the TSS, negative otherwise
5. maf: minor allele frequency observed in the set of donors for a given tissue
6. pval_nominal: nominal p-value associated with the most significant variant for this gene
7. slope: regression slope
8. slope_se: standard error of the regression slope
9. pval_beta: beta-approximated permutation p-value
10. pval_nominal_threshold: nominal p-value threshold for calling a variant-gene pair significant for the gene
11. ma_samples: number of samples carrying the minor allele
12. ma_count: total number of minor alleles across individuals
13. min_pval_nominal: smallest nominal p-value for the gene

Table 3: LIVER all used eQTL data

varian…	gene_id	tss_di…	ma_sam…	ma_count	maf	pval_n……7	slope	slope_se	pval_n……10
chr1_1…	ENSG00…	-282825	21	21	0.0504808	1.2263…	-0.992022	0.197055	7.3643…
chr1_5…	ENSG00…	-38486	3	3	0.0072…	1.4398…	1.9902	0.445336	4.4165…
chr1_1…	ENSG00…	819409	7	7	0.0168269	5.0290…	1.44172	0.27575	4.4165…
chr1_1…	ENSG00…	995083	77	86	0.206731	2.6972…	0.386875	0.08962	4.4165…
chr1_9…	ENSG00…	193015	3	3	0.0072…	3.7957…	-2.4096	0.504028	4.9560…
chr1_2…	ENSG00…	-510872	10	10	0.0240385	4.3979…	-0.97553	0.232511	4.4931…
chr1_9…	ENSG00…	158610	6	6	0.0144231	7.4378…	-1.47776	0.319499	4.4931…
chr1_9…	ENSG00…	170420	26	27	0.0652174	1.5197…	0.665794	0.149414	4.4931…
chr1_9…	ENSG00…	183319	27	28	0.0673077	8.0998…	0.680912	0.147854	4.4931…
chr1_7…	ENSG00…	-98104	45	49	0.117788	3.8806…	-0.430994	0.081567	4.4917…
chr1_7…	ENSG00…	-97896	44	48	0.115385	1.3477…	-0.408977	0.0815771	4.4917…
chr1_7…	ENSG00…	-97661	53	64	0.153846	2.6452…	0.343539	0.0794932	4.4917…
chr1_7…	ENSG00…	-66787	45	49	0.117788	1.1198…	-0.405186	0.0801673	4.4917…
chr1_7…	ENSG00…	-66695	45	48	0.115385	7.1610…	-0.421834	0.0818781	4.4917…
chr1_7…	ENSG00…	-28800	44	47	0.112981	4.1144…	-0.396941	0.0833519	4.4917…
…	…	…	…	…	…	…	…	…	…

6.4.2 Variant (与 DEGs 相关)

根据 DEGs 的基因名过滤 eQTL 数据：

Figure 4 (下方图) 为图 LIVER database of eQTL intersect with DEGs 概览。

(对应文件为 `Figure+Table/LIVER-database-of-eQTL-intersect-with-DEGs.pdf`)

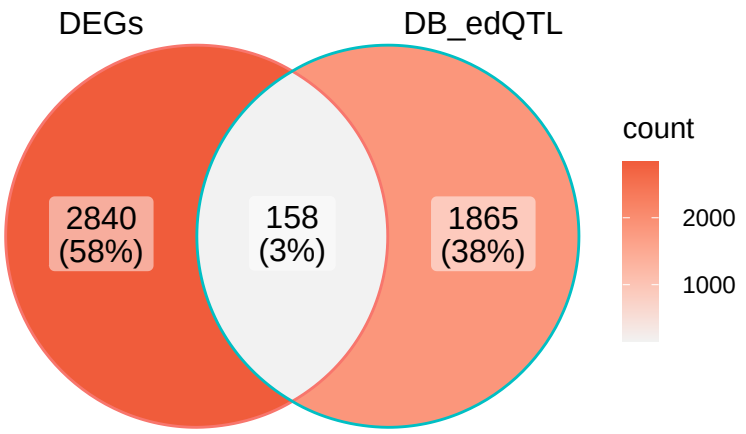

Figure 4: LIVER database of eQTL intersect with DEGs

Intersection :

ENHO, SLC22A3, STARD10, GNMT, TLCD2, PON1, SIK1, DDTL, APOC4, SPRYD3, GM2A, UGT3A2, RHPN2, SULT1C2, MRPL15, C9orf152, SLC39A4, RHOD, SAA4, SLC10A1, NUDT18, ANG, IPO7, PROB1, GNG12, MPZL3, OST4, ASAH2B, FOLR3, CTAGE15, CD1D, APOC2, KISS1, GOLT1A, PAQR9, DPP3, HUNK, MIF, HPR, HP, SULT2A1, IL17RB, FADS3, DELE1, SCCPDH, CYP2A6, CYP2A13, GNPNAT1, LINGO4, C14orf119, MRPL14, NEMP1, AP1M1, PXMP2, PSMB3, OTUD1, MBL2, RCN1, E2F2, TMEM238, EPPK1, CMTM6, IKBKE, ALDH16A1, C2orf42, TRABD, RTEL1, COL5A3, CTDSP1, DIPK2A, TXN2, IMMP2L, CA5A, GRIK5, TMEM176A, ATP23, GJC3, ZNF429, TPMT, OMD, RAP2C, CRCP, L3MBTL3, ACY3, PRMT6, CD2BP2, IL22RA1, RNF168, BRI3, ITPRIPL1, ORM2, KPNA2, CHMP4C, PPDPF, CCL15, OXSM, COPS7A, CYP3A7-CYP3A51P, MBLAC2, ACOT12, SEC11C, SURF6, TRUB1, ELOVL2, MLYCD, MARVELD3, ZBTB33, PPIL1, AMIGO1, CYC1, C11orf96, ITGB3, GLUD2, TMEM134, DHRS3, PRRG4, LLPH, GLO1, IL1RAP, NOCT, NTAQ1, VSIG10L, LRRC46, AMDHD1, LRRC57, SERPINA12, UFL1, CCL27, FAM136A, RAB22A, FCGR2C, ZFPM1, TMEM218, GCNT4, TADA1, GNG10, ANKRD9, DECR1, ZNF408, TCEA3, DSG1, INMT, IVD, LARS2, CYP27A1, PLIN3, TMEM47, CYSTM1, FXYD1, EVI5L, NME6, NSA2, GTF2I, LCMT2, PPP1CB, AGXT, CLDN1, PARG

(上述信息框内容已保存至 `Figure+Table/LIVER-database-of-eQTL-intersect-with-DEGs-content`)

Table 4 (下方表格) 为表格 LIVER database of eQTL intersect with DEGs DATA 概览。

(对应文件为 `Figure+Table/LIVER-database-of-eQTL-intersect-with-DEGs-DATA.csv`)

注：表格共有 9785 行 13 列，以下预览的表格可能省略部分数据；表格含有 9455 个唯一 'variant_id'。

1. hgnc_symbol: 基因名 (Human)
2. gene_id: GENCODE/Ensembl gene ID
3. variant_id: variant ID in the format {chr}_{pos_first_ref_base}_{ref_seq}_{alt_seq}_b38
4. tss_distance: distance between variant and transcription start site (TSS). Positive when variant is downstream of the TSS, negative otherwise
5. maf: minor allele frequency observed in the set of donors for a given tissue
6. pval_nominal: nominal p-value associated with the most significant variant for this gene
7. slope: regression slope
8. slope_se: standard error of the regression slope
9. pval_beta: beta-approximated permutation p-value
10. pval_nominal_threshold: nominal p-value threshold for calling a variant-gene pair significant for the gene
11. ma_samples: number of samples carrying the minor allele
12. ma_count: total number of minor alleles across individuals
13. min_pval_nominal: smallest nominal p-value for the gene

Table 4: LIVER database of eQTL intersect with DEGs DATA

varian…	gene_id	tss_di…	ma_sam…	ma_count	maf	pval_n……7	slope	slope_se	pval_n……10
chr1_1…	ENSG00…	-837790	32	35	0.0841346	1.7084…	0.446491	0.100836	3.1656…
chr1_1…	ENSG00…	-808267	24	25	0.0600962	2.8012…	-0.470158	0.109147	3.1656…
chr1_1…	ENSG00…	-270870	11	11	0.0264423	5.3189…	0.79694	0.169447	3.1656…
chr1_1…	ENSG00…	-270849	11	11	0.0264423	5.3189…	0.79694	0.169447	3.1656…
chr1_1…	ENSG00…	-193795	13	13	0.03125	4.2127…	0.739458	0.155452	3.1656…
chr1_1…	ENSG00…	-124521	13	13	0.03125	4.2127…	0.739458	0.155452	3.1656…
chr1_1…	ENSG00…	-94331	13	13	0.03125	4.2127…	0.739458	0.155452	3.1656…
chr1_2…	ENSG00…	-829148	4	4	0.0096…	8.5729…	0.771259	0.167959	4.9989…
chr1_2…	ENSG00…	-828022	4	4	0.0096…	8.5729…	0.771259	0.167959	4.9989…
chr1_2…	ENSG00…	40717	62	70	0.168269	4.0212…	0.173992	0.0412498	4.9989…
chr1_2…	ENSG00…	356991	11	11	0.0264423	6.9192…	-1.03253	0.222426	3.5636…
chr1_2…	ENSG00…	-135634	14	15	0.0360577	3.6543…	-0.674426	0.158994	4.9389…
chr1_2…	ENSG00…	-52692	20	23	0.0552885	3.2148…	-0.512818	0.119997	4.9389…
chr1_2…	ENSG00…	-25624	21	23	0.0552885	2.0117…	-0.553662	0.126165	4.9389…
chr1_2…	ENSG00…	-23866	19	21	0.0504808	4.0686…	-0.550719	0.130654	4.9389…
…	…	…	…	…	…	…	…	…	…

6.5 GWAS 数据：寻找与突变类型显著关联的肠道微生物或代谢物

6.5.1 GWAS 数据

以下为使用的 GWAS 数据 (代谢物或微生物与 variant 的显著关系，来源见 3.1)：

'PUBLISHED MendelianRandoLiuX2022' 数据已全部提供。

(对应文件为 `Figure+Table/PUBLISHED-MendelianRandoLiuX2022`)

> 注：文件夹 Figure+Table/PUBLISHED-MendelianRandoLiuX2022 共包含 2 个文件。
>
> 1. 1_snp_microbiota.csv
> 2. 2_snp_metabolite.csv

以下，结合 Tab. 4，根据 variant_id 筛选上述数据。

6.5.2 Microbiota

Figure 5 (下方图) 为图 LIVER filtered eQTL data intersect with microbiota related 概览。

(对应文件为 `Figure+Table/LIVER-filtered-eQTL-data-intersect-with-microbiota-related.pdf`)

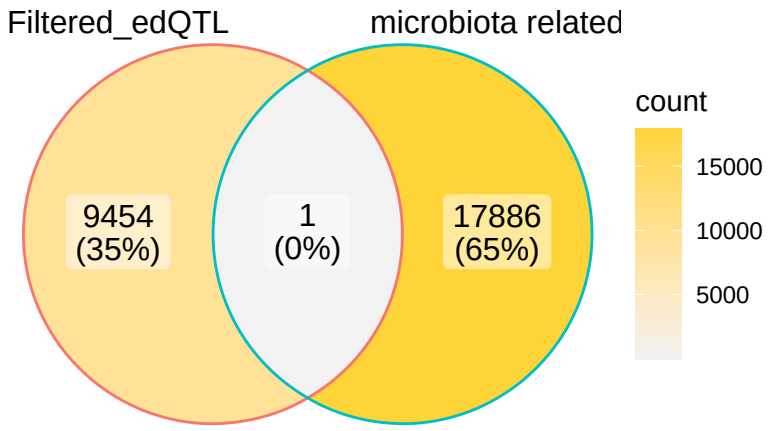

Figure 5: LIVER filtered eQTL data intersect with microbiota related

Intersection :

chr9_110149941_A_G_b38

(上述信息框内容已保存至 `Figure+Table/LIVER-filtered-eQTL-data-intersect-with-microbiota-related-content`)

Table 5 (下方表格) 为表格 LIVER filtered eQTL data intersect with microbiota related DATA 概览。

(对应文件为 `Figure+Table/LIVER-filtered-eQTL-data-intersect-with-microbiota-related-DATA.csv`)

> 注：表格共有 1 行 3 列，以下预览的表格可能省略部分数据；表格含有 1 个唯一 'variant_id'。

1. hgnc_symbol: 基因名 (Human)
2. variant_id: variant ID in the format {chr}_{pos_first_ref_base}_{ref_seq}_{alt_seq}_b38

Table 5: LIVER filtered eQTL data intersect with microbiota related DATA

variant_id	Microbiome.features	hgnc_symbol
chr9_110149941_A_G_b38	s_Mobiluncus_mulieris	C9orf152

6.5.3 Metabolite

Figure 6 (下方图) 为图 LIVER filtered eQTL data intersect with metabolite related 概览。

(对应文件为 `Figure+Table/LIVER-filtered-eQTL-data-intersect-with-metabolite-related.pdf`)

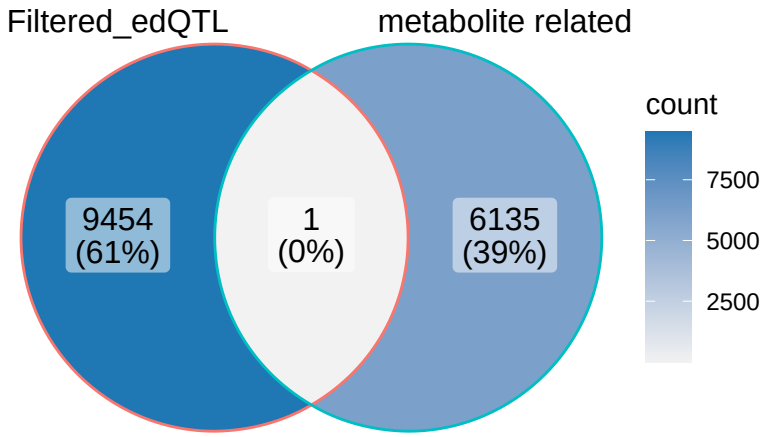

Figure 6: LIVER filtered eQTL data intersect with metabolite related

Intersection :

chr17_47247224_A_G_b38

(上述信息框内容已保存至 `Figure+Table/LIVER-filtered-eQTL-data-intersect-with-metabolite-related-content`)

Table 6 (下方表格) 为表格 LIVER filtered eQTL data intersect with metabolite related DATA 概览。

(对应文件为 `Figure+Table/LIVER-filtered-eQTL-data-intersect-with-metabolite-related-DATA.csv`)

注:表格共有 1 行 3 列,以下预览的表格可能省略部分数据;表格含有 1 个唯一 'variant_id'。

1. hgnc_symbol: 基因名 (Human)
2. variant_id: variant ID in the format {chr}_{pos_first_ref_base}_{ref_seq}_{alt_seq}_b38

Table 6: LIVER filtered eQTL data intersect with metabolite related DATA

variant_id	Metabolic.traits	hgnc_symbol
chr17_47247224_A_G_b38	Leucine	ITGB3

6.6 肠道菌和代谢物关联数据库筛选

在 6.5.2 和 6.5.3 中，分别筛选到了一组 SNP 与 microbiota 或者 SNP 与 metabolite 之间的关联。以下，以 gutMDisorder 数据库寻找与该 microbiota 或 metabolite 关联的其它 metabolite 或 microbiota。

6.6.1 以 Microbiota 筛选

无结果。

6.6.2 以 Metabolite 筛选

结果如下：

Table 7 (下方表格) 为表格 Liver gutMDisorder Matched metabolites and their related microbiota 概览。

(对应文件为 `Figure+Table/Liver-gutMDisorder-Matched-metabolites-and-their-related-microbiota.csv`)

> 注：表格共有 5 行 4 列，以下预览的表格可能省略部分数据；表格含有 1 个唯一 'Metabolite'。

Table 7: Liver gutMDisorder Matched metabolites and their related microbiota

Metabolite	Substrate	Gut.Microbiota	Classification
Leucine		Ruminococcus	genus
Leucine		Dorea	genus
Leucine		Blautia	genus
Leucine		Faecalibacterium	genus
Leucine		Faecalibacterium …	species

6.7 已有的胆结石 (gallstones, G) 的微生物和代谢物关联研究

6.7.1 ChangesAndCorChen2021

数据来源于[1]

Figure 7 (下方图) 为图 PUBLISHED ChangesAndCorChen2021 correlation heatmap 概览。

(对应文件为 `Figure+Table/PUBLISHED-ChangesAndCorChen2021-correlation-heatmap.pdf`)

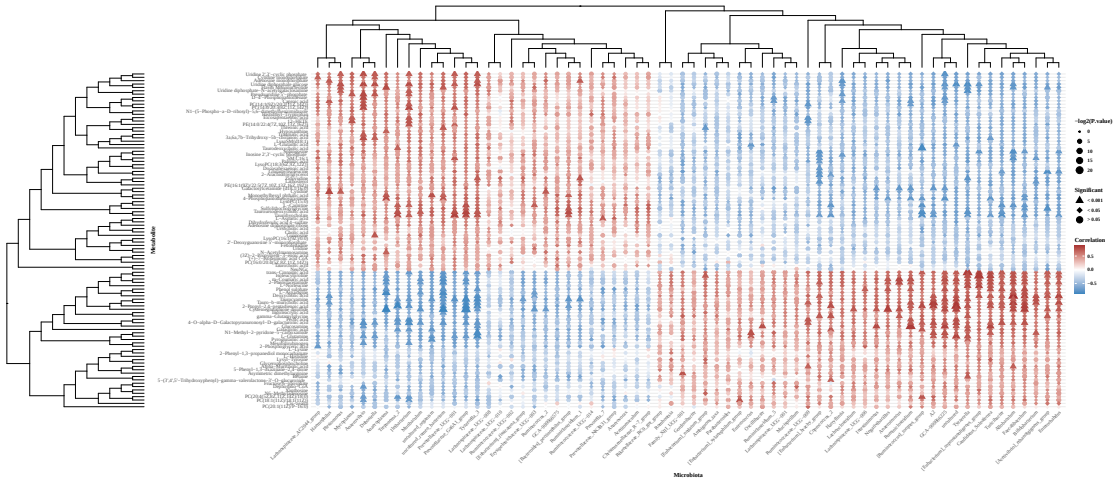

Figure 7: PUBLISHED ChangesAndCorChen2021 correlation heatmap

Table 8 (下方表格) 为表格 PUBLISHED ChangesAndCorChen2021 significant correlation 概览。

(对应文件为 `Figure+Table/PUBLISHED-ChangesAndCorChen2021-significant-correlation.xlsx`)

注：表格共有 3104 行 8 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 8: PUBLISHED ChangesAndCorChen2021 significant correlation

metabo...	microb...	cor	pvalue	AdjPvalue	-log2(...	signif...	sign
PE(16:...	Prevot...	0.6120...	0.0049...	0.0159...	7.6581...	< 0.05	*
PE(16:...	Alloba...	-0.559...	0.0115...	0.0218...	6.4339...	< 0.05	*
PE(16:...	[Eubac...	-0.461...	0.0419...	0.0636...	4.5738...	< 0.05	*
PE(16:...	A2	-0.514...	0.0218...	0.0428...	5.5171...	< 0.05	*
PE(16:...	Trepon...	0.5303...	0.0161...	0.0471...	5.9517...	< 0.05	*
PE(16:...	Anaero...	0.5185...	0.0191...	0.0383...	5.7051...	< 0.05	*
PE(16:...	Bifido...	-0.670...	0.0016...	0.0160...	9.2801...	< 0.05	*
PE(16:...	Entero...	-0.475...	0.0357...	0.0567...	4.8046...	< 0.05	*
PE(16:...	Turici...	-0.524...	0.0176...	0.0299...	5.8208...	< 0.05	*
PE(16:...	Tyzzer...	-0.568...	0.0100...	0.0197...	6.6310...	< 0.05	*
PE(16:...	[Eubac...	-0.478...	0.0345...	0.0931...	4.8570...	< 0.05	*
PE(16:...	GCA-90...	-0.498...	0.0252...	0.0406...	5.3097...	< 0.05	*
PE(16:...	Rumino...	-0.466...	0.0382...	0.0868...	4.7099...	< 0.05	*

metabo...	microb...	cor	pvalue	AdjPvalue	-log2(...	signif...	sign
PE(16:...	Tyzzer...	0.6169...	0.0037...	0.0096...	8.0559...	< 0.05	*
PE(16:...	[Rumin...	-0.472...	0.0370...	0.0699...	4.7527...	< 0.05	*
...	...	...	...	...	...	...	...

6.7.2 验证结果

将 Tab. 7 中的微生物在 Tab. 8 中搜索验证：

Table 9 (下方表格) 为表格 Liver gutMDisorder microbiota matched in PUBLISHED ChangesAndCorChen2021 概览。

(对应文件为 `Figure+Table/Liver-gutMDisorder-microbiota-matched-in-PUBLISHED-ChangesAndCorChen2021.xlsx`)

注：表格共有 104 行 8 列，以下预览的表格可能省略部分数据；表格含有 71 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 9: Liver gutMDisorder microbiota matched in PUBLISHED ChangesAndCorChen2021

metabo...	microb...	cor	pvalue	AdjPvalue	-log2(...	signif...	sign
PE(16:...	[Rumin...	-0.472...	0.0370...	0.0699...	4.7527...	< 0.05	*
PC(18:...	[Rumin...	0.5699...	0.0098...	0.0333...	6.6643...	< 0.05	*
PC(20:...	[Rumin...	0.7398...	0.0002...	0.0048...	11.751...	< 0.001	**
Tauroh...	[Rumin...	-0.8	2.8326...	0.0010...	15.107...	< 0.001	**
Tauroh...	Rumino...	0.4605...	0.0410...	0.1088...	4.6077...	< 0.05	*
trans-...	[Rumin...	0.7082...	0.0006...	0.0078...	10.522...	< 0.001	**
trans-...	Rumino...	-0.509...	0.0216...	0.0980...	5.5314...	< 0.05	*
L-Norl...	[Rumin...	0.5879...	0.0074...	0.0285...	7.0754...	< 0.05	*
L-Norl...	Rumino...	-0.456...	0.0427...	0.1088...	4.5465...	< 0.05	*
m-Coum...	[Rumin...	0.6390...	0.0030...	0.0148...	8.3672...	< 0.05	*
m-Coum...	Rumino...	-0.629...	0.0029...	0.0549...	8.4227...	< 0.05	*
Galact...	[Rumin...	0.7308...	0.0003...	0.0053...	11.378...	< 0.001	**
Hypoxa...	[Rumin...	-0.538...	0.0157...	0.0406...	5.9924...	< 0.05	*
L-Carn...	[Rumin...	-0.763...	0.0001...	0.0026...	12.864...	< 0.001	**
SM C16:1	[Rumin...	-0.562...	0.0110...	0.0335...	6.4991...	< 0.05	*
...	...	...	...	...	...	...	...

结果发现 Ruminococcus 这一微生物得到验证，属于胆结石 (gallstones, G) 的差异微生物。

Ruminococcus 向上对应：

Ruminococcus -> Leucine -> `chr17_47247224_A_G_b38` -> ITGB3

6.7.3 ITGB3、C9orf152 与 'Steroid biosynthesis' 通路的基因的关联性

(C9orf152 来源于 Tab. 5)

6.7.3.1 对应关系 (hgnc symbol 和 mgi symbol)

以下为这些基因的对应关系：

Table 10 (下方表格) 为表格 Mapping of ITGB3 and other genes from hgncSymbol to mgiSymbol 概览。

(对应文件为 `Figure+Table/Mapping-of-ITGB3-and-other-genes-from-hgncSymbol-to-mgiSymbol.csv`)

注：表格共有 13 行 11 列，以下预览的表格可能省略部分数据；表格含有 13 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 10: Mapping of ITGB3 and other genes from hgncSymbol to mgiSymbol

hgnc_s…	mgi_sy…	ensemb…	entrez…	descri…	logFC	AveExpr	t	P.Value	adj.P.Val
C9orf152	D63003…	ENSMUS…	242484	RIKEN …	0.8319…	3.6286…	4.4284…	0.0014…	0.0873…
ITGB3	Itgb3	ENSMUS…	16416	integr…	0.7621…	1.9664…	2.5048…	0.0324…	0.3350…
HSD17B7	Hsd17b7	ENSMUS…	15490	hydrox…	-1.949…	5.7472…	-7.173…	4.0766…	0.0170…
MSMO1	Msmo1	ENSMUS…	66234	methyl…	-4.130…	5.8215…	-6.801…	6.2586…	0.0210…
CYP51A1	Cyp51	ENSMUS…	13121	cytoch…	-2.839…	5.6728…	-5.878…	0.0001…	0.0351…
LSS	Lss	ENSMUS…	16987	lanost…	-1.865…	2.5679…	-5.855…	0.0002…	0.0351…
DHCR7	Dhcr7	ENSMUS…	13360	7-dehy…	-2.171…	4.9094…	-5.653…	0.0002…	0.0398…
DHCR24	Dhcr24	ENSMUS…	74754	24-deh…	-1.285…	9.3065…	-5.407…	0.0003…	0.0465…
TM7SF2	Tm7sf2	ENSMUS…	73166	transm…	-1.187…	6.0181…	-5.354…	0.0003…	0.0481…
EBP	Ebp	ENSMUS…	13595	phenyl…	-0.865…	7.0738…	-4.934…	0.0007…	0.0626…
NSDHL	Nsdhl	ENSMUS…	18194	NAD(P)…	-2.316…	4.5013…	-4.573…	0.0011…	0.0805…
SC5D	Sc5d	ENSMUS…	235293	sterol…	-0.953…	6.8768…	-4.064…	0.0025…	0.1117…

hgnc_s...	mgi_sy...	ensemb...	entrez...	descri...	logFC	AveExpr	t	P.Value	adj.P.Val
FDFT1	Fdft1	ENSMUS...	14137	farnes...	-4.499...	3.0868...	-3.658...	0.0048...	0.1480...

6.7.3.2 关联分析

Figure 8 (下方图) 为图 LIVER correlation heatmap 概览。

(对应文件为 `Figure+Table/LIVER-correlation-heatmap.pdf`)

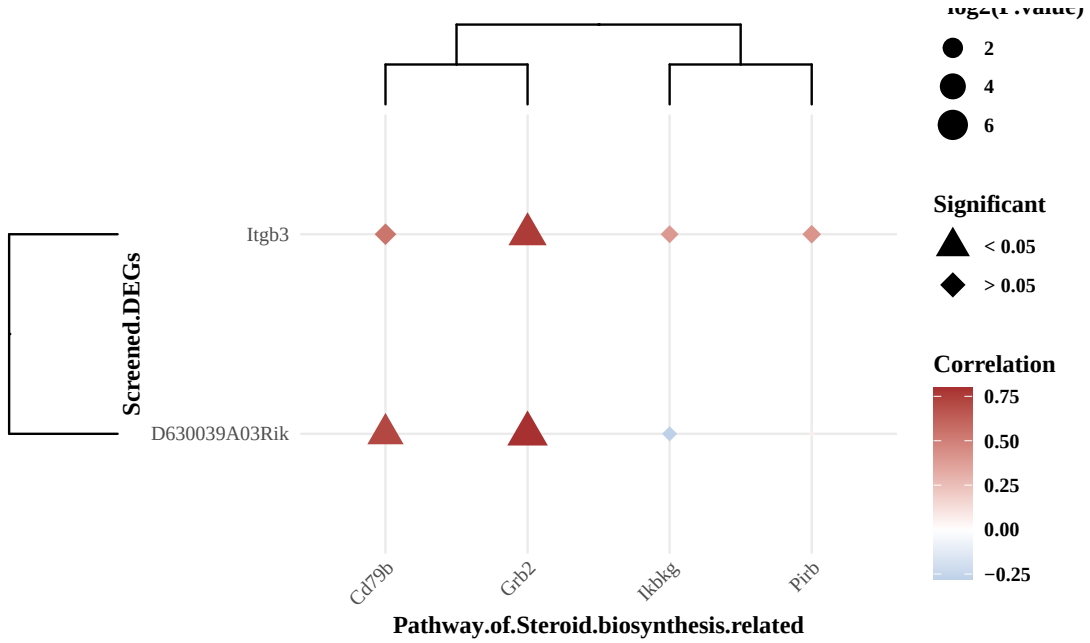

Figure 8: LIVER correlation heatmap

Table 11 (下方表格) 为表格 LIVER significant correlation 概览。

(对应文件为 `Figure+Table/LIVER-significant-correlation.csv`)

注：表格共有 16 行 7 列，以下预览的表格可能省略部分数据；表格含有 2 个唯一 'Screened.DEGs'。

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: LIVER significant correlation

Screened.DEGs	Pathway.of...	cor	pvalue	-log2(P.va...	significant	sign
Itgb3	Nsdhl	-0.68	0.0441	4.50307753...	< 0.05	*
D630039A03Rik	Nsdhl	-0.79	0.0107	6.54624539...	< 0.05	*

Screened.DEGs	Pathway.of...	cor	pvalue	-log2(P.va...	significant	sign
Itgb3	Cyp51	-0.71	0.0309	5.01624935...	< 0.05	*
D630039A03Rik	Cyp51	-0.81	0.0079	6.98393163...	< 0.05	*
Itgb3	Msmo1	-0.7	0.0375	4.73696559...	< 0.05	*
D630039A03Rik	Msmo1	-0.87	0.0022	8.82828076...	< 0.05	*
Itgb3	Sc5d	-0.78	0.0138	6.17918792...	< 0.05	*
Itgb3	Ebp	-0.71	0.0333	4.90833401...	< 0.05	*
D630039A03Rik	Ebp	-0.72	0.0296	5.07825901...	< 0.05	*
Itgb3	Tm7sf2	-0.82	0.0071	7.13796526...	< 0.05	*
D630039A03Rik	Tm7sf2	-0.67	0.0468	4.41734765...	< 0.05	*
Itgb3	Hsd17b7	-0.7	0.0373	4.74468055...	< 0.05	*
D630039A03Rik	Hsd17b7	-0.86	0.0027	8.53282487...	< 0.05	*
D630039A03Rik	Lss	-0.81	0.0083	6.91267294...	< 0.05	*
Itgb3	Dhcr24	-0.75	0.0197	5.66566056...	< 0.05	*
...	...	...	...	...	...	...

Reference

1. Chen, Y. *et al.* Changes and correlations of the intestinal flora and liver metabolite profiles in mice with gallstones. *Frontiers in physiology* **12**, (2021).

2. Liu, X. *et al.* Mendelian randomization analyses support causal relationships between blood metabolites and the gut microbiome. *Nature Genetics* **54**, (2022).

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. None, N. *et al.* The gtex consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).

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.