结肠炎和结肠癌的差异菌群

2024-03-18

LiChuang Huang

@ 立效研究院

Contents

List of Figures

List of Tables

1 摘要

需求：

结肠炎和结肠癌的差异菌

结果 (主要思路)：

- 获取结肠炎 (UC) 和结肠癌的差异菌分析数据。
- 以包含完整 logFC 和 p-value 数据为主数据 (其余作验证)，评估 Cancer vs UC，见 6.3.1 和 6.3.2。(因为 UC 的数据集包含多个国家的来源，因此这里也对各个国家都分析了一遍，但 Cancer 用的只是同一个数据集)。评估方式见 Fig. 2 下方注释。(**注意，为了应对不同数据集信息的不一致性，分析是以属 (genus) 为基本单位展开的**)。
- 对各个国家的结果取交集，获得 Cancer vs UC 的菌为上升 (Fig. 3) 的和下降 (Fig. 4) 的集。这些被整理于 Tab. 3。
- 尝试以更多数据集验证这些菌是否为 Cancer 或 UC 的差异菌 (前提)。最终结果见 Tab. 8

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 to TransplantationSinha2022[1].
- Supplementary file from article refer to TargetedSuppreFederi2022[2].
- Supplementary file from article refer to DepressionAndYuan2021[3].
- Supplementary file from article refer to AnIntegratedTRoelan2023[4].
- Supplementary file from article refer to LocationAndCoSambru2023[5].
- Supplementary file from article refer to FunctionalChanDaniel2017[6].

3.2 方法

Mainly used method:

- R version 4.3.2 (2023-10-31); 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 数据来源

6.1.1 结肠炎数据

6.1.1.1 TargetedSuppreFederi2022 结肠炎

'TargetedSuppreFederi2022 data' 数据已全部提供。

(对应文件为 `Figure+Table/TargetedSuppreFederi2022-data`)

> 注：文件夹 Figure+Table/TargetedSuppreFederi2022-data 共包含 4 个文件。
>
> 1. 1_Corrected France.csv
> 2. 2_Corrected Israel.csv
> 3. 3_Corrected US.csv
> 4. 4_Corrected Germany.csv

6.1.2 结肠癌数据

6.1.2.1 AnIntegratedTRoelan2023 结肠癌

Table 1 (下方表格) 为表格 AnIntegratedTRoelan2023 data 概览。

(对应文件为 `Figure+Table/AnIntegratedTRoelan2023-data.xlsx`)

> 注：表格共有 74 行 15 列，以下预览的表格可能省略部分数据；表格含有 1 个唯一 'sheet'。

Table 1: AnIntegratedTRoelan2023 data

sheet	Taxonomy	p_val	FDR	wilcox…	mean_N	mean_T	median_N	median_T	Direction
Supple…	D_0___B…	2.8875…	4.2447…	1965	0.0121…	0.0524…	0.0006…	0.0059…	Enrich…
Supple…	D_0___B…	6.6405…	4.8807…	657	0.0018…	0.0095…	0	0	Enrich…
Supple…	D_0___B…	7.1087…	3.4833…	21023.5	0.0190…	0.0126…	0.0128…	0.0067…	Enrich…
Supple…	D_0___B…	2.8160…	1.0348…	18604	0.0132…	0.0087…	0.0078…	0.0041…	Enrich…
Supple…	D_0___B…	1.7251…	5.0718…	11985	0.0070…	0.0043…	0.0034…	0.0019…	Enrich…
Supple…	D_0___B…	5.0924…	1.2476…	7227.5	0.0126…	0.0315…	0.0031…	0.0076…	Enrich…
Supple…	D_0___B…	9.2218…	1.9365…	223.5	0.0002…	0.0052…	0	0	Enrich…
Supple…	D_0___B…	1.5535…	2.8547…	16631.5	0.0106…	0.0076…	0.0064…	0.0041…	Enrich…
Supple…	D_0___B…	2.1475…	3.5075…	1494	0.0022…	0.0076…	0	0	Enrich…
Supple…	D_0___B…	4.3690…	6.2979…	62	0.0001…	0.0026…	0	0	Enrich…

sheet	Taxonomy	p_val	FDR	wilcox…	mean_N	mean_T	median_N	median_T	Direction
Supple…	D_0___B…	4.7127…	6.2979…	20341	0.0280…	0.0232…	0.0228…	0.0170…	Enrich…
Supple…	D_0___B…	7.9474…	9.7356…	237.5	0.0008…	0.0058…	0	0	Enrich…
Supple…	D_0___B…	1.6820…	1.9019…	4806.5	0.0040…	0.0026…	0	0	Enrich…
Supple…	D_0___B…	1.5690…	1.5376…	4009	0.0025…	0.0039…	0.0006…	0.0013…	Enrich…
Supple…	D_0___B…	1.5362…	1.5376…	12107.5	0.0075…	0.0059…	0.0017…	0.0012…	Enrich…
…	…	…	…	…	…	…	…	…	…

6.2 数据预处理

结肠炎 TargetedSuppreFederi2022 与结肠癌 AnIntegratedTRoelan2023 数据较为完整 (即，6.1.1.1, 6.1.2.1)，因此作为主要数据。

由于数据来源不同，格式不统一，需要根据微生物种属 (Taxonomy) 对信息补充或改动。

6.2.1 结肠炎

'Formated TargetedSuppreFederi2022' 数据已全部提供。

(对应文件为 `Figure+Table/formated-TargetedSuppreFederi2022`)

注：文件夹 Figure+Table/formated-TargetedSuppreFederi2022 共包含 4 个文件。

1. 1_Corrected France.csv
2. 2_Corrected Germany.csv
3. 3_Corrected Israel.csv
4. 4_Corrected US.csv

6.2.2 结肠癌

Table 2 (下方表格) 为表格 formated AnIntegratedTRoelan2023 概览。

(对应文件为 `Figure+Table/formated-AnIntegratedTRoelan2023.csv`)

注：表格共有 70 行 6 列，以下预览的表格可能省略部分数据；表格含有 70 个唯一 'Taxonomy'。

1. logFC: estimate of the log2-fold-change corresponding to the effect or contrast (for 'topTableF' there may be several columns of log-fold-changes)

Table 2: Formated AnIntegratedTRoelan2023

Taxonomy	Log2.Fold….	FDR	logFC	genus	taxon
D_0___Bacte…	2.11202492…	4.24476048…	2.11202492…	Fusobacterium	d___Bacteri…
D_0___Bacte…	2.34581040…	4.88079358…	2.34581040…	Campylobacter	d___Bacteri…

Taxonomy	Log2.Fold....	FDR	logFC	genus	taxon
D_0___Bacte...	-0.5903992...	3.48331113...	-0.5903992...	Parabacter...	d___Bacteri...
D_0___Bacte...	-0.6046987...	1.03488157...	-0.6046987...	Alistipes	d___Bacteri...
D_0___Bacte...	-0.7020994...	5.07188395...	-0.7020994...	Phascolarc...	d___Bacteri...
D_0___Bacte...	1.32218708...	1.24765615...	1.32218708...	Streptococcus	d___Bacteri...
D_0___Bacte...	4.25806815...	1.93658213...	4.25806815...	Leptotrichia	d___Bacteri...
D_0___Bacte...	-0.4918291...	2.85471118...	-0.4918291...	Fusicateni...	d___Bacteri...
D_0___Bacte...	1.78225934...	3.50758843...	1.78225934...	Gemella	d___Bacteri...
D_0___Bacte...	4.58804189...	6.29791077...	4.58804189...	Selenomonas	d___Bacteri...
D_0___Bacte...	-0.2729061...	6.29791077...	-0.2729061...	Blautia	d___Bacteri...
D_0___Bacte...	2.79144013...	9.73561642...	2.79144013...	Selenomonas	d___Bacteri...
D_0___Bacte...	0.63148802...	1.53763848...	0.63148802...	Lachnospir...	d___Bacteri...
D_0___Bacte...	-0.3387620...	1.53763848...	-0.3387620...	Barnesiella	d___Bacteri...
D_0___Bacte...	-0.6863772...	1.87594484...	-0.6863772...	Paraprevot...	d___Bacteri...
...	...	...	...	...	...

6.3 结肠炎与结肠癌差异菌比较

6.3.1 US (示例)

Figure 1 (下方图) 为图 US change detail 概览。

(对应文件为 `Figure+Table/US-change-detail.pdf`)

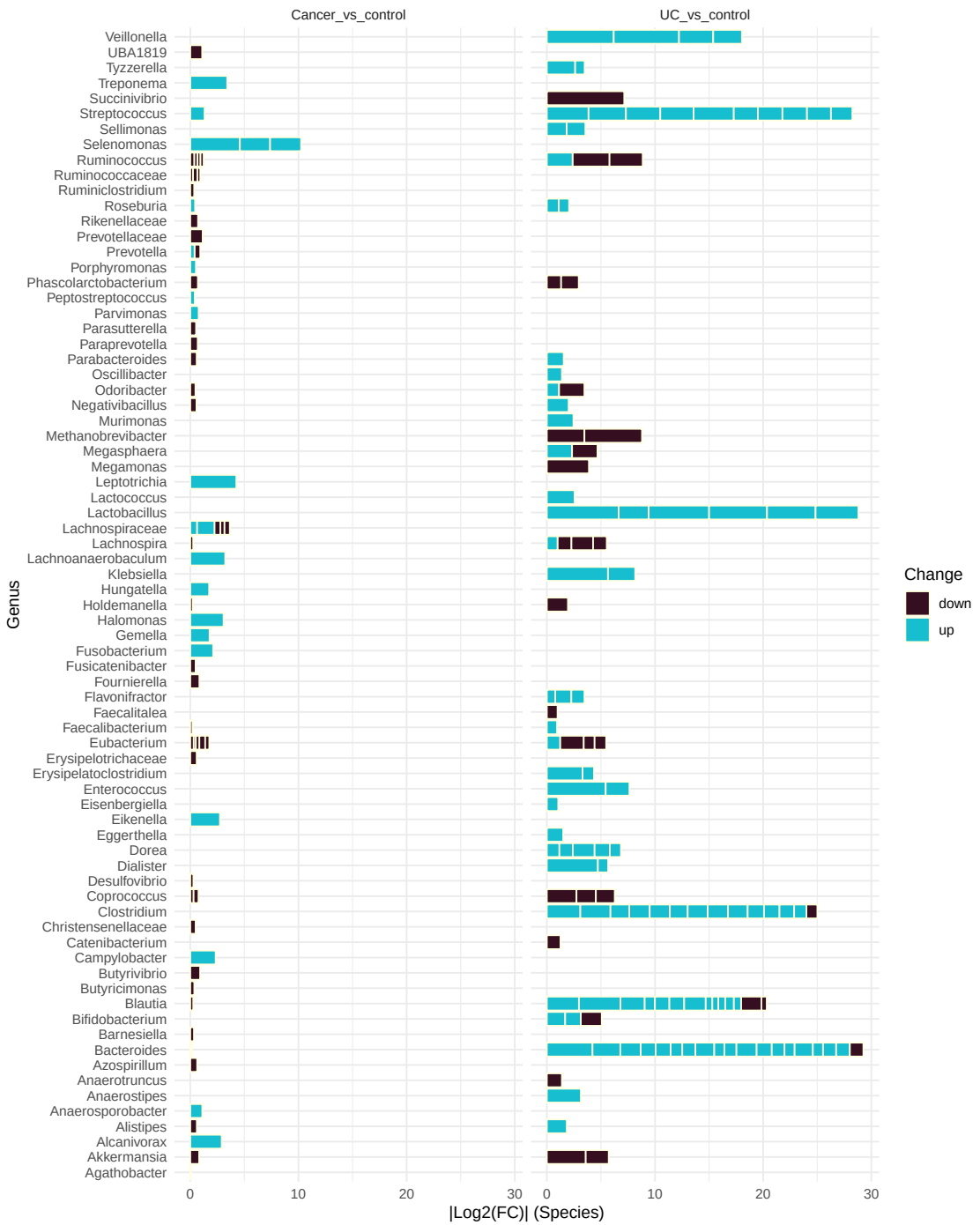

Figure 1: US change detail

Figure 2 (下方图) 为图 US change summary 概览。

(对应文件为 `Figure+Table/US-change-summary.pdf`)

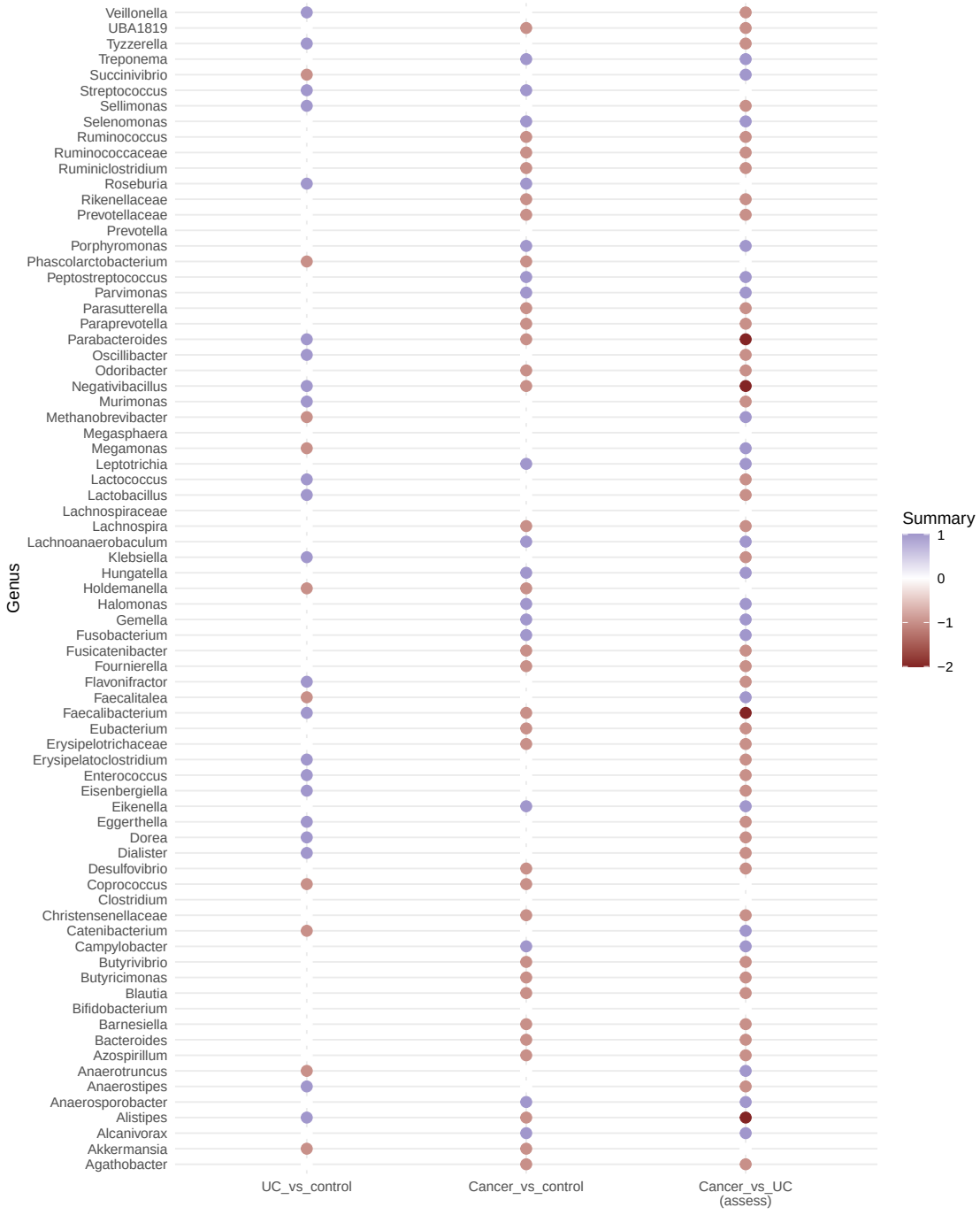

Figure 2: US change summary

说明:

- value 1 或 2, 代表 log2(FC) > 0, 该属 (genus) 包含差异菌为丰度升高的。
- value 0, 包含 log2(FC) > 0 或 log2(FC) < 0, 但该属 (genus) 整体不确定的 (因为不利于 Cancer vs UC 的推断)。
- value -1, 或 -2, 代表 log2(FC) < 0, 该属 (genus) 包含差异菌为丰度下降的。

6.3.2 其他

'Change detail' 数据已全部提供。

(对应文件为 `Figure+Table/Change-detail`)

> 注：文件夹 Figure+Table/Change-detail 共包含 4 个文件。
>
> 1. 1_Corrected France.pdf
> 2. 2_Corrected Germany.pdf
> 3. 3_Corrected Israel.pdf
> 4. 4_Corrected US.pdf

'Change summary' 数据已全部提供。

(对应文件为 `Figure+Table/Change-summary`)

> 注：文件夹 Figure+Table/Change-summary 共包含 4 个文件。
>
> 1. 1_Corrected France.pdf
> 2. 2_Corrected Germany.pdf
> 3. 3_Corrected Israel.pdf
> 4. 4_Corrected US.pdf

6.3.3 数据集的汇总

Figure 3 (下方图) 为图 UpSets Up 概览。

(对应文件为 `Figure+Table/UpSets-Up.pdf`)

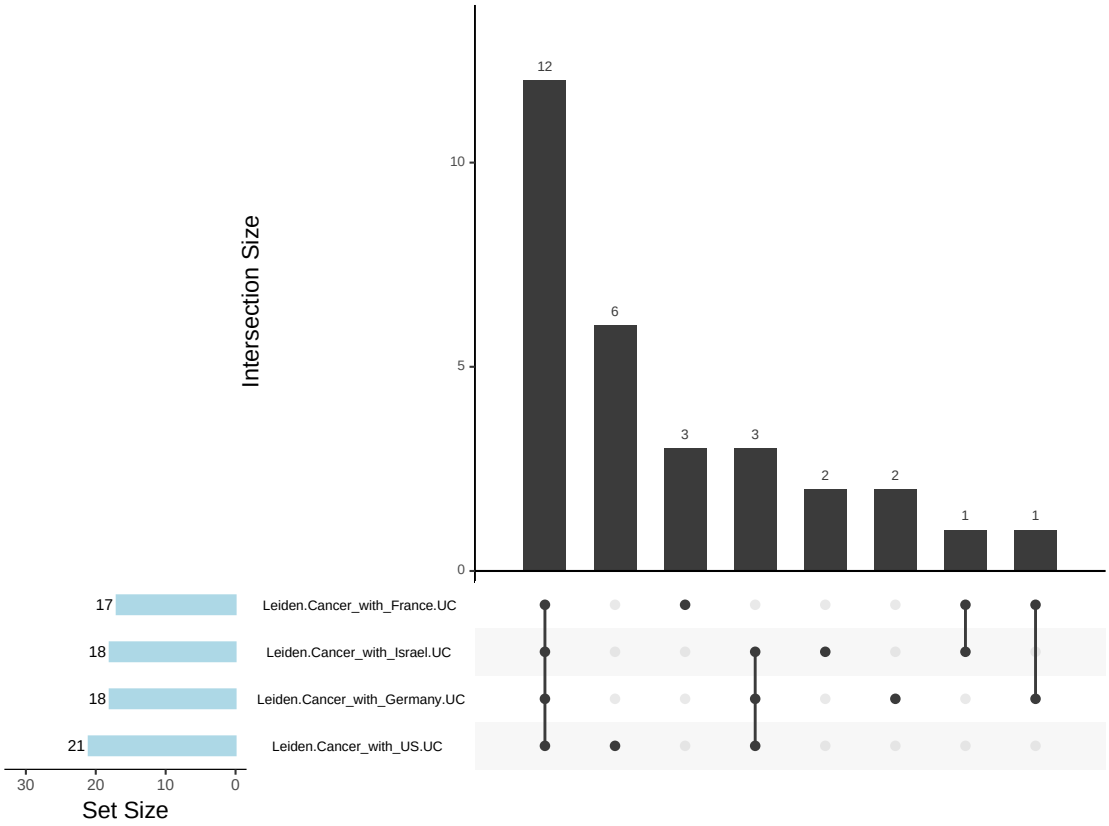

Figure 3: UpSets Up

All_intersection：

Alcanivorax, Anaerosporobacter, Campylobacter, Eikenella, Gemella, Halomonas, Hungatella, Lachnoanaerobaculum, Leptotrichia, Porphyromonas, Selenomonas, Treponema

(上述信息框内容已保存至 `Figure+Table/UpSets-Up-content`)

Figure 4 (下方图) 为图 UpSets down 概览。

(对应文件为 `Figure+Table/UpSets-down.pdf`)

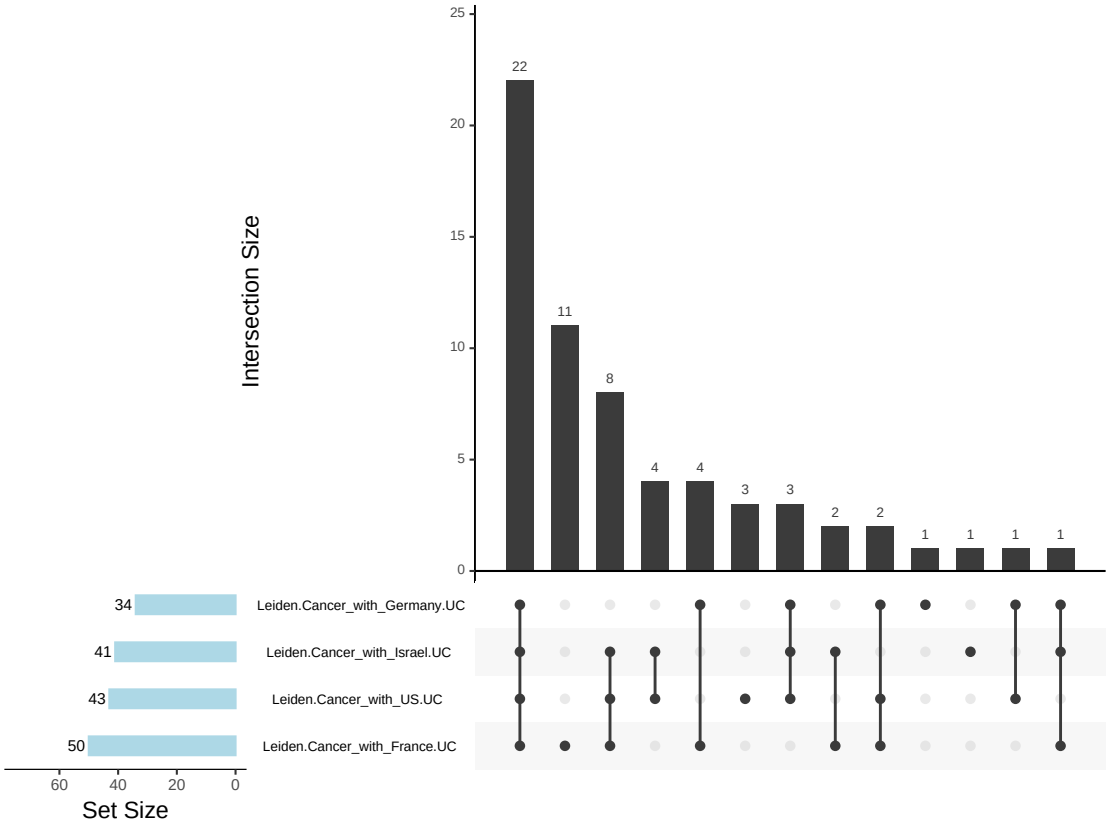

Figure 4: UpSets down

All_intersection：

Agathobacter, Azospirillum, Bacteroides, Blautia, Butyricimonas, Christensenellaceae, Desulfovibrio, Erysipelotrichaceae, Eubacterium, Flavonifractor, Fournierella, Fusicatenibacter, Klebsiella, Negativibacillus, Parabacteroides, Parasutterella, Prevotellaceae, Rikenellaceae, Ruminiclostridium, R...

（上述信息框内容已保存至 `Figure+Table/UpSets-down-content`）

Table 3 (下方表格) 为表格 All changed microbiota genus 概览。

（对应文件为 `Figure+Table/All-changed-microbiota-genus`）

注：表格共有 34 行 2 列，以下预览的表格可能省略部分数据；表格含有 34 个唯一 'name'。

Table 3: All changed microbiota genus

name	type
Agathobacter	Down
Azospirillum	Down

name	type
Bacteroides	Down
Blautia	Down
Butyricimonas	Down
Christensenellaceae	Down
Desulfovibrio	Down
Erysipelotrichaceae	Down
Eubacterium	Down
Flavonifractor	Down
Fournierella	Down
Fusicatenibacter	Down
Klebsiella	Down
Negativibacillus	Down
Parabacteroides	Down
…	…

6.4 在更多数据集验证

6.4.1 数据来源

6.4.1.1 DepressionAndYuan2021 结肠炎

Table 4 (下方表格) 为表格 DepressionAndYuan2021 data 概览。

(对应文件为 `Figure+Table/DepressionAndYuan2021-data.xlsx`)

> 注：表格共有 91 行 4 列，以下预览的表格可能省略部分数据；表格含有 91 个唯一 'Taxonomy'。

Table 4: DepressionAndYuan2021 data

Taxonomy	p.value	MRA.in.UC	MRA.in.HC
p___Gemmatimonadetes	0.00049553252446827	0.00171119907683608	0.000160190585722501
p___Actinobacteria	0.00834784591709094	0.0617033057472256	0.0389112516772091
p___Firmicutes	0.00478686441960355	0.589610871565727	0.499806265231797
p___Bacteroidetes	9.60337999115942e-05	0.250752809840626	0.403687121772228
p___unidentified	1.40986578996555e-07	0.000640301130981834	2.05372545798078e-06
c___Actinobacteria	0.00353055167754535	0.046725454341533	0.0237116295626934
c___Longimicrobia	7.16632518063251e-07	0.00145496100157866	0
c___Bacteroidia	6.38198131772293e-05	0.24518920382048	0.403287329883075
c___Deinococci	3.17450175630301e-06	0.00131123215732429	0
c___Bacilli	0.00330587139663399	0.0649165901033199	0.0154563377967633
c___Cytophagia	2.10636789003367e-05	0.0016664310495747	4.24436594649361e-05

Taxonomy	p.value	MRA.in.UC	MRA.in.HC
c___unidentified	2.64154914758447e-08	0.0016652538833675	8.83101946931736e-05
c___Flavobacteriia	5.26714172360768e-05	0.00167703418517471	5.95580382814426e-05
o___Oceanospirillales	0.00495867783478051	0.000461817587592187	2.05372545798077e-06
o___Bifidobacteriales	0.00354779153342515	0.0443044462799416	0.0224465346805772
...	...	...	...

6.4.1.2 TransplantationSinha2022 结肠炎

Table 5 (下方表格) 为表格 TransplantationSinha2022 data 概览。

(对应文件为 `Figure+Table/TransplantationSinha2022-data.csv`)

> 注：表格共有 141 行 12 列，以下预览的表格可能省略部分数据；表格含有 83 个唯一 'W'。

Table 5: TransplantationSinha2022 data

W	detect……2	detect……3	detect……4	detect……5	Kingdom	Phylum	Class	Order	Family	...
121	FALSE	TRUE	TRUE	TRUE	d___Bac...	Bacter...	Bacter...	Bacter...	Rikene...	...
89	FALSE	FALSE	FALSE	TRUE	d___Bac...	Bacter...	Bacter...	Bacter...	Rikene...	...
87	FALSE	FALSE	FALSE	TRUE	d___Bac...	Bacter...	Bacter...	Bacter...	Marini...	...
94	FALSE	FALSE	FALSE	TRUE	d___Bac...	Bacter...	Bacter...	Bacter...	Barnes...	...
125	FALSE	TRUE	TRUE	TRUE	d___Bac...	Bacter...	Bacter...	Bacter...	Barnes...	...
94	FALSE	FALSE	FALSE	TRUE	d___Bac...	Bacter...	Bacter...	Bacter...	Barnes...	...
95	FALSE	FALSE	FALSE	TRUE	d___Bac...	Bacter...	Bacter...	Bacter...	Barnes...	...
135	TRUE	TRUE	TRUE	TRUE	d___Bac...	Bacter...	Bacter...	Bacter...	Prevot...	...
138	TRUE	TRUE	TRUE	TRUE	d___Bac...	Bacter...	Bacter...	Bacter...	Prevot...	...
103	FALSE	FALSE	TRUE	TRUE	d___Bac...	Firmic...	Bacilli	Lactob...	Entero...	...
134	TRUE	TRUE	TRUE	TRUE	d___Bac...	Firmic...	Bacilli	Erysip...	Erysip...	...
118	FALSE	TRUE	TRUE	TRUE	d___Bac...	Firmic...	Bacilli	Erysip...	Erysip...	...
90	FALSE	FALSE	FALSE	TRUE	d___Bac...	Firmic...	Bacilli	Erysip...	Erysip...	...
87	FALSE	FALSE	FALSE	TRUE	d___Bac...	Firmic...	Bacilli	Erysip...	Erysip...	...
95	FALSE	FALSE	FALSE	TRUE	d___Bac...	Firmic...	Bacilli	Erysip...	Erysip...	...
...	...	...	...	...	...	...	...	...	...	...

6.4.1.3 LocationAndCoSambru2023 结肠癌

Table 6 (下方表格) 为表格 LocationAndCoSambru2023 data 概览。

(对应文件为 `Figure+Table/LocationAndCoSambru2023-data.csv`)

> 注: 表格共有 44 行 2 列, 以下预览的表格可能省略部分数据; 表格含有 44 个唯一 'tax_id'。

Table 6: LocationAndCoSambru2023 data

tax_id	taxon_name
40545	Sutterella_wadsworthensis
214856	Alistipes_finegoldii
328814	Alistipes_shahii
674529	Bacteroides_faecis
333367	[Clostridium]_asparagiforme
437898	Sutterella_parvirubra
74426	Collinsella_aerofaciens
1531	[Clostridium]_clostridioforme
239935	Akkermansia_muciniphila
901	Desulfovibrio_piger
1892897	Shigella_sp._FC569
68259	Streptomyces_purpurogeneisc…
1450439	Bacteroides_sp._UW
585543	Bacteroides_sp._D20
1581131	Actinomyces_sp._HMSC08A01
…	…

6.4.1.4 FunctionalChanDaniel2017 结肠癌

Table 7 (下方表格) 为表格 FunctionalChanDaniel2017 data 概览。

(对应文件为 `Figure+Table/FunctionalChanDaniel2017-data.xlsx`)

> 注: 表格共有 9 行 3 列, 以下预览的表格可能省略部分数据; 表格含有 9 个唯一 'Name(s)'。

Table 7: FunctionalChanDaniel2017 data

Name(s)	Relationship	genus
Citrobacter rodentium	Min mice inoculated with th…	Citrobacter
Enterococcus faecalis	Produces superoxide and hyd…	Enterococcus
Clostridium cluster XVIa (C…	Can produce secondary bile …	Clostridium
Acidovorax species	Associated with increased r…	Acidovorax
Enterotoxigenic Bacteroides…	Produces a toxin that cause…	Enterotoxigenic
Streptococcus gallolyticus	Present in approximately 20…	Streptococcus
Escherichia coli NC101	Produces genotoxic colibact…	Escherichia
Fusobacterium nucleatum	Induces hyperproliferation …	Fusobacterium
Akkermansia muciniphila	Mucin-degrading species wer…	Akkermansia

6.4.2 结果

Table 8 (下方表格) 为表格 change validated 概览。

(对应文件为 `Figure+Table/change-validated.csv`)

> 注: 表格共有 34 行 7 列, 以下预览的表格可能省略部分数据; 表格含有 34 个唯一 'name'。

Table 8: Change validated

name	type	Other_data…	Depression…	Transplant…	LocationAn…	Functional…
Agathobacter	Down	0	FALSE	FALSE	FALSE	FALSE
Azospirillum	Down	0	FALSE	FALSE	FALSE	FALSE
Bacteroides	Down	2	FALSE	TRUE	TRUE	FALSE
Blautia	Down	2	TRUE	TRUE	FALSE	FALSE
Butyricimonas	Down	2	TRUE	TRUE	FALSE	FALSE
Christense…	Down	1	FALSE	TRUE	FALSE	FALSE
Desulfovibrio	Down	3	TRUE	TRUE	TRUE	FALSE
Erysipelot…	Down	0	FALSE	FALSE	FALSE	FALSE
Eubacterium	Down	3	TRUE	TRUE	TRUE	FALSE
Flavonifra…	Down	2	TRUE	TRUE	FALSE	FALSE
Fournierella	Down	0	FALSE	FALSE	FALSE	FALSE
Fusicateni…	Down	1	TRUE	FALSE	FALSE	FALSE
Klebsiella	Down	0	FALSE	FALSE	FALSE	FALSE
Negativiba…	Down	1	FALSE	TRUE	FALSE	FALSE
Parabacter…	Down	2	FALSE	TRUE	TRUE	FALSE
…	…	…	…	…	…	…

Reference

1. Sinha, A. *et al.* Transplantation of bacteriophages from ulcerative colitis patients shifts the gut bacteriome and exacerbates the severity of dss colitis. *Microbiome* **10**, (2022).

2. Federici, S. *et al.* Targeted suppression of human ibd-associated gut microbiota commensals by phage consortia for treatment of intestinal inflammation. *Cell* **185**, 2879–2898.e24 (2022).

3. Yuan, X. *et al.* Depression and anxiety in patients with active ulcerative colitis: Crosstalk of gut microbiota, metabolomics and proteomics. *Gut microbes* **13**, (2021).

4. Roelands, J. *et al.* An integrated tumor, immune and microbiome atlas of colon cancer. *Nature medicine* **29**, 1273–1286 (2023).

5. Sambruni, G. *et al.* Location and condition based reconstruction of colon cancer microbiome from human rna sequencing data. *Genome medicine* **15**, (2023).

6. Daniel, S. G., Ball, C. L., Besselsen, D. G., Doetschman, T. & Hurwitz, B. L. Functional changes in the gut microbiome contribute to transforming growth factor β-deficient colon cancer. *mSystems* **2**, e00065–e00017 (2017).