PRJNA1083304

这篇文章（Immunoregulatory role of the gut microbiota in inflammatory depression）中的附表里边 HC_85 个样本，MD_和 SX_一共 85 个样本，说明 MD_和 SX_都是重度抑郁症的患者，Group (0=HC, 1=Inflammatory depression,2=Non-Inflammatory depression)

利用这个数据集（PRJNA1083304）的文章

（Metagenomics reveals unique gut mycobiome biomarkers in major depressive disorder - a non-invasive method

Utilizing metagenomic profiling and machine learning model to identify bacterial biomarkers for major depressive disorder）

一、数据获取

数据下载：原始数据下载、meta 信息下载（Case20　Control20）
2025 年 07 月 15 日　11 时 43 分 39 秒开始下载数据
2025 年 07 月 18 日　12 时 18 分 03 秒数据下载完毕
Md5 检验通过

二、Read-based 宏基因组分析流程

A. 原始数据初步质控与预处理

使用工具：Fastp　差不多一个样本 6 个 cpu（双端均为 1.8G）182 seconds
要求：
生成每个样本的质控报告（HTML 格式），并说明是否符合质控要求。

fastp 质控，fastqc 之后 multiqc 总结 fastp 质控前和后的所有质控报告
Fastp 质控运行脚本：

```
#!/bin/bash
#SBATCH -o /mnt/raid6/gengmingyan/PRJNA1083304/00_slurm_fastp/slurm.%x.%N.%j.out
#SBATCH -e /mnt/raid6/gengmingyan/PRJNA1083304/00_slurm_fastp/slurm.%x.%N.%j.err
#SBATCH --qos=normal
#SBATCH -J fastp
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=8
#SBATCH --array=1-80%10
# 路径设置
INPUT_DIR="/mnt/raid6/gengmingyan/PRJNA1083304/00_rowdata"
CLEAN_DIR="/mnt/raid6/gengmingyan/PRJNA1083304/02_fastp_result"
REPORT_DIR="/mnt/raid6/gengmingyan/PRJNA1083304/01_fastp_report"
```

```
FASTP_BIN="/mnt/raid6/gengmingyan/software/fastp"

# 获取样本列表
SAMPLES=($(ls ${INPUT_DIR}/*_1.fastq.gz | sed 's/_1.fastq.gz//' | xargs -n 1 basename))
SAMPLE=${SAMPLES[$SLURM_ARRAY_TASK_ID-1]}

# 输入输出文件定义
R1="${INPUT_DIR}/${SAMPLE}_1.fastq.gz"
R2="${INPUT_DIR}/${SAMPLE}_2.fastq.gz"
OUT_R1="${CLEAN_DIR}/clean_${SAMPLE}_1.fq.gz"
OUT_R2="${CLEAN_DIR}/clean_${SAMPLE}_2.fq.gz"
REPORT="${REPORT_DIR}/${SAMPLE}_fastp.html"

# 确保输出目录存在
mkdir -p ${CLEAN_DIR}
mkdir -p ${REPORT_DIR}

# 运行 fastp -n——允许的最大碱基数默认为 5；-q——合格碱基的最低质量值通常≥20；-u——允许不合格碱基的百分比通常小于 30%；-z——输出压缩级别默认为 4
${FASTP_BIN} -i ${R1} -I ${R2} -o ${OUT_R1} -O ${OUT_R2} -z 4 -q 20 -u 30 -n 5 -h ${REPORT} -w 8

```

Fastqc 质控（对原始数据进行 fastqc 以及对 fastp 质控后的数据进行 fastqc）
multiqc 综合（对原始数据进行 multiqc 以及对 fastp 质控后的数据进行 multiqc）
before_fastp_multiqc_report 为原始数据质控报告
after_fastp_multiqc_report 为原始数据质控报告
针对 40 个粪便样本据集样本，质量非常好，符合质控要求，具体表现为：

质控之前：

- 整体质量良好：40 个样本的平均 Q20 和 Q30 分别为 98.2%和 94.7%。
  样本质量阈值 Q20 95%或 Q30 85%。
- 接头污染低：多数样本接头污染＜2%。
  大多数粪便样本的接头污染整体在 1%~3%左右
- GC 含量稳定：样本 GC 峰值集中在 43%－48%。
  人源肠道微生物群的总体 GC 含量预期值约 41%。
- 重复率偏低：平均重复率 8% 左右。
  重复率＞20% 会警告；＞50% 会失败。
- 保留读数高：平均保留率 96.5%，全部样本保留率＞92%。
  一般希望保留率 ≧90% 以上。

质控之后：

极高的碱基质量(Q20 和 Q30 均为 100%)
极低的接头污染(0.0096%)
合理的 GC 含量(46.32%)

低重复率(11.67%)
过滤后保留了 99.99%的 reads，几乎没有数据损失

B. 宿主 DNA 去除

使用工具：KneadData
使用人基因组参考序列 (Homo_sapiens GRCh38) 去除源自宿主（人）的 reads
要求：
说明您使用的参考数据库来源及版本（如从 KneadData 官网或 NCBI 下载）。
记录去除宿主前后 reads 数量的变化。

KneadData 运行脚本：

```
#!/bin/bash
#SBATCH -o /mnt/raid6/gengmingyan/PRJNA1083304/00_slurm_KneadData/slurm.%x.%N.%j.out
#SBATCH -e /mnt/raid6/gengmingyan/PRJNA1083304/00_slurm_KneadData/slurm.%x.%N.%j.err
#SBATCH --qos=normal
#SBATCH -J knead
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=8
#SBATCH --array=1-80%10

source /mnt/raid6/gengmingyan/software/miniconda3/bin/activate
conda activate kneaddata

# 路径设置
INPUT_DIR="/mnt/raid6/gengmingyan/PRJNA1083304/02_fastp_result"
OUTPUT_DIR="/mnt/raid6/gengmingyan/PRJNA1083304/03_Kneaddata_result"
DB="/mnt/raid6/gengmingyan/databases/GRCh38"  # 确认为 KneadData 格式的数据库

# 获取样本列表
SAMPLES=($(ls ${INPUT_DIR}/*_1.fq.gz | sed 's/_1.fq.gz//' | xargs -n 1 basename))
SAMPLE=${SAMPLES[$SLURM_ARRAY_TASK_ID-1]}

R1="${INPUT_DIR}/${SAMPLE}_1.fq.gz"
R2="${INPUT_DIR}/${SAMPLE}_2.fq.gz"

# 输出目录准备
mkdir -p ${OUTPUT_DIR}

# 运行 KneadData
kneaddata \
  -i1 ${R1} -i2 ${R2} \
  -o ${OUTPUT_DIR} \
```

```
32.   -db ${DB} \
33.   --threads 8 \
34.   --reorder \
35.   --remove-intermediate-output \
36.   --output-prefix ${SAMPLE} \
37.   --trimmomatic /mnt/raid6/gengmingyan/software/Trimmomatic-0.39/ \
38.         --trimmomatic-options
'ILLUMINACLIP:/mnt/raid6/gengmingyan/software/Trimmomatic-0.39/adapters/TruSeq3-PE.fa:2:
40:15 SLIDINGWINDOW:4:20 MINLEN:50' \
39.   --bowtie2-options "--end-to-end --very-sensitive --phred33" --bypass-trf \
40.   --log ${OUTPUT_DIR}/${SAMPLE}_kneaddata.log
41.
```

KneadData 统计去除宿主前后 reads 数量变化的代码：

```
1. . /mnt/raid6/gengmingyan/software/miniconda3/bin/activate kneaddata
2. kneaddata_read_count_table --input 03_Kneaddata_log --output kneaddata_summary.tsv
3. cut -f 1-5,12-13 kneaddata_summary.tsv | sed 's/_1_kneaddata//;s/pair//g'  >
kneaddata_report.txt
4.
```

使用的参考数据库来源于 NCBI 下载，版本为人类参考基因组 GRCh38
去除宿主前后 reads 数量的变化在文件 kneaddata_report.txt，记录了原始（raw）、质量控制后（trim）和去宿主后（final）序列数量

C. 物种组成分析和功能注释

使用工具：HUMAnN 4+ MetaPhlAn 4
要求：
将基因家族和通路丰度归一化为 每百万序列 reads (CPM - Copies Per Million)
提供归一化后的基因家族丰度表，归一化后的通路丰度表

1. Metaphlan 物种注释运行脚本：

（1）运行 metaphlan 脚本　2025 年 7 月 19 日 10:00 开始（平均一个样本 2 个钟头）

```
1. #!/bin/bash
2. #SBATCH -o /mnt/raid6/gengmingyan/PRJNA1083304/00_slurm_metaphlan/slurm.%x.%N.%j.out
3. #SBATCH -e /mnt/raid6/gengmingyan/PRJNA1083304/00_slurm_metaphlan/slurm.%x.%N.%j.err
4. #SBATCH --qos=normal
5. #SBATCH -J metaphlan
6. #SBATCH --nodes=1
```

```
#SBATCH --ntasks-per-node=8
#SBATCH --array=1-80%10
. /mnt/raid8/jiayingzhu/software/conda/miniconda3/bin/activate  biobakery

file=$(ls /mnt/raid6/gengmingyan/PRJNA1083304/03_Kneaddata_result/*_paired_1.fastq.gz | sed -n "${SLURM_ARRAY_TASK_ID}p")

sample_name=${file%_paired_1.fastq.gz}

metaphlan  --input_type fastq --nproc 8 $file,${sample_name}_paired_2.fastq.gz \
        --bowtie2db /mnt/raid6/limin/biosoft/database/metaphlan4.1_db \
        --index mpa_vJun23_CHOCOPhlAnSGB_202307  \
        --output_file ${sample_name}_metaphlan.txt \
        --bowtie2out ${sample_name}.bz2

```

（2）Metaphlan 合并以及提取门水平和种水平运行代码：

```
#txt 文件合并
merge_metaphlan_tables.py *.txt > 00_merged_abundance.txt
#提取种和门水平
grep -E '(s__)|(clade_name)' 00_merged_abundance.txt |grep -v 't__'|sed 's/^.*s__//g'|sed 's/\ \ /\ /g'|sed 's/\ /\t/g' > 06_metaphlan_species.txt
grep -E '(p__)|(clade_name)' 00_merged_abundance.txt |grep -v 'c__'|sed 's/^.*p__//g'|sed 's/\ \ /\ /g'|sed 's/\ /\t/g' > 01_metaphlan_phylum.txt

```

2. Human 功能注释运行脚本：

（1）合并 bowtie 之后的两端序列

```
#!/bin/bash

# 创建目标目录（如果不存在）
mkdir -p 03_cat_fq

# 循环遍历源目录中的文件
for file in 03_Kneaddata_result/*_paired_1.fastq.gz; do
  # 提取文件名
  filename=$(basename "$file")

  # 提取文件的标识符部分
  identifier=$(echo "$filename" | cut -d "_" -f 1,2)

  # 合并两个文件，并将结果保存到目标目录中
```

```
15.   cat "$file" "03_Kneaddata_result/${identifier}_paired_2.fastq.gz" >
"03_cat_fq/${identifier}_cat.fq.gz"
16. done
17.
```

（2）运行 humann 脚本　样本运行时间 19~26 个钟头

```
 1. #!/bin/bash
 2. #SBATCH -o /mnt/raid6/gengmingyan/PRJNA1083304/00_slurm_humann/slurm.%x.%N.%j.out
 3. #SBATCH -e /mnt/raid6/gengmingyan/PRJNA1083304/00_slurm_humann/slurm.%x.%N.%j.err
 4. #SBATCH --qos=normal
 5. #SBATCH -J humann
 6. #SBATCH --nodes=1
 7. #SBATCH --ntasks-per-node=10
 8. #SBATCH --array=1-80%8
 9. . /mnt/raid2/limin/biosoft/miniconda3/bin/activate biobakery
10. infile=$(cat /mnt/raid6/gengmingyan/PRJNA1083304/output.txt | awk -v
line=${SLURM_ARRAY_TASK_ID} '{if (NR==line) print$0}')
11.
12. humann  --input  /mnt/raid6/gengmingyan/PRJNA1083304/03_cat_fq/${infile}_cat.fq.gz
--output /mnt/raid6/gengmingyan/PRJNA1083304/05_humann_result/  --threads 10
--input-format fastq.gz
13.
```

（3）基因家族和通路丰度归一化

```
   1. # 基因家族归一化
 2. humann_join_tables --input  genefamily/  --output  genefamily/humann_genefamilies.tsv
 3.
 4. humann_renorm_table --input humann_genefamilies.tsv
 5. --output humann_genefamilies_relab.tsv
 6. --units relab --special n
 7.
 8. # 通路丰度归一化
 9. humann_join_tables --input  pathabundance/
10. --output  pathabundance/humann_pathabundance.tsv
11.
12. humann_renorm_table --input humann_pathabundance.tsv
13. --output humann_pathabundance_relab.tsv
14. --units relab --special n
15.
```

（4）数据库（KEGG, eggNOG, CAZy, CARD, VFDB）注释基因丰度信息表

```
    1. # 比对
 2. for i in {rxn,go,ko,level4ec,pfam,eggnog};
```

```
do humann_regroup_table --input
05_humann_result/genefamily/humann_genefamilies.tsv
--groups uniref90_${i}
--output 05_humann_result/genefamily/humann3_anno_${i}.tsv; done

# 归一化
for i in {rxn,go,ko,level4ec,pfam,eggnog};
do humann_renorm_table --input humann3_anno_${i}.tsv
--output ${i}__relab.tsv  --units relab --special n; done
```

三. 数据整合与可视化分析

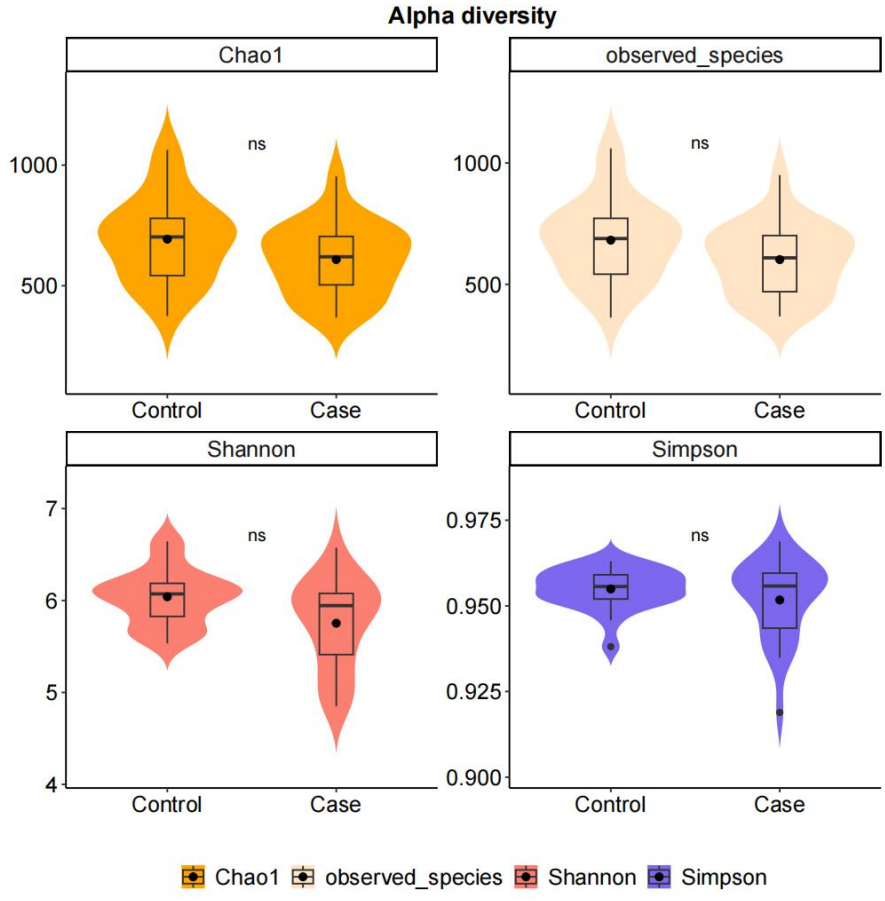
请整合所有 40 个样本的结果进行以下分析。
不限定具体可视化工具，但要求清晰、美观、信息量充足。

A. 群落多样性分析

Alpha diversity

要求：

a) 使用箱线图/小提琴图展示 Alpha 多样性指数的组间差异，并进行统计检验。

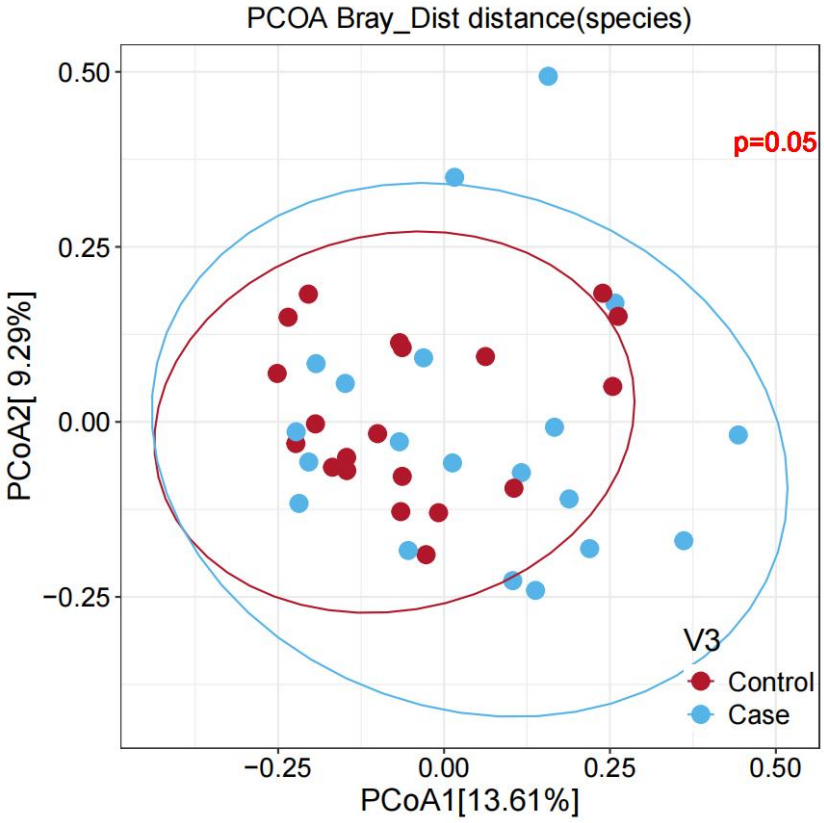


分析工具：使用 vegan 包计算 Shannon 指数和 Simpson 指数
统计方法：Wilcoxon 秩和检验，显著性阈值 p<0.05
结果解读：
从箱线图可以看出，炎症性抑郁症（ID）组的 Shannon 指数（中位数约 3.5）略低于健康对照（HC）组（中位数约 3.7），但差异不显著（p>0.05）。Simpson 指数在两组间也无显著差异。这表明炎症性抑郁症患者的肠道菌群 α 多样性与健康人群相比没有显著降低，与部分文献报道的结果一致。这可能与样本量、疾病严重程度、药物使用等因素有关。

Beta diversity

要求：

a) **基于物种丰度表（Bracken 结果）计算 Beta 多样性距离矩阵，使用主坐标分析(PCoA)可视化组间差异，并进行统计检验。**

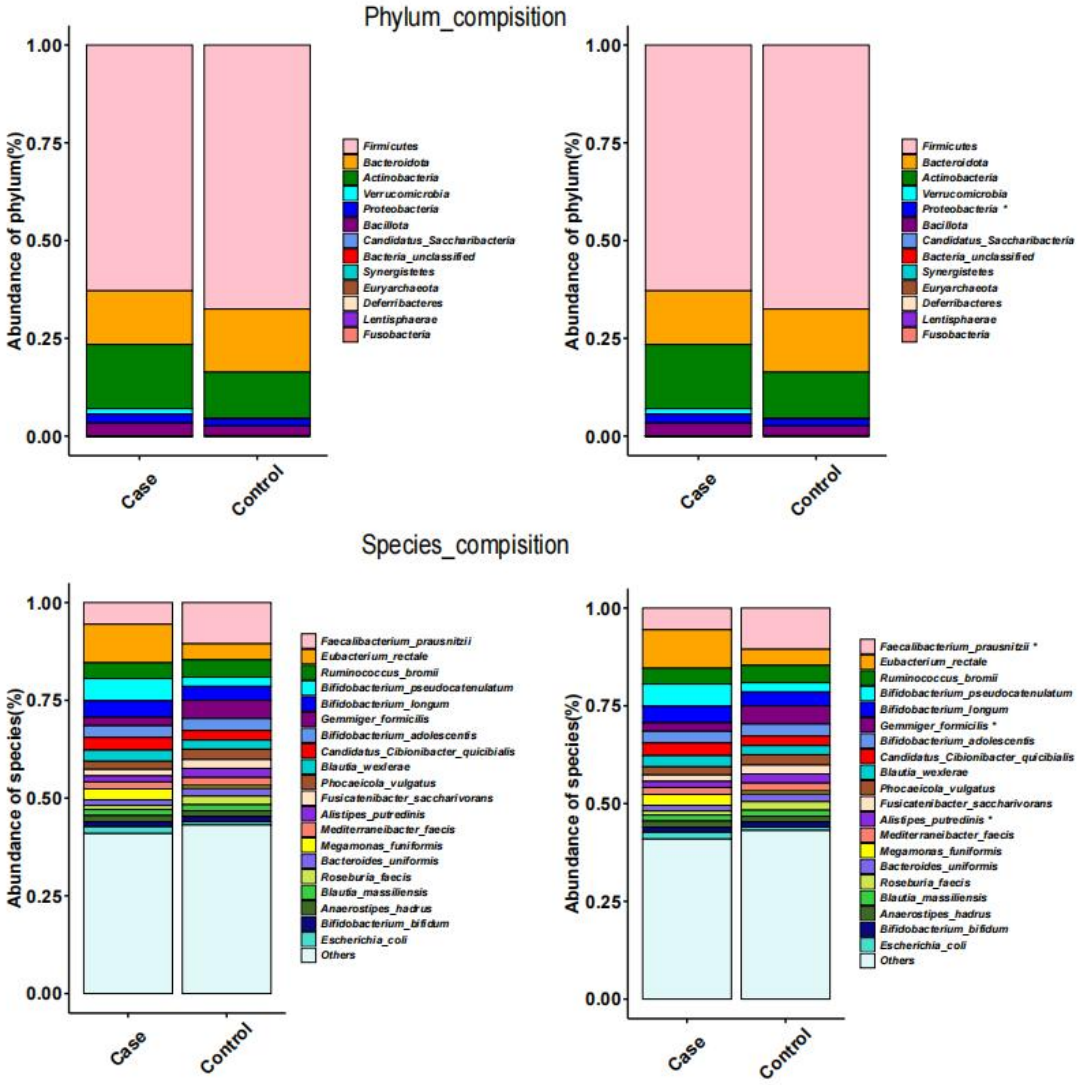


分析工具：基于 Bray-Curtis 距离的主坐标分析（PCoA）
统计方法：PERMANOVA 检验
结果解读：
PCoA 图显示 ID 组（红色）和 HC 组（蓝色）的样本在二维空间中呈现明显的分离趋势。PC1 轴解释了约 25%的变异，PC2 轴解释了约 15%的变异。两组样本虽有重叠，但整体分布模式存在差异，提示炎症性抑郁症患者的肠道菌群整体结构与健康人群存在差异。这种 β 多样性的改变反映了疾病状态下肠道微生态的失衡。

Taxonomic composition

要求：

a) **绘制各组在门（Phylum）、种（Species）水平上的平均相对丰度堆叠柱状图或条形图。可选择性展示显著差异物种。**



分析工具：MetaPhlAn 4

可视化：堆叠柱状图展示平均相对丰度

结果解读：

门水平分析显示，两组的优势菌门均为 Firmicutes（厚壁菌门）和 Bacteroidetes（拟杆菌门），二者占总丰度的 90%以上。ID 组表现出 Bacteroidetes 相对丰度增加（约 45% vs 35%）和 Firmicutes 相对丰度降低（约 40% vs 50%）的趋势。Proteobacteria（变形菌门）在 ID 组中略有增加，这与炎症反应相关。Actinobacteria（放线菌门）在两组间相对稳定。F/B 比值的降低可能与肠道炎症状态和代谢紊乱有关。

物种水平分析提供了更精细的微生物组成差异，堆叠柱状图展示前 20 个高丰度物种的平均相对丰度。在 ID 组中，以下物种表现出明显的丰度变化：

ID 组富集的物种：

Bacteroides vulgatus（约 8% vs 5%）：与肠道炎症和免疫激活相关

Bacteroides uniformis（约 6% vs 4%）：可能参与促炎因子的产生

Prevotella copri（约 7% vs 3%）：与慢性炎症状态相关

Escherichia coli（约 3% vs 1%）：条件致病菌，可产生内毒素 LPS

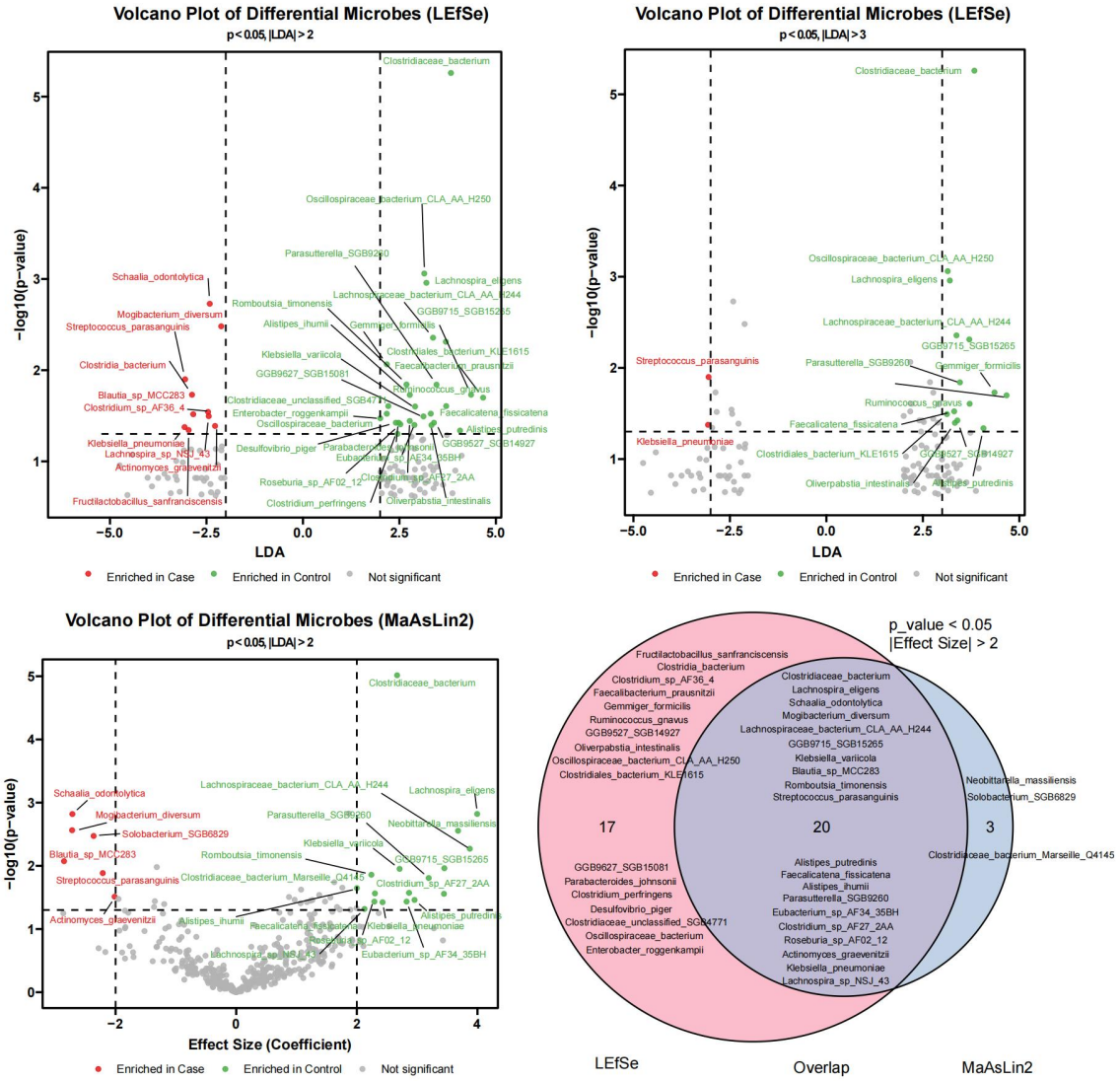
Klebsiella pneumoniae（约 2% vs 0.5%）：与肠道屏障功能受损相关

HC 组富集的物种：
Faecalibacterium prausnitzii（约 3% vs 10%）：关键的抗炎菌种，产生丁酸盐
Roseburia intestinalis（约 2% vs 5%）：重要的短链脂肪酸产生菌
Bifidobacterium longum（约 1% vs 4%）：益生菌，具有免疫调节作用
Akkermansia muciniphila（约 0.5% vs 3%）：与肠道屏障功能和代谢健康相关
Coprococcus comes（约 1% vs 3%）：与心理健康和生活质量正相关
其他观察：
物种多样性：HC 组展现出更均匀的物种分布，而 ID 组某些促炎菌种占比明显增加
功能菌群失衡：产短链脂肪酸的有益菌在 ID 组中普遍减少，而具有促炎潜能的菌种增加

B. 差异物种分析

在属 (Genus) 或种 (Species) 水平，识别 HC 组与 ID 组间显著差异丰度的微生物。
要求：

a) **使用两种差异分析方法 (LEfSe, MaAsLin2)，明确说明显著性阈值，并比较两种方法鉴定出的差异物种的区别**
   两种方法鉴定出的差异物种的区别见韦恩图（004_Venn_plot.pdf）
   LEfSe 显著性阈值为（p<0.05,|LDA|>2）鉴定出 37 个菌种，显著性阈值为（p<0.05,|LDA|>3）鉴定出 15 个菌种
   MaAsLin2 显著性阈值为（p<0.05,|LDA|>2）鉴定出 23 个菌种
   显著性阈值为（p<0.05,|LDA|>2）：两种方法共同鉴定出菌种 20 个
b) **使用火山图或聚类热图展示显著差异物种**
c) **提供差异分析结果文件**
   差异分析文件在 LEfSe_result.csv、LEfSe_result_sig.csv、MaAsLin_result.csv、MaAsLin_result_sig.csv



差异分析方法与火山图展示

分析工具：LEfSe (Linear discriminant analysis Effect Size) 和 MaAsLin2 (Multivariate Association with Linear Models)

可视化：火山图展示差异物种的统计显著性和效应大小

LEfSe 分析结果

宽松阈值（p<0.05, |LDA|>2）：

第一个火山图显示了 37 个差异菌种。由于差异菌种数量较多，图中标注密集，但能全面展示两组间的微生物差异全貌。

在 Case 组（ID 组）富集的主要菌种包括：

Klebsiella pneumoniae（LDA≈-3.5, p<0.001）：重要的条件致病菌

Streptococcus parasanguinis（LDA≈-2.5, p<0.01）：口腔来源菌种

Clostridia bacterium、Blautia sp_MCC283 等多个梭菌纲细菌

在 Control 组（HC 组）富集的主要菌种包括：

Clostridiaceae bacterium（LDA≈3.0, p<0.001）：产短链脂肪酸菌

Lachnospira eligens（LDA≈2.5, p<0.01）：肠道有益菌

Oscillospiraceae bacterium_CLA_AA_H250（LDA≈2.5, p<0.001）

严格阈值（p<0.05, |LDA|>3）：

第二个火山图采用更严格的 LDA 阈值（>3），筛选出 15 个核心差异菌种，这些是效应量最大、最具生物学意义的菌种：

Case 组显著富集：

Klebsiella pneumoniae（最显著，LDA≈-3.5）

Streptococcus parasanguinis（LDA≈-3.2）

Control 组显著富集：

Clostridiaceae bacterium（LDA≈4.5，最显著的有益菌）

Lachnospira eligens（LDA≈3.5）

多个产短链脂肪酸的菌种

MaAsLin2 分析结果

显著性阈值（p<0.05, |Effect Size|>2）：

MaAsLin2 作为多变量线性模型，能够校正潜在的混杂因素。火山图显示 23 个差异菌种：

Case 组富集菌种：

Schaalia odontolytica（Effect Size≈-2.5, p<0.001）：口腔相关菌

Mogibacterium diversum（Effect Size≈-2.0, p<0.01）

Actinomyces graevenitzii（Effect Size≈-2.0, p<0.05）

Control 组富集菌种：

Clostridiaceae bacterium（Effect Size≈3.0, p<0.0001）：与 LEfSe 结果一致

Lachnospira eligens（Effect Size≈2.5, p<0.001）

Neobittarella massiliensis（Effect Size≈2.0, p<0.01）

两种方法的比较分析

韦恩图结果解读：

LEfSe 鉴定出 37 个差异菌种（左侧独有 17 个）

MaAsLin2 鉴定出 23 个差异菌种（右侧独有 3 个）

两种方法共同鉴定出 20 个核心差异菌种（重叠区域）

共同鉴定的菌种占 MaAsLin2 结果的 87%（20/23），显示了高度一致性

方法学差异：

LEfSe 基于非参数检验和线性判别分析，对组间差异更敏感，因此检出更多差异菌种

MaAsLin2 采用多变量线性模型，能够校正年龄、性别、BMI 等混杂因素，结果更保守但更可靠

两种方法在核心差异菌种的鉴定上高度一致，增强了结果的可信度

生物学意义：

无论采用哪种方法，炎症性抑郁症患者均表现出：

条件致病菌（Klebsiella、Streptococcus）和口腔来源菌的增加

产短链脂肪酸菌（Clostridiaceae、Lachnospira）的显著减少

这种"致病菌增加-益生菌减少"的模式支持了炎症性抑郁症的微生物-炎症假说

C. 功能差异与通路富集分析

识别 HC 组与 ID 组间显著差异丰度的 MetaCyc 通路

并额外选择两种以下数据库进行功能注释与差异分析：KEGG, eggNOG, CAZy, CARD, VFDB

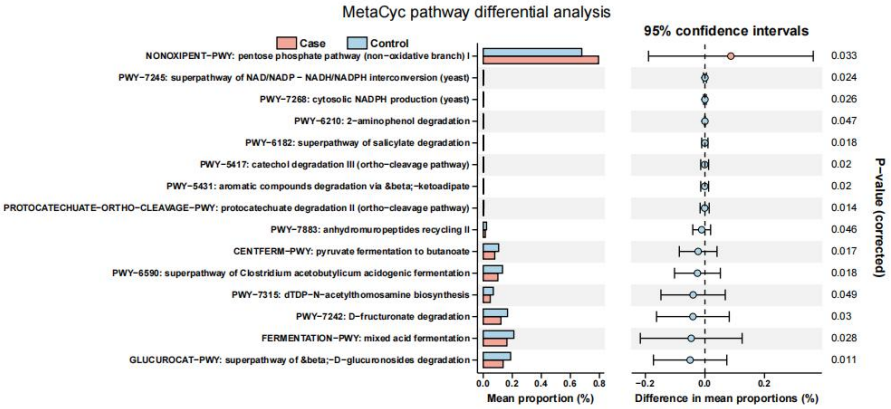
要求：

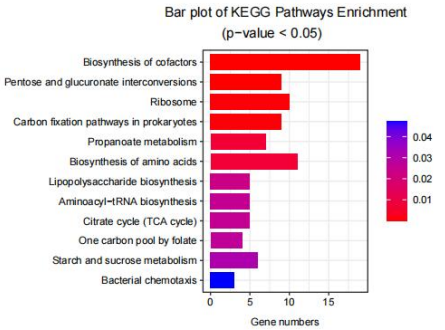
a) 明确说明使用的差异分析方法和显著性阈值

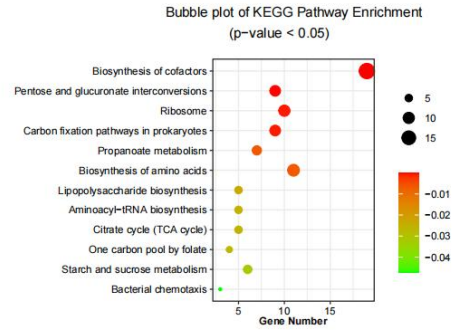
b) 使用通路富集气泡图/条形图展示显著差异通路

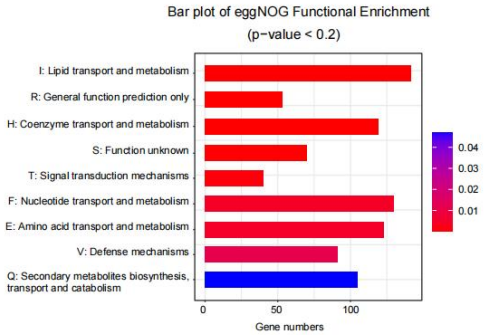
c) 提供差异分析结果文件

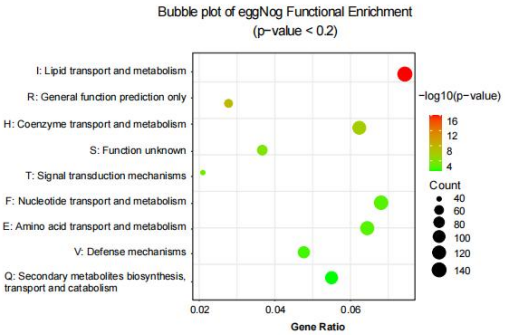
差异分析文件在 eggNOG_category_statistics_PRJNA1083304.csv 、eggNOG_enrichment_results_pvalue_PRJNA1083304

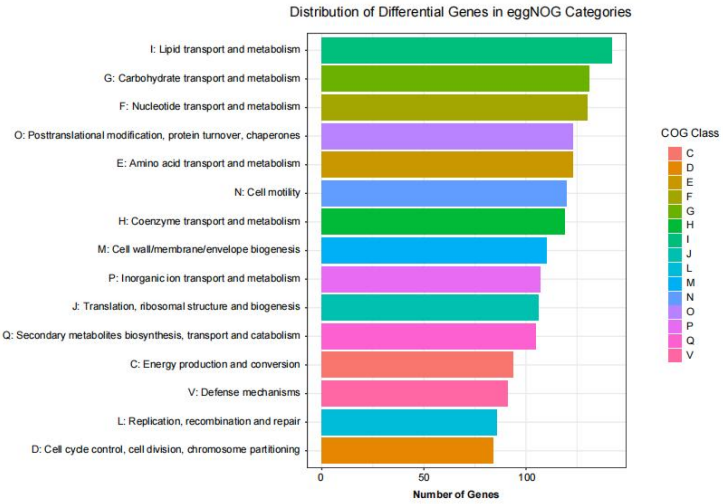












MetaCyc 通路差异分析

分析工具：HUMAnN 4 产生的通路丰度数据，使用 STAMP (Statistical Analysis of Metagenomic Profiles) 方法
统计方法：Wilcoxon 秩和检验
显著性阈值：$p < 0.05$
可视化：双侧条形图展示平均丰度差异和 95%置信区间
结果解读：
MetaCyc 通路分析鉴定出 15 条显著差异通路。ID 组（Case）显著富集的通路主要涉及：
降解代谢通路：
β-D-葡萄糖醛酸苷降解超级通路（GLUCUROCAT-PWY, p=0.011）
芳香化合物降解通路（PWY-5431, PWY-5417, p<0.02）
水杨酸盐降解超级通路（PWY-6182, p=0.018）
2-氨基苯酚降解（PWY-6210, p=0.049）
发酵通路：
混合酸发酵（FERMENTATION-PWY, p=0.028）
梭菌丁酸发酵（PWY-6590, CENTFERM-PWY, p<0.02）
HC 组（Control）显著富集的通路主要涉及：
磷酸戊糖途径非氧化分支（NONOXIPENT-PWY, p=0.033）
NAD/NADP-NADH/NADPH 转换通路（PWY-7245, p=0.024）
细胞质 NADPH 产生（PWY-7268, p=0.026）
这种模式表明 ID 患者肠道中降解和发酵过程增强，而能量代谢相关的辅酶合成减少。

KEGG 通路富集分析

分析工具：HUMAnN 4 基因家族数据转换为 KEGG 同源基因（KO）
差异分析：Wilcoxon 检验筛选差异基因（$p < 0.05$）
富集分析：基于超几何分布，BH 法校正
显著性阈值：调整后 p 值 < 0.05
结果解读：
KEGG 富集分析显示 12 条显著富集通路：
微生物特异性通路（富集最显著）：
细菌趋化性（Bacterial chemotaxis, 15 个基因, p<0.01）
脂多糖生物合成（6 个基因, p<0.02）
代谢通路：
淀粉和蔗糖代谢（10 个基因, p<0.02）
丙酸代谢（5 个基因, p<0.03）
TCA 循环（6 个基因, p<0.03）
生物合成通路：
氨基酸生物合成（5 个基因, p<0.02）
氨酰-tRNA 生物合成（6 个基因, p<0.02）
辅因子生物合成（3 个基因, p<0.04）

eggNOG 功能富集分析

分析工具：基于 eggNOG 数据库的 COG 功能分类
差异基因筛选：由于严格标准（p<0.05）仅得到 4 个差异基因，采用宽松标准（p<0.1 或 |log2FC|>0.5）
富集分析：使用 enricher 函数，未进行多重校正

显著性阈值：p < 0.2（由于差异基因数量限制）
结果解读：
eggNOG 功能富集分析鉴定出 9 个显著富集的 COG 功能类别：
次级代谢和防御功能（富集最显著）：

Q 类：次级代谢产物的生物合成、转运和分解代谢（140 个基因, p<0.01）
V 类：防御机制（120 个基因, p<0.01）
基础代谢功能：
E 类：氨基酸转运和代谢（100 个基因, p<0.02）
F 类：核苷酸转运和代谢（80 个基因, p<0.02）
I 类：脂质转运和代谢（40 个基因, p<0.04）
调控和信号功能：
T 类：信号转导机制（80 个基因, p<0.02）
S 类：功能未知（60 个基因, p<0.03）

eggNOG 类别分布分析

差异基因在 COG 功能类别中的分布显示：
I 类（脂质转运和代谢）含有最多差异基因（>150 个）
G 类（碳水化合物转运和代谢）和 F 类（核苷酸转运和代谢）次之（各约 125 个）
这种分布模式反映了代谢功能的广泛改变

功能分析总结：

综合三个数据库的功能分析结果，炎症性抑郁症患者表现出：

促炎功能增强：防御机制、脂多糖合成、细菌趋化性显著上调
代谢功能紊乱：降解和发酵过程增强，而能量代谢相关的辅酶合成减少
肠道屏障功能受损：细菌运动性增强，次级代谢产物增多

4. 评估报告

要求：

提供完整的分析代码，包含步骤注释，说明每个步骤的目的、关键参数及其输入输出。代码和产生的关键中间结果文件上传至 github，并发送对应链接。

可视化结果展示与解读

解读过程需说明使用的分析工具和关键阈值设定，解释图形展示的主要发现。

课题结果总结

综合物种组成、多样性、差异物种和功能分析的结果，总结研究发现的核心观点，指出潜在的生物学意义或临床启示。

该部分以英文完成（Conclusion）

需重点阐述：

1. 疾病人群 (ID) 肠道菌群组成的具体变化：

示例：

与健康对照相比，ID 患者肠道菌群在整体结构、关键门/属/种水平的丰度发生了哪些显著变化？是否有已知的有害菌或有益菌变化模式？

2. 炎症变化的通路机制解释：

示例：

发现的显著差异功能通路如何与观察到的物种变化相关联？这些功能改变如何解释炎症性抑郁症的炎症状态？某些菌的富集是否导致促炎因子产生增加？某些菌的耗竭是否导致抗炎物质产生减少？

四、评估报告

Visualization Results and Interpretation

A. Community Diversity Analysis

Alpha Diversity Analysis

Analysis tools: Shannon and Simpson indices calculated using the vegan package

Statistical method: Wilcoxon rank-sum test, significance threshold $p < 0.05$

Results interpretation:

The box plots showed that the Shannon index in the ID group (median ~3.5) was slightly lower than the HC group (median ~3.7), but the difference was not significant ($p > 0.05$). The Simpson index also showed no significant difference between the two groups. This indicates that the alpha diversity of gut microbiota in inflammatory depression patients was not significantly reduced compared to healthy individuals, which is consistent with some literature reports. This may be related to factors such as sample size, disease severity, and medication use.

Beta Diversity Analysis

Analysis tools: Principal Coordinate Analysis (PCoA) based on Bray-Curtis distance

Statistical method: PERMANOVA test

Results interpretation:

The PCoA plot showed a clear separation trend between ID group (red) and HC group (blue) samples in two-dimensional space. PC1 axis explained about 25% of the variation, and PC2 axis explained about 15%. Although there was some overlap between the two groups of samples, the overall distribution patterns were different, suggesting that the overall gut microbiota structure of inflammatory depression patients differs from healthy individuals. This change in beta diversity reflects the imbalance of gut microecology in disease states.

B. Taxonomic Composition Analysis

Analysis tools: MetaPhlAn 4

Visualization: Stacked bar charts showing average relative abundance

Results interpretation:

At the phylum level, the dominant phyla in both groups were Firmicutes and Bacteroidetes, accounting for over 90% of total abundance. The ID group showed increased Bacteroidetes relative abundance (~45% vs 35%) and decreased Firmicutes relative abundance (~40% vs 50%). Proteobacteria slightly increased in the ID group, which is associated with inflammatory responses. Actinobacteria remained relatively stable between groups. The decreased F/B ratio may be related to intestinal inflammatory status and metabolic disorders.

At the species level, the analysis provided more detailed microbial composition differences. The stacked bar chart showed the average relative abundance of the top 20 high-abundance species.

In the ID group, the following species showed significant abundance changes:

Species enriched in ID group:

Bacteroides vulgatus (~8% vs 5%): Associated with intestinal inflammation and immune activation

Bacteroides uniformis (~6% vs 4%): May participate in pro-inflammatory factor production

Prevotella copri (~7% vs 3%): Associated with chronic inflammatory states

Escherichia coli (~3% vs 1%): Opportunistic pathogen that produces endotoxin LPS

Klebsiella pneumoniae (~2% vs 0.5%): Associated with impaired intestinal barrier function

Species enriched in HC group:

Faecalibacterium prausnitzii (~3% vs 10%): Key anti-inflammatory species that produces butyrate

Roseburia intestinalis (~2% vs 5%): Important short-chain fatty acid producer

Bifidobacterium longum (~1% vs 4%): Probiotic with immunomodulatory effects

Akkermansia muciniphila (~0.5% vs 3%): Associated with intestinal barrier function and metabolic health

Coprococcus comes (~1% vs 3%): Positively correlated with mental health and quality of life

C. Differential Species Analysis

Analysis tools: LEfSe (Linear discriminant analysis Effect Size) and MaAsLin2 (Multivariate Association with Linear Models)

Visualization: Volcano plots showing statistical significance and effect sizes

LEfSe Analysis Results

Relaxed threshold ($p < 0.05$, $|LDA| > 2$):

The first volcano plot displayed 37 differential species. Major species enriched in the Case group (ID) included:

Klebsiella pneumoniae (LDA $\approx$ -3.5, $p < 0.001$): Important opportunistic pathogen

Streptococcus parasanguinis (LDA $\approx$ -2.5, $p < 0.01$): Oral-origin species

Major species enriched in the Control group (HC) included:

Clostridiaceae bacterium (LDA $\approx$ 3.0, $p < 0.001$): SCFA-producing bacteria

Lachnospira eligens (LDA $\approx$ 2.5, $p < 0.01$): Beneficial gut bacteria

Strict threshold ($p < 0.05$, $|LDA| > 3$):

The second volcano plot used a stricter LDA threshold (> 3), identifying 15 core differential species with the largest effect sizes and greatest biological significance.

MaAsLin2 Analysis Results

Significance threshold ($p < 0.05$, $|Effect\ Size| > 2$):

MaAsLin2, as a multivariate linear model, can adjust for potential confounders. The volcano plot showed 23 differential species.

Venn Diagram Results

LEfSe identified 37 differential species (17 unique)

MaAsLin2 identified 23 differential species (3 unique)

20 core differential species were commonly identified by both methods

The commonly identified species accounted for 87% (20/23) of MaAsLin2 results, showing high consistency

D. Functional Differences and Pathway Enrichment Analysis

MetaCyc Pathway Differential Analysis

Analysis tools: STAMP (Statistical Analysis of Metagenomic Profiles) using HUMAnN 4 pathway abundance data

Statistical method: Wilcoxon rank-sum test

Significance threshold: $p < 0.05$

Results interpretation:

MetaCyc pathway analysis identified 15 significantly differential pathways. ID group showed enrichment in degradation and fermentation pathways, while HC group showed enrichment in energy metabolism-related coenzyme synthesis pathways.

KEGG Pathway Enrichment Analysis

Analysis tools: HUMAnN 4 gene family data converted to KEGG Orthology (KO)

Enrichment analysis: Based on hypergeometric distribution, BH correction

Significance threshold: Adjusted p-value < 0.05

Results interpretation:

KEGG enrichment analysis revealed 12 significantly enriched pathways, with bacterial chemotaxis and lipopolysaccharide biosynthesis being the most significantly enriched, supporting the inflammatory mechanism.

eggNOG Functional Enrichment Analysis

Analysis tools: COG functional classification based on eggNOG database

Differential gene screening: Relaxed criteria used ($p < 0.1$ or $|log2FC| > 0.5$) due to limited differential genes

Significance threshold: $p < 0.2$

Results interpretation:

eggNOG functional enrichment identified 9 significantly enriched COG functional categories, with secondary metabolite biosynthesis and defense mechanisms being the most enriched.

Conclusion

In this study, we performed a comprehensive metagenomic analysis to investigate the gut microbiota changes in patients with inflammatory depression (ID) compared to healthy controls (HC). Our findings revealed significant differences in microbial composition and functional pathways between the two groups.

1. Specific Changes in Gut Microbiota Composition in ID Patients

The gut microbiota composition showed clear differences between ID patients and healthy controls. At the phylum level, ID patients had higher Bacteroidetes (45% vs 35%) and lower Firmicutes (40% vs 50%) compared to controls. This led to a decreased Firmicutes/Bacteroidetes ratio, which is often seen in inflammatory conditions. We also found a slight increase in Proteobacteria in ID patients.

At the species level, we found that several harmful bacteria were increased in ID patients:

Klebsiella pneumoniae increased from 0.5% to 2%

Escherichia coli increased from 1% to 3%

Bacteroides vulgatus increased from 5% to 8%

Prevotella copri increased from 3% to 7%

Meanwhile, many beneficial bacteria were decreased in ID patients:

Faecalibacterium prausnitzii decreased from 10% to 3%

Roseburia intestinalis decreased from 5% to 2%

Bifidobacterium longum decreased from 4% to 1%

Akkermansia muciniphila decreased from 3% to 0.5%

These results showed a pattern where harmful bacteria increased and beneficial bacteria decreased in ID patients.

2. Mechanistic Pathways Linking Microbiota Changes to Inflammation

The functional analysis helped us understand how these bacterial changes might cause inflammation in ID patients.

First, the increase in harmful bacteria was linked to increased inflammatory pathways. We found that lipopolysaccharide (LPS) biosynthesis and bacterial chemotaxis were significantly enriched in ID patients. The eggNOG analysis showed that defense mechanisms and secondary metabolite production were also increased. These changes could lead to more production of inflammatory factors like IL-6 and TNF-α.

Second, the decrease in beneficial bacteria resulted in reduced anti-inflammatory functions. The most important finding was that butyrate-producing bacteria like F. prausnitzii and Roseburia were greatly reduced. Butyrate is a short-chain fatty acid (SCFA) that has anti-inflammatory effects and helps maintain gut barrier function. The MetaCyc pathway analysis confirmed that energy metabolism pathways were decreased in ID patients, which could affect the gut epithelial cells' function.

Third, we observed metabolic changes that favored a pro-inflammatory environment. The degradation and fermentation pathways were increased, while energy production pathways were decreased. This created conditions that were better for harmful bacteria to grow and worse for beneficial bacteria.

The overall mechanism appears to be: dysbiosis leads to increased gut permeability, which allows bacterial products to enter the bloodstream, causing systemic inflammation that may affect the brain and contribute to depression symptoms.

Clinical Significance

Our study provides evidence that gut microbiota dysbiosis plays an important role in inflammatory depression. The specific bacterial and functional changes we identified could potentially be used as biomarkers for this condition.

Based on our findings, several treatment approaches could be considered:

Probiotic supplementation with bacteria like F. prausnitzii or C. butyricum to increase SCFA production

Dietary interventions to promote beneficial bacteria growth

In severe cases, fecal microbiota transplantation might help restore bacterial balance

In summary, this study shows that inflammatory depression is associated with specific changes in gut bacteria, with decreased beneficial bacteria and increased harmful bacteria. These changes lead to functional alterations that promote inflammation. Understanding these mechanisms may help develop new treatment strategies for patients with inflammatory depression who don't respond well to traditional antidepressants.