

Broliden_5325

16 February, 2021

Contents

Loading data and metadata	1
Calculate QC metrics	1
Computing differential expression across microbiome datasets	11
#Load libraries and other scripts	
#Defining some variables for the analysis	

Loading data and metadata

Calculate QC metrics

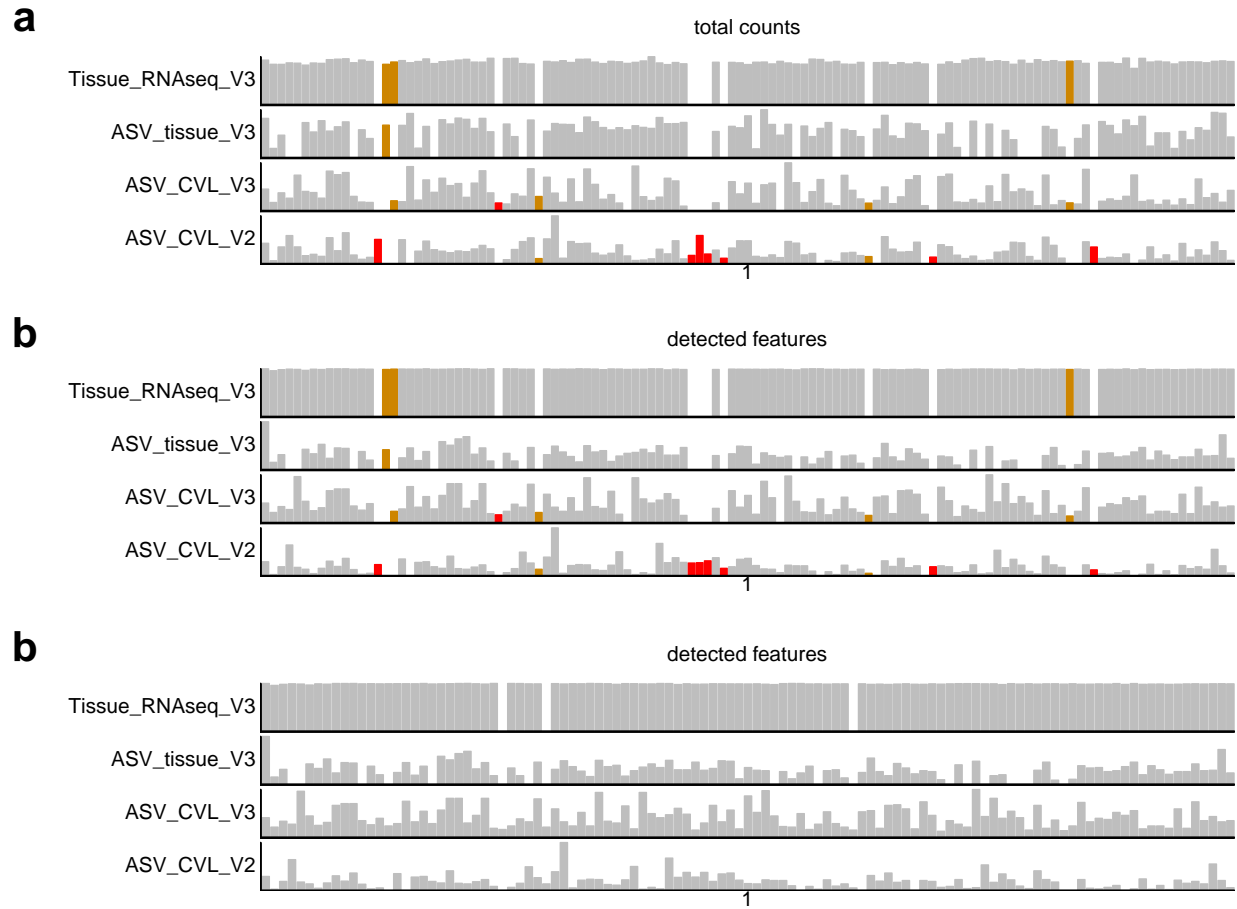
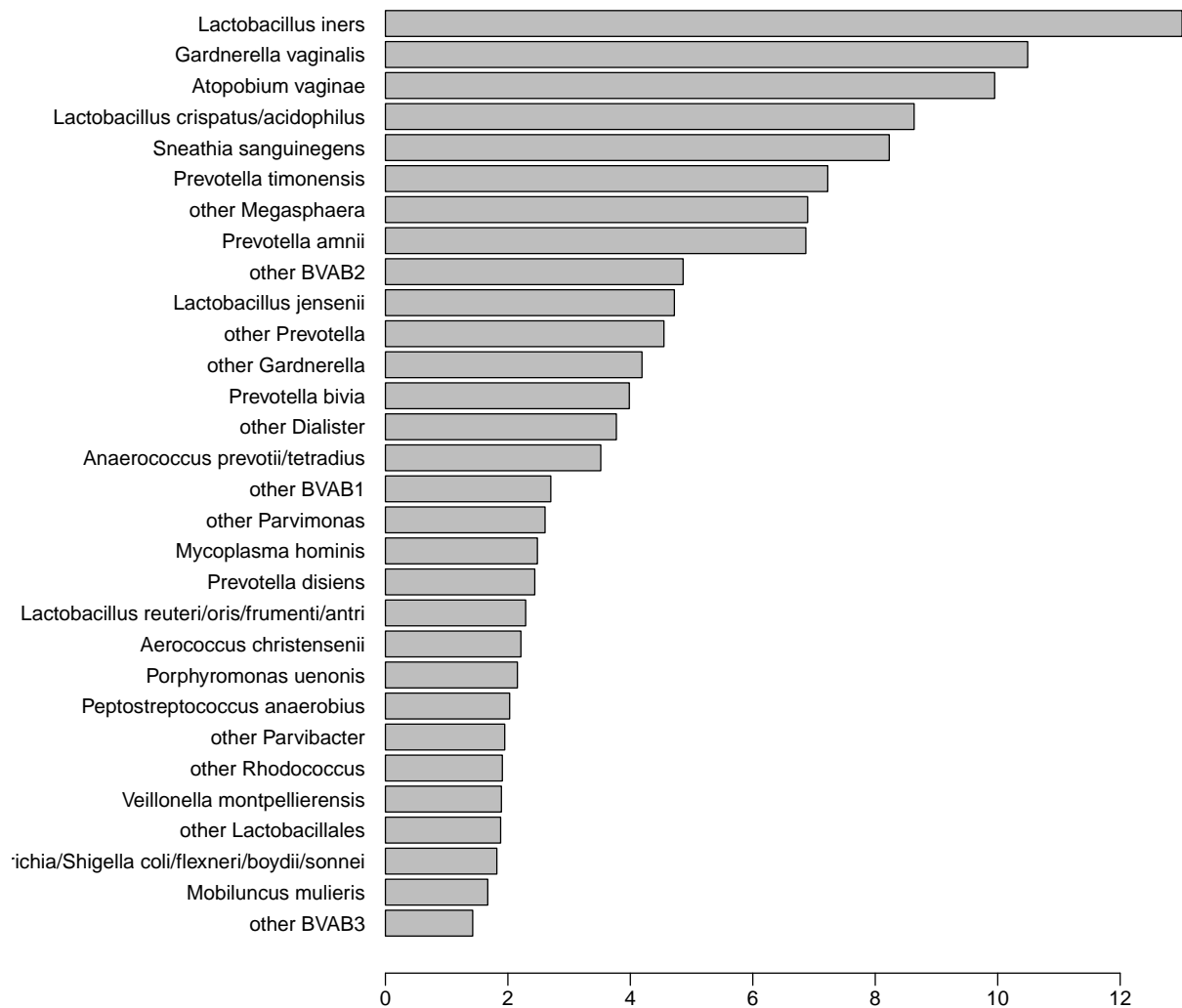


Figure 1. Comparative barplot for the **a)** total counts and **b)** number of non-zero detected features (genes / bacteria) for each of the sequencing datasets. Samples are ordered alphabetically according to the patient ID. Samples that are present in exactly two datasets are shown in orange (P016,P017,P035,P076,P101). Samples that are present in exactly two datasets are shown in red (P015,P030,P054,P055,P056,P058,P084,P104).



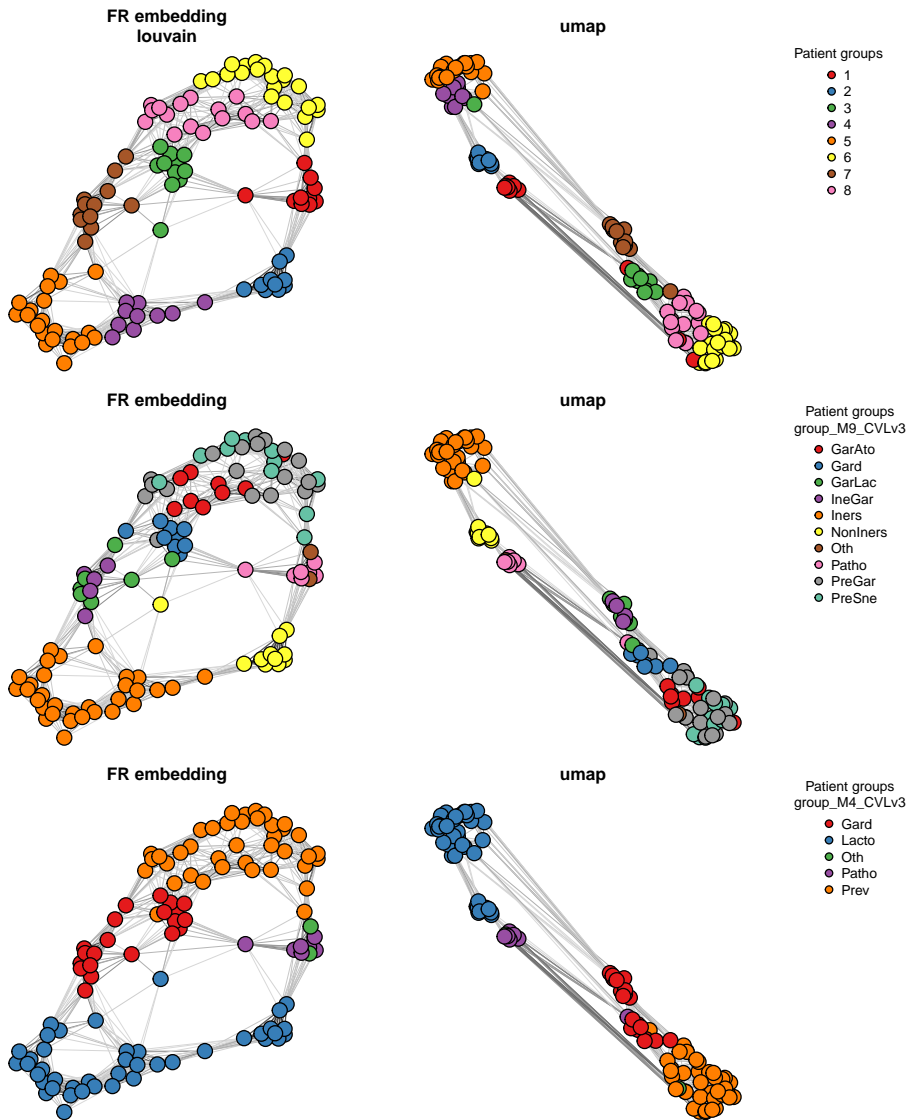
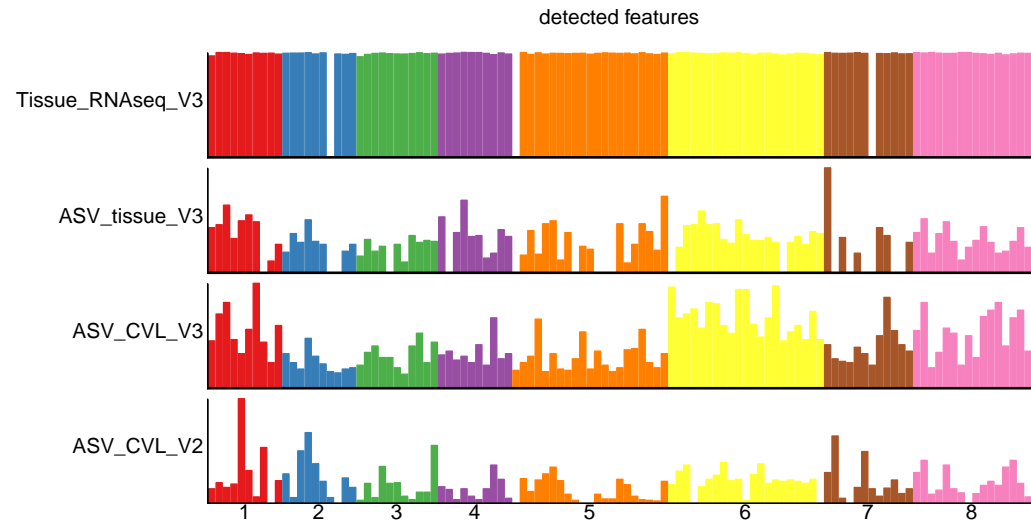


Figure 1b. Sample embedding of 10-SNN graph clustered using Louvain based on the normalized bacterial abundances across all microbiome datasets (including samples not present in all datasets).

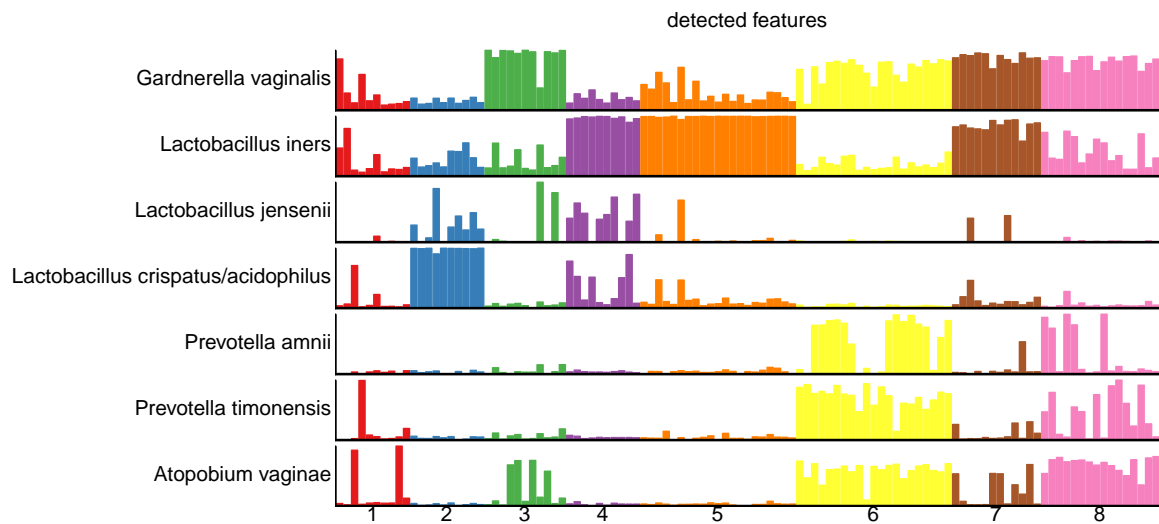
```
mypar(mar=c(2,15,2,2))

barlist( data = rbind( detected_counts),main = "detected features",
  genes = c(rownames(detected_counts)),labels = sub("_norm.*","",rownames(counts)),
  clustering = metadata$joint_clustering,
  col=pal,draw_mean_lines=F)
add_letter("b")
```

b



```
barlist( data = all_microbiome, main = "detected features",
  genes = c("Gardnerella vaginalis",
    "Lactobacillus iners",
    "Lactobacillus jensenii",
    "Lactobacillus crispatus/acidophilus",
    "Prevotella amnii",
    "Prevotella timonensis",
    "Atopobium vaginae"),
  clustering = metadata$joint_clustering,
  col=pal,draw_mean_lines=F)
```



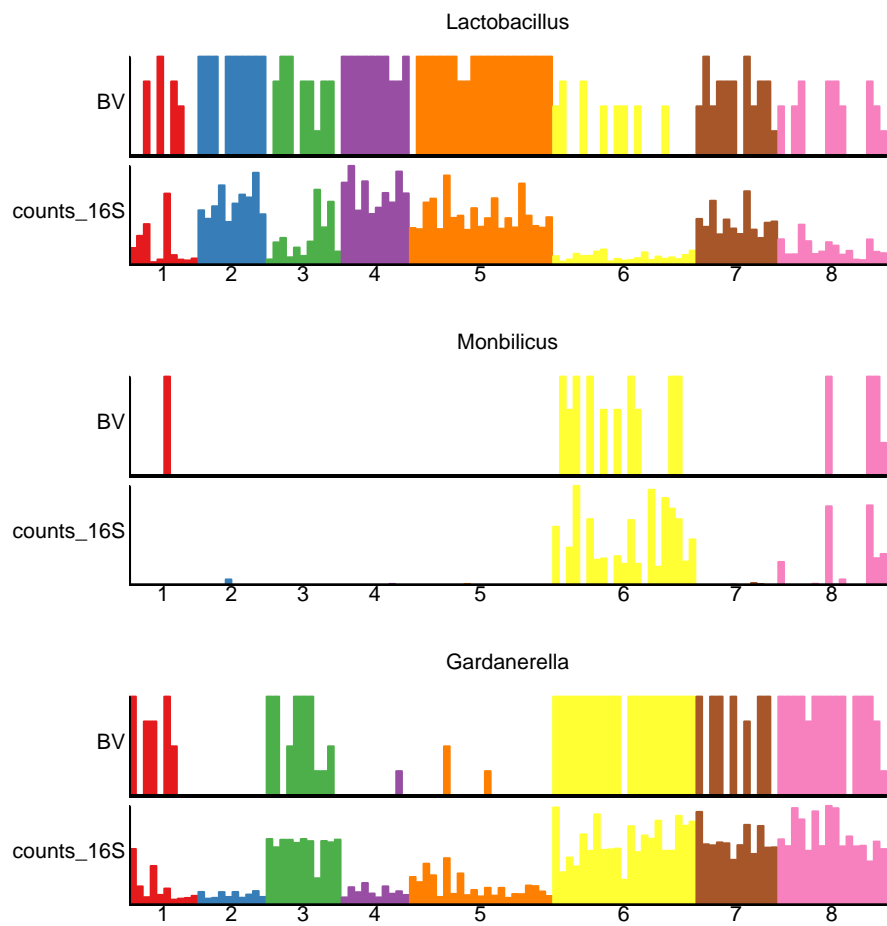


Figure 2. Differential bacterial abundance across joint clustered samples and compared across microbiome datasets. The results are shown both as **a)** Dot plots and **b)** barplots Bacteria with \log_2FC above 0.25 and p-value below 0.01 were considered significant and were sorted by the highest expression.

Figure 2. Differential bacterial abundance across microbiome datasets. The results are shown both as **a)** Dot plots and **b)** barplots Bacteria with log2FC above 0.25 and p-value below 0.01 were considered significant and were sorted by the highest expression.

Figure 2. Differential bacterial abundance across CVL2 and CVL3 microbiome datasets. The results are shown both as **a)** Dot plots and **b)** barplots Bacteria with log2FC above 0.25 and p-value below 0.01 were considered significant and were sorted by the highest expression.

Figure 2b. Differential bacterial abundance across tissue and CVL microbiome week3 datasets. The results are shown both as **a)** Dot plots and **b)** barplots Bacteria with log2FC above 0.25 and p-value below 0.01 were considered significant and were sorted by the highest expression.

Computing differential expression across microbiome datasets

Figure 4. Differential bacterial abundance across all groups and all microbiome datasets. The results are shown both as **a)** Dot plots and **b)** barplots Bacteria with \log_2FC above 0.25 and p-value below 0.01 were considered significant and were sorted by the highest expression.

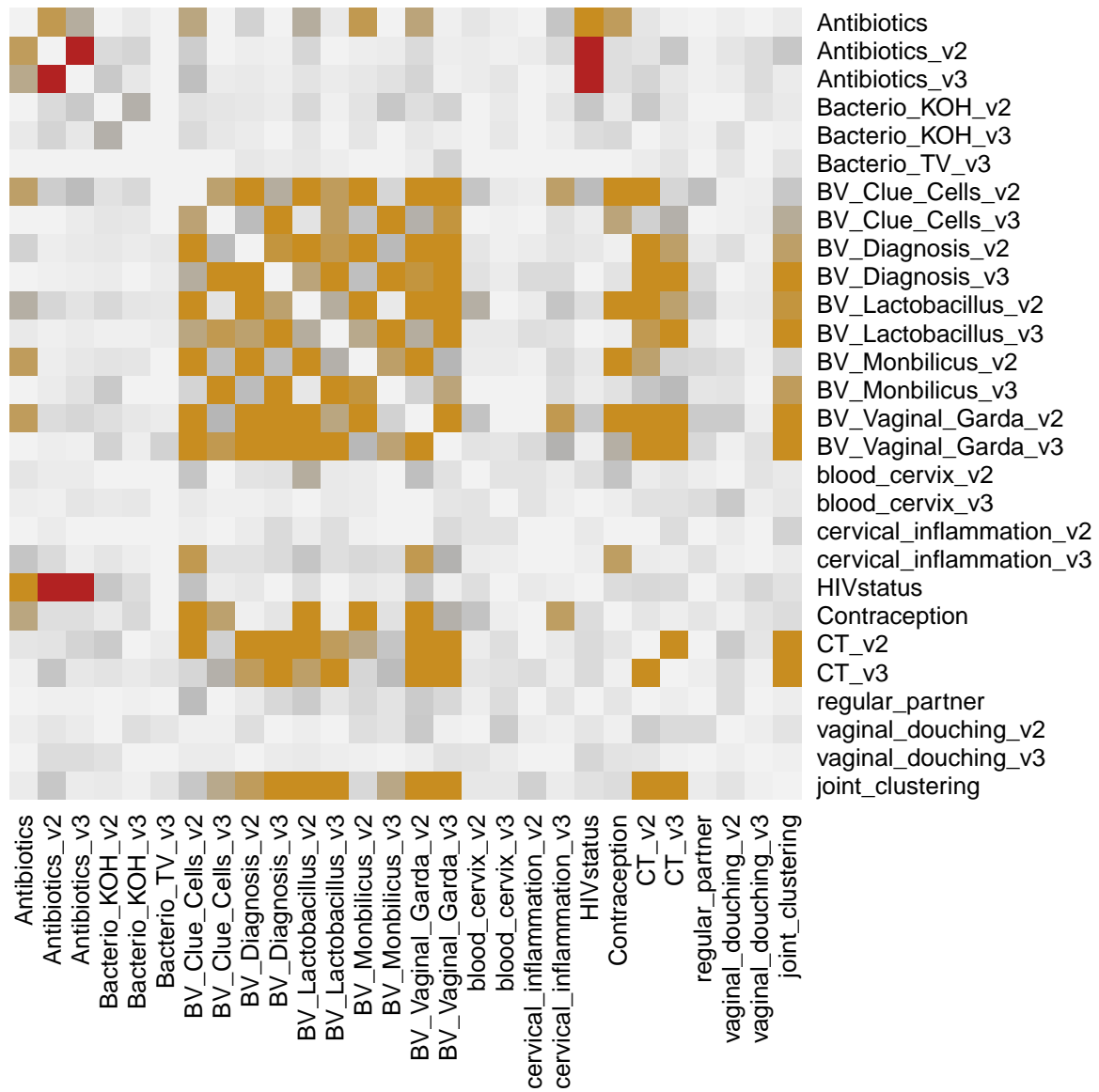


Figure 6. Association analysis across several patient categorical metadata parameters, including patient groupings annotations from microbiome.

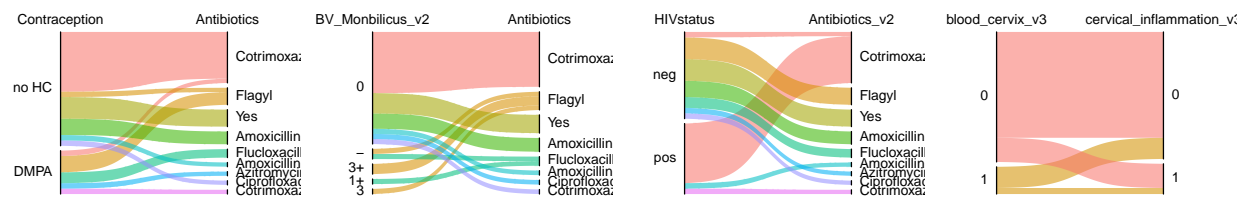
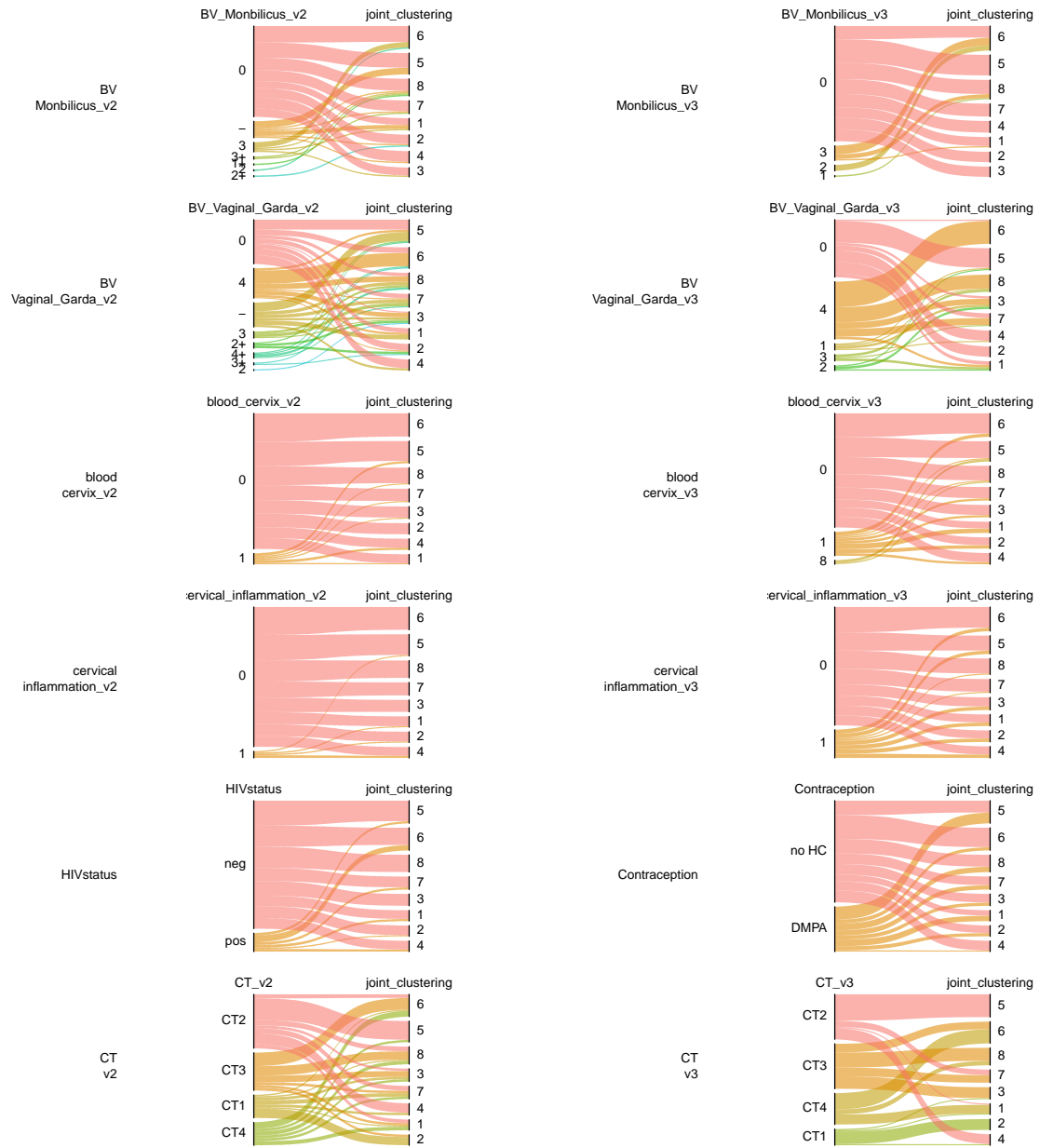


Figure 7. A few examples of significant association between metadata parameters shown as sankey plots.



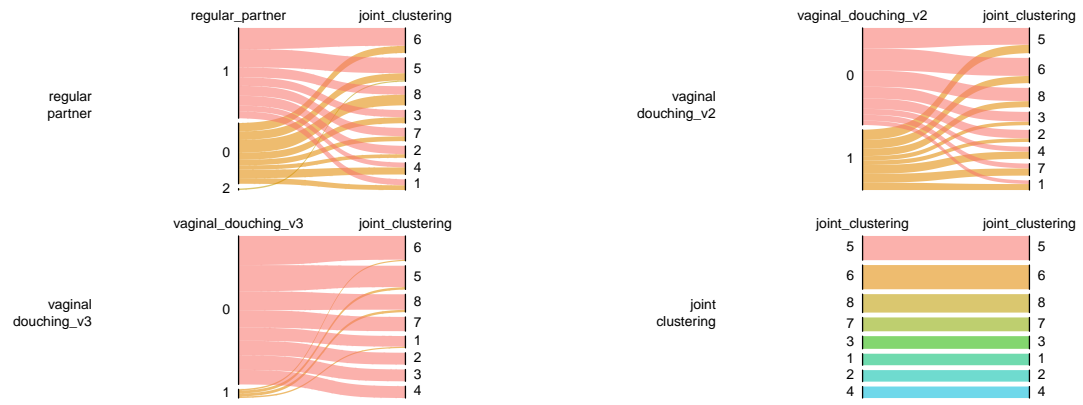


Figure 8. Sankey plots for all tested associations between the patient groups identified in in the microbiome datasets.

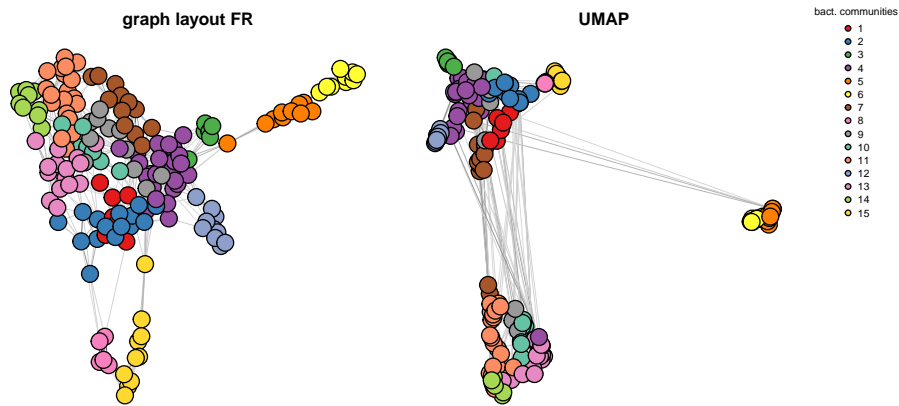


Figure 9a. Bacterial community embedding of 10-SNN graph clustered using walktrap community detection algorithm based on the normalized bacterial abundances across all microbiome datasets (including samples not present in all datasets).

community1	community2	community3	community4	community5	community6
Anaerococcus hydrogenalis Atopobium deliae Bifidobacterium bifidum Finegoldia magna other Anaerococcus other Eubacterium other Mobiluncus other Peptostreptococcus	Actinomyces turicensis Anaerococcus murchisonii Bacteria Dialister propionificiens Granulicatella elegans Mycoplasma hominis other Howardella other Olsenella other Veillonella Prevotella bivia Prevotella denticola Prevotella melaninogenica Streptococcus macedonicus/pasteuri Varibaculum cambriense Veillonella montpellierensis	Lactobacillus coleohominis Lactobacillus crispatus/acidophilus Lactobacillus iners Lactobacillus jensenii Lactobacillus reuteri/oris/frumenti/ant other Lactobacillales	Bifidobacterium adolescentis Bifidobacterium pseudocatenulatum/ Collinsella aerofaciens Corynebacterium jeikeium Corynebacterium pseudogenitalium Corynebacterium pyruviciproducens Corynebacterium sp/amycolatum Enhydrobacter aerosaccus Escherichia/Shigella coli/flexneri/boy Gemella mortuorum/sanguinis/haem Haemophilus parainfluenzae Lactobacillus crispatus Lactobacillus reuteri Lactobacillus salivarius Megaspheera elsdeni other Ezakiella other Gordonia other Lactobacillus other Listeria other Murdochella other Rhizobium other stricto Prevotella corporis Streptococcus saccharolyticus/capi Veillonella dispar Corynebacterium freneyi/sp/amycola	Cupriavidus metallidurans other Actinomyces other Brachybacterium other Dietzia other Intrasporangiaceae other Kocuria other Micrococcus other Nitrospira other Ornithinimicrobium other Sphingomonas	other Arthrobacter other Chryseobacterium other Exiguobacterium other Janibacter other Paracoccus other Phyllobacteriaceae other Psychrobacter other Rhodococcus other Neorhizobium
community7	community8	community9	community10	community11	community12
Aerococcus christensenii Bifidobacterium longum/breve Corynebacterium aurimucosum Corynebacterium sundsvallense/thor Corynebacterium tuberculoearicum Enterococcus villorum/hirae/laecum/ Lactobacillus gasseri/johnsonii/taiwai other Atopobium other Bacilli other Bifidobacteriaceae other Corynebacterium other Gemella	other Enterobacter other Ralstonia other Streptococcus other Veillonellaceae	other Actinomycetales other Aerococcus other Bacteroidales other Firmicutes other Methylobacterium other Porphyromonadaceae other Romboutsia Peptoniphilus duerdenii Peptoniphilus massiliensis Streptococcus agalactiae Ureaplasma parvum	Anaerococcus lactolyticus Anaerococcus obsiensis Campylobacter ureolyticus Fusobacterium nucleatum other Campylobacter other Porphyromonas other Sneathia Porphyromonadaceae bacterium Streptococcus infantis/oralis/mitis Sutterella sanguinus	Actinomyces hongkongensis Anaerococcus prevotii/tetradis Atopobium vaginae Gardnerella vaginalis Gemella asaccharolytica Mobiluncus mulieris other Arcanobacterium other Bulleidia other BVAB1 other BVAB2 other BVAB3 other Coriobacteriaceae other Falsiporphyromonas other Gardnerella Peptoniphilus harei Peptoniphilus lacrimalis Peptostreptococcus anaerobius Prevotella amnii Prevotella buccalis	Faecalibacterium prausnitzii other Alloprevotella other Bacillus other Blautia other Faecalibacterium other Roseburia other Ruminococcaceae other Ruminococcus2 Prevotella copri Streptococcus intermedius/constellat
community13	community14	community15			
Dialister microaerophilus Mobiluncus curtisii Moryella indoligenes other Clostridiales other Finegoldia other Parvibacter other Sutterella other Syntrophococcus Peptococcus niger Peptoniphilus coxi Peptostreptococcus stomatis Porphyromonas asaccharolytica Prevotella bergensis Prevotella disiens Prevotella sp	other Dialister other Lachnospiraceae other Megaspheera other Parvimonas other Peptoniphilus other Prevotella Porphyromonas uenonis Prevotella timonensis Sneathia sanguinegens	other Acinetobacter other Aquabacterium other Chitinophagaceae other Comamonas other Janthinobacterium other Neisseria other Pelomonas other Pseudomonas other Schlegella Thermus scotoductus			

Figure 9b. List of bacterial belongin to each community.

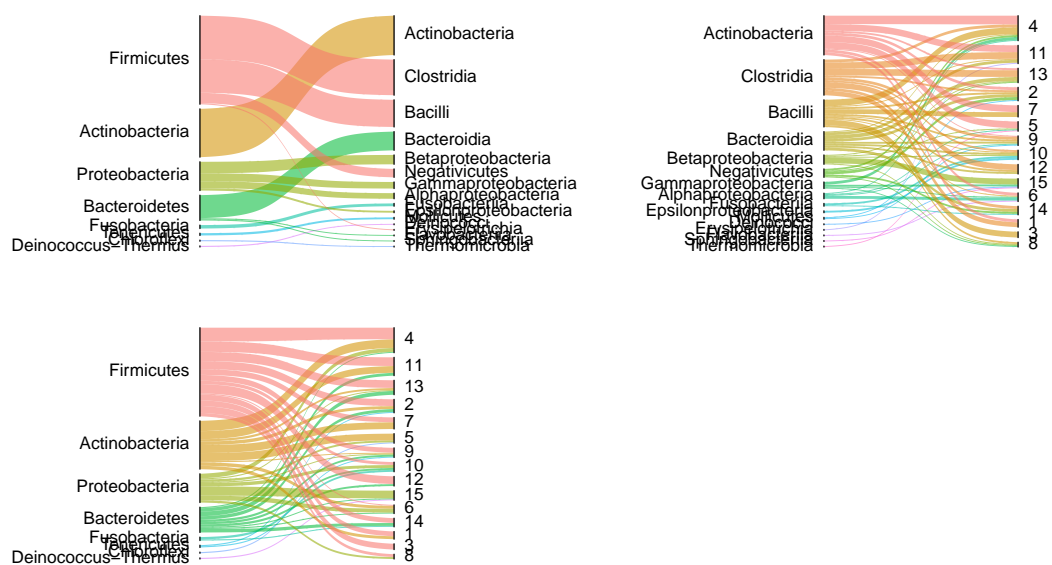


Figure 9b. Taxonomic annotation of bacterial communities.

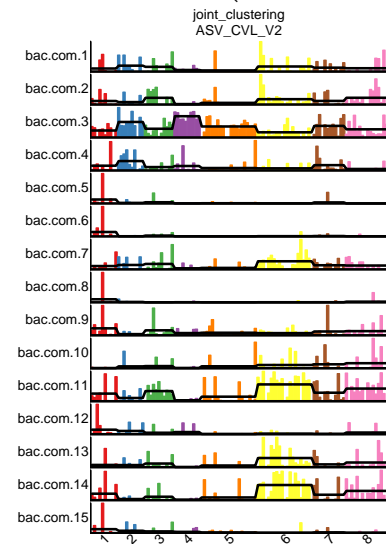
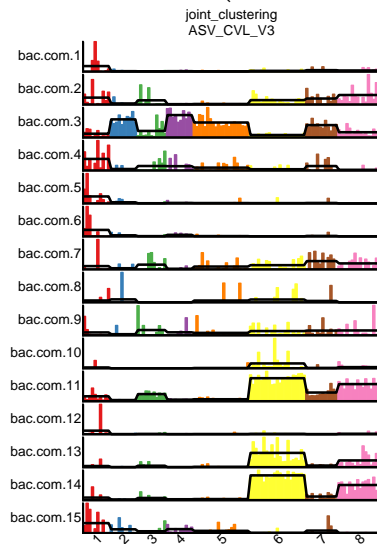
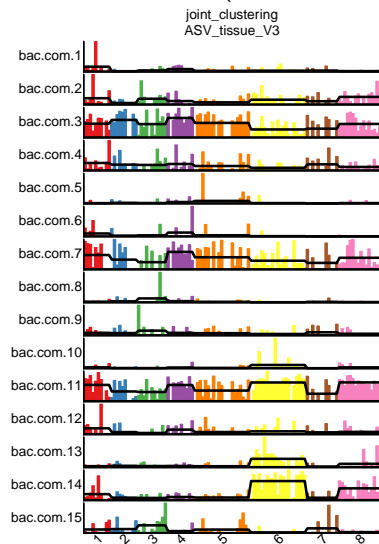
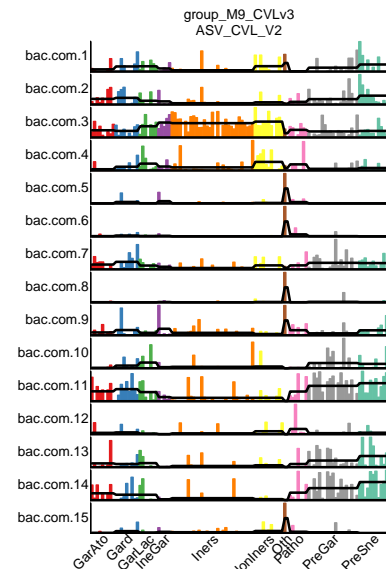
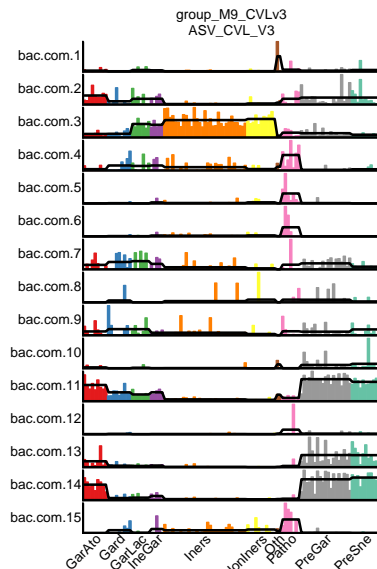
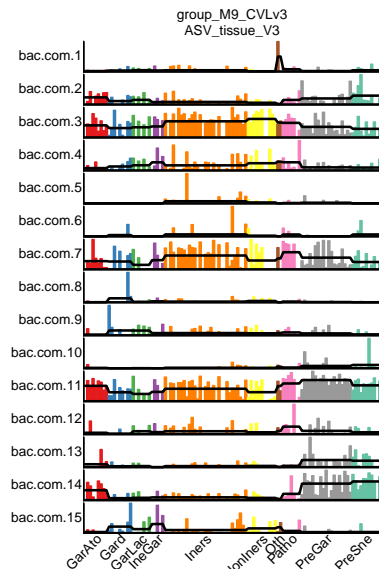
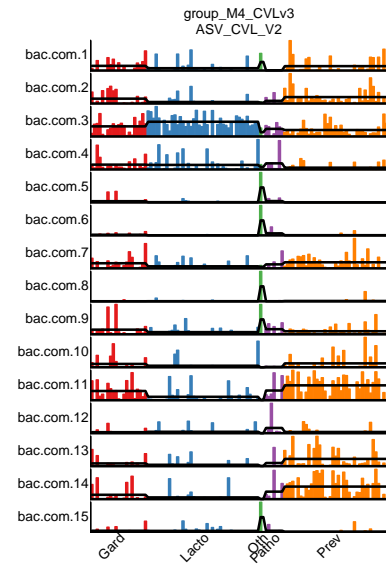
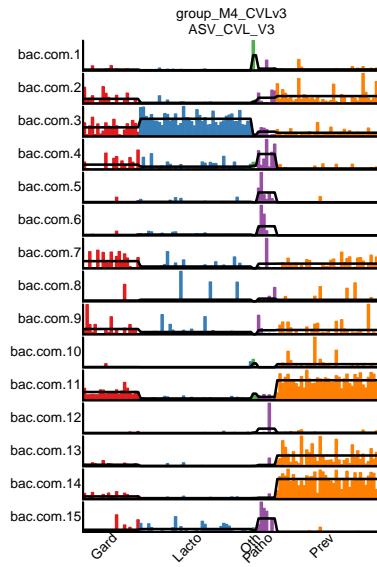
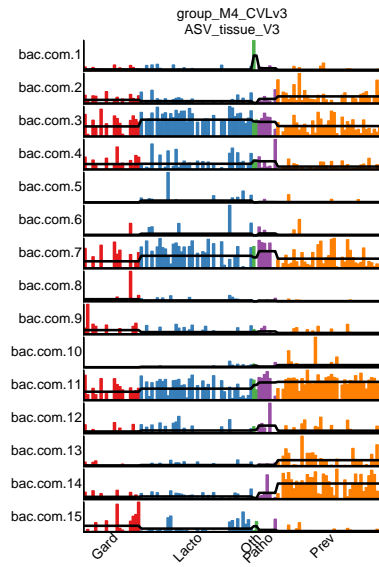


Figure 5. Comparisson of microbiome communities across all datasets and all clustering methods.





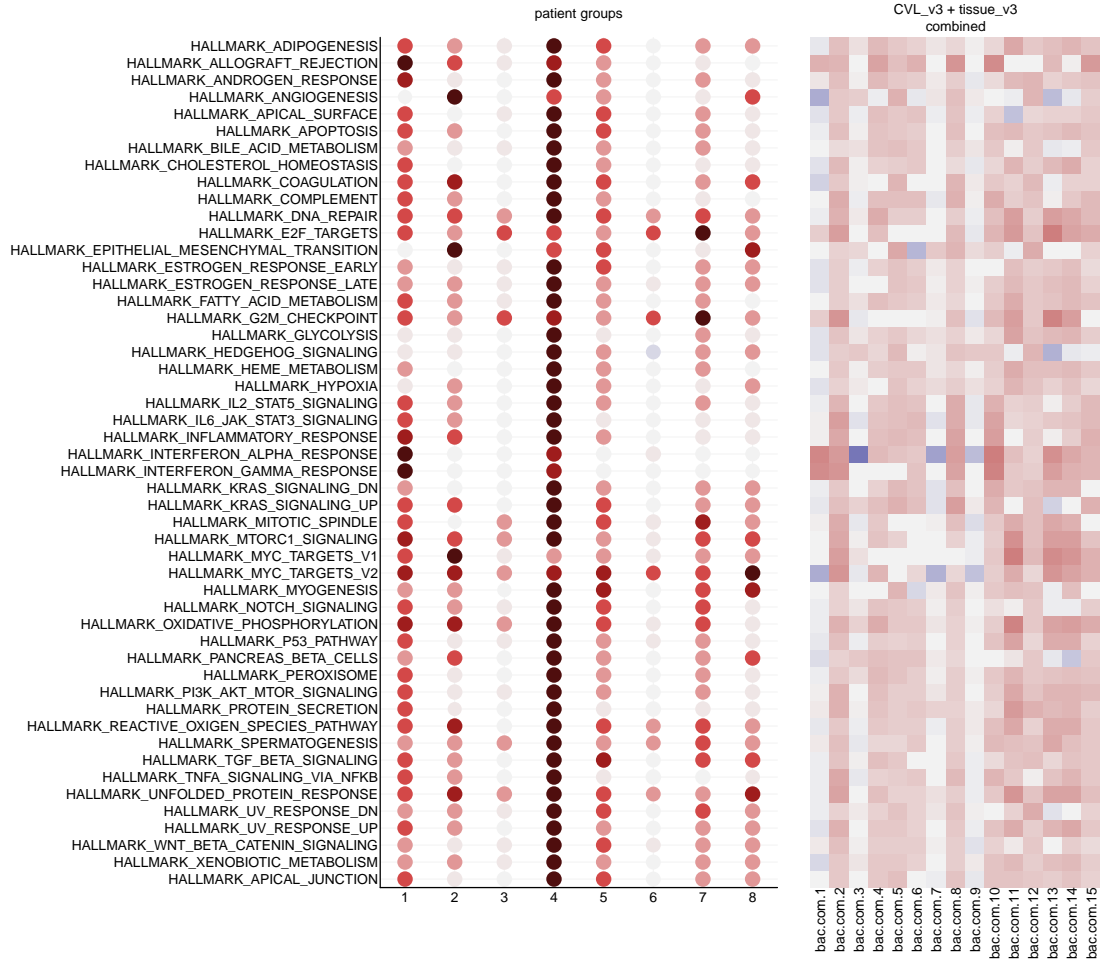


Figure 9. Heatmaps showing the functional association of microbiome datasets with the expression profiles in the RNAseq dataset. Briefly, bacterial abundances from each dataset were correlated with the gene expression of the top 5000 highly variable genes from the RNAseq dataset, generating a correlation matrix between bacteria and genes. Then for each bacteria, we rank genes based on their correlation to that bacteria and perform gene set enrichment analysis (GSEA) using the Hallmark MSigDB gene annotation database. This, in turn, will result in a matrix associating every bacteria with every Hallmark MSigDB process in the tissue. The heatmap shows the normalized enrichment score (NES). Only enrichments with pvalue below 0.05 are shown.

KEGG



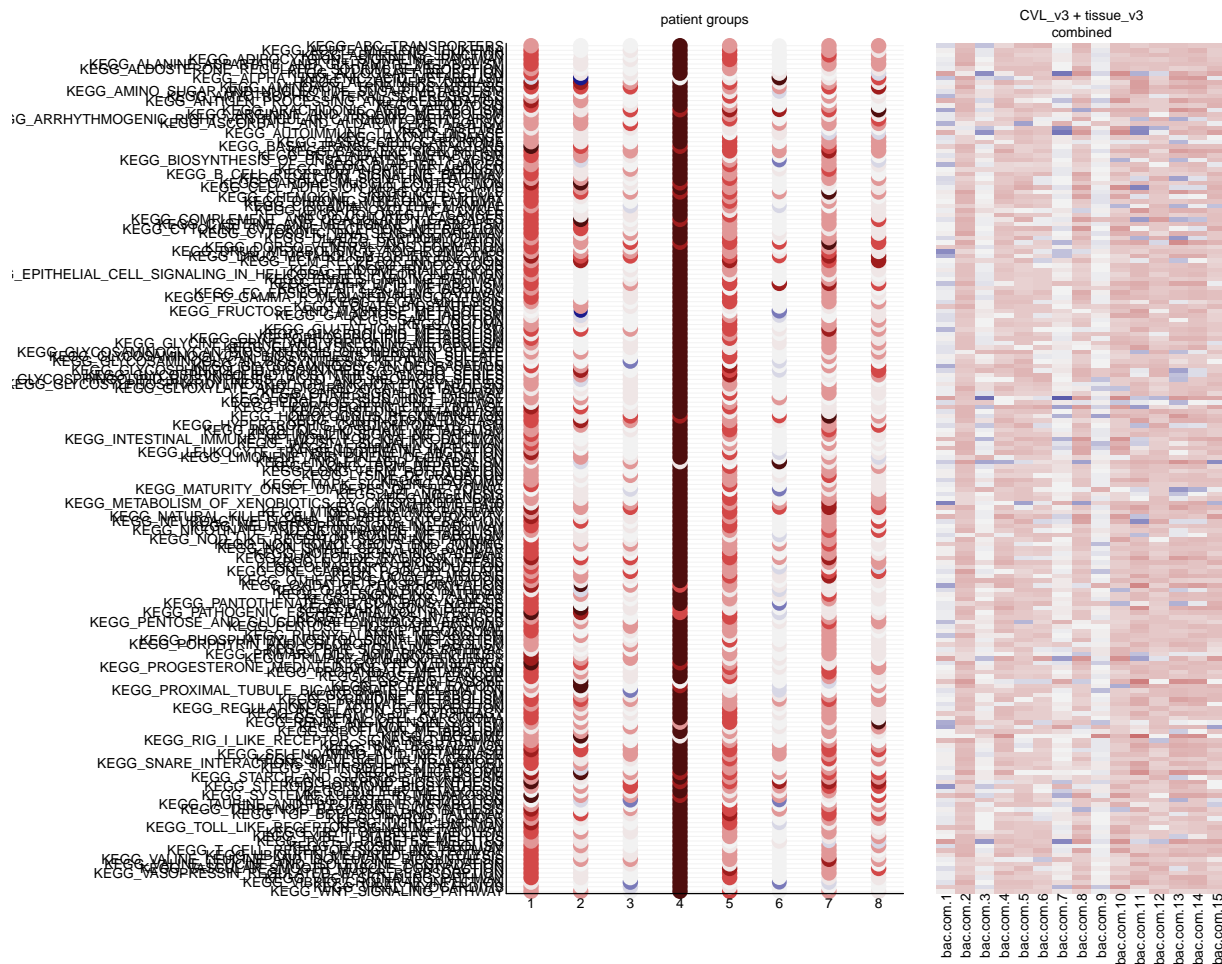


Figure 9. Heatmaps showing the functional association of microbiome datasets with the expression profiles in the RNAseq dataset. Briefly, bacterial abundances from each dataset were correlated with the gene expression of the top 5000 highly variable genes from the RNAseq dataset, generating a correlation matrix between bacteria and genes. Then for each bacteria, we rank genes based on their correlation to that bacteria and perform gene set enrichment analysis (GSEA) using the KEGG gene annotation database. This, in turn, will result in a matrix associating every bacteria with every KEGG process in the tissue. The heatmap shows the normalized enrichment score (NES). Only enrichments with pvalue below 0.05 are shown. Bacteria and pathways significant in less than 10 pathways and bacteria, respectively, were omitted.