

Broliden_5325

11 December, 2020

Contents

Loading data and metadata	1
Calculate QC metrics	1
Computing differential expression across microbiome datasets	10
#Load libraries and other scripts	
#Defining some variables for the analysis	

Loading data and metadata

Calculate QC metrics

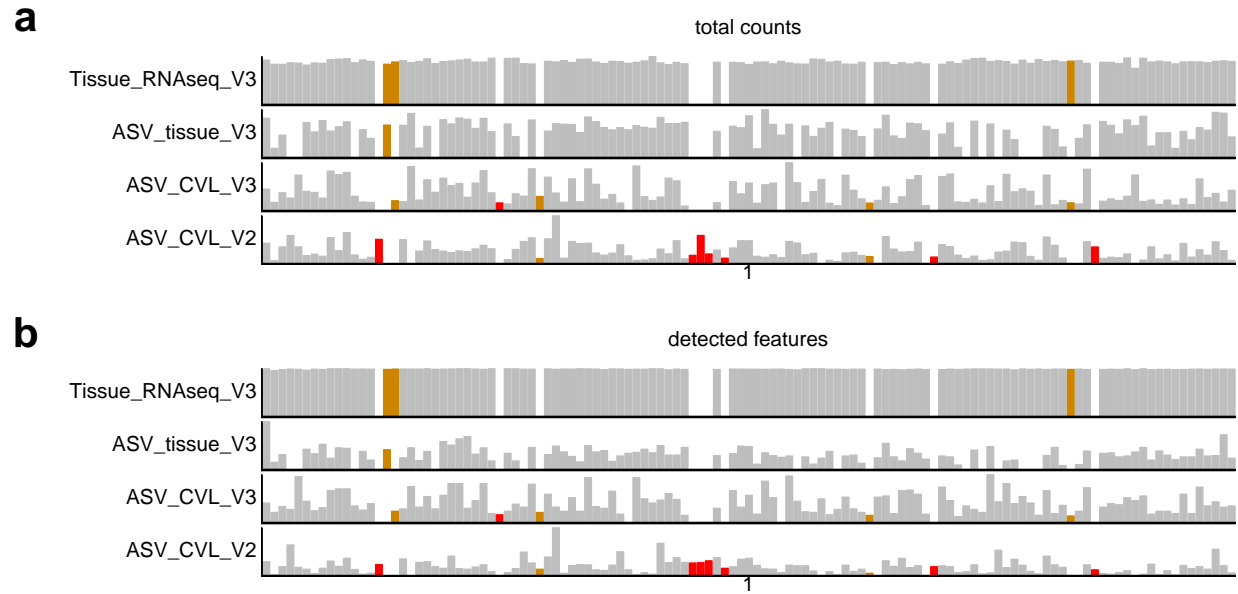
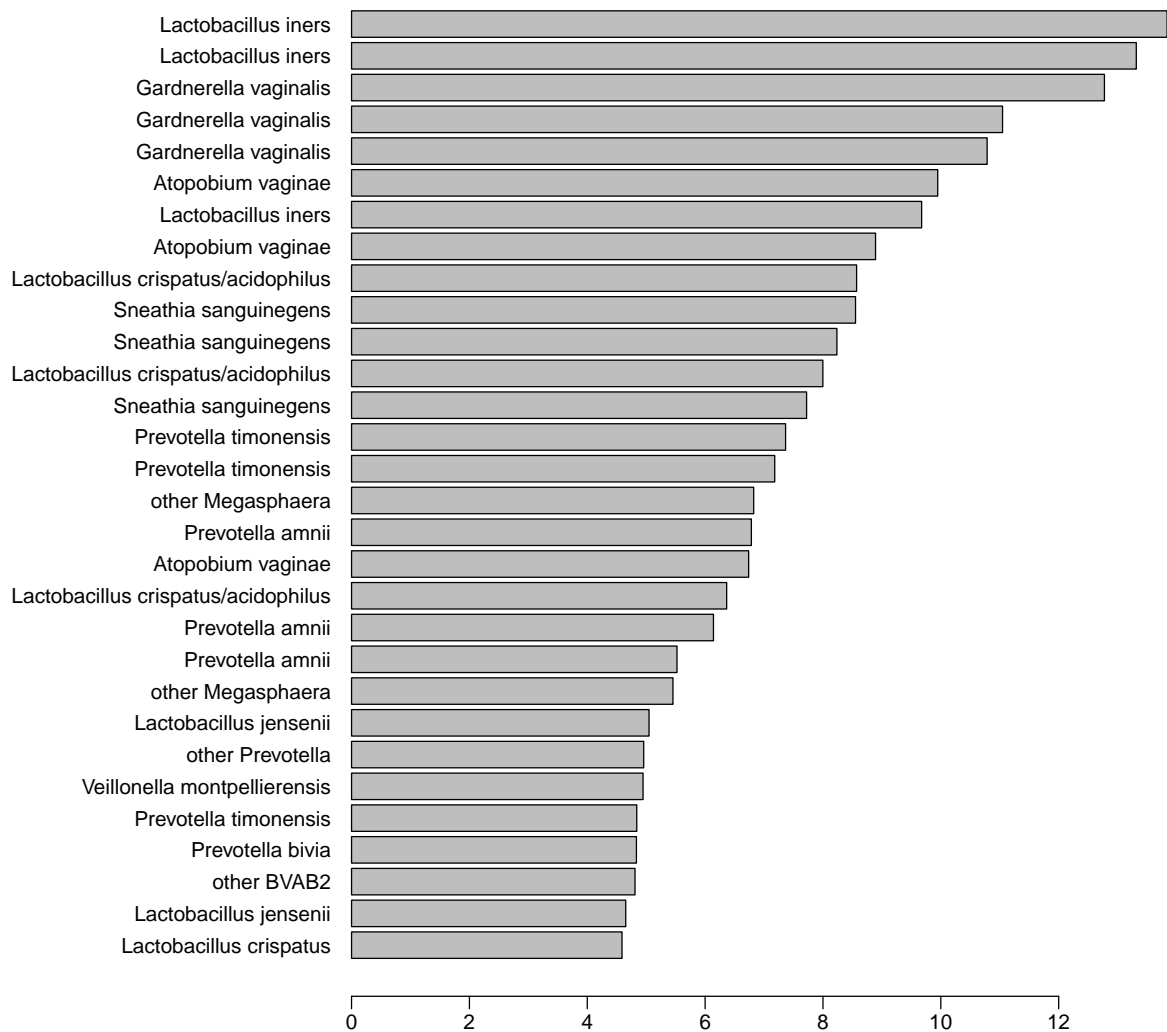


Figure 1. Comparative barplot for the **a)** total counts and **b)** number of non-zero detected features (genes / bacteria) for each of the sequencing datasets. Samples are ordered alfabetically according to the patient ID. Samples that are present in exactly two datasets are shown in orange (P016,P017,P035,P076,P101).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



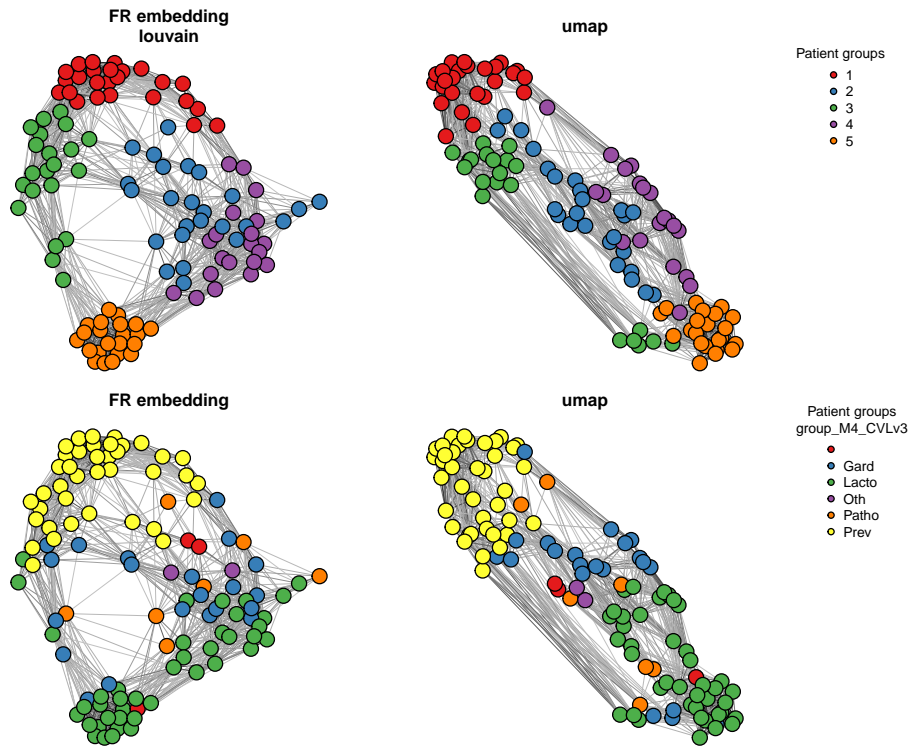


Figure 1b. Sample embedding of 20-SNN graph clustered using Louvain based on the normalized bacterial abundances across all microbiome datasets (including samples not present in all datasets).

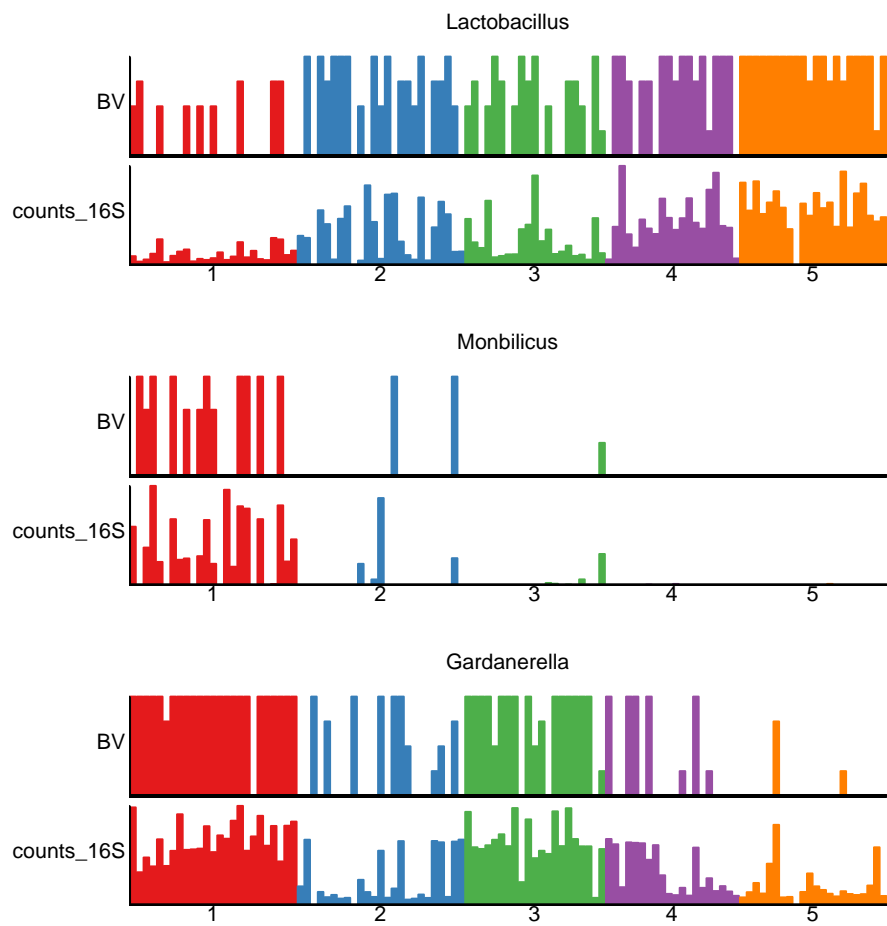


Figure 2. Differential bacterial abundance across joint clustered samples and compared across microbiome datasets. The results are shown both as **a)** Dot plots and **b)** barplots Bacteria with \log_2FC above 0.25 and p-value below 0.01 were considered significant and were sorted by the highest expression.

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 \log_2FC 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.

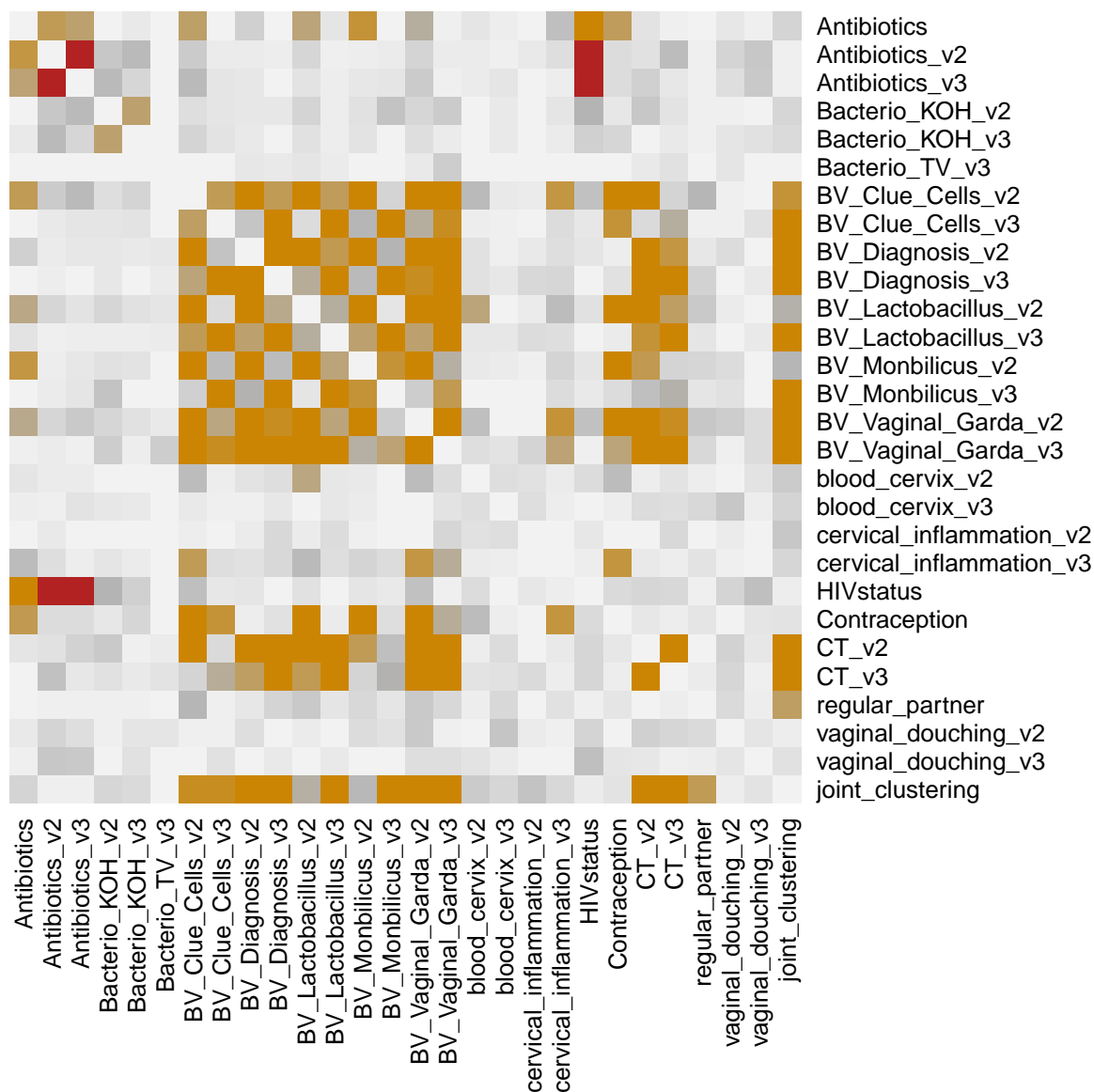


Figure 6. Association analysis across several patient categorical metadata parameters, including patient groupings annotations from microbiome.

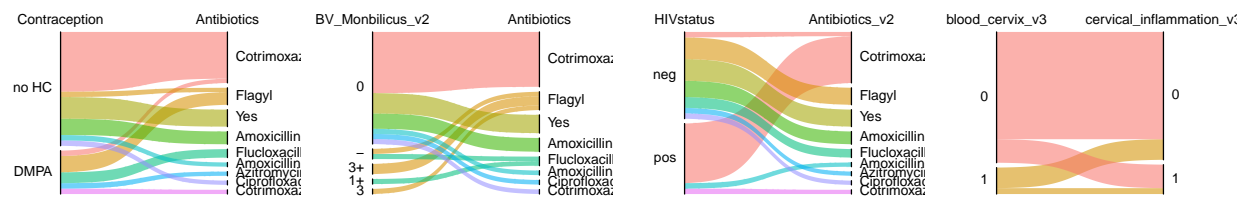
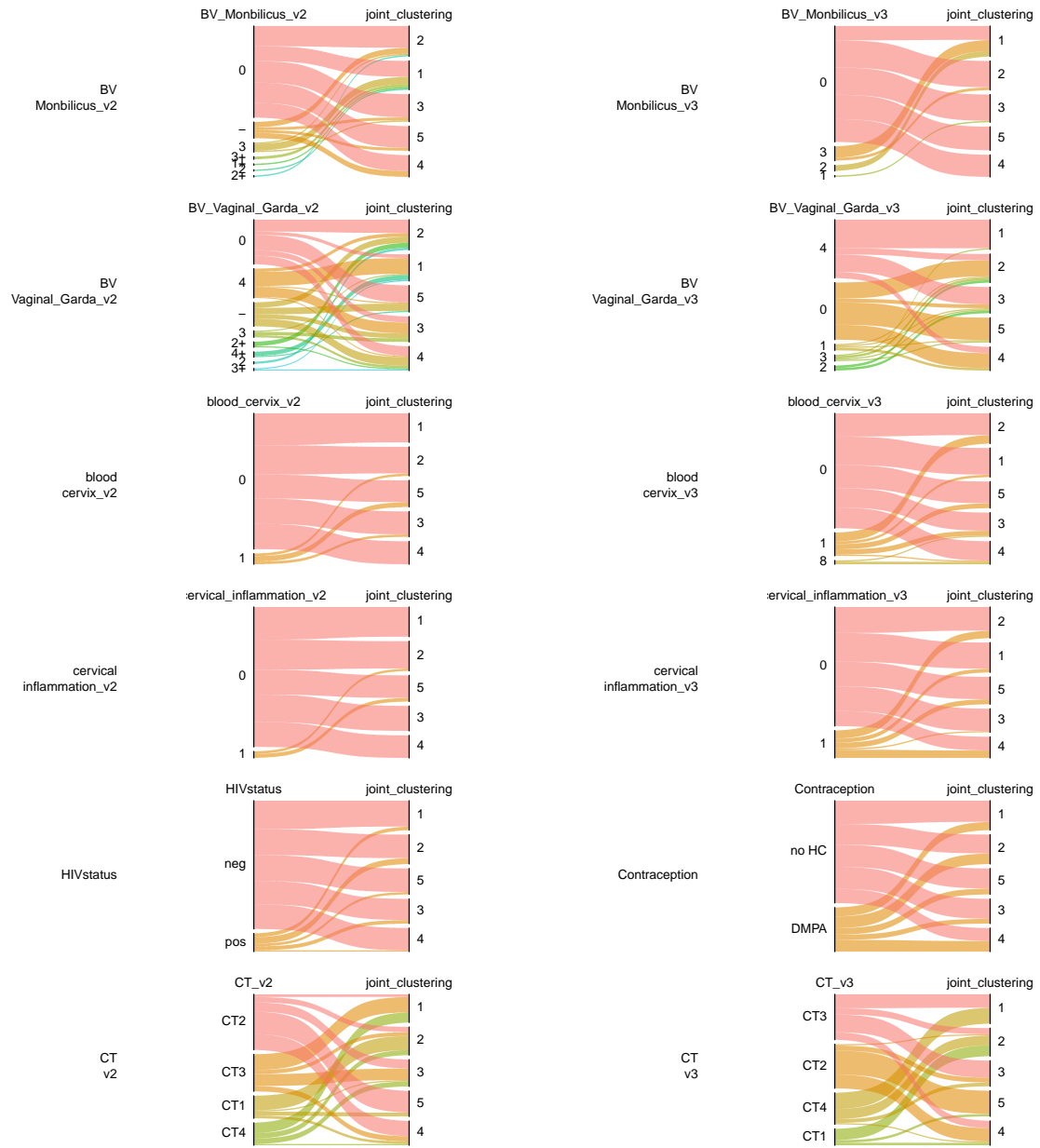


Figure 7. A few examples of significant association between metadata parameters shown as sankey plots.



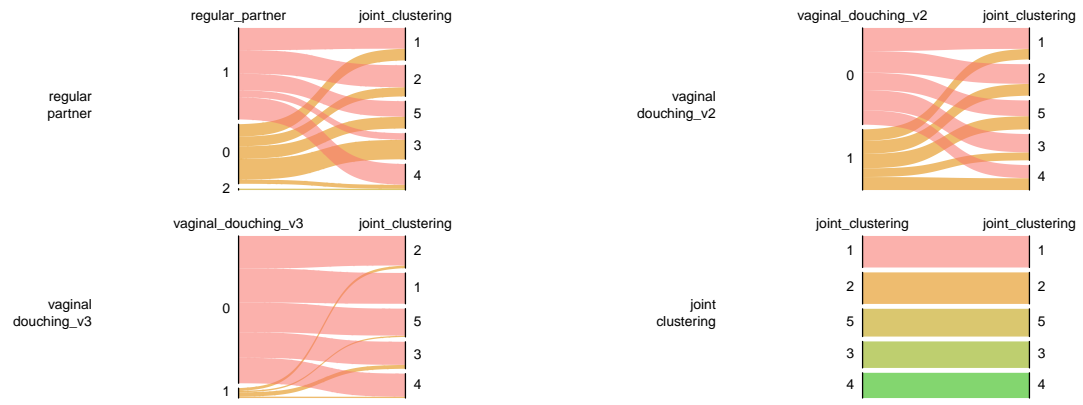


Figure 8. Sankey plots for all tested associations between the patient groups identified in in the microbiome datasets.

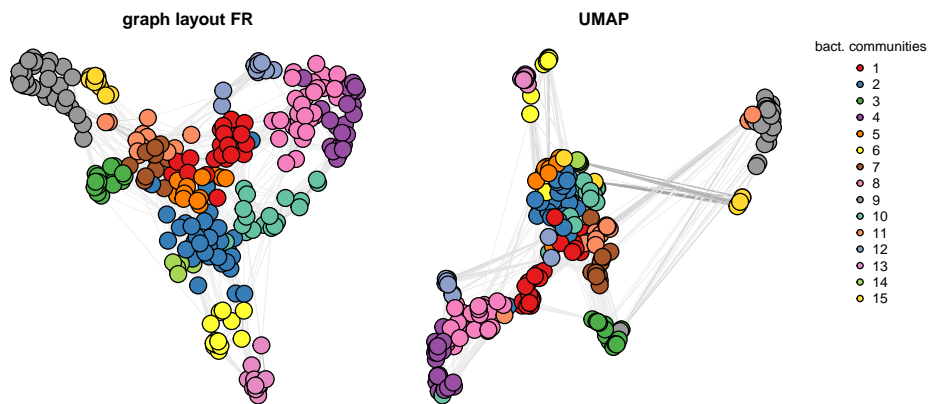


Figure 9a. Bacterial community embedding of 20-SNN graph clustered using walktrap community detection algorithm based on the normalized bacterial abundances across all microbiome datasets (including samples not present in all datasets).

community1	community2	community3	community4	community5	community6
Actinomyces neui Actinomyces turicensis Anaerococcus murchii Corynebacterium pyruviciproducens Dialister propionificiens Dialister succinatiphilus Finegoldia magna Granulicatella elegans Mycoplasma hominis other Burkholderia other Howardella other Hydrogenophaga other Olsenella other Peptostreptococcus other Romboutsia other Veillonella Prevotella bivia Prevotella melaninogenica Slackia exigua Solobacterium moorei Streptococcus agalactiae Streptococcus macedonicus/pasteuri Varibaculum cambriense Veillonella montpellierensis	Bifidobacterium adolescentis Brevibacterium ravenburgense Campylobacter hominis Chlamydia trachomatis Corynebacterium imitans Corynebacterium jeikeium Corynebacterium simulans/striatum Corynebacterium sp/amycolatum Corynebacterium spheniscorum Dermabacter hominis Dorea formicigenerans Gemmiger formicilis Lactobacillus crispatus Lactobacillus mucosae Lactobacillus reuteri Lactobacillus salivarius Megaspheara elsdeni other Bacteroidales other Bacteroidetes other Eubacterium other Ezakiella other Lactobacillales other Lactobacillus other Listeria other Murdochella other Nesterenkonia Prevotella denticola Staphylococcus saccharolyticus/capi Ureaplasma parvum Corynebacterium appendicis/imitans Methylobacterium radiotolerans other Bacteroides other Ruminococcus	Bacteria Bacteroides caccae Coprococcus comes Delftia tsuruhatensis other Aeromonas other Aquaspirillum Corynebacterium simularum other Comamonas other Enterobacter other Janthinobacterium other Neisseria other Pelomonas other Ralstonia other Schlegella other Staphylococcus other Streptococcus other Veillonellaceae Stenotrophomonas maltophilia Thermus scotoductus	Actinomyces hongkongensis Anaerococcus prevotii/tetradis Atopobium vaginale Gardnerella vaginalis Gemella asaccharolytica Mobiluncus mulieris other Arcanobacterium other Bulleidia other BVAB1 other BVAB2 other BVAB3 other Coriobacteriaceae other Dalister other Falsiporphyrionas other Gardnerella other Megasphaera other Parvimonas other Prevotella Peptoniphilus harei Peptoniphilus lacrimalis Peptostreptococcus anaerobius Prevotella amnii Prevotella timonensis Sneathia sanguinegens	Anaerococcus hydrogenalis Bifidobacterium bifidum Corynebacterium pseudogenitalium other Gordonia other Mogibacterium other Rhizobium Streptococcus dysgalactiae Streptococcus lutetiensis/equinus Brevundimonas mediterranea/interm Corynebacterium freneyi/sp/amycola Enterococcus villorum/hirae/faecium/ other Gammaproteobacteria	Bacteroides dorei Bacteroides vulgatus Bifidobacterium pseudocatenulatum/ Blautia obeum Dorea longicatena Escherichia/Shigella coli/flexneri/boy Faecalibacterium prausnitzii other Bacillus other Bradyrhizobium other Holdemania other Roseburia other XIVa Roseburia faecis
community7	community8	community9	community10	community11	community12
Collinsella aerofaciens Corynebacterium tuberculoostearicum Enhydrobacter aerosaccus Gemella morbillorum/sanguinis/haem Haemophilus parainfluenzae other Lactococcus other Rhodobacteriaceae other Streptophyta other XI Tapidomonas ardensis Veillonella dispar Veillonella parvula	Actinomyces urogenitalis Anaerococcus obsiensis Dialister microaerophilus Mobiluncus curtisii Moraxella indologenes other Anaerococcus other Barnesiella other Clostridiales other Finegoldia other Lachnospiraceae other Parvibacter other Peptococcus other Peptoniphilus other Porphyromonas other Sutterella other Syntrophococcus Peptococcus niger Peptoniphilus coxi Peptoniphilus duerdenii Peptoniphilus massiliensis Peptostreptococcus stomatis Porphyromonadaceae bacterium Porphyromonas asaccharolytica Porphyromonas uenonis Prevotella bergensis Prevotella buccalis Prevotella corporis Prevotella disiens Prevotella sp Sutterella sanguinis	Brachyobacterium paraconglomeratum Chryseobacterium hallense Corynebacterium freneyi Cupriavidus metallidurans Mesorhizobium loti other Acinetobacter other Actinomyces other Bergeyella other Brachyobacterium other Chryseobacterium other Clostridiaceae other Dietzia other Exiguobacterium other Flavobacteriaceae other Intrasporangiaceae other Janibacter other Kocuria other Leuconostoc other Micrococcus other Nitrospina other Nocardioides other Ornithinimicrobium other Parafilimonas other Pseudomonas other Rhodococcus other Riemerella other Rothia other Sphingomonas other Tetrasphaera other Xanthomonadaceae	Aerococcus christensenii Atopobium deltae Bifidobacterium longum/breve Brevibacterium paucivorans Corynebacterium aurimucosum Corynebacterium genitalium Corynebacterium sundsvallense/thor Enterococcus villorum/hirae/faecium/ Lactobacillus gasserii/johnsonii/taivai other Atopobium other Bacilli other Bifidobacteriaceae other Clostridia other Corynebacterium other Gemella other Mobiluncus other Paraeggerthella Peptoniphilus asaccharolyticus	Lactobacillus plantarum/fabiferment other Actinomycetales other Aerococcus other Facklamia other Firmicutes other Methylobacterium other Novosphingobium other Porphyromonadaceae other stricto other Succinivibrio Porphyromonas somerae Ruminococcus bromii	Anaerococcus lactolyticus Campylobacter ureolyticus Fusobacterium equinum Fusobacterium nucleatum other Campylobacter other Paraprevotella other Prevotellaceae other Proteobacteria other Sneathia Streptococcus infantilis/oralis/mitis Streptococcus salivarius/vesibularis/
community13	community14	community15			
Anaerostipes hadrus Lactobacillus vaginalis other Alloprevotella other Blautia other Faecalibacterium other IV other Oscillibacter other Ruminococcaceae other Ruminococcus2 Prevotella copri Streptococcus intermedius/constellat Fusicatenibacter saccharivorans	Lactobacillus coleohominis Lactobacillus crispatus/acidophilus Lactobacillus iners Lactobacillus jensenii Lactobacillus reuteri/oris/frumentii/ant other Arthrobacter other Bacillales other Paracoccus other Phyllobacteriaceae other Psychrobacter Microbacterium aoyamense/paludico other Aeromicrobium other Brevundimonas other Neorhizobium other Sphingobacterium				

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

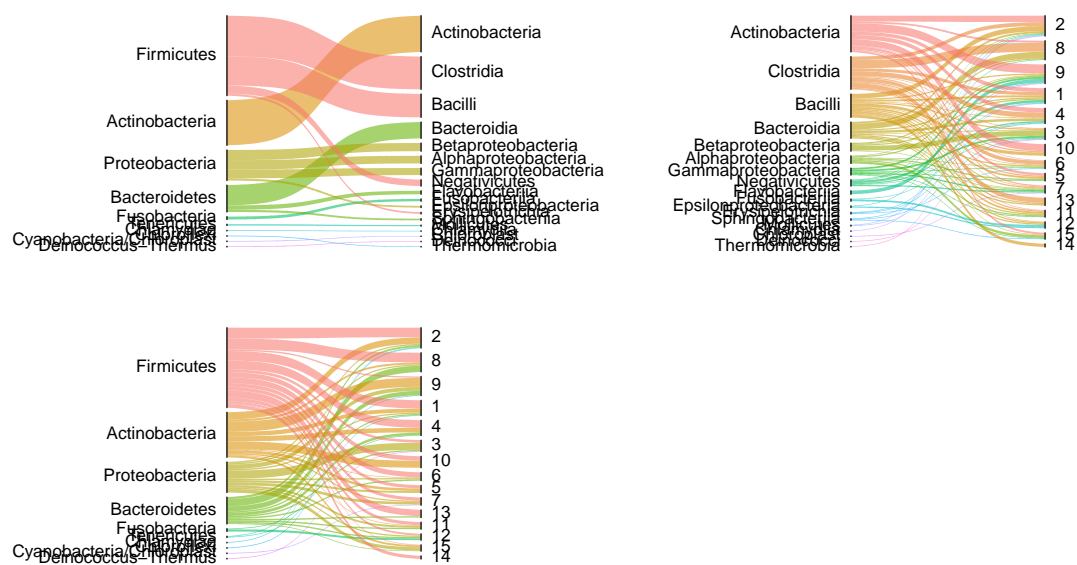


Figure 9b. Taxonomic annotation of bacterial communities.

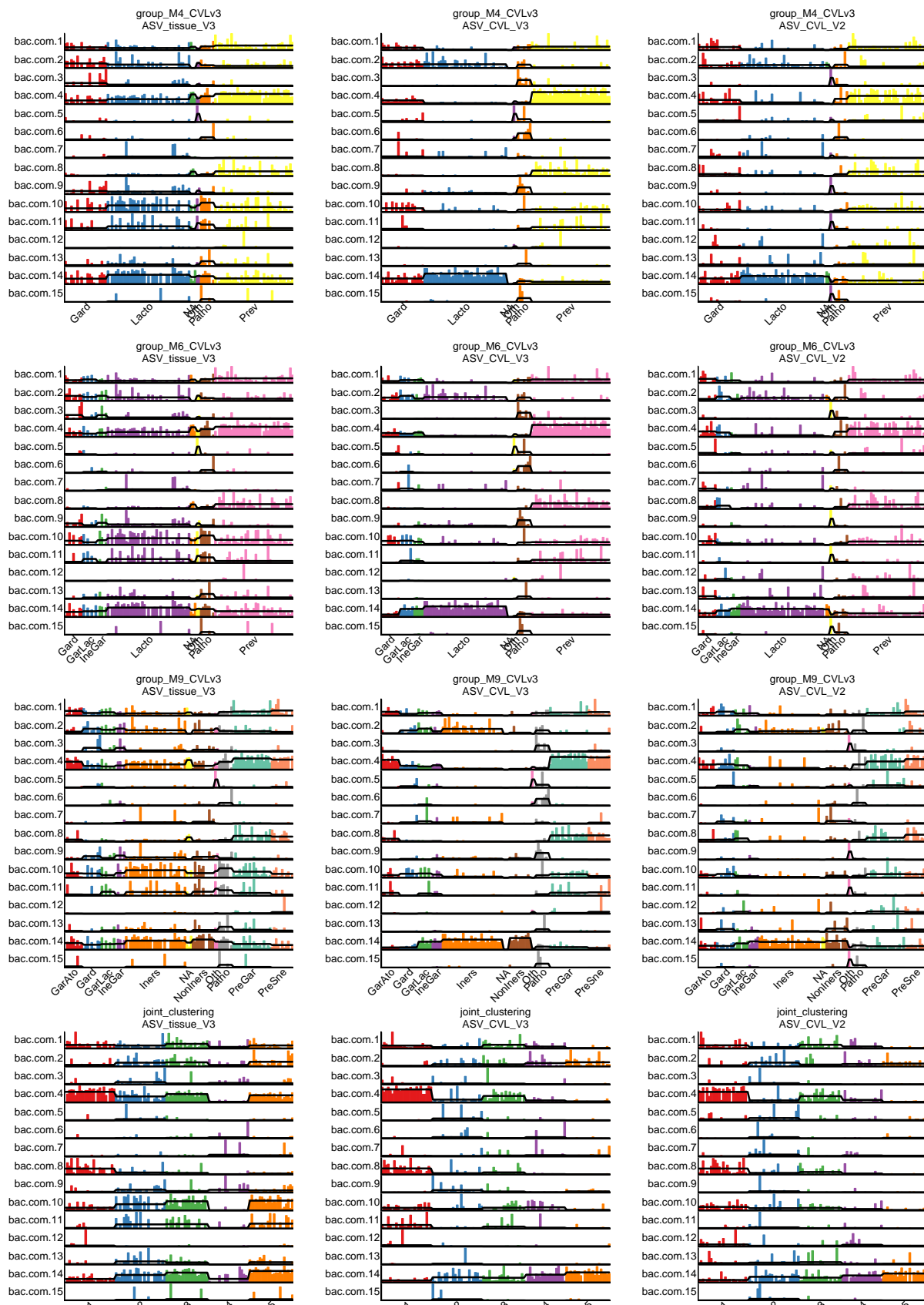
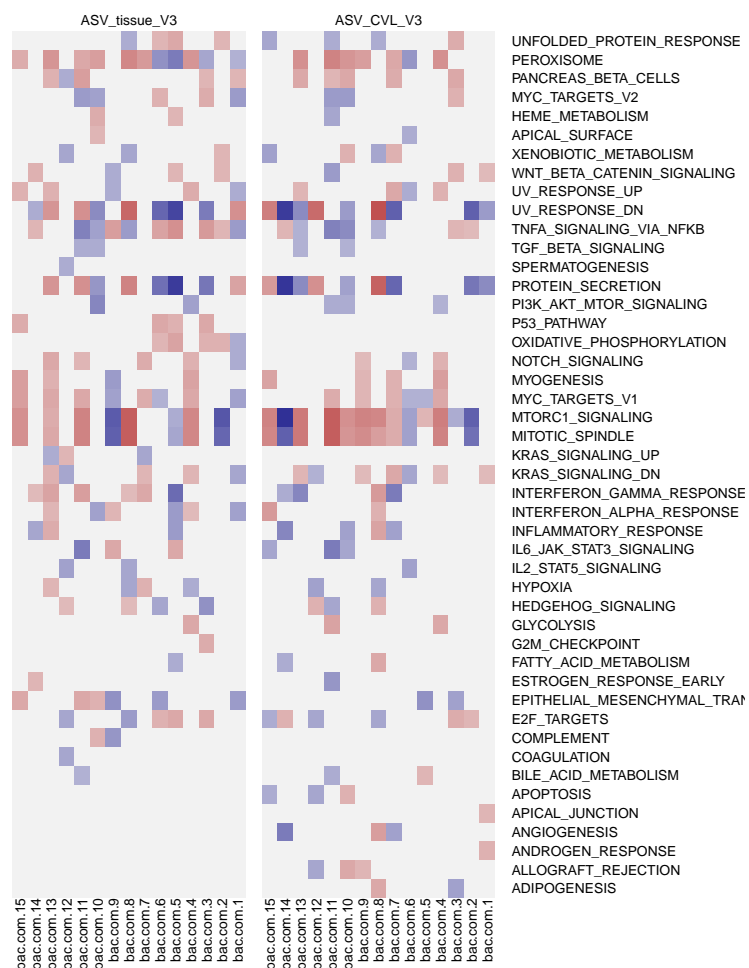
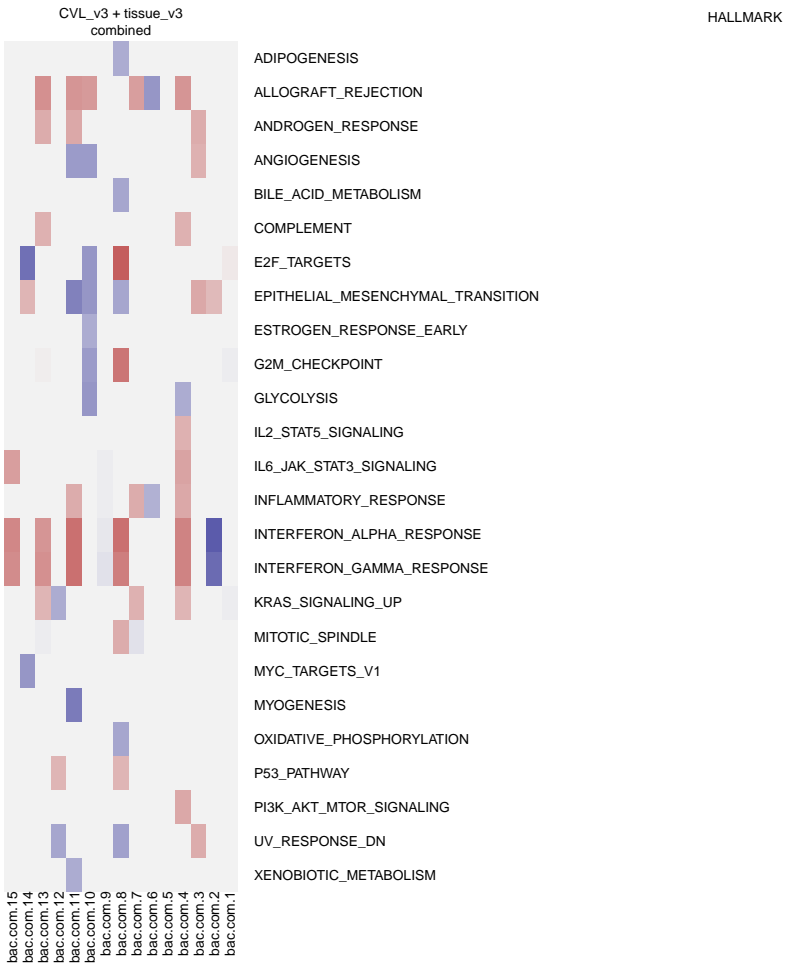


Figure 5. Comparisson of microbiome communities across all datasets and all clustering methods.





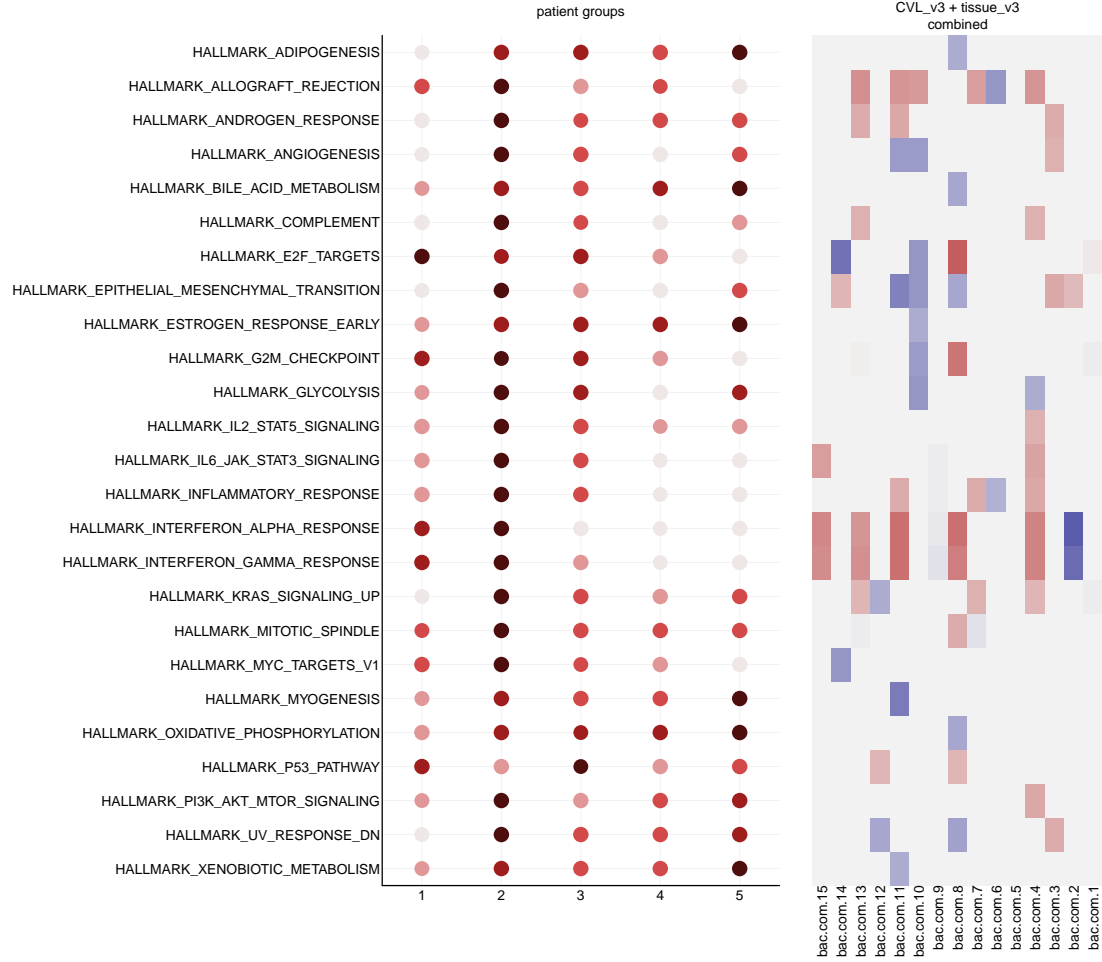


Figure 9. Heatmaps showing the functional association of microbiome datasets with the expression profiles in the RNAseq dataset. Briefly, bacterial abundances from each dataset were correlated with the gene expression of the top 5000 highly variable genes from the RNAseq dataset, generating a correlation 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.

