

Broliden_5325

24 November, 2020

Contents

Loading data and metadata	1
Calculate QC metrics	1
Computing differential expression across microbiome datasets	8
Visualise the data	18
#Load libraries and other scripts	
#Defining some variables for the analysis	

Loading data and metadata

Calculate QC metrics

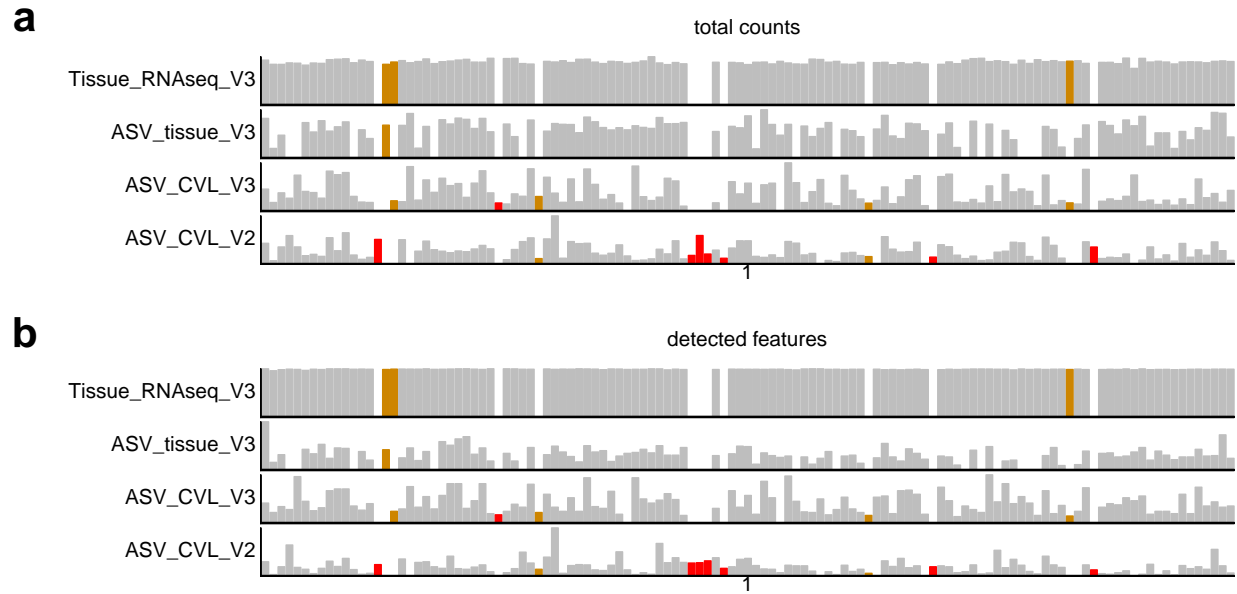


Figure 1. Comparative barplot for the **a)** total counts and **b)** number of non-zero detected features (genes / bacteria) for each of the sequencing datasets. Samples are ordered alfabetically according to the patient ID. Samples that are present in exactly two datasets are shown in orange (P016,P017,P035,P076,P101).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

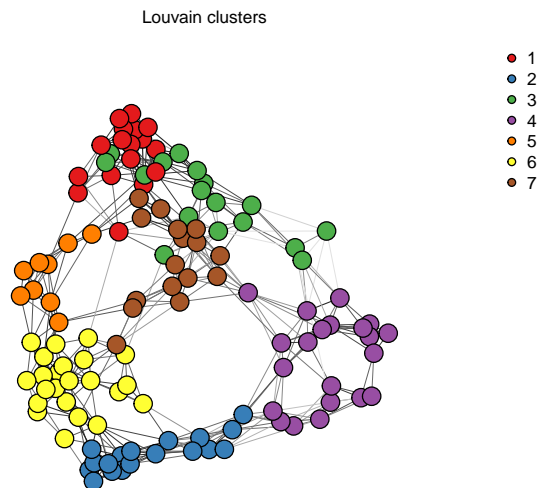


Figure 1b. Sample embedding of 5-SNN graph clustered using Louvain based on the normalized bacterial abundances across all microbiome datasets (including samples not present in all datasets).

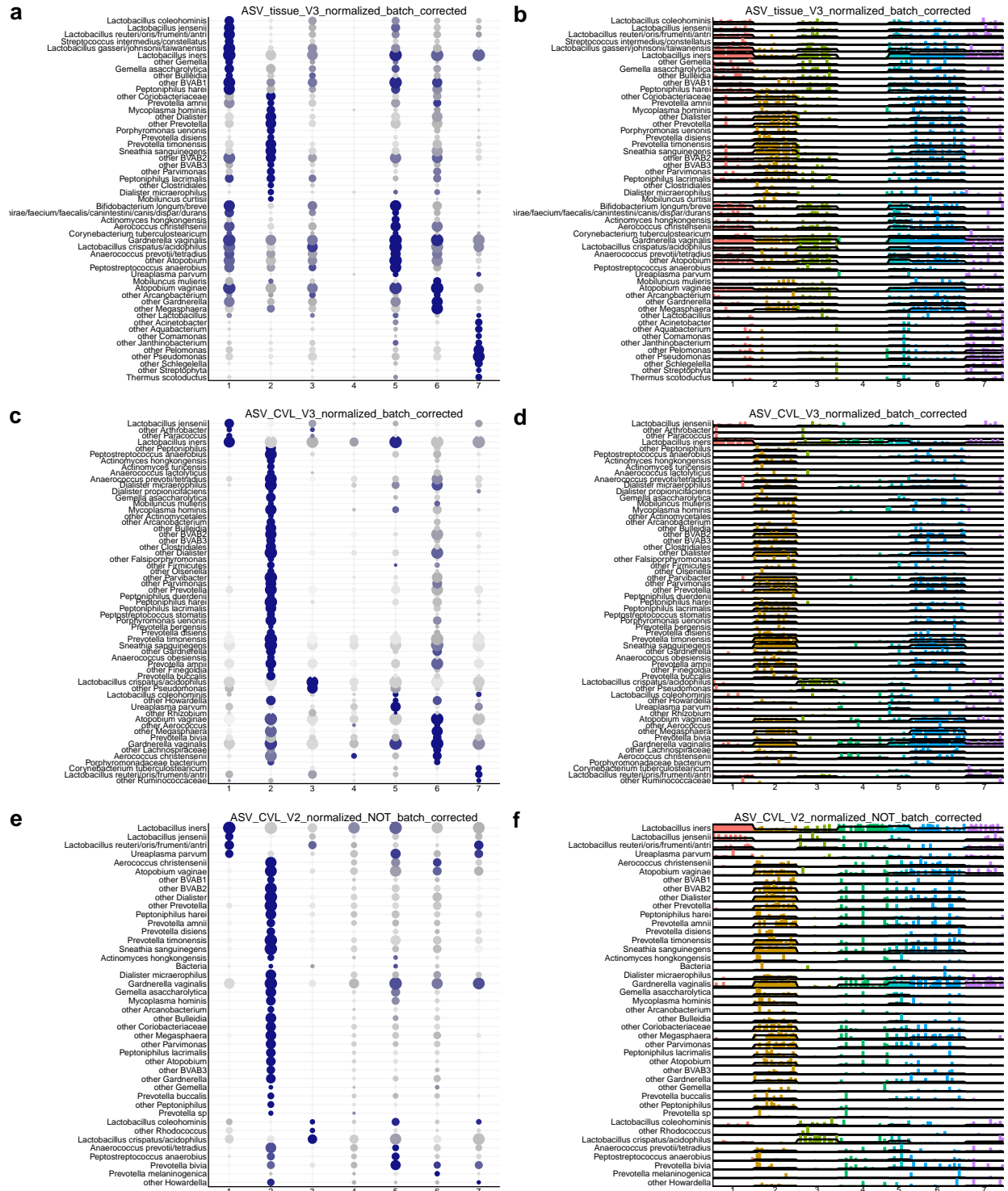


Figure 2. Differential bacterial abundance across joint clustered samples and compared across microbiome datasets. The results are shown both as **a)** Dot plots and **b)** barplots. Bacteria with log2FC above 0.25 and p-value below 0.01 were considered significant and were sorted by the highest expression.

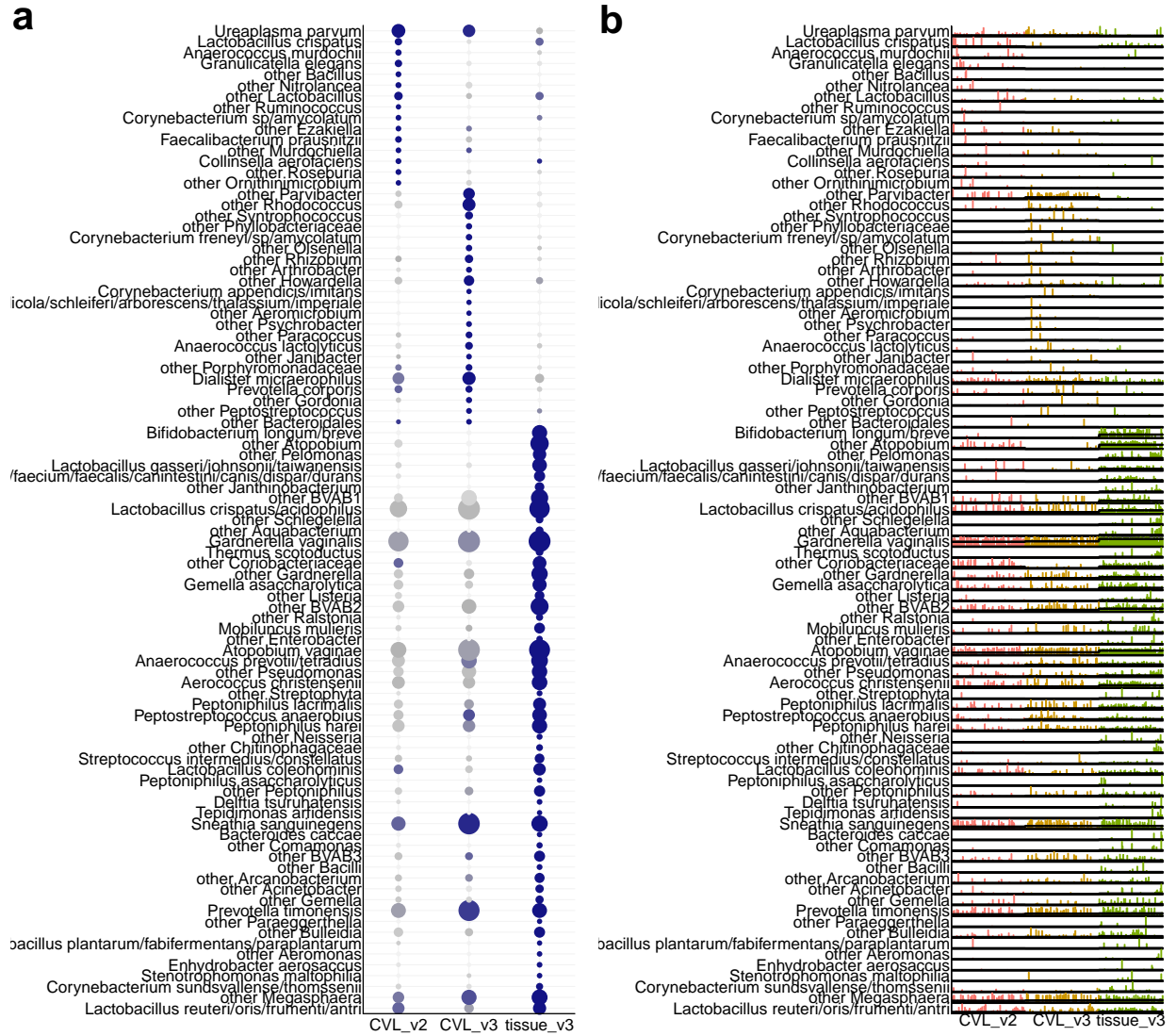


Figure 2. Differential bacterial abundance across microbiome datasets. The results are shown both as a) Dot plots and b) barplots. Bacteria with log₂FC above 0.25 and p-value below 0.01 were considered significant and were sorted by the highest expression.

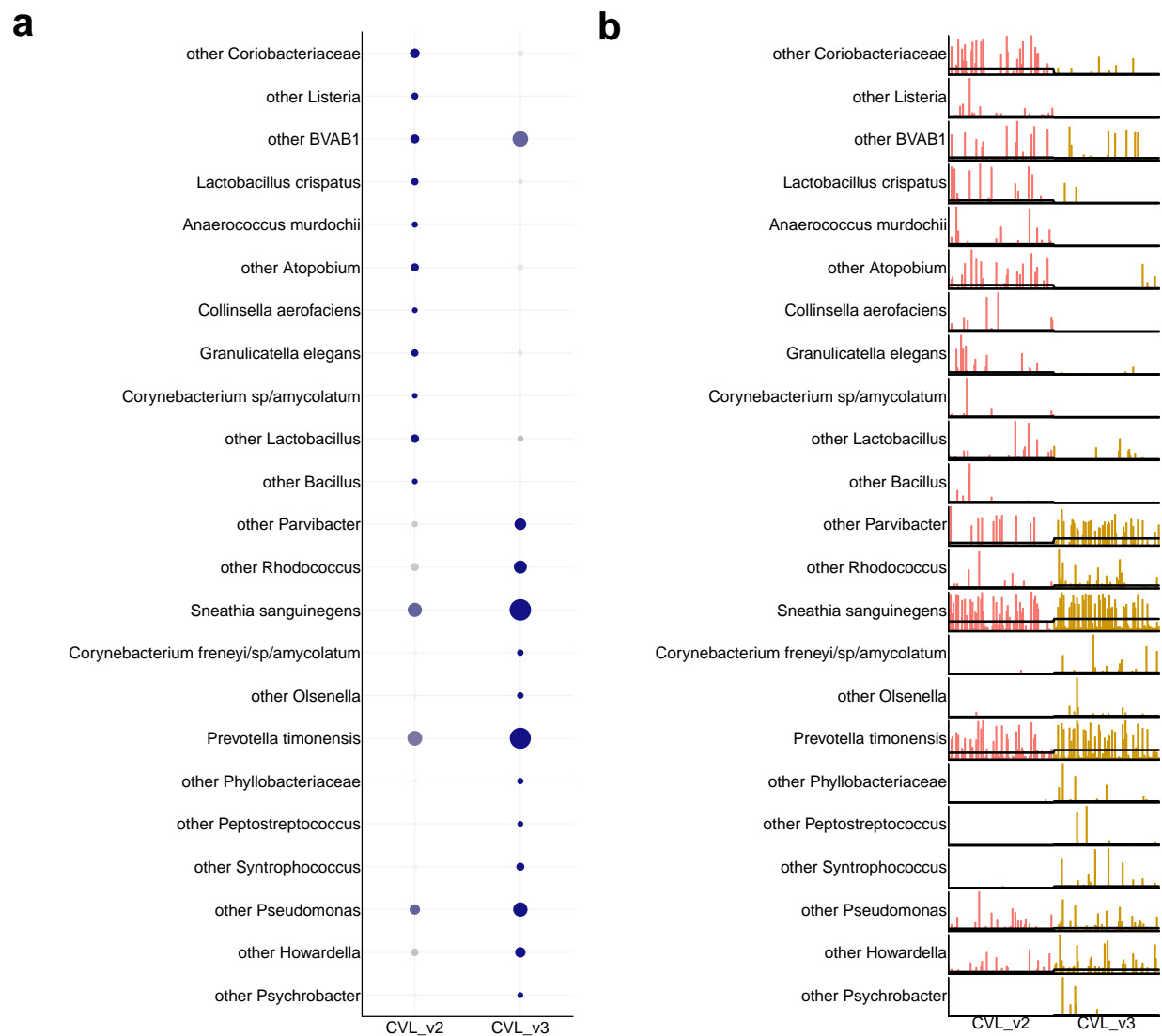


Figure 2. Differential bacterial abundance across CVL2 and CVL3 microbiome datasets. The results are shown both as **a)** Dot plots and **b)** barplots Bacteria with log2FC above 0.25 and p-value below 0.01 were considered significant and were sorted by the highest expression.

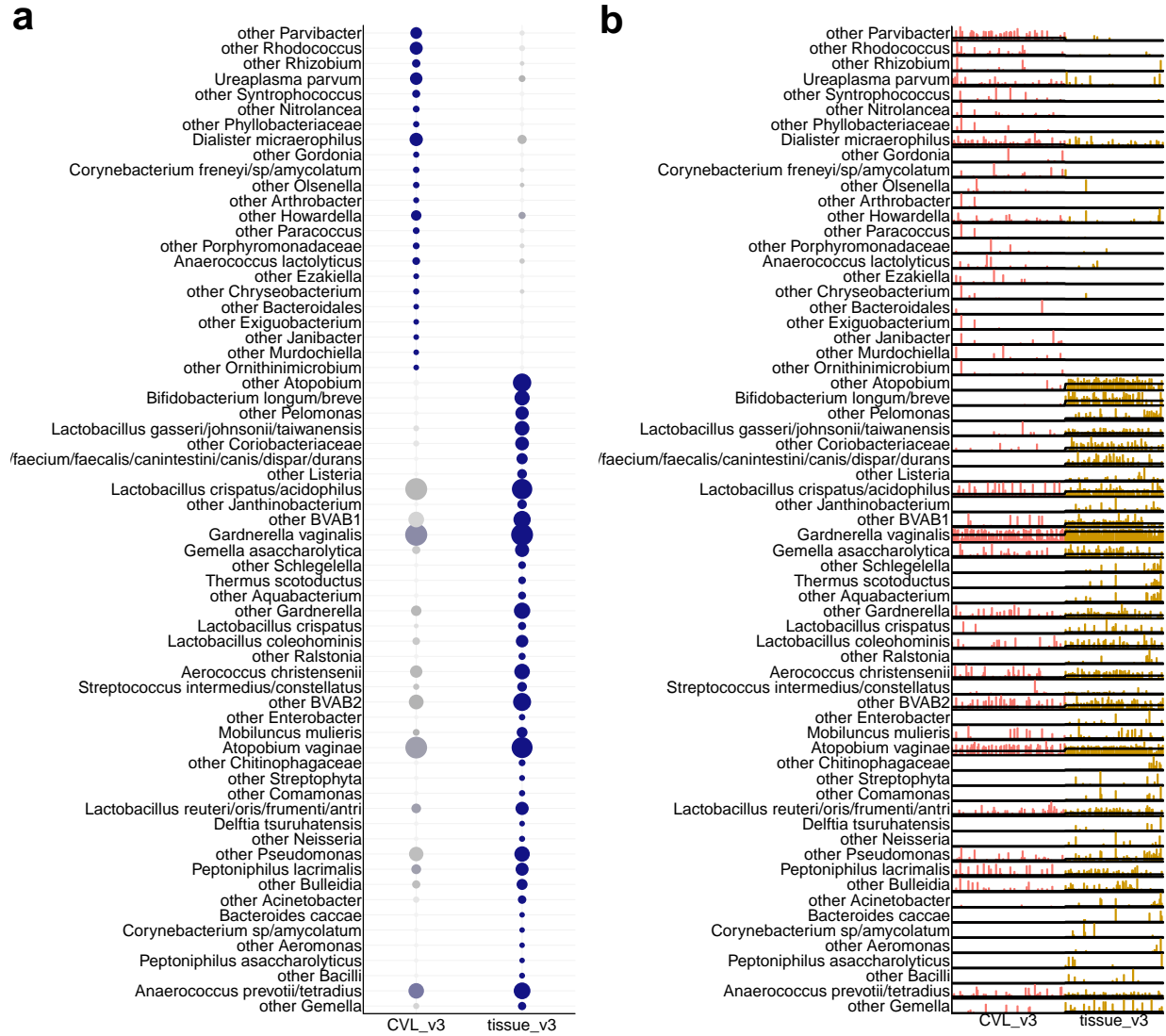


Figure 2b. Differential bacterial abundance across tissue and CVL microbiome week3 datasets. The results are shown both as **a)** Dot plots and **b)** barplots. Bacteria with log2FC above 0.25 and p-value below 0.01 were considered significant and were sorted by the highest expression.

Computing differential expression across microbiome datasets

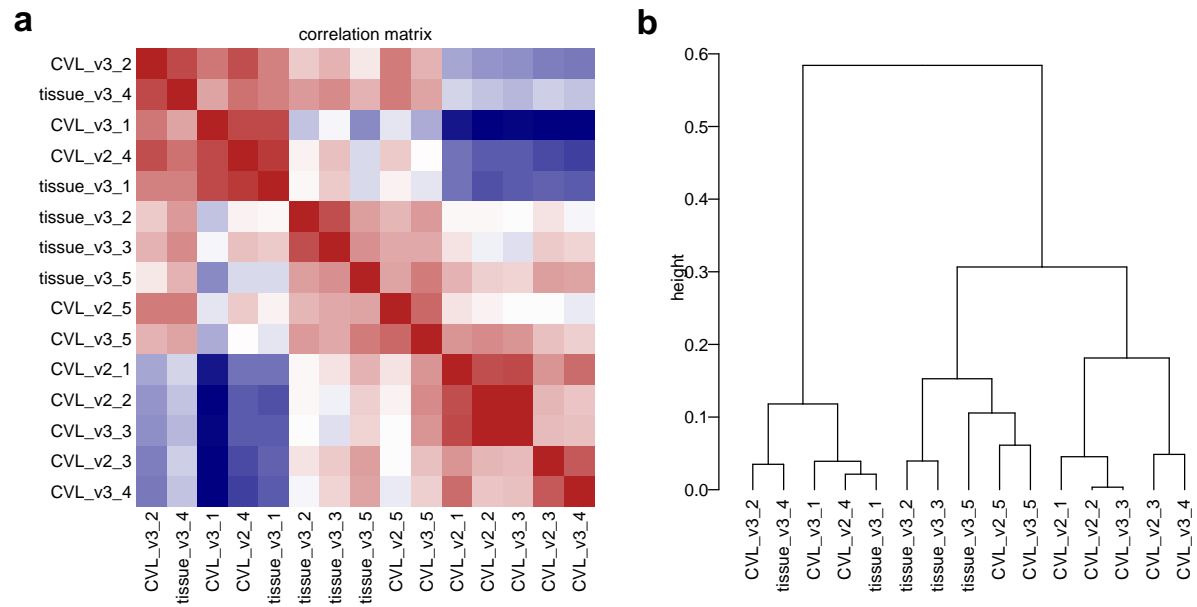


Figure 3. Comparisson among patient groups across datasets. **(a)** Correlation matrix across sample groups. **(b)** Hierarchical clustering of sample groups.

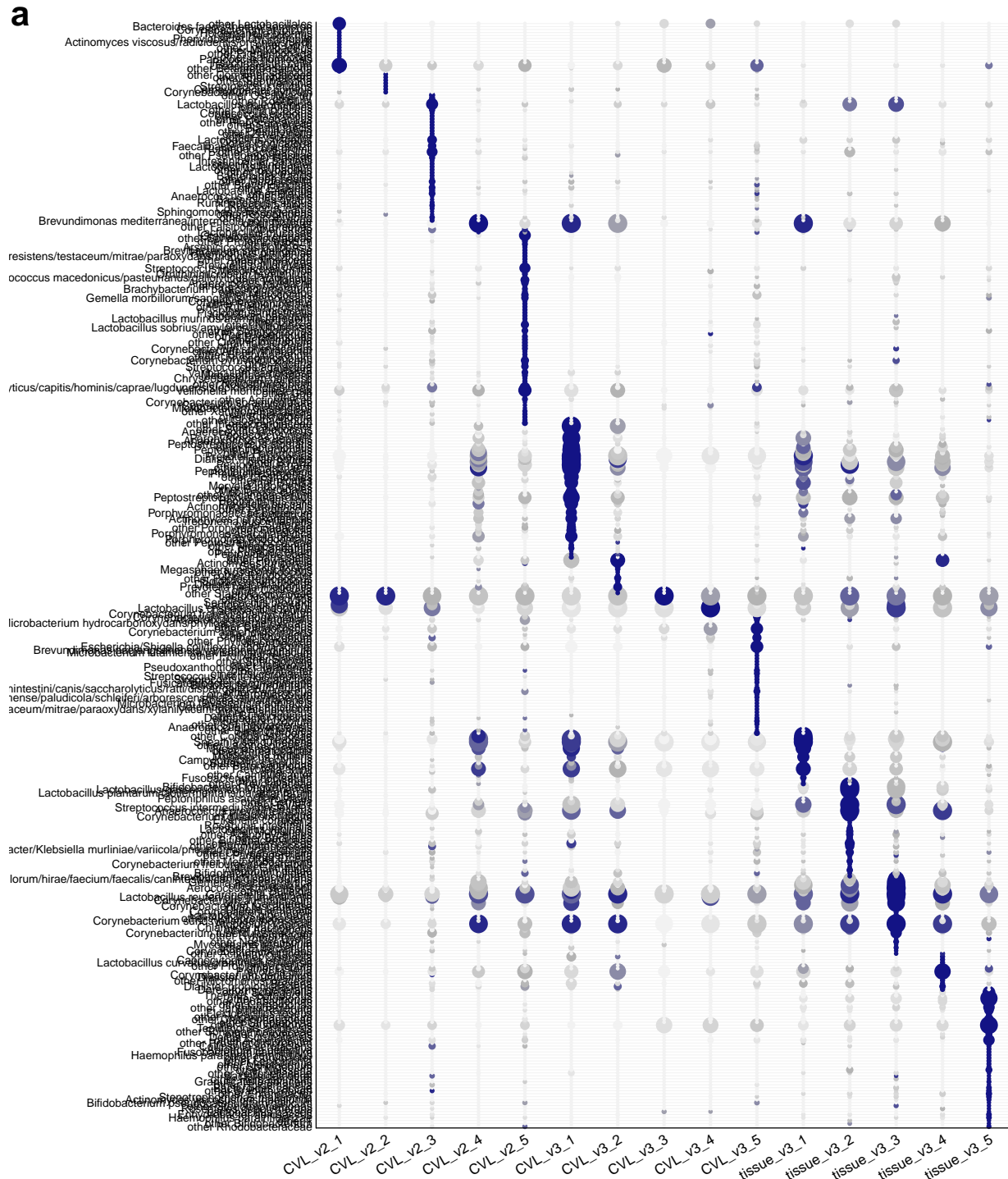


Figure 4. Differential bacterial abundance across all groups and all microbiome datasets. The results are shown both as **a)** Dot plots and **b)** barplots. Bacteria with log2FC above 0.25 and p-value below 0.01 were considered significant and were sorted by the highest expression.

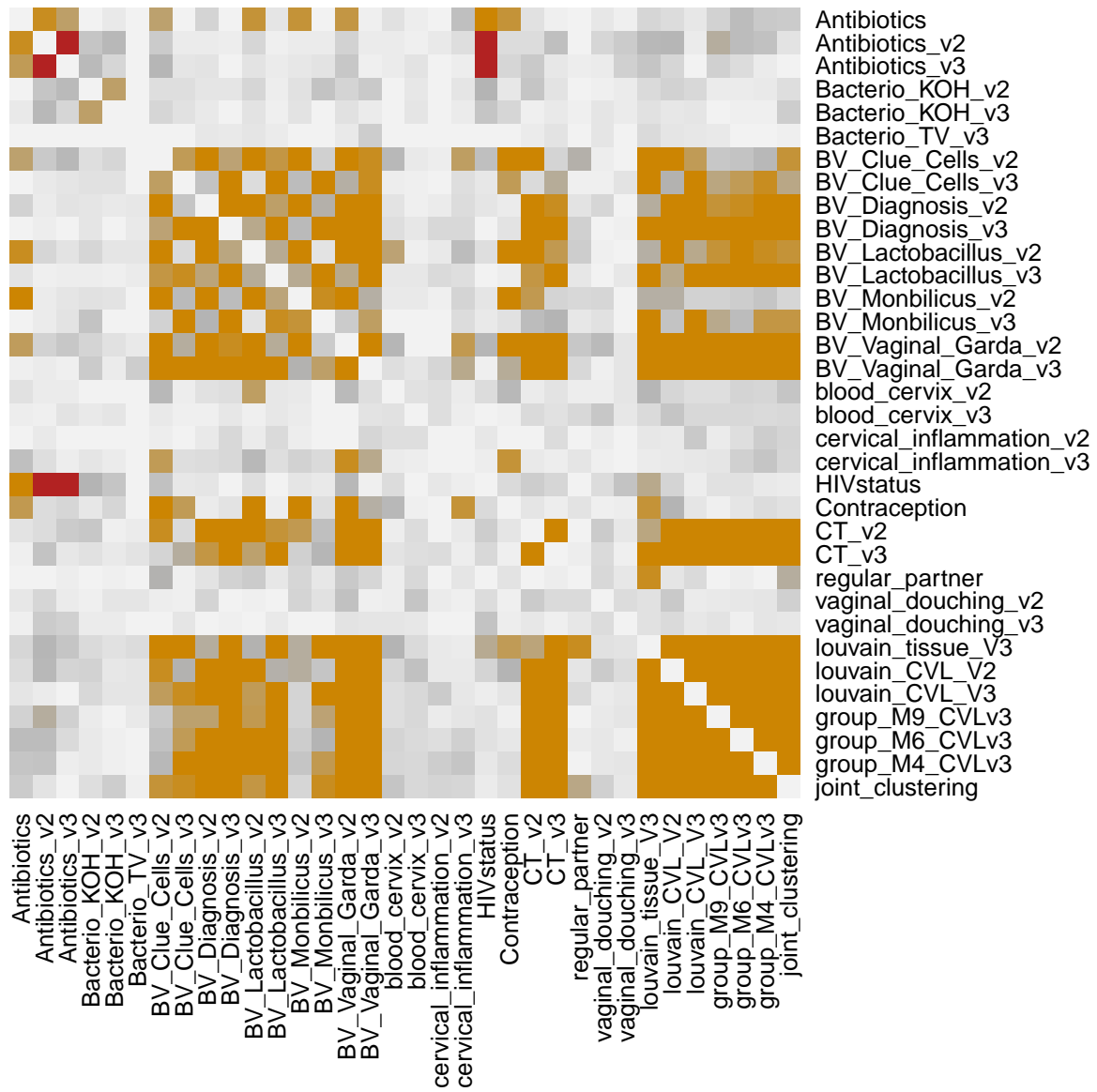


Figure 6. Association analysis across several patient categorical metadata parameters, including patient groupings annotations from microbiome.

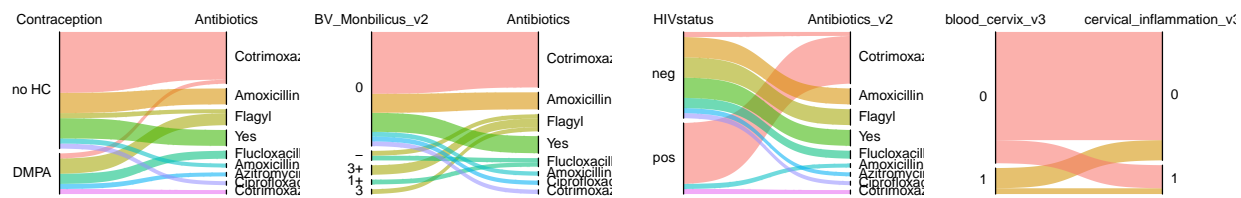
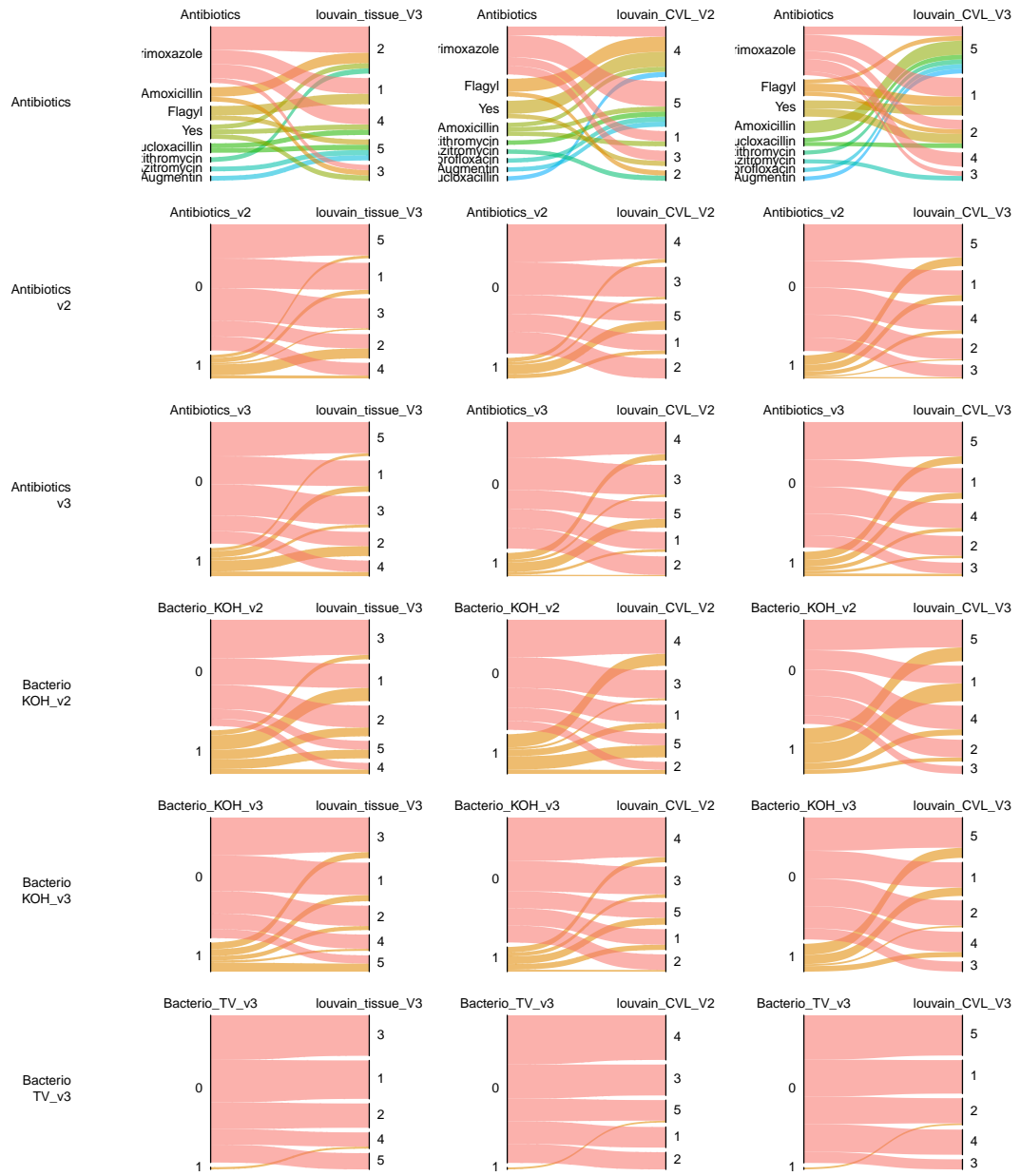
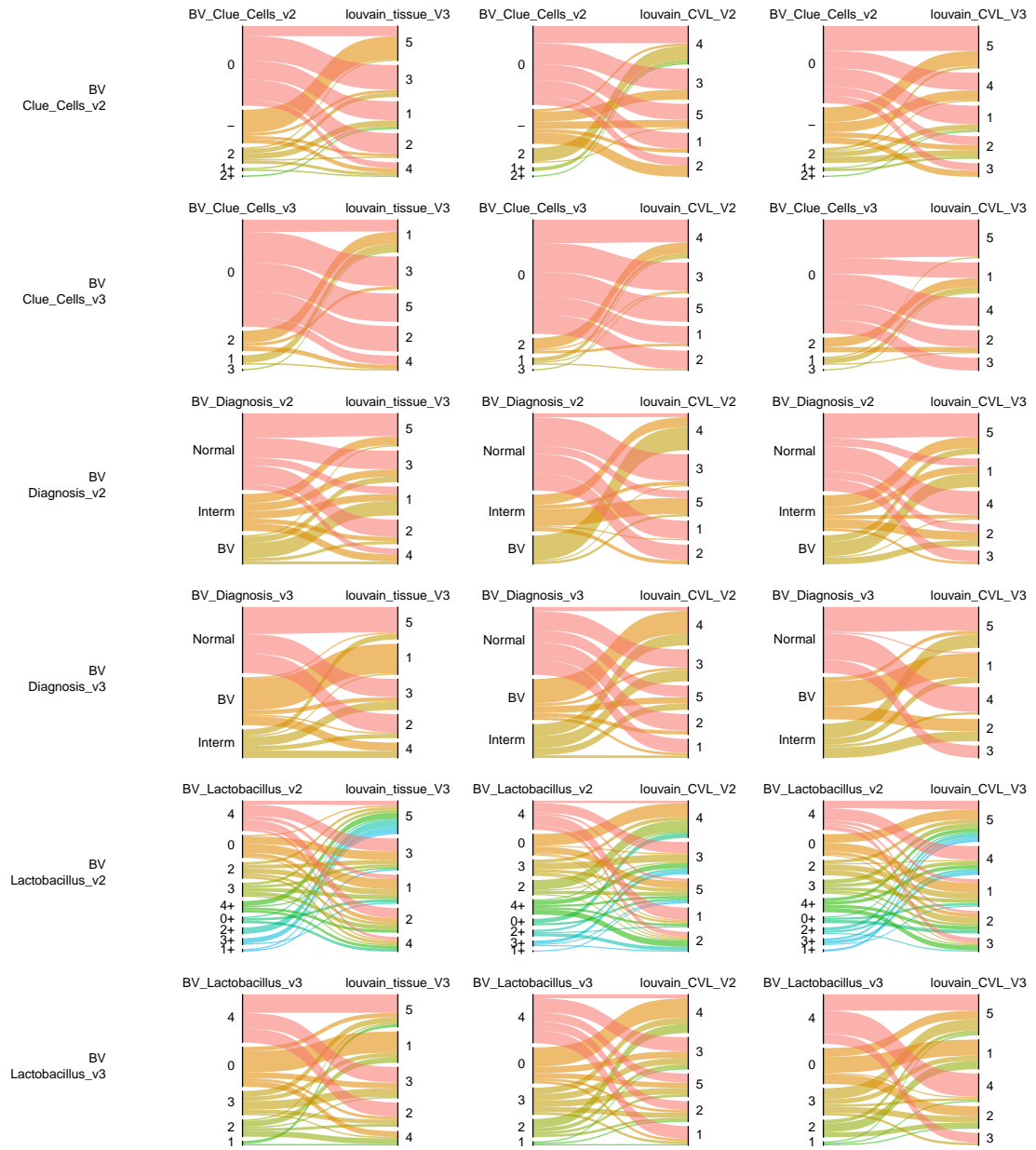
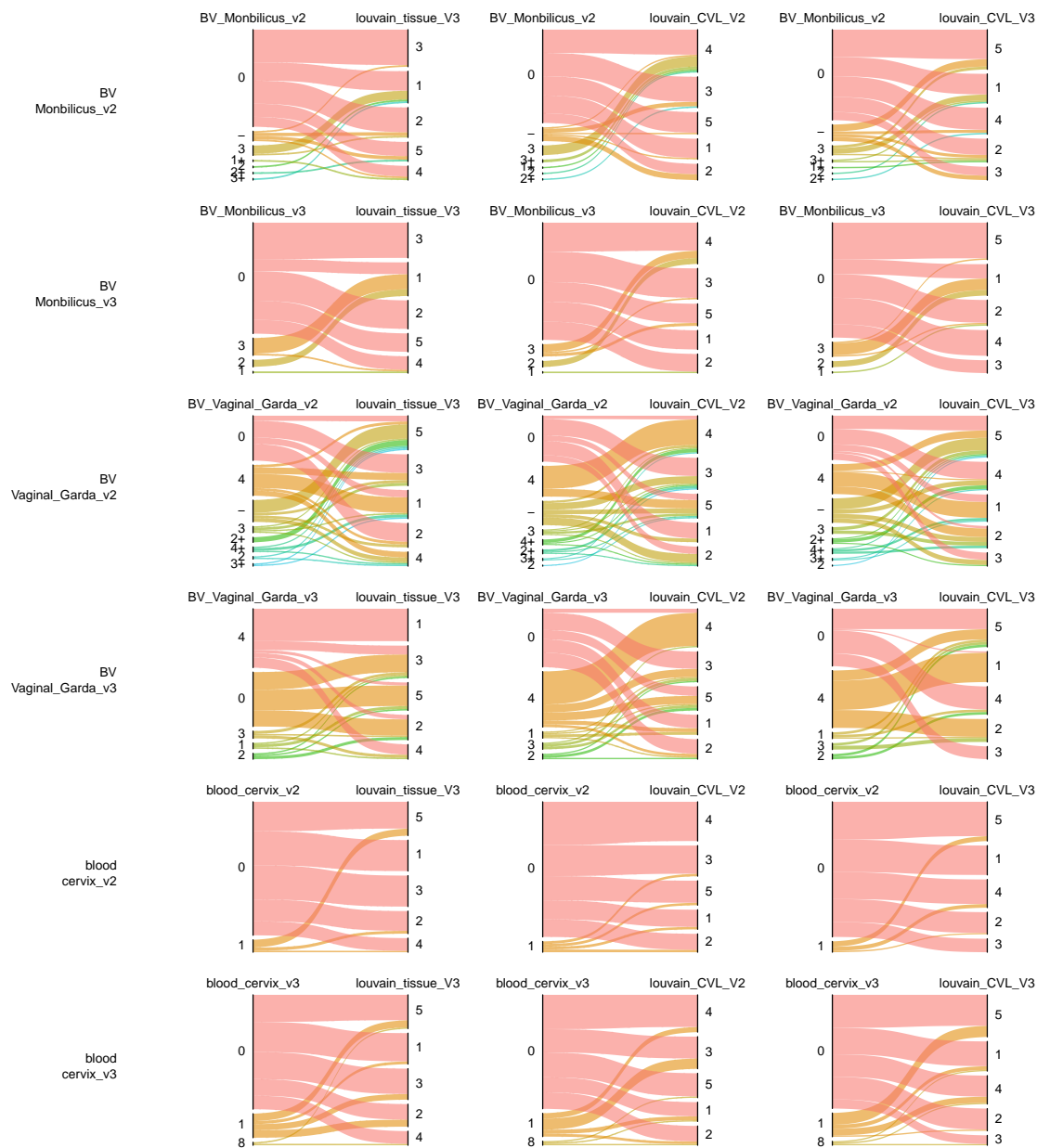
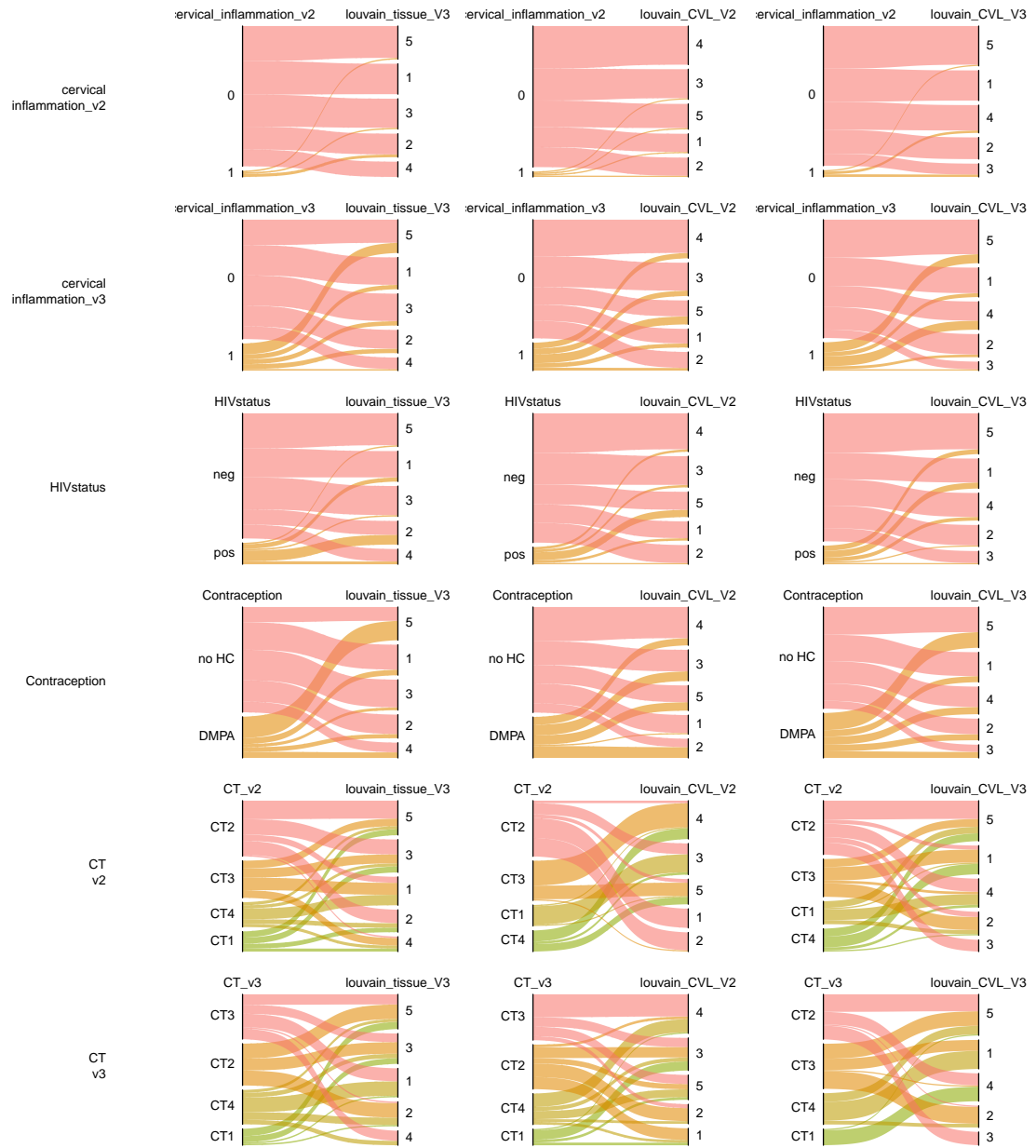


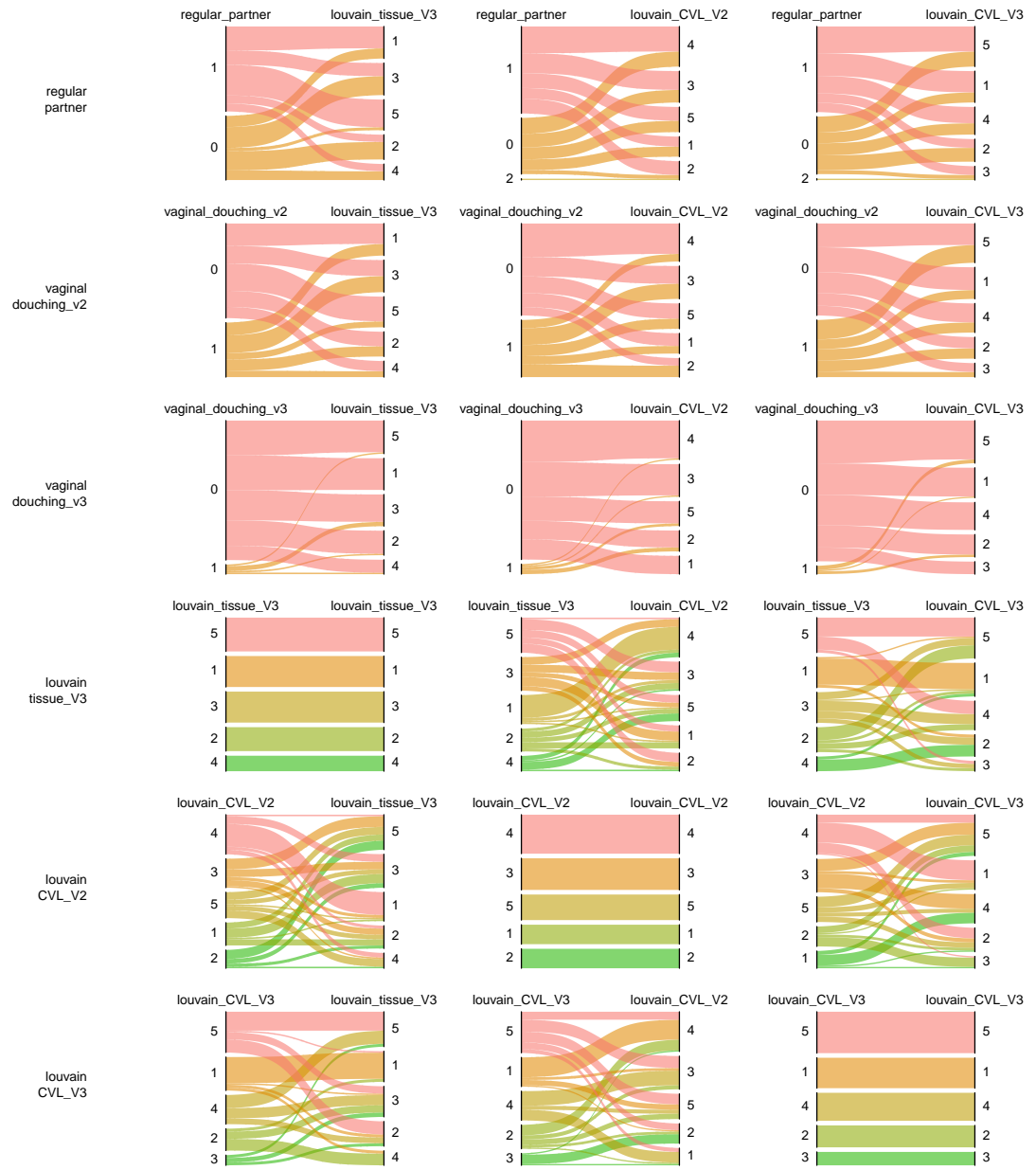
Figure 7. A few examples of significant association between metadata parameters shown as sankey plots.











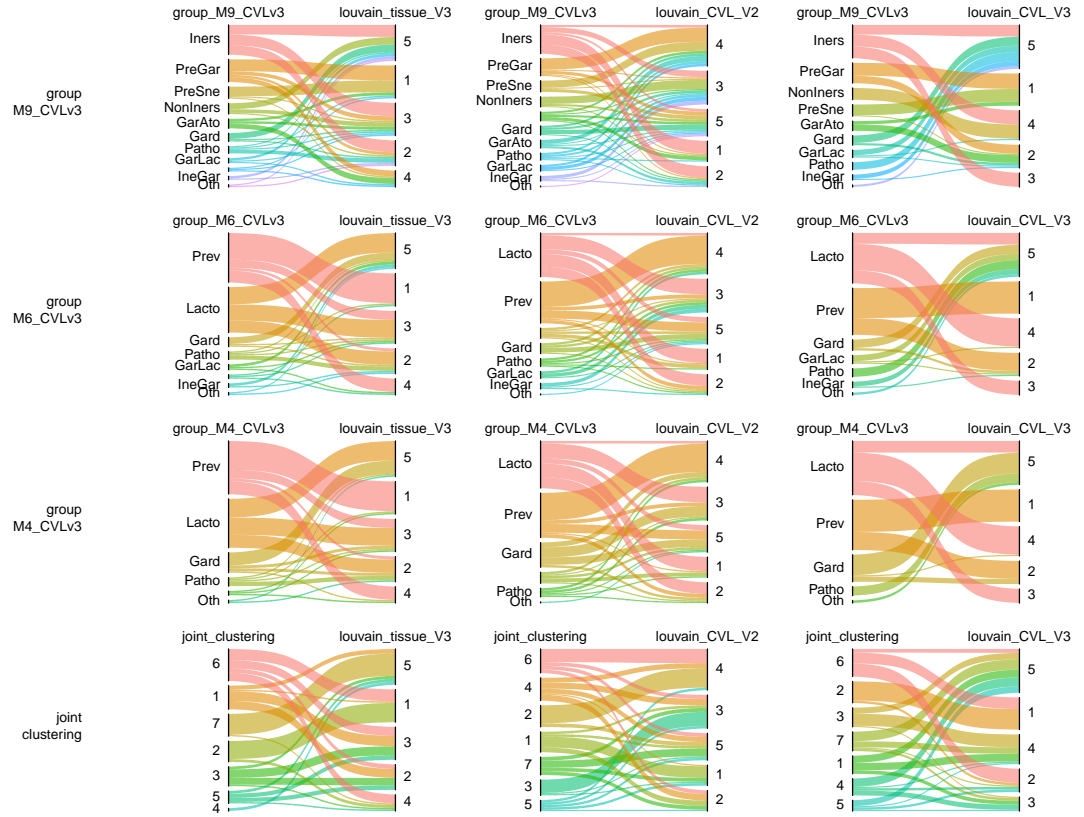


Figure 8. Sankey plots for all tested associations between the patient groups identified in in the microbiome datasets.

Visualise the data

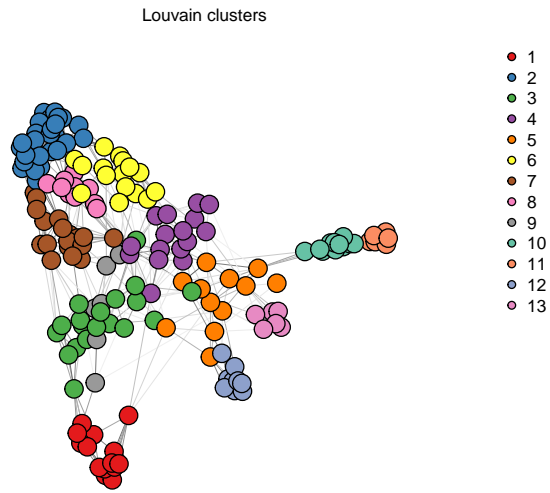


Figure 9a. Bacterial community embedding of 5-SNN graph clustered using walktrap community detection algorithm based on the normalized bacterial abundances across all microbiome datasets (including samples not present in all datasets).

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

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

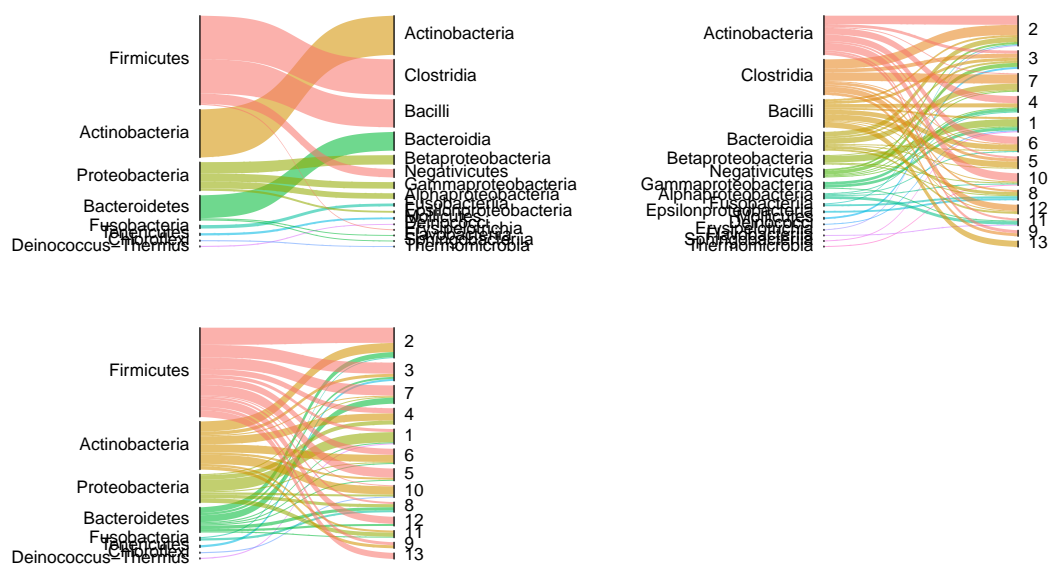
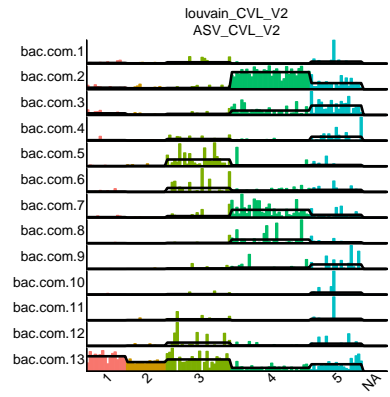
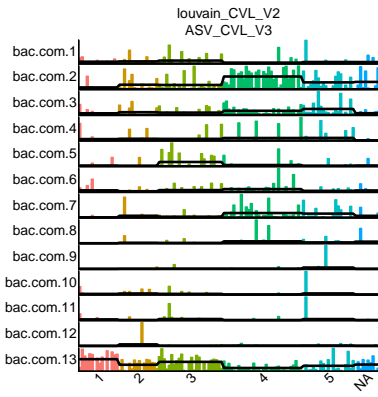
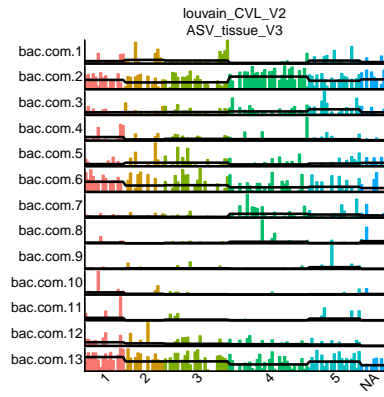
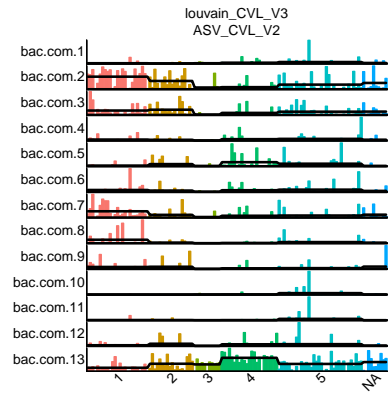
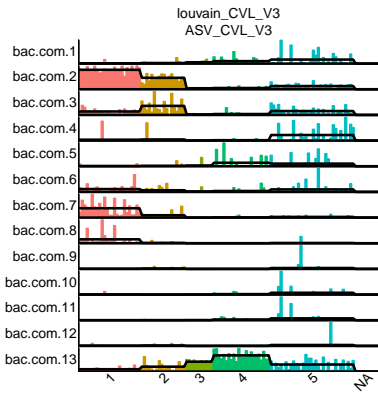
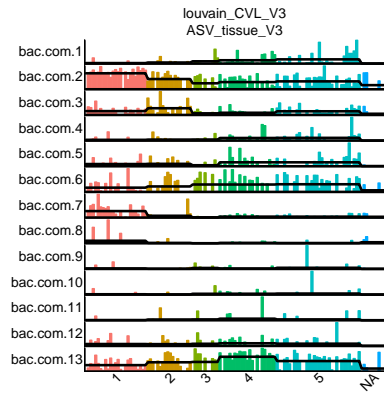
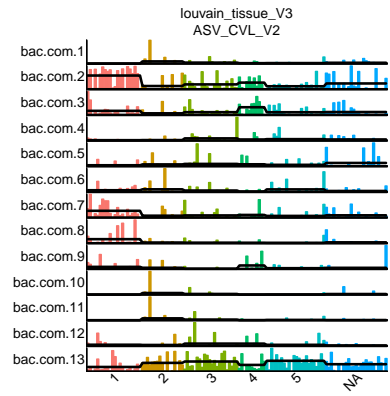
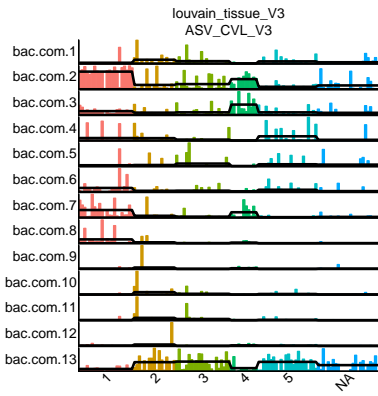
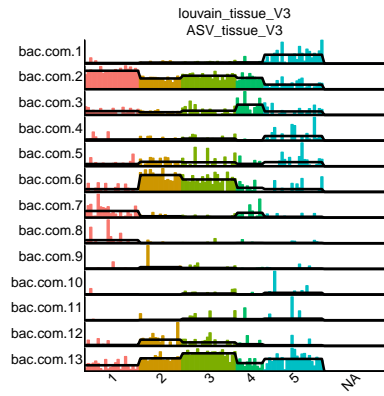


Figure 9b. Taxonomic annotation of bacterial communities.



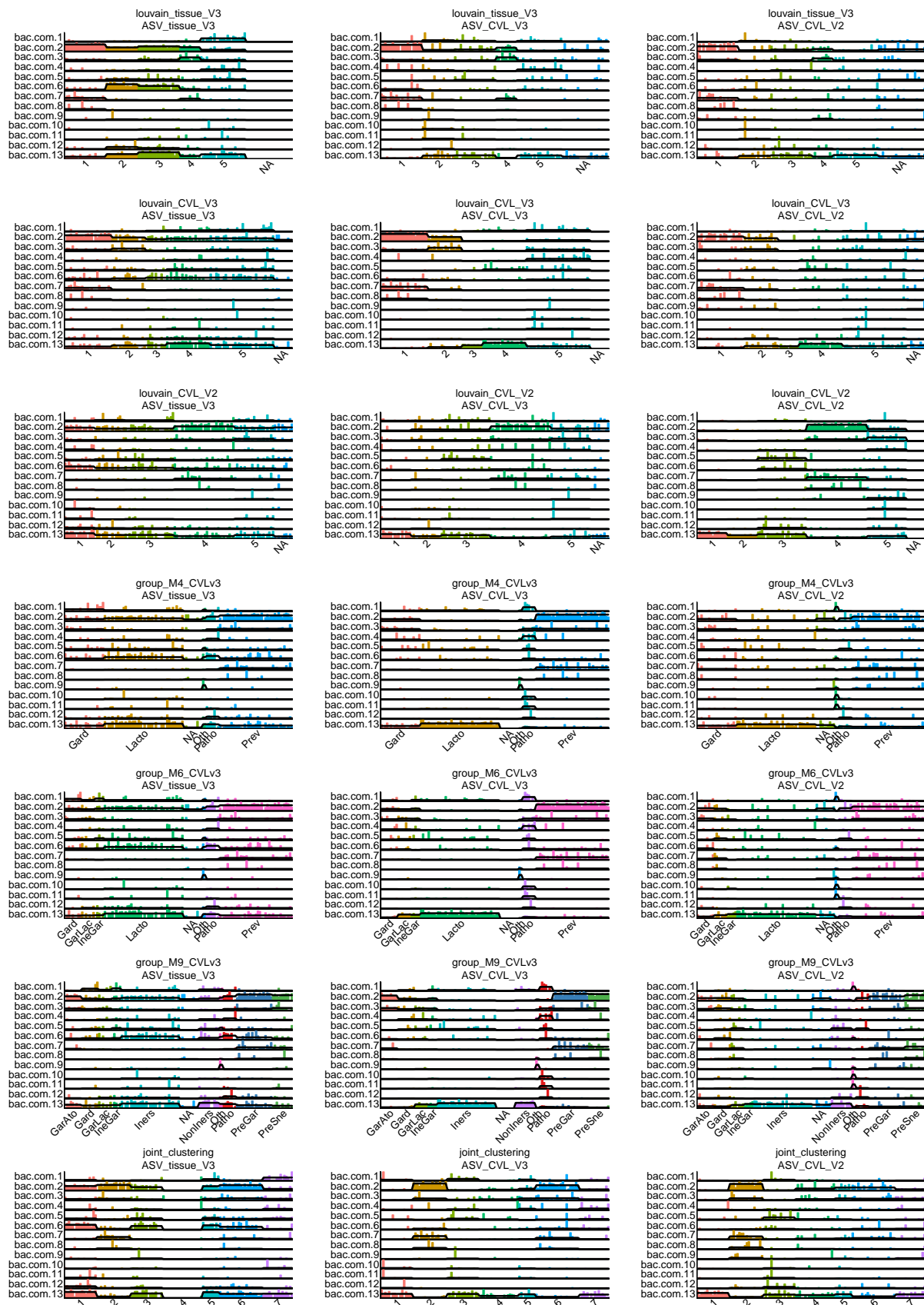
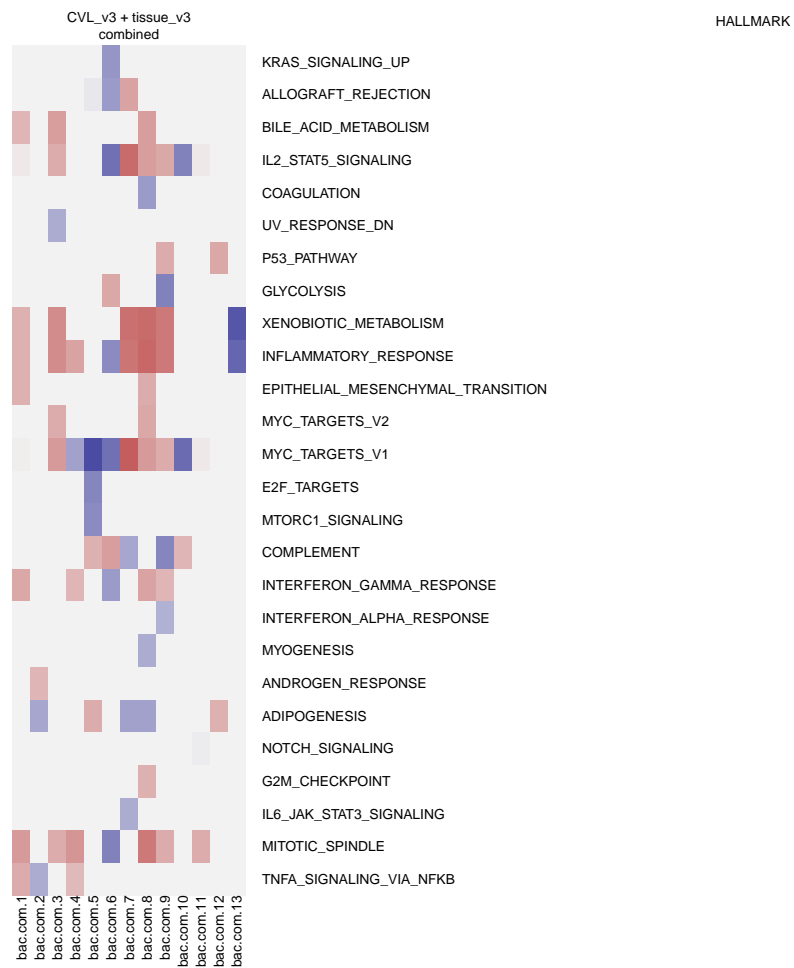


Figure 5. Comparisson of microbiome communities across all datasets and all clustering methods.





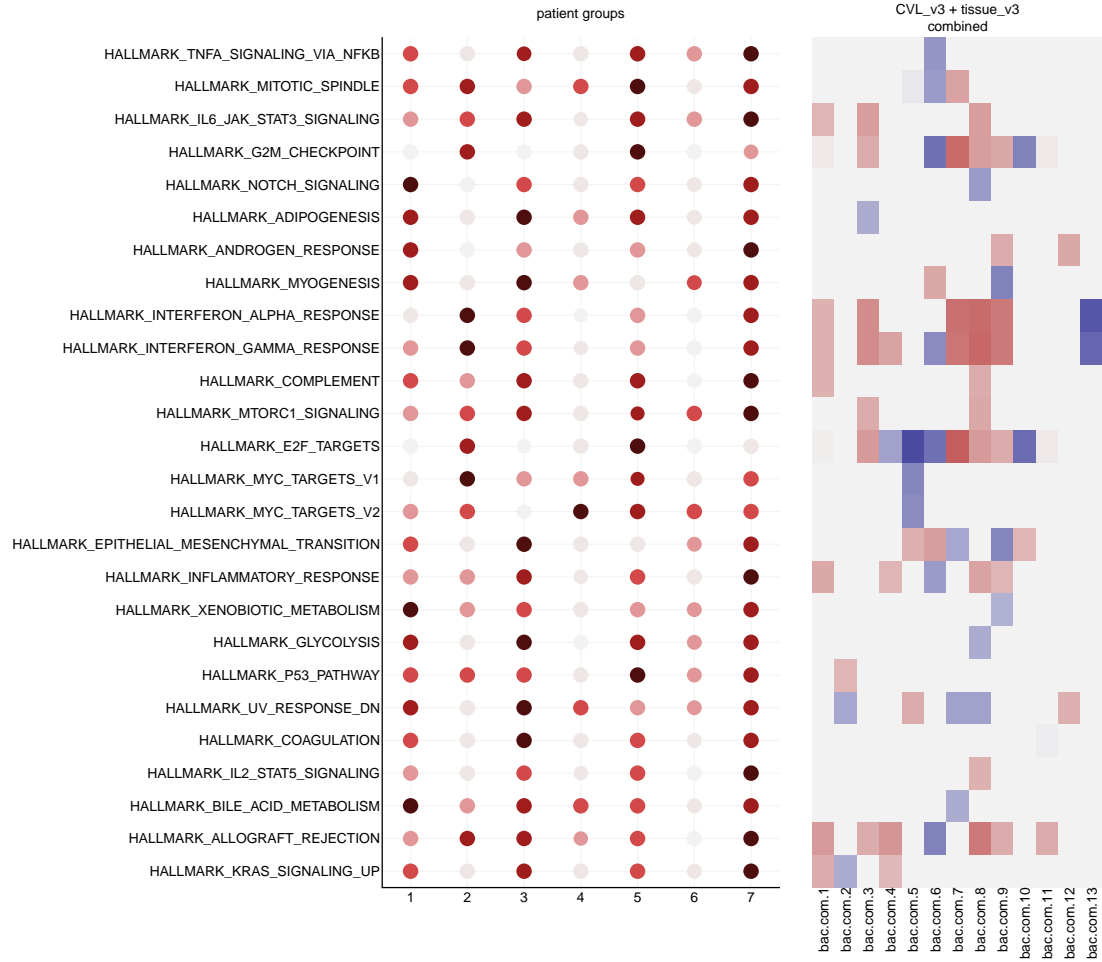


Figure 9. Heatmaps showing the functional association of microbiome datasets with the expression profiles in the RNAseq dataset. Briefly, bacterial abundances from each dataset were correlated with the gene expression of the top 5000 highly variable genes from the RNAseq dataset, generating a correlation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

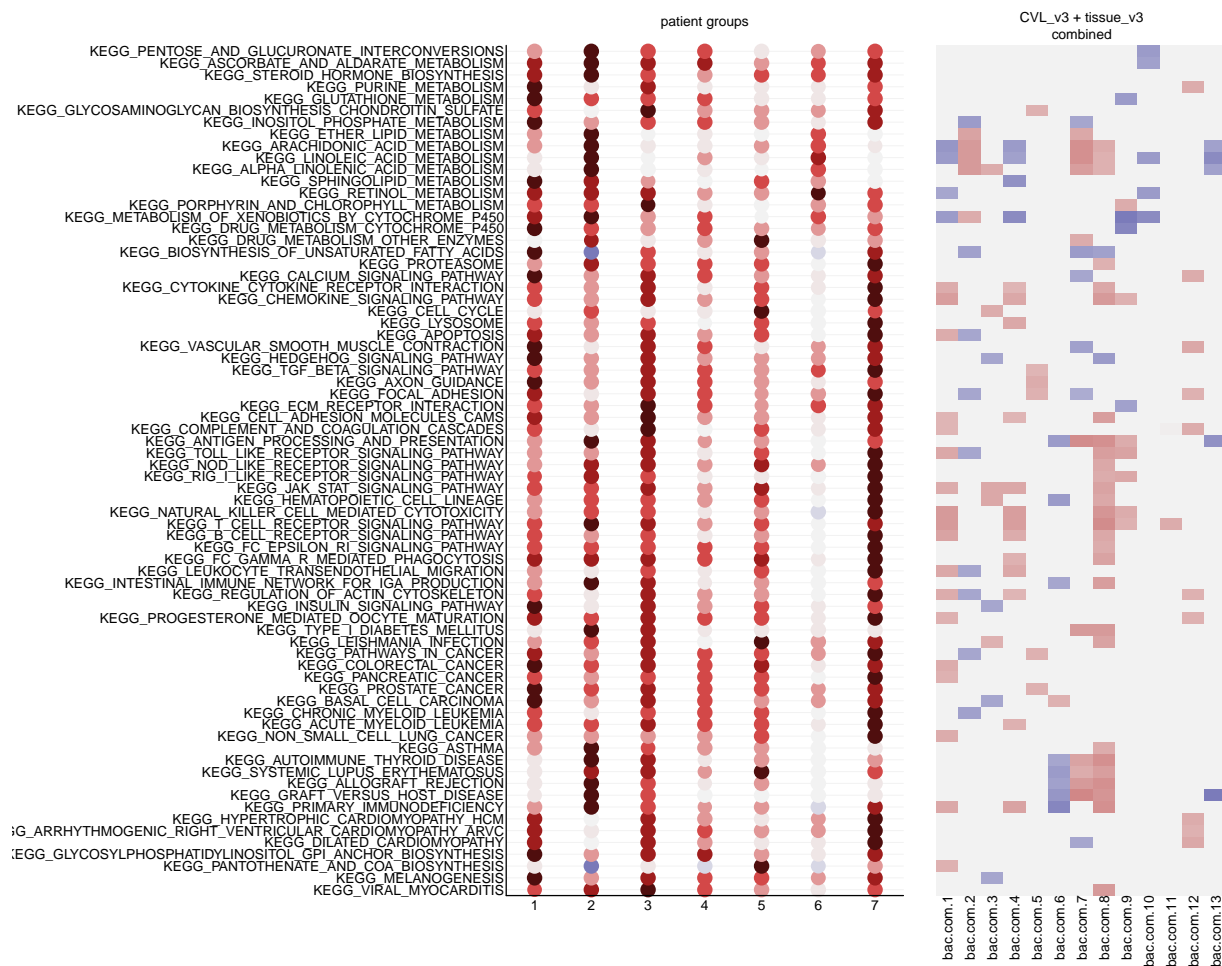


Figure 9. Heatmaps showing the functional association of microbiome datasets with the expression profiles in the RNAseq dataset. Briefly, bacterial abundances from each dataset were correlated with the gene expression of the top 5000 highly variable genes from the RNAseq dataset, generating a correlation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