

Broliden_5325

11 November, 2020

Contents

Loading data and metadata	1
Calculate QC metrics	1
Computing differential expression across microbiome datasets	6
Visualise the data	23
#Load libraries and other scripts	
#Defining some variables for the analysis	

Loading data and metadata

Calculate QC metrics

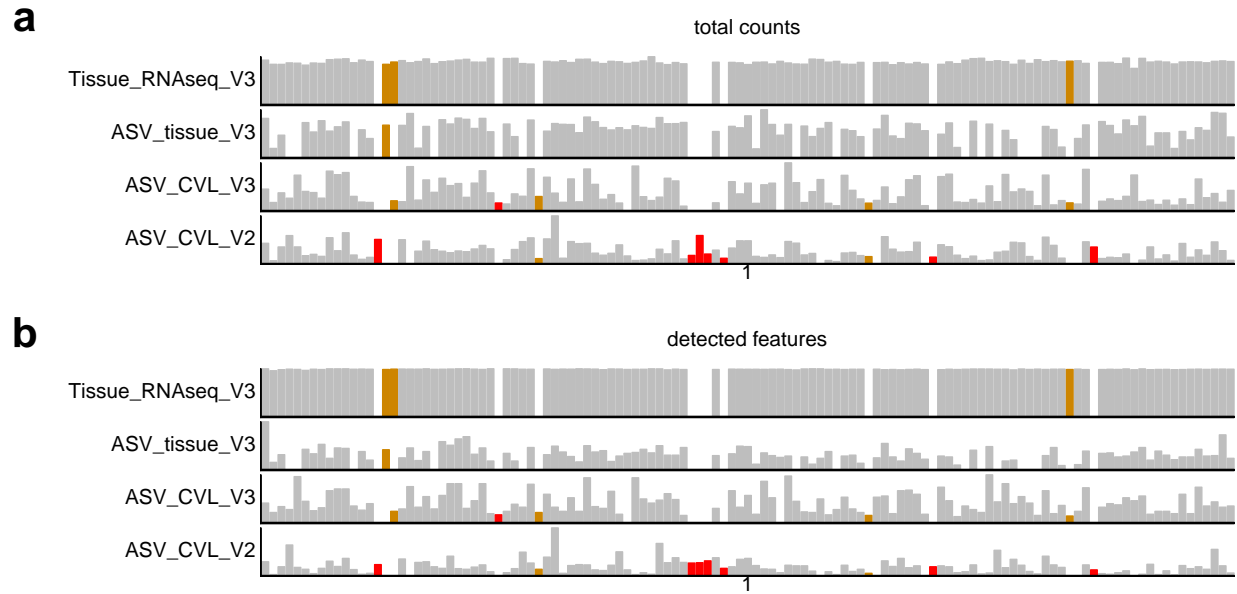


Figure 1. Comparative barplot for the **a)** total counts and **b)** number of non-zero detected features (genes / bacteria) for each of the sequencing datasets. Samples are ordered alfabetically according to the patient ID. Samples that are present in exactly two datasets are shown in orange (P016,P017,P035,P076,P101).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

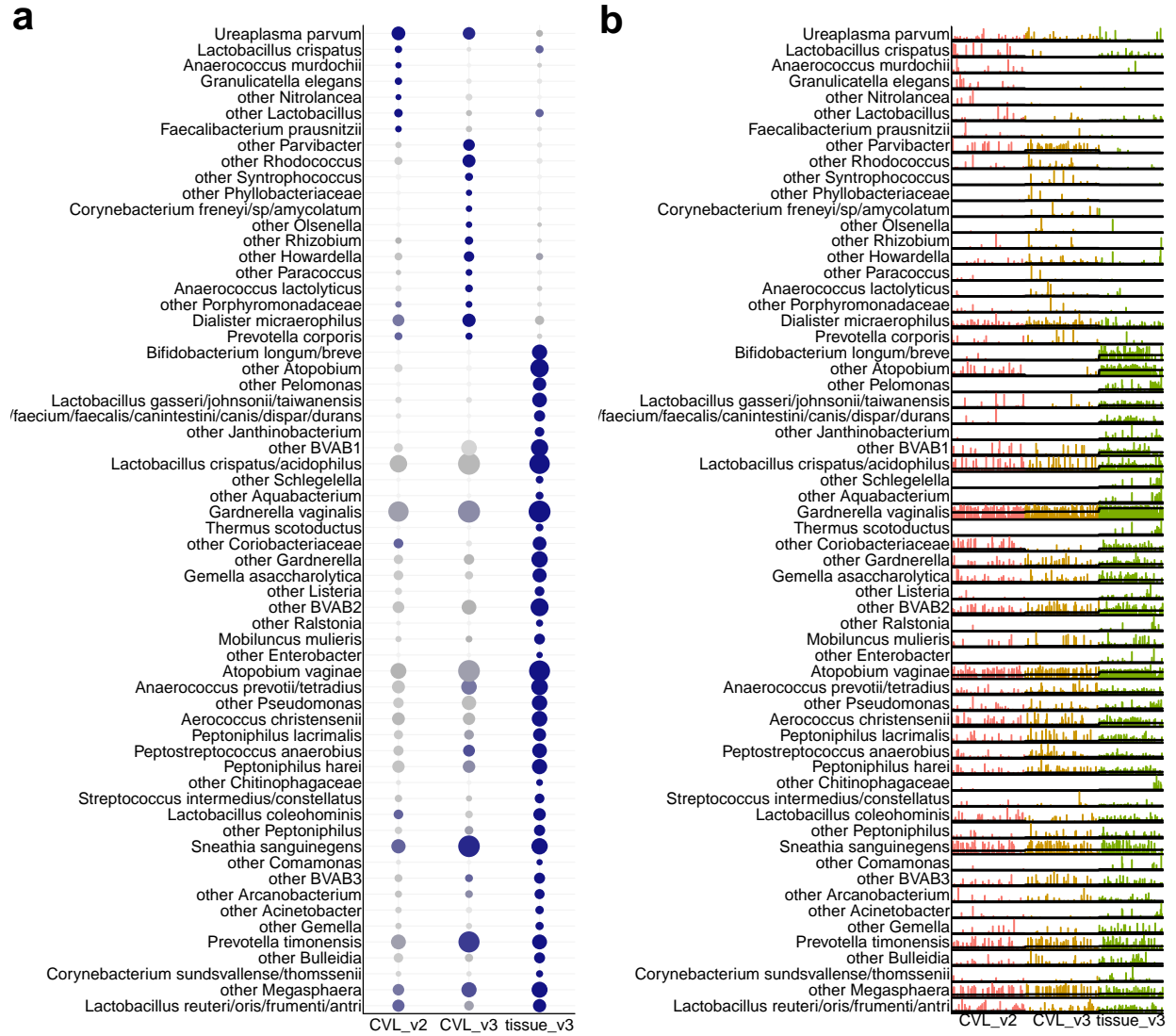


Figure 2. Differential bacterial abundance across microbiome datasets. The results are shown both as a) Dot plots and b) barplots. Bacteria with log2FC above 0.25 and p-value below 0.01 were considered significant and were sorted by the highest expression.

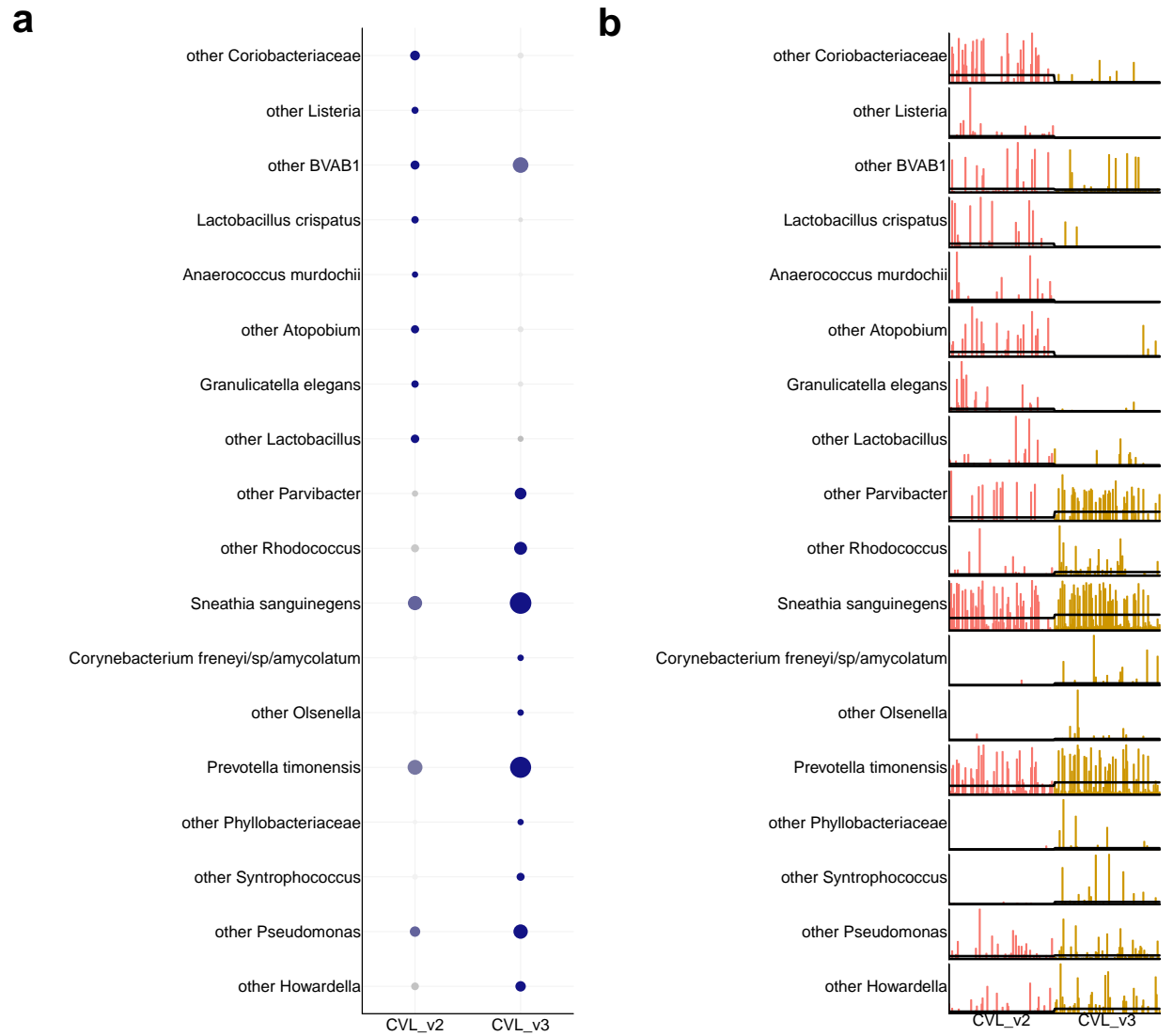


Figure 2. Differential bacterial abundance across CVL2 and CVL3 microbiome datasets. The results are shown both as **a)** Dot plots and **b)** barplots. Bacteria with log2FC above 0.25 and p-value below 0.01 were considered significant and were sorted by the highest expression.

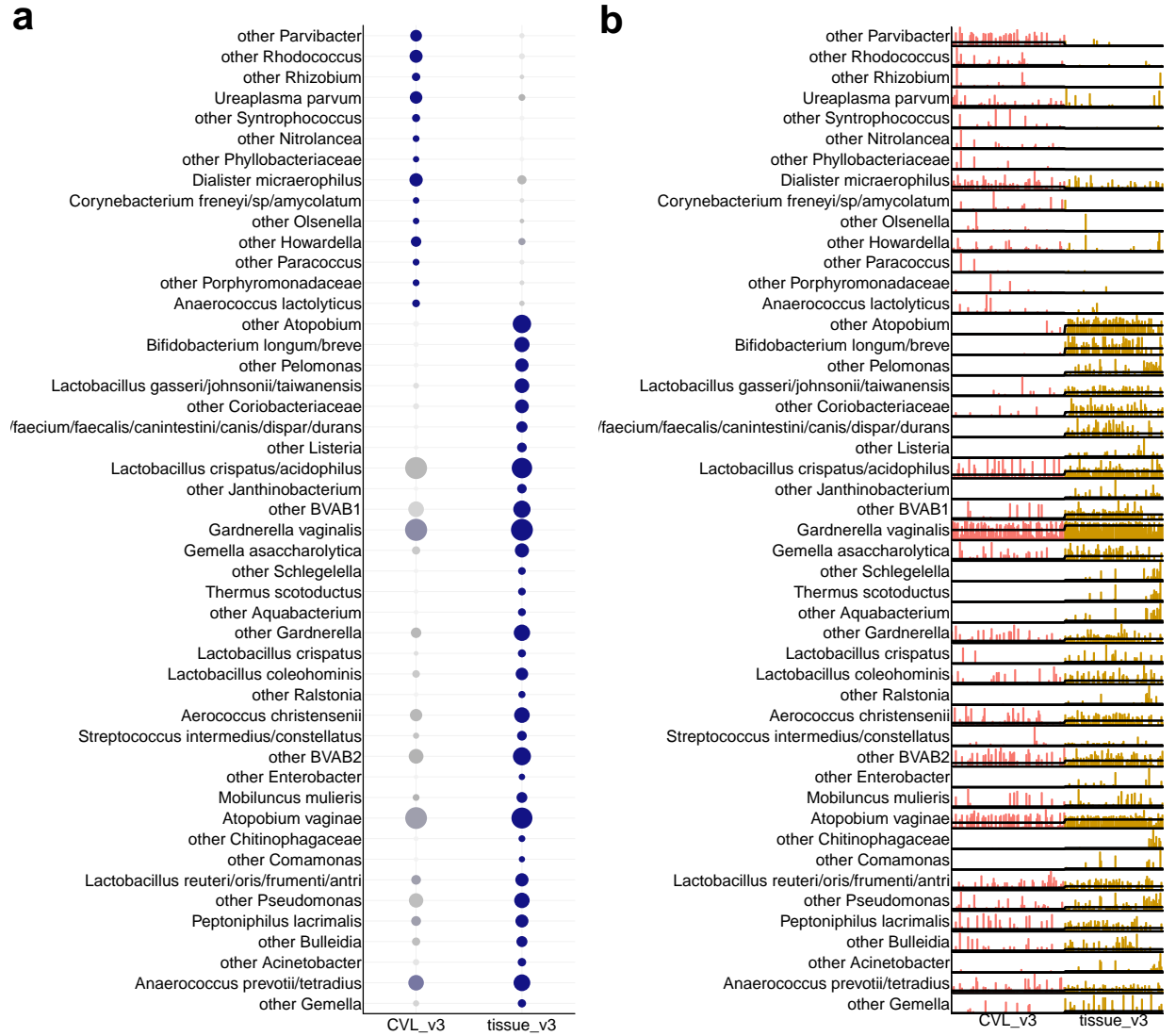


Figure 2b. Differential bacterial abundance across tissue and CVL microbiome week3 datasets. The results are shown both as **a)** Dot plots and **b)** barplots. Bacteria with log2FC above 0.25 and p-value below 0.01 were considered significant and were sorted by the highest expression.

Computing differential expression across microbiome datasets

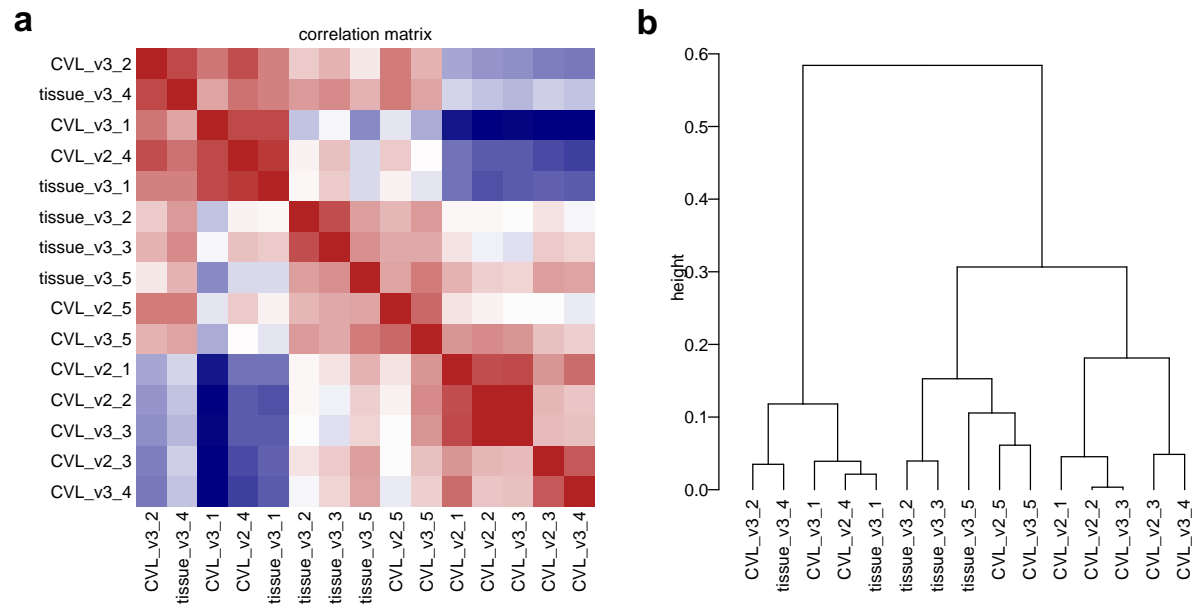


Figure 3. Comparisson among patient groups across datasets. **(a)** Correlation matrix across sample groups. **(b)** Hierarchical clustering of sample groups.

a

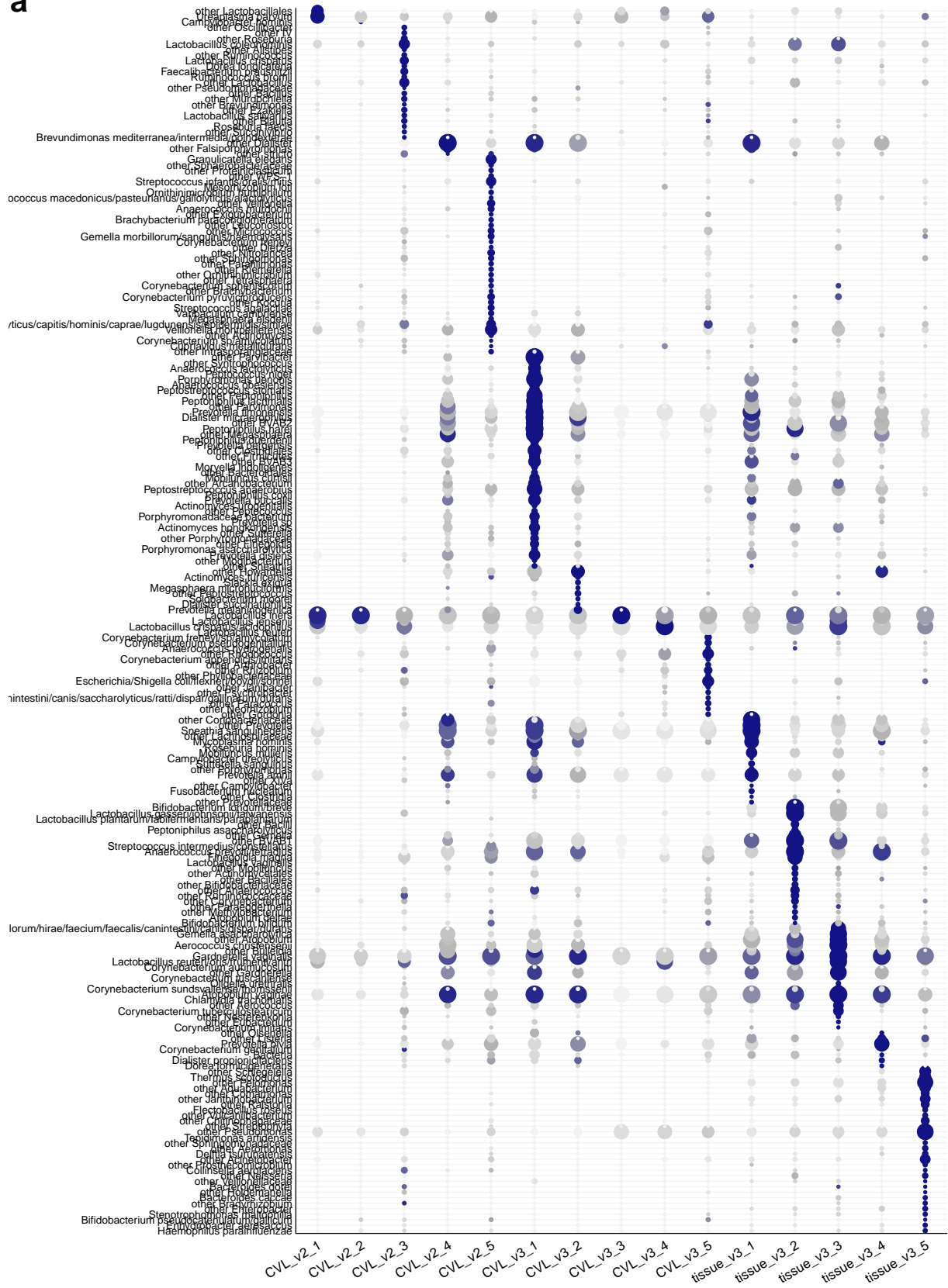


Figure 4. Differential bacterial abundance across all groups and all microbiome datasets. The results are shown both as **a)** Dot plots and **b)** barplots. Bacteria with \log_2FC above 0.25 and p-value below 0.01 were considered significant and were sorted by the highest expression.

Figure 5. Comparison of microbiome datasets, showing only the significant bacteria. Samples are ordered by the CVL3 groupings. The colors represent their respective bacterial groupings for each dataset.

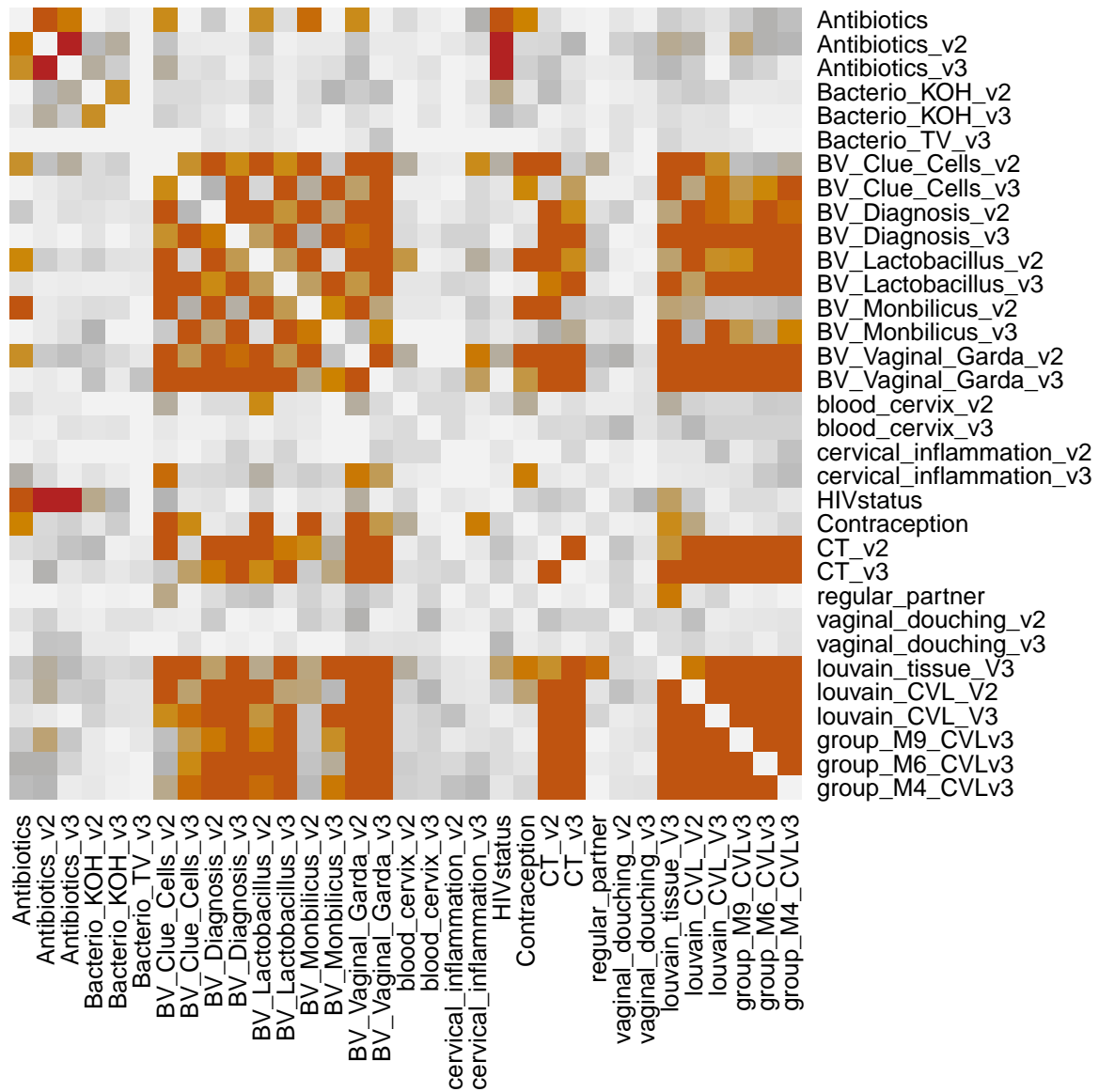


Figure 6. Association analysis across several patient categorical metadata parameters, including patient groupings annotations from microbiome.

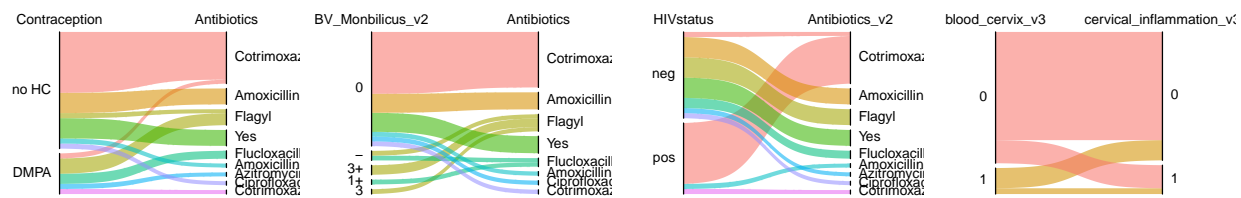
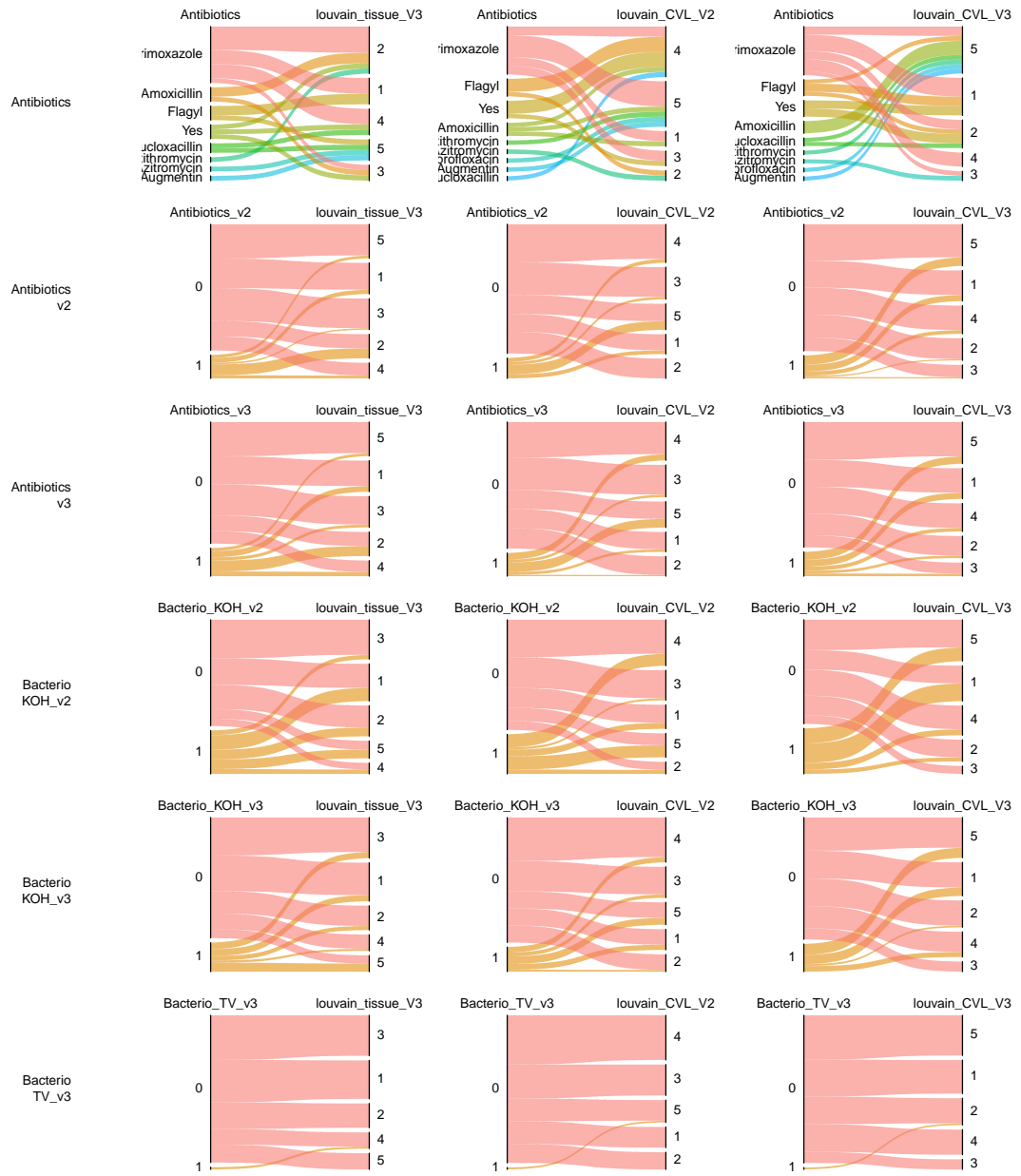
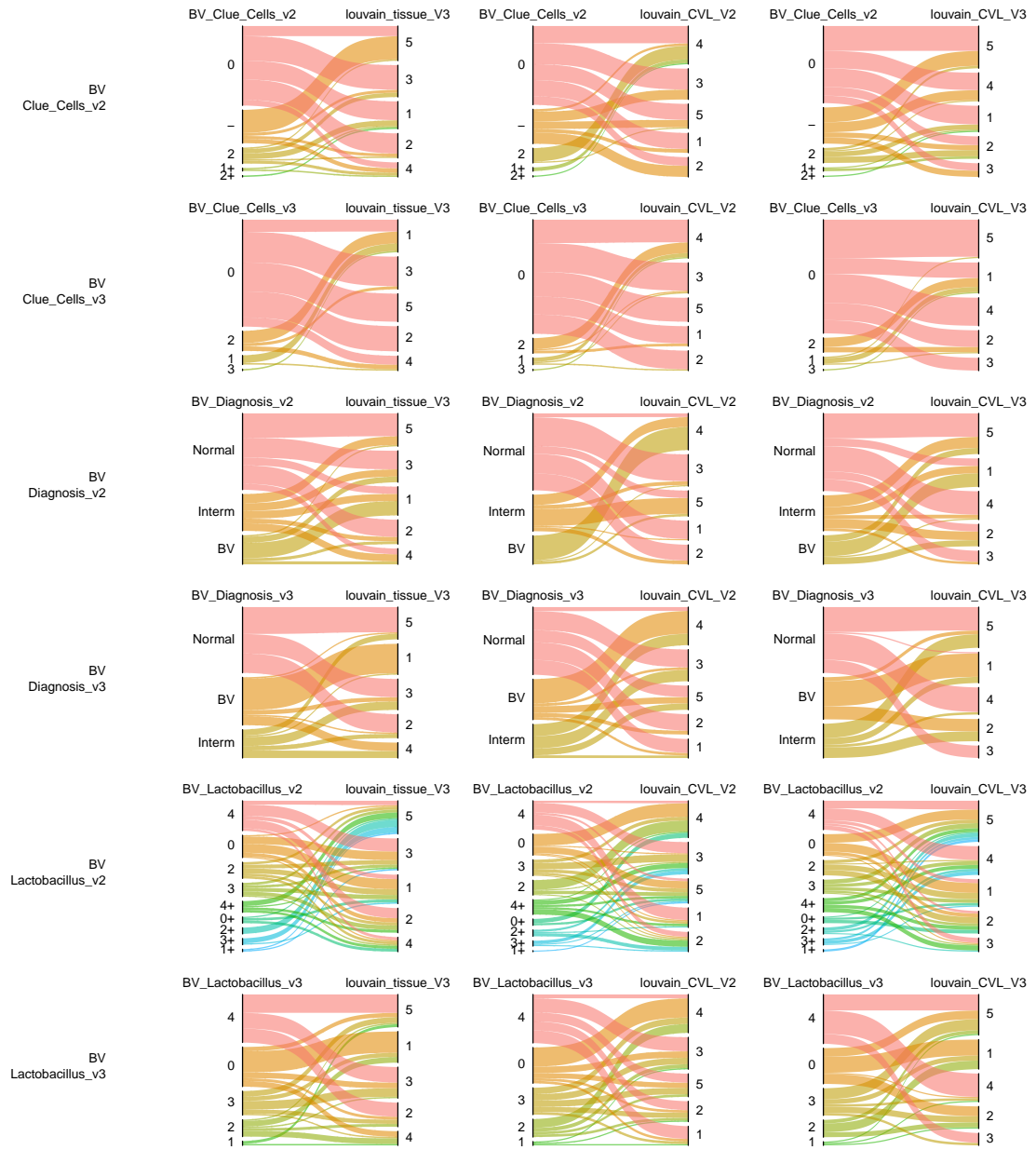
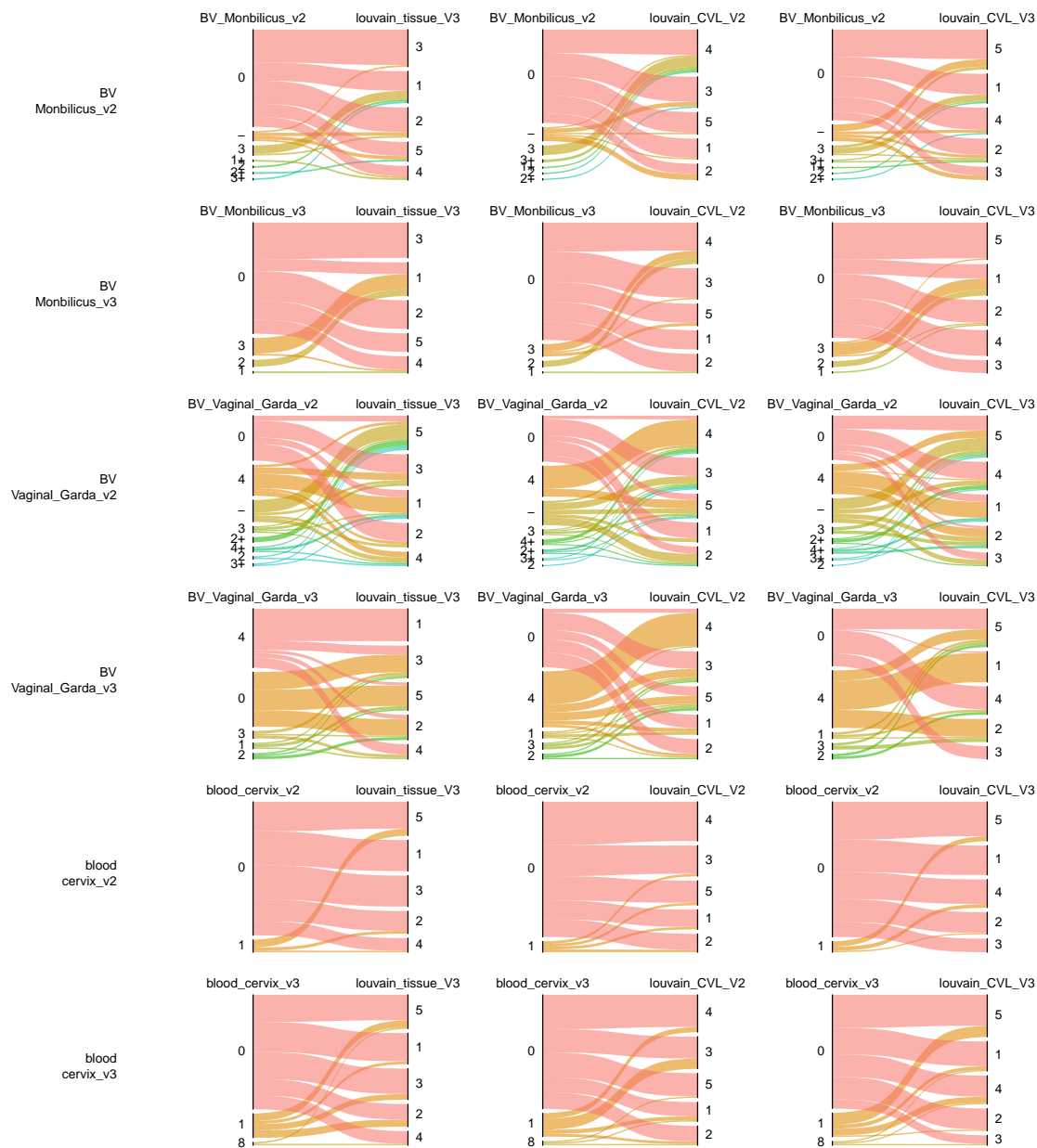
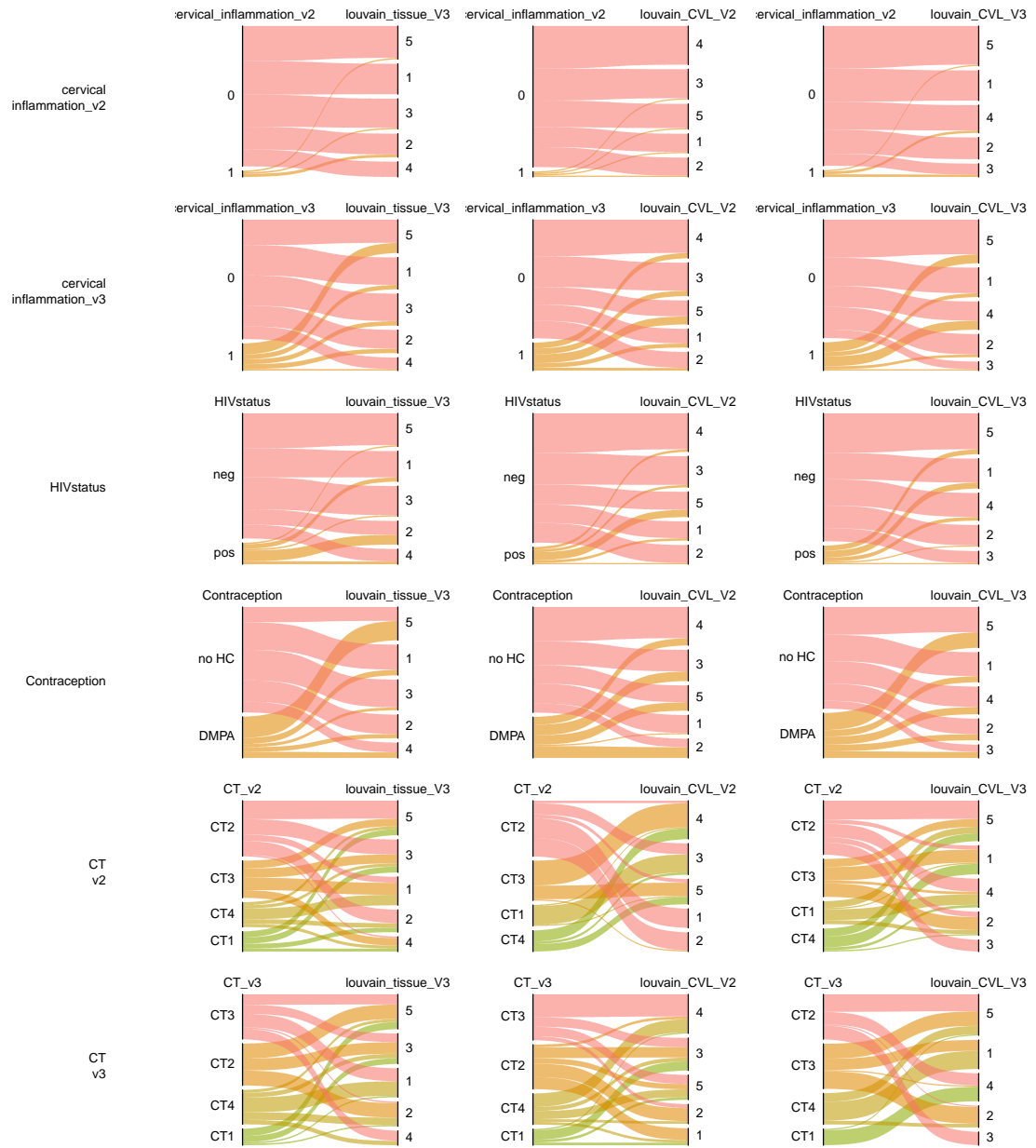


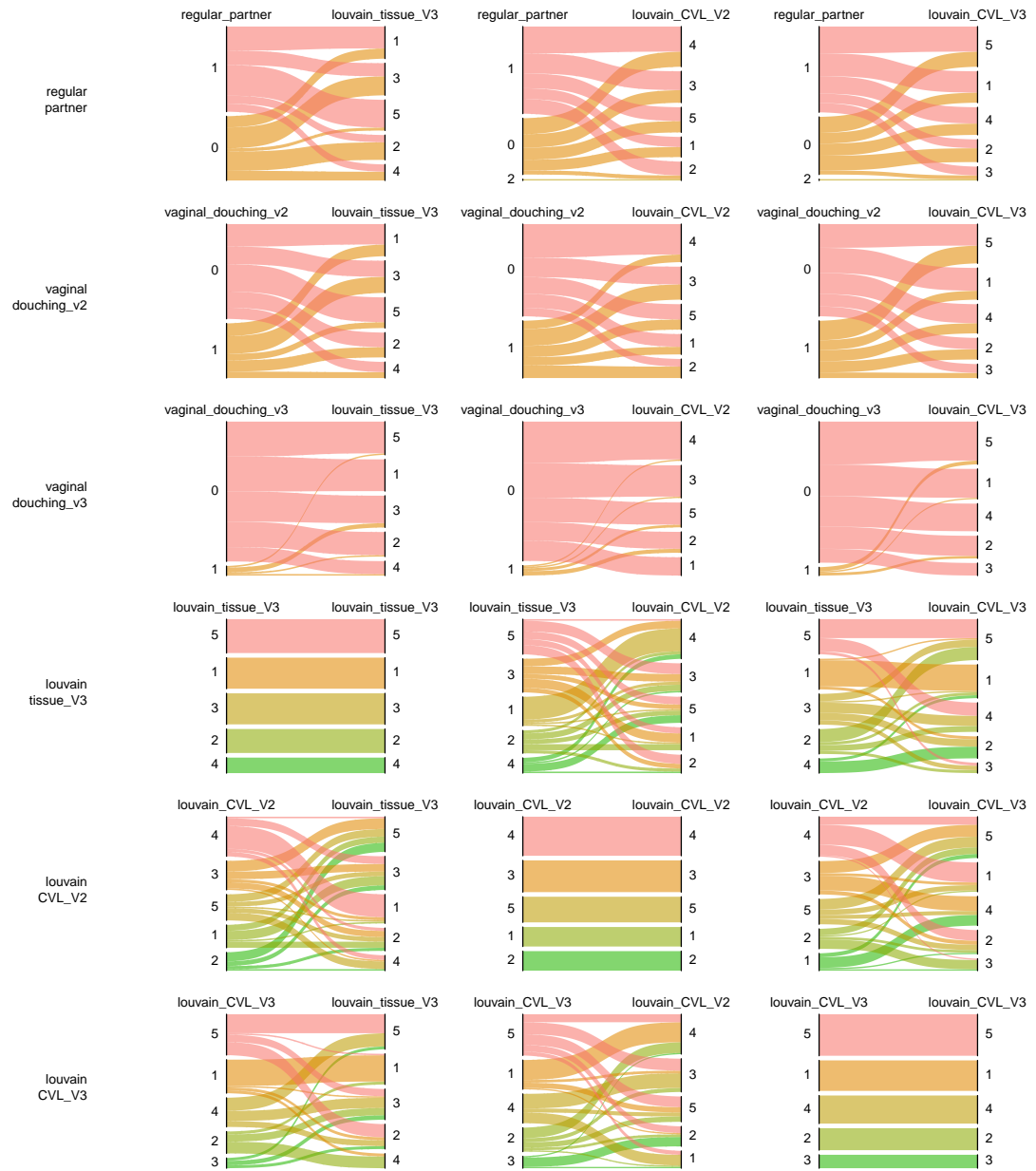
Figure 7. A few examples of significant association between metadata parameters shown as sankey plots.











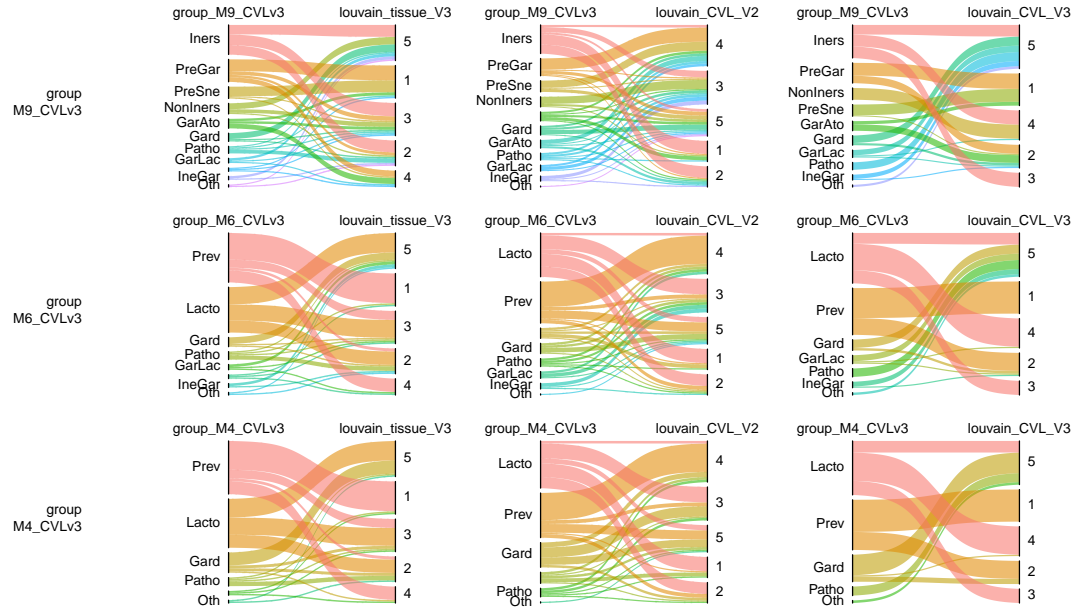


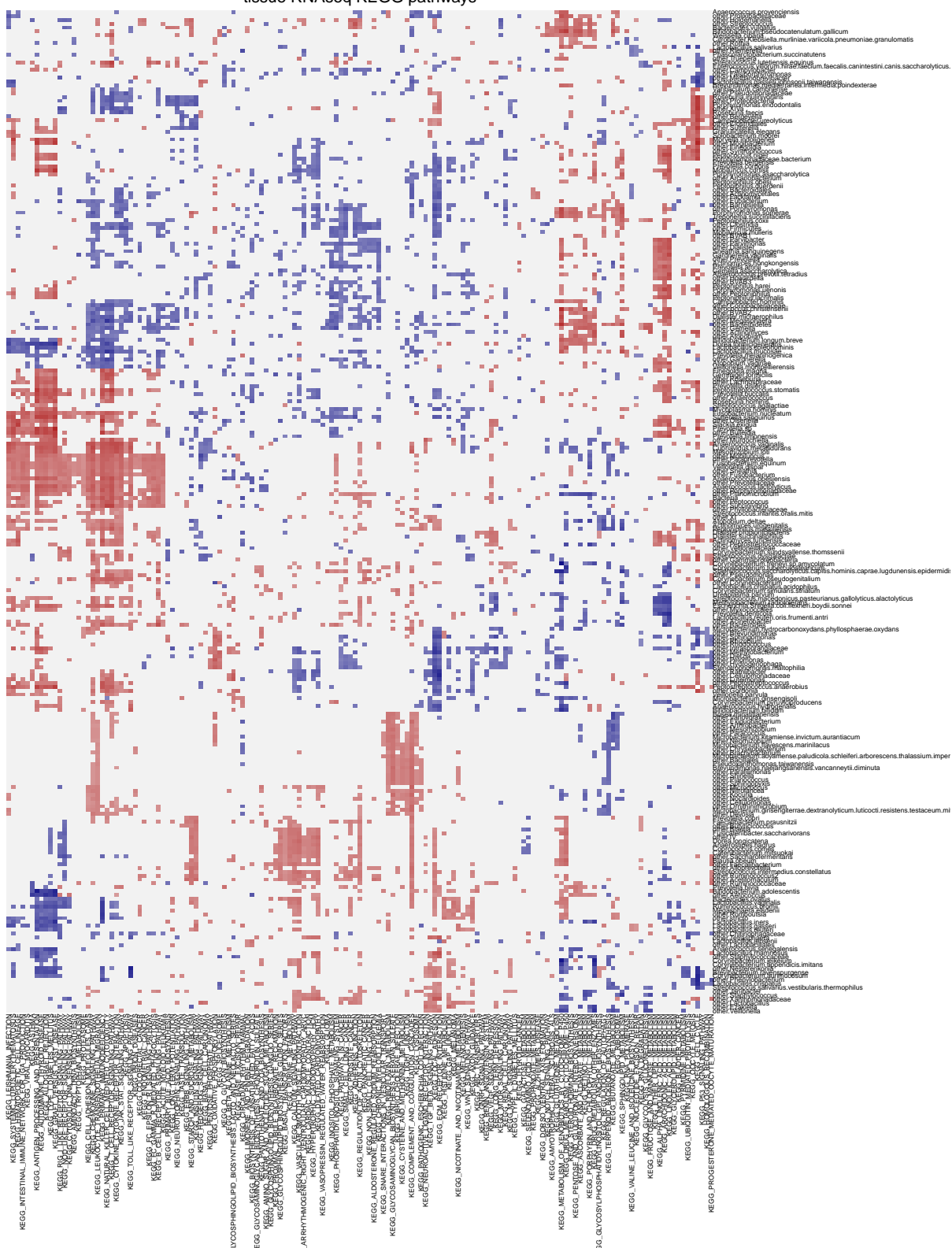
Figure 8. Sankey plots for all tested associations between the patient groups identified in in the microbiome datasets.

ASV_tissue_V3_normalized_batch_corrected

Heatmap showing the relative abundance of 1000 bacterial taxa across 1000 KEGG pathways. The color scale ranges from blue (low abundance) to red (high abundance). The taxa are listed on the right, and the pathways are listed on the bottom. The heatmap shows a complex pattern of abundance across the pathways, with some pathways showing high abundance for many taxa and others showing low abundance for most taxa.

ASV_CVL_V3_normalized_batch_corrected

tissue RNAseq KEGG pathways



C

tissue RNAseq KEGG pathways

ASV_CVL_V2_normalized_NOT_batch_corrected

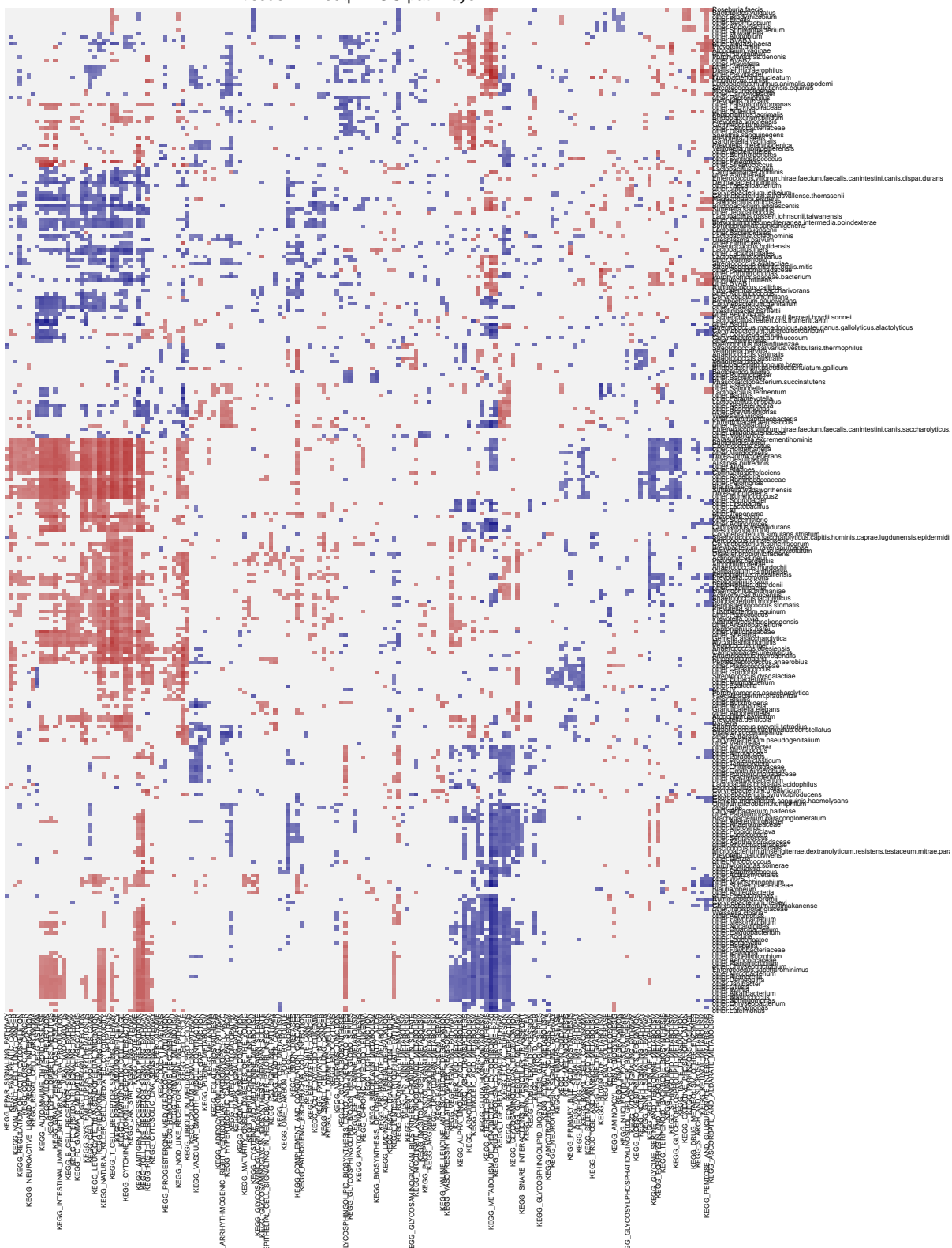
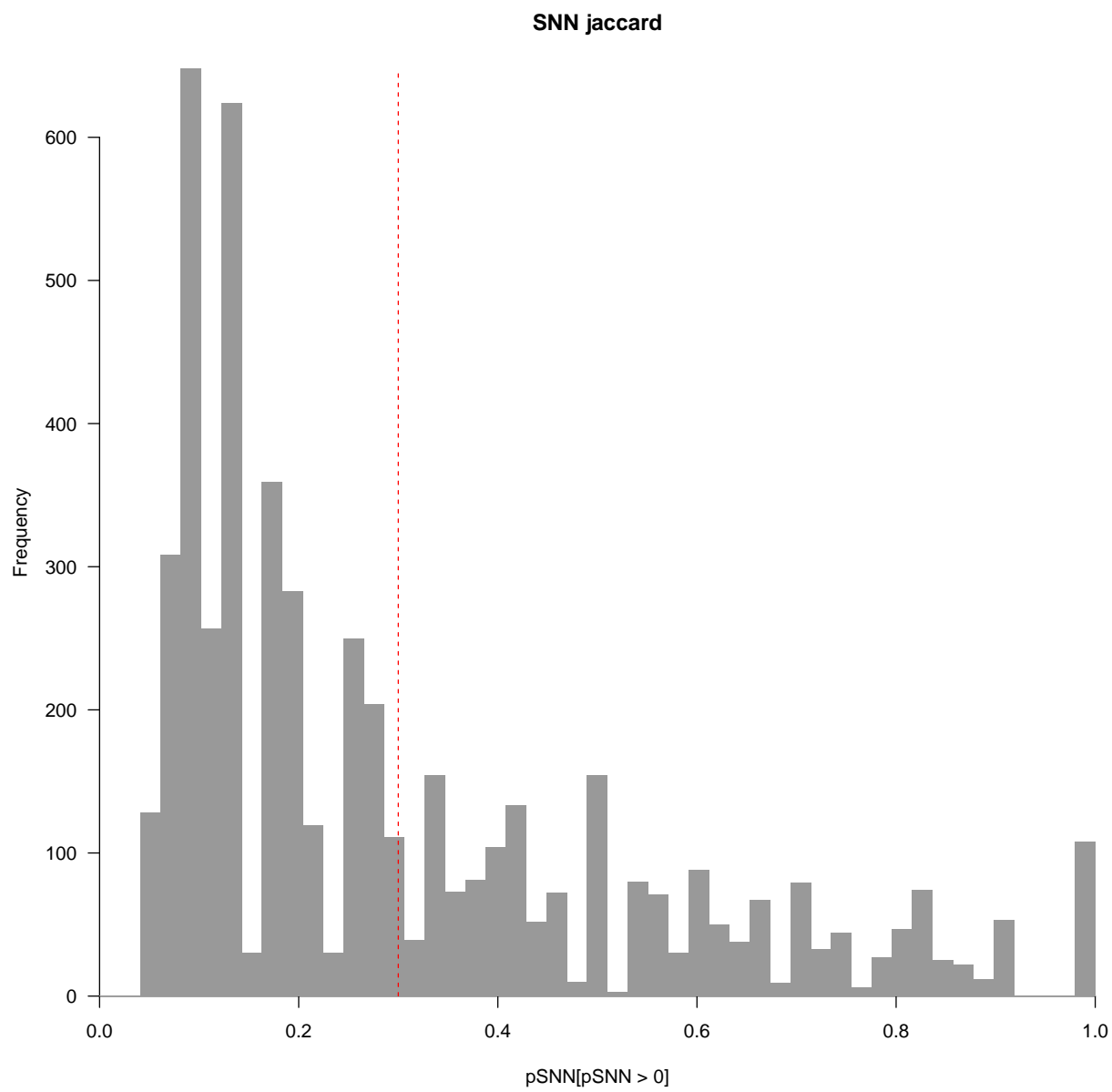
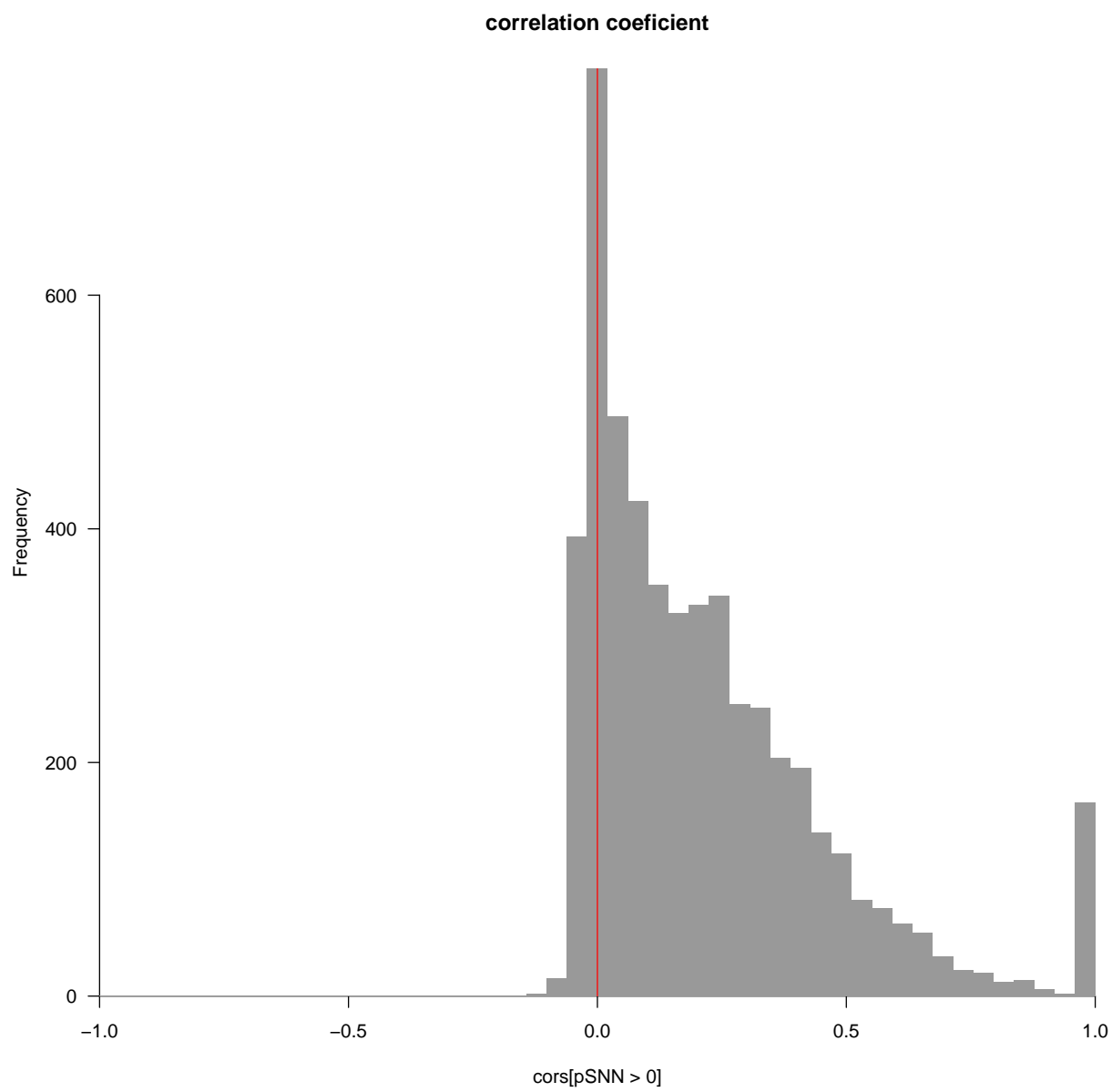


Figure 9. Heatmaps showing the functional association of microbiome datasets with the expression profiles in the RNAseq dataset. Briefly, bacterial abundances from each dataset were correlated with the gene expression

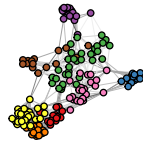
of the top 5000 highly variable genes from the RNAseq dataset, generating a correlation matrix between bacteria and genes. Then for each bacteria, we rank genes based on their correlation to that bacteria and perform gene set enrichment analysis (GSEA) using the KEGG gene annotation database. This, in 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.





Visualise the data

Louvain clusters



- 1
- 2
- 3
- 4
- 5
- 6
- 7

community1

Actinomyces hongkongensis
Aerococcus christensenii
Anaerococcus prevotii/tetradus
Atopobium vaginae
Gardnerella vaginalis
Gemella asaccharolytica
other Arcanobacterium
other Atopobium
other Bullidia
other BVAB1
other Coriobacteriaceae
other Firmicutes
other Gardnerella
Peptoniphilus harei

community2

other Acinetobacter
other Aquabacterium
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

community3

Atopobium deltae
Bifidobacterium longum/breve
Corynebacterium aurimucosum
Corynebacterium pseudogenitalium
Corynebacterium pyruviciproducens
Corynebacterium sp/amycolatum
Corynebacterium sundsvallense/thomsseni
Corynebacterium tuberculostearicum
Enterococcus villorum/hirae/faecium/faecalis/s
Escherichia/Shigella coliflexneri/boydii/sonne
Gemella morbillorum/sanguinis/haemolysans
Lactobacillus coleohominis
Lactobacillus crispatus
Lactobacillus crispatus/acidophilus
Lactobacillus gasseri/johnsonii/taiwanensis
Lactobacillus iners
Lactobacillus jensenii
Lactobacillus reuteri
Lactobacillus reuteri/oris/frumenti/antri
other Actinomycetales
other Aerococcus
other Bacilli
other Bifidobacteriaceae
other Corynebacterium
other Eubacterium
other Gemella
other Gordonia
other Lactobacillales
other Lactobacillus
other Mobiluncus
Staphylococcus saccharolyticus/capitis/homin
Ureaplasma parvum
Corynebacterium freneyi/sp/amycolatum

community4

Cupriavidus metallidurans
Lactobacillus salivarius
other Actinomyces
other Arthrobacter
other Brachybacterium
other Chryseobacterium
other Dietzia
other Exiguobacterium
other Intrasporangiaceae
other Kocuria
other Micrococcus
other Nitrolancea
other Ornithinimicrobium
other Paracoccus
other Phyllobacteriaceae
other Psychrobacter
other Rhizobium
other Rhodococcus
other Sphingomonas

community5

Dialister microaerophilus
Mobiluncus mulieris
other BVAB2
other BVAB3
other Dialister
other Falsiporphyromonas
other Lachnospiraceae
other Megasphaera
other Parvibacter
other Peptoniphilus
other Prevotella
Peptoniphilus lacrimalis
Porphyromonas unonisi
Prevotella amnii
Prevotella timonensis
Sneathia sanguinegens

community6

Anaerococcus lactolyticus
Anaerococcus obesiensis
Campylobacter ureolyticus
Fusobacterium nucleatum
Mobiluncus curtisi
Moryella indoligenes
other Anaerococcus
other Bacteroidales
other Clostridiales
other Finegoldia
other Porphyromonadaceae
other Porphyromonas
other Sneathia
other Sutterella
other Syntrophococcus
Peptococcus niger
Peptoniphilus coxi
Peptoniphilus duerdenii
Peptoniphilus massiliensis
Peptostreptococcus anaerobius
Peptostreptococcus stomatis
Porphyromonadaceae bacterium
Porphyromonas asaccharolytica
Prevotella bergensis
Prevotella buccalis
Prevotella corporis
Prevotella disiens
Prevotella sp
Sutterella sanguinus

community7

Collinsella aerofaciens
Corynebacterium jeikeium
Faecalibacterium prausnitzii
other Alloprevotella
other Blautia
other Ezakiella
other Faecalibacterium
other Methylobacterium
other Murdochella
other Roseburia
other Ruminococcaceae
other stricto
Prevotella copri
Prevotella denticola
Streptococcus intermedius/constellatus

community8

Actinomyces turicensis
Anaerococcus hydrogenalis
Anaerococcus murchii
Bacteria
Bifidobacterium bifidum
Dialister propionificiens
Finegoldia magna
Granulicatella elegans
Mycoplasma hominis
other Howardella
other Listeria
other Olsenella
other Peptostreptococcus
other Veillonella
Prevotella bivia
Prevotella melaninogenica
Streptococcus agalactiae
Streptococcus infantis/oralis/mitis
Streptococcus macedonicus/pasteurianus/gal
Varibaculum cambriense
Veillonella dispar
Veillonella montipallierensis

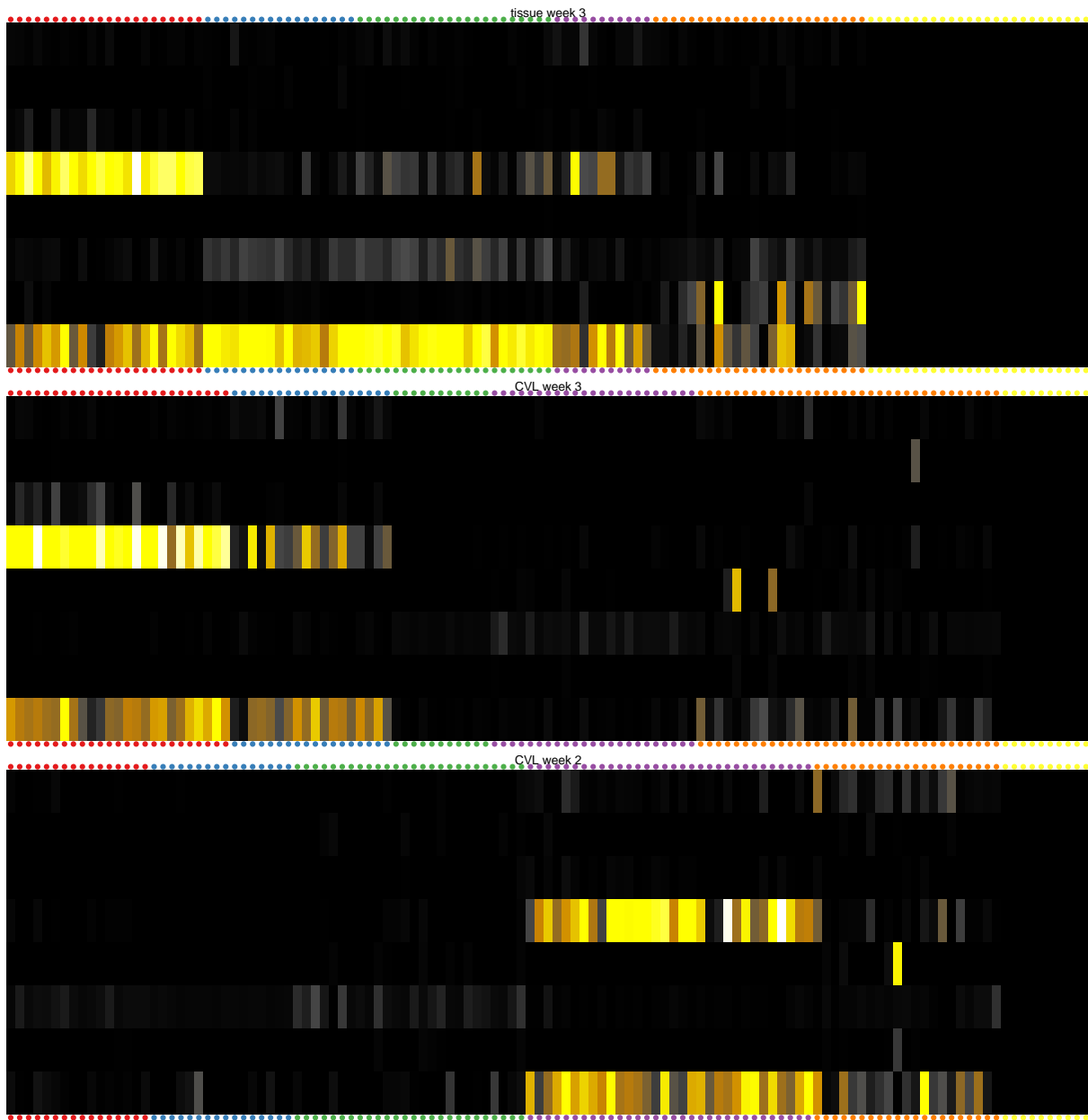
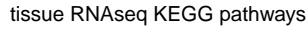


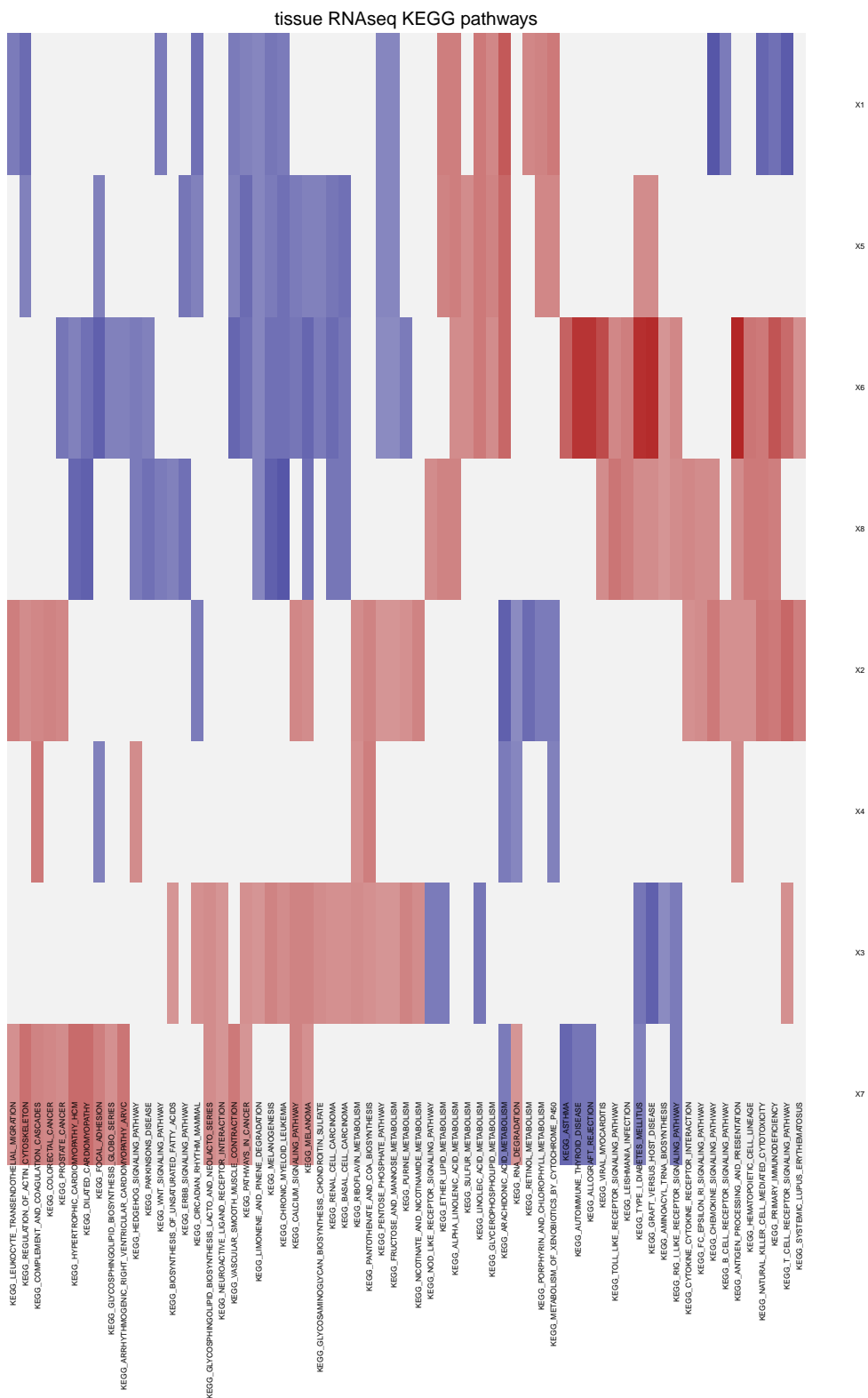
Figure 5. Comparisson of microbiome datasets, showing only the distribution of different bacterial communities (1-8). Samples are ordered by the CVL3 groupings. The colors represent their respective bacterial groupings for each dataset.

ASV_tissue_V3_normalized_batch_corrected



b

ASV_CVL_V3_normalized_batch_corrected



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.