Broliden_5325

28 October, 2020

Contents

Loading data and metadata

Calculate QC metrics

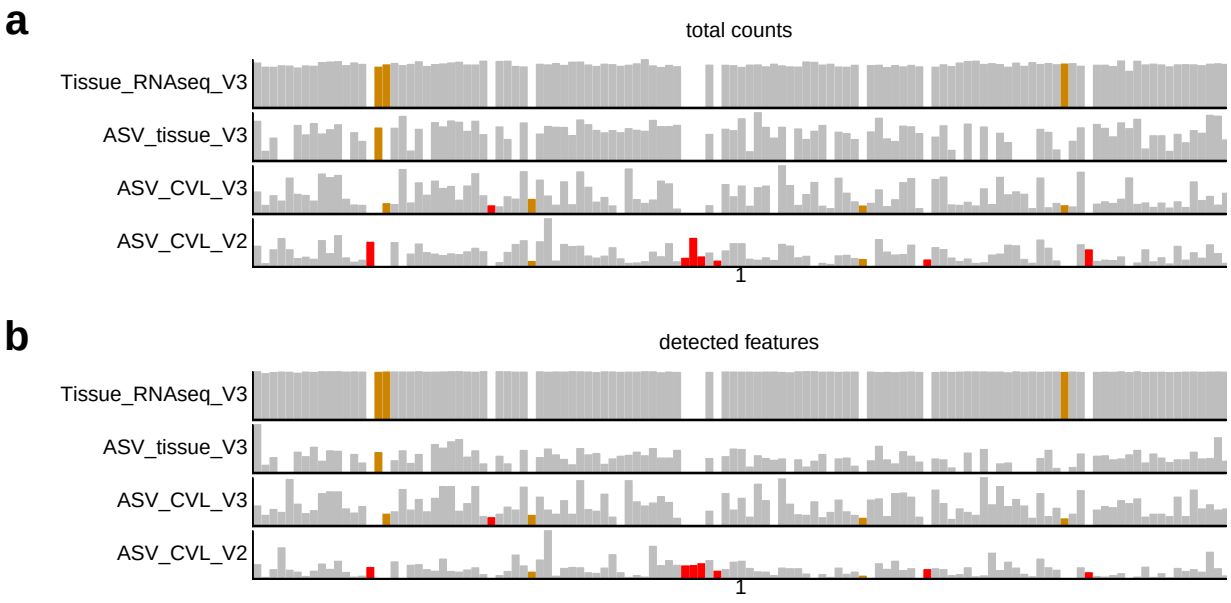


Figure 1. Comparative barplot for the **a)** total counts and **b)** number of non-zero detected features (genes / bacteria) for each of the sequencing datasets. Samples are ordered alfabetically according to the patient ID. Samples that are present in exactly two datasets are shown in orange (P015,P030,P054,P055,P056,P058,P084,P104).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

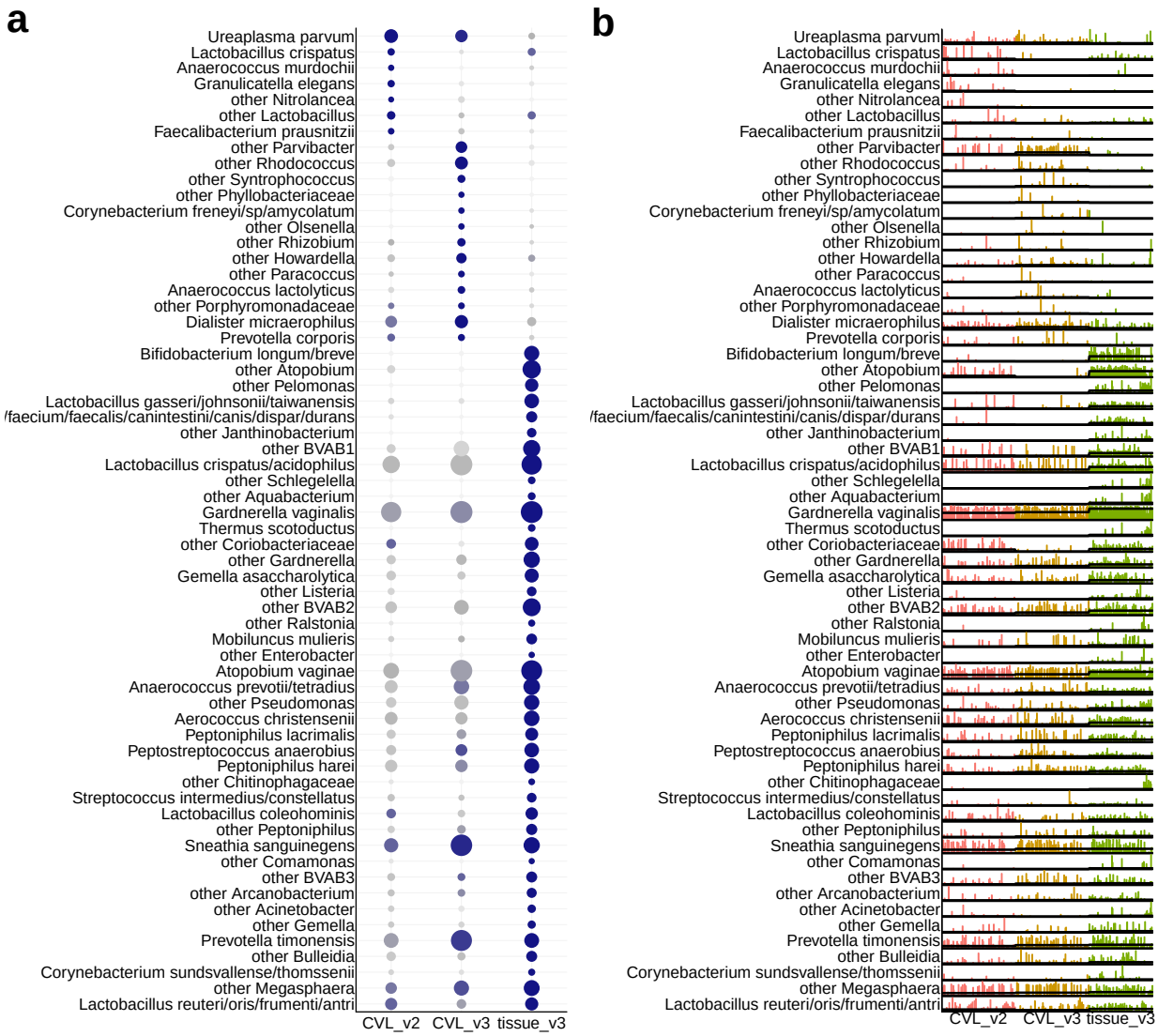


Figure 2. Differential bacterial abundance across microbiome datasets. The results are shown both as **a)** Dot plots and **b)** barplots Bacteria with log2FC above 0.25 and p-value below 0.01 were considered significant and were sorted by the highest expression.

Computing differential expression across microbiome datasets

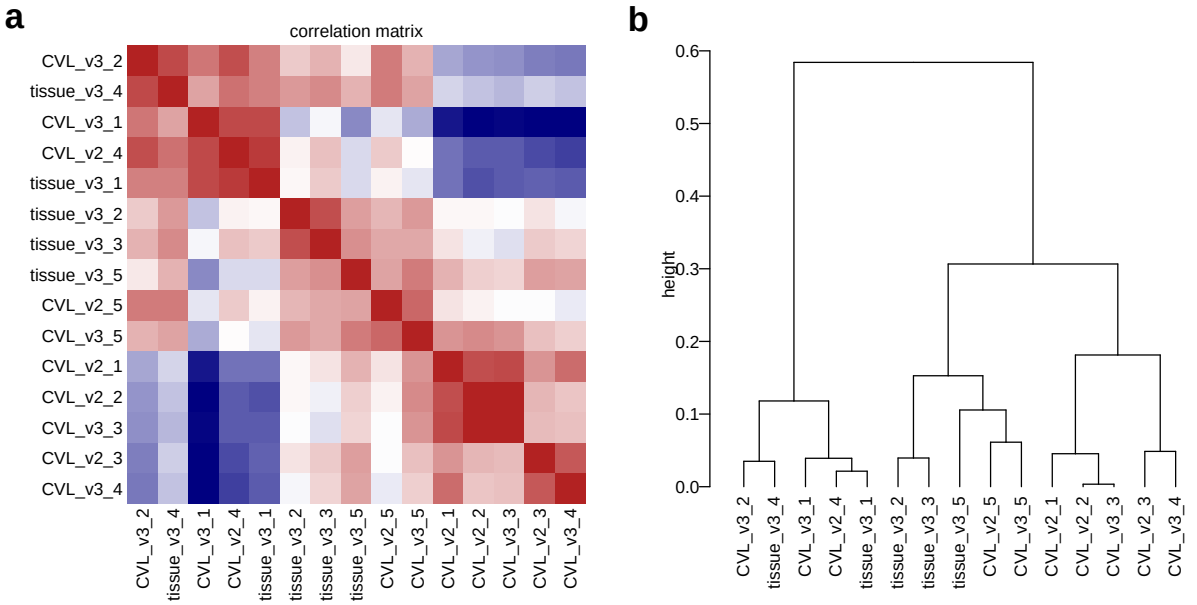


Figure 3. Comparrisson among patient groups across datasets. **(a)** Correlation matrix across sample groups. **(b)** Hierachical clustering of sample groups.

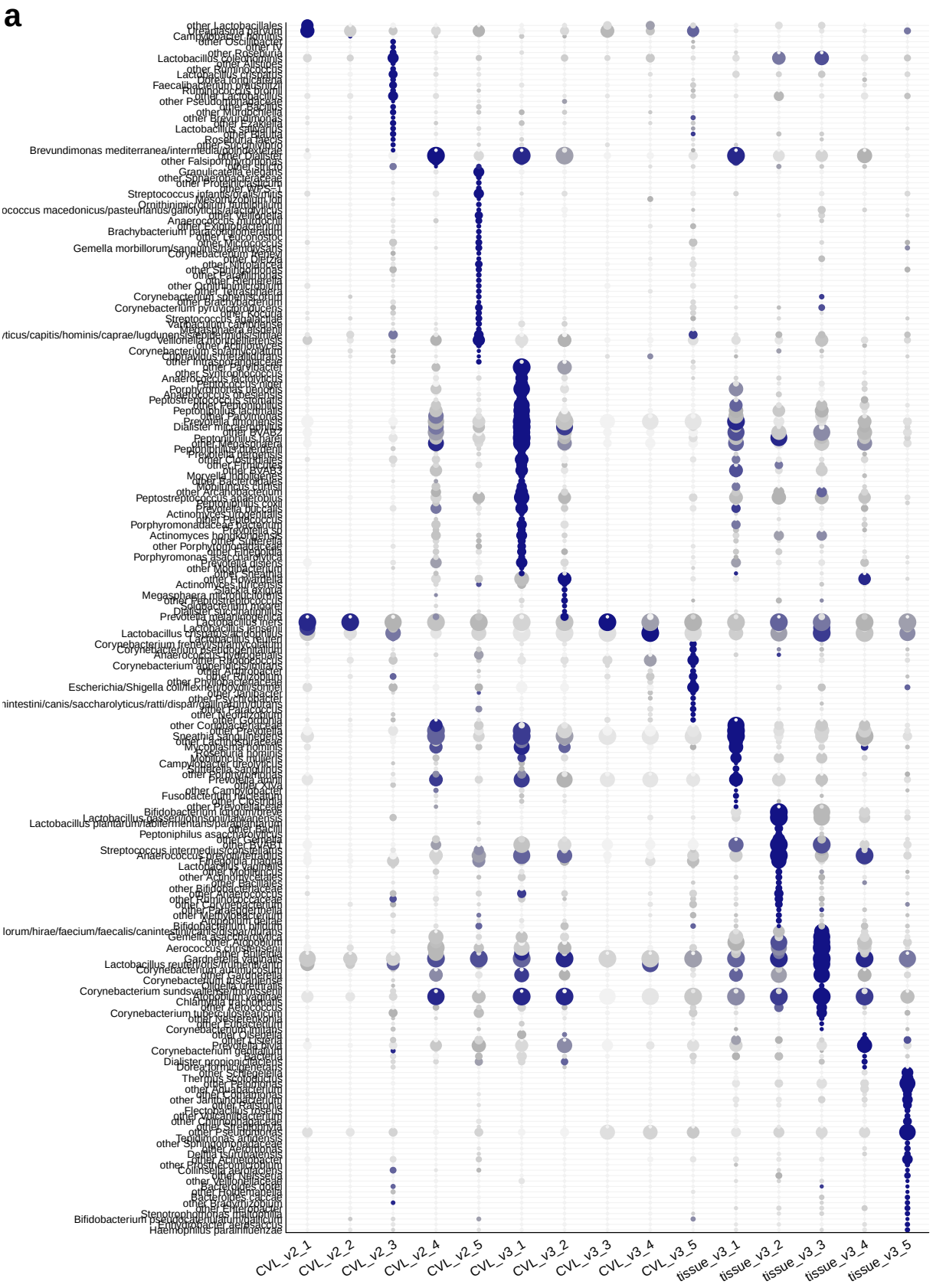

a
CVL_v2_1
CVL_v2_2
CVL_v2_3
CVL_v2_4
CVL_v2_5
CVL_v3_1
CVL_v3_2
CVL_v3_3
CVL_v3_4
CVL_v3_5
tissue_v3_1
tissue_v3_2
tissue_v3_3
tissue_v3_4
tissue_v3_5

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 5. Comparisson 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.

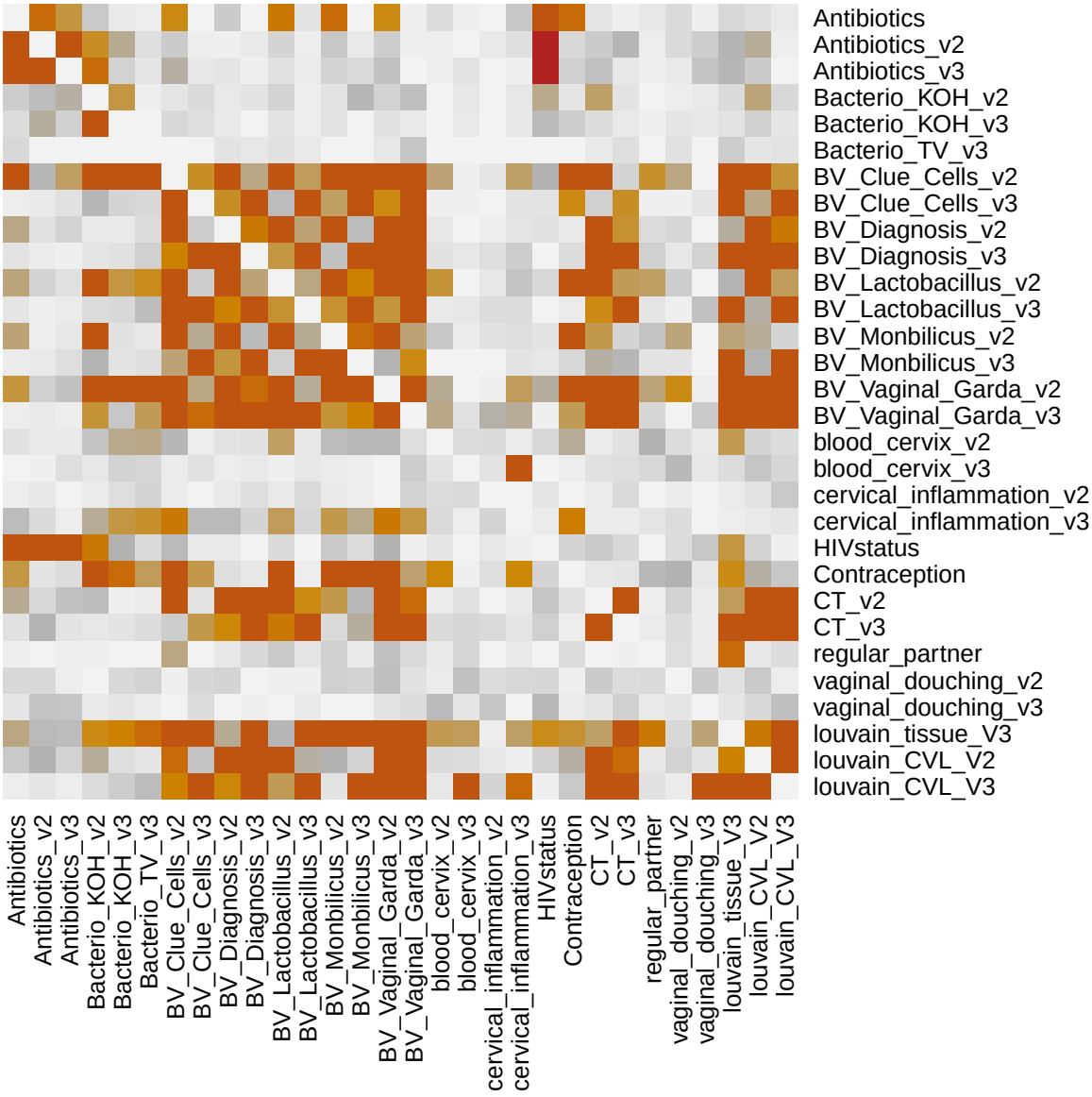


Figure 6. Association analysis across several patient categorical metadata parameters, including patient groupings annotations from microbiome.

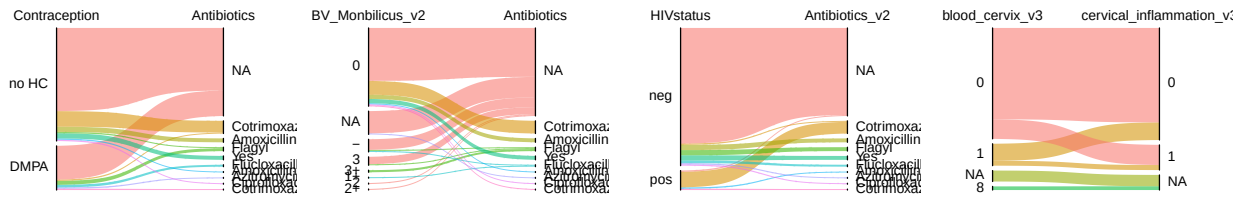
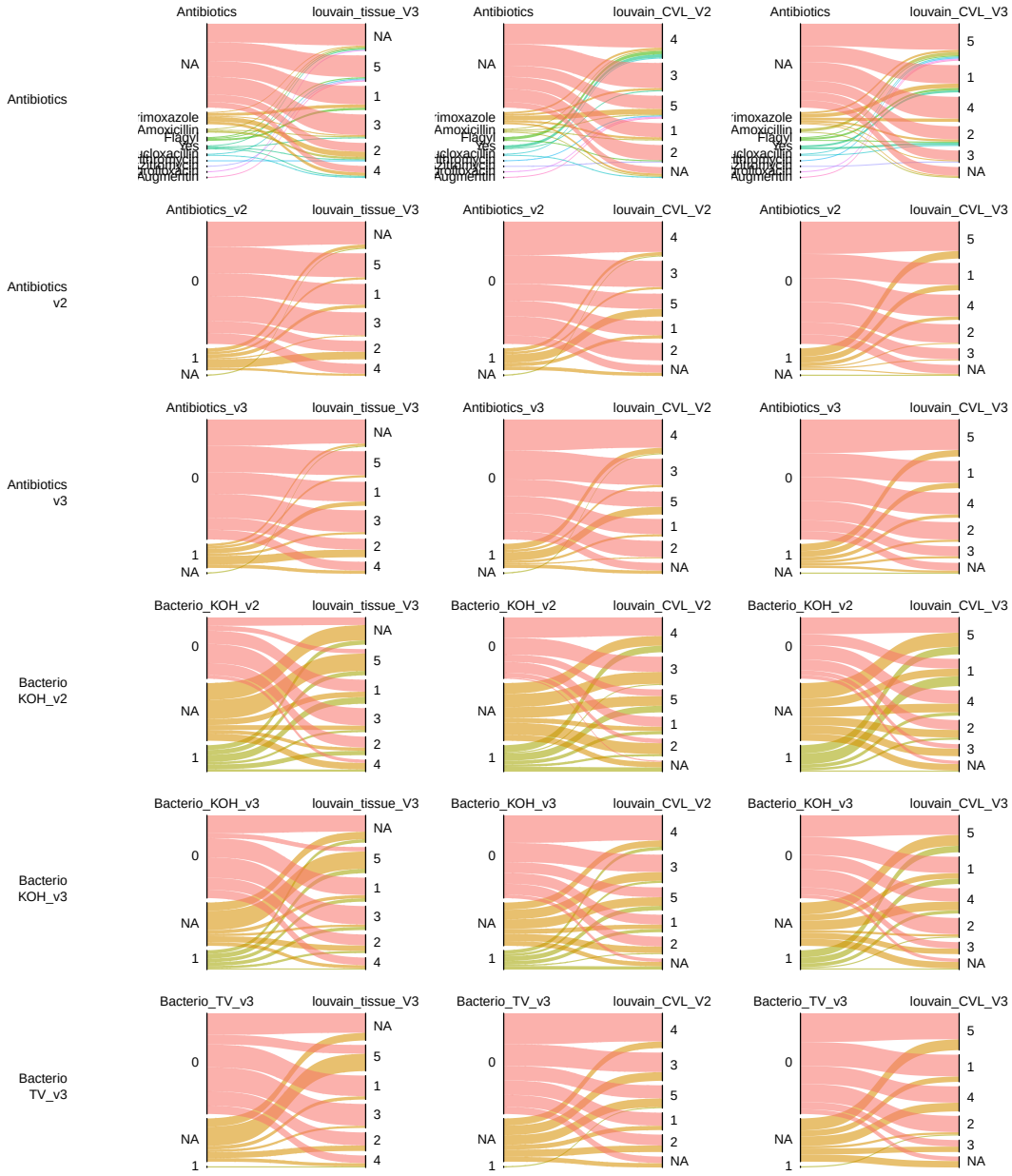
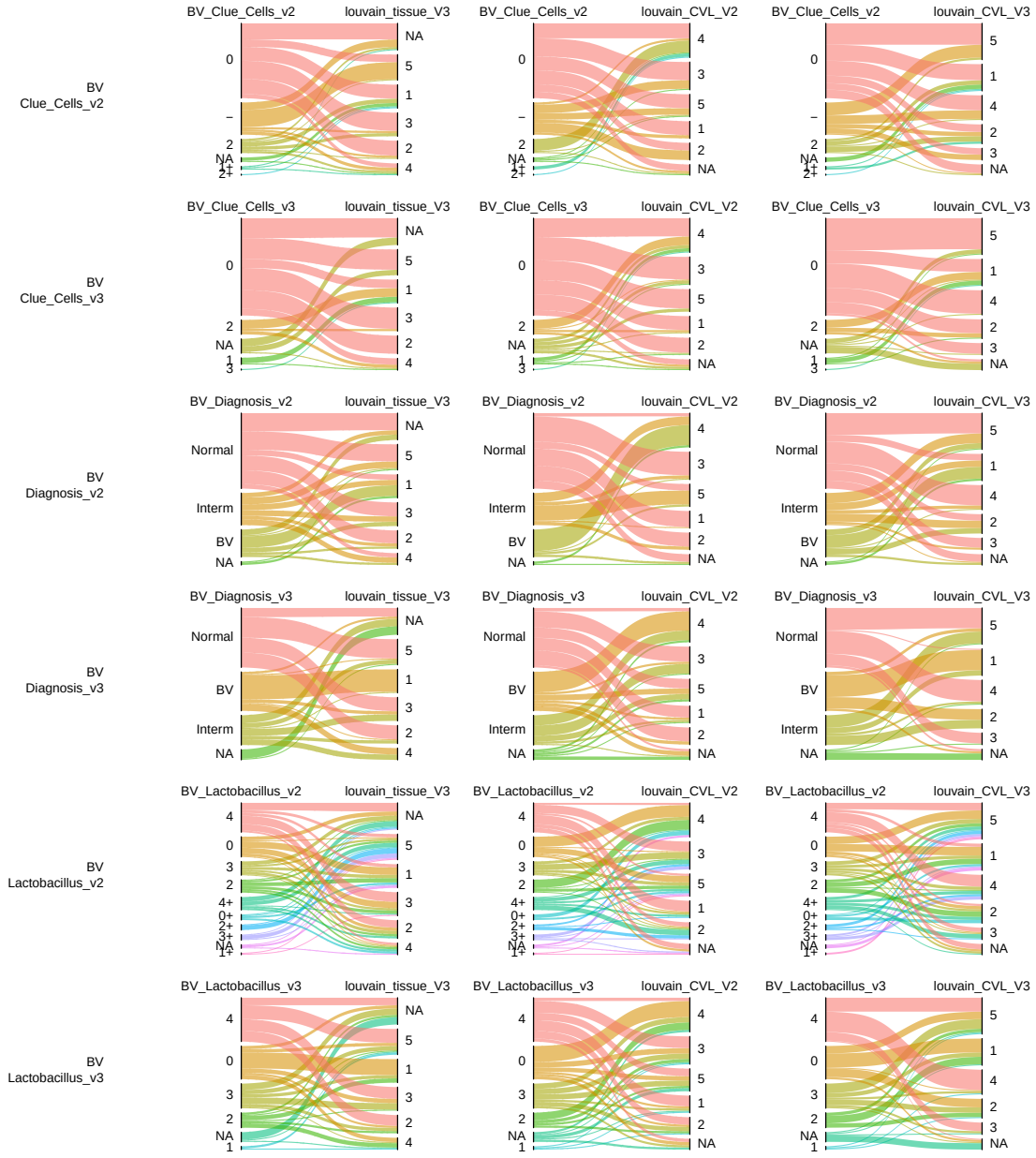
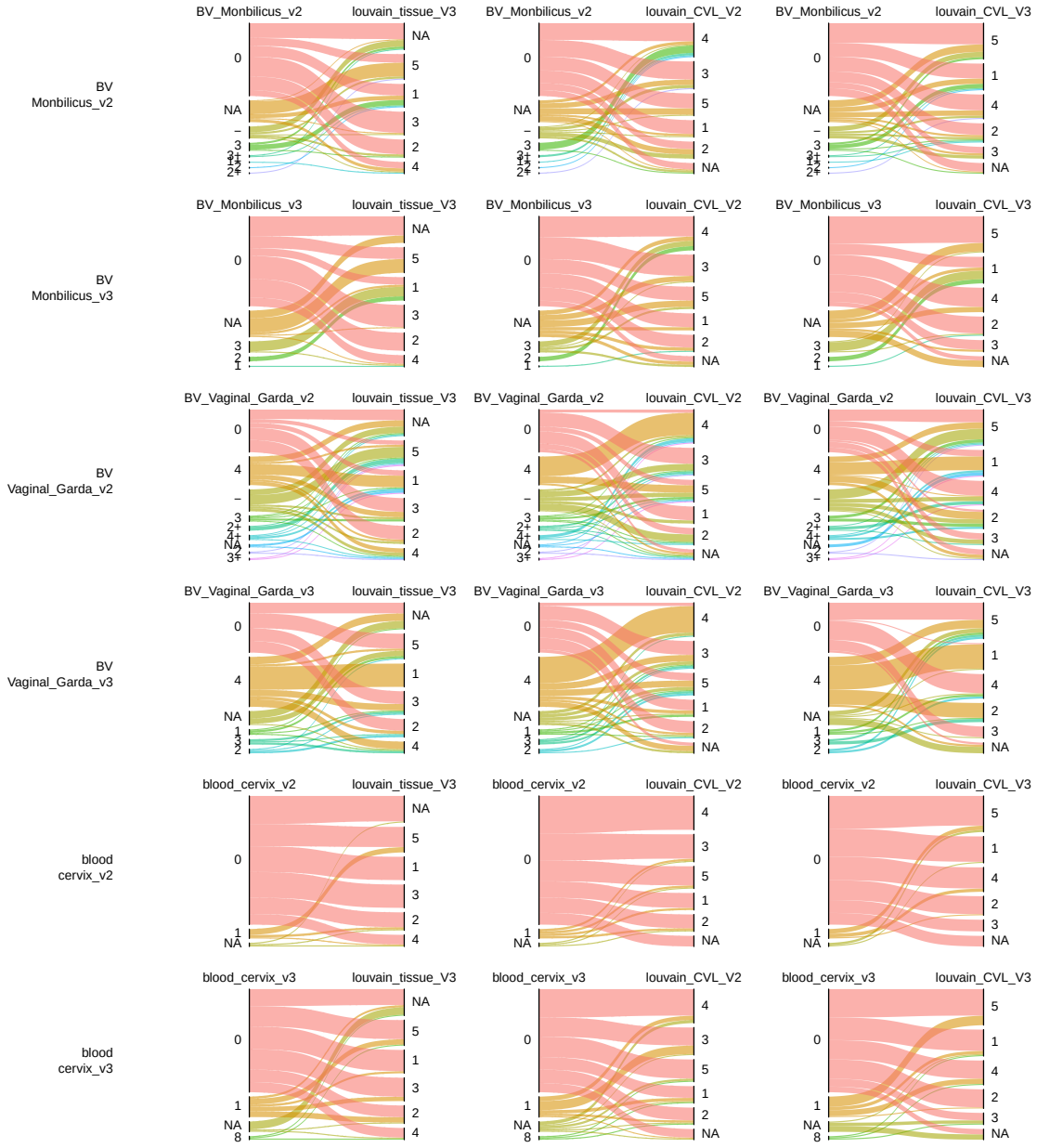
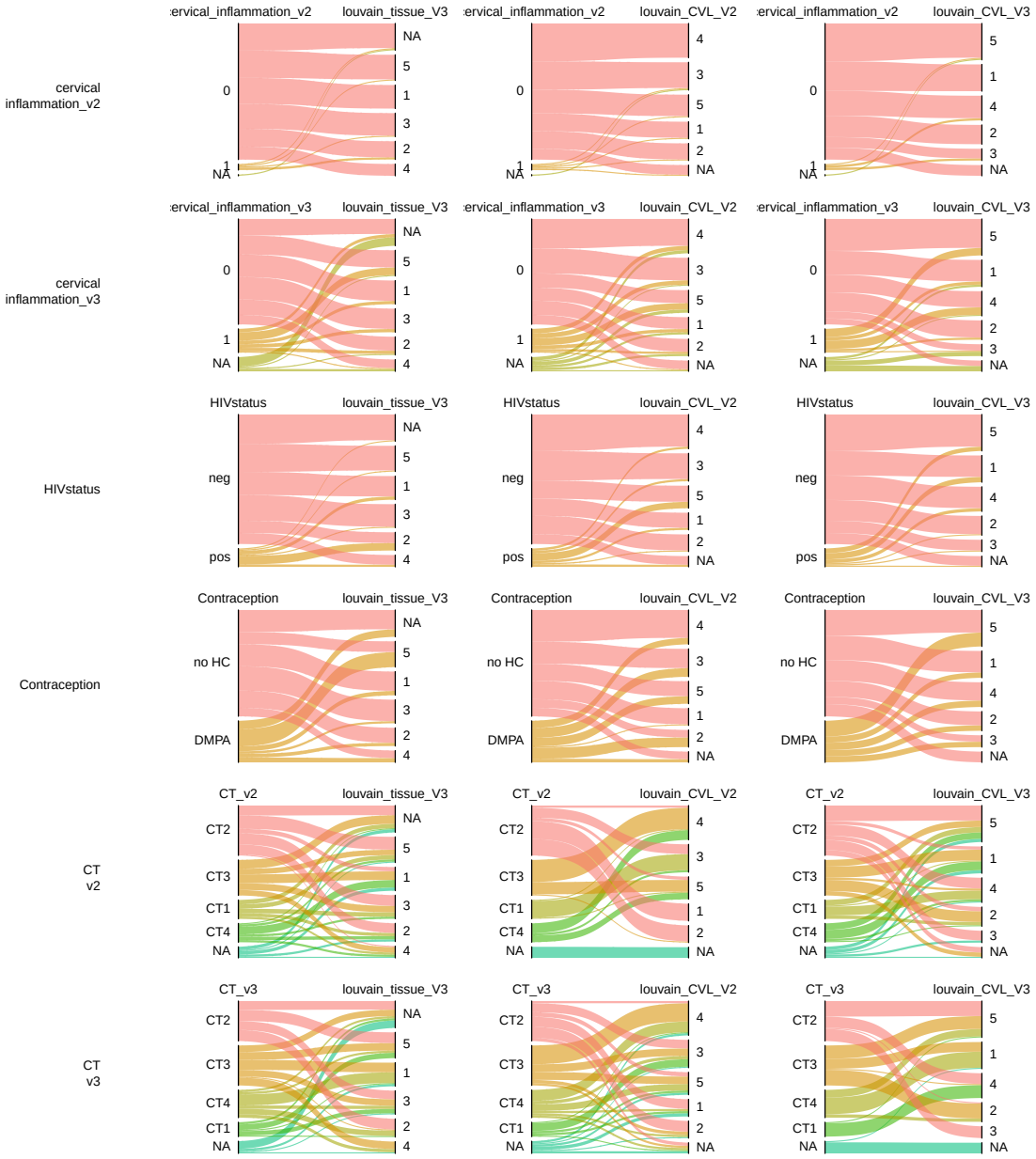


Figure 7. A few examples of significant association between metadata parameters shown as sankey plots.


Antibiotics
Antibiotics_v2
Antibiotics_v3
Bacterio_KOH_v2
Bacterio_KOH_v3
Bacterio_TV_v3
louvain_tissue_V3
louvain_CVL_V2
louvain_CVL_V3
NA
Amoxicillin
Flagyl
Augmentin


BV Clue_Cells_v2
BV_Clue_Cells_v2
louvain_tissue_V3
louvain_CVL_V2
louvain_CVL_V3
BV Clue_Cells_v3
BV_Clue_Cells_v3
BV Diagnosis_v2
BV_Diagnosis_v2
Normal
Interm
BV
NA
BV Diagnosis_v3
BV_Diagnosis_v3
BV Lactobacillus_v2
BV_Lactobacillus_v2
BV Lactobacillus_v3
BV_Lactobacillus_v3


BV_Monbilicus_v2
louvain_tissue_V3
BV_Monbilicus_v2
louvain_CVL_V2
BV_Monbilicus_v2
louvain_CVL_V3
BV Monbilicus_v2
BV_Monbilicus_v3
louvain_tissue_V3
BV_Monbilicus_v3
louvain_CVL_V2
BV_Monbilicus_v3
louvain_CVL_V3
BV Monbilicus_v3
BV_Vaginal_Garda_v2
louvain_tissue_V3
BV_Vaginal_Garda_v2
louvain_CVL_V2
BV_Vaginal_Garda_v2
louvain_CVL_V3
BV Vaginal_Garda_v2
BV_Vaginal_Garda_v3
louvain_tissue_V3
BV_Vaginal_Garda_v3
louvain_CVL_V2
BV_Vaginal_Garda_v3
louvain_CVL_V3
BV Vaginal_Garda_v3
blood_cervix_v2
louvain_tissue_V3
blood_cervix_v2
louvain_CVL_V2
blood_cervix_v2
louvain_CVL_V3
blood cervix_v2
blood_cervix_v3
louvain_tissue_V3
blood_cervix_v3
louvain_CVL_V2
blood_cervix_v3
louvain_CVL_V3
blood cervix_v3


cervical inflammation_v2
cervical inflammation_v3
HIVstatus
Contraception
CT v2
CT v3
louvain_tissue_V3
louvain_CVL_V2
louvain_CVL_V3
0
1
NA
neg
pos
no HC
DMPA
CT1
CT2
CT3
CT4
5
4
3
2

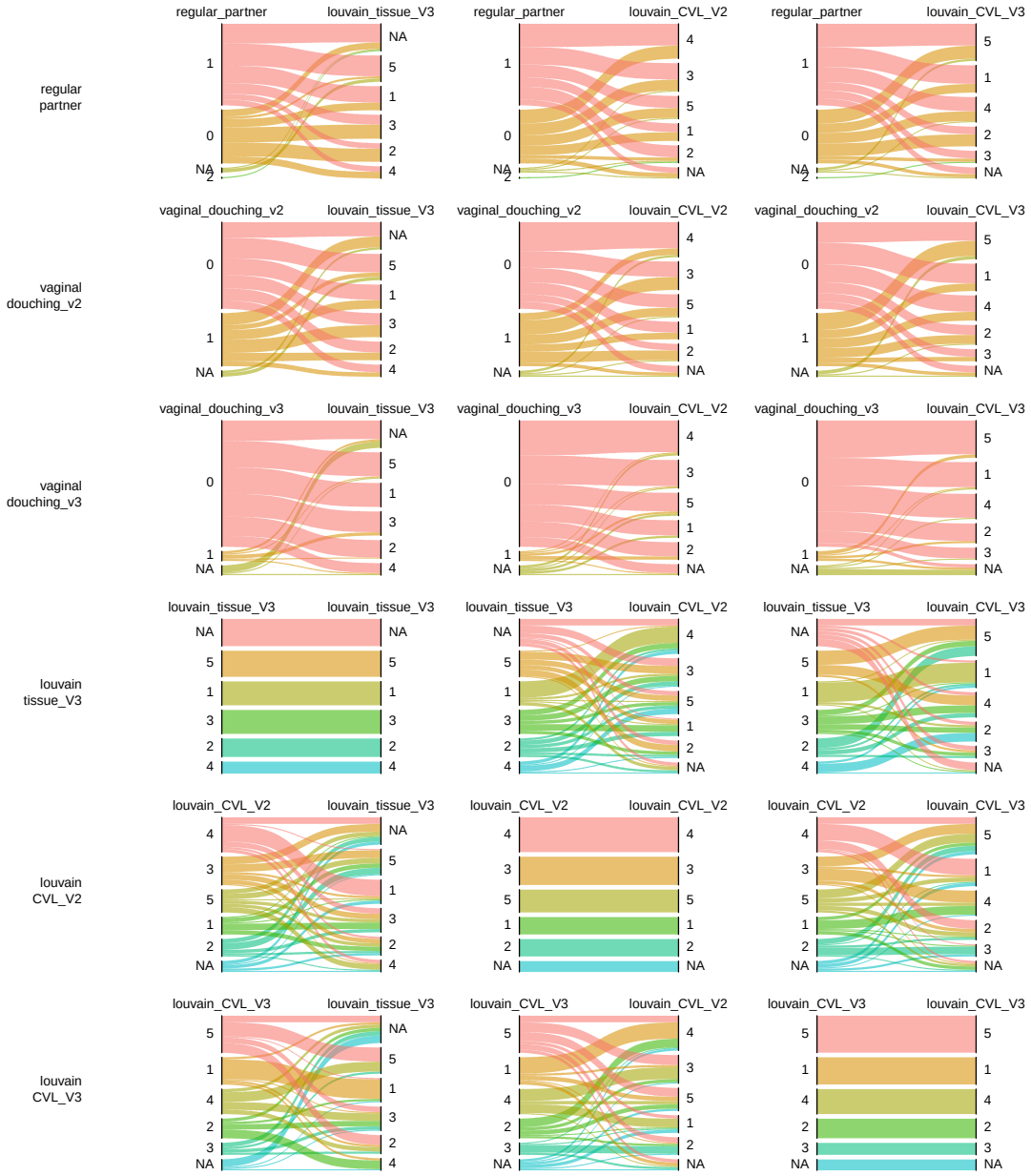
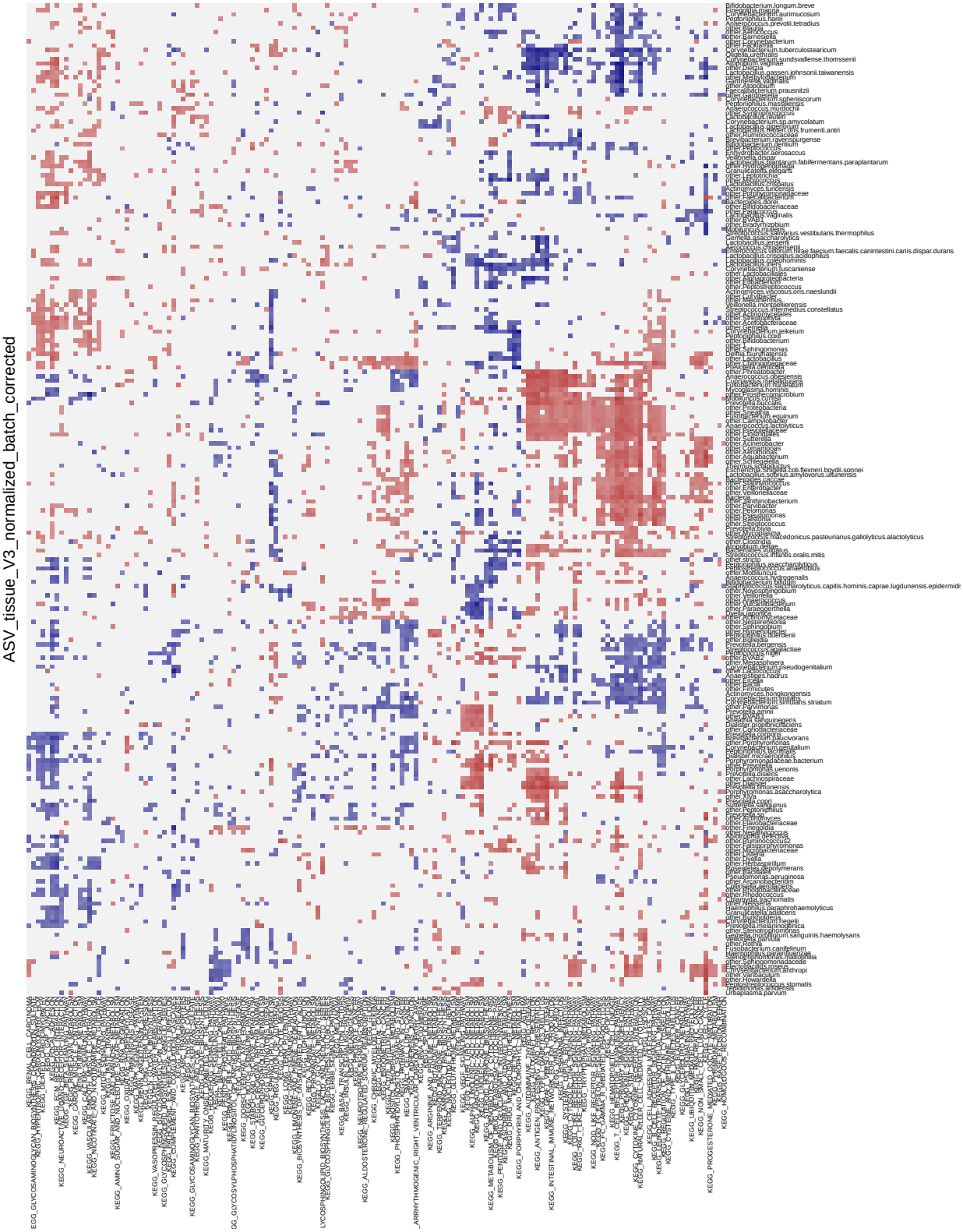
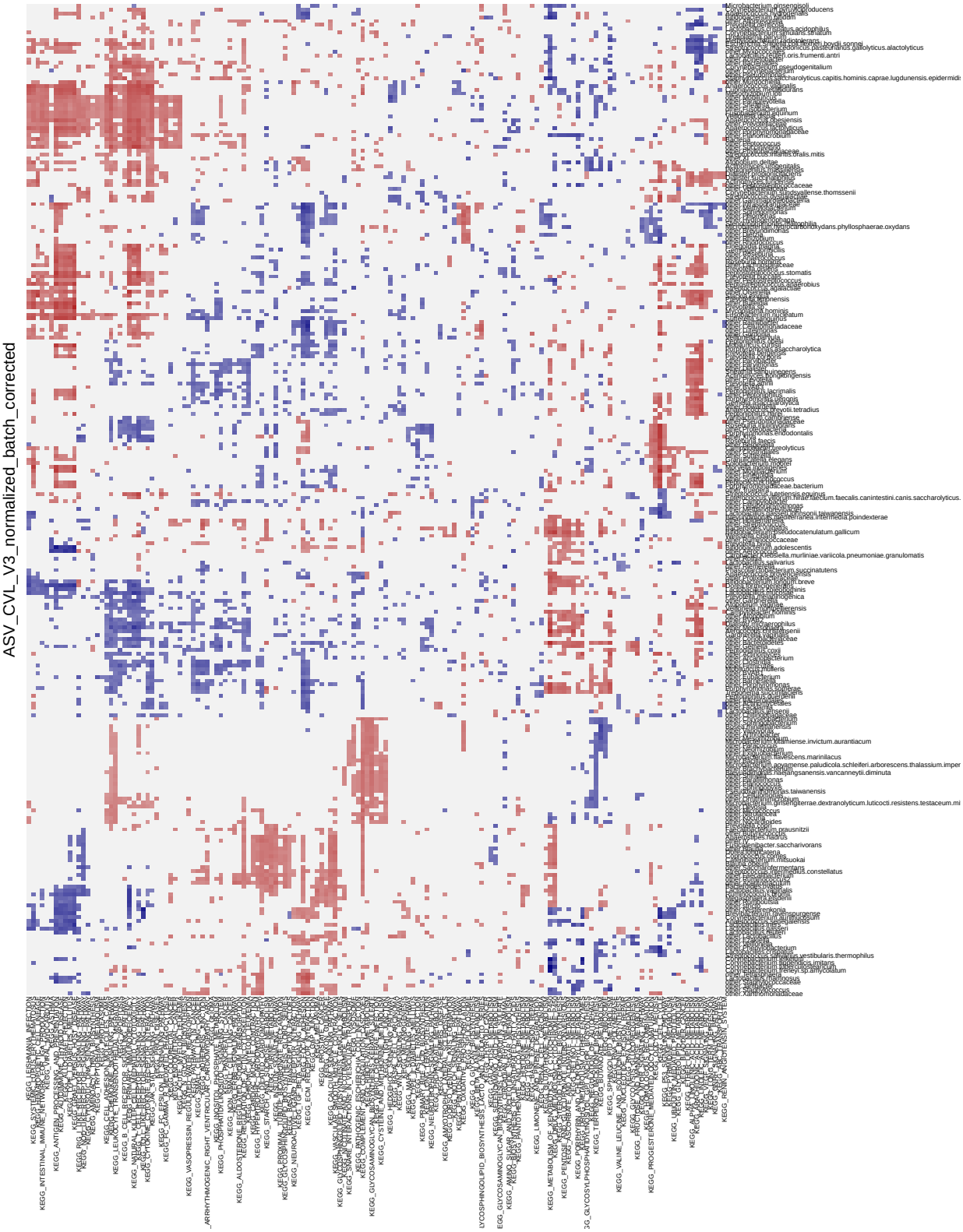


Figure 8. Sankey plots for all tested associations between the patient groups identified in in the microbiome datasets.

a


tissue RNAseq KEGG pathways
ASV_tissue_V3_normalized_batch_corrected

b


tissue RNAseq KEGG pathways
ASV_CVL_V3_normalized_batch_corrected

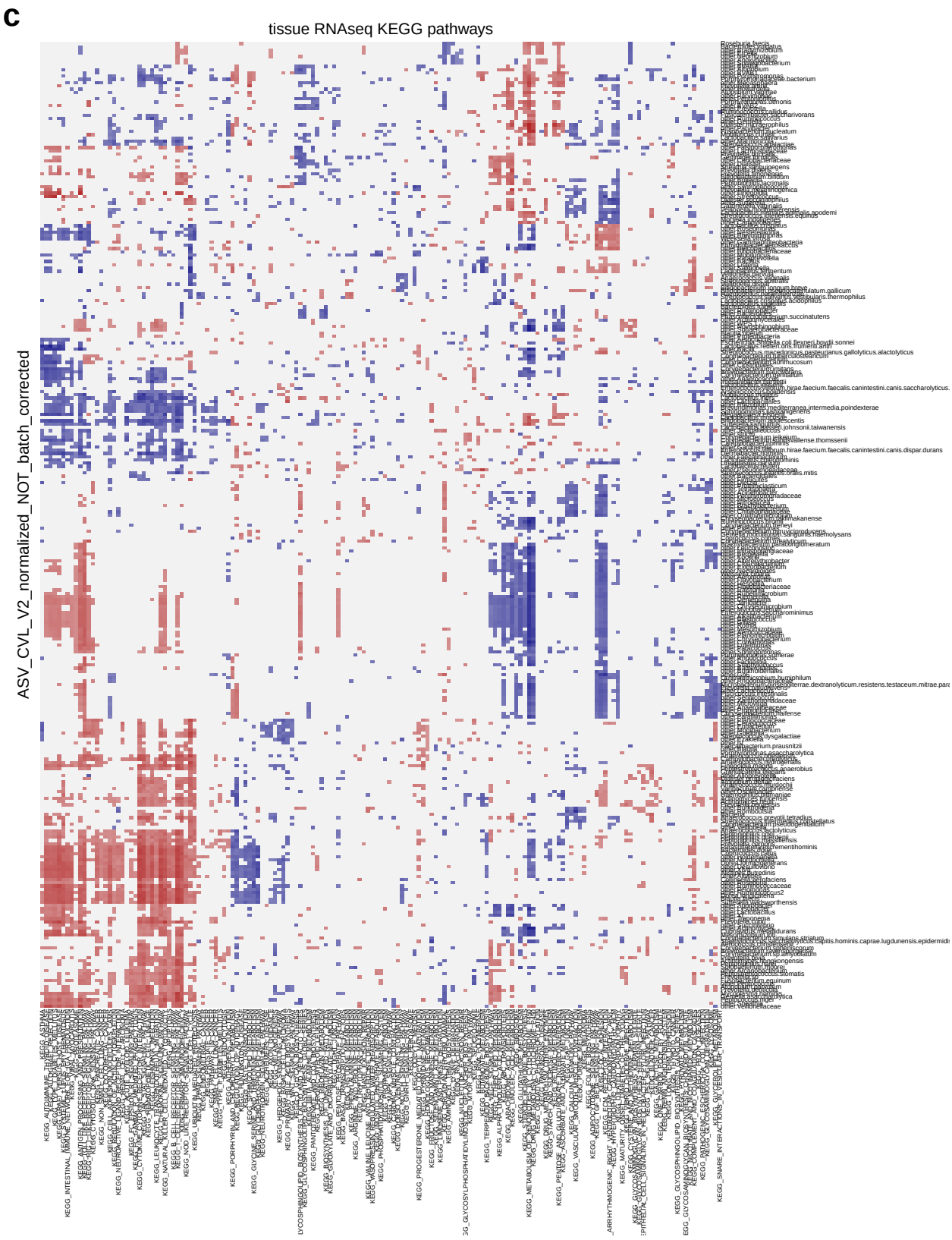


Figure 9. Heatmaps showing the functional association of microbiome datasets with the expression profiles in the RNAseq dataset. Briefly, bacterial abundances from each dataset were correlated with the gene expression

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 anlaysis (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 that 10 pathways and bacteria, respectivelly, were omitted.