Broliden_5325

10 December, 2020

Contents

#Load libraries and other scripts

#Defining some variables for the analysis

Loading data and metadata

Calculate QC metrics
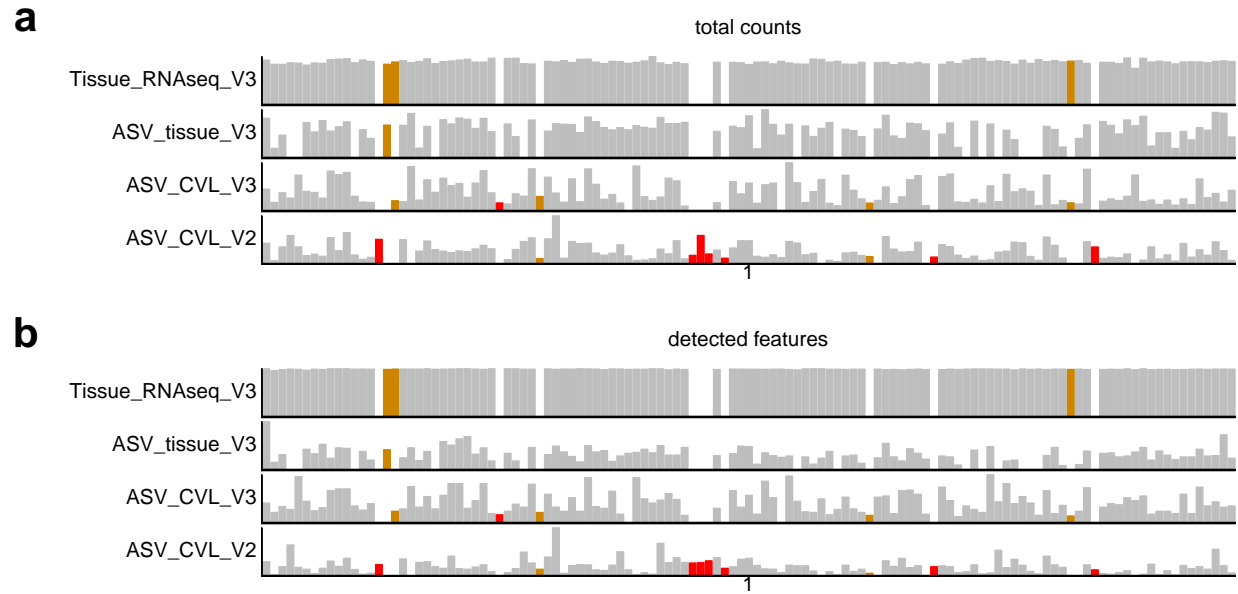
Figure 1. Comparative barplot for the **a)** total counts and **b)** number of non-zero detected features (genes / bacteria) for each of the sequencing datasets. Samples are ordered alfabetically according to the patient ID. Samples that are present in exactly two datasets are shown in orange (P016,P017,P035,P076,P101). Samples that are present in exactly two datasets are shown in red (P015,P030,P054,P055,P056,P058,P084,P104).
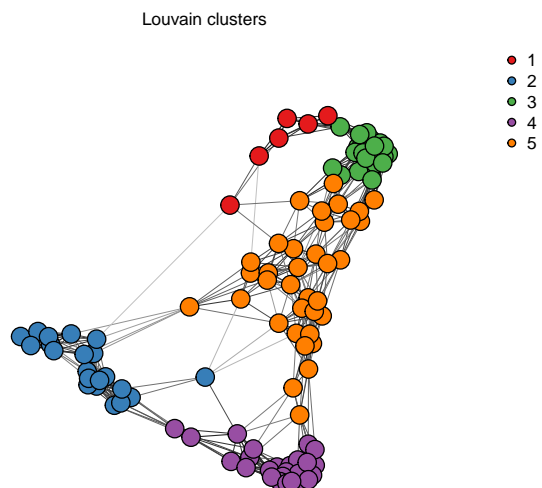
Figure 1b. Sample embedding of 20-SNN graph clustered using Louvain based on the normalized bacterial abundances across all microbiome datasets (including samples not present in all datasets).

Lactobacillus

Monbilicus

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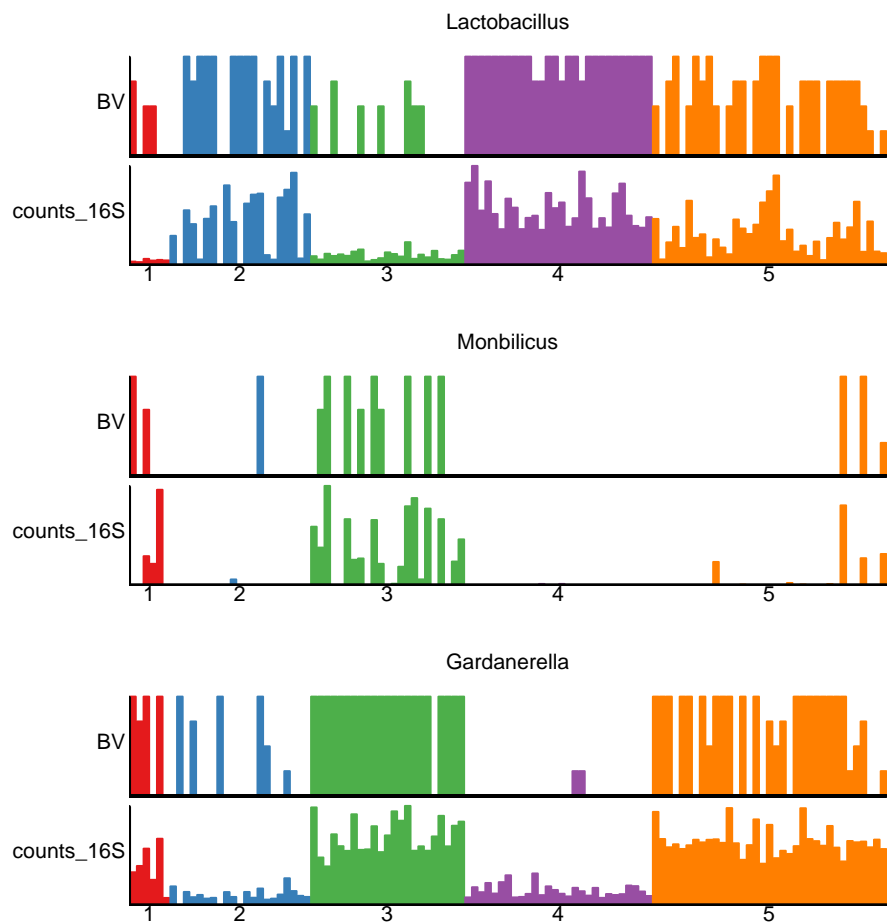
Figure 2. Differential bacterial abundance across joint clustered samples and compared across microbiome datasets. The results are shown both as **a)** Dot plots and **b)** barplots Bacteria with log2FC above 0.25 and p-value below 0.01 were considered significant and were sorted by the highest expression.

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 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.
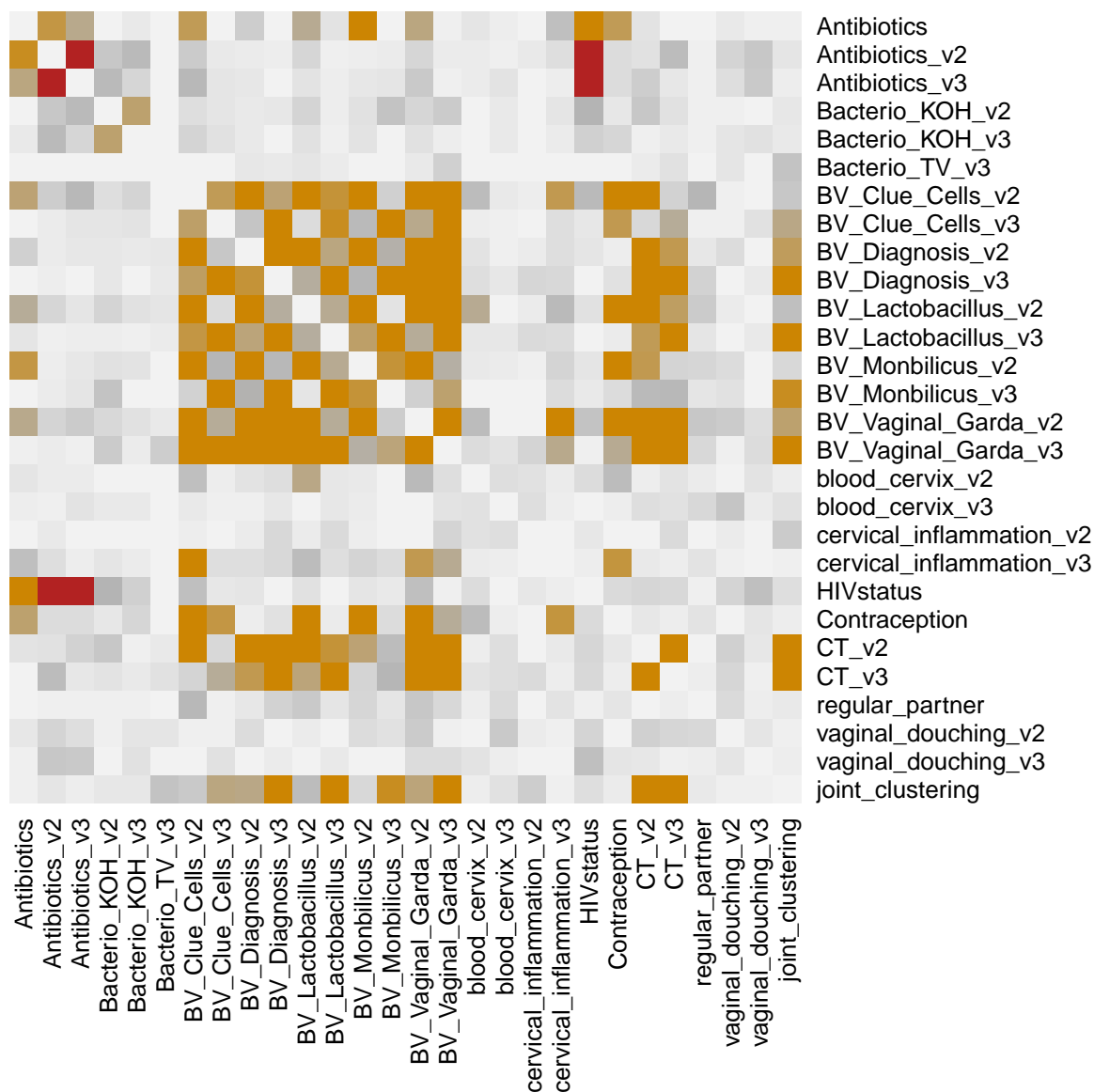
Figure 6. Association analysis across several patient categorical metadata parameters, including patient groupings annotations from microbiome.
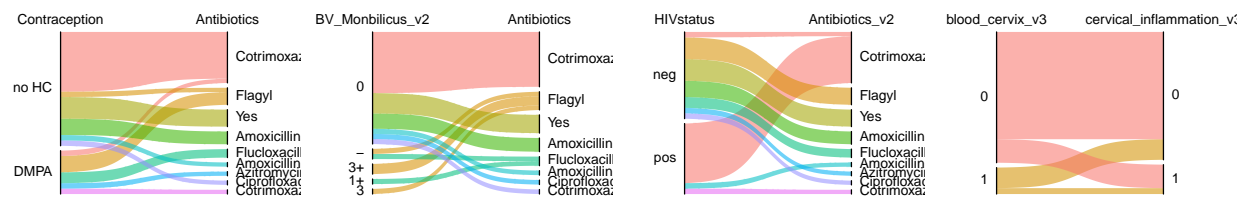
Figure 7. A few examples of significant association between metadata parameters shown as sankey plots.
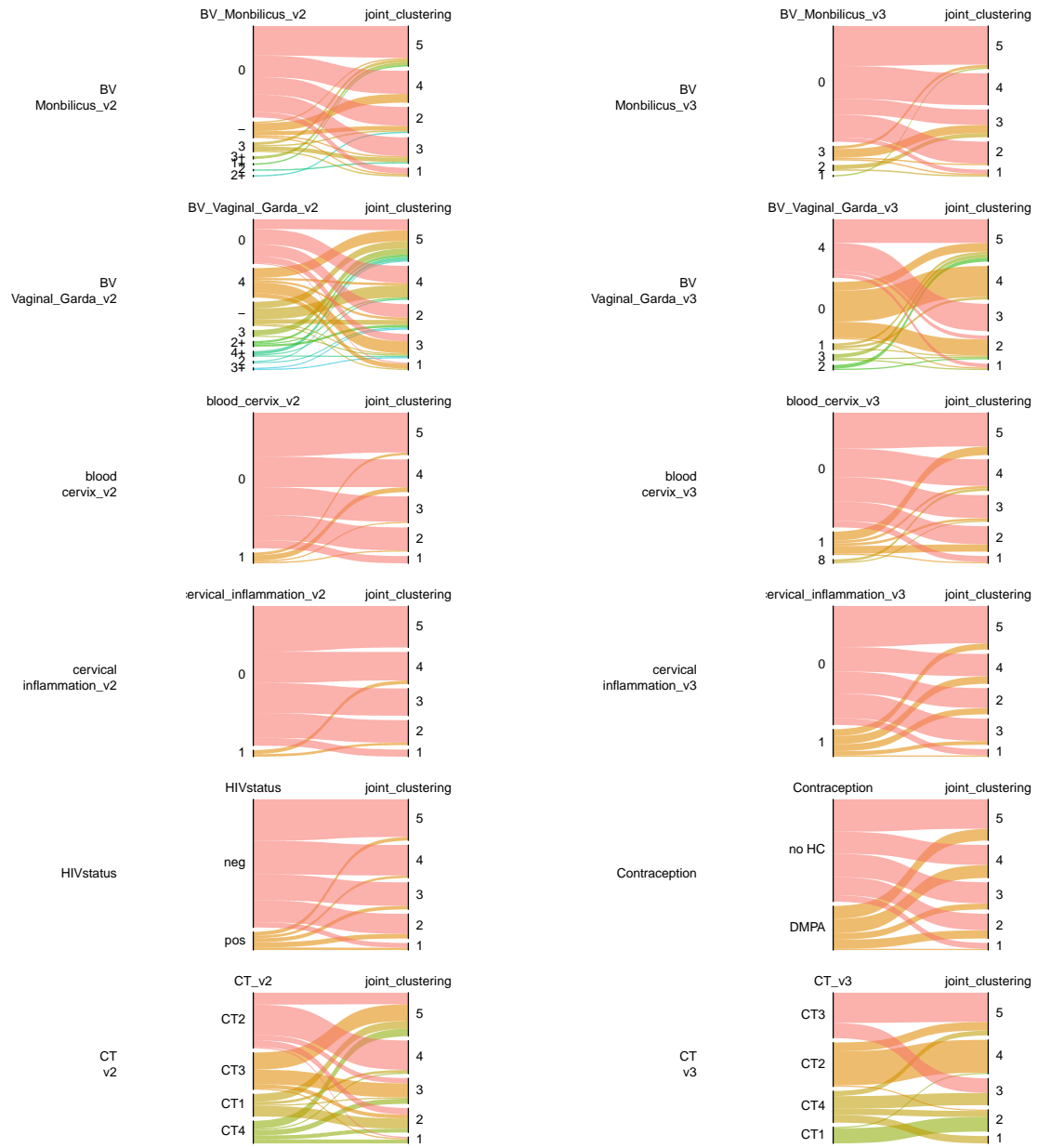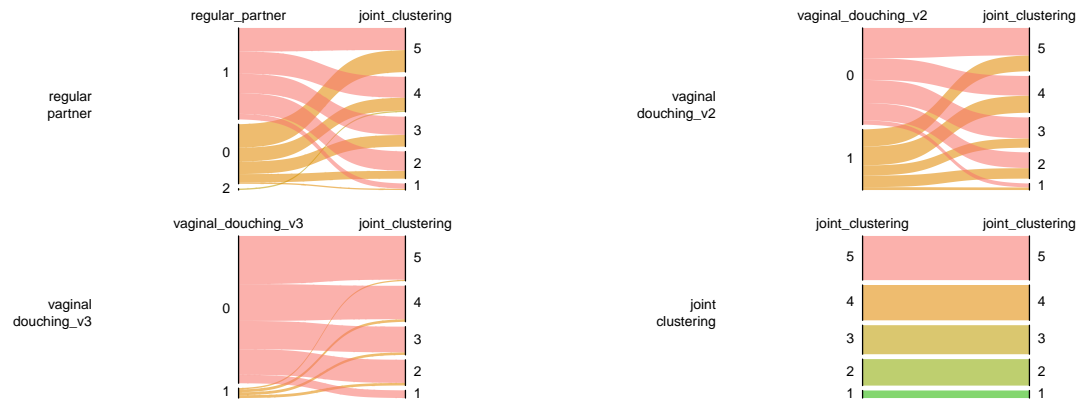
Figure 8. Sankey plots for all tested associations between the patient groups identified in in the microbiome datasets.

Visualise the data

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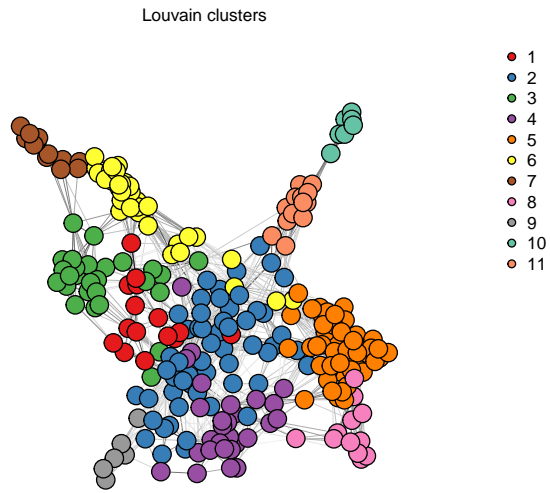
Figure 9a. Bacterial community embedding of 5-SNN graph clustered using walktrap community detection algorithm based on the normalized bacterial abundances across all microbiome datasets (including samples not present in all datasets).

community1

Collinsella aerofaciens
Corynebacterium pseudogenitalium
Enhydrobacter aerosaccus
Haemophilus parainfluenzae
other Gordonia
other Lactococcus
other Mogibacterium
other Rhodobacteraceae
other Streptophyta
other XI
Streptococcus dysgalactiae
Tepidimonas arfidensis
Veillonella dispar
Veillonella parvula
other Gammaproteobacteria

community2

Aerococcus christensenii
Atopobium deltae
Bifidobacterium adolescentis
Bifidobacterium longum/breve
Brevibacterium paucivorans
Brevibacterium ravenspurgense
Campylobacter hominis
Chlamydia trachomatis
Corynebacterium aurimucosum
Corynebacterium genitalium
Corynebacterium imitans
Corynebacterium jeikeium
Corynebacterium pyruviciproducens
Corynebacterium simulans/striatum
Corynebacterium sp/amycolatum
Corynebacterium spheniscorum
Corynebacterium sundsvallense/thor
Corynebacterium tuberculostearicum
Dermabacter hominis
Dorea formicigenerans
Enterococcus villorum/hirae/faecium/
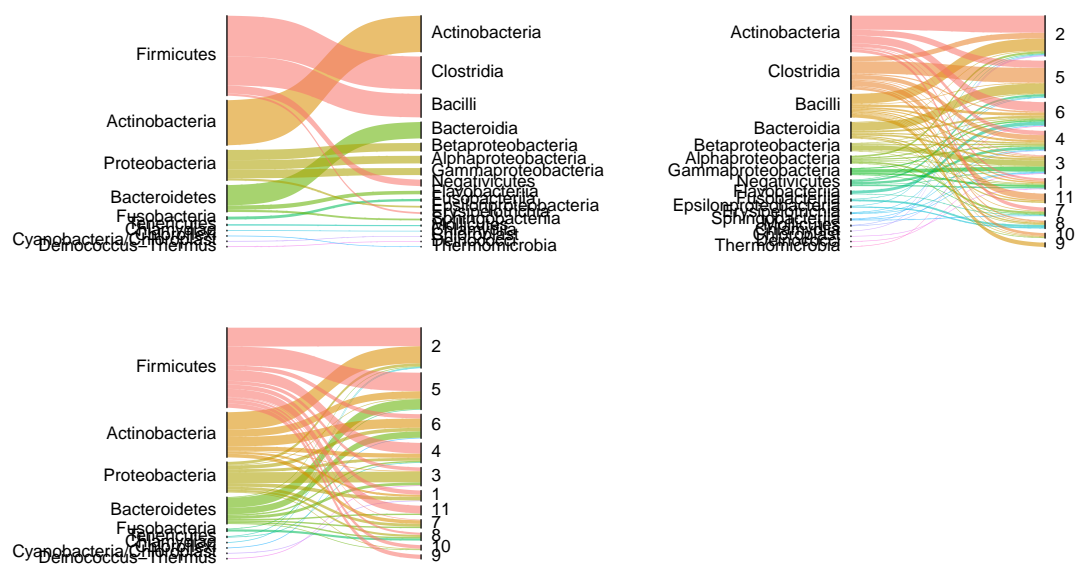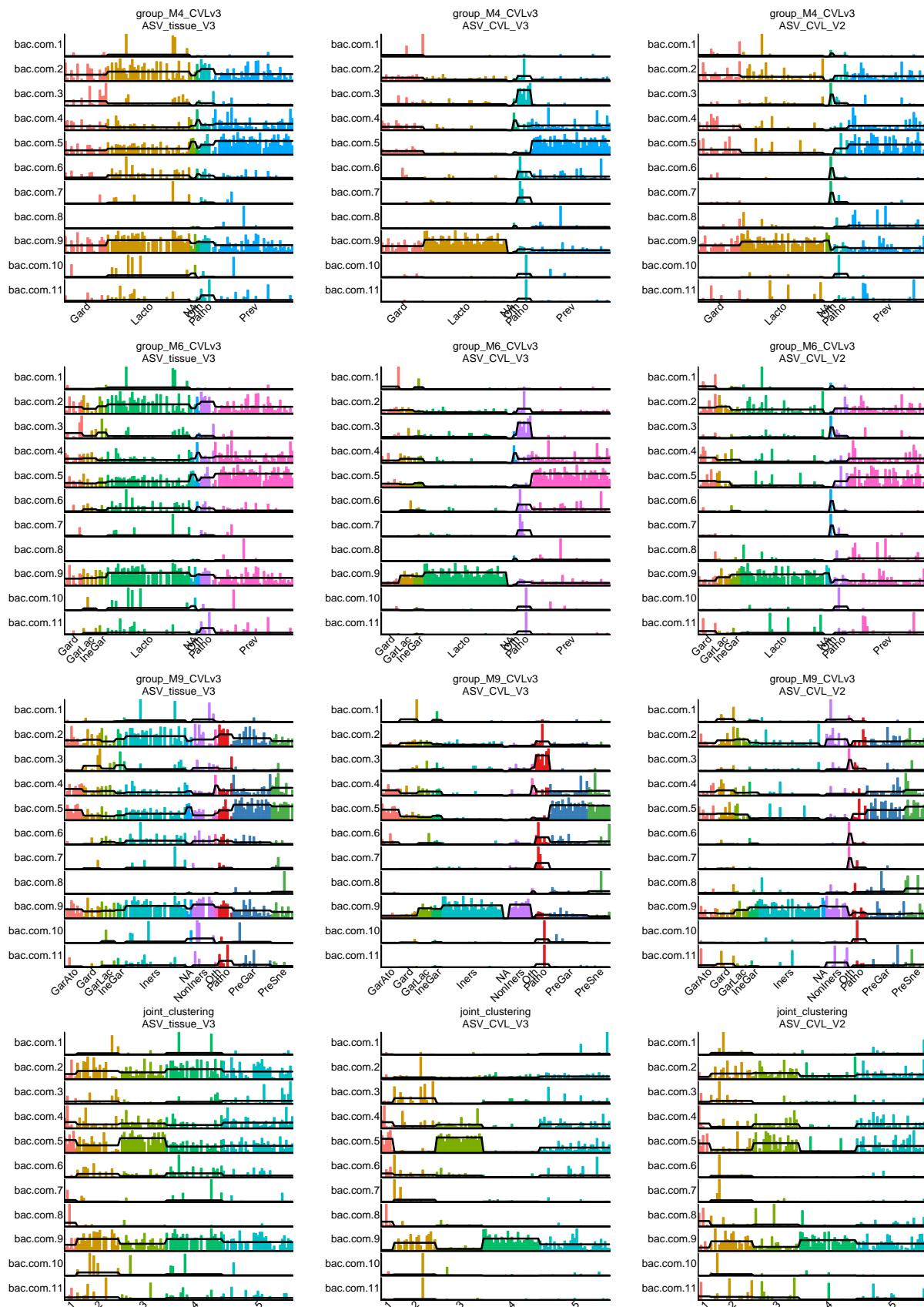Gemella morbillorum/sanguinis/haem
Gemmiger formicilis
Lactobacillus crispatus
Lactobacillus gasseri/johnsonii/taiwar
Lactobacillus mucosae
Lactobacillus reuteri
Lactobacillus salivarius
Megasphaera elsdenii
other Atopobium
other Bacillales
other Bacilli
other Bifidobacteriaceae
other Clostridia
other Corynebacterium
other Ezakiella
other Gemella
other Lactobacillus

community3

Bacteria
Bacteroides caccae
Bacteroides dorei
Bifidobacterium pseudocatenulatum/
Coprococcus comes
Delftia tsuruhatensis
Escherichia/Shigella coli/flexneri/boy
other Acinetobacter
other Aeromonas
other Aquabacterium
other Bacteroidetes
other Chitinophagaceae
other Comamonas
other Enterobacter
other Holdemanella
other Janthinobacterium
other Neisseria
other Novosphingobium
other Pelomonas
other Pseudomonas
other Ralstonia
other Schlegelella
other Staphylococcus
other Streptococcus
other Veillonellaceae
Prevotella denticola
Stenotrophomonas maltophilia
Thermus scotoductus

community4

Actinomyces neuii
Actinomyces turicensis
Anaerococcus hydrogenalis
Anaerococcus murdochii
Bifidobacterium bifidum
Dialister propionicifaciens
Dialister succinatiphilus
Finegoldia magna
Granulicatella elegans
Mycoplasma hominis
other Burkholderia
other Eubacterium
other Howardella
other Hydrogenophaga
other Methylobacterium
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
Methylobacterium radiotolerans

community5

Actinomyces hongkongensis
Actinomyces urogenitalis
Anaerococcus prevotii/tetradius
Atopobium vaginae
Dialister micraerophilus
Gardnerella vaginalis
Gemella asaccharolytica
Mobiluncus curtisii
Mobiluncus mulieris
Moryella indoligenes
other Anaerococcus
other Arcanobacterium
other Barnesiella
other Bulleidia
other BVAB1
other BVAB2
other BVAB3
other Clostridiales
other Coriobacteriaceae
other Dialister
other Falsiporphyromonas
other Finegoldia
other Gardnerella
other Lachnospiraceae
other Megasphaera
other Parvibacter
other Parvimonas
other Peptococcus
other Peptoniphilus
other Porphyromonas
other Prevotella
other Sutterella
other Syntrophococcus
Peptococcus niger
Peptoniphilus coxii
Peptoniphilus duerdenii
Peptoniphilus harei
Peptoniphilus lacrimalis

community6

Brachybacterium paraconglomeratun
Chryseobacterium haifense
Corynebacterium freneyi
Cupriavidus metallidurans
Lactobacillus plantarum/fabifermenta
Mesorhizobium loti
other Actinomyces
other Actinomycetales
other Aerococcus
other Bacteroidales
other Bergeyella
other Brachybacterium
other Cloacibacterium
other Dietzia
other Facklamia
other Firmicutes
other Flavobacteriaceae
other Intrasporangiaceae
other Kocuria
other Leuconostoc
other Micrococcus
other Nitrolancea
other Nocardioides
other Ornithinimicrobium
other Parafilimonas
other Porphyromonadaceae
other Riemerella
other Rothia
other Sphingomonas
other stricto
other Succinivibrio
other Tetrasphaera
other Xanthomonadaceae
Porphyromonas somerae
Ruminococcus bromii

community7

other Arthrobacter
other Chryseobacterium
other Exiguobacterium
other Janibacter
other Paracoccus
other Phyllobacteriaceae
other Psychrobacter
other Rhodococcus
Microbacterium aoyamense/paludico
other Aeromicrobium
other Neorhizobium
other Sphingobacterium

community8

Anaerococcus lactolyticus
Anaerococcus obesiensis
Campylobacter ureolyticus
Fusobacterium equinum
Fusobacterium nucleatum
other Campylobacter
other Paraprevotella
other Prevotellaceae
other Proteobacteria
other Sneathia
Streptococcus salivarius/vestibularis/
Sutterella sanguinus

community9

Lactobacillus coleohominis
Lactobacillus crispatus/acidophilus
Lactobacillus iners
Lactobacillus jensenii
Lactobacillus reuteri/oris/frumenti/ant
other Lactobacillales

community10

Bacteroides vulgatus
Blautia obeum
Faecalibacterium prausnitzii
other Bacillus
other Bradyrhizobium
other Roseburia
Roseburia faecis

community11

Anaerostipes hadrus
Dorea longicatena
Lactobacillus vaginalis
other Alloprevotella
other Blautia
other Faecalibacterium
other IV
other Oscillibacter
other Ruminococcaceae
other Ruminococcus2
Prevotella copri
Streptococcus intermedius/constellat
Fusicatenibacter saccharivorans

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

Figure 9b. Taxonomic annotation of bacterial communities.

group_M4_CVLv3 ASV_tissue_V3 · group_M4_CVLv3 ASV_CVL_V3 · group_M4_CVLv3 ASV_CVL_V2

group_M6_CVLv3 ASV_tissue_V3 · group_M6_CVLv3 ASV_CVL_V3 · group_M6_CVLv3 ASV_CVL_V2

group_M9_CVLv3 ASV_tissue_V3 · group_M9_CVLv3 ASV_CVL_V3 · group_M9_CVLv3 ASV_CVL_V2

joint_clustering ASV_tissue_V3 · joint_clustering ASV_CVL_V3 · joint_clustering ASV_CVL_V2

Figure 5. Comparisson of microbiome communities across all datasets and all clustering methods.
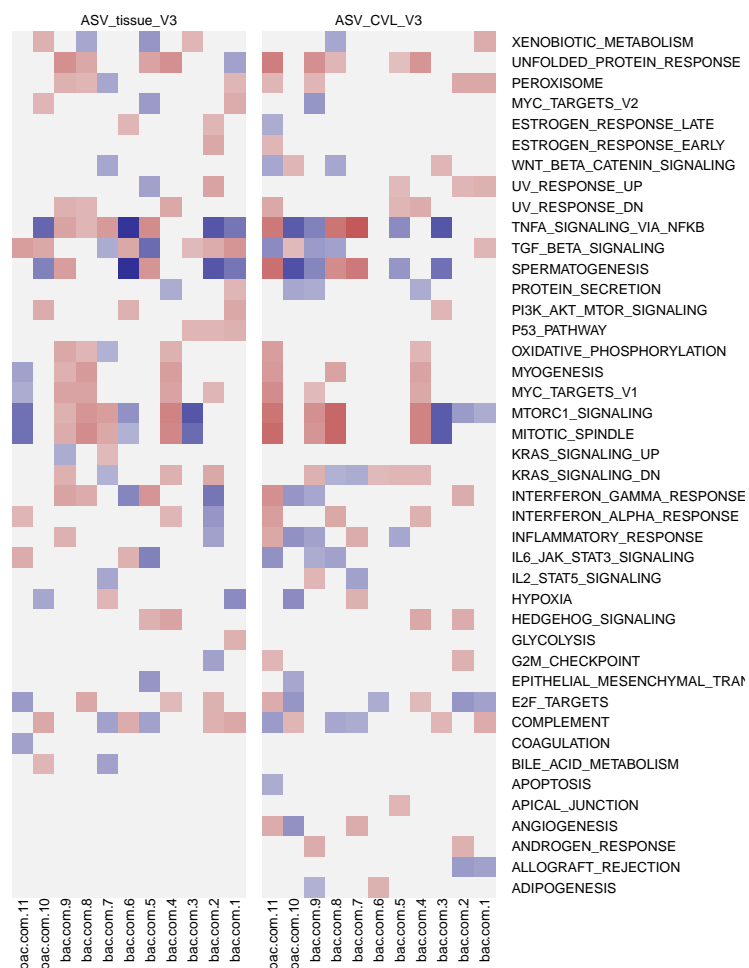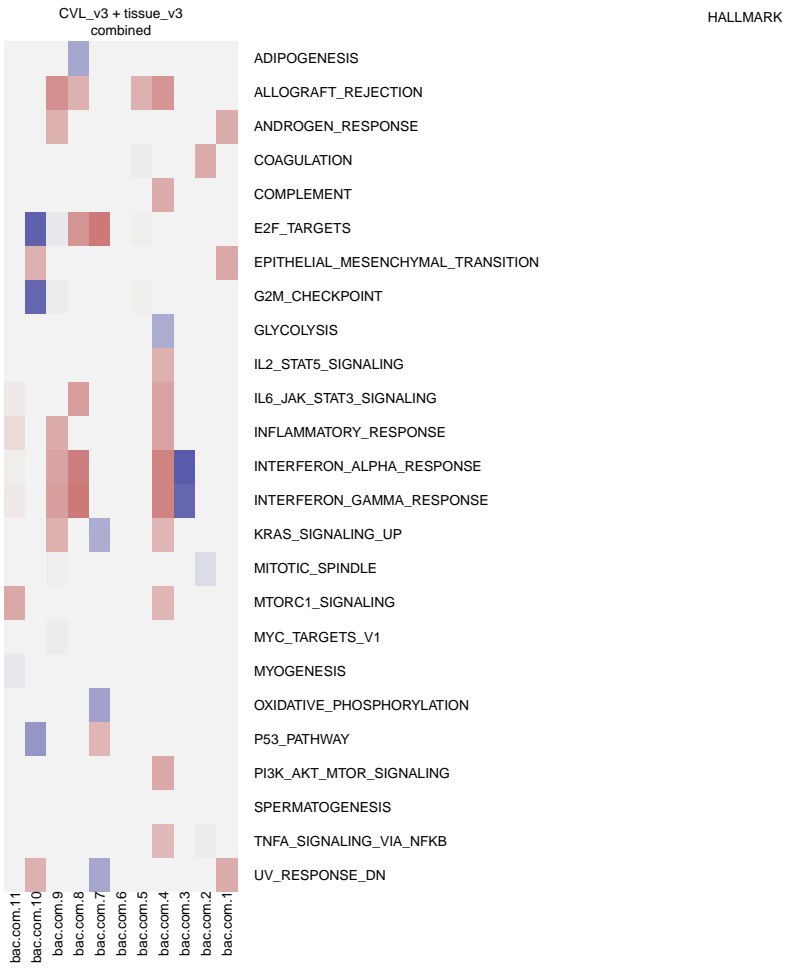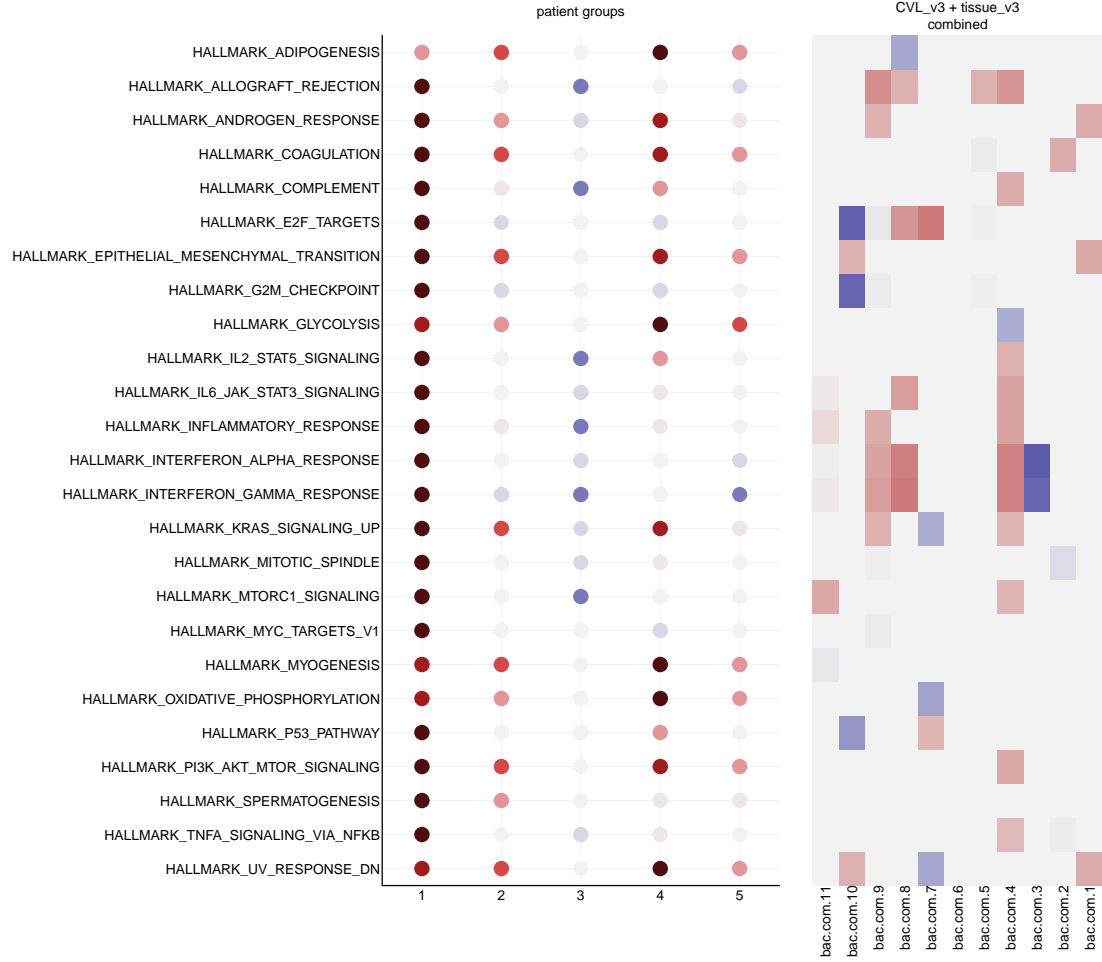
CVL_v3 + tissue_v3
combined

HALLMARK

ADIPOGENESIS
ALLOGRAFT_REJECTION
ANDROGEN_RESPONSE
COAGULATION
COMPLEMENT
E2F_TARGETS
EPITHELIAL_MESENCHYMAL_TRANSITION
G2M_CHECKPOINT
GLYCOLYSIS
IL2_STAT5_SIGNALING
IL6_JAK_STAT3_SIGNALING
INFLAMMATORY_RESPONSE
INTERFERON_ALPHA_RESPONSE
INTERFERON_GAMMA_RESPONSE
KRAS_SIGNALING_UP
MITOTIC_SPINDLE
MTORC1_SIGNALING
MYC_TARGETS_V1
MYOGENESIS
OXIDATIVE_PHOSPHORYLATION
P53_PATHWAY
PI3K_AKT_MTOR_SIGNALING
SPERMATOGENESIS
TNFA_SIGNALING_VIA_NFKB
UV_RESPONSE_DN

bac.com.11
bac.com.10
bac.com.9
bac.com.8
bac.com.7
bac.com.6
bac.com.5
bac.com.4
bac.com.3
bac.com.2
bac.com.1

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 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.

ASV_tissue_V3   ASV_CVL_V3   KEGG

ACUTE_MYELOID_LEUKEMIA
ADHERENS_JUNCTION
ALLOGRAFT_REJECTION
ALPHA_LINOLENIC_ACID_METABOLISM
ALZHEIMERS_DISEASE
AMINOACYL_TRNA_BIOSYNTHESIS
ANTIGEN_PROCESSING_AND_PRESENTATION
APOPTOSIS
ARACHIDONIC_ACID_METABOLISM
ARRHYTHMOGENIC_RIGHT_VENTRICULAR_CARDIOMYOPATHY_ARVC
ASCORBATE_AND_ALDARATE_METABOLISM
AUTOIMMUNE_THYROID_DISEASE
AXON_GUIDANCE
BASAL_CELL_CARCINOMA
BIOSYNTHESIS_OF_UNSATURATED_FATTY_ACIDS
CALCIUM_SIGNALING_PATHWAY
CELL_ADHESION_MOLECULES_CAMS
CHEMOKINE_SIGNALING_PATHWAY
CHRONIC_MYELOID_LEUKEMIA
COMPLEMENT_AND_COAGULATION_CASCADES
CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION
CYSTEINE_AND_METHIONINE_METABOLISM
DILATED_CARDIOMYOPATHY
DRUG_METABOLISM_CYTOCHROME_P450
DRUG_METABOLISM_OTHER_ENZYMES
ECM_RECEPTOR_INTERACTION
ENDOMETRIAL_CANCER
ETHER_LIPID_METABOLISM
FC_EPSILON_RI_SIGNALING_PATHWAY
FC_GAMMA_R_MEDIATED_PHAGOCYTOSIS
FOCAL_ADHESION
GAP_JUNCTION
GLIOMA
GLUTATHIONE_METABOLISM
GLYCOSAMINOGLYCAN_BIOSYNTHESIS_CHONDROITIN_SULFATE
GLYCOSPHINGOLIPID_BIOSYNTHESIS_LACTO_AND_NEOLACTO_SERIES
HEDGEHOG_SIGNALING_PATHWAY
HEMATOPOIETIC_CELL_LINEAGE
HYPERTROPHIC_CARDIOMYOPATHY_HCM
INOSITOL_PHOSPHATE_METABOLISM
INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION
JAK_STAT_SIGNALING_PATHWAY
LEISHMANIA_INFECTION
LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION
LIMONENE_AND_PINENE_DEGRADATION
LINOLEIC_ACID_METABOLISM
LONG_TERM_DEPRESSION
METABOLISM_OF_XENOBIOTICS_BY_CYTOCHROME_P450
NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY
NON_SMALL_CELL_LUNG_CANCER
NUCLEOTIDE_EXCISION_REPAIR
PATHWAYS_IN_CANCER
PENTOSE_AND_GLUCURONATE_INTERCONVERSIONS
PHOSPHATIDYLINOSITOL_SIGNALING_SYSTEM
PORPHYRIN_AND_CHLOROPHYLL_METABOLISM
PRIMARY_BILE_ACID_BIOSYNTHESIS
PRIMARY_IMMUNODEFICIENCY
PRION_DISEASES
PROGESTERONE_MEDIATED_OOCYTE_MATURATION
PURINE_METABOLISM
REGULATION_OF_ACTIN_CYTOSKELETON
RETINOL_METABOLISM
RIBOSOME
RIG_I_LIKE_RECEPTOR_SIGNALING_PATHWAY
SMALL_CELL_LUNG_CANCER
STEROID_HORMONE_BIOSYNTHESIS
SYSTEMIC_LUPUS_ERYTHEMATOSUS
TGF_BETA_SIGNALING_PATHWAY
TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY
TYPE_I_DIABETES_MELLITUS
T_CELL_RECEPTOR_SIGNALING_PATHWAY
UBIQUITIN_MEDIATED_PROTEOLYSIS
VASCULAR_SMOOTH_MUSCLE_CONTRACTION
VIRAL_MYOCARDITIS
WNT_SIGNALING_PATHWAY
BASE_EXCISION_REPAIR
BLADDER_CANCER
COLORECTAL_CANCER
CYTOSOLIC_DNA_SENSING_PATHWAY
ERBB_SIGNALING_PATHWAY
GLYCOSAMINOGLYCAN_BIOSYNTHESIS_KERATAN_SULFATE
GLYCOSPHINGOLIPID_BIOSYNTHESIS_GLOBO_SERIES
GLYOXYLATE_AND_DICARBOXYLATE_METABOLISM
INSULIN_SIGNALING_PATHWAY
MELANOGENESIS_AND_CERAMIDE_METABOLISM
NOD_LIKE_RECEPTOR_SIGNALING_PATHWAY
OOCYTE_MEIOSIS
P53_SIGNALING_PATHWAY
PANCREATIC_CANCER
RIBOFLAVIN_METABOLISM
STARCH_AND_SUCROSE_METABOLISM
TERPENOID_BACKBONE_BIOSYNTHESIS
TIGHT_JUNCTION
VASOPRESSIN_REGULATED_WATER_REABSORPTION

bac.com.1
bac.com.2
bac.com.3
bac.com.4
bac.com.5
bac.com.6
bac.com.7
bac.com.8
bac.com.9
bac.com.10
bac.com.11

bac.com.1
bac.com.2
bac.com.3
bac.com.4
bac.com.5
bac.com.6
bac.com.7
bac.com.8
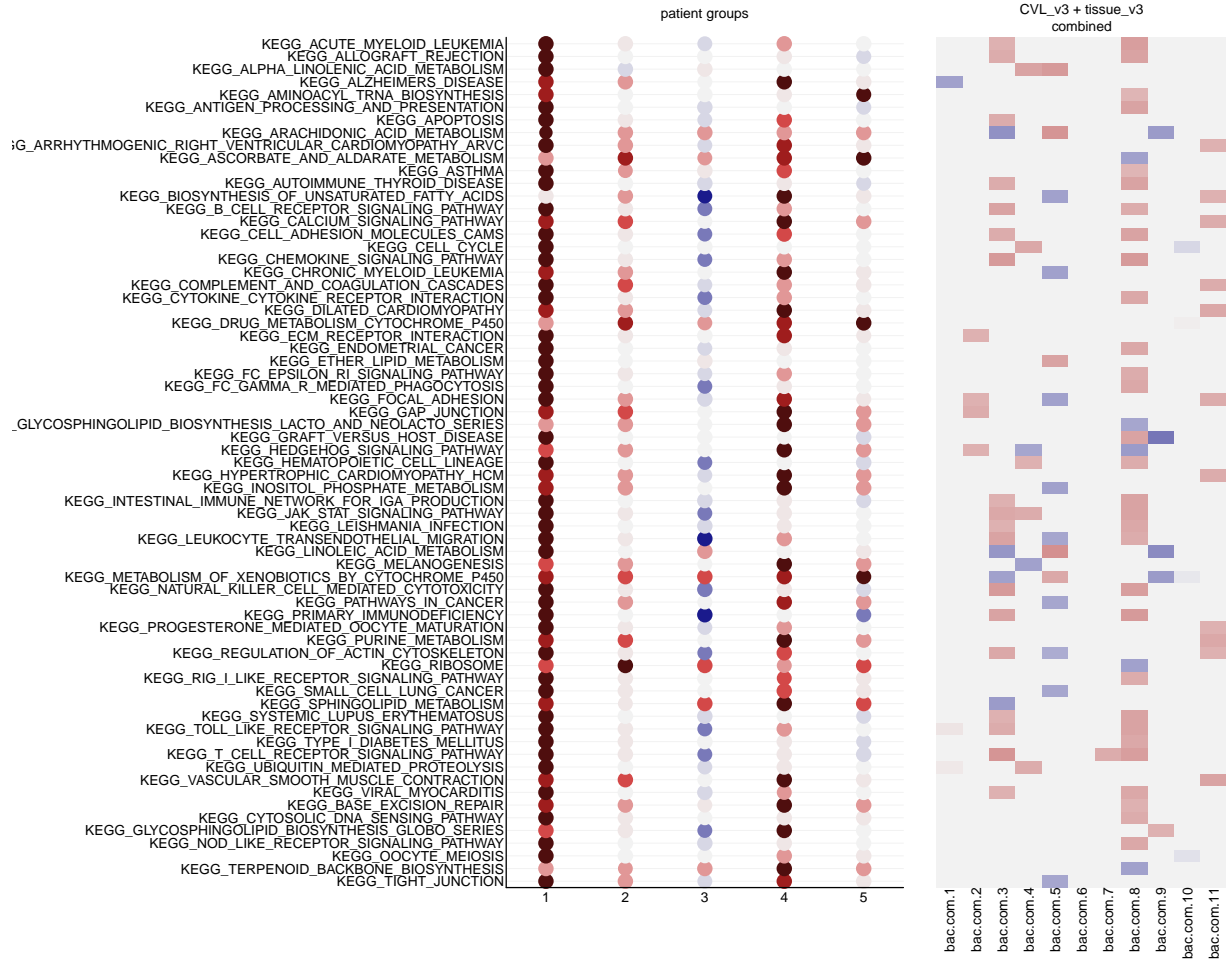bac.com.9
bac.com.10
bac.com.11

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.