**Supplementary Material:**

Table of Contents

[Materials/Patients and Methods: 2](#_Toc65738113)

[Supplementary Table 1: Clinical and biochemical data of study participants 6](#_Toc65738114)

[Supplementary Table 2: Genes that were significantly correlated with F. prausnitzii (Spearman correlation. p<0.001) 7](#_Toc65738115)

[Supplementary Table 3: Genes that were significantly correlated with Ruminococcus spp. (Spearman correlation. p<0.001) 14](#_Toc65738116)

[Supplementary Table 4: Genes that were significantly correlated with Coprococcus spp. (Spearman correlation. p<0.001) 22](#_Toc65738117)

[Supplementary Table 5: Results from the Venn diagram, showing genes that are significantly correlated (Spearman correlation, p<0.001) with F. prausnitzii, Ruminococcus spp. and/or Coprococcus spp. 27](#_Toc65738118)

[Supplementary Table 6: Results of the Gene Ontology analysis [12] for hepatic genes significantly correlated with F. prausnitzii (False Discovery Rate, FDR <0.05) 30](#_Toc65738119)

[Supplementary Table 7: Results of the Gene Ontology analysis [12] for hepatic genes significantly correlated with Ruminococcus spp. (False Discovery Rate, FDR <0.05) 32](#_Toc65738120)

[Supplementary Table 8: Results of the Gene Ontology analysis [12] for hepatic genes significantly correlated with Coprococcus spp. (False Discovery Rate, FDR <0.05 36](#_Toc65738121)

[Supplementary Table 9: Significantly upregulated pathways in NAFLD grouped into KEGG level II and level III. 38](#_Toc65738122)

[Supplementary Table 10: Differentially expressed hepatic genes significantly correlated with bacterial metabolic pathways carbohydrate, lipid, energy, and amino acid metabolism (p<0.05) in NAFLD patients. 44](#_Toc65738123)

[Supplementary Table 11: Differentially expressed hepatic genes significantly correlated with KEGG level II bacterial metabolic pathways carbohydrate, lipid, energy, and amino acid metabolism (p<0.05) in healthy controls. 51](#_Toc65738124)

[References: 53](#_Toc65738125)

#

# **Materials/Patients and Methods:**

*Participants*

Inclusion criteria were: male or female; ≥18; for NAFLD patients, a diagnostic liver biopsy; and for healthy controls (HC), presence of a normal liver (no steatosis or cirrhosis) on histology (1, 2). Exclusion criteria were: alcohol consumption >20g/d; smoking; in case of hyperlipidemia or diabetes, no stable drug regimen; other liver diseases; anticipated need for liver transplantation within one year or complications of liver disease; concurrent medical illnesses; any reasons contraindicating a liver biopsy; chronic gastrointestinal disease; previous gastrointestinal surgery modifying the anatomy; diabetes requiring insulin; use of medications known to precipitate steatohepatitis or regular intake of supplements (antioxidants, vitamins or n-3 fish oil, prebiotics, probiotics); ursodeoxycholic acid or any experimental drug in the 6 months prior to study entry; pregnancy or lactating state. The University Health Network Research Ethics Board approved the study protocol, which was performed in accordance with the guidelines of the 1975 Declaration of Helsinki and its revisions. This trial was registered at www.clinicaltrials.gov as NCT02148471. All participants provided written informed consent.

*Sample and data collection*

Patients with NAFLD provided a liver sample during diagnostic liver biopsy, and HC had a wedge biopsy taken during partial hepatectomy for liver donation. One part of the biopsy was preserved in 10% formalin for histology, and one part for gene expression analysis in RNA*later* (Qiagen, Hilden, Germany). Anthropometric and clinical data were collected as described previously (1, 2). For IM analysis, participants collected one stool sample, which was frozen immediately in their home freezer (−20 °C). The frozen sample was brought to the clinic within the next 24 hours in an insulated bag with cooling elements, and then stored at −80°C. For the processing, stool samples were thawed and homogenized in a sterile masticator bag using a masticator blender (IUL, S.A., Barcelona, Spain). The samples were aliquoted, placed in dry ice, and then stored at −80 °C until DNA extraction.

*Intestinal microbiota composition*

The details of the IM analysis composition were described previously (2). Briefly, bacterial DNA was extracted using the E.Z.N.A.TM Stool DNA Isolation Kit (Omega Bio-Tek, Doraville, GA, USA). Fecal DNA samples were stored at ‑80 C until being amplified by PCR using the Earth Microbiome V4 primer set. The amplified DNA was then sequenced on the Illumina MiSeq platform. The sequencing and pre-processing data for analysis was performed at the Microbiota Profiling Service, Department of Biochemistry, University of Western Ontario, London, Ontario, Canada. Based on 16S sequencing data analysis, the taxa identified as differentially abundant in patients with NAFLD (SS or NASH) compared to HC were further validated using quantitative PCR (qPCR). This showed lower abundance of *Ruminococcus* spp.*, F. prausnitzii* and *Coprococcus* spp.in NASH and SS patientscompared to HC (2), which was independent of insulin resistance and body mass index*.* These taxa are the focus of our secondary analysis.

The PCR Primers used for the analysis were: 5’- GAGTGAAGTAGAGGTAAGCGGAATTC -3’, 3’- GCCGTACTCCCCAGGTGG-5’ for *Ruminococcus* spp; 5’-TGTAAACTCCTGTTGTTGAGGAAGATAA -3’ and 3’-GCGCTCCCTTTACACCCA-5’ for *F. prausnitzii*; and 5’- CATCCTGATGACGGTTTCTTAACC -3’ and 3’- GTTGCGGGACTTAACCCA -5’ for *Coprococcus* spp. (reference 2).

(See supplementary Table 2, (2).)

*Liver histology*

Liver samples were stained with hematoxylin/eosin for morphologic evaluation and Prussian blue stain to rule out iron deposition. Sections of each liver sample were reviewed in a blinded manner by a single pathologist to diagnose SS or NASH (3), and the NAFLD activity score (NAS) was assessed (4).

*Hepatic gene expression*

The Illumina Human HT-12 V4 BeadChip with the Whole Genome Gene DASL HT Assay (Illumina Inc., San Diego, CA, USA) was used to measured hepatic gene expression levels. Briefly, RNA was extracted from hepatic biopsies with *mir*VanaTM miRNA Isolation kit, Life Technologies Corp., Carlsbad, CA, USA), assessing concentration and purity in a Thermo Scientific NanoDrop 1000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) and quality checked by spectrophotometry with an Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA, USA). The data analysis and whole gene expression profiles were described previously (1). The data are publicly available from the Gene Expression Omnibus (GEO) repository (https://www.ncbi.nlm.nih.gov/geo/, accession Nº GSE89632).

*Functional predictions of bacterial communities*

The functional potential of the fecal microbial communities was inferred with Piphillin (5) and the KEGG database (6) (Release 88.0) (<https://piphillin.secondgenome.com/>), using OTU abundance as an input and a 90% sequence identity cut-off. The statistical significance of differentially abundant pathways between NAFLD and healthy (for list of pathways see **Supplementary Table 9-10**) was analyzed using Wald test through DESeq2 package (version 1.28.1) in R (7).

*Blood biochemistry parameters*

Blood samples were obtained from all participants after an 8-hour fast at the Laboratory Medicine Program at the University Health Network. Blood biochemistry included transaminases, fasting plasma glucose, and fasting insulin. Insulin resistance was calculated according to the homeostasis model of assessment for insulin resistance (HOMA-IR) (8). Body mass index was calculated (BMI = weight (kg) / [height (m)]2) (1, 2).

*Statistical Analysis*

 Clinical parameters:

Clinical parameters are presented as median (Q1, Q3), or % of cases. Continuous variables were compared by Mann-Whitney test. Chi square and Fisher’s exact tests were applied for categorical variables.

 Hepatic gene expression and bacterial taxa:

The whole genome analysis published previously (1) identified 822 probes, corresponding to 732 genes that were differentially expressed among HC, SS and NASH and showed at least a 2‑fold up- or down-regulation between at least 2 of these groups. These genes were used for further analyses presented here. For the current report, the selection was modified to restrict the correlation analysis to only one probe per gene, resulting in a list of 728 genes for the analysis. Spearman correlations between these 728 differentially expressed genes in the liver and the normalized counts (relative abundances) of *F. prausnitzii, Ruminococcus* spp. or *Coprococcus spp.* were calculated. Spearman correlations with *p*<0.001 were considered significant. A heatmap was built to illustrate these correlations. The heatmap visualizing correlations between hepatic gene expressions and bacterial taxa in three disease categories (HC, SS or NASH) was plotted using R package Gplots (9). A functional annotation and gene ontology (GO) enrichment analysis was performed for each of the bacteria including genes that showed a significant correlation (*p*<0.001) with this bacterium of interest using the Database for Annotation, Visualization and Integrated Discovery (DAVID) v 6.8 (https://david.ncifcrf.gov) (10, 11).

Inferred functional metagenomic analysis:

The correlation between the three specific taxa and the KEGG pathways for bacterial communities were identified by Spearman correlation and plotted using linear model through R. A similar method was followed to identify the association between the bacterial pathways and hepatic gene expression. A heatmap was generated using pheatmap through R.

# **Supplementary Table 1:** Clinical and biochemical data of study participants

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter** | **n** | **HC** | **N** | **SS** | **n** | **NASH** |
| Age (years) | 6 | 37.3 ± 6.7 | 7 | 44.6 ± 10.0 | 8 | 44.3 ± 10.0 |
| Men (n (%)) | 6 | 3 (50%) | 7 | 5 (71%) | 8 | 5 (62%) |
| BMI (kg/m2) | 6 | 26.8 ± 4.2 A | 7 | 28.2 ± 3.8 | 8 | 33.6 ± 6.7 A |
| Steatosis Percentage | 6 | 1.0 (0 ;1.0) A,B | 7 | 20.0 (10.0; 50.0) A | 8 | 40.0 (17.5; 40.0) B |
| Fibrosis stage (% (n)) 0 1 2 3 4 | 6 | 6 (100%) A0 (0%)0 (0%)0 (0%)0 (0%) | 7 | 6 (85.7%) B1 (14.3%)0 (0%)0 (0%)0 (0%) | 8 | 2 (25%) A, B2 (25%)1 (12.5%)0 (12.5%)2 (25%) |
| NAS (0-8) | 6 | 0 (0; 0)  A, B | 7 | 1.0 (1.0; 2.0) A, C | 8 | 4.0 (4.0; 5.5) B, C |
| AST (U/L) | 6 | 20.7 ± 6.5  A | 7 | 25.7 ± 6.3 B | 8 | 65.3 ± 13.5 A, B |
| ALT (U/L) | 6 | 21.2 ± 12.0  A, B | 7 | 40.0 ± 14.1  A, C | 8 | 98.4 ± 42.7 B, C |
| Glucose (mmol/L) | 6 | 5.1 (4.8; 5.6) | 5 | 5.2 (4.8; 5.8) | 7 | 6.0 (4.3; 7.3) |
| Insulin (pmol/L) | 6 | 20.5 (15.0; 26.0) A | 6 | 27.5 (24.0; 33.0) B | 6 | 149.5 (56.0; 204.0) A, B |
| HOMA-IR | 6 | 0.85 (0.55; 1.04) A | 5 | 1.12 (0.85; 5.59) B | 6 | 7.16 (1.74;11.03) A, B |

Values given are mean ± SD, median (Q1; Q3) or % of valid cases. Values with identical uppercase letters are significantly different (p<0.05)

ALT: alanine transaminase, AST: aspartate transaminase, BMI: body mass index, HOMA-IR: homeostasis model of assessment for insulin resistance, NAS: non-alcoholic fatty liver disease activity score, NASH: non-alcoholic steatohepatitis, NS: not statistically significant, SS: simple steatosis.

# **Supplementary Table 2:** Genes that were significantly correlated with *F. prausnitzii* (Spearman correlation. p<0.001)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Symbol** | **EntrezID** | **Spearman rho** | **p-value** | **Definition (Gene name)** |
| TP53AP1 | 11257 | -0.861 | 5.43E-07 | TP53 target 1 |
| C6orf129 | 154467 | -0.819 | 5.50E-06 | coiled-coil domain containing 167 |
| C14orf80 | 283643 | -0.813 | 7.49E-06 | tubulin epsilon and delta complex 1 |
| PKIB | 5570 | -0.812 | 7.95E-06 | cAMP-dependent protein kinase inhibitor beta |
| HIST2H2AB | 317772 | -0.792 | 1.87E-05 | H2A clustered histone 21 |
| TNFSF10 | 8743 | -0.777 | 3.47E-05 | TNF superfamily member 10 |
| CEBPA | 1050 | -0.774 | 3.84E-05 | CCAAT enhancer binding protein alpha |
| CSRNP3 | 80034 | -0.774 | 3.84E-05 | cysteine and serine rich nuclear protein 3 |
| RTP4 | 64108 | -0.770 | 4.44E-05 | receptor transporter protein 4 |
| PCDH18 | 54510 | -0.764 | 5.62E-05 | protocadherin 18 |
| HIST1H1A | 3024 | -0.762 | 5.89E-05 | H1.1 linker histone, cluster member |
| DBI | 1622 | -0.758 | 6.76E-05 | diazepam binding inhibitor, acyl-CoA binding protein |
| PCDH24 | 54825 | -0.757 | 7.08E-05 | cadherin related family member 2 |
| SLC22A7 | 10864 | -0.755 | 7.74E-05 | solute carrier family 22 member 7 |
| RTP3 | 83597 | -0.755 | 7.74E-05 | receptor transporter protein 3 |
| ANKS4B | 257629 | -0.752 | 8.46E-05 | ankyrin repeat and sterile alpha motif domain containing 4B |
| PRSS3 | 5646 | -0.749 | 9.24E-05 | serine protease 3 |
| ZNHIT3 | 9326 | -0.744 | 1.10E-04 | zinc finger HIT-type containing 3 |
| KIAA1161 | 57462 | -0.744 | 1.10E-04 | myogenesis regulating glycosidase (putative) |
| EDAR | 10913 | -0.742 | 1.20E-04 | ectodysplasin A receptor |
| MGC13057 | 84281 | -0.742 | 1.20E-04 | chromosome 2 open reading frame 88 |
| TMEM169 | 92691 | -0.739 | 1.30E-04 | transmembrane protein 169 |
| MB | 4151 | -0.738 | 1.35E-04 | Myoglobin |
| RORC | 6097 | -0.736 | 1.41E-04 | RAR related orphan receptor C |
| FAM124B | 79843 | -0.736 | 1.41E-04 | family with sequence similarity 124 member B |
| KIF22 | 3835 | -0.734 | 1.53E-04 | kinesin family member 22 |
| AGBL5 | 60509 | -0.734 | 1.53E-04 | ATP/GTP binding protein like 5 |
| C1orf203 | 84852 | -0.732 | 1.60E-04 | ATP1A1 antisense RNA 1 |
| PAQR7 | 164091 | -0.730 | 1.73E-04 | progestin and adipoQ receptor family member 7 |
| GPER | 2852 | -0.726 | 1.95E-04 | G protein-coupled estrogen receptor 1 |
| P2RY2 | 5029 | -0.726 | 1.93E-04 | purinergic receptor P2Y2 |
| CYP7A1 | 1581 | -0.725 | 2.03E-04 | cytochrome P450 family 7 subfamily A member 1 |
| OGG1 | 4968 | -0.723 | 2.11E-04 | 8-oxoguanine DNA glycosylase |
| HEXIM2 | 124790 | -0.718 | 2.46E-04 | HEXIM P-TEFb complex subunit 2 |
| LAMA4 | 3910 | -0.717 | 2.55E-04 | laminin subunit alpha 4 |
| RAB26 | 25837 | -0.713 | 2.86E-04 | RAB26, member RAS oncogene family |
| LRRN3 | 54674 | -0.704 | 3.70E-04 | leucine rich repeat neuronal 3 |
| LOC728037 | 728037 | -0.704 | 3.70E-04 | kinesin-like DNA-binding protein pseudogene |
| MYL6B | 140465 | -0.701 | 3.97E-04 | myosin light chain 6B |
| EPB49 | 2039 | -0.700 | 4.11E-04 | dematin actin binding protein |
| GIMAP2 | 26157 | -0.700 | 4.11E-04 | GTPase, IMAP family member 2 |
| RMND1 | 55005 | -0.700 | 4.11E-04 | required for meiotic nuclear division 1 homolog |
| IDO2 | 169355 | -0.700 | 4.11E-04 | indoleamine 2,3-dioxygenase 2 |
| JMJD5 | 79831 | -0.699 | 4.26E-04 | lysine demethylase 8 |
| RFXDC2 | 64864 | -0.697 | 4.41E-04 | regulatory factor X7 |
| FRAT1 | 10023 | -0.696 | 4.57E-04 | FRAT regulator of WNT signaling pathway 1 |
| C10orf125 | 282969 | -0.696 | 4.57E-04 | fucose mutarotase |
| PMS2 | 5395 | -0.694 | 4.90E-04 | PMS1 homolog 2, mismatch repair system component |
| HIST2H2AC | 8338 | -0.692 | 5.07E-04 | H2A clustered histone 20 |
| HIST1H3C | 8352 | -0.692 | 5.07E-04 | H3 clustered histone 3 |
| C3orf14 | 57415 | -0.692 | 5.07E-04 | chromosome 3 open reading frame 14 |
| RBP5 | 83758 | -0.692 | 5.07E-04 | retinol binding protein 5 |
| NCAPG2 | 54892 | -0.691 | 5.25E-04 | non-SMC condensin II complex subunit G2 |
| PRKCZ | 5590 | -0.690 | 5.43E-04 | protein kinase C zeta |
| EPHA1 | 2041 | -0.688 | 5.62E-04 | EPH receptor A1 |
| ENHO | 375704 | -0.688 | 5.62E-04 | energy homeostasis associated |
| C10orf114 | 399726 | -0.688 | 5.62E-04 | MIR1915 host gene |
| USP18 | 11274 | -0.687 | 5.81E-04 | ubiquitin specific peptidase 18 |
| TMEM98 | 26022 | -0.687 | 5.81E-04 | transmembrane protein 98 |
| MGC12965 | 374408 | -0.686 | 6.01E-04 | similar to cytochrome c, somatic |
| RNF43 | 54894 | -0.682 | 6.64E-04 | ring finger protein 43 |
| LOC730101 | 730101 | -0.682 | 6.64E-04 | uncharacterised LOC730101 |
| NAT8B | 51471 | -0.679 | 7.09E-04 | N-acetyltransferase 8B (putative, gene/pseudogene) |
| TBX10 | 347853 | -0.679 | 7.09E-04 | T-box transcription factor 10 |
| DHRS4L1 | 728635 | -0.679 | 7.09E-04 | dehydrogenase/reductase 4 like 1 |
| ABCC6P2 | 730013 | -0.677 | 7.57E-04 | ATP binding cassette subfamily C member 6 pseudogene 2 |
| P8 | 26471 | -0.675 | 7.82E-04 | nuclear protein 1, transcriptional regulator |
| IFIT2 | 3433 | -0.674 | 8.07E-04 | interferon induced protein with tetratricopeptide repeats 2 |
| DMRTA1 | 63951 | -0.671 | 8.60E-04 | DMRT like family A1 |
| METRN | 79006 | -0.671 | 8.60E-04 | meteorin, glial cell differentiation regulator |
| C9orf125 | 84302 | -0.671 | 8.60E-04 | post-GPI attachment to proteins GalNAc transferase 4 |
| C6orf142 | 90523 | -0.671 | 8.60E-04 | muscular LMNA interacting protein |
| KBTBD3 | 143879 | -0.670 | 8.88E-04 | kelch repeat and BTB domain containing 3 |
| FEZ1 | 9638 | -0.669 | 9.16E-04 | fasciculation and elongation protein zeta 1 |
| ANXA9 | 8416 | -0.668 | 9.46E-04 | annexin A9 |
| C13orf15 | 28984 | 0.823 | 4.54E-06 | chromosome 13 open reading frame 15 |
| LEPR | 3953 | 0.814 | 7.04E-06 | leptin receptor |
| CASP4 | 837 | 0.812 | 7.95E-06 | caspase 4 |
| FLJ36031 | 168455 | 0.809 | 8.96E-06 | coiled-coil domain containing 71 like |
| NAMPT | 10135 | 0.809 | 8.96E-06 | nicotinamide phosphoribosyltransferase |
| NR5A2 | 2494 | 0.787 | 2.31E-05 | nuclear receptor subfamily 5 group A member 2 |
| CALCA | 796 | 0.773 | 4.03E-05 | calcitonin related polypeptide alpha |
| CDC42EP3 | 10602 | 0.770 | 4.44E-05 | CDC42 effector protein 3 |
| BIRC3 | 330 | 0.768 | 4.88E-05 | baculoviral IAP repeat containing 3 |
| NRG1 | 3084 | 0.766 | 5.12E-05 | neuregulin 1 |
| LYST | 1130 | 0.764 | 5.62E-05 | lysosomal trafficking regulator |
| LONRF2 | 164832 | 0.753 | 8.10E-05 | LON peptidase N-terminal domain and ring finger 2 |
| HIVEP1 | 3096 | 0.753 | 8.10E-05 | HIVEP zinc finger 1 |
| ATP2B4 | 493 | 0.748 | 9.65E-05 | ATPase plasma membrane Ca2+ transporting 4 |
| CHI3L1 | 1116 | 0.745 | 1.05E-04 | chitinase 3 like 1 |
| ERN1 | 2081 | 0.738 | 1.35E-04 | endoplasmic reticulum to nucleus signaling 1 |
| IL1RN | 3557 | 0.736 | 1.41E-04 | interleukin 1 receptor antagonist |
| RBPMS | 11030 | 0.735 | 1.47E-04 | RNA binding protein, mRNA processing factor |
| GRAMD4 | 23151 | 0.729 | 1.80E-04 | GRAM domain containing 4 |
| SNORD19 | 692089 | 0.726 | 1.95E-04 | small nucleolar RNA, C/D box 19 |
| MYADM | 91663 | 0.722 | 2.19E-04 | myeloid associated differentiation marker |
| CEBPD | 1052 | 0.719 | 2.37E-04 | CCAAT enhancer binding protein delta |
| MYC | 4609 | 0.716 | 2.65E-04 | MYC proto-oncogene, bHLH transcription factor |
| ZNF878 | 729747 | 0.713 | 2.86E-04 | zinc finger protein 878 |
| FAM115C | 285966 | 0.713 | 2.86E-04 | TRPM8 channel associated factor 2 |
| PTGER2 | 5732 | 0.709 | 3.20E-04 | prostaglandin E receptor 2 |
| AKAP12 | 9590 | 0.704 | 3.70E-04 | A-kinase anchoring protein 12 |
| IL6 | 3569 | 0.704 | 3.70E-04 | interleukin 6 |
| MIR21 | 406991 | 0.701 | 3.97E-04 | microRNA 21 |
| P4HA1 | 5033 | 0.699 | 4.26E-04 | prolyl 4-hydroxylase subunit alpha 1 |
| SOCS2 | 8835 | 0.697 | 4.41E-04 | suppressor of cytokine signaling 2 |
| IRAK3 | 11213 | 0.695 | 4.73E-04 | interleukin 1 receptor associated kinase 3 |
| SLC2A3 | 6515 | 0.695 | 4.73E-04 | solute carrier family 2 member 3 |
| PELI1 | 57162 | 0.691 | 5.25E-04 | pellino E3 ubiquitin protein ligase 1 |
| CDH23 | 64072 | 0.690 | 5.43E-04 | cadherin related 23 |
| ARID5B | 84159 | 0.688 | 5.62E-04 | AT-rich interaction domain 5B |
| IRS2 | 8660 | 0.687 | 5.81E-04 | insulin receptor substrate 2 |
| SDC4 | 6385 | 0.686 | 6.01E-04 | syndecan 4 |
| RALGDS | 5900 | 0.686 | 6.01E-04 | ral guanine nucleotide dissociation stimulator |
| FOSL2 | 2355 | 0.686 | 6.01E-04 | FOS like 2, AP-1 transcription factor subunit |
| STAT3 | 6774 | 0.684 | 6.21E-04 | signal transducer and activator of transcription 3 |
| THBD | 7056 | 0.683 | 6.42E-04 | Thrombomodulin |
| SIPA1L2 | 57568 | 0.682 | 6.64E-04 | signal induced proliferation associated 1 like 2 |
| VNN3 | 55350 | 0.679 | 7.09E-04 | vanin 3 |
| PMP22 | 5376 | 0.679 | 7.09E-04 | peripheral myelin protein 22 |
| VNN2 | 8875 | 0.678 | 7.33E-04 | vanin 2 |
| SOCS3 | 9021 | 0.677 | 7.57E-04 | suppressor of cytokine signaling 3 |
| PADI4 | 23569 | 0.675 | 7.82E-04 | peptidyl arginine deiminase 4 |
| FAM107A | 11170 | 0.675 | 7.82E-04 | family with sequence similarity 107 member A |
| CISH | 1154 | 0.675 | 7.82E-04 | cytokine inducible SH2 containing protein |
| RRS1 | 23212 | 0.674 | 8.07E-04 | ribosome biogenesis regulator 1 homolog |
| BCL6 | 604 | 0.674 | 8.07E-04 | BCL6 transcription repressor |
| PFKFB3 | 5209 | 0.673 | 8.33E-04 | 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3 |
| FOSB | 2354 | 0.673 | 8.33E-04 | FosB proto-oncogene, AP-1 transcription factor subunit |
| NCRNA00152 | 112597 | 0.671 | 8.60E-04 | cytoskeleton regulator RNA |
| KIAA0853 | 23091 | 0.671 | 8.60E-04 | zinc finger CCCH-type containing 13 |
| FAM169B | 283777 | 0.670 | 8.88E-04 | family with sequence similarity 169 member B |
| IVNS1ABP | 10625 | 0.670 | 8.88E-04 | influenza virus NS1A binding protein |
| OSMR | 9180 | 0.670 | 8.88E-04 | oncostatin M receptor |
| MMP19 | 4327 | 0.669 | 9.16E-04 | matrix metallopeptidase 19 |
| KCNE4 | 23704 | 0.668 | 9.46E-04 | potassium voltage-gated channel subfamily E regulatory subunit 4 |
| HAS2 | 3037 | 0.668 | 9.46E-04 | hyaluronan synthase 2 |
| GPR3 | 2827 | 0.668 | 9.46E-04 | G protein-coupled receptor 3 |

# **Supplementary Table 3:** Genes that were significantly correlated with *Ruminococcus* spp. (Spearman correlation. p<0.001)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **GeneName** | **EntrezID** | **Spearman rho** | **p-value** | **Definition (Gene name)** |
| FRMD7 | 90167 | -0.830 | 3.27E-06 | FERM domain containing 7 |
| CSTA | 1475 | -0.807 | 9.67E-06 | cystatin A |
| MAMDC4 | 158056 | -0.806 | 1.01E-05 | MAM domain containing 4 |
| MARVELD3 | 91862 | -0.805 | 1.07E-05 | MARVEL domain containing 3 |
| GLT8D1 | 55830 | -0.799 | 1.42E-05 | glycosyltransferase 8 domain containing 1 |
| C9orf91 | 203197 | -0.795 | 1.67E-05 | transmembrane protein 268 |
| RALGPS1 | 9649 | -0.794 | 1.77E-05 | Ral GEF with PH domain and SH3 binding motif 1 |
| PEG10 | 23089 | -0.792 | 1.87E-05 | paternally expressed 10 |
| TMEM63A | 9725 | -0.782 | 2.84E-05 | transmembrane protein 63A |
| ONECUT1 | 3175 | -0.779 | 3.14E-05 | one cut homeobox 1 |
| SNHG7 | 84973 | -0.770 | 4.44E-05 | small nucleolar RNA host gene 7 |
| SKP2 | 6502 | -0.768 | 4.88E-05 | S-phase kinase associated protein 2 |
| IFIT2 | 3433 | -0.766 | 5.12E-05 | interferon induced protein with tetratricopeptide repeats 2 |
| CASC5 | 57082 | -0.762 | 5.89E-05 | kinetochore scaffold 1 |
| RFXDC2 | 64864 | -0.756 | 7.40E-05 | regulatory factor X7 |
| HOPX | 84525 | -0.756 | 7.40E-05 | HOP homeobox |
| WNT5A | 7474 | -0.753 | 8.10E-05 | Wnt family member 5A |
| TMEM169 | 92691 | -0.753 | 8.10E-05 | transmembrane protein 169 |
| PCBP4 | 57060 | -0.749 | 9.24E-05 | poly(rC) binding protein 4 |
| ABCG8 | 64241 | -0.748 | 9.65E-05 | ATP binding cassette subfamily G member 8 |
| FANCL | 55120 | -0.745 | 1.05E-04 | FA complementation group L |
| OAS1 | 4938 | -0.738 | 1.35E-04 | 2'-5'-oligoadenylate synthetase 1 |
| PITPNM2 | 57605 | -0.735 | 1.47E-04 | phosphatidylinositol transfer protein membrane associated 2 |
| C18orf56 | 494514 | -0.735 | 1.47E-04 | TYMS opposite strand |
| CD3G | 917 | -0.732 | 1.60E-04 | CD3g molecule |
| JUB | 84962 | -0.732 | 1.60E-04 | ajuba LIM protein |
| APOL3 | 80833 | -0.730 | 1.73E-04 | apolipoprotein L3 |
| BTN3A1 | 11119 | -0.727 | 1.87E-04 | butyrophilin subfamily 3 member A1 |
| C3orf14 | 57415 | -0.725 | 2.03E-04 | chromosome 3 open reading frame 14 |
| TMEM45B | 120224 | -0.723 | 2.11E-04 | transmembrane protein 45B |
| CX3CR1 | 1524 | -0.722 | 2.19E-04 | C-X3-C motif chemokine receptor 1 |
| LRRC20 | 55222 | -0.717 | 2.55E-04 | leucine rich repeat containing 20 |
| CSRNP3 | 80034 | -0.717 | 2.55E-04 | cysteine and serine rich nuclear protein 3 |
| GNAO1 | 2775 | -0.714 | 2.75E-04 | G protein subunit alpha o1 |
| ABCC5 | 10057 | -0.714 | 2.75E-04 | ATP binding cassette subfamily C member 5 |
| C8orf47 | 203111 | -0.714 | 2.75E-04 | glutamate rich 5 |
| PALMD | 54873 | -0.714 | 2.79E-04 | palmdelphin |
| BCAR3 | 8412 | -0.713 | 2.86E-04 | BCAR3 adaptor protein, NSP family member |
| AKR1B10 | 57016 | -0.713 | 2.86E-04 | aldo-keto reductase family 1 member B10 |
| IFIT3 | 3437 | -0.710 | 3.08E-04 | interferon induced protein with tetratricopeptide repeats 3 |
| NINL | 22981 | -0.705 | 3.56E-04 | ninein like |
| MOGAT1 | 116255 | -0.705 | 3.56E-04 | monoacylglycerol O-acyltransferase 1 |
| HIST1H2AC | 8334 | -0.703 | 3.83E-04 | H2A clustered histone 6 |
| ACSM1 | 116285 | -0.703 | 3.83E-04 | acyl-CoA synthetase medium chain family member 1 |
| TRIM14 | 9830 | -0.697 | 4.41E-04 | tripartite motif containing 14 |
| ACE2 | 59272 | -0.696 | 4.57E-04 | angiotensin converting enzyme 2 |
| MIR564 | 693149 | -0.696 | 4.57E-04 | microRNA 564 |
| KANK4 | 163782 | -0.695 | 4.73E-04 | KN motif and ankyrin repeat domains 4 |
| CKAP2 | 26586 | -0.694 | 4.90E-04 | cytoskeleton associated protein 2 |
| ANKS4B | 257629 | -0.694 | 4.90E-04 | ankyrin repeat and sterile alpha motif domain containing 4B |
| SLC12A1 | 6557 | -0.692 | 5.07E-04 | solute carrier family 12 member 1 |
| CCDC150 | 284992 | -0.688 | 5.62E-04 | coiled-coil domain containing 150 |
| FMO1 | 2326 | -0.687 | 5.81E-04 | flavin containing dimethylaniline monoxygenase 1 |
| TYMS | 7298 | -0.687 | 5.81E-04 | thymidylate synthetase |
| ME1 | 4199 | -0.681 | 6.86E-04 | malic enzyme 1 |
| SLC25A34 | 284723 | -0.679 | 7.09E-04 | solute carrier family 25 member 34 |
| GRAP | 10750 | -0.678 | 7.33E-04 | GRB2 related adaptor protein |
| RNF43 | 54894 | -0.677 | 7.57E-04 | ring finger protein 43 |
| AFP | 174 | -0.675 | 7.82E-04 | alpha fetoprotein |
| CYP7A1 | 1581 | -0.675 | 7.82E-04 | cytochrome P450 family 7 subfamily A member 1 |
| ZWINT | 11130 | -0.675 | 7.82E-04 | ZW10 interacting kinetochore protein |
| SLC22A11 | 55867 | -0.675 | 7.82E-04 | solute carrier family 22 member 11 |
| CCDC121 | 79635 | -0.675 | 7.82E-04 | coiled-coil domain containing 121 |
| THAP2 | 83591 | -0.675 | 7.82E-04 | THAP domain containing 2 |
| PAQR4 | 124222 | -0.675 | 7.82E-04 | progestin and adipoQ receptor family member 4 |
| FRAT1 | 10023 | -0.671 | 8.60E-04 | FRAT regulator of WNT signaling pathway 1 |
| OAS2 | 4939 | -0.670 | 8.88E-04 | 2'-5'-oligoadenylate synthetase 2 |
| ZNF385B | 151126 | -0.670 | 8.88E-04 | zinc finger protein 385B |
| TMEM20 | 159371 | -0.668 | 9.46E-04 | solute carrier family 35 member G1 |
| C12orf27 | 283460 | -0.668 | 9.46E-04 | HNF1A antisense RNA 1 |
| FZD7 | 8324 | 0.814 | 7.04E-06 | frizzled class receptor 7 |
| RGS2 | 5997 | 0.814 | 7.04E-06 | regulator of G protein signaling 2 |
| AGPAT9 | 84803 | 0.810 | 8.44E-06 | glycerol-3-phosphate acyltransferase 3 |
| KLF11 | 8462 | 0.803 | 1.20E-05 | Kruppel like factor 11 |
| C2CD4B | 388125 | 0.795 | 1.67E-05 | C2 calcium dependent domain containing 4B |
| TUBB6 | 84617 | 0.795 | 1.67E-05 | tubulin beta 6 class V |
| PIM3 | 415116 | 0.782 | 2.84E-05 | Pim-3 proto-oncogene, serine/threonine kinase |
| PROK2 | 60675 | 0.782 | 2.84E-05 | prokineticin 2 |
| SERPINE1 | 5054 | 0.782 | 2.84E-05 | serpin family E member 1 |
| EPHA2 | 1969 | 0.782 | 2.84E-05 | EPH receptor A2 |
| MAP3K14 | 9020 | 0.779 | 3.14E-05 | mitogen-activated protein kinase kinase kinase 14 |
| MAK | 4117 | 0.779 | 3.14E-05 | male germ cell associated kinase |
| ROM1 | 6094 | 0.777 | 3.47E-05 | retinal outer segment membrane protein 1 |
| BCL3 | 602 | 0.775 | 3.65E-05 | BCL3 transcription coactivator |
| PPAN | 56342 | 0.770 | 4.40E-05 | peter pan homolog |
| RRP12 | 23223 | 0.768 | 4.88E-05 | ribosomal RNA processing 12 homolog |
| FOSL1 | 8061 | 0.768 | 4.88E-05 | FOS like 1, AP-1 transcription factor subunit |
| SLC11A1 | 6556 | 0.766 | 5.12E-05 | solute carrier family 11 member 1 |
| ELL | 8178 | 0.761 | 6.17E-05 | elongation factor for RNA polymerase II |
| NR4A1 | 3164 | 0.761 | 6.17E-05 | nuclear receptor subfamily 4 group A member 1 |
| LDLR | 3949 | 0.760 | 6.46E-05 | low density lipoprotein receptor |
| IER3 | 8870 | 0.758 | 6.76E-05 | immediate early response 3 |
| UPP1 | 7378 | 0.758 | 6.76E-05 | uridine phosphorylase 1 |
| C10orf10 | 11067 | 0.753 | 8.10E-05 | DEPP1 autophagy regulator |
| MT1A | 4489 | 0.753 | 8.10E-05 | metallothionein 1A |
| FOXO1 | 2308 | 0.752 | 8.46E-05 | forkhead box O1 |
| ARL14 | 80117 | 0.751 | 8.84E-05 | ADP ribosylation factor like GTPase 14 |
| CDH19 | 28513 | 0.751 | 8.84E-05 | cadherin 19 |
| SLC22A4 | 6583 | 0.751 | 8.84E-05 | solute carrier family 22 member 4 |
| SERTAD1 | 29950 | 0.744 | 1.10E-04 | SERTA domain containing 1 |
| S100A12 | 6283 | 0.744 | 1.10E-04 | S100 calcium binding protein A12 |
| IER5L | 389792 | 0.743 | 1.15E-04 | immediate early response 5 like |
| DHRS13 | 147015 | 0.742 | 1.20E-04 | dehydrogenase/reductase 13 |
| TNFRSF10D | 8793 | 0.742 | 1.20E-04 | TNF receptor superfamily member 10d |
| TNFRSF1A | 7132 | 0.739 | 1.30E-04 | TNF receptor superfamily member 1A |
| C17orf96 | 100170841 | 0.735 | 1.47E-04 | elongin BC and polycomb repressive complex 2 associated protein |
| ETS2 | 2114 | 0.734 | 1.53E-04 | ETS proto-oncogene 2, transcription factor |
| HBEGF | 1839 | 0.734 | 1.53E-04 | heparin binding EGF like growth factor |
| MSX1 | 4487 | 0.732 | 1.60E-04 | msh homeobox 1 |
| RCAN1 | 1827 | 0.731 | 1.66E-04 | regulator of calcineurin 1 |
| OSM | 5008 | 0.730 | 1.73E-04 | oncostatin M |
| MYADM | 91663 | 0.727 | 1.87E-04 | myeloid associated differentiation marker |
| LIF | 3976 | 0.727 | 1.87E-04 | LIF interleukin 6 family cytokine |
| NCOA7 | 135112 | 0.726 | 1.95E-04 | nuclear receptor coactivator 7 |
| S100P | 6286 | 0.725 | 2.03E-04 | S100 calcium binding protein P |
| CXCR7 | 57007 | 0.722 | 2.19E-04 | atypical chemokine receptor 3 |
| ODZ3 | 55714 | 0.722 | 2.19E-04 | teneurin transmembrane protein 3 |
| SOCS1 | 8651 | 0.722 | 2.19E-04 | suppressor of cytokine signaling 1 |
| S100A8 | 6279 | 0.722 | 2.19E-04 | S100 calcium binding protein A8 |
| SEMA4B | 10509 | 0.721 | 2.28E-04 | semaphorin 4B |
| RGS1 | 5996 | 0.721 | 2.28E-04 | regulator of G protein signaling 1 |
| ELF3 | 1999 | 0.719 | 2.37E-04 | E74 like ETS transcription factor 3 |
| MYBPH | 4608 | 0.718 | 2.46E-04 | myosin binding protein H |
| MMP19 | 4327 | 0.717 | 2.55E-04 | matrix metallopeptidase 19 |
| IGFBP2 | 3485 | 0.717 | 2.55E-04 | insulin like growth factor binding protein 2 |
| GADD45G | 10912 | 0.716 | 2.65E-04 | growth arrest and DNA damage inducible gamma |
| KIAA0040 | 9674 | 0.716 | 2.65E-04 | KIAA0040 |
| CSF3 | 1440 | 0.716 | 2.65E-04 | colony stimulating factor 3 |
| CLCF1 | 23529 | 0.713 | 2.86E-04 | cardiotrophin like cytokine factor 1 |
| OXCT2 | 64064 | 0.710 | 3.12E-04 | 3-oxoacid CoA-transferase 2 |
| C21orf63 | 59271 | 0.709 | 3.20E-04 | eva-1 homolog C |
| FOSB | 2354 | 0.708 | 3.31E-04 | FosB proto-oncogene, AP-1 transcription factor subunit |
| CNN1 | 1264 | 0.708 | 3.31E-04 | calponin 1 |
| C19orf23 | 148046 | 0.706 | 3.44E-04 | CIRBP antisense RNA 1 |
| AXUD1 | 64651 | 0.703 | 3.83E-04 | cysteine and serine rich nuclear protein 1 |
| SLITRK3 | 22865 | 0.703 | 3.83E-04 | SLIT and NTRK like family member 3 |
| DIO3OS | 64150 | 0.699 | 4.26E-04 | DIO3 opposite strand upstream RNA |
| RND1 | 27289 | 0.699 | 4.26E-04 | Rho family GTPase 1 |
| GFPT2 | 9945 | 0.699 | 4.26E-04 | glutamine-fructose-6-phosphate transaminase 2 |
| EMP1 | 2012 | 0.699 | 4.26E-04 | epithelial membrane protein 1 |
| EIF5 | 1983 | 0.699 | 4.26E-04 | eukaryotic translation initiation factor 5 |
| ACTG2 | 72 | 0.699 | 4.26E-04 | actin gamma 2, smooth muscle |
| WT1 | 7490 | 0.697 | 4.41E-04 | WT1 transcription factor |
| PPRC1 | 23082 | 0.696 | 4.57E-04 | PPARG related coactivator 1 |
| NRG1 | 3084 | 0.696 | 4.57E-04 | neuregulin 1 |
| MAFF | 23764 | 0.695 | 4.73E-04 | MAF bZIP transcription factor F |
| LARP1B | 55132 | 0.694 | 4.90E-04 | La ribonucleoprotein 1B |
| PADI4 | 23569 | 0.694 | 4.90E-04 | peptidyl arginine deiminase 4 |
| NFIL3 | 4783 | 0.694 | 4.90E-04 | nuclear factor, interleukin 3 regulated |
| SBNO2 | 22904 | 0.691 | 5.25E-04 | strawberry notch homolog 2 |
| PLAUR | 5329 | 0.687 | 5.81E-04 | plasminogen activator, urokinase receptor |
| ZNF331 | 55422 | 0.684 | 6.21E-04 | zinc finger protein 331 |
| FJX1 | 24147 | 0.684 | 6.21E-04 | four-jointed box kinase 1 |
| MYH11 | 4629 | 0.684 | 6.21E-04 | myosin heavy chain 11 |
| PNRC1 | 10957 | 0.683 | 6.42E-04 | proline rich nuclear receptor coactivator 1 |
| GPRC5A | 9052 | 0.683 | 6.42E-04 | G protein-coupled receptor class C group 5 member A |
| LPAR1 | 1902 | 0.683 | 6.42E-04 | lysophosphatidic acid receptor 1 |
| APOLD1 | 81575 | 0.682 | 6.64E-04 | apolipoprotein L domain containing 1 |
| SIPA1L2 | 57568 | 0.682 | 6.64E-04 | signal induced proliferation associated 1 like 2 |
| JUNB | 3726 | 0.682 | 6.64E-04 | JunB proto-oncogene, AP-1 transcription factor subunit |
| DAND5 | 199699 | 0.681 | 6.86E-04 | DAN domain BMP antagonist family member 5 |
| C13orf15 | 28984 | 0.681 | 6.86E-04 | regulator of cell cycle |
| ADAMTS1 | 9510 | 0.681 | 6.86E-04 | ADAM metallopeptidase with thrombospondin type 1 motif 1 |
| CCL2 | 6347 | 0.679 | 7.09E-04 | C-C motif chemokine ligand 2 |
| GRHL1 | 29841 | 0.678 | 7.33E-04 | grainyhead like transcription factor 1 |
| TMPRSS2 | 7113 | 0.677 | 7.57E-04 | transmembrane serine protease 2 |
| RALGDS | 5900 | 0.675 | 7.82E-04 | ral guanine nucleotide dissociation stimulator |
| MBNL2 | 10150 | 0.674 | 8.07E-04 | muscleblind like splicing regulator 2 |
| SOCS2 | 8835 | 0.674 | 8.07E-04 | suppressor of cytokine signaling 2 |
| FGF14 | 2259 | 0.673 | 8.33E-04 | fibroblast growth factor 14 |
| JUN | 3725 | 0.671 | 8.60E-04 | Jun proto-oncogene, AP-1 transcription factor subunit |
| SERPINB9 | 5272 | 0.670 | 8.88E-04 | serpin family B member 9 |
| DHODH | 1723 | 0.670 | 8.88E-04 | dihydroorotate dehydrogenase (quinone) |
| CREB5 | 9586 | 0.669 | 9.16E-04 | cAMP responsive element binding protein 5 |
| KCNK1 | 3775 | 0.669 | 9.16E-04 | potassium two pore domain channel subfamily K member 1 |
| FRMD6 | 122786 | 0.668 | 9.46E-04 | FERM domain containing 6 |

# **Supplementary Table 4:** Genes that were significantly correlated with *Coprococcus* spp.(Spearman correlation. p<0.001)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **GeneName** | **EntrezID** | **Spearman rho** | **p-value** | **Definition (Gene name)** |
| IFIT2 | 3433 | -0.806 | 1.03E-05 | interferon induced protein with tetratricopeptide repeats 2 |
| CX3CR1 | 1524 | -0.781 | 2.96E-05 | C-X3-C motif chemokine receptor 1 |
| CD3G | 917 | -0.772 | 4.09E-05 | CD3g molecule |
| MARVELD3 | 91862 | -0.762 | 5.98E-05 | MARVEL domain containing 3 |
| CENPQ | 55166 | -0.756 | 7.34E-05 | centromere protein Q |
| RTP4 | 64108 | -0.746 | 1.02E-04 | receptor transporter protein 4 |
| SLC12A1 | 6557 | -0.739 | 1.32E-04 | solute carrier family 12 member 1 |
| OSGIN1 | 29948 | -0.731 | 1.68E-04 | oxidative stress induced growth inhibitor 1 |
| HIST1H2AC | 8334 | -0.724 | 2.09E-04 | H2A clustered histone 6 |
| AMACR | 23600 | -0.722 | 2.22E-04 | alpha-methylacyl-CoA racemase |
| FANCL | 55120 | -0.720 | 2.31E-04 | FA complementation group L |
| C18orf56 | 494514 | -0.720 | 2.35E-04 | TYMS opposite strand |
| FBXO27 | 126433 | -0.718 | 2.49E-04 | F-box protein 27 |
| HOPX | 84525 | -0.715 | 2.68E-04 | HOP homeobox |
| MIR564 | 693149 | -0.713 | 2.84E-04 | microRNA 564 |
| PEG10 | 23089 | -0.712 | 2.95E-04 | paternally expressed 10 |
| NCAPG2 | 54892 | -0.709 | 3.23E-04 | non-SMC condensin II complex subunit G2 |
| FRMD7 | 90167 | -0.707 | 3.42E-04 | FERM domain containing 7 |
| SLC25A42 | 284439 | -0.703 | 3.81E-04 | solute carrier family 25 member 42 |
| CCDC121 | 79635 | -0.702 | 3.87E-04 | coiled-coil domain containing 121 |
| EDAR | 10913 | -0.694 | 4.79E-04 | ectodysplasin A receptor |
| PCDH24 | 54825 | -0.690 | 5.31E-04 | cadherin related family member 2 |
| PMS2 | 5395 | -0.689 | 5.49E-04 | PMS1 homolog 2, mismatch repair system component |
| SPTLC3 | 55304 | -0.688 | 5.68E-04 | serine palmitoyltransferase long chain base subunit 3 |
| APOL3 | 80833 | -0.688 | 5.68E-04 | apolipoprotein L3 |
| LGALS4 | 3960 | -0.684 | 6.28E-04 | galectin 4 |
| TMEM63A | 9725 | -0.684 | 6.28E-04 | transmembrane protein 63A |
| TRHDE | 29953 | -0.682 | 6.60E-04 | thyrotropin releasing hormone degrading enzyme |
| TMEM20 | 159371 | -0.681 | 6.71E-04 | solute carrier family 35 member G1 |
| SLC25A34 | 284723 | -0.681 | 6.82E-04 | solute carrier family 25 member 34 |
| ZWINT | 11130 | -0.676 | 7.65E-04 | ZW10 interacting kinetochore protein |
| TNFSF10 | 8743 | -0.675 | 7.90E-04 | TNF superfamily member 10 |
| KDELR3 | 11015 | -0.675 | 7.90E-04 | KDEL endoplasmic reticulum protein retention receptor 3 |
| SLC22A11 | 55867 | -0.675 | 7.90E-04 | solute carrier family 22 member 11 |
| MAL2 | 114569 | -0.675 | 7.90E-04 | mal, T cell differentiation protein 2 |
| RALGPS1 | 9649 | -0.674 | 8.03E-04 | Ral GEF with PH domain and SH3 binding motif 1 |
| ARID3C | 138715 | -0.673 | 8.29E-04 | AT-rich interaction domain 3C |
| TYMS | 7298 | -0.672 | 8.56E-04 | thymidylate synthetase |
| FADS2 | 9415 | -0.671 | 8.69E-04 | fatty acid desaturase 2 |
| RNF43 | 54894 | -0.671 | 8.69E-04 | ring finger protein 43 |
| SLC22A7 | 10864 | -0.668 | 9.41E-04 | solute carrier family 22 member 7 |
| ZFP36 | 7538 | 0.805 | 1.06E-05 | ZFP36 ring finger protein |
| LARP1B | 55132 | 0.787 | 2.35E-05 | La ribonucleoprotein 1B |
| EIF5 | 1983 | 0.787 | 2.29E-05 | eukaryotic translation initiation factor 5 |
| S100A8 | 6279 | 0.785 | 2.54E-05 | S100 calcium binding protein A8 |
| S100A12 | 6283 | 0.756 | 7.34E-05 | S100 calcium binding protein A12 |
| EPHA2 | 1969 | 0.753 | 8.03E-05 | EPH receptor A2 |
| ODZ3 | 55714 | 0.752 | 8.58E-05 | teneurin transmembrane protein 3 |
| PHLDA1 | 22822 | 0.751 | 8.77E-05 | pleckstrin homology like domain family A member 1 |
| DHODH | 1723 | 0.750 | 9.16E-05 | dihydroorotate dehydrogenase (quinone) |
| LDLR | 3949 | 0.744 | 1.11E-04 | low density lipoprotein receptor |
| AGPAT9 | 84803 | 0.742 | 1.16E-04 | glycerol-3-phosphate acyltransferase 3 |
| IL10 | 3586 | 0.738 | 1.33E-04 | interleukin 10 |
| SNORA67 | 26781 | 0.735 | 1.46E-04 | small nucleolar RNA, H/ACA box 67 |
| RGS2 | 5997 | 0.735 | 1.46E-04 | regulator of G protein signaling 2 |
| SERPINB9 | 5272 | 0.734 | 1.52E-04 | serpin family B member 9 |
| PPRC1 | 23082 | 0.730 | 1.72E-04 | PPARG related coactivator 1 |
| C19orf23 | 148046 | 0.728 | 1.86E-04 | CIRBP antisense RNA 1 |
| RBPMS | 11030 | 0.722 | 2.18E-04 | RNA binding protein, mRNA processing factor |
| TIPARP | 25976 | 0.720 | 2.31E-04 | TCDD inducible poly(ADP-ribose) polymerase |
| IL4R | 3566 | 0.719 | 2.40E-04 | interleukin 4 receptor |
| MIDN | 90007 | 0.717 | 2.54E-04 | Midnolin |
| C21orf63 | 59271 | 0.717 | 2.54E-04 | eva-1 homolog C |
| MLF1 | 4291 | 0.716 | 2.59E-04 | myeloid leukemia factor 1 |
| TNFSF14 | 8740 | 0.715 | 2.68E-04 | TNF superfamily member 14 |
| MAP3K14 | 9020 | 0.711 | 3.06E-04 | mitogen-activated protein kinase kinase kinase 14 |
| DHRS13 | 147015 | 0.706 | 3.48E-04 | dehydrogenase/reductase 13 |
| SNORD19 | 692089 | 0.705 | 3.54E-04 | small nucleolar RNA, C/D box 19 |
| CCRN4L | 25819 | 0.704 | 3.67E-04 | Nocturnin |
| NR4A1 | 3164 | 0.704 | 3.67E-04 | nuclear receptor subfamily 4 group A member 1 |
| SIPA1L2 | 57568 | 0.703 | 3.81E-04 | signal induced proliferation associated 1 like 2 |
| SOCS2 | 8835 | 0.698 | 4.39E-04 | suppressor of cytokine signaling 2 |
| BACH2 | 60468 | 0.698 | 4.31E-04 | BTB domain and CNC homolog 2 |
| ADAMTS1 | 9510 | 0.696 | 4.54E-04 | ADAM metallopeptidase with thrombospondin type 1 motif 1 |
| FOSB | 2354 | 0.696 | 4.54E-04 | FosB proto-oncogene, AP-1 transcription factor subunit |
| MYADM | 91663 | 0.695 | 4.70E-04 | myeloid associated differentiation marker |
| FAM107A | 11170 | 0.695 | 4.70E-04 | family with sequence similarity 107 member A |
| CMTM2 | 146225 | 0.694 | 4.87E-04 | CKLF like MARVEL transmembrane domain containing 2 |
| TGM2 | 7052 | 0.694 | 4.87E-04 | transglutaminase 2 |
| UPP1 | 7378 | 0.692 | 5.13E-04 | uridine phosphorylase 1 |
| SBNO2 | 22904 | 0.691 | 5.22E-04 | strawberry notch homolog 2 |
| KLF5 | 688 | 0.689 | 5.49E-04 | Kruppel like factor 5 |
| C2CD4B | 388125 | 0.687 | 5.87E-04 | C2 calcium dependent domain containing 4B |
| GRHL1 | 29841 | 0.685 | 6.18E-04 | grainyhead like transcription factor 1 |
| MMP25 | 64386 | 0.677 | 7.53E-04 | matrix metallopeptidase 25 |
| CXCR7 | 57007 | 0.677 | 7.40E-04 | atypical chemokine receptor 3 |
| AXUD1 | 64651 | 0.675 | 7.90E-04 | cysteine and serine rich nuclear protein 1 |
| LOC100131726 | 100131726 | 0.674 | 8.03E-04 |  |
| ARID5B | 84159 | 0.673 | 8.29E-04 | AT-rich interaction domain 5B |
| SLC11A1 | 6556 | 0.671 | 8.69E-04 | solute carrier family 11 member 1 |
| LYST | 1130 | 0.671 | 8.69E-04 | lysosomal trafficking regulator |
| IGFBP2 | 3485 | 0.668 | 9.26E-04 | insulin like growth factor binding protein 2 |

# **Supplementary Table 5:** Results from the Venn diagram, showing genes that are significantly correlated (Spearman correlation, p<0.001) with *F. prausnitzii, Ruminococcus* spp. and/or *Coprococcus* spp.

|  |
| --- |
| **108 genes correlated only with *F. prausnitzii*** |
| **Positively** |
| AKAP12, ATP2B4, BCL6, BIRC3, CALCA, CASP4, CDC42EP3, CDH23, CEBPD, CHI3L1, CISH, ERN1, FAM115C, FAM169B, FLJ36031, FOSL2, GPR3, GRAMD4, HAS2, HIVEP1, IL1RN, IL6, IRAK3, IRS2, IVNS1ABP, KCNE4, KIAA0853, LEPR, LONRF2, MIR21, MYC, NAMPT, NCRNA00152, NR5A2, OSMR, P4HA1, PELI1, PFKFB3, PMP22, PTGER2, RRS1, SDC4, SLC2A3, SOCS3, STAT3, THBD, VNN2, VNN3, ZNF878 |
| **Negatively** |
| ABCC6P2, AGBL5, ANXA9, C10orf114, C10orf125, C14orf80, C1orf203, C6orf129, C6orf142, C9orf125, CEBPA, DBI, DHRS4L1, DMRTA1, ENHO, EPB49, EPHA1, FAM124B, FEZ1, GIMAP2, GPER, HEXIM2, HIST1H1A, HIST1H3C, HIST2H2AB, HIST2H2AC, IDO2, JMJD5, KBTBD3, KIAA1161, KIF22, LAMA4, LOC728037, LOC730101, LRRN3, MB, METRN, MGC12965, MGC13057, MYL6B, NAT8B, OGG1, P2RY2, P8, PAQR7, PCDH18, PKIB, PRKCZ, PRSS3, RAB26, RBP5, RMND1, RORC, RTP3, TBX10, TMEM98, TP53AP1, USP18, ZNHIT3 |
| **112 genes correlated only with *Ruminococcus* spp.** |
| **Positively** |
| ACTG2, APOLD1, ARL14, BCL3, C10orf10, C17orf96, CCL2, CDH19, CLCF1, CNN1, CREB5, CSF3, DAND5, DIO3OS, ELF3, ELL, EMP1, ETS2, FGF14, FJX1, FOSL1, FOXO1, FRMD6, FZD7, GADD45G, GFPT2, GPRC5A, HBEGF, IER3, IER5L, JUN, JUNB, KCNK1, KIAA0040, KLF11, LIF, LPAR1, MAFF, MAK, MBNL2, MSX1, MT1A, MYBPH, MYH11, NCOA7, NFIL3, OSMR, OXCT2, PIM3, PLAUR, PNRC1, PPAN, PROK2, RCAN1, RGS1, RND1, ROM1, RRP12, S100P, SEMA4B, SERPINE1, SERTAD1, SLC22A4, SLITRK3, SOCS1, TMPRSS2, TNFRSF10D, TNFRSF1A, TUBB6, WT1, ZNF331 |
| **Negatively** |
| ABCC5, ABCG8, ACE2, ACSM1, AFP, AKR1B10, BCAR3, BTN3A1, C12orf27, C8orf47, C9orf91, CASC5, CCDC150, CKAP2, CSTA, FMO1, GLT8D1, GNAO1, GRAP, IFIT3, JUB, KANK4, LRRC20, MAMDC4, ME1, MOGAT1, NINL, OAS1, OAS2, ONECUT1, PALMD, PAQR4, PCBP4, PITPNM2, SKP2, SNHG7, THAP2, TMEM45B, TRIM14, WNT5A, ZNF385B |
| **28 genes correlated only with *Coprococcus* spp.** |
| **Positively** |
| BACH2, CCRN4L, CMTM2, IL10, IL4R, KLF5, LOC100131726, MIDN, MLF1, MMP25, PHLDA1, SNORA67, TGM2, TIPARP, TNFSF14, ZFP36 |
| **Negatively** |
| AMACR, ARID3C, CENPQ, FADS2, FBXO27, KDELR3, LGALS4, MAL2, OSGIN1, SLC25A42, SPTLC3, TRHDE |
|  |
| **12 genes correlated with *F. prausnitzii* spp. and *Ruminococcus* spp.** |
| **Positively** |
| C13orf15, MMP19, NRG1, PADI4, RALGDS |
| **Negatively** |
| ANKS4B, C3orf14, CSRNP3, CYP7A1, FRAT1, RFXDC2, TMEM169 |
|  |
| **12 genes correlated with *F. prausnitzii* and *Coprococcus* spp.** |
| **Positively** |
| ARID5B, FAM107A, LYST, RBPMS, SNORD19 |
| **Negatively** |
| EDAR, NCAPG2, PCDH24, PMS2, RTP4, SLC22A7, TNFSF10 |
|  |
| **46 genes correlated with *Ruminococcus* spp. and *Coprococcus* spp.** |
| **Positively** |
| ADAMTS1, AGPAT9, AXUD1, C19orf23, C21orf63, C2CD4B, CXCR7, DHODH, DHRS13, EIF5, EPHA2, GRHL1, IGFBP2, LARP1B, LDLR, MAP3K14, NR4A1, ODZ3, PPRC1, RGS2, S100A12, S100A8, SBNO2, SERPINB9, SLC11A1, UPP1 |
| **Negatively** |
| APOL3, C18orf56, CCDC121, CD3G, CX3CR1, FANCL, FRMD7, HIST1H2AC, HOPX, MARVELD3, MIR564, PEG10, RALGPS1, SLC12A1, SLC22A11, SLC25A34, TMEM20, TMEM63A, TYMS, ZWINT |
|  |
| **6 genes correlated with *F. prausnitzii*, *Ruminococcus* spp. and *Coprococcus* spp.** |
| **Positively** |
| FOSB, MYADM, SIPA1L2, SOCS2 |
| **Negatively** |
| IFIT2, RNF43 |

# **Supplementary Table 6:** Results of the Gene Ontology analysis (10) for hepatic genes significantly correlated with *F. prausnitzii* (False Discovery Rate, FDR <0.05)

| **Term** | **Count** | **%** | **p-value** | **Fold-enrichment** | **Benjamini** | **FDR** |
| --- | --- | --- | --- | --- | --- | --- |
| negative regulation of phosphorylation | 16 | 11.8 | 2.00E-07 | 5.5 | 6.40E-04 | 0.0004 |
| negative regulation of protein phosphorylation | 14 | 10.3 | 2.60E-06 | 5.2 | 2.10E-03 | 0.0048 |
| negative regulation of phosphorus metabolic process | 17 | 12.5 | 8.00E-07 | 4.6 | 1.30E-03 | 0.0015 |
| negative regulation of phosphate metabolic process | 17 | 12.5 | 8.00E-07 | 4.6 | 1.30E-03 | 0.0015 |
| cellular response to cytokine stimulus | 19 | 14 | 1.20E-06 | 3.9 | 1.30E-03 | 0.0022 |
| cellular response to hormone stimulus | 16 | 11.8 | 1.80E-05 | 3.8 | 3.50E-03 | 0.0320 |
| response to organonitrogen compound | 19 | 14 | 5.80E-06 | 3.5 | 2.30E-03 | 0.0110 |
| response to cytokine | 19 | 14 | 7.30E-06 | 3.5 | 2.60E-03 | 0.0130 |
| response to organic cyclic compound | 21 | 15.4 | 2.70E-06 | 3.4 | 1.70E-03 | 0.0049 |
| response to hormone | 20 | 14.7 | 4.40E-06 | 3.4 | 2.00E-03 | 0.0081 |
| response to nitrogen compound | 20 | 14.7 | 8.70E-06 | 3.3 | 2.80E-03 | 0.0160 |
| regulation of kinase activity | 18 | 13.2 | 2.20E-05 | 3.3 | 3.70E-03 | 0.0410 |
| regulation of protein phosphorylation | 24 | 17.6 | 1.70E-05 | 2.7 | 3.70E-03 | 0.0320 |
| response to endogenous stimulus | 28 | 20.6 | 3.50E-06 | 2.6 | 1.90E-03 | 0.0064 |
| regulation of phosphorylation | 25 | 18.4 | 1.70E-05 | 2.6 | 3.90E-03 | 0.0310 |
| regulation of protein modification process | 28 | 20.6 | 1.30E-05 | 2.5 | 3.60E-03 | 0.0240 |
| regulation of phosphate metabolic process | 27 | 19.9 | 2.20E-05 | 2.4 | 4.10E-03 | 0.0400 |
| regulation of phosphorus metabolic process | 27 | 19.9 | 2.20E-05 | 2.4 | 3.90E-03 | 0.0400 |
| cellular response to organic substance | 33 | 24.3 | 1.40E-05 | 2.2 | 3.50E-03 | 0.0260 |
| response to organic substance | 39 | 28.7 | 1.20E-05 | 2.0 | 3.40E-03 | 0.0210 |

FDR: False discovery rate

# **Supplementary Table 7:** Results of the Gene Ontology analysis (10) for hepatic genes significantly correlated with *Ruminococcus* spp. (False Discovery Rate, FDR <0.05)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Term** | **Count** | **%** | **p-value** | **Fold-enrichment** | **Benjamini** | **FDR** |
| response to lipopolysaccharide | 13 | 7.4 | 2.00E-05 | 4.7 | 1.80E-03 | 0.0370 |
| positive regulation of transcription from RNA polymerase II promoter | 31 | 17.6 | 1.30E-08 | 3.3 | 4.60E-05 | 0.0000 |
| response to cytokine | 24 | 13.6 | 9.20E-07 | 3.3 | 2.40E-04 | 0.0017 |
| response to lipid | 25 | 14.2 | 8.40E-07 | 3.2 | 2.40E-04 | 0.0015 |
| response to external biotic stimulus | 22 | 12.5 | 1.90E-05 | 2.9 | 1.70E-03 | 0.0350 |
| response to other organism | 22 | 12.5 | 1.90E-05 | 2.9 | 1.70E-03 | 0.0350 |
| circulatory system development | 24 | 13.6 | 1.40E-05 | 2.8 | 1.40E-03 | 0.0260 |
| cardiovascular system development | 24 | 13.6 | 1.40E-05 | 2.8 | 1.40E-03 | 0.0260 |
| positive regulation of intracellular signal transduction | 23 | 13.1 | 2.60E-05 | 2.8 | 2.20E-03 | 0.0480 |
| regulation of apoptotic process | 34 | 19.3 | 2.70E-07 | 2.7 | 1.60E-04 | 0.0005 |
| positive regulation of nucleic acid-templated transcription | 33 | 18.8 | 3.70E-07 | 2.7 | 1.40E-04 | 0.0007 |
| positive regulation of transcription, DNA-templated | 33 | 18.8 | 3.70E-07 | 2.7 | 1.40E-04 | 0.0007 |
| regulation of cell death | 36 | 20.5 | 1.60E-07 | 2.6 | 1.80E-04 | 0.0003 |
| regulation of programmed cell death | 34 | 19.3 | 3.40E-07 | 2.6 | 1.40E-04 | 0.0006 |
| positive regulation of RNA biosynthetic process | 33 | 18.8 | 5.20E-07 | 2.6 | 1.80E-04 | 0.0010 |
| positive regulation of RNA metabolic process | 33 | 18.8 | 1.10E-06 | 2.6 | 2.60E-04 | 0.0020 |
| anatomical structure formation involved in morphogenesis | 27 | 15.3 | 1.30E-05 | 2.6 | 1.40E-03 | 0.0250 |
| regulation of protein phosphorylation | 30 | 17 | 5.00E-06 | 2.5 | 8.50E-04 | 0.0091 |
| apoptotic process | 39 | 22.2 | 2.10E-07 | 2.4 | 1.50E-04 | 0.0004 |
| programmed cell death | 40 | 22.7 | 3.30E-07 | 2.4 | 1.60E-04 | 0.0006 |
| positive regulation of gene expression |   | 21 | 8.40E-07 | 2.4 | 2.60E-04 | 0.0015 |
| response to oxygen-containing compound | 32 | 18.2 | 6.60E-06 | 2.4 | 8.00E-04 | 0.0120 |
| regulation of phosphorylation | 30 | 17 | 1.90E-05 | 2.4 | 1.70E-03 | 0.0340 |
| response to external stimulus | 44 | 25 | 1.00E-07 | 2.3 | 1.70E-04 | 0.0002 |
| cell death | 42 | 23.9 | 1.70E-07 | 2.3 | 1.50E-04 | 0.0003 |
| regulation of intracellular signal transduction | 36 | 20.5 | 3.20E-06 | 2.3 | 6.80E-04 | 0.0058 |
| positive regulation of macromolecule biosynthetic process | 34 | 19.3 | 5.10E-06 | 2.3 | 8.30E-04 | 0.0093 |
| regulation of phosphate metabolic process | 34 | 19.3 | 6.00E-06 | 2.3 | 8.20E-04 | 0.0110 |
| regulation of phosphorus metabolic process | 34 | 19.3 | 6.10E-06 | 2.3 | 8.00E-04 | 0.0110 |
| regulation of cell proliferation | 32 | 18.2 | 2.00E-05 | 2.3 | 1.80E-03 | 0.0370 |
| transcription from RNA polymerase II promoter | 37 | 21 | 4.60E-06 | 2.2 | 8.70E-04 | 0.0084 |
| regulation of transcription from RNA polymerase II promoter | 37 | 21 | 5.50E-06 | 2.2 | 8.20E-04 | 0.0100 |
| positive regulation of biosynthetic process | 36 | 20.5 | 5.70E-06 | 2.2 | 8.20E-04 | 0.0100 |
| positive regulation of cellular biosynthetic process | 35 | 19.9 | 1.00E-05 | 2.2 | 1.20E-03 | 0.0190 |
| positive regulation of nucleobase-containing compound metabolic process | 34 | 19.3 | 1.10E-05 | 2.2 | 1.20E-03 | 0.0200 |
| positive regulation of macromolecule metabolic process | 51 | 29 | 1.30E-06 | 2 | 3.00E-04 | 0.0024 |
| positive regulation of metabolic process | 52 | 29.5 | 3.90E-06 | 1.9 | 7.90E-04 | 0.0072 |
| regulation of signal transduction | 48 | 27.3 | 4.90E-06 | 1.9 | 8.90E-04 | 0.0090 |
| response to organic substance | 49 | 27.8 | 5.30E-06 | 1.9 | 8.30E-04 | 0.0098 |
| positive regulation of cellular metabolic process | 49 | 27.8 | 6.30E-06 | 1.9 | 8.00E-04 | 0.0120 |
| cellular response to chemical stimulus | 46 | 26.1 | 1.20E-05 | 1.9 | 1.30E-03 | 0.0230 |
| regulation of signaling | 51 | 29 | 1.10E-05 | 1.8 | 1.20E-03 | 0.0200 |
| regulation of cell communication | 50 | 28.4 | 1.50E-05 | 1.8 | 1.50E-03 | 0.0280 |
| regulation of molecular function | 48 | 27.3 | 1.80E-05 | 1.8 | 1.70E-03 | 0.0340 |

FDR: False discovery rate

# **Supplementary Table 8:** Results of the Gene Ontology analysis (10) for hepatic genes significantly correlated with *Coprococcus* spp. (False Discovery Rate, FDR <0.05

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Term** | **Count** | **%** | **p-value** | **Fold-enrichment** | **Benjamini** | **FDR** |
| response to external stimulus | 26 | 28.3 | 9.90E-06 | 2.5 | 2.20E-02 | 0.0170 |

FDR: False discovery rat

**Supplementary Table 9:** Significantly upregulated pathways in NAFLD grouped into KEGG level II and level III.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **KEGG#** | **baseMean** | **log2FoldChange** | **lfcSE** | **stat** | **pvalue** | **padj** | **KEGG level II** | **KEGG level III** |
| K00030 | 64.9266777 | 0.99982519 | 0.45655505 | 2.18993347 | 0.02852906 | 0.69879016 | Carbohydrate metabolism | Citrate cycle (TCA cycle) |
| K00186 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Amino acid metabolism | Valine, leucine and isoleucine degradation |
| K00187 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Amino acid metabolism | Valine, leucine and isoleucine degradation |
| K00188 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Amino acid metabolism | Valine, leucine and isoleucine degradation |
| K00204 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Energy metabolism | Methane metabolism |
| K00205 | 15.2368773 | 4.46794763 | 1.62515985 | 2.74923579 | 0.00597344 | 0.42386389 | Energy metabolism | Methane metabolism |
| K00303 | 38.059664 | 1.33232501 | 0.58422805 | 2.2804879 | 0.02257877 | 0.64705382 | Amino acid metabolism | Glycine, serine and threonine metabolism |
| K00372 | 40.2081735 | 1.40183127 | 0.57316221 | 2.44578453 | 0.01445374 | 0.51144921 | Energy metabolism | Nitrogen metabolism |
| K00392 | 38.4979937 | 1.34065605 | 0.57367298 | 2.33696913 | 0.01944079 | 0.60187499 | Energy metabolism | Sulfur metabolism |
| K00442 | 11.2797909 | 5.09266192 | 1.62663064 | 3.13080412 | 0.00174328 | 0.26614017 | Energy metabolism | Methane metabolism |
| K00824 | 165.398609 | 0.56511232 | 0.28705793 | 1.96863509 | 0.04899501 | 0.94763182 | Amino acid metabolism | Lysine degradation |
| K00955 | 61.6344575 | 1.17102697 | 0.57545085 | 2.03497305 | 0.04185359 | 0.86376375 | Energy metabolism | Sulfur metabolism |
| K00966 | 61.5422346 | 1.17076994 | 0.57536378 | 2.03483427 | 0.04186755 | 0.86376375 | Carbohydrate metabolism | Fructose and mannose metabolism |
| K01114 | 38.3669093 | 1.2607678 | 0.58419541 | 2.15812684 | 0.03091797 | 0.72526427 |  Lipid metabolism | Glycerophospholipid metabolism |
| K01236 | 1.01382967 | 2.3095335 | 0.92428881 | 2.49871411 | 0.01246448 | 0.46073258 | Carbohydrate metabolism | Starch and sucrose metabolism |
| K01590 | 11.8185335 | 1.06507365 | 0.40753511 | 2.61345252 | 0.00896325 | 0.42386389 | Amino acid metabolism | Histidine metabolism |
| K01761 | 45.5171121 | 1.18630151 | 0.53966283 | 2.1982272 | 0.02793292 | 0.68694707 | Amino acid metabolism | Cysteine and methionine metabolism |
| K01777 | 46.1178689 | 1.16730177 | 0.5232484 | 2.23087497 | 0.02568941 | 0.67534364 | Amino acid metabolism | Arginine and proline metabolism |
| K02586 | 152.844074 | 1.32184049 | 0.56074058 | 2.35731199 | 0.01840777 | 0.58779589 | Energy metabolism | Nitrogen metabolism |
| K02591 | 195.069967 | 1.29636735 | 0.55463889 | 2.3373178 | 0.01942267 | 0.60187499 | Energy metabolism | Nitrogen metabolism |
| K02594 | 78.2578164 | 1.25425549 | 0.56210922 | 2.23133772 | 0.02565877 | 0.67534364 | Carbohydrate metabolism | Pyruvate metabolism |
| K03725 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Unclassified: metabolism |  |
| K06015 | 95.1161204 | 1.04172154 | 0.47182816 | 2.20784098 | 0.02725536 | 0.67573351 |  Unclassified: metabolism |  |
| K06044 | 1.01382967 | 2.3095335 | 0.92428881 | 2.49871411 | 0.01246448 | 0.46073258 | Carbohydrate metabolism | Starch and Sucrose metabolism |
| K06208 | 38.5630683 | 1.32216577 | 0.58740339 | 2.25086508 | 0.02439408 | 0.65831645 | Amino acid metabolism | Phenylalanine, tyrosine and tryptophan biosynthesis |
| K06324 | 10.6814737 | 1.47394367 | 0.71119014 | 2.07250295 | 0.03821856 | 0.82682103 | Unclassified: metabolism |  |
| K06350 | 41.1556901 | 1.33410931 | 0.55500711 | 2.40376978 | 0.01622699 | 0.53700372 |  Unclassified: signaling and cellular processes |  |
| K06351 | 41.1556901 | 1.33410931 | 0.55500711 | 2.40376978 | 0.01622699 | 0.53700372 |  Unclassified: signaling and cellular processes |  |
| K06415 | 38.4736054 | 1.26517363 | 0.58293552 | 2.1703492 | 0.0299804 | 0.71425966 |  Unclassified: signaling and cellular processes |  |
| K06439 | 37.8874972 | 1.32524673 | 0.5829931 | 2.27317739 | 0.02301549 | 0.64705382 |  Unclassified: signaling and cellular processes |  |
| K06440 | 38.0088502 | 1.33039166 | 0.57945423 | 2.29593916 | 0.02167936 | 0.63643057 |  Unclassified: signaling and cellular processes |  |
| K06862 | 11.2797909 | 5.09266192 | 1.62663064 | 3.13080412 | 0.00174328 | 0.26614017 | Unclassified: Energy metabolism |  |
| K06886 | 1.09272514 | 2.81155851 | 1.41577524 | 1.98587913 | 0.04704675 | 0.92860228 | Poorly characterized |  |
| K06906 | 6.1176458 | 1.62335611 | 0.79692297 | 2.03703015 | 0.04164702 | 0.86376375 | Poorly characterized |  |
| K06933 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Poorly characterized |  |
| K07129 | 38.0596456 | 1.30528554 | 0.5820761 | 2.24246545 | 0.02493131 | 0.66985038 | Poorly characterized |  |
| K09126 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 |  Poorly characterized |  |
| K10793 | 42.9859555 | 1.18972994 | 0.55504408 | 2.14348732 | 0.03207399 | 0.74663838 | Amino acid metabolism | Arginine and proline metabolism |
| K10794 | 41.6788358 | 1.27200549 | 0.56983134 | 2.23224912 | 0.0255985 | 0.67534364 | Amino acid metabolism | Arginine and proline metabolism |
| K13015 | 38.616895 | 1.29525195 | 0.57223732 | 2.26348737 | 0.02360566 | 0.64851772 | Carbohydrate metabolism | Amino sugar and nucleotide sugar metabolism |
| K13019 | 37.9010289 | 1.32576963 | 0.58382509 | 2.27083361 | 0.02315705 | 0.64705382 | Carbohydrate metabolism | Amino sugar and nucleotide sugar metabolism |
| K13942 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Energy metabolism | Methane metabolism |
| K14080 | 15.4337016 | 4.26351264 | 1.61624018 | 2.63792021 | 0.00834162 | 0.42386389 | Energy metabolism | Methane metabolism |
| K14081 | 9.53916579 | 4.38469273 | 1.65136019 | 2.6552007 | 0.00792612 | 0.42386389 | Energy metabolism | Methane metabolism |
| K14092 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Unclassified: metabolism |  |
| K14093 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Unclassified: metabolism |  |
| K14094 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Unclassified: metabolism |  |
| K14095 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Unclassified: metabolism |  |
| K14096 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Unclassified: metabolism |  |
| K14097 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Unclassified: metabolism |  |
| K14098 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Unclassified: metabolism |  |
| K14099 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Unclassified: metabolism |  |
| K14100 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Unclassified: metabolism |  |
| K14101 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Unclassified: metabolism |  |
| K14102 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Unclassified: metabolism |  |
| K14103 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Unclassified: metabolism |  |
| K14104 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Unclassified: metabolism |  |
| K14105 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Unclassified: metabolism |  |
| K14106 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Unclassified: metabolism |  |
| K14107 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Unclassified: metabolism |  |
| K14108 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Unclassified: metabolism |  |
| K15553 | 287.905221 | 0.54575683 | 0.25790344 | 2.11612855 | 0.03433388 | 0.78427832 | Energy metabolism | Sulfur metabolism |
| K15555 | 131.031848 | 1.01482031 | 0.4967065 | 2.0430985 | 0.04104269 | 0.86376375 | Energy metabolism | Sulfur metabolism |
| K15898 | 8.08811074 | 4.15118484 | 1.45238705 | 2.85818084 | 0.00426077 | 0.36092314 | Carbohydrate metabolism | Amino sugar and nucleotide sugar metabolism |
| K15976 | 75.7279012 | 1.32063267 | 0.57148104 | 2.31089498 | 0.02083866 | 0.62460353 | Unclassified: metabolism |  |
| K16705 | 37.8673529 | 1.3243756 | 0.58419155 | 2.26702286 | 0.02338883 | 0.64705382 | Unclassified: metabolism |  |
| K17830 | 13.0591605 | 5.30591292 | 1.64276416 | 3.22986892 | 0.00123847 | 0.26614017 |  Lipid metabolism | Glycerophospholipid metabolism |
| K19244 | 7.84941533 | 4.10201973 | 1.58872764 | 2.58195277 | 0.0098243 | 0.42386389 | Amino acid metabolism | Alanine, aspartate and glutamate metabolism |
| K19266 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Carbohydrate metabolism | Pyruvate metabolism |
| K19817 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Unclassified: metabolism |  |
| K19824 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Unclassified: metabolism |  |
| K20025 | 39.989175 | 1.18133701 | 0.56908152 | 2.07586605 | 0.03790634 | 0.82568135 | Unclassified: metabolism |  |
| K20026 | 39.989175 | 1.18133701 | 0.56908152 | 2.07586605 | 0.03790634 | 0.82568135 | Unclassified: metabolism |  |
| K20882 | 39.989175 | 1.18133701 | 0.56908152 | 2.07586605 | 0.03790634 | 0.82568135 | Unclassified: metabolism |  |
| K21909 | 37.8462382 | 1.32349831 | 0.58404664 | 2.26608327 | 0.02344629 | 0.64705382 | Unclassified: metabolism |  |
| K21935 | 37.8874972 | 1.32524673 | 0.5829931 | 2.27317739 | 0.02301549 | 0.64705382 | Unclassified: metabolism |  |
| K22223 | 7.79534734 | 4.09190271 | 1.58755998 | 2.57747912 | 0.00995239 | 0.42386389 | Carbohydrate metabolism | Glyoxylate and dicarboxylate metabolism |

# **Supplementary Table 10: Differentially expressed hepatic genes significantly correlated with bacterial metabolic pathways carbohydrate, lipid, energy, and amino acid metabolism (p<0.05) in NAFLD patients.**

|  |  |
| --- | --- |
| Genes |  |
| **Carbohydrate metabolism** |
|  | rho | p.value |
| CDC42EP3 | 0.7472527 | 0.004823 |
| FMO1 | 0.6978022 | 0.01033 |
| MAMDC4 | 0.5604396 | 0.04989 |
| NR5A2 | 0.5604396 | 0.04989 |
| RALGPS1 | 0.7307692 | 0.006323 |
| TMEM63A | 0.7032967 | 0.009557 |
| C8orf47 | 0.5714286 | 0.04489 |
| CD3G | 0.5549451 | 0.05253 |
| FADS2 | 0.7252747 | 0.006892 |
| FBXO27 | 0.6263736 | 0.02527 |
| KCNE4 | 0.5604396 | 0.04989 |
| ME1 | 0.6813187 | 0.01293 |
| PEG10 | 0.6263736 | 0.02527 |
| SLC25A42 | 0.5769231 | 0.04254 |
| ACE2 | 0.6098901 | 0.03029 |
| C6orf129 | -0.5659341 | 0.04734 |
| DBI | -0.7197802 | 0.007498 |
| HIST2H2AC | -0.5769231 | 0.04254 |
| LOC100131726 | -0.8351648 | 0.0006517 |
| ODZ3 | -0.7527473 | 0.004387 |
| RBP5 | -0.6098901 | 0.03029 |
| **Lipid Metabolism** |
| CDC42EP3 | 0.7087912 | 0.008829 |
| FMO1 | 0.6868132 | 0.01201 |
| MAMDC4 | 0.8296703 | 0.0007779 |
| NR5A2 | 0.6648352 | 0.016 |
| RALGPS1 | 0.7032967 | 0.009557 |
| TMEM63A | 0.5714286 | 0.04489 |
| C8orf47 | 0.5164835 | 0.07404 |
| CD3G | 0.2307692 | 0.4478 |
| FADS2 | 0.4010989 | 0.1758 |
| FBXO27 | 0.3791209 | 0.2024 |
| KCNE4 | 0.5384615 | 0.06109 |
| ME1 | 0.2307692 | 0.4478 |
| PEG10 | 0.4450549 | 0.1298 |
| SLC25A42 | 0.4835165 | 0.09704 |
| CHI3L1 | 0.5549451 | 0.05253 |
| FLJ36031 | 0.8681319 | 0.0001246 |
| MARVELD3 | 0.7307692 | 0.006323 |
| MYC | 0.5989011 | 0.03403 |
| NAMPT | 0.6428571 | 0.0209 |
| PCBP4 | 0.6813187 | 0.01293 |
| SDC4 | 0.6868132 | 0.01201 |
| STAT3 | 0.6978022 | 0.01033 |
| ACE2 | 0.6428571 | 0.0209 |
| AGBL5 | -0.5714286 | 0.04489 |
| ANXA9 | -0.7417582 | 0.00529 |
| C10orf114 | -0.6868132 | 0.01201 |
| C14orf80 | -0.5714286 | 0.04489 |
| C2CD4B | -0.6483516 | 0.01958 |
| C6orf129 | -0.6428571 | 0.0209 |
| DHRS4L1 | -0.7032967 | 0.009557 |
| DIO3OS | -0.6043956 | 0.03212 |
| GIMAP2 | -0.7912088 | 0.002082 |
| HIST2H2AC | -0.8021978 | 0.001624 |
| KIF22 | -0.6758242 | 0.0139 |
| MAK | -0.7307692 | 0.006323 |
| MGC12965 | -0.5879121 | 0.03811 |
| MT1A | -0.5549451 | 0.05253 |
| ODZ3 | -0.6208791 | 0.02687 |
| RGS2 | -0.5824176 | 0.04028 |
| RMND1 | -0.7197802 | 0.007498 |
| ROM1 | -0.5989011 | 0.03403 |
| ZNHIT3 | -0.7472527 | 0.004823 |
| **Energy Metabolism** |
| CDC42EP3 | 0.8131868 | 0.001239 |
| FMO1 | 0.6428571 | 0.0209 |
| MAMDC4 | 0.6318681 | 0.02374 |
| NR5A2 | 0.6263736 | 0.02527 |
| RALGPS1 | 0.6648352 | 0.016 |
| KCNE4 | 0.5989011 | 0.03403 |
| CHI3L1 | 0.5714286 | 0.04489 |
| FLJ36031 | 0.6648352 | 0.016 |
| MARVELD3 | 0.7582418 | 0.003981 |
| MYC | 0.5879121 | 0.03811 |
| PCBP4 | 0.6428571 | 0.0209 |
| SDC4 | 0.5659341 | 0.04734 |
| ANXA9 | -0.6483516 | 0.01958 |
| C10orf114 | -0.6978022 | 0.01033 |
| C14orf80 | -0.5824176 | 0.04028 |
| C2CD4B | -0.5824176 | 0.04028 |
| C6orf129 | -0.5824176 | 0.04028 |
| DHRS4L1 | -0.6648352 | 0.016 |
| DIO3OS | -0.5824176 | 0.04028 |
| GIMAP2 | -0.6978022 | 0.01033 |
| HIST2H2AC | -0.7857143 | 0.002341 |
| IL1RN | 0.6813187 | 0.01293 |
| IVNS1ABP | 0.5604396 | 0.04989 |
| KIF22 | -0.5769231 | 0.04254 |
| MGC12965 | -0.7417582 | 0.00529 |
| OAS2 | 0.5659341 | 0.04734 |
| ODZ3 | -0.7417582 | 0.00529 |
| PRKCZ | -0.5824176 | 0.04028 |
| RTP3 | -0.5659341 | 0.04734 |
| S100A12 | -0.5659341 | 0.04734 |
| ZNF385B | 0.5714286 | 0.04489 |
| ZNHIT3 | -0.7307692 | 0.006323 |
| **Amino acid metabolism** |
| CDC42EP3 | 0.7637363 | 0.003602 |
| FMO1 | 0.6043956 | 0.03212 |
| MAMDC4 | 0.9010989 | 2.20E-16 |
| NR5A2 | 0.6208791 | 0.02687 |
| RALGPS1 | 0.5769231 | 0.04254 |
| TMEM63A | 0.6208791 | 0.02687 |
| C8orf47 | 0.6373626 | 0.02229 |
| FLJ36031 | 0.7692308 | 0.00325 |
| MARVELD3 | 0.7032967 | 0.009557 |
| MYC | 0.7142857 | 0.008143 |
| NAMPT | 0.7087912 | 0.008829 |
| ACE2 | 0.6758242 | 0.0139 |
| ANXA9 | -0.6703297 | 0.01492 |
| ATP2B4 | 0.5714286 | 0.04489 |
| BIRC3 | 0.5934066 | 0.03603 |
| C10orf114 | -0.6868132 | 0.01201 |
| C14orf80 | -0.7362637 | 0.005789 |
| C6orf129 | -0.7197802 | 0.007498 |
| CASC5 | 0.5604396 | 0.04989 |
| CASP4 | 0.6153846 | 0.02854 |
| CEBPA | -0.5989011 | 0.03403 |
| CISH | 0.6098901 | 0.03029 |
| CMTM2 | -0.5659341 | 0.04734 |
| DHRS4L1 | -0.6263736 | 0.02527 |
| FAM124B | -0.6153846 | 0.02854 |
| GIMAP2 | -0.6648352 | 0.016 |
| HIST2H2AB | -0.5604396 | 0.04989 |
| HIST2H2AC | -0.6648352 | 0.016 |
| IL1RN | 0.6153846 | 0.02854 |
| KIF22 | -0.6208791 | 0.02687 |
| LONRF2 | 0.7912088 | 0.002082 |
| MAK | -0.6538462 | 0.01832 |
| MGC12965 | -0.5769231 | 0.04254 |
| MIR21 | 0.6263736 | 0.02527 |
| OGG1 | -0.6153846 | 0.02854 |
| OSMR | 0.5604396 | 0.04989 |
| P8 | -0.6043956 | 0.03212 |
| PPAN | -0.6758242 | 0.0139 |
| RMND1 | -0.6263736 | 0.02527 |
| RRP12 | -0.5714286 | 0.04489 |
| TMEM45B | 0.5934066 | 0.03603 |
| TNFSF10 | -0.6098901 | 0.03029 |
| TP53AP1 | -0.5824176 | 0.04028 |
| ZNHIT3 | -0.7527473 | 0.004387 |

**Supplementary Table 11:** Differentially expressed hepatic genes significantly correlated with KEGG level II bacterial metabolic pathways carbohydrate, lipid, energy, and amino acid metabolism (p<0.05) in healthy controls.

|  |  |
| --- | --- |
| Genes |  |
| **Carbohydrate Metabolism** |
|  | **rho** | **p.value** |
| C13orf15 | 0.8333333 | 0.01538 |
| ENHO | -0.7619048 | 0.03676 |
| HOPX | -0.7380952 | 0.04583 |
| LRRN3 | -0.8095238 | 0.02178 |
| TNFSF10 | -0.7619048 | 0.03676 |
| VNN2 | -0.7619048 | 0.03676 |
| **Lipid Metabolism** |
| C13orf15 | 0.8095238 | 0.02178 |
| ENHO | -0.7380952 | 0.04583 |
| LRRN3 | -0.7380952 | 0.04583 |
| **Energy Metabolism** |
| ACTG2 | -0.7619048 | 0.03676 |
| AKR1B10 | -0.7619048 | 0.03676 |
| C9orf125 | -0.7380952 | 0.04583 |
| FRMD7 | -0.8095238 | 0.02178 |
| IFIT2 | -0.8333333 | 0.01538 |
| IFIT3 | -0.7857143 | 0.02793 |
| LPAR1 | -0.7380952 | 0.04583 |
| NR4A1 | -0.7380952 | 0.04583 |
| PCBP4 | -0.8333333 | 0.01538 |
| PROK2 | -0.7619048 | 0.03676 |
| **Amino acid Metabolism** |
| C13orf15 | 0.7380952 | 0.04583 |
| ACTG2 | -0.7619048 | 0.03676 |
| AKR1B10 | -0.7619048 | 0.03676 |
| BCAR3 | -0.7857143 | 0.02793 |
| C9orf125 | -0.7380952 | 0.04583 |
| CREB5 | -0.8095238 | 0.02178 |
| FRMD7 | -0.8809524 | 0.007242 |
| IFIT2 | -0.8333333 | 0.01538 |
| IFIT3 | -0.7380952 | 0.04583 |
| LRRN3 | -0.7857143 | 0.02793 |
| MAMDC4 | -0.7380952 | 0.04583 |
| NR4A1 | -0.8333333 | 0.01538 |
| PCBP4 | -0.7380952 | 0.04583 |
| SLC22A4 | -0.7619048 | 0.03676 |
|  |  |  |

# **References:**

1. Arendt BM, Comelli EM, Ma DW, et al. Altered hepatic gene expression in nonalcoholic fatty liver disease is associated with lower hepatic n-3 and n-6 polyunsaturated fatty acids. Hepatology 2015;61:1565-78.

2. Da Silva HE, Teterina A, Comelli EM, et al. Nonalcoholic fatty liver disease is associated with dysbiosis independent of body mass index and insulin resistance. Sci Rep 2018;8:1466.

3. Brunt EM, Janney CG, Di Bisceglie AM, et al. Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. Am J Gastroenterol 1999;94:2467-74.

4. Kleiner DE, Brunt EM, Van Natta M, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology 2005;41:1313-21.

5. Iwai S, Weinmaier T, Schmidt BL, et al. Piphillin: Improved Prediction of Metagenomic Content by Direct Inference from Human Microbiomes. PLoS One 2016;11:e0166104.

6. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27-30.

7. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.

8. Matthews DR, Hosker JP, Rudenski AS, et al. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia 1985;28:412-9.

9. Wolfs MG, Gruben N, Rensen SS, et al. Determining the association between adipokine expression in multiple tissues and phenotypic features of non-alcoholic fatty liver disease in obesity. Nutr Diabetes 2015;5:e146.

10. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44-57.

11. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009;37:1-13.