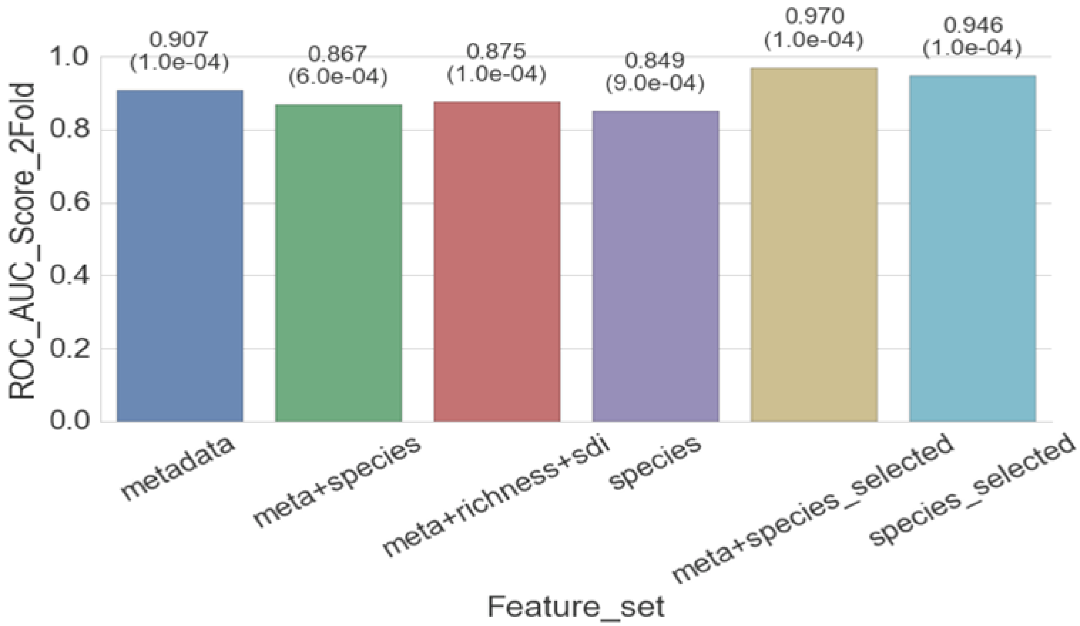
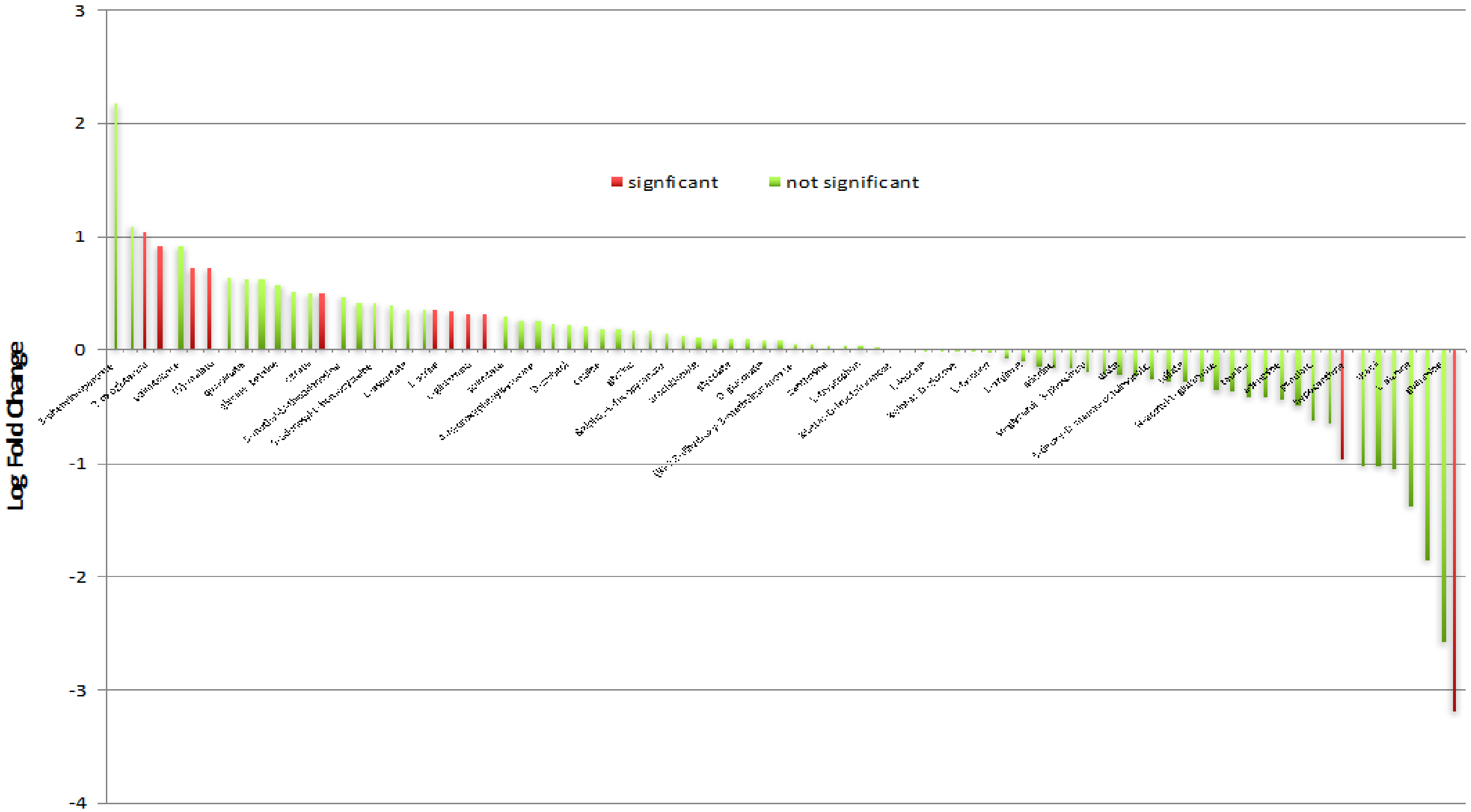
Figure S1. SVM performance of combination of input features, Related to STAR Methods
Metadata refers to patients metadata (Age, Sex, Race and BMI). SDI refers to Shannon diversity index



Supplemental Figure 1. SVM performance of combination of input features. Metadata refers to patient metadata (Age, Sex, Race, and BMI). SDI refers to Shannon diversity index.

Figure S2. **Log fold change of compounds, Related to Figure 3**
Log$_2$ ratios (G2/G1) using the median intensities of compounds detected in serum samples and associated with pathways reconstructed in species bins



Supplemental Figure 2. Log fold change of compounds. Log$_2$ ratios (G2/G1) using the median intensities of compounds detected in serum samples and associated with pathways reconstructed in species bins.

Table S2. Metagenome based relative abundances of enzymes associated with SCFA and ethanol metabolism, and their enrichment in group G2 (log fold change is given in the last column), Related to Table 4.

EC	Metabolite	Function	p-val	#g1 reconstructed	#g2 reconstructed	median g1 abundance	median g2 abundance	g2 enrichment *
1.1.1.1	Ethanol	Alcohol dehydrogenase	0.37	72	14	3.04E-03	3.33E-03	1.34E-01
1.1.1.2	Ethanol	Alcohol dehydrogenase (NADP(+))	0.16	54	7	8.95E-06	7.50E-07	-3.58E+00
1.1.2.7	Ethanol / Methanol	Methanol dehydrogenase	Not found	0	0	na	na	na
1.1.2.8	Ethanol	Alcohol dehydrogenase (cytochrome c)	0.87	2	0	0.00E+00	0.00E+00	na
1.1.1.27	L-Lactate	L-lactate dehydrogenase	0.06	72	14	1.72E-03	1.12E-03	-6.15E-01
1.1.1.28	D-Lactate	D-lactate dehydrogenase	0.03	72	14	3.08E-04	6.29E-04	1.03E+00
2.3.1.54	Formate	Formate C-acetyltransferase	0.52	72	14	1.42E-03	1.32E-03	-1.07E-01
1.1.99.33	Formate Utilization	Formate dehydrogenase (acceptor)	0.44	1	2	0.00E+00	0.00E+00	na
3.6.1.7	Acetate	Acylphosphate phosphohydrolase	0.03	72	14	4.08E-04	2.90E-04	-4.93E-01
2.7.2.12	Acetate	Acetate kinase (diphosphate)	Not found					na
6.2.1.13	Acetate	Acetate--CoA ligase (ADP-forming)	0.58	12	1	0.00E+00	0.00E+00	na
2.7.2.1	Acetate	Acetate kinase	0.03	72	14	3.20E-03	2.76E-03	-2.16E-01
1.3.5.4	Succinate	Fumarate reductase (quinol)	0.97	30	6	0.00E+00	0.00E+00	na
1.3.5.1	Succinate	Succinate dehydrogenase (quinone)	0.89	72	13	2.51E-04	2.94E-04	2.30E-01
2.8.3.8	Butyrate / Acetate	Acetate CoA-transferase	0.43	72	14	4.64E-04	5.53E-04	2.52E-01
2.7.2.7	Butyrate	Butyrate kinase	0.93	72	14	9.11E-04	9.61E-04	7.74E-02
2.8.3.1	Propionate / Buterate	Propionate CoA-transferase	0.93	4	1	0.00E+00	0.00E+00	na
2.7.2.15	Propionate	Propionate kinase	0.27	39	8	2.05E-06	2.06E-05	3.33E+00

Table S3. Pathways with top fold change differences (in relative abundance) across the groups G1 and G2, Related to Figure 3.

Metacyc Class	Pathway	MetaCycID	Log FC	p-val	#G2	G2 median	#G1	G1 median
Steroids-Degradation	androstenedione degradation	PWY-6944	-2.81	0.18	7	2.08E-03	50	1.45E-02
NUCLEO-DEG	thymine degradation	PWY-6430	-2.47	0.21	7	6.31E-04	39	3.49E-03
NUCLEO-DEG	uracil degradation I (reductive)	PWY-3982	-2.47	0.21	7	6.31E-04	39	3.49E-03
Nucleotide-Biosynthesis	pyrimidine deoxyribonucleotides de novo biosynthesis III	PWY-6545	-2.46	0.03	7	5.97E-03	63	3.28E-02
SECONDARY-METABOLITE-BIOSYNTHESIS	rhizocticin A and B biosynthesis	PWY-7510	-2.33	0.64	7	1.14E-03	43	5.72E-03
Cofactor-Biosynthesis	trans, trans-farnesyl diphosphate biosynthesis	PWY-5123	-2.18	0.01	8	1.17E-02	67	5.32E-02
Alcohol-Degradation	L-1,2-propanediol degradation	PWY-7013	-1.7	0.84	7	6.31E-04	37	2.05E-03
Amino-Acid-Degradation	methionine degradation II	PWY-701	-1.7	0.85	7	6.31E-04	37	2.05E-03
Polyamine-Biosynthesis	putrescine biosynthesis II	PWY-43	-1.68	0.50	7	1.66E-03	40	5.29E-03
Nucleotide-Biosynthesis	adenine and adenosine salvage II	PWY-6605	-1.3	0.36	7	5.87E-03	51	1.45E-02
CARBOXYLATES-DEG	glycolate and glyoxylate degradation I	GLYCOLATEMET-PWY	1.95	0.14	12	1.50E-02	43	3.89E-03
Noncarbon-Nutrients	nitrate reduction IV (dissimilatory)	PWY-5674	1.95	0.23	9	8.66E-03	37	2.24E-03
AT5G03860-MONOMER	glyoxylate cycle	GLYOXYLATE-BYPASS	2.15	0.16	10	9.92E-03	37	2.24E-03
CHEMOAUTOTROPHIC-ENERGY-METABOLISM	hydrogen oxidation I (aerobic)	P283-PWY	2.15	0.16	10	9.92E-03	37	2.24E-03
NUCLEO-DEG	pyrimidine deoxyribonucleotides dephosphorylation	PWY-7206	2.21	0.08	11	1.77E-02	40	3.82E-03
Carbohydrates-Biosynthesis	CMP-N-acetylneuraminate biosynthesis II (bacteria)	PWY-6139	2.32	0.19	8	2.58E-02	39	5.18E-03
Cofactor-Biosynthesis	glutathione redox reactions II	GLUT-REDOX-PWY	2.39	0.12	10	1.58E-02	39	3.03E-03
Alcohol-Degradation	ethanol degradation I	ETOH-ACETYLCOA-ANA-PWY	2.52	0.12	10	1.50E-02	38	2.61E-03
AROMATIC-COMPOUNDS-DEGRADATION	p-cymene degradation to p-cumate	PWY-741	3.35	0.14	10	9.92E-03	36	9.74E-04

Lipid-Biosynthesis	phospholipid remodeling (phosphatidate, yeast)	PWY-7417	3.35	0.14	10	9.92E-03	36	9.74E-04

Table S4. Species selected by a Random Forest model trained with data from an age-balanced group of patients. Related to STAR Methods

Species in bold overlap with the 37 species selected by our original model. Mean decrease in Gini index and log ratios of median species abundances in G2 and G1 are also listed,

Species	Mean Decrease Gini	Log2(G2/G1)
Oscillibacter_sp._CAG.241	2.41	-5.11
Firmicutes_bacterium_CAG.129	1.25	-4.57
Firmicutes_bacterium_CAG.170	1	-3.73
[Ruminococcus]_obeum	1.62	-3.68
[Bacteroides]_pectinophilus	0.36	-3.02
Oscillibacter_sp._1.3	0.94	-2.78
Holdemania_filiformis	1.21	-2.47
Firmicutes_bacterium_CAG.83	1.47	-2.33
Firmicutes_bacterium_CAG.103	0.79	-1.44

Table S5. Top features selected by Linear SVM, Related to Figure 2

Species in bold are those that are also included in the Random Forest model.

Rank	Feature	Coefficient
1	***Streptococcus parasanguinis***	1.02
2	***Firmicutes* bacterium CAG:129**	-0.90
3	*Firmicutes* bacterium CAG:24	-0.89
4	Age	0.83
5	Asian	-0.80
6	***Clostridium* sp. CAG:43**	-0.80
7	***Oscillibacter* sp. CAG:241**	-0.80
8	*Firmicutes* bacterium CAG:103	0.77
9	***Ruminococcus* sp. CAG:90**	-0.77
10	***Ruminococcus obeum* CAG:39**	-0.76
11	***Eubacterium rectale***	-0.72
12	*Alistipes shahii*	0.72
13	*Clostridium symbiosum*	0.66
14	***Clostridium* sp._7_3_54FAA**	0.61
15	*Clostridium* sp._CAG:58	-0.58
16	***Bacteroides* sp._1_1_30**	-0.56
17	*Bacteroides* sp._1_1_14	-0.43
18	Female	0.42
19	*Escherichia coli*	-0.40
20	***Bacteroides finegoldii***	0.38
21	**Blautia sp._CAG:37**	-0.35
22	Hispanic	0.14

Table S6. Species that have correlated abundances to representative species that are shown in bold, Related to Figure 3

- **Clostridium_sp._SS2.1**
  - Lachnospiraceae_bacterium_5_1_63FAA
  - Lachnospiraceae_bacterium_CAG.25
  - Anaerostipes_hadrus
  - butyrate.producing_bacterium_SSC.2
- **Bacteroides_sp._D2**
  - Lachnospiraceae_bacterium_5_1_63FAA
  - Bacteroides_uniformis_CAG.3
- **Bacteroides_vulgatus**
  - Lachnospiraceae_bacterium_5_1_63FAA
  - Bacteroides_vulgatus_CAG.6
- **Dorea_longicatena**
  - Dorea_longicatena_CAG.42
- **Eubacterium_hallii**
  - Eubacterium_hallii_CAG.12
- **Eubacterium_rectale**
  - Eubacterium_rectale_CAG.36
- **Lachnospiraceae_bacterium_3_1_46FAA**
  - Lachnospiraceae_bacterium_1_1_57FAA
  - Lachnospiraceae_bacterium_8_1_57FAA
- **Lachnospiraceae_bacterium_2_1_58FAA**
  - Ruminococcus_gnavus
- **Oscillibacter_sp._1.3**
  - Oscillibacter_sp._CAG.155
- **Ruminococcus_sp._5_1_39BFAA**
  - Ruminococcus_sp._CAG.9
- **Streptococcus_parasanguinis**
  - Streptococcus_salivarius