

MOMS-PI

2018-12-28

Settings

HMP2 Data

<https://ibdmdb.org/tunnel/public/summary.html> - 178 samples

Metadata

```
[1] "Data dimensions"
```

```
[1] 5533 490
```

Project	External ID	Participant ID	site_sub_coll	data_type
C3001CSC1_BP	206615	C3001	C3001CSC1	biopsy_16S
C3001CSC2_BP	206614	C3001	C3001CSC2	biopsy_16S
C3002CSC1_BP	206617	C3002	C3002CSC1	biopsy_16S
C3002CSC2_BP	206619	C3002	C3002CSC2	biopsy_16S
C3002CSC3_BP	206616	C3002	C3002CSC3	biopsy_16S

```
[1] "External ID"      "Participant ID"
[3] "site_sub_coll"    "data_type"
[5] "week_num"         "date_of_receipt"
[7] "interval_days"    "visit_num"
[9] "Research Project" "PDO Number"
[11] "GSSR IDs"         "Product"
[13] "LCSET"            "Aggregated Lanes"
[15] "WR ID"            "# Lanes in Aggregation"
[17] "reads_raw"        "reads_filtered"
[19] "reads_qc_fail"    "reads_human"
```

16S

taxonomic_profiles.tsv.gz

```
[1] "Data dimensions"
```

```
[1] 982 180
```

#OTU ID	206646	224324	206619	224326
IP8BSoli	0	0	0	0
UncTepi3	0	0	0	0

#OTU ID	206646	224324	206619	224326
Unc004ii	0	0	0	0
Unc00re8	0	0	0	0
Unc018j2	1	0	0	0

Serology

hmp2_serology__Compiled_ELISA_Data.tsv

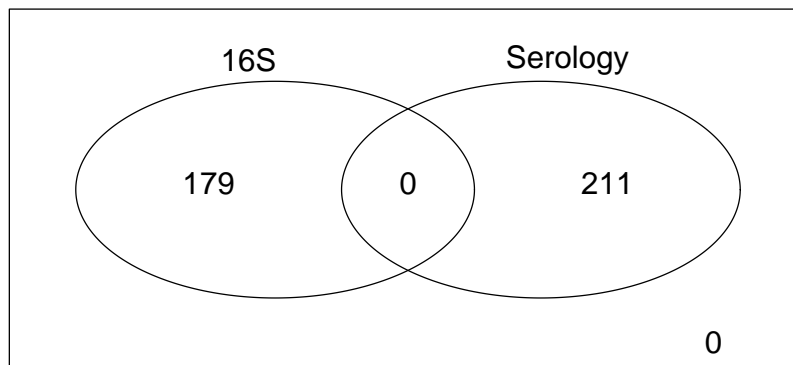
```
[1] "Data dimensions"
```

```
[1] 14 212
```

Serum ID	206454	206458	206459	206460
Site	Harvard	Harvard	Harvard	Harvard
Plate	1	1	1	1
Sample	1	2	3	4
IgA ASCA EU	0	0	0	49
IgA ASCA Pos.	0	0	0	1

Overlap between samples

Venn diagram



Metagenomes

taxonomic_profiles.tsv

[1] "Data dimensions"

[1] 1479 1639

#SampleID	CSM5FZ4M	CSM5MCUO	CSM5MCVIC	CSM5MCVN
k__Archaea	0	0	0	0
k__Archaea p__Euryarchaeota	0	0	0	0
k__Archaea p__Euryarchaeota c__Methanobacteria	0	0	0	0
k__Archaea p__Euryarchaeota c__Methanobacteria o__Methanobacteriales	0	0	0	0
k__Archaea p__Euryarchaeota c__Methanobacteria o__Methanobacteriales f__Methanobacteriaceae	0	0	0	0

pathabundances.tsv

[1] "Data dimensions"

[1] 10884 1639

# Pathway	CSM5FZ4M	CSM5MCUO	CSM5MCVIC	CSM5MCVN
1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis	0.01581	0.01017	0.01674	0.018
1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis g__Akkermansia.s__Akkermansia_muciniphila	0	0	0	0
1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis g__Bacteroides.s__Bacteroides_barnesi	0	0	0	0
1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis g__Bacteroides.s__Bacteroides_caccae	0	0.0001569	0	0
1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis g__Bacteroides.s__Bacteroides_cellulosilyticus	0	0	0	0

ecs.tsv

[1] "Data dimensions"

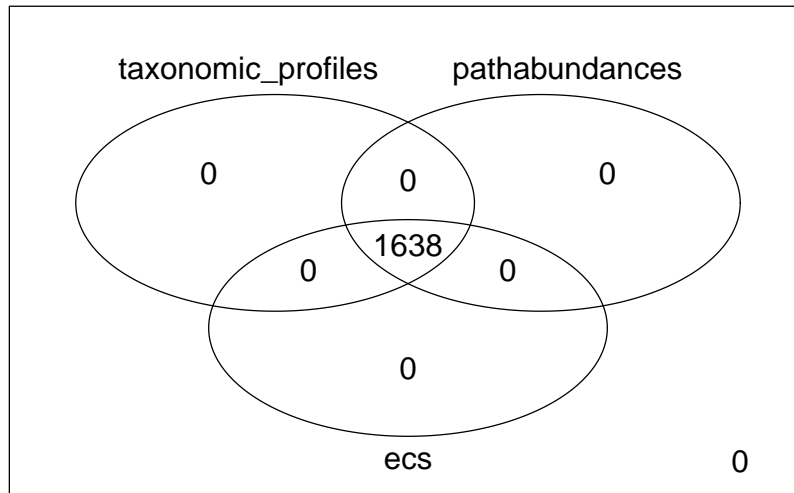
[1] 108433 1639

# Gene Family	CSM5FZ4M	CSM5MCUO	CSM5MCVIC	CSM5MCVN
1.1.1.100: 3-oxoacyl-[acyl-carrier-protein] reductase	0.001835	0.001472	0.001571	0.001173
1.1.1.100: 3-oxoacyl-[acyl-carrier-protein] reductase g__Aggregatibacter.s__Aggregatibacter_segnis	0	0	0	0
1.1.1.100: 3-oxoacyl-[acyl-carrier-protein] reductase g__Alloprevotella.s__Alloprevotella_tanner	0	0	0	0
1.1.1.100: 3-oxoacyl-[acyl-carrier-protein] reductase g__Anaerococcus.s__Anaerococcus_obesi	0	0	0	0

# Gene Family	CSM5FZ4M	CSM5MCU	CSM5MCV	CSM5MCVN
1.1.1.100: 3-oxoacyl-[acyl-carrier-protein] reductase g__Anaerococcus.s__Anaerococcus_vaginalis	0	0	0	0

Overlap between samples

Venn diagram



Proteomics

HMP2_proteomics_ecs.tsv

[1] "Data dimensions"

[1] 910 451

Gene	CSM5FZ3N	CSM5FZ3T	CSM5FZ44	CSM5FZ48
UNGROUPED	1769	1128	2765	1183
1.1.1.1: Alcohol dehydrogenase	0	0	0	0
1.1.1.100:	0	0	0	0
3-oxoacyl-[acyl-carrier-protein] reductase				
1.1.1.103: L-threonine 3-dehydrogenase	0	0	0	0
1.1.1.127:	0	0	0	0
2-dehydro-3-deoxy-D-gluconate				
5-dehydrogenase				

HMP2_proteomics_kos.tsv

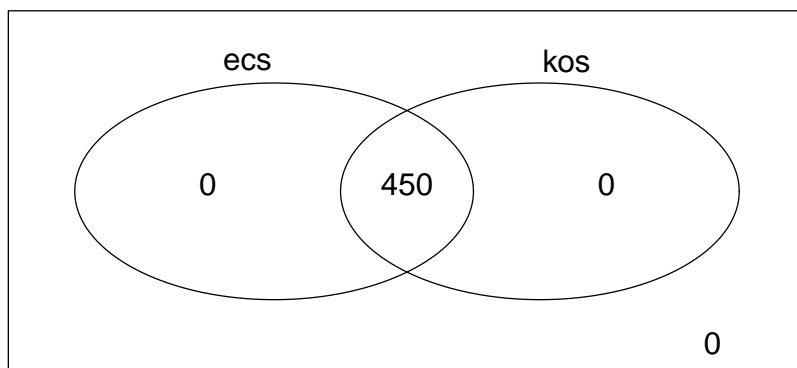
[1] "Data dimensions"

[1] 1823 451

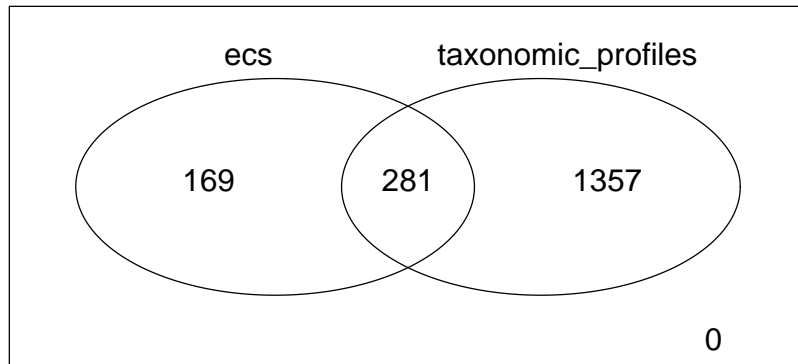
KO	CSM5FZ3N	CSM5FZ3T	CSM5FZ44	CSM5FZ48
UNGROUPED	1595	1134	2771	954
K00003: homoserine dehydrogenase [EC:1.1.1.3]	0	0	0	0
K00005: glycerol dehydrogenase [EC:1.1.1.6]	0	0	0	0
K00008: L-itol 2-dehydrogenase [EC:1.1.1.14]	0	0	0	0
K00009: mannitol-1-phosphate 5-dehydrogenase [EC:1.1.1.17]	0	0	0	0

Overlap between samples

Venn diagram



Venn diagram



Viromics

taxonomic_profiles.tsv.gz

```
[1] "Data dimensions"
```

```
[1] 982 180
```

#OTU ID	206646	224324	206619	224326
IP8BSoli	0	0	0	0
UncTepi3	0	0	0	0
Unc004ii	0	0	0	0
Unc00re8	0	0	0	0
Unc018j2	1	0	0	0

virome_virmap_analysis.tsv.gz

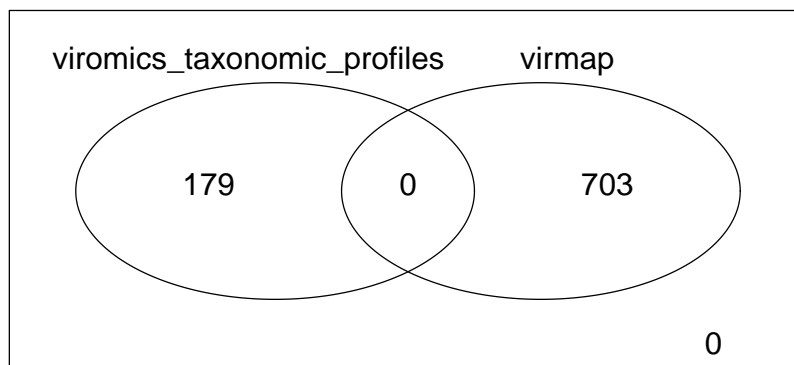
```
[1] "Data dimensions"
```

```
[1] 260 704
```

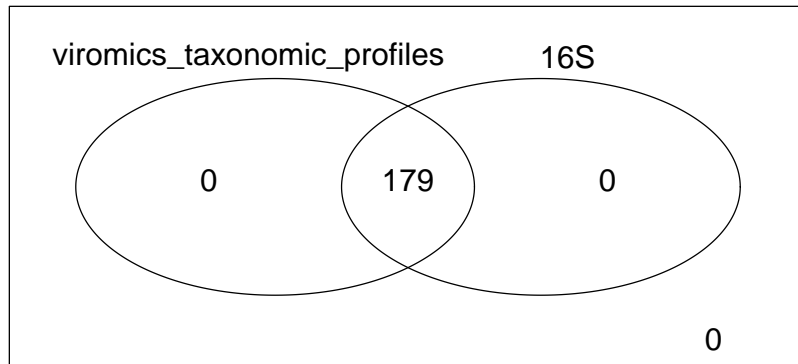
Virus	MSM5LLD	MSM5FZ4	MSM5FZB	MSM5MD5B
superkingdom=Viruses;dsDNA viruses, no RNA	0	0	0	0
stage;family=Adenoviridae;genus=Mastadenovirus;species=Human mastadenovirus A;taxId=129875				
superkingdom=Viruses;dsDNA viruses, no RNA	0	0	0	0
stage;family=Baculoviridae;genus=Alphabaculovirus;species=Autographa californica multiple nucleopolyhedrovirus;taxId=307456				
superkingdom=Viruses;dsDNA viruses, no RNA	0	0	0	0
stage;family=Baculoviridae;genus=Alphabaculovirus;taxId=558016				
superkingdom=Viruses;dsDNA viruses, no RNA	0	0	0	0
stage;order=Caudovirales;family=Myoviridae;genus=Felixo1 virus;species=Salmonella phage FelixO1;taxId=77775				
superkingdom=Viruses;dsDNA viruses, no RNA	0	0	0	0
stage;order=Caudovirales;family=Myoviridae;genus=Felixo1 virus;unclassified FelixO1likevirus;species=Escherichia phage vB_EcoM_AYO145A;taxId=1636202				

Overlap between samples

Venn diagram



Venn diagram



Metabolites

HMP2__metabolomics.csv.gz

```
[1] "Data dimensions"
```

```
[1] 81867    553
```

Method	Pooled QC sample CV	m/z	RT	HMDB (*Representative ID)
C18-neg	0.02719	313.2	9.75	HMDB04705
C18-neg	0.02641	313.2	9.95	HMDB04704
C18-neg	0.04684	115.1	5.79	HMDB00535
C18-neg	0.07408	129.1	7.43	HMDB00666
C18-neg	0.02414	149.1	5.81	HMDB00764

Metatranscriptomes

genefamilies.tsv.gz

- Large file, 2164739 x 736
- Row names like “UniRef90_A0A009DJS4” “UniRef90_A0A009DJS4|unclassified” “UniRef90_A0A009DPX5”
- Column names like “CSM5FZ4M_Abundance-RPKs” “CSM5MCUO_Abundance-RPKs” “CSM5MCVN_Abundance-RPKs”

ecs.tsv.gz

[1] "Data dimensions"

[1] 70711 736

# Gene Family	CSM5FZ4M	CSM5MCUO	CSM5MCVN	CSM5MCW6
1.1.1.100: 3-oxoacyl-[acyl-carrier-protein] reductase	0.003216	0.001697	0.002806	3.503e-05
1.1.1.100: 3-oxoacyl-[acyl-carrier-protein] reductase g__Aggregatibacter.s__Aggregatibacter_segnis	0	0	0	0
1.1.1.100: 3-oxoacyl-[acyl-carrier-protein] reductase g__Anaerostipes.s__Anaerostipes_hadrus	0	0	0	0
1.1.1.100: 3-oxoacyl-[acyl-carrier-protein] reductase g__Anaerotruncus.s__Anaerotruncus_colihominis	0	0	0	0
1.1.1.100: 3-oxoacyl-[acyl-carrier-protein] reductase g__Bacteroides.s__Bacteroides_caccae	0	0.0002605	0	0

pathabundances.tsv.gz

[1] "Data dimensions"

[1] 6061 736

# Pathway	CSM5FZ4M	CSM5MCUO	CSM5MCVN	CSM5MCW6
1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis	0.01575	0.007513	0.02524	0.01701
1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis g__Akkermansia.s__Akkermansia_muciniphila	0	0	0	0
1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis g__Bacteroides.s__Bacteroides_barnesiae	0	0	0	0
1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis g__Bacteroides.s__Bacteroides_caccae	0	0	0	0
1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis g__Bacteroides.s__Bacteroides_cellulosilyticus	0	0	0	0

Host Transcriptomes

host_tx_counts.tsv.gz

[1] "Data dimensions"

[1] 55765 254

CSM5FZ1F	CSM5FZ1G	CSM5FZ1H	CSM5FZ1I	CSM5FZ1J
5S_rRNA	0	1	2	1
7SK	76	37	47	105
A1BG	0	1	0	1

CSM5FZ1F	CSM5FZ1G	CSM5FZ1H	CSM5FZ1I	CSM5FZ1J
A1BG-AS1	31	22	14	66
A1CF	3638	2225	2360	1923