

HMP2Data: Integrative Human Microbiome Data R Bioconductor package

<https://github.com/katiasmirn/HMPDataWorkshop>

Ekaterina Smirnova, Virginia Commonwealth University

Ni Zhao, Johns Hopkins University

Mikhail Dozmorov, Virginia Commonwealth University

Levi Waldron, Graduate School of Public Health and Health Policy,
CUNY

Integrative Human Microbiome Project (iHMP)

<https://hmpdacc.org/ihmp/>



Pregnancy & Preterm Birth

The Vaginal Microbiome Consortium team at Virginia Commonwealth University is conducting the Multi-Omic Microbiome Study: Pregnancy Initiative (MOMS-PI) in



Onset of Inflammatory Bowel Disease (IBD)

The Inflammatory Bowel Disease (IBD) Multi'omics Data (IBDMDB) research team, led by Ramnik Xavier and Curtis



Onset of Type 2 Diabetes

Type 2 diabetes mellitus (T2D) is a significant health problem facing our nation. In a collaborative effort to systematically understand diabetes and its etiology, the

Novel data: first results just out of press

[MENU](#) ▾

nature
International journal of science

[Search](#)

[E-alert](#)

[Submit](#)

[Login](#)

Perspective | [OPEN](#) | Published: 29 May 2019

The Integrative Human Microbiome Project

[The Integrative HMP \(iHMP\) Research Network Consortium](#)

[Nature](#) **569**, 641–648 (2019) | [Download Citation](#) ⬇

Abstract

The NIH Human Microbiome Project (HMP) has been carried out over ten years and two phases to provide resources, methods, and discoveries that link interactions between humans and their microbiomes to health-related outcomes. The recently completed second phase, the Integrative Human Microbiome Project, comprised studies of dynamic changes in the microbiome and host under three conditions: pregnancy and preterm birth; inflammatory bowel diseases; and stressors that affect individuals with prediabetes. The associated

0
Citations

180
Altmetric

[Article metrics >>](#)

Associated Content

Special
[Human Microbiome Project, part 2](#)

[Nature](#) | [Article](#) | [OPEN](#)
[Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases](#)
Jason Lloyd-Price, Cesar Arze [...] Curtis Huttenhower

[Nature Medicine](#) | [Article](#)
[Racioethnic diversity in the dynamics of the vaginal microbiome during pregnancy](#)
Myrna G. Serrano, Hardik I. Parikh [...] Gregory A. Buck

3

Human Microbiome Project Data Portal (<https://portal.hmpdacc.org/>)

The screenshot displays the Human Microbiome Project Data Portal interface. The top navigation bar includes the HMP logo and links for Home, Studies, Data, and Analysis. User options for Login, Cart, and App are on the right.

Filters:

- Samples:**
 - Projects:
 - Integrative Human Microbiome Project: 18,024
 - Human Microbiome Project (HMP): 14,187
 - Body Site:
 - feces: 5,958
 - buccal mucosa: 4,406
 - vagina: 3,995
 - rectum: 2,813
 - blood cell: 2,615
 - Studies:
 - MOMS-PI: 11,001
 - T2D: 4,848
 - 16S-PP1: 4,536
 - WGS-PP1: 4,148
 - IBDMDB: 2,375
 - 16S-PP2: 1,884
 - Gender:
 - female: 22,030
 - male: 9,438
 - unknown: 723
- Search:** Clear, Study Name, IS, MOMS-PI, Advanced

Summary: Samples (11,001), Files (39,785)

Samples (Showing 1 - 20 of 11,001 samples)

Cart	Sample ID	Subject ID	Visit Number	Body Site	Study Name
	2341f5d5c8b48ad8e3ba519f217d30ce	EP368562	3	vagina	MOMS-PI
	2341f5d5c8b48ad8e3ba519f217d34e1	EP716673	3	vagina	MOMS-PI
	2341f5d5c8b48ad8e3ba519f217d3836	EP222037	3	vagina	MOMS-PI
	2341f5d5c8b48ad8e3ba519f217d3c94	EP824738	4	vagina	MOMS-PI
	2341f5d5c8b48ad8e3ba519f217d44ab	EP376326	3	vagina	MOMS-PI
	2341f5d5c8b48ad8e3ba519f217d54a7	EP376326	4	vagina	MOMS-PI
	2341f5d5c8b48ad8e3ba519f217d5651	EP424615	3	vagina	MOMS-PI
	2341f5d5c8b48ad8e3ba519f217d5687	EP557758	4	vagina	MOMS-PI
	2341f5d5c8b48ad8e3ba519f217d5f15	EP733790	3	vagina	MOMS-PI
	2341f5d5c8b48ad8e3ba519f217d6126	EP483881	4	vagina	MOMS-PI
	2341f5d5c8b48ad8e3ba519f217d68f5	EP521049	2	vagina	MOMS-PI
	2341f5d5c8b48ad8e3ba519f217d6dfb	EP482764	4	vagina	MOMS-PI
	2341f5d5c8b48ad8e3ba519f217d75a9	EP575820	3	vagina	MOMS-PI
	2341f5d5c8b48ad8e3ba519f217d7c0b	EP061373	4	vagina	MOMS-PI

Open data, but...

- Need **Aspera client** (or similar) to download the data
 - not every lab/researcher has these expertise
- After download still need to:
 1. map taxonomy ids from the database
 2. merge with meta data – not available on the DAC portal
 3. construct phylogenetic tree
 4. merge -omics modalities

HMP2Data package

- Motivated by HMP16SData ([Bioconductor link](#))
- Currently under review by Bioconductor
- When accepted can be installed using

```
BiocManager::install("HMP2Data")
```

Development version

Install from John Stansfield's GitHub

<https://github.com/jstansfield0/HMP2Data>

```
if(!require("HMP2Data")){  
  BiocStyle::CRANpkg("devtools")  
  devtools::install_github(  
    "jstansfield0/HMP2Data")  
}  
  
library(HMP2Data)
```

Multi-omics microbiome study pregnancy initiative (MOMS-PI)



The vaginal microbiome,
pregnancy and preterm birth

(<http://vmc.vcu.edu/momspi>)

Analysis opportunities

- Early results: term and pre-term birth
- Only a subset of samples was used
- More available through DAC portal
- Novel longitudinal and multi-omics models not explored
- Many opportunities for statistical models development and data exploration!
- 16S and cytokines data

Detailed study design:

- <https://www.nature.com/articles/s41591-019-0450-2>

Data structure

Available as:

- Matrix format for metadata, 16S, taxa table
- Phyloseq
- SummarizedExperiment for cytokines
- MultiAssayExperiment:
 - combine 16S and cytokines

Data objects

16S data as Phyloseq object

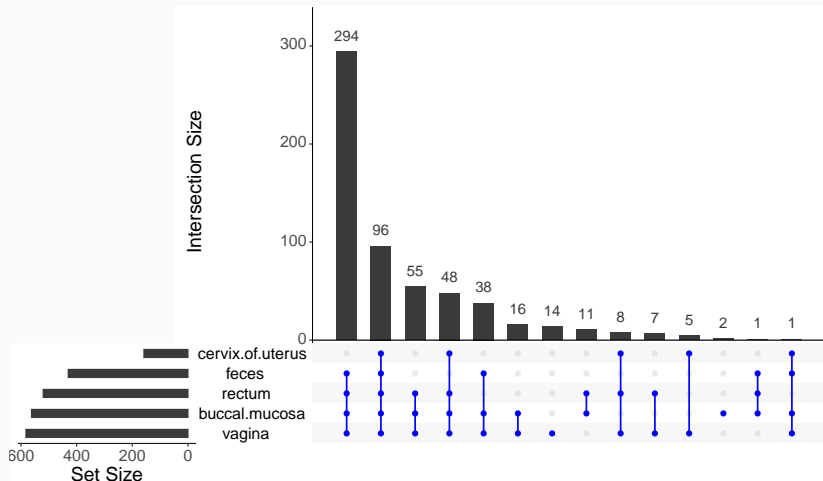
```
momspi16S_phyloseq <- momspi16S()
momspi16S_phyloseq
#> phyloseq-class experiment-level object
#> otu_table() OTU Table: [ 7665 taxa and 9107 samples ]
#> sample_data() Sample Data: [ 9107 samples by 13 sample variables ]
#> tax_table() Taxonomy Table: [ 7665 taxa by 7 taxonomic ranks ]
```

Cytokine data as SummarizedExperiment object

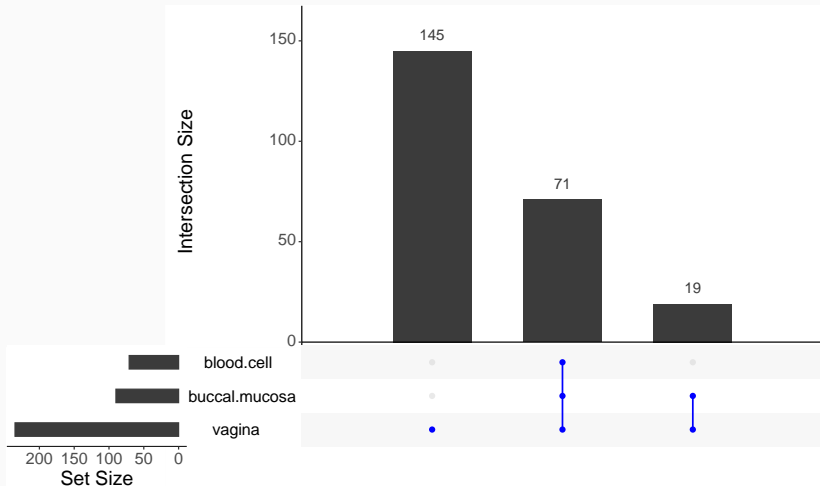
```
momspiCyto <- momspiCytokines()
momspiCyto
#> class: SummarizedExperiment
#> dim: 29 1396
#> metadata(0):
#> assays(1): ''
#> rownames(29): Eotaxin FGF ... FGF basic IL-17
#> rowData names(1): cytokine
#> colnames(1396): EP004835_K10_MVAX EP004835_K20_MVAX ...
#> EP996091_K40_MVAX EP996091_K60_MVAX
#> colData names(13): file_id md5 ... study_full_name project_name
```

	MOMS-PI 16S		MOMS-PI Cytokines		IBD 16S		T2D 16S	
	N	%	N	%	N	%	N	%
Body Site								
buccal mucosa	3313	36.38	311	22.28	0	0	0	0
cervix of uterus	162	1.78	0	0	0	0	0	0
feces	765	8.4	0	0	178	100	1041	47.15
rectum	2679	29.42	0	0	0	0	0	0
unknown	146	1.6	0	0	0	0	0	0
vagina	2042	22.42	979	70.13	0	0	0	0
blood cell	0	0	106	7.59	0	0	0	0
nasal cavity	0	0	0	0	0	0	1167	52.85
Sex								
male	0	0	0	0	84	47.19	1248	56.52
female	9107	100	1396	100	94	52.81	947	42.89
unknown	0	0	0	0	0	0	13	0.59
Race								
african american	0	0	0	0	5	2.81	117	5.3
asian	0	0	0	0	0	0	235	10.64
caucasian	0	0	0	0	164	92.13	1657	75.05
ethnic other	0	0	0	0	8	4.49	73	3.31
hispanic or latino	0	0	0	0	0	0	126	5.71
american indian or alaska native	0	0	0	0	1	0.56	0	0
unknown	9107	100	1396	100	0	0	0	0
total samples	9107	100	1396	100	178	100	2208	100

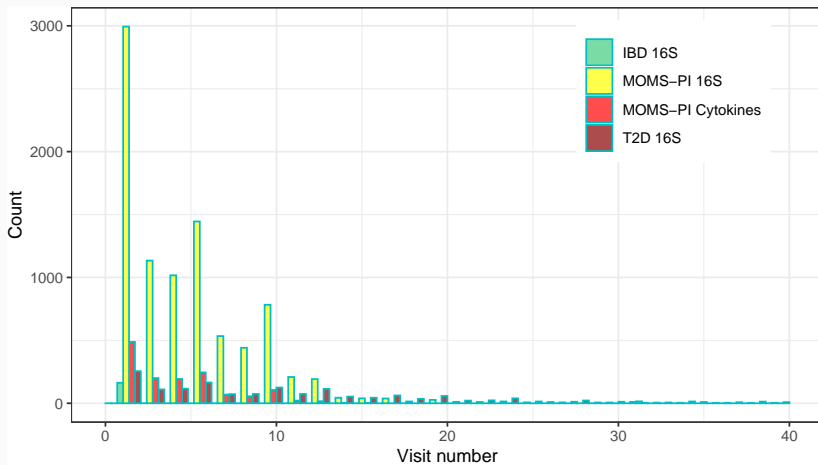
MOMS-PI 16S common samples



MOMS-PI Cytokines common samples



Multiple visits



Metadata (MOMS-PI) “dbGap”



Multi-Omic Microbiome Study-Pregnancy Initiative (MOMS-PI)

dbGaP Study Accession: phs001523.v1.p1

[Request Access](#)

Authorized Access

- **Data access provided by:** [dbGaP Authorized Access](#)
- **Release Date:** May 02, 2019
- **Embargo Release Date:** May 02, 2019
- [Data Use Certification Requirements \(DUC\)](#)
- **Use Restrictions**

Consent group	Is IRB required?	Data Access Committee	Number of participants
Disease-Specific (Pregnancy Complications, IRB, PUB, MDS) 	Yes	NICHD-DAC (HD-DAC@mail.nih.gov)	597

- [List of components](#) downloadable from [Authorized Access](#)

Publicly Available Data (Public ftp)

Connect to the [public download site](#). The site contains release notes and manifests. If available, the site also contains data dictionaries, variable summaries, documents, and truncated analyses.

Metadata (IBD): available through the package

Published results:

<https://www.nature.com/articles/s41586-019-1237-9>

```
Table1Var <- c("subject_gender", "Age.at.diagnosis", "race", "visit_number",  
              "site_name", "Education.Level", "Antibiotics", "BMI")  
  
demo_data = sample_data(IBD)[ ,c("sample_id", "subject_id", Table1Var, "diagnosis")] %>%  
  data.frame()  
  
#all <- CreateTableOne(vars = Table1Var, data = demo_data)  
#kableone(all)  
  
stratified = tableone::CreateTableOne(  
  vars = Table1Var,  
  data = summarytools::unlabel(demo_data), strata = "diagnosis", includeNA = TRUE)  
stratified <- print(stratified, printToggle = FALSE, showAllLevels = FALSE)  
stratified[,!(colnames(stratified) %in% "test")] %>%  
  knitr::kable(format = "html", caption = "Characteristics of the Study Cohort",  
              col.names = c("Crohn's disease (CD)", "nonIBD",  
                            "ulcerative colitis (UC)", "p-value")) %>%  
  kable_styling("striped", full_width = T)
```

	CD	nonIBD	UC	P-value
n	86	46	46	
subject_gender = male (%)	43 (50.0)	24 (52.2)	17 (37.0)	0.264
Age.at.diagnosis (mean (SD))	21.34 (11.31)	NaN (NA)	25.65 (15.21)	0.067
race (%)				0.002
American Indian or Alaska Native	1 (1.2)	0 (0.0)	0 (0.0)	
Black or African American	0 (0.0)	0 (0.0)	5 (10.9)	
More than one race	0 (0.0)	2 (4.3)	2 (4.3)	
Other	4 (4.7)	0 (0.0)	0 (0.0)	
White	81 (94.2)	44 (95.7)	39 (84.8)	
visit_number (mean (SD))	5.33 (10.80)	2.35 (6.39)	1.67 (4.57)	0.034
site_name (%)				<0.001
Cedars-Sinai	21 (24.4)	2 (4.3)	23 (50.0)	
Cincinnati	28 (32.6)	20 (43.5)	11 (23.9)	
MGH	24 (27.9)	18 (39.1)	6 (13.0)	
MGH Pediatrics	13 (15.1)	6 (13.0)	6 (13.0)	

Getting started: useful links

1. Bioconductor 2019 workshop materials:

http://biocworkshops2019.bioconductor.org.s3-website-us-east-1.amazonaws.com/page/MicrobiomeWorkshop__MicrobiomeWorkshop/

2. Rmarkdown files for the workshop materials

<https://github.com/waldronlab/MicrobiomeWorkshop>

3. HMP2Data package

<https://bioconductor.org/packages/devel/data/experiment/html/HMP2Data.html>

4. Slides for this presentation

<https://github.com/katiasmirn/HMPDataWorkshop>

Current and future directions

- Revise the package to include MOMS-PI dbGap data
- Already implemented in HMP16SData
- Package vignettes:
 1. HMP2Data package introduction
 2. Analysis workflow for MOMS-PI, IBD and T2D studies
 3. Multi-omics analysis combining 16S and cytokines data
- June 25: Bioconductor 2019 workshop on HMP16SData and HMP16SData packages