ERCC9 Data Analysis Tutorial

created by the Data Management and Resource Repository (DMRR)

Use Case 1: Exogenous exRNA in plasma of patients with Colorectal Cancer and Ulcerative Colitis

updated March 2018

PACIFIC NORTHWEST
RESEARCH
INSTITUTE

Data kindly provided by David Galas,
Pacific Northwest Research Institute (PNRI)

Use Case 1: Exogenous exRNA

The goal of this tutorial is to demonstrate how to

- Access the exRNA Atlas
- Select a group of datasets
- Examine pre-computed results from the exceRpt small RNA-seq analysis pipeline to answer biological questions of interest.

Wang K., Hong L., Yuan Y., Etheridge A., Zhou Y., Huang D., Wilmes P., & Galas D. (2012) The Complex Exogenous RNA Spectra in Human Plasma: An Interface with Human Gut Biota? PLoS ONE 7: e51009.

Use Case 1: Exogenous exRNA

PACIFIC NORTHWEST
RESEARCH
INSTITUTE

Background: Comparison of human plasma small RNA profiles of patients with colorectal cancer to those with ulcerative colitis indicated that a large fraction of reads were not mapping to the human genome. This finding raises the question – what is the origin of those small RNAs?

Results: Mapping suggested that a significant fraction of small RNA reads were derived from bacterial, fungal, and plant sources.

Wang K., Hong L., Yuan Y., Etheridge A., Zhou Y., Huang D., Wilmes P., & Galas D. (2012) The Complex Exogenous RNA Spectra in Human Plasma: An Interface with Human Gut Biota? PLoS ONE 7: e51009.

Use Case 1: Exogenous exRNA

We will use the exRNA Atlas and the pre-calculated results from the exceRpt small RNA-seq analysis pipeline there to answer the following questions:

1. Do all plasma small RNAs map to the human genome?
2. What are the sources of small RNAs found in human plasma that do not map to the human genome?
3. Which miRNAs are normally present in human plasma?

Wang K., Hong L., Yuan Y., Etheridge A., Zhou Y., Huang D., Wilmes P., & Galas D. (2012) The Complex Exogenous RNA Spectra in Human Plasma: An Interface with Human Gut Biota? PLoS ONE 7: e51009.

Finding datasets at the exRNA Atlas

PACIFIC NORTHWEST RESEARCH INSTITUTE

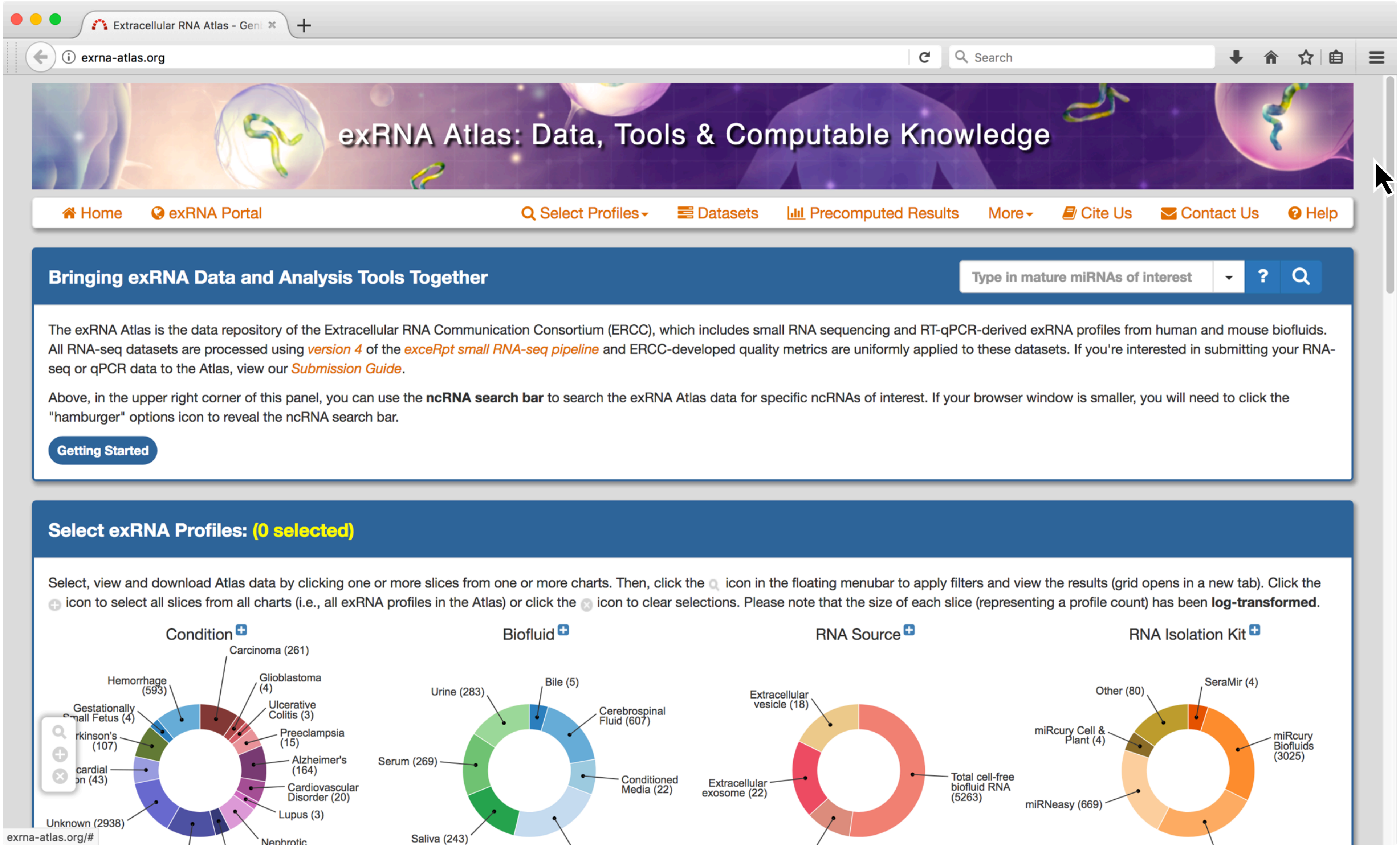
First, how can we find
the datasets from this publication
in the exRNA Atlas?

We need to find the dataset associated with this publication:

Wang K., Hong L., Yuan Y., Etheridge A., Zhou Y., Huang D., Wilmes P., & Galas D. (2012) The Complex Exogenous RNA Spectra in Human Plasma: An Interface with Human Gut Biota? PLoS ONE 7: e51009.

Finding datasets at the exRNA Atlas

1) Go to https://exRNA-Atlas.org



Bringing exRNA Data and Analysis Tools Together

The exRNA Atlas is the data repository of the Extracellular RNA Communication Consortium (ERCC), which includes small RNA sequencing and RT-qPCR-derived exRNA profiles from human and mouse biofluids. All RNA-seq datasets are processed using *version 4* of the *exceRpt small RNA-seq pipeline* and ERCC-developed quality metrics are uniformly applied to these datasets. If you're interested in submitting your RNA-seq or qPCR data to the Atlas, view our *Submission Guide*.

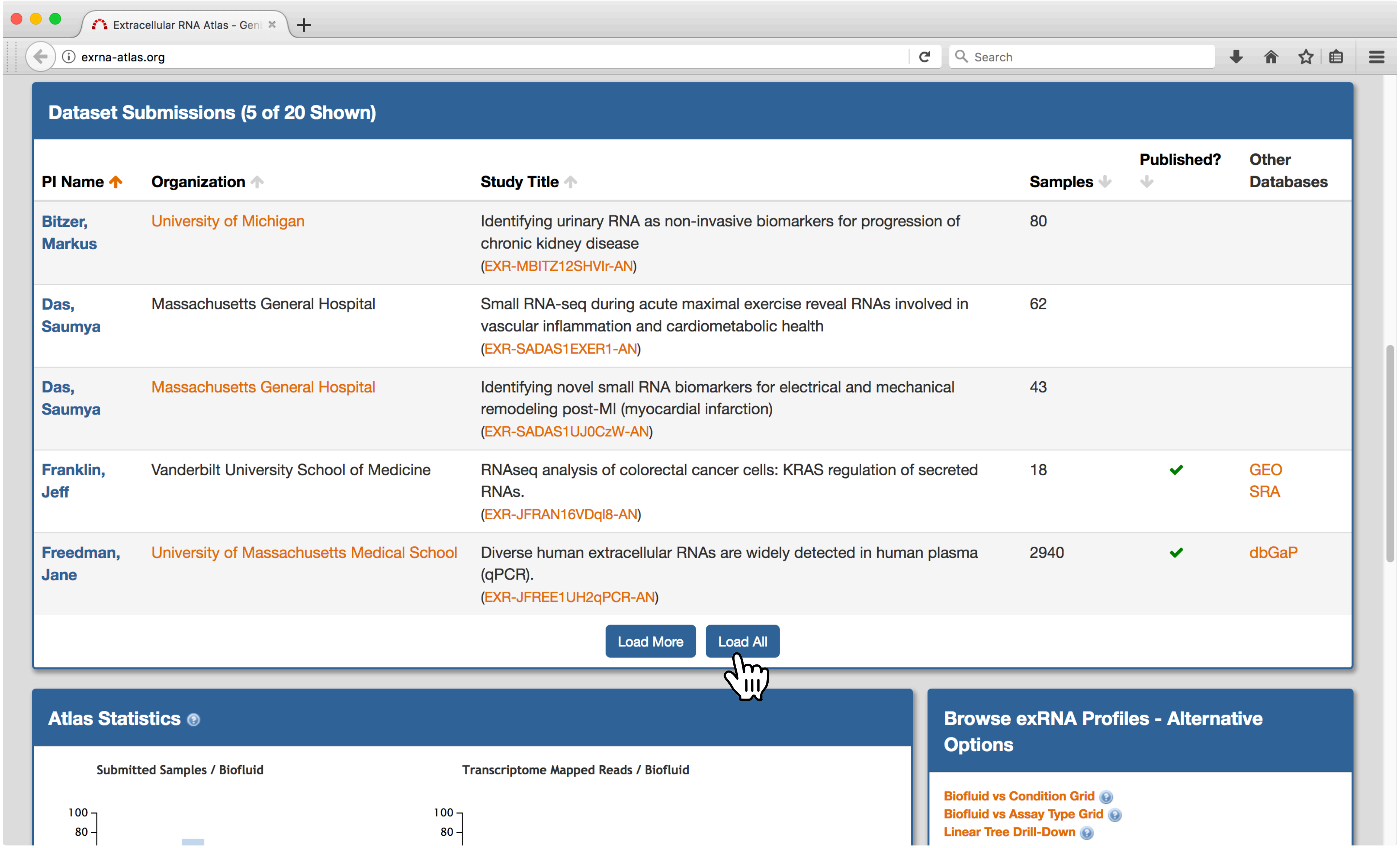
Above, in the upper right corner of this panel, you can use the **ncRNA search bar** to search the exRNA Atlas data for specific ncRNAs of interest. If your browser window is smaller, you will need to click the "hamburger" options icon to reveal the ncRNA search bar.

Select exRNA Profiles: (0 selected)

Select, view and download Atlas data by clicking one or more slices from one or more charts. Then, click the 🔍 icon in the floating menubar to apply filters and view the results (grid opens in a new tab). Click the ⊕ icon to select all slices from all charts (i.e., all exRNA profiles in the Atlas) or click the ⊗ icon to clear selections. Please note that the size of each slice (representing a profile count) has been **log-transformed**.

Finding datasets at the exRNA Atlas

2) Scroll down to **Dataset Submissions** and click **Load All**



Dataset Submissions (5 of 20 Shown)

PI Name	Organization	Study Title	Samples	Published?	Other Databases
Bitzer, Markus	University of Michigan	Identifying urinary RNA as non-invasive biomarkers for progression of chronic kidney disease (EXR-MBITZ12SHVIr-AN)	80		
Das, Saumya	Massachusetts General Hospital	Small RNA-seq during acute maximal exercise reveal RNAs involved in vascular inflammation and cardiometabolic health (EXR-SADAS1EXER1-AN)	62		
Das, Saumya	Massachusetts General Hospital	Identifying novel small RNA biomarkers for electrical and mechanical remodeling post-MI (myocardial infarction) (EXR-SADAS1UJ0CzW-AN)	43		
Franklin, Jeff	Vanderbilt University School of Medicine	RNAseq analysis of colorectal cancer cells: KRAS regulation of secreted RNAs. (EXR-JFRAN16VDql8-AN)	18	✓	GEO SRA
Freedman, Jane	University of Massachusetts Medical School	Diverse human extracellular RNAs are widely detected in human plasma (qPCR). (EXR-JFREE1UH2qPCR-AN)	2940	✓	dbGaP

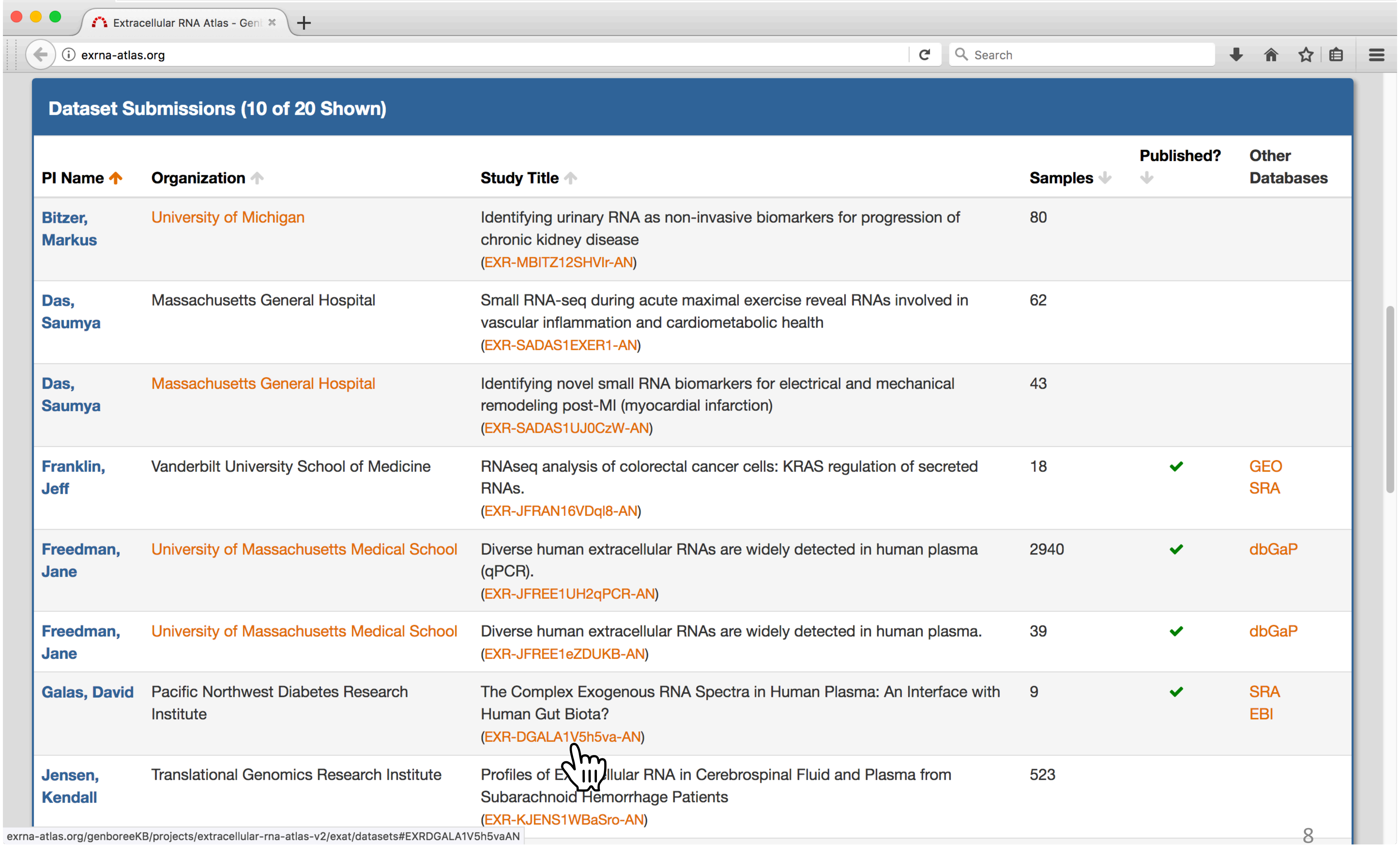
Atlas Statistics

Browse exRNA Profiles - Alternative Options

- Biofluid vs Condition Grid
- Biofluid vs Assay Type Grid
- Linear Tree Drill-Down

Finding datasets at the exRNA Atlas

3) Click on the ID number of the appropriate dataset (EXR-DGALA1V5h5va-AN)

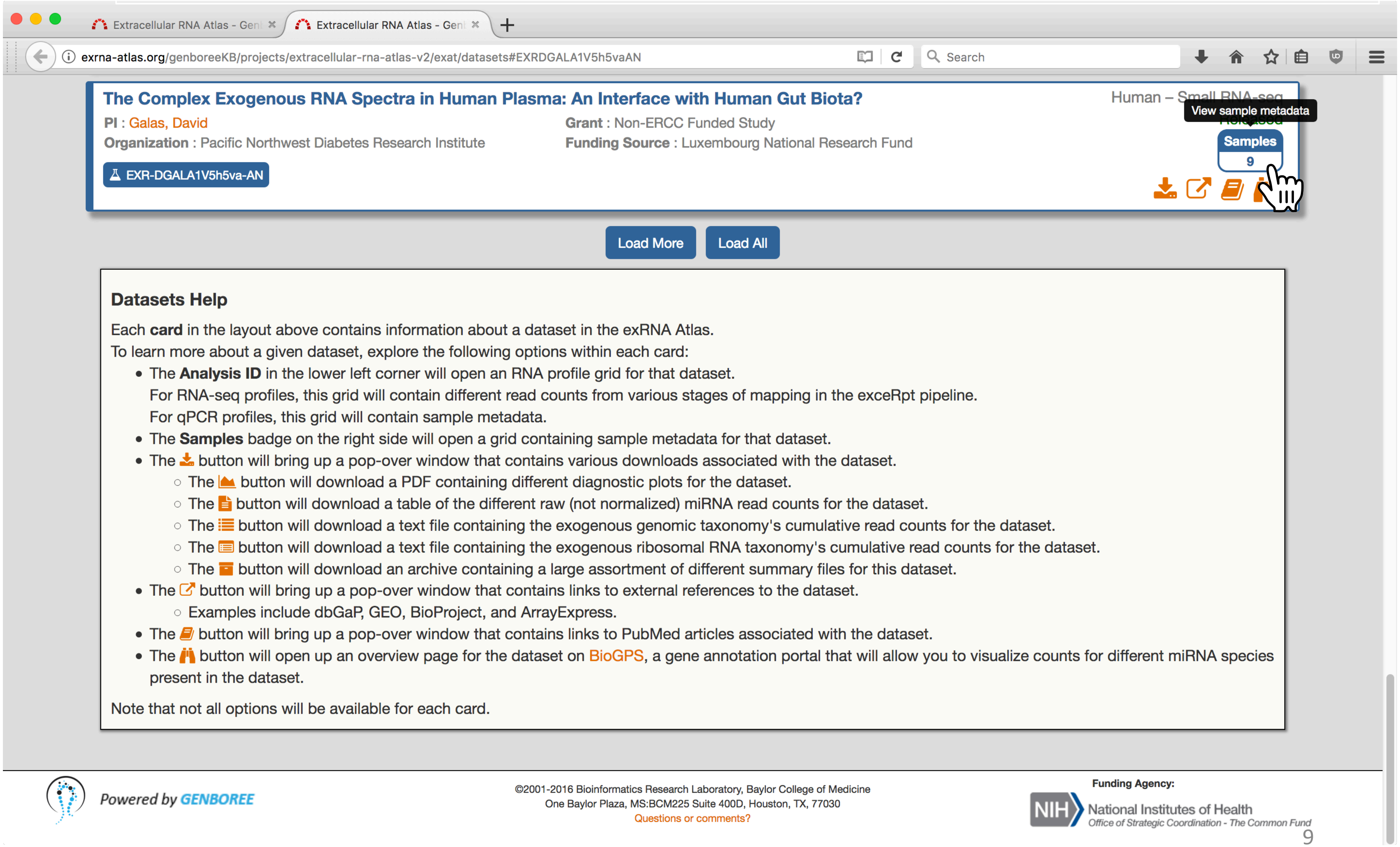


Dataset Submissions (10 of 20 Shown)

PI Name ↑	Organization ↑	Study Title ↑	Samples ↓	Published? ↓	Other Databases
Bitzer, Markus	University of Michigan	Identifying urinary RNA as non-invasive biomarkers for progression of chronic kidney disease (EXR-MBITZ12SHVIr-AN)	80		
Das, Saumya	Massachusetts General Hospital	Small RNA-seq during acute maximal exercise reveal RNAs involved in vascular inflammation and cardiometabolic health (EXR-SADAS1EXER1-AN)	62		
Das, Saumya	Massachusetts General Hospital	Identifying novel small RNA biomarkers for electrical and mechanical remodeling post-MI (myocardial infarction) (EXR-SADAS1UJ0CzW-AN)	43		
Franklin, Jeff	Vanderbilt University School of Medicine	RNAseq analysis of colorectal cancer cells: KRAS regulation of secreted RNAs. (EXR-JFRAN16VDql8-AN)	18	✓	GEO SRA
Freedman, Jane	University of Massachusetts Medical School	Diverse human extracellular RNAs are widely detected in human plasma (qPCR). (EXR-JFREE1UH2qPCR-AN)	2940	✓	dbGaP
Freedman, Jane	University of Massachusetts Medical School	Diverse human extracellular RNAs are widely detected in human plasma. (EXR-JFREE1eZDUKB-AN)	39	✓	dbGaP
Galas, David	Pacific Northwest Diabetes Research Institute	The Complex Exogenous RNA Spectra in Human Plasma: An Interface with Human Gut Biota? (EXR-DGALA1V5h5va-AN)	9	✓	SRA EBI
Jensen, Kendall	Translational Genomics Research Institute	Profiles of E[illegible]llular RNA in Cerebrospinal Fluid and Plasma from Subarachnoid Hemorrhage Patients (EXR-KJENS1WBaSro-AN)	523		

Finding datasets at the exRNA Atlas

4) Click on Samples 9 to view sample metadata



The Complex Exogenous RNA Spectra in Human Plasma: An Interface with Human Gut Biota?

Human – Small RNA-seq

PI : Galas, David

Organization : Pacific Northwest Diabetes Research Institute

Grant : Non-ERCC Funded Study

Funding Source : Luxembourg National Research Fund

EXR-DGALA1V5h5va-AN

Datasets Help

Each **card** in the layout above contains information about a dataset in the exRNA Atlas.
To learn more about a given dataset, explore the following options within each card:

- The **Analysis ID** in the lower left corner will open an RNA profile grid for that dataset.
  For RNA-seq profiles, this grid will contain different read counts from various stages of mapping in the exceRpt pipeline.
  For qPCR profiles, this grid will contain sample metadata.
- The **Samples** badge on the right side will open a grid containing sample metadata for that dataset.
- The ⬇ button will bring up a pop-over window that contains various downloads associated with the dataset.
  - The ▲ button will download a PDF containing different diagnostic plots for the dataset.
  - The 📄 button will download a table of the different raw (not normalized) miRNA read counts for the dataset.
  - The ☰ button will download a text file containing the exogenous genomic taxonomy's cumulative read counts for the dataset.
  - The ▤ button will download a text file containing the exogenous ribosomal RNA taxonomy's cumulative read counts for the dataset.
  - The ▬ button will download an archive containing a large assortment of different summary files for this dataset.
- The ↗ button will bring up a pop-over window that contains links to external references to the dataset.
  - Examples include dbGaP, GEO, BioProject, and ArrayExpress.
- The 📖 button will bring up a pop-over window that contains links to PubMed articles associated with the dataset.
- The 🔭 button will open up an overview page for the dataset on BioGPS, a gene annotation portal that will allow you to visualize counts for different miRNA species present in the dataset.

Note that not all options will be available for each card.

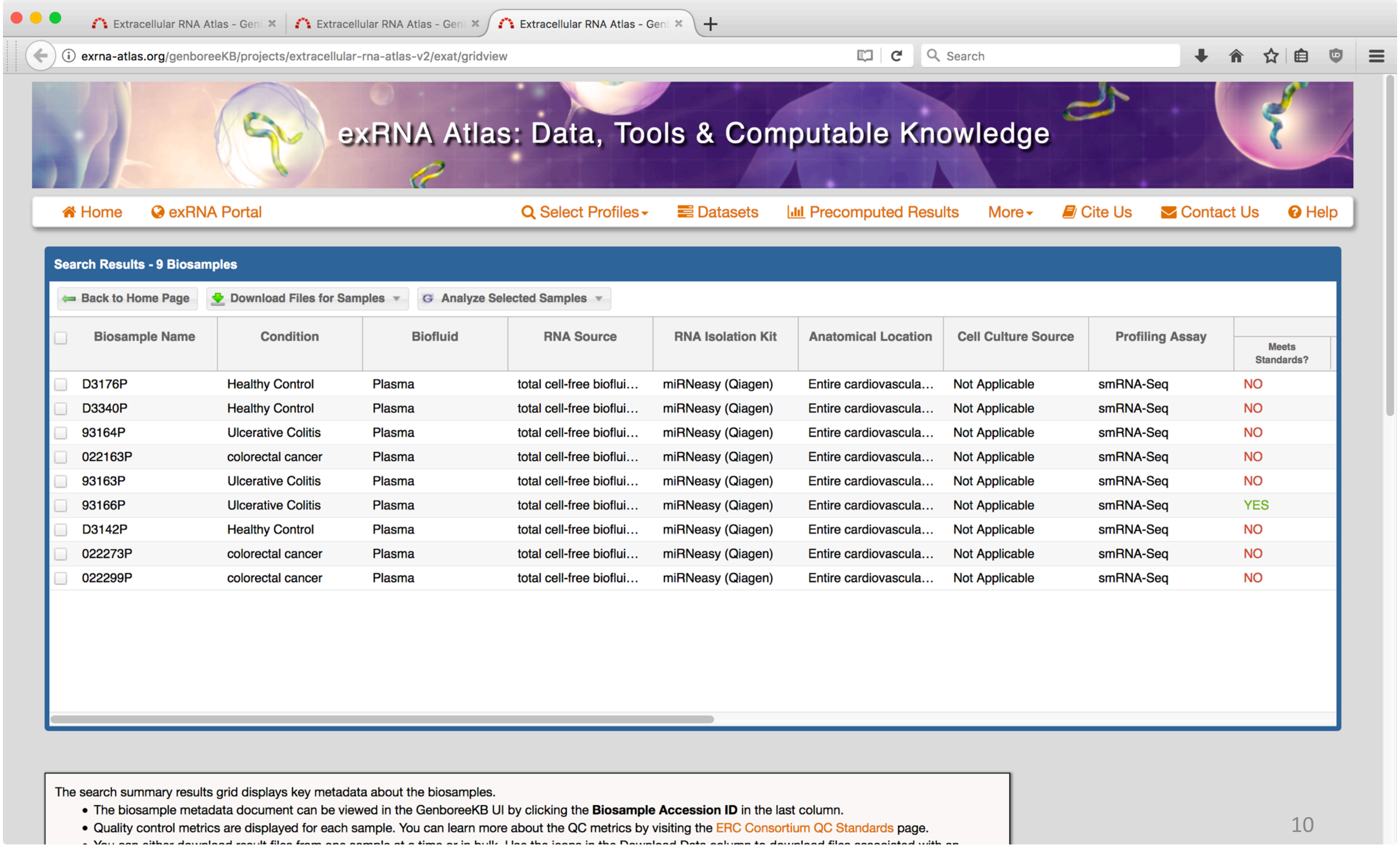
Powered by GENBOREE

©2001-2016 Bioinformatics Research Laboratory, Baylor College of Medicine
One Baylor Plaza, MS:BCM225 Suite 400D, Houston, TX, 77030
Questions or comments?

Funding Agency:
NIH National Institutes of Health
Office of Strategic Coordination - The Common Fund

Finding datasets at the exRNA Atlas

Here is the dataset (set of samples) we want to analyze.



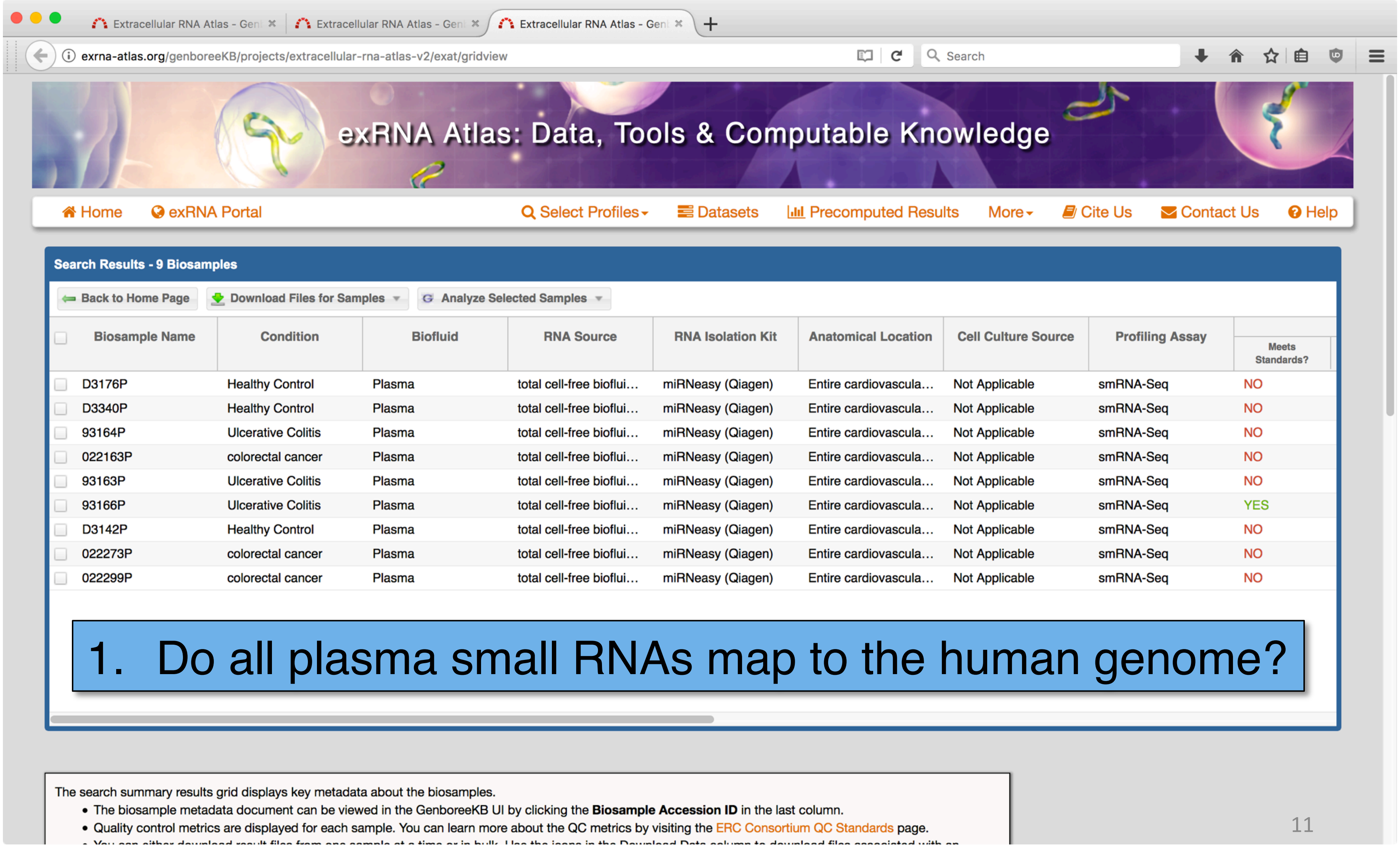
Search Results - 9 Biosamples

☐	Biosample Name	Condition	Biofluid	RNA Source	RNA Isolation Kit	Anatomical Location	Cell Culture Source	Profiling Assay	Meets Standards?
☐	D3176P	Healthy Control	Plasma	total cell-free biofluі…	miRNeasy (Qiagen)	Entire cardiovascula…	Not Applicable	smRNA-Seq	NO
☐	D3340P	Healthy Control	Plasma	total cell-free bioflui…	miRNeasy (Qiagen)	Entire cardiovascula…	Not Applicable	smRNA-Seq	NO
☐	93164P	Ulcerative Colitis	Plasma	total cell-free bioflui…	miRNeasy (Qiagen)	Entire cardiovascula…	Not Applicable	smRNA-Seq	NO
☐	022163P	colorectal cancer	Plasma	total cell-free bioflui…	miRNeasy (Qiagen)	Entire cardiovascula…	Not Applicable	smRNA-Seq	NO
☐	93163P	Ulcerative Colitis	Plasma	total cell-free bioflui…	miRNeasy (Qiagen)	Entire cardiovascula…	Not Applicable	smRNA-Seq	NO
☐	93166P	Ulcerative Colitis	Plasma	total cell-free bioflui…	miRNeasy (Qiagen)	Entire cardiovascula…	Not Applicable	smRNA-Seq	YES
☐	D3142P	Healthy Control	Plasma	total cell-free bioflui…	miRNeasy (Qiagen)	Entire cardiovascula…	Not Applicable	smRNA-Seq	NO
☐	022273P	colorectal cancer	Plasma	total cell-free bioflui…	miRNeasy (Qiagen)	Entire cardiovascula…	Not Applicable	smRNA-Seq	NO
☐	022299P	colorectal cancer	Plasma	total cell-free bioflui…	miRNeasy (Qiagen)	Entire cardiovascula…	Not Applicable	smRNA-Seq	NO

The search summary results grid displays key metadata about the biosamples.
- The biosample metadata document can be viewed in the GenboreeKB UI by clicking the **Biosample Accession ID** in the last column.
- Quality control metrics are displayed for each sample. You can learn more about the QC metrics by visiting the ERC Consortium QC Standards page.
- You can either download result files from one sample at a time or in bulk. Use the icons in the Download Data column to download files associated with an

Finding datasets at the exRNA Atlas

Here is the dataset (set of samples) we want to analyze.



Search Results - 9 Biosamples

☐	Biosample Name	Condition	Biofluid	RNA Source	RNA Isolation Kit	Anatomical Location	Cell Culture Source	Profiling Assay	Meets Standards?
☐	D3176P	Healthy Control	Plasma	total cell-free biofluі…	miRNeasy (Qiagen)	Entire cardiovascula…	Not Applicable	smRNA-Seq	NO
☐	D3340P	Healthy Control	Plasma	total cell-free biofluі…	miRNeasy (Qiagen)	Entire cardiovascula…	Not Applicable	smRNA-Seq	NO
☐	93164P	Ulcerative Colitis	Plasma	total cell-free biofluі…	miRNeasy (Qiagen)	Entire cardiovascula…	Not Applicable	smRNA-Seq	NO
☐	022163P	colorectal cancer	Plasma	total cell-free biofluі…	miRNeasy (Qiagen)	Entire cardiovascula…	Not Applicable	smRNA-Seq	NO
☐	93163P	Ulcerative Colitis	Plasma	total cell-free biofluі…	miRNeasy (Qiagen)	Entire cardiovascula…	Not Applicable	smRNA-Seq	NO
☐	93166P	Ulcerative Colitis	Plasma	total cell-free biofluі…	miRNeasy (Qiagen)	Entire cardiovascula…	Not Applicable	smRNA-Seq	YES
☐	D3142P	Healthy Control	Plasma	total cell-free biofluі…	miRNeasy (Qiagen)	Entire cardiovascula…	Not Applicable	smRNA-Seq	NO
☐	022273P	colorectal cancer	Plasma	total cell-free biofluі…	miRNeasy (Qiagen)	Entire cardiovascula…	Not Applicable	smRNA-Seq	NO
☐	022299P	colorectal cancer	Plasma	total cell-free biofluі…	miRNeasy (Qiagen)	Entire cardiovascula…	Not Applicable	smRNA-Seq	NO

1. Do all plasma small RNAs map to the human genome?

The search summary results grid displays key metadata about the biosamples.

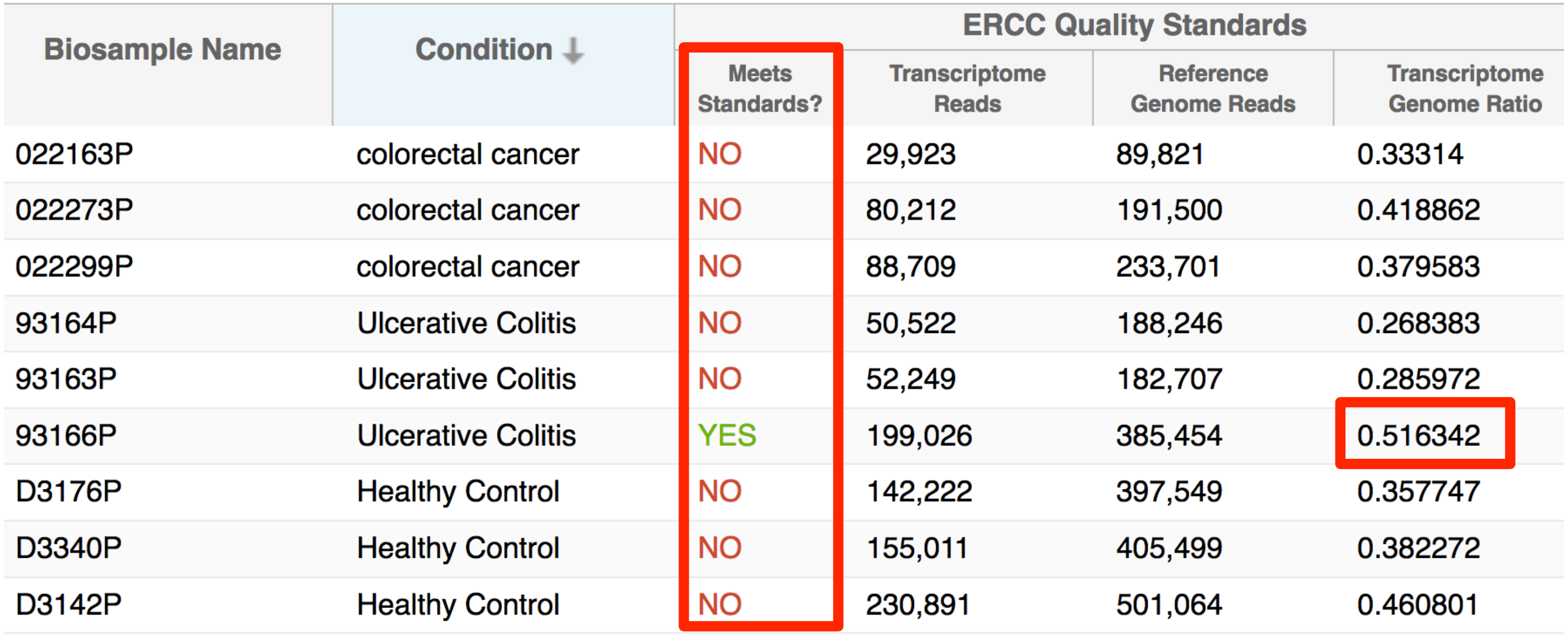
- The biosample metadata document can be viewed in the GenboreeKB UI by clicking the **Biosample Accession ID** in the last column.
- Quality control metrics are displayed for each sample. You can learn more about the QC metrics by visiting the ERC Consortium QC Standards page.
- You can either download result files from one sample at a time or in bulk. Use the icons in the Download Data column to download files associated with an

Use Case 1: Exogenous exRNA

PACIFIC NORTHWEST RESEARCH INSTITUTE

1. Do all plasma small RNAs map to the human genome?

- These plasma samples all have a high fraction of exogenous RNA, i.e. RNA that comes from organisms other than the human host.
- ERCC quality control (QC) standards require that over half the reads in a sample map to the human genome. All but one of these samples do not meet that criterion.



Biosample Name	Condition ↓	ERCC Quality Standards			
		Meets Standards?	Transcriptome Reads	Reference Genome Reads	Transcriptome Genome Ratio
022163P	colorectal cancer	NO	29,923	89,821	0.33314
022273P	colorectal cancer	NO	80,212	191,500	0.418862
022299P	colorectal cancer	NO	88,709	233,701	0.379583
93164P	Ulcerative Colitis	NO	50,522	188,246	0.268383
93163P	Ulcerative Colitis	NO	52,249	182,707	0.285972
93166P	Ulcerative Colitis	YES	199,026	385,454	0.516342
D3176P	Healthy Control	NO	142,222	397,549	0.357747
D3340P	Healthy Control	NO	155,011	405,499	0.382272
D3142P	Healthy Control	NO	230,891	501,064	0.460801

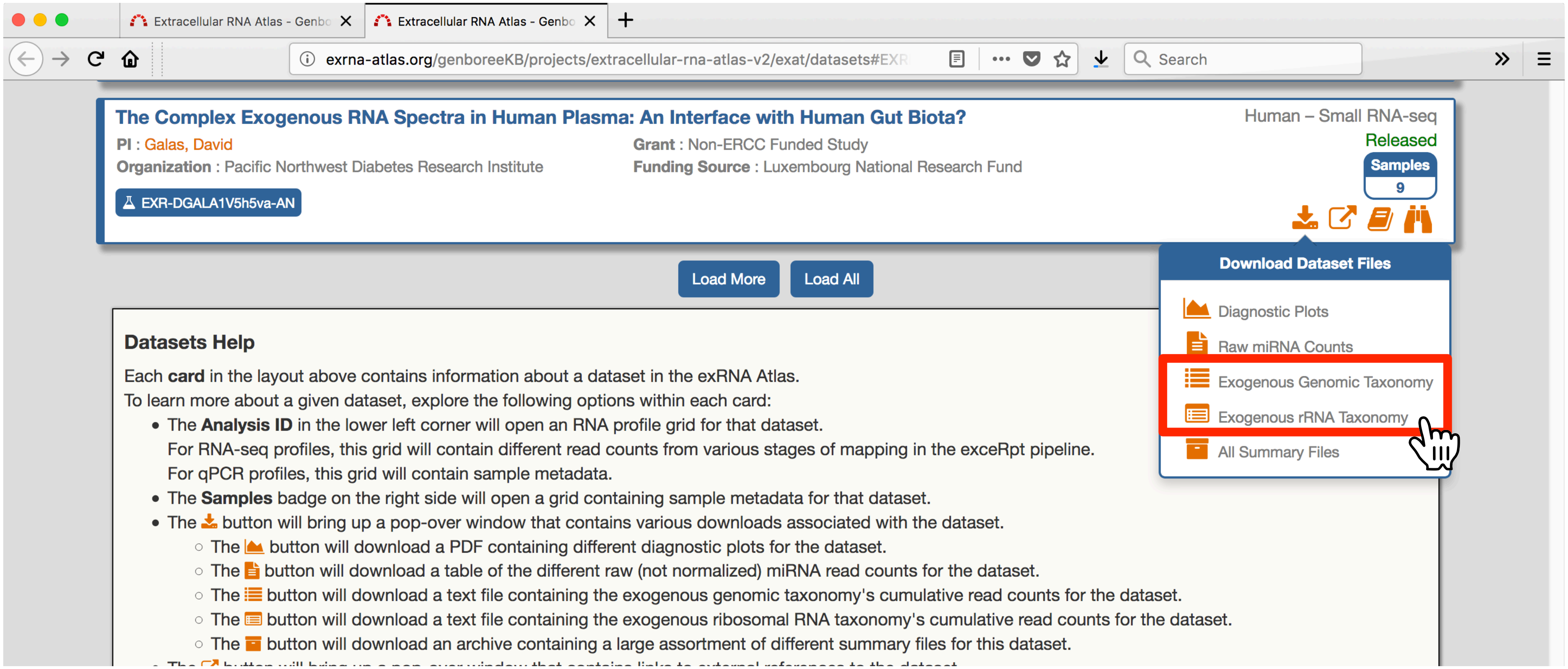
NO!

Wang K., Hong L., Yuan Y., Etheridge A., Zhou Y., Huang D., Wilmes P., & Galas D. (2012) The Complex Exogenous RNA Spectra in Human Plasma: An Interface with Human Gut Biota? PLoS ONE 7: e51009.

Use Case 1: Exogenous exRNA

2. Given that many small RNAs in these plasma samples are non-human, what is their source?

- The exceRpt small RNA-seq analysis pipeline automatically maps non-human RNAs to the genomes of all sequenced bacteria, viruses, and a wide selection of eukaryotes.
- You can download read counts of those mappings on the Dataset page.



The Complex Exogenous RNA Spectra in Human Plasma: An Interface with Human Gut Biota?

Human – Small RNA-seq

Released

PI : Galas, David

Grant : Non-ERCC Funded Study

Organization : Pacific Northwest Diabetes Research Institute

Funding Source : Luxembourg National Research Fund

EXR-DGALA1V5h5va-AN

Diagnostic Plots

Raw miRNA Counts

Exogenous Genomic Taxonomy

Exogenous rRNA Taxonomy

All Summary Files

Datasets Help

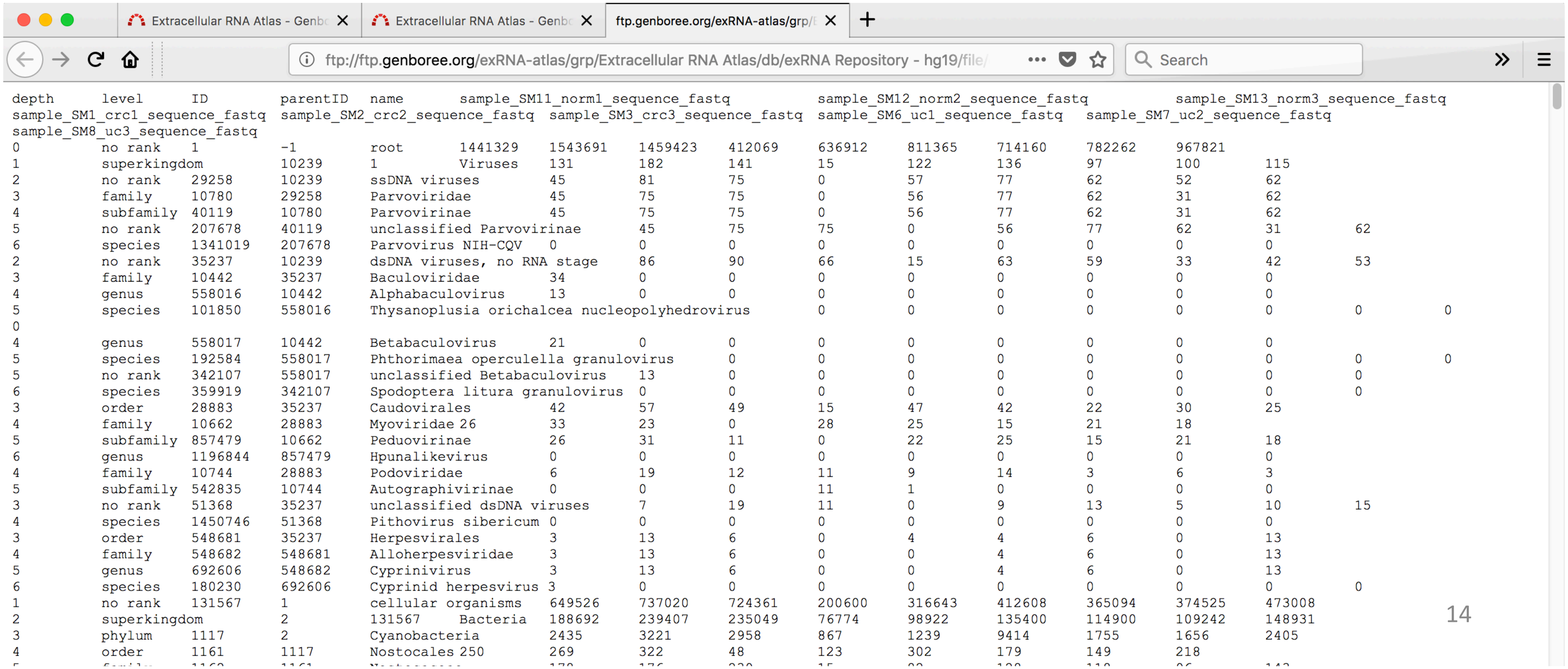
Each **card** in the layout above contains information about a dataset in the exRNA Atlas.
To learn more about a given dataset, explore the following options within each card:

- The **Analysis ID** in the lower left corner will open an RNA profile grid for that dataset.
  For RNA-seq profiles, this grid will contain different read counts from various stages of mapping in the exceRpt pipeline.
  For qPCR profiles, this grid will contain sample metadata.
- The **Samples** badge on the right side will open a grid containing sample metadata for that dataset.
- The ⬇ button will bring up a pop-over window that contains various downloads associated with the dataset.
  - The 📈 button will download a PDF containing different diagnostic plots for the dataset.
  - The 📄 button will download a table of the different raw (not normalized) miRNA read counts for the dataset.
  - The ☰ button will download a text file containing the exogenous genomic taxonomy's cumulative read counts for the dataset.
  - The ▤ button will download a text file containing the exogenous ribosomal RNA taxonomy's cumulative read counts for the dataset.
  - The 🗄 button will download an archive containing a large assortment of different summary files for this dataset.
- The ↗ button will bring up a pop-over window that contains links to external references to the dataset.

Use Case 1: Exogenous exRNA

2. Given that many small RNAs in these plasma samples are non-human, what is their source?

- You can download read counts of those mappings on the Dataset page.
- The result is a text file table in a new browser tab.



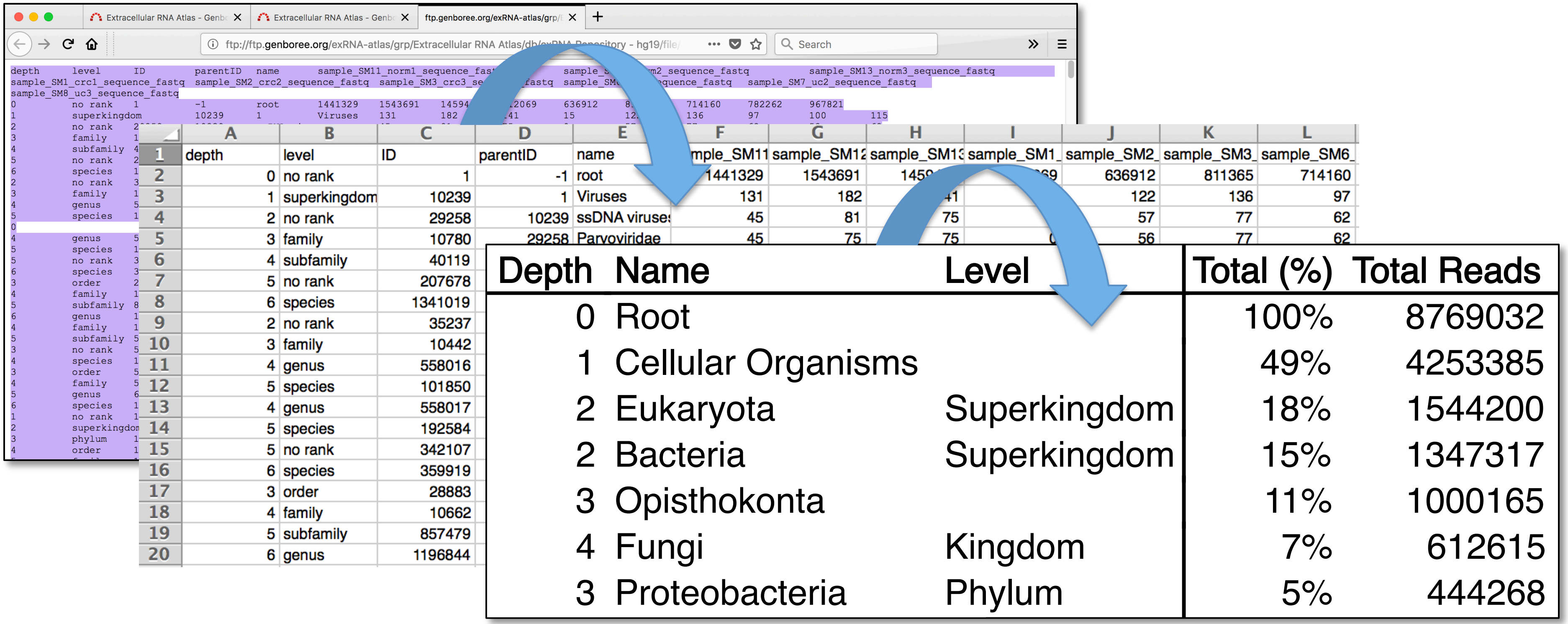
```
depth      level      ID         parentID   name       sample_SM11_norm1_sequence_fastq            sample_SM12_norm2_sequence_fastq            sample_SM13_norm3_sequence_fastq
sample_SM1_crc1_sequence_fastq   sample_SM2_crc2_sequence_fastq   sample_SM3_crc3_sequence_fastq   sample_SM6_uc1_sequence_fastq    sample_SM7_uc2_sequence_fastq
sample_SM8_uc3_sequence_fastq
0          no rank    1          -1         root       1441329    1543691    1459423    412069     636912     811365     714160     782262     967821
1          superkingdom          10239      1          Viruses    131        182        141        15         122        136        97         100        115
2          no rank    29258      10239      ssDNA viruses         45         81         75         0          57         77         62         52         62
3          family     10780      29258      Parvoviridae          45         75         75         0          56         77         62         31         62
4          subfamily  40119      10780      Parvovirinae          45         75         75         0          56         77         62         31         62
5          no rank    207678     40119      unclassified Parvovirinae        45         75         75         0          56         77         62         31         62
6          species    1341019    207678     Parvovirus NIH-CQV    0          0          0          0          0          0          0          0          0
2          no rank    35237      10239      dsDNA viruses, no RNA stage      86         90         66         15         63         59         33         42         53
3          family     10442      35237      Baculoviridae         34         0          0          0          0          0          0          0          0
4          genus      558016     10442      Alphabaculovirus      13         0          0          0          0          0          0          0          0
5          species    101850     558016     Thysanoplusia orichalcea nucleopolyhedrovirus          0          0          0          0          0          0          0          0
0
4          genus      558017     10442      Betabaculovirus       21         0          0          0          0          0          0          0          0
5          species    192584     558017     Phthorimaea operculella granulovirus        0          0          0          0          0          0          0          0          0
5          no rank    342107     558017     unclassified Betabaculovirus     13         0          0          0          0          0          0          0          0
6          species    359919     342107     Spodoptera litura granulovirus   0          0          0          0          0          0          0          0          0
3          order      28883      35237      Caudovirales          42         57         49         15         47         42         22         30         25
4          family     10662      28883      Myoviridae 26         33         23         0          28         25         15         21         18
5          subfamily  857479     10662      Peduovirinae          26         31         11         0          22         25         15         21         18
6          genus      1196844    857479     Hpunalikevirus        0          0          0          0          0          0          0          0          0
4          family     10744      28883      Podoviridae           6          19         12         11         9          14         3          6          3
5          subfamily  542835     10744      Autographivirinae     0          0          0          11         1          0          0          0          0
3          no rank    51368      35237      unclassified dsDNA viruses       7          19         11         0          9          13         5          10         15
4          species    1450746    51368      Pithovirus sibericum  0          0          0          0          0          0          0          0          0
3          order      548681     35237      Herpesvirales         3          13         6          0          4          4          6          0          13
4          family     548682     548681     Alloherpesviridae     3          13         6          0          0          4          6          0          13
5          genus      692606     548682     Cyprinivirus          3          13         6          0          0          4          6          0          13
6          species    180230     692606     Cyprinid herpesvirus  3          0          0          0          0          0          0          0          0          0
1          no rank    131567     1          cellular organisms    649526     737020     724361     200600     316643     412608     365094     374525     473008
2          superkingdom          2          131567     Bacteria   188692     239407     235049     76774      98922      135400     114900     109242     148931
3          phylum     1117       2          Cyanobacteria         2435       3221       2958       867        1239       9414       1755       1656       2405
4          order      1161       1117       Nostocales 250        269        322        48         123        302        179        149        218
```

Use Case 1: Exogenous exRNA

PACIFIC NORTHWEST RESEARCH INSTITUTE

2. Given that many small RNAs in these plasma samples are non-human, what is their source?

- Copy and paste that read count table into Excel (Paste Special – as Text).
- Then sum the read counts for all the samples and filter to show only the top levels of the tree with nodes where more than 5% of the reads map.



	A	B	C	D	E	F	G
1	depth	level	ID	parentID	name	mple_SM11	sample_SM12
2	0	no rank	1	-1	root	1441329	1543691
3	1	superkingdom	10239	1	Viruses	131	182
4	2	no rank	29258	10239	ssDNA viruses	45	81
5	3	family	10780	29258	Parvoviridae	45	75
6	4	subfamily	40119				
7	5	no rank	207678				
8	6	species	1341019				
9	2	no rank	35237				
10	3	family	10442				
11	4	genus	558016				
12	5	species	101850				
13	4	genus	558017				
14	5	species	192584				
15	5	no rank	342107				
16	6	species	359919				
17	3	order	28883				
18	4	family	10662				
19	5	subfamily	857479				
20	6	genus	1196844				

Depth	Name	Level	Total (%)	Total Reads
0	Root		100%	8769032
1	Cellular Organisms		49%	4253385
2	Eukaryota	Superkingdom	18%	1544200
2	Bacteria	Superkingdom	15%	1347317
3	Opisthokonta		11%	1000165
4	Fungi	Kingdom	7%	612615
3	Proteobacteria	Phylum	5%	444268

Use Case 1: Exogenous exRNA

PACIFIC NORTHWEST RESEARCH INSTITUTE

2. Given that many small RNAs in these plasma samples are non-human, what is their source?

The main sources of exogenous small RNA are bacteria (15%) – specifically proteobacteria (5%) – and fungi (7%).

Depth	Name	Level	Total (%)	Total Reads
0	Root		100%	8769032
1	Cellular Organisms		49%	4253385
2	Eukaryota	Superkingdom	18%	1544200
2	Bacteria	Superkingdom	15%	1347317
3	Opisthokonta		11%	1000165
4	Fungi	Kingdom	7%	612615
3	Proteobacteria	Phylum	5%	444268

Use Case 1: Exogenous exRNA

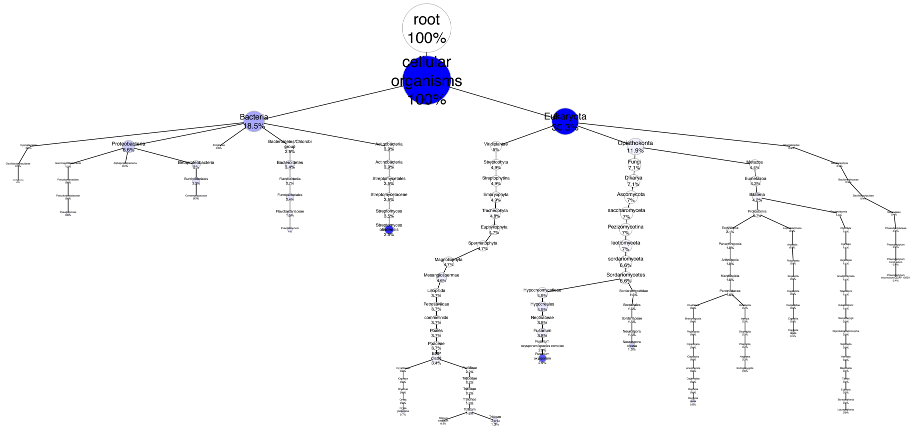
PACIFIC NORTHWEST RESEARCH INSTITUTE

2. Given that many small RNAs in these plasma samples are non-human, what is their source?

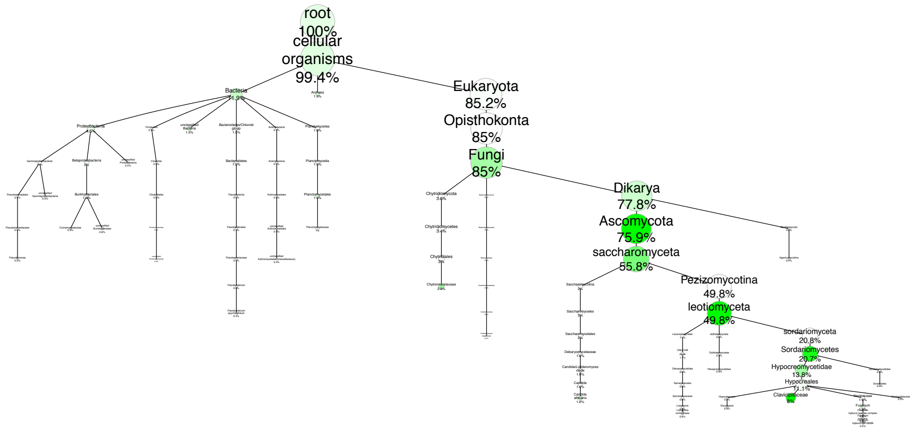
Exogenous phylogenetic trees

- The exceRpt pipeline generates visualizations of the phylogenetic tree with reads mapped (a) to exogenous genomes, or (b) to ribosomal RNAs.

exogenous genomes

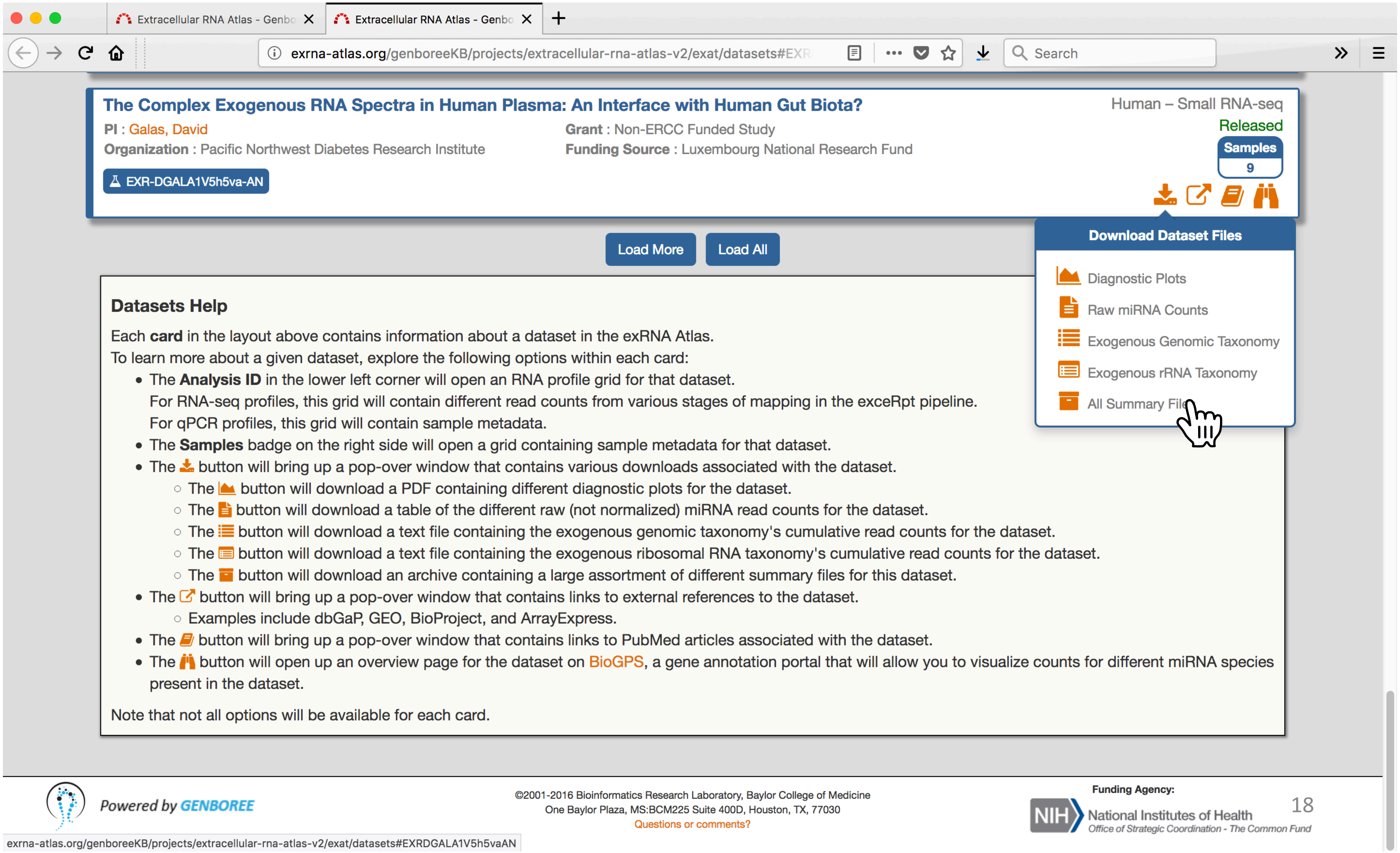


ribosomal RNA



Exogenous phylogenetic trees

- To see the phylogenetic tree visualizations, (1) click on ▬ All Summary Files on the Dataset page to download an archive of all the results of the exceRpt pipeline.



The Complex Exogenous RNA Spectra in Human Plasma: An Interface with Human Gut Biota?

Human – Small RNA-seq

Released

Samples 9

PI : Galas, David

Organization : Pacific Northwest Diabetes Research Institute

Grant : Non-ERCC Funded Study

Funding Source : Luxembourg National Research Fund

Datasets Help

Each **card** in the layout above contains information about a dataset in the exRNA Atlas.
To learn more about a given dataset, explore the following options within each card:

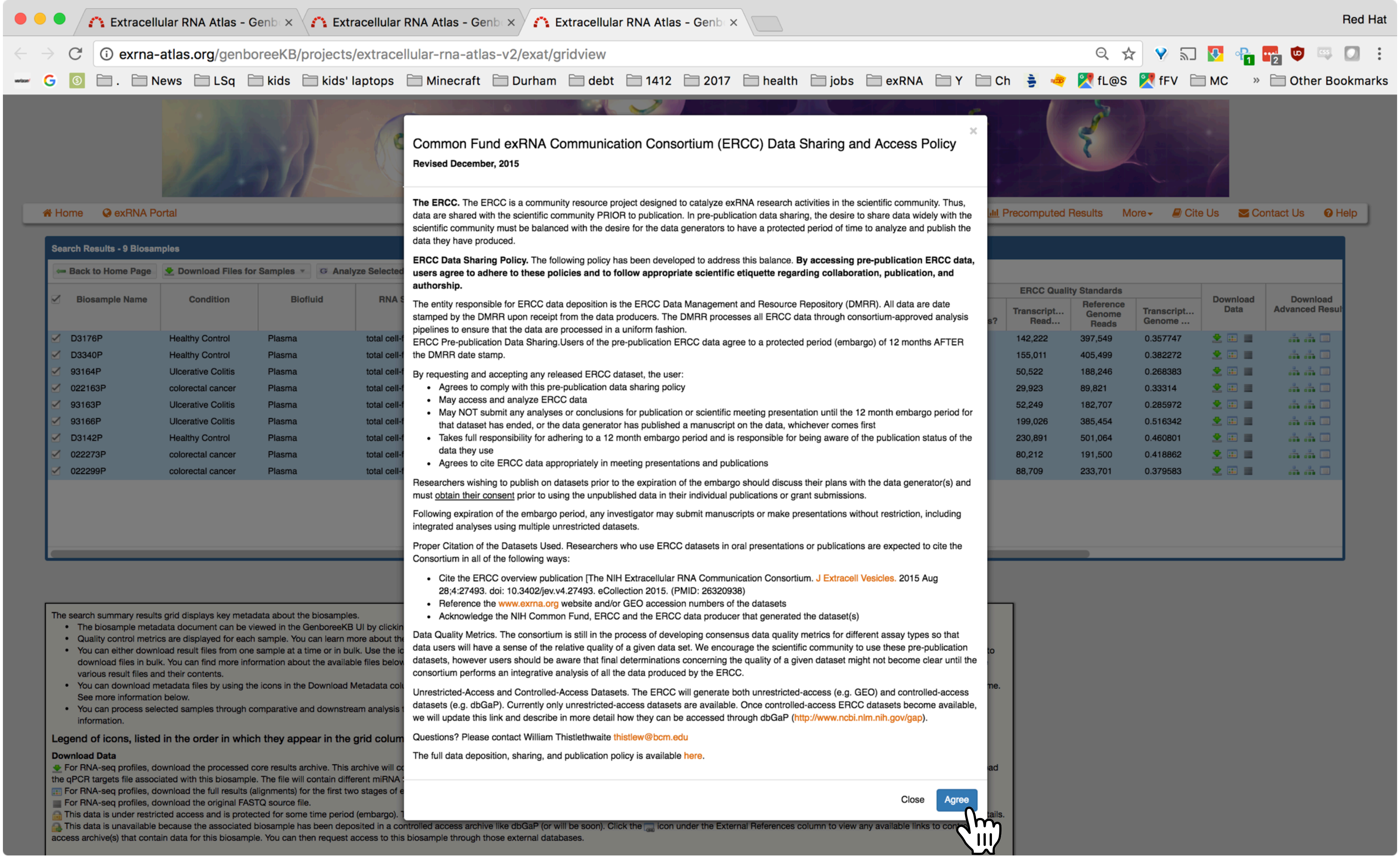
- The **Analysis ID** in the lower left corner will open an RNA profile grid for that dataset.
  For RNA-seq profiles, this grid will contain different read counts from various stages of mapping in the exceRpt pipeline.
  For qPCR profiles, this grid will contain sample metadata.
- The **Samples** badge on the right side will open a grid containing sample metadata for that dataset.
- The ⤓ button will bring up a pop-over window that contains various downloads associated with the dataset.
  - The ▲ button will download a PDF containing different diagnostic plots for the dataset.
  - The 🗎 button will download a table of the different raw (not normalized) miRNA read counts for the dataset.
  - The ☰ button will download a text file containing the exogenous genomic taxonomy's cumulative read counts for the dataset.
  - The ▤ button will download a text file containing the exogenous ribosomal RNA taxonomy's cumulative read counts for the dataset.
  - The ▬ button will download an archive containing a large assortment of different summary files for this dataset.
- The ↗ button will bring up a pop-over window that contains links to external references to the dataset.
  - Examples include dbGaP, GEO, BioProject, and ArrayExpress.
- The 📖 button will bring up a pop-over window that contains links to PubMed articles associated with the dataset.
- The 🔭 button will open up an overview page for the dataset on BioGPS, a gene annotation portal that will allow you to visualize counts for different miRNA species present in the dataset.

Note that not all options will be available for each card.

©2001-2016 Bioinformatics Research Laboratory, Baylor College of Medicine
One Baylor Plaza, MS:BCM225 Suite 400D, Houston, TX, 77030
Questions or comments?

Exogenous phylogenetic trees

2) Agree to the ERCC Data Sharing and Access Policy.



Common Fund exRNA Communication Consortium (ERCC) Data Sharing and Access Policy

Revised December, 2015

The ERCC. The ERCC is a community resource project designed to catalyze exRNA research activities in the scientific community. Thus, data are shared with the scientific community PRIOR to publication. In pre-publication data sharing, the desire to share data widely with the scientific community must be balanced with the desire for the data generators to have a protected period of time to analyze and publish the data they have produced.

ERCC Data Sharing Policy. The following policy has been developed to address this balance. **By accessing pre-publication ERCC data, users agree to adhere to these policies and to follow appropriate scientific etiquette regarding collaboration, publication, and authorship.**

The entity responsible for ERCC data deposition is the ERCC Data Management and Resource Repository (DMRR). All data are date stamped by the DMRR upon receipt from the data producers. The DMRR processes all ERCC data through consortium-approved analysis pipelines to ensure that the data are processed in a uniform fashion.
ERCC Pre-publication Data Sharing.Users of the pre-publication ERCC data agree to a protected period (embargo) of 12 months AFTER the DMRR date stamp.

By requesting and accepting any released ERCC dataset, the user:
- Agrees to comply with this pre-publication data sharing policy
- May access and analyze ERCC data
- May NOT submit any analyses or conclusions for publication or scientific meeting presentation until the 12 month embargo period for that dataset has ended, or the data generator has published a manuscript on the data, whichever comes first
- Takes full responsibility for adhering to a 12 month embargo period and is responsible for being aware of the publication status of the data they use
- Agrees to cite ERCC data appropriately in meeting presentations and publications

Researchers wishing to publish on datasets prior to the expiration of the embargo should discuss their plans with the data generator(s) and must obtain their consent prior to using the unpublished data in their individual publications or grant submissions.

Following expiration of the embargo period, any investigator may submit manuscripts or make presentations without restriction, including integrated analyses using multiple unrestricted datasets.

Proper Citation of the Datasets Used. Researchers who use ERCC datasets in oral presentations or publications are expected to cite the Consortium in all of the following ways:

- Cite the ERCC overview publication [The NIH Extracellular RNA Communication Consortium. J Extracell Vesicles. 2015 Aug 28;4:27493. doi: 10.3402/jev.v4.27493. eCollection 2015. (PMID: 26320938)
- Reference the www.exrna.org website and/or GEO accession numbers of the datasets
- Acknowledge the NIH Common Fund, ERCC and the ERCC data producer that generated the dataset(s)

Data Quality Metrics. The consortium is still in the process of developing consensus data quality metrics for different assay types so that data users will have a sense of the relative quality of a given data set. We encourage the scientific community to use these pre-publication datasets, however users should be aware that final determinations concerning the quality of a given dataset might not become clear until the consortium performs an integrative analysis of all the data produced by the ERCC.

Unrestricted-Access and Controlled-Access Datasets. The ERCC will generate both unrestricted-access (e.g. GEO) and controlled-access datasets (e.g. dbGaP). Currently only unrestricted-access datasets are available. Once controlled-access ERCC datasets become available, we will update this link and describe in more detail how they can be accessed through dbGaP (http://www.ncbi.nlm.nih.gov/gap).

Questions? Please contact William Thistlethwaite thistlew@bcm.edu

The full data deposition, sharing, and publication policy is available here.

Exogenous phylogenetic trees

The result is an archive file with the name template

<dataset>-<date-processed>_exceRpt_postProcessedResults_<exceRpt-version>.tgz

In this case:

DGALA1-gut-plasma-**2016-10-21**_exceRpt_postProcessedResults_**v4.6.3**.tgz

Opening the archive file (the app to use varies with your operating system) generates a large set of files, among which are phylogenetic trees, for all samples in the dataset in aggregate as well as for each sample, visualizing

1) all reads mapped to exogenous genomes
    name template: **<dataset>**-**<date-processed>**_exceRpt_exogenousGenomes_TaxonomyTrees_
        aggregateSamples.pdf
        perSample.pdf

2) all reads mapped to exogenous ribosomal RNA
    name template: **<dataset>**-**<date-processed>**_exceRpt_exogenousRibosomal_TaxonomyTrees_
        aggregateSamples.pdf
        perSample.pdf

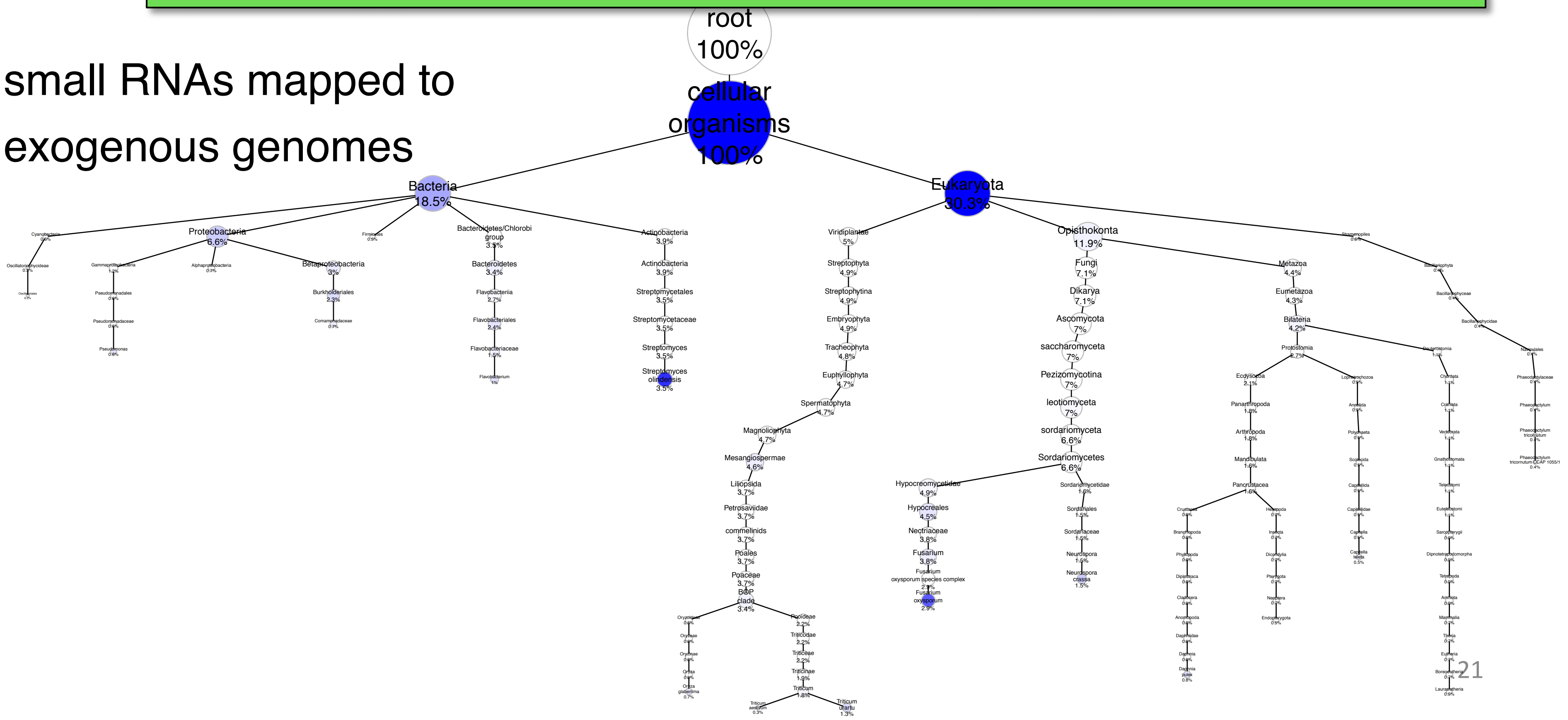
The following slides show **DGALA1-gut-plasma**-2016-10-21_
exceRpt_exogenousGenomes_TaxonomyTrees_aggregateSamples.pdf

Exogenous phylogenetic trees

2. Given that many small RNAs in these plasma samples are non-human, what is their source?

The main sources of exogenous small RNA are bacteria (**18%**) – specifically proteobacteria (**7%**) – and fungi (7%).

small RNAs mapped to exogenous genomes

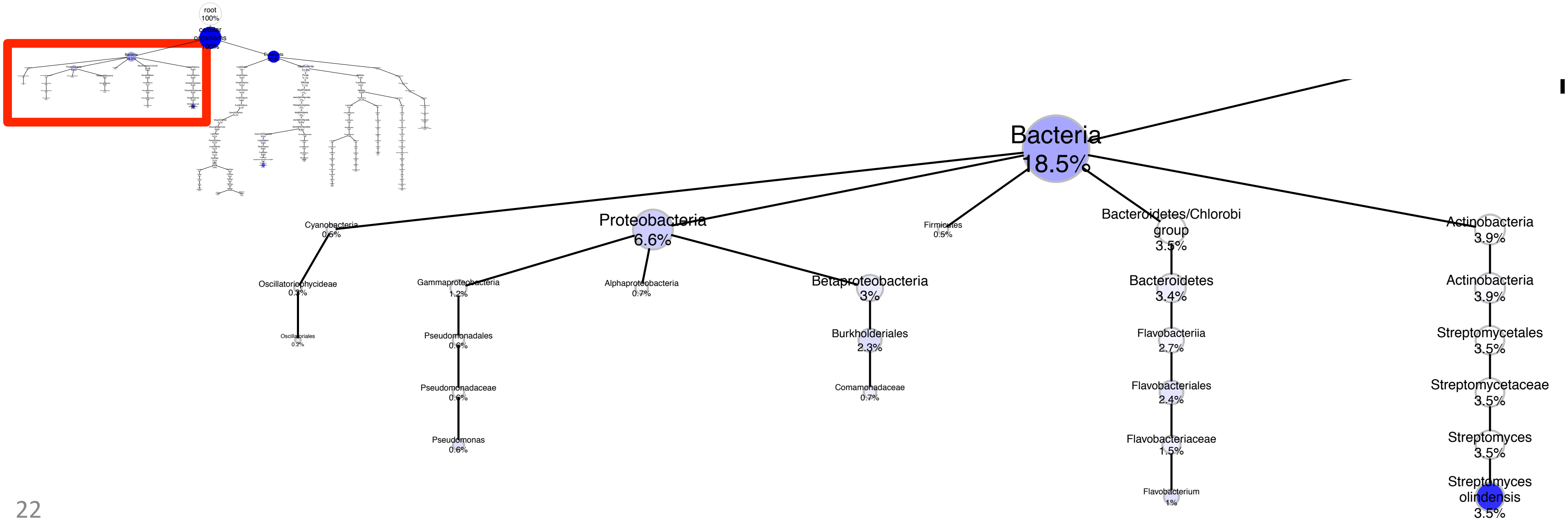


Exogenous phylogenetic trees

2. Given that many small RNAs in these plasma samples are non-human, what is their source?

The main sources of exogenous small RNA are bacteria (**18%**) – specifically proteobacteria (**7%**) – and fungi (7%).

small RNAs mapped to exogenous genomes

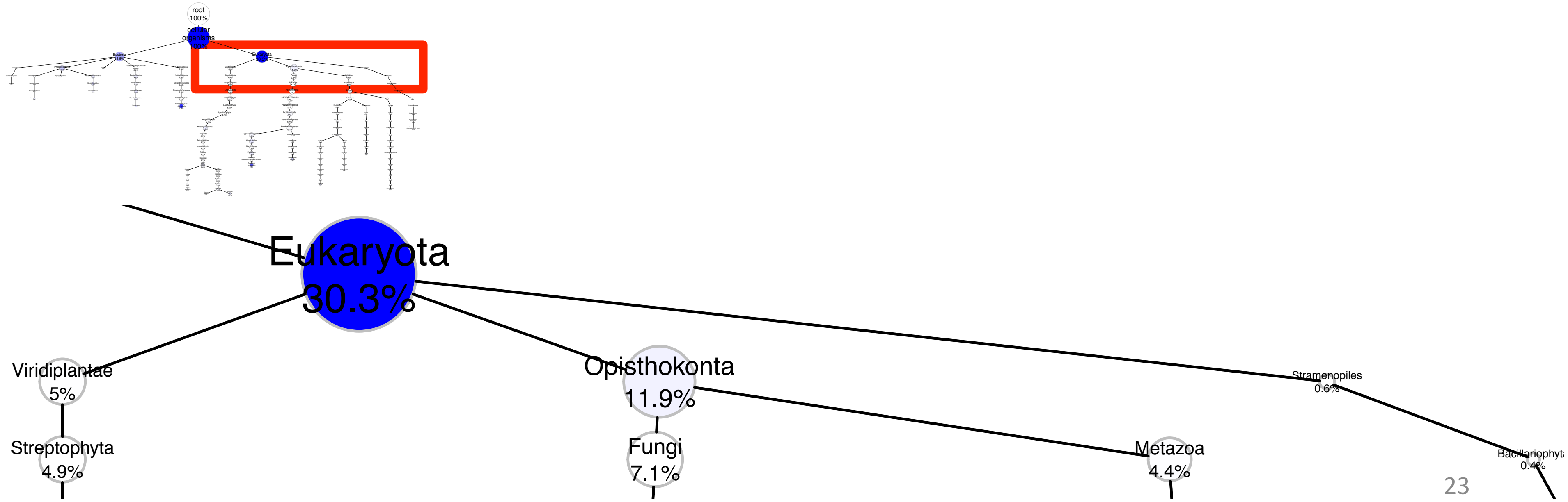


Exogenous phylogenetic trees

2. Given that many small RNAs in these plasma samples are non-human, what is their source?

The main sources of exogenous small RNA are bacteria (**18%**) – specifically proteobacteria (**7%**) – and fungi (7%).

small RNAs mapped to exogenous genomes



Exogenous phylogenetic trees

PACIFIC NORTHWEST
RESEARCH
INSTITUTE

2. Given that many small RNAs in these plasma samples are non-human, what is their source?

The main sources of exogenous small RNA are bacteria (**18%**) – specifically proteobacteria (**7%**) – and fungi (7%).

- Why do the numbers in the tree images differ from those in the table?

Depth	Name	Level	Total (%)	Total Reads
0	Root		100%	8769032
1	Cellular Organisms		49%	4253385
2	Eukaryota	Superkingdom	18%	1544200
2	Bacteria	Superkingdom	15%	1347317
3	Opisthokonta		11%	1000165
4	Fungi	Kingdom	7%	612615
3	Proteobacteria	Phylum	5%	444268

Exogenous phylogenetic trees

The main sources of exogenous small RNA are bacteria (18%) – specifically proteobacteria (7%) – and fungi (7%).

- Why do the numbers in the tree images differ from those in the table?
- The tree images are built by summing read counts from each specific node with the cumulative number of reads mapped to nodes below it in the tree.
- The tables downloaded from the Datasets pull-down are cumulative read counts.
- Summing the cumulative and specific read counts from the full exceRpt results generates the same numbers seen in the trees.

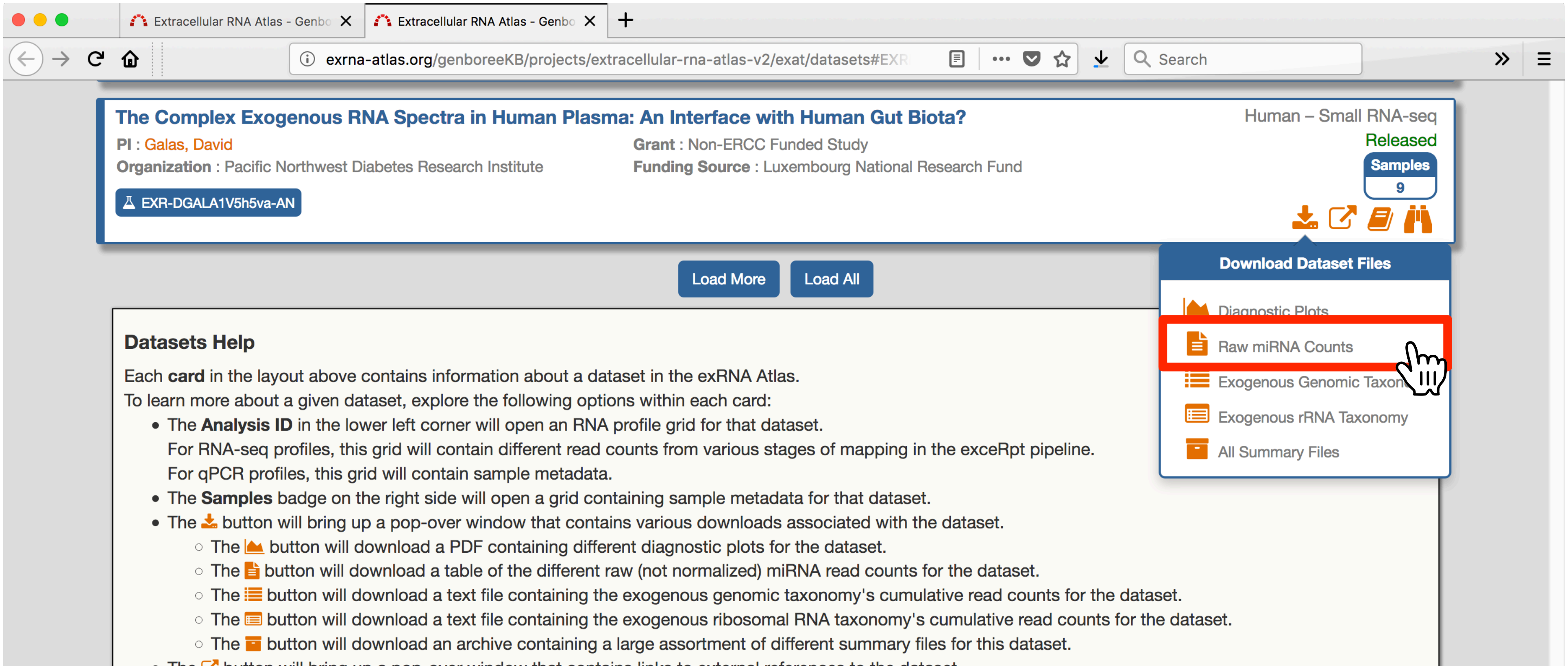
			Cumulative + Specific		Cumulative		Node-specific
Depth	Name	Level	Total (%)	Reads	Total (%)	Reads	Reads
0	Root		100%	8769067	100%	8769032	35
1	Cellular Organisms		100%	8767993	49%	4253385	4514608
2	Eukaryota	Superkingdom	30%	2660137	18%	1544200	1115937
2	Bacteria	Superkingdom	18%	1590635	15%	1347317	243318
3	Opisthokonta		12%	1036156	11%	1000165	35991
4	Fungi	Kingdom	7%	612615	7%	612615	0
3	Proteobacteria	Phylum	7%	572707	5%	444268	128439

Use Case 1: Exogenous exRNA

PACIFIC NORTHWEST RESEARCH INSTITUTE

3. Which miRNAs are normally present in human plasma?

- We can look for the answer to this question by downloading **Raw miRNA read counts** from the Dataset Files pull-down.



The Complex Exogenous RNA Spectra in Human Plasma: An Interface with Human Gut Biota?

Human – Small RNA-seq

Released

PI : Galas, David

Organization : Pacific Northwest Diabetes Research Institute

Grant : Non-ERCC Funded Study

Funding Source : Luxembourg National Research Fund

EXR-DGALA1V5h5va-AN

Datasets Help

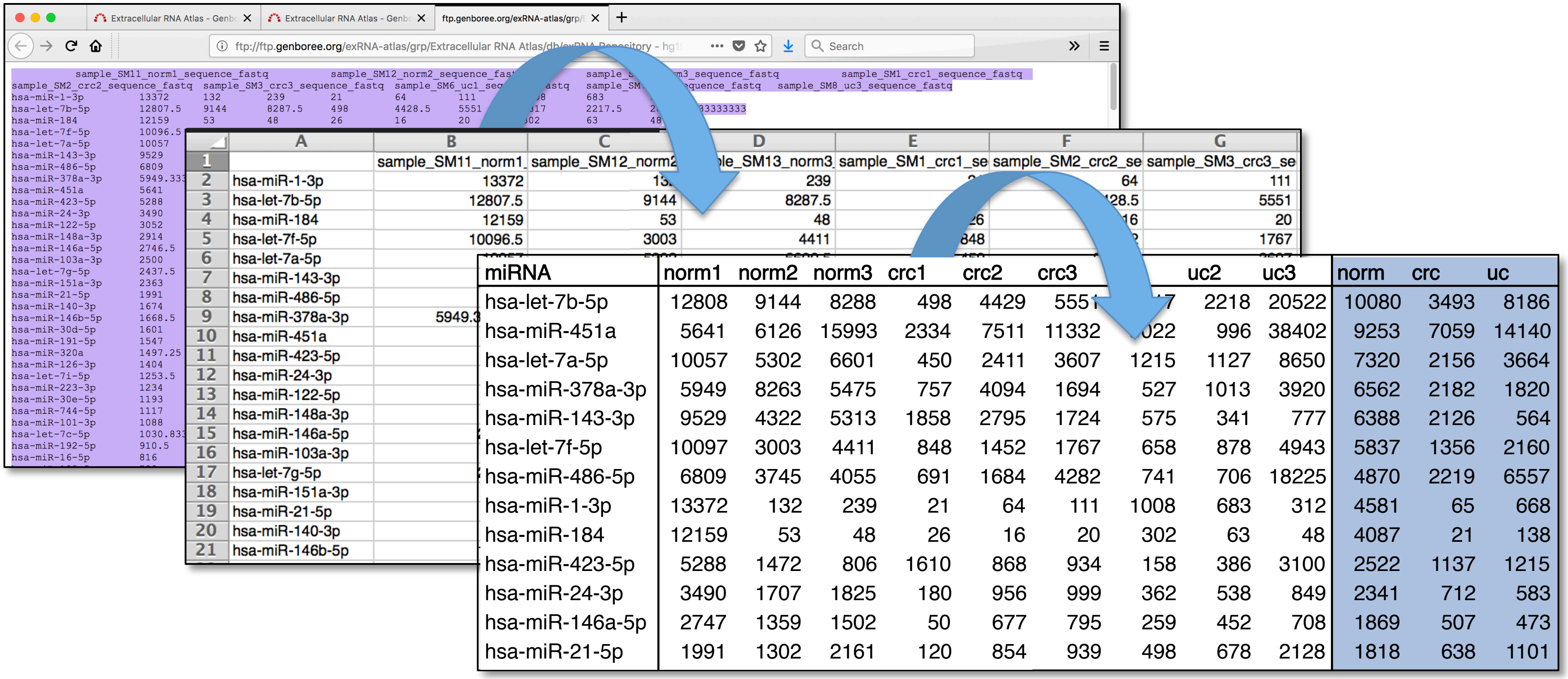
Each **card** in the layout above contains information about a dataset in the exRNA Atlas.
To learn more about a given dataset, explore the following options within each card:

- The **Analysis ID** in the lower left corner will open an RNA profile grid for that dataset.
  For RNA-seq profiles, this grid will contain different read counts from various stages of mapping in the exceRpt pipeline.
  For qPCR profiles, this grid will contain sample metadata.
- The **Samples** badge on the right side will open a grid containing sample metadata for that dataset.
- The ⤓ button will bring up a pop-over window that contains various downloads associated with the dataset.
  - ○ The 📈 button will download a PDF containing different diagnostic plots for the dataset.
  - ○ The 📄 button will download a table of the different raw (not normalized) miRNA read counts for the dataset.
  - ○ The ☰ button will download a text file containing the exogenous genomic taxonomy's cumulative read counts for the dataset.
  - ○ The ▤ button will download a text file containing the exogenous ribosomal RNA taxonomy's cumulative read counts for the dataset.
  - ○ The ▣ button will download an archive containing a large assortment of different summary files for this dataset.

Use Case 1: Exogenous exRNA

3. Which miRNAs are normally present in human plasma?

- Copy and paste that read count table into Excel (Paste Special – as Text).
- Then average the read counts for the 3 conditions – healthy controls (norm), colorectal cancer (crc), and ulcerative colitis (uc) – and sort by the average in the normal samples.



```
        sample_SM11_norm1_sequence_fastq          sample_SM12_norm2_sequence_fas            sample_S      m3_sequence_fastq          sample_SM1_crc1_sequence_fastq
sample_SM2_crc2_sequence_fastq  sample_SM3_crc3_sequence_fastq  sample_SM6_uc1_seq        astq  sample_SM      equence_fastq  sample_SM8_uc3_sequence_fastq
hsa-miR-1-3p        13372      132     239     21     64      111         8   683
hsa-let-7b-5p       12807.5    9144    8287.5  498    4428.5  5551      317   2217.5   2     3333333
hsa-miR-184         12159      53      48      26     16      20         02   63       48
hsa-let-7f-5p       10096.5
hsa-let-7a-5p       10057
hsa-miR-143-3p      9529
hsa-miR-486-5p      6809
hsa-miR-378a-3p     5949.333
hsa-miR-451a        5641
hsa-miR-423-5p      5288
hsa-miR-24-3p       3490
hsa-miR-122-5p      3052
hsa-miR-148a-3p     2914
hsa-miR-146a-5p     2746.5
hsa-miR-103a-3p     2500
hsa-let-7g-5p       2437.5
hsa-miR-151a-3p     2363
hsa-miR-21-5p       1991
hsa-miR-140-3p      1674
hsa-miR-146b-5p     1668.5
hsa-miR-30d-5p      1601
hsa-miR-191-5p      1547
hsa-miR-320a        1497.25
hsa-miR-126-3p      1404
hsa-let-7i-5p       1253.5
hsa-miR-223-3p      1234
hsa-miR-30e-5p      1193
hsa-miR-744-5p      1117
hsa-miR-101-3p      1088
hsa-let-7c-5p       1030.833
hsa-miR-192-5p      910.5
hsa-miR-16-5p       816
```

	A	B	C	D	E	F	G
1		sample_SM11_norm1_	sample_SM12_norm2	le_SM13_norm3	sample_SM1_crc1_se	sample_SM2_crc2_se	sample_SM3_crc3_se
2	hsa-miR-1-3p	13372	13	239		64	111
3	hsa-let-7b-5p	12807.5	9144	8287.5		128.5	5551
4	hsa-miR-184	12159	53	48	26	16	20
5	hsa-let-7f-5p	10096.5	3003	4411	848		1767
6	hsa-let-7a-5p						
7	hsa-miR-143-3p						
8	hsa-miR-486-5p						
9	hsa-miR-378a-3p	5949.3					
10	hsa-miR-451a						
11	hsa-miR-423-5p						
12	hsa-miR-24-3p						
13	hsa-miR-122-5p						
14	hsa-miR-148a-3p						
15	hsa-miR-146a-5p						
16	hsa-miR-103a-3p						
17	hsa-let-7g-5p						
18	hsa-miR-151a-3p						
19	hsa-miR-21-5p						
20	hsa-miR-140-3p						
21	hsa-miR-146b-5p						

miRNA	norm1	norm2	norm3	crc1	crc2	crc3		uc2	uc3	norm	crc	uc
hsa-let-7b-5p	12808	9144	8288	498	4429	5551	7	2218	20522	10080	3493	8186
hsa-miR-451a	5641	6126	15993	2334	7511	11332	022	996	38402	9253	7059	14140
hsa-let-7a-5p	10057	5302	6601	450	2411	3607	1215	1127	8650	7320	2156	3664
hsa-miR-378a-3p	5949	8263	5475	757	4094	1694	527	1013	3920	6562	2182	1820
hsa-miR-143-3p	9529	4322	5313	1858	2795	1724	575	341	777	6388	2126	564
hsa-let-7f-5p	10097	3003	4411	848	1452	1767	658	878	4943	5837	1356	2160
hsa-miR-486-5p	6809	3745	4055	691	1684	4282	741	706	18225	4870	2219	6557
hsa-miR-1-3p	13372	132	239	21	64	111	1008	683	312	4581	65	668
hsa-miR-184	12159	53	48	26	16	20	302	63	48	4087	21	138
hsa-miR-423-5p	5288	1472	806	1610	868	934	158	386	3100	2522	1137	1215
hsa-miR-24-3p	3490	1707	1825	180	956	999	362	538	849	2341	712	583
hsa-miR-146a-5p	2747	1359	1502	50	677	795	259	452	708	1869	507	473
hsa-miR-21-5p	1991	1302	2161	120	854	939	498	678	2128	1818	638	1101

Use Case 1: Exogenous exRNA

3. Which miRNAs are normally present in human plasma?

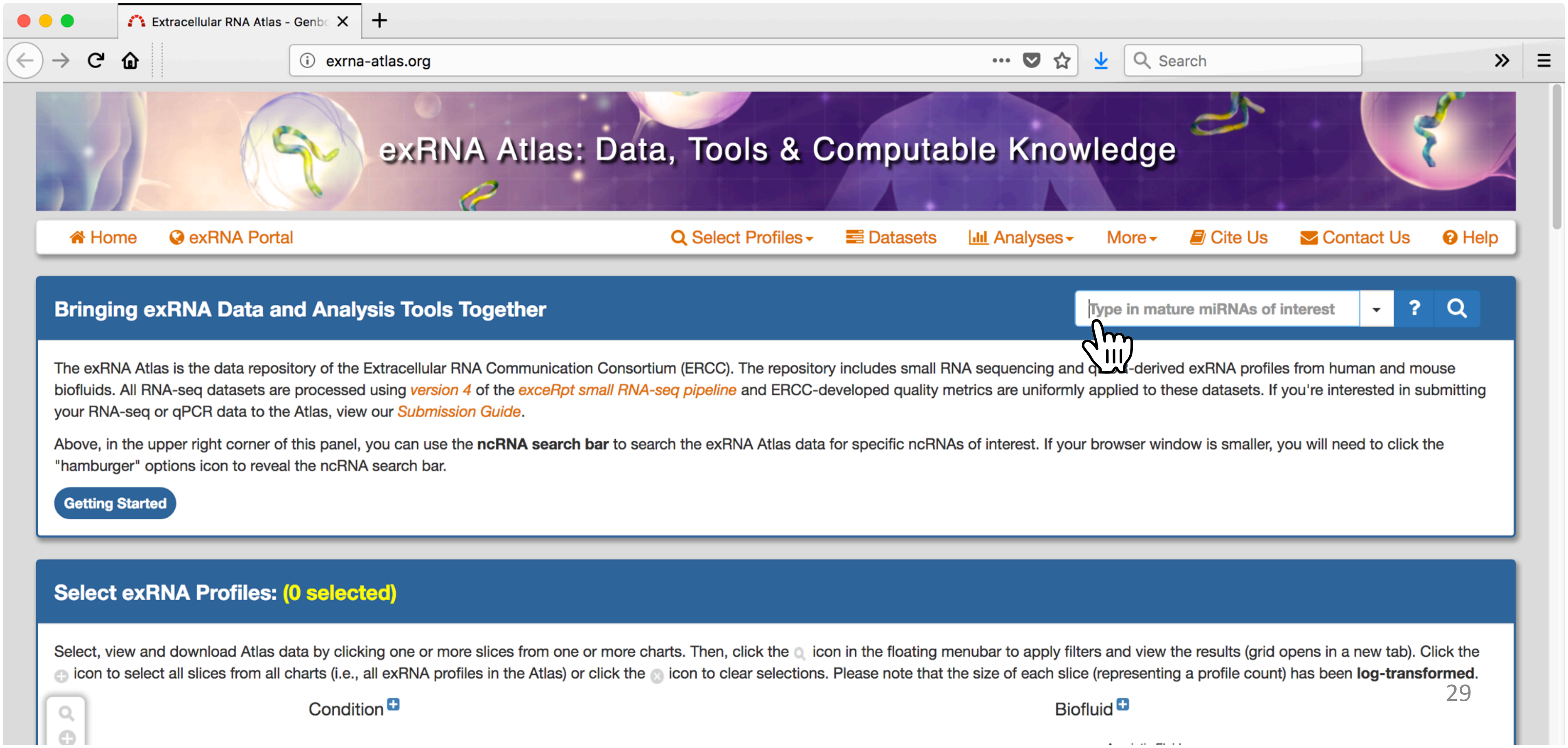
- Back at the exRNA Atlas home page, we can study the pathways in which a group of miRNAs of interest are involved.

miRNA	norm1	norm2	norm3	crc1	crc2	crc3	uc1	uc2	uc3	norm	crc	uc
hsa-let-7b-5p	12808	9144	8288	498	4429	5551	1817	2218	20522	10080	3493	8186
hsa-miR-451a	5641	6126	15993	2334	7511	11332	3022	996	38402	9253	7059	14140
hsa-let-7a-5p	10057	5302	6601	450	2411	3607	1215	1127	8650	7320	2156	3664
hsa-miR-378a-3p	5949	8263	5475	757	4094	1694	527	1013	3920	6562	2182	1820
hsa-miR-143-3p	9529	4322	5313	1858	2795	1724	575	341	777	6388	2126	564
hsa-let-7f-5p	10097	3003	4411	848	1452	1767	658	878	4943	5837	1356	2160
hsa-miR-486-5p	6809	3745	4055	691	1684	4282	741	706	18225	4870	2219	6557
hsa-miR-1-3p	13372	132	239	21	64	111	1008	683	312	4581	65	668
hsa-miR-184	12159	53	48	26	16	20	302	63	48	4087	21	138
hsa-miR-423-5p	5288	1472	806	1610	868	934	158	386	3100	2522	1137	1215
hsa-miR-24-3p	3490	1707	1825	180	956	999	362	538	849	2341	712	583
hsa-miR-146a-5p	2747	1359	1502	50	677	795	259	452	708	1869	507	473
hsa-miR-21-5p	1991	1302	2161	120	854	939	498	678	2128	1818	638	1101
hsa-miR-140-3p	1674	1485	2202	305	1127	1281	355	225	5378	1787	904	1986
hsa-miR-122-5p	3052	496	1675	17	749	400	173	1518	3381	1741	389	1691
hsa-miR-148a-3p	2914	583	1661	105	821	823	214	504	1257	1719	583	658

Use Case 1: Exogenous exRNA

3. Which miRNAs are normally present in human plasma?

- Paste the set of complete miRNA identifiers into the ncRNA search box.



Bringing exRNA Data and Analysis Tools Together

The exRNA Atlas is the data repository of the Extracellular RNA Communication Consortium (ERCC). The repository includes small RNA sequencing and q...-derived exRNA profiles from human and mouse biofluids. All RNA-seq datasets are processed using *version 4* of the *exceRpt small RNA-seq pipeline* and ERCC-developed quality metrics are uniformly applied to these datasets. If you're interested in submitting your RNA-seq or qPCR data to the Atlas, view our *Submission Guide*.

Above, in the upper right corner of this panel, you can use the **ncRNA search bar** to search the exRNA Atlas data for specific ncRNAs of interest. If your browser window is smaller, you will need to click the "hamburger" options icon to reveal the ncRNA search bar.

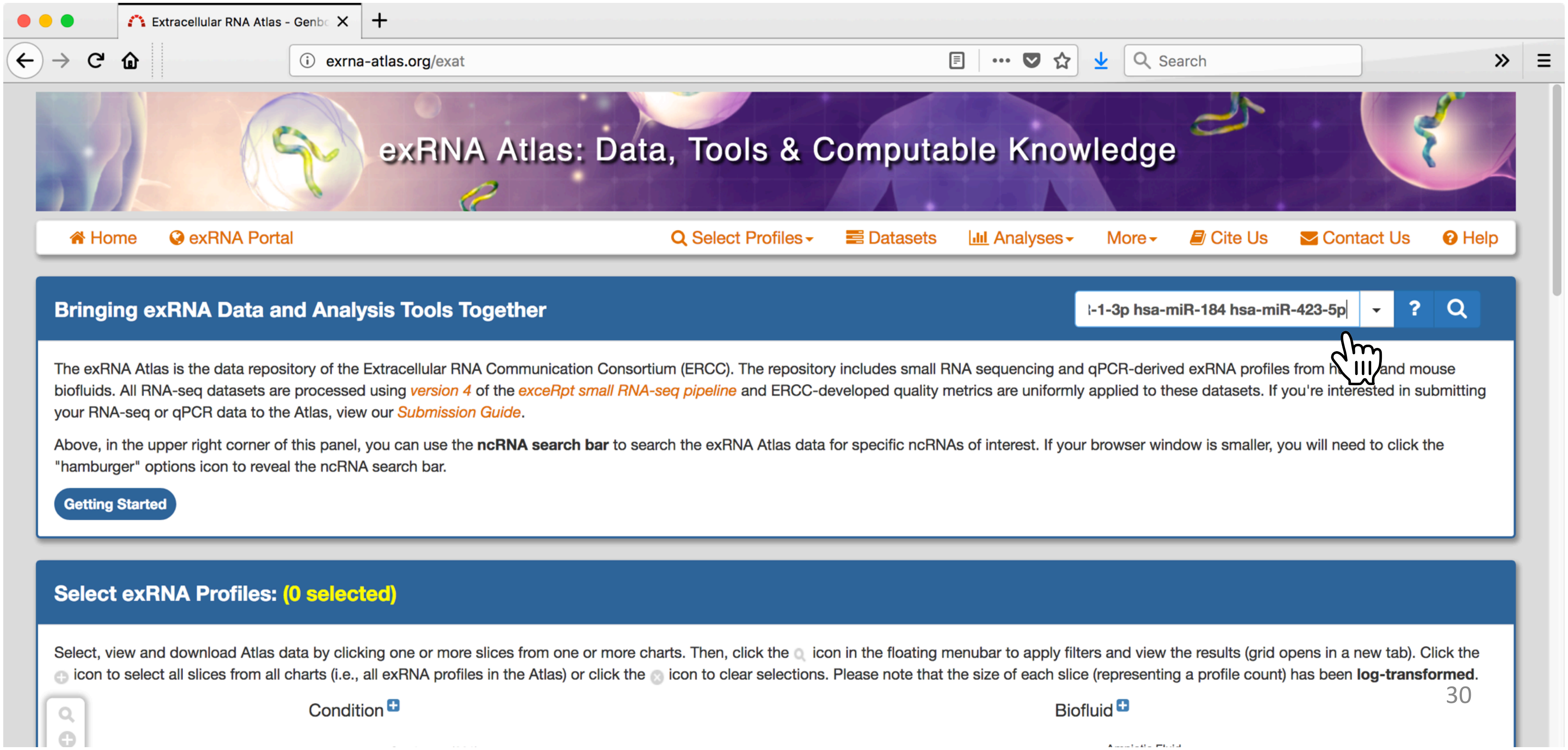
Select exRNA Profiles: (0 selected)

Select, view and download Atlas data by clicking one or more slices from one or more charts. Then, click the icon in the floating menubar to apply filters and view the results (grid opens in a new tab). Click the icon to select all slices from all charts (i.e., all exRNA profiles in the Atlas) or click the icon to clear selections. Please note that the size of each slice (representing a profile count) has been **log-transformed**.

Use Case 1: Exogenous exRNA

3. Which miRNAs are normally present in human plasma?

- Paste the set of complete miRNA identifiers into the ncRNA search box.



The exRNA Atlas is the data repository of the Extracellular RNA Communication Consortium (ERCC). The repository includes small RNA sequencing and qPCR-derived exRNA profiles from h and mouse biofluids. All RNA-seq datasets are processed using *version 4* of the *exceRpt small RNA-seq pipeline* and ERCC-developed quality metrics are uniformly applied to these datasets. If you're interested in submitting your RNA-seq or qPCR data to the Atlas, view our *Submission Guide*.

Above, in the upper right corner of this panel, you can use the **ncRNA search bar** to search the exRNA Atlas data for specific ncRNAs of interest. If your browser window is smaller, you will need to click the "hamburger" options icon to reveal the ncRNA search bar.

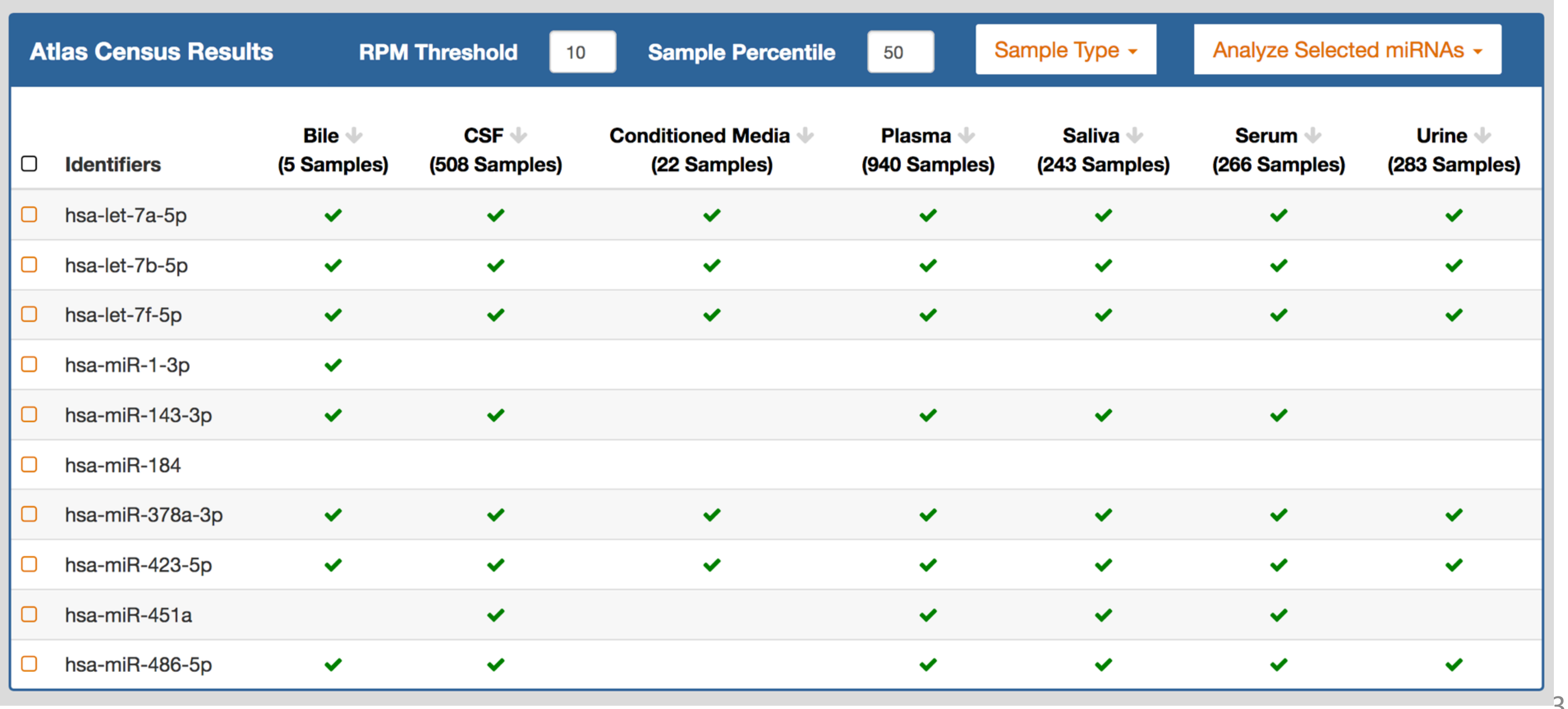
Select exRNA Profiles: (0 selected)

Select, view and download Atlas data by clicking one or more slices from one or more charts. Then, click the 🔍 icon in the floating menubar to apply filters and view the results (grid opens in a new tab). Click the ⊕ icon to select all slices from all charts (i.e., all exRNA profiles in the Atlas) or click the ⊗ icon to clear selections. Please note that the size of each slice (representing a profile count) has been **log-transformed**.

Use Case 1: Exogenous exRNA

3. Which miRNAs are normally present in human plasma?

- The result shows the frequency of the selected ncRNAs in the Atlas' RNA-seq data.



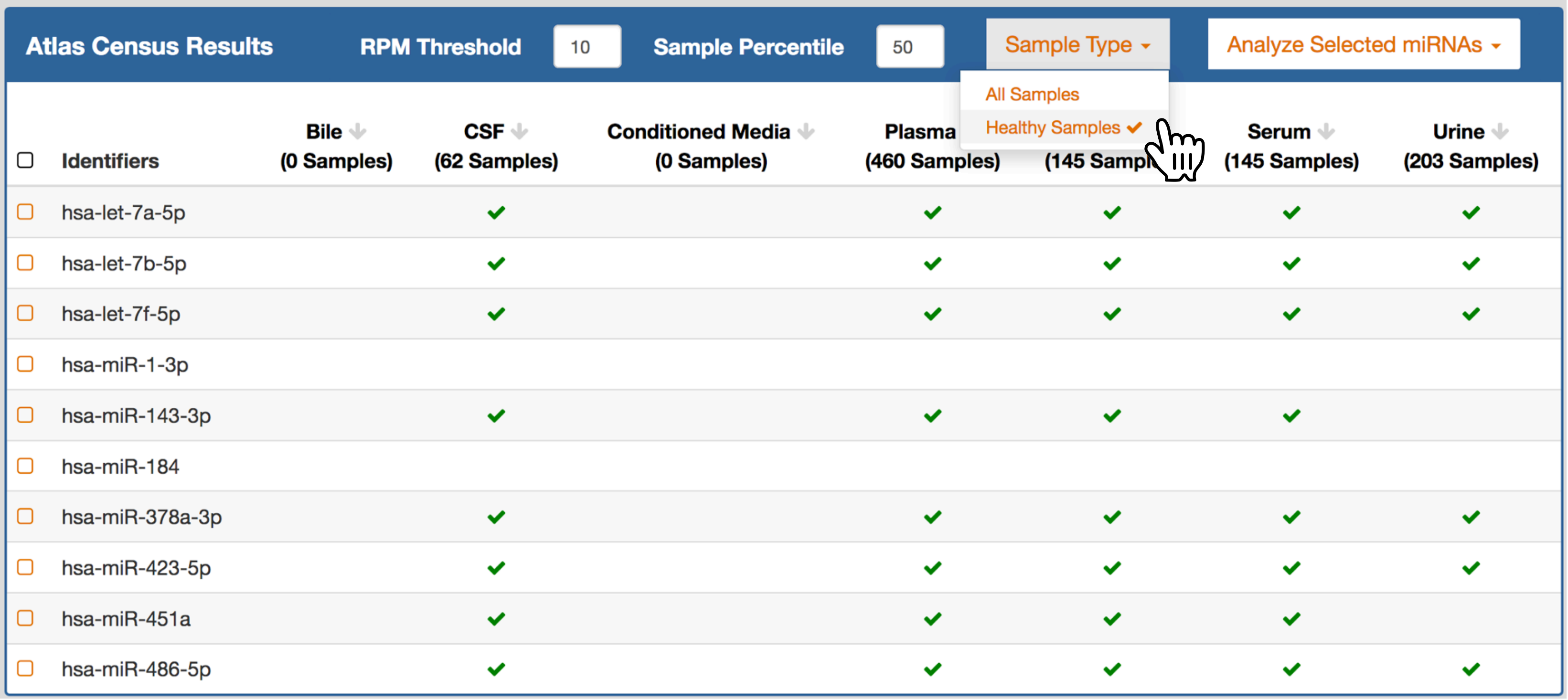
Atlas Census Results **RPM Threshold** 10 **Sample Percentile** 50 Sample Type ▾ Analyze Selected miRNAs ▾

☐	Identifiers	Bile ↓ (5 Samples)	CSF ↓ (508 Samples)	Conditioned Media ↓ (22 Samples)	Plasma ↓ (940 Samples)	Saliva ↓ (243 Samples)	Serum ↓ (266 Samples)	Urine ↓ (283 Samples)
☐	hsa-let-7a-5p	✓	✓	✓	✓	✓	✓	✓
☐	hsa-let-7b-5p	✓	✓	✓	✓	✓	✓	✓
☐	hsa-let-7f-5p	✓	✓	✓	✓	✓	✓	✓
☐	hsa-miR-1-3p	✓						
☐	hsa-miR-143-3p	✓	✓		✓	✓	✓	
☐	hsa-miR-184							
☐	hsa-miR-378a-3p	✓	✓	✓	✓	✓	✓	✓
☐	hsa-miR-423-5p	✓	✓	✓	✓	✓	✓	✓
☐	hsa-miR-451a		✓		✓	✓	✓	
☐	hsa-miR-486-5p	✓	✓		✓	✓	✓	✓

Use Case 1: Exogenous exRNA

3. Which miRNAs are normally present in human plasma?

- The result shows the frequency of the selected ncRNAs in the Atlas' RNA-seq data.
- You can focus just on healthy samples.



Atlas Census Results

RPM Threshold: 10

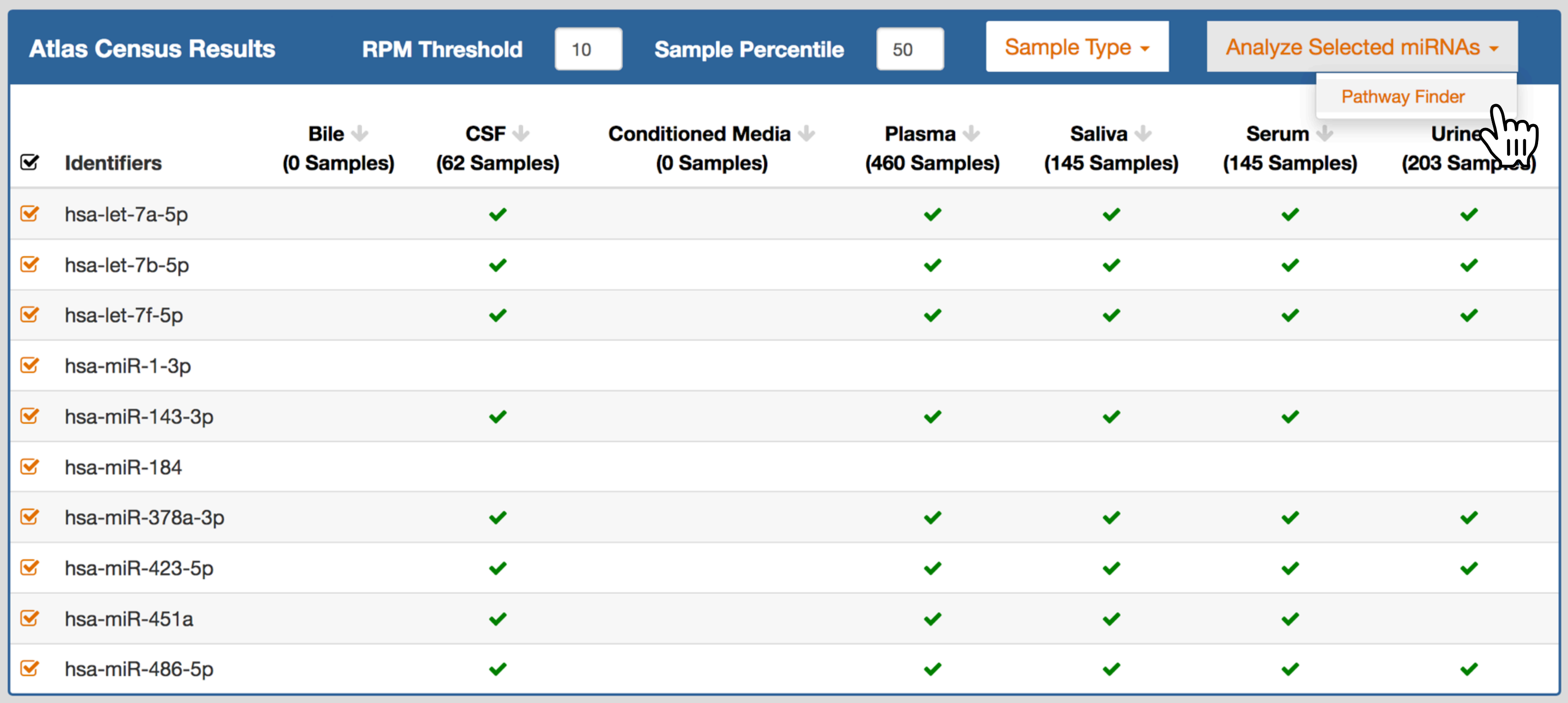
Sample Percentile: 50

☐	Identifiers	Bile (0 Samples)	CSF (62 Samples)	Conditioned Media (0 Samples)	Plasma (460 Samples)	(145 Sampl	Serum (145 Samples)	Urine (203 Samples)
☐	hsa-let-7a-5p		✓		✓	✓	✓	✓
☐	hsa-let-7b-5p		✓		✓	✓	✓	✓
☐	hsa-let-7f-5p		✓		✓	✓	✓	✓
☐	hsa-miR-1-3p							
☐	hsa-miR-143-3p		✓		✓	✓	✓	
☐	hsa-miR-184							
☐	hsa-miR-378a-3p		✓		✓	✓	✓	✓
☐	hsa-miR-423-5p		✓		✓	✓	✓	✓
☐	hsa-miR-451a		✓		✓	✓	✓	
☐	hsa-miR-486-5p		✓		✓	✓	✓	✓

Use Case 1: Exogenous exRNA

3. Which miRNAs are normally present in human plasma?

- The result shows the frequency of the selected ncRNAs in the Atlas' RNA-seq data.
- You can focus just on healthy samples...
- and use Pathway Finder to to examine which signaling pathways they affect.



Atlas Census Results

RPM Threshold: 10

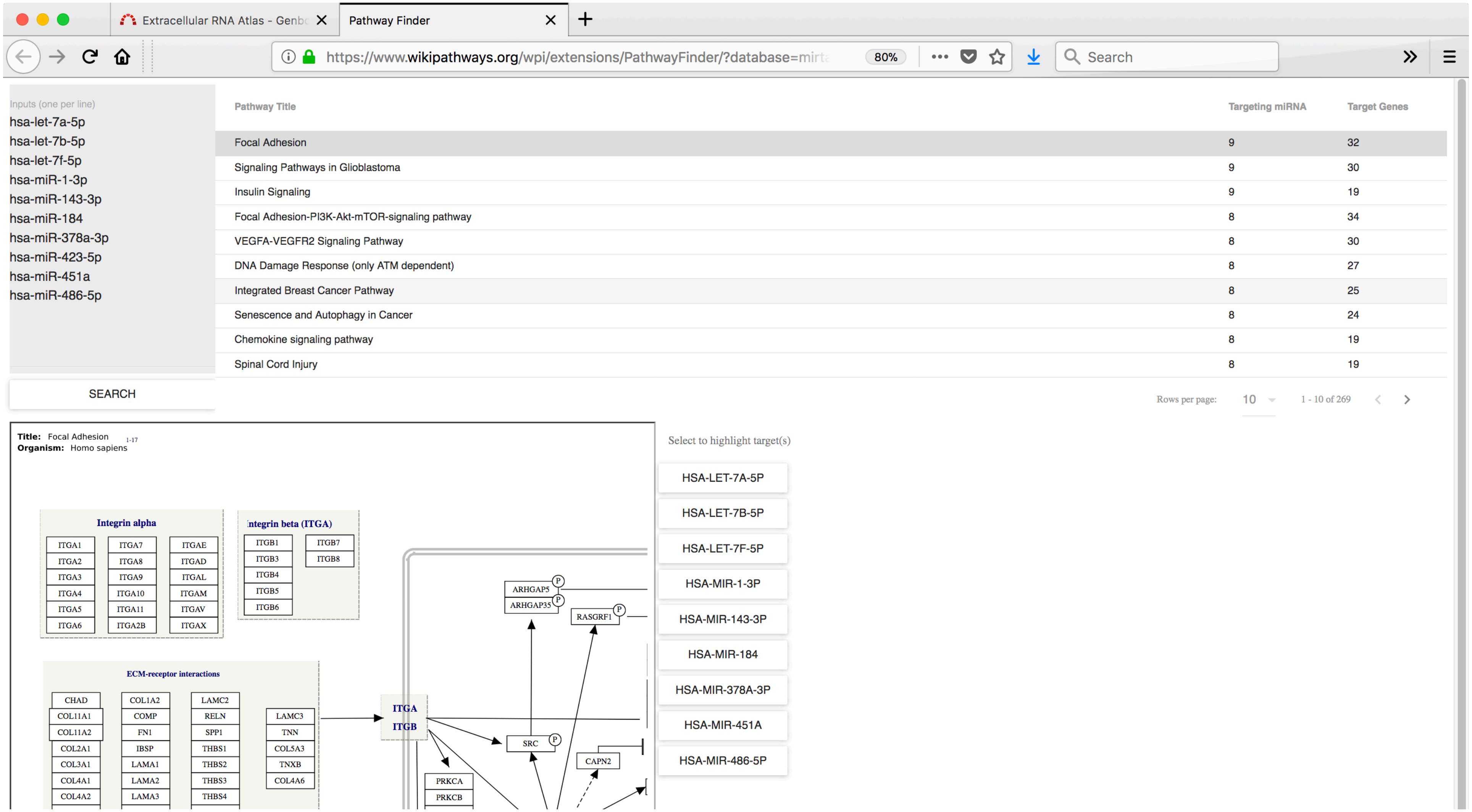
Sample Percentile: 50

☑	Identifiers	Bile (0 Samples)	CSF (62 Samples)	Conditioned Media (0 Samples)	Plasma (460 Samples)	Saliva (145 Samples)	Serum (145 Samples)	Urine (203 Samp s)
☑	hsa-let-7a-5p		✓		✓	✓	✓	✓
☑	hsa-let-7b-5p		✓		✓	✓	✓	✓
☑	hsa-let-7f-5p		✓		✓	✓	✓	✓
☑	hsa-miR-1-3p							
☑	hsa-miR-143-3p		✓		✓	✓	✓	
☑	hsa-miR-184							
☑	hsa-miR-378a-3p		✓		✓	✓	✓	✓
☑	hsa-miR-423-5p		✓		✓	✓	✓	✓
☑	hsa-miR-451a		✓		✓	✓	✓	
☑	hsa-miR-486-5p		✓		✓	✓	✓	✓

Use Case 1: Exogenous exRNA

PACIFIC NORTHWEST RESEARCH INSTITUTE

Pathway Finder



Inputs (one per line)
hsa-let-7a-5p
hsa-let-7b-5p
hsa-let-7f-5p
hsa-miR-1-3p
hsa-miR-143-3p
hsa-miR-184
hsa-miR-378a-3p
hsa-miR-423-5p
hsa-miR-451a
hsa-miR-486-5p

SEARCH

Pathway Title	Targeting miRNA	Target Genes
Focal Adhesion	9	32
Signaling Pathways in Glioblastoma	9	30
Insulin Signaling	9	19
Focal Adhesion-PI3K-Akt-mTOR-signaling pathway	8	34
VEGFA-VEGFR2 Signaling Pathway	8	30
DNA Damage Response (only ATM dependent)	8	27
Integrated Breast Cancer Pathway	8	25
Senescence and Autophagy in Cancer	8	24
Chemokine signaling pathway	8	19
Spinal Cord Injury	8	19

Rows per page: 10 1 - 10 of 269

Title: Focal Adhesion
Organism: Homo sapiens

Select to highlight target(s)

Use Case 1: Summary

1. Do all plasma small RNAs map to the human genome?

   No, in this set of plasma samples, more than 50% of small RNAs sequenced map to non-human genomes.

2. What are the sources of small RNAs found in human plasma that do not map to the human genome?

   In tables and phylogenetic trees, the exceRpt small RNAseq pipeline shows us that the major components of those non-human parasites are fungi and bacteria, specifically proteobacteria.

3. Which miRNAs are normally present in human plasma?

   We used the Pathway Finder tool just to scratch the surface of analyzing the miRNA content in the sample set.

Wang K., Hong L., Yuan Y., Etheridge A., Zhou Y., Huang D., Wilmes P., & Galas D. (2012) The Complex Exogenous RNA Spectra in Human Plasma: An Interface with Human Gut Biota? PLoS ONE 7: e51009.