



ERCC Data Analysis Workshop

Use Case 2 Part 2: exRNA Discovery and Functional Analysis using the exRNA Atlas and Network Analysis Resources on the Genboree Workbench

Organized and Hosted by the Data Management and Resource Repository (DMRR)

Sunday 17 April, 2016 6-9pm

Data kindly provided by Kendall Jensen, Translational Genomics Research Institute

Burgos K., et al. (2014) Profiles of Extracellular miRNA in Cerebrospinal Fluid and Serum from Patients with Alzheimer's and Parkinson's Diseases Correlate with Disease Status and Features of Pathology. PLoS ONE 9: e94839.





ERCC Data Analysis Workshop

Topics covered in Use Case 2 Part 2

- **Accessing samples in the exRNA Atlas**
- **Comparative analysis of exRNA samples using DESeq / exceRpt pipeline results and tools in the Genboree Workbench**
- **Finding information about exRNAs of interest**
 - **on Exocarta and Vesiclepedia**
 - **on GeneWiki and Wikidata**
- **Pathway analysis using Cytoscape & WikiPathways in the Genboree Workbench**



Use Case 2: **miRNA Changes** in Alzheimer's and Parkinson's Disease



One goal of this use case is to show that the Genboree Workbench and the exceRpt small RNA-seq analysis pipeline can replicate the results of Burgos et al.

We have developed these pipelines because as the number of datasets in the exRNA Atlas grows, it is vital that they be analyzed in a reproducible and comparable way.

Burgos K., et al. (2014) Profiles of Extracellular miRNA in Cerebrospinal Fluid and Serum from Patients with Alzheimer's and Parkinson's Diseases Correlate with Disease Status and Features of Pathology. PLoS ONE 9: e94839.



Use Case 2: **miRNA Changes** in Alzheimer's and Parkinson's Disease



Biological Question of Interest

Alzheimer's and Parkinson's are two neurodegenerative diseases that are difficult to assess by biopsy, since the affected tissue is inaccessible.

In their study, Burgos et al. assessed the miRNA content in cerebrospinal fluid (CSF) and serum from postmortem subjects with full neuropathology evaluations. The goal of the study was to identify extracellular miRNA biomarkers that correlate with disease status and progression.

In this use case, we focus on biomarkers that correlate with disease status by replicating the differential expression analysis in Burgos et al.

Burgos K., et al. (2014) Profiles of Extracellular miRNA in Cerebrospinal Fluid and Serum from Patients with Alzheimer's and Parkinson's Diseases Correlate with Disease Status and Features of Pathology. PLoS ONE 9: e94839.



ERCC Data Analysis Workshop

Topics covered in Use Case 2 Part 2

- **Accessing samples in the exRNA Atlas**
- Comparative analysis of exRNA samples using DESeq / exceRpt pipeline results and tools in the Genboree Workbench
- Finding information about exRNAs of interest
 - on Exocarta and Vesiclepedia
 - on GeneWiki and Wikidata
- Pathway analysis using Cytoscape & WikiPathways in the Genboree Workbench

Accessing samples in the exRNA Atlas for discovery and analysis

Sai Lakshmi Subramanian

ERC Consortium Data Management & Resource Repository (DMRR)
Baylor College of Medicine, Houston, TX

Outline

Introduction to the exRNA Atlas

Overview of exRNA profiling data flow

Faceted sub-selection of biosamples

Biosample partition grids

Linear Tree drill-down sub-selection of samples

Summary grid with exRNA profiling studies

Informational bar charts – Atlas Statistics

Introduction to the **exRNA Atlas**

Data repository of the Extracellular RNA Communication Consortium (ERCC)

Developed and maintained by the Data Management and Resource Repository (DMRR)

Contains the exRNA profiles from various biofluids, conditions

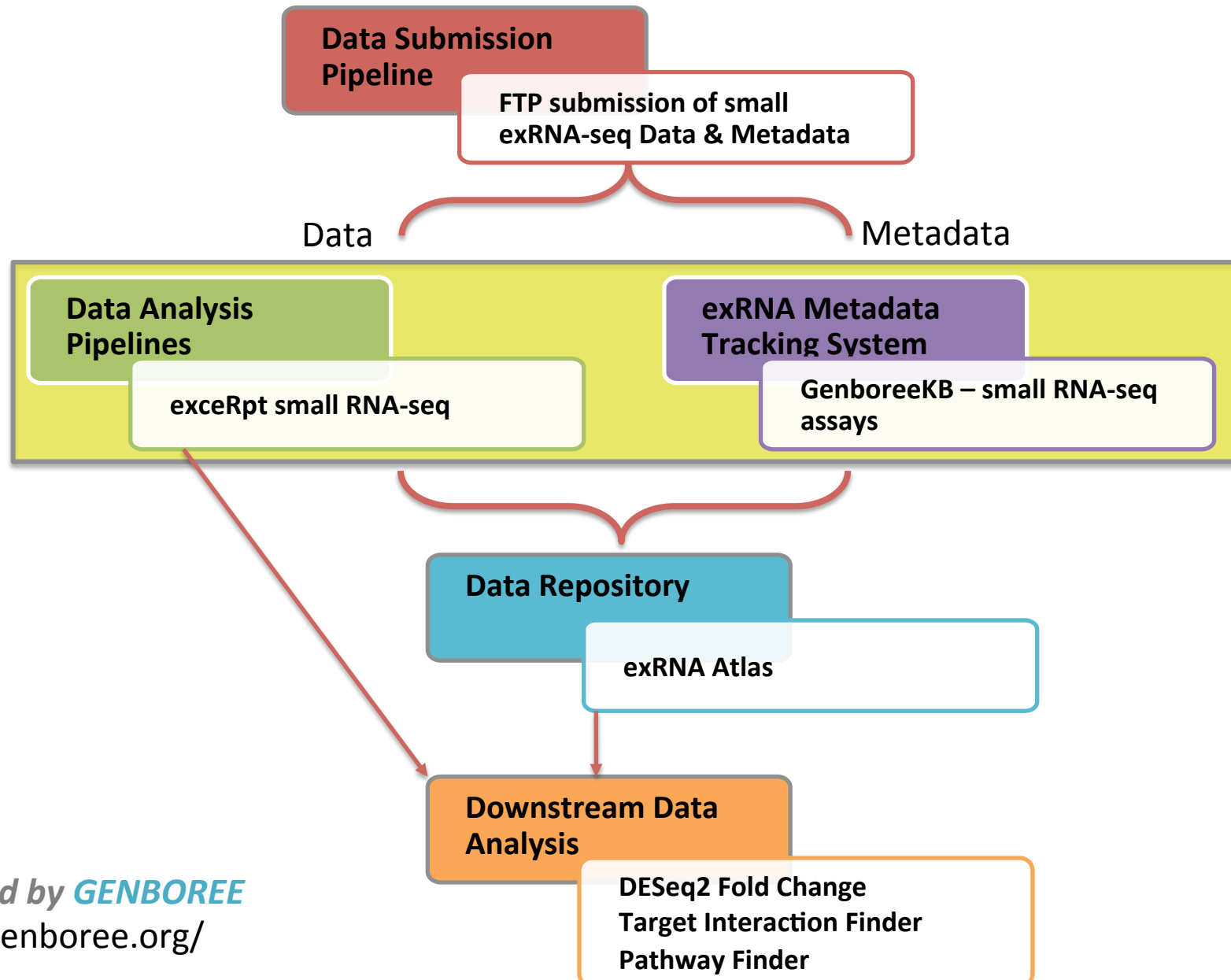
Faceted sub-selection and data navigation tools are available in the Atlas

Currently, stores samples profiled using small RNA-sequencing assays

All datasets are uniformly processed using the exceRpt small RNA-seq pipeline

As of today – **751** exRNA profiles are available in the **Public Atlas** and **1658** profiles are available in the **Consortium Atlas**

Overview of exRNA Profiling Data Flow



exRNA Atlas

- Go to the exRNA Portal at exrna.org
- Scroll down to the “Quick Links” panel
- Click “exRNA Atlas”

known to travel outside of cells and
discovered mechanisms of cell-to-

Learn More

Full Video ▶

Link to exRNA Atlas

01.27.16 by Leonora Balaj and Louise C Laurent

First CLIA-validated exosome-based
clinical liquid biopsy test is
launched



Exosome Diagnostics, Inc. [has announced](#) the
launch of ExoDx Lung(ALK), the first ever
CLIA-validated exosome based blood test. This test

Quick Links

[ExRNA Atlas](#)
[ExRNA @ BioGPS](#)
[ExRNA @ WikiPathways](#)
[ExoCarta](#)
[Vesiclepedia](#)
[miRandola](#)

NIH National Institutes of Health
Turning Discovery Into Health

Getting Started – Tutorials, Video

Open the exRNA Atlas page → Click the link “**Getting Started**” below the charts, to open the tutorial Wiki in a new tab.

Select, view and download extracellular RNA profiling data by using the various charts above. Alternatively, choose a grid view or drill-down search from the panel below to select exRNA profiles of interest.

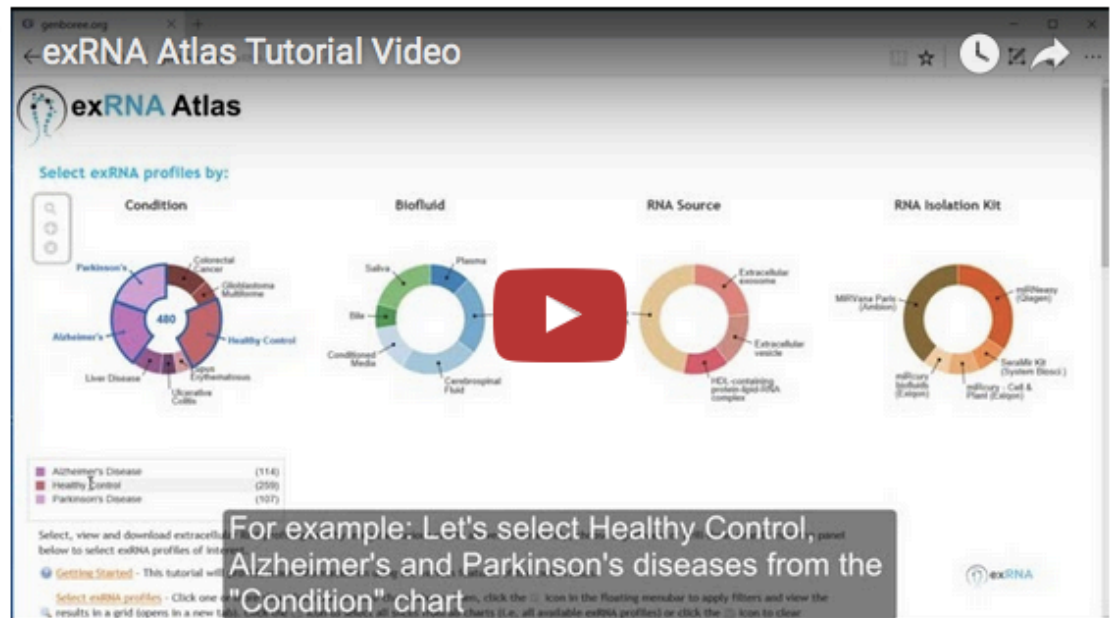
Getting Started - This tutorial will provide more information on using the various features of the exRNA Atlas.

Select exRNA profiles - Click one or more slices from one or more charts, click the **+** icon to select results in a grid (opens in a new tab). Click the **+** icon to select from all slices.

View tutorial video to learn more about the features in the Atlas

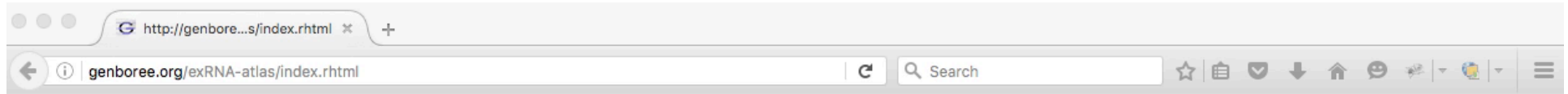
Introduction to the exRNA Atlas

Watch Video Tutorial on all of the available features in the exRNA Atlas.

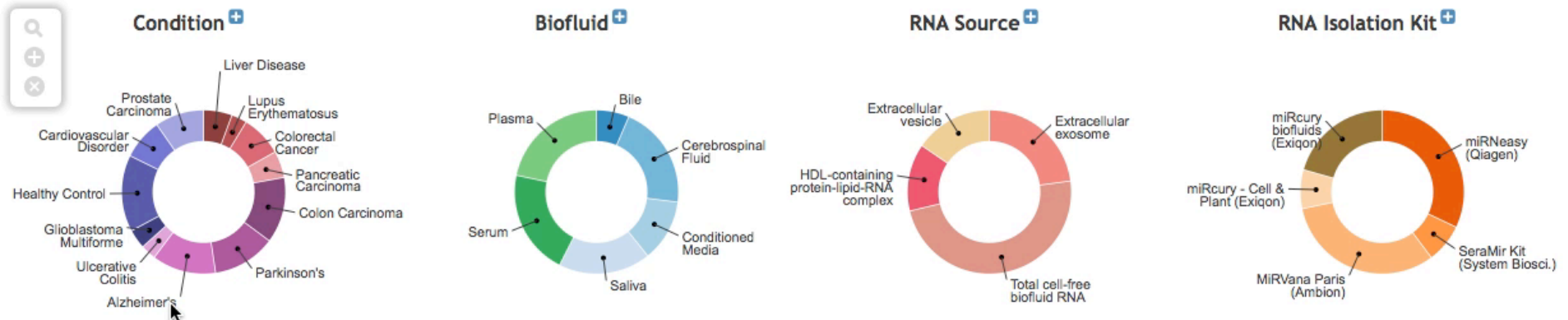


The screenshot shows a video player titled "exRNA Atlas Tutorial Video". The video content displays the exRNA Atlas web interface. At the top, it says "Select exRNA profiles by:". Below this are four circular charts: "Condition", "Biofluid", "RNA Source", and "RNA Isolation Kit". The "Condition" chart is highlighted with a red circle and a red arrow pointing to it from the text "View tutorial video to learn more about the features in the Atlas". The "Condition" chart shows slices for "Alzheimer's", "Parkinson's", "Healthy Control", "Liver Disease", "Lipid Dysregulation", "Ulcerative Colitis", "Colon Cancer", "Glioblastoma Multiforme", and "Healthy Control". A legend below the charts shows: Alzheimer's Disease (114), Healthy Control (259), and Parkinson's Disease (107). A text box at the bottom of the video player says: "For example: Let's select Healthy Control, panel Alzheimer's and Parkinson's diseases from the 'Condition' chart".

New Features in Atlas



Select exRNA profiles: (0 selected)



Select, view and download extracellular RNA profiling data by using the various charts above. Alternatively, choose a grid view or drill-down search from the panel below to select exRNA profiles of interest.

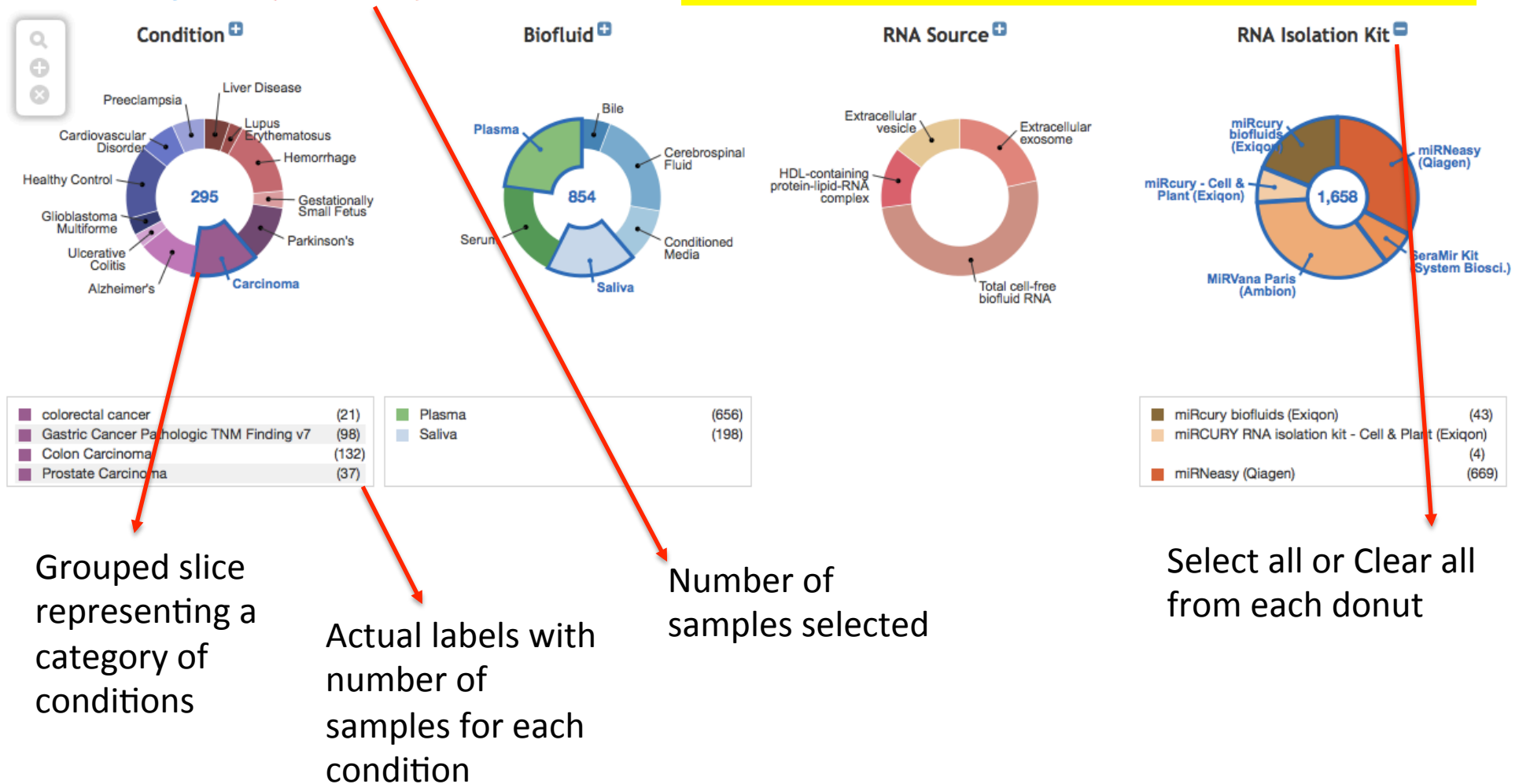
[Getting Started](#) - This tutorial will provide more information on using the various features of the exRNA Atlas.

[Select exRNA profiles](#) - Click one or more slices from one or more charts above. Then, click the  icon in the floating menubar to apply filters and view the results in a grid (opens in a new tab). Click the  icon to select all slices from all charts (i.e. all available exRNA profiles) or click the  icon to clear selections from all slices.

Faceted sub-selection of samples

Select exRNA profiles: (277 selected)

Faceted sub-selection of submitted samples



Sample Sub-selection in exRNA Atlas

Search Results - 277 Biosamples

[Back to Search Page](#)
[Download Samples](#)
[Go To Genboree Workbench](#)

Biosample Name	Condition	Anatomical Location	Biofluid Name	exRNA Source	RNA Isolation Kit	Meets Standard	ERCC Quality Standards		
							Reference Genome Reads	Transcriptome Genome Ratio	Transcriptome Reads
☐ Sample_2S23	Colon Carcinoma	Entire cardiovascular system	Plasma	total cell-free biofluid RNA	miRNeasy (Qiagen)	YES	↑ Sort Ascending	106531	8,203,188
☐ Sample_PC29	Prostate Carcinoma	Entire cardiovascular system	Plasma	total cell-free biofluid RNA	miRNeasy (Qiagen)	YES	↓ Sort Descending	371009	6,451,598
☐ Sample_2S7	Colon Carcinoma	Entire cardiovascular system	Plasma	total cell-free biofluid RNA	miRNeasy (Qiagen)	YES	Columns	352528	5,654,350
☐ Sample_PC32	Prostate Carcinoma	Entire cardiovascular system	Plasma	total cell-free biofluid RNA	miRNeasy (Qiagen)	YES	Filters	YES	
☐ Sample_2S13	Colon Carcinoma	Entire cardiovascular system	Plasma	total cell-free biofluid RNA	miRNeasy (Qiagen)	YES		10,014,349	0.898957
☐ Sample_PC16	Prostate Carcinoma	Entire cardiovascular system	Plasma	total cell-free biofluid RNA	miRNeasy (Qiagen)	YES		7,709,655	0.898957
☐ Sample_N9	Colon Carcinoma	Entire cardiovascular system	Plasma	total cell-free biofluid RNA	miRNeasy (Qiagen)	YES		7,629,281	0.86262
☐ Sample_3S5	Colon Carcinoma	Entire cardiovascular system	Plasma	total cell-free biofluid RNA	miRNeasy (Qiagen)	YES		6,248,010	0.838497
☐ Sample_PC14	Prostate Carcinoma	Entire cardiovascular system	Plasma	total cell-free biofluid RNA	miRNeasy (Qiagen)	YES		8,358,103	0.888418
☐ Sample_1S19	Colon Carcinoma	Entire cardiovascular system	Plasma	total cell-free biofluid RNA	miRNeasy (Qiagen)	YES		7,917,799	0.894843
☐ Sample_4S12	Colon Carcinoma	Entire cardiovascular system	Plasma	total cell-free biofluid RNA	miRNeasy (Qiagen)	YES		5,985,837	0.877731
☐ Sample_4S17	Colon Carcinoma	Entire cardiovascular system	Plasma	total cell-free biofluid RNA	miRNeasy (Qiagen)	YES		8,794,760	0.872674
☐ Sample_4S21	Colon Carcinoma	Entire cardiovascular system	Plasma	total cell-free biofluid RNA	miRNeasy (Qiagen)	YES		9,271,599	0.884286
☐ Sample_N44	Colon Carcinoma	Entire cardiovascular system	Plasma	total cell-free biofluid RNA	miRNeasy (Qiagen)	YES			8,198,743

Sub-select or filter samples of interest and **Go To Genboree Workbench** to submit Downstream Analysis Tools

Filter each column based on column values (keywords, numbers) to restrict number of samples selected

Download Data	Download Metadata	Actions	Biosample Metadata Accession
			EXR-TPATE1NEBLIB2S23-BS
			EXR-TPATE1NEBLIBPC29-BS
			EXR-TPATE1NEBLIB2S7-BS
			EXR-TPATE1NEBLIBPC32-BS
			EXR-TPATE1NEBLIB2S13-BS
			EXR-TPATE1NEBLIBPC16-BS
			EXR-TPATE1NEBLIBN9-BS
			EXR-TPATE1NEBLIB3S5-BS
			EXR-TPATE1NEBLIBPC14-BS
			EXR-TPATE1NEBLIB1S19-BS
			EXR-TPATE1NEBLIB4S12-BS
			EXR-TPATE1NEBLIB4S17-BS
			EXR-TPATE1NEBLIB4S21-BS
			EXR-TPATE1NEBLIBN44-BS

Partition Grids in exRNA Atlas

Biofluid vs Disease

BioFluids vs Diseases		
	Parkinson's Disease	Healthy Control
Serum	50	91
Saliva		100
Plasma		6
Cerebrospinal fluid	57	62

Biofluid vs Experiment Type

BioFluids vs Experiment Types	
	smRNA-Seq
Plasma	15
Serum	216
Cerebrospinal fluid	438
Bile	5
Saliva	198
Blood	266

BioFluids vs Experiment Types

	smRNA-Seq	mRNA-Seq	qRT-PCR
Plasma	15		
Serum	216		
Cerebrospinal fluid	438		
Bile	5		
Saliva	198		
Blood	266		

BioFluids vs Diseases

	Alzheimer's Disease	Chronic Maternal Hypertension with Superimposed Preeclampsia	Fetus Small for Gestational Age	Subarachnoid Hemorrhage	Healthy Control
Plasma					6
Serum	52	2	4		91
Cerebrospinal fluid	62			257	62
Bile					
Saliva					100
Blood				266	

Metadata about Biosamples:
smRNA-Seq » Bile

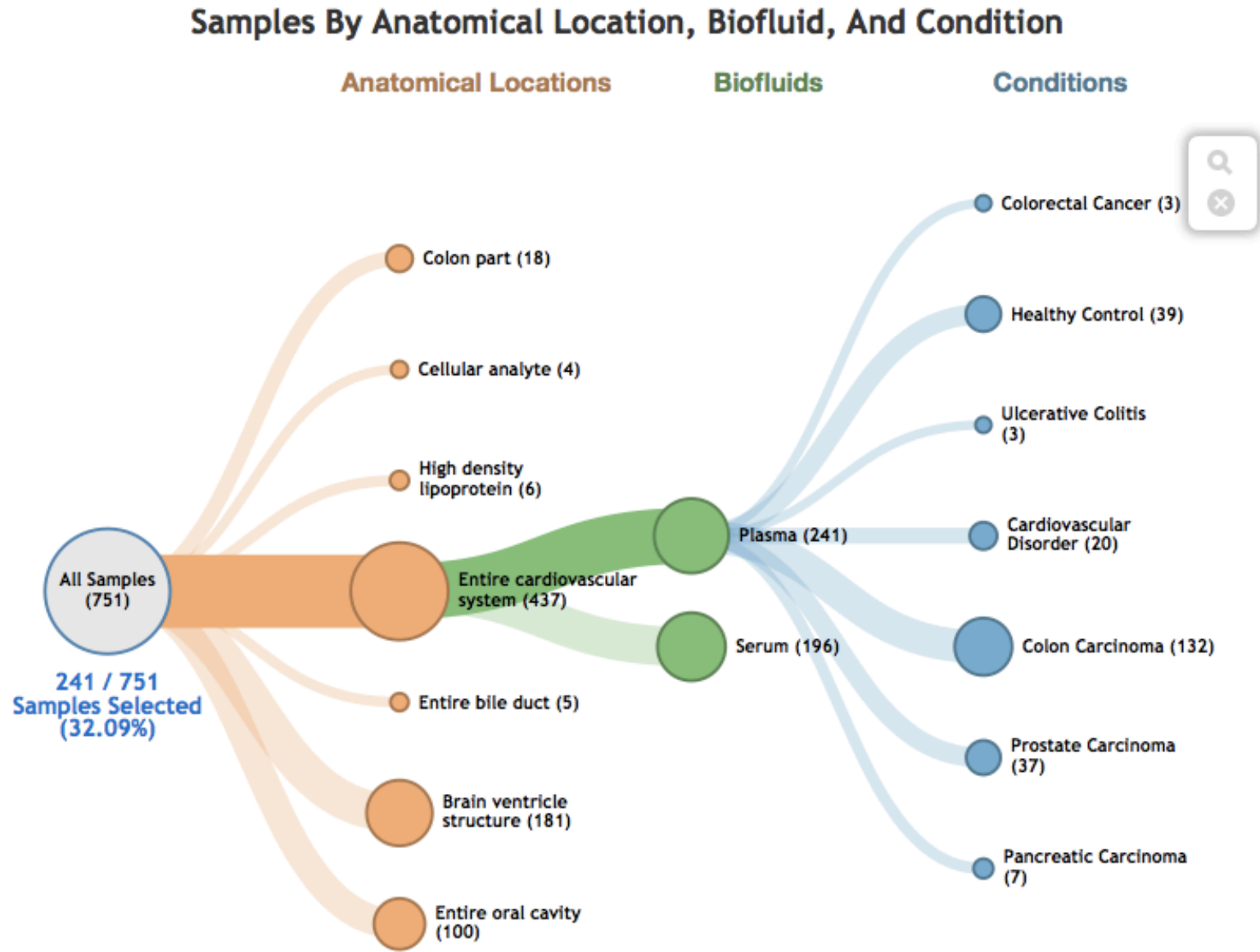
[Back to Search Page](#)

[Download Samples](#)

[Go To Genboree Workbench](#)

Biosample Name	Anatomical Location	exRNA Source	ERCC Quality Standards				Download Data	Download Metadata	Actions
			Meets Standard	Reference Genome Reads	Transcriptome Genome Ratio	Transcriptome Reads			
IYB2	Entire bile duct	extracellular vesicle	NO	5,168,243	0.434522	2,245,...			
IYB4	Entire bile duct	extracellular vesicle	YES	1,620,149	0.585237	948,171			
IYB1	Entire bile duct	extracellular vesicle	YES	994,960	0.770116	766,235			
IYB3	Entire bile duct	extracellular vesicle	YES	6,400,885	0.732687	4,689,...			
IYB5	Entire bile duct	extracellular vesicle	YES	1,723,941	0.607476	1,047,...			

Drill-down search in exRNA Atlas



Submission Summaries in Atlas

ExRNA Profiling Studies

ExRNA Profiling Studies	
PI Name	Study Title
Kendall Jensen	aSAH Study from
Tushar Patel	Human exRNA Se
Anna Krichevsky	AK-exosome RNA
Kasey Vickers	High-Density Lipop
Jeff Franklin	RNAseq analysis c
David Galas	The Complex Exog

Grid for exRNA Profiling Studies

Analysis Accession ID ↑	Study Title	PI Name	Funding Source	Grant Name/Number	Organization
EXR-AKRIC1AKGBMexo-AN	AK-exosome RNA	Anna Krichevsky	NIH Common Fund	1U19CA179563-01	BWH
EXR-DGALA1GUTPLASM-...	The Complex Exogenous RNA Spectra i...	David Galas	NIH Common Fund	1U01HL126496-01	Pacific Northwest Diabetes Research Ins.
EXR-DWONG1TMDW78D-AN	GC ExRNA	David Wong	NIH	1UH2TR000923-01	University of California-Los Angeles
EXR-JFRAN1CRCDL00-AN	RNAseq analysis of colorectal cancer ce...	Jeff Franklin	Common Fund/NIH: U19 CA179514-02 (...)	1U19CA179514-01	Vanderbilt University School of Medicine
EXR-KJENS1ADPD0000-AN	Profiles of Extracellular miRNA in Cerebr...	Kendall Jensen	Michael J Fox Foundation for Parkinsons...	Non-ERCC Funded Study	Translational Genomics Research Institut
EXR-KJENS1aSAHPR-AN	aSAH Study from Jensen lab	Kendall Jensen	NIH	1UH2TR000891-01	Translational Genomics Research Institut
EXR-KVICK1KCVSLE00-AN	High-Density Lipoproteins ? small RNA			19CA179514-01	Vanderbilt University School of Medicine
EXR-LLAUR1M4TD4M0N-AN	Sept2014_ExRNA			H2TR000906-01	University of California-San Diego
EXR-TPATE15bHLB4-AN	Human exRNA Sequencing Study			H2TR000884-01	Mayo Clinic
EXR-TPATE1SwLbJ-AN	Mouse Extracellular Small RNA Sequen			H2TR000884-01	Mayo Clinic

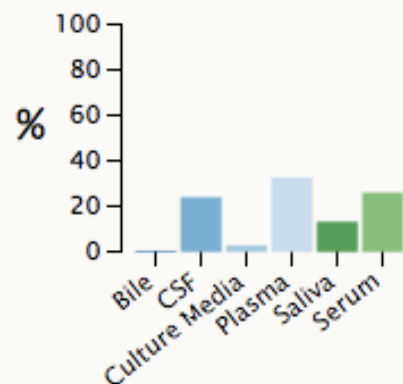
RNA PROFILE GRID	
Property	Value
Analysis	EXR-DWONG1TMDW78D-AN
Genome Version	hg19
Type	Reference Alignment
Alignment Method	exceRpt smallRNA-seq Pipeline Version 2.2.8
Grid View	

Grid with RNA Profile of Biosamples

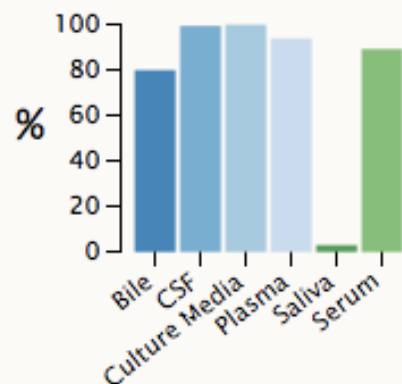
Biosamples	Actions	Input Reads	After Clipping	Failed Quality Filter	Failed Homopolymer Filter	Calibrator	UniVec Contaminants	rRNAs	Reads Used for Alignment	Genome	miRNAs Sense	miRNAs Antisense	piRNAs Sense
EXR-DWONG1GCB22S5-BS ExperimentType: smRNA-Seq Biofluids: Saliva	 	26,351,419	22,840,990	966,719	16,730	29,734	352,126	1,474,434	20,001,247	12,236,321	2,365,823	237	69,437
EXR-DWONG1GCB12S4-BS ExperimentType: smRNA-Seq Biofluids: Saliva	 	21,336,449	19,440,864	936,622	49,707	1,123,315	451,420	2,776,712	14,103,088	6,527,116	397,875	2,152	339,072
EXR-DWONG1GCB16S8-BS ExperimentType: smRNA-Seq Biofluids: Saliva	 	30,127,548	28,309,474	2,919,653	34,181	3,627,535	590,215	489,461	20,648,429	7,934,985	106,331	1,010	67,747
EXR-DWONG1GCB09S3-BS ExperimentType: smRNA-Seq Biofluids: Saliva	 	18,811,513	15,330,513	712,469	17,901	1,283,739	414,625	605,702	12,296,077	4,760,247	31,594	870	26,709
EXR-DWONG1GCB20S3-BS ExperimentType: smRNA-Seq Biofluids: Saliva	 	27,327,328	25,037,737	1,636,895	19,692	2,064,483	549,676	1,717,042	19,049,949	7,415,962	69,261	1,122	54,075
EXR-DWONG1GCB07S3-BS ExperimentType: smRNA-Seq Biofluids: Saliva	 	33,396,244	30,418,227	2,701,076	22,045	4,561,878	1,315,529	390,196	21,427,503	7,359,546	62,495	1,283	53,539
EXR-DWONG1GCB05S4-BS ExperimentType: smRNA-Seq	 	30,164,451	27,685,268	2,137,689	9,486	2,781,277	344,879	1,892,810	20,519,127	7,818,525	172,231	2,165	43,874

Submission Summaries in Atlas

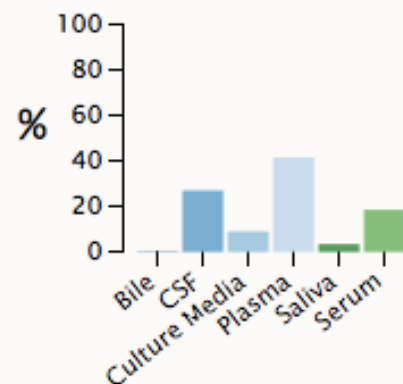
Submitted Samples / Biofluid



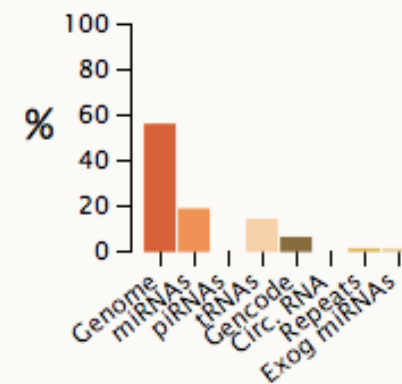
Samples Passed QC / Biofluid



Transcriptome Mapped Reads / Biofluid



Read Mappings / RNA Type

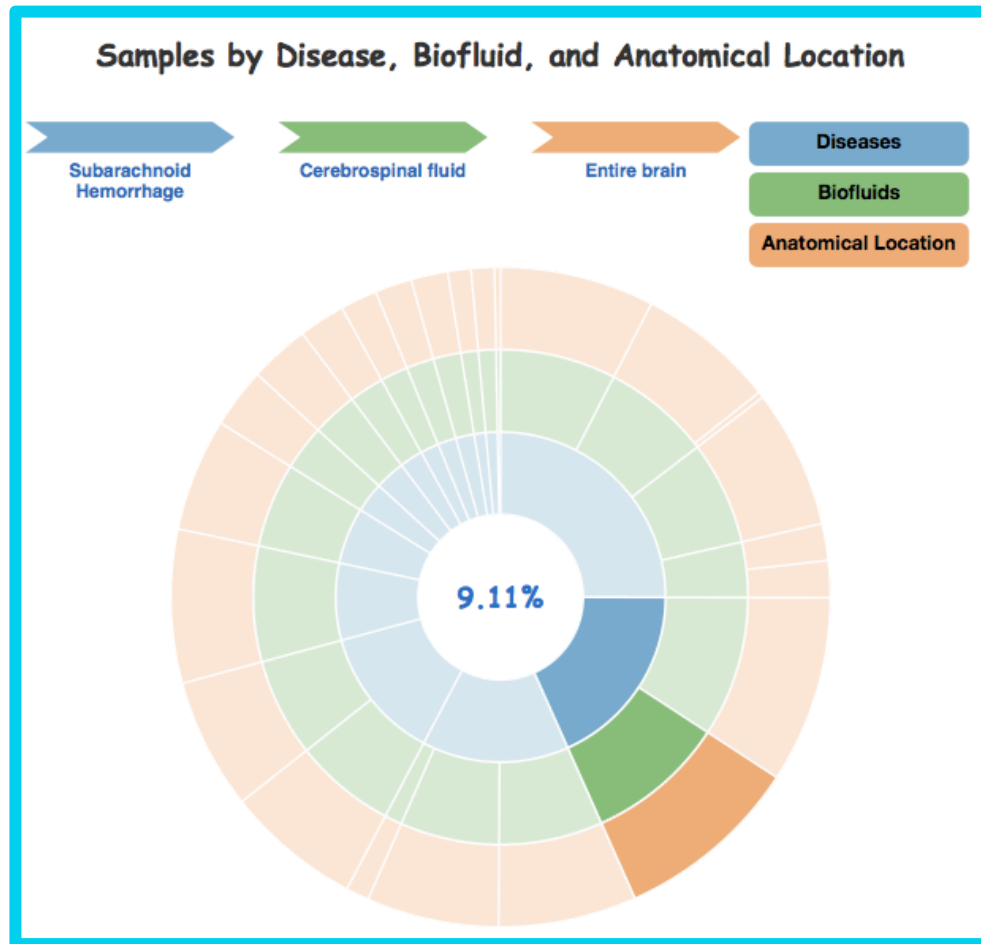


DCC Submission Summary

ERCC DCC Report		
Group: Job Date	Group: RFA Title	
PI Name	Number of Samples	Grant Number
- (U012-U013) Clinical Utility of Extracellular		
Liville C. Laurent	48	1U01HL126493-01
Taylor Patel	3	1U01HL126493-01
Kendall Jensen	345	5U01HL126493-02

Group: Submission Month/Year		Group: RFA Title								
Job Date	PI Name	Submitter's Name	Grant Number	Mode of Submission	Organization of PI	Co PI Names	Processing Pipeline	Submission Category	Project Registered by PI with dbGaP?	Number of Submitted Samples
+ Group: Data Management and Resource Repository (DMRR) on Extracellular RNA (U54) RFA-RM-12-010										
- Group: Defining A Comprehensive Reference Profile of Circulating Human Extracellular RNA (U01) RFA-RM-13-014										
2015-06	David J. Erle	Paula Godoy	1U01HL126493-01	Genboree Workbench	University of California San Francisco	[Not applicable]	exceRpt small RNA-seq	Samples Not Meant for Submission to DCC	Not Applicable	76
2015-07	David J. Erle	Paula Godoy	1U01HL126493-01	Genboree Workbench	University of California San Francisco	[Not applicable]	exceRpt small RNA-seq	Samples Not Meant for Submission to DCC	Not Applicable	4
2015-08	David J. Erle	Paula Godoy	1U01HL126493-01	Genboree Workbench	University of California San Francisco	[Not applicable]	exceRpt small RNA-seq	Samples Not Meant for Submission to DCC	Not Applicable	57
2015-09	Muneesh Tewari	Ryan Spengler	1U01HL126499-01	Genboree Workbench	University of Michigan at Ann Arbor	[Not applicable]	exceRpt small RNA-seq	Samples Not Meant for Submission to DCC	Not Applicable	22
+ Group: Extracellular RNA Biogenesis, Biodistribution, Uptake, and Effector Function (U19) RFA-RM-12-012										
- Group: Non-ERCC Submission										
2015-06	Non-ERCC PI	Alex H	Non-ERCC Funded Study	Genboree Workbench		[Not applicable]	exceRpt small RNA-seq	Samples Not Meant for Submission to DCC	Not Applicable	22

Drill-down search in **exRNA Atlas**



- Use the circular partition diagram (or "sunburst" diagram) to interactively drill down into different subsets of biosamples.
- If you hover over a colored segment, you will see the **Disease » Biofluid » Anatomical Location** values for the biosamples in that subset.
- The percentage of samples falling into that subset will also be displayed.
- If you click a specific segment, you will **zoom** into that subset.
- This drill-down tactic is the best approach for low-population subsets that are hard to select when zoomed out.
- Clicking the **Search** icon in the floating menubar will open the tabular view of your biosample subset.
- Click the center circle to zoom out to the previous level.
- Your last hovered path is always visible. To clear it, click outside the circle.

Tools available at the DMRR

Data Analysis Pipelines & Tools

- exceRpt small RNA-seq Pipeline v3.3 – [Genboree Workbench & FTP Submission](#)
- RSEQtools long RNA-seq Pipeline – [Genboree Workbench](#)
- Target Interaction Finder – [Genboree Workbench](#)
- Pathway Finder – [Genboree Workbench](#)
- Fold Change Calculation using DESeq2 – [Genboree Workbench](#)
- KNIFE – Known and Novel Circular and Linear Isoform Explorer – [Genboree Workbench](#)

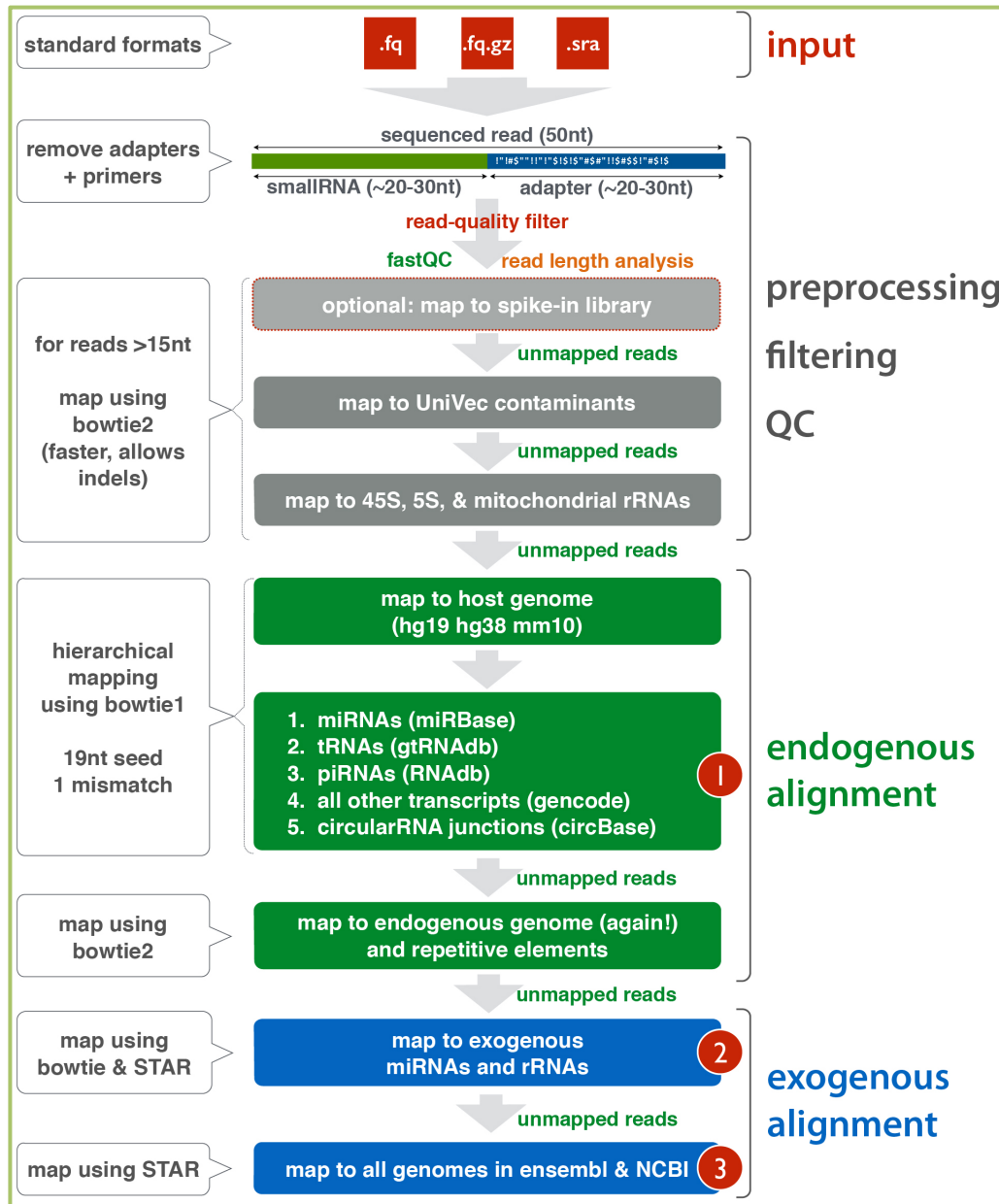
Data Submission Pipelines

- exRNA FTP data submission pipeline for small RNA profiling data

exRNA Metadata Tracking - [GenboreeKB](#)

- Small RNA-Seq Assays
- Long RNA-Seq Assays

exceRpt v3.1.9 Workflow



small RNA-seq
analysis toolkit

exceRpt

quantification (RPM) of
miRNA, tRNA, piRNA,
snoRNA, circularRNA,
& long transcript fragments



Developed by Rob Kitchen at the
Gerstein Lab, Yale University

exceRpt v3.1.9 in Genboree Workbench

GENBOREE TOOL MENU BAR (CONTAINS ALL GENBOREE TOOLS) Baylor College of Medicine

System/Network Data Genomes **Transcriptome** Cistrome Epigenome Metagenome Visualization Help

CLICK HERE TO USE EXCERPT

Welcome to the Genboree

Data Selector

Refresh

Examples and Test Data

Databases

- Atlas SNP2 - Example Data
- Bowtie - Example Data
- BWA hg19 - Example data
- CreateHub hg19 - Example Data
- FastQC - Example Data
- Import Samples - Example Data
- RSEQtools hg18 - Example Data
- RSEQtools hg19 - Example Data
- smallRNA-seq Pipeline - Example Data

Tracks

Lists & Selections

SampleSets

Samples

Files

- exceRptPipeline_v3.1.9
- Placental smallRNAs SRA SRP018255 4-samples
- placental_serum_plasma_SRA_Study_SRP018255_4_samples.tar.gz
- smallRNAseqPipeline_v2.2.8
- spikeinLibraries

Group

Database

Files

Input Data

placental_serum_plasma_SRA_Study_SRP018255_4_samples.tar.gz

DRAG YOUR INPUT DATA FILES HERE

Output Targets

Your Database

DRAG YOUR OUTPUT DATABASE HERE - YOUR RESULT FILES WILL BE UPLOADED TO THIS DATABASE

Tool Settings

Output Location:

Database: smallRNA-seq Pipeline - Example Data Group: Examples and Test Data

exceRpt Pipeline Settings

Analysis Name: exceRpt Pipeline-2015-10-30-11:40:36

Genome Version: hg19

3' Adapter Sequence Options

Adapter Clipped ☐

Input FASTQ File

3' Adapter Sequence

Random Barcode Options

Random Barcode Length: 0

Random Barcode Location: -5p -3p

Oligo (Spike-in) Library Options

Select Library: No custom oligo library

Endogenous Alignment Options

Below, you can select your order of preference for endogenous library alignment. Numbers are listed in order of priority ('1' is higher priority than '2', etc.). By default, the quantification engine will first align to miRNA, then tRNA, then piRNA, then Gencode annotations, and then circular RNA. You can change the order of priority by altering the numbers below. If you do not want to align to a particular library, erase the number for that particular library. You may not choose the same priority for multiple libraries.

miRNA: 1

Remove miRNA Mapping

tRNA: 2

Remove tRNA Mapping

piRNA: 3

Remove piRNA Mapping

Gencode: 4

Remove Gencode Mapping

Circular RNA: 5

Remove circRNA Mapping

Current Priority List: miRNA > tRNA > piRNA > Gencode > circRNA

Maximum Number of Mismatches Allowed: 1

Exogenous Alignment Options

Below, you can select your preference for exogenous library alignment. The first choice, endogenous-only, will disable mapping to exogenous miRNAs. The second choice, endogenous + exogenous (miRNA), will make the pipeline map to exogenous miRNAs in miRBase (i.e., from plants and viruses). Finally, the third choice, endogenous + exogenous (miRNA + Genome), will make the pipeline map to exogenous miRNAs in miRBase AND the genomes of all sequenced species in Ensembl/NCBI. Note that if you choose either the second or third option, then you cannot turn off any of the endogenous mappings above. If you have already turned off any mappings above, then you cannot select either of these options.

Exogenous Alignment: Endogenous-only

Maximum Number of Mismatches Allowed: 1

Advanced Options

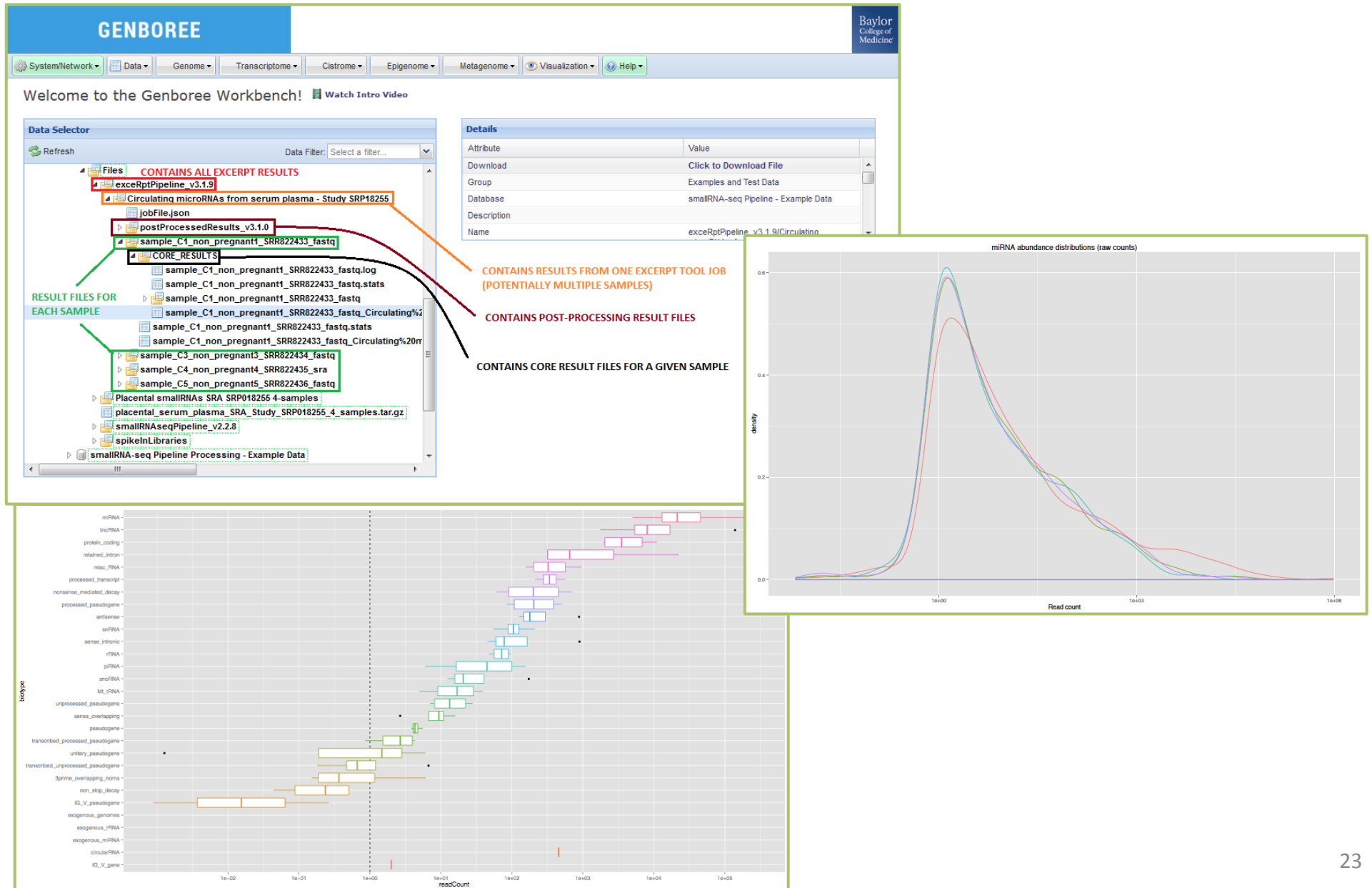
Bowtie Seed Length: 19

Submit Cancel

Tool Settings

- Provide 3' adapter sequence, if known
- Set options for random barcodes – length and location
- Upload custom oligo sequences or use previously uploaded oligo sequences
- Select endogenous libraries for mapping
- Change order of endogenous libraries
- Set mapping options – mismatches, bowtie seed length
- Choose exogenous library alignments

exceRpt v3.1.9 in Genboree Workbench





ERCC Data Analysis Workshop

Topics covered in Use Case 2 Part 2

- Accessing samples in the exRNA Atlas
- **Comparative analysis of exRNA samples using DESeq / exceRpt pipeline results and tools in the Genboree Workbench**
- Finding information about exRNAs of interest
 - on Exocarta and Vesiclepedia
 - on GeneWiki and Wikidata
- Pathway analysis using Cytoscape & WikiPathways in the Genboree Workbench

Using the exRNA Atlas to Perform Comparative and Downstream Analysis of exRNA Samples in the Genboree Workbench

William Thistlethwaite

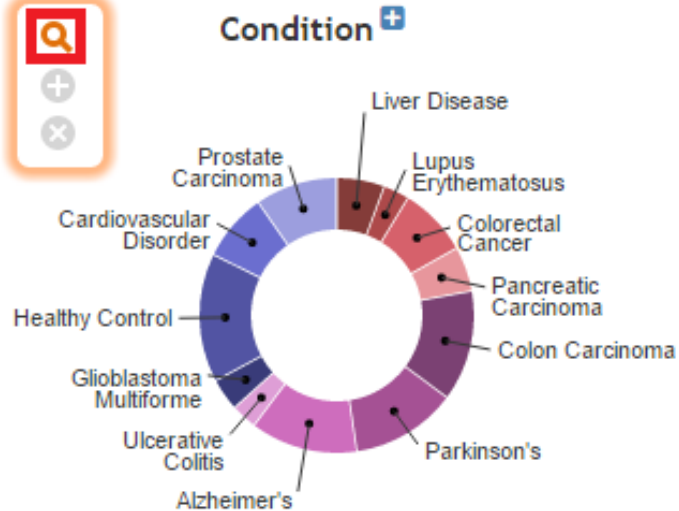
ERC Consortium Data Management & Resource Repository (DMRR)
Baylor College of Medicine, Houston, TX

Selecting Samples of Interest

- Go to the exRNA Atlas website
 - **Public:**
 - <http://genboree.org/exRNA-atlas/index.rhtml>
 - **Private (ERCC Only):**
 - <http://genboree.org/java-bin/exRNA-atlas/index.jsp>
- Use the faceted charts to select your samples of interest.
 - Click the appropriate facets and then click the magnifying glass icon to show corresponding samples in a grid.

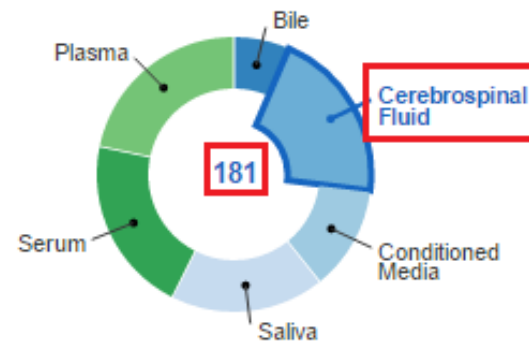
Selecting Samples of Interest

Select exRNA profiles: (181 selected)

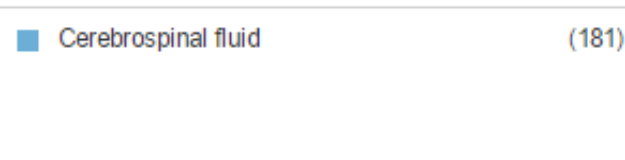


SECOND, CLICK THE MAGNIFYING GLASS ABOVE TO CREATE A GRID WITH YOUR SAMPLES

Biofluid⁺



FIRST, SELECT exRNA PROFILES BY CLICKING THE DESIRED FACETS



Picking Out Samples for Further Analysis

- Pick out the desired samples using the checkboxes to the left of each sample.
 - To select all samples, click the checkbox in the upper left corner of the grid.
- The different metadata columns (Condition, Anatomical Location, etc.) should help you figure out your samples of interest.
- Once you've picked out your samples, click the "**Go to Genboree Workbench**" button to see available tools.

Picking Out Samples for Further Analysis

Search Results - 181 Biosamples					
← Back to Search Page Download Samples Go To Genboree Workbench					
☐	Biosample Name	Condition		Biofluid Name	exRNA Source
☐	0758 Healthy Control CSF	Healthy Control		Cerebrospinal fluid	total cell-free biofluid RNA
☐	9948 Alzheimer's disease CSF	Alzheimer's Disease	Brain ventricle structure	Cerebrospinal fluid	total cell-free biofluid RNA
☐	0888 Healthy Control CSF	Healthy Control	Brain ventricle structure	Cerebrospinal fluid	total cell-free biofluid RNA
☐	1004 Parkinson's Disease CSF	Parkinson's Disease	Brain ventricle structure	Cerebrospinal fluid	total cell-free biofluid RNA
☐	0146 Healthy Control CSF	Healthy Control	Brain ventricle structure	Cerebrospinal fluid	total cell-free biofluid RNA
☒	0855 Healthy Control CSF	Healthy Control	Brain ventricle structure	Cerebrospinal fluid	total cell-free biofluid RNA
☒	9618 Alzheimer's disease CSF	Alzheimer's Disease	Brain ventricle structure	Cerebrospinal fluid	total cell-free biofluid RNA
☒	0438 Healthy Control CSF	Healthy Control	Brain ventricle structure	Cerebrospinal fluid	total cell-free biofluid RNA
☒	0511 Alzheimer's disease CSF	Alzheimer's Disease	Brain ventricle structure	Cerebrospinal fluid	total cell-free biofluid RNA
☐	1016 Alzheimer's disease CSF	Alzheimer's Disease	Brain ventricle structure	Cerebrospinal fluid	total cell-free biofluid RNA
☐	9914 Healthy Control CSF	Healthy Control	Brain ventricle structure	Cerebrospinal fluid	total cell-free biofluid RNA
☐	0207 Alzheimer's disease CSF	Alzheimer's Disease	Brain ventricle structure	Cerebrospinal fluid	total cell-free biofluid RNA
☐	0552 Alzheimer's disease CSF	Alzheimer's Disease	Brain ventricle structure	Cerebrospinal fluid	total cell-free biofluid RNA
☐	9635 Healthy Control CSF	Healthy Control	Brain ventricle structure	Cerebrospinal fluid	total cell-free biofluid RNA

Going from Atlas to Workbench

- Once you've clicked the "Go to Genboree Workbench" button, you can choose a tool to run on your selected samples.
- You will be prompted to log into the Genboree Workbench once you choose a tool. This means that you must have a Genboree account to use the tools.
 - If you have an account already, you can fill in your login information and then click the "Login" button.
 - If you don't have an account, you can click the "Register here" link to create one.
 - Once you've logged in, you won't need to log in again for that Atlas session.

Going from Atlas to Workbench

ExRNA Atlas Log in

GENBOREE

The Genboree page you are trying to access requires you to be logged in.

- If you have access permission, you will be redirected to the page after login.
- Otherwise, there will be an opportunity to ask the group administrators for access to their group.

[GENBOREE HOME](#)

Login Name:

Password:

(Forgot your password?)

New to Genboree? [Register here!](#)



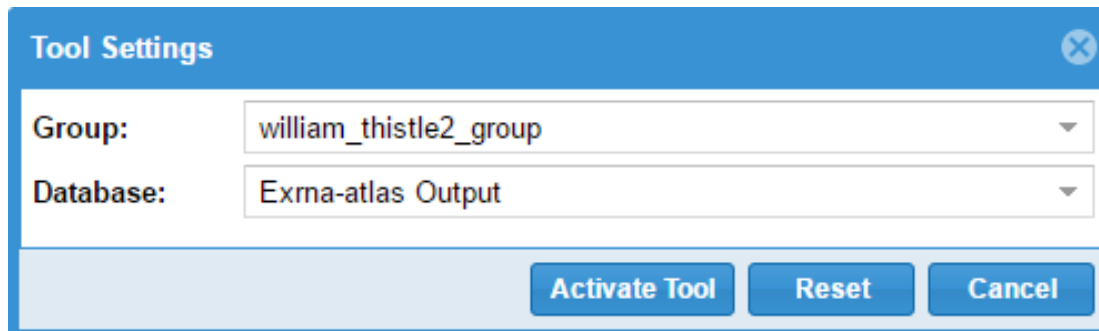
B
R
L

Genboree is built & maintained by the **Bioinformatics Research Laboratory** at **Baylor College of Medicine**.

Genboree is a hosted service. Free open source code is available [here](#).

Going from Atlas to Workbench

- After you've logged in, you'll be able to select the Group and Database which you want to use to store your output files for that tool run.
 - Each Genboree account starts out with a Group (named after your username), but you will need to create a Database in order to run any tools.
 - If the group you select doesn't have a Database, we will offer to create a Database for you (named "Exrna-atlas Output").
 - Pathway Finder does not require a Database to run.



Tool Settings

Group: william_thistle2_group

Database: Exrna-atlas Output

Activate Tool Reset Cancel

Going from Atlas to Workbench

- Once you click "Activate Tool", you will be taken to the Genboree Workbench.
- Your Input Data panel and Output Targets panel will be filled out automatically by the Atlas.
- You can then select your tool of interest, fill out the appropriate settings, and then launch a tool job.
- We will now go into more depth on running
 - the exceRpt pipeline post-processing tool and
 - the DESeq2 differential expression tool

Run Post-Processing Tool

- The post-processing tool will take results from any number of exceRpt small RNA-seq pipeline runs and will condense valuable information into an easy-to-read format.
 - Instead of looking at read counts from 400 different files (each generated by a different sample), look at **1** file that contains all of the most important information!
 - The tool also generates useful plots and histograms for comparative analysis.

Run Post-Processing Tool

- In order to run the tool, select "Run Post-Processing Tool" from the Atlas and choose your Group and Database.
- Then, choose the post-processing tool in the tool menu:
 - Transcriptome -> Analyze Small RNA-Seq Data -> exRNA Data Analysis -> exceRpt small RNA-seq Post-processing
- Finally, fill out your Analysis Name (used for organizing your results on the Workbench) and click "Submit".
- You can follow the progress of your Job using the "Job Summary" tool (found under System/Network -> Jobs).

Run Post-Processing Tool

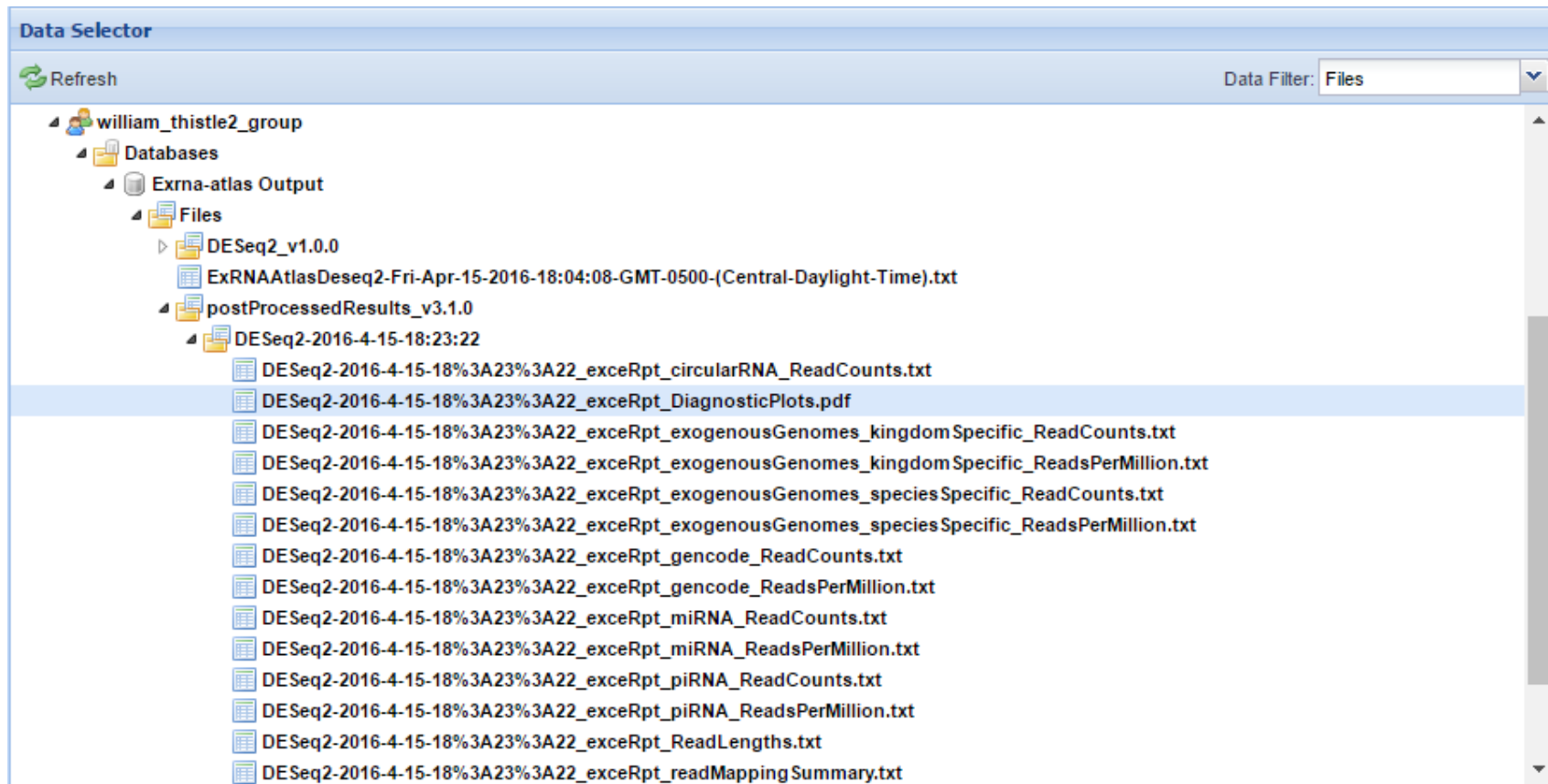
The screenshot displays the GENBOREE web interface, which is part of the Baylor College of Medicine's Genomics and Bioinformatics Research Environment. The interface is organized into several sections:

- Header:** The top of the page features the "GENBOREE" logo on the left and the "Baylor College of Medicine" logo on the right.
- Navigation Bar:** A horizontal bar contains tabs for different data types: "System/Network", "Data", "Genome", "Transcriptome", "Cistrome", "Epigenome", "Metagenome", "Visualization", and "Help".
- Welcome Section:** Below the navigation bar, a "Welcome to the Genboree" message is displayed.
- Data Selector:** On the left, a "Data Selector" panel allows users to browse data sources. It includes a "Refresh" button and a "Data Filter" dropdown. The data sources are organized into a tree structure under "genboree.org", including "Epigenome Informatics Demo Output Data", "Epigenome ToolSet Demo Input Data", "Epigenomics Roadmap Repository", "Examples and Test Data", "exRNAAtlas", "Extracellular RNA Atlas", "Genboree Videos", "GenboreeKB Test", "Registries", "ROI Repository", and "william_thistle2_group". A "Databases" section is also visible, containing "Exrna-atlas Output".
- Transcriptome Menu:** The "Transcriptome" tab is selected, revealing a dropdown menu with options: "Analyze RNA-Seq Data", "Analyze Small RNA-Seq Data", and "Differential Expression Analysis".
- Workflow Selection:** The "Analyze Small RNA-Seq Data" option is selected, leading to a workflow selection screen. This screen shows a list of tools: "exRNA Data Analysis", "Filter Reads", "Map Reads by Pash", and "Profile Combined Coverage". The "exRNA Data Analysis" tool is further expanded, showing sub-options: "exceRpt small RNA-seq Pipeline", "exceRpt small RNA-seq Post-processing" (highlighted with a red box), and "KNIFE (Known and Novel IsoForm Explorer)".
- Input Data:** Below the workflow selection, the "Input Data" section shows a list of sample files: "sample_SAMPLE_0329_PD_CSF_fastq_KJENS1-Alzheimers_Parkinsons-2", "sample_SAMPLE_9948_AD_CSF_fastq_KJENS1-Alzheimers_Parkinsons-2", "sample_SAMPLE_1016_AD_CSF_fastq_KJENS1-Alzheimers_Parkinsons-2", and "sample_SAMPLE_0321_AD_CSF_fasta_KJENS1-Alzheimers_Parkinsons-2".
- Output Targets:** The "Output Targets" section shows a list of output files, including "Exrna-atlas Output".

Outputs from Post-Processing Tool

- Once your job has completed, you can download the result files from the Workbench, using the Data Selector Panel.
- You should navigate to the Group and Database you chose earlier in the Atlas.
- Then, navigate to the following directory:
 - Files -> postProcessedResults_v3.1.0 -> [Your Analysis Name]
- In order to download a given file, click on it in the Data Selector panel and then click the "Click to Download File" link in the Details panel.

Outputs from Post-Processing Tool



Fold Change Calculation Using DESeq2

- This tool will test exRNA samples for differential expression of miRNAs using DESeq2 (version 1.6.3).
- Currently, the tool allows you to test a given factor (disease, for example) across two different factor levels (Alzheimer's disease versus control, for example).
- We will continue to develop this tool and will add new features (like allowing analysis over multiple factors) in the coming months.

Fold Change Calculation Using DESeq2

- In order to run the tool, select "Fold Change Calculation Using DESeq2" from the Atlas and choose your Group and Database.
- Then, choose the DESeq2 tool in the tool menu:
 - Transcriptome -> Differential Expression Analysis -> Fold Change Calculation Using DESeq2

Fold Change Calculation Using DESeq2

The screenshot displays the GENBOREE web application interface. At the top, a blue header bar contains the 'GENBOREE' logo. To the right of the header is the 'Baylor College of Medicine' logo. Below the header is a navigation bar with tabs for 'System/Network', 'Data', 'Genome', 'Transcriptome', 'Cistrome', 'Epigenome', 'Metagenome', 'Visualization', and 'Help'. The 'Transcriptome' tab is currently selected, and its dropdown menu is open, showing three options: 'Analyze RNA-Seq Data', 'Analyze Small RNA-Seq Data', and 'Fold Change Calculation Using DESeq2'. The 'Fold Change Calculation Using DESeq2' option is highlighted with a red rectangular box. Below the navigation bar, the main content area is divided into two columns. The left column, titled 'Data Selector', contains a 'Refresh' button and a 'Data Filter' dropdown set to 'No Filter'. Below these are several data sources listed under 'genboree.org', including 'Epigenome Informatics Demo Output Data', 'Epigenome ToolSet Demo Input Data', 'Epigenomics Roadmap Repository', 'Examples and Test Data', 'exRNAAtlas', 'Extracellular RNA Atlas', 'Genboree Videos', 'GenboreeKB Test', 'Registries', 'ROI Repository', 'william_thistle2_group', and 'Databases'. The right column, titled 'Welcome to the Genboree', contains a table with two columns: 'Attribute' and 'Value'. Below this table are two sections: 'Input Data' and 'Output Targets'. The 'Input Data' section contains a list of sample names, including 'sample_SAMPLE_0329_PD_CSF_fastq_KJENS1-Alzheimers_Parkinsons-2', 'sample_SAMPLE_9948_AD_CSF_fastq_KJENS1-Alzheimers_Parkinsons-2', 'sample_SAMPLE_1016_AD_CSF_fastq_KJENS1-Alzheimers_Parkinsons-2', and 'sample_SAMPLE_0321_AD_CSF_fastq_KJENS1-Alzheimers_Parkinsons-2'. The 'Output Targets' section contains a single entry: 'Exrna-atlas Output'.

Fold Change Calculation Using DESeq2

- After choosing DESeq2, you will need to fill out your tool settings in the dialog box that pops up.
- You can use the Atlas grid to choose your factor name / factor levels.
 - Factor name is column name (Condition or Biofluid Name for now)
 - Factor levels can be found within those columns (Alzheimer's Disease, Healthy Control, etc.).
- Second factor level is used as base level (should probably be your control).

Fold Change Calculation Using DESeq2

Search Results - 181 Biosamples

[Back to Search Page](#)
[Download Samples](#)
[Go To Genboree Workbench](#)

☐	Biosample Name	Condition	FACTOR NAMES	Biofluid Name
☐	0758 Healthy Control CSF	Healthy Control	Cerebrospinal fluid	
☐	9948 Alzheimer's disease CSF	Alzheimer's Disease	Cerebrospinal fluid	
☐	0888 Healthy Control CSF	Healthy Control	Cerebrospinal fluid	
☐	1004 Parkinson's Disease CSF	Parkinson's Disease	Cerebrospinal fluid	
☐	0146 Healthy Control CSF	Healthy Control	Cerebrospinal fluid	
☐	0855 Healthy Control CSF	Healthy Control	Cerebrospinal fluid	
☐	9618 Alzheimer's disease CSF	Alzheimer's Disease	Cerebrospinal fluid	
☐	0438 Healthy Control CSF	Healthy Control	Cerebrospinal fluid	
☐	0511 Alzheimer's disease CSF	Alzheimer's Disease	Cerebrospinal fluid	
☐	1016 Alzheimer's disease CSF	Alzheimer's Disease	Cerebrospinal fluid	
☐	9914 Healthy Control CSF	Healthy Control	Cerebrospinal fluid	
☐	0207 Alzheimer's disease CSF	Alzheimer's Disease	Cerebrospinal fluid	
☐	0552 Alzheimer's disease CSF	Alzheimer's Disease	Cerebrospinal fluid	
☐	9635 Healthy Control CSF	Healthy Control	Cerebrospinal fluid	

FACTOR LEVELS

FACTOR LEVELS

Fold Change Calculation Using DESeq2

- If you are an ERCC member, you should also fill out the ERCC Submission Options section.
 - This section helps us keep track of tool usage, and we can use the information to improve the tools and get valuable user feedback.
- Finally, you can choose to have your results uploaded to our FTP server via the "Remote Storage Area" option in the "Advanced Options" section. To learn more, follow this link:
 - [http://genboree.org/theCommons/projects/exrna-tools-may2014/wiki/Using_Remote_\(FTP\)_Storage_for_exceRpt](http://genboree.org/theCommons/projects/exrna-tools-may2014/wiki/Using_Remote_(FTP)_Storage_for_exceRpt)

Fold Change Calculation Using DESeq2

Tool Settings

DESeq2 Settings

Analysis Name

DESeq2-2016-4-15-23:30:55

Analysis Type

Below, you can select the factor name / factor levels for your DESeq2 analysis. For example, if your sample descriptor document has a "disease" factor, you can select that factor and then choose a pair of factor levels to consider ("Alzheimer's Disease" versus "Control", for example). Please note that the second factor level is used as the base level in your DESeq2 analysis. Thus, if one of your factor levels is a control, then we recommend using that factor level as the second factor level.

Factor Name

Condition

Factor Level 1

Alzheimer's Disease

Factor Level 2 (Base)

Healthy Control

ERCC Submission Options

If you are not a member of the ERCC, ignore this section. If you are a member of the ERCC, choose the appropriate options for your submission with respect to your grant number and anticipated data repository. If your submission does not fall under an ERCC grant, then choose the 'Non-ERCC Funded Study' option. If you are an ERCC member and your PI / grant numbers are not showing up properly, please email Sai at sailakss@bcm.edu with your PI's name so you can be added to our database as a submitter.

ERCC PI

Aleksandar Milosavljevic

ERCC Grant Number

Non-ERCC Funded Study

Anticipated Data Repository

None

Advanced Options

Below, you can select your preference for advanced options that don't fit into the other categories above. The 'Remote Storage Area' option will allow you to choose a remote storage (FTP) area where your result files will be uploaded. You can learn more by visiting the [tutorial](#).

Remote Storage Area

None Selected

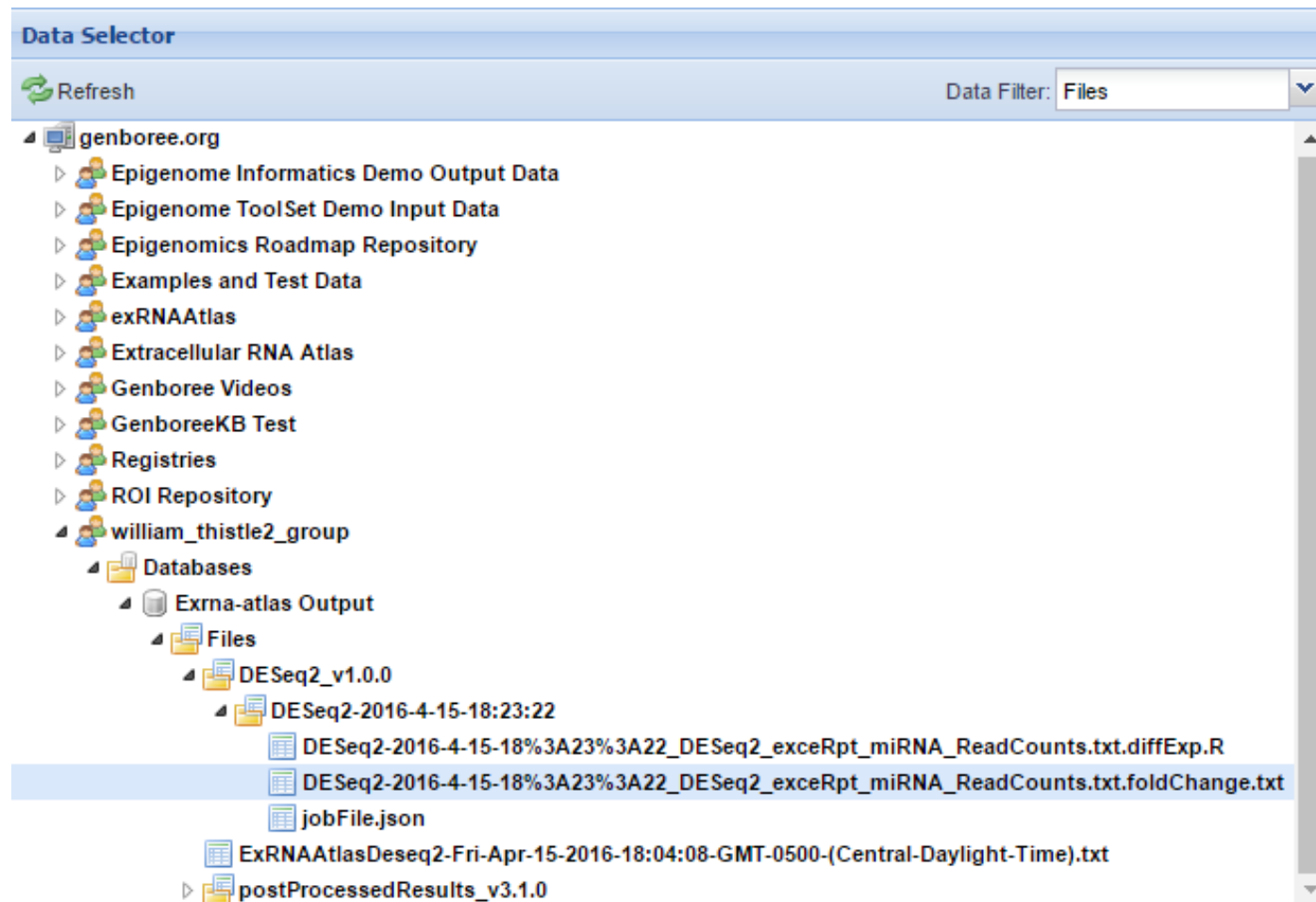
Submit

Cancel

Outputs from DESeq2

- Once your job has completed, you can download the result files from the Workbench, using the Data Selector Panel.
- You should navigate to the Group and Database you chose earlier in the Atlas.
- Then, navigate to the following directory:
 - Files -> DESeq2_v1.0.0 -> [Your Analysis Name]
- In order to download a given file, click on it in the Data Selector panel and then click the "Click to Download File" link in the Details panel.

Outputs from DESeq2

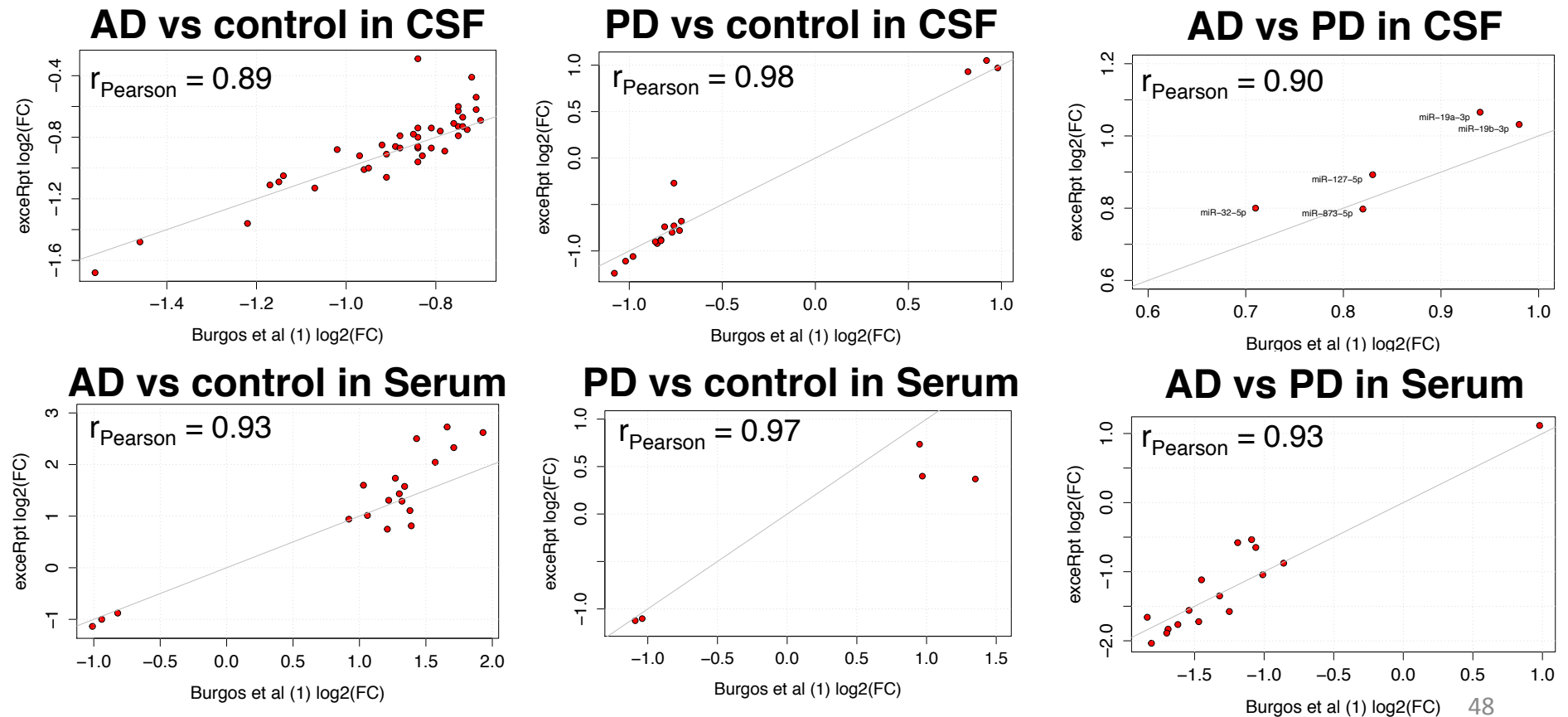




Use Case 2: miRNA Changes in Alzheimer's and Parkinson's Disease



Below is a comparison of the differential expression analysis results from Burgos et al. with the results from Genboree Workbench, namely running DESeq2 on the output of the exceRpt pipeline.

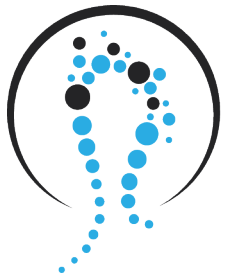




ERCC Data Analysis Workshop

Topics covered in Use Case 2 Part 2

- Accessing samples in the exRNA Atlas
- Comparative analysis of exRNA samples using DESeq / exceRpt pipeline results and tools in the Genboree Workbench
- **Finding information about exRNAs of interest**
 - on Exocarta and Vesiclepedia
 - on GeneWiki and Wikidata
- Pathway analysis using Cytoscape & WikiPathways in the Genboree Workbench



Vesiclepedia and Exocarta

You can access both Vesiclepedia and Exocarta from the Quick Links sidebar at exRNA.org.



exRNA.org

Quick Links

ExRNA Atlas

ExRNA @ BioGPS

ExRNA @ WikiPathways

ExoCarta

Vesiclepedia

miRandola



Vesiclepedia and Exocarta



- a compendium of proteins, RNA and lipids found in exosomes
- manually curated from the scientific literature
- available since 2009

Mathivanan et al. *Proteomics* 2009

Keerthikumar et al. *J. Mol. Biol.* 2016



- a compendium of proteins, RNA, and lipids found in all types of EVs
- manually curated from the scientific literature
- focuses strongly on community annotation
- available since 2012

Kalra et al. *PLoS Biology* 2012



Vital Statistics



STATISTICS	
Studies	286
Protein entries	41,860
Proteins	9,769
mRNA entries	4,946
mRNA	3,408
miRNA	2,838
Lipid entries	1,116



STATISTICS	
Studies	538
Protein entries	92,897
mRNA entries	27,642
miRNA entries	4,934
Lipid entries	584
Species	33

	Exocarta	Vesiclepedia	In Common
All Publications	132	271	108
Human Publications	100	190	78



Exocarta

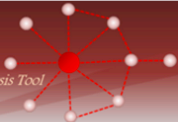
[Credits](#) | [Contact](#) | [Vesiclepedia](#)

ExoCarta

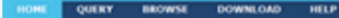
[HOME](#) | [QUERY](#) | [BROWSE](#) | [DOWNLOAD](#) | [DATA SUBMISSION](#) | [HELP](#)

EXOSOME PROTEIN, RNA AND LIPID DATABASE

Exosomes are 30-150 nm membrane vesicles of endocytic origin secreted by most cell types *in vitro*. ExoCarta, an exosome database, provides with the contents that were identified in exosomes in multiple organisms.



Perform bioinformatics analysis of your extracellular vesicle data set using FunRich, a open access standalone tool.



Vesiclepedia - Compendium on extracellular vesicles

STATISTICS

Studies	286
Protein entries	41,860
Proteins	9,769
mRNA entries	4,946
mRNA	3,408
miRNA	2,838
Lipid entries	1,116

EXOSOMAL MARKERS

Statistics of proteins that are identified more often in exosomes.

[download ...](#)



Vesiclepedia



A community compendium for extracellular vesicles

Extracellular vesicles (EVs) are membraneous vesicles released by a variety of cells into the extracellular microenvironment. Based on the mode of biogenesis, EVs can be classified into three broad classes (i), ectosomes or shedding microvesicles (ii), exosomes and (iii), apoptotic bodies. Recent studies have ignited significant interest on EVs by elucidating their role in intercellular communication, pathogenesis, drug, vaccine and gene-vector delivery and as possible reservoirs of biomarkers. With such immense interest, the amount of data generated has increased exponentially. Here, we describe Vesiclepedia, a manually curated compendium of molecular data (lipid, RNA and protein) identified in different classes of EVs. Currently, Vesiclepedia comprises 35,264 protein, 18,718 mRNA, 1,772 miRNA and 342 lipid entries encompassed from 341 independent studies that were published over the past several years. Even though databases are indispensable resources for the scientific community, recent studies have shown that more than 50% of the databases are not updated for a long time. In addition, more than 20% of the database links are not active after its initial publication. To prevent such database decay and keep them updated, for the first time, we have initiated a continuous community annotation project with the active involvement of the EV researchers with long standing (> 10 years) and newly found reputation in the field. The principal investigators working on EVs have vowed to contribute data generated by their group in a regular and continuous fashion making this effort significantly unique. By undertaking such initiatives, we foresee the EV research community as trendsetters in data sharing. We expect, Vesiclepedia to evolve as a primary resource for EV research.

STATISTICS	
Studies	538
Protein entries	92,897
mRNA entries	27,642
miRNA entries	4,934
Lipid entries	584
Species	33

LATEST NEWS

09/01/2015 - Vesiclepedia Version 3.1 released!!

09/01/2015 - Addition of new datasets to Vesiclepedia.

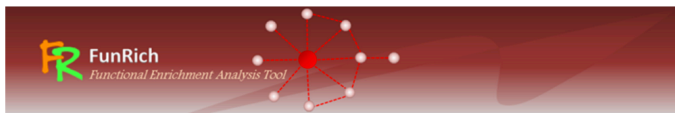
01/09/2013 - Vesiclepedia Version 2.1 released!!

01/09/2013 - Addition of new datasets to Vesiclepedia.

01/04/2013 - Vesiclepedia highlighted in Nature Methods

01/04/2013 - Manuscript describing Vesiclepedia was highlighted in the tool section of Nature Methods.

[PubMed](#)



Perform bioinformatics analysis of your extracellular vesicle data set using [FunRich](#), a open access standalone tool.



Differential Expression of miRNAs in CSF

Control vs Alzheimer's

miR Name	Burgos et al		exceRpt + DESeq2	
	log2(FC)	P-value	log2(FC)	P-value
miR-124-3p	-1.56	1.00E-06	-1.68	2.00E-06
miR-138-5p	-1.46	1.00E-06	-1.48	7.00E-06
miR-127-3p	-1.17	1.90E-05	-1.11	2.55E-04
miR-132-3p	-0.89	1.90E-05	-0.86	1.29E-04
miR-127-5p	-1.15	3.10E-05	-1.09	3.07E-04
miR-136-3p	-1.02	4.00E-05	-0.88	1.16E-03
miR-381(-3p)*	-1.14	4.90E-05	-1.05	9.86E-04
miR-101-5p	-0.92	6.70E-05	-0.85	9.86E-04
miR-199b-5p	-1.22	6.70E-05	-1.36	1.29E-04
miR-136-5p	-0.91	1.91E-04	-0.91	7.73E-04

We will focus on the
top 10 most differentially expressed miRNAs
in CSF of Alzheimers' patients vs. healthy controls.

Burgos K., et al. (2014) Profiles of Extracellular miRNA in Cerebrospinal Fluid and Serum from Patients with Alzheimer's and Parkinson's Diseases Correlate with Disease Status and Features of Pathology.

PLoS ONE 9: e94839.





Querying Exocarta for miRNAs

http://exocarta.org/mirna_summary?species=Homo%20sapiens&mirna_id=miR-124-3p

Control vs Alzheimer's

miR Name

miR-124-3p

miR-138-5p

miR-127-3p

miR-132-3p

miR-127-5p

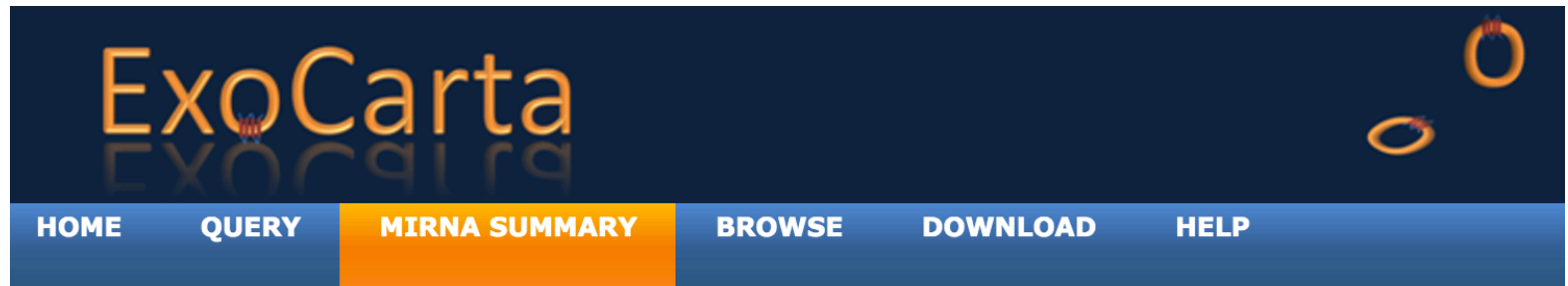
miR-136-3p

miR-381(-3p)

miR-101-5p

miR-199b-5p

miR-136-5p



>> miRNA description for miR-124-3p

miRNA accession	miR-124-3p
Species	Homo sapiens

>> miR-124-3p identified in exosomes derived from the following tissue/cell type

Colorectal cancer cells	Unpublished / Not applicable
Colorectal cancer cells	Unpublished / Not applicable
Melanoma cells	26176991 PubMed
Plasma	23663360 PubMed

Burgos K., et al. (2014) Profiles of Extracellular miRNA in Cerebrospinal Fluid and Serum from Patients with Alzheimer's and Parkinson's Diseases Correlate with Disease Status and Features of Pathology.

PLoS ONE 9: e94839.





Querying Exocarta for miRNAs

http://exocarta.org/mirna_summary?species=Homo%20sapiens&mirna_id=miR-124-3p

Control vs Alzheimer's

miR Name

miR-124-3p

miR-138-5p

miR-127-3p

miR-132-3p

miR-127-5p

miR-136-3p

miR-381(-3p)

miR-101-5p

miR-199b-5p

miR-136-5p

>> Experiment description of studies that identified miR-124-3p in exosomes

Experiment ID	218
Identified molecule	miRNA 
Identification method	Human miRNA Microarray Kit
PubMed ID	Unpublished / Not applicable
Organism	Homo sapiens
Experiment description	Characterization of colorectal cancer cell exosomal microRNA content in vitro
Authors	-
Journal name	Not applicable
Publication year	0
Sample	Colorectal cancer cells
Sample name	HCT116 HT29
Molecules identified in the study	miRNA
Methods used in the study	miRCURY LNA microRNA Array 7th Gen
Comments	NOTFOUND NOTFOUND

Burgos K., et al. (2014) Profiles of Extracellular miRNA in Cerebrospinal Fluid and Serum from Patients with Alzheimer's and Parkinson's Diseases Correlate with Disease Status and Features of Pathology.

PLoS ONE 9: e94839.





Querying Exocarta for miRNAs

http://exocarta.org/mirna_summary?species=Homo%20sapiens&mirna_id=miR-124-3p

Control vs Alzheimer's

miR Name

miR-124-3p

miR-138-5p

miR-127-3p

miR-132-3p

miR-127-5p

miR-136-3p

miR-381(-3p)

miR-101-5p

miR-199b-5p

miR-136-5p

Experiment ID	265
Identified molecule	miRNA 
Identification method	RNA deep sequencing
PubMed ID	26176991 
Organism	Homo sapiens
Experiment description	Small RNA deep sequencing discriminates subsets of extracellular vesicles released by melanoma cells – evidence of unique microRNA cargos
Authors	Lunavat TR, Cheng L, Kim DK, Bhadury J, Jang SC, Lasser C, Sharples RA, Lopez MD, Nilsson J, Gho YS, Hill AF, Lotvall J.
Journal name	RNA Biology
Publication year	2015
Sample	Melanoma cells
Sample name	MML-1
Molecules identified in the study	Protein miRNA snRNA snoRNA mtRNA Ro associated Y-RNA MiscRNA
Methods used in the study	Western blotting RNA deep sequencing
Comments	NOTFOUND NOTFOUND

Burgos K., et al. (2014) Profiles of Extracellular miRNA in Cerebrospinal Fluid and Serum from Patients with Alzheimer's and Parkinson's Diseases Correlate with Disease Status and Features of Pathology.

PLoS ONE 9: e94839.





Querying Exocarta for miRNAs

http://exocarta.org/mirna_summary?species=Homo%20sapiens&mirna_id=miR-124-3p

Control vs Alzheimer's

miR Name

miR-124-3p

miR-138-5p

miR-127-3p

miR-132-3p

miR-127-5p

miR-136-3p

miR-381(-3p)

miR-101-5p

miR-199b-5p

miR-136-5p

Experiment ID	222
Identified molecule	miRNA 
Identification method	RT-PCR
PubMed ID	23663360 
Organism	Homo sapiens
Experiment description	Characterization of human plasma-derived exosomal RNAs by deep sequencing
Authors	Huang X, Yuan T, Tschannen M, Sun Z, Jacob H, Du M, Liang M, Dittmar RL, Liu Y, Liang M, Kohli M, Thibodeau SN, Boardman L, Wang L
Journal name	BMC Genomics
Publication year	2013
Sample	Plasma
Sample name	Plasma - Normal
Molecules identified in the study	miRNA tRNA snRNA snoRNA rRNA piRNA lncRNA
Methods used in the study	RT-PCR
Comments	NOTFOUND NOTFOUND

Burgos K., et al. (2014) Profiles of Extracellular miRNA in Cerebrospinal Fluid and Serum from Patients with Alzheimer's and Parkinson's Diseases Correlate with Disease Status and Features of Pathology.

PLoS ONE 9: e94839.





Querying Vesiclepedia for miRNAs

http://microvesicles.org/mirna_summary?species=Homo%20sapiens&mirna_id=miR-124-3p

Control vs Alzheimer's

miR Name

miR-124-3p

miR-138-5p

miR-127-3p

miR-132-3p

miR-127-5p

miR-136-3p

miR-381(-3p)

miR-101-5p

miR-199b-5p

miR-136-5p



Vesiclepedia

AGCTCTGCGCTG

[QUERY](#) [MIRNA SUMMARY](#) [BROWSE](#) [DOWNLOAD](#) [HELP](#)

**miRNA description for miR-124-3p**

miRNA accession	miR-124-3p
Species	Homo sapiens

**miR-124-3p identified in exosomes derived from the following tissue/cell type**

Urine [Exosomes]	24352158 
------------------	--

**Experiment description of studies that identified miR-124-3p in exosomes**

Experiment ID	357
Identified molecule	miRNA 
Identification method	RNA deep sequencing
PubMed ID	24352158 
Organism	Homo sapiens
Experiment description	Characterization and deep sequencing analysis of exosomal miRNA in human urine
Authors	Lesley Cheng, Xin Sun, Bradley M. Coleman, Andrew F. Hill
Journal name	Not applicable
Publication year	0
Sample	Urine
Sample name	Urine - Normal
Molecules identified in the study	miRNA
Methods used in the study	RNA deep sequencing



Downloading Vesiclepedia and Exocarta

You can download both the Exocarta and Vesiclepedia databases as structured text files.

ExoCarta Download - 5 (Release date: 29 July 2015)	
Protein/mRNA data	<u>ExoCarta protein mRNA details 5.txt</u>
miRNA data	<u>ExoCarta miRNA details 5.txt</u>
Lipid data	<u>ExoCarta lipid details 5.txt</u>
Exosome study details	<u>ExoCarta experiment details 5.txt</u>
Gene details	<u>ExoCarta gene details 5.txt</u>
Top 100 proteins	<u>ExoCarta top100 protein details 5.txt</u>

Vesiclepedia Download - Version 3 (Release date: 9 Jan 2015)	
Protein/mRNA data	<u>Vesiclepedia protein mRNA details.txt</u>
miRNA data	<u>Vesiclepedia miRNA details.txt</u>
Lipid data	<u>Vesiclepedia lipid details.txt</u>
Experiment study details	<u>Vesiclepedia experiment details.txt</u>
Gene details	<u>Vesiclepedia gene details.txt</u>
PTM details	<u>Vesiclepedia PTM details.txt</u>



Downloading Vesiclepedia and Exocarta

The miRNA and Experiment structured text files have the following fields.

<u>miRNA</u>	<u>Experiment</u>
CONTENT ID	EXPERIMENT ID
CONTENT TYPE	PUBMED ID
MIRNA ID	SPECIES
SPECIES	EXPERIMENT DESCRIPTION
EXPERIMENT ID	SAMPLE
METHODS	SAMPLE SOURCE
ENTREZ GENE ID	SAMPLE NAME
COMMENTS	IDENTIFICATIONS
	METHODS
	YEAR
	ISOLATION METHOD
	VESICLE TYPE (Vesiclepedia only)



Downloading Vesiclepedia and Exocarta

Control vs Alzheimer's miR Name

10 input miRs

0 miRs with no info in Exocarta

7 miRs with no info in Vesiclepedia

miRNA	Experiments where found (E=Exocarta, V=Vesiclepedia)					
miR-124-3p	e218	e222	e265	v357		
miR-138-5p	e218	e222	v357			
miR-127-3p	e222					
miR-132-3p	e222	e223				
miR-127-5p	e222					
miR-136-3p	e222					
miR-381	e226					
miR-101-5p	e222					
miR-199b-5p	e117	e115	e116	e218	e222	
	v117	v115	v116			
miR-136-5p	e218	e222				

After a bit of Perl scripting, we find the above associations between studies cataloged in Exocarta and Vesiclepedia and the top 10 differentially expressed miRNAs in the AD vs control comparison.



Downloading Vesiclepedia and Exocarta

Control vs Alzheimer's

miR Name

miR-124-3p
miR-138-5p
miR-127-3p
miR-132-3p
miR-127-5p
miR-136-3p
miR-381(-3p)
miR-101-5p
miR-199b-5p
miR-136-5p

Exp#	PMID	Cell Type	miRNAs found		
e115	21505438	T	miR-199b-5p		
e116	21505438	B	miR-199b-5p		
e117	21505438	dendritic	miR-199b-5p		
e218	--	colorectal cancer	miR-124-3p	miR-138-5p	
e222	23663360	plasma	miR-199b-5p	miR-136-5p	
			miR-124-3p	miR-138-5p	miR-127-3p
			miR-132-3p	miR-127-5p	miR-136-3p
			miR-101-5p	miR-199b-5p	miR-136-5p
e223	25330373	colon carcinoma	miR-132-3p		
e226	26027894	endothelial	miR-381		
e265	26176991	melanoma	miR-124-3p		
v115	21505438	T	miR-199b-5p		
v116	21505438	B	miR-199b-5p		
v117	21505438	dendritic	miR-199b-5p		
v357	24352158	urine	miR-124-3p	miR-138-5p	

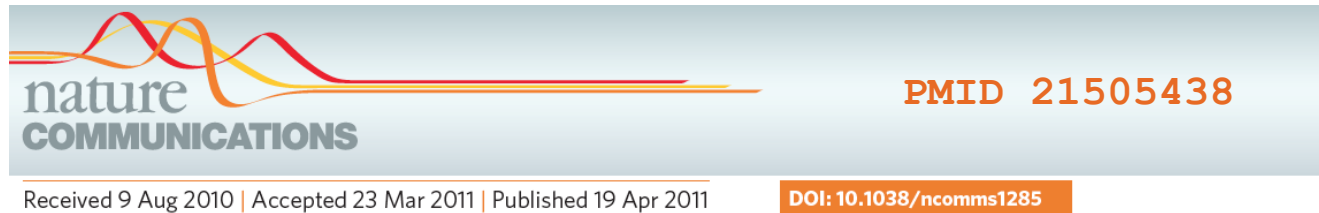
Here we have listed the Pubmed IDs of each publication, along with the miRNAs associated with each one.



Results from Vesiclepedia and Exocarta

Control vs Alzheimer's

miR Name	Burgos et al		exceRpt + DESeq2	
	log2(FC)	P-value	log2(FC)	P-value
miR-199b-5p	-1.22	6.70E-05	-1.36	1.29E-04



Unidirectional transfer of microRNA-loaded exosomes from T cells to antigen-presenting cells

María Mittelbrunn^{1,*}, Cristina Gutiérrez-Vázquez^{1,*}, Carolina Villarroja-Beltrí¹, Susana González¹, Fátima Sánchez-Cabo¹, Manuel Ángel González¹, Antonio Bernad¹ & Francisco Sánchez-Madrid^{1,2}

miR-199b-5p

Comparison	Rank	FC(log2)	Adjusted p-value
Raji Cells vs Exosomes	42	-2.08	2.11E-02
DC Cells vs Exosomes	116	2.37	1.38E-02
J77 Cells vs Exosomes	166	1.79	1.49E-02



Results from Vesiclepedia and Exocarta



Characterization of human plasma-derived exosomal RNAs by deep sequencing

Xiaoyi Huang¹, Tiezheng Yuan¹, Michael Tschannen², Zhifu Sun³, Howard Jacob², Meijun Du¹, Meihua Liang⁴, Rachel L Dittmar¹, Yong Liu⁵, Mingyu Liang⁵, Manish Kohli⁶, Stephen N Thibodeau⁷, Lisa Boardman⁶ and **Liang Wang¹***

Exp#	PMID	Cell Type	miRNAs found		
e222	23663360	plasma	miR-124-3p	miR-138-5p	miR-127-3p
			miR-132-3p	miR-127-5p	miR-136-3p
			miR-101-5p	miR-199b-5p	miR-136-5p

miR-124-3p and **miR-127-3p** were among the 20 most abundant RNAs found in plasma exosomes.

“The five most abundant miRNAs in the libraries were miR-99a-5p, miR-128, **miR-124-3p**, miR-22-3p, and miR-99b-5p, which together accounted for 48.99% of all detectable miRNAs.”



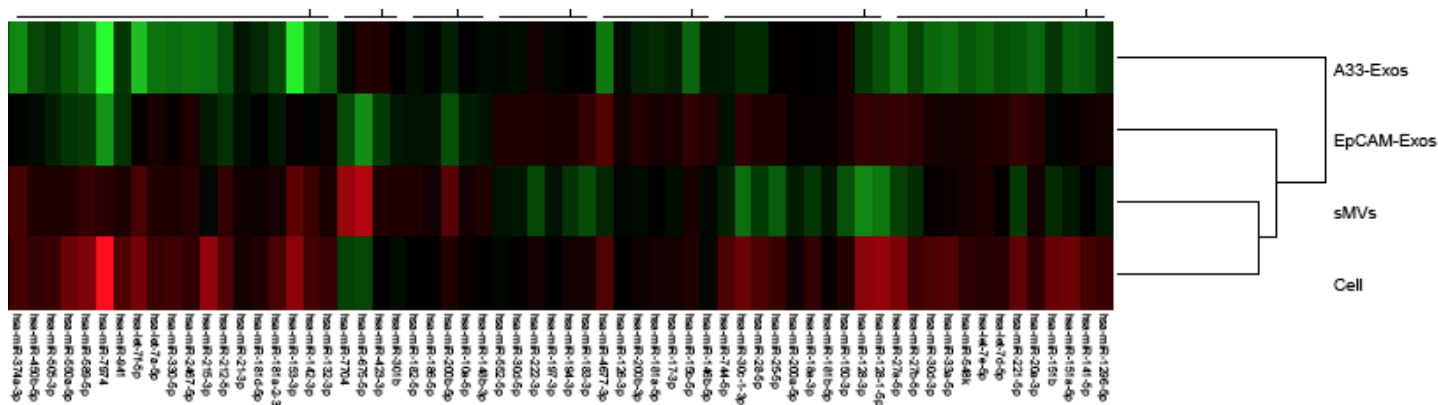
Results from Vesiclepedia and Exocarta

Exp#	PMID	Cell Type	miRNAs found
e223	25330373	colon carcinoma	miR-132-3p



October 2014 | Volume 9 | Issue 10 | e110314

Deep sequencing of RNA from three different EV subtypes released from the human LIM1863 colon cancer cell line uncovers distinct miRNA-enrichment signatures



miR-132-3p is depleted in immuno-affinity isolated A33 exosomes



Results from Vesiclepedia and Exocarta

Exp#	PMID	Cell Type	miRNAs found
e226	26027894	endothelial	miR-381

Table I. Top 15 miRNAs in cells

ID	Rank in cells	Rank in exosomes
miR-10B	1	1
miR-30A	2	3
miR-27B	3	2
miR-191	4	4
miR-411	5	7
miR-92B	6	16
miR-LET7I	7	5
miR-222	8	15
miR-100	9	10
miR-30E	10	17
miR-381	11	13



Journal of
Extracellular Vesicles

Quantitative and qualitative analysis of small RNAs in human endothelial cells and exosomes provides insights into localized RNA processing, degradation and sorting

Bas W. M. van Balkom^{1*}, Almut S. Eisele¹, D. Michiel Pegtel², Sander Bervoets³ and Marianne C. Verhaar¹

¹Department of Nephrology and Hypertension, UMC Utrecht, Utrecht, the Netherlands; ²Exosomes Research Group, VU University Medical Center, Amsterdam, the Netherlands; ³ServiceXS B.V., Leiden, the Netherlands



Results from Vesiclepedia and Exocarta

Exp#	PMID	Cell Type	miRNAs found	
v357	24352158	urine	miR-124-3p	miR-138-5p

Characterization and deep sequencing analysis of exosomal and non-exosomal miRNA in human urine

Kidney International 2014 86: 433

In a technical study of methods for extracting miRNA from exosomes from urine, the authors found that **miR-124-3p** was only found when purified by a Norgen Biotek Urine Exosomal RNA kit, while **miR-138-5p** was not found by that kit but was found by ultracentrifugation.



FunRich



Finding relevant literature about small sets of miRNAs in Vesiclepedia is a piecemeal approach.

For a more integrative method, you can try FunRich. Currently a Windows application, it will be released as a tool on the Genboree Workbench this year.



Vesiclepedia enables **community annotation**

Membership on three conditions:

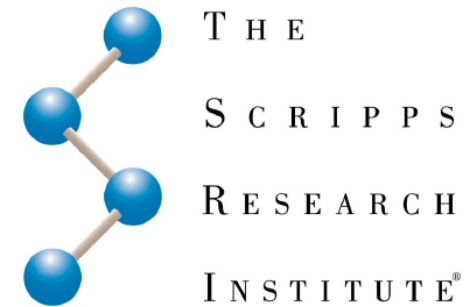
- 1) The laboratory head or group leader needs to sign up.
- 2) You must agree to submit datasets pertaining to extracellular vesicles continuously to Vesiclepedia, either before or after publication.
- 3) When reviewing articles, please mandate or request investigators to submit datasets on extracellular vesicles to Vesiclepedia.

Credit

All members were co-authors on the initial Vesiclepedia manuscript.

<http://www.microvesicles.org>

Kalra et al. PLoS Biology, 2012



Wikidata for Interpretation of miRNA RNA-seq data

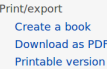
Sebastian Burgstaller-Muehlbacher, PhD

 @sebotic

<https://commons.wikimedia.org/wiki/File:Wikidata-logo-en.svg>

Contents

- Wikidata basics
 - Graph database
 - Document based
 - Semantic web integration
- Example
 - Data: Jensen Alzheimers and Parkinson disease data set.
 - Aim: Retrieve extracellular miRNA target genes and Gene Ontology terms.



- What links here
- Related changes
- Special pages
- Permanent link
- Page information
- Concept URI
- Cite this page



- [Main page](#)
- [Contents](#)
- [Featured content](#)
- [Current events](#)
- [Random article](#)
- [Donate to Wikipedia](#)
- [Wikipedia store](#)

- [Help](#)
- [About Wikipedia](#)
- [Community portal](#)
- [Recent changes](#)
- [Contact page](#)

[What links here](#)
[Related changes](#)
[Upload file](#)
[Special pages](#)
[Permanent link](#)
[Page information](#)
[Wikidata item](#)
[Cite this page](#)

Print/export

- Create a book
- Download as PDF
- Printable version

Languages
 العربية
 Ελληνικά
 Français
 日本語
 Polski

 Edit links

pharmaceutical drug
MDX-1106 | ONO-4538 | BMS-936558

▼ In more languages

Language	Label	Description	Also known as
English	Nivolumab	pharmaceutical drug	MDX-1106 ONO-4538 BMS-936558
German	Nivolumab	monoklonaler PD-1 Antikörper	anti-PD-1
French	Nivolumab	No description defined	

CAS registry number **946414-94-4** [edit](#)

[1 reference](#)

[+ add](#)

UNII  31Y063LBSN  [edit](#)

[▶ 1 reference](#)

[+ add](#)

instance of	chemical compound edit ↳ 1 reference
	pharmaceutical drug edit ↳ 1 reference
	biopharmaceutical edit ↳ 1 reference
	monoclonal antibodies edit ↳ 1 reference
	+ add

ChEMBL ID  CHEMBL2108738  [edit](#)

[▶ 1 reference](#)

[+ add](#)

Wikipedia (6 entries)  [edit](#)

ar	نیفولوماب
el	Nivolumab
en	Nivolumab
fr	Nivolumab
ja	ニボルマブ
pl	Anti-PD-1

Wikibooks (0 entries)  [edit](#)

Wikinews (0 entries)  [edit](#)

Wikiquote (0 entries)  [edit](#)

From Wikipedia, the free encyclopedia

Nivolumab (nye voh-lee mab; **ONO-4538**, **BMS-936558**, or **MDX1106**), marketed as **Opdivo**, is a human IgG4 anti-**PD-1 monoclonal antibody** developed by Ono Pharmaceutical and **Medarex** (later acquired by **Bristol-Myers Squibb**) for the treatment of **cancer**.^{[1][2]} Nivolumab acts as an **immunomodulator** by blocking ligand activation of the **programmed cell death 1** (PD-1) receptor on activated **T cells**.

Nivolumab is approved by the [Food and Drug Administration](#) for treatment of patients with [unresectable](#) or [metastatic melanoma](#) who no longer respond to other drugs.^[3] In addition, it is approved for the treatment of [squamous non-small cell lung cancer](#).^[4]

Contents [hide]

- 1 Mechanism of action
- 2 Clinical trials
 - 2.1 Metastatic melanoma
 - 2.2 Lung cancer
 - 2.3 Hodgkin's lymphoma
- 3 External links
- 4 References

Mechanism of action [\[edit \]](#)

Nivolumab is an inhibitory ligand blocking antibody against the **programmed death receptor**. In contrast to traditional chemotherapies and targeted anti-cancer therapies, which exert their effects by direct cytotoxic or tumor growth inhibition, nivolumab acts by blocking a negative regulator of **T-cell** activation and response thus allowing the immune system to attack the tumor.^{[5][6]} This is an example of **immune checkpoint blockade**.

PD-1 is a protein on the surface of activated T cells. If another molecule, called [programmed cell death 1 ligand 1](#) or programmed cell death 1 ligand 2 (PD-L1 or PD-L2), binds to PD-1, the T cell becomes inactive. This is one way that the body regulates the immune system, to avoid an overreaction. Many cancer cells make PD-L1, which inhibits T cells from attacking the tumor. Nivolumab blocks PD-L1 from binding to PD-1, allowing the T cell to work.^{[5][6]}

PD-1 blockade reinvigorates immune cells that are able to target cancer cells, which is speculated but not proven to reduce side effects.^[5]

Clinical trials [[edit](#)]

Metastatic melanoma [\[edit \]](#)

A phase 1 dose-optimization trial of nivolumab was performed in people with melanoma, lung cancer, kidney cancer and other cancers. Among the 107 melanoma patients in this trial, nivolumab demonstrated one-, two-year and three year survival rates of 62%, 48%, and 41%.

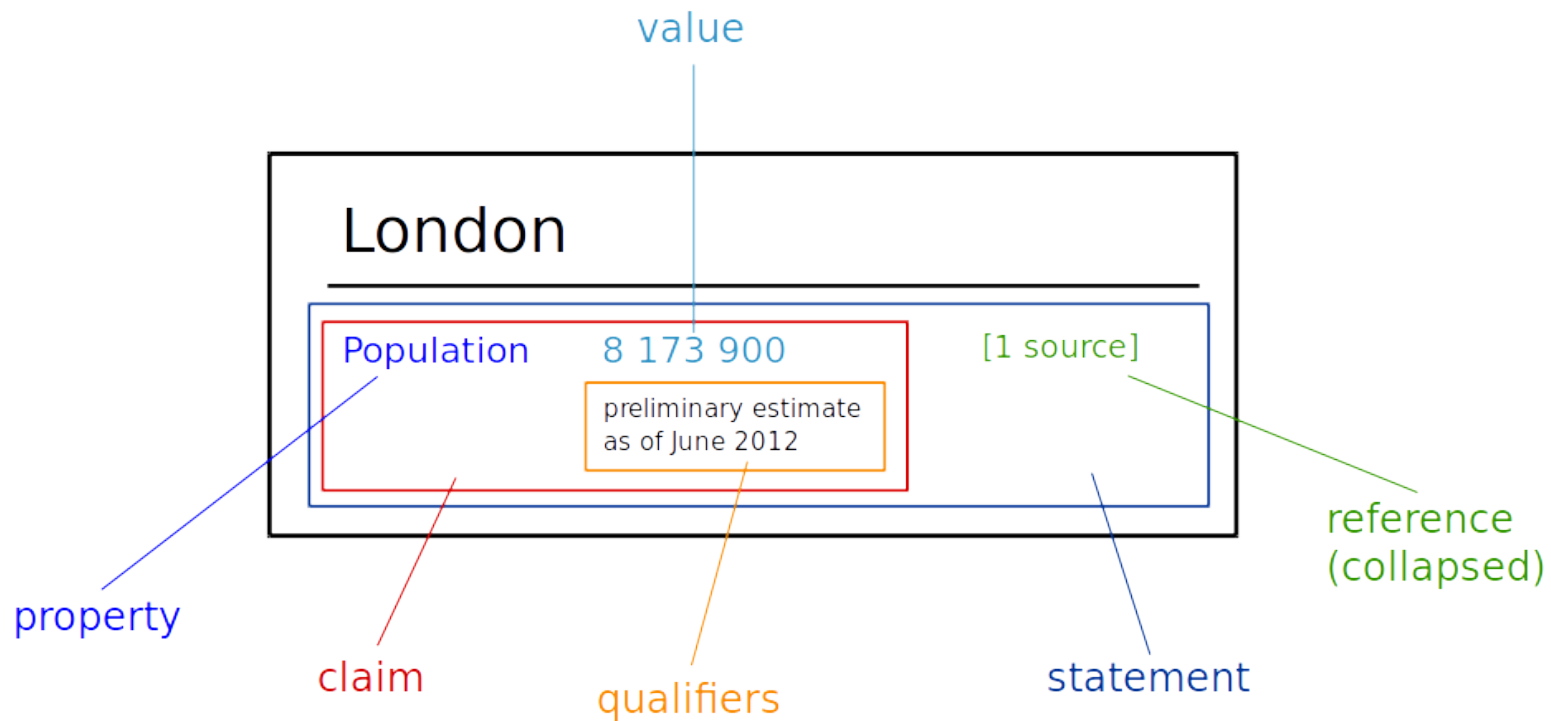
Toxicities, which were not cumulative and which mostly occurred during the first 6 months of therapy, included **pneumonitis**, low grade fatigue, diarrhea, pruritus, nausea, and decreased appetite. Pneumonitis was the most important adverse effect, leading to 3 deaths. Twenty-two percent of people in the trial experienced a treatment-related Grade 3 or Grade 4 toxicity.^{[7][8]}

A second phase 1 trial examined combination treatment of metastatic melanoma with nivolumab and **ipilimumab**. Among 53 people treated concurrently with the highest dose (1 mg/kg nivolumab and 3 mg/kg ipilimumab), 53% had an objective response, all of whom saw their tumors shrink by 80% or more. An updated report from this trial, including an additional 41 people treated with the highest dose, stated that 79% were still alive after 2 years. [\[8\]\[9\]](#)

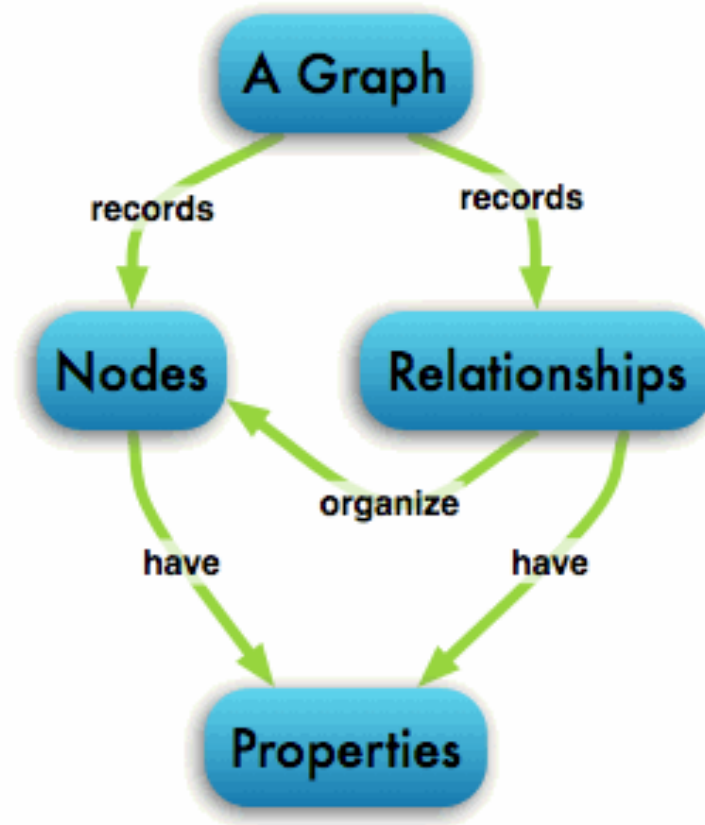
In a phase 3 trial comparing nivolumab monotherapy to treatment with traditional chemotherapy, nivolumab gave

Nivolumab	
Monoclonal antibody	
Type	Whole antibody
Source	Human
Target	PD-1
Clinical data	
Trade names	Opdivo
Identifiers	
CAS Number	946414-94-4
ATC code	L01XC17
Chemical data	
Formula	C ₃₆₃₂ H ₉₈₆₂ N ₁₇₁₂ O ₁₉₉₅ S ₄₂
Molecular mass	143.6 kDa

A Wikidata Statement



Wikidata is a Graph database



From neo4j.com

Wikidata API and query endpoints

- Three ways to access data:
 - Wikidata API allows read, write and full text search.
(www.wikidata.org/w/api.php)
 - REST endpoint for fast, direct data access.
(queryr.wmflabs.org/)
 - Wikidata query service (WDQS) as a SPARQL endpoint for complex queries.
(query.wikidata.org/)

-

Example

<https://www.wikidata.org/wiki/Q23839066>

<http://query.wikidata.org>

hsa-miR-127-3p (Q23839066)

mature human miRNA

No aliases defined

 edit

▼ In more languages

Language	Label	Description	Also known as
English	hsa-miR-127-3p	mature human miRNA	
German	No label defined	No description defined	
French	No label defined	No description defined	

Statements

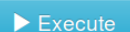
subclass of	 mature microRNA	 edit
	▼ 0 references	+ add reference
	 Mir-127	 edit
	▼ 0 references	+ add reference
		+ add
found in taxon	 human	 edit
	▼ 0 references	+ add reference
		+ add
regulates (molecular biology)	 PR domain containing 1, with ZNF domain	 edit
	▼ 0 references	+ add reference
	 X-box binding protein 1	 edit
	▼ 0 references	+ add reference
	 B-cell CLL/lymphoma 6	 edit
	▼ 0 references	+ add reference
	 SET domain containing (lysine methyltransferase) 8	 edit
	▼ 0 references	



```
1 PREFIX wd: <http://www.wikidata.org/entity/>
2 PREFIX wdt: <http://www.wikidata.org/prop/direct/>
3 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
4
5 select distinct * where {
6   ?qid rdfs:label 'hsa-miR-127-3p'@en .
7   ?qid wdt:P128 ?gene .
8   ?gene rdfs:label ?label filter (lang(?label) = "en") .
9 }
```



Press **[CTRL-SPACE]** to activate auto completion. Data last updated: 6:52:44 AM PDT, Apr 17, 2016



Execute



Clear

23 Results in 147 ms



Display ▾



Download ▾



Link ▾

qid	gene	label
Q23839066	Q18028126	potassium voltage-gated channel, shaker-related subfamily, member 6
Q23839066	Q18030532	serpin peptidase inhibitor, clade B (ovalbumin), member 9
Q23839066	Q18030787	mitogen-activated protein kinase 4
Q23839066	Q18042716	repulsive guidance molecule family member a
Q23839066	Q18042959	ARP3 actin-related protein 3 homolog B (yeast)
Q23839066	Q18060686	ARP3 actin-related protein 3 homolog C (yeast)
Q23839066	Q17847411	B-cell CLL/lymphoma 6
Q23839066	Q17859718	septin 7
Q23839066	Q17916835	solute carrier family 29 (equilibrative nucleoside transporter), member 1
Q23839066	Q18029207	matrix metalloproteinase 13 (collagenase 3)
Q23839066	Q18036143	ZW10 interacting kinetochore protein



```

1 PREFIX wd: <http://www.wikidata.org/entity/>
2 PREFIX wdt: <http://www.wikidata.org/prop/direct/>
3 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
4
5 select distinct ?protein ?label ?go_id ?go_label ?go where {
6   wd:Q23839066 wdt:P128 [wdt:P688 ?protein] .
7   ?protein wdt:P681 ?go .
8   ?protein rdfs:label ?label filter (lang(?label) = "en") .
9   ?go wdt:P686 ?go_id .
10  ?go rdfs:label ?go_label filter (lang(?go_label) = "en") .
11 } order by ?label

```



Press [CTRL-SPACE] to activate auto completion. Data last updated: 7:10:52 PM PDT, Apr 16, 2016



Execute



Clear

154 Results in 191 ms



Display ▾



Download ▾



Link ▾

protein	label	go_id	go_label	go
Q21130866	Actin-related protein 3B	GO:0005737	cytoplasm	Q79899
Q21130866	Actin-related protein 3B	GO:0005856	cytoskeleton	Q154626
Q21130866	Actin-related protein 3B	GO:0042995	cell projection	Q14353100
Q21130866	Actin-related protein 3B	GO:0070062	extracellular exosome	Q14864178
Q21130866	Actin-related protein 3B	GO:0005885	Arp2/3 protein complex	Q21102007
Q21140508	Actin-related protein 3C	GO:0005885	Arp2/3 protein complex	Q21102007
Q21108216	Actin-related protein 3C	GO:0070062	extracellular exosome	Q14864178
Q21108216	Actin-related protein 3C	GO:0005885	Arp2/3 protein complex	Q21102007
Q21116345	B-cell lymphoma 6 protein	GO:0005657	replication fork	Q14817957
Q21116345	B-cell lymphoma 6 protein	GO:0005634	nucleus	Q21541775
Q21111733	BolA-like protein 1	GO:0005739	mitochondrion	Q39572

Statistics

WD items added	
Human genes	59,727
Mouse genes	73,352
Human proteins	27,194
Mouse proteins	16,172
Gene Ontology terms	42,281
Human miRNAs	2,619
Human diseases	6,782
FDA approved drugs	2,079

# of Properties	
Biology	34
Molecular biology	35
Medicine	28
Drugs/Chemistry	34

In total, our bots maintain ~220,000 items with each typically >10 statements. This results in a current extent of ~2,200,000 values being maintained on Wikidata by our bots.



ERCC Data Analysis Workshop

Topics covered in Use Case 2 Part 2

- Accessing samples in the exRNA Atlas
- Comparative analysis of exRNA samples using DESeq / exceRpt pipeline results and tools in the Genboree Workbench
- Finding information about exRNAs of interest
 - on Exocarta and Vesiclepedia
 - on GeneWiki and Wikidata
- **Pathway analysis using Cytoscape & WikiPathways in the Genboree Workbench**



Pathway and Network Analysis from the Genboree Workbench

Welcome to the Genboree Workbench! [Watch Intro Video](#)

Data Selector

Refresh Data Filter: Select a filter...

- Epigenome Informatics Demo Output Data
- Epigenome ToolSet Demo Input Data
- Epigenomics Roadmap Repository
- Examples and Test Data
- exRNA
- exRNA KJENS1
 - Databases
 - ADPDb
 - Tracks
 - Lists & Selections
 - Lists of Files
 - AD CSF first40 results.zip
 - AD CSF results.zip
 - SampleSets
 - Samples
 - Files
 - OriginalDiffExpResults
 - KJ_diffExp_InputList.txt
 - PostProcessedRuns
 - ProcessPipelineRuns
 - smalRNAseqPipeline
 - Projects
 - exRNA Metadata Standards

Details

Attribute	Value
Group	exRNA KJENS1
Role	author
Name	ADPDb
Description	Database for the Alzheimer's and Parkinson's Disease Project from Kendall Jensen's lab at TGen - ERCC
Species	Homo sapiens

Input Data

KJ_diffExp_InputList.txt

Output Targets

Genboree is built & maintained by the **Bioinformatics Research Laboratory** at **Baylor College of Medicine**.

Genboree is a hosted service. Code is available **free** for academic use.

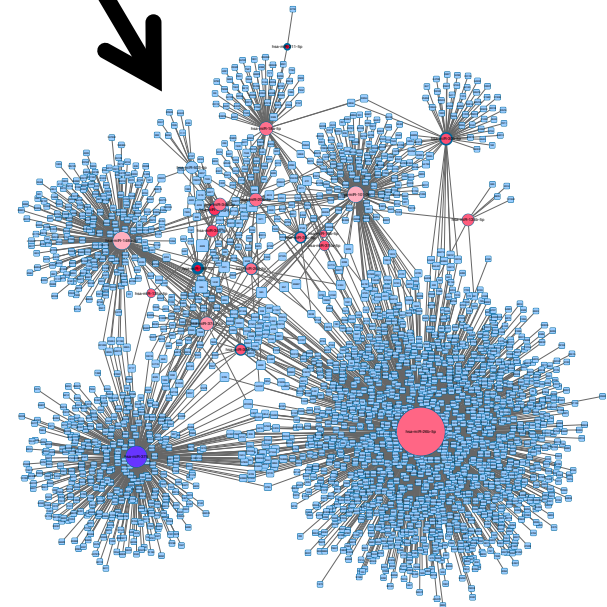
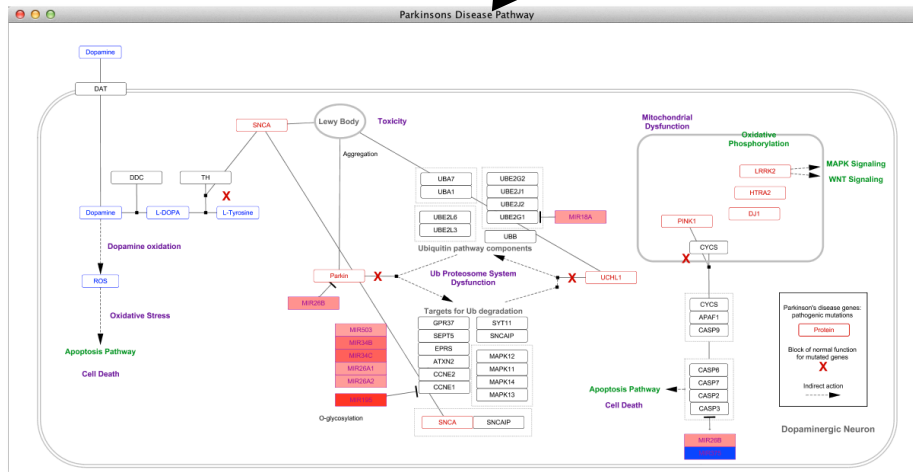
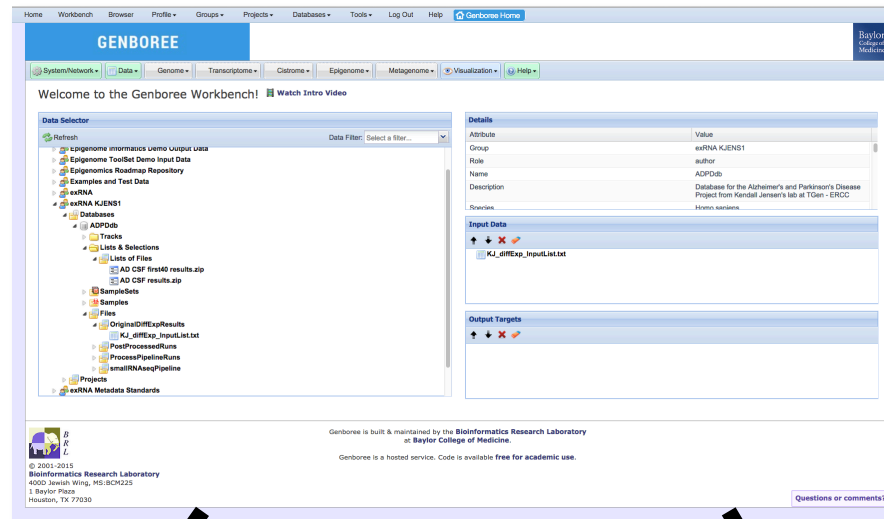
© 2001-2015
Bioinformatics Research Laboratory
400D Jewish Wing, MS:BCM225
1 Baylor Plaza
Houston, TX 77030

[Questions or comments?](#)

Alex Pico, PhD
Gladstone Institutes, San Francisco
April 17, 2016



Pathway and Network Analysis





Pathway and Network Analysis

1. *Why* would you do this?

2. *What* would you do?

3. *How* would you do it?

4. *Who* can do this?



Pathway and Network Analysis: Why?

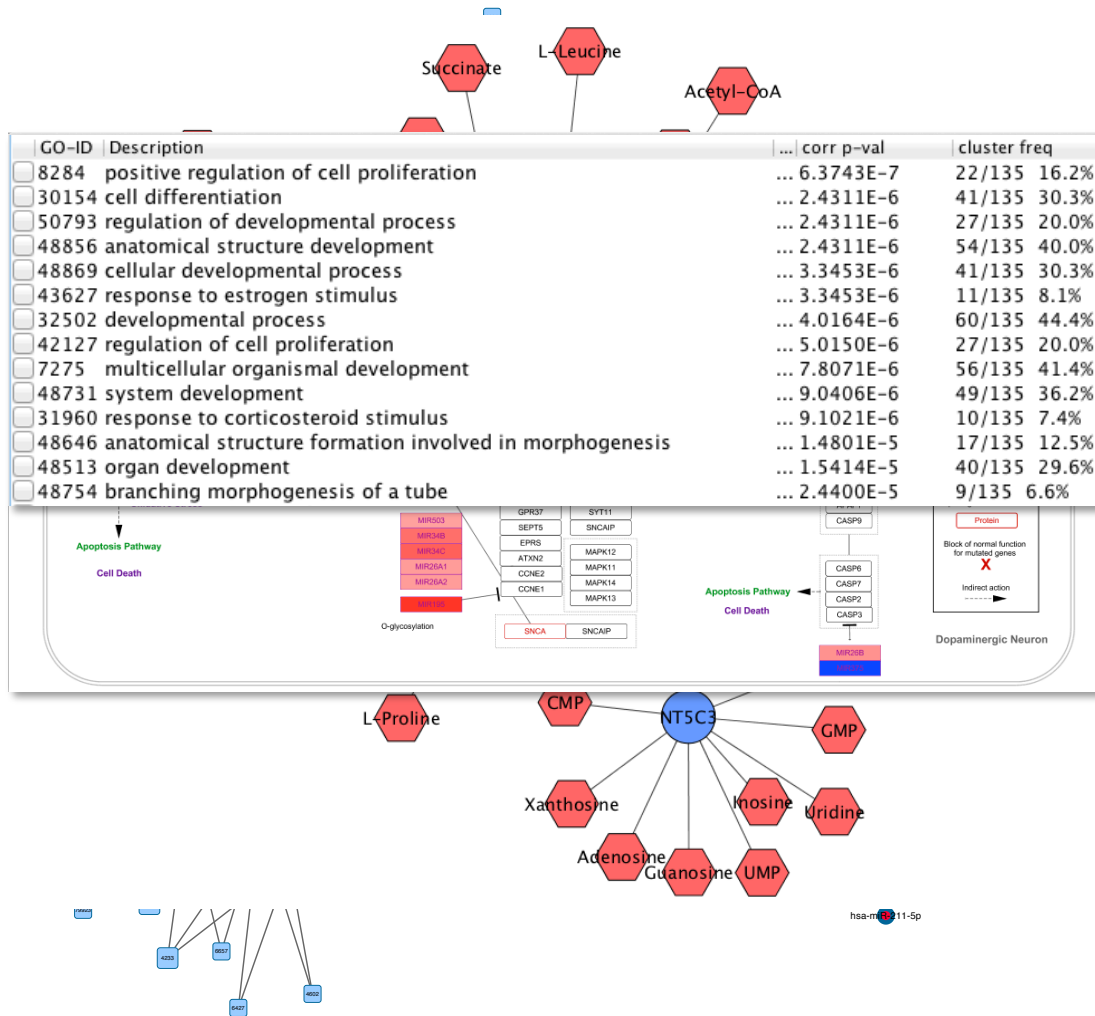
...a list of differentially expressed miRNA

identifier	baseMean	log2FoldChai	lfcSE	pvalue	padj
hsa-miR-195-5p	96.7700199	2.44841391	0.49366515	7.06E-07	0.00018136
hsa-miR-30e-3p	702.455493	2.02502471	0.41203155	8.89E-07	0.00018136
hsa-miR-211-5p	49.2191661	2.44930572	0.51596328	2.06E-06	0.00028071
hsa-miR-30b-5p	87.5770117	1.93973337	0.41974017	3.81E-06	0.00036691
hsa-miR-30c-5p	177.914843	1.7910964	0.39047141	4.50E-06	0.00036691
hsa-miR-30a-3p	727.311039	1.88682103	0.41998768	7.04E-06	0.00047852
hsa-miR-664-3p	9.70560333	2.35032438	0.53633997	1.18E-05	0.00068493
hsa-miR-26a-5p	4601.81791	1.69854746	0.40517223	2.76E-05	0.00140918
hsa-miR-34c-5p	635.267228	2.12361583	0.52355429	4.99E-05	0.00226158
hsa-miR-582-3p	32.7961329	1.8613787	0.46851513	7.10E-05	0.00289653
hsa-miR-26b-5p	755.55888	1.76395192	0.4495387	8.71E-05	0.00323141
hsa-miR-1298	739.885163	2.05340959	0.54460665	0.00016296	0.00554052
hsa-let-7f-1-3p	4.75127804	1.89744027	0.51372912	0.00022122	0.00694289
hsa-miR-375	1005.63367	-1.610137	0.44283469	0.00027693	0.00807049
hsa-miR-1247-5p	10.2790818	-1.8194525	0.51240356	0.00038403	0.00921681
hsa-miR-1911-5p	99.7237089	1.98039539	0.55768118	0.00038359	0.00921681
hsa-miR-374a-5p	10.8650312	1.89672368	0.53132051	0.00035721	0.00921681
hsa-miR-34b-5p	41.0116005	2.00714578	0.5692402	0.00042187	0.00956246
hsa-miR-152	77.4552781	1.35732907	0.3898756	0.00049872	0.01070941
hsa-miR-7-1-3p	2.63312729	1.67557636	0.49030805	0.00063223	0.01289758
hsa-miR-135a-5p	11.9855253	1.86274748	0.56392061	0.00095585	0.01695599
hsa-miR-18b-5p	8.77201213	1.78357817	0.53584358	0.00087302	0.01695599
hsa-miR-23c	103.658692	1.36705145	0.4124759	0.00091887	0.01695599
hsa-miR-135b-5p	8.74843581	1.82445524	0.55621122	0.00103753	0.01763794
hsa-miR-374a-3p	7.55916002	1.62705861	0.50745634	0.00134453	0.02194265
hsa-miR-148b-3p	812.441206	1.2714604	0.39939394	0.00145519	0.02204724
hsa-miR-18a-5p	10.7773676	1.70424536	0.53683176	0.00150025	0.02204724
hsa-miR-204-5p	65558.9447	1.71577961	0.54088519	0.00151305	0.02204724
hsa-miR-374b-5p	17.1285677	1.46507029	0.47932833	0.00223932	0.03150498
hsa-miR-3960	24.1059322	1.43616846	0.47370107	0.00243098	0.03306127
hsa-miR-3605-5p	33.4311818	-1.4371226	0.47567792	0.00251767	0.03313582
hsa-miR-101-3p	2484.00912	1.2855992	0.42889521	0.00272233	0.03470976
hsa-miR-34b-3p	33.7603604	1.70785085	0.57338237	0.00289611	0.03475334
hsa-miR-503	2.88157858	1.66640756	0.55795946	0.00282089	0.03475334

- Identify expected hits
- Look up each one
- Prioritize for validation and further investigation



Pathway and Network Analysis: Why?



- **Networks**
 - Target protein interactions
 - TF or drug interactions
 - Functional characterization
- **Pathways**
 - Known mechanism
 - New hypotheses
- **Prioritize for validation and further investigation**



Pathway and Network Analysis: What?

Network Analysis

1. Find a database with miRNA-target protein interaction
2. Lookup each of your miRNA of interest
3. Merge binary interactions into a single network model
4. Apply layout algorithm on network
5. Visualize your data on network

Pathway Analysis

1. Find a database of pathway diagrams
2. Annotate pathways with miRNA based on target protein database
3. Lookup each of your miRNA of interest
4. Produce ranked list of relevant pathways
5. Load pathway into visualization tool
6. Visualize your data on network



Pathway and Network Analysis: What?

Network Analysis

1. Find a data
2. Lookup each
3. Merge bin
4. Apply layout algorithm on network
5. Visualize your data on network

GENBOREE

Target Interaction Finder

tein interaction

network model



Pathway Analysis

1. Find a database of pathway diagrams
2. Annotate pathways with miRNA based on target protein database
3. Lookup each of your miRNA of interest
4. Produce ranked list of relevant pathways
5. Load pathway into visualization tool
6. Visualize your data on network



Pathway and Network Analysis: What?

Network Analysis

1. Find a data
2. Lookup each
3. Merge bin
4. Apply layout algorithm on network
5. Visualize your data on network

GENBOREE

Target Interaction Finder

tein interaction

network model



Pathway Analysis

1. Find a database of pathway diagrams
2. Annotate path
3. Lookup each o
4. Produce ranked list of relevant pathways
5. Load pathway into visualization tool
6. Visualize your data on network

GENBOREE

Pathway Finder

based on target protein database

interest





Pathway and Network Analysis: How?

Home Workbench Browser Profile Groups Projects Databases Tools Log Out Help [Genboree Home](#)

GENBOREE

System/Network Data Genome Transcriptome Cistrome Epigenome Metagenome Visualization Help

Welcome to the Genboree Workbench! [Watch Intro Video](#)

Data Selector

Refresh Data Filter: Select a filter...

- Epigenome Informatics Demo Output Data
- Epigenome ToolSet Demo Input Data
- Epigenomics Roadmap Repository
- Examples and Test Data
- exRNA
 - exRNA KJENS1
 - Databases
 - ADPDb
 - Tracks
 - Lists & Selections
 - Lists of Files
 - AD CSF first40 results.zip
 - AD CSF results.zip
 - SampleSets
 - Samples
 - Files
 - OriginalDiffExpResults
 - KJ_diffExp_InputList.txt
 - PostProcessedRuns
 - ProcessPipelineRuns
 - smallRNAseqPipeline

- Projects
- exRNA Metadata Standards

Details

Attribute	Value
Group	exRNA KJENS1
Role	author
Name	ADPDb
Description	Database for the Alzheimer's and Parkinson's Disease Project from Kendall Jensen's lab at TGen - ERCC
Species	Homo sapiens

Input Data

KJ_diffExp_InputList.txt

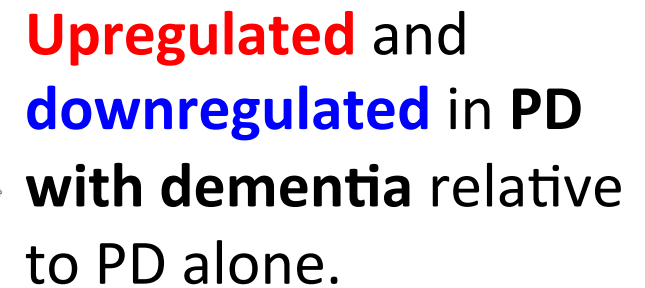
Output Targets

Genboree is built & maintained by the **Bioinformatics Research Laboratory** at **Baylor College of Medicine**.

Genboree is a hosted service. Code is available **free for academic use**.

© 2001-2015 **Bioinformatics Research Laboratory**
400D Jewish Wing, MS:BCM225
1 Baylor Plaza
Houston, TX 77030

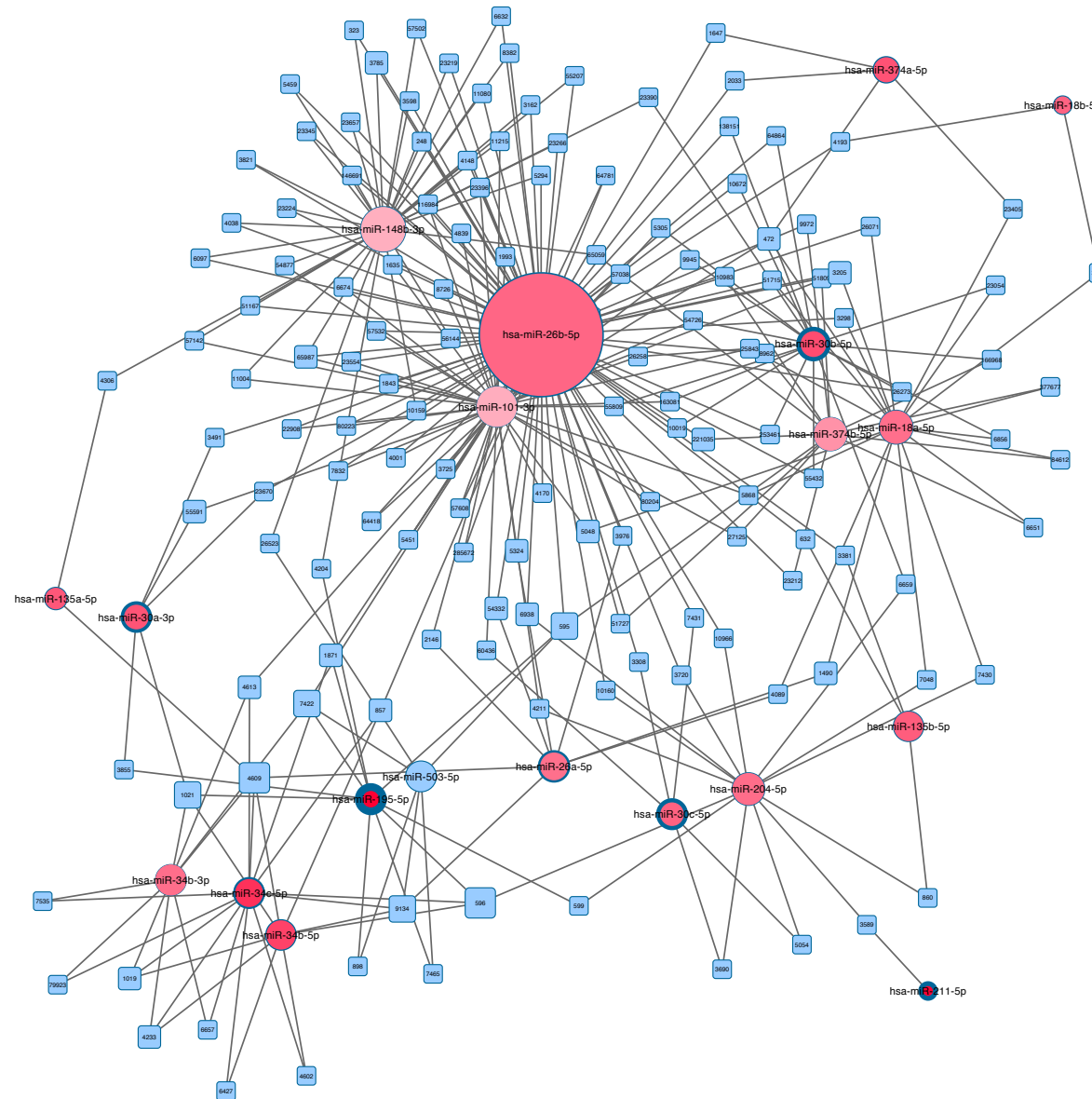
[Questions or comments?](#)



93

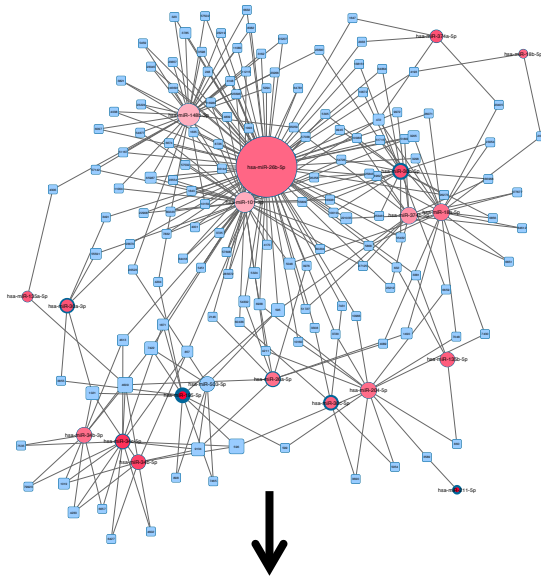


Pathway and Network Analysis: How?





Pathway and Network Analysis: How?



BiNGO

BiNGO

3.0+

Calculates overrepresented GO terms in the network and display them as a network of significant GO terms.

★★★★☆ (54) 19530 downloads

GO-ID	Description	...	corr p-val	cluster freq
☐ 8284	positive regulation of cell proliferation	...	6.3743E-7	22/135 16.2%
☐ 30154	cell differentiation	...	2.4311E-6	41/135 30.3%
☐ 50793	regulation of developmental process	...	2.4311E-6	27/135 20.0%
☐ 48856	anatomical structure development	...	2.4311E-6	54/135 40.0%
☐ 48869	cellular developmental process	...	3.3453E-6	41/135 30.3%
☐ 43627	response to estrogen stimulus	...	3.3453E-6	11/135 8.1%
☐ 32502	developmental process	...	4.0164E-6	60/135 44.4%
☐ 42127	regulation of cell proliferation	...	5.0150E-6	27/135 20.0%
☐ 7275	multicellular organismal development	...	7.8071E-6	56/135 41.4%
☐ 48731	system development	...	9.0406E-6	49/135 36.2%
☐ 31960	response to corticosteroid stimulus	...	9.1021E-6	10/135 7.4%
☐ 48646	anatomical structure formation involved in morphogenesis	...	1.4801E-5	17/135 12.5%
☐ 48513	organ development	...	1.5414E-5	40/135 29.6%
☐ 48754	branching morphogenesis of a tube	...	2.4400E-5	9/135 6.6%



Pathway and Network Analysis: How?

Home Workbench Browser Profile Groups Projects Databases Tools Log Out Help [Genboree Home](#)

GENBOREE

System/Network Data Genome Transcriptome Cistrome Epigenome Metagenome Visualization Help

Welcome to the Genboree Workbench! [Watch Intro Video](#)

Data Selector

Refresh Data Filter: Select a filter...

- Epigenome Informatics Demo Output Data
- Epigenome ToolSet Demo Input Data
- Epigenomics Roadmap Repository
- Examples and Test Data
- exRNA
 - exRNA KJENS1
 - Databases
 - ADPDdb
 - Tracks
 - Lists & Selections
 - Lists of Files
 - AD CSF first40 results.zip
 - AD CSF results.zip
 - SampleSets
 - Samples
 - Files
 - OriginalDiffExpResults
 - KJ_diffExp_InputList.txt
 - PostProcessedRuns
 - ProcessPipelineRuns
 - smallRNAseqPipeline

- Projects
- exRNA Metadata Standards

Details

Attribute	Value
Group	exRNA KJENS1
Role	author
Name	ADPDdb
Description	Database for the Alzheimer's and Parkinson's Disease Project from Kendall Jensen's lab at TGen - ERCC
Species	Homo sapiens

Input Data

KJ_diffExp_InputList.txt

Output Targets

Genboree is built & maintained by the **Bioinformatics Research Laboratory** at **Baylor College of Medicine**.

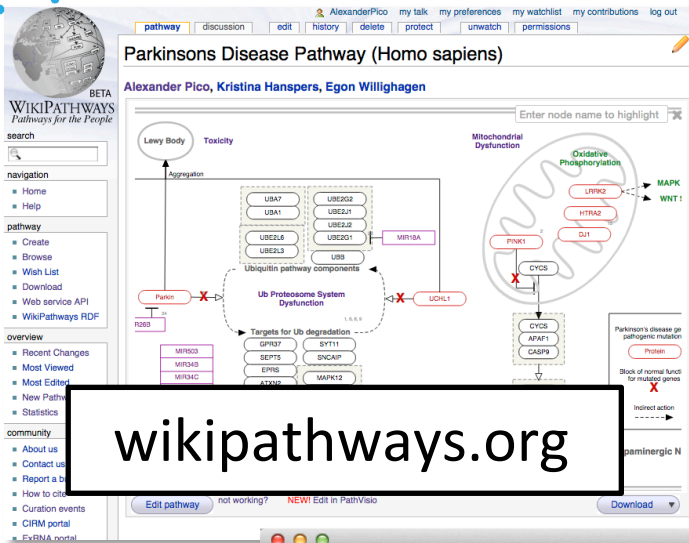
Genboree is a hosted service. Code is available **free for academic use**.

© 2001-2015 **Bioinformatics Research Laboratory**
400D Jewish Wing, MS:BCM225
1 Baylor Plaza
Houston, TX 77030

[Questions or comments?](#)



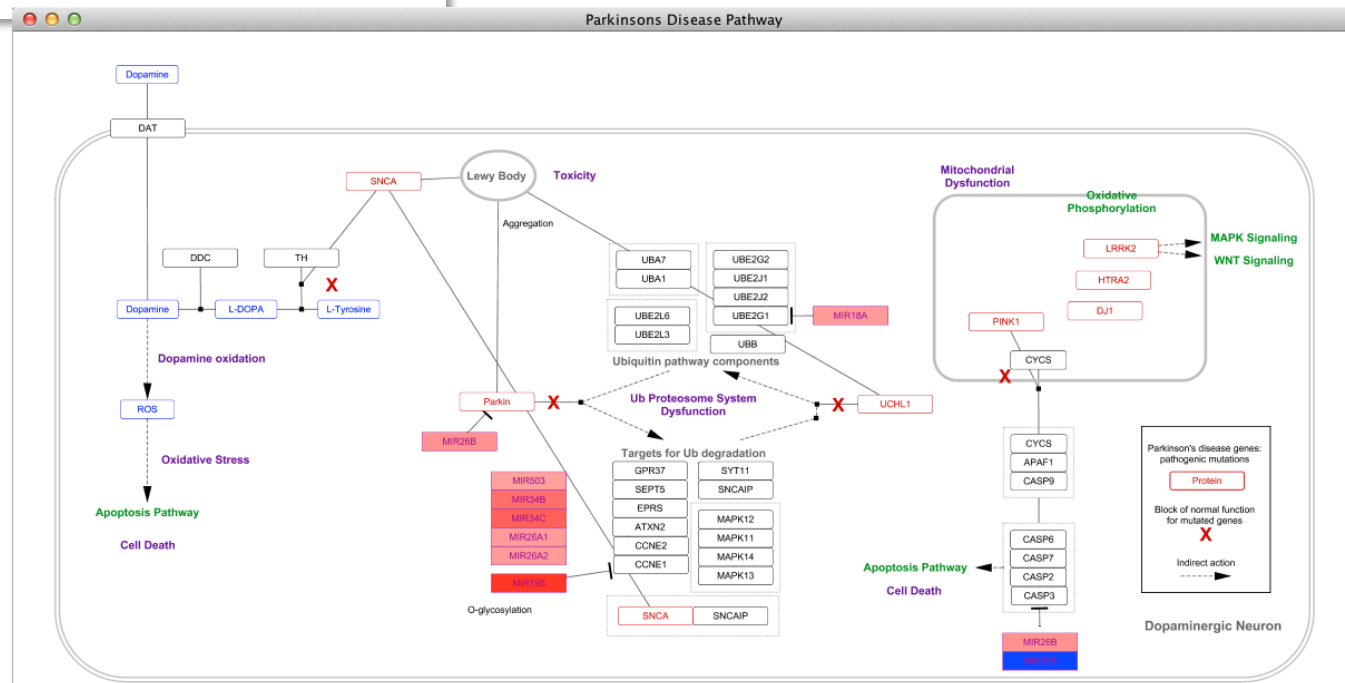
Pathway and Network Analysis: How?



WikiPathways 3.0+

WikiPathways web service client
and GPML file format importer

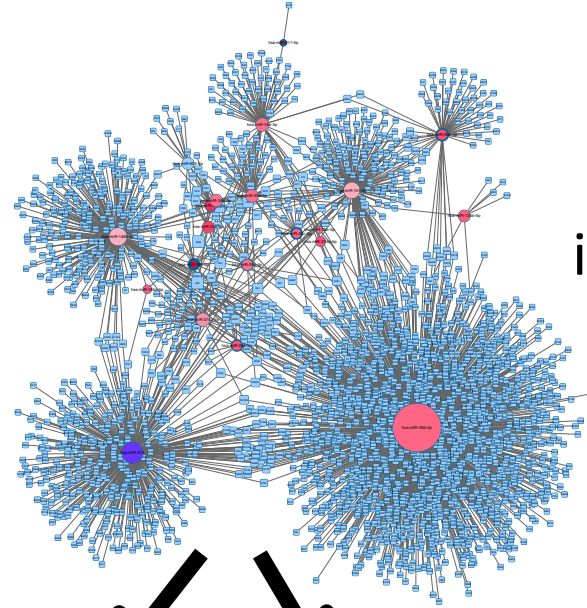
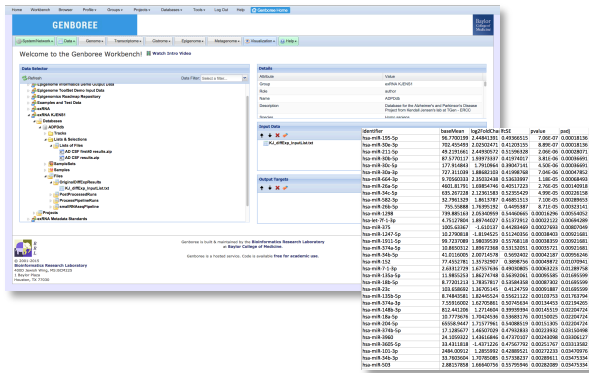
★★★★★ (8) 5548 downloads





Pathway and Network Analysis: How?

Genoree

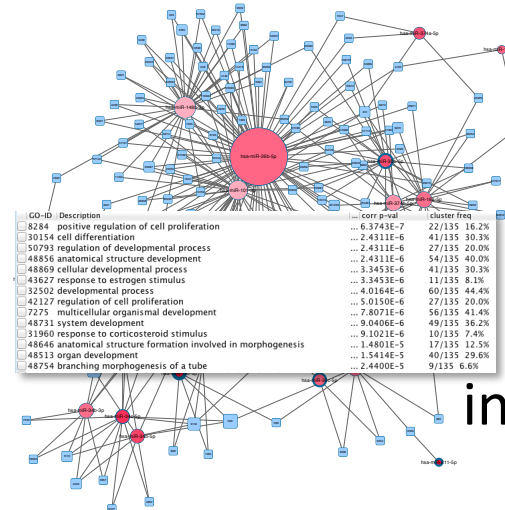
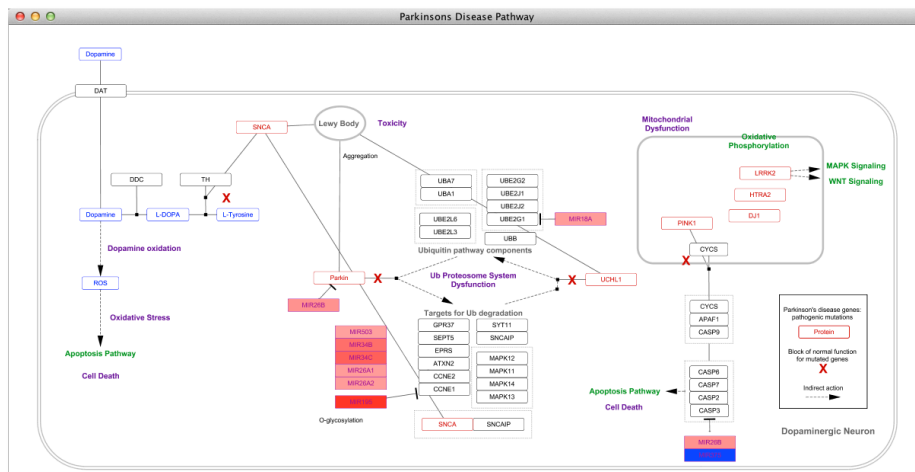


All known interactions

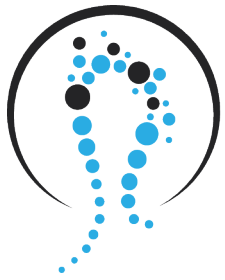


PD vs PDD diff exp
Burgos K., et al. (2014) PLoS ONE

Pathways



Focused set of interactions



Pathway and Network Analysis: **Who?**

1. Why would you do this?

2. What would you do?

3. How would you do it?

4. Who can do this?



Pathway and Network Analysis: Who?

Login:

GENBOREE

Welcome to the Genboree Workbench! [Watch Intro Video](#)

Data Selector

- Epigenome Informatics Demo Output Data
- Epigenome ToolSet Demo Input Data
- Epigenomics Roadmap Repository
- Examples and Test Data
- exRNA
- exRNA KJENS1
- Databases
 - ADPDdb
- Tracks
- Lists & Selections
 - Lists of Files
 - AD CSF first40 results.zip
 - AD CSF results.zip
- SampleSets
- Samples
- Files
 - OriginalDiffExpResults
 - KJ_diffExp_InputList.txt
- PostProcessedRuns
- ProcessPipelineRuns
- smRNAseqPipeline
- Projects
- exRNA Metadata Standards

Details

Attribute	Value
Group	exRNA KJENS1
Role	author
Name	ADPDdb
Description	Database for the Alzheimer's and Parkinson's Disease Project from Kendall Jensen's lab at TGen - ERCC
Species	Homo sapiens

Input Data

- KJ_diffExp_InputList.txt

Output Targets

Genboree is built & maintained by the **Bioinformatics Research Laboratory** at **Baylor College of Medicine**.

Genboree is a hosted service. Code is available **free for academic use**.

© 2001-2015
Bioinformatics Research Laboratory
400D Jewish Wing, MS-BCM225
1 Baylor Plaza
Houston, TX 77030

[Questions or comments?](#)

Email: alex.pico@gladstone.ucsf.edu

Useful **Links**

exRNA Portal

- <http://exrna.org>
- **Portal Data & Software Resources**
 - <http://exrna.org/resources/data/>
 - <http://exrna.org/resources/software>

exRNA Atlas

- <http://genboree.org/exRNA-atlas/>
- **Atlas Tutorials**
 - <http://genboree.org/theCommons/projects/exrna-mads/wiki/exRNA%20Atlas>

ERCC Data Coordination Center Wiki

- <http://genboree.org/theCommons/projects/exrna-mads/wiki>

exRNA Data Analysis Tools Wiki

- <http://genboree.org/theCommons/projects/exrna-tools-may2014/wiki>



Presenters

exRNA Atlas

Sai Lakshmi Subramanian, Baylor College of Medicine, Houston, TX

DESeq / exceRpt pipelines

William Thistlethwaite, Baylor College of Medicine, Houston, TX

Exocarta and Vesiclepedia

Roger P. Alexander, Pacific NW Diabetes Research Institute, Seattle, WA

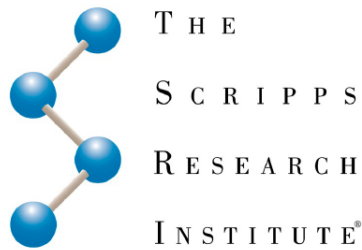
GeneWiki and Wikidata

Sebastian Burgstaller-Muehlbacher, Scripps Research Inst., San Diego, CA

Pathway analysis

Alex Pico, Gladstone Institutes, San Francisco, CA

Acknowledgments



Andrew Su
Benjamin Good
Julia Turner
Tim Putman
(TSRI)

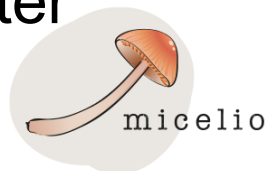
Elvira Mitraka
Lynn Schriml
(Disease Ontology, U Baltimore)



Gang Fu
(NIH, PubChem)



Andra Waagmeester
(Micelio.be)





Acknowledgements



LA TROBE
UNIVERSITY



LA TROBE
INSTITUTE FOR
MOLECULAR SCIENCE

La Trobe University, Melbourne, Australia

Suresh Mathivanan

Dr Shivakumar Keerthikumar

Mohashin Pathan

David Chisanga



Acknowledgements

All data submitters from the ERC Consortium

Baylor College of Medicine, Houston, TX

Aleksandar Milosavljevic, Director
Matthew Roth, Co-Director
Elke Norwig-Eastaugh



Genboree Dev Team at Baylor

Sai Lakshmi Subramanian
William Thistlethwaite
Andrew Jackson
Neethu Shah
Sameer Paithankar
Aaron Baker



*Bioinformatics
Research
Laboratory*

Sponsored by:



National Institutes of Health
Office of Strategic Coordination - The Common Fund

Grant **1U54DA036134** from the NIH Common Fund, through the Office of Strategic Coordination/Office of the NIH Director

Yale University, New Haven, CT

Mark Gerstein
Joel Rozowsky
Rob Kitchen
Fabio Navarro



Gladstone Institutes, San Francisco, CA

Alexander Pico
Anders Riutta
Kristina Hanspers



Pacific Northwest Diabetes Research Institute, Seattle, WA

David Galas
Roger Alexander



Stanford University, Stanford, CA

Julia Salzman
Linda Szabo