

Cloud-hosted tools for the analysis of extracellular RNA and the construction of the exRNA Atlas



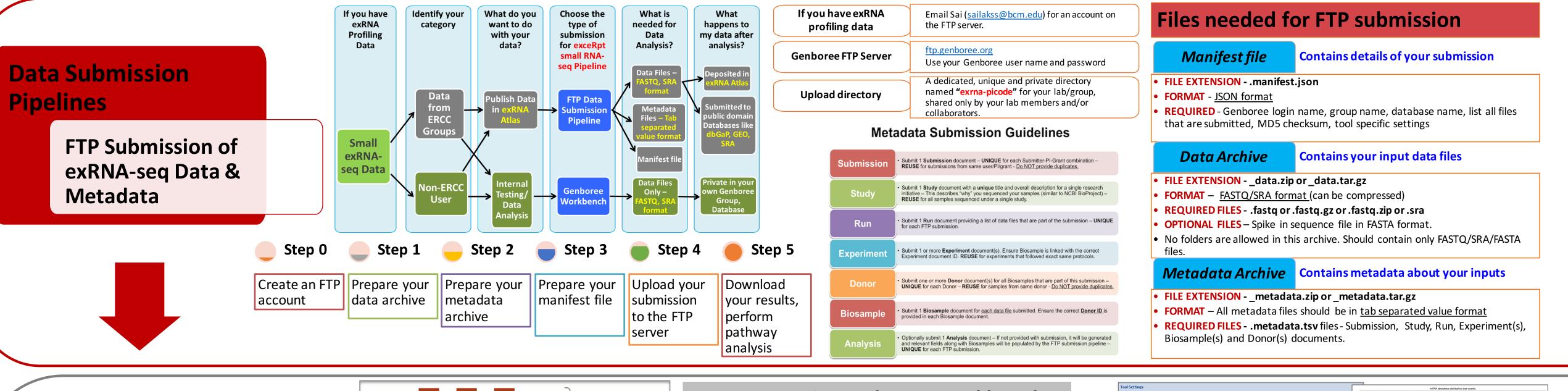
Baylor College of Medicine

Rocco D. Lucero¹, Sai Lakshmi Subramanian¹, William Thistlethwaite¹, Robert Kitchen², Fabio Navarro², Joel Rozowsky², Alexander Pico³, Roger Alexander⁴, David Galas⁴, Matthew E. Roth¹, Mark Gerstein², Aleksandar Milosavljevic¹

1 Baylor College of Medicine, Houston, TX; 2 Yale University, New Haven, CT; 3 Gladstone Institutes, San Francisco, CA; 4 Pacific Northwest Diabetes Research Institute, Seattle, WA

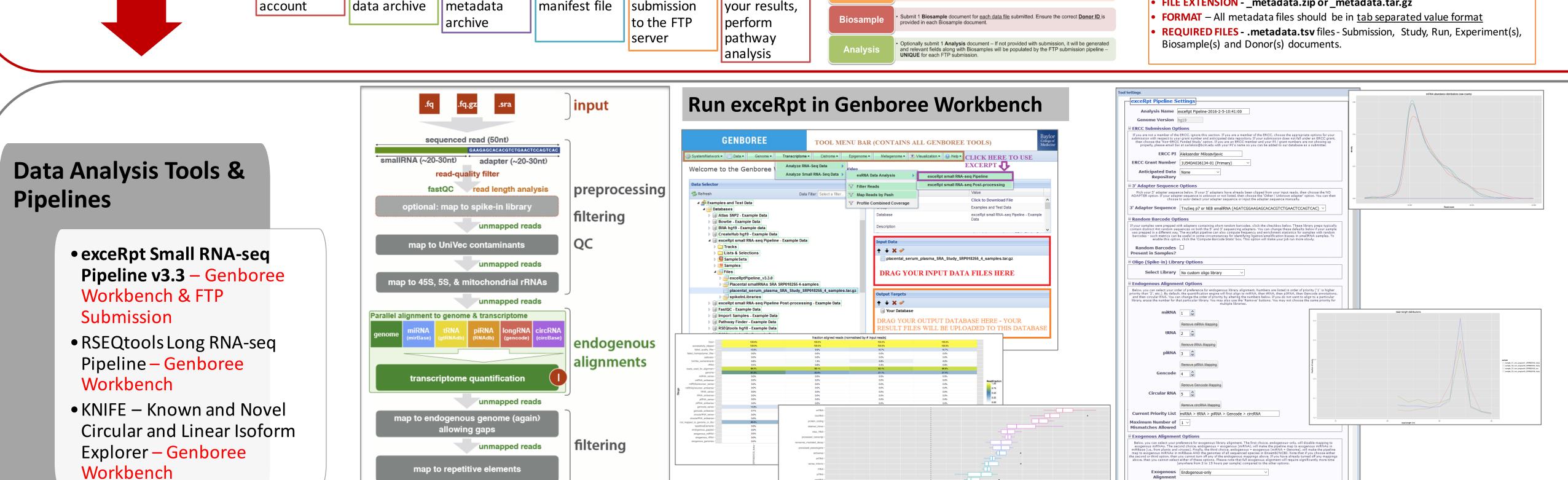
Tools Available at the Data Management and Resource Repository (DMRR) for Integrative Analysis of exRNA Profiling Data

To support the research activities and goals of ERC Consortium (ERCC) members as well as the entire scientific community, we provide three categories of tools for submission and analysis of exRNA profiling data, all of which have been built using the Genboree framework. First, the **Data Submission Pipelines** are used for submission of exRNA profiling datasets to be published in the exRNA Atlas. The submitted data is then uniformly processed using the Data Analysis Tools & Pipelines, and sample metadata is stored in the GenboreeKB exRNA Metadata system, both of which are used to build the exRNA Atlas. The exRNA Atlas consists of 751 exRNA profiles generated from about 8.8 billion small RNA-seq reads. The Atlas allows faceted sub-selection of samples of interest based on sample metadata, for further downstream analysis. Uniform quality metrics agreed by the ERCC are applied to all datasets.



The FTP data submission pipeline has been implemented for submitting small exRNA-seq datasets (with associated metadata) for processing through the exceRpt small RNA-seq pipeline.

- Users submit required files to the shared directory created for their lab on the Genboree FTP server.
- After files are processed through exceRpt, results will be deposited in the Genboree Workbench and FTP server and will also be accessible through the exRNA Atlas.
- Data will be deposited in the public domain databases like dbGaP, GEO and SRA.

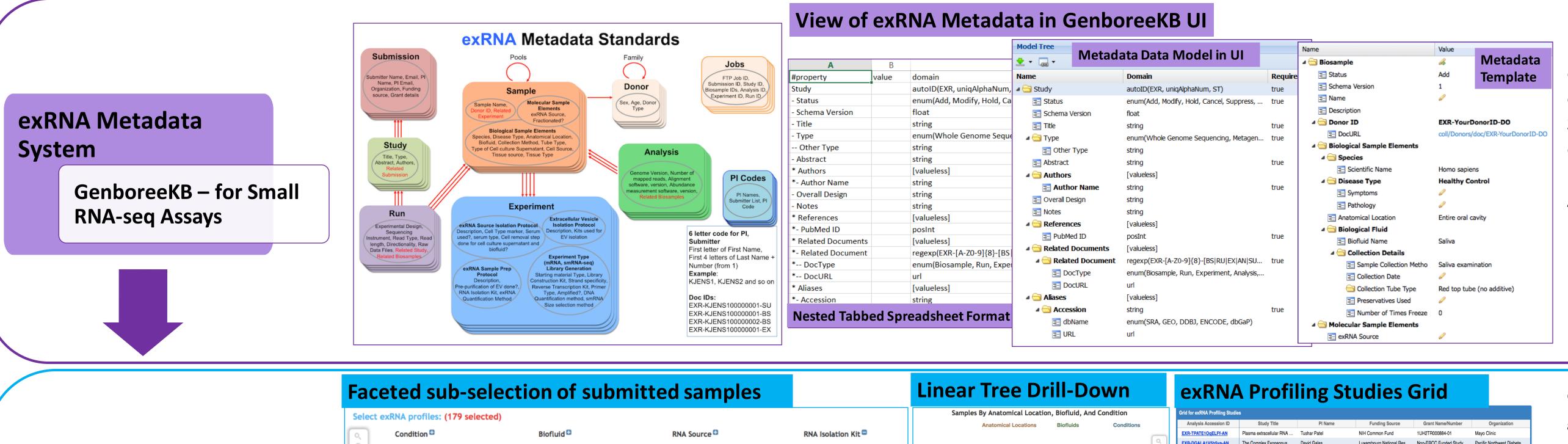


unmapped read

map to all genomes in ensembl & NCBI

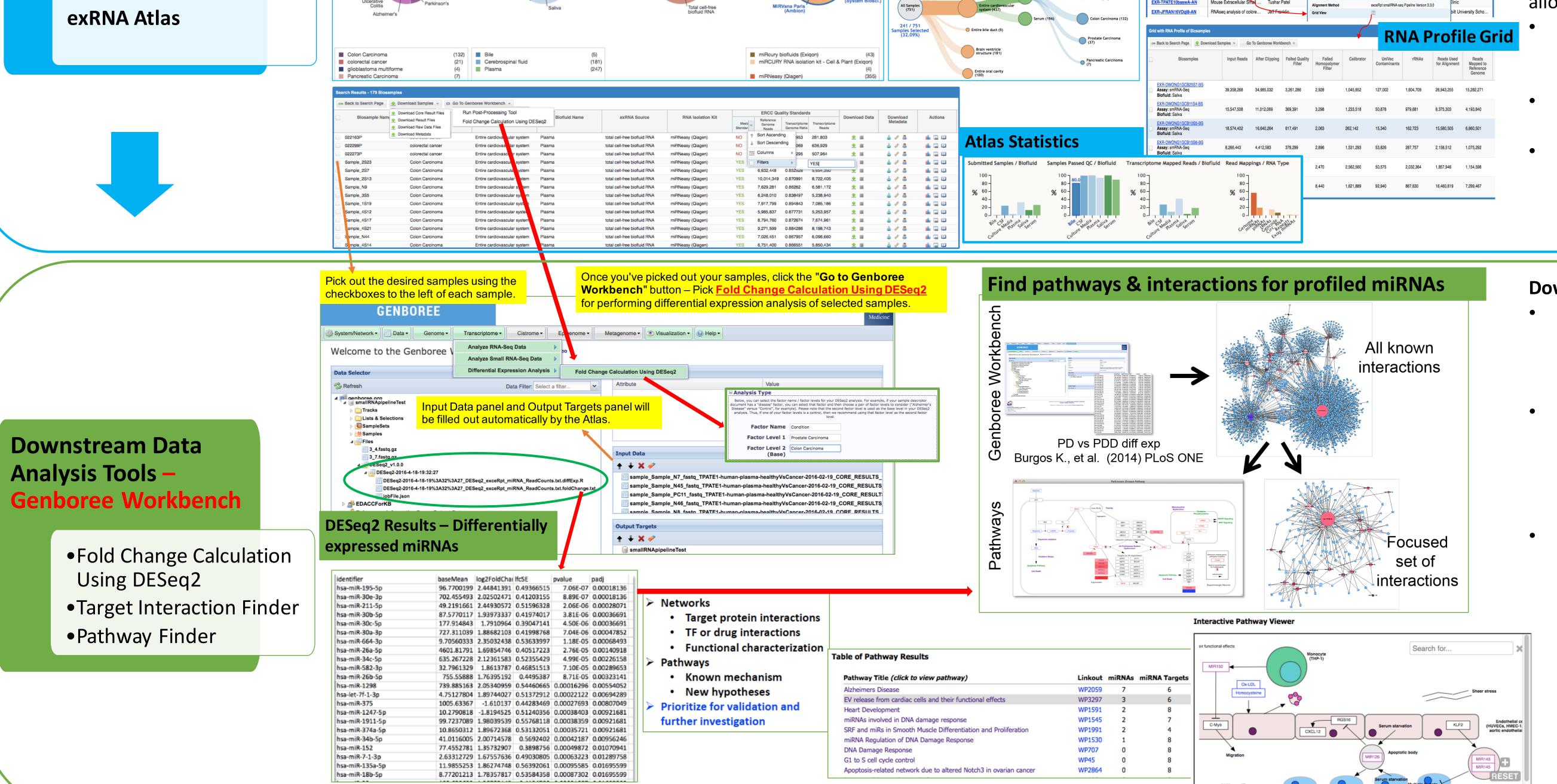
Data analysis tools and pipelines such as the exceRpt small RNA-seq pipeline, RSEQtools long RNAseq pipeline and KNIFE are available in the Genboree Workbench.

- The exceRpt small RNA-seq pipeline generates a variety of sample-level quality control metrics, produces abundance estimates for various small RNA species, and makes available detailed alignment information for visualization and validation.
- The RSEQtools pipeline performs gene-expression quantification, visualizes signal tracks of mapped reads, calculates mapping bias, and computes annotation coverage.
- **KNIFE** performs statistically based splicing detection for circular and linear isoforms from RNA-Seq data.



Metadata standards have been defined by the ERCC members to effectively annotate exRNA profiles, ensure reproducibility of experiments and allow data comparisons. The GenboreeKB exRNA Metadata System allows for submission, tracking and editing of exRNA metadata for various categories including Biosamples, Donors, Experiments, Runs, Studies and Analyses.

- All metadata submitted through the FTP data pipeline is validated against agreed ontologies and data domains can be viewed in GenboreeKB.
- Templates and questionnaires are available to facilitate metadata entry.



exRNA Atlas contains data generated by various groups that were analyzed using the exceRpt pipeline. Data and metadata are displayed in grids or tabular views that aggregate submitted biosamples by biofluids and disease or experiment type. The Atlas allows users to:

- browse, filter and select submitted samples across various facets including biofluids, diseases, RNA isolation kits, exRNA Source.
- download raw FASTQ files and processed results from datasets of interest.
- view summaries of exRNA profiling studies from various labs and other statistics including reads passing QC, transcriptome mapped reads and read mappings per library type.

Downstream Data Analysis Tools:

- The **DESeq2** tool tests samples for differential expression and allows you to test a given factor (say, disease) across two different factor levels (say, prostate carcinoma vs colon carcinoma).
- The **Target Interaction Finder** tool generates miRNA-protein target interaction files for a set of miRNA identifiers, which can be imported into downstream tools, such as Cytoscape, for network analysis and visualization.
- The **Pathway Finder** tool performs a search for pathways either containing your miRNAs of interest or protein targets of your miRNAs. A table of pathway results and an interactive pathway viewer are displayed. The 1st column of the table lists a clickable pathway title that updates the viewer. The second column lists pathway identifiers that link to WikiPathways.org. The list is sorted by the number of "miRNAs" (primary) and by "miRNA Targets" (secondary) found on each pathway. The top 20 results are listed.

- exRNA Atlas http://genboree.org/exRNA-atlas/ 2. exRNA Portal – http://exrna.org

Data Repository

3. DCC Wiki – http://genboree.org/theCommons/projects/exrna-mads/wiki Data Analysis Tools Wiki - http://genboree.org/theCommons/projects/exrna-tools-may2014/wiki

© Bioinformatics Research Laboratory, Baylor College of Medicine

- 5. Software Resources All links to data analysis tools and learning materials can be found under the "Resources" tab in the exRNA Portal. http://exrna.org/resources/ 6. Acknowledgements: Genboree Dev Team at Baylor: Andrew Jackson, Sameer Paithankar, Neethu Shah, Aaron Baker; Downstream Analysis Pathway/Network Tools: Anders Riutta, Kristina Hanspers





Access the exRNA **Portal** by scanning this QR Code.