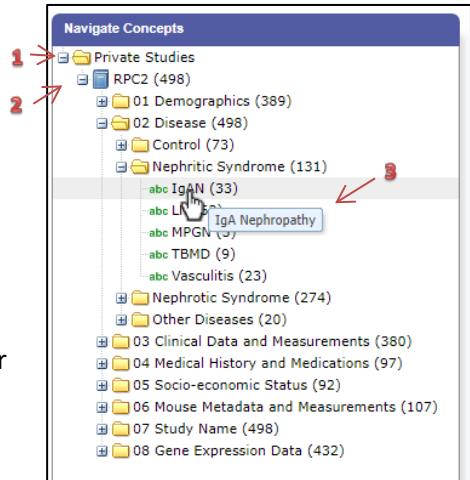


The RPC2 tranSMART instance can be accessed here: <https://rpc2-transmart.med.umich.edu/>

Do **NOT** use Internet Explorer; **Chrome**, **Firefox**, or **Safari** work fine. The first time you log in (and once every 90 days after that), you will be asked to attest to a Data Use Statement.

## Overview

To explore the data available for each study, click on the plus sign (+) next to “Private Studies” in the **Navigate Concepts** section (1). Then, click on the plus sign next to “RPC2” to expand the terms (2).

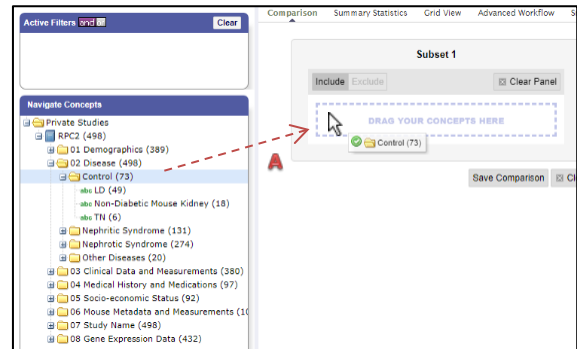


The tree provides additional useful information:

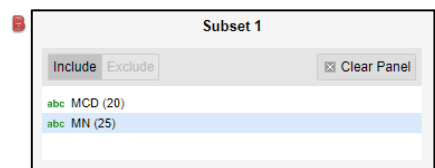
- The numbers after the terms are the number of samples that have data for that term.
- The **123** or **abc** before the term indicates if it is a numeric or character response.
- Hovering over a node, or right clicking on it, will bring up a definition pop-up for some terms (3).

## Selecting Data

The first step is to establish the Subset(s) of data you wish to explore. Drag and drop the terms from the tree into the **Subset** boxes on the right (A).



- If an entire folder is dragged over, all samples with a value for that data point are included in the subset regardless of the value (e.g. dragging the folder “Control” would include all patients with values of “LD”, “Non-Diabetic Mouse Kidney”, and “TN”).
- If only a single term is included, then only samples having that value are included (e.g. dragging “MN” would include only those patients assigned a value of “MN”).
- If a numeric term is selected, values can be further restricted by choosing an option from the pop-up menu. Select “No Value” to use all samples with any value.
- Terms in the same subset box are part of an “OR” statement: “Disease = MN OR MCD” (B).
- Separate subset boxes are joined together in “AND” statements: “Disease = FSGS” AND “Male”.
- Two different subsets (**Subset 1** and **Subset 2**) can be defined and compared. In the example below, Subset 1: “MCD” OR “MN” AND “Male” is compared to Subset 2: “FSGS” AND “Male” (C).



### Analyzing Data

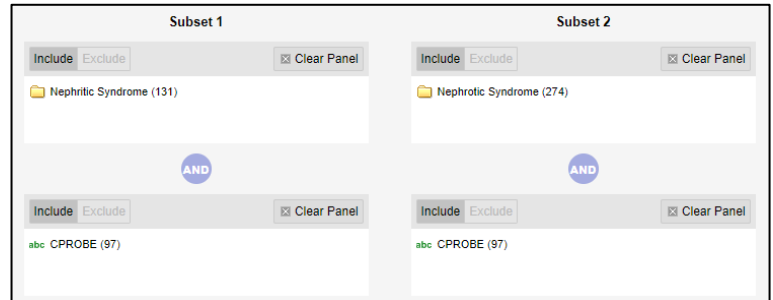
After establishing your subset(s), clicking on the **Summary Statistics** tab will bring up an overview of the subset(s) by Age, Sex, and Race. Scroll through the results to see summaries and breakdowns of the terms used in the subset statements.



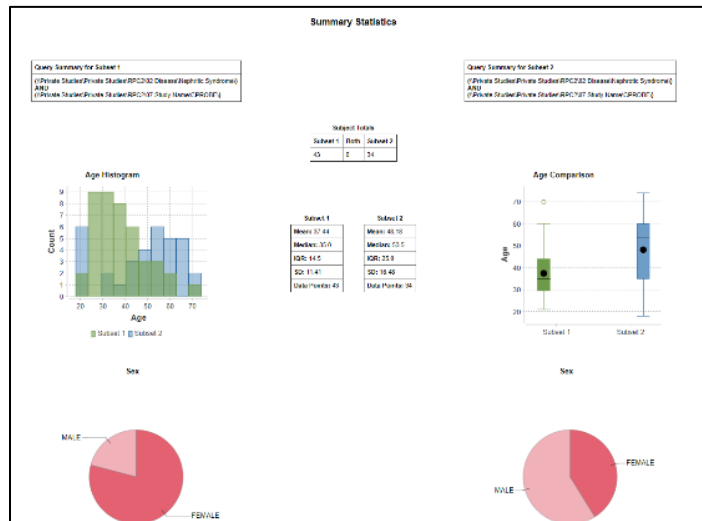
For the Subsets created, compare a concept across those Subsets by dragging an additional term from the tree directly onto the **Summary Statistics** results page.

#### Example 1: Compare GFR at baseline between Nephritic Syndrome and Nephrotic Syndrome in CPROBE data.

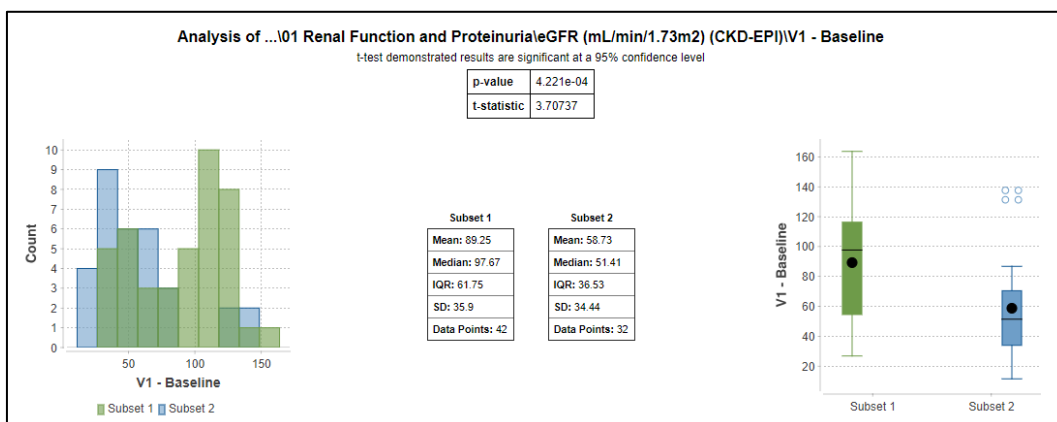
- From the “02 Disease” section of the **Navigate Concepts** tree, drag and drop “Nephritic Syndrome” into Subset 1 and “Nephrotic Syndrome” into Subset 2.
- From the “07 Study Name” section of the tree, drag and drop “CPROBE” into the second boxes of both Subset 1 and Subset 2.



- Click on **Summary Statistics**. At the top is a comparison of the two subsets by age, sex, and race:



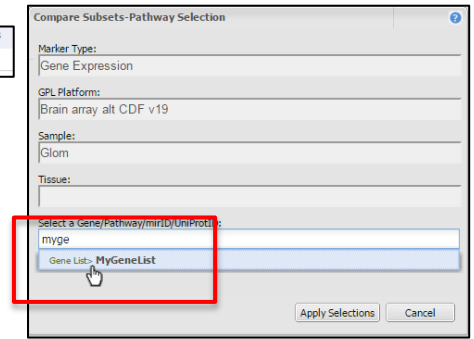
- Drag and drop “V1 – Baseline” from under the “eGFR (mL/min/1.73m2)(CKD-EPI)” folder onto the **Summary Statistics** results page. A box plot comparing GFR values in each subset will appear at the bottom of the page:



### Using Gene Expression Data

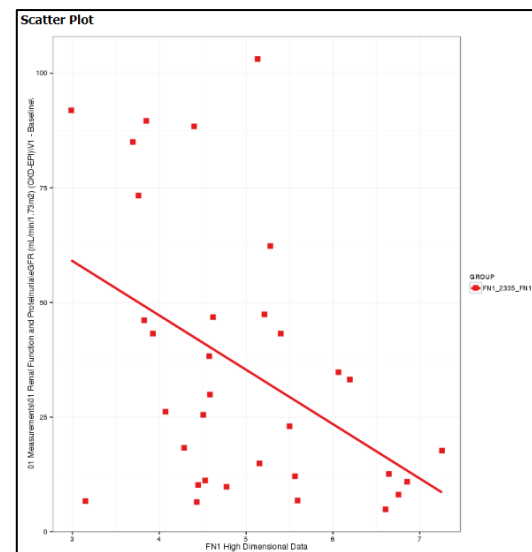
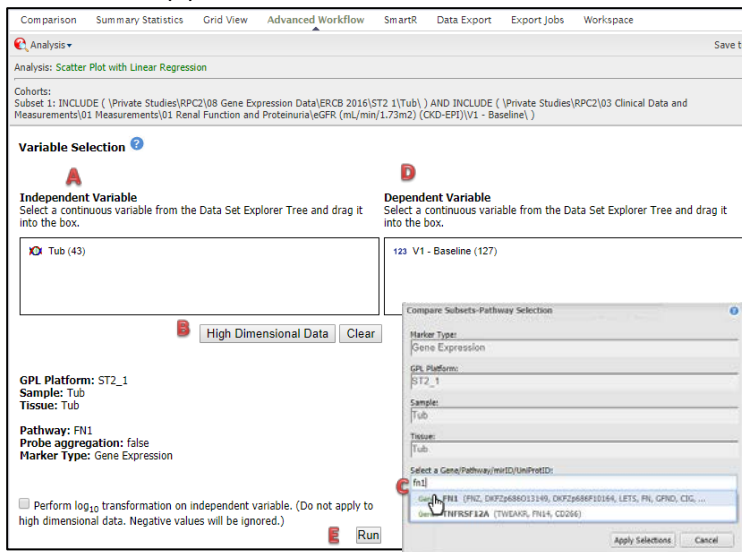


Explore genes of interest in gene expression data using the **Advanced Workflow** tab. Use a single gene or upload a gene list to compare multiple genes. To create a gene list, use the **Gene Signature/Lists** link at the top right-hand side of the application and follow the instructions (only fields with an asterisk are required). To use a gene list, type the name of the list into the “Select a Gene” box in the “High Dimensional Data” popup window and select it from the dropdown options.



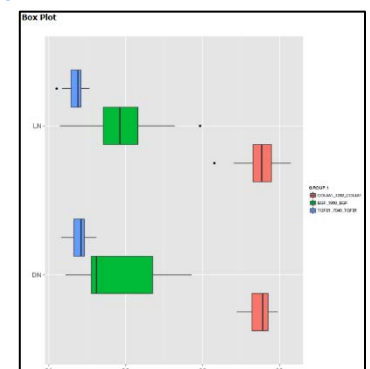
### Example 2: Create a scatter plot of GFR data against Gene Expression levels of FN1 in ERCB\_2016 data.

- Subset to desired patients on the **Comparison** tab:
  - Drag “Tub” from the ERCB 2016/ST2 1 folder in the tree into Subset 1. Select “No filter”. Drag “V1 – Baseline” from the “eGFR (mL/min/1.73m2)(MDRD)” folder into a second Subset 1 box.
- Click on the **Advanced Workflow** tab and select “Scatter Plot with Linear Regression” from the “Analysis” menu.
- Drag “Tub” from the tree into the “Independent Variable” box (A).
- Click on the “High Dimensional Data” button under the “Independent Variable” box (B).
- Type “FN1” in the gene box and then select it from the dropdown list (C). Apply Selections.
- Drag “V1 - Baseline” from the tree into the “Dependent Variable” box (D).
- Click “Run” (E).



### Example 3: Create a box plot to show expression of genes in ERBC disease cohorts of DN and LN.

- Create a gene list with Col4A1, EGF, and TGFB1 in the **Gene Signatures/Lists** tab as explained above.
- On the **Comparison** tab, subset to “DN” and “LN” in the first Subset box, and “Glom” (from ERCB/Brain array alt CDF v19 folder) into the second Subset 1 box.
- Click on the **Advanced Workflow** tab and select “Box Plot with ANOVA”. Drag “Glom” into the “Independent Variable”. Click on the “High Dimensional Data” button. Type the name of your gene list into the gene box and select it. Drag “DN” and “LN” into the Dependent Variable box.
- Click Run.

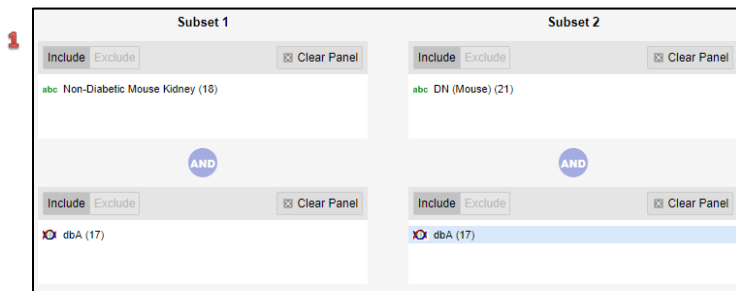
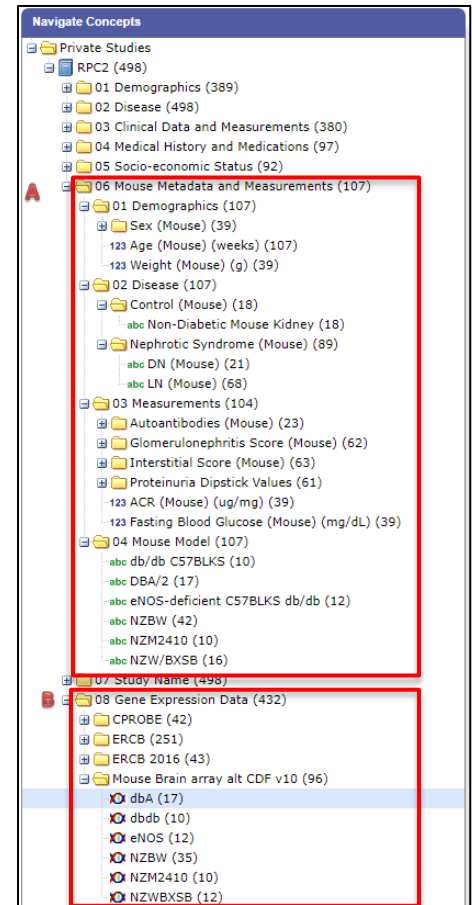


Working with Mouse data

The RPC2 tranSMART instance includes gene expression data for 6 mouse models. Metadata for mouse models can be found in the Navigate Concepts pane under “06 Mouse Metadata and Measurements” (A), and the gene expression data is available by model under “08 Gene Expression Data/Mouse Brain array alt CDF v10” (B).

Example 4: Find differential expression of genes between Non-Diabetic Mouse Kidney Control samples and DN (Mouse) samples in the DBA/2 mouse model.

- From the **Navigate Concepts** tree, drag “Non-Diabetic Mouse Kidney” into Subset 1 and “DN (Mouse)” into Subset 2. Then from under “08 Gene Expression Data/Mouse Brain array alt CDF v10”, drag, “dbA” into both Subset 1 and Subset 2. (1)



- Click on the **Advanced Workflow** tab and select “Marker Selection (Differential Expression)”.
- Drag “dbA” into the variable selection box.
- Click on the “High Dimensional Data” button. Type 2 or more mouse genes into the “Select a Gene...” box, selecting them from the auto-complete options and separating the genes with a comma. (2)
- Click Run to view the resulting heatmap. (3)

