

## Visualizing Expression data using the Co-expression Tool Web service and TableView

**Introduction.** TableView was written by James (Jim) E. Johnson and colleagues at the University of Minnesota Center for Computational Genomics and Bioinformatics.

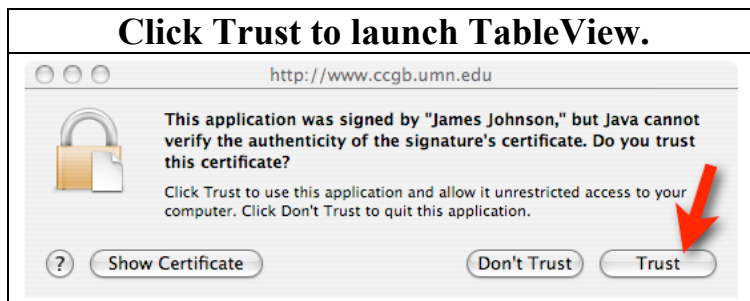
See: Johnson et al, "TableView: portable genomic data visualization" (2003) Bioinformatics Vol. 19, No. 10: <http://bioinformatics.oxfordjournals.org/cgi/content/abstract/19/10/1292>

We started using the program because it has interactive scatter plots, and so the rest of this tutorial will focus on this feature.

As of March, 2009, the University of Minnesota no longer supports TableView, but Jim (first author on the paper above) has kindly provided us a copy the program to distribute. To download your own copy, <http://igb.bioviz.org/links.shtml> and click the "Start TableView" with Java Web Start" button.

**How it works:** When you click the button, your browser will download the JNLP (Java Web Start) file. The browser will read the file, which tells it where to obtain the compiled code for the TableView program. It then will download the code and try to run it. You may then see a new window that looks something like the figure below.

To continue loading the software, you must click "Trust."

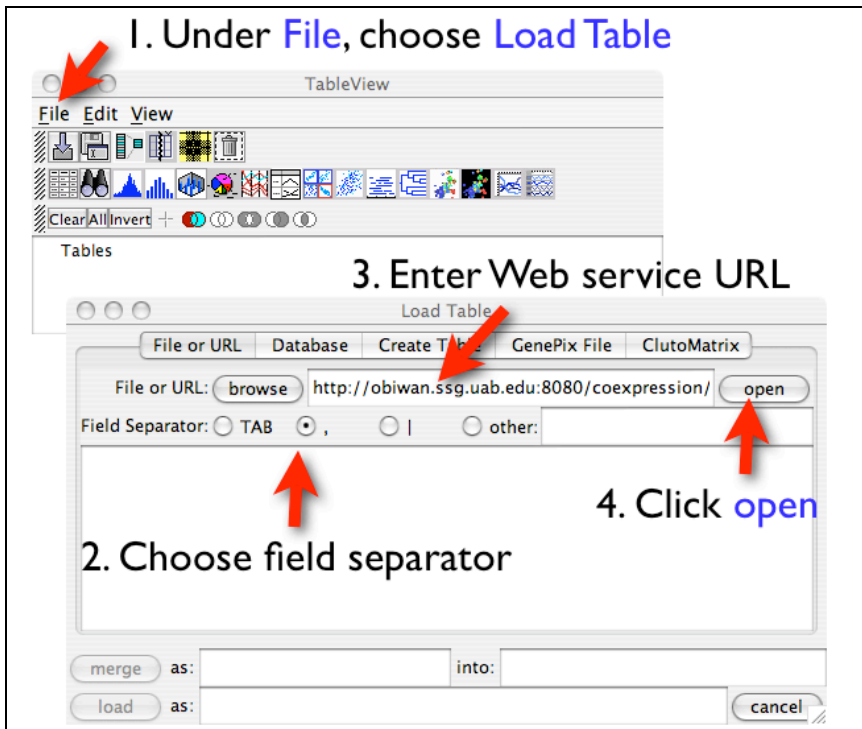


**The TableView Console.** Once the software downloads and starts up, a new window named "TableView" should open your desktop. Under the **File** menu, choose **Load Table**. A new window will open. Choose " ," as the field separator. Enter the Web service URL :

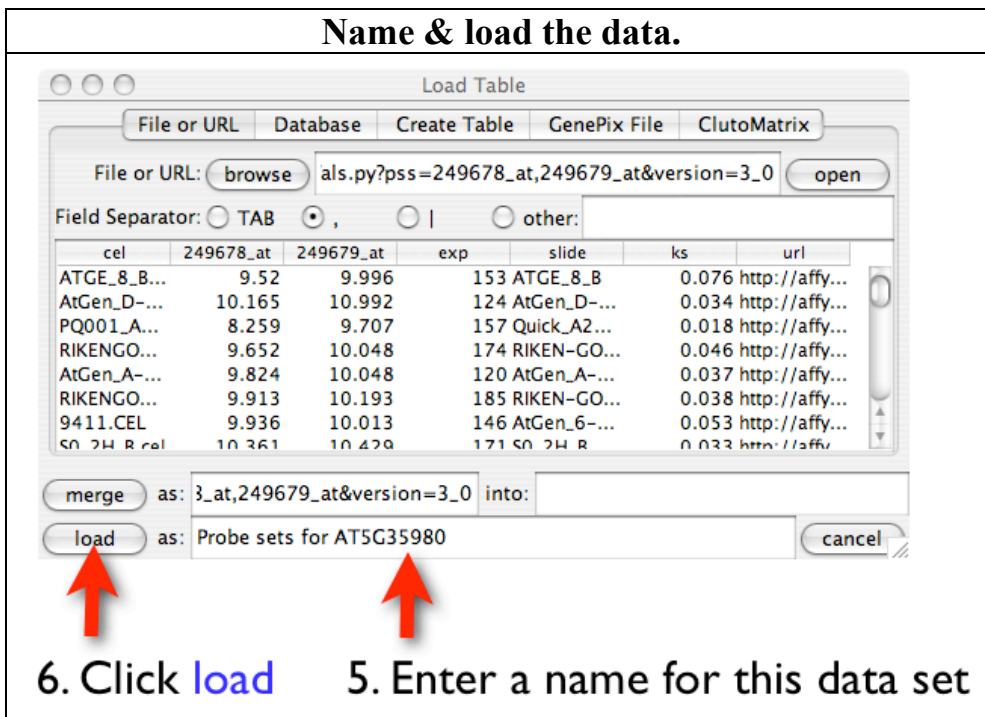
[http://www.cressexpress.org/cgi-bin/getExpVals.py?pss=249678\\_at,249679\\_at&version=3\\_0](http://www.cressexpress.org/cgi-bin/getExpVals.py?pss=249678_at,249679_at&version=3_0)

This URL will retrieve data release 3.0 expression values for ATH1 probe sets 249678\_at and 249679\_at, which have the same target gene: AT5G35980.

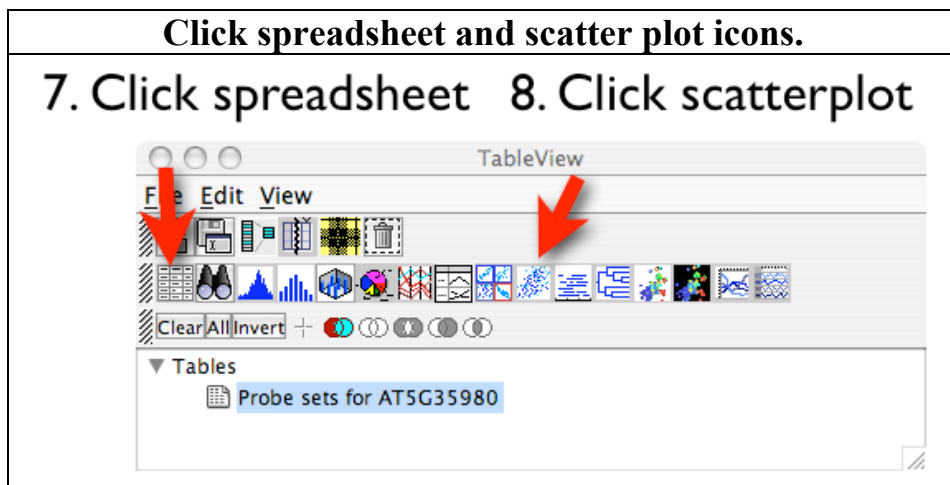
Note that probe set 249678\_at targets the five-prime region of the gene, and 249679\_at targets the 3-prime region of the gene. (Skip ahead to the last figure if you would like to see what the gene looks like.) According to TAIR version 7 annotations, this gene generates two different transcript variants with alternative three prime ends. The probe set 249678\_at can hybridize with both transcripts, but the probe set 249679\_at can hybridize with the longer form, the variant with an extended 3-prime end relative to the other.



When you click the “open” button, some newly-loaded data should appear in the central text area of the “Load Table” window. You can make working with the data a little easier later on by typing something easy to remember into the text box next to the load button. Whatever you type here will appear in the main TableView window. Then, click “load” to import the data into TableView.



The TableView window should now list the data set under the tables menu. Now click the spreadsheet and scatter plot icons.



This will open two new windows: one showing a spreadsheet with data from the selected table, and another window showing an empty plotting space.

To group all the slides from the same experiment, click the top or bottom triangle next to the exp column label.

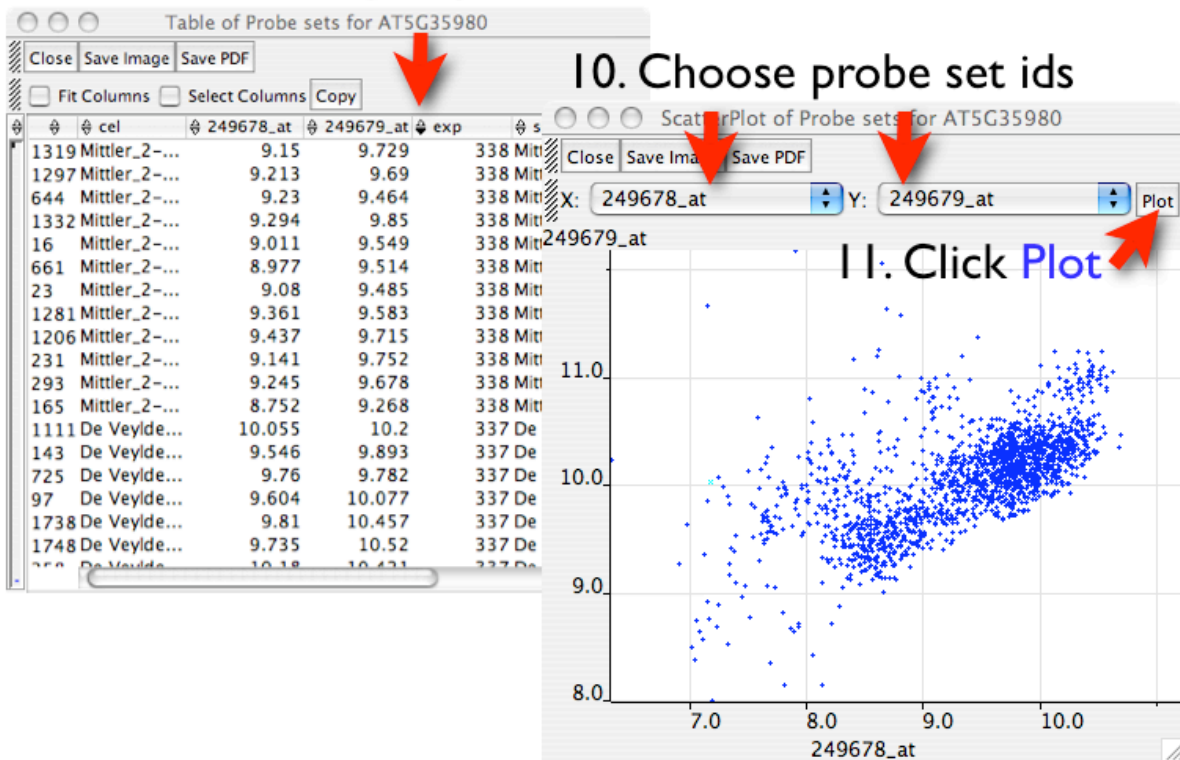
**Note:** Sorting in the spreadsheet view can be a bit tricky! For this to work, the cursor has to change from an arrow to another type of cursor – it looks like an “H” lying on its side. To see what we mean, hover the cursor over the triangles in the header cell and wait for the cursor to change. When it does, click. Until the cursor changes, the program will ignore your clicks.

If funding and time permit, we hope the TableView authors or other interested developers might look into making this sorting operation in the spreadsheet view a bit easier. Note that the source code for the program is freely available on the TableView Web site: <http://trac.cbri.umn.edu/tableview>. In theory, anyone with the right skills has the freedom to try his or her hand at improving this aspect of the program.

To view a scatter plot that shows the expression pattern of the two query probe sets plotted against each other, choose the probe sets in the scatter plot menus and then click **Plot**.

Sort slides by experiment, and then plot the expression data.

## 9. Click to group slides



Note that there are some slides along the upper left region of the plot where y-axis probe set values seem unusually high relative to the other probe set. These might be examples of experiments where the ratio of target transcripts relative to each other varies, perhaps due to differential regulation of alternative splicing.

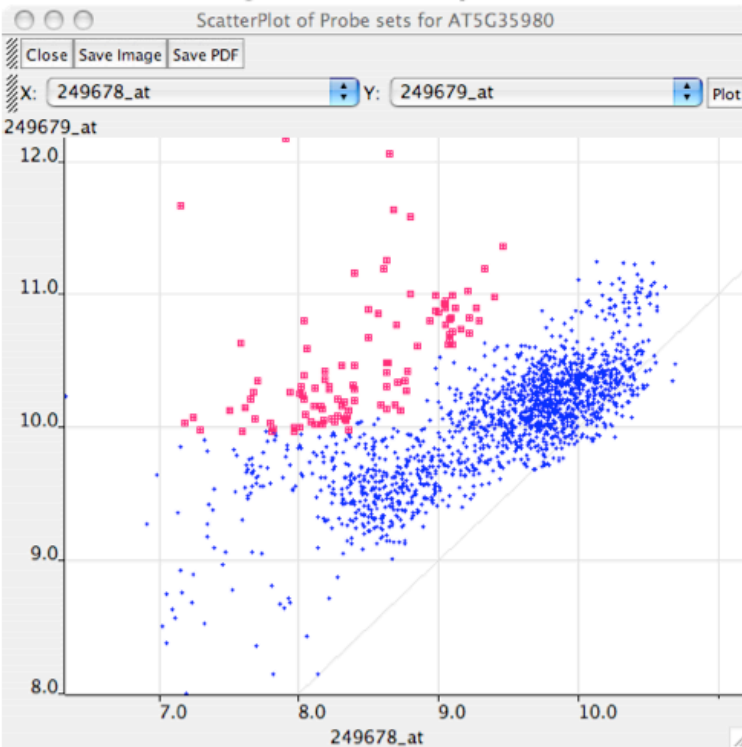
To investigate these points in detail, click drag over the upper left region of the plot. Rows corresponding to these points should now be selected in the spreadsheet view. To add new points to the selection, click and drag again, but with the **SHIFT** key depressed.

Now check the spreadsheet view. Use the scroller to move up and down through the spreadsheet.

Note the blue markings in the scrollbar gutter on the left side of the spreadsheet window and how the black bar moves over these as you scroll up and down. The blue bars stay in one place – these correspond to selected rows, and their relative position tells you where they are located within the larger spreadsheet. The black bar moves in concert with the scrollbar, and corresponds to the rows that are currently visible. In this way, TableView helps you stay oriented within this large spreadsheet of data.

## Selecting apparent outlier points.

### 12. Click-drag and SHIFT-click-drag to select points



Look carefully at the column labeled “slide” in the spreadsheet view.

The names of the slides in the “slide” column provide some clues as to the types of samples they represent, and the “url” column gives the Web address for the NASCArrays Web site that describes the the experiment the slides were part of.

If you enter this Web address into a Web browser, you can find out more information about the sample types and about how the original experiment was designed. However, this can sometimes be a bit difficult. Unfortunately, the names at NASC don’t always match up with the slide names in the spreadsheet. This is due to the fact that many slide names have changed since we imported the data into the co-expression tool database, which supplies data to the Web service URLs. This is a huge problem; so far as we can tell, there is no stable naming system for NASCArrays slides. We would love to be able to make it possible to “link out” to pages at NASCArrays where you could investigate individual slides and experiments in detail, but until NASCArrays provides a stable way to cross-reference slide names, this is very difficult. However, in most cases, it’s usually relatively clear which arrays are replicates of each other, and which sample types on the NASCArrays Web site correspond to slide names in the spreadsheet. We hope that some community curation efforts will make this easier in future. Alternatively, we are considering switching to the Gene Expression Omnibus as our source

of data, mainly because the naming schemes and conventions appear to be much easier to work with, at least for the mammalian data sets we have used in other projects.

**Selecting points on the scatter plot view highlights the corresponding rows in the spreadsheet view.**

### 13. Use scroller to view selected arrays.

	cel	249678_at	249679_at	exp	slide	ks	url
1723	AM002_A...	9.239E0	9.772E0	89	McCormac_2-8_gun-mutant-Dpretreated_Rep2_ATH1	1.09E-1	http://a
813	AM002_A...	8.576E0	9.544E0	89	McCormac_2-5_wildtype-Frprecon_Rep2_ATH1	1.09E-1	http://a
782	AM002_A...	9.148E0	9.546E0	89	McCormac_2-2_wildtype-Dpretreated_Rep1_ATH1	5.02E-2	http://a
1591	AM002_A...	8.705E0	9.705E0	89	McCormac_2-9_phyAmutant-FRprecon_Rep1_ATH1	5.77E-2	http://a
1062	AM002_A...	8.967E0	9.573E0	89	McCormac_2-1_wildtype-Frprecon_Rep1_ATH1	6.67E-2	http://a
1477	AM002_A...	8.974E0	9.586E0	89	McCormac_2-3_gun-mutant-Frprecon_Rep1_ATH1	7.15E-2	http://a
1674	JN001_AT...	9.541E0	1.046E1	85	Newbury_1-9_Halleri-control-leaves(HLO)_Rep3_ATH1	6.32E-2	http://a
1045	JN001_AT...	9.255E0	1.044E1	85	Newbury_1-12_Halleri-highZn-leaves(HLH)_Rep3_ATH1	4.92E-2	http://a
305	JN001_AT...	8.801E0	1.158E1	85	Newbury_1-1_Halleri-control-roots(HRO)_Rep1_ATH1	3.67E-2	http://a
872	JN001_AT...	8.499E0	1.067E1	85	Newbury_1-20_Petraea-control-leaves(PLO)_Rep2_ATH1	6.77E-2	http://a
232	JN001_AT...	8.68E0	1.164E1	85	Newbury_1-4_Halleri-highZn-roots(HRH)_Rep1_ATH1	3.82E-2	http://a
369	JN001_AT...	8.624E0	1.126E1	85	Newbury_1-14_Petraea-control-roots(PRO)_Rep2_ATH1	2.36E-2	http://a
80	JN001_AT...	7.623E0	9.667E0	85	Newbury_1-10_Halleri-highZn-leaves(HLH)_Rep1_ATH1	5.71E-2	http://a
1160	JN001_AT...	8.979E0	1.088E1	85	Newbury_1-5_Halleri-highZn-roots(HRH)_Rep2_ATH1	3.66E-2	http://a
480	JN001_AT...	9.585E0	1.062E1	85	Newbury_1-11_Halleri-highZn-leaves(HLH)_Rep2_ATH1	7.67E-2	http://a
265	JN001_AT...	9.102E0	1.1E1	85	Newbury_1-23_Petraea-highZn-leaves(PLH)_Rep2_ATH1	5.16E-2	http://a
549	JN001_AT...	8.982E0	1.099E1	85	Newbury_1-6_Halleri-highZn-roots(HRH)_Rep3_ATH1	7.52E-2	http://a
1743	JN001_AT...	8.564E0	1.086E1	85	Newbury_1-16_Petraea-highZn-roots(PRH)_Rep1_ATH1	2.48E-2	http://a
1100	JN001_AT...	8.636E0	1.049E1	85	Newbury_1-17_Petraea-highZn-roots(PRH)_Rep2_ATH1	9.18E-2	http://a
537	JN001_AT...	9.217E0	1.045E1	85	Newbury_1-18_Petraea-highZn-roots(PRH)_Rep3_ATH1	4.68E-2	http://a
1046	JN001_AT...	9.952E0	1.072E1	85	Newbury_1-24_Petraea-highZn-leaves(PLH)_Rep3_ATH1	6.27E-2	http://a
293	JN001_AT...	7.712E0	1.035E1	85	Newbury_1-19_Petraea-control-leaves(PLO)_Rep1_ATH1	3.35E-2	http://a
1210	JN001_AT...	9.332E0	1.12E1	85	Newbury_1-13_Petraea-control-roots(PRO)_Rep1_ATH1	1.94E-2	http://a
977	JN001_AT...	8.493E0	1.089E1	85	Newbury_1-3_Halleri-control-roots(HRO)_Rep1_ATH1	1.009E-1	http://a
1398	JN001_AT...	8.796E0	1.1E1	85	Newbury_1-15_Petraea-control-roots(PRO)_Rep3_ATH1	8.37E-2	http://a
1408	JN001_AT...	9.46E0	1.136E1	85	Newbury_1-2_Halleri-control-roots(HRO)_Rep2_ATH1	5.17E-2	http://a
409	JN001_AT...	7.63E0	9.718E0	85	Newbury_1-7_Halleri-control-leaves(HLO)_Rep1_ATH1	4.02E-2	http://a
1093	JN001_AT...	9.476E0	1.052E1	85	Newbury_1-21_Petraea-control-leaves(PLO)_Rep3_ATH1	6.72E-2	http://a
726	JN001_AT...	8.701E0	1.078E1	85	Newbury_1-8_Halleri-control-leaves(HLO)_Rep2_ATH1	6.65E-2	http://a
695	JN001_AT...	8.393E0	1.047E1	85	Newbury_1-22_Petraea-highZn-leaves(PLH)_Rep1_ATH1	5.23E-2	http://a
801	RW001_A...	8.251E0	9.59E0	84	Walters_A-09-Kruger-ML3_REP3	4.03E-2	http://a
562	RW001_A...	8.397E0	9.61E0	84	Walters_A-06-Kruger-WH3_REP3	1.46E-2	http://a

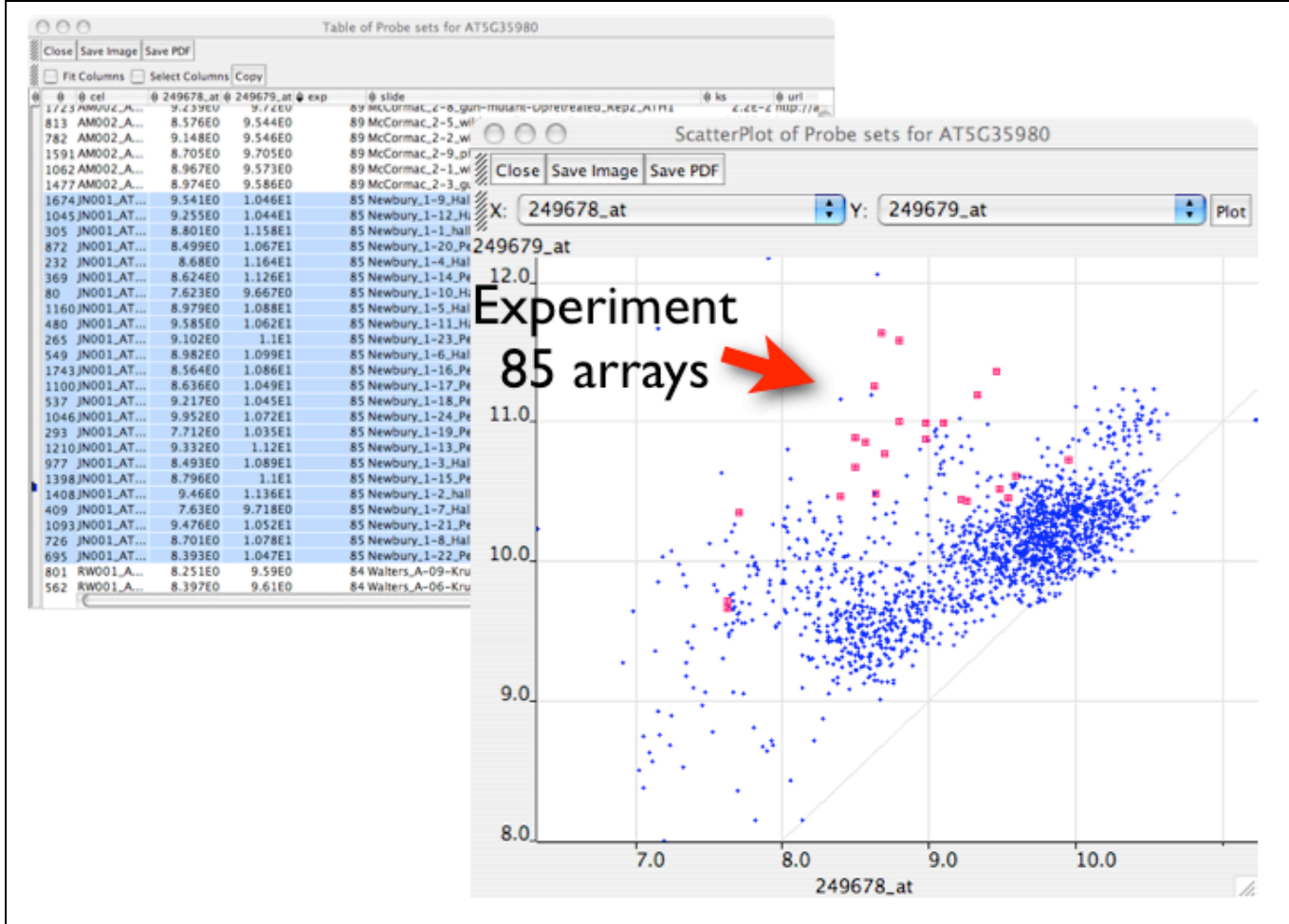
**Black bar shows position of current view in the entire spreadsheet.**

The selection process works both ways. To select rows, click somewhere on the spreadsheet to deselect everything, and then click-drag over a set of rows to select them. Then, view the highlighted arrays plotted on the scatter plot view.

This is the feature we called “interactive scatter plots” earlier in the tutorial – when you interact with the spreadsheet view, the scatter plot responds, and vice versa.

Other views in TableView behave the same way. You should try some of these out. This tutorial is limited to just the spreadsheet and scatter plot views, but these two views are just a small part of what the program can do.

**Highlighting slides in the spreadsheet view selects the corresponding points on the plot. All the slides from NASCArrays Experiment #85 are selected.**



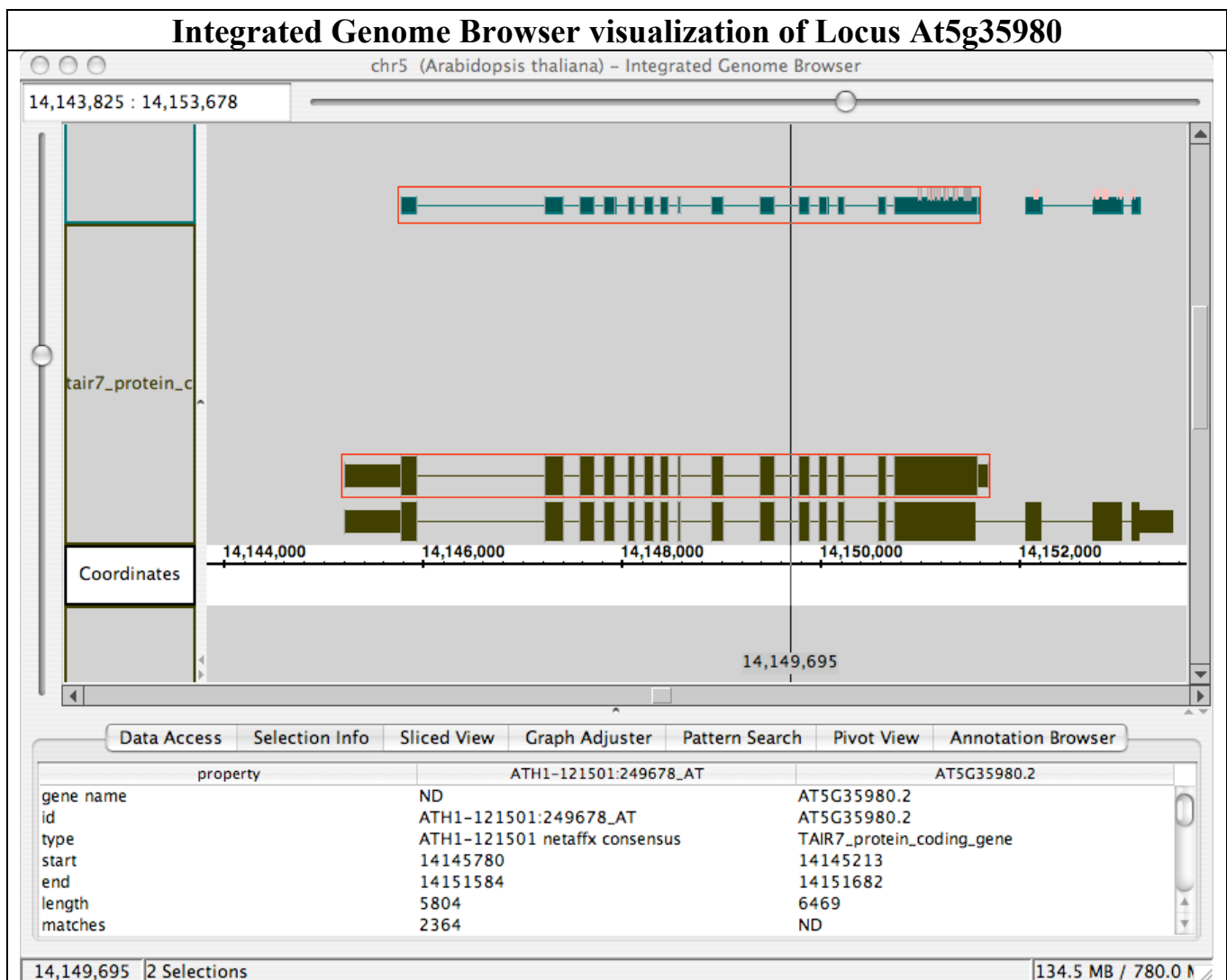
Note again the upward sweep of points on the left of the plot and the corresponding lack of outliers occupying the opposite, bottom right corner. This suggests that the readings from the probe set on the x axis are unusually low when compared to the y-axis probe set, and this trend is apparent for a large number of slides.

One explanation for this asymmetric pattern of outliers might be RNA degradation. It is well-known that measurements from Affymetrix (and other types of arrays) are sensitive to RNA degradation, which tends to impact 5-prime probe sets more so than 3-prime probe sets, relative to a common target transcript. It is possible that the x-axis probe set is more vulnerable to degradation. As a result, points that would be further to the right (higher x-axis values) seem swept back to the left. That is, the y values (y-axis probe set) for those points are just fine, but the x values are too small.

Alternatively, it may be that these seeming outlier points indicate some type of regulated shift in the relative amounts of splice variants detected by the different probe sets. Imagine a scenario in which the y-axis probe set target (the longer variant, with an extended 3-prime end) is expressed at a higher level relative to the other, shorter transcript. In that case, the y values on the plot might seem unusually large, giving rise to an upward "sweep." In this case, the x axis values are just where they should be, but the y axis values are a bit too large, relative to the other points on the plot.

It is very difficult to tease apart these two possibilities. For this locus, the TAIR version 7 genome annotations report two distinct transcript variants with alternative three prime ends. One probe set (the more 3-prime probe set) only detects one of these transcripts, while the more five-prime probe set detects both. However, the distance separating the five-prime probe set's target region from the three prime terminus is different for each of the two target transcripts. This means that the readings from the five prime probe set when detecting the longer transcript are perhaps less reliable in some samples, due to RNA degradation problems.

To help you consider the possibilities and think about the relative locations of the probe sets relative to the gene (DNA) and their putative target transcripts (RNA), here is a screen shot from yet another visualization tool, the Integrated Genome Browser. The screen shot shows the probe sets (green), probes (light pink), and gene models (brown) they are designed to interrogate. For more information on IGB and how to use it, visit <http://igb.bioviz.org>.



Comments and suggestions on this tutorial should go to:

Ann Loraine  
 aloraine@gmail.com  
 aloraine@uncc.edu