Guide for Image and Data Formats for eFP Browser/ePlant Setup

22 November 2017, Nicholas Provart

To create your own eFP Browser, you'll need two things: one, an **image** that depicts how you set up your experiment to obtain RNA, and second, **gene expression measurements** for all of the parts or treatments depicted in your image. Note that we can use the bitmap image you create for an eFP Browser to fairly easily generate an SVG (vector) image to use in ePlant. The data feed will be the same.

1. Creating an image

Use Photoshop or a similar program to put together your depiction of your experimental setup. We recommend using simple line drawings, which can be done by hand and then scanned. We have had success too with using photographs and radically changing the contrast and brightness to make a line-drawing-like representation. The goal is to have areas of the image that are easy to "flood fill" with a single colour, as each colour acts as a key to the corresponding sample data (for more information on this, see the original eFP Browser paper in PLoS ONE by Winter et al., 2007). Set the image resolution to 72 dpi, as this is screen resolution. Set the colour mode to be 8 bit RGB. A typical size is 800 x 600 pixels, but the image may be wider or longer as needed, see **Figure 1**. The image may be anti-aliased to create smooth lines.

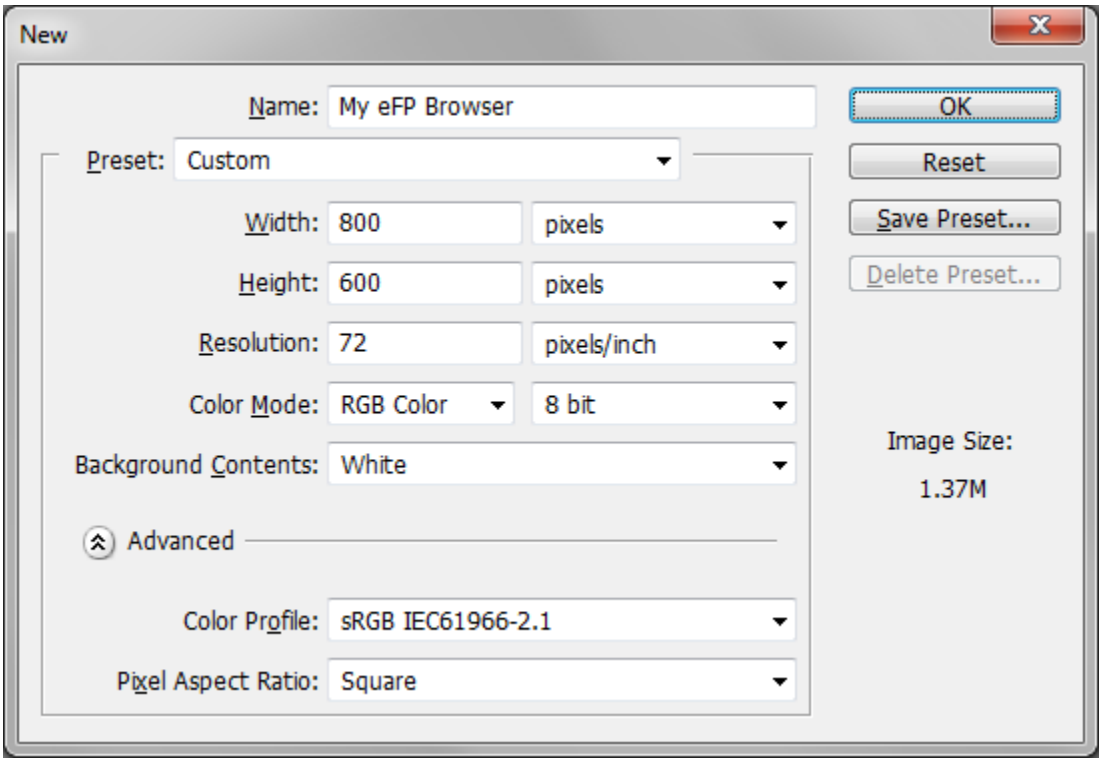


Figure 1: Adobe Photoshop settings for creating a new eFP image.

Once you have an image assembled, use the flood fill function to fill in the areas that correspond to the tissues you sampled for extracting RNA. This is accomplished using the Paint Bucket tool of Adobe Photoshop. Check the "Contiguous" option but ensure that the "Anti-aliased" option is unchecked, see

Figure 2. It is crucial to deactivate the "Anti-aliased" option, otherwise multiple hues of the selected colour will be present at the bounding areas, and these won't be cleanly replaced by the eFP Browser engine, see **Figure 3**. Each depiction of the tissues from which RNA was isolated should be filled with a unique colour (don't worry, there are more than 65,000,000 to choose from). One final point is to leave a 50 pixel space along the top left of the image – this will be where gene information is added when your eFP Browser is functional. Also, **do not resize the image after you have done the colour flood filling with the paint bucket tool**, as this will cause the colours at the edge areas to be dithered, creating pixels that won't get replaced by the eFP Browser engine.

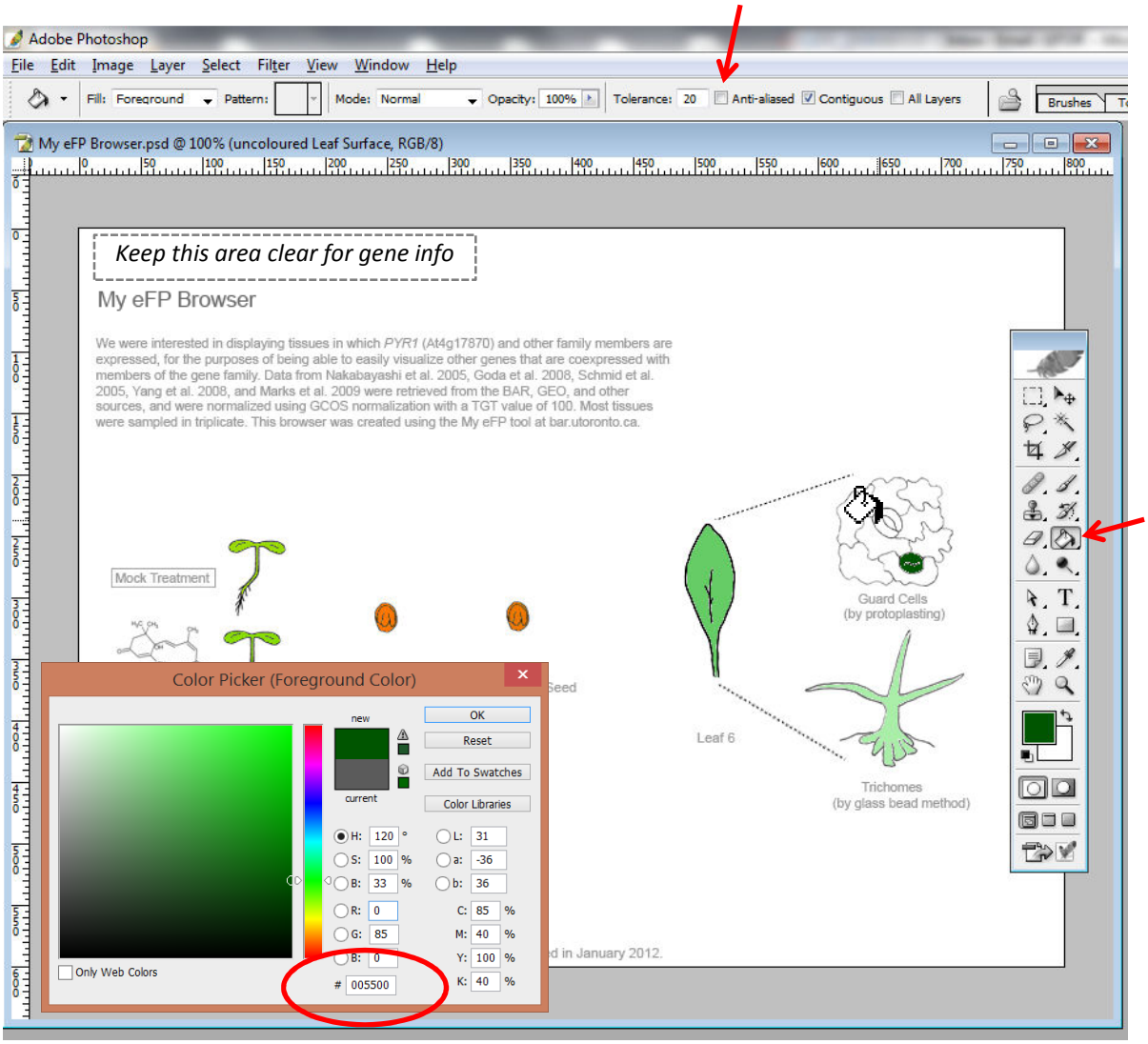


Figure 2: Using the Paint Bucket tool to fill a region on the image with a specified colour. Ensure that the "Anti-aliased" option is deactivated. Record the hexadecimal RGB colour from the Color Picker (#005500 for the green being used to fill the guard cell here) in the sample descriptions sheet (see 3.) for use later.

You can play around with the "Tolerance" value when using the paint bucket tool. Increasing this value will perhaps improve the colour fill within a given area.

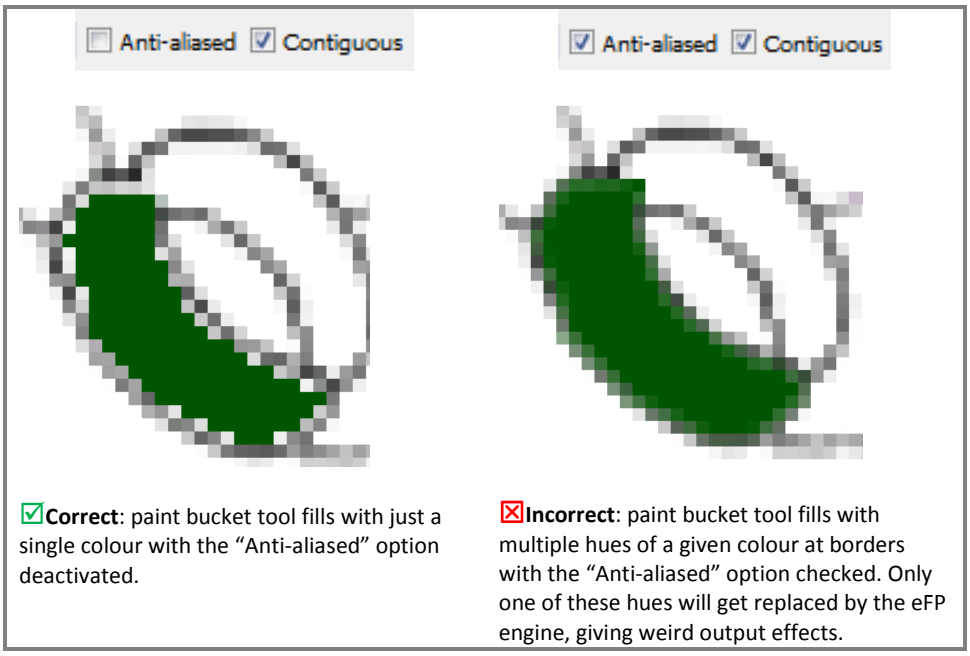


Figure 3: Effect of not deactivating the "Anti-aliased" option.

Once you have suitable image file, "Save as…" a PNG file, with no interlacing.

2. Creating a Data File

For every part or treatment depicted in your image, you require at least one set of measurements. It is typical, however, to have biological replicates, i.e. duplicated or triplicated measurements. You will be associating these measurements (samples) with the parts or treatments you have depicted in your image. You can use any normalization/summarization method you wish. For RNA-seq data we recommend FPKM, and for microarray data GCOS with a target value of 100 (this is how many of the existing data sets have been processed so comparison with other views is easier).

Sample names

Probe sets/ gene IDs

Tissue_Hormone_mini_matrix.csv - Excel

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW ADD-INS Sign in

A1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		RIKEN-NAKABAYASHI1A	RIKEN-NAKABAYASHI1B	RIKEN-NAKABAYASHI2A	RIKEN-NAKABAYASHI2B	ATGE_14_A	ATGE_14_B	ATGE_14_C	JS85	JS33	ColtrichomeArd1	ColtrichomeArd2	ColtrichomeMN12	Coltric meMN
2	244901_at	305.8	331.5	9.1	18.9	20.5	28.0	28.2	267.3	734.8	51.8	48.5	98.6	9
3	244902_at	194.8	168.0	12.5	19.9	27.3	25.4	25.8	285.6	810.1	32.6	44.4	56.8	7
4	244903_at	370.6	374.1	19.6	32.0	84.0	96.6	76.8	160.7	803.5	244.9	338.6	185.3	11
5	244904_at	44.9	53.8	6.4	7.6	16.6	35.2	28.6	45.1	55.6	31.2	45.0	21.9	1
6	244905_at	15.9	3.1	11.5	8.1	5.2	12.6	14.1	26.3	19.3	16.3	8.8	18.1	1
7	244906_at	227.2	127.1	24.7	39.4	57.7	47.2	49.8	408.5	2845.3	382.8	311.9	430.5	32
8	244907_at	1.7	8.0	0.6	0.3	4.2	1.2	1.8	21.6	58.2	6.0	3.9	5.7	
9	244908_at	0.6	5.3	0.5	0.9	0.8	0.5	2.4	5.0	13.2	0.5	0.2	0.5	
10	244909_at	22.7	6.9	7.1	2.1	18.6	10.9	13.4	43.0	56.1	7.6	12.7	28.8	2

Figure 4: Data matrix of expression values with one gene/probe set per row, and one sample per column. Cells contain the expression level for a given combination. Here the first 2 data columns are biological replicates for the dry seed sample. Note that sample names must match up with those in the Excel sheet you create in Section 3.

It should be fairly easy to retrieve such data from NCBI's GEO, ArrayExpress, or other primary data repositories, and to create such a data file, perhaps by processing the data with BioConductor. Save this file as a CSV (comma separated values) file. The eFP engine uses a linear colour scale scheme so don't submit log-transformed values.

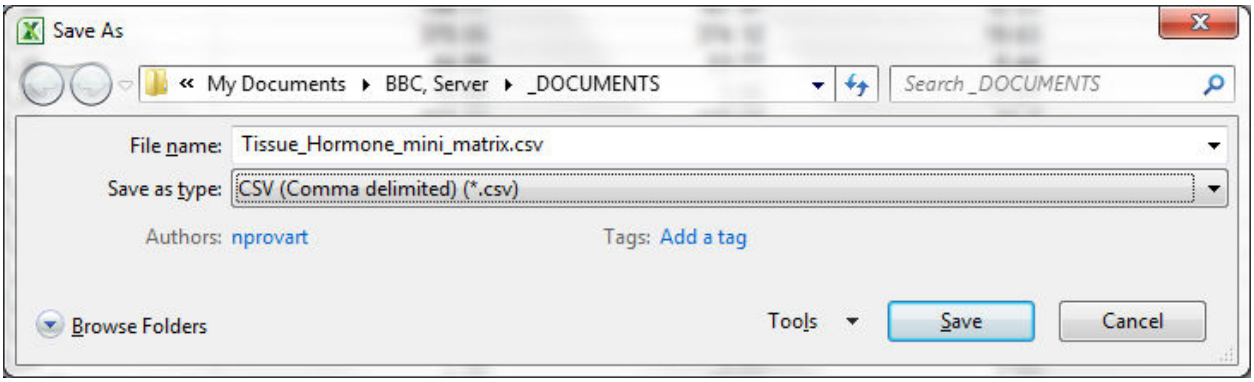
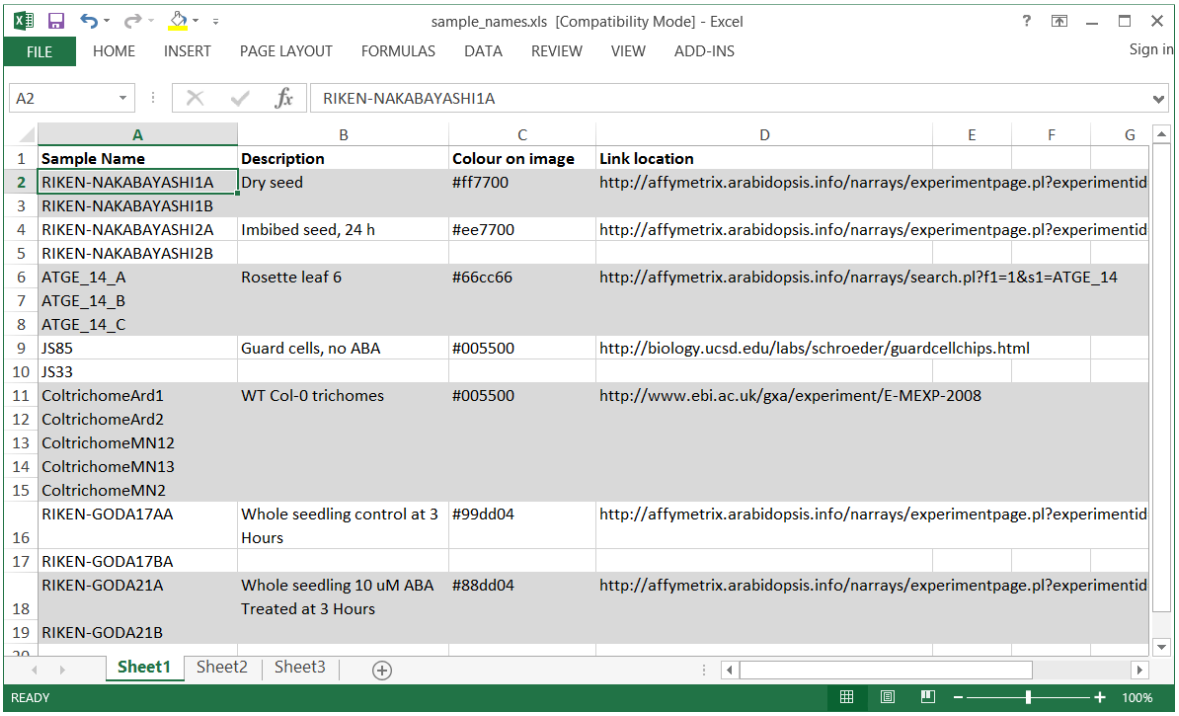


Figure 5: Saving the expression data in CSV format.

3. Bringing It All Together

With these two files in hand, we are close to being able to create your own eFP view. One final "nice-to-have" for us for making your eFP Browser is a list of the sample names, the corresponding part / treatment information as a description, and if desired, a linkout to a location on the internet (e.g., GEO/SRA record URL – copy this from the address bar of your browser), along with the colour you used.



	Sample Name	Description	Colour on image	Link location
2	RIKEN-NAKABAYASHI1A	Dry seed	#ff7700	http://affymetrix.arabidopsis.info/narrays/experimentpage.pl?experimentid
3	RIKEN-NAKABAYASHI1B			
4	RIKEN-NAKABAYASHI2A	Imbibed seed, 24 h	#ee7700	http://affymetrix.arabidopsis.info/narrays/experimentpage.pl?experimentid
5	RIKEN-NAKABAYASHI2B			
6	ATGE_14_A	Rosette leaf 6	#66cc66	http://affymetrix.arabidopsis.info/narrays/search.pl?f1=1&s1=ATGE_14
7	ATGE_14_B			
8	ATGE_14_C			
9	JS85	Guard cells, no ABA	#005500	http://biology.ucsd.edu/labs/schroeder/guardcellchips.html
10	JS33			
11	ColtrichomeArd1	WT Col-0 trichomes	#005500	http://www.ebi.ac.uk/gxa/experiment/E-MEXP-2008
12	ColtrichomeArd2			
13	ColtrichomeMN12			
14	ColtrichomeMN13			
15	ColtrichomeMN2			
16	RIKEN-GODA17AA	Whole seedling control at 3 Hours	#99dd04	http://affymetrix.arabidopsis.info/narrays/experimentpage.pl?experimentid
17	RIKEN-GODA17BA			
18	RIKEN-GODA21A	Whole seedling 10 uM ABA Treated at 3 Hours	#88dd04	http://affymetrix.arabidopsis.info/narrays/experimentpage.pl?experimentid
19	RIKEN-GODA21B			

Figure 6: Sample descriptions table. Note that sample names must match those in the data table, see **Figure 4**.

With these files in hand, we will work with you to set up an eFP Browser instance, which we can then convert to an ePlant instance. Basically, the last thing that needs to happen is to create an XML configuration file, that would look like this, based on the above data:

```
<?xml version="1.0" encoding="UTF-8"?><!DOCTYPE specimen SYSTEM
"http://www.bar.utoronto.ca/efp/cgi-bin/data/efp_config.dtd">
<specimen name="Arabidopsis">
<view name="all" class="" db="efpexpressiondata" dbGroup="all"
img="My_eFP_Test_20120216140211_expr.png" table="My_eFP_Test_20120216140211">
<coords graphX="508" graphY="152" graphWidth="60" graphHeight="80" legendX="621"
legendY="17"/>
<group name="RIKEN-GODA17AA;RIKEN-GODA17BA">
<control sample="RIKEN-GODA17AA"/>
<control sample="RIKEN-GODA17BA"/>
<tissue colorKey="#99dd04" name="Whole seedling control at 3 Hours">
<link
url="http://affymetrix.arabidopsis.info/narrays/experimentpage.pl?experimentid=175"/>
<area coords="111,243,177,246,157,315,117,312"/>
<sample name="RIKEN-GODA17AA" />
<sample name="RIKEN-GODA17BA" />
</tissue>
<tissue colorKey="#88dd04" name="Whole seedling 10 uM ABA Treated at 3 Hours">
<link
url="http://affymetrix.arabidopsis.info/narrays/experimentpage.pl?experimentid=176"/>
<area coords="108,319,167,319,167,401,108,401"/>
<sample name="RIKEN-GODA21A" />
<sample name="RIKEN-GODA21B" />
</tissue>
</group>
… (repeat <group> </group> for each experiment you are depicting, using appropriate
controls)
</view>
</specimen>
```

This XML file goes in the same DATA directory as the image. The “db” name is the name of the SQL database where you’ve stored your expression data. Sample names must match those in the DB. The “area coords” defined the (x,y) coordinates (top left of the image is 0,0) of a polygon on the image that is clickable in the eFP Browser output, linking to the URL specified. Again, see Winter et al. (2007) paper for full details.

Here’s how the eFP Browser would look before and after entering a query gene:

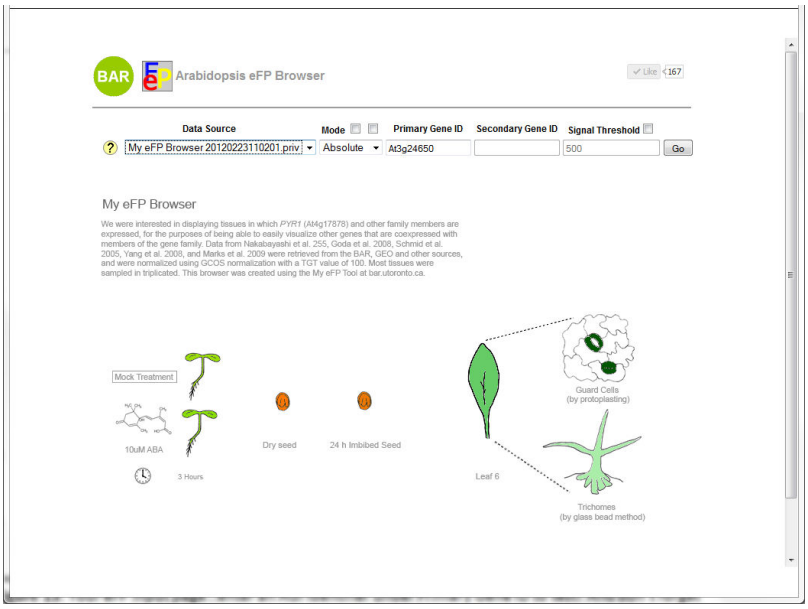


Figure 7: Your "My eFP Browser" input page.

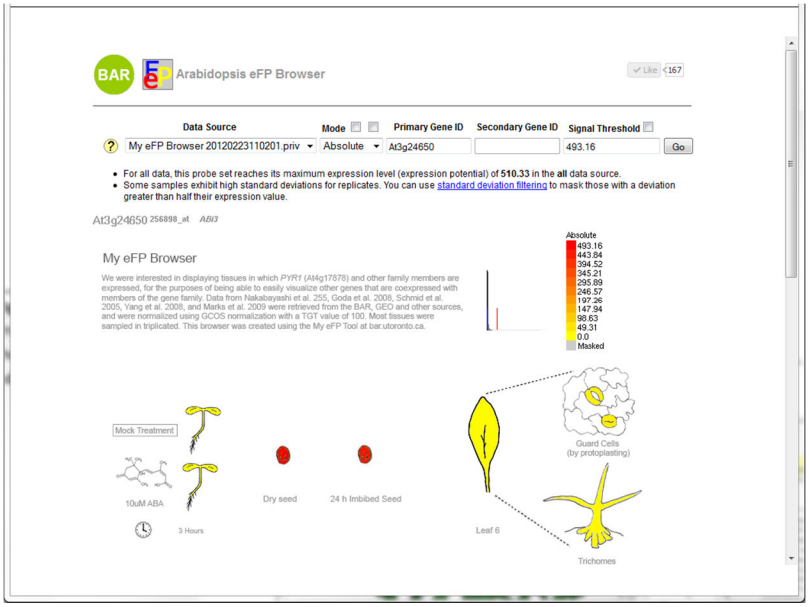


Figure 8: My eFP Browser output for the *ABI3* gene, At3g24650, which is strongly expressed in dry and imbibed seeds.