**Exercise: Finding expression data for your candidate genes**

***In this exercise we will use the ExpVIP database of wheat expression data, from*** [www.wheat-expression.com](http://www.wheat-expression.com) and <http://bar.utoronto.ca/efp_wheat/cgi-bin/efpWeb.cgi>***, to look at the expression patterns of the candidate genes within your region of interest.***

***We have a list of candidate genes which are located in a QTL for grain size on chromosome 6A. Can we use expression data to narrow down potential candidate genes?***

1. Open the list of candidate genes (“gene\_ID\_list\_for\_practical.xlsx”) and navigate to www.wheat-expression.com
2. Paste the list of candidate genes into the “Multiple Genes” box, making sure that the “TGACv1” reference is chosen in the dropdown box, and click on the “heatmap” button.
3. Look at expression levels for the genes in different metadata categories. Which metadata categories are most useful for this trait? Which are not?
4. Based on the expression patterns and what we know about our trait of interest, which genes look like good candidates? Why?
5. Now we’ll have a look at the expression levels of different genes in specific samples. To do this, you can either hover over the shaded boxes to see the log2 TPM of each or download the data to look at in Excel. Which gene is most highly expressed *overall* in the grain? What about in the *aleurone layer*?
6. Now, for one of the genes in the candidate list, paste its name in the “search” box at the top right of the screen and hit “search”. Once you have the bargraphs of expression data, tick the box next to “homoeologs” to see the expression levels of the homoeologs. How do their expression levels differ? Do the same for at least one more of the candidate genes—do the patterns differ?
7. Now go back to the initial heatmap. Using the different metadata categories, can you download the TPM data for the candidate genes by tissue and age? What metadata categories would also be useful to include?

While there are a number of genes within the list of candidates that would be worth following up if this was your real region of interest, for the purposes of this practical we are going to focus on two candidates: **TRIAE\_CS42\_6AS\_TGACv1\_490017\_AA1577030** and **TRIAE\_CS42\_6AS\_TGACv1\_485817\_AA1552750**

We want to check their expression levels in the eFP browser (http://bar.utoronto.ca/efp\_wheat/cgi-bin/efpWeb.cgi) but to do this we need to have their RefSeqv1.0 IDs. The easiest way to get these is to download the CDS for the TGAC genes from Ensembl and then BLAST these CDS at www.wheat-expression.com to get the RefSeqv1.0 gene IDs. These can then be used in the eFP browser.

1. Download the CDS for **TRIAE\_CS42\_6AS\_TGACv1\_490017\_AA1577030** and **TRIAE\_CS42\_6AS\_TGACv1\_485817\_AA1552750** from ensembl plants (<http://plants.ensembl.org/index.html>). (Search for the gene, then click the “transcript” tab, in the lefthand menu click on “cDNA” and download the sequence for the CDS).
2. BLAST these CDS against the RefSeqv1.0 gene models at www.wheat-expression.com. What are the gene IDs in RefSeqv1.0 format?
3. Use the eFP browser (http://bar.utoronto.ca/efp\_wheat/cgi-bin/efpWeb.cgi) to look at the expression of these two genes separately. Make sure to input the gene ID rather than transcript ID otherwise the eFP browser will not find your gene of interest. Which tissues are these genes most highly expressed in?