Expression browsers

The reduction in cost of next generation sequencing has generated an explosion of gene expression data from RNA-seq studies in wheat in the last 5 years. However until recently most of this data has been inaccessible to wheat researchers and breeders due to the complicated bioinformatics analysis required to unlock this information. The development of gene expression atlases for wheat allows researchers to access information about the expression of genes in difference circumstances. For example a list of candidate genes within a mapping interval can be narrowed down using the expression patterns of these genes.

Currently two expression browsers exist which will be described below: WheatExp and expVIP.

a) WheatExp

WheatExp (available at http://wheat.pw.usda.gov/WheatExp/) includes gene expression data from 6 studies:

- different tissues over development
- grain development
- grain layers
- senescence timecourse
- drought and heat stress
- photomorphogenesis

Each study is displayed separately with a graphic interface describing the particular study that is being viewed. For example expression of the gene Traes_4DS_1CF5D9107.1 in the "grain layer" study reveals higher expression in the endosperm than in the pericarp (Figure 1). The display of multiple genes side by side is possible and the database can be searched either by gene ID or a BLAST interface.
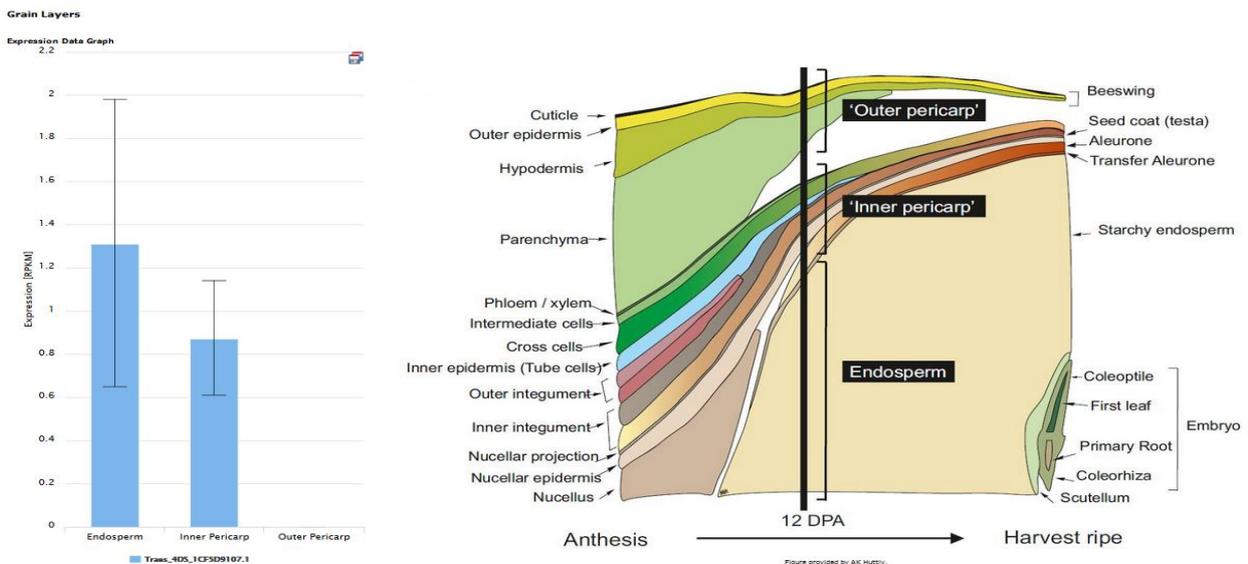

Figure 1: Expression of Traes_4DS_1CF5D9107.1 in the "grain layer" study on WheatExp.

WheatExp website and citation

WheatExp website: http://wheat.pw.usda.gov/WheatExp/

WheatExp reference: Pearce *et al.*, 2015. DOI: 10.1186/s12870-015-0692-1

b) expVIP

expVIP is a flexible RNA-seq analysis platform which can produce an interactive display of RNA-seq data for any species with a transcriptome. expVIP was used to generate a wheat expression browser (www.wheat-expression.com) which includes data from 16 studies including:

- different tissues over development
- grain development
- grain layers
- senescence
- abiotic stresses (phosphate starvation, drought and heat stress)
- disease infection (yellow rust, powdery mildew, septoria and fusarium)

expVIP displays expression of your gene(s) of interest in all 16 studies either as a bar graphs (for a single gene, Figure 2) or as heatmaps (for multiple genes, Figure 3). The heatmap display mode is particularly useful for determining the expression of candidate genes underlying a QTL. In the example shown in Figure 3, the expression patterns of six genes underlying a QTL for pre-harvest sprouting show that two of these candidate genes are expressed in the grain, as would be expected for a germination related trait. Therefore these two candidate genes could be prioritised for further study.
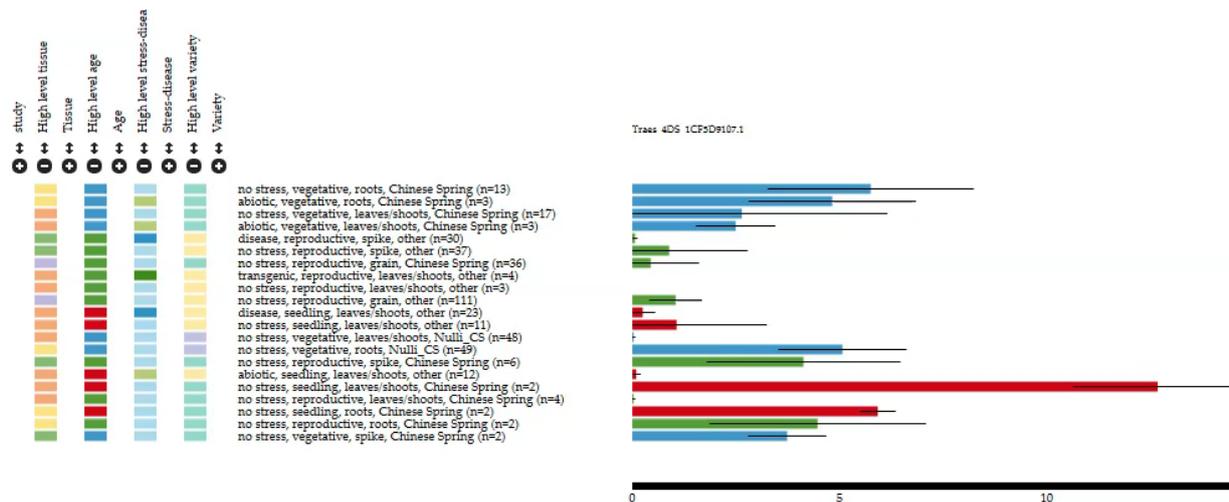
Figure 2: Expression of Traes_4DS_1CF5D9107.1 in 16 studies displayed by expVIP.
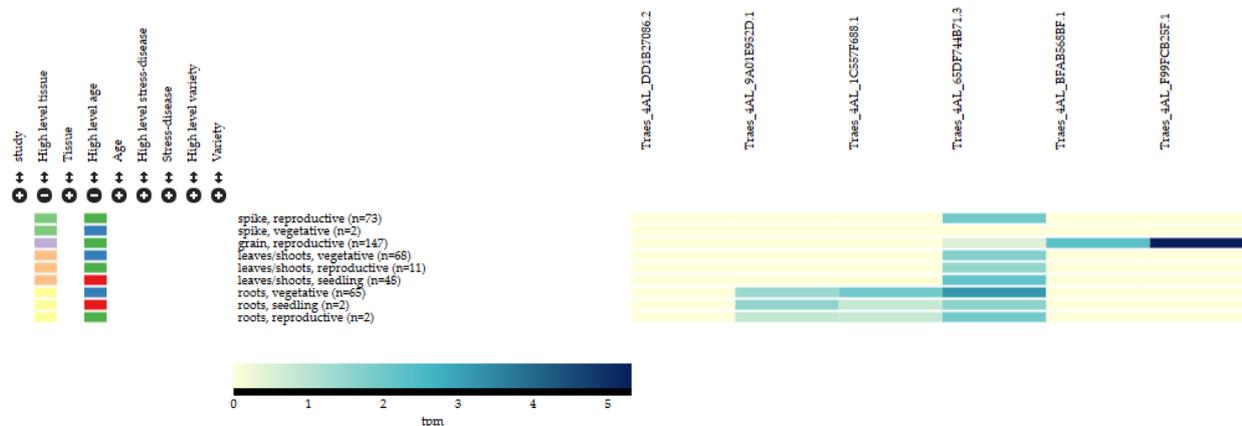

Figure 3: Expression of candidate genes underlying a pre-harvest sprouting QTL.

Metadata on expVIP

All samples within expVIP are categorised according to 4 high level categories: variety, tissue, developmental stage and stress which are divided into subcategories. For example a Chinese Spring starchy endosperm grain sample harvested 20 days post anthesis would be categorised as shown in the green text in Figure 4. Users can filter what data they want to see on the website dynamically, for example displaying only grain samples or only samples which had a disease infection.

Variety

High level	Variety
Other	CS Nullitetra
	Bobwhite
	Avalon
	Cadenza
	...
Chinese Spring	**Chinese Spring**

Tissue

High level	Tissue
Grain	**Endosperm**
	Transfer cells
	...
Shoots	Flag leaf
	2nd leaf
	...
Spike	Spikelets
	Stamen
	...
Roots	roots

Dev. stage

High level	Dev stage
Seedling	7 days
	14 days
	...
Vegetative	3 leaf stage
	5 leaf stage
	...
Reproductive	Anthesis
	5 dpa
	20 dpa
	...

Stress

High level	Stress
None	**None**
Disease	Septoria 4 dpi
	Septoria 10 dpi
	Mildew 24 h
	Yellow rust 3 dpi
	...
Abiotic	Heat (1 h)
	Drought (6h)
	P starvation (10 d)
	...

**Figure 4: Metadata allows categorisation of samples in expVIP.**

Virtual machine

expVIP is a highly flexible platform which requires 3 inputs: metadata (as described above), reads from RNA-seq experiments and a reference transcriptome (Figure 5). This allows expVIP to be used not just on wheat but any species with a reference transcriptome.

A user-friendly virtual machine which runs the entire expVIP platform on a desktop PC is available (https://github.com/homonecloco/expvip-web/wiki/LoadingVM) and has an extensive tutorial available at the same site, putting RNA-seq analysis and visualisation within the reach of non-bioinformaticians.

The virtual machine comes in two versions:

- with the existing wheat data pre-loaded so a user can add in their own wheat samples to compare
- empty so that RNA-seq data for any species can be used
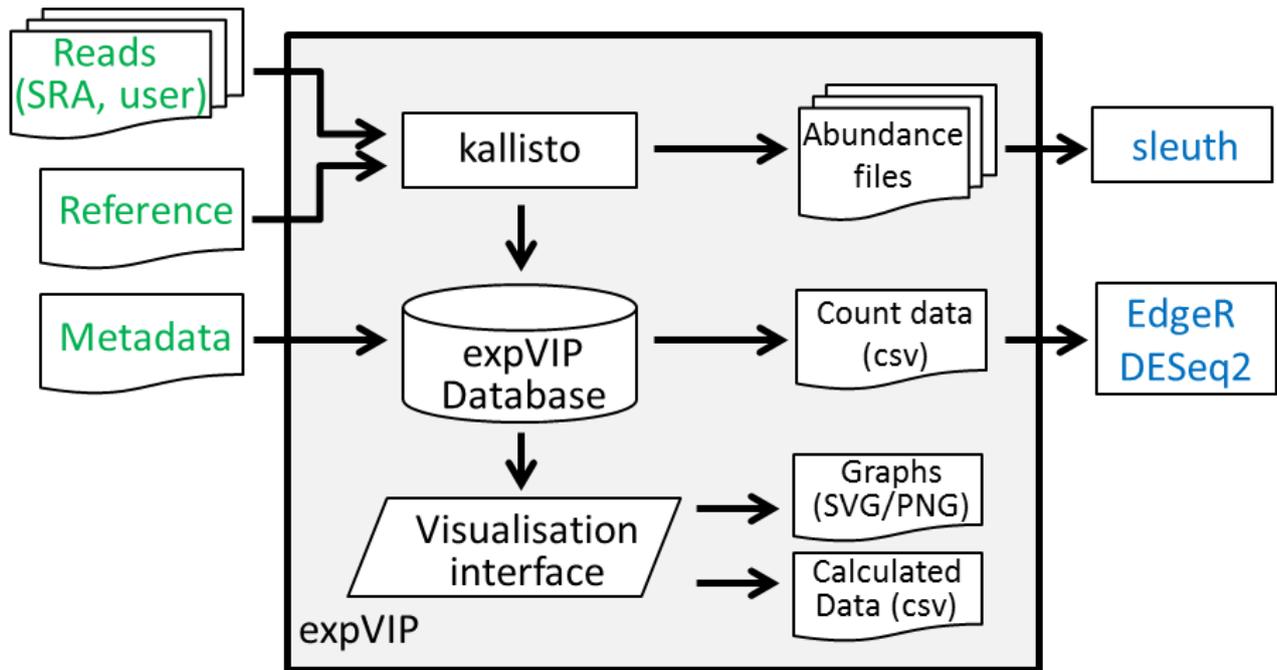
Figure 5: expVIP platform.

The virtual machine generates a visual display exactly the same as found at www.wheat-expression.com but including the users own samples, which are stored locally (user data is not made public). The expVIP platform also generates data which is ready to use in downstream differential gene expression analysis, for example using DESeq2, EdgeR or sleuth.
Graphs can be saved from the website display in both png and svg format, and all data can be extracted at the push of a button.

Tutorials

A tutorial explaining how to use expVIP's features is available both as a video:
https://www.youtube.com/watch?v=axDfuVn6GIk

And as a walk-through pdf:
https://github.com/homonecloco/expvip-web/wiki/Tutorial-expVIP-Graphical-Interface-(Wheat-Expression-Browser-example)

expVIP website and citation

expVIP website: www.wheat-expression.com
expVIP reference: Borrill et al., 2016. doi: http://dx.doi.org/10.1104/pp.15.01667