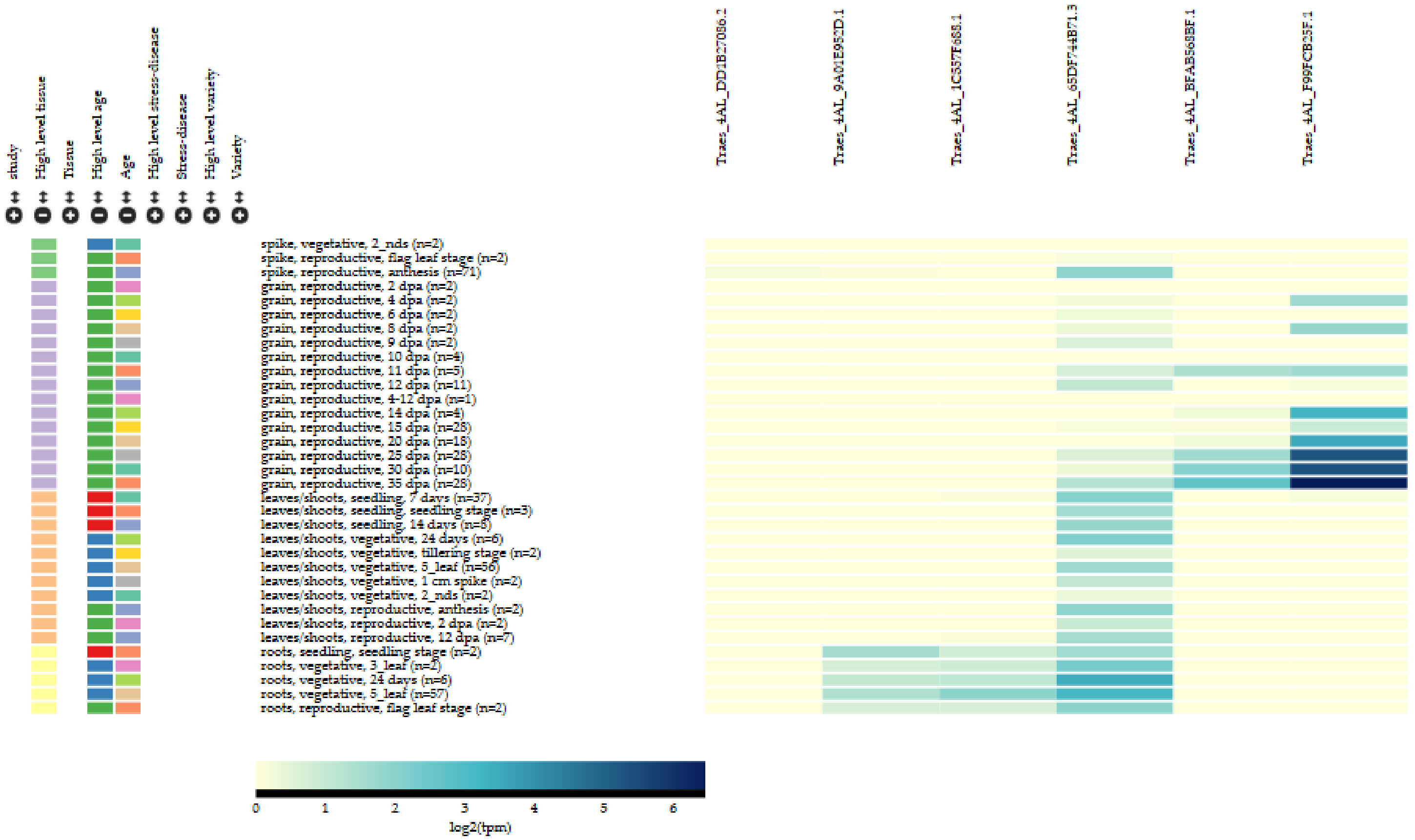
Wheat Gene Expression Browser

Implemented by the expVIP customisable RNA-Seq data analysis and visualisation platform

John Innes Centre

Cristobal Uauy (cristobal.uauy@jic.ac.uk)

@CristobalUauy



In the past 18 months...

Pearce et al 2015 BMC Plant Biol

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

Borrill et al *under review*

www.wheat-expression.com

IWGSC 2014 *Science*

Chapman et al. 2015 *Genome Biology*

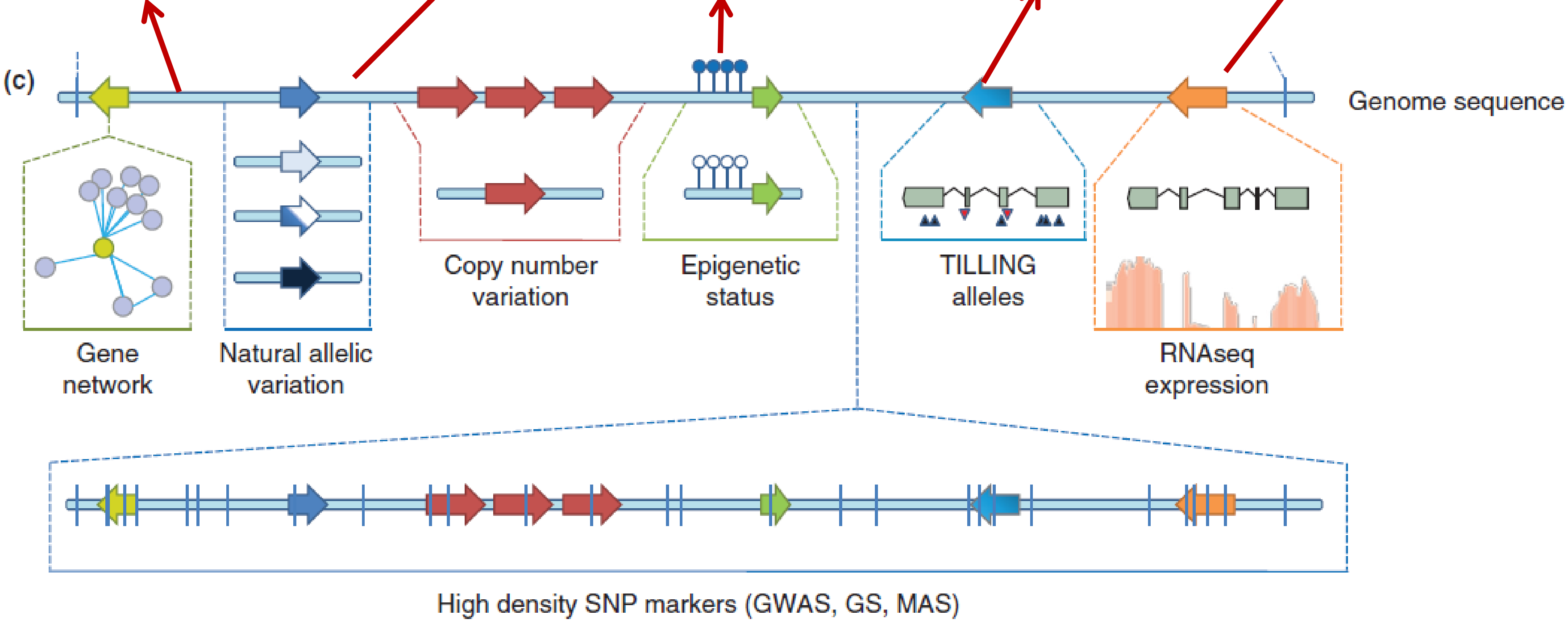
NRGene 2016

TGAC-UK 2016

HapMap: Jordan et al. 2015 *Genome Biology*

Gardiner et al. 2015 *Genome Biology*

in silico TILLING: Krasileva et al *in prep*



CerealsDB
Tools for the analysis of the wheat genome

polymarker

Ramirez-Gonzalez et al 2015 *Bioinformatics*

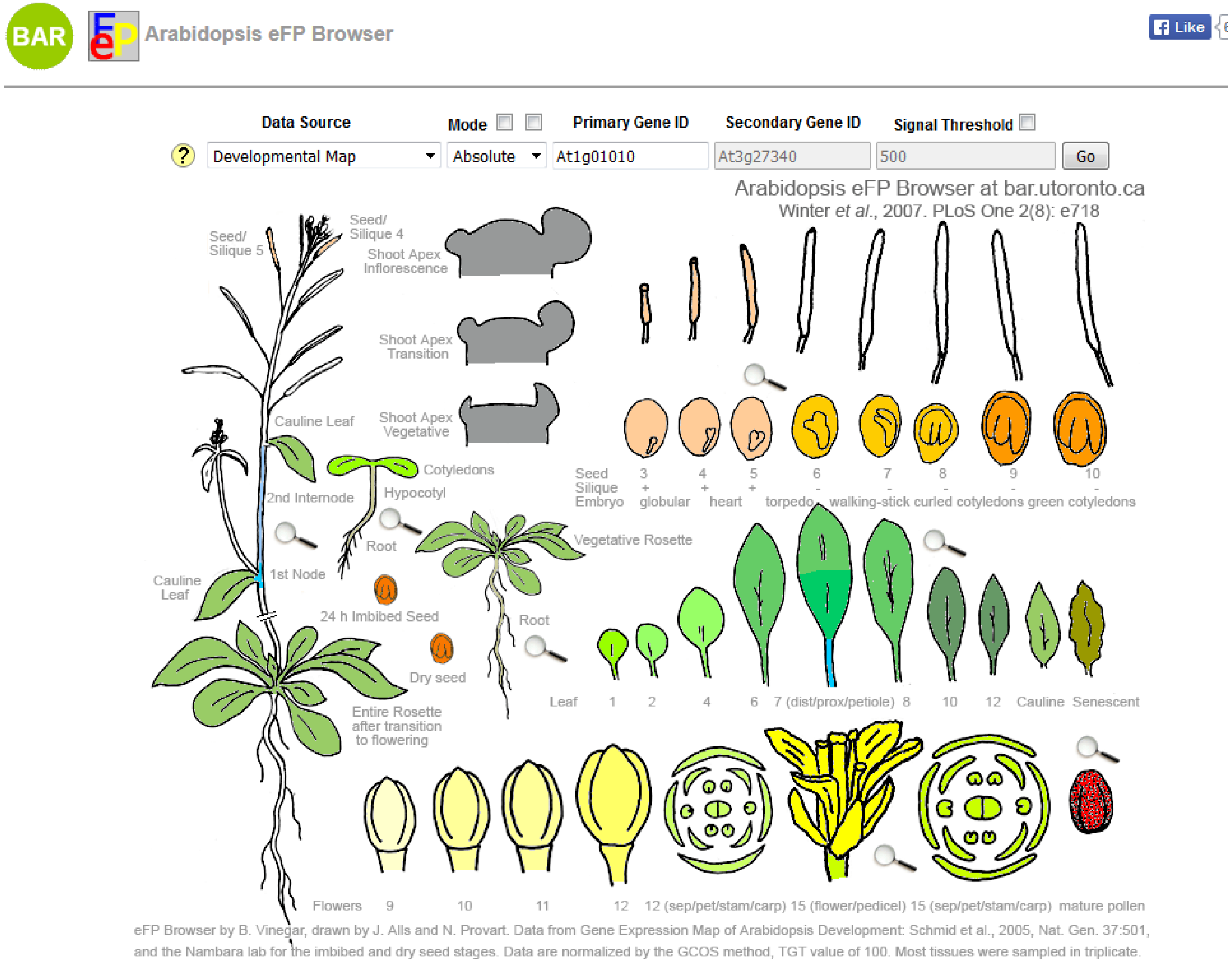
Uauy and Bevan 2014 *Genome Biology*

Over 10,000 RNA-seq samples from crops

Species (common name)	Samples in SRA database	Ploidy (Recent WGD)
Saccharum officinarum (sugarcane)	46	8x/10x
Zea mays (maize)	3,514	2x (WGD)
Oryza sativa (rice)	1,264	2x
***Triticum aestivum* (wheat)**	**799**	**6x**
Solanum tuberosum (potato)	337	4x
Manihot esculenta (cassava)	61	2x
Glycine max (soybean)	972	2x (WGD)
Beta vulgaris (sugar beet)	32	2x
Solanum lycopersicum (tomato)	830	2x
Hordeum vulgare (barley)	269	2x
Musa acuminata (banana)	73	2x/3x (WGD)
Sorghum bicolor (sorghum)	128	2x
Brassica ssp (mustard and oilseed rape)	835	2x/4x
Phaseolus vulgaris (common bean)	106	2x
Gossypium hirsutum (cotton)	468	4x
Vitis vinifera (grape)	448	2x

NCBI - Short Reads Archive: as of 5th August 2015

Where and when is my gene expressed?



Arabidopsis eFP Browser

Data Source: Developmental Map

Mode ☐ ☐: Absolute

Primary Gene ID: At1g01010

Secondary Gene ID: At3g27340

Signal Threshold ☐: 500

Go

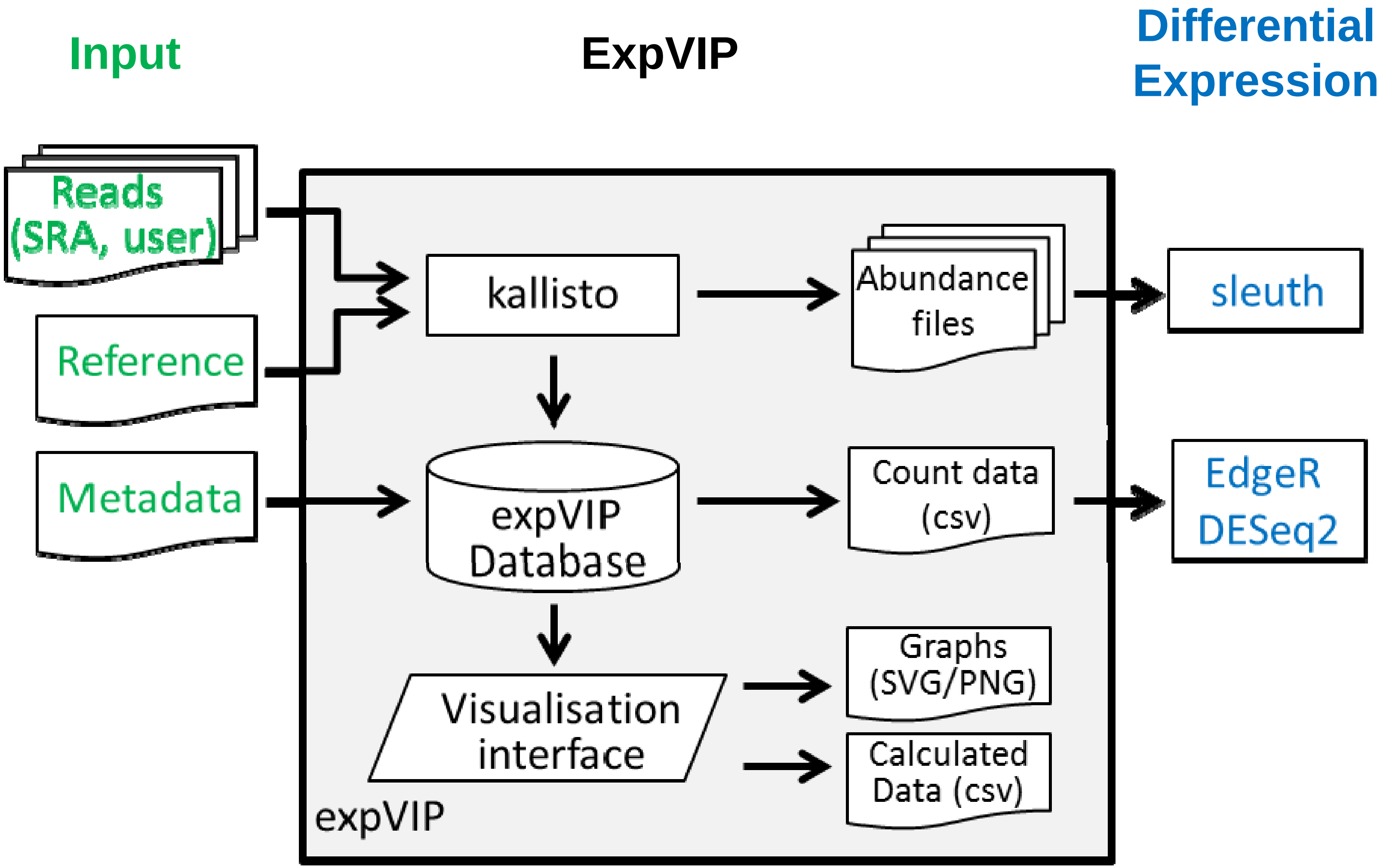
Arabidopsis eFP Browser at bar.utoronto.ca
Winter *et al*., 2007. PLoS One 2(8): e718

eFP Browser by B. Vinegar, drawn by J. Alls and N. Provart. Data from Gene Expression Map of Arabidopsis Development: Schmid et al., 2005, Nat. Gen. 37:501, and the Nambara lab for the imbibed and dry seed stages. Data are normalized by the GCOS method, TGT value of 100. Most tissues were sampled in triplicate.

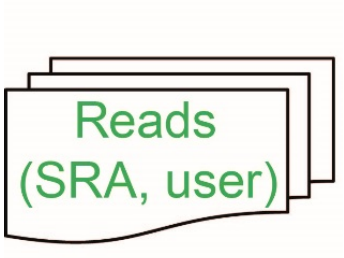
expVIP (expression Visualisation and Integration Platform)

- User-friendly
- Intuitive
- Dynamic/interactive interface (easy to ask questions)
- Flexible (more data can be added)
- Simple inputs

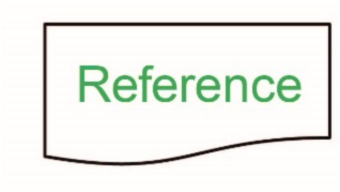
expVIP (expression Visualisation and Integration Platform)



Input files



- Fastq files from SRA (418 RNA-seq samples)
- Internal/personal files



- IWGSC v 2.26 transcriptome (103,274 genes)

(Choulet *et al.*, 2014; IWGSC, 2014)

- Information to classify samples
  (tissue, developmental stage, stress, variety)

Input SRA reads

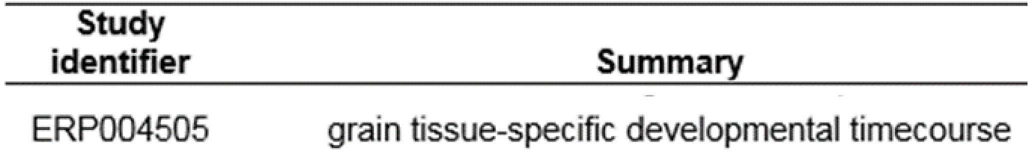
Study identifier	Summary	Total reads	Mapped reads (%)	Ref
DRP000768	phosphate starvation in roots and shoots	118,053,746	84,529,715 (72 %)	(Oono *et al.*, 2013b)
ERP003465	fusarium head blight infected spikelets	1,827,362,091	1,357,197,955 (74 %)	(Kugler *et al.*, 2013)
ERP004505	grain tissue-specific developmental timecourse	873,709,556	475,184,621 (54 %)	(Pfeifer *et al.*, 2014)
SRP004884	flag leaf downregulation of *GPC*	209,427,573	121,855,143 (58 %)	(Cantu *et al.*, 2011)
SRP013449	grain tissue-specific developmental timecourse	132,702,451	82,417,257 (62 %)	(Gillies *et al.*, 2012)
SRP017303	stripe rust infected seedlings	33,361,836	13,732,210 (41 %)	(Cantu *et al.*, 2013)
SRP022869	*Septoria tritici* infected seedlings	100,582,632	63,155,877 (63 %)	(Yang *et al.*, 2013)
SRP028357	shoots and leaves of nulli tetra group 1 and group 5	3,304,500,117	2,258,692,000 (68 %)	(Leach *et al.*, 2014)
SRP029372	grain tissue-specific developmental timecourse	101,477,759	17,525,439 (17 %)	(Li *et al.*, 2013)
SRP038912	comparison of stamen, pistil and pistilloidy expression	217,315,378	153,009,134 (70 %)	(Yang *et al.*, 2015)
SRP041017	stripe rust and powdery mildew infection timecourse	395,463,786	272,228,560 (69 %)	(Zhang *et al.*, 2014)
SRP041022	developmental time-course of synthetic hexaploid	134,641,113	84,583,556 (63 %)	(Li *et al.*, 2014)
ERP008767	grain tissue-specific expression at 12 days post anthesis	45,213,827	26,420,708 (58 %)	(Pearce *et al.*, 2015)
SRP045409	drought and heat stress time-course in seedlings	921,578,806	533,928,182 (58 %)	(Liu *et al.*, 2015)
ERP004714	developmental time-course of Chinese Spring	1,536,051,415	1,066,712,760 (69 %)	(Choulet *et al.*, 2014)
SRP056412	grain developmental timecourse with 4A dormancy QTL	1,875,916,011	808,809,053 (43 %)	(Barrero *et al.*, 2015)

16 studies; 418 RNA-seq samples; 11 billion reads

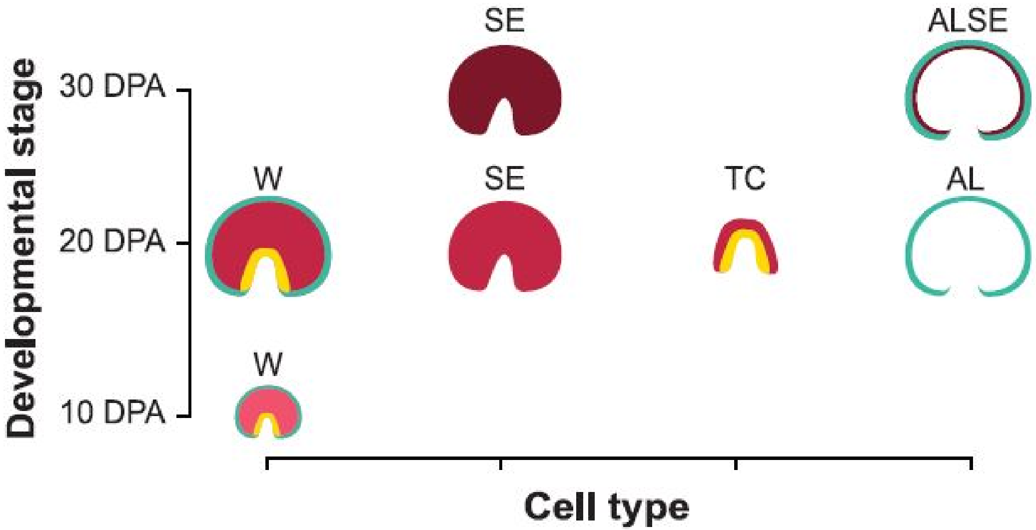
Metadata: tissue, developmental stage, stress, variety

Study identifier	Summary	Total reads	Mapped reads (%)	Ref
DRP000768	phosphate starvation in roots and shoots	118,053,746	84,529,715 (72 %)	(Oono *et al.*, 2013b)
ERP003465	fusarium head blight infected spikelets	1,827,362,091	1,357,197,955 (74 %)	(Kugler *et al.*, 2013)
ERP004505	grain tissue-specific developmental timecourse	873,709,556	475,184,621 (54 %)	(Pfeifer *et al.*, 2014)
SRP004884	flag leaf downregulation of *GPC*	209,427,573	121,855,143 (58 %)	(Cantu *et al.*, 2011)
SRP013449	grain tissue-specific developmental timecourse	132,702,451	82,417,257 (62 %)	(Gillies *et al.*, 2012)
SRP017303	stripe rust infected seedlings	33,361,836	13,732,210 (41 %)	(Cantu *et al.*, 2013)
SRP022869	*Septoria tritici* infected seedlings	100,582,632	63,155,877 (63 %)	(Yang *et al.*, 2013)
SRP028357	shoots and leaves of nulli tetra group 1 and group 5	3,304,500,117	2,258,692,000 (68 %)	(Leach *et al.*, 2014)
SRP029372	grain tissue-specific developmental timecourse	101,477,759	17,525,439 (17 %)	(Li *et al.*, 2013)
SRP038912	comparison of stamen, pistil and pistilloidy expression	217,315,378	153,009,134 (70 %)	(Yang *et al.*, 2015)
SRP041017	stripe rust and powdery mildew infection timecourse	395,463,786	272,228,560 (69 %)	(Zhang *et al.*, 2014)
SRP041022	developmental time-course of synthetic hexaploid	134,641,113	84,583,556 (63 %)	(Li *et al.*, 2014)
ERP008767	grain tissue-specific expression at 12 days post anthesis	45,213,827	26,420,708 (58 %)	(Pearce *et al.*, 2015)
SRP045409	drought and heat stress time-course in seedlings	921,578,806	533,928,182 (58 %)	(Liu *et al.*, 2015)
ERP004714	developmental time-course of Chinese Spring	1,536,051,415	1,066,712,760 (69 %)	(Choulet *et al.*, 2014)
SRP056412	grain developmental timecourse with 4A dormancy QTL	1,875,916,011	808,809,053 (43 %)	(Barrero *et al.*, 2015)

Metadata



Study identifier	Summary
ERP004505	grain tissue-specific developmental timecourse



Tissue:	Grain	aleurone, endosperm, etc
Dev. Stage	Reproductive	10, 20, 30 dpa
Stress	None	none
Variety	CS	CS

Pfeifer *et al* 2014 Science

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
	10 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)
	…

Tissue

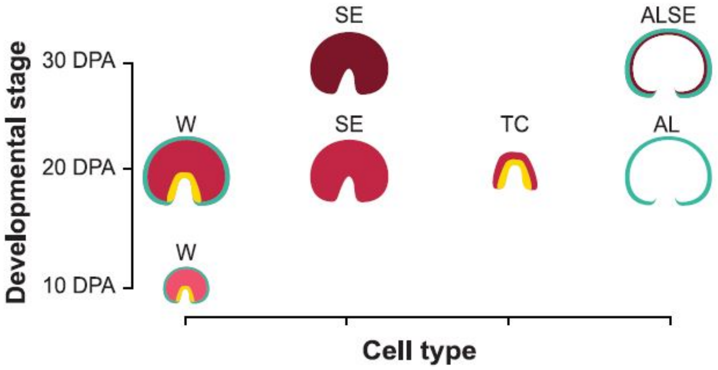
High level	tissue
grain	**Endosperm**
	Transfer cells
	…
shoots	Flag leaf
	2nd leaf
	…
spike	Spikelets
	Stamen
	…
roots	roots

Dev. stage

High level	Dev stage
seedling	7 days
	14 days
	…
Vegetative	3 leaf stage
	5 leaf stage
	…
Reproductive	Anthesis
	5 dpa
	10 dpa
	…

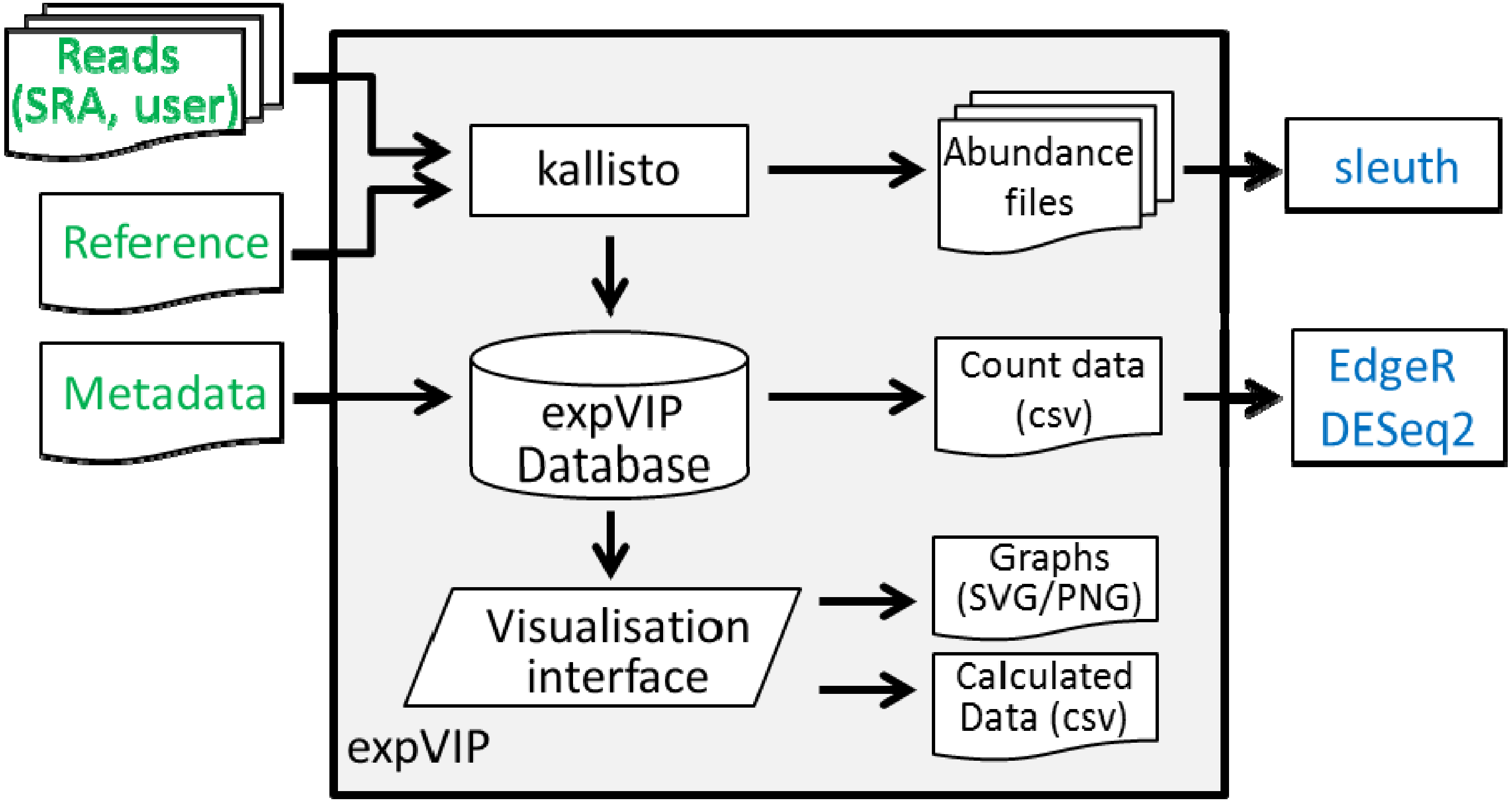
Stress

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)
	…



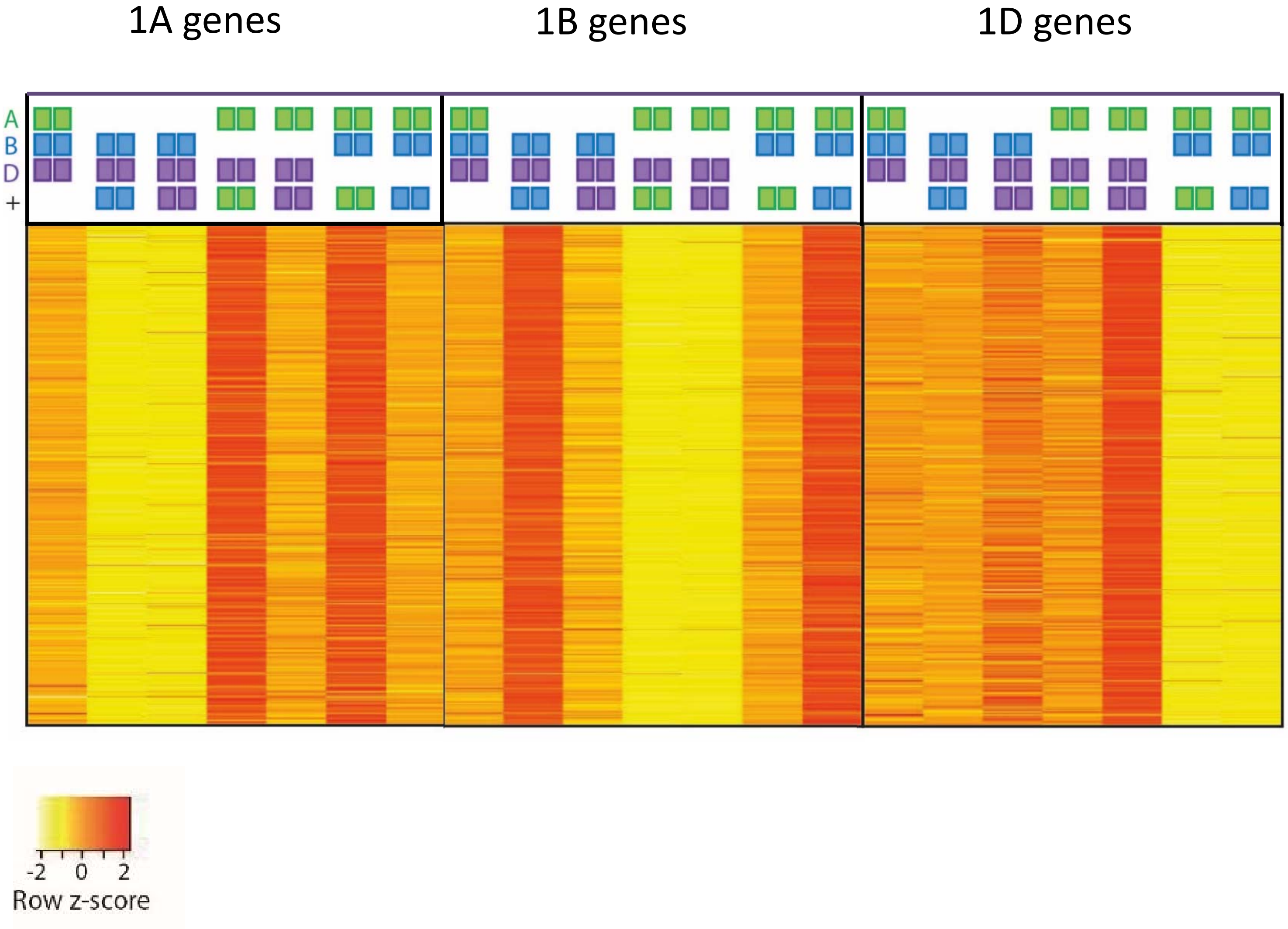
kallisto

Bray *et al* 2015 arXiv



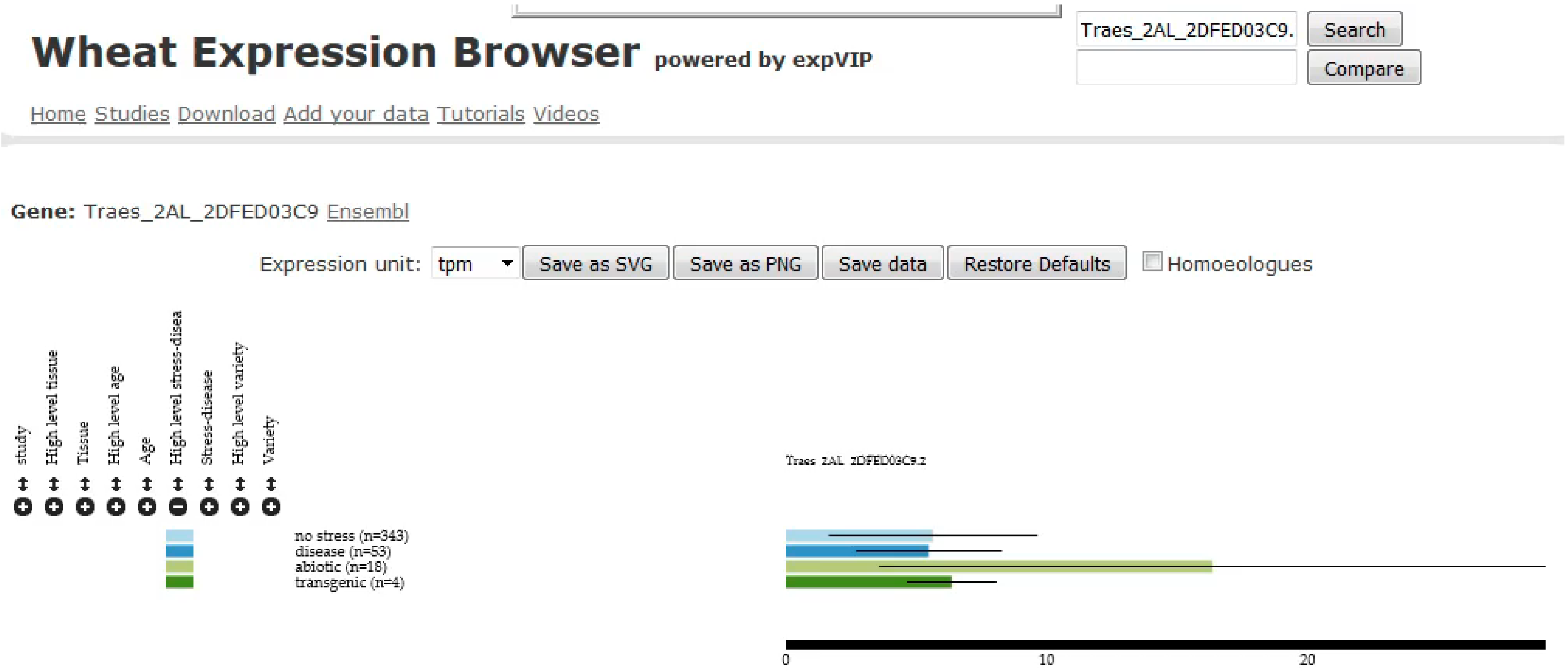
kallisto: accurate read mapping in polyploid wheat

Chromosome 1 nullitetrasomic lines; root RNA-seq samples



NCBI: SRP028357 from Leach et al., 2014 *BMC Genomics* 2014, 15:276

Visual interface: www.wheat-expression.com



Wheat Expression Browser powered by expVIP

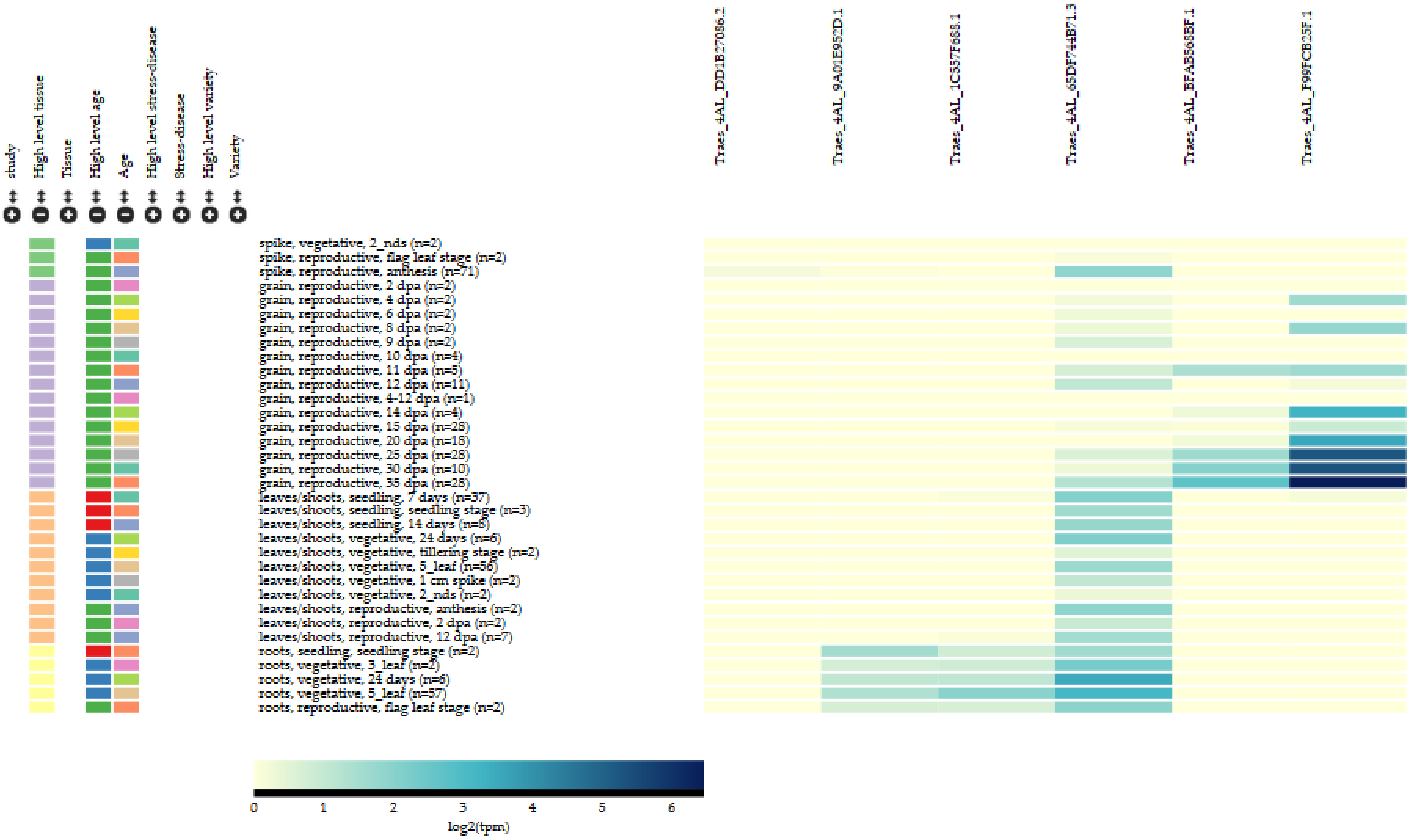
Gene: Traes_2AL_2DFED03C9 Ensembl

Expression unit: tpm

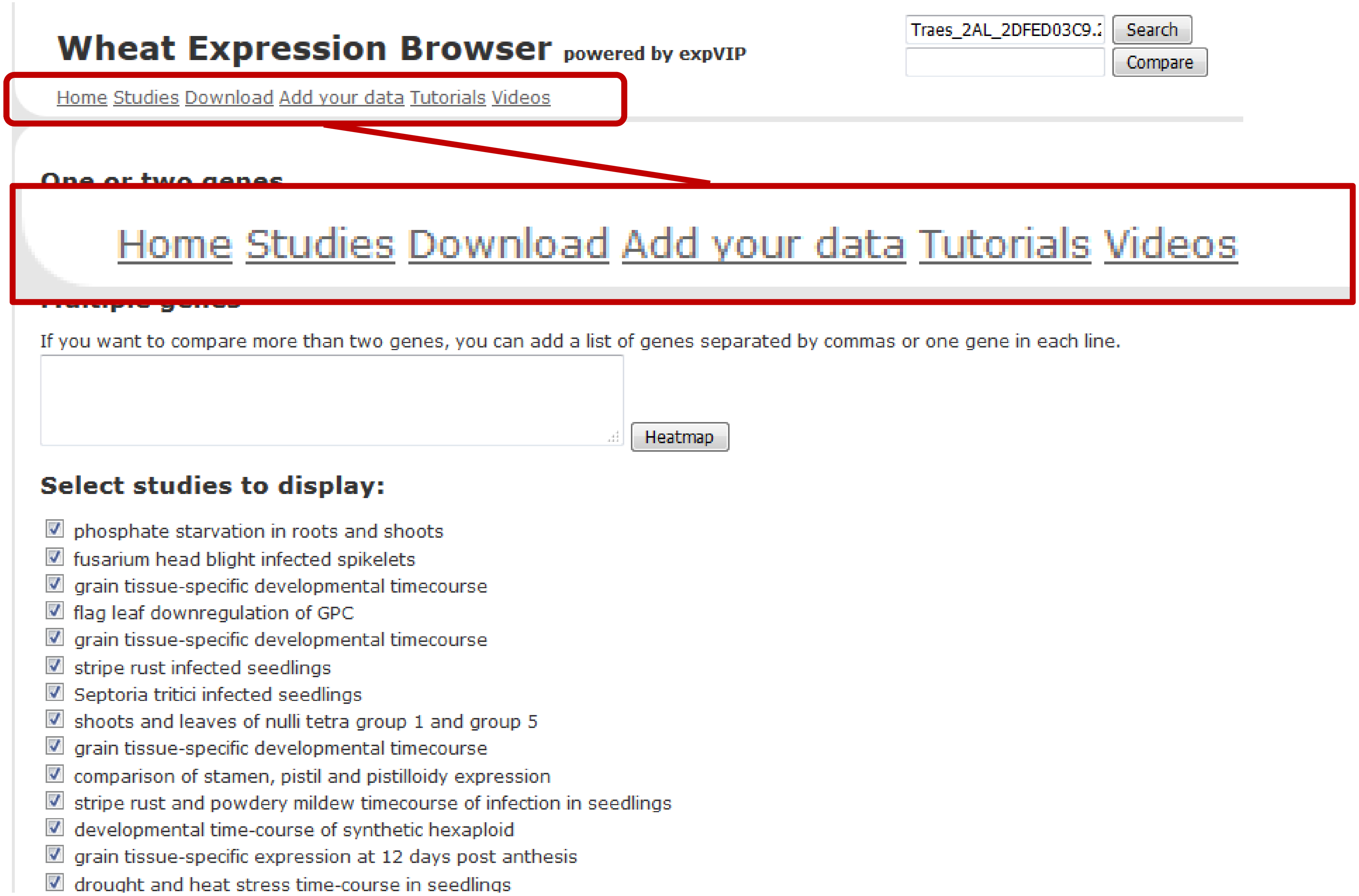
☐ Homoeologues

no stress (n=343)
disease (n=53)
abiotic (n=18)
transgenic (n=4)

Visual interface: heatmap



www.wheat-expression.com



Wheat Expression Browser powered by expVIP

One or two genes

Multiple genes

If you want to compare more than two genes, you can add a list of genes separated by commas or one gene in each line.

Select studies to display:

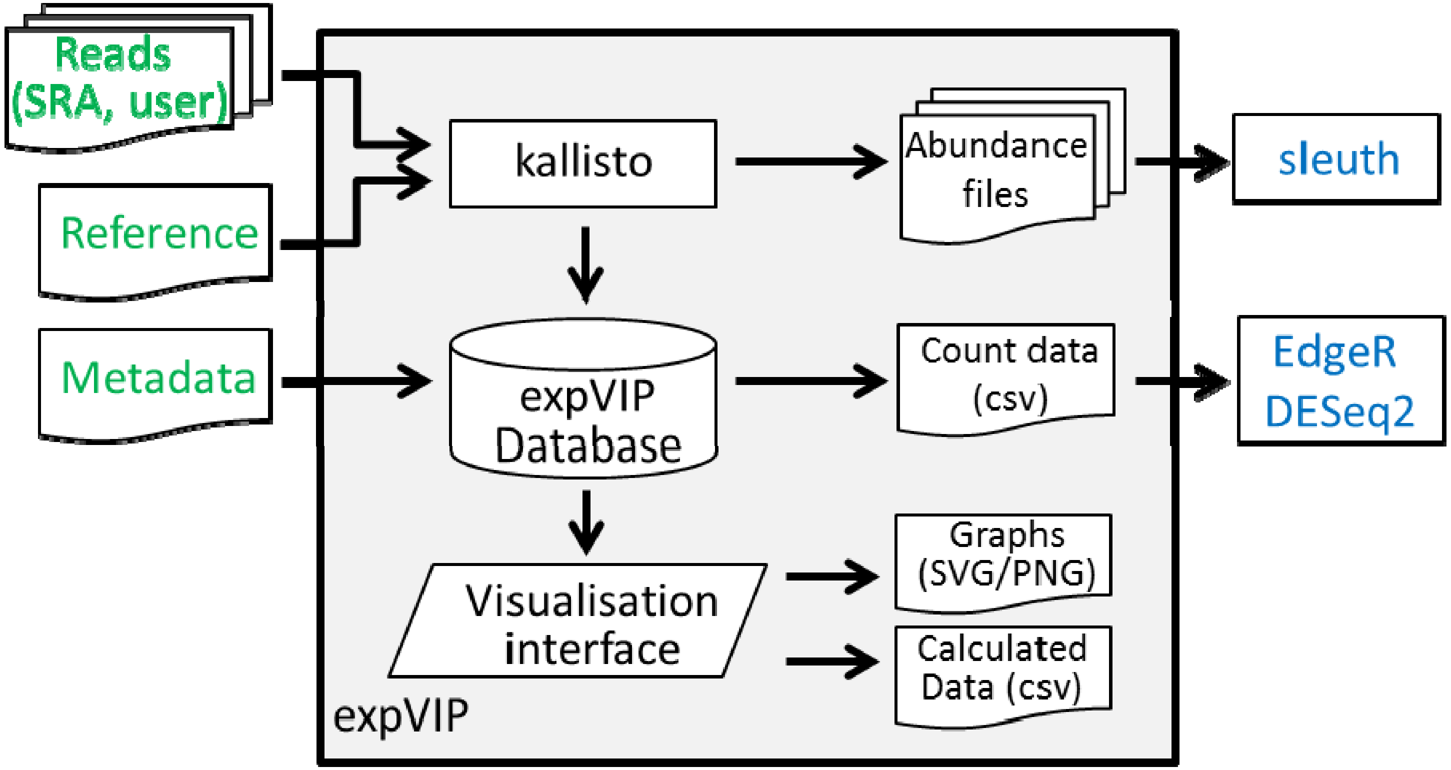
- ☑ phosphate starvation in roots and shoots
- ☑ fusarium head blight infected spikelets
- ☑ grain tissue-specific developmental timecourse
- ☑ flag leaf downregulation of GPC
- ☑ grain tissue-specific developmental timecourse
- ☑ stripe rust infected seedlings
- ☑ Septoria tritici infected seedlings
- ☑ shoots and leaves of nulli tetra group 1 and group 5
- ☑ grain tissue-specific developmental timecourse
- ☑ comparison of stamen, pistil and pistilloidy expression
- ☑ stripe rust and powdery mildew timecourse of infection in seedlings
- ☑ developmental time-course of synthetic hexaploid
- ☑ grain tissue-specific expression at 12 days post anthesis
- ☑ drought and heat stress time-course in seedlings

Home Studies Download Add your data Tutorials Videos

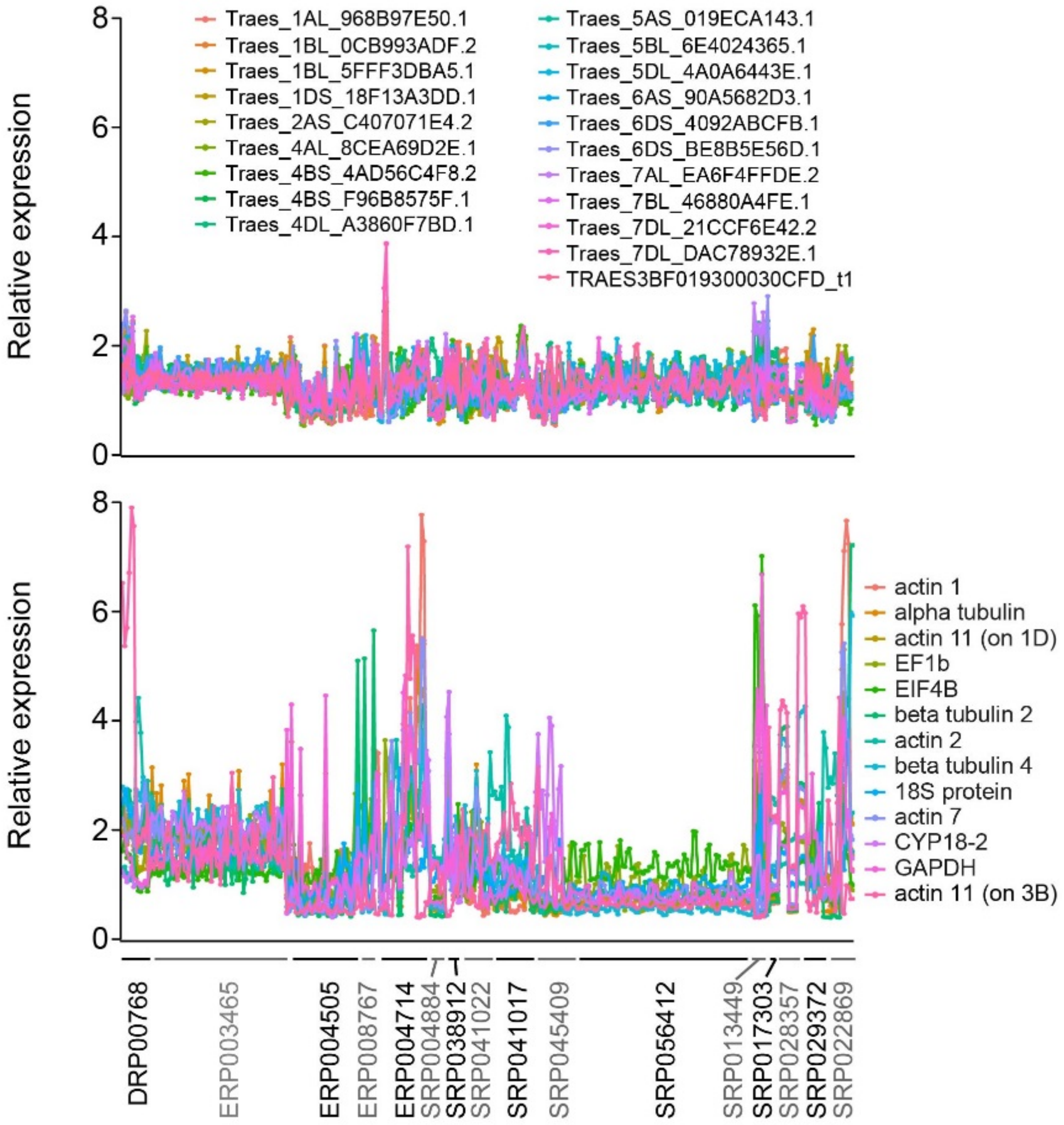
Download: All expression value, metadata, etc

Abundance files (differential expression Sleuth)

Counts (differential expression DeSeq, EdgeR)

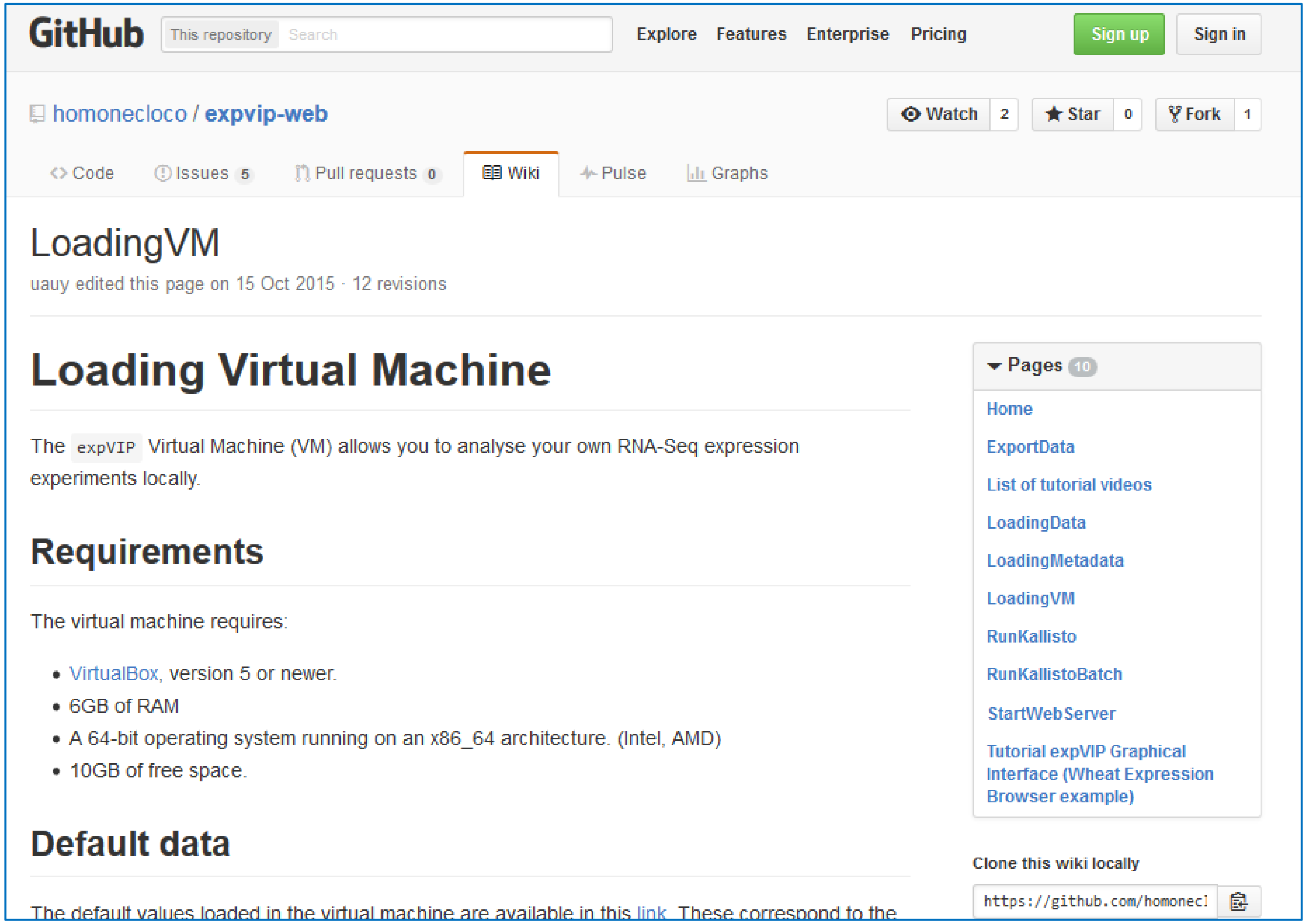


Reference genes for qRT-PCR



Add your data: download Virtual Machine

- With all wheat data
- Empty (any transcriptome)



LoadingVM

uauy edited this page on 15 Oct 2015 · 12 revisions

Loading Virtual Machine

The `expVIP` Virtual Machine (VM) allows you to analyse your own RNA-Seq expression experiments locally.

Requirements

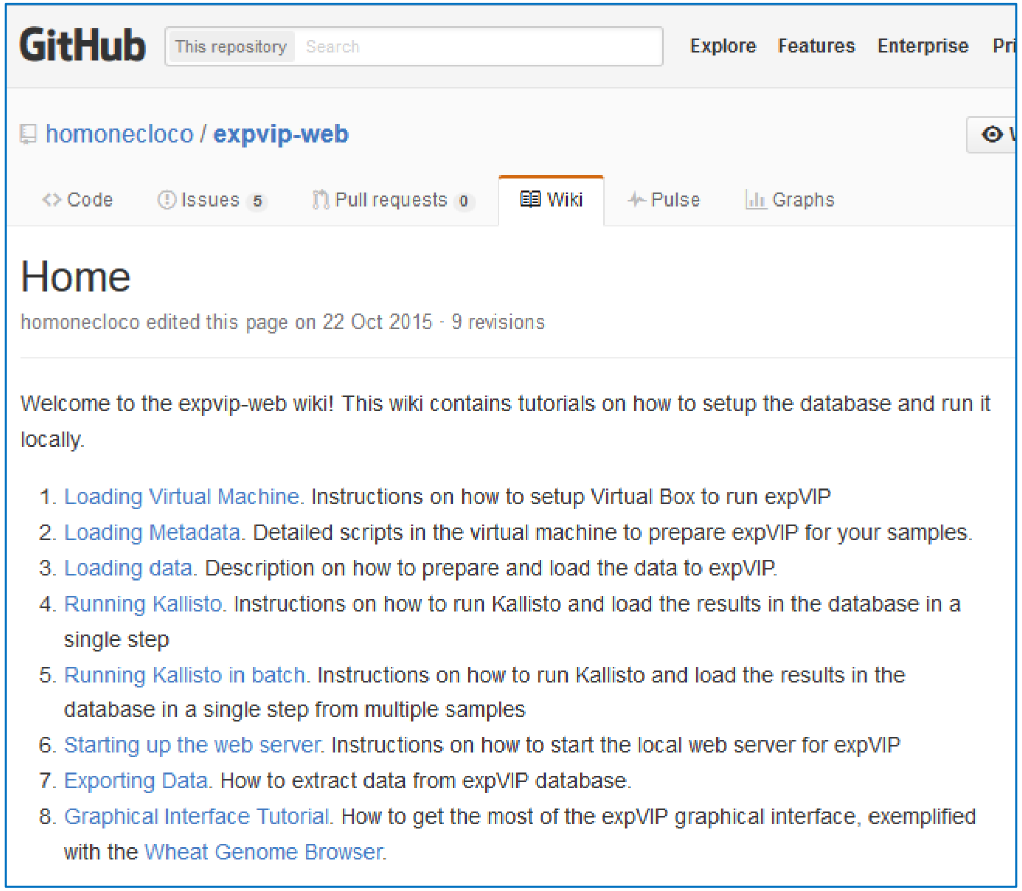
The virtual machine requires:

- VirtualBox, version 5 or newer.
- 6GB of RAM
- A 64-bit operating system running on an x86_64 architecture. (Intel, AMD)
- 10GB of free space.

Default data

The default values loaded in the virtual machine are available in this link. These correspond to the

Home Studies Download Add your data Tutorials Videos



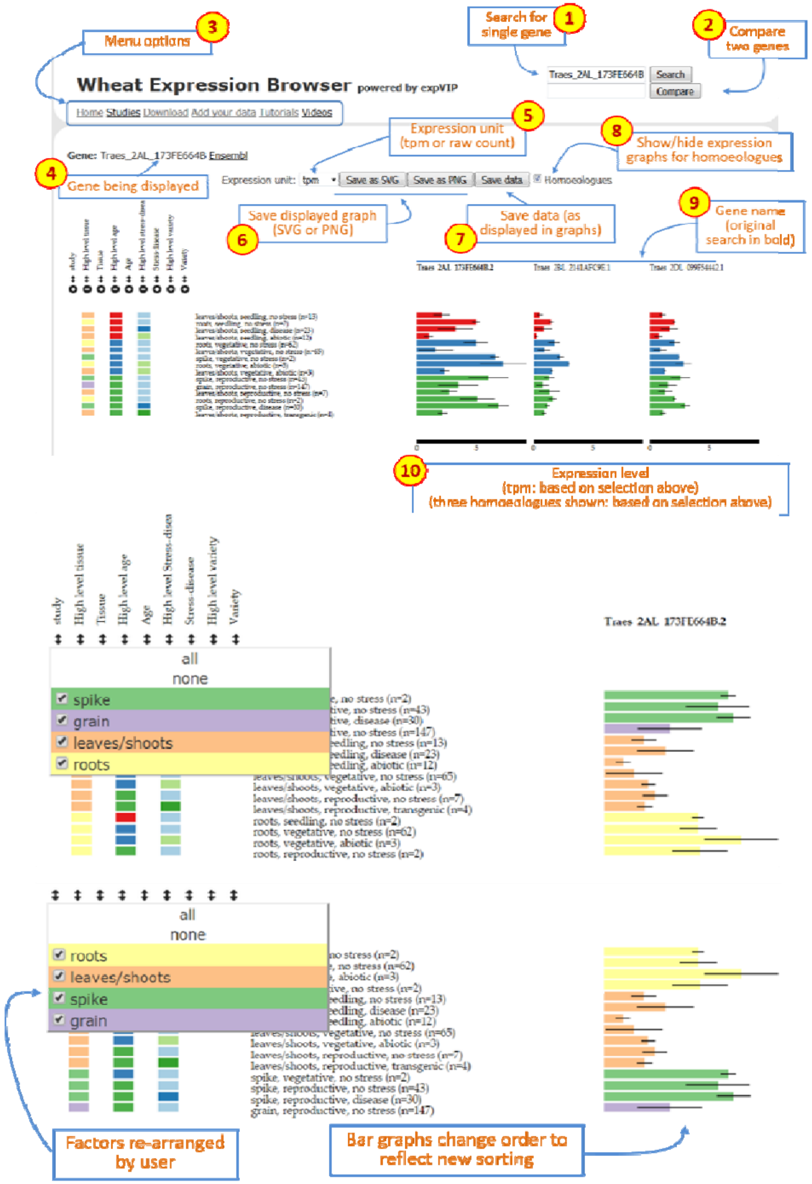
homonecloco / **expvip-web**

Home

homonecloco edited this page on 22 Oct 2015 · 9 revisions

Welcome to the expvip-web wiki! This wiki contains tutorials on how to setup the database and run it locally.

1. Loading Virtual Machine. Instructions on how to setup Virtual Box to run expVIP
2. Loading Metadata. Detailed scripts in the virtual machine to prepare expVIP for your samples.
3. Loading data. Description on how to prepare and load the data to expVIP.
4. Running Kallisto. Instructions on how to run Kallisto and load the results in the database in a single step
5. Running Kallisto in batch. Instructions on how to run Kallisto and load the results in the database in a single step from multiple samples
6. Starting up the web server. Instructions on how to start the local web server for expVIP
7. Exporting Data. How to extract data from expVIP database.
8. Graphical Interface Tutorial. How to get the most of the expVIP graphical interface, exemplified with the Wheat Genome Browser.



expVIP (expression Visualisation and Integration Platform)

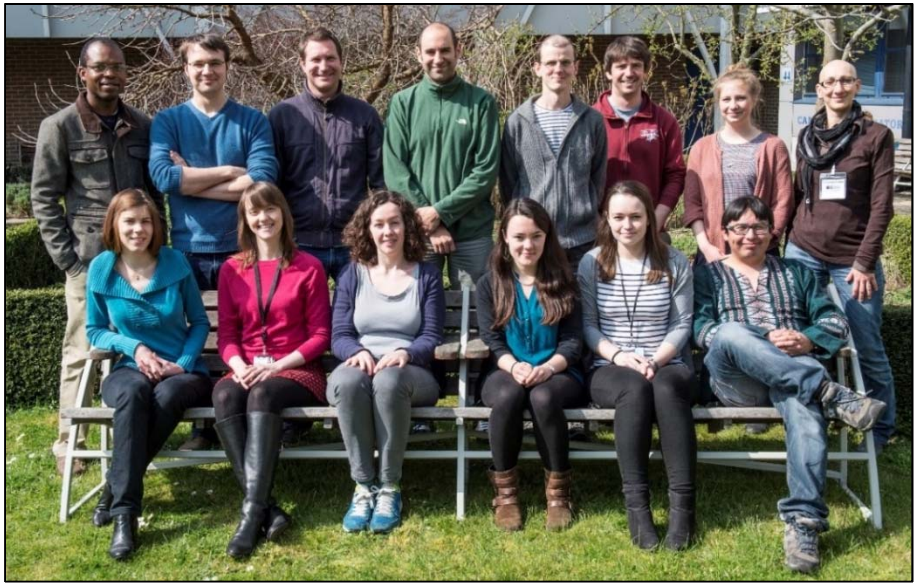
- User-friendly
- Intuitive
- Dynamic/interactive interface (easy to ask questions)
- Flexible (more data can be added)
- Simple inputs

Community feedback very welcomed!

John Innes Centre TGAC The Genome Analysis Centre™

www.wheat-expression.com

Philippa Borrill
Ricardo Ramirez-Gonzalez
Rob Davey
NRP Computing



Wheat (and human) diversity

Source code:
Database and interface setup:
https://github.com/homonecloco/expvip-web

BioJS visualisation component:
http://biojs.io/d/bio-vis-expression-bar

@CristobalUauy

IWYP International Wheat Yield Partnership
Research to Deliver Wheat for the Future

BBSRC bioscience for the future

HGCA