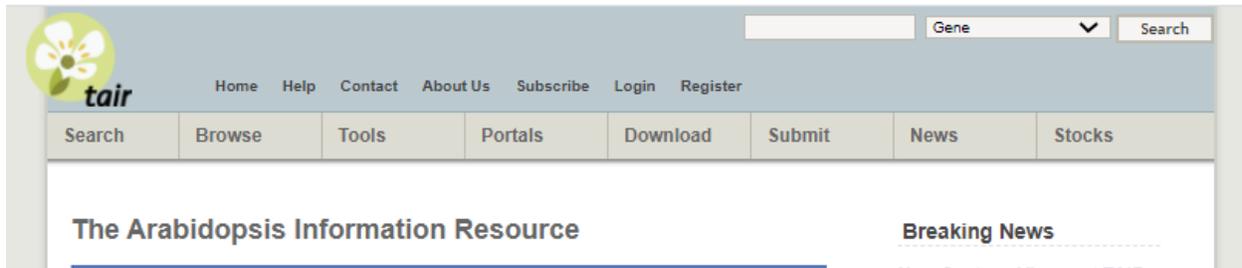


Part 3. Gene Ontology (GO) Biological Process

1. Copy the first column with the AGI#'s into a new Excel sheet. Do not copy the column header. Save the file as a comma delimited file (CSV).
2. Go to <https://www.arabidopsis.org/>. Click Search and select Gene Ontology annotations from the drop down menu.

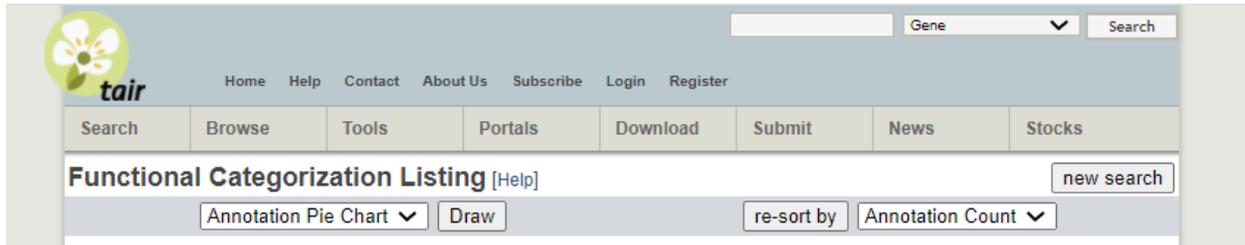


3. Click Choose file. Select your CSV file. Click Functional Categorization.

The image displays the 'Whole Genome Categorization' tool interface. It is organized into three main sections. The first section, 'Whole Genome Categorization', has a 'Get:' label and a button labeled 'Whole Genome Categorization'. The second section, 'GO Annotations', includes a 'Locus Identifiers:' label and a large empty text box. Below this is an 'Upload file:' label and a 'Choose File' button, with the text 'No file chosen' next to it. The third section, 'Output Options:', has a 'Select output type:' label and two radio buttons: 'HTML' (which is selected) and 'Text'. A note below the radio buttons states: '(Please note that if more than 1000 loci are entered, only text output will be given)'. At the bottom of the interface are three buttons: 'Reset', 'Get all GO Annotations', and 'Functional Categorization'.

4. Click Draw next to "Annotation Pie Chart". This will generate 3 pie charts: GO Cellular Component, GO Biological Process, and GO Molecular Function. You will include the GO Biological Process chart in

your paper. Copy and paste that into your Word file for your paper. When you write your paper, you should discuss anything that stands out to you as particularly interesting given your chosen topic. You do not need to discuss every single category of information appearing in these charts. You may include the other two charts in your paper if there is something in particular that you wish to highlight or tie into your discussion section of the paper but you are not required to do so.

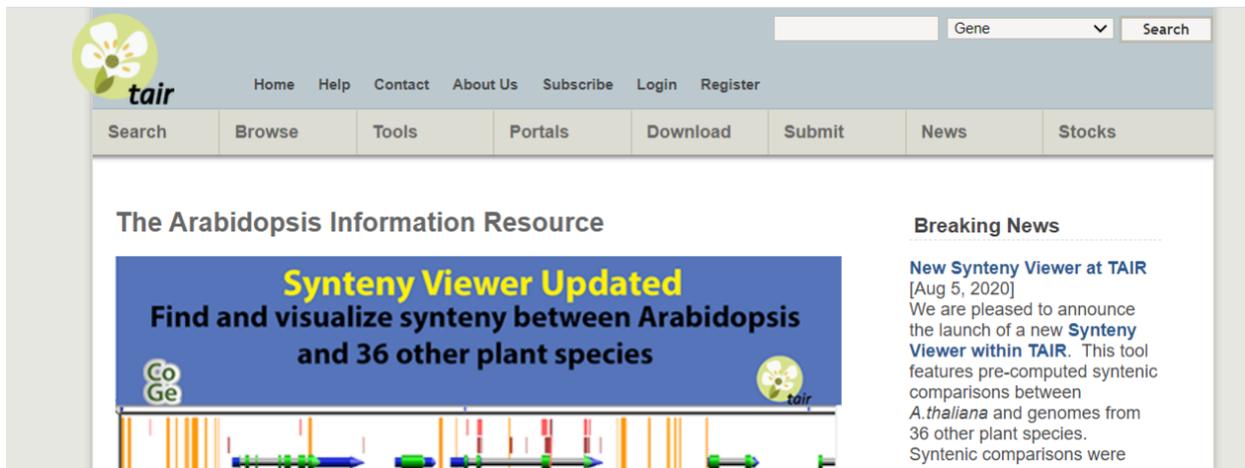


Part 4.

Selecting a gene of interest for detailed study.

Information is continuously being added to our knowledge base. Many genes have been identified since the data in this particular data set was first collected. If you want to see if more information is available for a particular gene that has a particularly striking fold change, you can check TAIR, the Arabidopsis Information Resource at <https://www.arabidopsis.org/>.

Click Search:



Click Microarray Element from the dropdown box. Enter your locus identifier in the box (example: At5g01810). Make sure Affymetrix ATH1 is selected (this is the type of chip our data set is from) and click "Get Microarray Elements".

To get detailed information on a gene of interest.



Home Help Contact About Us Subscribe Login Register

Search Browse Tools Portals Download Submit News Stocks

Home > Tools > Bulk Data Retrieval > Microarray Elements

Microarray Elements Search and Download [\[Help\]](#)

This tool allows you to find information about the microarray elements (Arabidopsis Genome ATH1 array, CATMA GSTs, and Agilent oligos) contained on the ATH1, CATMA and Agilent arrays. This includes their mapping to Arabidopsis locus identifies. This data was obtained from ENSEMBL Gramene (<http://ensembl.gramene.org>). The complete data files can be downloaded from [TAIR Microarray Downloads](#).

Paste locus identifiers (e.g., At5g01810), or array element names (e.g., 251059_at, A_84_P12978, CATMA1a00010, A_84_P10014) in the textbox below and press the submit button. Separate identifiers by tabs, commas or carriage returns. Alternatively, a file with a list of identifiers may also be uploaded. Choose the output type text if you want to save the results into your local computer.

Get Microarray Elements

Locus identifiers:

Upload file: No file chosen

Search Against: Affymetrix ATH1 CATMA Agilent (G2519F-021169)

Output Options:

Select output type: HTML (1316 genes max) Text

In this example, information about the gene can be found under the annotation.



Home Help Contact About Us Subscribe Login Register

Search Browse Tools Portals Download Submit News Stocks

Microarray Elements Search Results [\[Help\]](#)

Array Element	Locus Identifier	Annotation	Organism	Probe Type	Is Control
251059_at	AT5G01810	Encodes a CBL-interacting serine/threonine protein kinase, also has similarities to SOS2 kinase.	Arabidopsis thaliana	oligonucleotide	no

Genome Mapping Date: 04/29/2020, Araport11 release

You will want to select a gene that from your dataset that is strongly up or down-regulated (a fold change of 3 is preferred but you may go as low as 1.5-fold if necessary for the purpose of this assignment). You need to select a gene that has been studied in the past. Skip ones that are listed as unknown function in both our data set and when you look it up in the search above.

Next, click the search box in the top left corner again and this time select "Genes". Enter your locus ID (example At5g01810) in the "starts with" box under the Search by Name or Phenotype section. Scroll to the bottom of the page and hit "Submit Query". Select your locus from the list by clicking on the blue locus identifier.

The screenshot shows the TAIR Gene Search Results page. At the top left is the TAIR logo. A navigation bar contains links for Home, Help, Contact, About Us, Subscribe, Login, and Register. Below this is a menu with Search, Browse, Tools, Portals, Download, Submit, News, and Stocks. The main content area is titled "TAIR Gene Search Results" and includes buttons for "new search", "download all", "download checked", "get all sequences", and "get checked sequences". A text block explains the search query: "Your query for genes where gene name starts with the term At5g01810, taxon name is Arabidopsis thaliana, keyword term is any and keyword types of null resulted in 1 locus match with 3 distinct gene model associated to the keyword or keyword children terms." It also states "Displaying 1 - 1." and "To see ESTs associated with your gene of interest, click on the Locus link." Below this are "Check All" and "Uncheck All" buttons. A table with two columns, "Locus" and "Description", lists one result: locus AT5G01810, which is checked with a blue box. The description for AT5G01810 is "Other names: CBL-INTERACTING PROTEIN KINASE 15, CIPK15. Encodes a CBL-interacting serine/threonine protein kinase, also has similarities to SO..."

	Locus	Description
<input checked="" type="checkbox"/>	1 AT5G01810	Other names: CBL-INTERACTING PROTEIN KINASE 15, CIPK15 Encodes a CBL-interacting serine/threonine protein kinase, also has similarities to SO...

TAIR Gene Search

Genes may be searched by name, keywords, features, and/or location. In TAIR, a Gene Model is defined as any description of a gene product from a variety of sources including computational prediction, mRNA sequencing, or genetic characterization. A locus is defined as the genomic sequence corresponding to a transcribed unit (e.g. AT2G03340) in the genome. In TAIR, many gene models can exist for a given locus, therefore a search for a gene may result in multiple hits for the same gene name.

Search by Name or Phenotype

Species:

Gene name starts with

(leaving the input box blank will retrieve all entries)

Include obsoleted genes

Search by Associated Keyword

Keyword Term starts with

GO/PO ID (exact match only)

Keyword Type

- Any
- GO Molecular Function
- GO Biological Process
- GO Cellular Component

Evidence

- Any
- inferred from experiment
- inferred from direct assay
- inferred from electronic annotation

Restrict by Features

Gene Model Type

- Any
- pre tma
- transposable element gene
- protein coding

Advanced

- gene structure predicted
- has associated literature
- is sequenced
- is not sequenced

Time Restriction only search last 2 months

Restrict by Map Locations

Chromosome

Map Type

Range and

Output Options

Number of records per page

Sort records by

If the gene has been previously studied, a wealth of information will be available on the next page. Information to include in your paper:

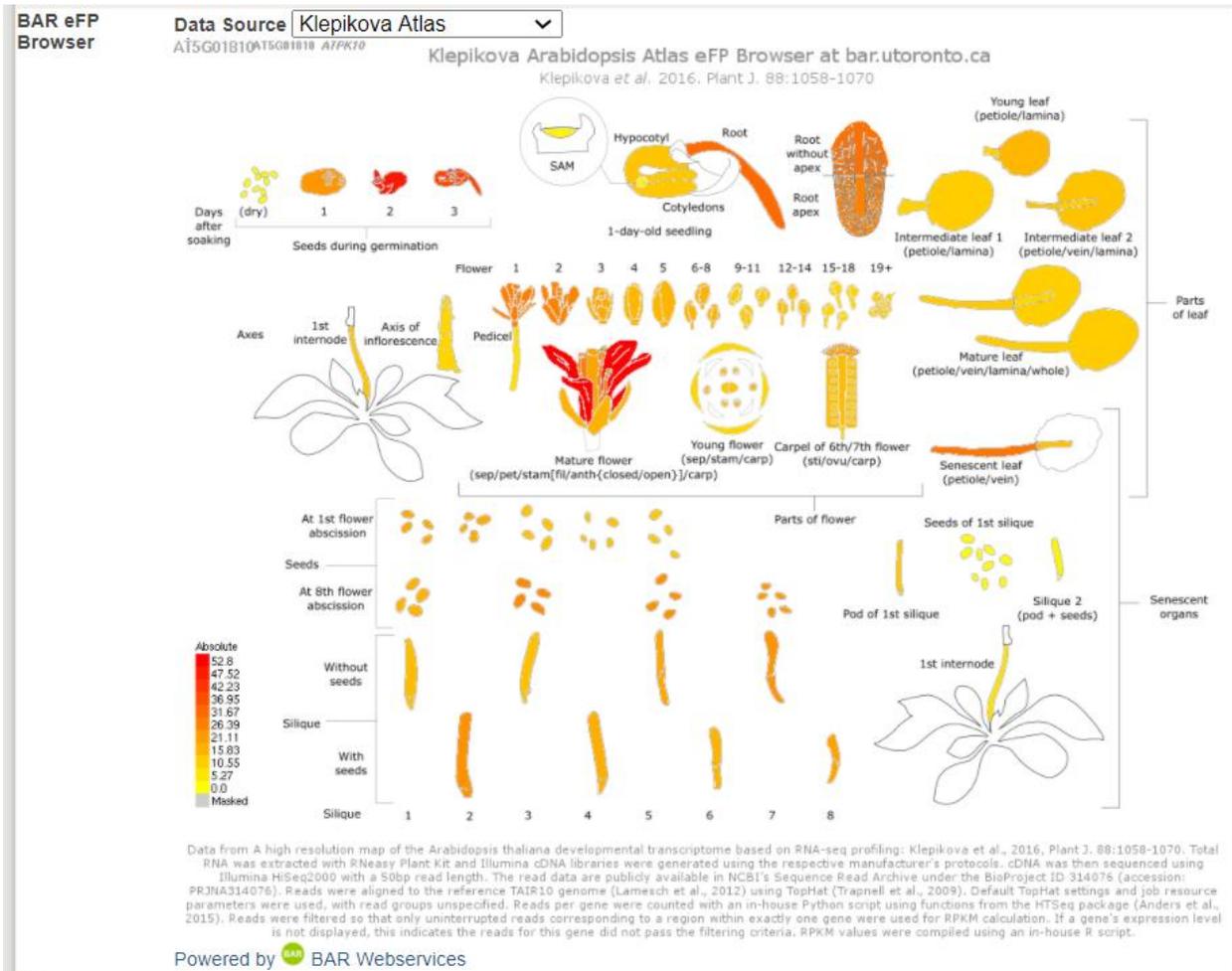
1. Gene locus
2. Other names for the gene:
3. Biological Processes in which the gene plays a role (GO Biological Process)
4. The cellular component in which the protein product is expressed (GO Cellular Component)
5. Growth and developmental stages when the gene is expressed
6. The plant structures where the protein product of the gene is expressed

The screenshot shows the TAIR website interface for the locus AT5G01810. The top navigation bar includes links for Home, Help, Contact, About Us, Subscribe, Login, and Register. Below this is a search bar and a menu with options like Search, Browse, Tools, Portals, Download, Submit, News, and Stocks. The main content area is titled "Locus: AT5G01810" and includes a "What's new on this page" button and an "Add a Comment" button. The "Representative Gene Model" section shows "AT5G01810.1" with a "Gene Model Type" of "protein_coding". The "Other names" section lists several names: ATPK10, CBL-INTERACTING PROTEIN KINASE 15, CIPK15, PKS3, PROTEIN KINASE 10, SIP2, SNF1-RELATED PROTEIN KINASE 3.1, SNRK3.1, and SOS3-INTERACTING PROTEIN 2. The "Description" section states: "Encodes a CBL-interacting serine/threonine protein kinase, also has similarities to SOS2 kinase." The "Other Gene Models" section lists "AT5G01810.2" and "AT5G01810.3" as splice variants. The "Map Detail Image" section shows a genomic map with protein-coding genes represented by horizontal bars. The "Annotations" section is a table with columns for category, relationship type, and keyword.

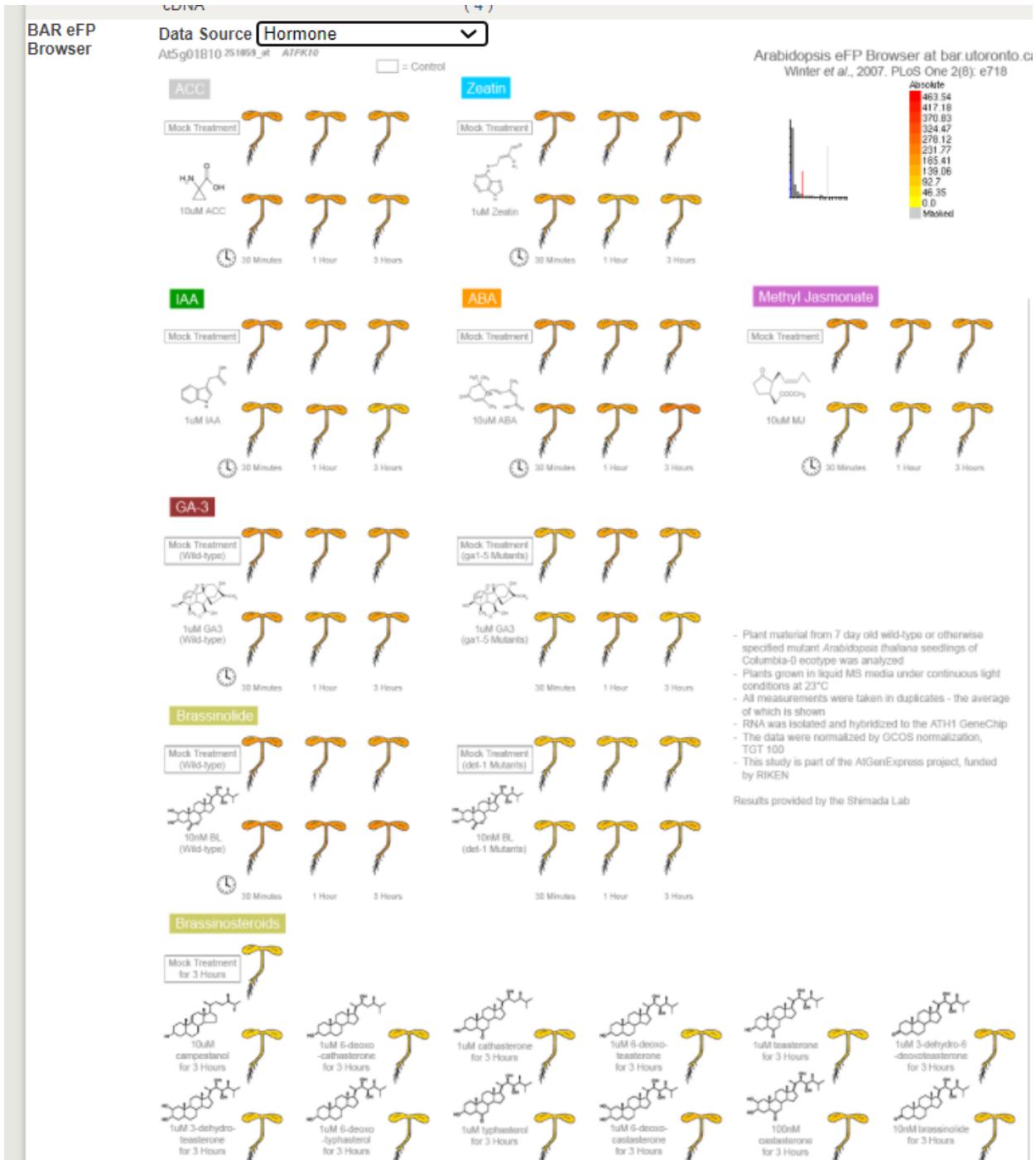
category	relationship type	keyword
GO Biological Process	involved in	abscisic acid-activated signaling pathway, negative regulation of abscisic acid-activated signaling pathway, protein phosphorylation, signal transduction
GO Cellular Component	located in	nucleus, plasma membrane
GO Molecular Function	has	protein binding, protein kinase activity
Growth and Developmental Stages	expressed during	L mature pollen stage, LP.02 two leaves visible stage, LP.04 four leaves visible stage, LP.06 six leaves visible stage, LP.08 eight leaves visible stage, LP.10 ten leaves visible stage, LP.12 twelve leaves visible stage, M germinated pollen stage, flowering stage, mature plant embryo stage, petal differentiation and expansion stage, plant embryo bilateral stage, plant embryo cotyledonary stage, plant embryo globular stage, vascular leaf senescent stage
Plant structure	expressed in	carpel, cauline leaf, collective leaf structure, cotyledon, flower, flower pedicel, guard cell, hypocotyl, inflorescence meristem, leaf apex, leaf lamina base, petal, petiole, plant embryo, plant sperm cell, pollen, pollen tube cell, root, seed, sepal, shoot apex, shoot system, stamen, stem, vascular leaf

Take a look at the BAR eFP (The Bio-Analytic Resource for Plant Biology electronic fluorescent pictograph) data. This is a browser engine that "paints" data from genomic data sets, such as

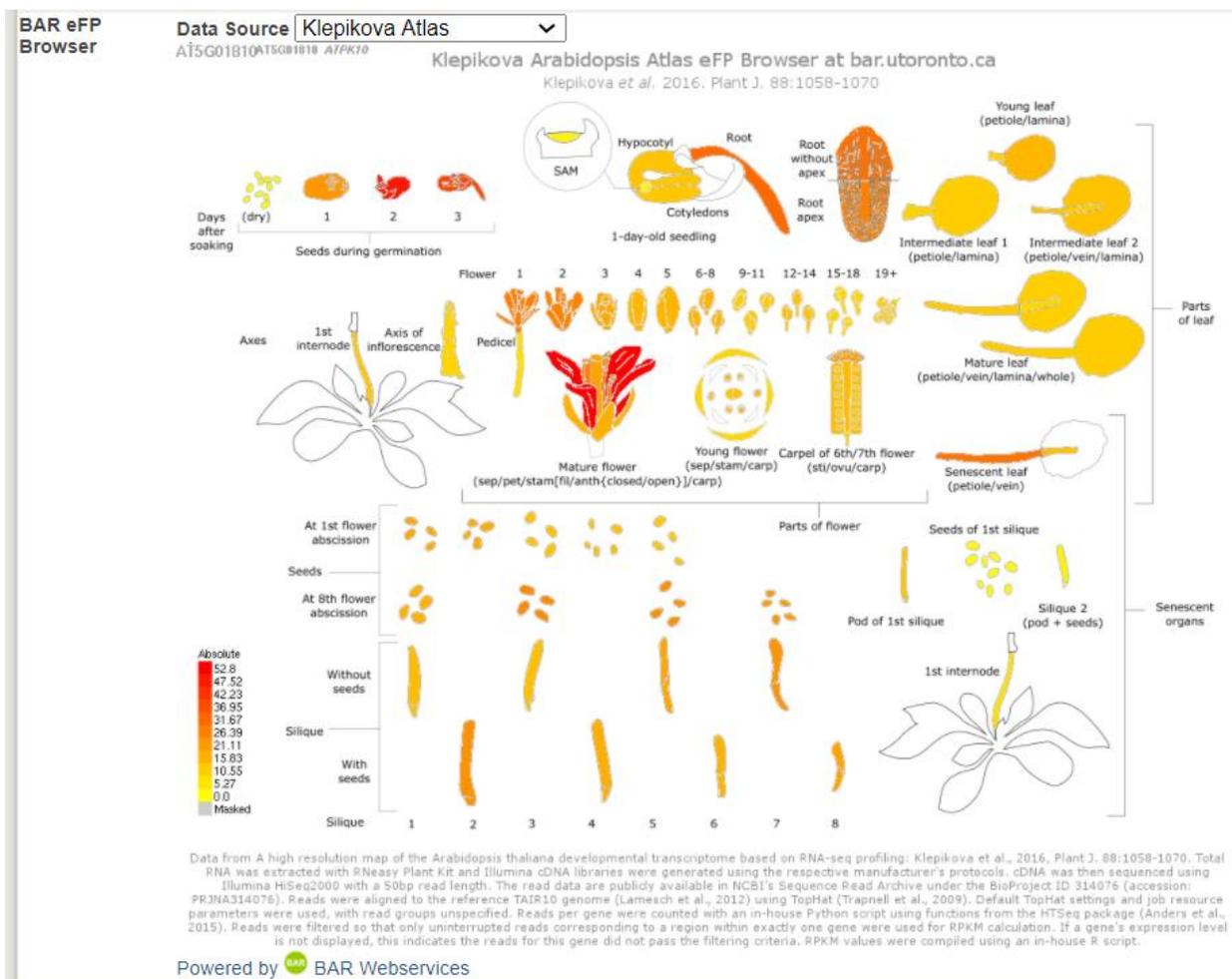
microarrays, ont pictographs that represent the experimental samples used to generate the data set. The purpose of the tool is to help researchers develop testable hypothesis based on the enormous amount of data generated by genomics projects. If you click the Data source you have options you can select that will provide you with information on experimental work others have conducted to study this gene. The information will be in a nicely illustrated summary form. The original reference will be included on the page as well.



Another example for the same gene:



This is a great place to look for information on your gene to use in your narrative. You should cite the original papers if you use the information in this section. You may need to go back to the original paper for details or clarity.



Under the Protein Data section, you will find the following information to include in your paper:

1. Protein Length
2. Molecular weight
3. Isoelectric point
4. List of InterPro domains: Create a table of the domains and their function (if the function is known). Click on the links. This will take you out to the InterPro site where you will find info on the domain. The information in the description might provide some useful information to include in your manuscript. In the table, you should indicate a very BRIEF description of whatever you think is most relevant about this particular domain (think about what your microarray experiment was to help you decide what might be the most useful information to include in the table) and the biological process, molecular function or cellular component that is applicable to the domain (see under GO terms). If no information is available, record "none" in your table. Example:

Domain	Brief Description	Biological Process	Molecular Function	Cellular Component
NAF/FISL_domain: IPR018451	Serine-threonine protein kinase that itneracts with calcineurin B-like calsium sensor proteins	Signal transduction	None	none

Table 1. Domain ontology from <http://www.ebi.ac.uk/interpro/entry/InterPro/IPR018451/>.

Chromosome 5					
Protein Data					
	name	length(aa)	molecular weight	isoelectric point	INTERPRO domains
	AT5G01810.1	421	47928.1	9.18	NAF/FISL_domain:IPR018451 ITPase:IPR016253 Protein_kinase_ATP_BS:IPR017441 Prot_kinase_dom:IPR000719 NAF_dom:IPR004041 Ser/Thr_kinase_AS:IPR008271 Ca/CaM-dep_Ca-dep_prot_Kinase:IPR020636 KA1_dom:IPR001772 Kinase-like_dom_sf:IPR011009 ITPase:IPR017442 ITPase:IPR002290 ITPase:IPR020660

All the way at the bottom of the TAIR page, you will find a list of publications related to the gene. Use these publications as references for your paper.

Part 4. Write your microarray paper.

Your microarray paper should contain the following components:

2. **Introduction:** Be sure to state the purpose of the study, why the experiment was conducted, review previous works of others in the field (integrated seamlessly, not one reference after another). How a microarray works is not needed here. Assume your reader is familiar with this now long-standing, common-place technique. Focus on your topic (osmotic signaling, sugar signaling, phytohormone signaling, or the interplay between sugar and phytohormone signaling).

4. **Discussion:**

a. Recap your results. Take a look at the descriptions for the genes that are up or down regulated. Now look at the review of literature you selected for homework. Are there genes on the list that you would expect to see based on the literature? Looking at the descriptions, are there genes that make sense to see? If you are looking at sugar, are there genes that are obviously part of sugar metabolic pathways or involved in photosynthesis? If you are looking at phytohormones, do the receptors to your chosen phytohormone appear on the list? You might want to pull up journal articles on some of the genes appearing on the list to explain why they might be appearing on your list. Include a few suggestions for future experiments that could be conducted to expand our understanding of your