Journal of Pharmacognosy and Phytochemistry
Available online at www.phytojournal.com

E-ISSN: 2278-4136
P-ISSN: 2349-8234
www.phytojournal.com
JPP 2020; 9(5): 987-995
Received: 05-06-2020
Accepted: 22-07-2020

Rutwick Surya U
Department of Life Science, Christ, Deemed to be University, Hosur road, Bangalore, Karnataka, India

Corresponding Author:
Rutwick Surya U
Department of Life Science, Christ, Deemed to be University, Hosur road, Bangalore, Karnataka, India

In-silico characterization of At5g18130 gene in Arabidopsis thaliana with emphasis on its expression patterns and functional aspects

Rutwick Surya U

DOI: https://doi.org/10.22271/phyto.2020.v9.i5n.12358

Abstract
Presently, Genomics research has flourished due to advances in bioinformatics technology. There is however still a lot of unknowns when it comes to the genomic characterization of the model plant Arabidopsis thaliana. In this study, one such uncharacterized gene At5g18130 is chosen and its functional, expressional, and biochemical characters are annotated using online tools. With all the information gathered, it was concluded that At5g18130 was a gene coading for a transmembrane protein that played a role primarily in leaf senescence in and to a minor extent in seed germination, fruit and embryo maturation and shoot apex elongation, At5g18130 also modulated physiological changes to long time exposure of osmotic stress.

Keywords: Arabidopsis, gene ontology, Coexpressed gene, At5g1813

Introduction
Bioinformatics is a vastly developing field and houses numerous scientists empowered by newer tools developed and made available in an ever so increasing magnitude of interest and dedicated developers. This empowers scientists to characterize data revolving around genomics, proteomics, metabolomics, transcriptomics and many more, all *in silico*. In the present study, such tools were used to identify and charecterize a gene and hence its corresponding protein known for its poor annotation, AT5G18130 belonging to the plant species *Arabidopsis thaliana*. Basic gene and its corresponding protein data, were collected from various databases to identify homologues of the gene At5g18130 and its protein, because, very little is know about it. The expression patterns of At5g18130 and its homologues, along with its coexpressors, were characterized with the bioinformatics tools for expression and coexpression analysis. To identify the transcription factors that bind to and regulate the expression of At5g18130 and genes which exhibit similar tissue-specific patterns of expression patterns and coexpress with At5g18130, the promoters of these genes were examined for the presence of significant ciselements with promoter analysis tools. The functional aspects of the At5g18130 gene and hence its corresponding protein in the tissues where its expression is abundant was explored by applying tools that leverage the guilt-by-association principle to recognize the gene ontology (GO) enriched terms and annotations for lists of coexpressed genes. A pathway mapping tool was used to identify pathways in which At5g18130 and its top 50 coexpressed genes were involved. Finally, the neighbours of At5g18130 were identified using network exploration methods – those that form a direct network connexion and eswtablish interaction with At5g18130 via protein- interactions were found out, although this only led to one potential interaction that could possibly be part of its biological background. Lastly it was attempted to identify potential gene transcription regulators but none such were documented. With all the information gathered, it was concluded that At5g18130 was a gene coading for a transmembrane protein that played a role primarily in leaf senescence in and to a minor extent in seed germination, fruit and embryo maturation and shoot apex elongation, At5g18130 also modulated physiological changes to long time exposure of osmotic stress.

Methods
Initial impressions
A brief review of literature on the gene was done. The gene sequence information was accessed from its genbank data on NCBI [1]. More comprehensive gene structural information was accessed from Araport at https://www.araport.org [2].

Ensemble Plants was used to find phylogenetic relationships between the gene of interest and its homologs athttp://plants.ensembl.org/. Phylogenetic trees curated by Gramene and Ensembl Plants were used to plot homologs and their phylogenetic relationships to At5g18130 at http://www.gramene.org and http://plants.ensembl.org, respectively [3, 4]. All the homologues and orthologues of At5g18130 in Arabidopsis were identified using PLAZA 4.0 [5]. These tools were accessed on 30st March 2020.

Expression analysis

The tissues in which At5g18130 were highest expressed were identified using the developmental map and specific tissue related data in the Arabidopsis eFP Browser [6] at http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi and JBrowse from Araport at https://apps.araport.org/jbrowse/?data=arabidopsis viewing the available read alignment tracks [2] were used to understand the results from the eFP Browser. All of the above resources were accessed on 30th march 2020.

Coexpression analysis

Expression Angler [7] at http://bar.utoronto.ca/ExpressionAngler/ was used to arrive at the top 50 coexpressed genes of At5g18130 with respect to the expression levels in At5g18130 as well as the expression level patterns in its homologs, these results were compared for various tissues in which the expression was earlier observed. Additionally, a "Custom Expression Pattern" was defined to identify genes that showed expression patterns similar to At5g18130 using a tissue specific search. Genes with the greatest mutual rank with At5g18130 were determined with ATTED-II [8] at http://atted.jp/. All these tools were accessed on 31st March 2020.

Promoter analysis

The earlier obtained list of coexpressed genes with a Custom Expression Pattern like At5g18130 from the Expression Angler tool that were in common with the ATTED II tool's set of genes with high mutual rank with At5g18130 were analyzed to spot common motifs using the Cistome tool [7] at http://bar.utoronto.ca/cistome/cgi-bin/BAR_Cistome.cgi. The *TAIR* Upstream (TSS/TrSS) 1000 bp promoter data set was used, and for Step 3 "Paste in your own PSSMs or consensus sequences and/or choose precomputed motifs from various sources for de novo mapping" was cutoff > 2 (other settings were kept to their respective defaults). The top coexpressed genes from just the Expression Angler "Tissue Specific" output were also investigated.

Functional classification and pathway visualization

Enriched GO terms were identified with AgriGO (reference genome: Affymetrix ATH1 Genome Array (blast); test: hypergeometric test with Hochberg FDR correction) at http://systemsbiology.cau.edu.cn/agriGOv2/species_analysis.php?SpeciseID=1&latin=Arabidopsis_thaliana [9] using the top coexpressed genes showing the same "Tissue Specific" expression pattern as At5g18130 from Expression Angler. The tool g:Profiler [10] at https://biit.cs.ut.ee/gprofiler/ (with Statistical domain scope set to "Custom over all known genes", background set to "AFFY_ATH1_121501" and Significance threshold set to "Benjamini-Hochberg FDR", to be as close to the AgriGO settings as possible (the species was selected to be Arabidopsis thaliana, and all other settings were left as their defaults)) was used to identify and further confirm the enriched GO results, clearly the results complemented and supported each other as discussed in the next section.

The potential pathways of this set of genes were investigated through the Cellular Overview feature of Aracyc [10] at https://pmn.plantcyc.org/overviewsWeb/celOv.shtml. All the above resources were accessed 31st April, 2020.

Network tools

Interactions with At5g18130 were investigated with Arabidopsis Interactions Viewer 2 using default parameters [11] at https://bar.utoronto.ca/interactions2/. Enriched transcription factors for the set of top coexpressed genes showing similar expression patterns as At5g18130 were identified with TF2Network [12] at http://bioinformatics.psb.ugent.be/webtools/TF2Network/ ePlant [13] was also used, with default settings. Neither of these tools gave out any possible network or connection among any of the coexpressed genes except one. All above resources were accessed 31st April 2020.

Results and Discussion

The gene At5g18130 is a 2,176 base pair long linear DNA sequence primarily located on chromosome 5 from the position 5,995,323 to 5,997,498 bases on the forword strand of *Arabidopsis thaliana* (Thale cress) belonging to the ecotype Columbia. The lineage of *A. Thaliana* is - Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae; Pentapetalae; rosids; malvids; Brassicales; Brassicaceae; Camelineae; Arabidopsis. The gene At5g18130 is also known as F20H23.8 and F20H23_8. Its function is primarily described as a gene coding Transmembrane Protein [1].

The gene is known to have 2 transcripts consisting of 3 exons and 2 introns. The 2 proteins annotated for this gene are Q3E9G1 and Q9FK57 which are distributed in eukaryotic lineage as transmembrane proteins with uncharacterized functions.

The gene At5g18130 was found to have two 1:1 orthologues namely g13283 and fgenesh2_kg.6__1813__AT5G18130.1 as per the data on Ensembl Plants at http://plants.ensembl.org/. Both these were found be originating from a common ancestor and under a common speciation node. PLAZA further showed a family of 46 genes that were homologous to the query gene At5g18130, namely HOM03D004032.

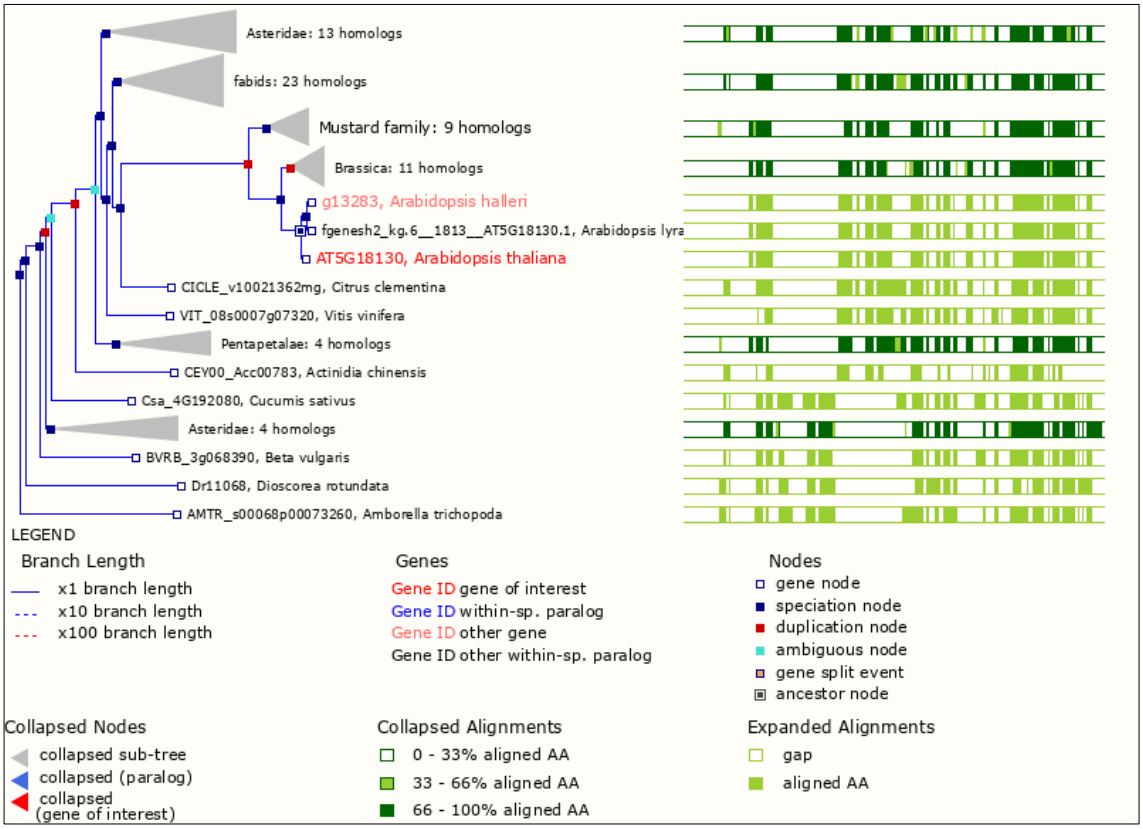

Fig 1: Phylogenetic tree as curated by Ensembl Plants. AT5G18130 gene is seen in red, along with its Paralogues

The eFP browser was used to analyse the tissue specific expression levels of our gene or intrest. At5g18130 showed highest expression levels in senescent leaves, dry seeds and minor expression levels in mature floral whorls and the abiotic stress specific expression levels seemed to increase with longer duration (12 to 24 hours) or osmotic stress [15]. The tissue specific expression levels were higher for mature green embryo showing a gradual increase [16] while the reverse was true with germination, that is, the expression levels decreased with progressing germination, the highest expression

level for seeds were observed to be in the dry ungerminated seed stage. Silique showed significant expression only in DPA 9 and 12 stages [17].

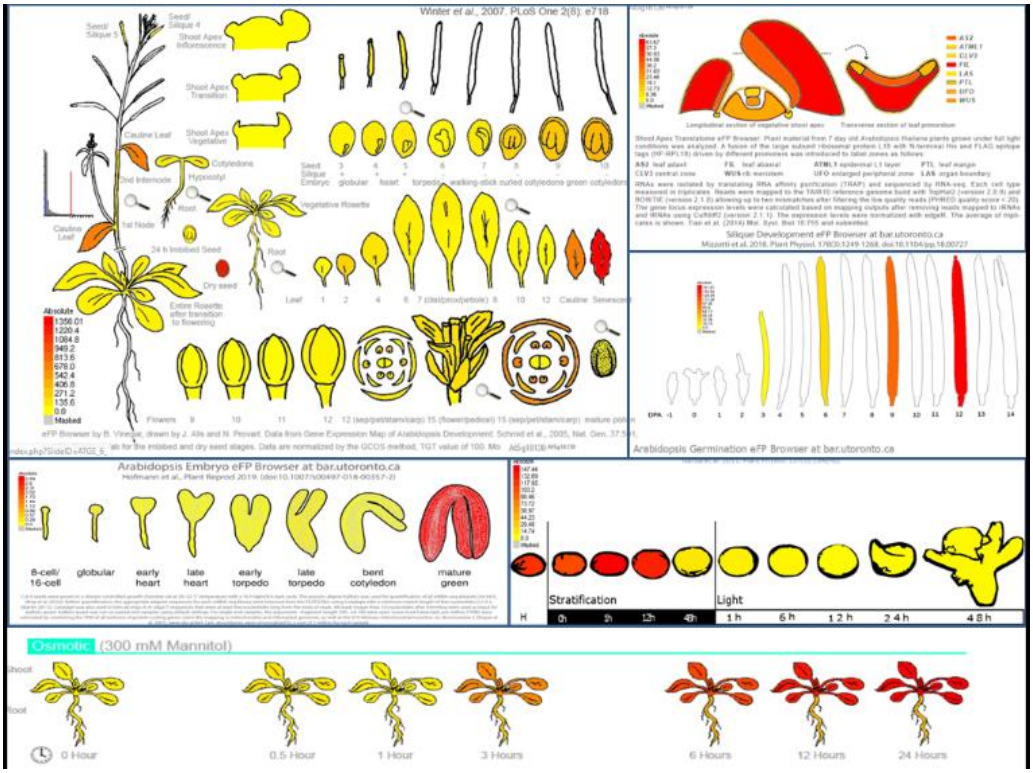

Fig 2: Expression levels of the gene At5g18130 in various tissues at various instances. Top left corner - Tissue specific expression; Top right corner - expression levels in shoot apex; Below the Top right corner - expression levels in silique; Bottom most - expression levels in osmotic stress; Bottom left -expression in devoloping embryo; Bottom right –expression in germinating seed.

Table 1: Expression levels in tissue specific manner for At5g18130 from eFP Browser

Tissue	Expression Level	Standard Deviation	Samples
Dry seed	1256.38	160.09	RIKEN-NAKABAYASHI1A,RIKEN-NAKABAYASHI1B,
Imbibed seed, 24 h	30.36	4.59	RIKEN-NAKABAYASHI2A,RIKEN-NAKABAYASHI2B,
1st Node	34.76	3.06	ATGE_28_A2,ATGE_28_B2,ATGE_28_C2,
Flower Stage 12, Stamens	275.21	5.53	ATGE_36_A,ATGE_36_B,ATGE_36_C,
Cauline Leaf	678.45	22.67	ATGE_26_A,ATGE_26_B,ATGE_26_C,
Cotyledon	108.6	11.94	ATGE_1_A,ATGE_1_B,ATGE_1_C,
Root	116.93	19.52	ATGE_9_A,ATGE_9_B,ATGE_9_C,
Entire Rosette After Transition to Flowering	78.53	9.1	ATGE_23_A,ATGE_23_B,ATGE_23_C,
Flower Stage 9	22.58	2.61	ATGE_31_A2,ATGE_31_B2,ATGE_31_C2,
Flower Stage 10/11	22.98	3.58	ATGE_32_A2,ATGE_32_B2,ATGE_32_C2,
Flower Stage 12	59.51	3.6	ATGE_33_A,ATGE_33_B,ATGE_33_C,
Flower Stage 15	167.78	12.47	ATGE_39_A,ATGE_39_B,ATGE_39_C,
Flower Stage 12, Carpels	27.0	6.93	ATGE_37_A,ATGE_37_B,ATGE_37_C,
Flower Stage 12, Petals	22.05	2.23	ATGE_35_A,ATGE_35_B,ATGE_35_C,
Flower Stage 12, Sepals	108.71	11.19	ATGE_34_A,ATGE_34_B,ATGE_34_C,
Flower Stage 15, Carpels	51.25	4.46	ATGE_45_A,ATGE_45_B,ATGE_45_C,
Flower Stage 15, Petals	341.18	17.26	ATGE_42_B,ATGE_42_C,ATGE_42_D,
Flower Stage 15, Sepals	625.21	31.56	ATGE_41_A,ATGE_41_B,ATGE_41_C,
Flower Stage 15, Stamen	656.35	24.83	ATGE_43_A,ATGE_43_B,ATGE_43_C,
Flowers Stage 15, Pedicels	47.95	4.77	ATGE_40_A,ATGE_40_B,ATGE_40_C,
Leaf 1 + 2	39.11	4.01	ATGE_5_A,ATGE_5_B,ATGE_5_C,
Leaf 7, Petiole	27.76	3.63	ATGE_19_A,ATGE_19_B,ATGE_19_C,
Leaf 7, Distal Half	89.33	8.08	ATGE_21_A,ATGE_21_B,ATGE_21_C,
Leaf 7, Proximal Half	50.21	7.43	ATGE_20_A,ATGE_20_B,ATGE_20_C,
Hypocotyl	53.55	6.52	ATGE_2_A,ATGE_2_B,ATGE_2_C,
Root	105.4	3.07	ATGE_3_A,ATGE_3_B,ATGE_3_C,
Rosette Leaf 2	237.03	11.55	ATGE_12_A,ATGE_12_B,ATGE_12_C,
Rosette Leaf 4	134.71	16.17	ATGE_13_A,ATGE_13_B,ATGE_13_C,
Rosette Leaf 6	86.08	15.6	ATGE_14_A,ATGE_14_B,ATGE_14_C,
Rosette Leaf 8	50.36	2.75	ATGE_15_A,ATGE_15_B,ATGE_15_C,
Rosette Leaf 10	54.03	0.85	ATGE_16_A,ATGE_16_B,ATGE_16_C,
Rosette Leaf 12	35.99	2.91	ATGE_17_A,ATGE_17_B,ATGE_17_C,
Senescing Leaf	1356.01	111.38	ATGE_25_A,ATGE_25_B,ATGE_25_C,
Shoot Apex, Inflorescence	25.86	4.11	ATGE_29_A2,ATGE_29_B2,ATGE_29_C2,
Shoot Apex, Transition	17.68	2.25	ATGE_8_A,ATGE_8_B,ATGE_8_C,
Shoot Apex, Vegetative	19.34	1.91	ATGE_6_A,ATGE_6_B,ATGE_6_C,
Stem, 2nd Internode	168.13	14.7	ATGE_27_A,ATGE_27_B,ATGE_27_C,
Mature Pollen	30.56	19.14	ATGE_73_A,ATGE_73_B,ATGE_73_C,
Seeds Stage 3 w/ Siliques	67.31	11.19	ATGE_76_A,ATGE_76_B,ATGE_76_C,
Seeds Stage 4 w/ Siliques	116.3	6.1	ATGE_77_D,ATGE_77_E,ATGE_77_F,
Seeds Stage 5 w/ Siliques	112.59	4.88	ATGE_78_D,ATGE_78_E,ATGE_78_F,
Seeds Stage 6 w/o Siliques	52.78	7.7	ATGE_79_A,ATGE_79_B,ATGE_79_C,
Seeds Stage 7 w/o Siliques	107.06	5.82	ATGE_81_A,ATGE_81_B,ATGE_81_C,
Seeds Stage 8 w/o Siliques	257.36	32.6	ATGE_82_A,ATGE_82_B,ATGE_82_C,
Seeds Stage 9 w/o Siliques	301.38	20.17	ATGE_83_A,ATGE_83_B,ATGE_83_C,
Seeds Stage 10 w/o Siliques	381.93	30.23	ATGE_84_A,ATGE_84_B,ATGE_84_D,
Vegetative Rosette	51.26	0.73	ATGE_89_A,ATGE_89_B,ATGE_89_C,

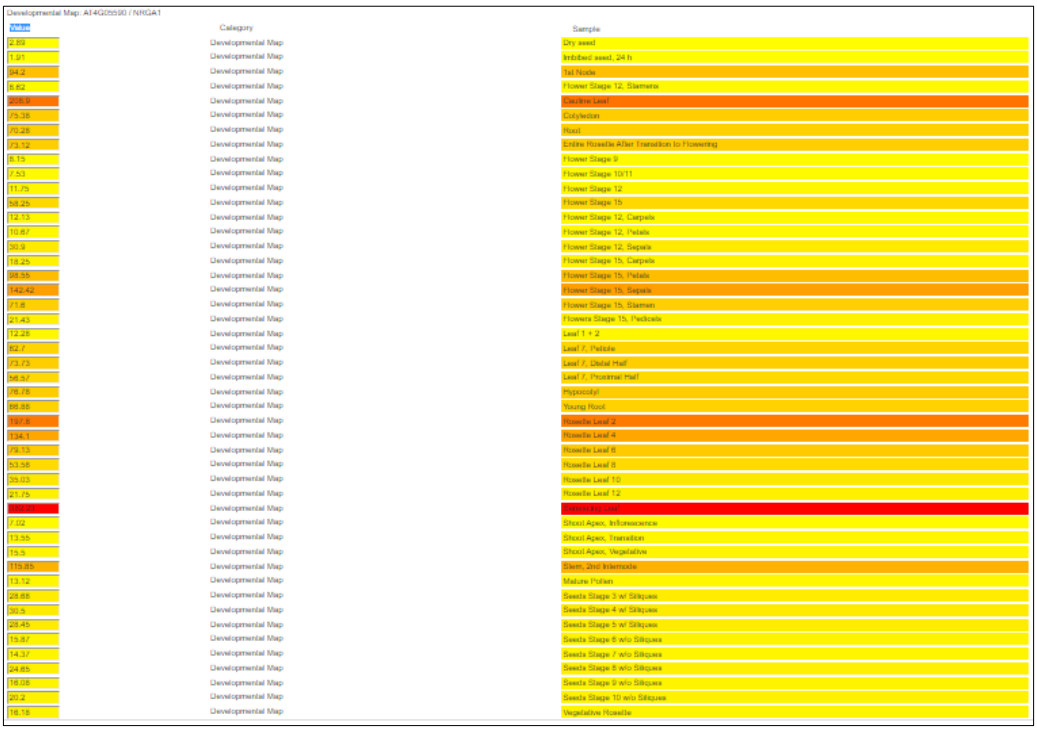

Fig 3: The top 50 genes along with their heatmap profiles found to have significant coexpression along with AT5G18130 when this gene is selected and screened under developmental map

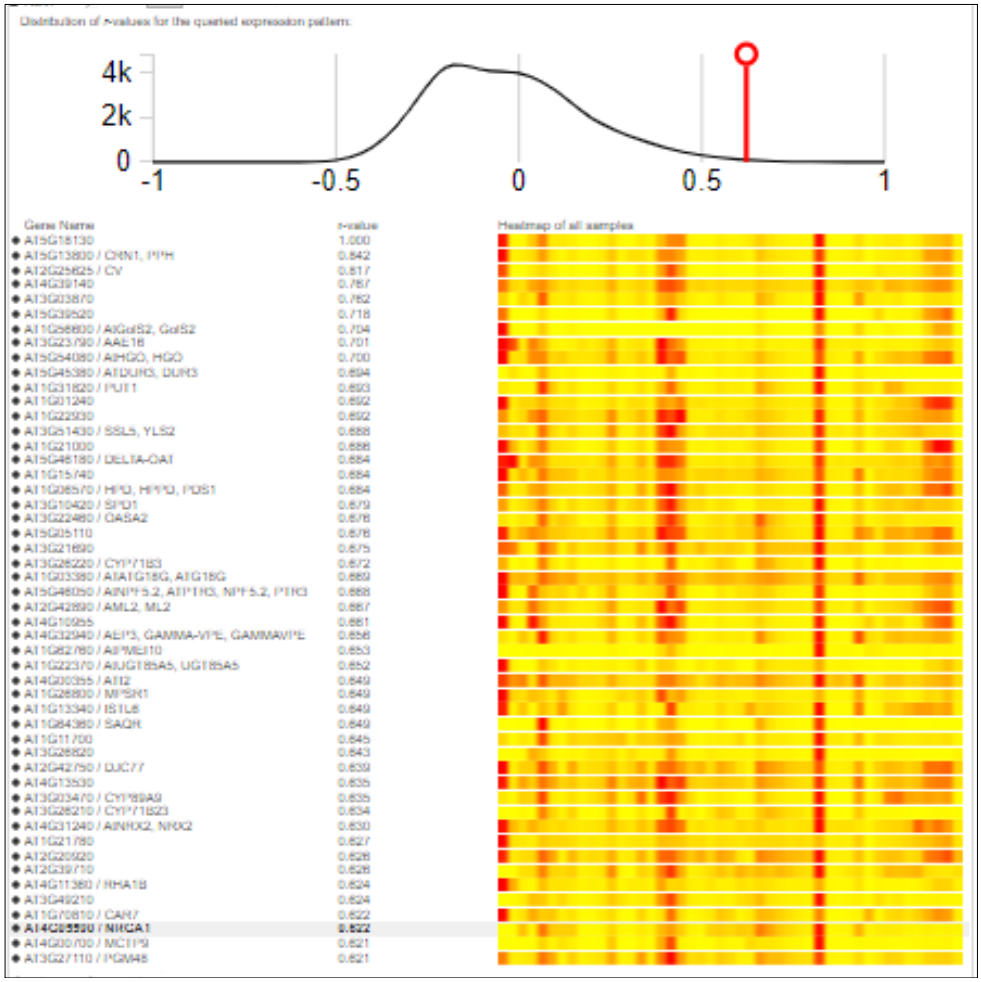

Fig 4: Heatmap profiles of the top 50 genes found by Expression angler to have significant level of coexpression using custom expression patterns set to senescent leaves

Coexpressed genes in direct contact with At5g18130 was found to be At3g03870, At1g76590 and At4g32940 showing the highest MR (mutual rank) values (table 1). Further top 50 genes coexpressed with At5g18130 were analyzed in developmental map view (Figure 2). The maximum expression levels were seen in senescent leaves which is also the tissue showing a maximum expression of our gene of interest At5g18130. A "custom expression pattern" was further constructed by defining the tissues showing high expression for At5g18130 gene which was senescent leaves (Figure 3). The obtained results were similar, with other half of the coexpressed genes identical in both the searches. These common genes are essentially those that is co-expressed with At5g18130 and has expression levels similar to At5g18130 in the chosen tissue ie. Senescent leaves.

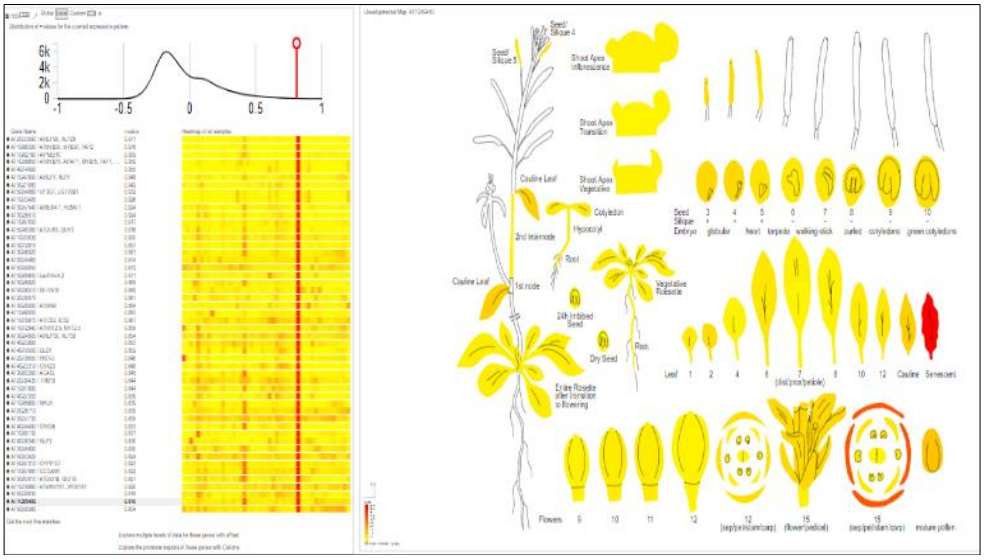

Fig 5: Co expressional genes under custom expression pattern set to high relative expression in senescent leaves using Expression Angler tool. The table at the right shows the heatmap of these co-expressed genes and the developmental map at the left shows the resultant expression levels in various tissues.

Through Aranet, it was found that the gene ontology biological function and molecular function was not known for At5g18130 while it was listed as a cellular component of nucleolus. Through Aranet, it was not possible to view the expression profiles of association supported by coexpressed data.

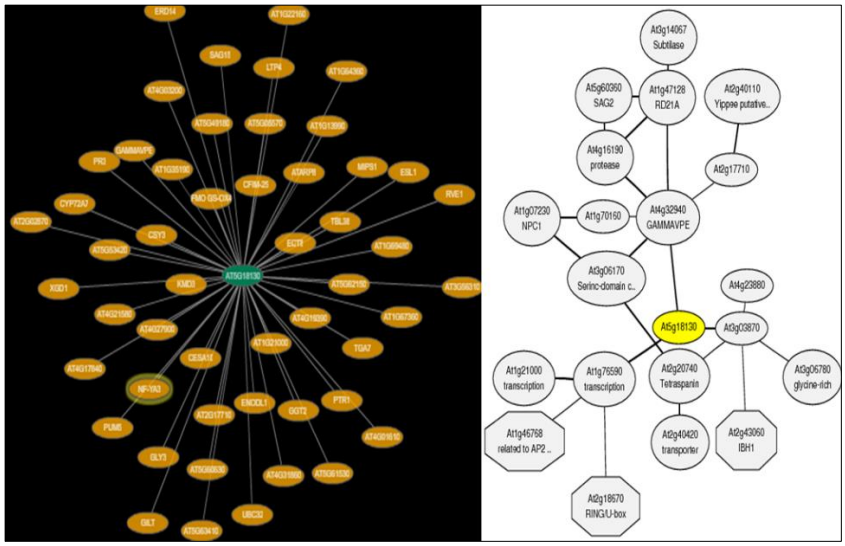

Fig 6: Gene association as described by Aranet for At5g18130 (Right) and genes coexpressed with At5g18130 as described by ATTED II (Left)

The PLACE database of the Cistome tool with TAIR Upstream (TSS/TrSS) chosen for a 1000 base pair promoter dataset to find motifs with Ze cutoff more than 2.0 for the set of genes that were found to be coexpressed with At5g18130 and common in the findings of the top 50 coexpressed gene sets generated by ATTED II and Expression Angler "Tissue specific" which are - AT2G33080, AT1G66390, AT1G62760, AT1G56650, AT4G140900, AT1G478900, AT3G210800, AT5G54060, AT1G034950, AT3G57540, AT3G285100, AT1G670000, AT5G45380, AT1G330300, AT3G129100, AT3G480200, AT2G044600, AT5G590500, AT5G65600, AT3G268200, AT5G06510, AT2G350700, AT3G29590, AT1G490000, AT1G18870, AT1G12940, AT3G24900, AT4G238800, AT4G10500, AT2G18050, AT4G23310, AT3G63380 AT2G39435 AT1G518900 AT4G273000 AT1G65690 AT2G287100, AT3G077300, AT4G04490, AT1G801300, AT4G38340, AT3G558900, AT5G534200, AT5G67310, AT1G67980, AT3G63010, AT1G18860, AT5G390300 and AT5G55560 respectively.

This resulted in 2 motifs with disappointingly low Ze cutoff values of 2.06 and 2.08 respectively for the motifs ACGTATERD1 with a consensus sequence of ACGT and the motif POLLEN1LELAT52 with a consensus sequence of AGAAA to be found. The motifs showed significant positional disequilibrium in its arrangement and location across the set of coexpressed genes of At5g18130. The transcription factor identified here was found not to be expressed in appreciable levels in the senescent leaves or dry seeds according to the eFP browser's data.

Fig 7: Motifs ACGTATERD1 with a consensus sequence of ACGT (green) and the motif POLLEN1LELAT52 with a consensus sequence of AGAAA (Red) plotted on their respective locations across the set of genes co-expressed with At5g18130.

The agriGO analysis (Figure 8) didn't show any significant enrichment in biological processes for any of the top coexpressed genes submitted for analysis, on the other hand there were no significant GO terms for determination of pathways under cellular components and molecular function.

The results from g:Profiler revealed that the top enriched GO:BP terms were response to oxygen containing terms and acids, while the top enriched GO:MF terms were protein kinase and phophotransferase activity respectively.

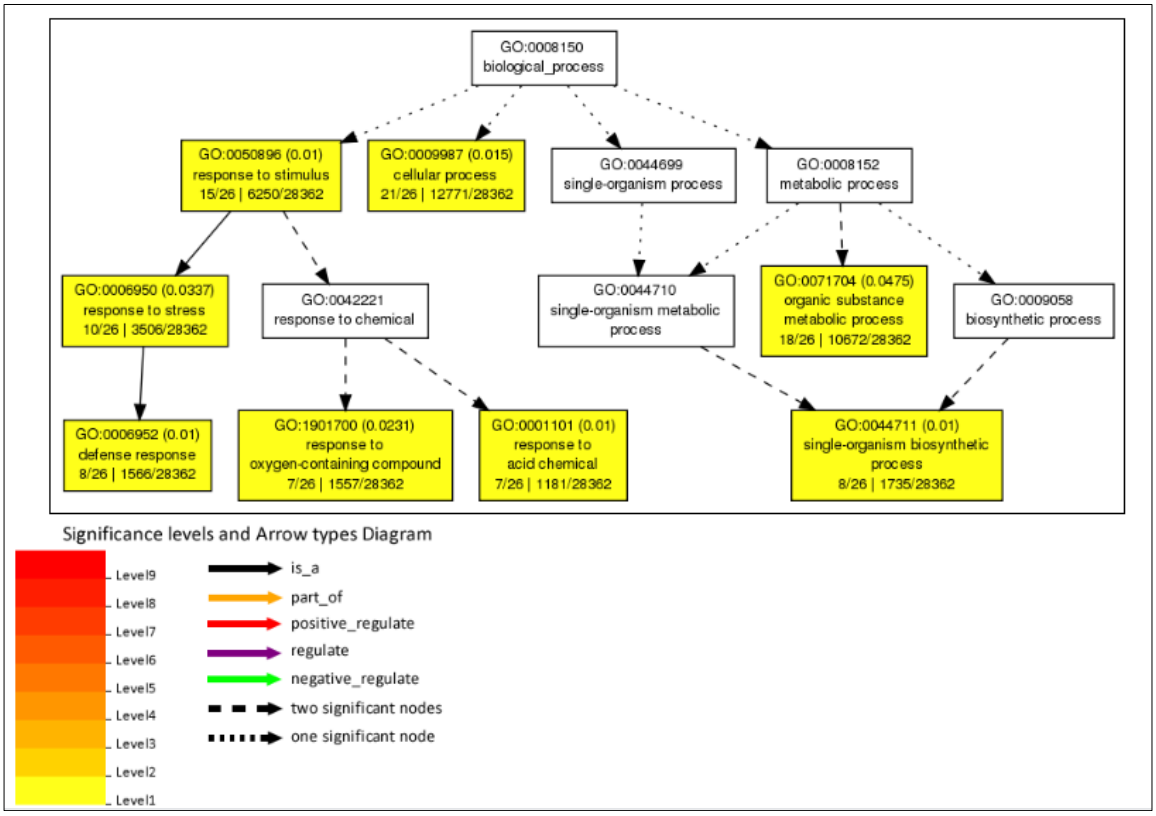

Fig 8: AgriGO analysis of the pathway derived from the top 50 genes coexpressed with At5g18130

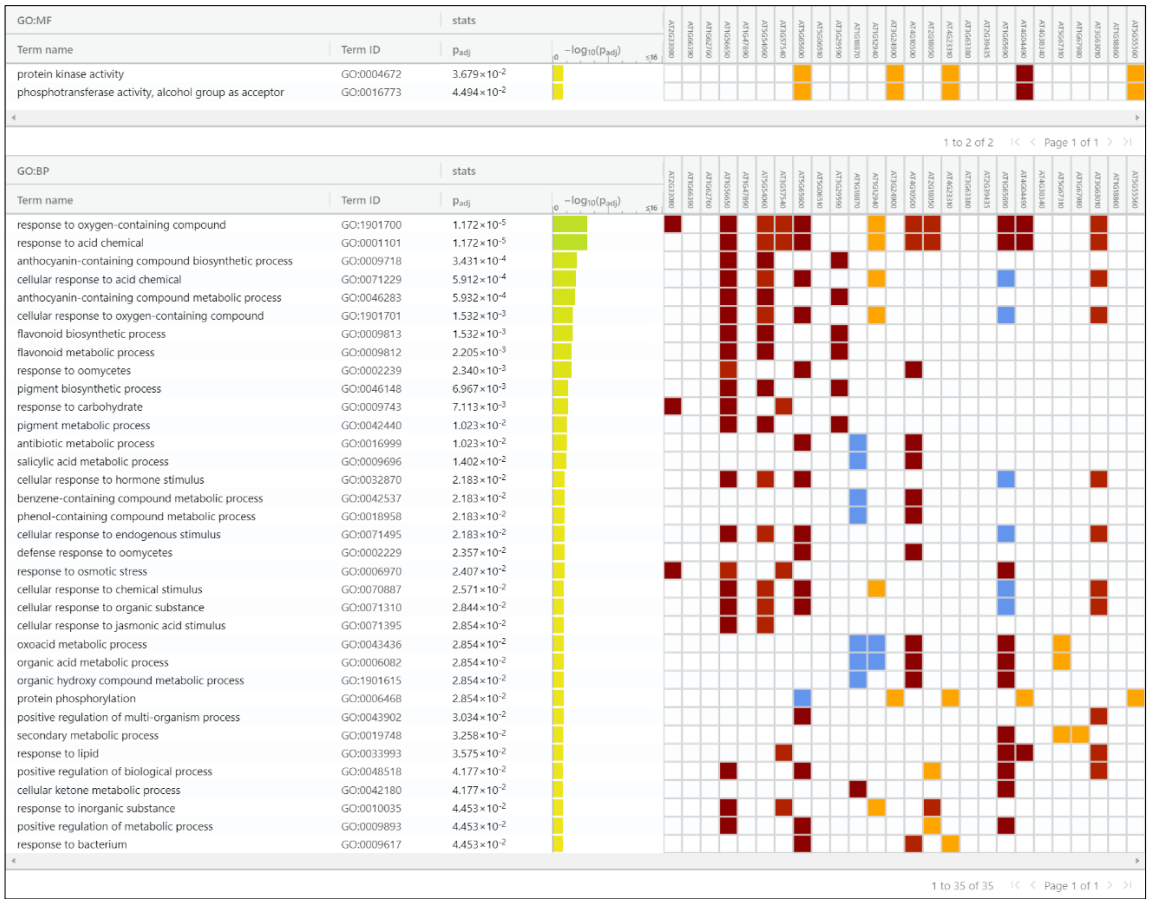

Fig 9: Profileer mapping of the genes with expression pattern similar to At5g18130 showing GO:MP (top) and GO:BP (bottom).

The gene At5g18130 barely showed any interactions or networks when analysed in Arabidopsis Interactions Viewer 2 and eplant except for an interaction with At3g63260 in the plasma membrane. It had no enriched regulators in TF2Network either.

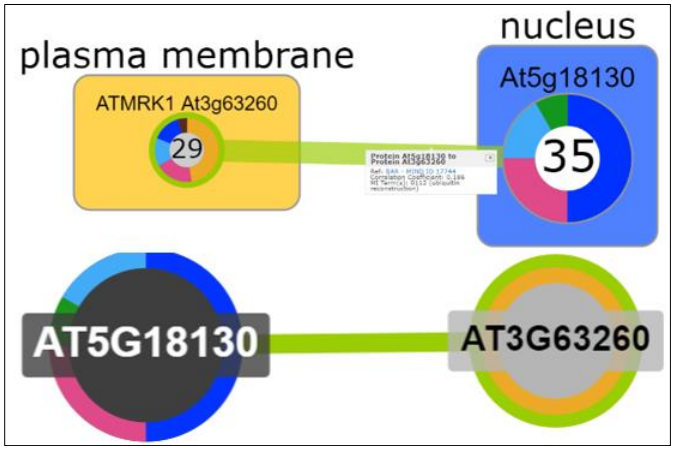

Fig 10: At5g18130 interacting with At3g63260 as shown by Arabidopsis Interaction viewer 2 (Top) and eplant (Bottom).

Summary

The gene At5g18130 has over 50 homologs in many plant species especially in the Brassicaceae family and hence it may be assumed that it has an important role to play in the pants. In *Arabidopsis thaliana* the gene At5g18130 arrived from a recent speciation node. At5g18130 is preferentially expressed in senescent leaves of this species, however, high expression levels in dry seeds, mature embryo, shoot apex and mature fruit isn't uncommon. The expression was seen to increase with embryo maturation and decrease with seed germination. It was also seen to be highly expressed in the plant undifferentially after longer durations of osmotic stress. But it must be noted that further experimental data is essential to substantiate these expression patterns. It is also highly suggested to induce stress (especially osmotic) in the plants expressing At5g18130 to identify physiochemical changes that complement its expression in these instances. One might also consider scrutinizing these data for other biotic and abiotic stress induced gene expression patterns and its complement phenotypic changes. The role of the gene At5g18130 in senescent leaf is also not known, it is hence suggested to induce gene expression in leaves and stimulate the changes that would occur on doing so, if this causes any signs of senescence in leaves, the gene's relation with the plant growth regulators may then be probed.

Apart from the expressional aspects, I was able to gather very little information for At5g18130, for instance the genes that co-express with At5g18130 are not many, and its highly suggested that these coexpressional genes be characterized. They may also be "turned off" or knocked off and the relative changes of At5g18130 may be observed and vice versa. The coexpressed genes may also be potential regulators of At5g18130 either directly or indirectly, which can only be identified by studying each of these genes with a comparison of its expression patterns in the absence and presence of the other coexpressed genes. GO enrichment results were consistent with At5g18130 and its top 50 coexpressed genes. The absolute functional aspects of At5g18130 and its relative functional aspects in the presence of its coexpressed genes were similar. The GO data however wasn't sufficient in relation to CC terms, and barely gave a few genes in MF terms while a plethora of genes in BP was characterised. The coexpressed neighbours were identified to be similar transmembrane proteins, they too showed expression in similar pattern in senescent leaves, but a surprising level of expression was also seen in floral whorls which was absent in At5g18130. It wasn't possible to construct any pathway or interaction among the coexpressed genes with the given databases, perhaps in the future, with more of these genes characterized; some essential networks may be constructed.

References

1. Sequence and analysis of chromosome 5 of the plant Arabidopsis thaliana. Nature. 2000; 408:823-826.
2. Krishnakumar V *et al*. Araport: the Arabidopsis Information Portal. Nucleic Acids Research. 2014; 43:D1003-D1009.
3. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J *et al*. Ensembl 2018. Nucleic Acids Res. 2017; 46(D1):D754–D761. doi:10.1093/nar/gkx1098.
4. Tello-Ruiz MK, Naithani S, Stein JC, Gupta P, Campbell M, Olson A *et al*. Gramene 2018: unifying comparative genomics and pathway resources for plant research. Nucleic Acids Res. 2018; 46(D1):D1181–D1189. doi:10.1093/nar/gkx1111.
5. Van Bel M, Diels T, Vancaester E, Kreft L, Botzki A, Van de Peer Y *et al*. PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. Nucleic Acids Res. 2017; 46(D1):D1190–D1196. doi:10.1093/nar/gkx1002.
6. Austin RS, Hiu S, Waese J, Ierullo M, Pasha A, Wang TT *et al*. New BAR tools for mining expression data and exploring Cis-elements in Arabidopsis thaliana. Plant J. 2016; 88(3):490-504. doi:10.1111/tpj.13261
7. Obayashi T, Aoki Y, Tadaka S, Kagaya Y, Kinoshita K. ATTED-II in 2018: A Plant Coexpression Database Based on Investigation of the Statistical Property of the Mutual Rank Index. Plant Cell Physiol. 2017; 59(1):e3–e3. doi:10.1093/pcp/pcx191.
8. Du Z, Zhou X, Ling Y, Zhang Z, Su Z. AgriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res. 2010; 38:W64–W70. doi:10.1093/nar/gkq310.
9. Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, Vilo J. g:Profiler-a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res. 2016; 44(W1):W83–W89. doi:10.1093/nar/gkw199.
10. Mueller LA, Zhang P, Rhee SY. AraCyc: A Biochemical Pathway Database for Arabidopsis. Plant Physiol. 2003; 132(2):453-460. doi:10.1104/pp.102.017236.
11. Dong S, Lau V, Song R, Ierullo M, Esteban E, Wu Y *et al*. Proteome-wide, Structure-Based Prediction of Protein-Protein Interactions/New Molecular Interactions Viewer. Plant Physiol. 2019; 179(4):1893-1907. doi: 10.1104/pp.18.01216.
12. Kulkarni SR, Vaneechoutte D, Van de Velde J, Vandepoele K. 2017. TF2Network: predicting transcription factor regulators and gene regulatory networks in Arabidopsis using publicly available binding site information. Nucleic Acids Res. 46(6):e31–e31. doi:10.1093/nar/gkx1279.
13. Waese J, Fan J, Pasha A, Yu H, Fucile G, Shi R *et al*. ePlant: Visualizing and Exploring Multiple Levels of Data for Hypothesis Generation in Plant Biology. Plant Cell. 2017; 29(8):1806-1821. doi: 10.1105/tpc.17.00073.

14. Winter D *et al*. An Electronic Fluorescent Pictograph Browser for Exploring and Analyzing Large-Scale Biological Data Sets. PLoS ONE. 2007; 2:e718.
15. Hofmann F, Schon M, Nodine M. The embryonic transcriptome of Arabidopsis thaliana. Plant Reproduction. 2019; 32:77-91.
16. Mizzotti C *et al*. Time-Course Transcriptome Analysis of Arabidopsis Siliques Discloses Genes Essential for Fruit Development and Maturation. Plant Physiology. 2018; 178:1249-1268.