Infernet: Gene Function Inference By Leveraging Large, Organ-Specific Expression Datasets And Validation Of Non-Redundant Regulators

Kranthi Varala[1]* (kvarala@purdue.edu), Rajeev Ranjan[1], Ying Li[1], and Karen Hudson[2]
[1]Purdue University, West Lafayette, IN; [2] USDA-ARS, West Lafayette, IN

https://www.purdue.edu/hla/sites/varalalab/infernet

Project Goals

This project combines computational approaches, e.g., machine learning, network inference and phylogenomics, with molecular approaches, e.g., metabolite profiling and ChIP-Seq, to find novel transcription factors (TF) that regulate traits of agronomic or biofuel interest. This project focuses on the biofuel trait of seed oil synthesis as a proof of concept that is extensible to any agronomic/biofuel trait of interest. This project focuses on regulation of a biological process of interest (e.g., lipid biosynthesis) in an organ specific manner (e.g., in seeds) and by estimating the likelihood of a given TF being redundant in its function (Aim 1). We then validate our functional predictions, using transgenic lines (Aim 2), via phenotypic assays (Aim 3a) and by identifying the specific targets these TFs regulate (Aim 3b). Finally, we translate the validated TF regulation knowledge gained in a model species (*Arabidopsis*) to biofuel crops (e.g., *Camelina sativa*) (Aim 4).

Abstract

Gene regulator network inference [1] from public RNA-Seq data (Aim 1) identified the top predicted regulators of seed lipid biosynthesis. This list included four known regulators of this process in the top 10 predicted TFs and many novel TFs that are predicted to have a strong influence on seed lipid biosynthesis. We have identified and collected mutant lines in most of the

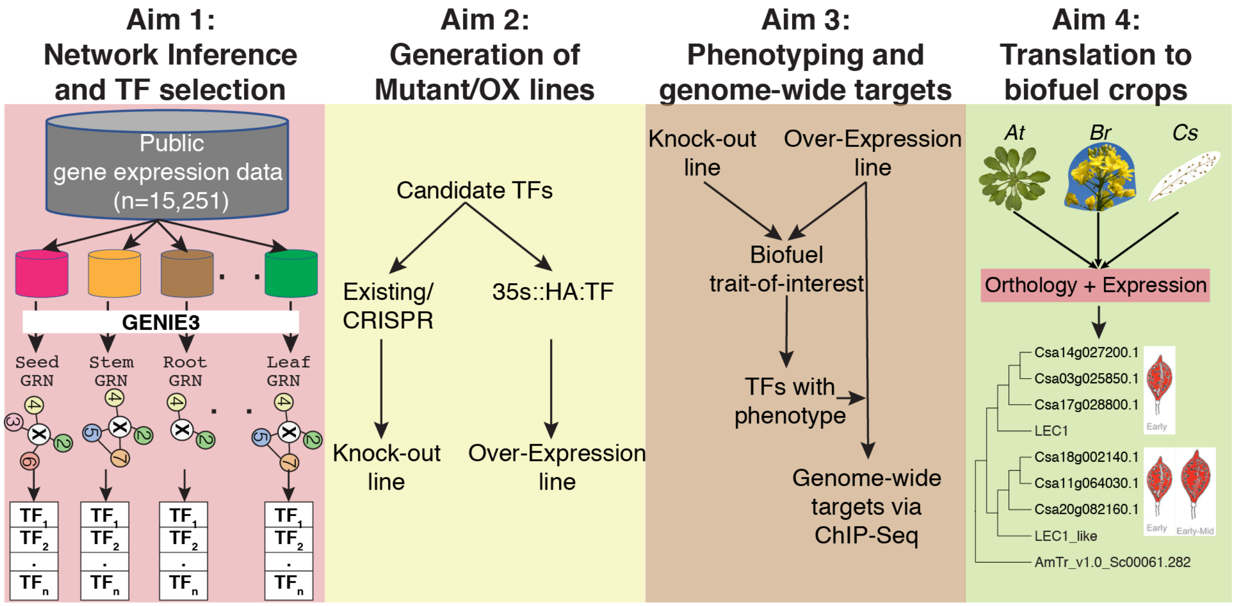
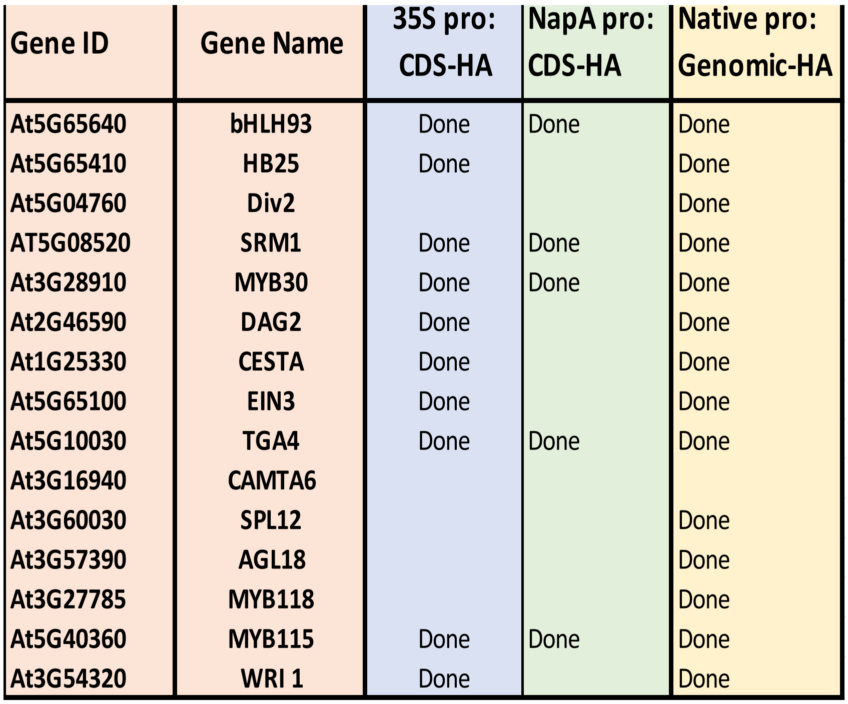


Table 1 Over-expression lines for each candidate TF. All TFs were conjugated with a HA tag to enable ChIP-Seq assays to identify global targets of TF binding.



Gene ID	Gene Name	35S pro: CDS-HA	NapA pro: CDS-HA	Native pro: Genomic-HA
At5G65640	bHLH93	Done	Done	Done
At5G65410	HB25	Done		Done
At5G04760	Div2			Done
AT5G08520	SRM1	Done	Done	Done
At3G28910	MYB30	Done	Done	Done
At2G46590	DAG2	Done		Done
At1G25330	CESTA	Done		Done
At5G65100	EIN3	Done		Done
At5G10030	TGA4	Done	Done	Done
At3G16940	CAMTA6			
At3G60030	SPL12			Done
At3G57390	AGL18			Done
At3G27785	MYB118			Done
At5G40360	MYB115	Done	Done	Done
At3G54320	WRI 1	Done		Done

novel TFs and are in the process of phenotyping them (see below) to detect changes in seed lipid profile. In addition, we generated over-expression lines for each of these candidate TFs using a set of 3 promoters: i. a generic 35S promoter for robust plant-wide over-expression ii. a seed-specific NapinA promoter for robust seed-specific over-expression and iii. native promoter driven expression. In all three constructs the TF was tagged with a HA epitope to allow downstream ChIP-Seq studies (see Table 1). In addition to the proposed lipid profiling of mature seeds from the mutant and over-expression lines, seed size and shape were assayed and compared to the wild-type mature seeds.

The GRN inference used in Aim 1 only considered TFs as regulators of gene expression. Subsequently, we repeated the inference pipeline using TFs + epigenetic (i.e., DNA and histone) modifiers as potential regulators of gene expression. Using the shoot and root apical meristems as use cases we identified many known and novel non-TF regulators of plant development. A manuscript describing this study was submitted for peer review and is currently under revision.

Publications

McCoy, R.M., Julian, R., Kumar, S.R.V., Ranjan, R., Varala, K., Li, Y. A systems biology approach to identify essential epigenetic regulators for specific biological processes in plants. Plants, 2021 (under revision)

References

1. Huynh-Thu, V.A., A. Irrthum, L. Wehenkel and P. Geurts, Inferring regulatory networks from expression data using tree-based methods. PLoS One, 2010. 5(9).

The InferNet grant is supported by the Genomics Science program within the Office of Biological and Environment Research in the DOE Office of Science.