Title: Cell-type specific analysis reveals the spatial and temporal transcriptome of the Sorghum bicolor stem

Authors: Jie Fu[1, 2]* (jief3@illinois.edu), Lye Meng Markelie[3], John Mullet[4], Kankshita Swaminathan[1, 5], and **Amy Marshall-Colon**[1, 2]

Institutions: [1]DOE Center for Advanced Bioenergy and Bioproducts Innovation (CABBI), University of Illinois, Urbana-Champaign, Urbana, [2]Department of Plant Biology, University of Illinois, Champaign-Urbana, Illinois; [3]Environmental Molecular Sciences Laboratory (EMSL), Richland, Washington; [4]Department of Biochemistry & Biophysics, Texas A&M University, College Station, Texas; [5]HudsonAlpha Biotechnology Institute, Huntsville, Alabama

Project Goals:

The specific aim of this project is to achieve a comprehensive understanding of sorghum stem biology by constructing molecular atlas on the scale of each specific stem cell type as well as bulk stem tissue by exploring their transcriptome, proteome, and metabolome data. Our goals of this project are:

1. Construct sorghum stem cell-type specific gene regulatory network and advance our knowledge of gene regulatory mechanism and stem function on the scale of individual cell types. To implement such study on the single cell type scale, we develop a customized sweet sorghum cell type separation protocol.

2. Exploit this information to facilitate sorghum molecular engineering to improve plant desirable properties, especially on the aspects of either reducing degradation difficulty of cellulosic biomass, or converting non-structural carbohydrate to more energy-intense substance, like triacylglycerol.

Abstract Text:

Sorghum bicolor is currently an important source of food, forage, and feedstock for bioenergy. Due to its advantage conferred by C_4 photosynthesis, it will play a more critical role in the foreseeable future where population growth and environmental pressure will demand higher grain production and more sustainable bioenergy. Many studies have explored sorghum genome information and there already exist abundant resources of different sorghum genotype genome sequence data. However, the majority of studies related to gene expression and regulation were conducted on the scale of plant tissue type, and there is no study investigating this information with a finer resolution, like single cell or single cell type. In this study, we focus on stem cell-type specific transcriptome data at vegetative and reproductive stage, which could advance our knowledge of sorghum stem biology on the finer scale over time and space.

Transcriptome data of different stem cell types were achieved by laser capture microdissection (LCM) followed by RNA-Seq. Differentially expressed genes (DEGs) between two growing

stages were identified by *edgeR* (Robinson et al., 2010) pipeline and cell-type specific genes at each growing stage were identified by combining Tau (τ) index (Kryuchkova-Mostacci et al., 2017) and Wilcoxon test. Enriched GO termed were pulled out by *clusterProfiler* (Yu et al., 2012) for genes in each category identified as either DEGs or cell-type specific genes to indicate their specific functions. Finally, gene regulatory network (GRN) for each cell type at each developmental stage was constructed by integrating co-expression strength of cell-type specific genes and their predictive regulatory relationship. The transcriptome data achieved by the protocol developed in this study indicates that different stem cell types had distinct gene expression profiles, which underlay their morphological and physiological differentiation. By comparing two growing stages, we find that most of genes were not differentially expressed over time. These genes may execute basic functions that are indispensable for all the cell types in plant life. Furthermore, enriched GO term for these DEGs indicates that different cell type had both common and unique functions at each growing stage. Tau, combined with Wilcoxon test, successfully identified cell-type specific genes at each growing stage. The enriched functions of these cell type specific genes were consistent with their characteristic physiological functions, which added another layer of confidence for the precision and feasibility of this single cell separation protocol. Gene regulatory network of pith parenchyma is of great interest among all the cell types since this is the stem location where plant stores most of non-structural carbohydrate after anthesis and it focuses most engineering attention from the perspective of either converting cellulosic biomass to bioproducts or transforming carbohydrate to triacylglycerol (TAG) that have more than twofold energy than that of carbohydrate. Transcription factor hubs, which were identified by node outgoing degree number, had a large control over GRN, indicating these genes are potentially most influential regulators in pith parenchyma.

References/Publications

1. Kryuchkova-Mostacci, N., & Robinson-Rechavi, M. (2017). A benchmark of gene expression tissue-specificity metrics. Briefings in bioinformatics, 18(2), 205-214.
2. Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 26(1), 139-140.
3. Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: a journal of integrative biology, 16(5), 284-287.

Funding Statement:

This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.