A multi-omics approach to understand cotton fiber development.

Youngwoo Lee[1,2], Pengcheng Yang[3], Corrinne E Grover[4], Sivakumar Swaminathan[5], Heena Rani[1,2], Eileen Mallery[1,2], Jun Xie[3], Olga A Zabotina[5], Jonathan F Wendel[4], Daniel B. Szymanski[1,2]
[1]Center for Plant Biology; [2]Department of Botany and Plant Pathology; [3]Department of Statistics, Purdue University; [4]Department of Ecology, Evolution and Organismal Biology; [5]Roy J Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University

The highly polarized cotton fiber cell that emerges from the seed coat surface is the foundation of a multi-billion-dollar textile industry. Important traits such as fiber diameter, length, and strength are defined by the growth of individual cells. At present, the ability to predict and control fiber traits is limited by our lack of understanding regarding the primary controls governing the rate, duration, and patterns of cell growth. To address this, we are conducting detailed multi-omics developmental analyses of the gene expression patterns, protein complexes, and cell wall features that underlie cotton fiber development and integrate these data to create predictive control networks. In our project, we collected single cell-type fiber samples daily from 5 to 24 Days Post-Anthesis (DPA). This novel, dense sampling regime of 20-time points captures most of the elongation phase and the transition to secondary cell wall synthesis. We then integrate multi-omics datasets to build a knowledge base that will facilitate the genetic engineering of cotton crops with improved traits. To gain insights into the compartmentalized functions of proteins in cotton fiber cells, we developed a quantitative shotgun proteomic analysis pipeline. We fractionated the fiber proteome into apoplast (APOT), total membrane-associated (P200), and cytosolic (S200) fractions. Subsequently, proteins were identified and quantified using protein mass spectrometry, and then their localizations were predicted using bioinformatics approaches in conjunction with conventional statistical methods. This method effectively identified true apoplastic proteins, and many of them are shown to reside in extracellular vesicles. The protein data also were projected onto biochemical pathways, and these analyses suggest a biochemical functionality of soluble and membrane associated apoplastic proteins. Further, we applied this method to generate abundance profiles across fiber development for the three subcellular fractions. Data filtering and machine learning methods were created to deal with false negative values and identify reliable protein abundance profiles. The protein expression groups and their membership provide important clues about systems-level functions during developmental transitions and insights into post-transcriptional control. We are organizing the data from our multi- omics approach so that it is findable and useful to the community with the goal of accelerating the genetic engineering of cotton fiber traits.