Transcriptomic analyses of bulk and single cell *Chalmydomonas* RNA-seq data reveal new gene functions and cell state heterogeneity

Patrice Salome, Sabeeha Merchant, Matteo Pellegrini

UCLA-DOE Institute for Genomics and Proteomics

Project Goals: Research in the UCLA-DOE Institute for Genomics and Proteomics includes major efforts in the elucidation of algal biology using genomics approaches.

The unicellular green alga *Chlamydomonas reinhardtii* is a choice reference system for the study of photosynthesis, lipid and starch metabolism and metal homeostasis. It is also a valuable model for the study of algal biofuel production. Despite decades of research, the function of thousands of genes remains largely unknown, and new approaches are needed to categorically assign genes to cellular pathways. Growing collections of transcriptome and proteome data now allow a systematic approach based on integrative co-expression analysis. We have used a dataset comprising 518 RNAseq samples derived from 58 independent experiments to identify potential co-expression relationships. While on a global scale random gene-pairs are not co-expressed, the determination of the co-expression profile of gene lists (manually curated from the literature) revealed high-confidence candidate genes with roles in cell cycle control, photosynthesis and respiration. Another striking observation was the clustering of nuclear genes largely as a function of their diurnal phase, even after the removal of all RNAseq samples collected over a diurnal cycle. A closer look at the remaining samples uncovered partial but frequent diurnal synchronization, although these samples had been collected from cultures maintained in constant light. We are now also exploring the potential heterogeneity of Chlamydomonas cultures grown in constant light by single-cell RNAseq: we have sequenced ~73,000 cells from two independent experiments representing two genotypes and three biological conditions, which we can identify clearly during visualization of this high-dimension dataset. We are now investigating the sources of heterogeneity within each culture by overlaying informative gene sets from our bulk RNA-seq data, to attempt to characterize diverse cellular states within large populations of cells. In the future, understanding the sources of variability could allow us, for example, to select states that are optimal for biofuel production.

This material is based upon work supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research program under Award Number DE-FC02-02ER63421