BERKELEY LAB
Bringing Science Solutions to the World

U.S. DEPARTMENT OF ENERGY

DOE Bioenergy Technologies Office (BETO)
2019 Project Peer Review

Genetic Blueprint of Microalgae Carbon Productivity

Igor Grigoriev (LBNL),
Shawn Starkenburg (LANL)
Collaborators:
Michael Guarnieri (NREL), Michael Huesemann (PNNL)

March 5, 2019

Goal Statement

Develop an algal functional genomics pipeline for production and interpretation of multi-omics measurements from multi-state perturbations experiments to identify gene targets for strain improvement and commercialization at accelerated (50% reduced) development time.

Quad Chart Overview

Timeline

Project start date 10/1/2016

Project end date 9/30/2019

Percent complete ~75%

	Total Costs Pre FY17	FY 17 Costs	FY 18 Costs	Total Planned Funding (FY 19-Project End Date)
DOE Funded	0	$350K	$350K	$400K
Project Cost Share*	0	LBNL (35%) LANL (65%)	LBNL (35%) LANL (65%)	LBNL (50%) LANL (50%)

•Partners: If multiple DOE recipients are involved in the project, please list level of involvement, expressed as percentages of project funding from FY 17-18. [(i.e. NREL (70%); INL (30%)]

Barriers addressed

Aft-C. Biomass Genetics and Development: The productivity and robustness of algae strains against temperature and seasonality could be improved by selection, screening, breeding, mixing cultures, and/or genetic engineering. These approaches require extensive ecological, genetic, and biochemical information.

Objective

Develop a functional genomics pipeline to baseline, standardize, and accelerate strain improvement and commercialization

End of Project Goal

Using the newly developed pipeline identify gene targets for strain improvement in *Scenedesmus sp.* NREL 46B-D3(NREL) within the first 2 years and then in *Monoraphidium minutum* 26B-AM(PNNL) within one (3rd) year, achieving target identification at 50% time reduction

BERKELEY LAB

1. Project Overview

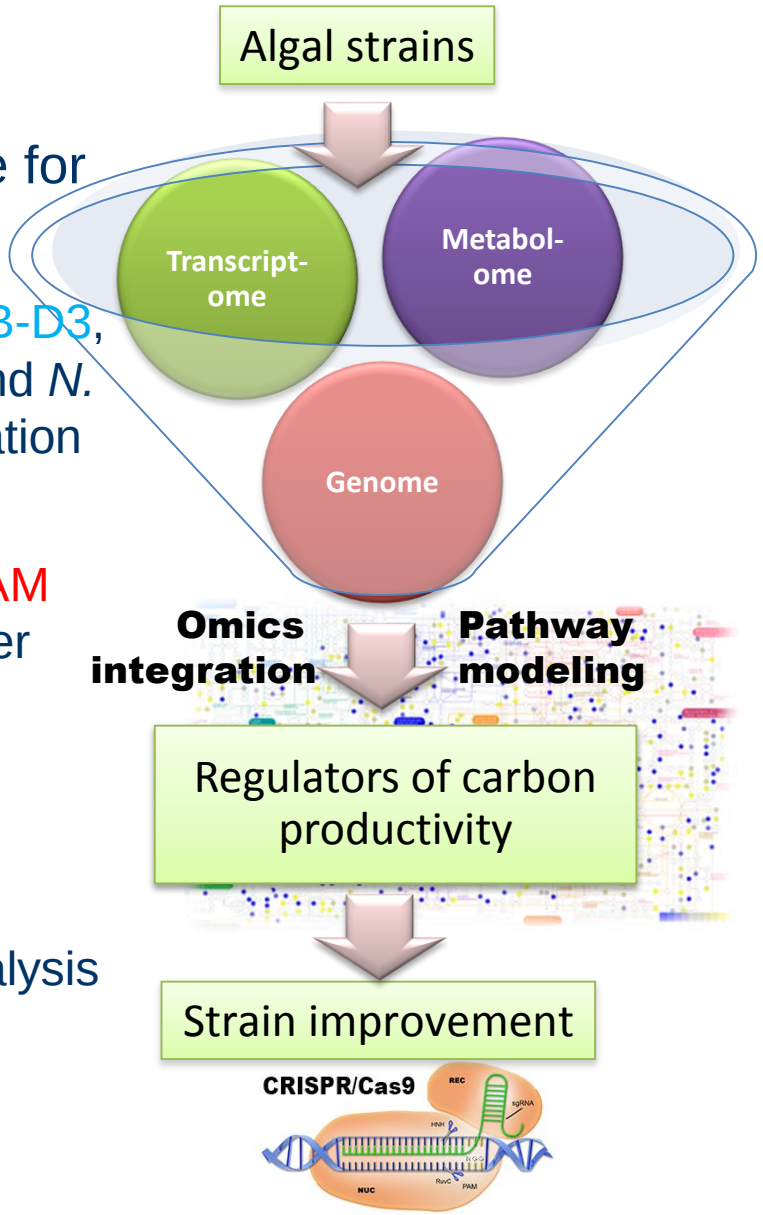
Goal: Develop a functional genomics pipeline for strain improvement & commercialization

1st target (FY17-18): *Scenedesmus sp.* NREL 46B-D3, which outperforms both *Nannochloropsis salina* and *N. oceanica* in lab-based and outdoor summer cultivation (NREL, ATP^3)

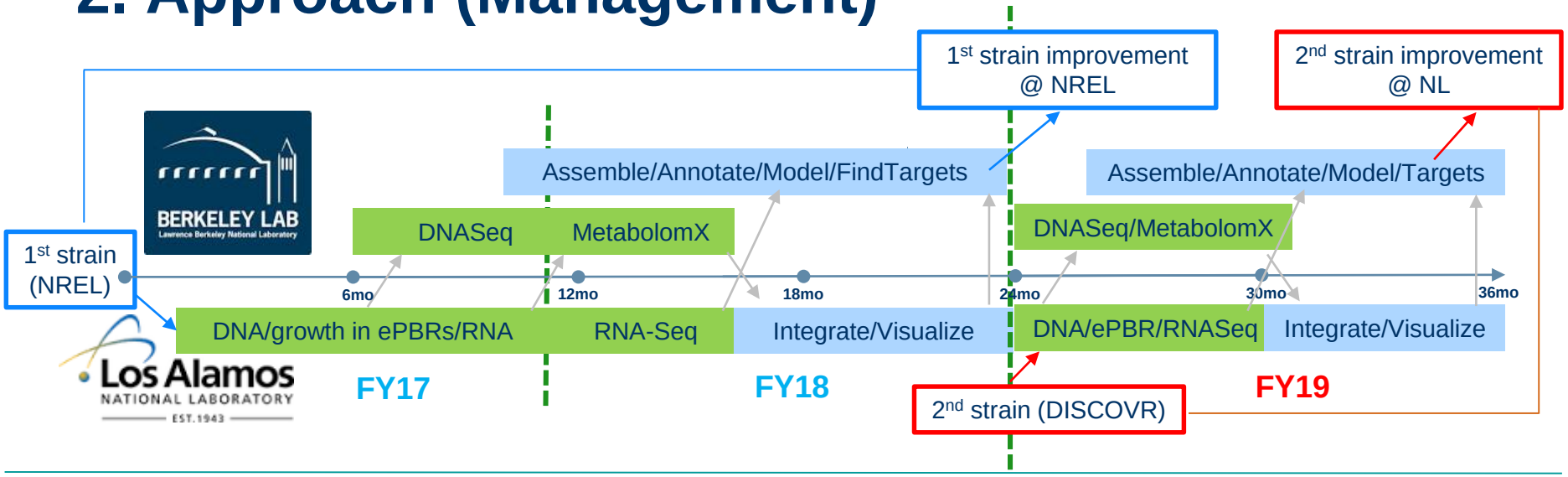
2nd target (FY19): *Monoraphidium minutum* 26B-AM from DISCOVR, a top performer in lab-based winter cultivation and in outdoor growth in AZ.

Objectives:

- Produce high quality sequence, assembly, and annotation of a target genome
- Generate omics data for functional genome analysis and modeling
- Identify gene targets for strain mutagenesis to improve biomass production rates



2. Approach (Management)



LBNL/LANL: Multi-omics data production, integration, analysis & target identification for two algal strains (1st - complete, 2nd - in progress)

Scenedesmus sp. NREL 46B-D3: using genes identified in this study ***NREL*** preforms targeted mutagenesis (WBS *1.3.2.102*; Guarnieri, PI)

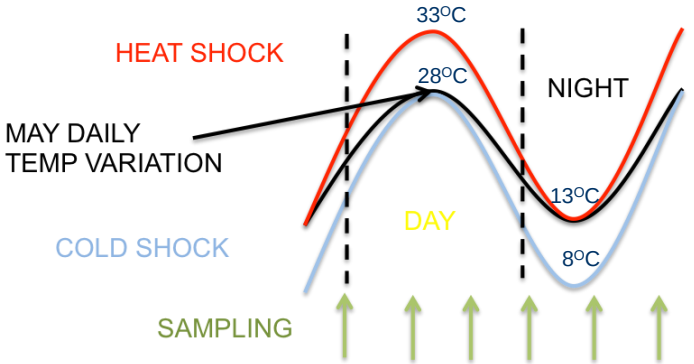
Monoraphidium minutum 26B-AM: ***PNNL*** identified and delivered the DISCOVR strain (WBS 1.3.2.502; Huesemann, PI) to ***LANL*** to produce DNA and RNA, which are in library prep for sequencing at ***JGI/LBNL***

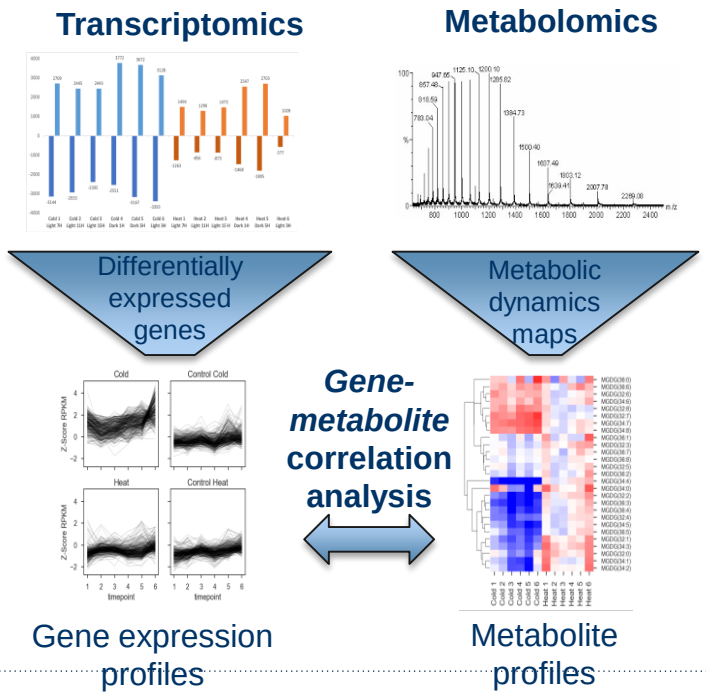
1. Approach: Temperature Perturbations

Algal strains are grown under 3 conditions (6 time points) in ePBRs (LANL):

- Informed by the growth conditions of outdoor cultivation studies at Mesa, AZ
- To compare the growth response to two perturbation simulations: low and a high temperature perturbations
- To determine the ability to tolerate suboptimal environmental conditions at the 'cross over' point of the crop rotation (spring/fall)

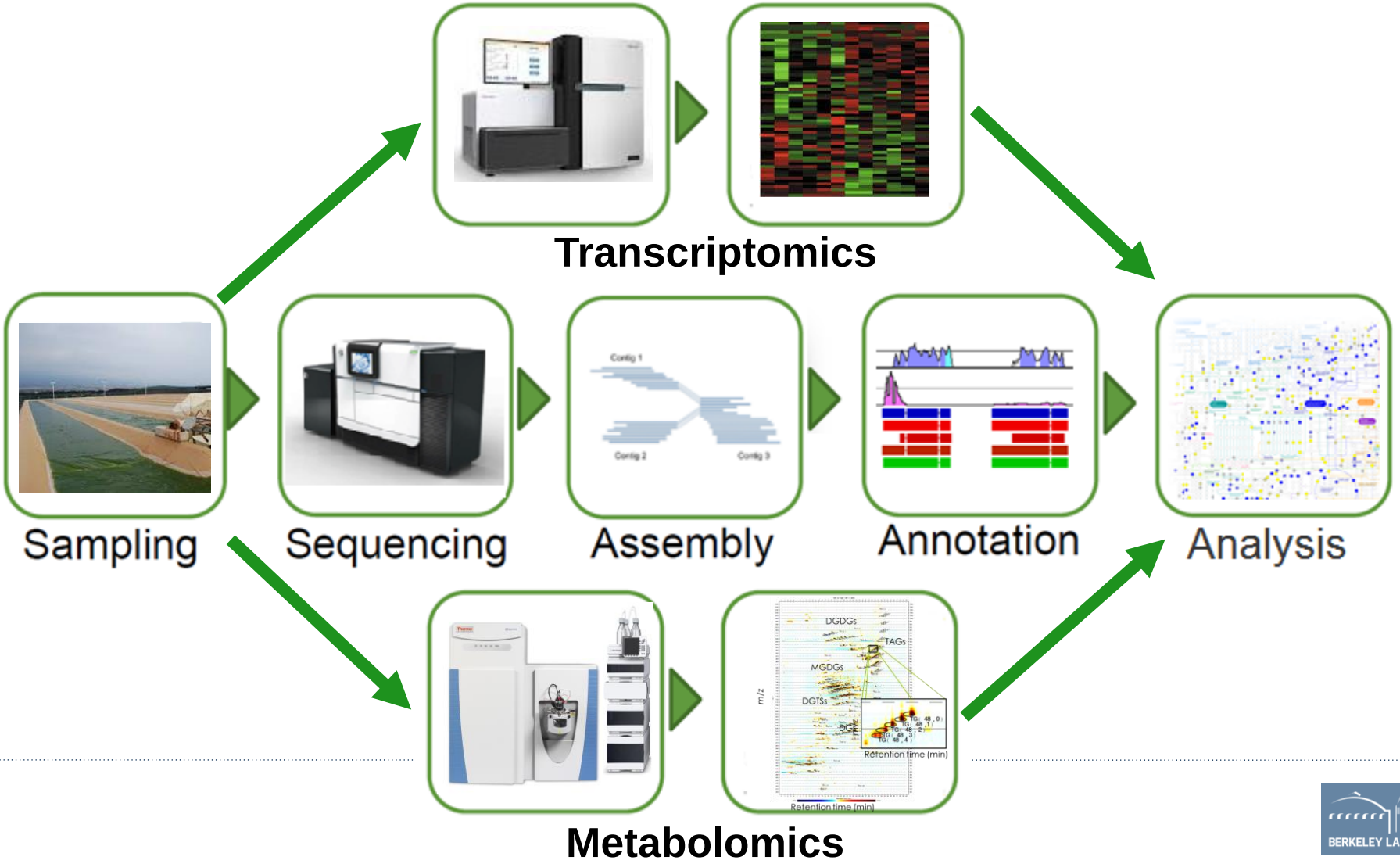
Integrated multi-omics data enable metabolic modeling, gene network analysis, and gene-metabolite correlation analysis to identify gene targets for strain improvement





2. Approach: Functional Genomics Pipeline

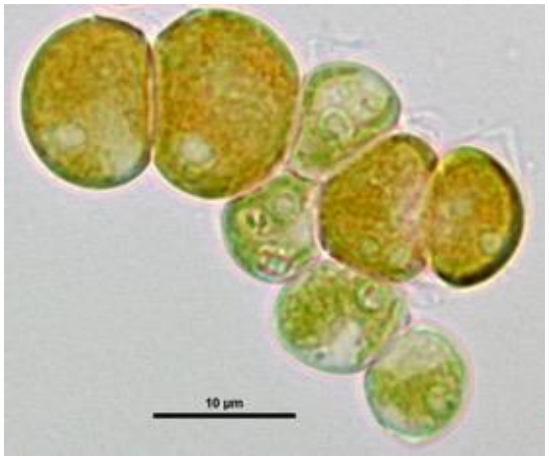
Genome sequencing, multi-omics profiling of perturbed system, metabolic modeling and analysis of integrated data to identify gene targets for strain improvement

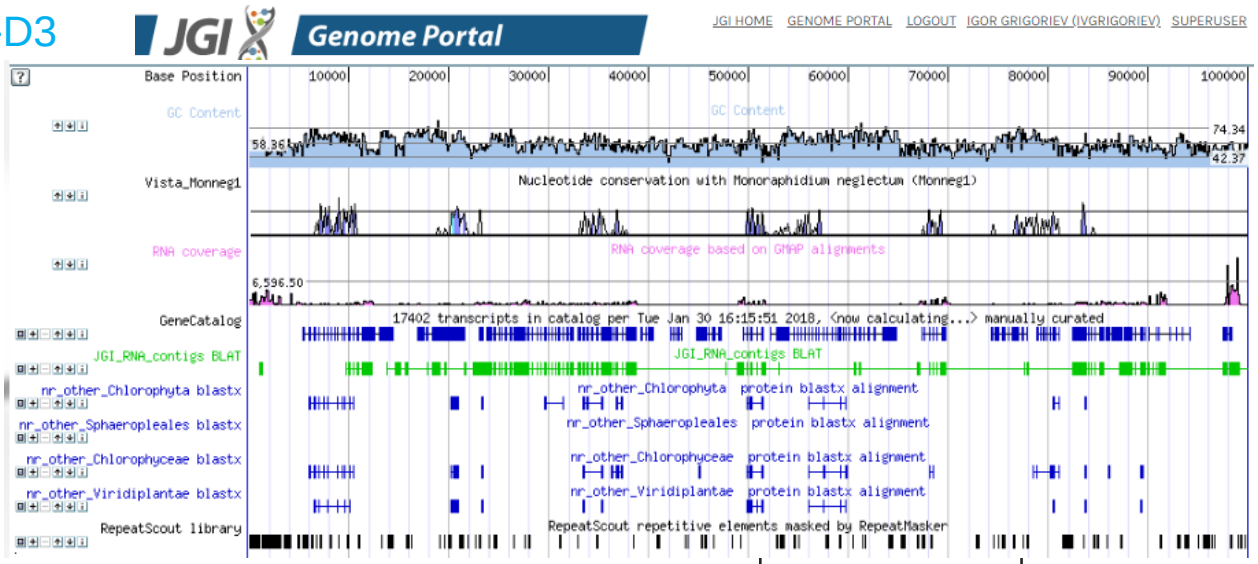


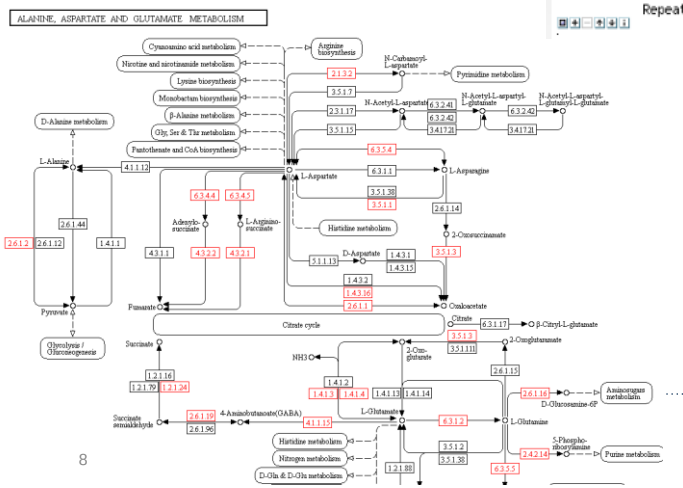
3. Results: Sequenced and Annotated Genome

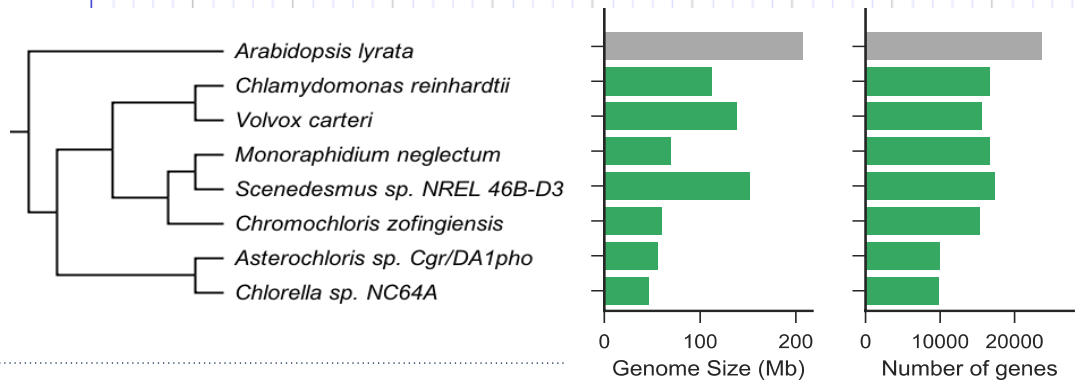
~150Mbp PacBio assembly of NREL 46B-D3 in 2,685 contigs; 17,399 genes predicted; 97% complete (CEGMA/RNA) – *a blueprint to improve strains*

Scenedesmus sp. NREL 46B-D3









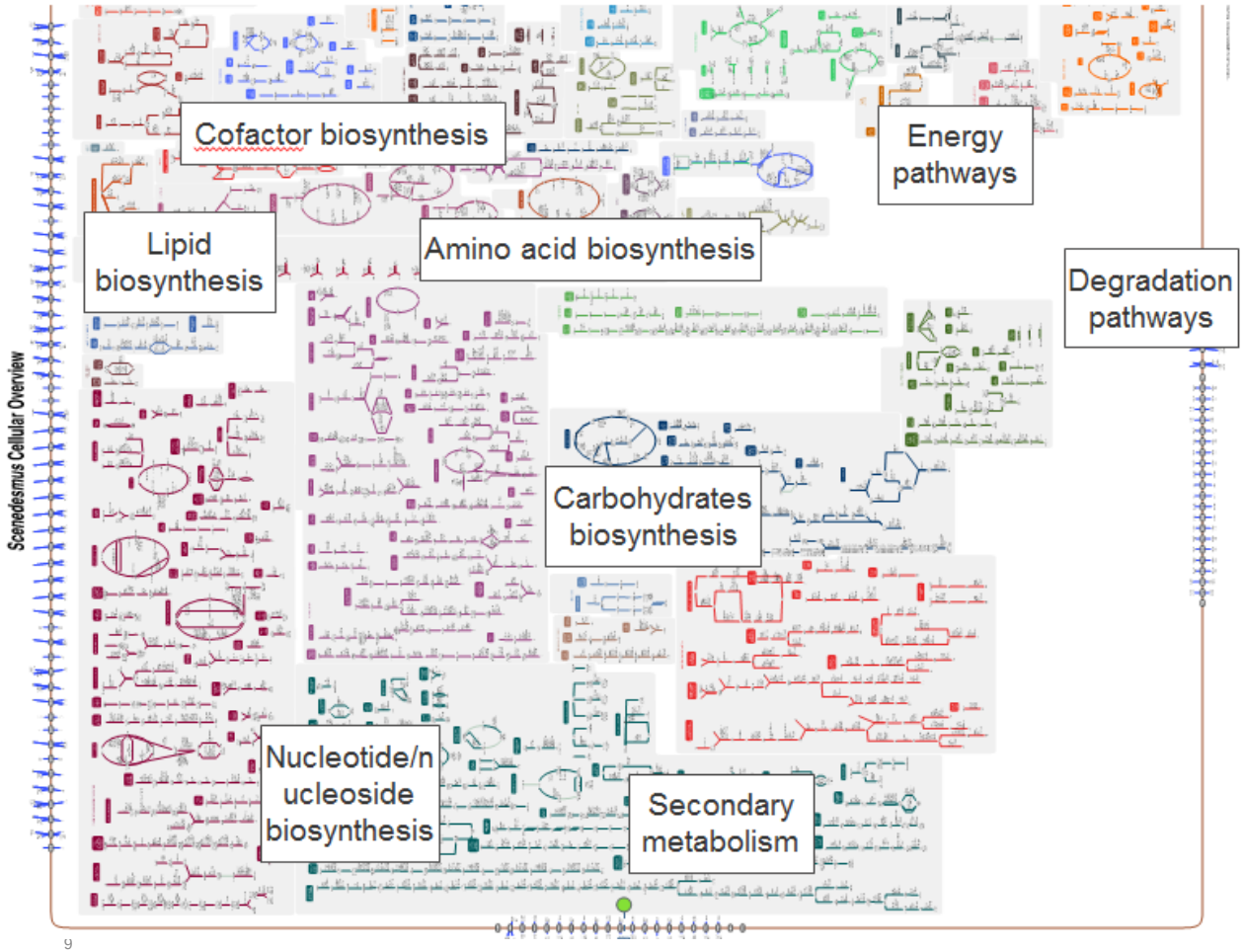
BERKELEY LAB

3. Results: Metabolic Modeling

17399 gene models
6291 GO annotations
4529 KEGG/PRIAM annotations
8084 KOG annotations

→

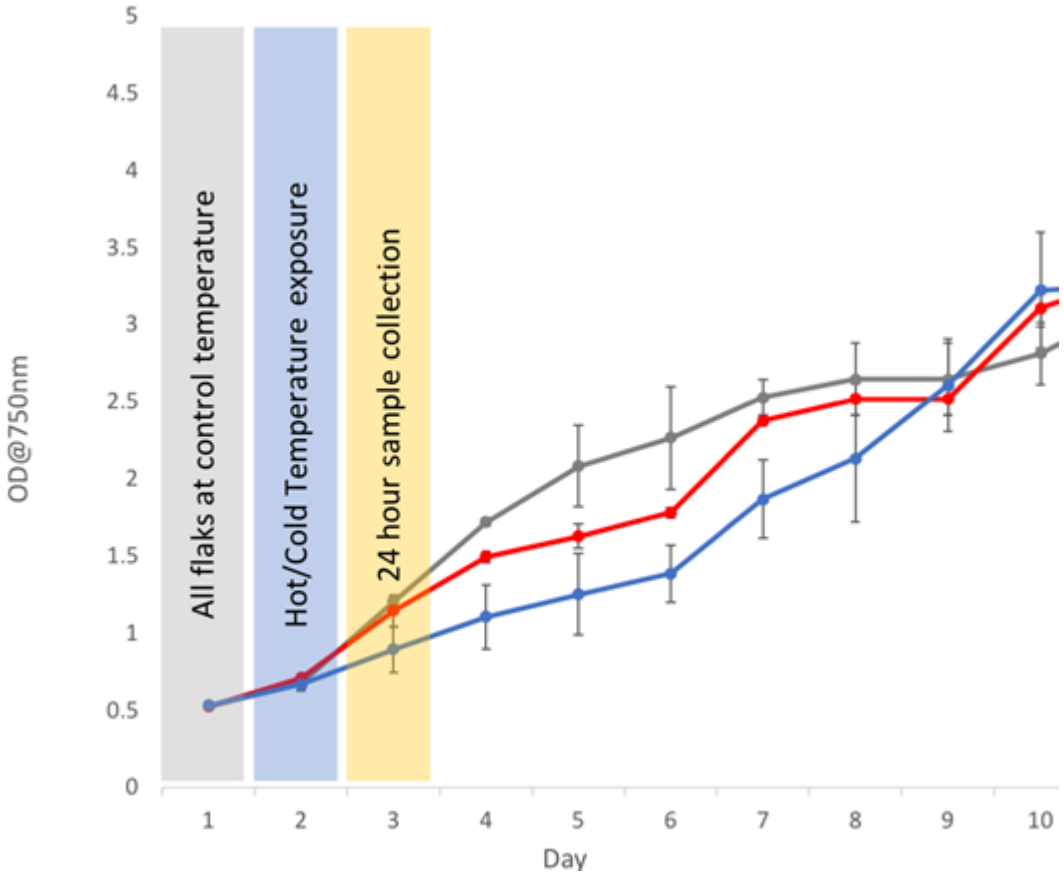
Scenedesmus **model**
2319 metabolic reactions
3009 metabolic genes



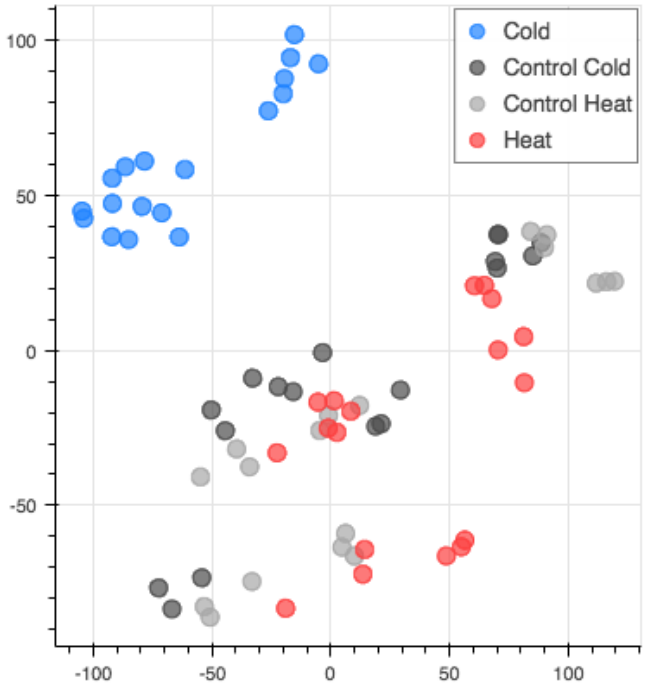
NREL 46B-D3 metabolic model constructed using *PathwayTools* and manually curated.

BERKELEY LAB

3. Results: Growth Response to Temperature Shock

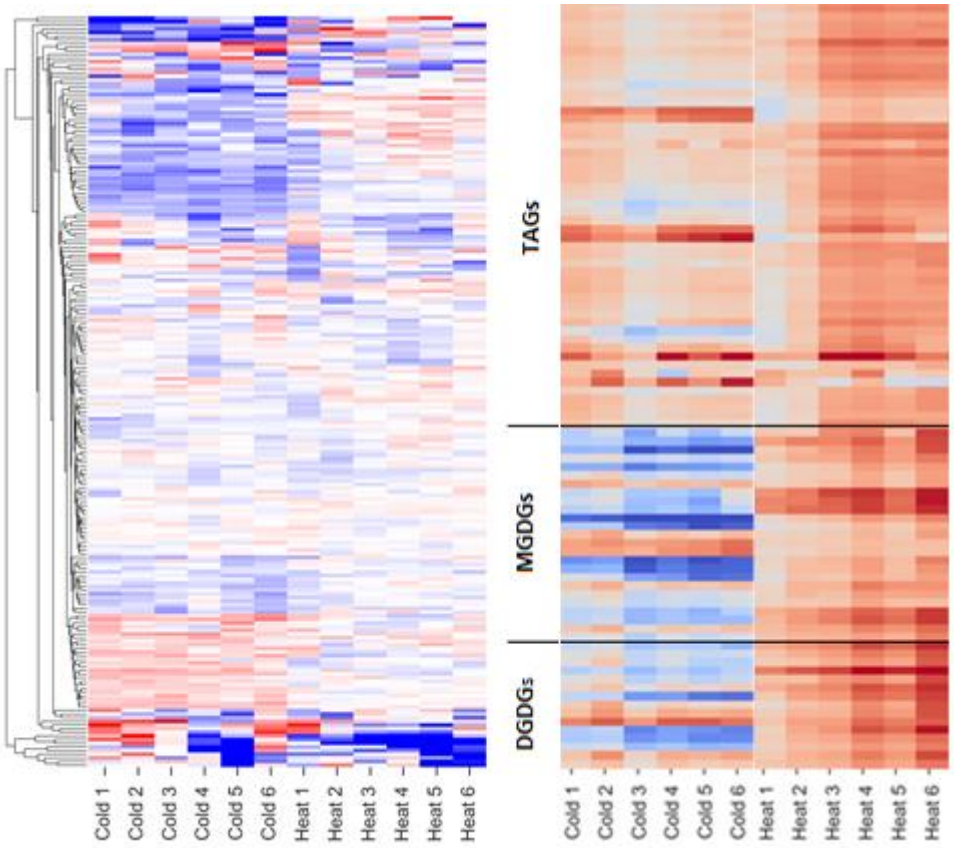


Cold samples have distinct expression profiles



BERKELEY LAB

3. Results: Multi-omics Analysis



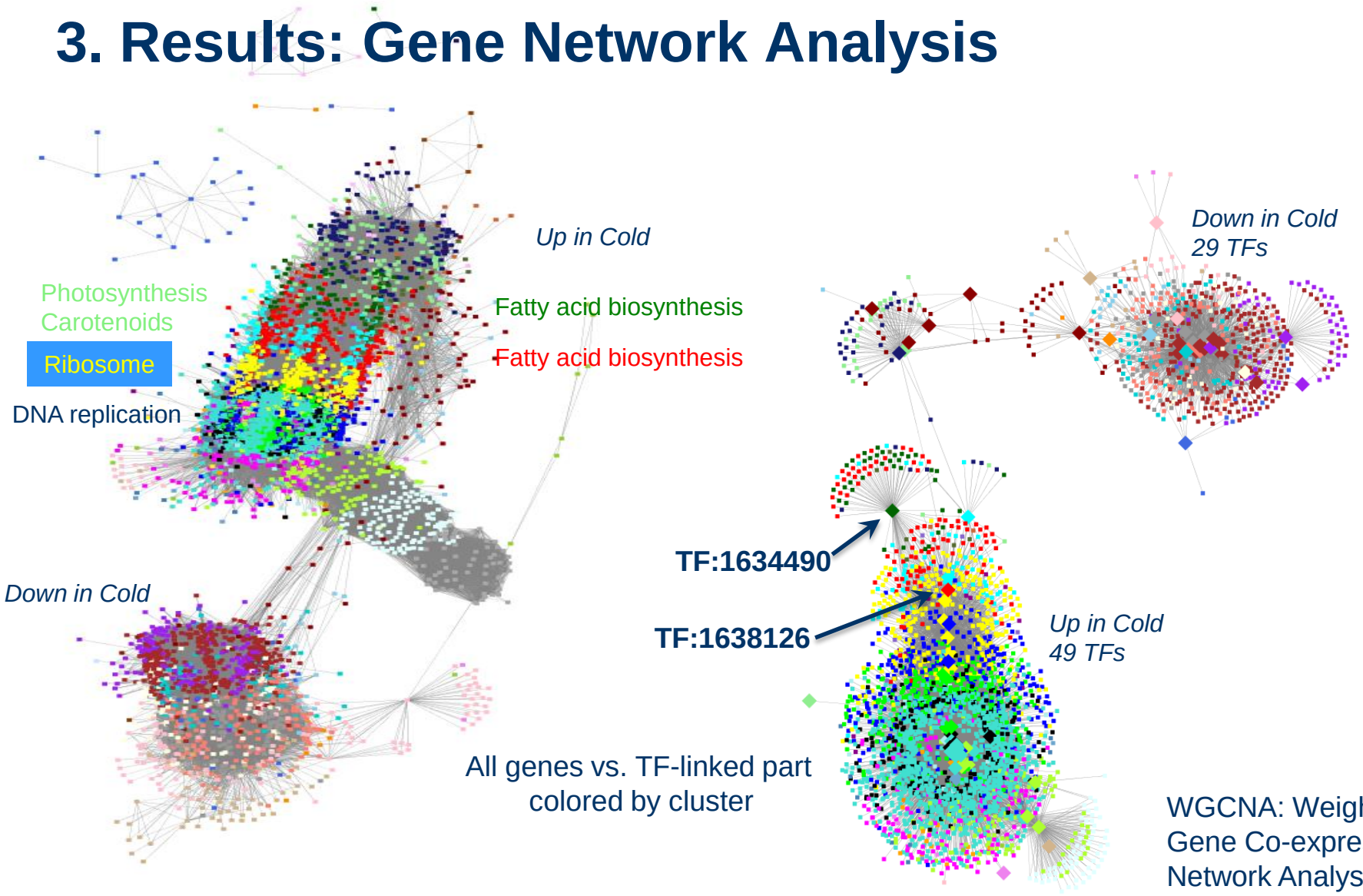
TF transcriptomics Lipidomics

Changes in both gene expression and metabolite production are more dramatic during cold vs. heat shock.

During cold shock large groups of transcription factors (left panel) are up and down regulated as well as MGDGs, DGDGs, some TAGs (right panel) and amino acids.

TAG: Triacylglycerol
MGDG: Monogalactosyldiacylglycerol
DGDG: Digalactosyldiacylglycerol

3. Results: Gene Network Analysis



All genes vs. TF-linked part colored by cluster

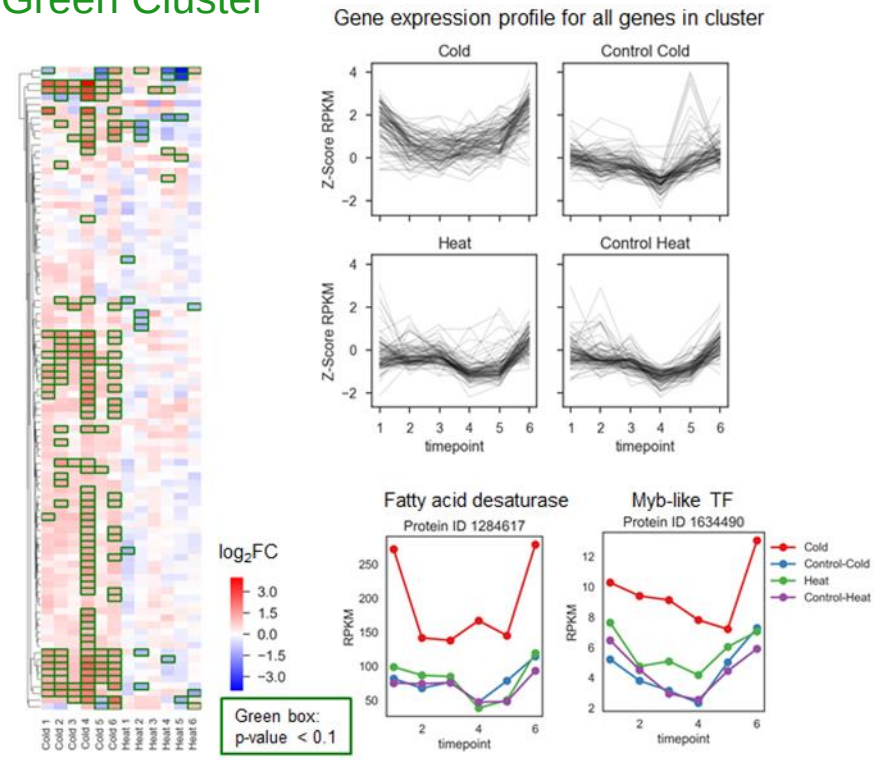
WGCNA: Weighted Gene Co-expression Network Analysis

5,511 genes & 1M edges after removal of low expressed and poorly correlated genes (left) and a subset 3,502 genes correlated w/ TF (right)

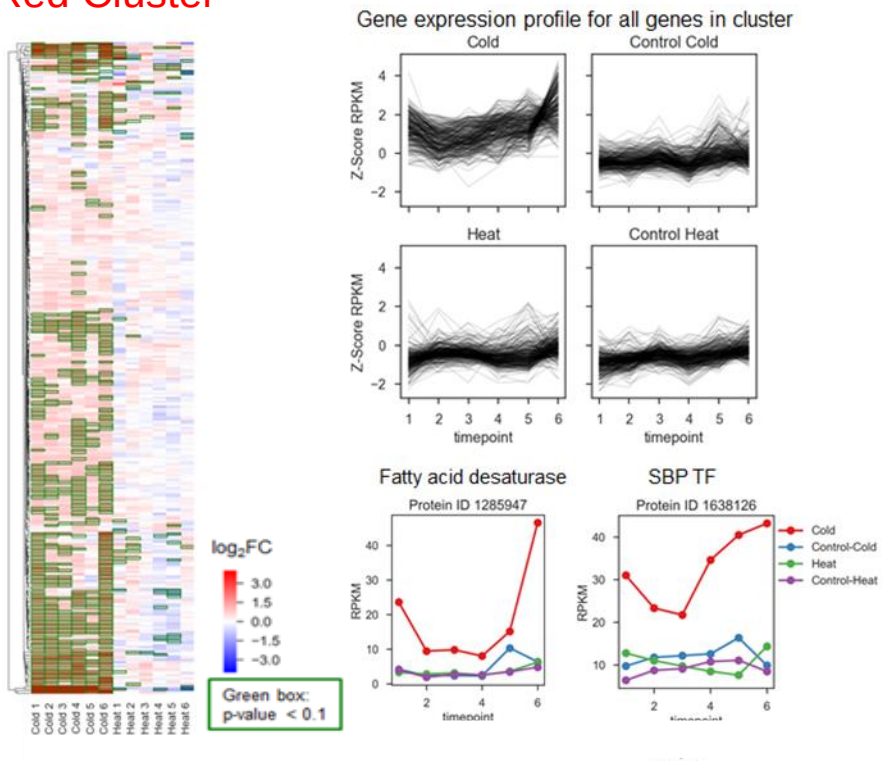
BERKELEY LAB

3. Results: Fatty Acid Desaturases and TFs

Green Cluster

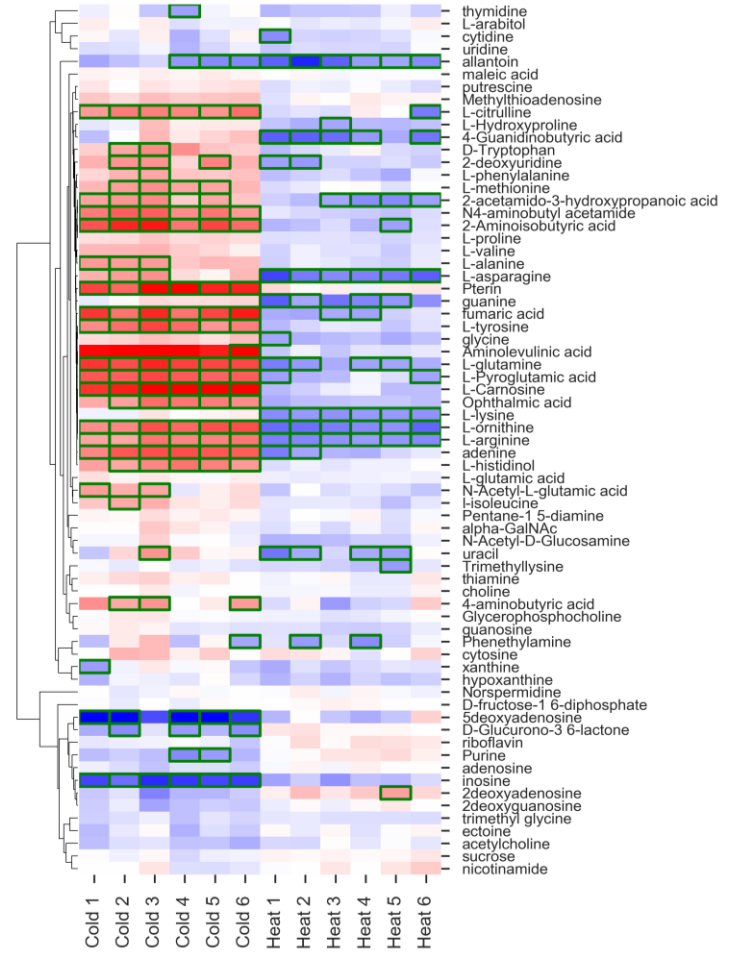


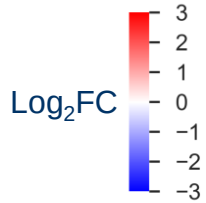
Red Cluster



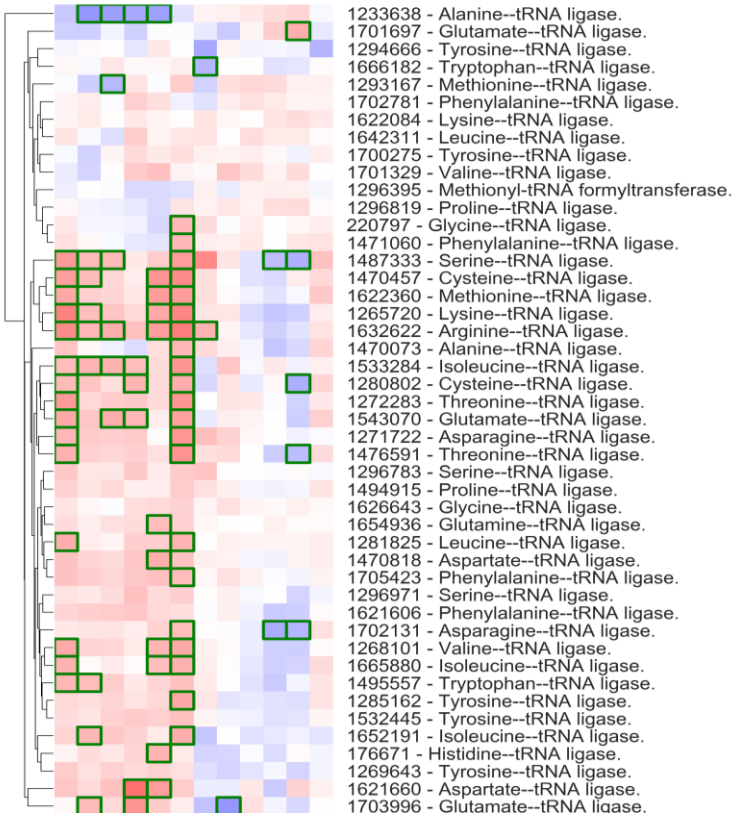
- Two clusters enriched for GeneOntology term “fatty acid biosynthetic process”
- Contain Fatty Acid Desaturases (FAD) 1284617 & 1285947 upregulated in cold
- Transcription factors 1634490(Myb-like) and 1638126(SBP) co-expressed w/ FAD
- FAD adds double bonds, alters membrane lipid profile, proposed as a cold adaptation strategy in some plants & algae, and affects biofuel composition

3. Results: Amino Acid Biosynthesis in Cold Stress

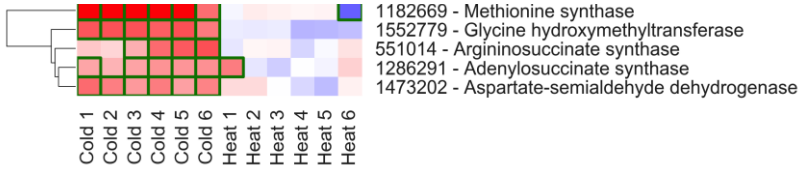




Green box:
p-value < 0.1



Enzymes related to amino acid biosynthesis



Polar Metabolites:

- Amino acids increase in cold stress
- Decrease in Arg, Lys, Asn, Gln, ornithine in heat stress

Cold stress-specific increased expression of:

- Amino acid biosynthesis genes
- Aminoacyl-tRNA ligases

BERKELEY LAB

4. Relevance

Challenge: Rotation of 'seasonal' strains adapted to various environmental conditions is required to maximize/stabilize biomass production throughout the year. Improving the productivity and robustness of algal strains against perturbations requires extensive advanced genetic, genomic, and molecular biology tools, which are currently lacking for most algal species.

Goal: Develop a functional genomics pipeline for production and interpretation of multi-omics measurements from multi-state perturbations experiments to identify gene targets for strain improvement and commercialization at accelerated development time.

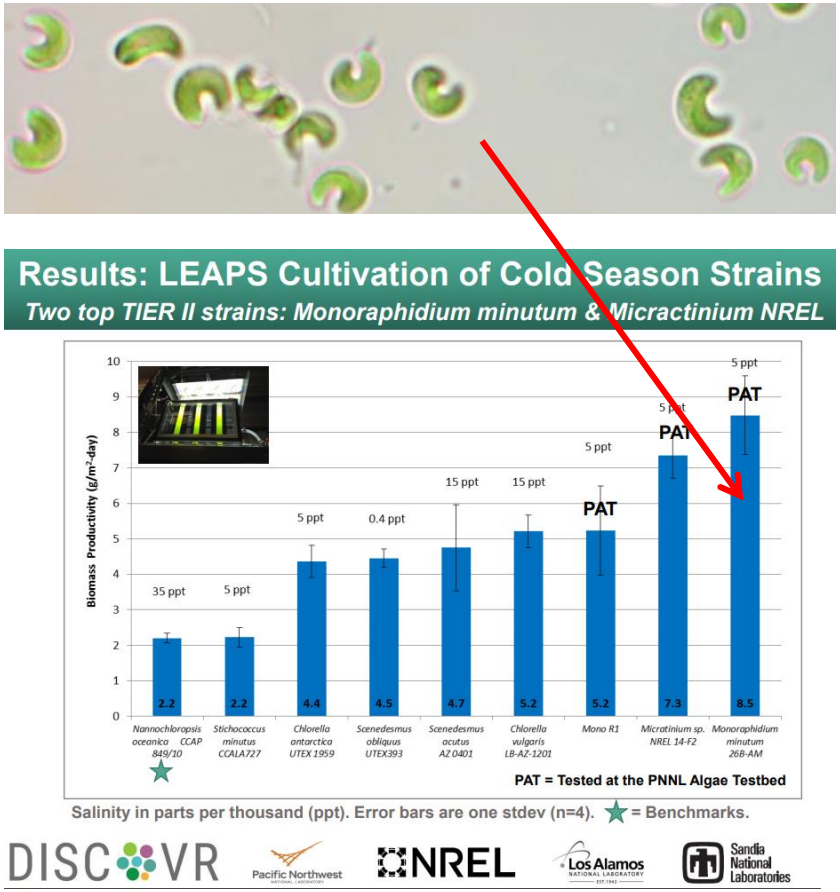
This is directly connected to BETO Aft-C (Biomass Genetics & Development) to enable improving productivity of algal strains by selection, screening, breeding, mixing cultures, and/or genetic engineering and to MYPP performance goal to increase the summer seasonal areal productivity (from 13.3 g/m^2/d in 2016 to 25 g/m^2/d in 2025).

Outcomes: Algal strains with sequenced and annotated genomes equipped with customized modeling tools will advance the state of technology and accelerate strain development and commercial viability of biomass and biofuels.

5. Future Work: FY19 deliverables

- Sequence and annotate the genome of *Monoraphidium minutum* 26B-AM
- Produce and integrate *M.minutum* 26B-AM transcriptomes and metabolomes for modeling and target identification
- Identify *M.minutum* 26B-AM gene targets to follow up with strain improvement at PNNL

Monoraphidium minutum 26B-AM



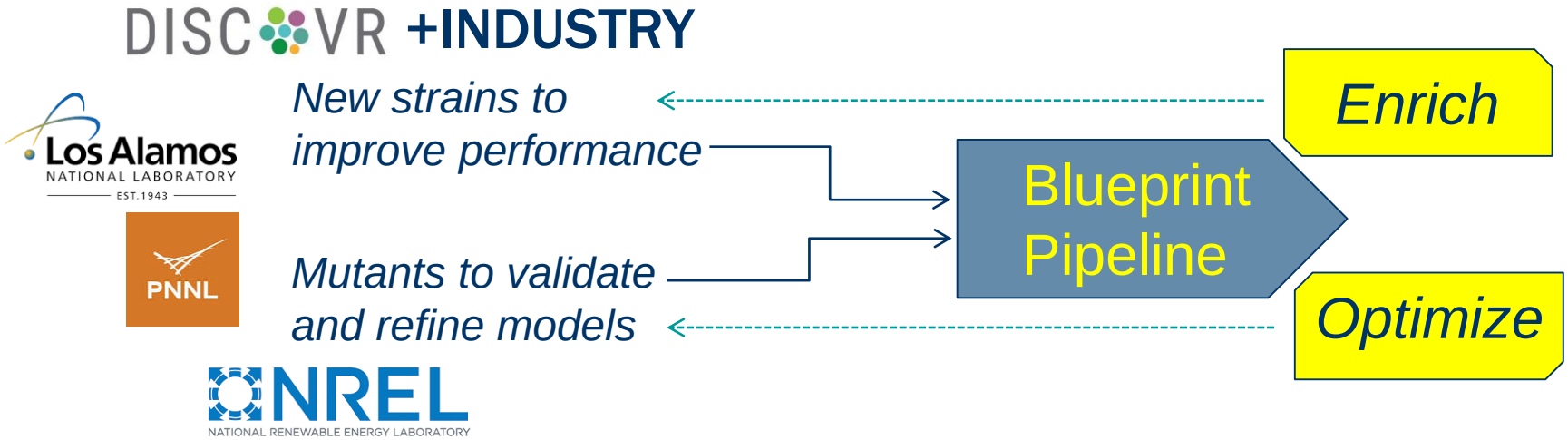
Courtesy of Michael Huesemann

5. Future Work: FY19 Project Milestones

Lab	Performance Measure	Completion	
LANL	*M.minutum* strain acquired from PNNL; high molecular weight genomic DNA obtained and transferred to LBNL for sequencing	11/1/2018	✓
LBNL	DNA quality assessed to initiate sequencing	12/31/2018	✓
LANL	2 perturbation studies with *M. minutum* in triplicate exposed to a 5^{o}C temperature challenge at the day time temperature maxima and 0.5X seawater salinity challenge	12/31/2018	✓
LBNL	Genome sequenced using PacBio and assembled	3/31/2019	
LBNL	Metabolomics data produced for 12 samples with preliminary analysis demonstrating fold changes (LBNL)	3/31/2019	
LANL	12 transcriptomes sequenced and analyzed	3/31/2019	
LBNL/ LANL	Genome annotated and integrated w/ multi-omics data	6/30/2019	
LBNL/ LANL	Initial metabolic model built and improved using multi-omics analysis to identify at least two target genes (over or under expression) for strain improvement of *M. minutum*	9/30/2019	

5. Future Work: After FY19

Objective: optimize, enrich, and apply the Blueprint functional genomics pipeline to produce genomes, multi-omes, and customized modeling/analysis tools for additional BETO approved strains



Genetic Engineering work can be completed at NREL, LBNL as needed based on Blueprint targets identified.

BERKELEY LAB

Summary

1. **Goal**: Develop a functional genomics pipeline for production and interpretation of multi-omics measurements from multi-state perturbations experiments to identify gene targets for strain improvement and commercialization at accelerated (50% reduced) development time.
2. **Approach**: Genome sequencing and multi-omics profiling during temperature perturbation experiments to enable metabolic modeling and gene network analysis
3. **Results:** Gene targets for strain improvement identified based on analysis and modeling with sequenced genome and multi-omics data, automated data production, and custom analyses
4. **Relevance:** Addresses Aft-C barrier (Biomass Genetics and Development) by genome-based strain improvement and MYPP performance goal to increase the summer seasonal areal productivity (from 13.3 $g/m^2/d$ in 2016 to 25 $g/m^2/d$ in 2025).
5. **FY19 Deliverables:** complete the same SOW for 2nd target strain within 1 year (i.e., 2x faster than the 1st target strain)
6. **After FY19:** optimize, enrich, and apply the Pipeline to new strains

BERKELEY LAB