

Chromatin landscaping in algae reveals novel regulation pathway for biofuels production

Chew Yee Ngan¹, Chee-Hong Wong¹, Cindy Choi¹, Abhishek Pratap¹, James Han¹, Chia-Lin Wei¹

¹ Department of Energy Joint Genome Institute // LBNL - Walnut Creek, CA

^aTo whom correspondence may be addressed. E-mail: cyngan@lbl.gov

February 19, 2013

ACKNOWLEDGMENTS:

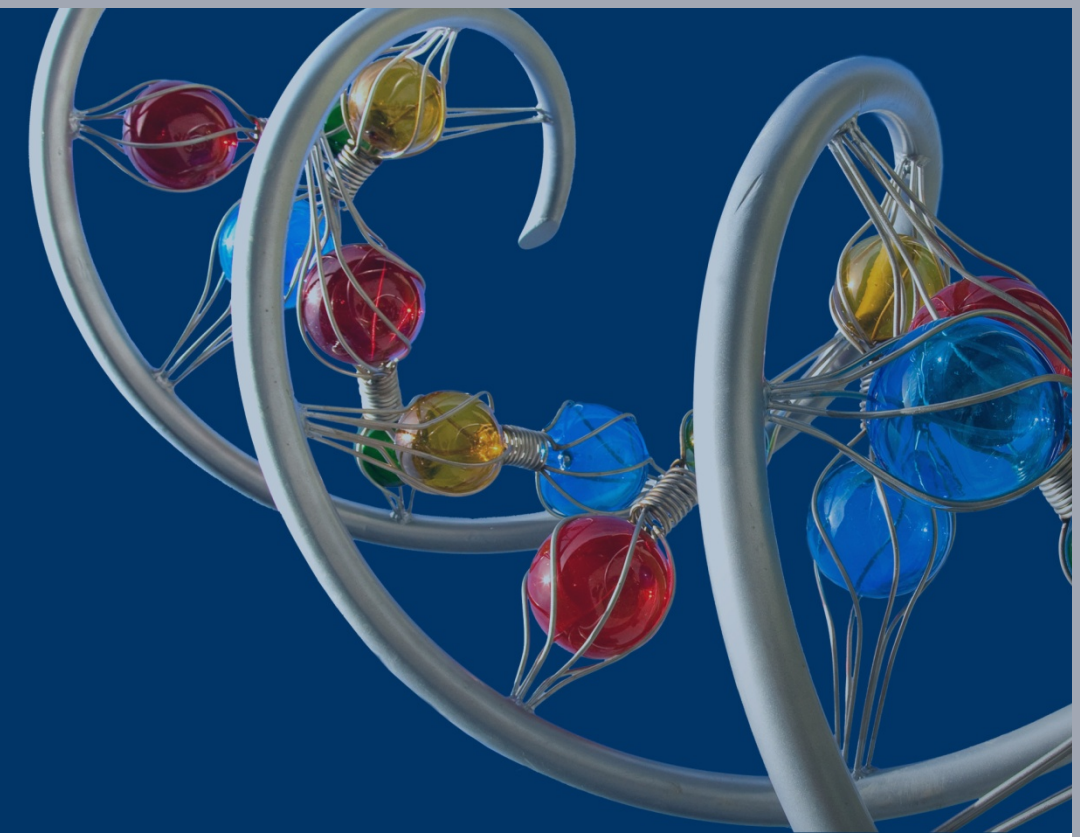
The work conducted by the US Department of Energy (DOE) Joint Genome Institute is supported by the Office of Science of the DOE under Contract Number DE-AC02-05CH11231. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government, or any agency thereof, or the Regents of the University of California.

DISCLAIMER:

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California

Chromatin landscaping in algae reveals novel regulation pathway for biofuels production

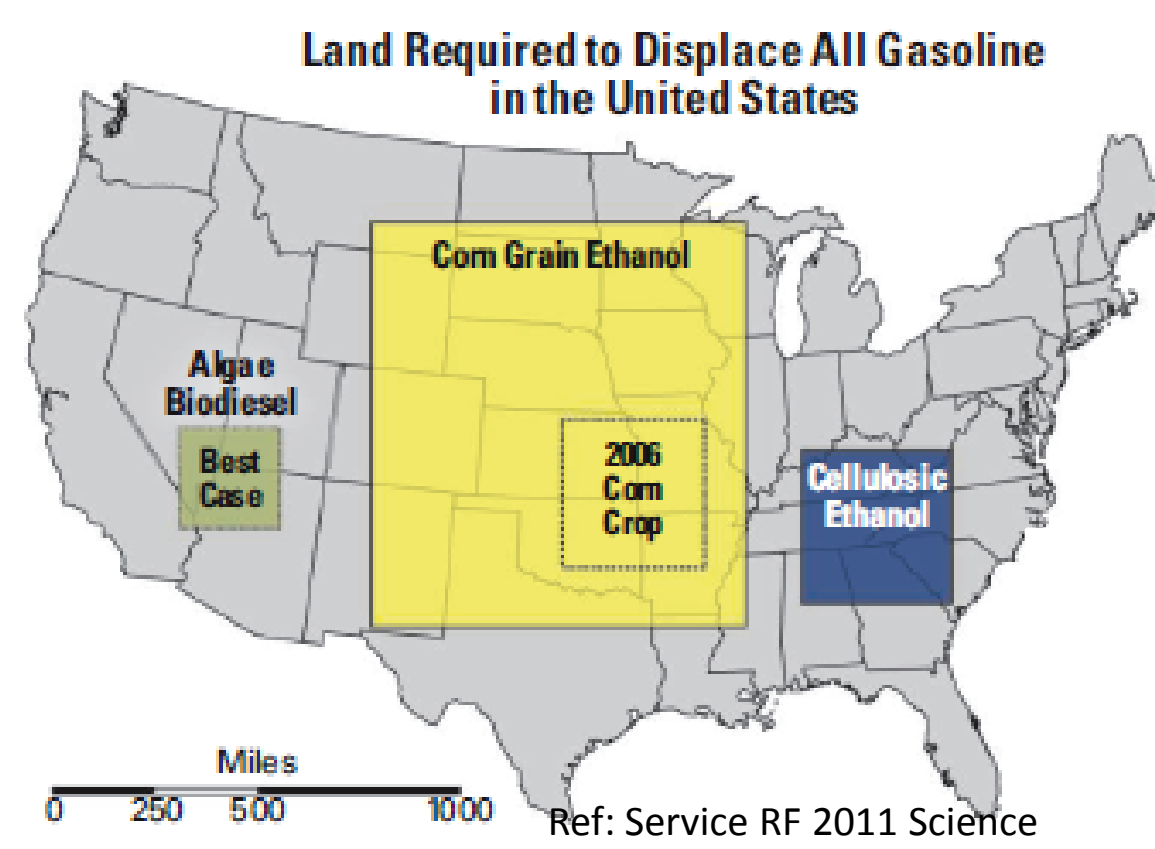
Chew Yee Ngan, Chee-Hong Wong, Cindy Choi, Abhishek Pratap, James Han, Chia-Lin Wei



Summary

The diminishing reserve of fossil fuels calls for the development of biofuels. Biofuels are produced from renewable resources, including photosynthetic organisms, generating clean energy. Microalgae is one of the potential feedstock for biofuels production. It grows easily even in waste water, and poses no competition to agricultural crops for arable land. However, little is known about the algae lipid biosynthetic regulatory mechanisms. Most studies relied on the homology to other plant model organisms, in particular Arabidopsis or through low coverage expression analysis to identify key enzymes. This limits the discovery of new components in the biosynthetic pathways, particularly the genetic regulators and effort to maximize the production efficiency of algal biofuels. Here we report an unprecedented and de novo approach to dissect the algal lipid pathways through disclosing the temporal regulations of chromatin states during lipid biosynthesis. We have generated genome wide chromatin maps in chlamydomonas genome using ChIP-seq targeting 7 histone modifications and RNA polymerase II in a time-series manner throughout conditions activating lipid biosynthesis. To our surprise, the combinatory profiles of histone codes uncovered new regulatory mechanism in gene expression in algae. Coupled with matched RNA-seq data, chromatin changes revealed potential novel regulators and candidate genes involved in the activation of lipid accumulations. Genetic perturbation on these candidate regulators further demonstrated the potential to manipulate the regulatory cascade for lipid synthesis efficiency. Exploring epigenetic landscape in microalgae shown here provides powerful tools needed in improving biofuel production and new technology platform for renewable energy generation, global carbon management, and environmental survey.

Chlamydomonas as model for algal biofuel feedstock



Microalgae potentially produces the highest oil yield with the smallest land area required.

Unique Advantages of Algae as Biofuel Feedstock

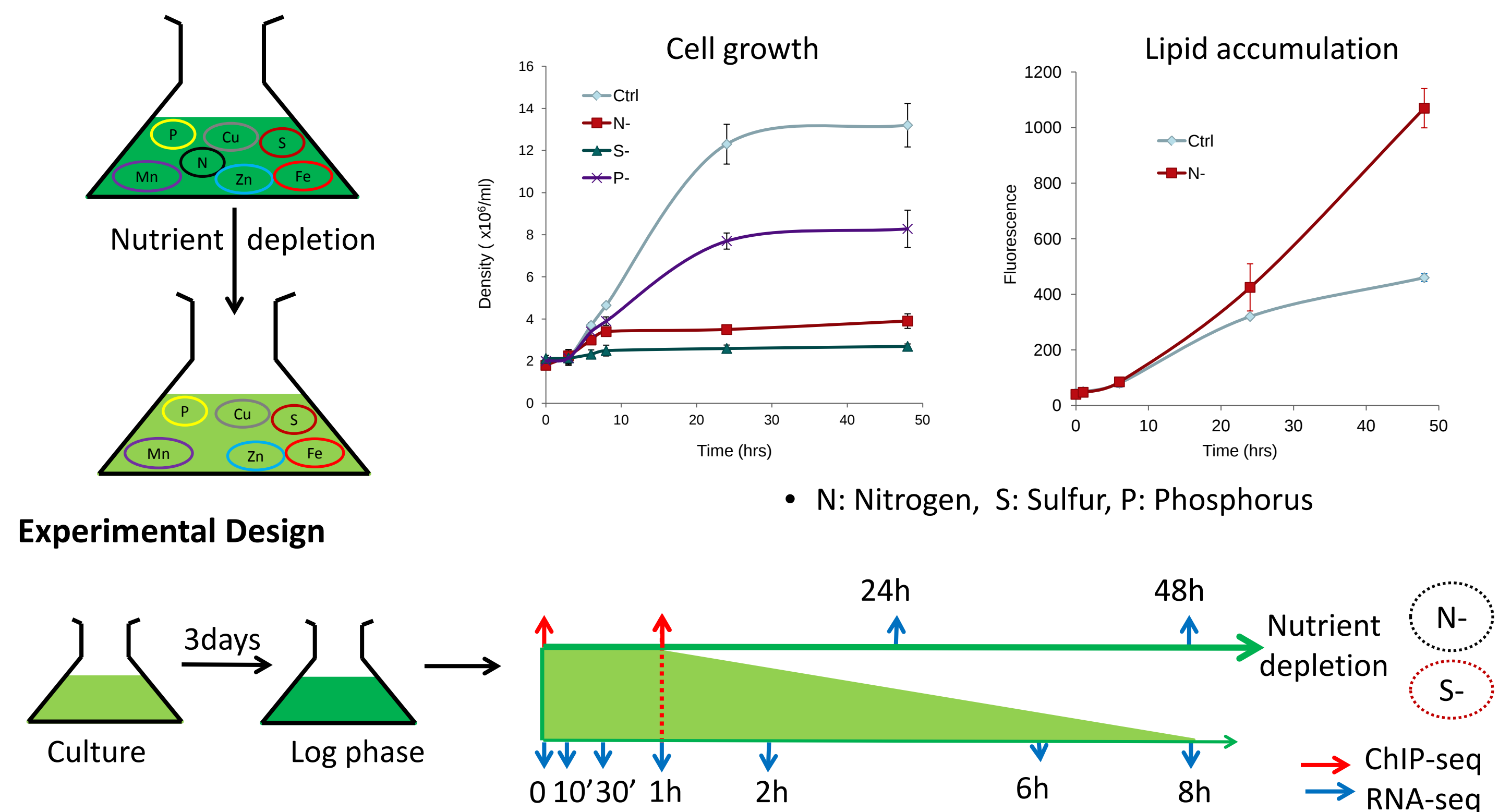
- High area productivity
- Non-seasonal, fast growing
- Minimum competition with agriculture
- Flexibility on water quality
- Renewable energy
- Recycles stationary emissions of CO₂

Ref: DOE National Algal Biofuels Technology Roadmap 2010

Chlamydomonas reinhardtii as a reference model

- Unicellular green algae (Chlorophyta), found all over the world, in soil, fresh water, oceans
- The most widely used laboratory species is *Chlamydomonas reinhardtii*
- Cells can grow on a simple medium of inorganic salts, using photosynthesis to provide energy
- Reference genome available

Lack of nutrients induces lipid accumulation chlamydomonas.

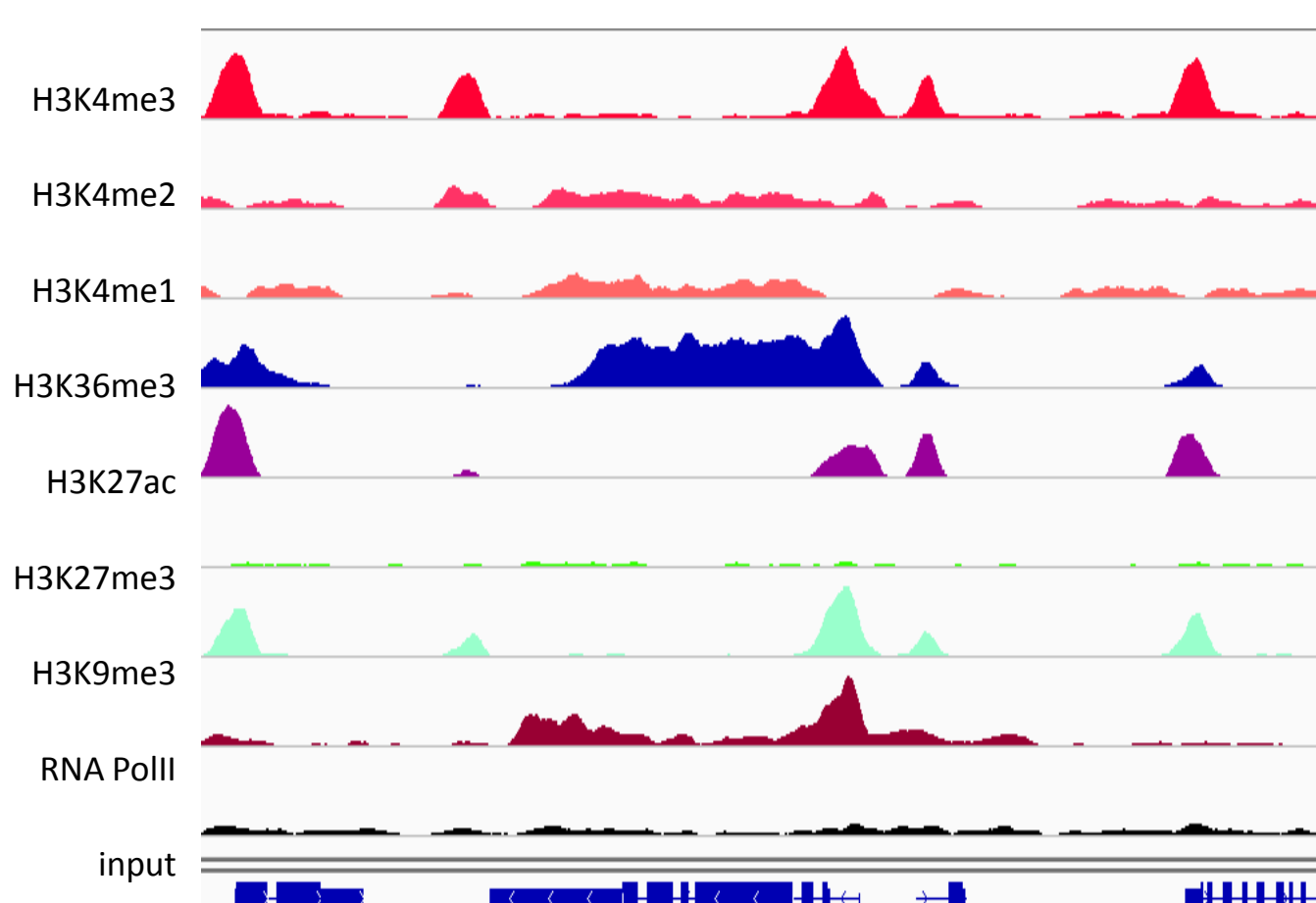


Chromatin Landscape

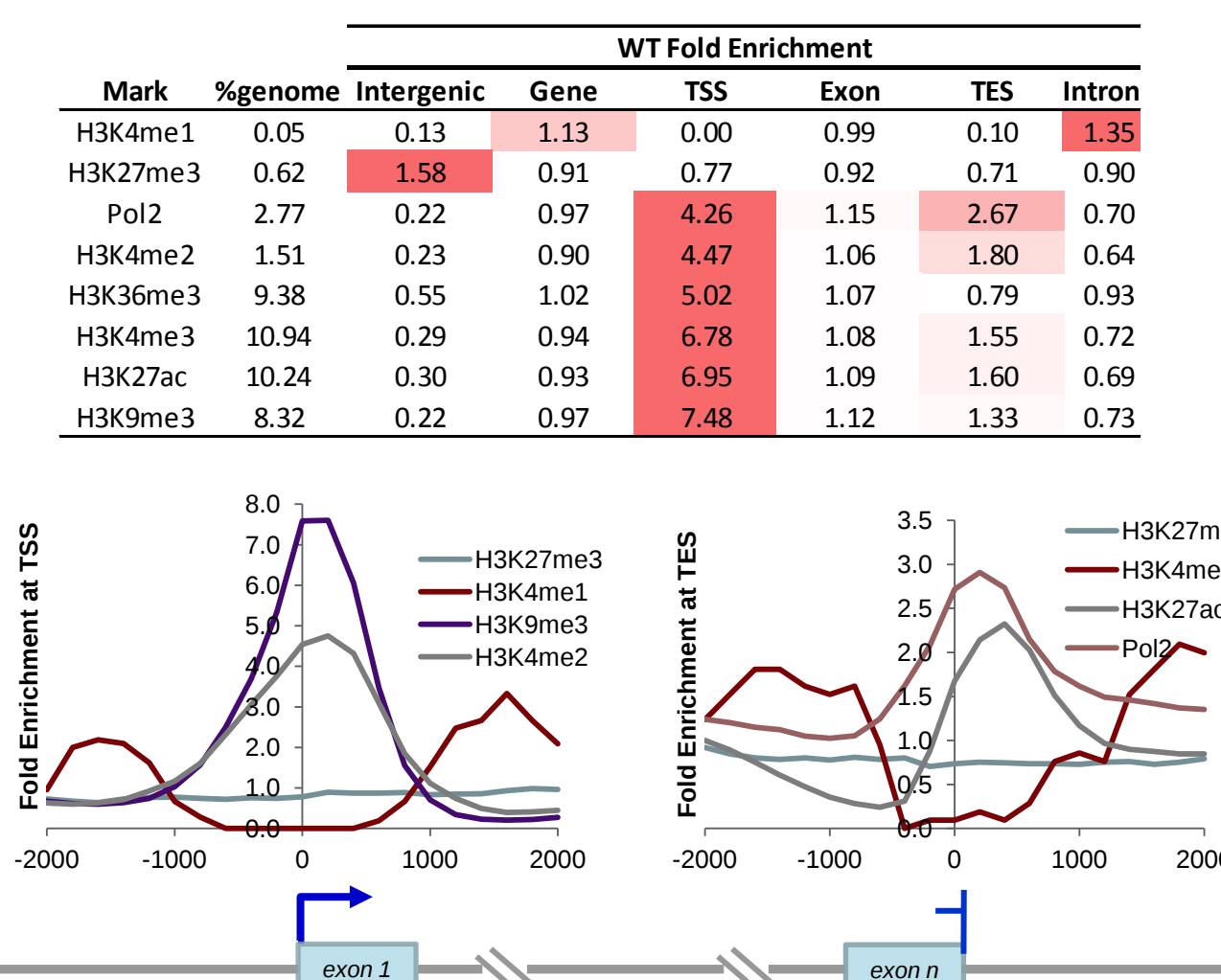
ChIP-seq Data summary

		Wild Type (0hr)			Nitrogen Depletion (1hr)			Sulfur Depletion (1hr)		
Target	Type	# of reads	# of dupl unique reads	#MACS2 peaks*	# of reads	# of dupl unique reads	#MACS2 peaks*	# of reads	# of dupl unique reads	#MACS2 peaks*
H3K4me3	IP	6,991,217	5,146,811	11,487	3,977,733	3,074,149	12,821	6,860,877	4,651,112	12,868
	Input	6,743,218	4,040,258		3,719,293	2,216,073		6,382,232	3,638,293	
H3K27ac	IP	4,112,025	2,924,426	11,992	4,060,312	3,028,160	12,244	7,591,542	4,914,747	12,623
	Input	5,394,178	3,158,292		3,719,293	2,216,073		6,131,473	2,868,969	
H3K4me1	IP	4,897,309	4,067,191	106	4,798,606	3,844,505	2	6,858,495	4,895,951	808
	Input	4,090,951	2,714,973		3,757,818	2,404,500		7,755,521	3,867,641	
H3K4me2	IP	4,503,995	3,510,032	2,482	4,301,190	3,250,887	996	6,791,655	4,832,831	9,620
	Input	4,090,951	2,714,973		3,757,818	2,404,500		6,382,232	3,638,293	
H3K9me3	IP	5,174,636	3,214,285	10,575	4,091,263	2,851,236	8,120	8,704,508	3,107,982	2,293
	Input	3,797,702	2,558,492		4,369,910	2,942,246		7,722,693	3,787,568	
H3K27me3	IP	2,541,255	1,744,210	484	5,786,549	3,321,905	521	4,983,273	1,057,077	510
	Input	4,099,370	2,529,618		3,867,019	2,411,860		6,541,621	3,509,535	
H3K36me3	IP	4,339,861	3,371,955	8,873	4,710,003	3,554,242	9,188	8,481,474	4,456,946	8,528
	Input	4,070,390	2,681,361		4,369,910	2,942,246		7,722,693	3,787,568	
RNA Pol II	IP	3,791,640	2,077,194	3,143	3,877,267	2,119,854	3,138	6,729,049	1,945,188	4,385
	Input	2,475,900	1,550,148		3,649,425	2,221,803		6,382,232	3,638,293	

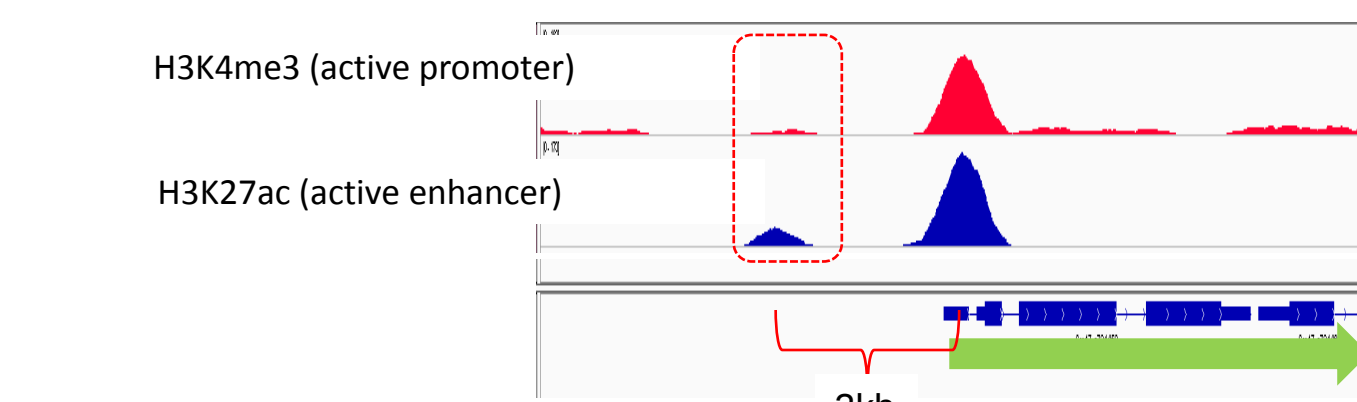
Genomic features



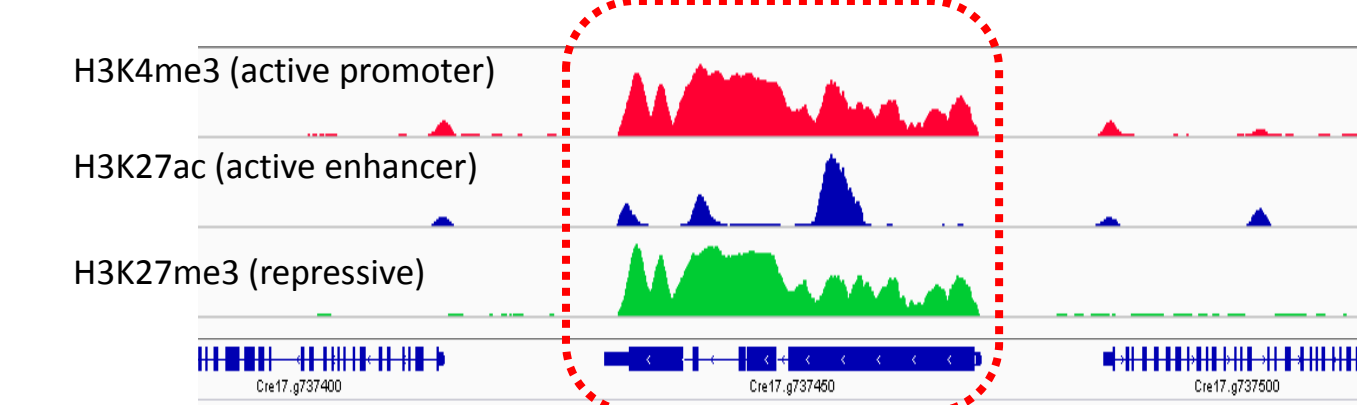
- Histone marks indicate a promoter-centric regulation.



- H3K27me3 marks heterochromatin.



- H3K27ac-specific marks indicates potential distal regulatory elements/enhancers.



- Simultaneous H3K27me3 and H3K4me3 marking indicates potential bivalent domain.

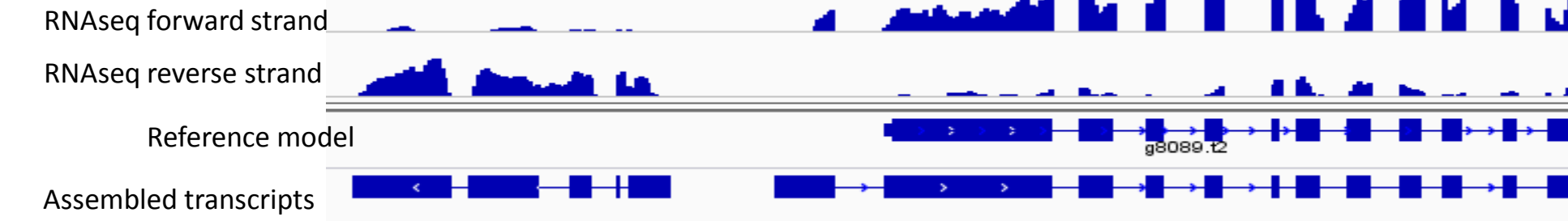
Transcriptome

Data summary

Total 9 timepoints x 2 conditions, Total reads of >50m per lib

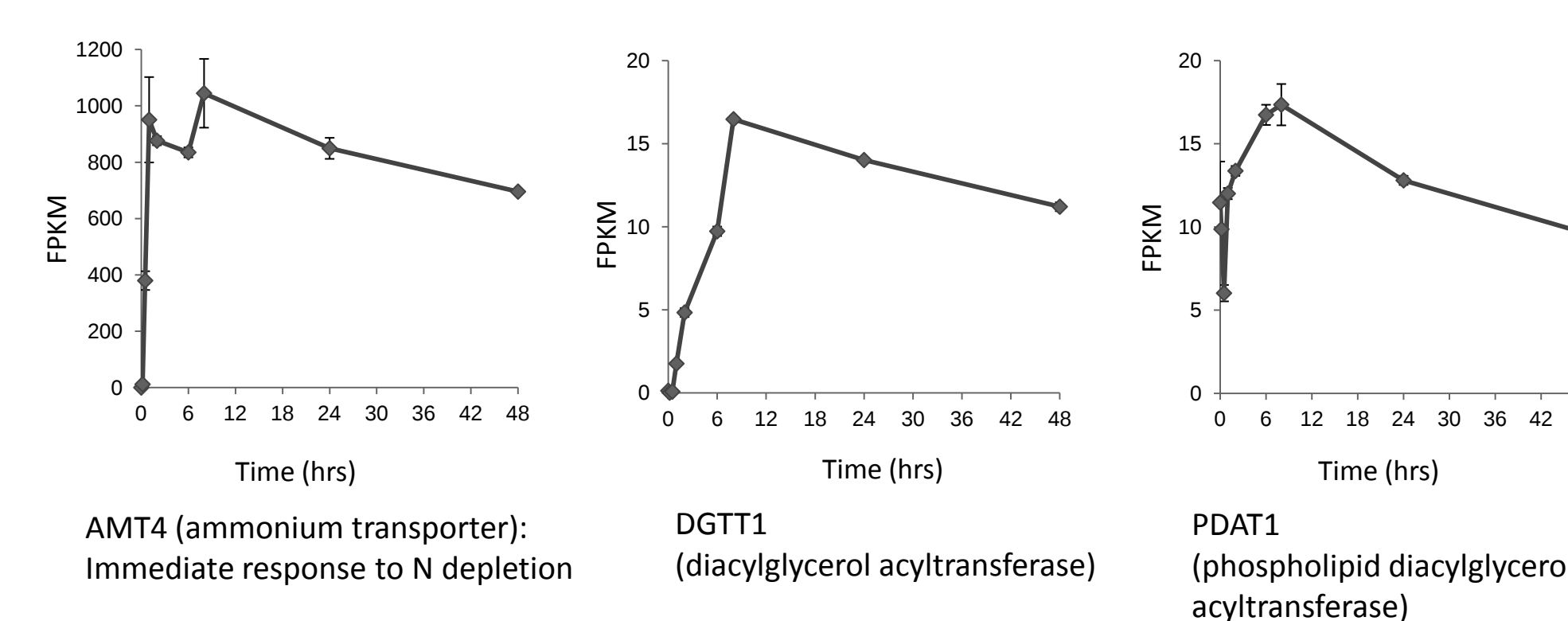
Genome ver5	Reference	Cufflink	Modified/filtered
Genes	18,773	17,730	16,467
Transcripts	19,529	33,929	32,395
Unique Transcriptome size (Mb)	94.6	97.8	96.1
Mean transcript length	5411	5922	6203

RNA-seq (coverage)



- Novel genes (184)
- Antisense genes (1727)
- Novel exon (544)
- Extension (6100)
- Merged (685)

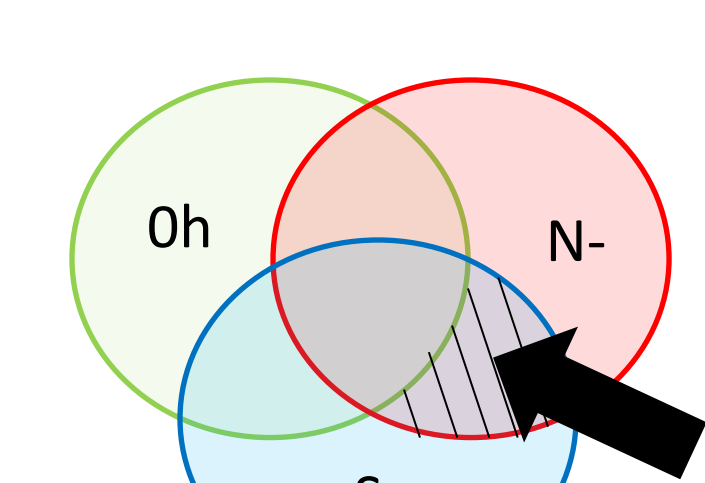
Control genes expression profile



Data mining: Searching for common players in lipid biosynthesis pathway

Multiple approaches targeting common differentially regulated components in nitrogen and sulfur depleted conditions.

- Simple differential modified histone marks

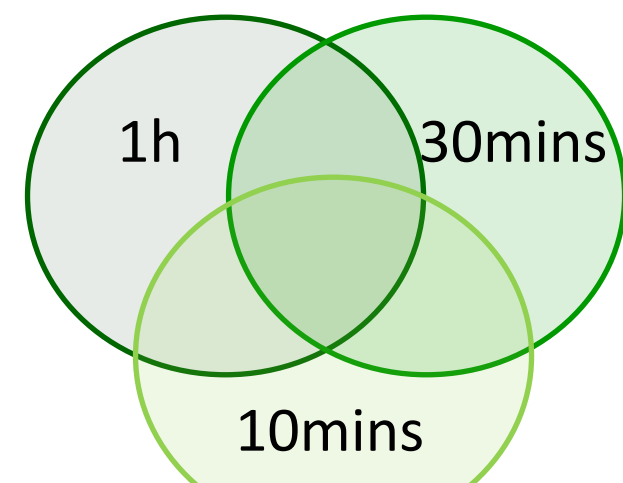


H3K4me3

RNA Pol II

H3K36me3

- Differential gene expression (targeting early response)



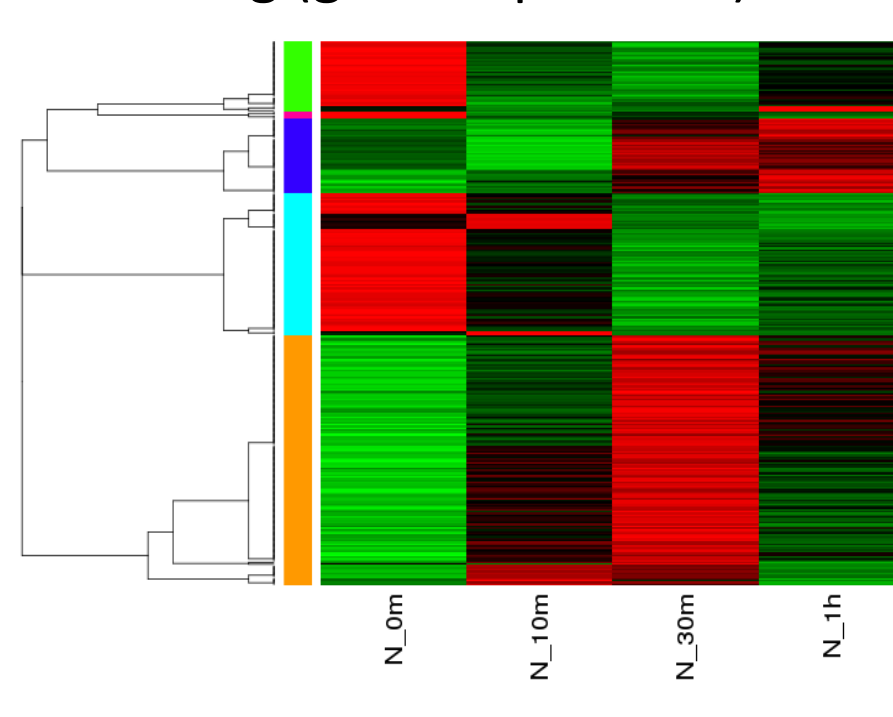
Nitrogen-depleted

Sulfur-depleted

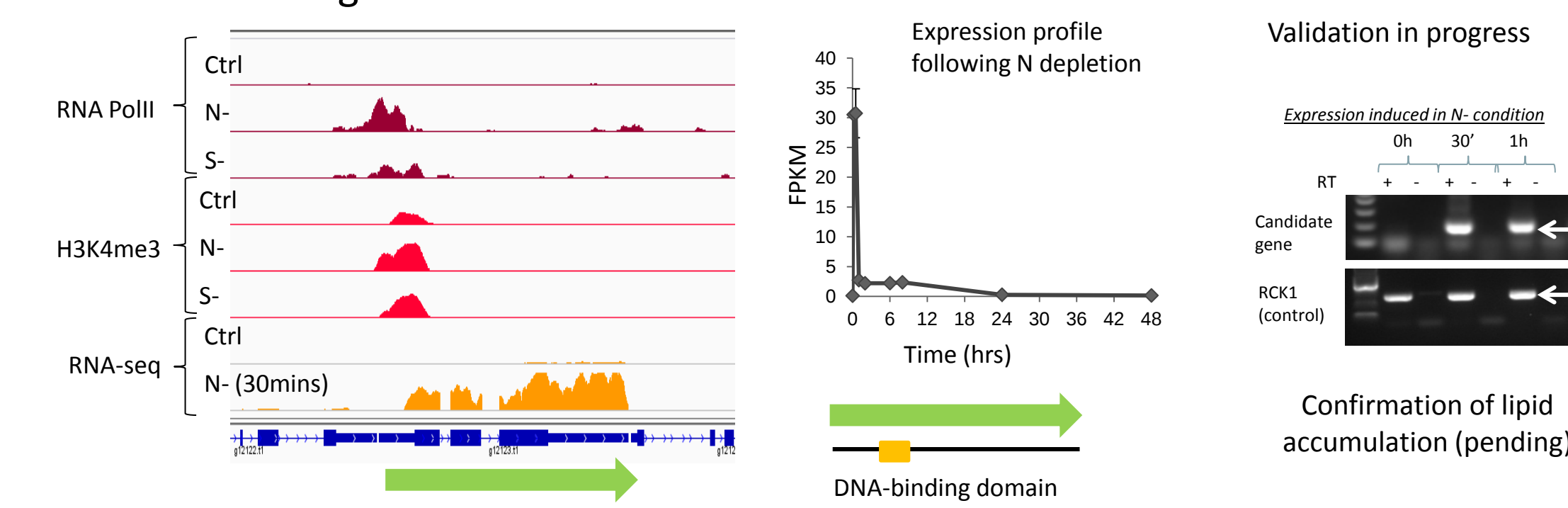
- Differential chromatin state (ChromHMM)

State	H3K27	H3K4	H3K4	H3K4	H3K36	H3K27	H3K9	Pol2	%genome	#segments	Intergenic	Gene	TSS	Exon	TES	Intron
14	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	84.32	13,641	1.39	1.01	0.26	0.99	0.89	1.04
3	0.6%	0.6%	1.0%	1.3%	98.8%	0.0%	2.6%	0.0%	1.96	3,982	1.53	1.12	0.07	1.00	0.16	1.31
4	3.2%	1.0%	3.1%	3.3%	100.0%	0.0%	5.4%	97.7%	0.31	613	0.10	1.08	0.27	1.01	0.91	1.21
1	0.1%	0.0%	0.5%	0.4%	0.0%	86.1%	10.4%	0.5%	0.19	302	0.29	0.88	0.54	0.91	0.78	0.84
2	0.1%	0.0%	1.8%	0.9%	98.8%	98.7%	11.9%	0.4%	0.42	204	0.25	0.93	0.66	0.93	0.74	0.93
5	2.0%	0.2%	0.9%	1.5%	0.0%	0.0%	0.4%	97.2%	0.91	1,205	0.12	0.95	0.91	1.29	0.58	0.41
10	84.8%	0.0%	0.0%	91.3%	78.8%	0.0%	0.0%	0.0%	0.33	1,707	0.15	1.01	1.22	0.84	0.67	1.28
13	89.4%	0.0%	0.0%	0.0%	0.0%	0.0%	2.3%	1.3%	0.52	2,385	0.33	0.71	2.41	0.85	2.65	0.49
12	24.0%	0.1%	95.5%	66.9%	0.6%	0.5%	0.0%	1.2%	1.07	2,085	0.63	0.88	4.15	1.04	1.97	0.62
9	80.1%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	3.1%	1.72	6,093	0.33	0.82	4.62	0.96	2.50	0.62
11	63.8%	0.0%	99.9%	99.3%	58.0%	10.3%	78.8%	0.0%	0.39	1,019	0.60	0.94	5.88	1.12	1.32	0.66
6	98.2%	0.0%	0.5%	99.7%	95.3%	0.0%	94.8%	98.8%	1.43	2,010	0.49	0.98	7.13	1.10	1.29	0.80
7	97.1%	0.0%	0.0%	99.7%	100.0%	0.0%	100.0%	0.0%	4.78	7,519	1.00	0.99	7.44	1.13	0.93	0.78
8	94.3%	0.0%	0.0%	100.0%	0.0%	0.0%	100.0%	0.1%	1.65	4,422	1.12	0.89	8.65	1.11	2.50	0.56
									Pol2	0.22	0.97	4.26	1.15	2.67	0.70	

- Clustering (gene expression)



- Potential regulators



Discussion

Here, we report a pilot attempt to understand the regulation of algae lipid biosynthetic pathways by genome wide chromatin landscaping. Deep sequencing of transcriptome also provide supporting data to further dissect the regulatory pathways. Computational analysis have enabled us to narrow into a smaller group of highly potential regulators. In-depth analysis is still on going to map out a potential interaction or inter-regulatory pathways. Effort to further validate these findings in the biological system is currently underway. This approach shall provide a more feasible way of dissecting regulatory elements in algae and potentially overcome the limitation of the laborious and low-throughput mutagenesis screening approach.