

TRANSCRIPTIONAL RESPONSE OF *Chlamydomonas reinhardtii* TO SMALL
LIPID-INDUCING MOLECULES

by

Boqiang Tu

A THESIS

Presented to the Faculty of

The College of Agricultural Sciences and Natural Resources at the University of Nebraska

In Partial Fulfilment of Requirements

For the Degree of Bachelor of Science in Biochemistry
with Highest Distinction

Major: Biochemistry

Under the Supervision of Professor Concetta C. DiRusso

Lincoln, Nebraska

December, 2015

TRANSCRIPTIONAL RESPONSE OF *Chlamydomonas reinhardtii* TO SMALL
LIPID-INDUCING MOLECULES

Boqiang Tu, B.S.

University of Nebraska, 2015

Adviser: Concetta C. DiRusso

In response to the energy crisis in recent years, large-scale investments and government mandates have led to an expansion in renewable energy research. This situation provides incentive for research in microalgae as a feedstock for biofuel production. Nitrogen starvation is commonly known for stimulating lipid production in algae while causing severe cellular stress and limiting growth. Our laboratory performed a large scale in vivo high throughput screen to identify small molecules that induce lipid accumulation in the model organism *Chlamydomonas reinhardtii*. Four of these lipid-inducing compounds were selected for targeted gene expression analysis using quantitative PCR. Relative expression of 18 genes were analyzed via qPCR from 3 biological replicates. The final growth of compound treated cells was on average 90% over controls, 2-fold higher than nitrogen starved cells, and lipid accumulation was up to 6-fold higher than controls. It was shown that compound treatment and nitrogen starvation have different effects on lipid metabolism-related gene expression. Three compounds were selected for gene expression analysis using next-generation sequencing technique. Samples were collected after 72h of treatment with 3 biological replicates for each compound. mRNA was isolated and sequenced on Illumina Hi-seq 2000. The sequencing reads were mapped to the genome using Tophat2 and assembled as transcriptome using Cufflinks. Pathway analysis revealed significant metabolic shift under compound treatment. Changes in TCA cycle favor the shift of energy metabolism towards lipid biosynthesis related pathways via citrate efflux. Down-regulation of anabolic

pathway and up-regulation of ER protein processing/degradation suggests a significant amount of carbon flux used for TAG biosynthesis is from recycling of cellular components, but unlike nitrogen starvation, compound treatments do not severely limit photosynthesis. Differential expression of genes involved in plastid membrane lipid metabolism suggests that unlike nitrogen starvation, SQDG and PG may play more important roles as lipid pools than MGDG and DGDG. At the level of transcription regulation, compound treatments induced changes that are opposite to that induced by N starvation, particularly RWP-RK domain transcription factors and SBP domain transcription factors. This study has demonstrated that small lipid-inducing molecules are useful for identifying components and mechanisms that regulate lipid synthesis and may be utilized for biofuel production.

ACKNOWLEDGMENTS

Over the past 3.5 years, I have received guidance and support from a great number of individuals. This work would not have been possible without the supervision of my advisor, Dr. Concetta DiRusso, who has been a great mentor and friend to me during this exciting research experience. I would like to thank Dr. Nishikant Wase, who thoughtfully trained me to conduct scientific research with critical thinking and state-of-the-art technologies. I would like to thank Dr. Paul Black and Dr. Istvan Ladunga, who served on my thesis committee. As lab manager, Mark Behrens provided indispensable technical support and taught me important work ethics. I would also like to thank all other members of FATT Lab and the faculty and staff of the Department of Biochemistry for their intellectual, technical, logistic and emotional support.

Appendix B "Parallelization on distributed memory using MPI of transcriptome assembly tool Cufflinks" is a term project of CSCE 456: parallel programming (Spring 2015). I would like to thank the instructors: Dr. David Swanson, Dr. Adam Caprez and Dr. Brian Bockelmann and the Holland Computing Center.

I gratefully acknowledge the Nebraska Center for Energy Science Research (NCESR) and the National Science Foundation (NSF-EPSCoR, EPS-1004094 and 1264409, CBET-1402896 and ACI-1053575) for support that made this work possible.

Contents

Contents	v
List of Abbreviations	ix
List of Figures	x
List of Tables	xii
1 Introduction	1
1.1 Microalgae as a feedstock for biofuel production	1
1.2 Methods in the literature for lipid induction in microalgae	2
1.3 High-throughput screening for lipid-inducing small molecules	3
1.4 Next-generation sequencing and bioinformatics	3
2 Targeted gene expression analysis using quantitative PCR	7
2.1 Growth and lipid accumulation	7
2.1.1 Growth assessment	7
2.1.2 Lipid droplets estimation	8
2.1.3 FAMES analysis	9
2.1.4 TLC analysis of TAGs	10
2.2 Amplification efficiency and specificity assessment	10

2.3	Analysis of differential gene expression and potential impact on metabolism	12
2.3.1	Summary of differential gene expression	12
2.3.2	Selected genes in TCA and Glyoxylate Cycle	14
2.3.3	Fatty acid and TAG biosynthesis	20
2.3.4	RuBisCO and light reaction components in photosynthesis	20
2.3.5	Assessment of expression of the BIP and APG8 genes	21
3	Transcriptome analysis using next-generation RNA sequencing	22
3.1	Quality assurance of RNA-seq experiment	22
3.2	Summary of differential gene expression analysis	24
3.2.1	Summary of overall differential gene expression patterns	24
3.2.2	Summary of the most significant differences in gene expression af- ter treatment with the 3 compounds	26
3.3	Enrichment and pathway analysis	33
3.3.1	Summary of coordinately changed GO processes	33
3.3.2	Transcriptional regulation and signal transduction	40
3.3.3	Carbon metabolism	43
3.3.3.1	Glycolysis	43
3.3.3.2	TCA cycle	49
3.3.3.3	Carbon fixation	50
3.3.3.4	Starch metabolism	50
3.3.4	<i>de novo</i> fatty acid biosynthesis and desaturation	51
3.3.5	Lipid metabolism	52
3.4	Validation of differential expression with qPCR	55
4	Role of exogenous citrate in lipid production	57
4.1	Introduction	57

4.2	Growth and lipid accumulation	58
4.2.1	Growth assessment	58
4.2.2	Lipid droplets estimation	58
4.2.3	FAMES analysis	61
4.3	Uptake of citrate from the media	63
5	Discussion	65
5.1	Effects of compounds on lipid accumulation and growth	66
5.2	Advantages and limitations of RNA-seq and computational methods	66
5.3	Overview of the transcriptional responses leading to TAG accumulation	67
5.4	Cell signaling and transcription regulation	70
5.5	Future directions	72
6	Methods and Material	73
6.1	Culture conditions and growth assessment	73
6.2	Lipid droplets estimation using Nile red	74
6.3	FAMEs and metabolite analysis using GC/MS	74
6.3.1	Lipid extraction and transesterification	74
6.3.2	Metabolite extraction and derivatization	75
6.3.3	GC/MS parameters	75
6.3.4	Data analysis	76
6.4	Lipid analysis using TLC	76
6.4.1	Lipid extraction	76
6.4.2	Chromatography	76
6.4.3	Data analysis	77
6.5	Gene expression analysis using qPCR	77
6.5.1	Isolation and quality assessment of RNA	77

6.5.2	qPCR amplification	78
6.5.3	Data analysis	79
6.6	Transcriptome analysis using RNA-seq	79
6.6.1	Mapping reads to a reference genome	79
6.6.2	Transcriptome assembly and quantification	81
6.6.3	Differential expression analysis	81
6.6.4	Gene set enrichment and pathway analysis	82
A	Scripts used in RNA-seq data analysis	83
B	Parallelization on distributed memory using MPI of transcriptome assembly tool Cufflinks	84
B.1	Introduction	85
B.2	Existing pthread Implementation	85
B.3	MPI Implementation	89
B.4	Performance Evaluation	93
B.5	Conclusions and Future Work	94
B.6	Source Code and Testing Data Accession	95
References		100

List of Abbreviations

CPM Counts per million mapped reads

DMSO Dimethyl sulfoxide

FAME Fatty acid methyl ester

FDR False discovery rate

FPKM Fragments per kilobase of transcript per million mapped reads

GC/MS Gas chromatography/mass spectrometry

GO Gene ontology

GSEA Gene set enrichment analysis

MTBE Methyl tert-butyl ether

OD Optical density

qPCR Quantitative Polymerase Chain Reaction

TAP Tris-acetate-phosphate

TLC Thin-layer chromatography

TMSH Trimethyl sulfonium hydroxide

List of Figures

2.1	Growth under compound treatment	8
2.2	Lipid droplet accumulation during compound treatment	9
2.3	TLC plate image of TAGs	12
2.4	Amplification and melting curves	13
2.5	Summary of differential gene expression in qPCR	16
2.6	Expression of selected genes under different treatments	18
3.1	RNA gel electrophoresis image	23
3.2	Quality check of raw sequencing reads and transcripts	25
3.3	Summary of global gene expression profiles	28
3.4	Expression profiles of most significantly changed transcripts	32
3.5	Expression profile of transcription factors	41
3.6	Expression profile of transcripts in coordinately shifted major metabolic pathways	48
3.7	Comparison of the expression of selected genes using qPCR	56
4.1	Growth during citrate supplementation	59
4.2	Lipid droplets accumulation during citrate supplementation	60
4.3	FAMEs profiles during citrate supplementation	62
4.4	Change in intra- and extracellular citrate concentration over time	64

B.1	The Cufflinks algorithm of transcriptome assembly	96
B.2	Strong scaling of MPI parallelized code on Crane at HCC	97
B.3	Weak scaling of single-threaded MPI implementation	97
B.4	Idealized strong scaling of MPI parallelized code	98
B.5	Strong scaling of pure pthread implementation on a single node	98
B.6	Correctness check of the output from MPI parallelized implementation against the original implementation	98
B.7	Two schemes of improving the performance of this MPI implementation	99

List of Tables

1.1	Structural information of small molecules identified in HTS	4
2.1	Quantification of fatty acid species under compound treatment	11
3.1	Numbers of differentially expressed transcripts	29
3.2	Summary of significantly changed GO categories	35
6.1	List of primers used in qPCR experiments	80

Chapter 1

Introduction

1.1 Microalgae as a feedstock for biofuel production

The recent global challenges of climate change and energy crisis caused by use and depletion of fossil fuels have led to increased interest in renewable energy. The production of corn ethanol in the US has increased by 10 fold during the last 10 years, but its life-cycle energy consumption, greenhouse gas emission have raised concerns [1]. Produced from oleaginous crops, mainly soybean, rapeseed, sunflower and palm, biodiesel has recently gained attention for its versatility in uses and the higher energy efficiency in its production. Unfortunately, these crops are also used for the production of vegetable oils, which directly competes with the production of biodiesel for land and other resources [2]. Microalgae, a group of unicellular photosynthetic organisms, can be used as alternative feedstock for biodiesel production because numerous species contain high amounts of oil and can grow at a high rate in fresh and brackish waters on arable lands [3]. However, due to the costs of biomass harvest and lipid extraction, commercial production of algal biodiesel requires higher lipid content in most suitable algal species [4].

1.2 Methods in the literature for lipid induction in microalgae

To induce lipid accumulation in microalgae, most approaches reported in the literature require an environmental stress, particularly nutrient deprivation of nitrogen [5], sulfur [6] or metallic micronutrients [7, 8]. Under deprivation of nitrogen after entering stationary phase, model alga *Chlamydomonas reinhardtii* produces abundant cytoplasmic lipid bodies. After 48 h of N starvation in the presence of acetate, the content of lipid bodies increased by up to 15 folds in the cell[9, 5, 6]. It has been shown that after 4 days of sulfur starvation, significant amount of TAG lipid bodies is produced and the amount of TAG does not decrease after 4 days as occurs in N starvation [6].

Current genetic engineering approaches mainly focus on the diversion of carbon from the starch biosynthetic pathway [10, 11, 12, 13]. In the *sta6* mutant in which starch biosynthesis was blocked, the lipid body content increased up to 30-fold under nitrogen starvation [5]. However, it has been reported that the interactions between different biosynthetic pathways were more complex. Blocking the starch biosynthesis does not increase lipid accumulation under normal condition but under specific stress conditions [12].

Lipid induction by nutrient limitation was often accomplished by rapid autophagic processes[5], manifested by increased protein recycling, decreased protein synthesis, degradation of chloroplast and ribosome, and turnover of membrane lipids, including the plastid galactolipids [9, 14, 15].

1.3 High-throughput screening for lipid-inducing small molecules

High-throughput screening (HTS) of chemical compounds to identify modifiers of molecular targets and cellular processes has become a central component of the drug discovery process. Phenotype-based HTS has long been used to identify inhibitors of yeast and mammalian cells [16, 17, 18] and recently effectors of microalgae growth and lipid accumulation [19, 20]. In a previous work [21], our lab developed a small molecule phenotypic screen to identify compounds that induced lipid accumulation while maintaining growth and minimizing induction of stress response pathways. To evaluate this screening method, we employed two test libraries obtained from the National Cancer Institute (NCI), Diversity Set III and Natural Products Set II. We identified structurally diverse groups of small molecules using model organism *Chlamydomonas reinhardtii*, which induce lipid accumulation with minimal impact on biomass accumulation. Four compounds were selected for further analysis in this study (Table 1.1).

1.4 Next-generation sequencing and bioinformatics

Compared to microarray analysis, next-generation mRNA sequencing (RNA-seq) enables the measurement of transcript expression and discovery of novel transcripts of the entire transcriptome in a single assay, without the need for hybridization of probes synthesized on the basis of a particular transcriptome. With the advanced technology and lowered cost, RNA-seq has become a revolutionary tool for molecular biology[22]. However, for the big data set to become biologically meaningful, efficient and robust algorithms built upon appropriate statistical principles must be employed in the analysis. In this study, we employed various software tools developed by the bioinformatics community such as

Table 1.1: Structural information of small molecules identified in HTS

ND ^a	ChemBridge ID ^b	IUPAC Name	Structure	Lab ID ^c
30	5345030	1-(3S)-1-[4-(Methoxycarbonyl)phenyl]-2,5-dioxo-3-pyrrolidinyl-4-(4-methoxyphenyl)piperazin-1-ium		WD30030
42	5950542	3-[1-adamantyl(methyl)amino]propanenitrile hydrochloride		WD20542
67	5234067	1-(2-adamantyl)-4-(methylsulfonyl)piperazine		WD20067
84	6718784	1-[1-(bicyclo[2.2.1]hept-5-en-2-ylmethyl)-4-piperidinyl]-4-(2-methoxyphenyl)piperazine oxalate		WD10784

^aID of compound used in this study

^bUnique identifier used by ChemBridge (supplier of compounds)

^cDesignation used in our lab based on chemical scaffolds

Bowtie, TopHat, Cufflinks, eXpress and GSEA to process and analyze the large amount of data generated by the RNA-seq experiment.

Since the genome of *Chlamydomonas reinhardtii* has been sequenced and partially annotated [23], this study employed reference-guided transcriptome assembly. TopHat is a efficient tool that utilizes Bowtie 2 as the underlying algorithm to align short reads to the genome at annotated loci and enables the discovery of novel splice site. It was reported that TopHat pipeline is much faster than previous systems [24]. Based on Dilworth's theorem, Cufflinks reduces the transcript assembly problem to maximizing compatibilities among fragments in a weighted bipartite graph. Without relying on the existing gene annotation, Cufflinks enables the discovery of novel transcripts [25]. With millions of short reads as input, multiple alignment to a single loci is unavoidable and resolving such ambiguity is critical for accurate transcript abundance quantification. We also employed eXpress, a tool based on an online algorithm for fragment assignment that processes one fragment at a time to directly utilize estimated counts rather than relative abundances. In the algorithm, each fragment is initially mapped to an arbitrary number of target sequences and reassigned according to previously estimated counts. Parameter estimates for the fragment-length distribution, the sequence bias and a sequencing error model for reads are simultaneously updated [26]. Gene expression analysis focusing on identifying differential expression of individual genes often fails to discover the systematic changes in biological processes, which are often consisted of very subtle changes at the level of individual genes. GSEA identifies pathways and processes represented by gene sets which are more reproducible and more interpretable than individual genes. Based on the strong cross-correlation of members of a gene set, GSEA can reveal modest changes in individual genes by boosting the signal-to-noise ratio. By clustering the genes in the leading edges of GSEA results, important subsets that are responsible for shifts in biological processes can be identified [27]. The Geno Ontology Project provides a comprehensive source of computational rep-

resentation of the functional information of genes using controlled vocabulary of molecular function, biological process and cellular component, which can be used to annotate a wide range of transcripts [28]. While the numbers of genes annotated in KEGG database is less, it provides information of biological processes in the form of one-dimensional graph with substrates as nodes and enzyme as edges, allowing the analysis and visualization of generalized protein interaction network [29].

Chapter 2

Targeted gene expression analysis using quantitative PCR

2.1 Growth and lipid accumulation

2.1.1 Growth assessment

Cells grown at mid-log phase in liquid culture were washed and treated with compounds at a concentration of 20 μM each. We note that this concentration is 2-fold higher than that used in the initial high-throughput screening. Our rationale was to maximize differences between compound-treated and non-treated samples. Three biological replicates were used for each condition. Incubation was continued for 72 h and its optical density was monitored daily (Figure 2.1.) At 20 μM , reduced growth rate was observed for most compounds after 24 h compared to control (TAP N+ with equivalent amount of DMSO). Except for compound 84, the impact on growth of compound treatment was not as severe as nitrogen starvation, which also reduced the amount of photosynthetic pigments as the culture became bleached. Again with the exception of compound 84, no significant culture bleaching

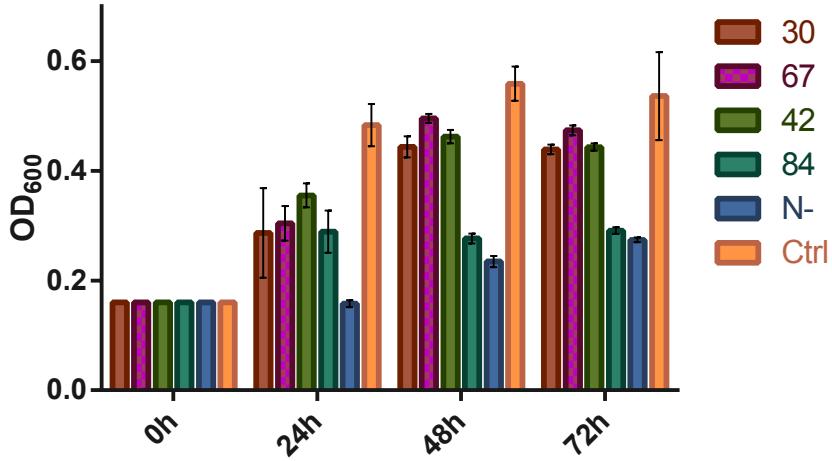


Figure 2.1: Growth under compound treatment. Cell density at 0h, 24h, 48h and 72h was estimated by optical density at 600 nm for each treatment and control. Three biological replicates were used for each condition.

was observed amount compound treatments. The dose response patterns of these compounds were obtained by treating the cell at different concentrations and the EC 50 was determined to be no greater than 5 μ M for each compound [30]. Therefore, to maximize lipid accumulation while minimizing cellular stress, the concentration of 5 μ M was used for the RNA-seq study 2.2.

2.1.2 Lipid droplets estimation

To monitor the lipid accumulation during treatment, an aliquot of cell culture was collected every 24 h and cells were stained with Nile red to estimate the amount of lipid droplet by measuring the fluorescence. As shown in Figure 2.2A, the amount of lipid droplet increased 2-fold after 24 h of treatment with each compound. After 72 h of treatment, the amount of lipid droplet increased up to 6-fold with compound 84 and 3- to 5-fold with compound 30, 42 and 67. To visually assess the effect of compounds on morphological features, stained

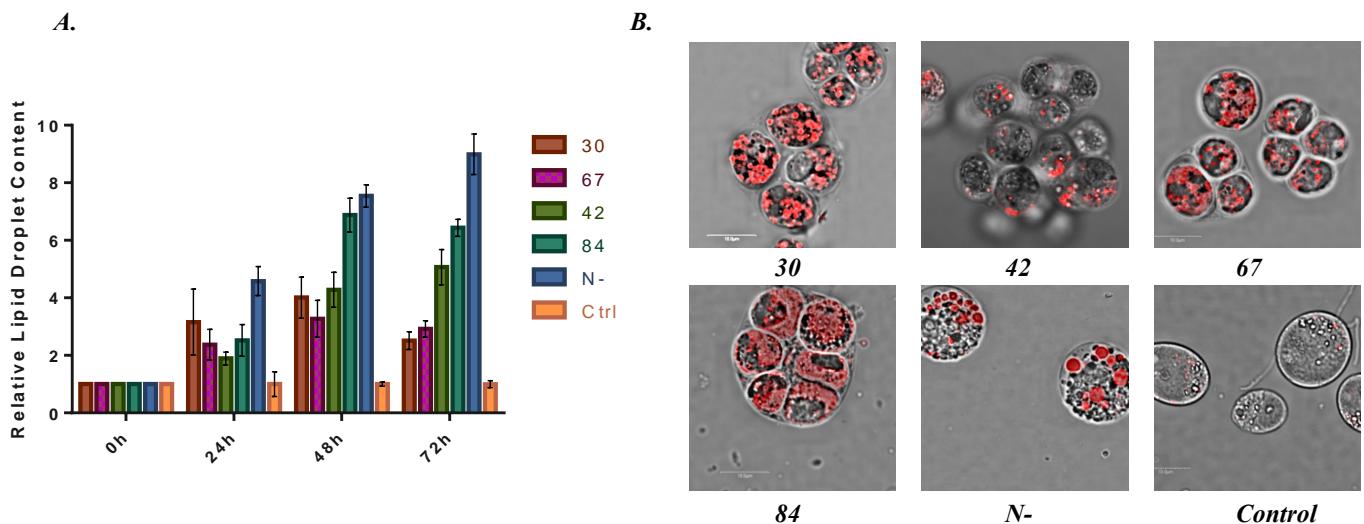


Figure 2.2: Lipid droplet accumulation during compound treatment. **A.** Lipid droplet content at 0h, 24h, 48h and 72h was estimated by fluorescence intensity after cells were stained with Nile red for each treatment and control. The fold change of fluorescence intensity relative to control was normalized by cell density in OD₆₀₀. Three biological replicates were used for each condition. **B.** Representative confocal microscopic images of cells stained with Nile red after 72h of incubation

cells were subjected to fluorescent confocal microscopy. As shown in Figure 2.2B, all 4 compounds and N- induced significant accumulation of lipid droplets. Interestingly, unlike N- or control, an extracellular matrix linking the cells seemed to have formed as cells were observed growing in clusters, particularly for compound 84.

2.1.3 FAMES analysis

To confirm total lipid accumulation and examine the composition of fatty acids, lipids were extracted and transesterified for analysis using GC/MS (Table 2.1 on page 11). The amount of total fatty acid increased significantly by treatment with all 4 compounds compared to control ($p < 0.05$) up to 5-fold. Palmitic acid (C16:0) and alpha-linolenic acid (C18:3), which increased up to 4-fold and 7-fold, respectively, were two major species that con-

tributed to the overall increase of fatty acids. Interestingly, after nitrogen starvation [9], C16:3, but not C18:3, increased by a large amount, contributing to the overall increase in total amount of fatty acid.

2.1.4 TLC analysis of TAGs

To visually confirm the accumulation of TAGs after compound treatment, fresh lipid extracts were analyzed on a silica gel TLC plate with olive oil as standard (Figure 2.3.) As expected, cells grown under nitrogen starvation accumulation the highest amount of TAGs, with cells treated with compound 84 being second highest in TAG accumulation, followed by compound 42. This results is consistent with the lipid droplet estimation by Nile red staining. As estimated using ImageJ quantification software, the density of the TAG band after treatment with compound 84 was approximately 4-fold higher than that of control cells. A subtle difference in the bands of N- and compound 84 suggested different TAG composition was induced under compound treatment.

2.2 Amplification efficiency and specificity assessment

To assure the quality of qPCR amplification and thus the reliability of the expression data, the amplification efficiency was estimated from the raw fluorescence data using the LinReg-PCR [31] algorithm and amplification specificity assessed using the melting curve method (Figure 2.4.) The amplification curve of each sample in log-scale was shown in Figure 2.4C, from which the baseline and amplification efficiency was determined from a common window-of-linearity. As shown in Figure 2.4D, the majority of the samples had satisfactory efficiencies above 70%. A representative set of melting curves of a single amplicon was shown in Figure 2.4A. The single peak of the first derivative of melting curve indicates high specificity of the amplification.

Table 2.1: Quantification of fatty acid species under compound treatment

Condition	Fatty acid content (µg/millioncells)							p-value ^a
	C16:0	C16:1	C16:3	C16:4	C18:0	C18:1	C18:2	
30	23.64 ± 8.88	0.98 ± 0.38	0.78 ± 0.24	5.20 ± 1.54	2.74 ± 0.92	3.82 ± 1.50	8.66 ± 3.20	18.16 ± 6.52
67	16.48 ± 5.46	0.86 ± 0.32	0.64 ± 0.26	2.54 ± 0.98	2.38 ± 0.66	2.22 ± 0.92	5.16 ± 1.94	10.60 ± 3.70
42	13.16 ± 2.10	0.78 ± 0.12	0.50 ± 0.06	2.42 ± 0.24	2.24 ± 0.34	2.18 ± 0.20	5.10 ± 0.68	10.24 ± 1.42
84	13.78 ± 4.74	0.22 ± 0.40	0.22 ± 0.40	1.06 ± 0.92	1.16 ± 0.42	0.98 ± 0.86	2.44 ± 1.00	8.26 ± 3.42
Control	5.50 ± 0.18	0.26 ± 0.04	0.24 ± 0.02	2.22 ± 0.30	0.56 ± 0.04	1.18 ± 0.06	1.34 ± 1.38	2.54 ± 1.70
								13.82 ± 1.00

^at-test between the total fatty acid content of each treatment and control.

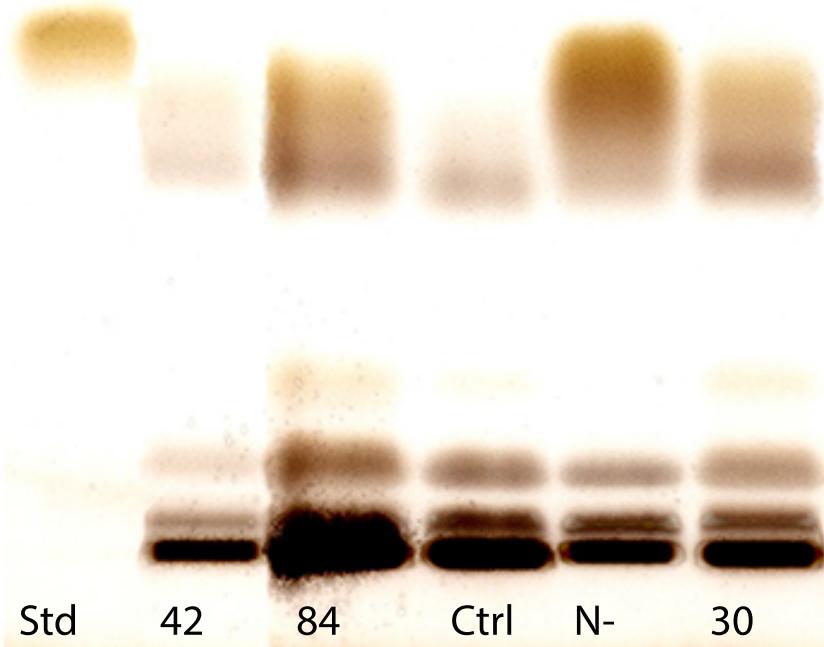


Figure 2.3: TLC plate image of TAGs. Total lipid extracts of each treatment as indicated and the control were analyzed on a TLC plate using olive oil as standard (Std). Different amount of TAGs was present in cells in each condition. One representative sample of each condition was used for this visualization.

2.3 Analysis of differential gene expression and potential impact on metabolism

2.3.1 Summary of differential gene expression

To assess the impact of compounds on the level of transcription, 16 genes were selected from different metabolic pathways that may be important for the process of lipid accumulation, of which ACLA1, CIS1 and MDH, encoding ATP-citrate lyase, citrate synthase and malate dehydrogenase, respectively, belong to TCA cycle; ACX, FAB2 and PLSB1, encoding acetyl-CoA carboxylase, plastid acyl-ACP desaturase and glycerol-3-phosphate

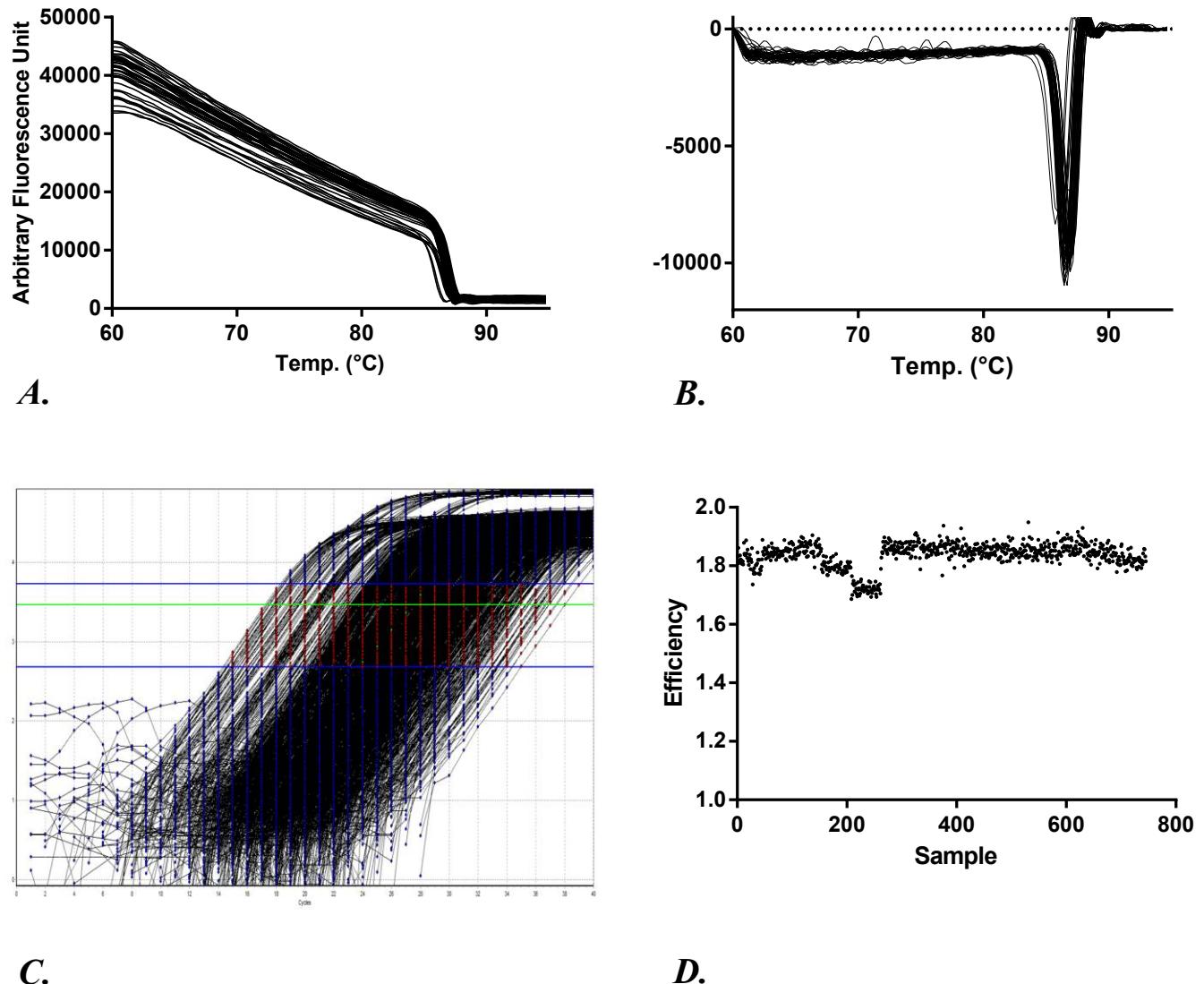
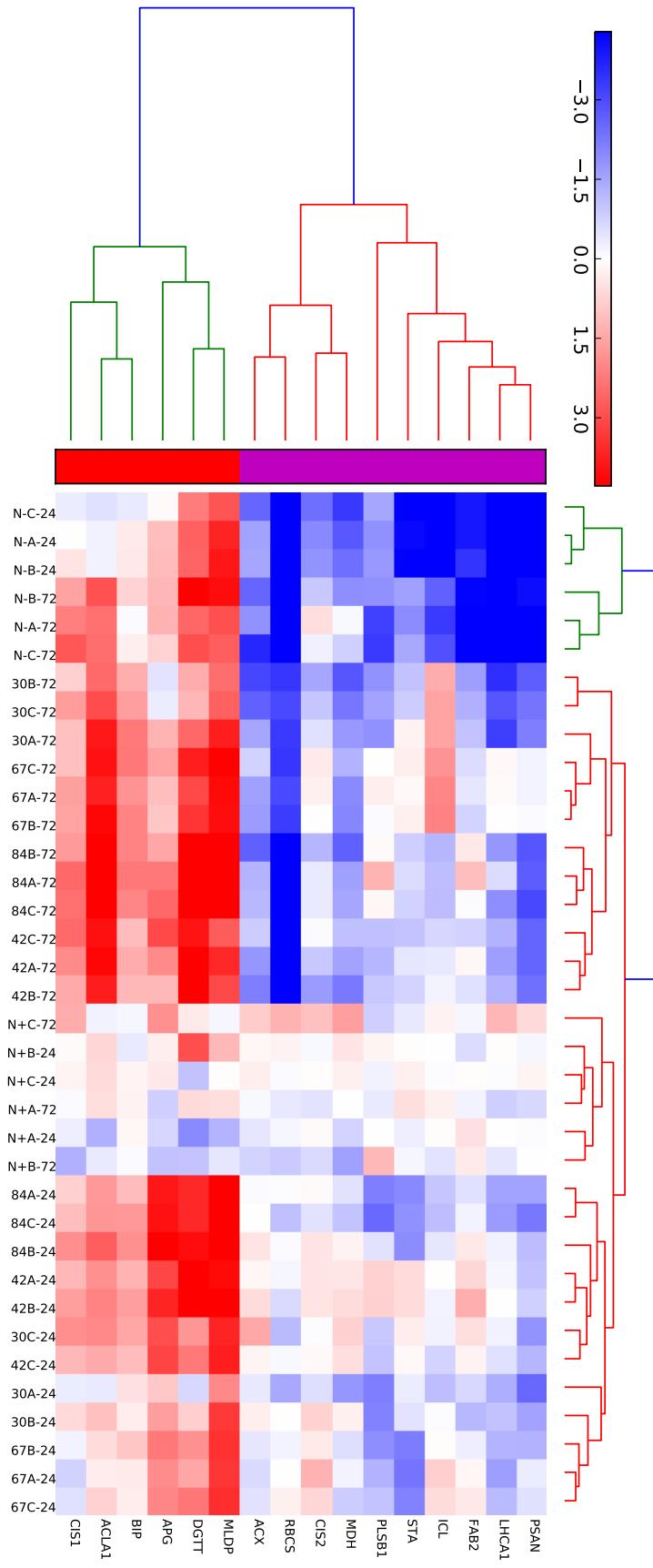


Figure 2.4: Amplification and melting curves. **A.** Representative melting curve of a single amplicon. **B.** First derivative of the melting curve shown in A. with respect to temperature. The presence of only a single peak typically indicates the high specificity of amplification. **C.** The log-transformed amplification curve plotted from raw fluorescence data. Points in red color are within the window-of-linearity and used for the estimation of amplification efficiency. **D.** Plot of the values of amplification efficiency of each individual reaction as shown in C.

acyltransferase, respectively, belong to *de novo* fatty acid biosynthesis; CIS2, encoding citrate synthase, belongs glyoxylate pathway; LHCA, PSAN and RBCS, encoding light-harvesting protein of photosystem I, photosystem I reaction center subunit N and RuBisCo small subunit, respectively, belong to photosynthesis; STA, encoding ADP-glucose pyrophosphorylase, belongs to starch synthesis; APG and BIP, encoding autophagy-related protein 8 and binding protein 1, belong to autophagy/stress response; DGTT and MLDP, encoding diacylglycerol acyltransferase and major lipid droplet protein, belong to TAG biosynthesis. To determine the temporal effects of compound treatment on gene expression, samples were collected at 24 h and 72 h time points. In general, genes up-regulated at 24 h were further up-regulated at 72 h and genes down-regulated at 24 h were further down-regulated at 72 h. Hierarchical clustering 2.5 suggest that the selected genes DGTT and MLDP in TAG biosynthesis that were increased in expression by compound treatment may be co-expressed with stress response genes APG and BIP, as well as ACL and CIS1 in citrate metabolism, while genes encoding photosynthetic enzymes were coordinately down-regulated. Principal component analysis has revealed that compound strucure and time indeed had major impacts on the pattern of gene expression. While there was very little temporal change in the gene expression pattern of control cells, compound treated cells experienced pronounced transcriptional change over time, with the change in compound 84-treated cells being most drastic.

2.3.2 Selected genes in TCA and Glyoxylate Cycle

ATP-citrate lyase has been identified as a key enzyme in algae and higher plants, which catalyzes the formation of cytosolic acetyl-CoA and oxaloacetate from citrate[32, 33]. It has been reported in a proteomic study [9] that expression of ACL increases under nitrogen starvation and is associated with induced lipid accumulation. Unlike mammalian cells,

A.

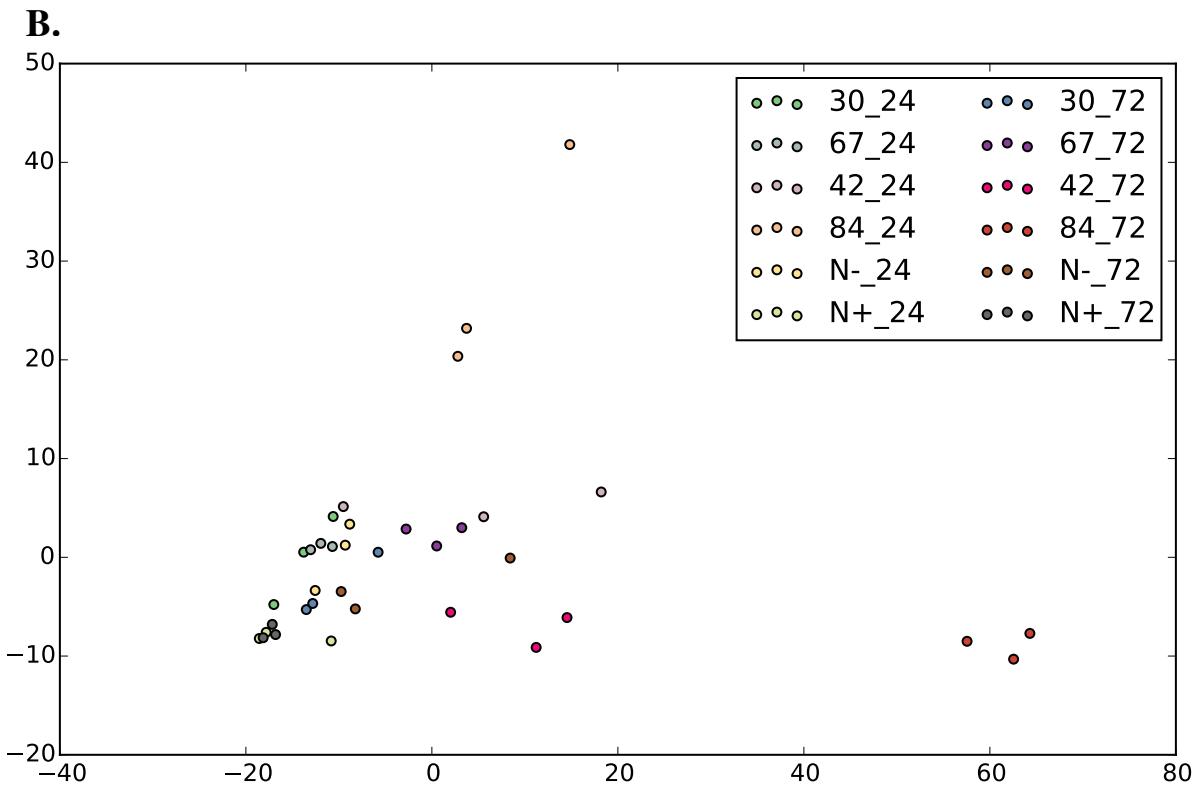
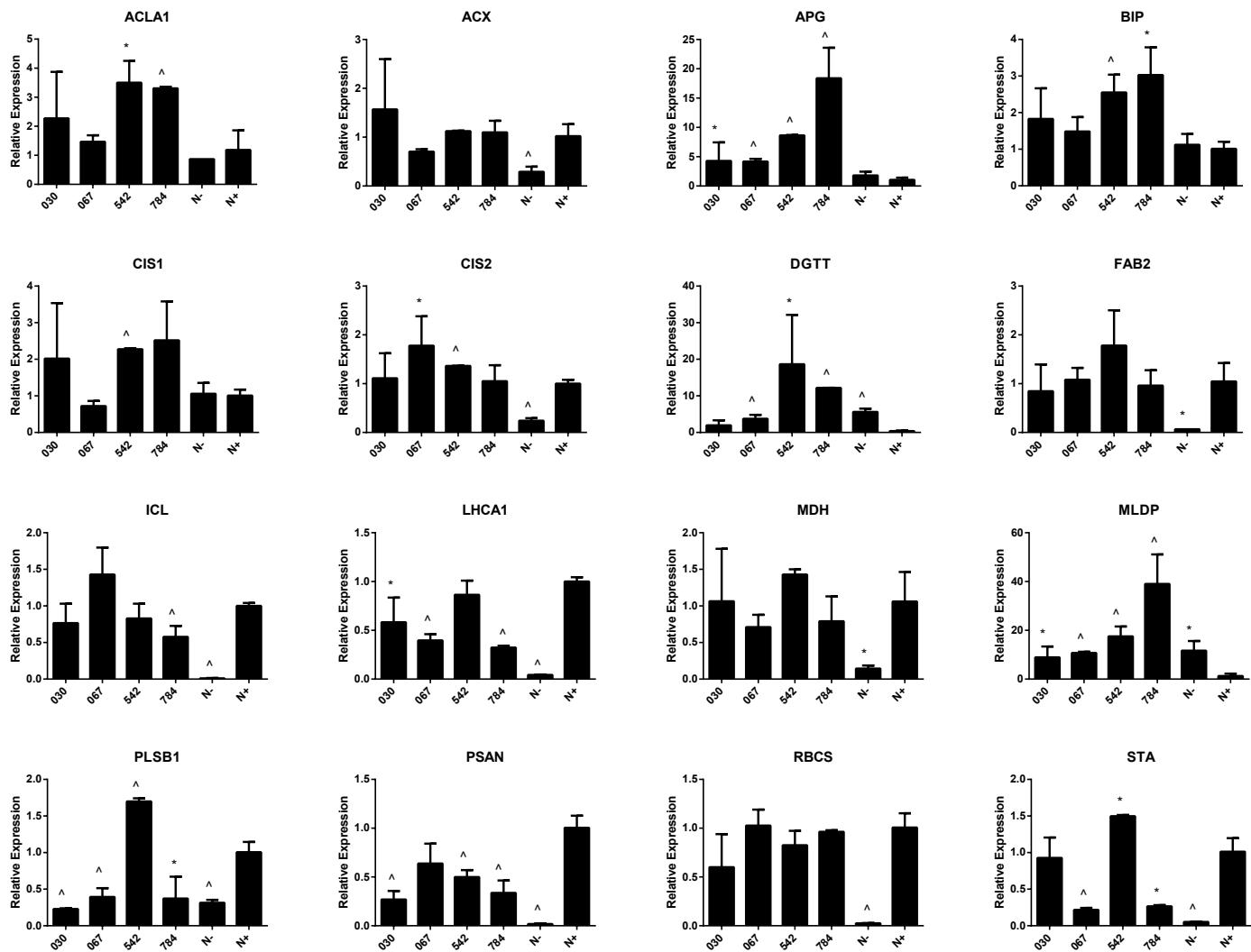


Figure 2.5: Summary of differential gene expression in qPCR. **A.** Heatmap of gene expression relative to control. The log₂ fold change of expression of each sample was plotted on a color scale normalized with the 10th and 90th percentile of all samples. Hierarchical clustering was performed on columns using city block metric and on rows using correlation metric. Major clusters are designated with different colors. Samples were collected at 24h, 48h and 72h of incubation and 3 biological replicates were used for each condition. **B.** PCA plot of each condition at different time points. Two principle components were identified after dimension reduction on genes

Chlamydomonas has heterodimeric ATP-citrate lyase, consisted of ACLA and ACLB. ACLA1 was targeted in this work. We found that after 24 hours of compound treatment, ACLA1 was significantly up-regulated in cells treated with compound 42 ($p < 0.05$) and 84 ($p < 0.01$), 3- to 4-fold. After 72 hours of compound treatment, ACLA1 was up-regulated in cells treated with each compounds as well as nitrogen-starved cells. Expression was increased up to 8-fold in cells treated with compound 30 ($p < 0.01$), 14-fold in both 67 and 42 ($p < 0.01$) and the highest 21-fold increase in expression was observed in cells

A.

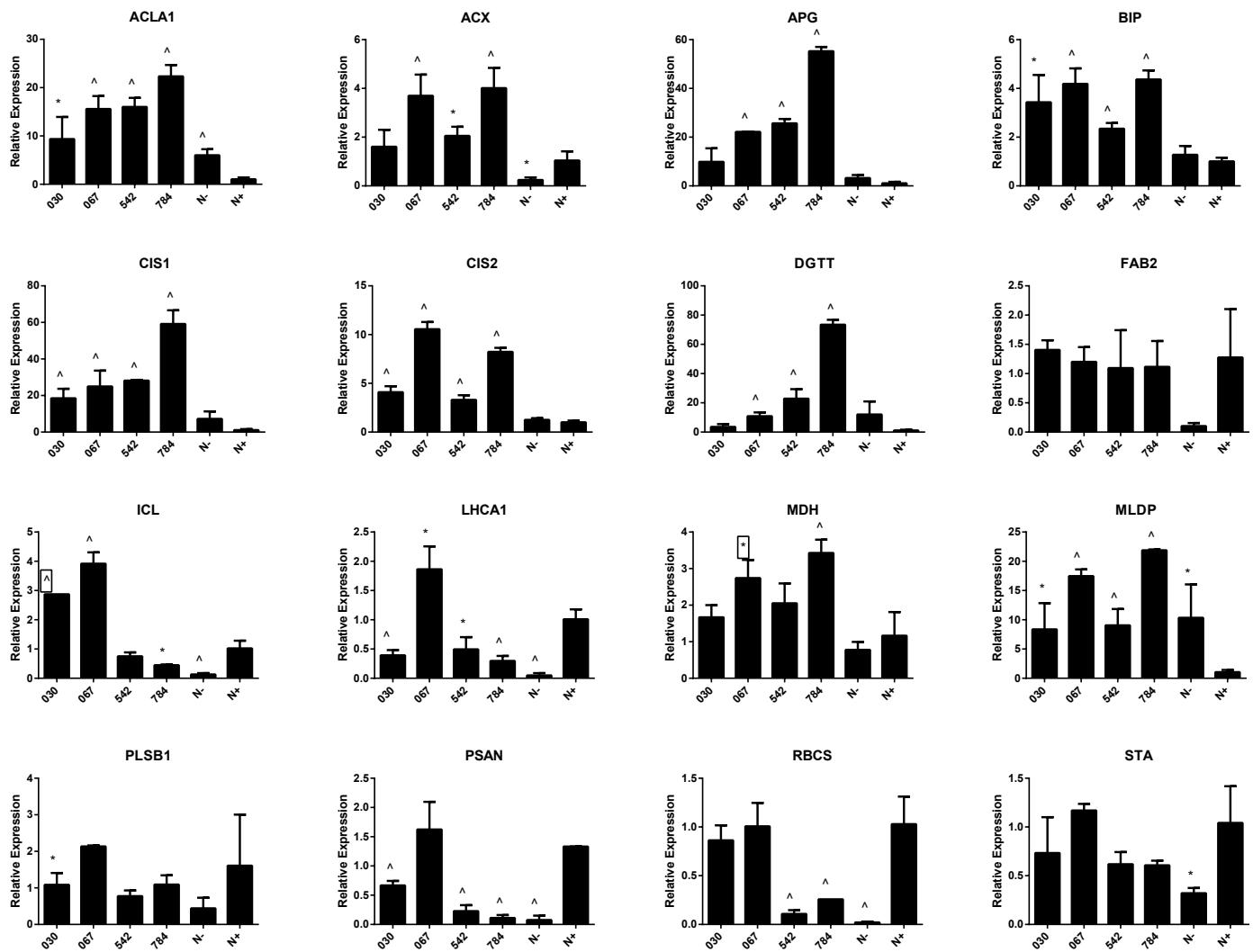
B.

Figure 2.6: Expression of selected genes under different treatments. Relative expression is the fold change of each condition relative to control (N+) normalized with the relative expression of reference gene RACK1. * indicates $p < 0.05$ and ^ indicates $p < 0.01$. **A.** Expression of selected genes after 24h of incubation. **B.** Expression of selected genes after 72h of incubation.

treated with compound 84($p < 0.01$). In comparison, expression of ACL in N-starved cells increased 4.9-fold ($p < 0.01$). These data suggest that ACL plays an important role in providing cytosolic acetyl-CoA for fatty acid biosynthesis.

There are also two isoforms of citrate synthase in Chlamydomonas, encoded by CIS1, the product of which is localized in mitochondrial and CIS2, the product of which is localized in glyoxysome or cytosol. CIS1 plays an important role in fatty acid biosynthesis since TCA cycle only occurs inside mitochondria, where citrate efflux can occur via the citrate transporters. As expected the expression of CIS1 was increased dramatically in both compound treated cells and N-starved cells. After 72 h of treatment, a 20-fold increase was observed in cells treated with compounds 30, 67, and 42, while a 58-fold increase was observed in cells treated with compound 84. In comparison, N starvation increased expression 6-fold ($p < 0.01$). The expression of CIS2 was decreased in N- ($p < 0.01$) cells after 24 h of incubation but was unchanged at 72 h. Importantly, in all compound-treated cells expression of CIS2 increased up to 10-fold, suggesting that the pathways involved in lipid accumulation in compound-treated cells is different from that in nitrogen-starved cells. As the glyoxylate cycle plays an anabolic role especially in carbohydrate biosynthesis, the expression of the gene STA6, encoding ADP-glucose pyrophosphorylase small subunit, a key enzyme required for starch synthesis was measured. Although the carbohydrate levels in the cell increased when treated with any of these compounds (Wase, Tu et al., in submission), STA6 did not increase. Expression of ICL, encoding isocitrate lyase and MDH, encoding malate dehydrogenase varied in response to treatment with the different compounds, suggesting different effects of compounds due to their structural differences. Further analysis presented in Chapter 3 helped elucidate the role of TCA cycle in lipid accumulation.

2.3.3 Fatty acid and TAG biosynthesis

ACX1, the gene encoding acetyl carboxylase, is a rate limiting enzyme that catalyzes the formation of malonyl-CoA through carboxylation of acetyl-CoA. In compound-treated cells, the expression of ACX increased from 24 h to 72 h of incubation, 3- to 4-fold in cells treated with compound 67 and 84 ($p < 0.01$) and less than 2-fold in cells treated with compound 42 ($p < 0.05$). Another gene FAB2, encoding an acyl-ACP desaturase in the fatty acid biosynthetic pathway, was unaffected by compound treatment but was suppressed significantly by N starvation by 24 h of incubation. Unlike N starvation, compound treatment did not seem to suppress *de novo* fatty acid biosynthesis.

TAG is the major form of storage lipid in algae but the process of its biosynthesis has yet been completely elucidated. Diacylglycerol:acyl-CoA acyltransferease, which is considered to be a rate-limiting enzyme that catalyzes the committed step in TAG synthesis [34], was found to be significantly up-regulated during lipid accumulation induced by nitrogen starvation [35]. In the present study, DGTT1 was significantly up-regulated in cells treated by all 4 compounds, 4-fold, 11-fold, 23-fold and 73-fold in cell treated with compound 30, 42, 67 and 84, respectively. It has been previously shown that the major lipid droplet protein, encoded by MLDP, plays a major role in assisting the formation of lipid droplet and affects the accumulation of TAG [36]. The expression of MLDP was increased up to 40-fold in compound-treated cells after 24 h of treatment but was only increased 20-fold after 72 h of treatment, suggesting that the formation of lipid droplets mainly occurs between 24h to 72h of incubation.

2.3.4 RuBisCO and light reaction components in photosynthesis

One of the main concerns with lipid accumulation induced by nutrient starvation is that photosynthetic components are often degraded under global cellular stress. Therefore, one

of the main goals of this study was to identify compounds that induce lipid accumulation with minimal impact on photosynthesis. We found that unlike N-starved cells, ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo) small subunit subunit 2, encoded by RBCS, was not significantly down-regulated after 24 h of treatment with any of the 4 compounds. After 72 h of treatment, the expression of RBCS2 decreased by more than 50% in cells treated with compounds 42 and 67, but was still much higher than that of cells starved from N. The expression of RBCS in cells treated with compounds 30 and 67 remained unchanged. Although the expression of LHCA1 and PSAN, which encode light-harvesting complex and photosystem I reaction center subunit N, decreased by %50 in cells treated with all compounds except 67, compared to N- condition, the light reaction components were not severely affected.

2.3.5 Assessment of expression of the BIP and APG8 genes

It is well known that lipid accumulation induced by nitrogen starvation is usually accompanied by autophagy and increased protein turnover [5, 35]. To examine the stress response of the cells to compounds, the expression levels of APG8, encoding a key autophagy regulator CrATG8 [37] and BIP, a heat shock protein associated with ER stress, were measured. Although nitrogen starvation is known to cause a wide range of stress responses, BIP and APG8 were not significantly changed in cells deprived of nitrogen in this experiment. By comparison, all 4 of the selected compounds increased expression of BIP 3- to 5- fold and APG8 by 7- to 55-fold. These data suggest that cells treated with compounds may respond to stress in a different mechanism from nitrogen starvation. Although these two biomarkers for cellular stress were significantly up-regulated, the growth of cells treated with compounds was not severely affected, suggesting that regulation of the protein level might be involved.

Chapter 3

Transcriptome analysis using next-generation RNA sequencing

3.1 Quality assurance of RNA-seq experiment

It has been reported that the RNA sample with poor quality can have profound effects on the outcome of the experiment [38, 39]. Therefore, to ensure the quality of the input RNA, samples were analyzed on a non-denaturing agarose gel before library preparation. As shown in (Figure 3.1,) no significant degradation was observed in any of the samples. A NanoDrop 1000 was also used to check the A_{260}/A_{280} and A_{260}/A_{230} ratios to ensure little protein or other organic contaminant was carried over. Contamination was not detection in any of the samples (data not shown).

To ensure the quality of the raw sequencing read is suitable for subsequent analysis, FastQC software was used to perform statistical analysis on raw reads. The quality of each sample satisfied the recommended requirements. A representative plot of sequence quality scores is shown in Figure 3.2A.

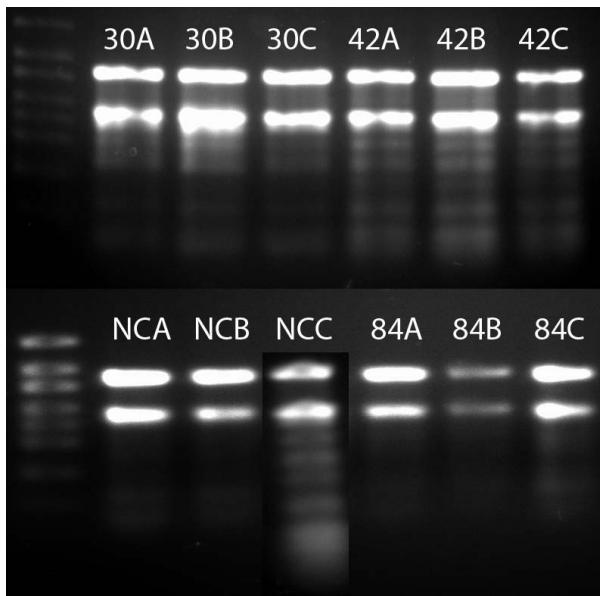


Figure 3.1: RNA gel electrophoresis image. Samples of RNA isolated from cells were checked for integrity on a non-denaturing agarose gel. 28S and 18S rRNA were present in the top two bands, respectively, with an approximate density ratio of 2:1. Smear of degraded RNA was not observed.

TopHat2 was used to map the all the raw reads to the genome to generate a single alignment (bam) file as described in 6.6.2. Among 168820835 bp of raw reads, 148533368 bp were mapped to the JGI v5.5 *Chlamydomonas reinhardtii* genome [23], giving an overall mapping rate of 88 %. Cufflinks was used to assemble the transcriptome from the alignment file in the reference guided mode, with the optional parameter '–max-bundle-frags 2500000' to ensure loci with highly abundant transcripts are included in the assembled transcriptome. Before transcript quantification and differential analysis, the distribution of transcript lengths was examined to ensure the proper functioning of the assembler (Figure 3.2B). Including novel isoforms, a total of 24837 transcripts were assembled, among which 19502 were annotated in Phytozome, which has a total of 19526 transcripts annotated for Chlamydomonas. The raw reads of each sample were separately aligned back to the assembled transcriptome using Bowtie2, giving an average mapping rate of 89%. These alignment files were used as inputs for eXpress to quantify the abundances of individual

transcripts in each sample.

3.2 Summary of differential gene expression analysis

3.2.1 Summary of overall differential gene expression patterns

To identify differentially expressed genes, the unnormalized counts from eXpress output were used as input for edgeR. The output of edgeR, including log2 fold change and false discovery rate (FDR) for each treatment and count per million (CPM) for each replicates, was combined as a single table with functional information from existing annotation from JGI and a complete interProScan v5.12 output. An in-house program (Appendix A) was created using the Scipy and Pyplot libraries to query data from this table and perform statistics.

To identify the numbers of significantly differentially expressed genes which are biologically meaningful, a set of criteria for "significance" must be defined. A single criterion on FDR usually results in many genes that are attributed to be statistically significantly differentially expressed but have fold changes too small to have an identifiable impact on biological processes [35]. In addition, some genes expressed at low abundance have coverage which is too low for the quantification results to be certain of validity. Therefore, in this study, the criteria of "significance" was defined as log2 fold change greater than 2 or less than -2, FDR less than 0.01 and total CPM of all samples greater than 20.

As shown in Table 3.1, according to the above criteria, 1216 (6.24%), 1413(7.25%) and 3144(16.12%) transcripts were differentially expressed in cells treated with compound 30, 42 and 84, respectively. While only 89 (0.46%) and 80 (0.41%) transcripts were uniquely

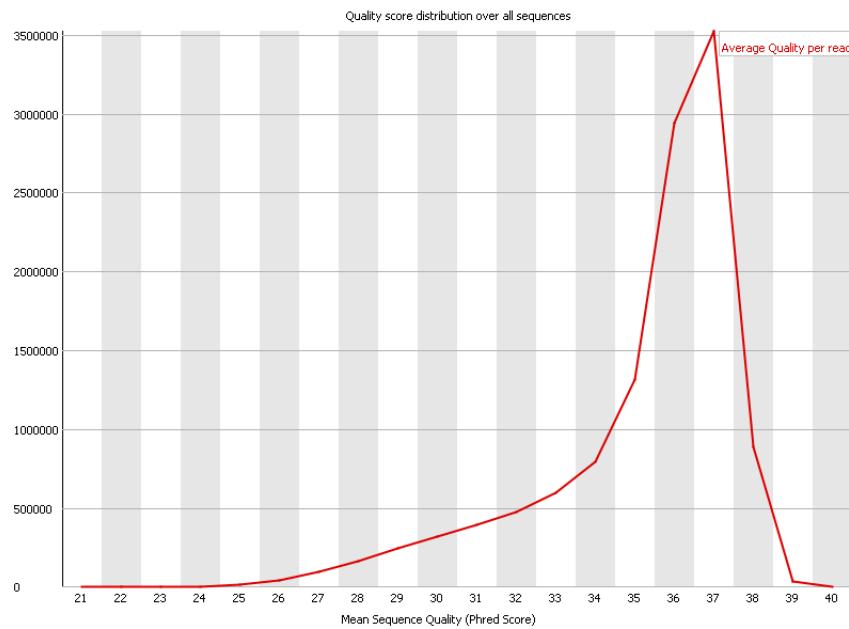
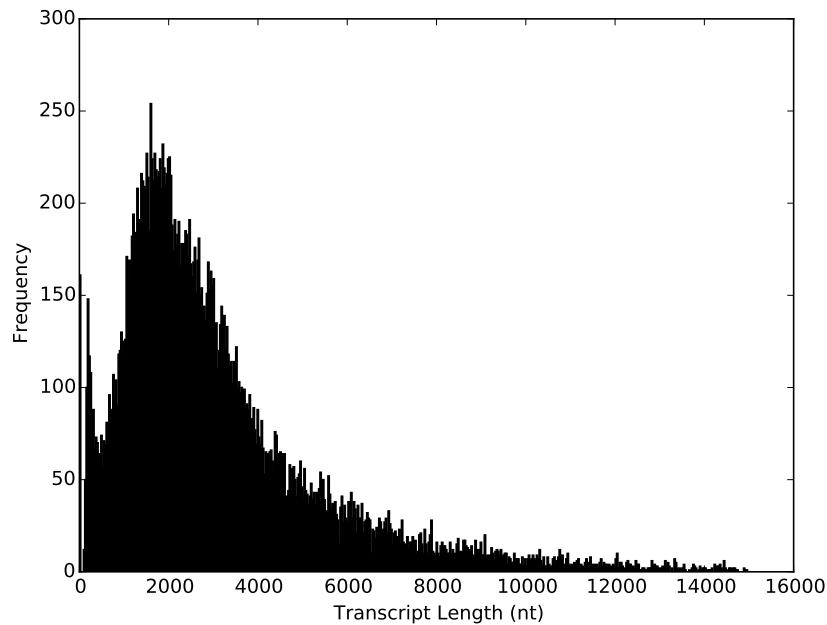
A.**B.**

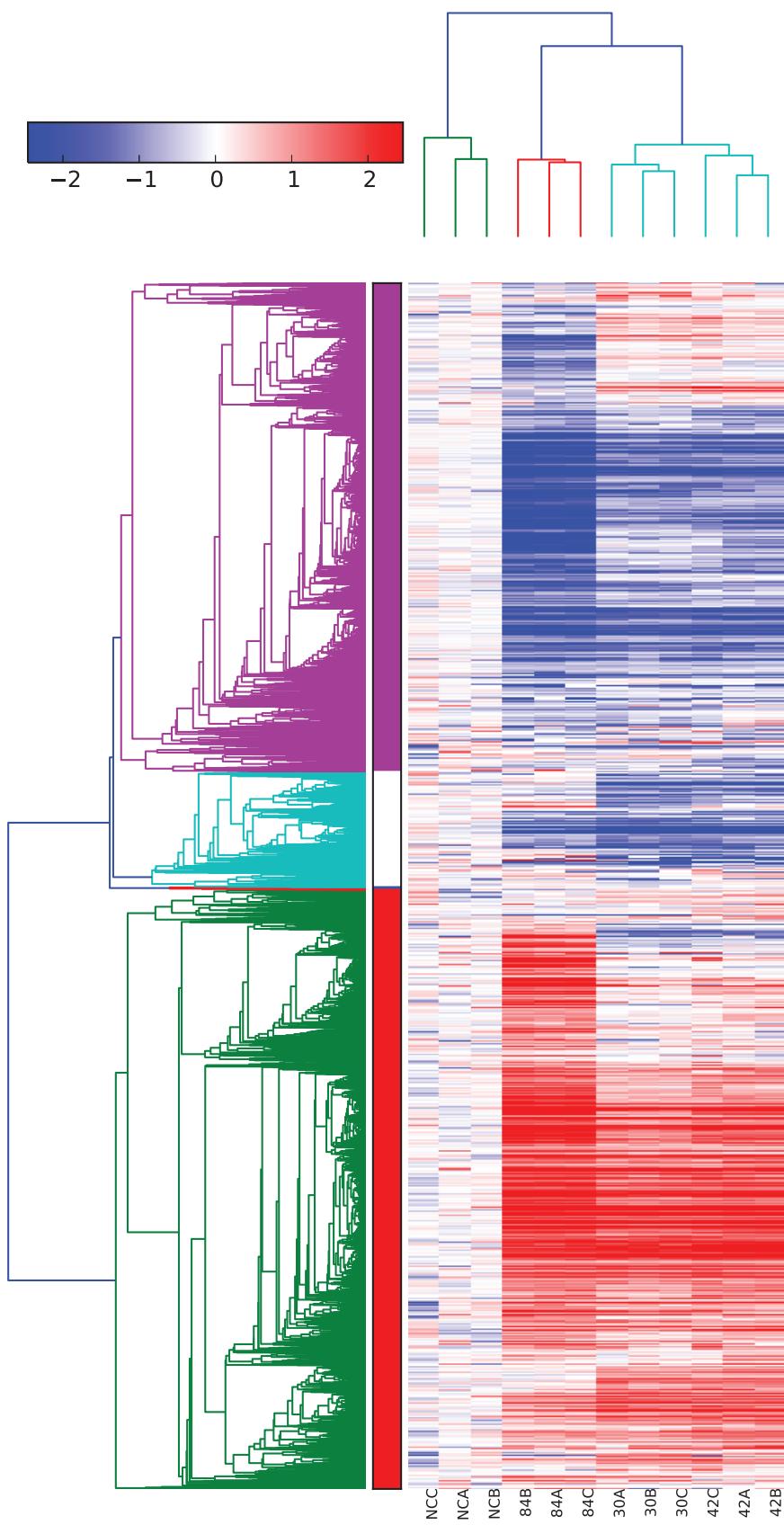
Figure 3.2: Quality check of raw sequencing reads and transcripts. **A.** Quality score distribution over all raw sequencing reads. **B.** Transcript length distribution over the assembly

differentially expressed in cells treated with compound 30 and 42, respectively, 806(9.41%) transcripts were uniquely differentially expressed in cells treated with compound 84. Of transcripts differentially expressed relative to control, more than 50% were found differentially expressed only in treatment with compound 84 while less than 10% were found only in treatment with compound 30 or 42. These different patterns of differential gene expression are also shown on the heat map (Figure 3.3A). These data suggest that while many transcriptional changes occurring in cells treated with compound 30 and 42 were shared by cells treated by compound 84, compound 84 had many unique transcriptional responses that might have contributed to its more pronounced lipid accumulation as well as cellular stress.

After dimension reduction on gene expression levels (Figure 3.3B) in principal component analysis (PCA), four clusters appeared on the plot, with each corresponding to a treatment/control condition. Interestingly, while clusters of compound 30 and 42 are close to each other, they are distant from the control cluster and the compound 84 cluster in two orthogonal dimensions. As a result, the 84 cluster and the control cluster are distant from each other in both dimensions. This clustering pattern suggest that the differential expression of some genes is shared by in cells treated with compound 30, 42 and 84, while other genes distinguish cells treated by compound 84 from cells treated by other two compounds.

3.2.2 Summary of the most significant differences in gene expression after treatment with the 3 compounds

To identify genes with most significant differences in expression between compound treatment and control, data entries were filtered by their FDR values and the data entries with the 50 lowest values were selected. In addition to FDR, selection criteria also included combined CPM > 20 and only annotated genes were selected for the convenience of anal-

A.

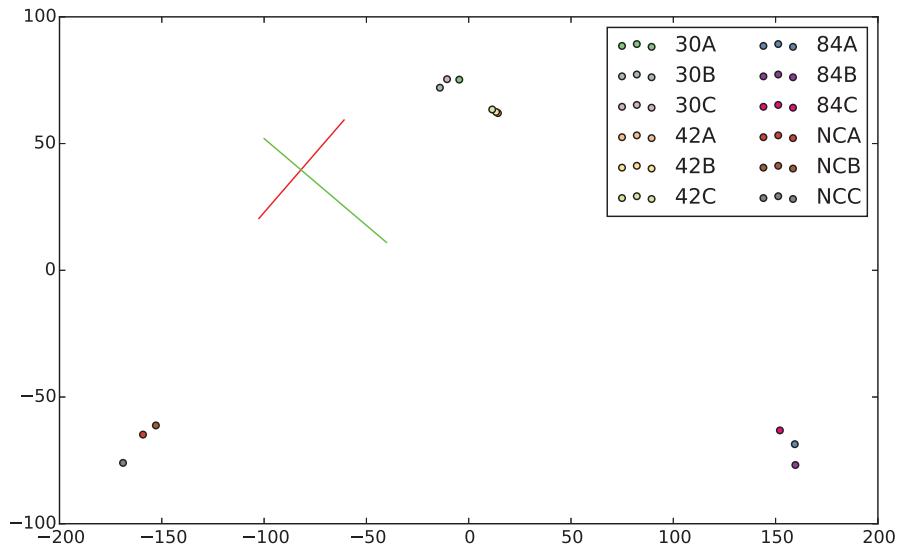
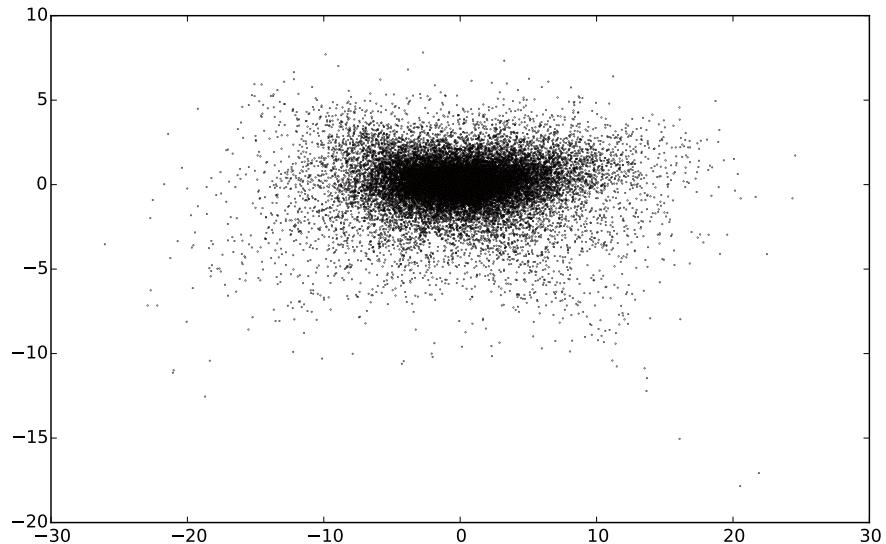
B.**C.**

Figure 3.3: Summary of global gene expression profiles. **A.** Heatmap of gene expression relative to control. The log₂ fold change of CPM of each sample was plotted on a color scale normalized with the 10th and 90th percentile of all samples. Hierarchical clustering was performed on columns using city block metric and on rows using correlation metric. Major clusters are designated with different colors. **B.** PCA plot of different conditions. Two orthogonal principal components, indicated by green and red lines, were identified after dimension reduction on transcripts. **C.** PCA plot of different genes. Two principal components were identified after dimension reduction on conditions.

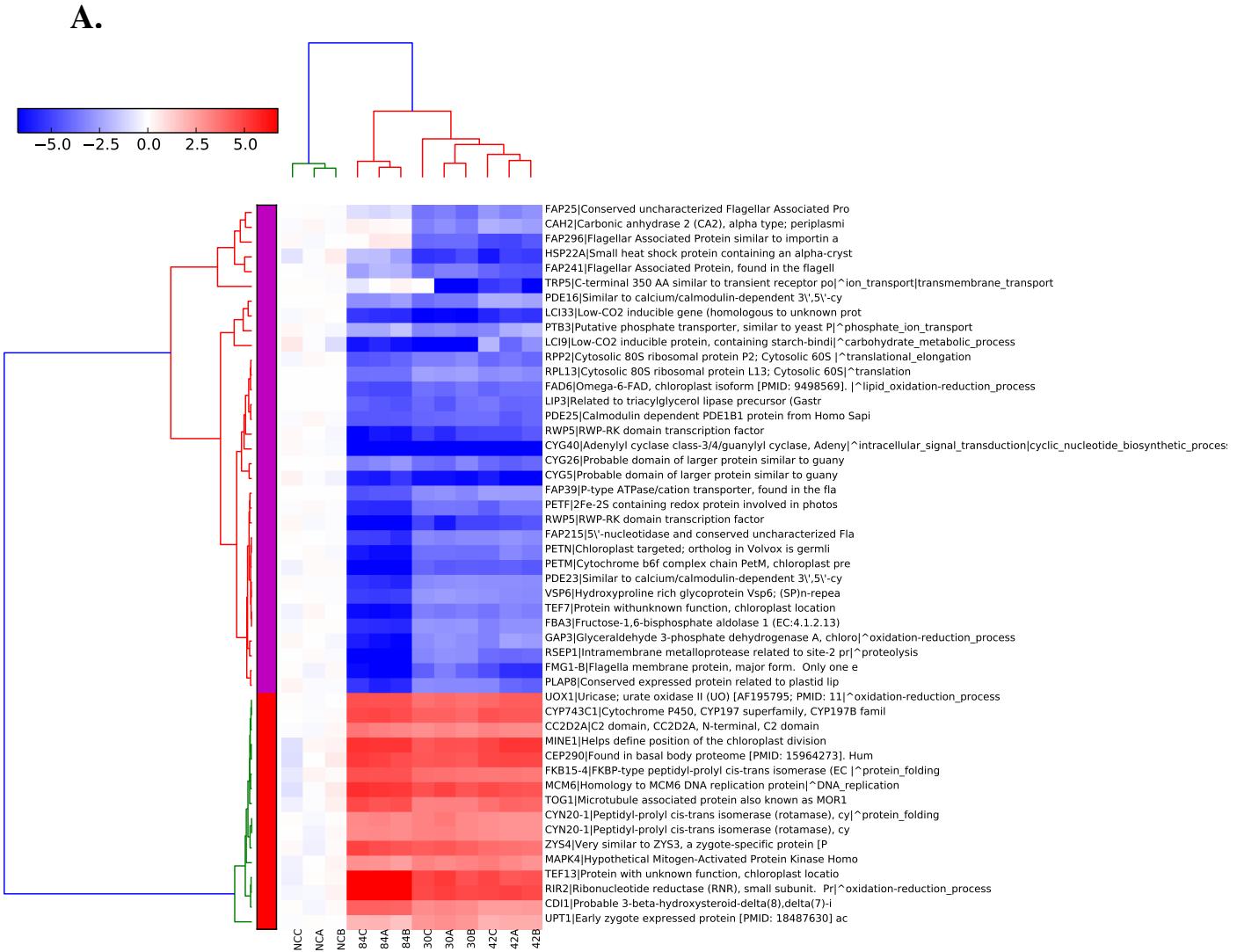
Table 3.1: Numbers of differentially expressed transcripts

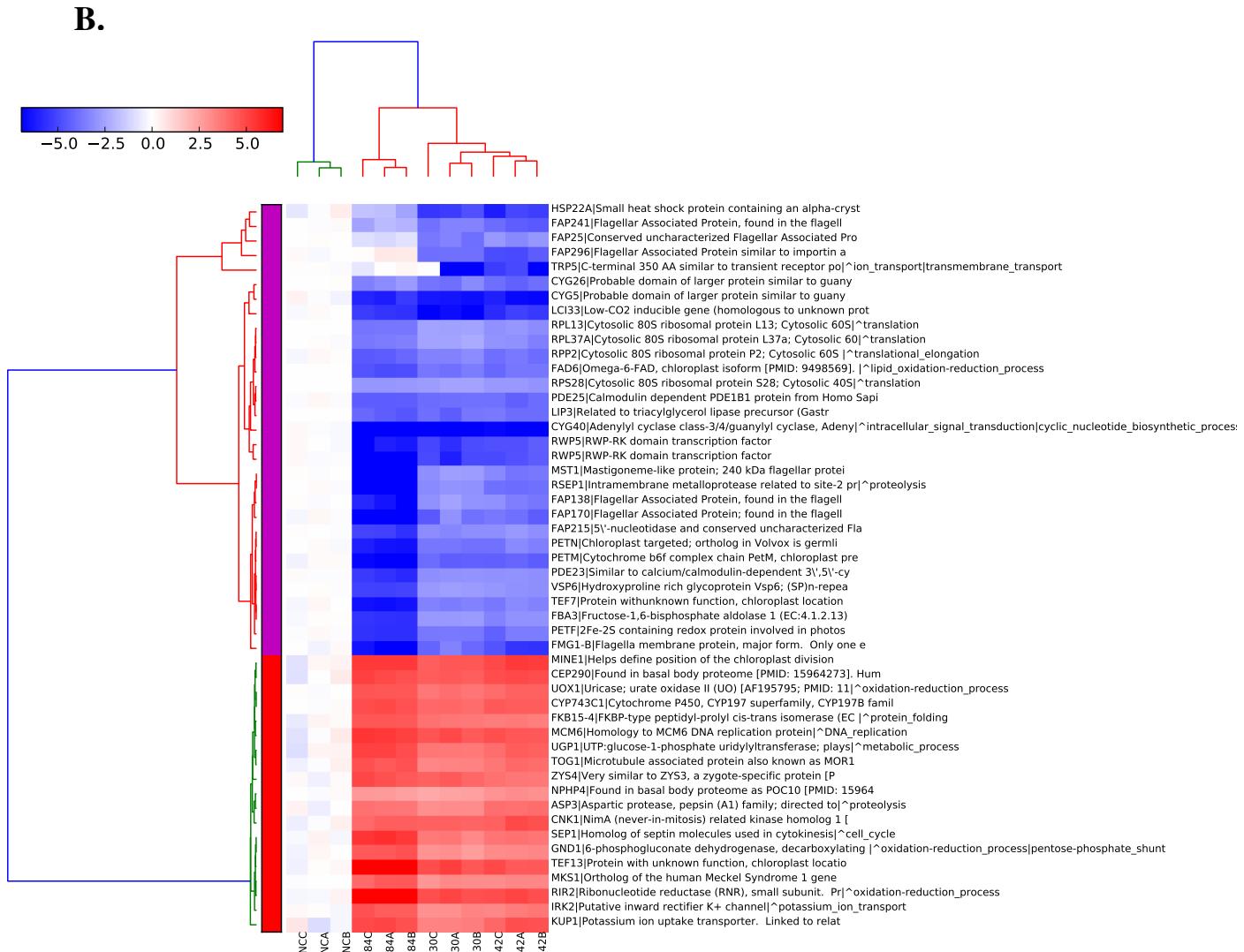
	Up	Down	Total	Percentage
30	754	462	1216	6.24%
42	917	496	1413	7.25%
84	1644	1500	3144	16.12%
30 ^{*a}	36	53	89	0.46%
42*	51	29	80	0.41%
84*	806	1029	1835	9.41%
30+42*	69	62	131	0.67%
30+84*	36	66	102	0.52%
42+84*	183	125	308	1.58%
Common	613	280	893	4.58%

^a * indicates that transcripts were differentially expressed only in the condition named in the current row. For example, the row of '42+84*' shows that 183 transcripts are significantly up-regulated in response to 42 and 84 but not 30.

ysis. Unannotated genes which provide additional opportunities in the analysis of lipid induction mechanism were considered in the Chapter 5 discussion.

Among the most significantly down-regulated genes in all 3 compounds are genes encoding flagellar associated proteins, ribosomal proteins, proteins localized in chloroplast such as cytochrome b6f and plastidic aldolase, triacylglycerol lipases, nucleotide cyclases, cyclic nucleotide phosphodiesterases and RWP-RK domain transcription factors. Among the most significantly up-regulated genes in all 3 compounds are genes encoding proteins involved in cell cycle and chloroplast division, DNA replication proteins, ribonucleotide reductase, cytochrome P450, and enzymes regulating redox homeostasis.





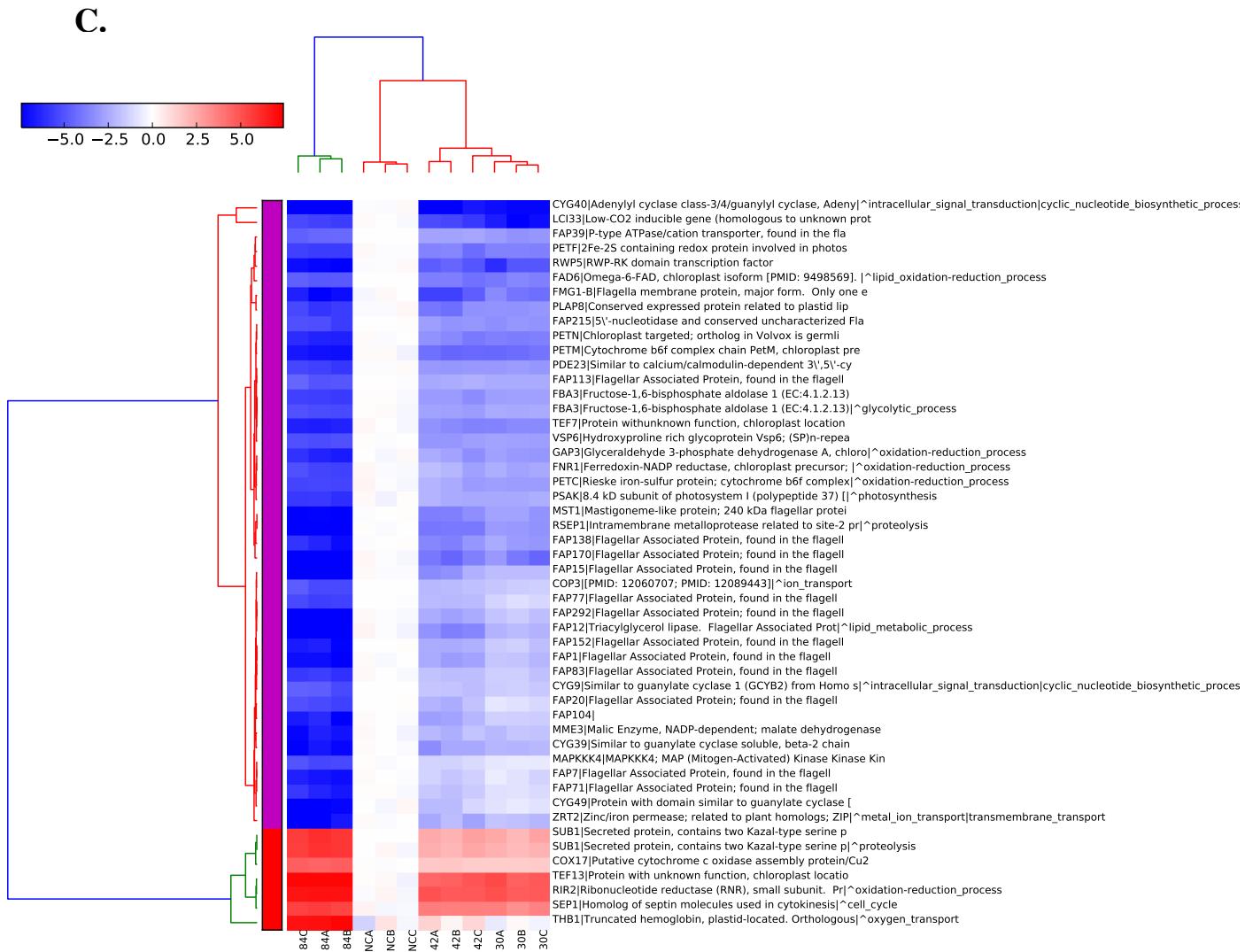


Figure 3.4: Expression profiles of most significantly changed transcripts. 50 annotated transcripts with lowest FDR in each condition were selected for analysis. The log2 fold change of CPM of each sample was plotted on a color scale normalized with the 10th and 90th percentile of all samples. Hierarchical clustering was performed on columns using city block metric and on rows using correlation metric. Major clusters are designated with different colors. **A.** Transcripts most significantly changed in response to compound 30. **B.** Transcripts most significantly changed in response to compound 42. **C.** Transcripts most significantly changed in response to compound 84.

3.3 Enrichment and pathway analysis

3.3.1 Summary of coordinately changed GO processes

As recommended by the author of GSEA algorithm, the criterion for identifying coordinately up- or down-regulated gene sets was FDR being less than 0.25 and nominal p-value less than 0.05 [27]. In Figure 3.4, several nucleotide cylases were among the most significantly down-regulated genes. As shown in Figure 3.2, the intracellular signal transduction which is closely related to the biosynthesis of cyclic nucleotide, was significantly down-regulated in all three treatments. A total of 181 transcripts were enriched in the category 'cyclic nucleotide biosynthesis process', whose coordinate shift could have very pronounced impacts on cell signaling and subsequence metabolic processes. Details of cell signaling will be examined in the next section. Although the gene set of proteins involved in translation was down-regulated in all three treatments, the gene set of tRNA-aminoacylation for protein translation was up-regulated in all three treatments, which may be a compensatory mechanism for the overall reduction of translational activities. In all three treatments, the up-regulation of genes required for protein glycosylation and ubiquitination was accompanied by the up-reguation of genes involved in intracellular protein transport/vesicle-mediated transport and proteolysis, suggesting that a systematic program of cellular component recycling and protein degradation was activated under compound treatment. The suppression of genes encoding photosynthetic proteins was also observed in all three treatment, which can be a result of autophagy and protein degradation. With the up-regulation of gene sets involved in DNA replication and various repair mechanisms, the increased expression of genes responsible for the regulation of mitotic metaphase-anaphase transition and microtubule-based movement suggests that the compounds have profound effects on the cell cycle. Unlike other two compounds, compound 84 possibly induced up-regulation of RNA-dependent DNA replication, down-regulation of protein methylation as

well as up-regulation of a large set of genes involved in oxidation-reduction process.

Table 3.2: Summary of significantly changed GO categories

GO term	Size	30			42			84		
		NES ^a	FDR	Up/down	NES	FDR	Up/down	NES	FDR	Up/down
(1→3)-beta-D-glucan biosynthetic process	8	1.70*	0.05	5/3	1.67*	0.05	5/3	0.83	0.90	6/2
ATP hydrolysis coupled proton transport	21	1.70* ^b	0.05	19/2	1.61*	0.07	18/3	1.71*	0.05	20/1
Base-excision repair	13	1.57*	0.12	8/5	1.77*	0.02	9/4	1.63*	0.07	8/5
Carbohydrate metabolic process	146	1.21	0.45	81/65	1.51*	0.13	80/66	1.32	0.29	79/67
Cell cycle	11	1.09	0.54	8/3	1.64*	0.07	8/3	1.50	0.14	6/5
Cellular metabolic process	28	1.83*	0.01	21/7	1.93*	0.00	21/7	1.81*	0.02	24/4
cGMP biosynthetic process	9	-1.92*	0.01	0/9	-1.85*	0.02	0/9	-1.93*	0.00	0/9
Cyclic nucleotide biosynthetic process	181	-2.52*	0.00	37/144	-2.50*	0.00	44/137	-2.76*	0.00	33/148
Defense response	7	-1.77*	0.03	1/6	-1.76*	0.04	1/6	-1.08	0.64	3/4
DNA integration	33	1.36	0.29	22/11	1.28	0.34	17/16	1.56*	0.11	19/14
DNA metabolic process	6	1.57*	0.12	4/2	1.34	0.28	5/1	1.24	0.38	4/2
DNA recombination	30	1.71*	0.05	18/12	1.68*	0.05	19/11	1.72*	0.04	20/10
DNA repair	78	1.17	0.49	49/29	1.29	0.34	57/21	1.62*	0.07	51/27

Table 3.2: (continued)

GO term	Size	30			42			84		
		NES	FDR	Up/down	NES	FDR	Up/down	NES	FDR	Up/down
DNA replication	47	2.76*	0.00	44/3	2.59*	0.00	45/2	2.59*	0.00	40/7
DNA topological change	8	1.55*	0.13	8/0	1.51*	0.12	8/0	1.70*	0.05	8/0
Glycerol metabolic process	14	1.57*	0.13	11/3	1.60*	0.07	11/3	1.94*	0.00	11/3
Glycerol-3-phosphate catabolic process	6	0.89	0.80	4/2	1.40	0.22	5/1	1.53*	0.12	4/2
Glycerol-3-phosphate metabolic process	7	0.95	0.74	5/2	1.43	0.19	5/2	1.60*	0.08	5/2
Glycolytic process	34	-1.40	0.27	13/21	-1.50*	0.16	13/21	-1.45	0.30	14/20
Histidine biosynthetic process	9	-1.45	0.22	1/8	-1.63*	0.11	2/7	-0.49	0.99	3/6
Intracellular protein transport	66	2.26*	0.00	55/11	2.22*	0.00	51/15	1.98*	0.00	49/17
Intracellular signal transduction	187	-2.41*	0.00	43/144	-2.36*	0.00	50/137	-2.66*	0.00	39/148
Ion transport	74	-1.46	0.25	21/53	-1.63*	0.12	17/57	-1.51*	0.22	18/56
Metabolic process	473	1.15	0.48	253/220	1.40*	0.22	252/221	1.28	0.34	265/208
Metal ion transport	39	-1.22	0.40	19/20	-1.07	0.71	20/19	-1.71*	0.06	19/20
Microtubule cytoskeleton organization	7	1.51*	0.14	6/1	1.46	0.17	6/1	1.26	0.37	5/2
Microtubule-based movement	77	2.17*	0.00	50/27	1.62*	0.07	41/36	1.78*	0.02	44/33

Table 3.2: (continued)

GO term	Size	30			42			84		
		NES	FDR	Up/down	NES	FDR	Up/down	NES	FDR	Up/down
Mismatch repair	17	2.02*	0.00	13/4	2.21*	0.00	14/3	2.03*	0.00	17/0
Multicellular organismal development	6	1.54*	0.13	5/1	1.63*	0.07	3/3	0.96	0.74	2/4
Nucleosome assembly	57	-1.04	0.60	33/24	-1.52*	0.18	25/32	1.64*	0.07	38/19
Oxidation-reduction process	475	1.15	0.49	262/213	1.24	0.37	260/215	1.39*	0.23	259/216
Oxygen transport	13	-1.55*	0.19	4/9	-1.25	0.44	5/8	-1.29	0.38	5/8
Phospholipid biosynthetic process	9	1.49	0.15	8/1	1.62*	0.07	7/2	1.21	0.40	8/1
Phosphorelay signal transduction system	30	-1.90*	0.01	3/27	-2.17*	0.00	2/28	-2.37*	0.00	1/29
Photosynthesis	41	-1.83*	0.02	10/31	-1.57*	0.15	10/31	-2.14*	0.00	5/36
Photosynthesis light harvesting	30	-2.72*	0.00	3/27	-2.47*	0.00	6/24	-2.39*	0.00	4/26
Protein folding	120	1.44*	0.20	69/51	0.93	0.74	70/50	0.69	0.96	73/47
Protein glycosylation	12	2.02*	0.00	10/2	1.78*	0.03	11/1	1.84*	0.01	11/1
Protein methylation	9	-1.46	0.24	1/8	-1.56	0.16	1/8	-1.71*	0.05	1/8
Protein polymerization	10	1.62*	0.09	8/2	1.06	0.61	6/4	1.15	0.47	6/4
Protein transport	31	1.73*	0.04	24/7	1.63*	0.07	26/5	1.88*	0.01	24/7

Table 3.2: (continued)

GO term	Size	30			42			84		
		NES	FDR	Up/down	NES	FDR	Up/down	NES	FDR	Up/down
Protein ubiquitination	48	1.50*	0.14	36/12	1.61*	0.07	40/8	1.69*	0.05	38/10
Proteolysis involved in cellular protein catabolic process	18	1.88*	0.01	18/0	1.97*	0.00	18/0	1.99*	0.00	18/0
Proton transport	7	1.53*	0.14	7/0	1.20	0.40	7/0	1.40	0.22	7/0
Regulation of mitotic metaphase anaphase transition	6	1.65*	0.07	5/1	1.59*	0.07	6/0	1.53*	0.13	5/1
Riboflavin biosynthetic process	13	-1.74*	0.04	2/11	-1.51*	0.18	3/10	-1.45	0.29	4/9
RNA-dependent DNA replication	122	1.15	0.50	76/46	1.33	0.28	83/39	1.83*	0.02	92/30
rRNA processing	25	-1.52*	0.22	5/20	-2.08*	0.00	1/24	-1.79*	0.03	3/22
Signal transduction	84	-2.29*	0.00	16/68	-2.42*	0.00	15/69	-2.56*	0.00	10/74
Translation	255	-1.91*	0.01	93/162	-2.27*	0.00	98/157	-1.54*	0.19	106/149
tRNA aminoacylation for protein translation	39	1.51*	0.14	25/14	1.97*	0.00	25/14	2.26*	0.00	29/10
tRNA processing	26	-1.52*	0.20	7/19	-1.94*	0.01	2/24	-1.42	0.31	11/15

Table 3.2: (continued)

GO term	Size	30			42			84		
		NES	FDR	Up/down	NES	FDR	Up/down	NES	FDR	Up/down
Ubiquitin-dependent protein catabolic process	44	1.53*	0.13	32/12	1.59*	0.07	32/12	1.40	0.23	33/11
Vesicle docking involved in exocytosis	8	1.67*	0.06	8/0	1.73*	0.03	8/0	1.68*	0.05	8/0
Vesicle-mediated transport	44	2.29*	0.00	39/5	2.26*	0.00	38/6	2.10*	0.00	38/6

^aNormalized enrichment score for the condition specified. A positive score indicates up-regulation of the named gene set. A negative score indicates down-regulation.

^b* indicates nominal $p < 0.05$ and FDR < 0.25 . Only gene sets that were significantly changed in at least one condition were selected for this table.

3.3.2 Transcriptional regulation and signal transduction

As shown in Figure 3.5, all three compounds induced systematic shifts in expression of certain transcription factors. Among the most significantly down-regulated transcription factors are RWP-RK domain containing transcription factors, which were identified in Chlamydomonas and other higher plants as key regulators of nitrogen responses and of gametophyte development [40]. During nitrogen starvation, nearly all RWP-RK transcription factors were up-regulated [40, 41, 42], among which RWP4, RWP5 and RWP6 were most significantly down-regulated in cells treated with any of the 3 compounds. The expression of RWP5 was significantly decreased in cells treated with compound 30, 42 and 84, 32-fold, 16-fold and 64-fold, respectively. The expression of RWP7 was significantly decreased in cells treated with compound 30 and 84, but not in cells treated with compound 42, whereas the expression of RWP3 and RWP11 was significantly decreased in cells treated with compound 30 and 42 not in cells treated with compound 84. The expression of NIT2, another RWP-RK domain containing transcription factor was significantly decreased in cells treated with compound 30, 2-fold and in cells treated with compound 84 3-fold but not significantly in cells treated with compound 42. It has been reported that the gametogenesis in the Chlamydomonas reinhardtii minus mating type is controlled by MID, which is up-regulated during nitrogen starvation. With RNAi knockdown of MID, vegetative cells were unable to differentiate into gametes [42]. Although cells with plus mating type were used in this experiment, it has been demonstrated that plus type cells utilize a similar system for gametogenesis, regulated by a different RWP-RK transcription factor homologous to MID[43]. Additionally, in higher plants such as *A.thaliana*, RWP-RK proteins which belong to the same family of CrMID regulate early embryo development, which is also sensitive to nitrate [44], suggesting the highly conserved function of RWP-RK transcription factors. In the current experiment, the sequence of RWP5 is most similar to that

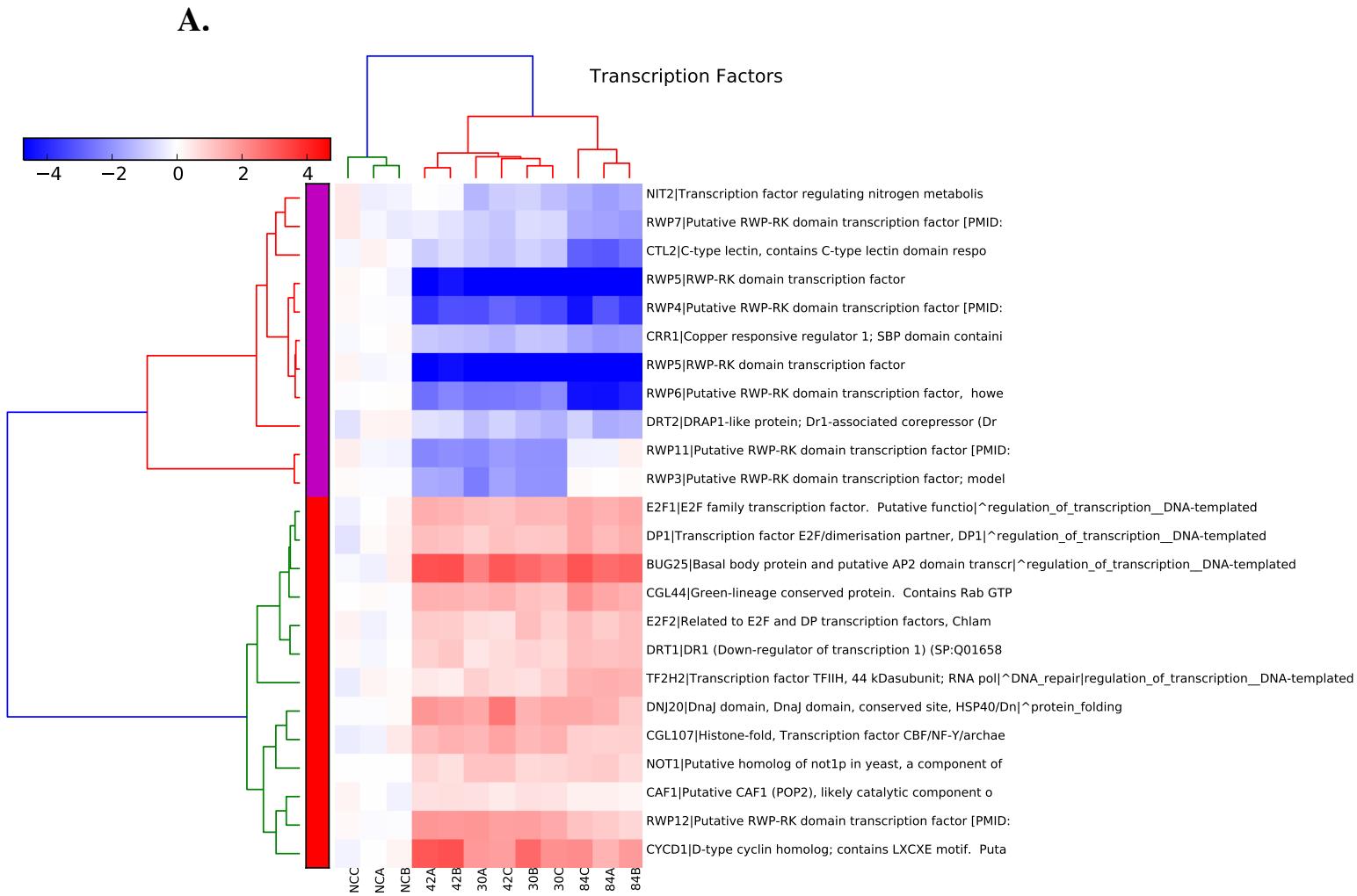


Figure 3.5: Expression profile of transcription factors. Annotated transcripts encoding transcription factors with total CPM > 10 and FDR < 0.01 in at least one condition were selected for analysis. The log2 fold change of CPM of each sample was plotted on a color scale normalized with the 10th and 90th percentile of all samples. Hierarchical clustering was performed on columns using city block metric and on rows using correlation metric. Major clusters are designated with different colors.

of MID. The very significant down-regulation of RWP5 and other RWP-RK transcription factors indicates that unlike nitrogen starvation which induces gametogenesis, compound treatment inhibits gametogenesis through transcriptional regulation.

NIT2, another RWP-RK domain transcription factor, has been identified as a transcription factor that binds to the promoter of NIA1, encoding a nitrate reductase. Its expression is maximized during nitrogen starvation and suppressed by ammonium [41]. A recent study [45] has shown that NO_3^- not only functions as a signaling molecule in nitrogen assimilation but also regulates carbon metabolism. When grown in medium containing NO_3^- , nitrate reductase deficient mutant nia1 significantly accumulated more TAG and starch as well as NO_3^- compared to wild-type control. In the current experiment, the expression of NIA1(Cre10.g453600) was significantly decreased in cells treated with compound 30, 42 and 84, 1.5-fold, 2.1-fold and 2.5-fold, respectively. These data suggest that NO_3^- signaling played an important role in the lipid accumulation induced by compounds.

Another significantly down-regulated transcription factor was CRR1, the expression of which was significantly decreased in cells treated with compound 30, 42 and 84, 2.2-fold, 2.4-fold and 3.4-fold, respectively. CRR1 encodes copper response regulator, a transcription factor with a plant-specific SBP DNA-binding domain [46]. A number of other SBP domain transcription factors with unknown functions were also significantly down-regulated in this experiment. When Cu was depleted, CRR1 transcripts only increased marginally, suggesting that the regulation of CRR1 is mainly on the protein level [46]. Nonetheless, the down-regulation of CRR1 led to significantly decreased expression of a number of fatty acid desaturases, which are targets of CRR1 [8]. For example, FAB2 and FAD3 were significantly down-regulated by 3-5 folds. In response to nitrogen starvation, NRR1, a SBP domain transcription factor, was identified as the primary regulator of the expression of DGTT1 [34]. Interestingly, NRR1 was not significantly changed during compound treatment, suggesting that another regulatory mechanism was responsible for

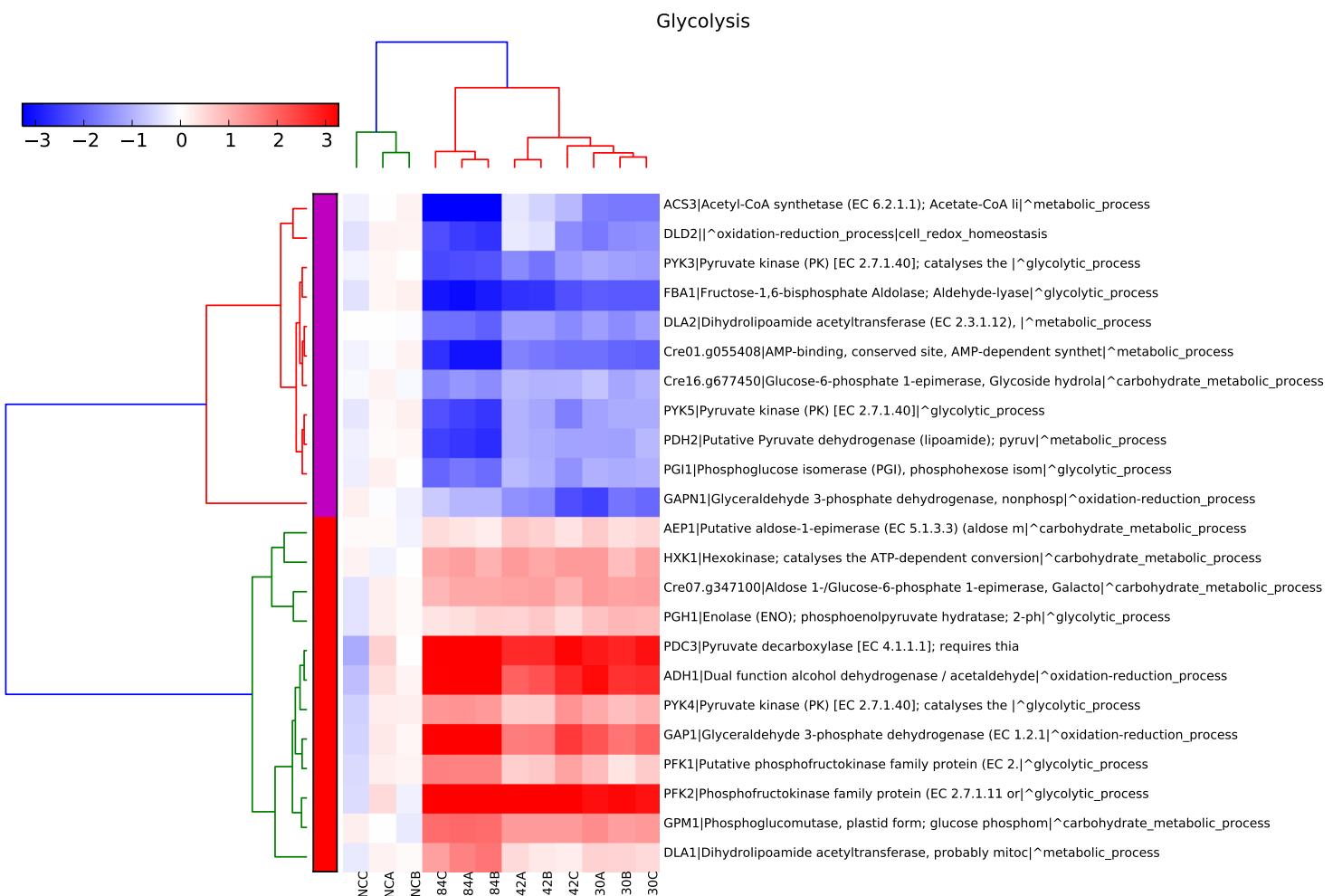
the up-regulation of DGTT1 and DGTT3.

Among significantly up-regulated transcription factors are regulators of cell cycle progression [47]: E2F1, E2F2 and DP1. The expression of E2F1 was significantly increased in cells treated with compound 30, 42 and 84, 2.4-fold, 2.4-fold and 3.0-fold, respectively. The expression of E2F2 was significantly increased in cells treated with compound 30, 42 and 84, 1.9-fold, 1.7-fold and 2.2-fold, respectively. The expression of DP1 was significantly increased in cells treated with compound 30, 42 and 84, 1.9-fold, 2.1-fold and 2.8-fold, respectively. Although the values of fold change were very small, all FDR values were less than 1×10^{-3} while the FDR values for DP1 and E2F1 were all less 1×10^{-5} , indicating the changes were very significant statistically. These data were consistent with significant changes in GO categories related to cell cycle progression.

3.3.3 Carbon metabolism

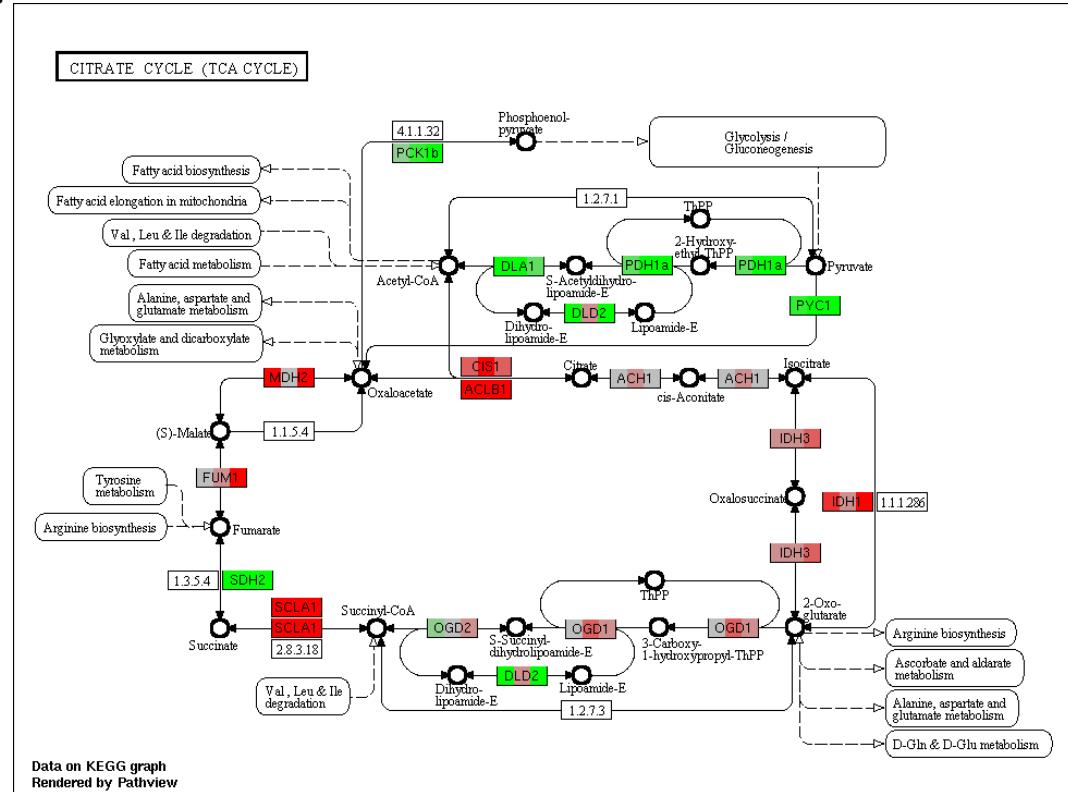
3.3.3.1 Glycolysis

The effect of compound treatment varied for different genes encoding glycolysis enzymes (Figure 3.6.) However, significant up-regulation was observed for PFK, encoding the rate-limiting enzyme phosphofructokinase. The expression of PFK1 was significantly increased in cells treated with compound 30, 42 and 84, 8-fold, 10-fold and 20-fold, respectively. Genes encoding other important enzymes such as hexokinase and glyceraldehyde-3-phosphate dehydrogenase were also significantly up-regulated in all three treatments. Interestingly, genes encoding pyruvate decarboxylase and alcohol dehydrogenase were also significantly up-regulated. Meanwhile, genes encoding pyruvate dehydrogenase complex were significantly down-regulated. The expression of PDC3, encoding pyruvate decarboxylase was significantly increased in cells treated with compound 30, 42 and 84, 7-fold, 7-fold and 19-fold, respectively and the expression of ADH, encoding alcohol dehydro-

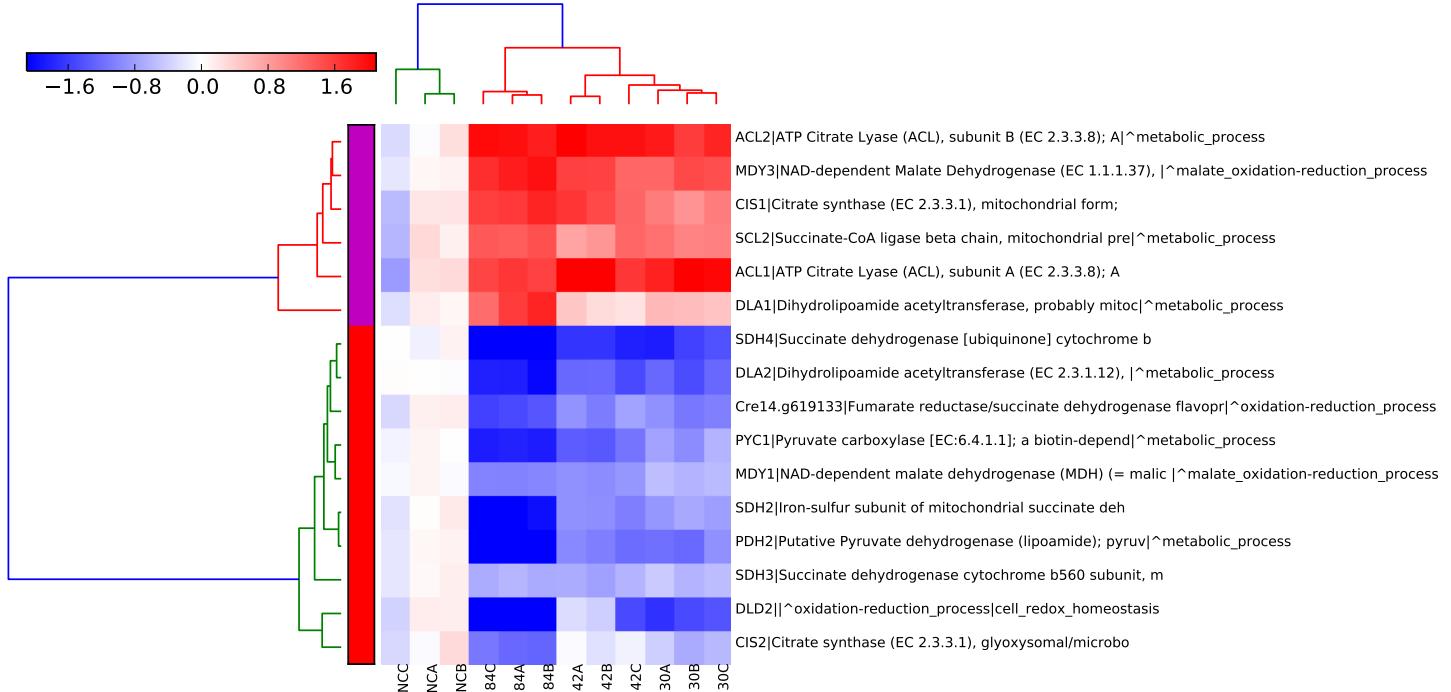
A.

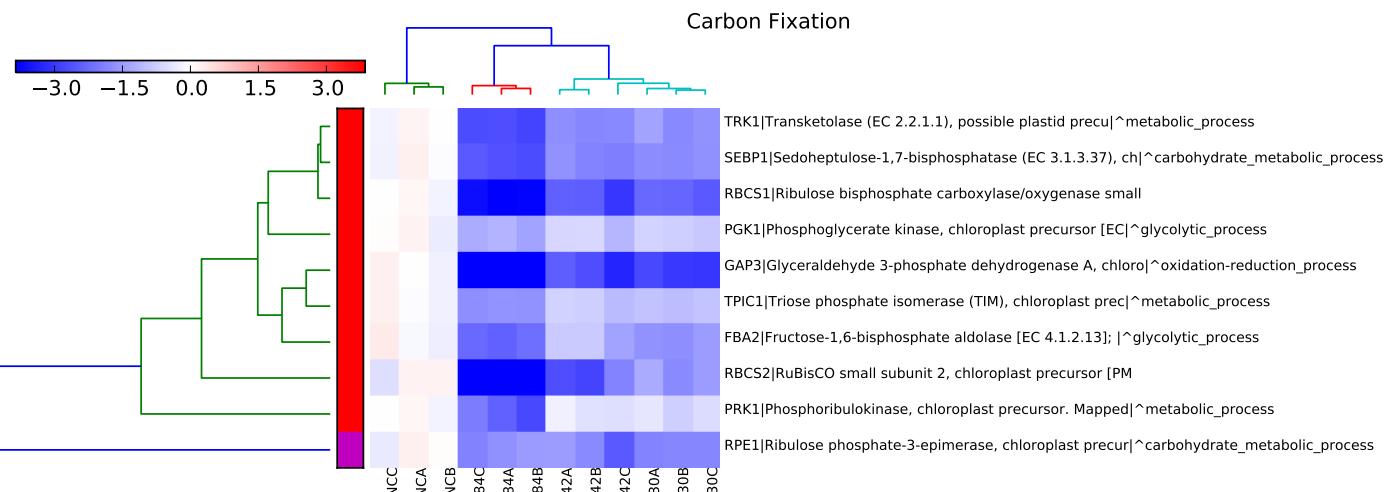
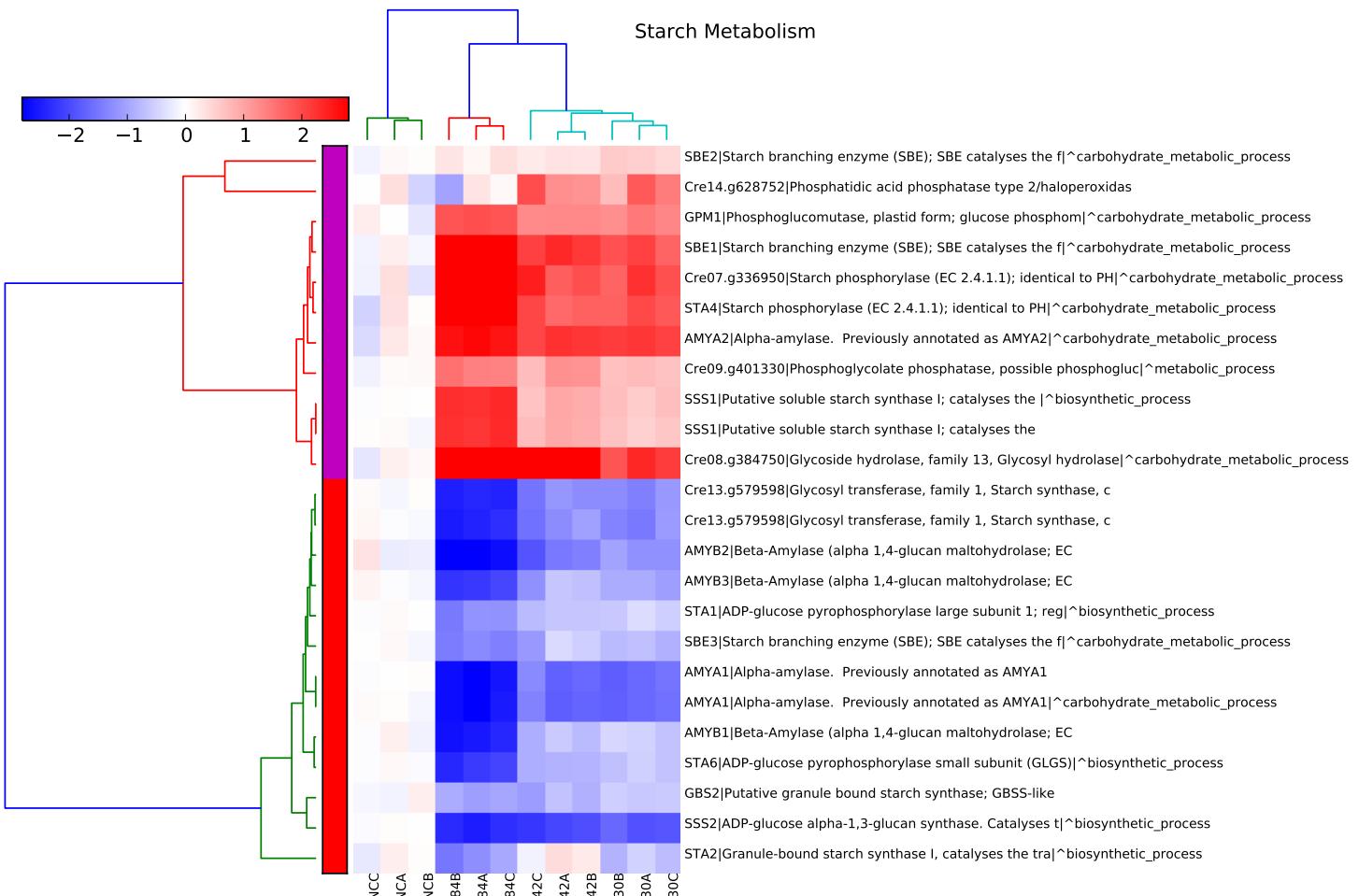
genase was significantly increased in cells treated with compound 30, 42 and 84, 7-fold, 5-fold and 11-fold, respectively. These data suggest that during compound treatment, glycolytic process was up-regulated and the carbon flux was mainly directed to fermentation rather than TCA cycle.

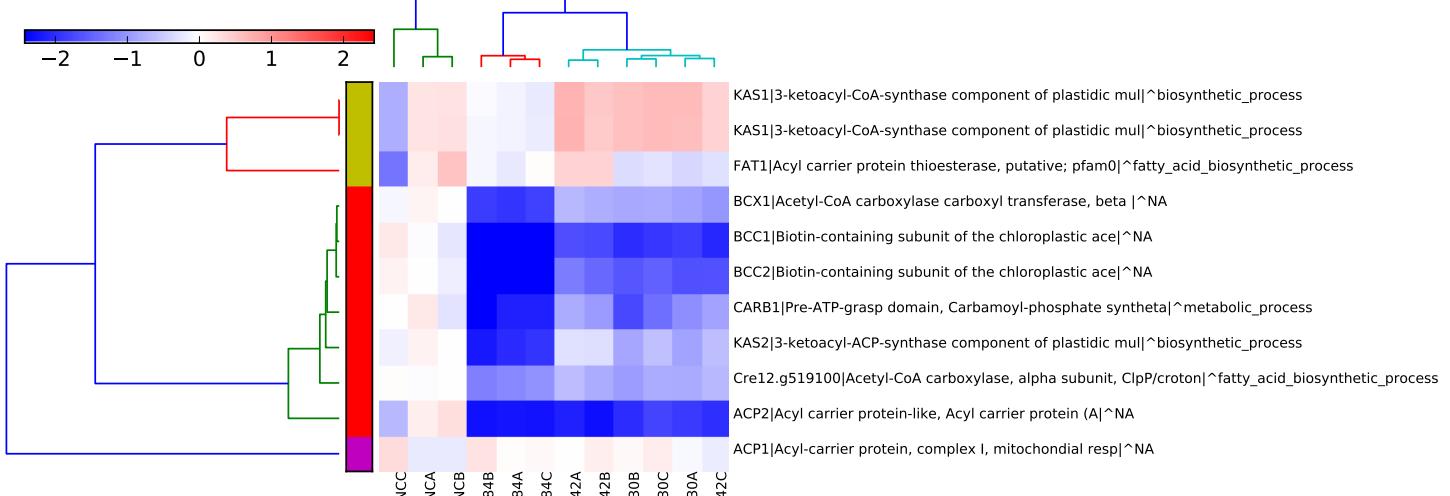
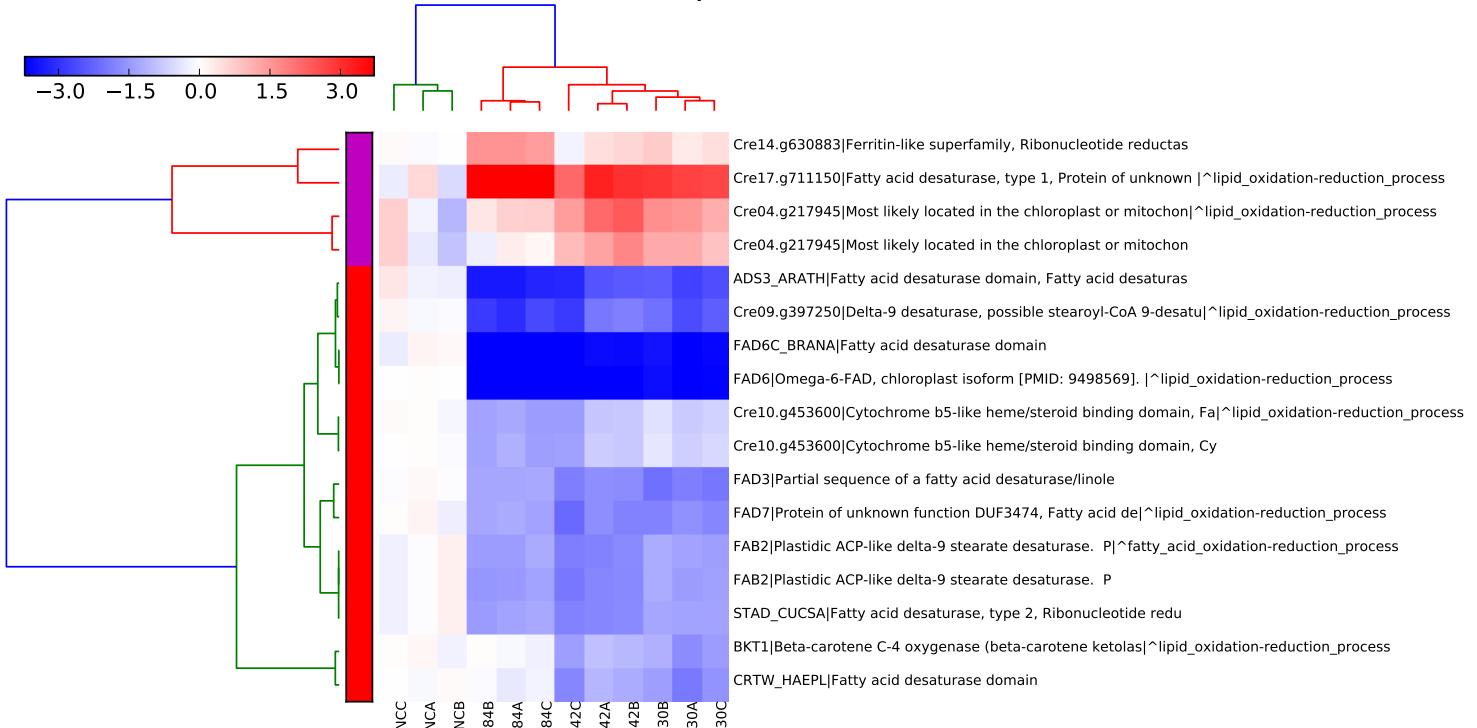
Unlike nitrogen starvation, which significantly increased the expression of ACS3 [48],

B.

Tca



C.**D.**

E.**Fatty Acid Biosynthesis****Fatty Acid Desaturation**

F.

Lipid Metabolic Process

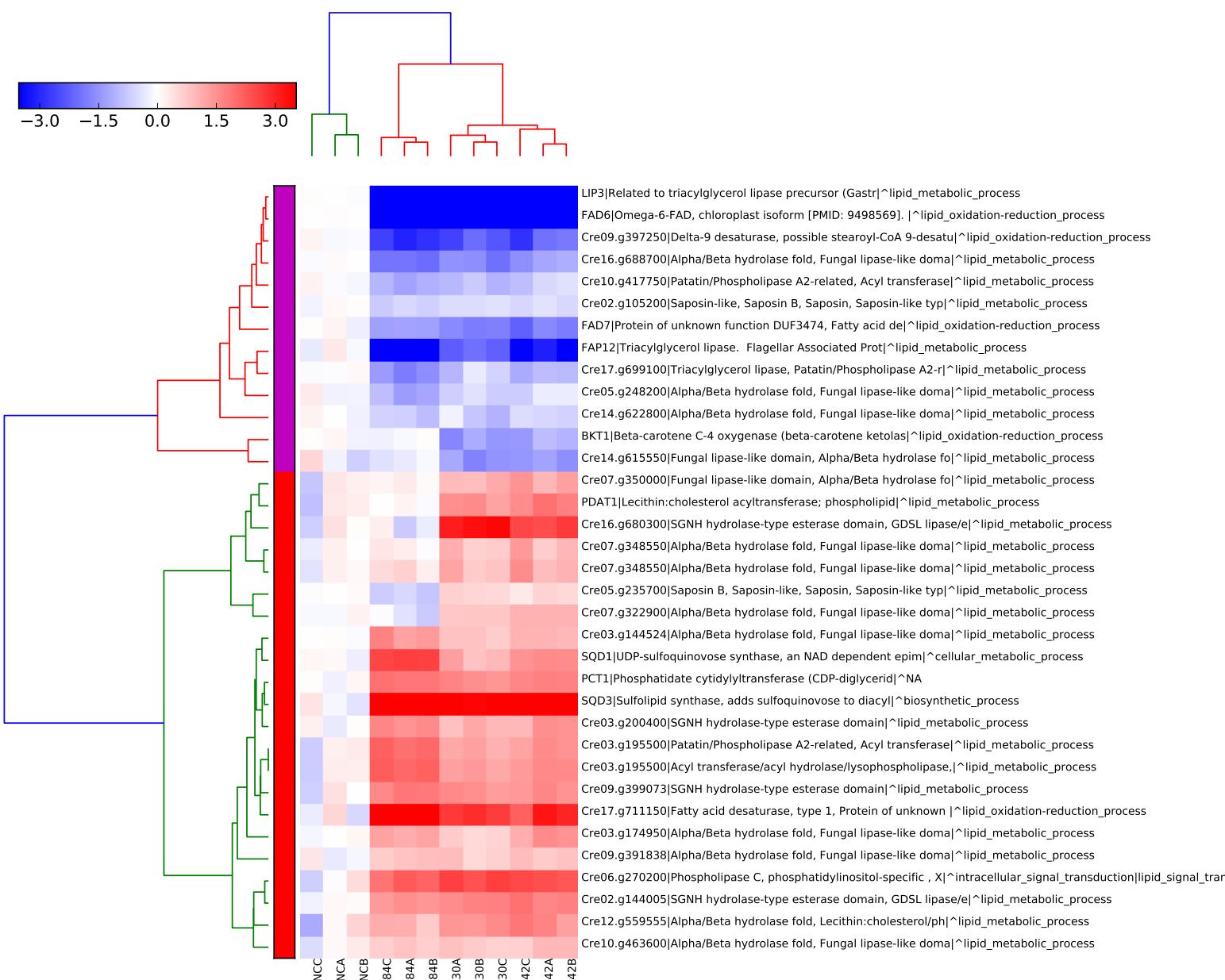


Figure 3.6: Expression profile of transcripts in coordinately shifted major metabolic pathways. Annotated transcripts encoding enzymes involved in specific pathways with total CPM > 10 and FDR < 0.01 in at least one condition were selected for analysis. The log₂ fold change of CPM of each sample was plotted on a color scale normalized with the 10th and 90th percentile of all samples. Hierarchical clustering was performed on columns using city block metric and on rows using correlation metric. Major clusters are designated with different colors. **A.** Expression profile of transcripts in glycolysis. **B.** Expression profile of transcripts in TCA cycle. **C.** Expression profile of transcripts in carbon fixation. **D.** Expression profile of transcripts in starch metabolism. **E.** Expression profiles of transcripts in fatty acid biosynthesis and desaturation. **F.** Expression profiles of transcripts in lipid metabolism.

encoding acetyl-CoA synthetase, which is responsible for the utilization of acetate in the media, the expression of ACS3 was significantly decreased in cells treated with compound 30, 42 and 84, 3.3-fold, 1.6-fold and 33-fold, respectively. Although acetate is the carbon source in TAP media used for cell culture, these data indicate that cells treated with compounds, in particular 84, decreased their acetate utilization and the carbon input for glycosis is likely from the degradation of starch, as the gene encoding phosphoglucomutase was also significantly up-regulated.

3.3.3.2 TCA cycle

While the expression many genes in the TCA cycle, including IDH, which encodes the rate-limiting enzyme isocitrate dehydrogenase, was not significantly change, genes encoding pyruvate dehydrogenase complex (PDH, DLD and DLA) and succinate dehydrogenase (SDH2, SDH3 and SDH4) were significantly down-regulated, particularly in cells treated with compound 84. The down-regulation of genes encoding PDC is consistent with the observation of the transcriptional change in glycosis where fermentation was favored. Succinate dehydrogenase is part of the electron transport chain in oxidative phosphorylation, where other co-expressed genes were also down-regulated. The expression of SDH2 was significantly decreased in cells treated with compound 30, 42 and 84, 1.7-fold, 2.0-fold and 4.2-fold, respectively. These relatively small changes are not likely to affect the TCA cycle activity to a considerable extent.

In contrast, genes required for the synthesis and efflux of citrate from mitochondria were significantly up-regulated. A significant up-regulation of 4 folds was observed in the expression of ACL1 and ACL2, encoding the heterodimeric ATP-citrate lyase, in cells treated with any of the three compounds. The expression of CIS1, encoding mitochondrial citrate synthase was significantly increased in cells treated with compound 30, 42 and 84, 2-fold, 2.5-fold and 3.3-fold, respectively and the expression of MDY3(MDH3), encod-

ing mitochondrial malate dehydrogenase was significantly increased in cells treated with compound 30, 42 and 84, 1.4-fold, 1.4-fold and 1.8-fold, respectively.

These data suggest when cells were treated with any of the three compounds, the TCA cycle played an important role in lipid accumulation by providing cytosolic acetyl-CoA for *de novo* fatty acid biosynthesis, as more oxaloacetate was accumulated, which was then converted to citrate and transported out to the cytosol. It is likely that the carbon flux was channeled from other cellular compounds since the pyruvate dehydrogenase complex was down-regulated.

3.3.3.3 Carbon fixation

The transcription of gene encoding nearly all the important enzymes involved in carbon fixation (Calvin cycle) was significantly down-regulated, including genes RBCS1 and RBCS2, which encode the heterodimeric small subunit of RuBisCo, the rate-limiting enzyme in Calvin cycle. However, the suppression of genes involved in Calvin cycle was less pronounced in cells treated with compound 30 and 42. The expression of RBCS1 was significantly decreased in cells treated with compound 30, 42 and 84, 5-fold, 6-fold and 13-fold, respectively and the expression of RBCS2 was significantly decreased in cells treated with compound 30, 42 and 84, 2-fold, 5-fold and 21-fold, respectively. These results were expected as many genes involved in autophagy were also significantly up-regulated.

3.3.3.4 Starch metabolism

A number of studies [48, 5, 12, 35] have shown that starchless mutant sta6 of *Chlamydomonas reinhardtii* accumulates more TAG during nitrogen starvation. When cells were treated with compounds, a number of genes involved in starch biosynthesis were down-regulated. The expression of STA6, the gene targeted in starchless mutants, which encodes small subunit of ADP-glucose pyrophosphorylase was significantly decreased in

cells treated with compound 30, 42 and 84, 1.6-fold, 1.9-fold and 4.5-fold, respectively. STA1, encoding the large subunit of ADP-glucose pyrophosphorylase, was also significantly down-regulated in cells treated with compound any of the three compounds. The expression of SSS2, encoding the major soluble starch synthase which favors elongation of amylopectin [49] was significantly decreased in cells treated with compound 30, 42 and 84, 1.6-fold, 1.9-fold and 4.5-fold, respectively, whereas expression of SSS1, encoding the minor soluble starch synthase which favors elongation of unbranched starch [49] was significantly increased in cells treated with compound 30, 42 and 84, 1.5-fold, 1.7-fold and 4.9-fold, respectively. Since highly-branched amylopectin needs additional enzyme for debranching, the unbranched starch synthesized by the gene product of SSS1 is more readily degradable for utilization in catabolic pathways. Several other genes encoding granule-bound starch synthases were also down-regulation in compound treated cells. Also significantly up-regulated in compound treated cells were STA4, encoding starch phosphorylase, AMY2A, encoding the major form of alpha-amylase and GPM1, encoding phosphoglucomutase. The up-regulation of these genes can facilitate the breakdown of starch as glucose which can then be activated by hexokinase, encoded by a gene which was also up regulated as seen in the previous section. The glucose-1-phosphate produced by starch phosphorylase can then be readily converted to glucose-6-phosphate for utilization in glycolysis. These data suggest that during compound treatment starch was store temporarily and its breakdown was highly favored. This is in contrast to nitrogen starvation in which starch storage is highly favored.

3.3.4 *de novo* fatty acid biosynthesis and desaturation

During compound treatment, a number of genes were significantly down-regulated, including Cre12.g519100 and BCX1, which encode the alpha and beta subunits of the car-

boxyltransferase subunit of acetyl-CoA carboxylase, the rate-limiting enzyme of fatty acid biosynthesis; BCC1 and BCC2, which encode biotin carboxyl carrier proteins; CARB2, which encodes the biotin carboxylase; KAS2, which encodes 3-ketoacyl-ACP-synthase; and ACP2, which encodes the major form of acyl-carrier protein. These changes were more pronounced in cells treated with compound 84, which induced a 16-fold decrease in the expression of BCC1, 3.6-fold decrease in BCX and 4.7-fold decrease in CARB1, whereas compound 30 or 42 induced a 3.8-fold decrease in BCC1, 2.8-fold decrease in BCX and 1.6-fold decrease in CARB1. Therefore, cells treated with compound 30 and 42 are likely still be able to synthesize fatty acid *de novo* at a moderate rate.

The vast majority of genes encoding plastidic fatty acid desaturases were significantly down-regulated under compound treatment, including FAB2, FAD3, FAD6 and FAD7. expression of FAD6, which encodes an omega-6 desaturase was significantly decreased in cells treated with compound 30, 42 and 84, 13-fold, 15-fold and 27-fold, respectively. In contrast, Cre17.g711150, a microsomal omega-6 desaturase was significant increased by 6 to 16 folds. However, this increase is unlikely to affect the lipid composition because the plastidic FAD6 is nearly 1000 fold higher in abundance.

3.3.5 Lipid metabolism

As expected, genes encoding diacylglycerol acyltransferases, which are required for TAG synthesis and accumulation, were significantly up-regulated in cells treated with any of the three compounds. Among the 5 DGTTs, DGTT 3, with an abundance of 33 CPM in control, was the major form of diacylglycerol acyltransferase in this experiment, the expression of which was significantly increased in cells treated with compound 30, 42 and 84, 2.0-fold, 4.3-fold and 4.5-fold, respectively. The expression of DGTT1 was significantly increased in cells treated with compound 30, 42 and 84, 7.0-fold, 21-fold and 70-fold, respectively, al-

though the abundance of this isoform in control was only 0.30 CPM. No significant change was observed in other isoforms, including DGAT1. While DGTT1 is not likely to significantly contribute to TAG synthesis under normal condition due to its low abundance, under compound treatment, the abundance of this isoform increased drastic and may have played a major role in TAG biosynthesis along with DGTT3. This pattern of change in the expression of different isoforms is similar to the observations in cells under nitrogen starvation, except for DGAT1 which was significant up regulated during nitrogen starvation[48, 35].

It has been reported that PDAT1, encoding phospholipid diacylglycerol acyltransferase, is also a major contributor of lipid accumulation under nitrogen starvation [34]. Unlike DGAT1 and DGTTs, phospholipid diacylglycerol acyltransferase does not use acyl-CoA as a substrate and converts diacylglycerol to triacylglycerol by transferring an acyl group to its sn-3 position from the sn-2 position of a phospholipid molecule [34]. Therefore, differences in the expression patterns for these genes, DGAT1 and DGTTs, can be expected. Indeed, only in cells treated with compound 30 and 42 was PDAT1 significant up-regulated, up to 3.2-fold. Along with PDAT1, a number of genes encoding putative lipases were also significantly changed only in cells treated with compound 30 and 42, suggesting that these genes might be co-regulated. A recent study has identified a novel phospholipase from the PDAT family in Arabidopsis, suggesting that these possibly co-regulated genes may have shared functions [50].

Among the most significantly down-regulated genes are FAP12 and LIP3, which encode two triacylglycerol lipases, respectively. The expression of FAP12 was significantly decreased in cells treated with compound 30, 42 and 84, 4-fold, 11-fold and 307-fold, respectively and the expression of LIP3 was significantly decreased in cells treated with compound 30, 42 and 84, 16-fold, 15-fold and 20-fold, respectively. The majority of other down-regulated genes encoding putative lipases are likely to be functionally similar to triacylglycerol lipases since the accumulation TAG suggests the overall decrease of TAG

lipase activities. Among up-regulated putative lipases, a number of them can be identified as phospholipase. Phospholipase A cleaves phospholipid to release free fatty acid while phospholipase C cleaves phospholipid to release diacylglycerol. These two products can be readily used for TAG synthesis in different steps.

Interestingly, two genes encoding enzymes utilizing phosphatidate as substrate were both significantly up-regulated. The expression of SQD3, encoding sulfolipid synthase was significantly increased in cells treated with compound 30, 42 and 84, 11.5-fold, 19.6-fold and 42.3-fold, respectively, while the expression of PCT, encoding CDP-diacylglycerol synthase was significantly increased in cells treated with compound 30, 42 and 84, 2.9-fold, 3.1-fold and 3.8-fold, respectively. Sulfolipid synthase catalyzes the synthesis of sulfoquinovosyldiacylglycerol (SQDG), while the product of CDP-diacylglycerol synthase reaction is eventually converted to phosphatidylglycerophosphate (PG). SQDG and PG are both important to the conformational stability and functional integrity of photosystem I (PSI) and photosystem II (PSII), while PG is required for the biosynthesis of PSI, SQDG is not. The latter, however, plays an important role in the functional and structural integrity of PSII [51]. Under P starvation, SQDG partially substitutes for PG in photosynthetic components increased as phospholipids were degraded to provide phosphorus for other essential cellular compounds [52]. The drastic difference between the fold change of SQD3 and PCT suggest that during compound treatment, a significant amount of SQDG was synthesized to partially replace degraded PG. Under S starvation, the level of sulfur acclimation protein, a protein kinase central to S response and one of its targets arylsulfatase, which releases sulfur from organic compounds, increased significantly [53]. Under the treatment of any of the three compounds, both SAC3, encoding sulfur acclimation protein and ASR1,3 encoding arylsulfatases were up-regulated. These data suggest that the drastic increase of SQDG synthesis may have partially depleted intracellular S.

3.4 Validation of differential expression with qPCR

To preliminarily confirm the accuracy of differential expression results computed from the sequencing data, four genes of interest were selected for qPCR assay using the methods described section 6.5. As shown in (Figure 3.7,) most of the relative expression values measured by RNA-seq were consistent with those measured by qPCR, with the exception of cells treated by compound 84. However, these discrepancies were most likely caused by the significant 4-fold down-regulation of the reference gene RACK1(CBLP) in cells treated with compound 84.

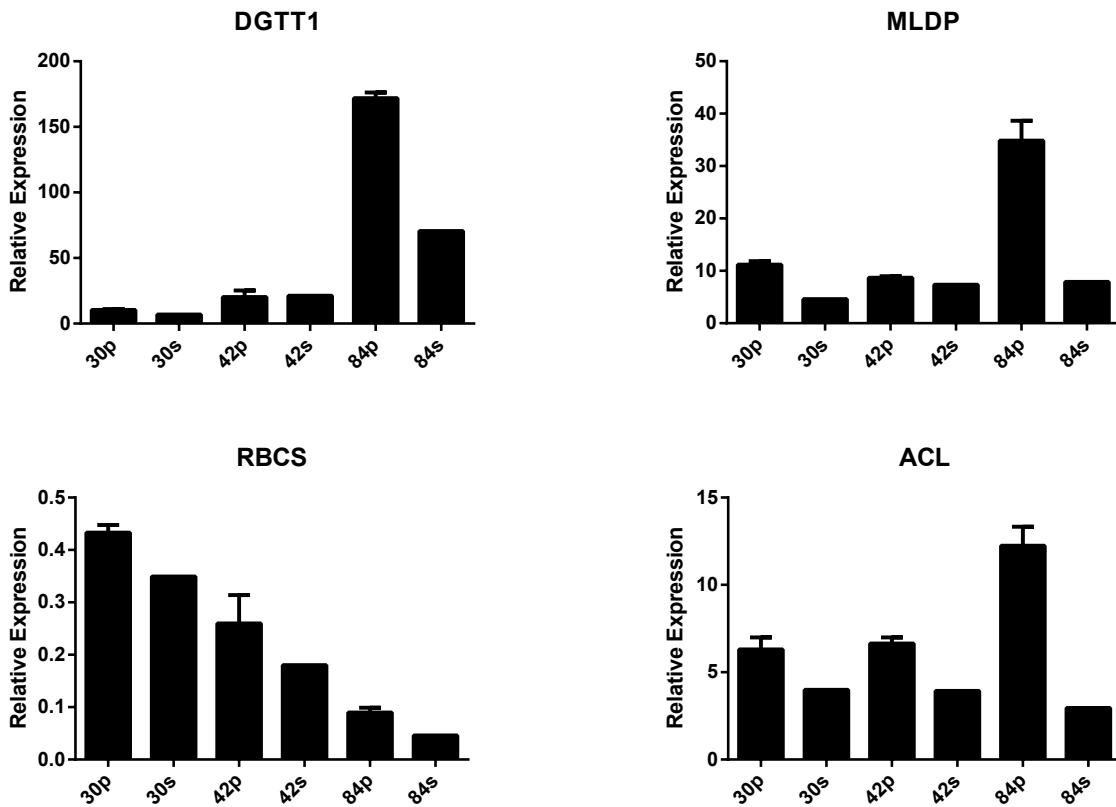


Figure 3.7: Comparison of the expression of selected genes using qPCR. In qPCR measurements (indicated with suffix "p"), relative expression is the fold change of each condition relative to control (N+) normalized with the relative expression of reference gene RACK1. In RNA-seq measurements (indicated with suffix "s"), relative expression is the fold change of CPM in each condition relative to control.

Chapter 4

Role of exogenous citrate in lipid production

4.1 Introduction

In previous chapters, genes involved in citrate efflux from mitochondria, such as ACL, encoding ATP-citrate lyase, CIS1, encoding mitochondrial citrate synthase and MDH1/3, encoding mitochondrial malate dehydrogenase, were found to be significantly up-regulated during compound treatment, suggesting that cytosolic citrate may play an important role in fatty acid biosynthesis, which leads to TAG accumulation. Here, we provided cells with citrate supplementation in the media and showed that increase in cytosolic led to significant increase in TAG accumulation.

4.2 Growth and lipid accumulation

4.2.1 Growth assessment

To assess the effect of exogenous citrate on cell growth, different concentrations of citrate was added to the media. Cells in midlog phase were wash 2× with TAP N+ and TAP N- media respectively. 100 mM citrate stock buffered with Tris base (pH 7.0) was diluted to give final concentrations between 1 and 10mM in 96-well plate cell cultures, with 8 replicates for each concentration. As shown in (Figure 4.1A), the growth of cells supplemented with different concentration of citrate was not different from control during the first 36 h of incubation. As the incubation continued, cells supplemented with citrate showed enhanced growth compared to control and the difference in growth was correlated to citrate concentration, suggesting that the supplemented citrate was transported into the cytosol and utilized as carbon/energy source for growth. Cultures deprived of N were inoculated at a higher cell density to provide sufficient biomass available for subsequent assays.

A representative color image, taken at 48h of incubation, of cultures with different concentrations of citrate in the media was shown in Figure 4.1B. Bleaching of culture was observed when exogenous citrate concentration were above 5 mM. This could be due to osmotic stress induced by citrate and Tris.

4.2.2 Lipid droplets estimation

To estimate the accumulation of lipid droplets, cell cultures were collected after 48 h of growth and stained with the lipophilic dye Nile red for fluorescence intensity measurement. As shown in 4.2, the fold change of fluorescence intensity normalized to 1.0 OD₆₀₀, which approximates the fold change of lipid droplets per cell, showed a sigmoidal response to the concentration of exogenous citrate. Up to 8-fold increase in lipid droplets was induced at

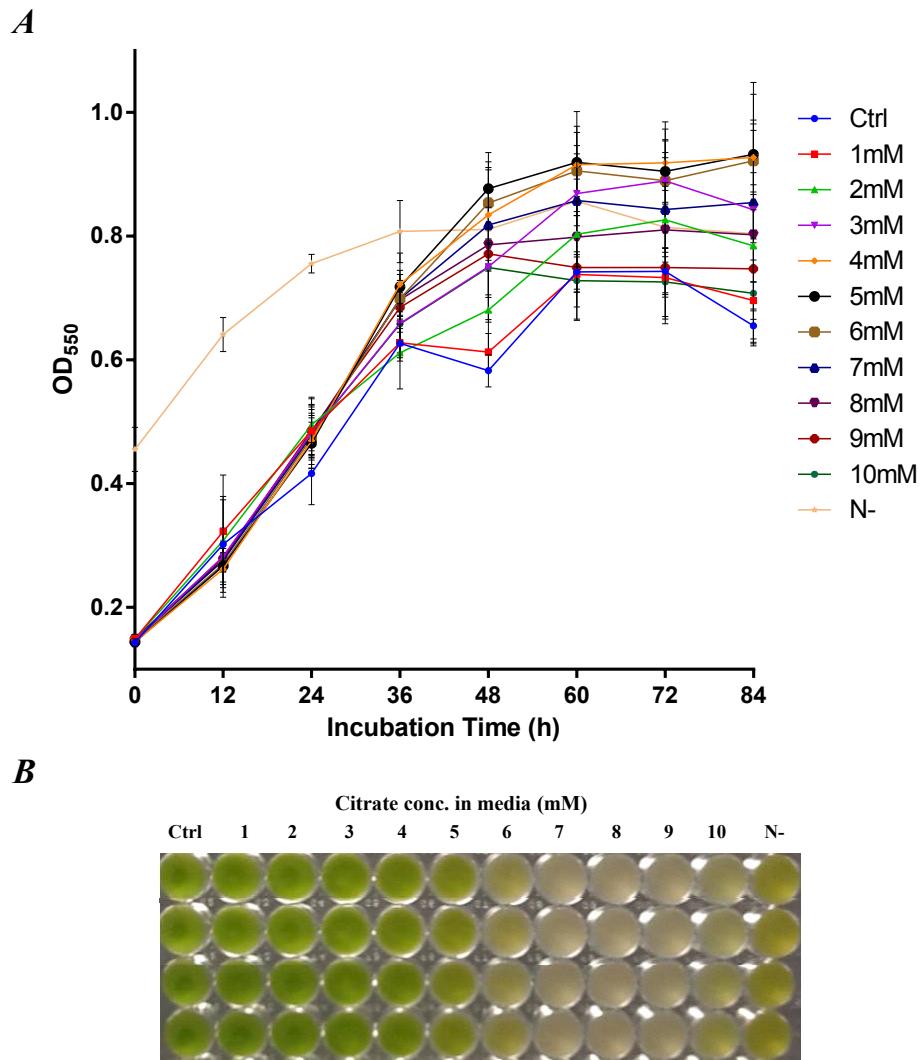


Figure 4.1: Growth during citrate supplementation. **A.**Cell growth at different concentrations of citrate in the media. Cells in midlog phase were wash 2 with TAP N+ and TAP N- media respectively. 100 mM citrate stock buffered with Tris base (pH = 7.0) was diluted to final concentrations of 1-10mM in each column of the 96-well plate and to a final volume of 200 L. The plate was sealed with BreathEasy membrane and placed on a light shelf and shaken vigorously every 12h. **B.** Representative image of cultures with different concentrations of citrate in the media after 48h of incubation.

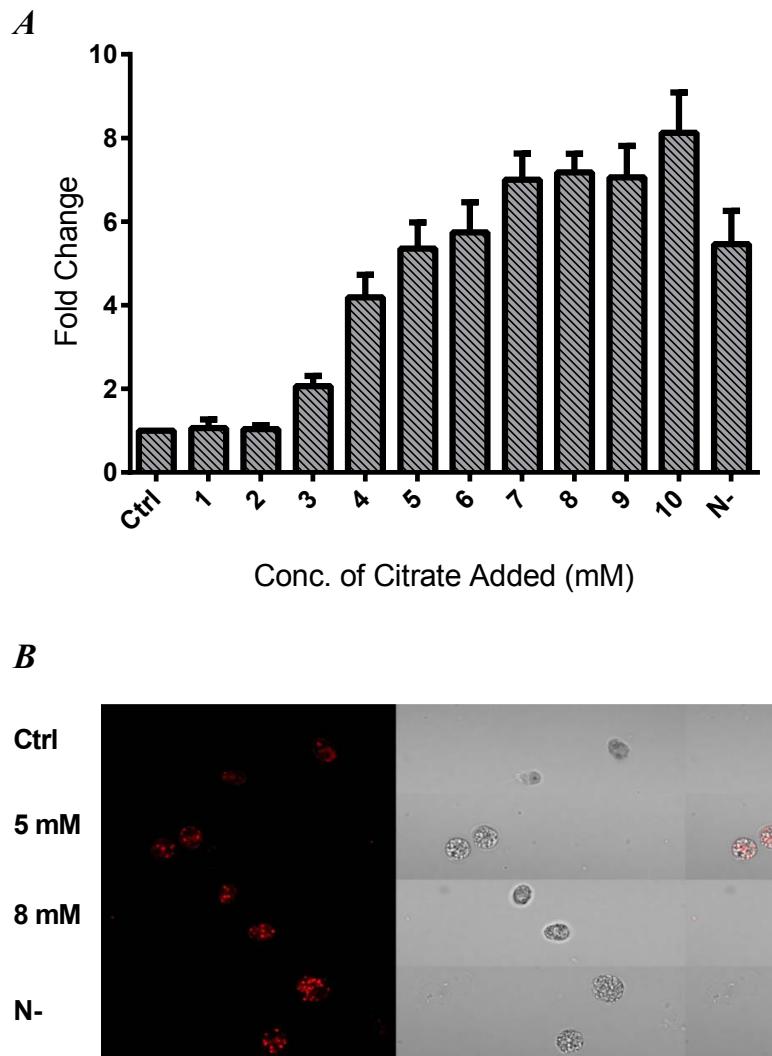


Figure 4.2: Lipid droplets accumulation during citrate supplementation. **A.** Lipid droplet content was estimated by fluorescence intensity after cells treated with different concentrations of citrate were stained with Nile red. The fold change of fluorescence intensity relative to control was normalized with cell density in OD₆₀₀. **B.** Representative confocal microscopic image of cells stained with Nile red after 72h of incubation.

the highest concentration. At 5 mM of citrate, a 5-fold induction was observed and this concentration was used for subsequent assays.

To visually assess the effect of compounds on morphological features, stained cells

were subjected to fluorescent confocal microscopy. As shown in Figure 4.2B, the amount of lipid droplets was positively correlated with the concentration of citrate and the morphology of cells were not severely affected.

4.2.3 FAMES analysis

To confirm total lipid accumulation and examine the composition of fatty acids in total lipid, lipids were extracted and transesterified for analysis using GC/MS. To determine the optimal concentration of citrate supplemented, at which the total lipid content in a unit volume of culture is maximized, same volumes of cell culture at different concentrations with 3 replicates each were directly subjected to FAMES sample preparation. As shown in Figure 4.2A, the total amount of fatty acids increased with the concentration of supplemented citrate up to 4 mM and a sharp decrease was observed beyond 6 mM. This sharp decrease was caused by decreased total biomass due to the stress induced by high concentration of citrate. A relative increase of C18:1 was observed at treatment conditions with 4-7 mM of citrate.

To quantitatively determine the amount of total FA and various FA species in dry biomass, cells were grown in 100 mL of liquid cultures in flasks at 5 mM of supplemented citrate with 3 biological replicates for each condition. The amount of total fatty acids increased significantly by 1.5 folds in cells supplemented with 5 mM of citrate compared to control ($p < 0.05$) and is comparable to N- (Figure 4.3C). The FA composition profile (Figure 4.3B) has shown significant increase in C18:1 in citrate-treated cells by nearly two folds and a small but significant increase in C16:0 and C16:1($p < 0.05$). Moreover, the FA profile of citrate-supplemented cells is significantly different from that of N starved cells as evidenced by significant changes in C16:1, C16:4, C18:0, C18:1 and C18:3 ($p < 0.05$).

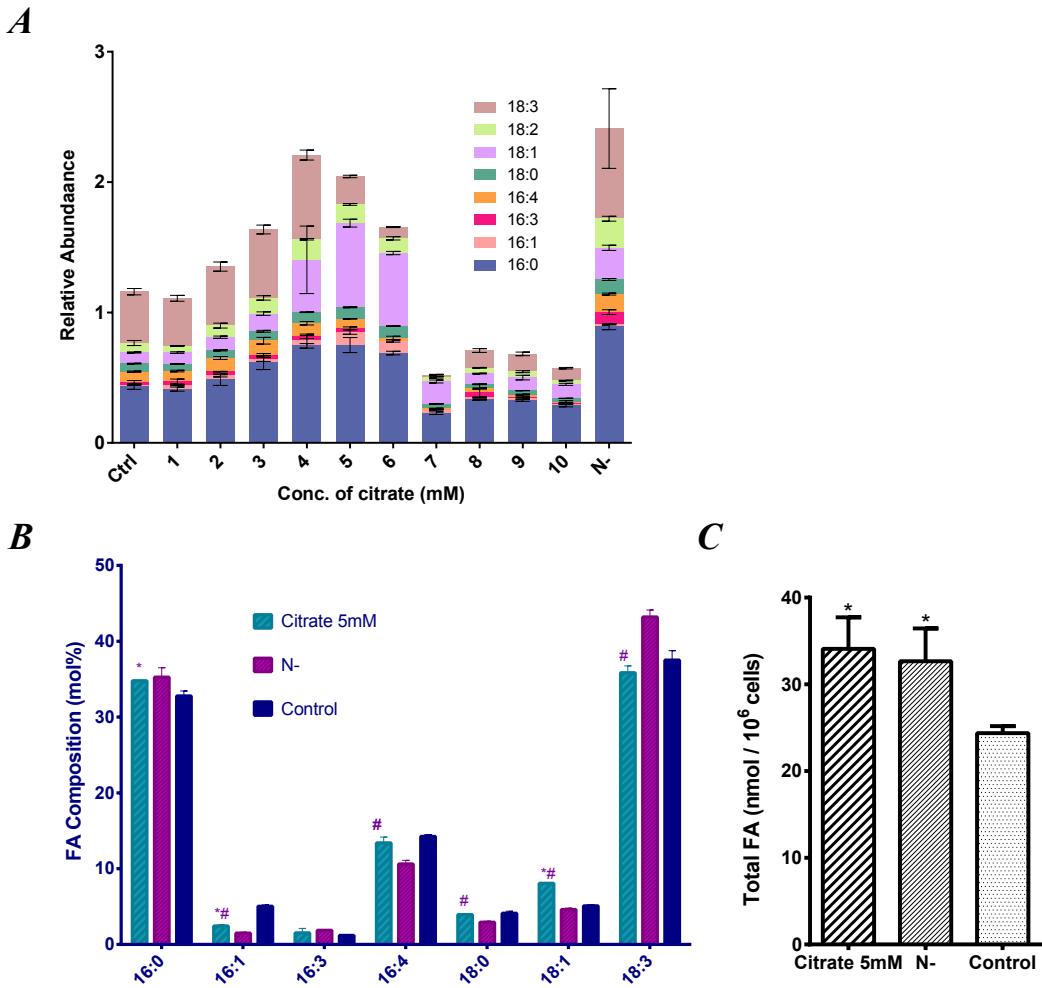


Figure 4.3: FAMEs profiles during citrate supplementation. **A.** Relative abundance of FA species in liquid cultures with different citrate concentration. Three replicates of 200 μ L cultures were centrifuged and directly used for lipid extraction and FAMEs analysis. Peak area of each species normalized to internal standard is used as relative abundance. **B.** FA composition after 72h of incubation. **C.** Total FA after 72h of incubation, normalized to cell count.

These data suggest that the FAs in citrate-supplemented cells undergo metabolic processes distinct from N starvation.

4.3 Uptake of citrate from the media

To measure the uptake of citrate, total metabolites were extracted from the cells at 0, 24, 36 and 48 hours of incubation and derivatized with MSTFA for GC/MS analysis. As expected, the intracellular concentration of citrate increased 1.5-fold during N starvation most likely due to the efflux of citrate from mitochondria, which may then be utilized for FA biosynthesis (Figure 4.4.) After 24 h of incubation with 5 mM of citrate, the intracellular level of citrate increased significantly by 30-fold relative to control ($p < 0.05$). At 36 h of incubation, a dramatic 600-fold increase was observed ($p < 0.01$). At 48 h, the intracellular level of citrate decreased. To confirm the uptake of citrate by the cells from the medium, cell-free medium was collected at 0, 24 and 36 h of incubation by passing through 0.22 μm filters, dried and derivatized with MSTFA for GC/MS analysis. The concentration of citrate decreased over time (Figure 4.4C) and a significant decrease in citrate concentration in the medium by 22% was observed at 36 h of incubation ($p < 0.05$). These data suggest that exogenous citrate was readily utilized by cells in metabolic processes.

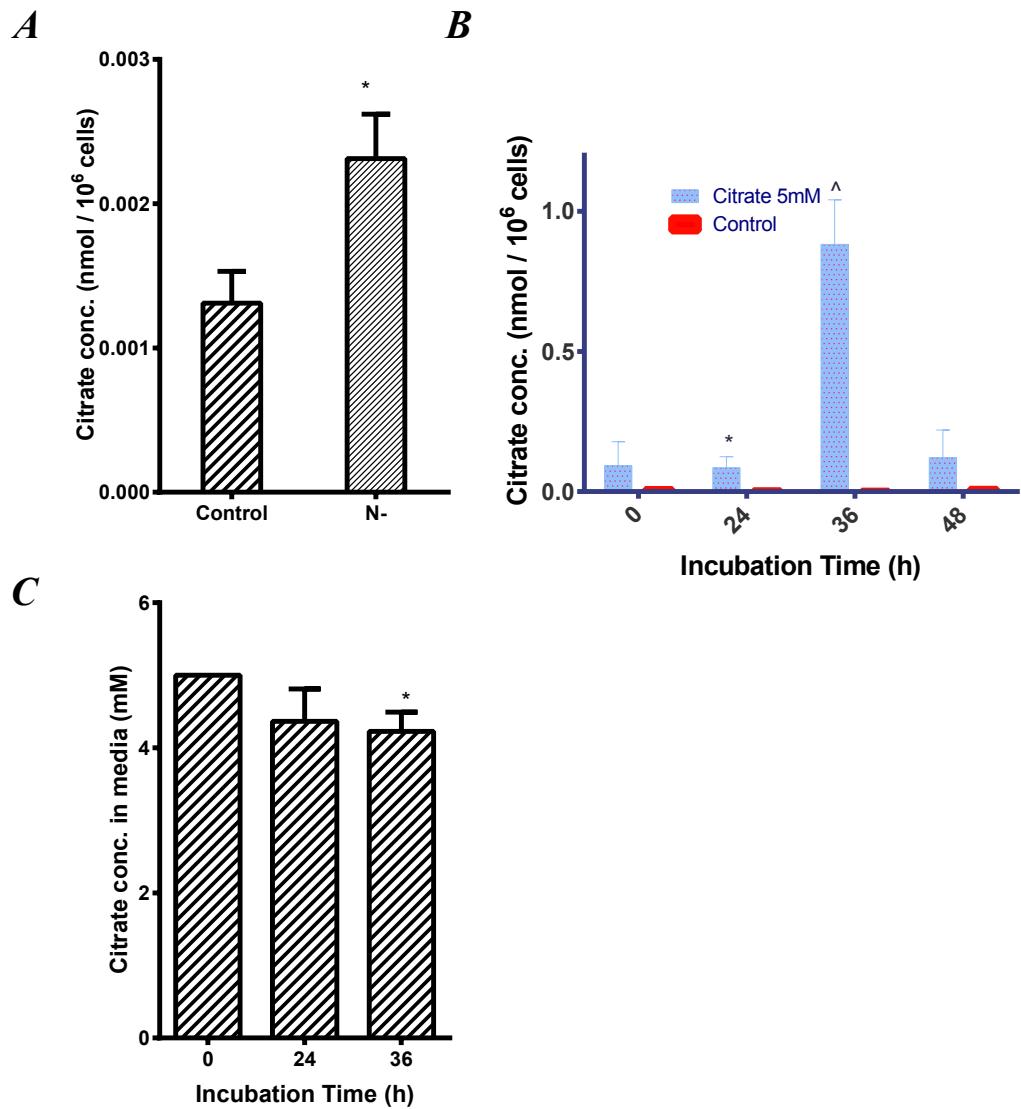


Figure 4.4: Change in intra- and extracellular citrate concentration over time. * indicates $p < 0.05$ and ^ indicates $p < 0.01$. **A.** Effect of N starvation on the concentration of intracellular citrate. **B.** Change in citrate concentration in the cells during the first 48h of incubation. **C.** Change in citrate concentration in the media during the first 36h of incubation.

Chapter 5

Discussion

This study has shown that each of the 4 compounds selected from a high-throughput screening induce significant increase in TAG accumulation in *Chlamydomonas reinhardtii* without severe impact on growth, accompanied by systematic changes in its transcriptional program. The transcriptional response of cells to three of these compounds were examined using RNA-seq. Using carefully selected sequencing protocol and computational tools, we have demonstrated that at the mRNA level, lipid accumulation response of cells to compounds is distinct from that to nitrogen starvation and the response to each compound is different. A unique set of genes are differentially expressed in cells treated with compound 84 but not in cells treated with compound 30 and 42. Treatment with any of the three compounds induces the up-regulation of glycolysis and fermentation in the cells but not acetate assimilation. Although genes encoding pyruvate dehydrogenase complex are down-regulated, citrate efflux is highly favored due to the changes in TCA cycle, which may play an important role in the *de novo* fatty acid synthesis. Differential expression of genes involved in plastid membrane lipid metabolism suggests that unlike nitrogen starvation, SQDG and PG may play more important roles as lipid pools than MGDG and DGDG. At the level of transcription regulation, compound treatments induced changes that are op-

posite to that induced by N starvation, particularly RWP-RK domain transcription factors and SBP domain transcription factors.

5.1 Effects of compounds on lipid accumulation and growth

At the final concentration of 20 μM , all four compounds induced significant increase in lipid accumulation in *Chlamydomonas reinhardtii*. Compound 84 induced a 6-fold increase in lipid droplet accumulation, measured by Nile red staining, the highest among all three compounds. Compound 30, 42 and 67 induced 3- to 5-fold increase in lipid droplet accumulation. The accumulation of TAG was visually confirmed by TLC analysis. FAMEs analysis quantitatively confirmed lipid accumulation and identified palmitic acid and α -linolenic acid as two major species contributing to the overall increase of fatty acids.

With the exception of compound 84, which reduced the growth by 45% at 72h of incubation, compound treatment did not severely affect growth as the growth reduction was only between 11% to 18%, at the concentration of 20 μM for each compound. The EC50 of each compound was subsequently determined to be less than 5 μM (data not shown). Therefore, compound treatment is able to induce significant increase in lipid accumulation without severely limiting growth.

5.2 Advantages and limitations of RNA-seq and computational methods

RNA-seq experiment provides large amount of sequencing data of short reads. With appropriate computational methods, these data can be used to assemble the whole transcriptome,

measure the expression of individual transcripts and detect novel transcripts at a high accuracy. One of the most popular next-generation sequencing technology Illumina HiSeq was employed in this study due to its affordable cost and high reliability [54, 55]. To balance sequencing depth and number of replicates under limited budget, we chose to run 12 samples using 100 bp single-end read mode with approximately 30 million reads per sample. Although paired-end reads are often preferred over single reads due to its lowered ambiguity in sequence mapping, large numbers of single reads are sufficient for differential gene expression analysis [56, 25]. The read length of 100 bp allows more accurate assembly than shorter read lengths and 30 million reads for each sample allow the quantification of transcripts expressed at low levels due to the small genome size of *Chlamydomonas reinhardtii* of 112 Mb. In this study, 19502 out of 19526 transcripts identified by JGI were detected and quantified, indicating the completeness of the transcriptome assembly. In addition, 5335 putative novel transcripts were also assembled. These putative novel transcripts provide potential opportunities for identifying unique patterns of gene expression regulation in response to compound treatment. Differentially expressed novel transcripts could be the key to uncovering part of the mechanism of lipid accumulation induced by compounds. However, since pair-end reads were not used in this experiment, analysis of alternative splicing, which requires mapping to splicing junctions at a high certainty, was not performed.

5.3 Overview of the transcriptional responses leading to TAG accumulation

Previous studies have shown that under environmental stresses, such as nutrient starvation and high salinity, metabolic fluxes in microalgae are channeled to TAG accumulation [11, 12, 14, 6]. This systematic shift in metabolic fluxes can be attributed to dramatic

transcriptional changes in response to environmental factors, as more than 4000 genes significantly up- or down-regulated more than 2-fold with FDR < 0.05 during nitrogen starvation [57]. Several studies [57, 35, 58] have suggested that during N starvation, acetate in TAP media is no longer mainly utilized in anabolic pathways such as glyoxylate cycle and gluconeogenesis, but is funneled into fatty acid biosynthesis, with evidence being the significant up-regulation of ACS3, encoding acetyl-CoA synthetase and down-regulation of genes involved in glyoxylate cycle and gluconeogenesis. A chloroplast pathway for the *de novo* fatty acid biosynthesis was identified as a major process contributing to the lipid accumulation during nitrogen starvation [58]. In addition to *de novo* fatty acid biosynthesis, the turnover of the pool of chloroplast membrane lipids, especially galactolipids, also plays an important role in the overall accumulation of TAG [58, 59, 14]. Under nitrogen stress, digalactosyldiacylglycerol (DGDG) was found to increase concurrently with TAG while the most abundant plastidic lipid species monogalactosyldiacylglycerol(MGDG) slightly decreased, suggesting DGDG was rapidly synthesized under nitrogen stress and utilized for TAG synthesis through a pool of MGDG [58]. It has also been reported that PDG1, encoding a galactolipid lipase required for TAG synthesis [59], was significantly up-regulated during nitrogen starvation [35].

Although the accumulation of significant amount of TAG was induced by compounds, the transcriptional response of the cells suggests that the metabolic fluxes are channeled to TAG accumulation in a mechanism distinct from nitrogen starvation. The glycolytic process is up-regulated in compound treated cells and the carbon flow is likely to be directed to mainly fermentation rather than TCA cycle, as evidenced by significant up-regulation of genes encoding phosphofructokinase, pyruvate decarboxylase, alcohol dehydrogenase and significant down-regulation of genes encoding pyruvate dehydrogenase complex. In contrast to nitrogen starvation, compound treatment did not induce increase in the expression ACS3, encoding acetyl-CoA synthase, but the expression of ACS3 was significantly

decreased in cells treated with compound 30, 42 and 84, 3.3-fold, 1.6-fold and 33-fold, respectively. Meanwhile, genes encoding enzymes involved in the degradation and utilization of starch were significantly up-regulated during compound treatment. These data suggest that the carbon from starch but not acetate is utilized in the glycolytic process, which, followed by fermentation, serves as an important source of cellular energy during compound treatment. Although the vast majority of genes encoding enzymes involved in carbon fixation were down-regulated, many of these enzymes, including the rate-limiting enzyme RuBisCo were not decreased at the protein level (Wase, Tu et al., in submission). Therefore, starch synthesized in the chloroplast via the carbon fixation may play an important role in the catabolic processes that release energy for cellular functions during compound treatment.

Similar to the TAG accumulation mechanism during nitrogen starvation, *de novo* fatty acid biosynthesis played an important role in the TAG accumulation during compound treatment. Although genes directly involved in fatty acid biosynthesis, such as those encoding acetyl-CoA carboxylase subunits and 3-ketoacyl-ACP synthase, were significantly down-regulated in compound treated cells, with compound 84 having the most pronounced effects, genes involved in citrate efflux, which readily provides cytosolic acetyl-CoA, were significantly up-regulated. The role of citrate efflux in TAG accumulation was further established by supplementation of citrate in the media, which induced up to 5-fold increase in lipid droplet accumulation. With the exception of genes encoding pyruvate dehydrogenase complex, genes involved in TCA cycle were not significantly down-regulated. These data suggest that TCA cycle may serve as a shuttle for the channeling of carbon from other cellular components to fatty acid biosynthesis.

Recycling of membrane lipids also played an important role in TAG accumulation response of cells treated with compounds. In contrast to nitrogen starvation, compound treatment does not induce significant up-regulation of genes involved in galactolipid biosynthe-

sis and degradation, such as DGD1, encoding a DGDG synthase and PGD1, encoding a galactolipid lipase. Interestingly, genes involved in the synthesis of sulfoquinovosyldiacylglycerol (SQDG) and phosphatidylglycerophosphate (PG) were significantly up-regulated. For example, the expression of SQD3, encoding sulfolipid synthase was significantly increased in cells treated with compound 30, 42 and 84, 11.5-fold, 19.6-fold and 42.3-fold, respectively. These data suggest the turnover of SQDG and PG may play a more important role in the TAG accumulation than DGDG and MGDG during compound treatment. In addition, compound treatment led to stronger transcriptional inhibition of TAG lipases. For example, The expression of FAP12 was significantly decreased in cells treated with compound 30, 42 and 84, 4-fold, 11-fold and 307-fold, respectively and the expression of LIP3 was significantly decreased in cells treated with compound 30, 42 and 84, 16-fold, 15-fold and 20-fold, respectively.

5.4 Cell signaling and transcription regulation

A number of genes encoding transcription factors have been identified in *Chlamydomonas reinhardtii* as differentially expressed during nitrogen starvation and are associated with the regulation of genes responsible for TAG accumulation [60]. The identification of differentially expressed transcription factor allows us to further establish the mechanism of the unique transcriptional response to compound treatment leading to TAG accumulation. Genes encoding transcription factors (TF) which are up-regulated in response to nitrogen starvation are significantly down-regulated in compound treated cells, including RWP-RK domain TFs, which are key regulators of nitrogen responses and gametogenesis and SBP domain TFs, whose regulatory targets include genes involved in fatty acid and lipid biosynthesis. The down-regulation of RWP-RK domain TFs suggests that unlike nitrogen starvation, which strongly activates gametogenesis, compound treatment leads to the suppression

sion of gametogenesis, which is potentially beneficial to vegetative growth. In addition, a subset of RWP-RK domain TFs were differentially expressed between treatment with compound 30 or 42 and treatment with compound 84, suggesting subtly different mechanism in RWP-RK domain TF-mediated cell signaling. Although nitrogen starvation and compound treatment share some patterns of changes in lipid biosynthetic processes, the regulatory mechanism of these transcriptional changes may be different. During nitrogen starvation, increased expression of NRR1, a SBP domain TF, leads to the up-regulation DGATs and PDAT, which are also significantly up-regulated during compound treatment, which does not increase the expression of NRR1.

Interestingly, genes encoding transcription factors regulating cell cycle progression, such as E2F1, E2F2 and DP1, were significantly up-regulated in cells treated with compounds. This may partially explain the moderate growth rate of cells despite of the cellular stress induced by the compounds.

The differential expression of specific transcript factors also reveals the possible roles of NO_3^- and sulfur in the cell signalling processes that regulate the TAG accumulation in response to compound treatment. The decreased expression of NIT2, encoding a nitrogen response TF, led to the down-regulation of NIA1, which encodes a nitrate reductase that converts NO_3^- to NO_2^- . This decrease in transcriptional activation of NIA1 may lead to increase in the concentration of NO_3^- , which has been identified as a signalling molecule involved in the regulation of nitrogen assimilation and carbon metabolism [45]. The increase in the expression of sulfur response genes, SAC3, encoding sulfur acclimation protein and ASR1/3 encoding arylsulfatases suggest that sulfur is partially depleted during compound treatment. Such depletion of S, possibly caused by strong transcriptional activation of SQDG synthesis, may induce TAG accumulation response similar to that induced by sulfur starvation.

5.5 Future directions

This study has demonstrated that in response to a set of small molecules, *Chlamydomonas reinhardtii* cells adjust to a new transcriptional program that lead to TAG accumulation while maintaining moderate growth. However, the following issue still need to be addressed: 1) the levels of protein, metabolite and mRNA do not always agree, and thus in order to confirm the shift in biological processes, a targeted proteomic and a targeted metabolic study are warranted; 2) the lipid trafficking and biosynthetic process is highly complex, and thus in order to fully understand the unique TAG accumulation response during compound treatment, a lipidomic study is warranted; 3) the cell signalling process in *Chlamydomonas reinhardtii* during TAG accumulation is poorly understood, and thus a study of the components of signalling cascade, such as protein kinases, phosphatases and small effectors, will reveal important information of the TAG accumulation mechanisms; and 4) the direct molecular target(s) of these small molecules is currently unknown, and thus a protein-ligand interaction study combined with *in silico* approaches such as molecular dynamic simulation will be useful for the more complete understanding of the biological processes involved in TAG accumulation in response to these small molecules.

Chapter 6

Methods and Material

6.1 Culture conditions and growth assessment

Cells of *Chlamydomonas reinhardtii* CC-125 strain were routinely maintained on plates of TAP medium with 5% agar at 25 °C with a photon flux of 54 $\mu\text{mol m}^{-2} \text{s}^{-1}$ [61]. To start a liquid culture, a loopful of cells was inoculated into 100 mL of TAP medium in a 250 mL Erlenmeyer flask placed in an ventilated incubator shaking at 120 RPM and at the temperature of 25 °C. For each treatment experiment, after cells reaching their mid-log phase, the liquid culture was centrifuged at $2000 \times g$, media removed, cells rinsed in fresh TAP medium for 2×, and then resuspended in the same medium at a final cell density of 5×10^5 cells mL^{-1} . For the compound treatment in the qPCR experiment presented in Chapter 2, the 10 mM stock solution in DMSO of each compound was added to each flask to give a final concentration of 20 μM . For the compound treatment in the RNA-seq experiment presented in Chapter 3, each compound dissolved in DMSO was added to each flask to reach a final concentration of 5 μM . For the citrate supplementation experiment, different volumes of 1 M citric acid solution, with pH adjusted to 7.0 using Tris base, was added to the liquid culture to reach final concentrations of 1-10 mM. To assess the growth

of the cells, 200 μL of culture was collected, transferred to a 96-well microtiter plate and optical density at 600 nm measured using a Synergy plate reader (BioTek, Winooski, VT).

6.2 Lipid droplets estimation using Nile red

To estimate the accumulation of lipid droplets in the cells, 200 μL of culture was collected and transferred to a black 96-well microtiter plate, immediately followed by a measurement of OD_{600} as described in the previous section. 5 μL of 300 mM Nile red (dissolved in DMSO) was added and the mixture was incubated at 37°C for 30 min. The fluorescence intensity in arbitrary unit was measured using the said plate reader at 485/590 nm (excitation/emission) and normalized by cell density of OD_{600} to give fold change relative to control (cells grown in regular TAP medium).

6.3 FAMEs and metabolite analysis using GC/MS

6.3.1 Lipid extraction and transesterification

To measure the level of FAMEs, approximately 1×10^7 cells were harvested by centrifugation and the pellets flash frozen and stored at -80 °C until analysis. Before lipid extraction, the sample was dried by lyophilization overnight. Total lipids were extracted using the MTBE method as described by Kurzchalia et al. 2008. Briefly, 400 μL ice cold 75% methanol was added to the dried biomass along with 100 g of nonadecanoic acid as internal standard. The samples were then sonicated on ice for 2 min to lyse the cells. Then In 1 mL of MTBE was added and the samples were shaken for 1 h at room temperature and 250 L of water was added for phase separation. The top organic phase was collected and dried under nitrogen stream. Samples were transesterified in 50 μL of MTBE with 10 μL of TMSH for 30 min at room temperature.

6.3.2 Metabolite extraction and derivatization

Approximately 1 mL of culture was centrifuged at selected time points to yield 5×10^6 cells. The supernatant was carefully removed with a needle and then filtered through a 0.22 μm syringe driven filter. The cell pellets were washed 2 \times with water, flash frozen with liquid nitrogen and stored at -80 °C. The cell pellet was lysed in 600 μL of MCW (MeOH:CHCl₃:H₂O at 10:3:1) stored at -80 °C with glass beads on a TissueLyser LT(Qiagen, Hilden, Germany) for 2 min and 1 μg of sorbitol was added as internal standard. The tubes were then vigorously shaken in ice for more than 30 min and centrifuged at 4 °C. 500 μL of supernatant was removed and dried in a SpeedVac(Thermo Scientific, Waltham, MA). 100 μL of the filtered supernatant (media) was also dried for the determination of citrate level in the media. 5 μL of 40 mg/mL methoxyamine hydrochloride was added and incubated at 30 °C, 400 rpm for 90 min. 45 μL of MSTFA was added and the sample incubated at 37 °C, 400 rpm for 30min prior to analysis by GC/MS.

6.3.3 GC/MS parameters

One microliter of the sample was injected in splitless mode onto a DB-5 ms column of an Agilent 6890 GC/MS system (Agilent Technologies, Santa Clara, CA). The injection port was set at 250 °C and the auxillary temperature was set at 280 °C. For detecting FAMEs, temperature ramp was started at 70 °C and held for 2 min, then increased to 140 °C at 20 °C min⁻¹ and held for 2 min. The second ramp was from 140 °C to 280 °C at 10 °C min⁻¹ and then held for 3 min. A constant flow of helium (99.99%) was maintained at a flow rate of 1.5 mL min⁻¹. Mass spectra were acquired using an Agilent 5973 mass selective detector with a scan range from 50 to 550 amu at 70 eV. Compounds were identified using the NIST (2008) library applied to relevant spectra. For detecting citrate, the temperature was held at 70 °C for 2 min and then increased to 130 °C at the rate of 10 °C min⁻¹,

followed by an increase to 210 °C at 5 °C min⁻¹. Selected ion monitoring (SIM) mode was enabled to detect ions with mass-to-charge ratios of 273.1 and 147.1. The abundance of these ion were used to determine the level of citrate in the sample.

6.3.4 Data analysis

The chromatogram of each sample was integrated to determine the area of each peak, identified to be of a single species according to it mass spectrum. The peak area of each species was then normalized with the peak area of the internal standard. The amount of each species was then determined by its normalized peak area and the known amount of internal standard.

6.4 Lipid analysis using TLC

6.4.1 Lipid extraction

Approximately 5×10^7 cells were harvested by centrifugation and pellet flash frozen and stored at -80 °C until analysis. Prior to lipid extraction, the sample was dried by lyophilization. 5 mg of dry biomass was resuspended in a mixture of 1 mL of 50 mM K₂HPO₄, 1 mL of CHCl₃ and 2 mL of MeOH, which was then vortexed for 30 min with glass beads. An additional 2 mL of CHCl₃ was added to the mixture before centrifugation at 3800 × g for 10 min. The lower organic layer was transferred to new tube, evaporated under nitrogen stream and resuspended in 25 µl of CHCl₃.

6.4.2 Chromatography

The running solvent was a mixture of 220 mL of hexane, 80 mL of diethyl ether and 1 mL of glacial acetic acid. Before the plate was spotted, silica TLC plate was activated at

105 °C for 1 h. The solvent was poured into the tank with paper for saturation. 10 µl of samples and standards were spotted in each lane. The plate was developed in the tank until the solvent front reached half of the plate and then dried in the hood. Dried plate was submerged in the charring solution for a few seconds and then incubate at 100 °C for 30 min. The charring solution was a mixture of 0.63 g of MnCl₂·H₂O, 60 mL of water, 60 mL of methanol and 4 mL of concentrated H₂SO₄.

6.4.3 Data analysis

Using ImageJ, the intensity of each band was estimated by integration of pixels on the image. A fold change was calculated for each sample relative to the intensity of the control.

6.5 Gene expression analysis using qPCR

6.5.1 Isolation and quality assessment of RNA

To assess impact on gene expression by selected compounds, approximately 1×10^7 cells treated with compounds for 72 h in the culture condition described in 6.1 were harvested by centrifugation. A total volume of 250 µl of cell pellet along with small amount of media in a 2 mL micro-centrifuge tube was flash-frozen in liquid nitrogen and stored at -80 °C until analysis. To the sample placed on ice, 750 µl of TRIzol LS (Invitrogen, Waltham, MA) and approximately 600 µl of acid-washed glass beads were added and the sample tube was immediately transferred to a TissueLyser for homogenization at 50 Hz for 5 min. The homogenate was transferred to a new tube and centrifuged at $10000 \times g$ for 15 min. The supernatant without the top layer of lipids was collected and pellet discarded. 200 µl of CHCl₃ was mixed with the supernatant after being shaken vigorously for 15s. The mixture was then left at room temperature for 3 min, followed by centrifugation at $10000 \times g$ for

15 min. The upper aqueous phase was transferred to a new tube with 1 equal volume of 70% EtOH and mixed by pipetting. The mixture with any precipitate was transferred to an RNeasy Mini spin column (Qiagen, Hilden, Germany) and centrifuged at $8000 \times g$ for 15 s and washed with 500 μ l RPE buffer (provided in the kit). To the center of the column, 50 μ l of nuclease-free water was added and RNA was eluted by centrifugation at $8000 \times g$ for 1 min. To eliminate genomic DNA contamination, DNase digestion was performed according to manufacturer's instruction (5 PRIME, Gaithersburg, MD). The digested sample was then mixed with 350 μ l of provided RLT buffer and 250 μ l of 100% EtOH, transferred back to the column, washed with buffer RPE 2 \times according to manufacturer's instruction. The purified RNA was eluted with 50 μ l of nuclease-free water with 1U/ μ L of RiboLock RNase inhibitor (Thermo Scientific) and stored at -80 °C.

The absorbance at 260 nm, 280 nm and 230 nm of the above purified sample was measured on a NanoDrop 1000 spectrophotometer (Thermo Scientific) and the 260/280 and 260/230 ratios were checked to assure the purity of sample.

For qPCR assays, 500 μ g of purified RNA from each sample was used for first-strand cDNA synthesis using iScript cDNA Synthesis Kit (Bio-Rad, Hercules, CA) as per the manufacturers protocol.

6.5.2 qPCR amplification

With the above synthesized cDNA, qPCR was performed on a Mastercycler ep realplex 2S (Eppendorf, Hamburg, Germany) using ABsolute qPCR SYBR Green Mix (Thermo Scientific). The PCR program consisted of an initial denaturation step at 95 °C for 15 min, followed by 40 cycles of denaturation at 95 °C for 15 s, annealing at 59 °C for 30s

and extension at 72 °C for 15s. Melting curve steps were added after amplification with a resolution of 0.5 °C. RACK1 (receptor of activated protein kinase) was used as reference gene for normalization. The primers used for this study were given in Table 1. No-template controls were included for each primer pair and melting curve analysis was performed to check the specificity of amplification of gene of interest.

6.5.3 Data analysis

The amplification efficiency of each gene and quantification cycle (C_q) of each sample were determined from the raw fluorescence data using the LinRegPCR algorithm [53]. Fold change of each sample relative to control normalized with the expression level of reference gene was calculated according to a widely used relative quantification algorithm [62] briefly described as follows. $RQ_{X,G} = E^{C_{q,c,A} + C_{q,c,B} + C_{q,c,C} - C_{q,t,X}}$, where $RQ_{X,G}$, relative quantity of amplicon G is the fold change of each biological replicate X relative to control. E is the amplification efficiency of the specific gene of interest. $C_{q,t,X}$ is the quantification cycle of the specific gene of each biological replicate of treated sample and $C_{q,c,i}$ is the quantification cycle of one of the three control samples. $NRQ_{X,G} = \frac{RQ_{X,G}}{RQ_{X,ref}}$, where $NRQ_{X,G}$, normalized relative quantity of amplicon G, is the fold change of each biological replicate X relative to control normalized with the selected reference gene.

6.6 Transcriptome analysis using RNA-seq

6.6.1 Mapping reads to a reference genome

The raw reads in FASTQ format were used as input for alignment to the genome (Joint Genome Institute C. reinhardtii v5.5) using Tophat2/Bowtie2 [24], in a reference annotation-guided mode. The raw reads from all samples were aligned in a single run, generating a

Table 6.1: List of primers used in qPCR experiments

Gene	Protein	Accession	Forward Primer	Reverse Primer
ACLA1	ATP citrate lyase, subunit A	XM_001700848	TGAGTTCCCACTGCCCTTC	TGGGGTTGAGAACTGTGAA
ACX1	acetyl-CoA carboxylase	XM_001696893	CACCACGACGCTTGAGTTG	CGGCCTGTAAGTCTCCTTCC
APG8	Autophagy-related protein 8	XM_001699138	TCTCCACAATGGTGGCTCC	CCTCCGCTTCCGCTATCT
BIP1	binding protein 1	XM_001701633	TTGGGGCTGAGGAGGAAG	CGCAATAATCTCCACACGGC
CIS1	citrate synthase	XM_001702931	TGTCGGTCCCTGCTTGGTT	CGAATTACCCCCATCCTCGG
CIS2	citrate synthase	XM_001695519	GTTGTTGGCAATGCGATGTT	GGCATTCATGATGTCACCGC
DGTT1	diacylglycerol acyltransferase	JN815266	CAGAGCCGTGATGTTGCA	TGCACGTAAGCAATGTGCC
ICL1	Isocitrate Lyase	XM_001695279	GTTGTACGGGTGCAGTTGTG	CTGTTCAAGCCAGAACGAGA
LHCA	light-harvesting protein of photosystem I	XM_001695283	ATGGCCCGTTATGTGGAGG	TGTTACACGGATGGCACGA
MDH1	malate dehydrogenase	XM_001693066	GGGCCTCGACTTTGTCAAGA	CCCTGCCTGCATCACAAAT
MLDP	major lipid droplet protein	XM_001697616	TGGAAAGCCTCTGAAGCACC	GTCTTGGCCTCTGGTAGCC
PSAN	photosystem I reaction center subunit N	XM_001701648	GCGCTCCAACAAATGCTG	CCACTCCAACAGCTAGACCG
RBCS	ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit 2	XM_001702356	TGTAAATGGAGGGGCTCGTT	AAGCTACCGCTCAGCACTT
STA6	ADP-glucose pyrophosphorylase small subunit	XM_001691802	GGGCTCCAAGATCCACAACT	CACTCCAGGGTCTCGTA

single file (bam) of mapped reads.

6.6.2 Transcriptome assembly and quantification

The mapped reads were then used for transcriptome assembly using Cufflinks [25] in the reference guide mode with maximum bundle fragments adjusted to 2500000 to accommodate highly abundant transcripts. The transcriptome in FASTA format was generated by extracting sequences from the genome according to the loci specified in the GTF output of Cufflinks. Sequences without reference annotation were annotated with interProScan and Blast. Using Bowtie2, the raw reads from each sample were aligned to the above transcriptome to generate separate mapped reads files. eXpress was used to quantify the abundance of each transcript with raw reads and mapped reads as input for each sample. The FPKM output of eXpress was normalized with library sizes.

6.6.3 Differential expression analysis

As recommended by the author of eXpress, the transcript count values in its output were used as input for differential expression analysis using edgeR. The differential expression between each pair of treatment/control groups were analyzed using an adapted Fischer's exact test. The output of edgeR included fold change, FDR and CPM, which, along with FPKM generated by eXpress, were combined into a data matrix including all samples for further analysis. An in-house program (Appendix A) was created to facilitate the fast extraction and analysis of subsets by ID, gene name, annotation and the above numerical variables.

6.6.4 Gene set enrichment and pathway analysis

Gene matrix files (gmt) of GO categories and KEGG pathways were generated using the GO id and KO id (mapped to pathway) assigned to each transcript in the annotation information, respectively. The CPM of each transcript was used as input for GSEA, with ranking metric of log2 ratio of classes, number of permutation of gene set of 10000, gene set maximum of 500, minimum of 5 and other default parameters. The KEGG pathway enrichment results were plotted on KEGG pathway maps using R package pathview.

Appendix A

Scripts used in RNA-seq data analysis

Files of source code and executable are available at https://github.com/btu6626/RNA-seq_Processing

Appendix B

Parallelization on distributed memory using MPI of transcriptome assembly tool Cufflinks

Cufflinks is one of the most popular tools for transcriptome assembly, a memory and CPU time intensive problem which can generate large amount of useful data in biological studies. The current implementation only allows parallelization on shared memory. In this project, the Cufflinks algorithm was parallelized on distributed memory using MPI, which resulted in speed-up by up to 6 folds. Although the scaling of this implementation still needs improvement, it allows user to run jobs in a more flexible environment, not necessarily cores on a single node, potentially reducing queue time. This implementation will be further optimized by using different strategies in the processing of data structures.

B.1 Introduction

RNA-seq has become an important technique in the studies of differential gene expression, splicing and other RNA-mediated regulatory mechanisms. The lowering prices of next-generation sequencing technologies allow researcher to large amount of shotgun sequence data, which consist of short reads that need to be aligned and assembled to produce biologically meaningful results. Therefore, processing of RNA-seq data is computationally intensive, requiring large amount of memory and CPU time. Cufflinks is one of the most popular tools for transcriptome assembly from short sequencing reads aligned to the genome. Using the Boost library with wrappers based pthreads, the Cufflinks algorithm was parallelized on shared memory, leading to significant speed-up at strong scaling efficiency greater than 0.5. However, shared memory only allows user to run the job on a single computing node, which has limited CPU and memory resources, potentially leading to longer queue time. Parallelization with MPI can distribute the work to multiple nodes and lead to significant speed-up and less memory usage on single node. The embarrassingly parallel nature of a large part of this algorithm allows the parallelization with reasonable effort.

B.2 Existing pthread Implementation

The author of Cufflinks has parallelized this algorithm with Boost threads, a wrapper of pthreads. Each "bundle" as shown in Figure 1, is passed to a single pthread launched in a thread pool for transcript assembly from mapped reads and abundance estimation. This implementation utilizes mutex lock mechanism on the output data file to prevent concurrent writing on a single file. Since bundles often have different numbers of hits (mapped reads), load balancing issue might arise and affect the scaling efficiency.

```

while(true)
{
    HitBundle* bundle_ptr = new HitBundle();
    \\\Find the next bundle
    if (!bundle_factory.next_bundle(*bundle_ptr , \
true))
    {
        delete bundle_ptr;
        break;
    }

    HitBundle& bundle = *bundle_ptr;

    ....

#if ENABLE_THREADS

while(1)
{
    thread_pool_lock.lock();
    if (curr_threads < num_threads)
    {
        thread_pool_lock.unlock();
        break;
    }

    thread_pool_lock.unlock();
}

```

```

sleep(boost::posix_time::milliseconds \
(5));

}

#endif

.....
\\launch a boost thread in the pool to assemble current bundle
#if ENABLE_THREADS

    thread_pool_lock.lock();
    curr_threads++;
    thread_pool_lock.unlock();

    thread asmb1(assemble_bundle,
                 boost:: cref(rt),
                 bundle_ptr,
                 bundle_factory.read_group_properties(),
                 bl_ptr,
                 ftranscripts,
                 fgene_abundances,
                 ftrans_abundances,
                 fskipped);

    ....
}

\\Prevent concurrent writing to a single file

```

```
#if ENABLE_THREADS

    out_file_lock.lock();

#endif

    for (size_t i = 0; i < genes.size(); ++i)
    {
        ...

        for (size_t j = 0; j < isoforms.size(); ++j)
        {
            ...

            for (size_t g = 0; g < isoform_exon_recs.size(); ++g)
            {
                fprintf(ftranscripts, "%s", isoform_exon_recs[g]\
                    .c_str());
            }

            fflush(ftranscripts);

            ...
        }
    }

    fprintf(fgene_abundances, ...);
    fflush(fgene_abundances);

    ...
}

#endif
```

```

    out_file_lock.unlock();

#endif

```

B.3 MPI Implementation

In this current MPI Implementation, in order to find its own starting bundle, each process must open the same data file and sequentially pass through all the reads before its starting position. Bundles contiguously are grouped into larger groups and each single group with equal number of bundles are assigned sequentially to MPI processes according to their ranks. As a result, the process with the highest rank will always waste the largest amount of time looking for the starting bundle. The final result is combined when all processes finish execution.

```

\\Processes other than rank 0 needs to pass through the data
\\file to find the starting bundle

if (mpi_rank !=0){

    HitBundle* bundle_ptr;

    bundle_ptr = new HitBundle();

    while (counter < mpiBundleStart){

        bundle_ptr ->~HitBundle();

        new(bundle_ptr) HitBundle();

        get_next_gene_id();

        bundle_factory.next_bundle(*bundle_ptr);

        counter++;

    }
}
```

```

delete bundle_ptr;
}

\\process ends after (mpiBundleEnd-mpiBundleStart)
\\ bundles are assembled

while( counter<mpiBundleEnd)
{
    counter++;
    HitBundle* bundle_ptr = new HitBundle();
    if (!bundle_factory.next_bundle(*bundle_ptr))
    {
        delete bundle_ptr;
        break;
    }
    ....
#if ENABLE_THREADS
    thread_pool_lock.lock();
    curr_threads++;
    thread_pool_lock.unlock();
    thread_asmbl(assemble_bundle,
        ....);
}

\\Bundles are evenly distributed to processes
mpiBundleStart=bundle_factory->num_bundles()\
/n_proc*mpi_rank;
mpiBundleEnd=bundle_factory->num_bundles()\

```

```

    / n_proc *( mpi_rank +1);

\\ensure the last process finishes the residual bundles
    if ( mpi_rank==n_proc -1) mpiBundleEnd=100000000;
\\assemble_hits calls assemble_bundle
    assemble_hits (*bundle_factory , bl_ptr , \
    mpiBundleStart , mpiBundleEnd );

int mpiAllProcDone=0;
int mpiProcDone=0;
\\enable multi thread support
    MPI_Init_thread(&argc ,&argv , MPI_THREAD_MULTIPLE \
    ,&provided );
    MPI_Comm_size( MPLCOMM_WORLD, &n_proc );
    MPI_Comm_rank( MPLCOMM_WORLD, &mpi_rank );
\\driver calls assemble_hits which calls assemble_bundle
    driver(sam_hits_file_name , ref_gtf , mask_gtf );
    mpiProcDone=1;
\\Check whether all processes done
    MPI_Reduce(&mpiProcDone ,&mpiAllProcDone ,1 ,MPI_INT \
    ,MPI_LAND ,0 ,
    MPLCOMM_WORLD);

    if ( mpi_rank==0){

\\Wait until all processes done
    while (! mpiAllProcDone){

    }
}

```

```
....  
    ofstream of_transcripts( string( output_dir +"\\"  
        "transcripts.gtf"  
    ).c_str(), ios_base::app | ios_base::out);  
  
....  
\\Read files written by each process and concatenate  
for (int i=0; i<n_proc; i++){  
    ifstream if_transcripts( (string( output_dir  
        +"/"+boost::lexical_cast<std::string>(i)+\  
        "transcripts.gtf"  
    )).c_str(), ios_base::in);  
  
....  
    of_transcripts << if_transcripts.rdbuf();  
  
....  
    remove((string( output_dir +"/"+  
        boost::lexical_cast<std::string>(i)  
        +"transcripts.gtf")).c_str());  
  
....  
}  
....  
}  
MPI_Finalize();
```

B.4 Performance Evaluation

Methods

Strong scaling efficiency, speed-up and weak scaling efficiency were determined to evaluate the performance of this MPI parallelization. The data used for testing is from a real RNA-seq experiment of differential gene expression in green alga *Chlamydomonas reinhardtii*. The raw data from the sequencing instrument was aligned to genome with Tophat to produce 30 million mapped reads, used as input for Cufflinks. Fractions of this data file was randomly subsampled using Samtools for different testing purposes. Correctness were check by comparing the data in the output files.

Batch jobs were submitted to SLURM and each pthread was executed on a single core.

Results

As shown in Figure 2, the strong scaling efficiency of MPI parallelized code is satisfactory up to 16 cores with different MPI/pthread hybridization. The maximum speed-up is about 6 fold. Increased data size resulted in increased speed due to the nature of the algorithm as the number of bundles is not necessarily proportional to number of mapped reads, which in turn led to very high weak scaling efficiency in shown in Figure 3. To estimate the potential improvement of this parallelized implementation, the time each process wasted on finding the starting bundle was subtracted from total execution time. As shown in Figure 4, the idealized case shows increases in strong scaling efficiency by up to 2 folds, leading to a maximum speed up about 10 folds. The idealized case suggests that improvement can be made to approach the performance of the original pthread code shown in Figure 5. Figure 6 has shown that both the position and abundance of transcripts are perfectly consistent between serial and parallel code.

B.5 Conclusions and Future Work

Conclusions

Parallelization with MPI of Cufflinks significantly reduced the its running time compared to serial implementation, by up to 6 fold on a real data set. Although this MPI implementation does not scale as well as the original shared memory implementation, it allows user choose flexible environment and significantly reduce queue time. The scaling data suggest that pthread and MPI process should have comparable efficiency in this embarrassingly parallel algorithm.

Schemes of improved parallelization

Two different schemes are proposed for performance improvement as illustrated in Figure 7.

Scheme 1: MPI process with rank 0 reads the data files and determines the mapping parameters. It then sends the bundles reorganized on contiguous memory location as messages to other processes. The process with rank 0 will wait until all the processes including itself finish and combine all the temporary files written to the disk.

Scheme 2: MPI processes call the collective function MPI_FILE_OPEN, with offsets such that each process opens equal amount of file. Processes with adjacent ranks then communicate with each other and send/receive reads to ensure the completeness of chromosome (scaffold) in the memory of each process. Incomplete chromosomes will then be removed from the memory. The process with rank 0 will wait until all the processes including itself finish and combine all the temporary files written to the disk. This approach assumes that no transcript can be assembled across two chromosomes, which is biologically correct.

B.6 Source Code and Testing Data Accession

The source code, job submission script, input files, output files and raw testing data can be found on Crane under the directory: "/work/pblack/btu2/cufflinks-2.2.1/src-submit". The directory with write permission is "/work/pblack/btu2/cufflinks-2.2.1/src_rw". Please read the README file under the directory for instructions on opening/executing/compiling files.

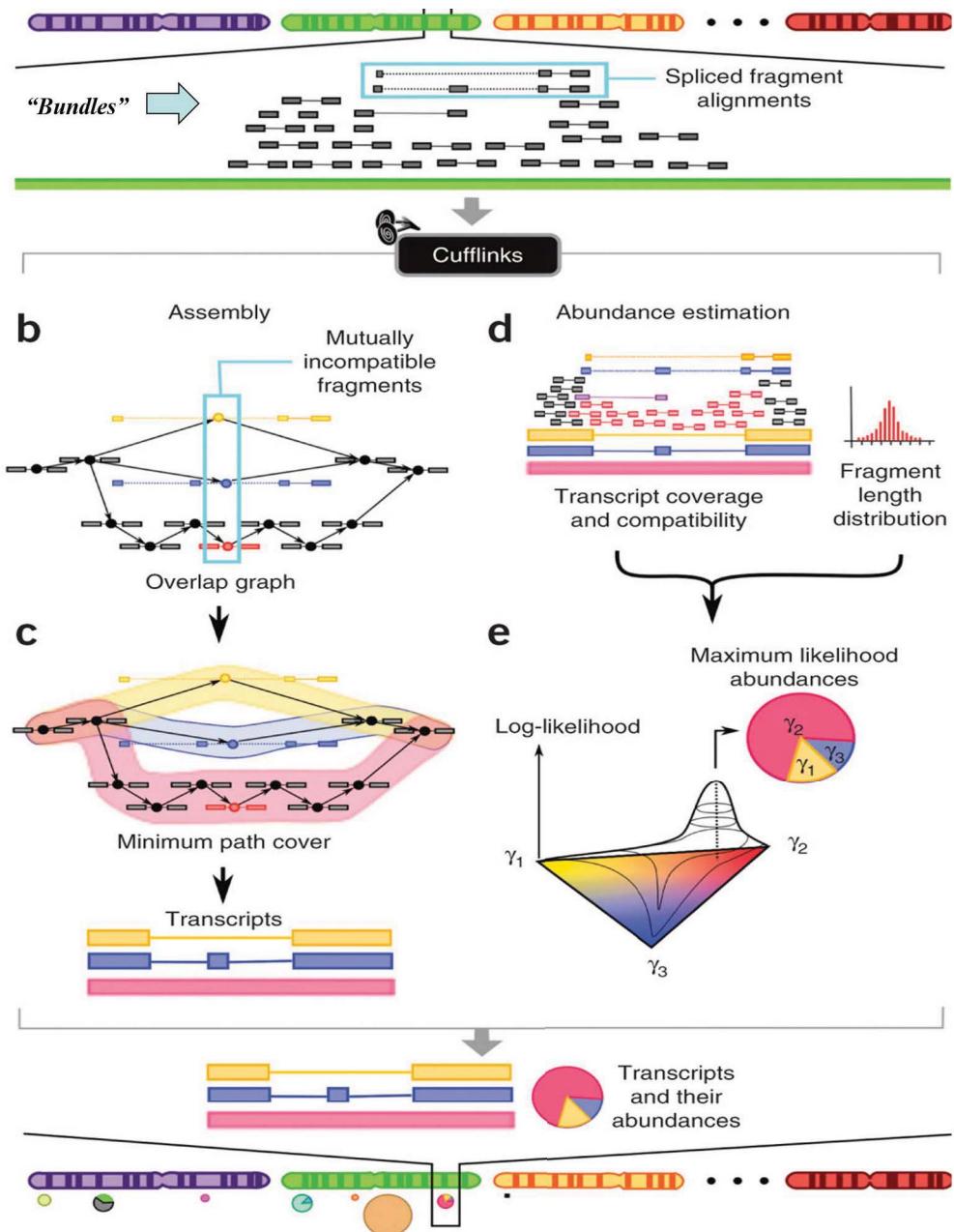


Figure B.1: The Cufflinks algorithm of transcriptome assembly. Mapped reads are grouped into bundles, which are processed by separated pthreads in the original implementation without communication (Adapted from Trapnell et al., 2011).

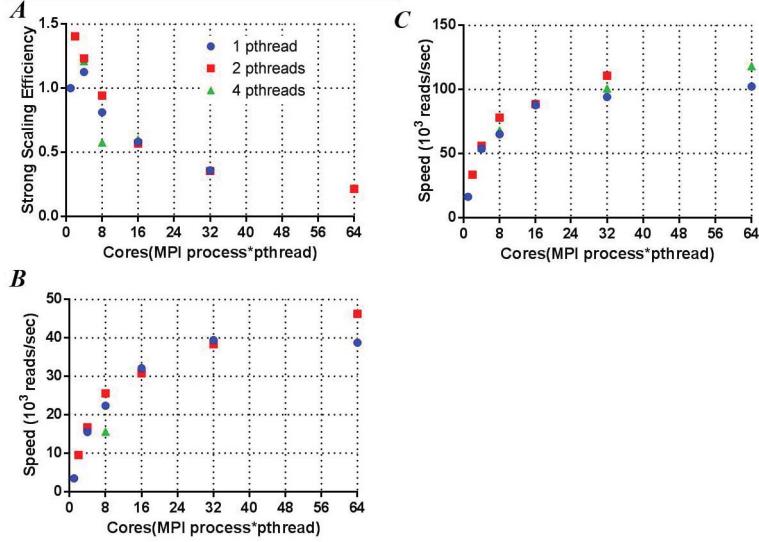


Figure B.2: Strong scaling of MPI parallelized code on Crane at HCC. The code was compiled with intel compiler 13 and mvapich2 1.9. The pthread was implemented with Boost library. **A.** Strong scaling efficiency of MPI code with different numbers of pthreads within each process. **B.** Speed-up on a real data file with 3 million reads. **C.** Speed-up on a real data file with 30 million reads.

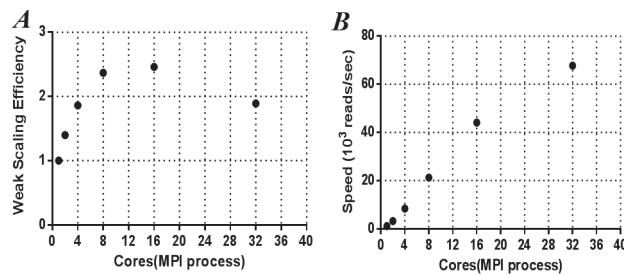


Figure B.3: Weak scaling of single-threaded MPI implementation. Multithreaded codes have similar weak scaling (data not shown). The high weak scaling efficiency is determined by the nature of the overlap algorithm. **A.** Weak scaling efficiency measured from 0.3 to 10 million reads with corresponding core numbers. **B.** Speed-up measured from 0.3 to 10 million reads with corresponding core numbers.

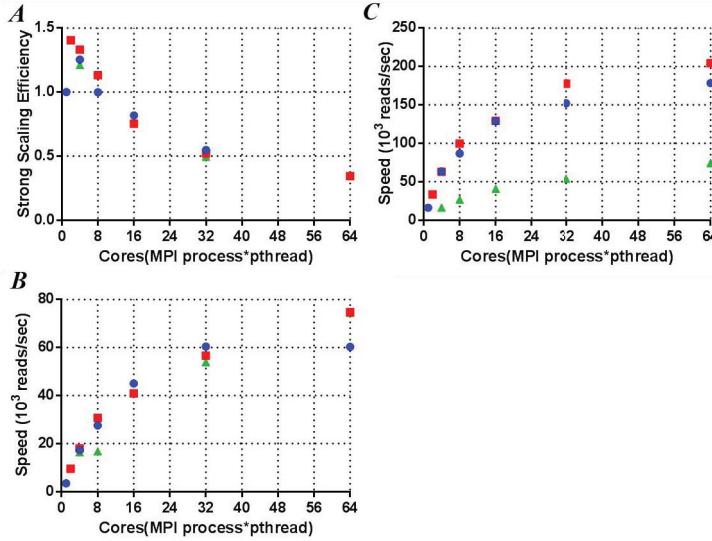


Figure B.4: Idealized strong scaling of MPI parallelized code. The time each process wasted on sequentially reading through the data file was subtracted from the total run time. **A.** Idealized strong scaling efficiency of MPI code with different numbers of pthreads within each process **B.** Idealized speed-up on a real data file with 3 million reads **C.** Idealized speed-up on a real data file with 30 million reads.

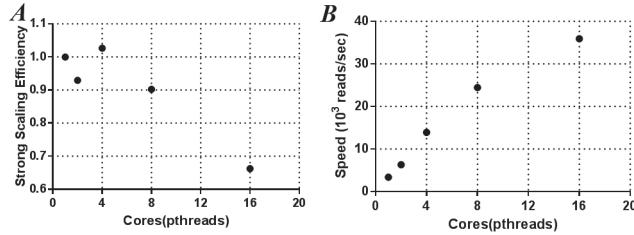


Figure B.5: Strong scaling of pure pthread implementation on a single node. **A.** Strong scaling efficiency **B.** Speed-up on a real data file with 3 million reads

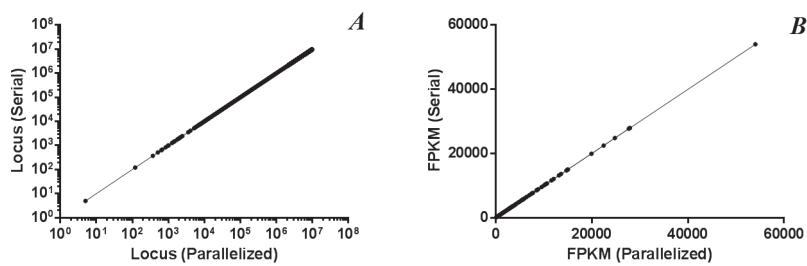


Figure B.6: Correctness check of the output from MPI parallelized implementation against the original implementation. **A.** Scatter plot of starting nucleotide position of loci from two output files. **B.** Scatter plot of transcript abundances from two output files. FPKM: fragments per kilobase of exon per million fragments mapped.

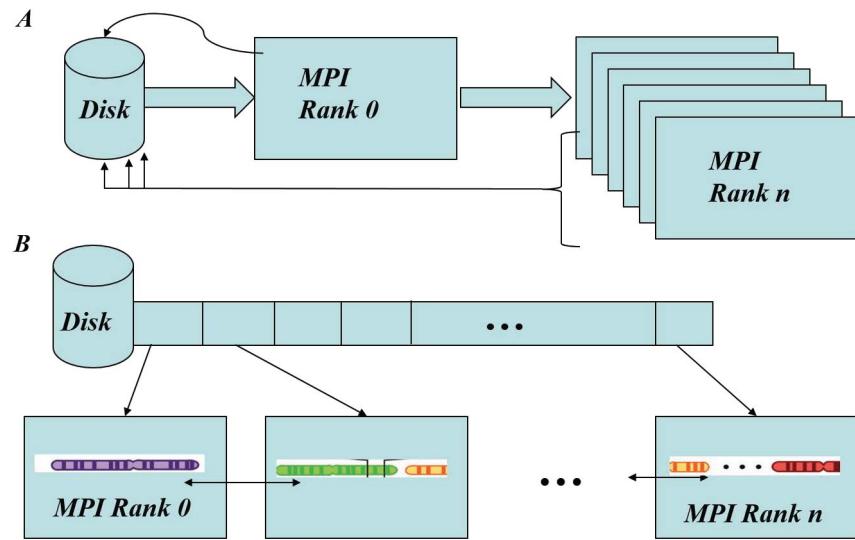


Figure B.7: Two schemes of improving the performance of this MPI implementation. **A.** Rank 0 reads data from disk and send data on contiguous memory to other processes. Temporary files from processes are combined in the end. **B.** Each rank collectively open the data file using parallel I/O and communicate with each other to obtain necessary data for completeness.

References

- [1] Wang M, Han J, Dunn J, Cai H, Elgowainy A. 2012. Well-to-wheels energy use and greenhouse gas emissions of ethanol from corn, sugarcane and cellulosic biomass for us use. *Environmental Research Letters* 7
- [2] Chisti Y. 2007. Biodiesel from microalgae. *Biotechnology Advances* 25:294–306
- [3] Gouveia L, Oliveira A. 2009. Microalgae as a raw material for biofuels production. *Journal of Industrial Microbiology and Biotechnology* 36:269–274
- [4] Scott S, Davey M, Dennis J, Horst I, Howe C, et al. 2010. Biodiesel from algae: Challenges and prospects. *Current Opinion in Biotechnology* 21:277–286
- [5] Wang Z, Ullrich N, Joo S, Waffenschmidt S, Goodenough U. 2009. Algal lipid bodies: Stress induction, purification, and biochemical characterization in wild-type and starchless chlamydomonas reinhardtii. *Eukaryotic Cell* 8:1856–1868
- [6] Cakmak T, Angun P, Demiray Y, Ozkan A, Elibol Z, Tekinay T. 2012. Differential effects of nitrogen and sulfur deprivation on growth and biodiesel feedstock production of chlamydomonas reinhardtii. *Biotechnology and Bioengineering* 109:1947–1957
- [7] Caccavo Jr. F, Ramsing N, Costerton J. 1996. Morphological and metabolic responses to starvation by the dissimilatory metal-reducing bacterium shewanella alga bry. *Applied and Environmental Microbiology* 62:4678–4682

- [8] Castruita M, Casero D, Karpowicz S, Kropat J, Vieler A, et al. 2011. Systems biology approach in chlamydomonas reveals connections between copper nutrition and multiple metabolic steps. *Plant Cell* 23:1273–1292
- [9] Wase N, Black P, Stanley B, Dirusso C. 2014. Integrated quantitative analysis of nitrogen stress response in chlamydomonas reinhardtii using metabolite and protein profiling. *Journal of Proteome Research* 13:1373–1396
- [10] Radakovits R, Jinkerson R, Darzins A, Posewitz M. 2010. Genetic engineering of algae for enhanced biofuel production. *Eukaryotic Cell* 9:486–501
- [11] Courchesne N, Parisien A, Wang B, Lan C. 2009. Enhancement of lipid production using biochemical, genetic and transcription factor engineering approaches. *Journal of Biotechnology* 141:31–41
- [12] Siaut M, Cuin S, Cagnon C, Fessler B, Nguyen M, et al. 2011. Oil accumulation in the model green alga chlamydomonas reinhardtii: Characterization, variability between common laboratory strains and relationship with starch reserves. *BMC Biotechnology* 11
- [13] Rosenberg J, Oyler G, Wilkinson L, Betenbaugh M. 2008. A green light for engineered algae: redirecting metabolism to fuel a biotechnology revolution. *Current Opinion in Biotechnology* 19:430–436
- [14] Allen J, DiRusso C, Black P. 2015. Triacylglycerol synthesis during nitrogen stress involves the prokaryotic lipid synthesis pathway and acyl chain remodeling in the microalgae cocomymxa subellipsoidea. *Algal Research* 10:110–120

- [15] Goncalves E, Johnson J, Rathinasabapathi B. 2013. Conversion of membrane lipid acyl groups to triacylglycerol and formation of lipid bodies upon nitrogen starvation in biofuel green algae chlorella utex29. *Planta* 238:895–906
- [16] Gupta P, Onder T, Jiang G, Tao K, Kuperwasser C, et al. 2009. Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* 138:645–659
- [17] Sandoval A, Chokshi A, Jesch E, Black P, DiRusso C. 2010. Identification and characterization of small compound inhibitors of human fatp2. *Biochemical Pharmacology* 79:990–999
- [18] Lemieux G, Liu J, Mayer N, Bainton R, Ashrafi K, Werb Z. 2011. A whole-organism screen identifies new regulators of fat storage. *Nature Chemical Biology* 7:206–213
- [19] Alfred S, Surendra A, Le C, Lin K, Mok A, et al. 2012. A phenotypic screening platform to identify small molecule modulators of chlamydomonas reinhardtii growth, motility and photosynthesis. *Genome Biology* :R105
- [20] Franz A, Danielewicz M, Wong D, Anderson L, Boothe J. 2013. Phenotypic screening with oleaginous microalgae reveals modulators of lipid productivity. *ACS Chemical Biology* 8:1053–1062
- [21] Wase N, Tu B, Black P, DiRusso C. 2015a. Phenotypic screening identifies brefeldin a/ascotoxin as an inducer of lipid storage in the algae chlamydomonas reinhardtii. *Algal Research* 11:74–84
- [22] Wang Z, Gerstein M, Snyder M. 2009. Rna-seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10:57–63

- [23] Merchant S, Prochnik S, Vallon O, Harris E, Karpowicz S, et al. 2007. The chlamydomonas genome reveals the evolution of key animal and plant functions. *Science* 318:245–251
- [24] Trapnell C, Pachter L, Salzberg S. 2009. Tophat: Discovering splice junctions with rna-seq. *Bioinformatics* 25:1105–1111
- [25] Trapnell Cc, Williams B, Pertea G, Mortazavi A, Kwan G, et al. 2010. Transcript assembly and quantification by rna-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* 28:511–515
- [26] Roberts A, Pachter Lc. 2013. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nature Methods* 10:71–73
- [27] Subramanian A, Tamayo P, Mootha V, Mukherjee S, Ebert B, et al. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 102:15545–15550
- [28] Ashburner M, Ball C, Blake J, Botstein D, Butler H, et al. 2000. Gene ontology: Tool for the unification of biology. *Nature Genetics* 25:25–29
- [29] Kanehisa M, Goto S, Kawashima S, Nakaya A. 2002. Thed kegg databases at genomenet. *Nucleic Acids Research* 30:42–46
- [30] Wase N, Tu B, Black P, DiRusso C. 2015b. Promoting algal obesity: Identification of compounds that stimulate lipid accumulation during growth in microalgae. *Under review*

- [31] Ruijter J, Ramakers C, Hoogaars W, Karlen Y, Bakker O, et al. 2009. Amplification efficiency: Linking baseline and bias in the analysis of quantitative pcr data. *Nucleic Acids Research* 37
- [32] Fatland B, Anderson M, Nikolau B, Wurtele E. 2000. Molecular biology of cytosolic acetyl-coa generation. *Biochemical Society Transactions* 28:593–595
- [33] Fatland B, Ke J, Anderson M, Mentzen W, Li W, et al. 2002. Molecular characterization of a heteromeric atp-citrate lyase that generates cytosolic acetyl-coenzyme a in arabidopsis. *Plant Physiology* 130:740–756
- [34] Boyle N, Page M, Liu B, Blaby I, Casero D, et al. 2012. Three acyltransferases and nitrogen-responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in chlamydomonas. *Journal of Biological Chemistry* 287:15811–15825
- [35] Goodenough U, Blaby I, Casero Dj, Gallaher S, Goodson C, et al. 2014. The path to triacylglyceride obesity in the sta6 strain of chlamydomonas reinhardtii. *Eukaryotic Cell* 13:591–613
- [36] Moellering E, Benning C. 2010. Rna interference silencing of a major lipid droplet protein affects lipid droplet size in chlamydomonas reinhardtii. *Eukaryotic Cell* 9:97–106
- [37] Prez-Prez M, Florencio F, Crespo J. 2010. Inhibition of target of rapamycin signaling and stress activate autophagy in chlamydomonas reinhardtii. *Plant Physiology* 152:1874–1888

- [38] Feng H, Zhang X, Zhang C. 2015. Mrin for direct assessment of genome-wide and gene-specific mrna integrity from large-scale rna-sequencing data. *Nature Communications* 6
- [39] Gallego Romero I, Pai A, Tung J, Gilad Y. 2014. Rna-seq: Impact of rna degradation on transcript quantification. *BMC Biology* 12
- [40] Chardin C, Girin T, Roudier F, Meyer C, Krapp A. 2014. The plant rwp-rk transcription factors: Key regulators of nitrogen responses and of gametophyte development. *Journal of Experimental Botany* 65:5577–5587
- [41] Camargo A, Llamas a, Schnell R, Higuera J, Gonzlez-Ballester D, et al. 2007. Nitrate signaling by the regulatory gene nit2 in chlamydomonas. *Plant Cell* 19:3491–3503
- [42] Lin H, Goodenough U. 2007. Gametogenesis in the chlamydomonas reinhardtii minus mating type is controlled by two genes, mid and mtd1. *Genetics* 176:913–925
- [43] Nozaki H, Mori T, Misumi O, Matsunaga S, Kuroiwa T. 2006. Males evolved from the dominant isogametic mating type. *Current Biology* 16:R1018–R1020
- [44] Almagro A, Shan H, Yi F. 2008. Characterization of the arabidopsis nitrate transporter nrt1.6 reveals a role of nitrate in early embryo development. *Plant Cell* 20:3289–3299
- [45] Remacle C, Eppe G, Coosemans N, Fernandez E, Vigeolas H. 2014. Combined intracellular nitrate and nit2 effects on storage carbohydrate metabolism in chlamydomonas. *Journal of Experimental Botany* 65:23–33
- [46] Kropat J, Tottey S, Birkenbihl R, Depge N, Huijser P, Merchant S. 2005. A regulator of nutritional copper signaling in chlamydomonas is an sbp domain protein that recognizes the gtac core of copper response element. *Proceedings of the National Academy of Sciences of the United States of America* 102:18730–18735

- [47] Bisova K, Krylov D, Umen J. 2005. Genome-wide annotation and expression profiling of cell cycle regulatory genes in chlamydomonas reinhardtii. *Plant Physiology* 137:475–491
- [48] Blaby I, Glaesener A, Mettler T, Fitz-Gibbon S, Gallaher S, et al. 2013. Systems-level analysis of nitrogen starvation-induced modifications of carbon metabolism in a chlamydomonas reinhardtii starchless mutant. *Plant Cell* 25:4305–4323
- [49] Bulon A, Gallant D, Bouchet B, Mouille G, D'Hulst C, et al. 1997. Starches from a to c: Chlamydomonas reinhardtii as a model microbial system to investigate the biosynthesis of the plant amylopectin crystal. *Plant Physiology* 115:949–957
- [50] Chen G, Greer M, Lager I, Lindberg Yilmaz J, Mietkiewska E, et al. 2012. Identification and characterization of an lcat-like arabidopsis thaliana gene encoding a novel phospholipase a. *FEBS Letters* 586:373–377
- [51] Sato N. 2004. Roles of the acidic lipids sulfoquinovosyl diacylglycerol and phosphatidylglycerol in photosynthesis: Their specificity and evolution. *Journal of Plant Research* 117:495–505
- [52] Riekhof W, Ruckle M, Lydic T, Sears B, Benning C. 2003. The sulfolipids 2'-o-acyl-sulfoquinovosyldiacylglycerol and sulfoquinovosyldiacylglycerol are absent from a chlamydomonas reinhardtii mutant deleted in sqd1. *Plant Physiology* 133:864–874
- [53] Irihimovitch V, Stern D. 2006. The sulfur acclimation sac3 kinase is required for chloroplast transcriptional repression under sulfur limitation in chlamydomonas reinhardtii. *Proceedings of the National Academy of Sciences of the United States of America* 103:7911–7916

- [54] Minoche A, Dohm J, Himmelbauer H. 2011. Evaluation of genomic high-throughput sequencing data generated on illumina hiseq and genome analyzer systems. *Genome Biology* 12
- [55] Caporaso J, Lauber C, Walters W, Berg-Lyons D, Huntley J, et al. 2012. Ultra-high-throughput microbial community analysis on the illumina hiseq and miseq platforms. *ISME Journal* 6:1621–1624
- [56] Li B, Dewey C. 2011. Rsem: Accurate transcript quantification from rna-seq data with or without a reference genome. *BMC Bioinformatics* 12
- [57] Miller R, Wu G, Deshpande R, Vieler A, Gartner K, et al. 2010. Changes in transcript abundance in chlamydomonas reinhardtii following nitrogen deprivation predict diversion of metabolism. *Plant Physiology* 154:1737–1752
- [58] Fan J, Andre C, Xu C. 2011. A chloroplast pathway for the de novo biosynthesis of triacylglycerol in chlamydomonas reinhardtii. *FEBS Letters* 585:1985–1991
- [59] Li X, Moellering Ed, Liu B, Johnny C, Fedewa M, et al. 2012. A galactoglycerolipid lipase is required for triacylglycerol accumulation and survival following nitrogen deprivation in chlamydomonas reinhardtii. *Plant Cell* 24:4670–4686
- [60] Gargouri M, Park JJ, Holguin F, Kim MJ, Wang H, et al. 2015. Identification of regulatory network hubs that control lipid metabolism in chlamydomonas reinhardtii. *Journal of Experimental Botany* 66:4551–4566
- [61] Harris Gc, Scanlan D, Geider R. 2009. Responses of emiliania huxleyi (prymnesiophyceae) to step changes in photon flux density. *European Journal of Phycology* 44:31–48

- [62] Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J. 2007. qbase relative quantification framework and software for management and automated analysis of real-time quantitative pcr data. *Genome biology* 8:R19