

**HOW TO EAT AND LOSE WEIGHT:  
A SECOND LOOK**

by

Sue Dunham

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Science in Bioinformatics and Computational Biology

Summer 2014

© 2014 Sue Dunham  
All Rights Reserved

**HOW TO EAT AND LOSE WEIGHT:  
A SECOND LOOK**

by

Sue Dunham

Approved: \_\_\_\_\_

Eleftherios T. Papoutsakis, Ph.D.

Professor in charge of thesis on behalf of the Advisory Committee

Approved: \_\_\_\_\_

Fats Domino, Ph.D.

Chair of the Department of Nutrition

Approved: \_\_\_\_\_

Minnesota Fats, Ph.D.

Dean of the College of Hard Knockwurst

Approved: \_\_\_\_\_

James G. Richards, Ph.D.

Vice Provost for Graduate and Professional Education

## ACKNOWLEDGMENTS

I write this paper with unending thanks for my mother Donna, father Thomas, and sister Allison for their love, support, and encouragement throughout this adventure. With special thanks for my grandfather George for his love, support, encouragement, and inspiration for my pursuit of science and chemistry. With special thanks for my grandmother Winnie for her love, support, encouragement, and optimism. With special thanks for my uncle Kevin for his support, encouragement and inspiration. With special thanks for my brother-in-law Matt for his love, support, and encouragement. With special thanks for my nieces Violet and Ruby for their love, for all that they teach me, and for the energy and optimism that they give. Many thanks for Madeline for her love, open ear, curiosity, support, and encouragement. Many thanks to my best friend Andrew for his friendship, curiosity, support and encouragement. With many thanks to the rest of my wonderful and supportive family and friends for their love and support.

Many thanks for my mentors, Drs. Venkataramanan and Papoutsakis, for their mentorship, support, critique, and understanding throughout this project. Many thanks for each of the the many members of the Papoutsakis lab for their friendship, support, and critique. Many thanks to Bruce, Olga, and Summer for their work in the Sequencing and Genotyping Center.

*Ad maiorem dei gloriam.*

## TABLE OF CONTENTS

<b>LIST OF TABLES</b> . . . . .	<b>v</b>
<b>LIST OF FIGURES</b> . . . . .	<b>vi</b>
<b>ABSTRACT</b> . . . . .	<b>vii</b>
 <b>Chapter</b>	
<b>1 METHODS</b> . . . . .	<b>1</b>
1.1 Culture . . . . .	1
1.2 RNA preparation . . . . .	1
1.3 RNA concentration and RNA-seq library preparation . . . . .	2
1.4 Data Analysis . . . . .	2
1.5 Molecular Methods . . . . .	2
<b>2 RESULTS</b> . . . . .	<b>4</b>
<b>3 DISCUSSION</b> . . . . .	<b>5</b>
<b>4 CONCLUSIONS</b> . . . . .	<b>6</b>
<b>BIBLIOGRAPHY</b> . . . . .	<b>7</b>

## LIST OF TABLES

## LIST OF FIGURES

## ABSTRACT

## Chapter 1

### METHODS

#### 1.1 Culture

Wild type *Clostridium acetobutylicum* ATCC 824 was cultured anaerobically in 4L New Brunswick Scientific BioFlo 310 bioreactors at 37 °C, pH 5.0, 200 mL min<sup>-1</sup> N<sub>2</sub> and 200rpm agitation in a defined *Clostridia* growth medium, as described previously (Venkataramanan 2013). When the cultures were grown to A<sub>600</sub>=1, the N<sub>2</sub> flow rate was decreased to 50 mL min<sup>-1</sup> and cultures were stressed to a final concentration of 90 mM *n*-butanol, 48 mM potassium butyrate, or left unstressed. 15 mL samples were acquired at 15, 75, 150, and 270 minutes after treatment and centrifuged at 8,000rpm, 4 °C for 20 minutes. After discarding the supernatant, cell pellets were then immediately frozen at -85 °C.

#### 1.2 RNA preparation

RNA was extracted by first washing the cell pellets in 1mL of RNase-free SET buffer (25% sucrose, 50 mM EDTA [pH 8.0], 50 mM Tris-HCl [pH 8.0]) before resuspending cells in a 220 mL solution of RNase-free SET buffer containing 4.55 U mL<sup>-1</sup> proteinase K and 20 mg mL<sup>-1</sup> lysozyme and incubating for 6 minutes. Resuspended cells were vortexed with 40mg of RNase-free glass beads ( $\leq 106 \mu\text{m}$ ) at maximum speed and room temperature for 4 minutes. Then each sample was mixed with 1 mL of ice-cold QIAzol (Qiagen, Valencia, CA, USA) and then 200  $\mu\text{L}$  of ice-cold chloroform. After a 3 minute room temperature incubation, samples were centrifuged at 11,000rpm for 15 minutes at 4 °C. The aqueous phase was then mixed with 1.3 mL of ice-cold ethanol before transferring to a miRNeasy Mini spin-column (Qiagen, Valencia, CA, USA), following the manufacturers instructions. Samples were then treated with the Turbo



DNA-free kit (Ambion, Austin, TX, USA) according to manufacturers instructions. Samples were then quantified with a Nanodrop , aliquoted for quality analysis with the BioAnalyzer platform (Agilent, Wilmington, DE, USA), and 10  $\mu$ g aliquots in 10  $\mu$ L samples were stored at  $-85^{\circ}\text{C}$ .

### 1.3 RNA concentration and RNA-seq library preparation

Ribosomal RNA was removed with the MicrobExpress kit (Ambion, Austin, TX, USA). Enriched RNA was assessed for quality with the BioAnalyzer platform (Agilent, Wilmington, DE, USA). High quality samples were used to prepare RNA-seq libraries with the ScriptSeq v2 library preparation kit and indexed PCR primers (Epicentre, Madison, WI, USA). Libraries were multiplexed and sequenced for 101 cycles over two lanes of an Illumina HiSeq 2500 at the University of Delaware Sequencing and Genotyping Center (Newark, DE, USA).

### 1.4 Data Analysis

Paired-end sequencing resulted in 100bp reads which are deposited in the Sequence Read Archive ( ). After removing adapters and low quality bases with Trimmomatic, reads were first mapped to the ribosomal RNA sequences before mapping to the *C. acetobutylicum* genome both with Bowtie 2.1.0 [parameters: ]. The resulting alignment was then cleaned, validated, sorted, and indexed with SAMtools and Picard. Unmapped and duplicate reads were flagged for omission from further analysis. Transcriptome assembly was performed with the Cufflinks suite. Transcriptional start sites were identified with two peak-finding algorithms TSSi and MACS. The fully merged transcriptome assembly were used to determine digital gene expression for visualization in R and Circos. Clustering was done with and promoter predictions were performed with MEME, and RSAT.

### 1.5 Molecular Methods

Fold changes were confirmed for randomly selected genes by qRT-PCR. Transcriptional start sites were verified by 5'RACE for randomly selected genes. Small

RNAs were confirmed by RT-PCR and Northern blot.

## **Chapter 2**

### **RESULTS**

**Chapter 3**  
**DISCUSSION**

**Chapter 4**  
**CONCLUSIONS**

## BIBLIOGRAPHY

- [1] Keith V Alsaker, Thomas R Spitzer, and Eleftherios T Papoutsakis. Transcriptional analysis of *spo0a* overexpression in *clostridium acetobutylicum* and its effect on the cell's response to butanol stress. *Journal of bacteriology*, 186(7):1959–1971, 2004.
- [2] S Andrews. Fastqc: A quality control tool for high throughput sequence data. *Reference Source*, 2010.
- [3] Dominik Antoni, Vladimir V Zverlov, and Wolfgang H Schwarz. Biofuels from microbes. *Applied microbiology and biotechnology*, 77(1):23–35, 2007.
- [4] Shota Atsumi, Anthony F Cann, Michael R Connor, Claire R Shen, Kevin M Smith, Mark P Brynildsen, Katherine JY Chou, Taizo Hanai, and James C Liao. Metabolic engineering of *escherichia coli* for 1-butanol production. *Metabolic engineering*, 10(6):305–311, 2008.
- [5] Timothy L Bailey, Mikael Boden, Fabian A Buske, Martin Frith, Charles E Grant, Luca Clementi, Jingyuan Ren, Wilfred W Li, and William S Noble. Meme suite: tools for motif discovery and searching. *Nucleic acids research*, 37(suppl 2):W202–W208, 2009.
- [6] Jacob R Borden and Eleftherios Terry Papoutsakis. Dynamics of genomic-library enrichment and identification of solvent tolerance genes for *clostridium acetobutylicum*. *Applied and environmental microbiology*, 73(9):3061–3068, 2007.
- [7] Yili Chen, Dinesh C Indurthi, Shawn W Jones, and Eleftherios T Papoutsakis. Small rnas in the genus *clostridium*. *MBio*, 2(1):e00340–10, 2011.
- [8] Patrik Dhaeseleer, Shoudan Liang, and Roland Somogyi. Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics*, 16(8):707–726, 2000.
- [9] A Gordon and GJ Hannon. Fastx-toolkit. *FASTQ short-reads pre-processing tools*, 2010.
- [10] Michael Hecker, Sandro Lambeck, Susanne Toepfer, Eugene Van Someren, and Reinhard Guthke. Gene regulatory network inference: data integration in dynamic modelsa review. *Biosystems*, 96(1):86–103, 2009.

- [11] Michael Hecker, Wolfgang Schumann, and Uwe Vlker. Heatshock and general stress response in bacillus subtilis. *Molecular microbiology*, 19(3):417–428, 1996.
- [12] Eric CH Ho, Michael E Donaldson, and Barry J Saville. *Detection of antisense RNA transcripts by strand-specific RT-PCR*, pages 125–138. Springer, 2010.
- [13] Jay D Keasling, Abraham Mendoza, and Phil S Baran. Synthesis: A constructive debate. *Nature*, 492(7428):188–189, 2012.
- [14] Donghyuk Kim, Jay Sung-Joong Hong, Yu Qiu, Harish Nagarajan, Joo-Hyun Seo, Byung-Kwan Cho, Shih-Feng Tsai, and Bernhard Palsson. Comparative analysis of regulatory elements between escherichia coli and klebsiella pneumoniae by genome-wide transcription start site profiling. *PLoS genetics*, 8(8):e1002867, 2012.
- [15] Clemens Kreutz, JS Gehring, D Lang, Ralf Reski, Jens Timmer, and Stefan A Rensing. Tssian r package for transcription start site identification from 5 mrna tag data. *Bioinformatics*, 28(12):1641–1642, 2012.
- [16] Ethan I Lan and James C Liao. Metabolic engineering of cyanobacteria for 1-butanol production from carbon dioxide. *Metabolic engineering*, 13(4):353–363, 2011.
- [17] Ben Langmead and Steven L Salzberg. Fast gapped-read alignment with bowtie 2. *Nature methods*, 9(4):357–359, 2012.
- [18] Joshua Z Levin, Moran Yassour, Xian Adiconis, Chad Nusbaum, Dawn Anne Thompson, Nir Friedman, Andreas Gnirke, and Aviv Regev. Comprehensive comparative analysis of strand-specific rna sequencing methods. *Nature methods*, 7(9):709–715, 2010.
- [19] Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–2079, 2009.
- [20] Ronny Lorenz, Stephan HF Bernhart, Christian Hoener Zu Siederdisen, Hakim Tafer, Christoph Flamm, Peter F Stadler, and Ivo L Hofacker. Viennarna package 2.0. *Algorithms for Molecular Biology*, 6(1):26, 2011.
- [21] Marloes H Maathuis, Diego Colombo, Markus Kalisch, and Peter Bhlmann. Predicting causal effects in large-scale systems from observational data. *Nature Methods*, 7(4):247–248, 2010.
- [22] Florian Markowetz, Dennis Kostka, Olga G Troyanskaya, and Rainer Spang. Nested effects models for high-dimensional phenotyping screens. *Bioinformatics*, 23(13):i305–i312, 2007.

- [23] G Najoshi. Sickle-a windowed adaptive trimming tool for fastq files using quality.
- [24] Sergios A Nicolaou, Stefan M Gaida, and Eleftherios T Papoutsakis. A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: from biofuels and chemicals, to biocatalysis and bioremediation. *Metabolic engineering*, 12(4):307–331, 2010.
- [25] Fatih Ozsolak and Patrice M Milos. Rna sequencing: advances, challenges and opportunities. *Nature reviews genetics*, 12(2):87–98, 2011.
- [26] Eleftherios T Papoutsakis. Engineering solventogenic clostridia. *Current opinion in biotechnology*, 19(5):420–429, 2008.
- [27] Christophe Pichon and Brice Felden. Small rna gene identification and mrna target predictions in bacteria. *Bioinformatics*, 24(24):2807–2813, 2008.
- [28] Yosef Prat, Menachem Fromer, Nathan Linial, and Michal Linial. Recovering key biological constituents through sparse representation of gene expression. *Bioinformatics*, 27(5):655–661, 2011.
- [29] Juan L Ramos, Estrella Duque, Mara-Trinidad Gallegos, Patricia Godoy, Mara Isabel Ramos-Gonzalez, Antonia Rojas, Wilson Tern, and Ana Segura. Mechanisms of solvent tolerance in gram-negative bacteria. *Annual Reviews in Microbiology*, 56(1):743–768, 2002.
- [30] Adam Roberts, Harold Pimentel, Cole Trapnell, and Lior Pachter. Identification of novel transcripts in annotated genomes using rna-seq. *Bioinformatics*, 27(17):2325–2329, 2011.
- [31] Tobias Sahr, Christophe Rusniok, Delphine Dervins-Ravault, Odile Sismeiro, Jean-Yves Coppee, and Carmen Buchrieser. Deep sequencing defines the transcriptional map of *l. pneumophila* and identifies growth phase-dependent regulated ncRNAs implicated in virulence. *RNA Biol*, 9(4):503–519, 2012.
- [32] Yogita Sardesai and Saroj Bhosle. Tolerance of bacteria to organic solvents. *Research in Microbiology*, 153(5):263–268, 2002.
- [33] Robert Schmieder and Robert Edwards. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27(6):863–864, 2011.
- [34] Olga A Soutourina, Marc Monot, Pierre Boudry, Laure Saujet, Christophe Pichon, Odile Sismeiro, Ekaterina Semenova, Konstantin Severinov, Chantal Le Bouguenec, and Jean-Yves Coppe. Genome-wide identification of regulatory RNAs in the human pathogen *clostridium difficile*. *PLoS genetics*, 9(5):e1003493, 2013.



- [35] Morgane Thomas-Chollier, Matthieu Defrance, Alejandra Medina-Rivera, Olivier Sand, Carl Herrmann, Denis Thieffry, and Jacques van Helden. Rsat 2011: regulatory sequence analysis tools. *Nucleic acids research*, 39(suppl 2):W86–W91, 2011.
- [36] Christopher A Tomas, Jeffrey Beamish, and Eleftherios T Papoutsakis. Transcriptional analysis of butanol stress and tolerance in *clostridium acetobutylicum*. *Journal of bacteriology*, 186(7):2006–2018, 2004.
- [37] Bryan P Tracy, Shawn W Jones, Alan G Fast, Dinesh C Indurthi, and Eleftherios T Papoutsakis. Clostridia: the importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications. *Current opinion in biotechnology*, 23(3):364–381, 2012.
- [38] Charles J Vaske, Stephen C Benz, J Zachary Sanborn, Dent Earl, Christopher Szeto, Jingchun Zhu, David Haussler, and Joshua M Stuart. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using paradigm. *Bioinformatics*, 26(12):i237–i245, 2010.
- [39] Keerthi P Venkataramanan, Shawn W Jones, Kevin P McCormick, Sridhara G Kunjeti, Matthew T Ralston, Blake C Meyers, and Eleftherios T Papoutsakis. The *clostridium* small rnome that responds to stress: the paradigm and importance of toxic metabolite stress in *c. acetobutylicum*. *BMC genomics*, 14(1):849, 2013.
- [40] Alec W. Picard, 2009.
- [41] E Gerhart H Wagner. Kill the messenger: bacterial antisense rna promotes mrna decay. *Nature structural and molecular biology*, 16(8):804–806, 2009.
- [42] Qinghua Wang, Keerthi Prasad Venkataramanan, Hongzhan Huang, Eleftherios T Papoutsakis, and Cathy H Wu. Transcription factors and genetic circuits orchestrating the complex, multilayered response of *clostridium acetobutylicum* to butanol and butyrate stress. *BMC systems biology*, 7(1):120, 2013.
- [43] Ka Yee Yeung and Walter L. Ruzzo. Principal component analysis for clustering gene expression data. *Bioinformatics*, 17(9):763–774, 2001.
- [44] Yong Zhang, Tao Liu, Clifford A Meyer, Jrme Eeckhoutte, David S Johnson, Bradley E Bernstein, Chad Nusbaum, Richard M Myers, Myles Brown, and Wei Li. Model-based analysis of chip-seq (macs). *Genome Biol*, 9(9):R137, 2008.
- [45] Kyle A Zingaro and Eleftherios Terry Papoutsakis. Toward a semisynthetic stress response system to engineer microbial solvent tolerance. *Mbio*, 3(5):e00308–12, 2012.

- [46] Kyle A Zingaro and Eleftherios Terry Papoutsakis. Groesl overexpression imparts *escherichia coli* tolerance to *i* *i*/*i*-, *i* *n*/*i*-, and 2-butanol, 1, 2, 4-butanetriol and ethanol with complex and unpredictable patterns. *Metabolic engineering*, 15:196–205, 2013.