TRANSCRIPTOME ANALYSIS IN MAMMALIAN CELL CULTURE: APPLICATIONS IN PROCESS DEVELOPMENT AND CHARACTERIZATION

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Anne Kantardjieff

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Wei-Shou Hu

August, 2009

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Prof. Wei-Shou Hu. He is a consummate teacher, who always puts the best interests of his students first. I am eternally grateful for all the opportunities he has given me and all that I have learned from him. I can only hope to prove as inspriring to others as he has been to me.

I would like to thank my thesis committee members, Prof. Kevin Dorfman, Prof. Scott Fahrenkrug, and Prof. Friedrich Srienc, for taking the time to serve on my committee.

It goes without saying that what makes the Hu lab a wonderful place to work are the people. I consider myself lucky to have joined what could only be described as a family. Thank you to all the Hu group members, past and present: Jongchan Lee, Wei Lian, Mugdha Gadgil, Sarika Mehra, Marcela de Leon Gatti, Ziomara Gerdtzen, Patrick Hossler, Katie Wlaschin, Gargi Seth, Fernando Ulloa, Joon Chong Yee, C.M. Cameron, David Umulis, Karthik Jayapal, Salim Charaniya, Marlene Castro, Nitya Jacob, Bhanu Mulukutla, Siguang Sui, Kartik Subramanian, Cornelia Bengea, Huong Le, Anushree Chatterjee, Jason Owens, Shikha Sharma, Kathryn Johnson, Eyal Epstein, ze Germans, Kirsten Keefe, Kim Coffee, Katherine Mattews and Jessica Raines-Jones.

This work was enabled by the great collaboration we have had with the Bioprocessing Technology Institute in Singapore. I would like to thank Prof. Miranda Yap, Peter Morin Nissom, Robin Philp, Kathy Wong, Song-Hui Chuah, Bernard Loo, Faraaz Yusufi, Eugene Koh and Eric Tan. I am happy to count you as friends. This work was also the result of a number of fruitful industrial collaborations, and I would like to take the opportunity to thank all of the wonderful people I had the chance to collaborate with.

Thank you to Scott McIvor and the people of his lab, especially Paul Score, who first taught me how to clone (when I had no idea what cloning was!). I would also like to thank Prof. Jennifer Maynard, who was a great teacher to me in my first few years. I would like to thank all of the people at the Minnesota Supercomputing Institute who were

invaluable to our bioinformatics efforts, including Zheng Jin Tu and Wen Dong. I would also like to thank Kevin Silverstein, for always doing his best to answer my endless questions.

I would like to thank all of the wonderful friends I have made during my time here. You are the reason grad school was so much fun. Thank you to Allison, Josh and Alex for a great way to spend Sunday nights! Finally, I would like to thank my family, on whose support I have always been able to count. Thank you for all your love and encouragement.

ABSTRACT

The advent of recombinant protein therapeutics more than two decades ago fundamentally transformed healthcare paradigms and has since improved the quality of life of millions of people. The production of these complex products is typically carried out in cultured mammalian cell lines, with a few cell lines accounting for the majority of production. Chief among these is the Chinese hamster ovary (CHO) cell line. Despite its importance, little genomic information is currently available in the public domain for this cell line. Consequently, our lab has devoted significant efforts to the development of genomic tools for CHO, including custom Affymetrix microarrays. These tools enable the global study of cellular gene expression. This thesis research has applied these transcriptome analysis tools to further understand the process of recombinant protein production.

The course of bringing a recombinant protein product to production scale involves a series of complex and often lengthy steps. As demand for these products continues to increase, there is a need to streamline process development efforts. One approach to facilitate this process is to increase our fundamental understanding of cell culture processes. Microarrays are well-suited to this application, and in this work, transcriptome analysis has been use to characterize multiple facets of cell culture process development, including cell line development and modulation of process parameters in fed-batch cultures. We found significant variation amongst clonal gene expression profiles during cell line development, an observation which could be exploited to develop gene expression-based clone screening protocols. We also uncovered the widespread impact process parameters can have on cellular gene expression. In particular, we found that raw material source, namely different hydrolysate lots, has a profound effect on the transcriptional signatures of fed-batch cell culture processes.

As next-generation sequencing technologies become increasingly mature and cost-effective, they are now being applied to the study of gene expression. We have used ultra high-throughput sequencing to investigate the deep transcriptome of CHO cells. We found that the technology correlated well with microarrays, and displayed a significantly broader detection range. Through this analysis, we also identified a number of transcriptionally-active regions in the CHO genome. The unprecedented depth achievable through next-generation sequencing now allows us to set genome sequencing firmly in our sights.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

Recombinant protein therapeutics have fundamentally altered the landscape of medicine, providing life-saving therapies for a broad range of indications, including cancer and rheumatological diseases. Due to their structural complexity, as well as their sometimes extensive post-translational modifications, these recombinant products are typically produced in cultured mammalian cells. There are currently 165 recombinant protein products on the market, with total annual sales of more than $44 billion (Aggarwal, 2008). The demand for these products is expected to continue to increase as a growing number of therapeutics enter clinical trials, and experts are predicting that the industry will continue to grow at near double digit rates (Chartrain and Chu, 2008). Consequently, there is a need for continued innovation and improvement in the field of cell culture process development.

However, the manufacturing challenges encountered in mammalian cell culture processes are much greater than those in bacterial cultures. The production of recombinant protein products at an industrial scale involves a series of technically complex and lengthy steps. Furthermore, many of these steps are currently treated as black boxes, with empirical optimization often driving process improvements. As the field continues to push for increased product yields, there is a drive to reduce the timelines required to bring a product to production scale. One approach to achieve this objective is to improve our fundamental understanding of cell culture processes, and to use this biological knowledge for the design of rational process improvement strategies. Also driving the need for an increased biological understanding of cell culture processes is the emergence of biosimilars, as patents for licensed recombinant therapeutics begin to expire. In order to stay competitive in a biogenerics market, it will be important to consistently deliver a high-quality product. By understanding which biological pathways are key during production, better process control strategies can be developed which ensure that product quality specifications stay within range.

The completion of the human genome sequencing project ushered in a new era of scientific research, and fundamentally transformed bioscience research and

biotechnology development. The availability of DNA sequence information spurred the development of various genomic tools, which have revolutionized traditional paradigms in research. One example is the development of microarray technology, a research tool which enables the global analysis of the expression of tens of thousands of transcripts in a single experiment. Transcriptome analysis has already yielded fundamental biological insights into numerous diseases, including cancer. While this technology has yet to be widely applied in the field of mammalian cell culture, transcriptome analysis can find multiple applications in this area, as shown in Figure 1. This thesis research focuses on the application of transcriptome analysis to multiple aspects of cell culture. One reason why transcriptome analysis has yet to be widely applied in cell culture has been the lack of genomic resources for one of the workhorses of the mammalian cell culture industry, the Chinese hamster ovary (CHO) cell. To address this issue, our lab has worked extensively in the past several years to develop transcriptome analysis tools for this widely used cell line. The development of these tools has been instrumental in my research, and our efforts to improve these tools are ongoing.

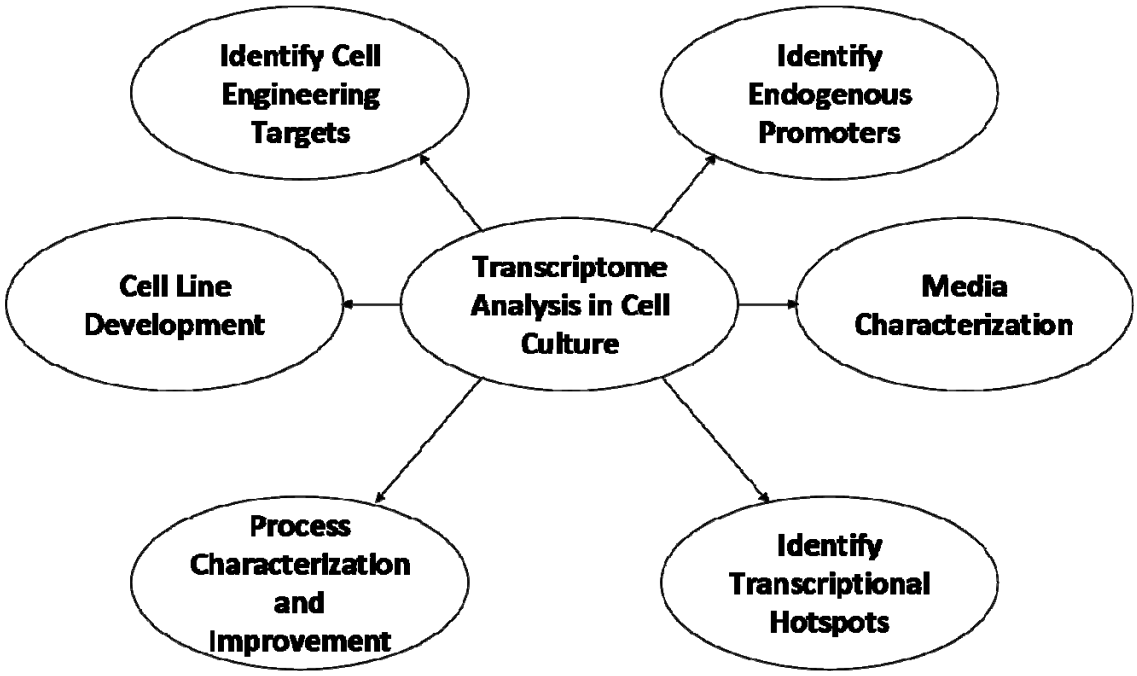


Figure 1 – Applications of transcriptome analysis in cell culture demonstrated in this research.

1.1 THESIS ORGANIZATION

This thesis demonstrates the multiple applications of transcriptome analysis in mammalian cell culture. A brief background of the cell culture field and transcriptome analysis is provided in Chapter 2. Furthermore, details regarding our development of custom Affymetrix platforms for Chinese hamster ovary cells are also provided in Chapter 2. The use of microarrays in the context of cell line development is the topic of Chapter 3. Next, transcriptome analysis was applied to characterize process conditions known to increase recombinant protein production, and is the subject of Chapter 4. Microarrays were then used for characterization of a CHO fed-batch process for recombinant protein production. The transcriptional impact of modifying process parameters was evaluated in this study, and is presented in Chapter 5. One particular process parameter which was studied, varying sources of raw materials in the growth medium, represents an application of transcriptome analysis for media characterization and is detailed in Section 5.6. An example of the use of microarrays for the identification of cell engineering targets is presented in Chapter 6. The application of transcriptome analysis to identify endogenous cellular promoter is described in Chapter 7. Finally, next-generation sequencing technology results were compared to microarray results and used to identify transcriptional hotspots in Chapter 8. A brief conclusion and future directions are presented in Chapter 9.

2 BACKGROUND

This chapter provides an overview of recombinant protein production in cultured mammalian cells. In particular, detailed attention is paid to the use of Chinese hamster ovary (CHO) cells, a cell line which accounts for the production of more than 90% of biologics currently on the market (Jayapal et al., 2007a). An overview of the process of developing a cell line for protein production is presented, and the opportunities which exist for improvement are highlighted. Next, the fundamentals of transcriptome analysis are described. Particular importance is placed on our development of genomic tools for CHO cells, for which very little sequence information is currently available in the public domain. Finally, a brief review of the progress being made in the field of transcriptome analysis is presented, with a focus on the use of next-generation sequencing technologies as an increasingly affordable alternative to microarrays.

2.1 RECOMBINANT PROTEIN PRODUCTION IN CULTURED MAMMALIAN CELLS

At the conclusion of the last century, the landscape of healthcare underwent a major transformation with the arrival of a new generation of therapeutics based on recombinant proteins. These biologics, as they are named to differentiate them from chemically-synthesized drugs, have been widely applied to the treatment of cancers, rheumatoid arthritis and bleeding disorders, among others (Aggarwal, 2008). Except for a few cases, these biologics require extensive post-translational modifications, such as glycosylation and disulfide bond formation, and are consequently produced using recombinant mammalian cells. More than 90% of these products are manufactured using CHO cells (Jayapal et al., 2007a), originally derived more than five decades ago (Puck et al., 1958). CHO cells secrete little protein in their native state, yet upon receiving the product transgene and subsequently undergoing gene amplification and other manipulations, can become high producers, secreting as much as or even more than professional secretors in our body, such as hepatocytes and antibody-producing plasma cells (Wurm, 2004). Along the route to becoming super producers, they also develop the required capacity for folding and glycosylation of the product protein. This process is likely to entail extensive reprogramming of their cellular regulation, possibly not unlike

the differentiation of endoderm cells to highly secretory hepatocytes in developing embryos. However, in spite of the economic importance and industrial significance of generating CHO-derived recombinant cells highly productive in the desired molecules, little is known about this biological transformation event.

2.1.1 Host Cell Lines for Recombinant Protein Production

Only a few cell lines account for the majority of therapeutic protein production, notably, Chinese hamster ovary (CHO) cells, mouse myeloma cells (NS0 and SP0), mouse hybridoma cells and baby hamster kidney (BHK) cells (Waterston et al., 2002). The parental CHO cell line (CHO-K1) was initiated from a biopsy of an adult Chinese hamster ovary by T.T. Puck in 1957 (Puck, 1957). Since then, a variety of established cell lines have been developed from this parental cell line. Critical to their widespread use in recombinant protein production however was the development of CHO cell lines deficient in the dihydrotetrafolate enzyme (DHFR). These CHO DHFR-minus cell lines were developed in the mid-1980s and two principal lineages have emerged which are commonly used for recombinant protein production: DG44, a double deletion mutant which contains no copies of the DHFR gene (Urlaub et al., 1986) and DXB11, which contains a point mutation in one DHFR allele, and a deletion mutation in the other (Urlaub and Chasin, 1980).

Another commonly used cell line in the production of recombinant protein products is NS0, a mouse myeloma cell line. NS0 cells originated from plasma cell neoplasms induced as a result of injecting mineral oil into the intraperitoneal region of BABL/c mice (Barnes et al., 2000). A continuous tissue culture cell line was developed from these neoplasms, and following a series of subsequent selection steps, the NS0 cell line commonly used today was obtained. Since NS0 cells are derived from immunoglobulin-secreting plasma cells, they are considered to be well-equipped for the production and secretion of recombinant protein products.

A summary of the recombinant antibody products currently on the market, and the cell lines used to produce them is shown in Table 1 (Chartrain and Chu, 2008). It should be noted however that biologics extend beyond recombinant antibodies to include growth

factors, hormones, fusion proteins, cytokines and several other drug classes (Aggarwal, 2008).

Table 1 – Commercialized monoclonal antibodies and their production platform.

CHO	*NS0*	*Sp2/0*	*Murine hybridomas*
Avastin, Camptah, Herceptin, Humira, Raptiva, Rituxan, Vectivbix, Xolair, Zevalin	Mylotarg, Soliris, Synagis, Tysabri, Zenapax	Erbitux, Remicade, Reopro, Simuleet	Bexaar, Orthoclone

2.1.2 Developing Production Cell Lines

The development of a production cell line begins with the introduction of the product coding sequence into the host cell line. This is supplied in a plasmid, which contains the required elements for replication and manipulation in *E. coli*. A selectable marker is also included in the plasmid to allow for selection of stably transfected cells. When expressing an antibody product, the heavy and light chain genes of IgG may be supplied in the same plasmid, or on separate expression vectors. This expression construct is introduced into the host cell using one of several transfection methodologies, all of which generate temporary destabilization of the membrane integrity, allow the DNA construct to enter the cell. Once inside the cytoplasm, the DNA construct must be transported inside the nucleus, where, to generate stable transfectants, it must be integrated into the host genome. This is a rare event, which usually occurs at a frequency of 1 in 10^4 cells (Mortensen and Kingston, 2009). As a consequence, cells must be grown in the presence of a selective agent in order to select for the subpopulation which has integrated the product gene. Selection relies on the principle that the selectable marker and the product gene will have co-integrated in the genome. Gene integration is a random event, and several copies of the expression construct may be integrated at different locations within the genome. The site of integration is known to have a profound effect on the resulting level of gene expression, a phenomenon known as the positional effect (Wurm and Jordan, 2003). There have been efforts to specifically target

gene integration to highly active regions in the genome, although these have not gained widespread use in industry (Chartrain and Chu, 2008).

The process of selection is generally combined with limited dilution cloning, in order to generate cell clones. Cloning is required to ensure that the resulting clonal population will be homogenous and that cultivation of the production clone will be consistent and reproducible. Furthermore, clonality is a requirement imposed by regulatory agencies. In limited dilution cloning, cells are diluted in micro-titer plates to ensure the presence of one or less cell per well. A proportion of wells will proliferate, and colonies will emerge. Supernatants from positive wells can be assayed for antibody titer, usually using ELISA (enzyme linked immunosorbent assay) and the top producing clones will be expanded for further characterization. These colonies are typically further subcloned in order to ensure clonality (i.e. that the population truly arose from a single cell). During industrial cell line development, it is typical to screen hundreds of clones, and this brute force approach is both time-consuming and labor-intensive. Consequently, significant opportunities exist within the cell line development process to improve efficacy and timelines.

It is also common to include a gene amplification step to increase the number of copies of the product gene and its transcript levels. This can facilitate the isolation of cell lines which secrete large amounts of recombinant protein. Several systems are commonly employed for gene amplification, including the CHO DHFR system and the NS0/CHO glutamine synthetase (GS) system. In these systems, the DHFR or GS gene is supplied alongside the product gene, and a selective agent is added which results in amplification of both the selectable marker and the gene of interest. Copy numbers on the order of several hundreds are typically observed for the DHFR system (Kaufman et al., 1985), while copy numbers between 4 and 10 have been reported for the GS system (Bebbington et al., 1992; Brown et al., 1992).

Once candidate clones have been expanded, they are evaluated for their ability to produce large quantities of quality antibody product. Quality testing typically involves ensuring correct glycoform profiles and amino acid sequence. Clones used for industrial production of recombinant protein need to be stable in terms of their growth, protein

production and quality of protein produced, over the life of the manufacturing process. The growth rate and morphological appearance of the cells should also be invariable with time. The level of protein production can be measured using assays like ELISA and HPLC, and needs to be demonstrated to be stable. Stability is generally tested for 50 to 100 generations. It is not uncommon to observe a decline in recombinant protein production over several generations, especially in cells where the product gene has been amplified. The molecular mechanisms for this instability are not completely understood. A decrease in transgene copy numbers has been reported (Kim et al., 1998a; Kim et al., 1998b; Fann et al., 2000), although in some cases, copy number was unchanged but transcriptional efficiency decreased with time (Kim et al., 1998b; Chusainow et al., 2009). Gene silencing is another mechanism which can lead to reduced productivity over time. Gene integration is a random event, and the site of integration could be susceptible to silencing. Mechanisms such as histone deacetylation, histone methylation and promoter methylation have been implicated in this phenomenon (Richards and Elgin, 2002; Mutskov and Felsenfeld, 2004).

Finally, the narrowed field of candidate clones must be evaluated in small-scale bioreactors models for growth performance under typical process conditions. Once a final production clone has been selected, process optimization work can proceed. A pictorial representation of the steps associated with cell line development and the typical increases in specific productivities observed are shown in Figure 2.

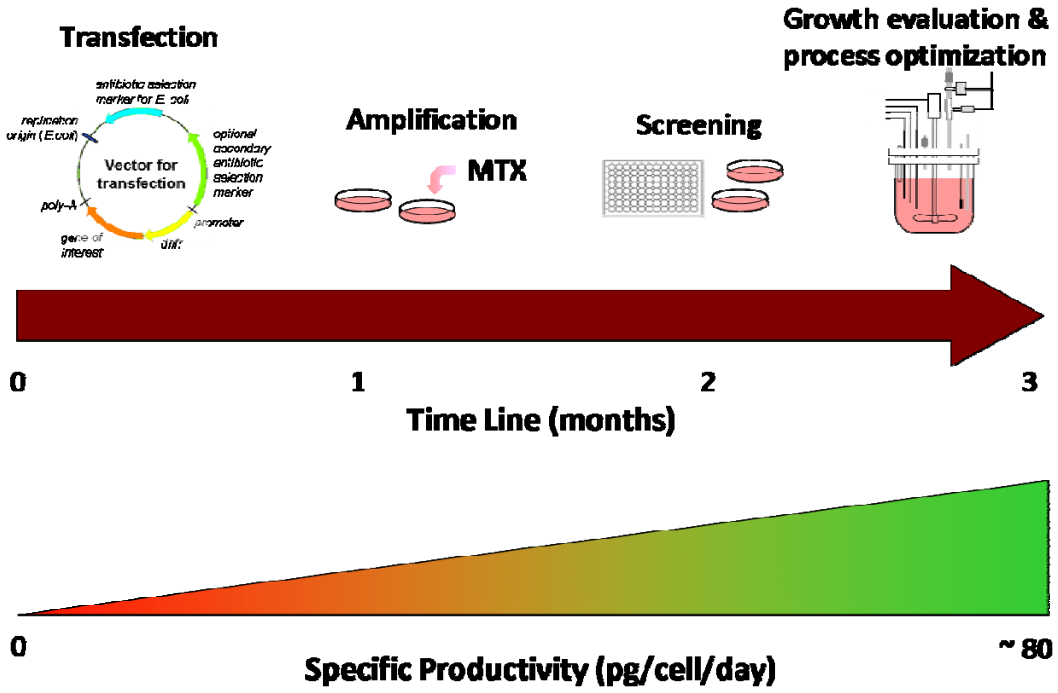


Figure 2 – Overview of cell line development.

2.1.3 Typical Culture Process for Recombinant Protein Production

Mammalian cells used for recombinant protein production are most often grown in bioreactors, where several process parameters are tightly controlled, including temperature, pH, O_2, CO_2 and agitation. Typical cell culture processes are operated in fed-batch mode, and cultures usually last on the order of two weeks without any byproduct removal. Concentrated nutrients, often including glucose and amino acids, are periodically added during the cultivation. This nutrient addition prolongs cell growth and allows cell viability to be maintained over an extended period of time. Eventually, the accumulation of waste byproducts, primarily lactate and ammonia, leads to a decline in cell growth and ultimately, cell death (Ozturk et al., 1992; Zhou et al., 1997). One of the primary challenges in designing effective cell culture processes is therefore to reduce metabolite accumulation, thereby increasing maximal cell densities, prolonging the culture period and consequently increasing achievable titers.

Many bioreactor processes employ a biphasic cultivation approach: during the first phase, conditions are optimized to allow for rapid cell growth and biomass accumulation; subsequently, conditions are optimized to maximize protein production.

Typical mammalian expression systems used for recombinant protein production do not employ inducible promoters, and therefore, optimal antibody production conditions are generally achieved through manipulation of process parameters. Commonly used approaches include modulation of culture temperature or pH, which minimize cell growth and shift cellular resources toward protein production. While these approaches are widely used, the biological mechanisms which make them effective are poorly understood. This represents an opportunity for greater understanding, which could potentially lead to a better optimization of process conditions.

2.2 MICROARRAYS FOR TRANSCRIPTOME ANALYSIS

The development of DNA microarray technology, first reported in 1995 (Schena et al., 1995), has enabled the global analysis of cellular behavior. Using this groundbreaking technology, the expression of thousands of transcripts could now be interrogated simultaneously in a single experiment. Microarrays make use of the highly selective base pairing of nucleic acids to detect the presence of individual target sequences in complex mixtures. Early DNA microarray technology emerged as a byproduct of sequencing projects, which generated cDNA libraries for expressed sequence tag (EST) sequencing (Adams et al., 1991). By amplifying each clone in a cDNA library using parallel PCR reactions, these amplified cDNA sequences could be spotted onto glass slides and used to interrogate transcript expression. One of the drawbacks of these cDNA microarrays is that they require hybridization of two samples onto the array (these are known as two-dye arrays). Two samples must be separately fluorescently labeled, and it is the ratio of fluorescence intensity that is subsequently compared. Consequently, elaborate experimental designs with multiple pair-wise combinations are required in order to compare expression across multiple samples. To circumvent this problem, one can use single-dye arrays, where a single sample is hybridized onto each array. This class of array is often also referred to as oligonucleotide arrays. In this implementation, short oligonucleotide sequences are designed and subsequently synthesized onto a solid surface.

One of the most widely used types of oligonucleotide array is the Affymetrix GeneChip® array. Affymetrix arrays are made by the direct synthesis of sets of 25-mer probes onto the chip using photolithography-based technology (Gramer and Goochee, 1993; Lockhart et al., 1996). The 25-mers are designed to span the 600 base pair region nearest to the 3' end of the sequence (Figure 3). In contrast to cDNA microarrays, a single sample is hybridized onto each array. Multiple samples can be compared by taking the ratio of their intensities for each set of data. The high level of chip-to-chip consistency using direct synthesis, as opposed to spotted DNA probes, allows a high level of consistency amongst technical replicate hybridizations.

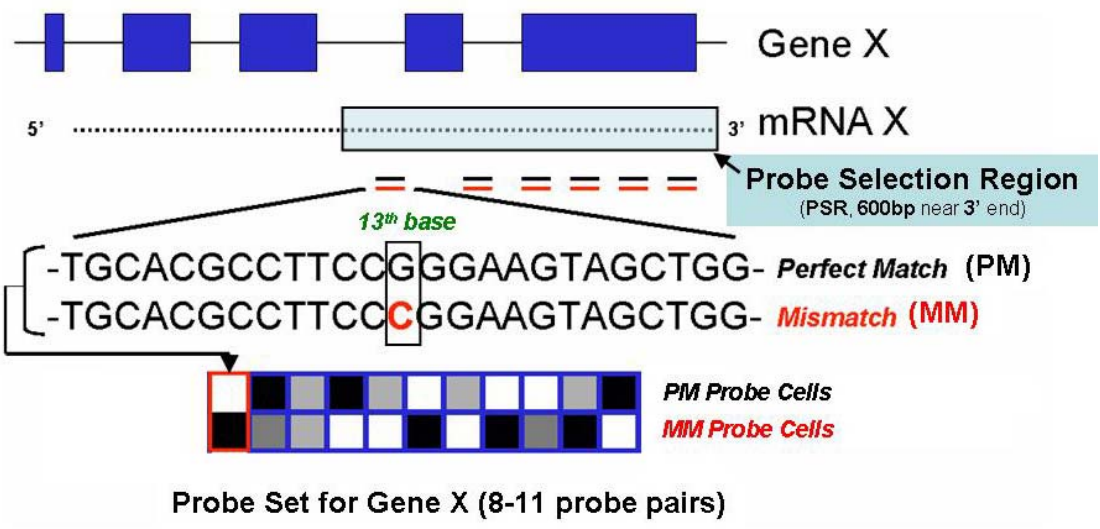


Figure 3 – Overview of Affymetrix microarray probe design.

An Affymetrix experiment begins with total or mRNA which is then reverse transcribed to cDNA. This cDNA is then transcribed using biotinylated ribonucleotide analogs to yield biotinylated cRNA. It is this cRNA which is hybridized onto the array. Following hybridization, the array is stained with a streptavidin-phycoerythrin conjugate, which allows fluorescent labeling of hybridized transcripts. Arrays are scanned at 570 nm and the amount of light emitted is proportional to the bound target at each location on the array. An overview of the Affymetrix experimental protocol is shown in Figure 4.

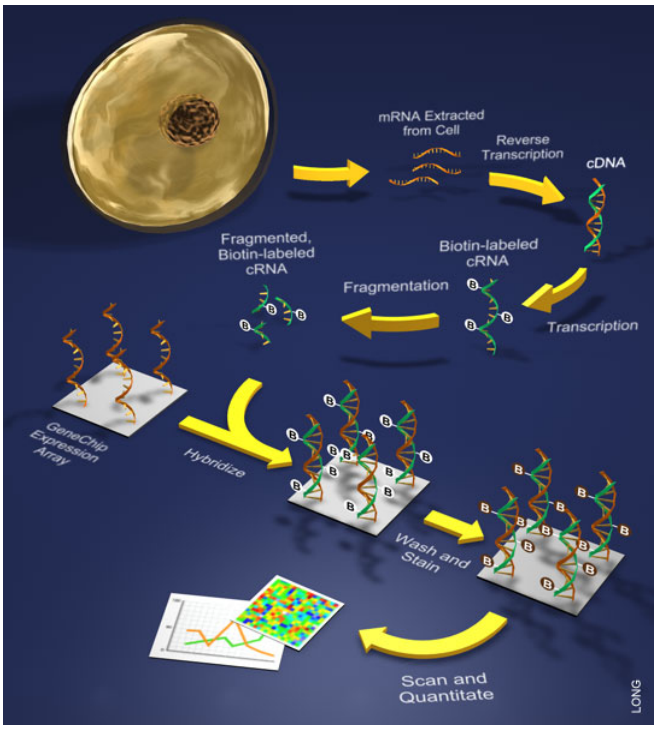


Figure 4 – Overview of experimental protocol for Affymetrix microarrays.

The analysis of microarray data hinges on the experimental design at hand. The statistical analytical tools employed to identify differentially expressed genes are different for static versus time series experimental designs. All of the studies presented here use time series experimental designs, therefore, the discussion will focus on this type of analysis. A general pipeline for time series transcriptome data analysis is shown in Figure 5. All microarray data must first be normalized in order to eliminate systematic variation which may be present in the data. Data may subsequently be pre-processed, depending on downstream analyses, and this could include log transformation and/or mean centering. In order to identify kinetically differentially expressed genes, two general approaches can be used: distance-based or regression-based. In the first approach, distance calculations are used to identify those genes which show different expression profiles across biological conditions. Several distance metrics can be employed, including Euclidean and Pearson distance, and each will yield different results. Euclidean distance should be used to identify genes with different expression levels across biological conditions, while Pearson distance should be used to identify genes with different expression trends across biological conditions. To define a threshold above

which genes are kinetically differentially expressed, the distance value distribution should be plotted and a value should be chosen which is at least one standard deviation above the mean.

A second approach to identify kinetically differentially expressed genes is to use regression-based calculations. In this methodology, the time-series expression profile of each gene is fitted with a curve, which could be first-order or higher. The fitted parameters are then compared for each gene across biological conditions, and genes with significantly different fitted parameter values are considered kinetically differentially expressed. Statistical testing is used to determine whether the difference is significant, therefore this approach necessitates biological replicates for each experimental condition, a requirement which can sometimes be cost-prohibitive. Several publicly-available algorithms which incorporate this methodology are available, including SAM (Statistical Analysis of Microarrays) (Tusher et al., 2001) and EDGE (Extraction and Analysis of Differential Gene Expression) (Leek et al., 2006).

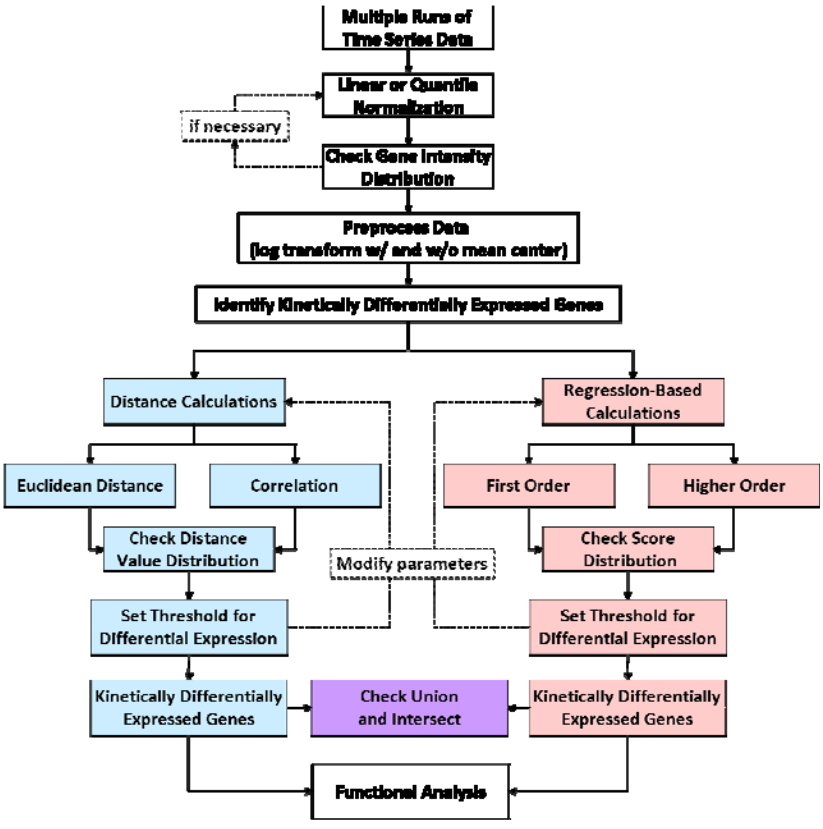


Figure 5 – General pipeline for analysis of time series microarray data.

In recent years, microarrays have gained popularity for cell culture applications, as reviewed in (Gupta and Lee, 2007). One example has been the transcriptional study of cell lines of varying productivity, in both CHO (Nissom et al., 2006; Doolan et al., 2008; Trummer et al., 2008) and NS0 (Khoo et al., 2007; Seth et al., 2007; Charaniya et al., 2008b). The impact of various culture parameters on the cellular transcriptome have also been studied, including reduced temperature cultivation (Baik et al., 2006; Swiderek and Al-Rubeai, 2007; Yee et al., 2009), sodium butyrate treatment (Yee et al., 2008a) and osmotic shock (Shen and Sharfstein, 2006). Furthermore, transcriptome analysis has been applied to the characterization of cell culture processes, including batch and fed-batch cultures (Wong et al., 2006; Krampe et al., 2008). These studies have generated cell engineering strategies which have resulted in increased productivity and enhanced growth characteristics. Additional potential applications include characterizing the transcriptional response of cell culture process systems to hydrolysate lot variation, as well as profiling the expression fingerprint of process scale-up.

2.2.1 Developing Genomic Tools for Chinese Hamster Ovary Cells

Very few genomic resources are currently available for the Chinese hamster, despite the pivotal role that CHO cells play in recombinant protein production. To this end, our laboratory began a collaboration more than 8 years ago with the Bioprocessing Technology Institute in Singapore to develop genomic tools for the Chinese hamster. Initial sequencing results indicated that the sequence identity of most genes isolated in CHO was substantially lower than the level reported to give rise to satisfactory cross species hybridization (e.g. chimpanzees and humans (Enard et al., 2002)). This later spurred the formation of the Consortium for Chinese Hamster Ovary Cell Genomics, under whose umbrella we continued our sequencing efforts in CHO (Jayapal et al., 2007a). This consortium resulted in the development of first and second generation CHO Affymetrix microarrays for use in transcriptional profiling.

Acquiring the DNA sequence repertoire necessary for the construction of DNA microarrays for CHO cells was the first aim of our efforts in developing genomic tools for CHO. Our work began with expressed sequence tag (EST) sequencing. This

methodology, originally developed in the late 1980s, was one of the cornerstones of early sequencing projects (Adams et al., 1991) and is the first step in characterizing an organism's transcriptome. Traditional EST sequencing projects begin with the construction of cDNA libraries that together, have a wide-ranging coverage of transcripts expressed by the organism under the physiological conditions of interest. The resulting libraries are collections of transcripts in the form of cDNAs, inserted into phage or *E. coli* vectors (Klickstein, 2001). These cDNAs are subsequently amplified for sequencing. This traditional way of EST sequencing was still the norm until recently, as Sanger technology remained the workhorse of DNA sequencing. One of the principal considerations of EST sequencing is to generate a wide diversity of transcript sources and representation.

It is likely that CHO cells under different culture conditions or cells which have undergone different selection conditions or treatments, may express different sets of genes. To increase the diversity of our sequence repertoire, we constructed cDNA libraries from a variety of Chinese hamster ovary cell sources (Table 2), including parental and recombinant clones of different lineages, as well as various cell lines under treatments known to incur profound transcriptional changes, including sodium butyrate, a known histone deacetylase inhibitor (Candido et al., 1978), tunicamycin, an inhibitor of N-linked glycosylation which triggers an unfolded protein response (Bush et al., 1994), and 5-azacytidine, a demethylation agent that induces transcription of silenced genes (Jones and Taylor, 1980).

CHO cells have been a laboratory cell line for more than 50 years since their original isolation (Puck et al., 1958). The population doublings they have since undergone may have generated extensive polymorphism amongst different sublines. Consequently, libraries were also prepared from a number of Chinese hamster tissues known to be transcriptionally active, including brain and spleen. The sequences derived from these tissues can serve as a common reference since the mutation rate in the animal is likely to be lower than in the aneuploid cell line which has been maintained nearly continuously in culture. Furthermore, we also used normalization, a process which aims

to eliminate redundant copies of transcripts within a cDNA library, for two of the constructed libraries.

Table 2 – cDNA libraries constructed for EST Sanger sequencing.

No.	*Cell or Tissue Source*	*Treatment*	*Normalization*
1	CHO DXB11 & CHO DXB11-IgG	---	---
2	CHO DXB11-IFNγ	---	---
3	CHO DG44-IgG	Sodium butyrate	---
4	CHO DG44-IgG	Tunicamycin	Yes
5	CHO DG44-IgG	5-azacytidine	Yes
6	Chinese hamster brain	---	---
7	Chinese hamster spleen	---	---

The purpose of constructing multiple cDNA libraries from diverse sources extends beyond the need to increase transcript representation. In any EST sequencing, more abundant transcripts are inevitably sequenced multiple times, while rare transcripts may not be reached at all. As the same library is sequenced in greater depth, the gene discovery rate quickly drops. By switching to new libraries after a decrease in gene discovery rate (which translates to an increased cost for every new gene identified), we were able to continue to discover new sequences. Figure 6 shows the number of unique sequences obtained as a function of the total number of ESTs sequenced from each cDNA library. The slope of each line at any given sequencing depth represents the gene discovery rate for each library. A decline in gene discovery rates for untreated, unnormalized libraries was observed when as little as 2000 ESTs had been sequenced. Transcriptionally active tissues, or libraries treated with chemical compounds known to induce transcriptional changes delayed this effect. Normalization resulted in a sustained gene discovery rate at the observed sequencing depths.

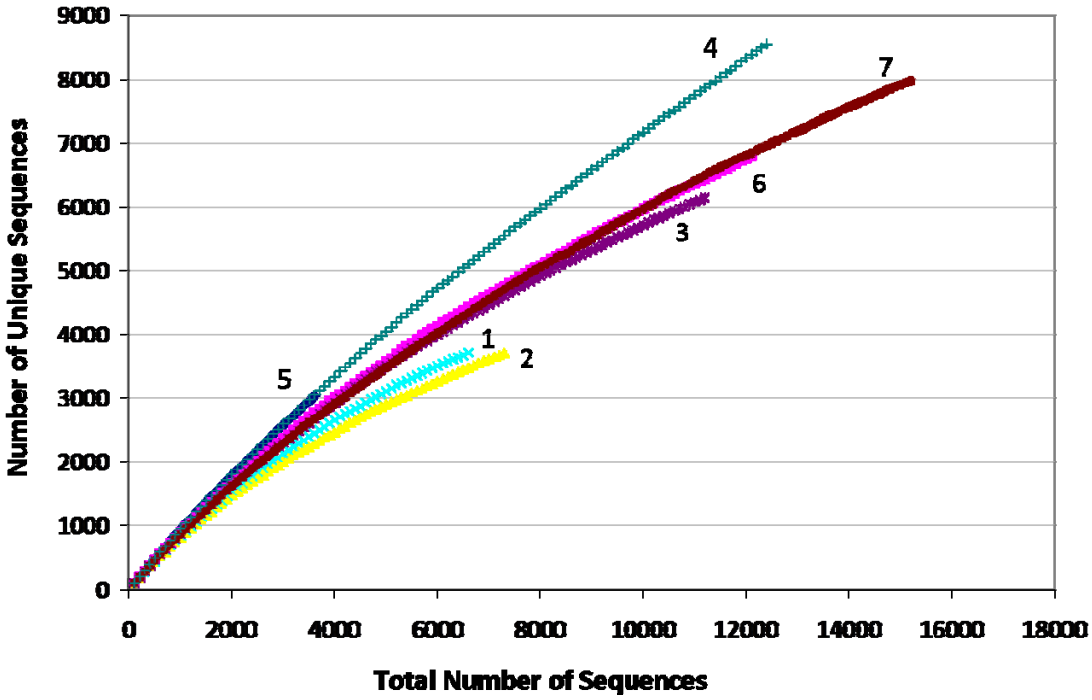


Figure 6 – Gene discovery in individual cDNA libraries.

2.2.1.1 Sequencing, Assembly and Annotation

In library-based sequencing efforts, a key step following sequence generation is to collapse the sequence reads that are derived from the same transcript or DNA segment together, in the order they are found in their native state. This process of assembling sequences converted our collection of more than 68,000 sequences obtained from Sanger technology into approximately 18,000 contigs and 10,000 singletons (this represents the set of sequences which occur only once in our repertoire, i.e. unique sequences). The reads generated by new sequencing technologies are substantially shorter than what can be obtained using Sanger technology. This makes the identification of overlapping regions required for the creation of contigs more difficult and consequently, the process of sequence assembly becomes an informatics hurdle. A number of short sequence assemblers have been specifically developed to handle these types of next-generation sequencing reads (Pop and Salzberg, 2008) and the methods will certainly continue to improve rapidly. At the rate at which sequencing costs are decreasing, there are likely to be more attempts to sequence CHO or other technologically important, yet genomic

resources poor organisms (such as baby hamster kidney (BHK) cells). However, the informatics challenges post-sequencing should not be underestimated.

The assignment of functional annotation to each sequence member in an assembly relies on the identification of orthologs in closely related genomes. We employed several public databases for the annotation of our CHO sequence repertoire, including a first-pass annotation using Ensembl Mouse (Hubbard et al., 2002) and FANTOM3. FANTOM3 represents a collection of more than 100,000 unique full-length cDNA transcripts obtained through extensive sequencing of multiple mouse cDNA libraries by the Riken Institute (Carninci et al., 2005). Transcripts which could not be assigned an annotation from either database were searched serially against Ensembl Rat, Ensembl Human and the GenBank nucleotide repository. Finally, any transcripts which could not be annotated via this pipeline were searched for the presence of open reading frames using a gene prediction algorithm (Crow and Retzel, 2005), and were further queried against public non-coding RNA (ncRNA) databases (Griffiths-Jones et al., 2008; He et al., 2008). Figure 7 shows the annotation assignment distribution for our most current set of Sanger-generated EST sequences. More than 45% of transcripts can be annotated using both public mouse databases. While these databases provide mostly overlapping information, they also each provide novel annotation, together enabling the annotation of an additional 8.2% of sequences. The FANTOM3 database annotation also yields additional information that is absent from Ensembl. As an example, FANTOM3 has identified more than 34,000 potential non-coding RNA sequences (ncRNAs) in its cDNA collection (Furuno et al., 2006). Our annotation identified 680 of these sequences within our CHO EST dataset.

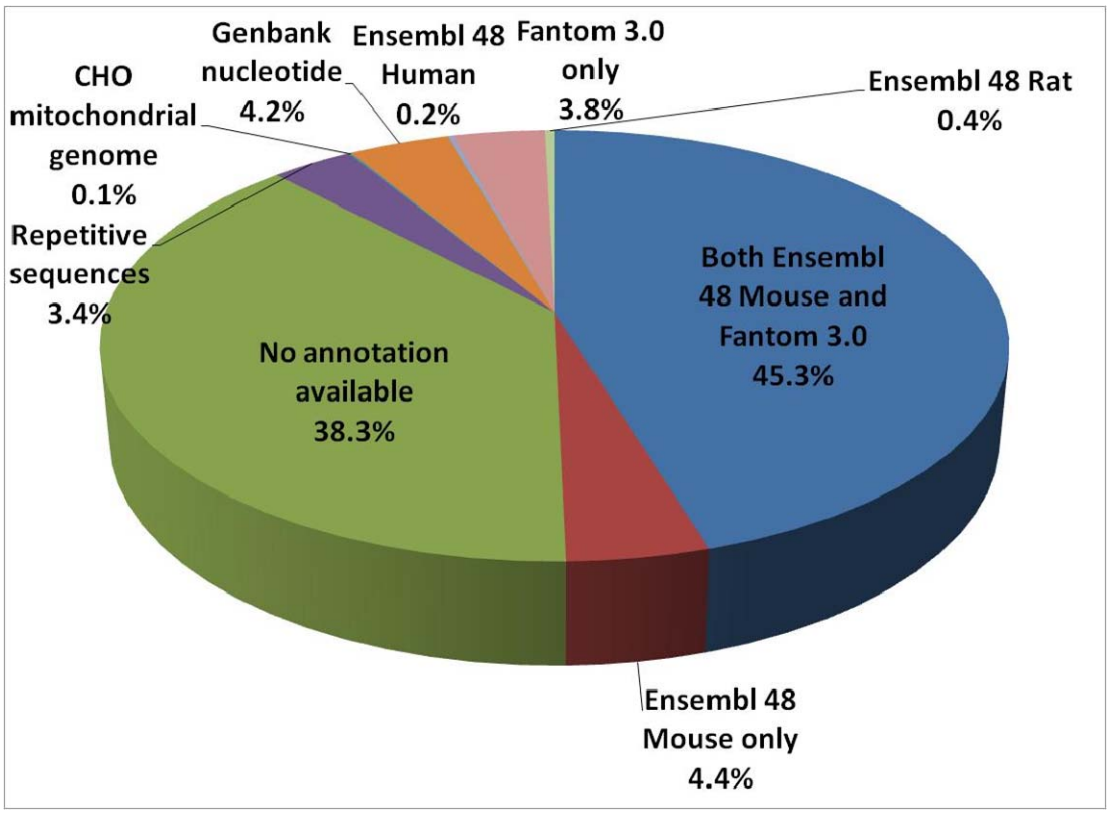


Figure 7 – Distribution of sequence annotation sources for unique CHO EST sequences.

We used gene ontology (GO) assignments (Gene Ontology Consortium [Online], 2008) to assess the degree of functional class coverage in our EST sequence repertoire. As can be seen in Table 3, our current coverage of a number of functional classes exceeds 60%. Of note, we have achieved significantly high coverage in a number of functional classes closely related to protein production, including protein secretion and vesicle-mediated transport. Conversely, coverage in a number of functional classes remains low, including cell signaling genes, genes whose protein products are destined for the extracellular space, and membrane proteins. Members of these functional classes are known to be difficult to capture in cDNA libraries. For example, the expression of cell signaling genes is highly tissue and cell-type specific. As a consequence, it becomes necessary to sample across a broader range of tissues and cell types in order to identify all gene members in this functional class.

Table 3 – Functional coverage of unique CHO EST sequences.

Gene Ontology Class	*No. Present in CHO*	*Total No. in GO*	*% Present*
Apoptosis	341	674	50.6%
Cell cycle	436	762	57.2%
Cell-cell signaling	135	538	25.1%
ER-to-Golgi vesicle-mediated transport	60	66	90.9%
Extracellular space	673	2052	32.8%
Lipid metabolic process	316	650	48.6%
mRNA processing	164	218	75.2%
Membrane	2288	6535	35.0%
Secretory pathway	149	229	65.1%
Vesicle-mediated transport	278	436	63.8%

Functional annotation of this CHO EST dataset also allowed us to evaluate the degree of sequence identity of the Chinese hamster to closely related species. Based on orthologous transcript annotations, we found that Chinese hamster coding sequences have an average sequence identity of 91% to both mouse and rat transcripts, and 88% to human transcripts (Figure 8). These results help solidify our annotation strategy which relies primarily on the mouse transcriptome for first-pass annotation.

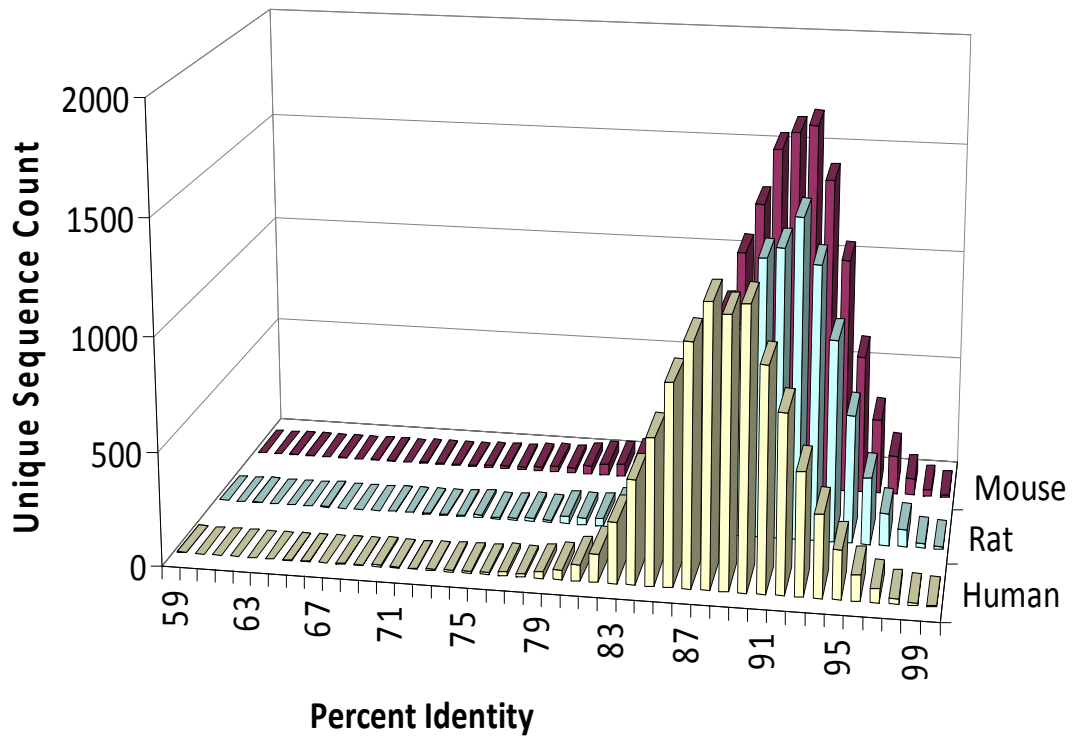


Figure 8 – Sequence identity distribution of CHO transcripts to orthologous mouse, human and rat sequences.

2.2.1.2 DNA Microarrays and Transcriptome Analysis

After an initial round of EST sequencing, we constructed cDNA microarrays using PCR-amplified EST clones (Wlaschin et al., 2005). Such cDNA microarrays are not optimized for sequence specificity or hybridization temperature. However, in the initial stages of the project, during which there was an insufficient amount of sequence available for probe design, they served as an invaluable investigation tool and allowed for an initial assessment of the feasibility of using mouse DNA arrays for Chinese hamster (De Leon Gatti et al., 2007). We consequently constructed synthetic oligonucleotide-based Affymetrix arrays and to date have released both a first and second generation array, whose salient features are shown in Table 4. Increasingly, other oligonucleotide-based arrays, such as those commercialized by Agilent and Nimblegen, are being used. These arrays rely on multiple oligonucleotide probes designed to minimize cross-hybridization. Consequently, this increases the power for interrogating and quantifying

the transcript levels of specific genes, and is often capable of differentiating closely related transcripts, or even alternatively spliced variants.

Table 4 – Summary of CHO Affymetrix microarrays.

Array version	*Total # of probesets*	*Probesets which can be annotated*	*Unannotatable probesets*
CHO version 1	10,072	9,001	1,071
CHO version 2	23,030	13,886	9,144

2.3 NEXT-GENERATION TECHNOLOGIES FOR TRANSCRIPTOME ANALYSIS

In the past few years, the field of sequencing has grown in leaps and bounds. The cost per base pair has dropped precipitously, decreasing by half approximately every 20 months, closely following Moore's law (Mardis, 2008). This is a consequence of tremendous advances in the field of sequencing, and the emergence of so-called next-generation sequencing technologies. Fueled by the quest for the $1000 genome, a number of new technologies have emerged, including those commercialized by 454, Illumina Solexa and ABI Solid. These technologies can now deliver tremendous sequencing depth at a fraction of the cost of traditional Sanger sequencing. This dramatic increase in output has been the result of their ability to massively parallelize the sequencing process. Furthermore, next-generation sequencing technologies have developed protocols which allow them to circumvent the traditional steps of creating libraries through vector-based cloning and amplification in *E. coli*. This eliminates cloning bias issues which are present when using traditional sequencing methodologies.

Given their ability to generate tens of gigabases in a single run, ultra-high throughput sequencing technologies are increasingly being explored as transcriptome profiling tools. In this approach, sequencing data can be used to provide an estimate of transcript abundance. Gene expression can be determined by recording the frequency at which each transcript is represented in the sequence data based on the number of sequencing reads which align to a particular transcript. Methodologies such as serial analysis of gene expression (SAGE) (Velculescu et al., 1995; Harbers and Carninci,

2005), and massively parallel signature sequencing (MPSS) (Brenner et al., 2000; Reinartz et al., 2002), were developed based on this principle. Both techniques create unique tags for each transcript in the cell. These are subsequently sequenced, and their abundance can be used to infer gene expression levels. However, these technologies were traditionally encumbered by often prohibitive sequencing costs and limited throughput (Wang et al., 2009). With the advent of ultra-high throughput sequencing methodologies, such as those commercialized by Illumina Solexa and ABI Solid, protocols have been developed which adapt these technologies for transcriptome analysis without the bias introduced by DNA arrays, a technique commonly referred to as RNA-Seq (reviewed in (Wang et al., 2009)). In its most direct implementation, cDNAs generated from a population of RNA are fragmented and sequenced. The resulting reads are then aligned to a reference genome or collection of transcripts to yield abundance levels. In the past year, several studies have been published which use RNA-Seq to investigate cellular transcriptomes in a number of organisms, including mouse (Mortazavi et al., 2008; Rosenkranz et al., 2008), human (Marioni et al., 2008; Mudge et al., 2008; Pan et al., 2008; Sultan et al., 2008; Wang et al., 2008), yeast (Nagalakshmi et al., 2008; Wilhelm et al., 2008) and *Arabidopsis thalinana* (Lister et al., 2008). These technologies represent an exciting opportunity to achieve unparalleled depth and coverage in transcriptome interrogation, and will no doubt render microarray technology all but obsolete within the next decade.

3 TRANSCRIPTOME ANALYSIS OF DHFR AMPLIFICATION

3.1 SUMMARY

This chapter describes our transcriptional study of gene amplification in Chinese hamster ovary (CHO) cells. This process can lead to dramatic increases in recombinant protein production, and is often used in the development of production cell lines for biologics. Indeed, the gene amplification process can transform a cell line of non-secretory origin to one which can rival the productivity of professional secretors. However, the physiological mechanisms of this process remain largely unknown, and a better understanding can provide opportunities for systematic improvements in cell line development. To this end, we used transcriptome analysis to study the amplification process in CHO cells. A parental CHO cell line deficient in dihydrofolate reductase (DHFR) activity was transfected with a vector expressing a DHFR gene and an antibody product gene. Methotrexate treatment was applied following transfection and initial selection. Samples were taken throughout the amplification process for transcriptome analysis. As a control, the same cell line was transfected with another vector carrying only the DHFR gene, thus providing a baseline by which to assess the importance of the expressed product in the transcriptional response. Clones isolated following transfection showed significant variation in their expression profiles, and also displayed varying degrees of gene amplification, as measured by mRNA levels of DHFR and light chain IgG. Methotrexate treatment elicited few changes at the transcriptome level, with little consistency observed amongst different clones. However, a comparison of amplified clones and their control counterpart revealed an upregulation of stress response genes and a downregulation of tRNA synthetases. This study provides physiological insights into the mechanisms involved in gene amplification, and can allow the design of a screening strategy for the selection of viable production cell line candidates.

3.2 INTRODUCTION

In the past few decades, cultured mammalian cells have become the workhorse of recombinant protein production. The titers achieved in these cells have increased more than 100-fold over the past twenty years, both as a result of improvements in media composition and process culture conditions, as well as increased efficiency in screening potential producing clones (Wurm, 2004). As demand for these therapeutic drugs continues to increase, there is an ever-increasing drive to shorten the timelines required to establish a production cell line and to increase achievable yields.

One commonly used mechanism to generate high levels of recombinant protein expression is the use of gene amplification systems, which co-amplify a selectable marker alongside the product gene. One such system is based on the dihydrofolate reductase (DHFR) gene, whose chemical antagonist, methotrexate (MTX) can be used for gene amplification. Dihydrofolate reductase is an enzyme which catalyzes the conversion of dihydrofolic acid to tetrahydrofolic acid, a compound required for the biosynthesis of glycine, thymidine monophosphate and purine. Chinese hamster ovary cells deficient in DHFR activity were derived using mutagenesis (Urlaub and Chasin, 1980; Urlaub et al., 1983) and the resulting cell lines require supplementation of hypoxanthine and thymidine (HT) to the culture media for growth. With the introduction of an exogenous DHFR gene, cells can be selected for in HT-minus media. Furthermore, gene amplification can be achieved through addition of the folate derivative, methotrexate (MTX), which inhibits DHFR activity. MTX treatment results in amplification of the DHFR gene, and between 10 – 10,000 kilobases of surrounding DNA (Stark et al., 1989; Coquelle et al., 1997). Therefore, by introducing a gene of interest (e.g. antibody product gene) alongside the DHFR gene, co-amplification of the product gene can be achieved. Using stepwise increases in MTX concentrations, gene copy numbers on the order of thousands have been reported (Wurm et al., 1986). MTX treatment has also been reported to lead to phenomena other than gene amplification, including altered MTX transport properties (Sirotnak et al., 1981) or the development of an MTX-resistant DHFR enzyme (Flintoff and Essani, 1980).

A number of molecular mechanisms have been proposed to explain the process of gene amplification (reviewed in (Wurm and Jordan, 2003)) and it is likely that more than one mechanism may be employed by the cell. However, very little information is currently available regarding the transcriptional changes incurred by the gene amplification process. This is particularly relevant when the gene being amplified is a secreted protein. Through the process of gene amplification, a cell line of non-secretory origin (i.e. ovary-derived) can transform into a highly secretory cell whose productivity rivals that of professional secretors *in vivo* (e.g. plasma cells). Such increases in recombinant protein production are likely to require an adaptation by the cell to cope with increased demands on the secretory machinery. DNA microarrays are well-suited for the study of such global transcriptional changes, and are currently gaining popularity in cell culture applications (reviewed in (Gupta and Lee, 2007).

In this study, we generated a series of monoclonal antibody (mAb)-producing clones via transfection with both chains of the immunoglobulin G (IgG) gene, as well as the DHFR gene. Using DNA microarrays, we tracked the transcriptional changes associated with MTX-mediated gene amplification and compared them to the changes incurred in a control population, expressing solely the DHFR gene. We also examined the transcriptional changes which occurred following transfection and selection. We observed a significant degree of variation in expression profiles amongst mAb-producing clones, while minimal transcriptional changes were associated with gene amplification. This report represents the first example of the application of transcriptome analysis to the temporal study of gene amplification.

3.3 MATERIALS AND METHODS

All transfections, amplifications, sample preparations and array hybridizations were performed by our industrial collaborators, and raw microarray data files were provided for our analysis.

3.3.1 Plasmid and Transfection

Chinese Hamster Ovary (CHO) DXB-11 cells were grown in MEMα medium containing 10% serum as well as nucleotides and nucleosides, in a 37°C incubator with 7.5% CO_2. Cells at ~ 80% confluency in T-75 flasks were transfected using the Lipofectamine kit (Invitrogen, Carlsbad, CA) with 8 μg of an expression vector containing cDNAs for a mammalian selectable marker (hygromycin), an amplification marker (DHFR) and the genes of interest (IgG heavy and light chains). Two days later, the transfected cells were subcloned into 96-well plates in the selection medium (i.e., growth medium without nucleotides and nucleosides (HT-minus), supplemented with 400 μg/mL hygromycin B). Three weeks after transfection, ELISA was performed on the supernatant samples from the wells containing single colonies to measure antibody concentration. The five most stable and productive antibody-producing clones, namely CHO-1, CHO-2, CHO-3, CHO-4, and CHO-5, were selected for gene amplification studies.

A control transfection was performed using the same protocol as described above, with the exception that the expression vector used only contained cDNAs for the hygromycin resistance gene and DHFR, but no antibody cDNAs; cells were propagated in T-75 flasks as a pool (instead of subcloning) in the selection medium after day 5.

3.3.2 Gene Amplification, Sample Preparation and Array Hybridization

Transfected cells were propagated in the selection medium in T-75 flasks to approximately 40% confluency, after which time they were incubated in the selection medium with 20 nM of methotrexate (Sigma-Aldrich, St. Louis, MO) for 0, 10, or 15 days. Total RNA of treated cells was extracted using TRIzol (Invitrogen, Carlsbad, CA) and the Qiagen RNeasy mini kit protocol. Total RNA (5 μg) from each sample was used to generate first-strand cDNA using a T7-Oligo (dT) primer. Following second-strand synthesis, in vitro transcription was performed with the BioArray, High Yield RNA Transcription Labeling Kit (Enzo Life Sciences, Inc) and 10 μg of fragmented cRNA were used for hybridization to custom Affymetrix microarray with 10,118 probe sets.

Microarrays were washed using the Affymetrix GeneChip Fluidics Station 450 and scanned using the Affymetrix GeneChip Scanner 3000.

3.3.3 Microarray Data Processing and Analysis

Array image files (.CEL) files were processed using the Genedata Expressionist software (Basel, Switzerland). Briefly, image files were diagnosed, corrected for any image defects and subsequently condensed using the Affymetrix MAS5 algorithm and linearly scaled to an average intensity of 500. Probesets with a detection p-value < 0.04 and intensity > 70 in at least one sample were retained for analysis.

Hierarchical clustering was performed using Spotfire DecisionSite 9.2 (Somerville, MA) on $\log_2$-mean centered intensity data. Data was filtered to exclude genes with non-dynamic behavior by applying a coefficient of variation threshold of ≥ 0.2. The UPGMA (unweighted average) clustering method was used, and correlation was chosen as the similarity measure.

When comparing $\log_2$ intensity ratios in the control population and IgG-producing clones, a z-score was calculated to provide statistical significance to the results. The z-score is defined as:

$$z = \frac{x - \mu}{\sigma}$$

where x is the $\log_2$ ratio of Day 0 to Host intensity values in the control population

μ is the average of $\log_2$ ratios of Day 0 to Host intensity values in the mAb-producing clones

σ is the standard deviation of $\log_2$ ratios of Day 0 to Host intensity values in the mAb-producing clones

Genes with an absolute z-score ≥ 2 were considered significant.

3.3.4 Functional Analysis

GenMAPP's MAPPfinder was used to integrate expression data with known biological pathways and determine confidence levels for differential expression within ontology groups (Doniger et al., 2003). Using this method, the significance of a gene class is calculated by an enrichment score, based on a null hypothesis described by a hypergeometric distribution. The enrichment score is defined as:

$$z = \frac{r - n\frac{R}{N}}{\sqrt{n\left(\frac{R}{N}\right)\left(1-\frac{R}{N}\right)\left(1-\frac{n-1}{N-1}\right)}}$$

where n is the number of genes in a gene class S, r is the number of differentially expressed genes in gene class S, N is the number of genes measured in the dataset and R is the number of differentially expressed genes in the dataset. Gene ontology definition files dated 01/09 were used for the analysis. Gene ontology terms with a significance p-value ≤ 0.05 were considered as enriched.

3.4 RESULTS

3.4.1 Experimental Design and Gene Amplification

A panel of five mAb-producing clones was generated by transfection with an antibody expression vector also containing the DHFR gene and was subjected to MTX treatment for amplification (Figure 9). Samples for microarray analysis were taken immediately prior to treatment, as well as 10 and 15 days following treatment. A control population, transfected with the same vector except without IgG genes, was generated in parallel from the same host cell line. Samples from the control population were taken for transcriptome analysis immediately before treatment, and 15 days following treatment. Triplicate samples of the host cell line were taken for microarray analysis and were also analyzed.

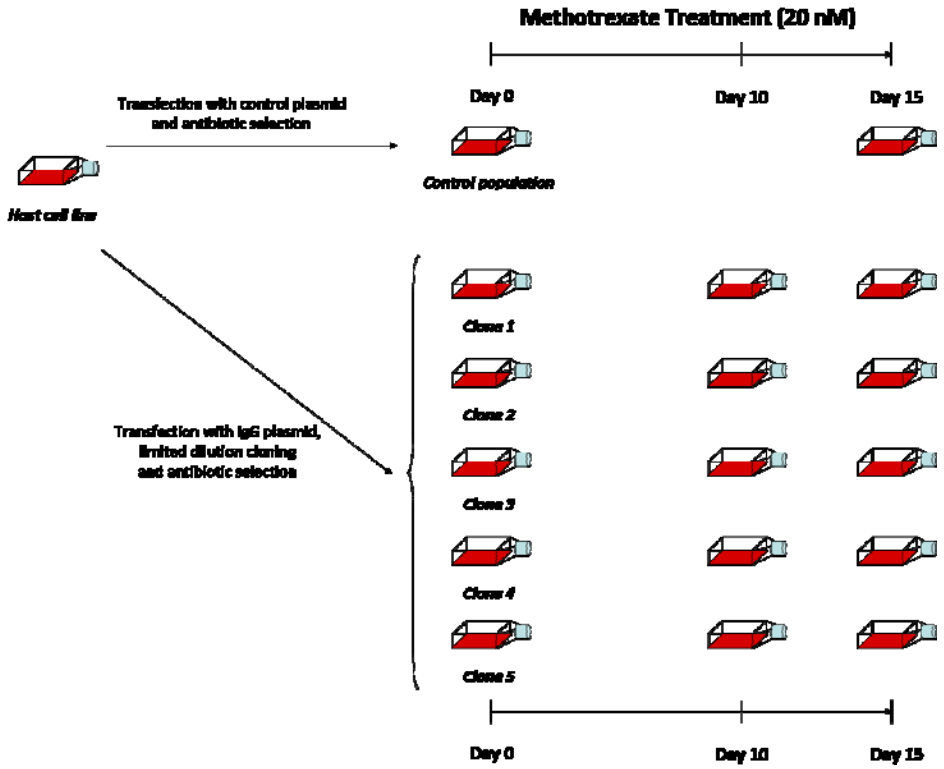


Figure 9 – Outline of experimental design. A DXB11-derived host cell line was transfected with a control plasmid encoding DHFR and an antibiotic resistance gene. Antibiotic selection was used to isolate a pool of stable transfectants. This pool was subjected to 15 days of 20 nM methotrexate treatment. In parallel, IgG-producing clones were generated via transfection of the same host cell line with a plasmid encoding DHFR, both chains of the IgG gene and an antibiotic resistance gene. Limited dilution cloning in selection media was used to isolate stable transfectants. The top 5 producing clones, as determined by ELISA, were selected for MTX treatment. Each clone was subjected to the same 15 days of 20 nM methotrexate treatment. In this figure, flasks represent samples which were taken for microarray analysis.

All five mAb-producing clones displayed increased expression of the DHFR and IgG light chain transcripts following the 15-day methotrexate treatment (Figure 10). The control population showed only a modest increase in DHFR expression as compared to the five IgG producing clones. This may be attributed to the heterogeneity of the control population which represents an amalgam of transfectants with multiple integration sites, each with their own expression level of the transgene.

The expression levels of the DHFR and IgG light chain genes were found to be different in each clone following selection. Furthermore, initial expression levels of either gene could not be used to predict amplification. For example, Clone 2 had the lowest IgG expression level following selection, yet had the highest IgG expression level

after MTX treatment (Figure 10). In addition, increases in DHFR gene expression were not directly correlated with increases in IgG expression. For instance, Clone 3 showed the highest increase in DHFR expression with methotrexate treatment (3.5-fold change), yet had the second lowest increase in IgG expression (3.8-fold change, see Figure 10).

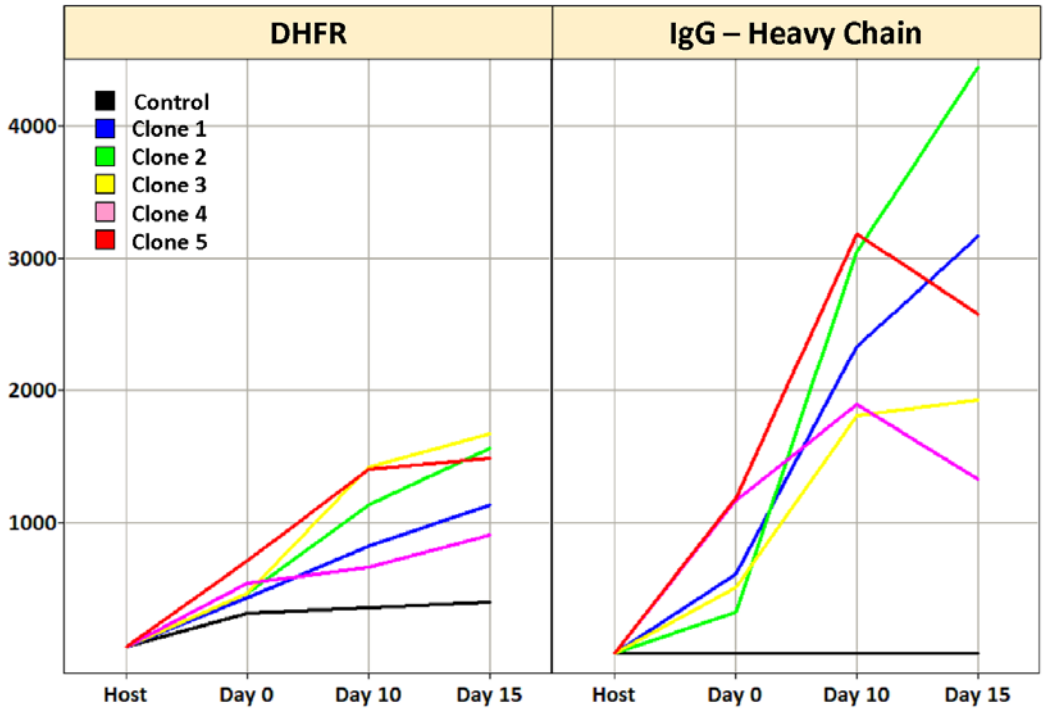


Figure 10 – Expression profiles of DHFR and heavy chain of IgG in host cell, control population and all IgG-producing clones throughout methotrexate treatment.

3.4.2 Global Analysis of Transcriptional Response

Unsupervised clustering was used to assess the transcriptional similarity present amongst all analyzed samples. The results of this hierarchical clustering are shown in Figure 11. Interestingly, a greater degree of similarity was found amongst individual clones, rather than with time following methotrexate treatment (see column dendograms in Figure 11). This suggests that each clone had its own unique response to the MTX treatment. A notable outlier appears to be the Day 0 sample of Clone 4, which possesses a transcriptional profile distinct from all other samples examined. These results also distinguish the mAb-producing clones from the control population, as well as the parental host cell line. This suggests that the expression of a secreted antibody product leads to distinct expression signatures, different from cells expressing solely DHFR or no product

at all. Furthermore, it is interesting to note that the control population, either before or after methotrexate treatment, shows a greater degree of transcriptional similarity to the host cell line than any of the mAb-producing clones.

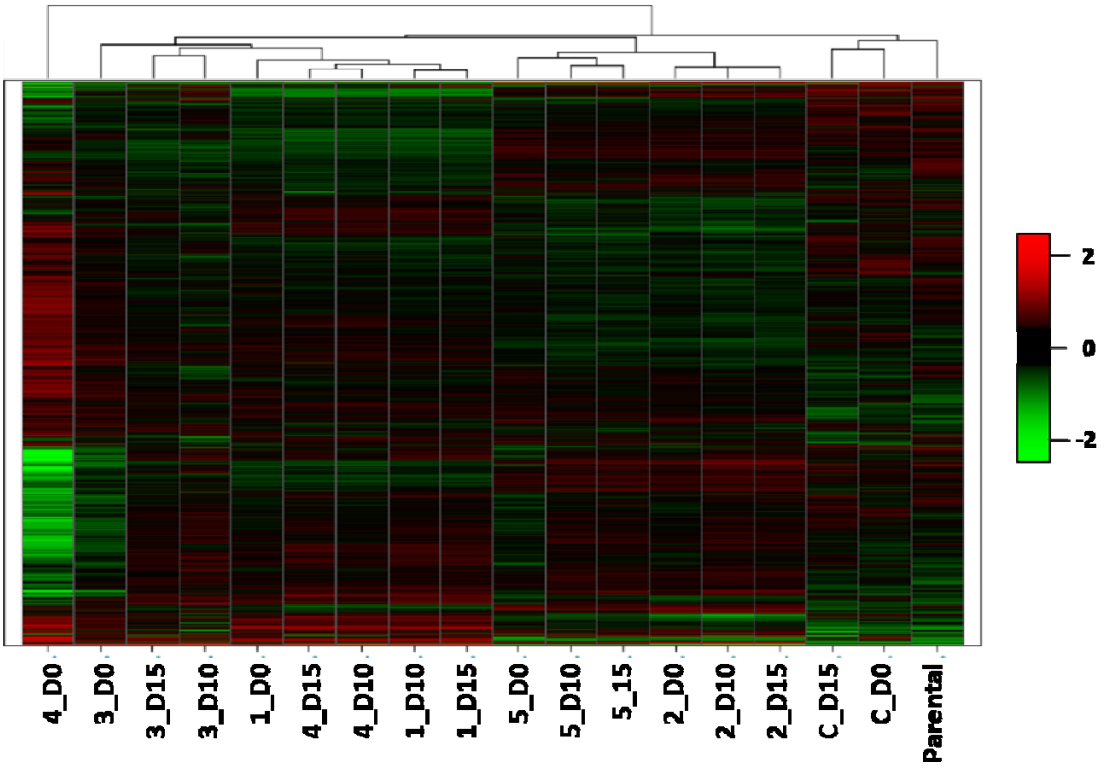


Figure 11 – Hierarchical clustering of all samples. Log_2-mean centered intensity data for all samples was clustered using a hierarchical clustering algorithm. Input data was restricted based on the following criteria: maximum intensity ≥ 70, minimum detection p-value ≤ 0.04, coefficient of variation ≥ 0.2 (see text for details). Sample nomenclature: Clone number_Sample Day. Samples taken from the host cell line prior to transfection are marked as "parental." The heat bar indicates log_2-mean centered intensity values.

3.4.3 Comparison to Host Cell Line

We characterized the gene expression changes following selection in the control population and IgG clones by comparing log_2 transformed ratios of Day 0 samples (following selection) to the host cell line (Figure 12). Only a small number of genes were differentially expressed by at least 2-fold: 428 genes were differentially expressed in the IgG-producing clones as compared to the host cell line (both up and downregulated), while 420 genes were differentially expressed in the control population (both up and downregulated). Amongst these, 67 genes were commonly upregulated in both the

control population and IgG-producing clones (Region I), while 58 genes were commonly downregulated (Region I'). The most prominently upregulated gene was secretory leukocyte peptidase inhibitor, which increased 141-fold in the control population and an average of 18-fold in IgG-producing clones. This serine protease inhibitor is known to protect tissues by inhibiting the action of many proteases, including elastase, cathepsin G and trypsin (Doumas et al., 2005). The most prevalent gene ontology terms identified amongst the 67 commonly upregulated genes were cell cycle, DNA repair and transcription. The most downregulated gene was solute carrier family 44, member 1, a choline transporter. This gene decreased by 3.9-fold in the control population and an average of 6.3-fold in IgG-producing clones. This transporter allows the transport of choline, a positively charged molecule which cannot freely diffuse into the cell, across the plasma membrane (Yuan et al., 2006). Protein modification process and cell cycle regulation were identified as the most prominent gene ontology terms amongst the commonly downregulated genes.

Genes changing solely in the IgG-producing clones by at least 2-fold (Figure 12 – Regions II and II') were examined. An additional differential expression criterion was imposed: a z-score of at least 2. This z-score criterion excludes genes whose $\log_2$ ratio in the control population is less than 2 standard deviations away from the average $\log_2$ ratio in IgG-producing clones. A total of 62 genes were upregulated solely in IgG-producing clones upon selection ($|\text{z-score}| \geq 2$), while 20 genes were downregulated only in IgG-producing clones ($|\text{z-score}| \geq 2$). As expected, the IgG light chain transcript was the most upregulated transcript, a consequence of its absence in the host cell line. Another gene which increased dramatically following selection was chemokine (C-C motif) ligand 2, with an average fold change of 7.5 amongst mAb-producing clones and an average intensity in the host cell line of 344. Of the 20 genes which were downregulated solely in mAb-producing clones, asparaginyl-tRNA synthetase was the most downregulated, with an average fold change of -2.8.

A small number of genes identified as differentially expressed showed opposing trends between the control and mAb-producing clones (Figure 12 – Regions III and III'). Among these eight genes, CD36 antigen was upregulated by more than 2.7-fold in the

control population, while being downregulated by an average of 6.2-fold in the IgG-producing clones. CD36 is a cell surface receptor, known to bind many ligands, including collagen, lipoproteins and long-chain fatty acids (Su and Abumrad, 2009). Increased insulin levels lead to an increased concentration of CD36 at the plasma membrane, resulting in enhanced fatty acid uptake (Schwenk et al., 2008).

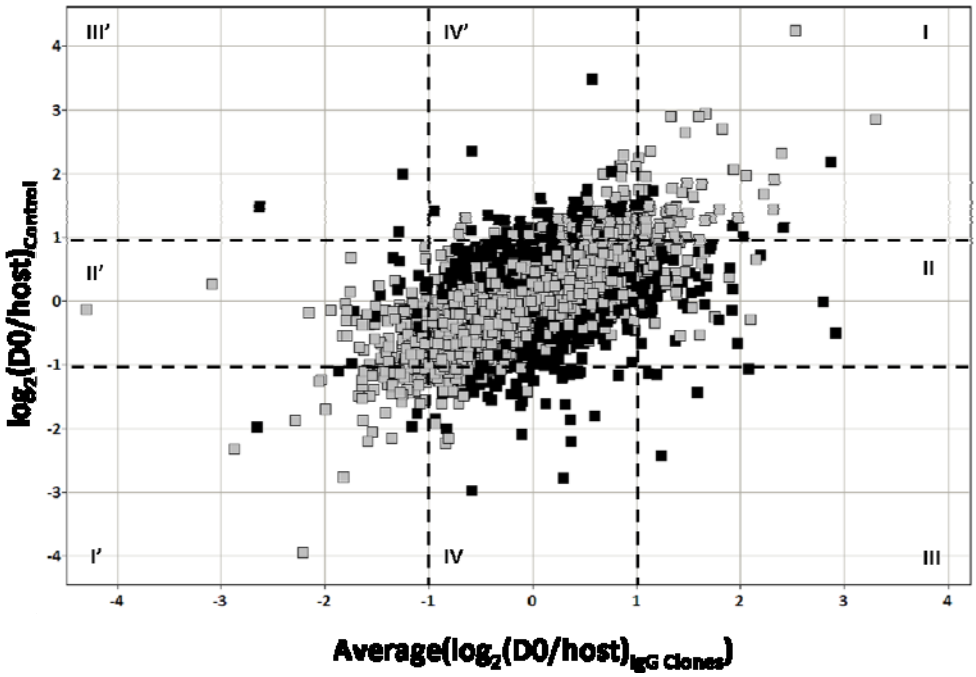


Figure 12 – Comparison of transcriptional changes in IgG-producing clones and control population at Day 0 with respect to the host cell line. On the x-axis, an average of $\log_2$ ratios in all five IgG-producing clones is shown. Dashed lines delineate regions of 2-fold change. Solid black squares indicate genes with control z-score ≥ 2, while grey squares indicate genes with control z-score < 2. Control z-score is defined in the materials and methods.

3.4.4 Transcriptional Response to Amplification

To further characterize the transcriptional dynamics in response to MTX-mediated amplification, we compared the $\log_2$ ratios of Day 15 (following methotrexate treatment) expression values to Day 0 expression values in both the IgG-producing clones and the control population (Figure 13). Only a small number of genes met the differentially expression criterion of 2-fold change or greater. In the control population, 194 genes were differentially expressed by at least 2-fold change upon MTX treatment, with 67 genes being upregulated and 127 genes being downregulated. Amongst the differentially expressed genes, cell cycle and tRNA aminoacylation were the most

prominent gene ontology terms identified. In the mAb-producing clones, only 65 genes were differentially expressed (average absolute fold change in all IgG clones greater than 2-fold) when comparing Day 0 and Day 15 (following methotrexate treatment). 62 genes were upregulated upon MTX treatment while only 3 were downregulated. The most commonly identified gene ontology terms amongst this set of genes were helicase activity, chromatin and stress response. Only 5 genes were commonly upregulated while none were commonly downregulated between the control population and IgG clones (Figure 13 – Regions I and I'). Of the five commonly upregulated genes, 3 are annotated: alpha thalassemia mental retardation syndrome X-linked homolog (Atrx), chromosome helicase DNA binding protein 6 (Chd6) and zinc finger protein 53 (Zfp 53). Both Atrx and Chd6 have helicase activity and all have DNA binding activity. The identification of helicase activity and chromatin genes is not unexpected, as MTX treatment leads to significant genome rearrangements.

No genes were found to be consistently changing more than 2-fold upon MTX treatment in all 5 clones. Reducing the fold-change criterion to 1.4-fold, 15 genes are changing (either up or downregulated) in all 5 clones upon MTX treatment. This supports the notion that each clone responds differently to gene amplification, and that each clone's transcriptional response may in part be dictated by its transcriptional signature at Day 0. Those commonly upregulated genes with an annotation are shown in Table VIII. Also noteworthy is the fact that expression of the light chain of IgG was not increased by more than 2-fold in all mAb-producing clones. As can be seen in Figure 10, IgG expression in Clone 4 and 5 decreased at Day 15, after reaching a peak at Day 10, accounting for this observation.

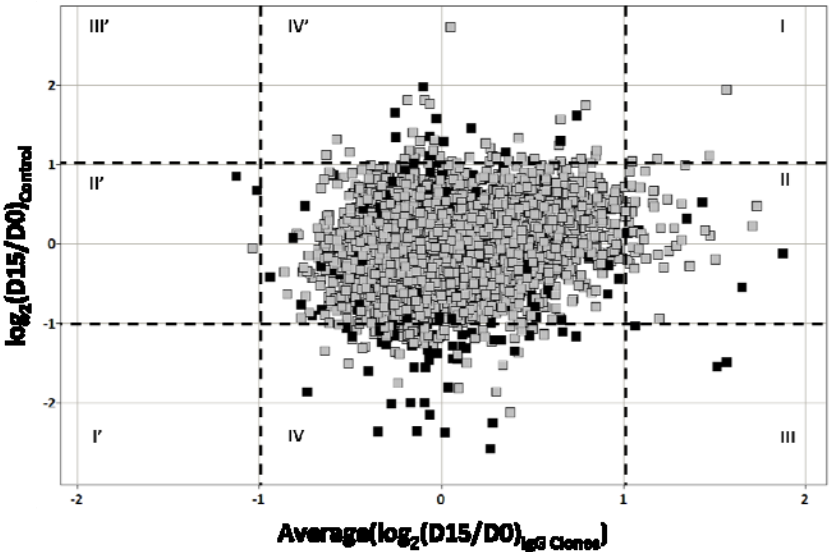


Figure 13 – Comparing the transcriptional response of the control population and IgG-expressing clones to methotrexate treatment. On the x-axis, an average of $\log_2$ ratios between Day 15 and Day 0 samples in all five IgG-producing clones is shown. Dashed lines delineate regions of 2-fold change. Solid black squares indicate genes with a Day 15 fold change between IgG clones and control population ≥ 2, while grey squares indicate genes with a Day 15 fold change between IgG clones and control population < 2.

3.4.5 Comparison between Control Population and IgG-Producing Clones

The observation that few genes were consistently changing upon MTX treatment in all mAb-producing clones led us to compare expression profiles at Day 15 between the control population and mAb-producing clones. Using this approach, we can attempt to identify the transcriptional differences between the amplified clones and their control counterpart. We imposed a differential expression criterion of $\left|\frac{average\ (IgG\_Day\ 15)}{average\ (Control\_Day\ 15)}\right| \geq 2$. This identified 210 genes as upregulated at Day 15 in mAb-producing clones, and 83 genes as downregulated at Day 15. These genes are shown as solid black squares in Figure 13 and their expression levels are summarized in Tables I – VII in the appendix, according to ontological classes. Gene class testing was used to identify those ontological classes significantly enriched (p-value ≤ 0.05) amongst differentially expressed genes, and are discussed below.

3.4.5.1 Regulation of Cell Cycle

A total of 15 genes involved in cell cycle regulation were differentially expressed by at least 2-fold at Day 15 between the control population and mAb-producing clones. Cyclin D1, a gene required for the transition from G1 to S phase of the cell cycle (Baldin et al., 1993), was upregulated in all mAb-producing clones by an average fold change of 3.5. DNA-damage inducible transcript 3, also known as CHOP, was upregulated in the control population by 2.95. This gene is a component of the ER stress-mediated apoptotic pathway, and is a downstream target of all three branches of the unfolded protein response. Its upregulation can lead to either cell cycle arrest or apoptosis (Oyadomari and Mori, 2004). Also upregulated in the control population (by 2.5-fold) was Gadd45a (growth arrest and DNA-damage-inducible 45 alpha). This gene inhibits Cdc2 kinase, thus blocking progression from the G2 phase to the M phase of the cell cycle. It is also a downstream target of p53 (Smith et al., 1994).

3.4.5.2 Response to Stress

Twenty genes associated with a stress response were identified as differentially expressed between the control population and IgG-producing clones. Fourteen of these were upregulated in the IgG-producing clones. The most upregulated gene was fatty acid binding protein 4 (Fabp4), which displayed on/off behavior between control and clones. This gene is a member of the fatty acid binding protein family, cytoplasmic proteins which bind fatty acids, eicosanoids and other lipids (Makowski and Hotamisligil, 2005). It has been suggested that these proteins play a role in protecting cellular membrane integrity by binding excess free fatty acids (Coe and Bernlohr, 1998). Also strongly upregulated in the IgG-producing clones was retinoic acid receptor responder 2 (Rarres2), whose expression was more than 18-fold greater than in the control population. This gene, also known as chemerin, is a soluble protein hormone whose expression is stimulated during adipogenesis (MacDougald and Burant, 2007).

3.4.5.3 Protease Inhibitors

Six genes with protease inhibitor properties were upregulated in mAb-producing clones at Day 15. Three of the six protease inhibitors had an intensity less than 100 (30th

percentile) in the control population, making them either absent or low intensity genes. One of the most upregulated gene was Timp2 (tissue inhibitor of metalloproteinase 2), which had an average intensity of 890 in the IgG-producing clones. This gene is a member of the tissue inhibitor of metalloproteinases family, who are inhibitors of matrix metalloproteinases (MMPs). MMPs are a family of proteins involved in the degradation of extracellular matrix components, and increased expression of Timp2 has been observed to prevent tumor metastasis (Stetler-Stevenson and Seo, 2005).

3.4.5.4 Rho Guanine Nucleotide-Exchange Factor Activity

Three members of this functional class were identified amongst differentially expressed genes and all were upregulated in the mAb-producing clones. Guanine nucleotide-exchange factors (GEFs) activate Rho-family GTPases, members of the Ras superfamily of small GTPases. These proteins exist in two distinct conformational states: an inactive form, bound to GDP, and an active, GTP-bound form which then mediates signal transduction mechanisms involved in a multitude of cellular processes, including adhesion, phagocytosis, cell growth and survival (Rossman et al., 2005). Guanine nucleotide-exchange factors catalyze the exchange of GDP for GTP, thus promoting Rho activity. The most upregulated GEFs gene was Farp1 (FERM, RhoGEF and plackstrin domain protein 1), which showed a 3.6-fold upregulation in IgG-producing clones. This gene has been shown to be necessary and sufficient for the dendritic growth of lateral motor column neurons (Zhuang et al., 2009).

3.4.5.5 tRNA Aminoacylation

Five tRNA synthetases were downregulated in mAb-producing clones at Day 15 (Lars, Nars, Sars, Tars and Yars). Aminoacyl-tRNA synthetases covalently attach each amino acid to its appropriate tRNA and are therefore key components of the cell's translational machinery. First, the amino acid is adenylated in an ATP-dependent reaction by the appropriate aminoacyl-tRNA synthetase. Subsequently, the adenylated amino acid, still complexed with the aminoacyl-tRNA synthetase, is transferred to its cognate tRNA molecules. The downregulation of a number of tRNA synthetases therefore suggests a general slowdown in cellular translation. A modulation of

aminoacyl-tRNA synthetases has been observed in batch and fed-batch cultures of 293 cells (Gupta and Lee, 2007), where they were downregulated in the late stages of fed-batch cultures. However, it should be noted that these cells were not expressing a secreted protein.

3.5 DISCUSSION

The process of selection in transfected mammalian cells allows for the isolation of cells which have stably integrated the heterologous gene into their genome. Expression constructs will integrate at random in the host chromosome and the site of integration is known to significantly affect the resultant expression, a phenomenon known as the "position effect" (Wilson et al., 1990). Integration of the product gene into a region of heterochromatin can lead to little or no expression, whereas integration into euchromatin can result in active transcription. This variation in integration site could possibly explain the diverse DHFR and IgG expression levels observed in Day 0 samples (following selection) of mAb-producing clones. Another contributing factor could be a variation in plasmid copy number following initial selection. This variation has been observed in CHO cells selected in HT-minus medium, where plasmid copy numbers were shown to vary from one to as many as 250 copies (Wurm and Petropoulos, 1994).

We observed that primary clones isolated following selection showed a varied propensity for gene amplification, as measured by mRNA expression levels of DHFR and the light chain of the IgG gene during methotrexate treatment (Figure 10). This enhanced ability of some clones to amplify exogenous genes was a phenomenon first described by Wahl and colleagues, who described it as a hyperamplifiable phenotype (Stark et al., 1989). In the current study, each clone showed varying degrees of gene amplification, possibly as a consequence of the integration site of exogenous genes. However, it is also possible that the transcriptional profile of each clone following selection could affect the cell's ability to efficiently amplify a secreted gene product. This notion is reinforced by the hierarchical clustering results we obtained, where a greater degree of similarity was observed amongst clones rather than across time. This suggests that the transcriptional response to amplification is in part dictated by the cell's expression profile following

selection. By studying the transcriptomes of a wider range of clones prior to amplification, one could possibly define an expression profile which can capture the cell's amplification potential. This could potentially be used as a screening tool to quickly identify promising clones.

Time series transcriptome analysis allowed us to track the mRNA levels of the DHFR gene and light chain of IgG during methotrexate treatment. While all five clones showed increased expression of the DHFR gene throughout the gene amplification process, two of the five clones exhibited a decrease in IgG light chain mRNA levels with prolonged methotrexate treatment. This suggests that the optimal duration of MTX treatment can vary amongst different clones, an observation which has been reported by others (Jun et al., 2005; Alahari, 2009). Furthermore, increases in DHFR and IgG mRNA levels were not always proportional within each clone. In the majority of clones, methotrexate treatment resulted in higher increases in IgG light chain mRNA level than DHFR, an observation which has been made by others (Chusainow et al., 2009).

Our analysis identified a modest number of genes as differentially expressed following methotrexate treatment. Furthermore, the differentially expressed genes induced by amplification were different in each of the mAb-producing clones observed, as well as in the control population. To our knowledge, this is the first study characterizing the temporal transcriptional response of mAb-producing cells to methotrexate treatment and gene amplification. However, in a study investigating the changes incurred at the proteome level upon stepwise MTX treatment and gene amplification, the authors examined the proteome of cell clones isolated at different MTX concentrations and found very few changes amongst the different clones (Hayduk and Lee, 2005). In another study, suppression subtractive hybridization was used to identify differentially expressed genes between a recombinant Ab-producing CHO cell line generated via gene amplification, and its parental host (Grillari et al., 2001). Five transcripts were identified as upregulated, including aspartate aminotransferase and lysyl-tRNA-transferase. The authors surmised that the recombinant cell line showed increased transcriptional and translational activity.

To evaluate the impact of the amplified gene on the cellular response, we compared the transcriptome profiles of cells which had solely amplified DHFR (control population) with those which had amplified both DHFR and IgG (mAb-producing clones). A few notable functional classes emerged. Among these was tRNA aminoacylation. We observed a downregulation of five tRNA synthetases in mAb-producing clones, suggesting a general slowdown in translation. A decrease in global translation is one of the hallmarks of the unfolded protein response, which is mediated via one of three sensor proteins, PERK (Schroder and Kaufman, 2005). However, this response attenuates global protein synthesis by abrogating translation initiation by blocking recruitment of initiator methionyl-tRNA to the ribosome (Harding et al., 1999). It is therefore difficult to conclude that overexpression of IgG induced an unfolded protein response in the mAb-producing clones. However, the identification of multiple stress response genes as differentially expressed between control population and mAb-producing clones suggests that expression of a secreted product did elicit a stress response in the amplified IgG clones.

3.6 CONCLUSIONS

In this study, we used time series transcriptome analysis to characterize the process of gene amplification during cell line development. Our experimental design allowed us to examine the transcript changes incurred following initial transfection and selection, as well as upon methotrexate-mediated gene amplification. Our study also included the use of a control population, expressing solely the DHFR gene, which allowed us to evaluate the effect of amplifying an antibody product gene, in addition to DHFR. We identified a significant degree of variation in each clone's expression profiles, as quantified by hierarchical clustering as well as varying degrees of gene amplification (as measured by mRNA levels of DHFR and light chain IgG). We identified an upregulation of stress response genes and a downregulation of tRNA synthetases in IgG-producing clones. By characterizing the transcriptome of additional clones, it may be possible to define a gene expression signature which could be used as a screening tool in cell line development.

4 TRANSCRIPTOME ANALYSIS OF ENHANCED PRODUCTIVITY CONDITIONS

4.1 SUMMARY

This chapter presents the results of a transcriptional analysis of recombinant CHO cells cultured under butyrate treatment at 33°C. The combined used of sodium butyrate treatment and reduced temperature cultivation were shown to increase specific productivity by more than 2-fold. By characterizing the transcriptional changes incurred under conditions of enhanced productivity, we can begin to better understand the hyperproducitivy phenotype. Using oligo-DNA microarrays, we investigated the gene expression profiles of cells cultured in batch mode under butyrate treatment at 33°C, and identified genes kinetically differentially expressed, as compared to a control 37°C culture. Through distance calculations, more than 900 genes were identified as kinetically differentially expressed. Furthermore, transcript levels of the heavy and light chain IgG genes increased upon treatment. Our results suggest that increased productivity under treatment is attributable to elevated secretory capacity, as supported by the results of gene set enrichment analysis (GSEA). Using this analysis, we found that elements of the secretory pathway, including Golgi apparatus, cytoskeleton protein binding and small GTPase-mediated signal transduction were enriched and thus may play a role in the increased recombinant protein production observed upon combined sodium butyrate treatment and reduced temperature cultivation. The biological insights garnered here can be used to design cell engineering strategies aimed at increasing recombinant protein production in CHO cells.

4.2 INTRODUCTION

Over the last decade, the production of recombinant protein therapeutics in cultured mammalian cells has seen a dramatic increase. These improvements have largely been the result of empirically-driven process enhancements. However, the advent of genomic and proteomic methodologies has facilitated the global study of increased productivity at the physiological level. Several approaches have been employed to

increase protein production through process enhancements, including the manipulation of feeding processes (Zhou et al., 1995; Wlaschin and Hu, 2006), media components (Dyring et al., 1994; Burteau et al., 2003; Hayduk and Lee, 2005) and culture conditions. Modified culture conditions reported to have a positive impact on achievable titers include reduced temperature cultivation (Barnabe and Butler, 1994; Fox et al., 2004), sodium butyrate treatment (Oh et al., 1993; Rodriguez et al., 2005; De Leon Gatti et al., 2007) and increased osmolarity (Shen and Sharfstein, 2006).

Sodium butyrate has been shown to increase productivity in a number of cultured cell lines, including Chinese hamster ovary cells (CHO) (Chotigeat et al., 1994; Kim and Lee, 2000; Hendrick et al., 2001; Mimura et al., 2001; Wang et al., 2002; Chun et al., 2003a) and hybridomas (Oh et al., 1993; Cherlet and Marc, 2000). Sodium butyrate is a known histone-deacetylase inhibitor, which blocks the action of histone deacetylases, preventing DNA condensation. As a result, the ability of histones to bind to DNA is decreased, allowing chromatin expansion and facilitating transcription (Sealy and Chalkley, 1978). Sodium butyrate has also been reported to promote growth arrest (Xiao et al., 1997). These mechanisms may contribute to the observed increase in product titer upon butyrate treatment. Conversely, the mechanisms by which reduced temperature cultivation leads to increased productivity are poorly understood. Cultivation at reduced temperature has been shown to increase productivity in some, but not all recombinant mammalian cell lines (Al-Fageeh et al., 2006). Several researchers have reported increased productivity in CHO cells upon temperature shift (Moore et al., 1997; Kaufmann et al., 1999; Yoon et al., 2003a; Yoon et al., 2003b; Fogolin et al., 2004; Trummer et al., 2006a). It has been suggested that prolonged exposure to sub-physiological temperatures results in a modulation of the cell cycle, transcription, translation and the cellular cytoskeleton (Baik et al., 2006).

In recent years, transcriptome analysis tools, such as microarrays, have gained popularity in cell culture applications, as reviewed in (Gupta and Lee, 2007). These applications have included the study of cell lines of varying productivities (Khoo et al., 2007; Seth et al., 2007), as well different culture conditions known to impact recombinant protein production (Baik et al., 2006; Shen and Sharfstein, 2006; De Leon Gatti et al.,

2007). These tools are valuable, as they can help provide insight into the underlying biological mechanisms of increased productivity. Such an understanding of the physiology of productivity can help guide rational cell engineering efforts as well as process enhancement strategies.

We have previously reported transcriptional surveys of the treatment of a Chinese hamster ovary (CHO) cell line to sodium butyrate and reduced temperature separately (Yee et al., 2008a; Yee et al., 2008b). In this study, we examined the combined effects of reduced temperature cultivation and sodium butyrate treatment on a high-producing CHO cell line in an effort to further our understanding of the complex trait of productivity. This combined approach led to improved productivities, compared to the individual treatments alone. Custom CHO Affymetrix arrays were used to study the cells' response in a time-dependent manner, and the dimensionality-reduction method of principal component analysis was used to identify major expression patterns in the data. Furthermore, we used gene set enrichment analysis to identify functional classes significantly enriched amongst kinetically differentially expressed genes. We identified Golgi apparatus, cytoskeleton protein binding and small GTPase-mediated signal transduction as some of the functional classes enriched under high productivity conditions. By studying the transcriptional response of this combined treatment, we are closer to defining the biological attributes which contribute to increased recombinant protein production.

4.3 MATERIALS AND METHODS

4.3.1 Cell Culture

The recombinant IgG-producing CHO cell line used in this study overexpresses two anti-apoptotic genes, Aven and E1B19K, and has been described previously (Figueroa et al., 2007). Cells were cultivated in CHO-S-SFM II media (Invitrogen, Carlsbad, CA) at 37°C in shaker flasks. For the treated condition, an exponentially growing culture was shifted to 33°C and sodium butyrate was added to a concentration of 2 mM 20 hr following temperature shift. The control culture was maintained at 37°C throughout without butyrate addition. An additional culture was also performed which

was shifted to 33°C during the exponential phase and maintained without butyrate addition. Viability and cell concentration were determined using the trypan blue dye exclusion method and hematocytometer cell counting. Cell pellets and supernatant were collected and kept frozen at –80°C and –20°C respectively until transcriptome analysis.

4.3.2 RNA Preparation for Target Labeling and Array Hybridization

Total RNA was isolated using RNeasy mini kits (Qiagen, Valencia, CA) according to the manufacturer's protocol. Biotin-labeled cRNA was prepared from 5 µg of total RNA from each sample using the one-cycle target labeling (Affymetrix, Santa Clara, CA) according to the manufacturer's instructions. Labeled, fragmented cRNA was hybridized on CHO Affymetrix arrays (version 1 with 10,118 probesets), washed and scanned at the University of Minnesota Affymetrix Microarray Core Facility. CEL files were processed using the GeneData Expressionist Refiner module (GeneData, San Francisco, CA), which was used to assess overall array quality and obtain a single intensity values for each probeset using the Microarray Analysis Suite Statistical Algorithm (MAS 5.0). The mean intensity values for all chips were linearly scaled to 500. Probesets with a detection p-value < 0.04 and intensity > 70 in at least one sample were retained for analysis.

4.3.3 Microarray Data Analysis

4.3.3.1 Identification of Kinetically Differentially Expressed Genes

Euclidean distance between the $\log_2$-transformed control and combined sodium butyrate and reduced temperature cultures at all five time points was calculated for every gene. Euclidean distance is defined as

$$d^i_{ab} = \sqrt{\sum_t (e^i_{at} - e^i_{bt})^2}$$

where d^i_{ab} is the distance between the expression pattern of gene i in conditions a and b; e^i_{at} is the transcript level of gene i at time t in condition a. Replicate cultures were averaged before calculating distance. Genes were called kinetically differentially

expressed if they met a Euclidean distance threshold greater than one standard deviation above the mean (Euclidean distance ≥ 2.2). A further criterion of 1.7 fold change between the treated and control cultures in at least one time point was also imposed.

4.3.3.2 Principal Components Analysis

Principal components analysis (PCA) was performed using the $\log_2$ mean-centered intensity values of the time-series transcriptome data (consisting of five time points) using Spotfire (Spotfire, Cambridge, MA) based on the algorithm described by Alter et. al. (Alter et al., 2000). Prior to PCA, genes with a low degree of covariance (CV ≤ 0.2) in the time-course profile were excluded.

4.3.3.3 Gene Ontology Enrichment Analysis

GenMAPP's MAPPfinder was used to integrate expression data with known pathways and determine confidence levels for differential expression within ontology groups (Doniger et al., 2003). Using this method, the significance of a gene class is calculated by an enrichment score, based on a null hypothesis described by a hypergeometric distribution. The enrichment score is defined as:

$$z = \frac{r - n\frac{R}{N}}{\sqrt{n\left(\frac{R}{N}\right)\left(1-\frac{R}{N}\right)\left(1-\frac{n-1}{N-1}\right)}}$$

where n is the number of genes in a gene class S, r is the number of differentially expressed genes in gene class S, N is the number of genes measured in the dataset and R is the number of differentially expressed genes in the dataset. Gene ontology definition files dated 09/08 were used for the analysis. Gene ontology terms with a significance p-value ≤ 0.05 were considered as enriched.

4.3.3.4 Gene Set Enrichment Analysis

All expressed genes were ranked by Euclidean distance and used as an input for Gene Set Enrichment Analysis (GSEA). A total of 239 gene sets constructed for the mouse were obtained from GenMAPP (www.genmapp.org) and used as the gene set

database for GSEA. The gene sets are derived from gene ontology annotations and KEGG metabolic pathways. For each gene set, GSEA uses a weighted Kolmogorov-Smirnov-like statistic to calculate an enrichment score (Subramanian et al., 2005). Using a gene set permutation procedure, an enrichment p-value is estimated to identify gene sets that have a strong correlation within the ranked list of genes. Gene sets with a nominal p-value ≤ 0.05 were considered as enriched.

4.3.4 Quantitative Real-Time PCR

Total RNA samples were treated with an RNAse-free DNAse kit (Ambion) to remove genomic DNA contamination. Reverse transcription of total RNA was performed according to the Superscript III Reverse transcriptase (Invitrogen, Carlsbad, CA) protocol for cDNA synthesis using 5 μg of total RNA and 1 μg of Oligo dT primers. Glyceraldehyde 3-phosphate dehydrogenase was used as a control for normalization. The forward and reverse primers for the antibody heavy and light chain constant region were designed using Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3) and a PCR product in the size range of 200 - 250 bp was specified in the design. The 5' and 3' primer for IgG heavy chain, constant region were GCTTCTATCCCAGCGACATC and CATCACGGAGCATGAGAAGA respectively. The 5' and 3' primer for IgG kappa light chain were CTGTTGTGTGCCTGCTGAAT and CCTAGCACTCTCCCCTGTTG.

SYBR Green I dye was used for detection of the PCR products using the Full Velocity SYBR green QPCR kit (Stratagene, La Jolla, CA). Each 12.5 μL PCR mix contained 6.25 μl of SYBR green PCR master mix (Applied Biosystems, Foster City, CA), 150 nM forward and reverse primers, and 50 ng template. Real-time PCR was performed using the Stratagene Mx3000P instrument (Stratagene). The reaction was started by incubating the samples at 94°C for 15 min, followed by 40 cycles of 94°C for 15 sec, 60°C for 30 sec, and 72°C for 30 sec. The PCR reaction was followed by incubation at 60°C for 20 sec and a 20 min ramping step. The threshold cycle number (C_t) was set at the minimum level of 10-fold the average standard deviation of the fluorescent intensity of baseline cycles. cDNA samples were run along with a no RT reaction control and a no cDNA template control. All samples were run in triplicates.

4.3.5 Antibody Quantification using ELISA

Antibody IgG concentrations in cell culture supernatants were measured by ELISA. Goat anti-human IgG Fc-specific and mouse anti-goat IgG Alkaline phosphatase (Sigma-Aldrich, St. Louis, MO) were used as primary and secondary detection antibodies, while *P*-nitro-phenyl phosphate was used as the enzyme substrate (Sigma-Aldrich, St. Louis, MO) and human IgG (Sigma-Aldrich, St. Louis, MO) was used as a standard. IgG concentration was determined by absorbance reading at 405 nm.

4.4 RESULTS

4.4.1 Effect of Low Temperature and Sodium Butyrate on Growth and IgG Productivity

The growth and antibody production kinetics under control (37°C) and treated conditions (33°C with 2 mM butyrate) are shown in Figure 14. An additional culture, maintained at 33°C throughout without butyrate addition is also shown. Three cultures were initiated at 37°C. After reaching a cell concentration of 5 x 10^5 cells/ml, two cultures were switched to 33°C while the other was maintained at 37°C. The effect of low temperature on cell growth was noticeable after 40 hr, whereby temperature-shifted cells reached a viable cell concentration of 8 x 10^5 cells/ml, while the control culture reached a viable cell concentration of 1.4 x 10^6 cells/ml (Figure 1A). At this point, sodium butyrate was added to one of the temperature-shifted cultures while keeping the other untreated. Viable cell concentration in the 37°C culture peaked at 1.9 x 10^6 cells/ml at 98 hr after which viability decreased. In comparison, the temperature-shifted and butyrate-treated culture reached a lower peak cell density of 1.4 x 10^6 cells/ml at the later time point of 120 hr, while the 33°C culture reached a maximal viable cell density of x cells/mL at y hr.

The combined temperature shift and butyrate treatment gave rise to a higher final IgG titer than in the control culture or under reduced temperature cultivation alone (Figure 14B). This elevated final concentration also translated to a three-fold increase in specific IgG productivity as compared to the control culture (Figure 14C). The increase

in IgG production was consistently observed in biological replicate runs (data not shown).

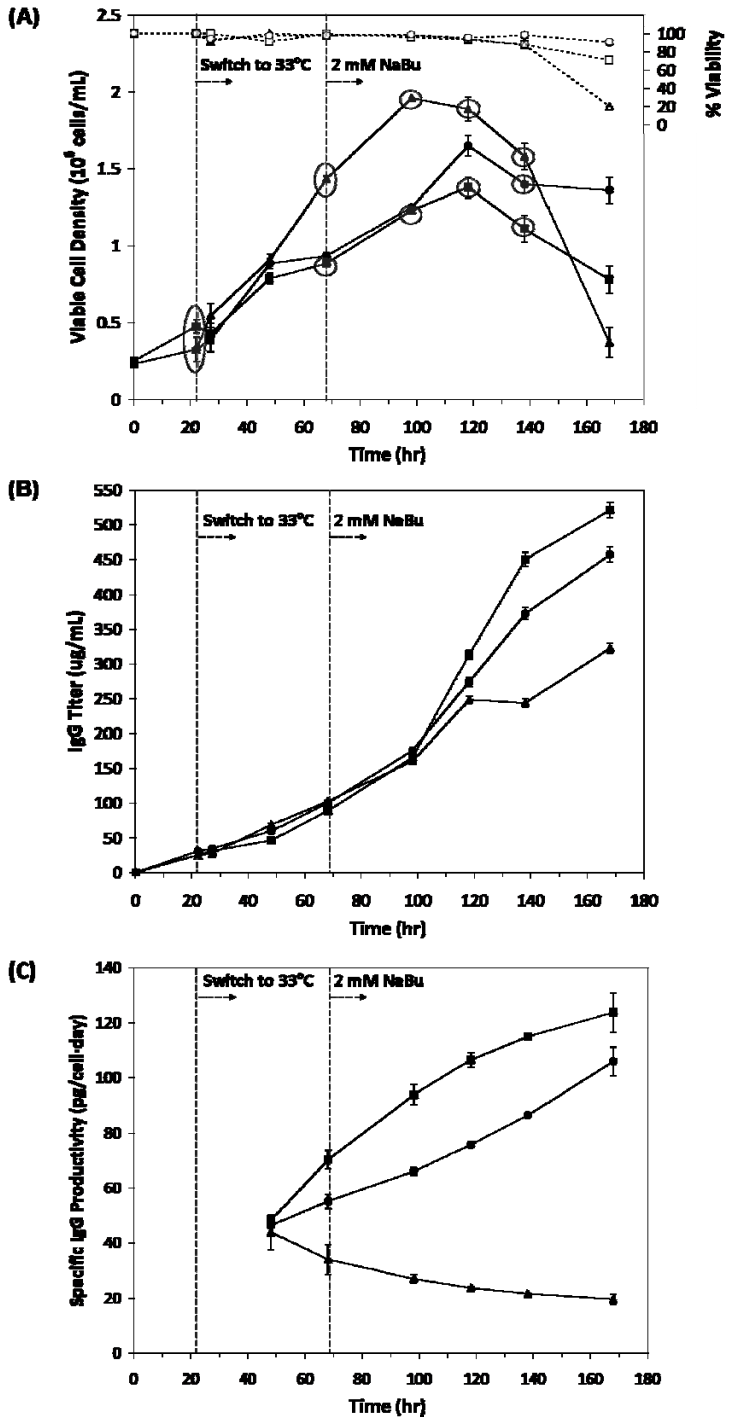


Figure 14 – (A) Viable cell density and viability (B) IgG titer and (C) specific IgG productivity of recombinant CHO cells under the following conditions: (▲) 37°C, (●) 33°C and (■) 33°C + butyrate. Shaded circles denote samples which were collected for transcriptome profiling.

4.4.2 Time Series Transcriptome Analysis

Samples were taken at 22, 68, 98, 118, and 138 hr (representing 0, 46, 76, 96 and 116 hr from the onset of the temperature shift, as indicated by shaded circles in Figure 14A) from both the control culture and combined low temperature/butyrate treated culture for microarray analysis. Additional samples at 22, 68 and 138 hr from biological replicate cultures were also analyzed. Hybridizations were also performed on 138 hr samples from biological replicate cultures grown at 33°C without butyrate addition (Figure 14A).

4.4.3 Transcriptome Dynamics

The transcriptional response of the treated and control cultures at different time points were compared to the initial time point at 22 hr (before temperature shift). A criterion of at least 2-fold change in transcript level was used to identify differentially expressed genes. The number of genes which met this criterion in each culture condition is shown in Table 5. The greatest degree of differential expression was seen under combined sodium butyrate treatment and reduced temperature cultivation, with the largest effect being observed at the 118 hr time point (96 hr following temperature shift and 50 hr following sodium butyrate addition). At this time point, 1089 genes were upregulated more than two fold with respect to the initial time point, while 1028 genes were downregulated. In comparison, at the same time point in the control culture, only 356 genes were upregulated and 444 genes were downregulated. When comparing the degree of overlap between differentially expressed genes, we find that 165 genes are commonly upregulated between the two cultures, while 268 genes are commonly downregulated. A significant degree of differential expression was also observed in the final time point of the culture (138 hr), as shown in Table 5. A similar degree of overlap between control and treated cultures was also observed at this time point, with 231 genes commonly upregulated and 342 genes commonly downregulated.

Slightly more genes were differentially expressed in the 33°C culture than in the control 37°C culture, but the number did not approach that observed for the butyrate-treated culture at 33°C. A comparison of the differentially expressed genes identified in

each culture found 138 genes to be commonly upregulated in all three cultures, while 264 genes were commonly downregulated.

Table 5 – Number of genes identified as differentially expressed in pair-wise comparisons.

	37°C		**33°C + butyrate**		**33°C**	
	Up	*Down*	*Up*	*Down*	*Up*	*Down*
68 hr / 22 hr	122	98	378	310	378	310
98 hr / 22 hr	184	101	725	915	N/A	N/A
118 hr / 22 hr	356	444	1089	1028	N/A	N/A
138 hr / 22 hr	451	543	1072	865	481	569

[1] *Differential expression is defined as absolute fold change ≥ 2 with respect to the 22 hr time point*

4.4.3.1 Principal Components Analysis

Principal components analysis (PCA) was used to reduce the dimensionality of the observed expression time profiles and identify major expression trends in both the combined butyrate-treated and temperature-shifted culture and the control culture. For the butyrate-treated culture, the first principal component obtained (PC 1) showed a monotonic trend in expression until the 118 hr time point. The second principal component (PC 2) showed an acute response to butyrate addition in the first 30 hr following treatment, and subsequently reverted the trend (Figure 15B). 78% of the observed variability in gene expression was captured by these first two principal components. In the control culture, the first principal component mirrored that of the treated culture, showing a monotonic increase or decrease with time. The second principal component displayed a peak in expression at 68 hr, subsequently reverting to a similar trend as PC 1 (Figure 15A). Together, these captured more than 85% of the variability in gene expression. This analysis suggests that gene expression trends are similar under both conditions, yet the extent of change for individual genes may be

different under treatment. This is consistent with the pair-wise comparison of transcriptome data shown above.

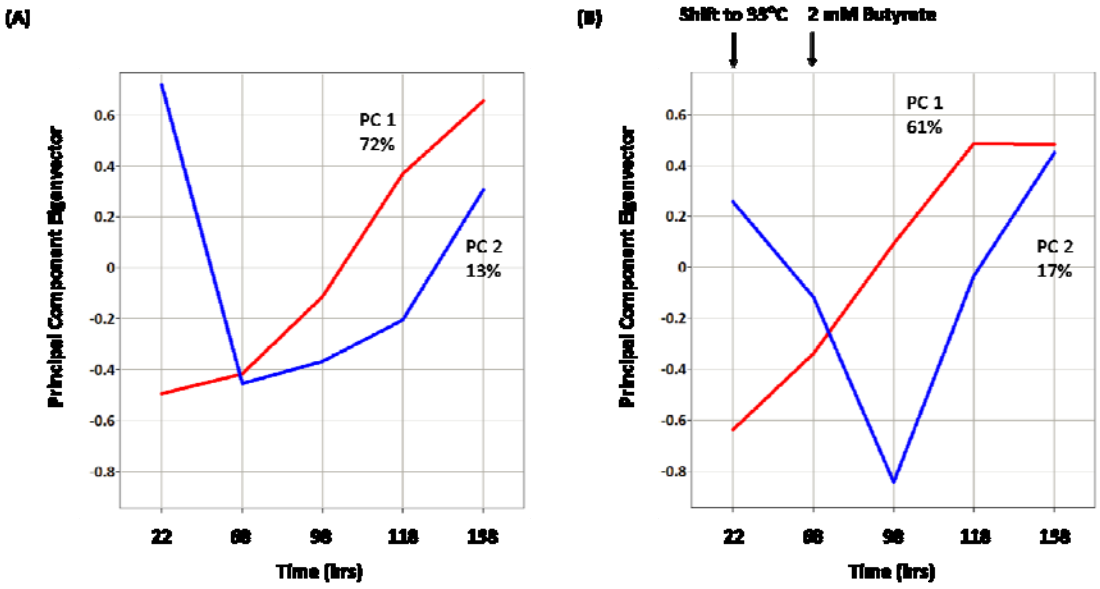


Figure 15 – Principal component analysis of time-series transcriptome data. (A) PCA results for 37°C control culture. (B) PCA results for sodium-butyrate treated and temperature-shifted culture. The percentages refer to the percent variance captured by each principal component.

4.4.3.2 Pair-Wise Distance Calculations

To identify kinetically differentially expressed genes upon butyrate treatment at 33°C, Euclidean distance (ED) was calculated between the expression profiles of the control and treated cultures. A higher Euclidean distance indicates that the expression profile of a gene differs between conditions. The criteria for kinetic differential expression used were ED $\geq$ 2.2, which represents a value of more than one standard deviation above the mean, and a fold change (treated/control) $\geq$ 1.7 in at least one time point. 978 genes met these kinetic differential expression criteria. Sixty-three of the kinetically differentially expressed genes showed a strong positive correlation to PC 1 in both the control and treated cultures (PC 1 $\geq$ 2, top percentile), while 111 genes displayed a strong negative correlation to PC 1 in both cultures. These genes therefore exhibit similar expression patterns, but different magnitudes of change upon treatment. Conversely, 179 genes had a strong positive correlation to PC 1 under one condition, and

a strong negative correlation under the other. Only 44 genes displayed a strong correlation, in either direction, to PC 2 (PC 2 $\geq$ 2, top percentile).

Among the genes with a strong positive correlation to PC 1 solely under butyrate treatment at 33°C, ovary-specific acidic protein showed the highest degree of fold change, with a 34-fold increase from 22 hr to 168 hr. Also strongly induced by combined butyrate treatment and reduced temperature was thioredoxin interacting protein, which changed by more than 23-fold during the course of the culture. Mcm5 (minichromosome maintenance deficient 5) was the most strongly downregulated by butyrate treatment at 33°C, with a 146-fold change (168 hr / 22 hr) in the treated culture and a 20-fold change in the control culture. Cyclin B1 displayed similar kinetics with 99-fold and 17-fold change in the treated and control culture, respectively. Lysosomal-associated protein transmembrane 4A showed the strongest positive correlation to PC 2 under butyrate treatment at 33°C. Inhibitor of DNA binding 2 showed a strong negative correlation to PC 2 in both cultures, with a peak fold change with respect to 22 hr of 2.2-fold and 2.5-fold, in the control and treated cultures, respectively.

4.4.4 IgG Transcript Expression

In order to investigate the transcription dynamics of IgG, quantitative PCR was used to probe the expression levels of heavy and light chain IgG transcripts. As shown in Figure 16, the expression of both heavy and light chains increased upon combined sodium butyrate treatment and reduced temperature cultivation. A maximum fold change of approximately 16-fold (with respect to the first time point) for both the heavy and light chain was observed at 98 hr, with expression levels subsequently decreasing. In contrast, IgG transcript levels in the control 37°C culture and the temperature-shifted 33°C culture each showed only a modest increase of approximately 2-fold, peaking at 98 hr.

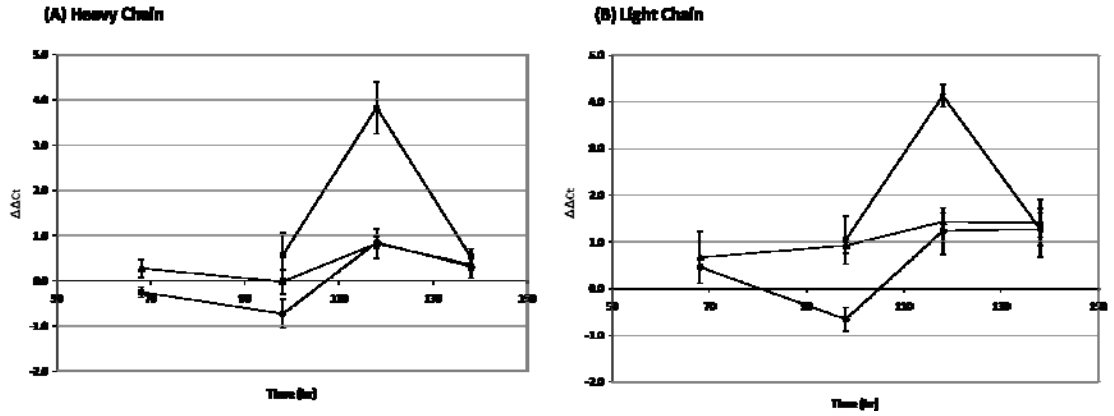


Figure 16 – Quantitative PCR results of heavy (A) and light (B) chain IgG expression under different culture conditions: (▲) 37°C, (●) 33°C and (■) 33°C + butyrate.

4.4.5 Functional Analysis of Temporally Differentially Expressed Genes

4.4.5.1 Gene ontology enrichment analysis

Gene ontology enrichment analysis was used to identify significantly enriched functional classes amongst kinetically differentially expressed genes (ED ≥ 2.2, absolute fold change (treated/control) ≥ 1.7 in at least one time point), using the MappFinder software (Doniger et al., 2003). This analysis identifies gene ontology terms with a higher-than-expected proportion of differentially expressed genes. Gene ontology terms which met an enrichment p-value cutoff of 0.05 are shown in Table 6, and are segregated according to their gene ontology classification, namely cellular component, molecular function and biological process. The number of genes which met the differential expression criteria in each gene ontology term, as well as the total number of genes within each term are also listed in Table 6.

Table 6 – Enriched gene ontology terms identified amongst temporally differentially expressed genes.

Gene ontology class	Number of differentially expressed genes[1]	Number of genes in GO class	Enrichment p-value
Cellular component			
Extracellular matrix	13	60	0.0
Cytoplasmic vesicle	18	93	0.029
Golgi apparatus	10	200	0.03

Molecular function			
Anion transport	5	30	0.0
Growth factor activity	7	25	0.0
Fatty acid beta-oxidation	8	21	0.038
Biological process			
Cell cycle	19	331	0.03
Regulation of small GTPase-mediated signaling	12	96	0.036

[1] *Differentially expressed genes are defined as those with a Euclidean distance of at least 2.2 and a maximum fold change across conditions of at least 1.7*

4.4.5.2 Gene set enrichment analysis

To complement the previous functional analysis, we also used gene set enrichment analysis (Subramanian et al., 2005). Genes were ranked according to their Euclidean distance and used as an input. Gene classes which met a nominal p-value threshold of 0.05 were considered significant. In this case, we focused our attention solely on gene sets enriched at the top of the ranked list (highest Euclidean distance), since we are interested in genes which are kinetically differentially expressed between the control and treated cultures. Gene classes which met this criterion are listed in Table 7.

Table 7 – Functional gene sets identified as significantly enriched by gene set enrichment analysis on genes ranked according to Euclidean distance between control and treated conditions.

Gene set	Number of differentially expressed genes[1]	Number of genes in gene set	Nominal p-value
Cell cycle	19	55	0.0
Extracellular matrix	13	49	0.0
Growth factor activity	7	19	0.001
Ion transport	5	18	0.037
Cytoplasmic membrane-bound vesicle	6	57	0.05
Cytoskeletal protein binding	6	82	0.05

[1] *Differentially expressed genes are defined as those with a Euclidean distance of at least 2.2 and a maximum fold change across conditions of at least 1.7*

Both functional analysis approaches yielded mostly overlapping results, with both methodologies identifying vesicle transport, cell cycle, growth factor activity, extracellular matrix components and ion transport as enriched functional classes under treated conditions. Furthermore, gene members of the Golgi apparatus, cytoskeletal protein binding and small GTPase-mediated signal transduction were also identified as functionally enriched under these conditions. Any method used to identify kinetically differentially expressed genes is not foolproof and will omit genes whose profiles are clearly differentially expressed when manually examined. Therefore, we examined all gene members of those functional classes which were identified through functional enrichment and imposed two additional criteria: firstly, an absolute PC 1 value in the treated culture greater than 0.88 (which represents the 15^{th} percentile) and an absolute PC 1 value in the control culture less than 0.88. This captures genes which have a strong monotonic increase or decrease in the treated culture, while remaining relatively constant in the control culture. The second criterion was an absolute fold-change greater than 1.4 fold between the treated and control culture in at least three consecutive time points. Gene members of enriched functional classes which passed either of these criteria were also considered differentially expressed and are included for discussion.

Gene expression profiles for members of the identified gene sets are represented using heat maps in the left panels of Figure 17 and Figure 18. Each row represents a transcript, and the color-shaded cell corresponds to the $\log_2$ mean-centered expression intensity at each timepoint. Additionally, bar graphs show the average gene expression fold change between the first and last time point in each culture. The full gene description, as well as differential expression criteria they met, average expression intensity at 0 hr, and average fold change in each culture are also listed in Tables IX – XI in the appendix. A list of genes which met the differential expression criteria, yet were not identified through functional analysis are listed in Table XI in the appendix.

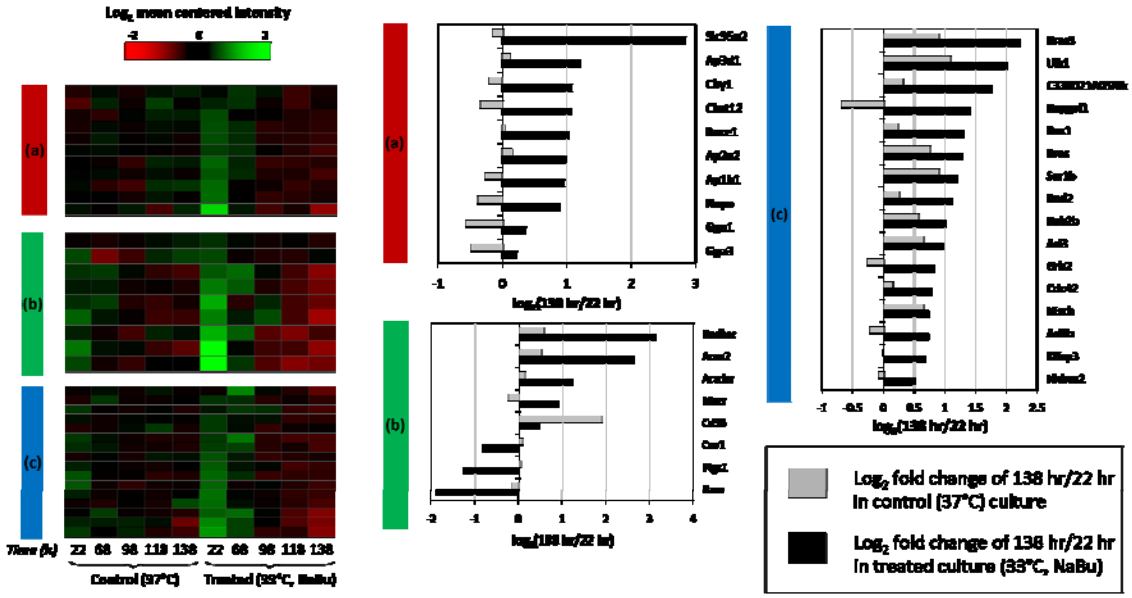


Figure 17 – Expression profiles of differentially expressed genes in enriched functional classes identified through gene ontology enrichment analysis: (a) Golgi apparatus, (b) Fatty acid beta-oxidation and (c) Regulation of small GTPase-mediated signaling. Heat maps in the left panel show time-series gene expression profiles of log2 mean centered intensity values. Bar graphs show pairwise comparisons of 138 hr / 22 hr in control (37°C) and treated (33°C + butyrate) cultures.

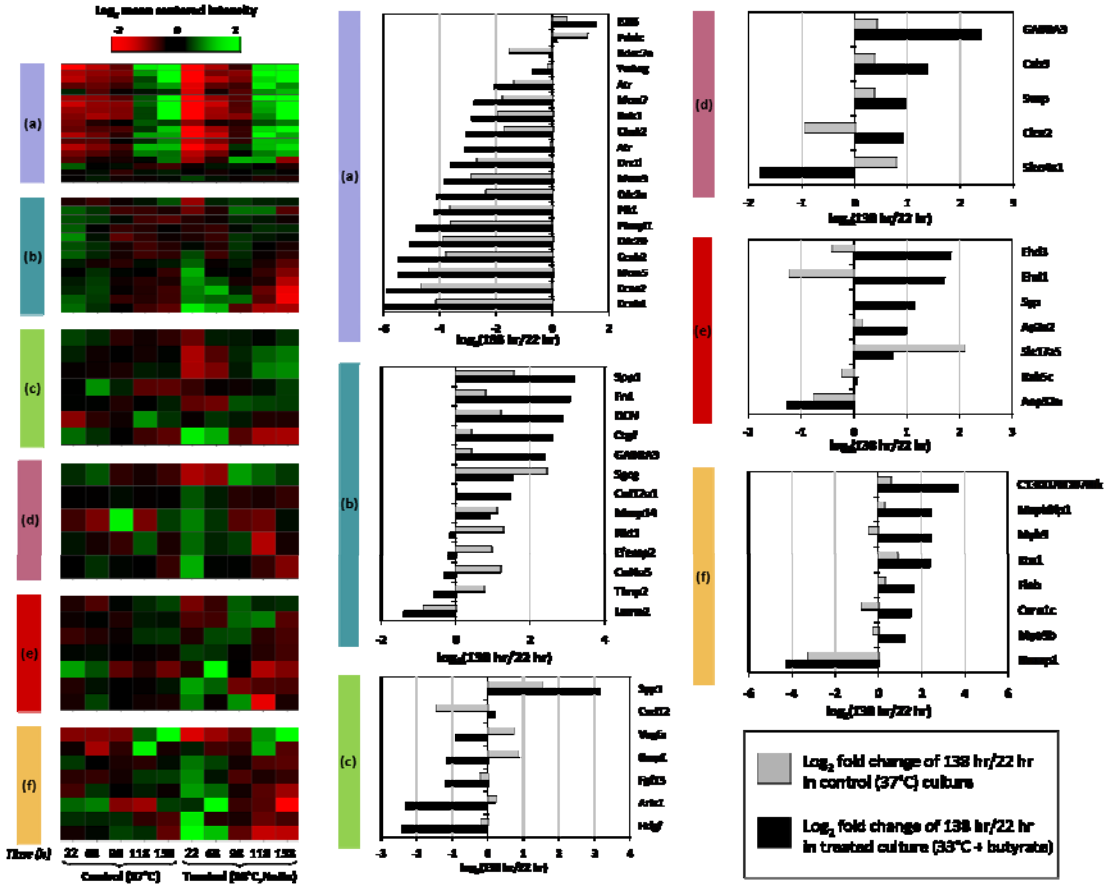


Figure 18 – Expression profiles of differentially expressed genes in enriched functional classes identified through gene set enrichment analysis: (a) Cell cycle, (b) Extracellular matrix, (c) Growth factor activity, (d) Ion transport, (e) Cytoplasmic membrane-bound vesicles and (f) Cytoskeletal protein binding. Heat maps in the left panel show time-series gene expression profiles of log2 mean centered intensity values. Bar graphs show pairwise comparisons of 138 hr / 22 hr in control (37°C) and treated (33°C + butyrate) cultures.

4.4.6 Golgi Apparatus

Genes localized to the Golgi apparatus were identified as significantly affected by combined low temperature and butyrate treatment. Many proteins destined for secretion are processed and sorted in the Golgi apparatus. The most prominently upregulated gene in this functional class was Slc35a2, which encodes for the UDP-galactose transporter. UDP-galactose is imported into the Golgi lumen through the UDP-galactose transporter and is used for the galactosylation of proteins and lipids (Sprong et al., 2003). Expression increased more than 7-fold under butyrate treatment, from an intensity of less than 15 to a moderately low expression level of 104. In contrast, this gene remains below the array detection limit throughout the control culture. For a comparison of abundance levels

corresponding to the range of signal intensities, the 25th, 50th and 75th percentiles represent signal intensities of 128, 500 and 1024 respectively.

Another gene prominently upregulated in this functional class was Ap3d1, which increased by almost 3-fold during the course of the treated culture, while remaining constant in the control culture. This gene is a subunit of the Ap3 adaptor-like complex which has been implicated in intracellular trafficking to lysosomal compartments (Simpson et al., 1997). In addition to Ap3d1, another adaptor-related protein, Ap2a2, was upregulated upon butyrate treatment at 33°C. This component of the AP2 adaptor protein complex plays a role in promoting the formation of clathrin-coated pits and vesicles (Stamnes and Rothman, 1993).

4.4.7 GTPase Signaling Elements

Consistent with the findings that transcripts involved in vesicle fusion/docking and budding are increased, twelve GTPase elements that activate these processes were differentially expressed. An Arl protein (Arl3) and a Sar1 GTPAse (Sar1b) were upregulated between 2.4 and 2.9 fold by butyrate treatment at 33°C. Collectively, the Arf (ADP ribosylation factor), Arl (ADP ribosylation-like) and Sar (Secretion associated and Ras-related protein) proteins form the Arf family of GTPases. These members regulate vesicular trafficking at multiple intracellular sites by recruiting coat proteins, regulating phospholipid metabolism and altering actin cytoskeleton networks at membrane surfaces (D'Souza-Schorey and Chavrier, 2006).

Like other Ras-related proteins, the Arf, Arl and Sar1 proteins cycle between their active-GTP bound and inactive-GDP-bound conformations. Hydrolysis of bound GTP is mediated by GTPase-activating proteins (GAPs), whereas the exchange of GDP for tri-phosphate nucleotide is mediated by guanine nucleotide-exchange factors (GEFs). Rapgef1, known as Rap guanine nucleotide exchange factor 1 is a member of the latter group (Tanaka et al., 1994). This protein has been shown to assemble into a signaling complex which may mediate insulin-stimulated actin rearrangement (Whiteman et al., 2002). The expression of this transcript increased by 2.9 fold during the course of the

culture under butyrate treatment at 33°C, while showing a small decrease in the control culture.

A number of downstream effectors of GTPase signaling were found to be kinetically differentially expressed including Rab factors (Rab5c, C30021A05Rik), Ras elements (Kndc1, Rsu1, Rras, Grb2, Ulk1) and Rho proteins (Ect2, Ngef, Plekhg2, Isn1, Rnd2, Cdc42). Collectively, these GTPases are responsible for regulating a diverse range of cellular activities, including changes in cell morphology, activation/deactivation of actin cytoskeleton networks and formation of focal adhesion.

4.4.8 Cytoplasmic Vesicles

Cytoplasmic vesicle gene members were identified as enriched, with eighteen of ninety-three genes meeting the kinetic differentially expressed criteria. Furthermore, GSEA analysis identified the subset of cytoplasmic membrane-bound vesicles as significantly enriched under butyrate treatment at 33°C. The majority of genes in this functional class showed increased expression under combined sodium butyrate treatment and reduced temperature cultivation. Among the most differentially expressed genes were two isoforms of the EH domain-containing gene, Ehd1 and Ehd3. Ehd1 expression increased 3.3-fold throughout the course of the treated culture, while Ehd3 showed a 3.6-fold increase in expression. Ehd proteins are known to play a role in the regulation of endocytotic recycling, including receptor recycling. Of note, loss of Edh1 in mammalian cells hinders transferrin receptor recycling. These genes are also known to interact with a number of Rab effector proteins, thus playing a role in the regulation of endocytotic trafficking events (Grant and Caplan, 2008).

4.4.9 Cytoskeletal Protein Binding

A number of cytoskeletal protein binding elements were identified as kinetically differentially expressed upon butyrate addition at 33°C. Among the upregulated genes, four (Coro1c, Flnb, Myh9, Myo9b) are involved specifically in the biogenesis of the actin cytoskeleton. Coro1c is a member of the coronin family which has been implicated in the remodeling of the actin cytoskeleton during endocytosis (de Hostos, 1999). Within the myosin family, Myo9b has been identified as an actin-based motor protein that relays

signal transduction based on the dynamics of the actin cytoskeleton. Specifically, Myo9b exhibits GTPase-activating properties on the Rho family of G proteins (Post et al., 1998).

4.4.10 Extracellular Matrix

A number of extracellular matrix components were identified as enriched between the treated and control culture. As shown in Figure 18, slightly more than half of the identified genes showed increased expression under butyrate treatment at 33°C, while the remainder showed increased expression in the control culture. Amongst the genes upregulated in the treated culture, Ctgf, or connective tissue growth factor, has been implicated in endothelial cell growth and survival (Brigstock, 2002). This gene increased by 4.1-fold under treated conditions, while remaining relatively constant in the control culture.

4.4.11 Cell Cycle

Cell cycle genes were identified as enriched amongst kinetically differentially expressed genes using both gene ontology enrichment and gene set enrichment analysis. As shown in Figure 18, the majority of these genes display a decreased expression profile with time under both control and treated conditions. Overall, members of this functional class showed the largest magnitude of fold change both across time and biological conditions, when compared to other identified functional gene sets. A number of cyclin genes were kinetically differentially expressed, including Cyclin A2, Cylin B1 and Cyclin B2. Cyclin A2, through binding and activation of Cdc2 or Cdk2 kinases, promotes G1/S and G2/M cell cycle transitions (Wheeler et al., 2008). Cyclins B1 and B2 are involved in mitosis and form a complex with p34, resulting in the assembly of the maturation-promoting factor (MTF) (Takizawa and Morgan, 2000). Decreased expression of these genes is consistent with an overall downregulation in the progression of the M phase. This is further supported by decreased expression levels of eight additional mitotic genes (Bub1, Ccna2, Ccnb2, Cdc20, Cdc2a, Pkmyt, Plk1, Ywhag).

4.4.12 Fatty Acid Beta-Oxidation

Eight of the twenty-one fatty acid beta-oxidation genes present on the CHO array were differentially expressed under combined butyrate treatment and low temperature cultivation. These genes showed a more pronounced increase in expression with time under butyrate treatment at 33°C, when compared to the control culture (Figure 17). Acsl1 and Acsl4 are isozymes involved in converting fatty acids of different chain lengths to fatty acyl-CoA, which is subsequently trans-esterified to fatty acyl-carnitine. Hadha and Hadhb are subunits of the mitochondrial trifunctional enzymes, which catalyzes the last three steps of fatty acid beta-oxidation, leading to the generation of acetyl-CoA and shortened acyl-CoA. Hadhsc, or L-3-hydroxyacyl-CoA dehydrogenase, short chain, catalyzes the penultimate reaction of fatty acid beta-oxidation, while Acaa2, or acetyl-coA acyltransferase 2, catalyzes the final step of fatty acid beta-oxidation.

4.4.13 Other Differentially Expressed Gene Members

Genes that met the differential expression criteria (ED $\geq$ 2.2, absolute fold change (treated/control) in at least one time point $\geq$ 1.7) but do not belong to a significantly enriched gene class identified in either functional analysis are listed separately in Table XI in the appendix. Based on gene ontology, these genes are broadly classified into three categories: protein processing and transport (glycosylation, protein transport, protein folding), metabolism (lipid metabolism, phosphoinositide metabolism), and cell growth and maintenance (regulation of DNA transcription and apoptosis).

4.5 DISCUSSION

The recombinant CHO cell line used in this study harbors exogenous anti-apoptotic genes and produces approximately 40 pg/cell/day of humanized IgG, a level that has been estimated for professional secretory cells in vivo. The combined butyrate and low temperature treatment further increased the IgG secretion rate to over 100 pg/cell/day. An increase in recombinant secretion of this magnitude likely represents a situation of overload in various aspects of cellular capacity. The dynamics of the transcriptome responding to the stimulus of treatment most likely carries the signature of the physiological alterations that lead to the drastically increased secretion rate.

Using quantitative PCR, both the heavy and light chain of the IgG transcript were found to be elevated under combined treatment, while under reduced temperature alone, the levels were largely similar to that seen in the control culture. At the time point of peak specific productivity, the transcript level of both light and heavy chain was nearly ten fold of that in the control culture. A similar increase in IgG transcript level has also been reported in sodium butyrate-treated CHO cells (Jiang and Sharfstein, 2008). The nearly ten-fold increase in transcript level observed in this study gave rise to an approximately three-fold increase in IgG secretion under combined treatment. It also interesting to note that under reduced temperature, the transcript levels of both heavy and light chain IgG remained at similar levels as that of the control culture, while the IgG secretion rate was increased by nearly two fold.

The synthesis of IgG molecules initiates with the translation of light and heavy chains independently in the cytosolic side of endoplasmic membrane. The nascent light chain and heavy chain molecules are translocated into the ER lumen where they are folded and assembled into IgG molecules. The assembled molecules, complete with disulfide bonds and glycans, are transported to the Golgi via membrane vesicles for further glycan processing. The final secretion takes place via secretory vesicles moving along the cytoskeletal assembly to the cytoplasmic membrane, where they exit to the extracellular space. Extensive biogenesis of this secretory machinery occurs during the development of secretory tissues, including liver, pancreas and immunoglobulin-secreting plasma cells. Presumably, a similar event takes place when recombinant CHO cells transform into protein secretory cells during the selection and amplification process of cell line development. It is not clear whether CHO cells which are already high secretors undergo further expansion of secretory capacity upon the stimulation of increased levels of heavy and light chain IgG transcripts.

To capture the physiological alterations coinciding with the increase in IgG production, we examined the transcriptome time dynamics of CHO cells undergoing these treatments. A challenge inherent to the analysis of time series gene expression data is the detection of differentially expressed genes that give biological relevance to the observed phenotype. With multiple arrays across different conditions from this time-

course study, the number of possible pair-wise comparisons can be daunting; furthermore, pair-wise comparisons, while convenient for comprehension and presentation, do not capture the dynamism of gene expression.

The specific productivity of both sodium butyrate/low temperature and control cultures changes monotonically over time, although in opposite directions. It is possible that genes related to increased productivity also exhibit a similar dynamic behavior. Using principal component analysis, we indentified a major gene expression pattern of monotonic increase and decrease in expression with time (PC 1) for both the butyrate-treated and low temperature culture and the control culture. A second component captured genes whose expression profiles differ somewhat between the two cultures. Note that through the observed time period, both cultures traverse from exponential growth into a decline phase. It is likely that the major principal components encompass characteristics related to this transition of growth. Genes in the same principal components in both cultures, although displaying similar dynamic behavior, may differ in their magnitude and/or direction of gene expression change and may contribute to the observed difference in productivity between cultures. We thus took a second approach to identify kinetically differentially genes across culture conditions. We used Euclidean distance of log2-transformed transcript abundance (intensity values) as a measure of the difference between the two time profiles of each gene. Using a criterion of a minimum fold change and a threshold value of Euclidean distance, 978 genes were identified as kinetically differentially expressed. The dynamic behavior of these genes span a wide spectra, with some being highly weighted in PC 1 or PC 2, and others exhibiting expression dynamics not captured by principal components analysis.

While pursuing the gene expression changes that may coincide with elevated IgG secretion, it is instructive to be reminded that there may be multiple routes to the same complex trait. It has been hypothesized that multiple functional alterations in energy metabolism, secretory machinery, redox balance, etc. lead to hyperproductivity of recombinant proteins, and multiple routes are possible to give rise to enhanced expression of each functional class. Hence, not all high producing cells or all hyperproductivity conditions may incur the same set of gene expression alterations (Seth et al., 2007).

Consequently, we examined changes in expression amongst gene classes, rather than focusing on individual genes. Using gene class testing and gene set enrichment analysis, we identified genes related to Golgi apparatus proteins, cytoplasmic membrane vesicles, cytoskeletal protein binding and GTPase signaling elements were identified as significantly enriched under combined sodium butyrate and reduced temperature cultivation. The increased expression of these gene members suggests that the cellular secretory pathway has been augmented under increased productivity conditions.

It can be useful to compare the results obtained in this study with those of previously published reports examining the transcriptional response of cells exposed to productivity-enhancing conditions. While previous studies differ in their methodology and experimental systems, the identification of common functional classes can further refine our physiological understanding of hyperproductivity. We previously reported a transcriptional analysis of CHO cells under low temperature cultivation (Yee et al., 2008). Differentially expressed genes induced by low temperature cultivation were found to belong to several key functional classes, including vesicle trafficking, endocytosis and cytoskeletal elements. While the current study found little overlap at the individual gene level (data not shown), it is interesting to note that similar functional classes were identified as enriched in both studies. Similarly, we compared the results of the current study to a previous report analyzing the transcriptome of CHO cells under sodium butyrate treatment (Yee et al., 2008). The analysis found 752 genes to be differentially expressed more than 1.4 fold at a p-value threshold of 0.05. Several genes related to vesicle trafficking were among the differentially expressed genes, and seven of those were also identified in the present study (Table 8). Both studies also identified genes involved in fatty acid beta oxidation to be changing under treated conditions. It has been reported that butyrate can be oxidized in human cell lines and its products subsequently incorporated into cellular lipids (Leschelle et al., 2000). This could potentially explain the observed increase in fatty acid oxidation genes under combined butyrate treatment and reduced temperature cultivation.

The similarities between the findings of the current study and previously published reports suggest that a core set of functional classes can be identified as

overlapping amongst multiple productivity-enhancing conditions. Interestingly, similar functional classes were also identified in a transcriptome study of several NS0 antibody-producing clones of varying productivities using a combination of gene class testing methodologies, including GSEA and gene set analysis (GSA). The functional classes identified include the Golgi apparatus, cytoskeleton, genes with GTPase regulatory activity and members of the cell cycle (Charaniya et al., 2009). This suggests that the biological attributes of increased protein production may be shared amongst recombinant cell lines of vastly different origins.

Table 8 – Comparison of differentially expressed transcripts from previous sodium butyrate study (Yee et al., 2008a) with the current study.

Gene symbol	Gene Description	Yee e. al. 2007	Current study
		Fold Change	Fold Change[1]
Pitpna	Phosphatidylinositol transfer protein	+ 1.7	+ 2.5
Timm8b	Translocase of inner mitochondrial membrane 8	+ 1.4	+ 2.4
Napg	N-ethylmaleimide sensitive fusion protein γ	+ 1.6	+ 2.0
Gga3	Golgi associated gamma adaptin ear containing 3	+ 1.6	+ 1.9
Arl1	ADP-ribosylation factor like 1	+ 1.6	+ 1.8
Vdp	Vesicle docking protein, p115	+ 1.5	+ 1.8
Ap1g1	Adaptor protein complex AP-1, γ	+ 1.7	+ 1.4
Clta	Clathrin polypeptide light chain	+ 1.5	N.C.
Sec61g1	Sec61g1	+ 1.8	N.C.
Timm10b	Translocase of inner mitochondrial membrane 10b	+ 1.7	N.C.

N.C. – no significant fold change.

[1] *In the current study, fold change is defined as 168 hr expression / 22 hr expression in the treated culture*

4.6 CONCLUSIONS

The microarray results reported here represent a transcriptional analysis of CHO-IgG under high productivity conditions stimulated by both low temperature and butyrate treatment. A large fraction of genes related to GTPase signaling, vesicle formation in the Golgi apparatus, and cytoskeletal protein binding were consistently upregulated upon low temperature induction followed by butyrate stimulation. The gene functional class alterations observed via transcriptional analysis suggest that elevated IgG production can be attributed to an overall increase in the protein secretion machinery, namely components of the ER, Golgi and associated proteins, akin to the transformation of resting B cells to professional antibody-secreting plasma cells (van Anken et al., 2003). The results presented here open up avenues for cellular engineering approaches which can enhance secretory capacity and increase antibody production.

5 PROCESS CHARACTERIZATION USING TRANSCRIPTOME ANALYSIS

5.1 SUMMARY

The production of recombinant antibodies in cultured mammalian cells is typically carried out in large-scale bioreactors operated in fed-batch mode. Optimization of these processes often involves manipulation of process parameters, including temperature and pH setpoints, as well as media improvement strategies. We have used oligo DNA microarrays to investigate the cellular transcriptome of CHO cells cultivated in fed-batch mode. Using time series microarray analysis, we evaluated the transcriptional impact of modulating process temperature and pH, as well as the effect of different hydrolysate lots, a supplement to the basal growth media, on cellular gene expression. In order to identify kinetically differentially expressed genes, we employed two approaches: a regression-based method, as well as a distance-based method. Modulation of each of these process parameters led to the kinetic differential expression of a large number of genes, and using gene ontology analysis, we identified a number of key biological pathways affected by each process parameter modulation. Finally, data mining techniques were applied to the entire transcriptome dataset in order to identify major trends. This analysis revealed that hydrolysate lot could effectively be used to classify the fed-batch cultures' transcriptional response. This discovery suggests that raw material source is a crucial component in process design and optimization, and the gene expression signatures identified could be used for rational culture medium design.

5.2 INTRODUCTION

In the previous two chapters, transcriptome analysis was used to investigate the process of cell line development and conditions known to increase recombinant protein production. We next applied microarray analysis to the characterization of cell culture processes used for recombinant protein production. As described in Section 2.1.3, typical mammalian cell culture production processes are operated in fed-batch mode, where periodic additions of a concentrated nutrient solution, usually containing glucose and amino acids, are used to achieve high viable cell concentrations. Such cultures are

normally carried out in bioreactors, where a number of process variables are controlled, including temperature, pH, O_2, CO_2 and agitation (Chartrain and Chu, 2008). While production runs are carried out at scales on the order of several thousand liters, it is common to employ a scale-down model for process optimization studies. Such scale-down models are commonly developed to mimic conditions at large-scale.

Manipulation of process parameters, including temperature and pH, is commonly used to enhance recombinant protein production (Chartrain and Chu, 2008). By optimizing process parameters, it is possible to significantly improve the performance of mammalian cell culture processes, in terms of cell growth, specific recombinant protein production, cell metabolism and viability. However, the molecular mechanisms by which modulation of these parameters leads to increased production remain largely unknown. Consequently, the use of microarrays is well-suited to characterize the molecular impact of manipulating process parameters and enable the development of more effective environmental process control strategies. However, to truly capture the culture's response to these modulations, it is necessary to characterize the transcriptome in a dynamic manner, thus necessitating a temporal experimental design. Time series microarray analysis is a useful approach to capture the dynamics of transcription in the cell.

The first process parameter to be investigated in this study was temperature. Modulation of culture temperature is commonly employed in industrial recombinant protein production to increase productivity (Birch and Racher, 2006). A moderate reduction in temperature has been shown to reduce cell growth, leading to reduced nutrient consumption and decreased accumulation of waste byproducts, such as lactate and ammonia (Yoon et al., 2003a; Yoon et al., 2003b). This can result in sustained viabilities, prolonging the duration of the culture. Consequently, it is common practice to employ a two-phase approach to protein production. Conditions are first optimized to allow maximal cell growth, allowing biomass to accumulate. Culture temperature is then reduced, allowing cells to redirect their metabolic resources to antibody production. Reduced temperature has been shown to increase specific productivity in some but not all

cell lines (Al-Fageeh et al., 2006). However, the mechanisms by which specific productivity is increased remain unclear.

The second process parameter under investigation was culture pH. Controlling culture pH is an integral part of ensuring an optimal environment for cell growth, as cells are known to be highly sensitive to changes in pH (Brusick, 1986). It has been shown that optimal pH values for recombinant protein-producing cell lines can vary (Osman et al., 2001), and that identifying the optimal pH can maximize cellular growth and productivity (Kurano et al., 1990; Borys et al., 1993; Link et al., 2004). Modulation of culture pH has also been employed as a biphasic growth strategy (Chartrain and Chu, 2008), analogous to the modulation of culture temperature. By gaining a deeper understanding of the biological impact of culture pH modulation, it may be possible to optimize cell culture process control strategies.

The third parameter to be studied was variation in raw material source, namely the use of different hydrolysate lots as supplements to the basal cell culture growth media. Hydrolysates have been common supplements in mammalian cell culture processes ever since the elimination of serum as a culture component. Hydrolysates are obtained by enzymatic hydrolysis of animal or plant material, including soybean meal and animal tissues. These hydrolysates contain a mixture of amino acids, small peptides, inorganic ions, carbohydrates and vitamins (Sung et al., 2004; Chun et al., 2005). While hydrolysates can be derived from both plant and animal sources, there has been a push by regulatory agencies in recent years to eliminate animal-derived components from cell culture processes to alleviate safety concerns, such as the introduction of animal viruses into the process (Lu et al., 2007). Hydrolysates are complex, i.e. undefined, nutrient supplements; therefore their exact chemical composition is unknown. However, it has been shown that these additives provide several advantages to the cells, and can reduce the time needed to optimize cellular growth medium. Given their complex nature, they are highly susceptible to variations in raw materials, and factors such as climate and geographic location can influence the composition of these supplements (Luo and Chen, 2007). Such variability has been shown to have a profound impact on the reproducibility and performance of cell culture processes (Schwartz et al., 2002). It can therefore be

highly beneficial to attempt to understand the biological mechanisms which are modulated by hydrolysates in order to potentially identify means to mitigate the impact their variability has on cell culture processes.

The field of data mining for knowledge discovery has grown leads and bounds in the past decade. As data of any kind is accumulated for a particular system, a number of data mining opportunities become accessible. Data mining techniques, including principal component analysis, supervised and unsupervised clustering, can be applied to both transcriptome (reviewed in (Nevins and Potti, 2007)) and bioprocess data (reviewed in (Charaniya et al., 2008a)) to help uncover correlations and trends, and lead to a more thorough understanding of cell culture processes. The use of data mining has been applied to manufacturing data from an antibody-production process and strong associations between several process parameters and antibody titer were uncovered (Charaniya et al., 2009). While no examples have yet been reported of data mining for transcriptome data from cell culture manufacturing processes, the techniques have been applied extensively in the field of oncology. By mining expression data obtained from multiple collections of human cancers, researchers were able to identify expression signatures representative of the activation status of several oncogenic pathways. This information was subsequently used to predict the sensitivity of therapeutic agents that target components of the identified pathways (Bild et al., 2006).

In this study, we systematically analyzed the transcriptional impact of modulating three different process parameters, temperature, pH and raw material source, on antibody-producing recombinant CHO fed-batch processes. Finally, the data from all three experiments was combined and data mining techniques were applied to uncover the global impact of these modulations on both process data and the cellular transcriptome.

5.3 MATERIALS AND METHODS

All fed-batch cultures, sample preparations and process parameter measurements were performed by our industrial collaborators, and process data was provided to us for further calculations and analysis. Microarray hybridizations were performed by Peter

Morin Nissom and Sze Wai Ng at the Bioprocessing Technology Institute in Singapore, and raw microarray data files were provided for our analysis.

5.3.1 Cell Culture

All studies described in this chapter were performed using an antibody-producing CHO cell line. The studies described in this chapter were performed in 10 L bioreactors, with a working volume of 2.5 L and operated in a fed-batch mode. Cells were seeded at 6 x 10^5 cells/mL in a starting volume of 2.3 L. The following feeding schedule was employed: on Day 6, 280 mL of basal media, containing 24 mM glucose and 4 mM glutamine was added to the system; and starting on Day 3, 60 mL of Feed Medium 1, containing 244 mM glucose and 50 mM glutamine, was added daily. Samples were taken daily and a number of process parameters were measured, listed in Table 9.

Table 9 – Process parameters measured during CHO fed-batch cultures.

Process Parameter	***Units***	***Process Parameter***	***Units***
Cell concentration	cells/mL	LDH	U/L
Viability	% viable cells	Glutamate	mg/L
Glucose	mg/L	Product	% of maximum titer
Glutamine	mg/L	Potassium	mg/L
Lactate	mg/L	Sodium	mg/L
Ammonia	mg/L	Osmolarity	mOsmol/kg

5.3.2 Calculation of Cumulative and Specific Rates

Calculations of cumulative and specific consumption and production rates are based on mass balance equations for all nutrients, metabolites and cells present in the culture. Cumulative consumption or production of nutrients and metabolites were calculated as follows:

$$\overline{S_{i,t}} = \int_0^t q_{i,t} \cdot x \cdot V \cdot dt \approx V_{t_0} \cdot S_{i,t_0} - V_t \cdot S_{i,t} + \sum_0^k V_{f_k} \cdot S_{f_k}$$

where:

$S_{i,t}$: cumulative amount of nutrient I consumed or produced at time t

V: culture volume

S_i: concentration of component i in the culture medium

V_f: volume of feed medium added

S_f: concentration of component i in the feed medium

k: the total number of feed medium additions up until time t

Specific consumption or production rates of nutrients and metabolites were calculated as follows:

$$q_{\bar{S}} = \frac{1}{x \cdot V} \frac{d\bar{S}}{dt}$$

A third-order polynomial function was fitted to each component's cumulative consumption data and the fitted equations were used to take the derivative and calculate the specific rate.

5.3.3 Microarray Data Processing

Samples for microarray analysis were taken on the following days: 3, 6, 8, 10, 13 and 15. In all cases, samples were taken from duplicate cultures at each experimental condition. Additional samples were taken at Day 4 for the pH shift study. For microarray analysis, 1 x 10^6 cells were sampled and total RNA was isolated using RNeasy Midi kit (Qiagen). Sample labeling and hybridization were performed as described in section 4.3.2 using CHO version 2 Affymetrix arrays, which interrogate expression with 23,020 probesets. The average intensity on each array was normalized by global scaling to a target intensity of 500. Probeset intensities were condensed using the Affymetrix Microarray Suite 5 (MAS5) algorithm. Subsequently, quantile normalization was used to align all intensity distributions. A representative graph of intensity distributions before and after quantile normalization is shown in Figure 19. Following quantile normalization, genes with a maximum intensity ≤ 70 and a detection p-value ≥ 0.04 across all samples in a given study were called absent and excluded from further analysis.

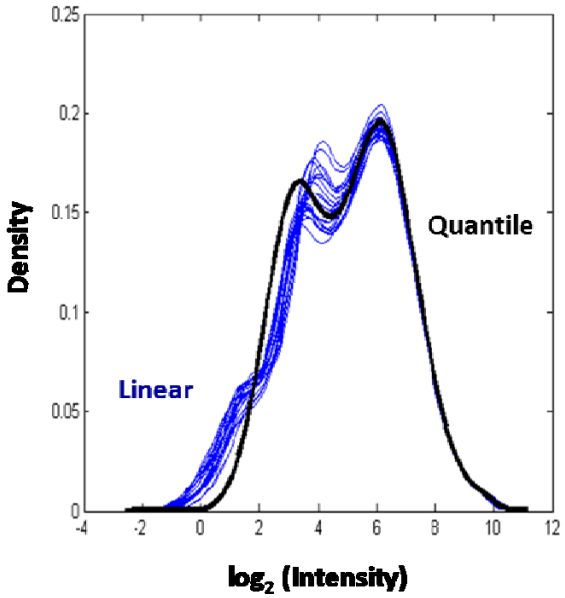


Figure 19 – Representative intensity distribution profiles before (linear) and after quantile normalization.

5.3.4 Microarray Data Analysis

5.3.4.1 Identifying Kinetically Differentially Expressed Genes

Gene expression as a function of time was analyzed using EDGE, an implementation in the statistical software R of the algorithm described by Storey et. al. (Storey et al., 2005). This algorithm allows for the identification of genes which are statistically different across time and experimental conditions. Although this method is statistically rigorous, it requires at least two replicates of time-series data for every experimental condition. A description of the algorithm is given below:

If y_{ijk} is the expression level of gene i in biological replicate j at the k^{th} time point, then

$$y_{ijk} = \mu_i(t_{jk}) + \gamma_{ij} + \varepsilon_{ijk}$$

where $\mu_i(t)$ represents the population average (average of all biological replicates) time curve for gene i

γ_{ij} accounts for the variation of biological replicate j from the population average

ε_{ijk} accounts for other sources of variation (including random error) in the measurement of gene i in replicate j at time k

The population average time curve $\mu_i(t)$ can be expressed as a p^{th} order polynomial function of time:

$$\mu_i(t) = \alpha_i + \beta_{i1}t + \beta_{i2}t^2 + \cdots + \beta_{ip}t^p$$

With this formulation, the hypothesis of kinetic differential expression of gene i can be tested in terms of the parameters (α and β). Thus, if the gene is not significantly differentially expressed between the two experimental conditions, then the null hypothesis (H_O) is:

$$\mu_i(t)_{group1} = \mu_i(t)_{group2}$$
$$\therefore \alpha_{i,group1} = \alpha_{i,group2}$$
$$\therefore \beta_{im,group1} = \beta_{im,group2} \quad \forall m = 1,2,\ldots,p$$

The alternative hypothesis (H_A) is that the parameters (α and β) are different in the two groups.

Prior to analysis, genes with a coefficient of variation ≤ 20% were removed. For each study, an EDGE input file with 24 samples, consisting of 2 biological replicates at each of two conditions over 6 time points was used for analysis. A third order polynomial function was specified for data fitting in the longitudinal model. Genes with a false discovery rate of ≤ 5% were considered kinetically differentially expressed

As a second approach, distance calculations, in this case, Euclidian distance, were used to identify kinetically differentially expressed genes. Details of this method can be found in Section 4.3.3.1.

5.3.4.2 Functional Analysis

Gene class testing was used to identify gene ontology terms significantly enriched amongst differentially expressed genes. GenMAPP's MappFinder application was used for this, and the details of the methodology are described in Section 4.3.3.3.

5.3.4.3 Similarity Analysis

Hierarchical clustering was used to cluster each of the 12 CHO fed-batch cultures based on their time series transcriptional profiles. First, the distance of a gene's temporal expression profile (g) was computed between every possible pair-wise combination of runs (*i* and *j*) using Pearson distance as the distance metric:

$$s_{ij}^{g} = 1 - \frac{(x_i - \overline{x_i})(x_j - \overline{x_j})'}{\sqrt{(x_i - \overline{x_i})(x_i - \overline{x_i})'}\sqrt{(x_j - \overline{x_j})(x_j - \overline{x_j})'}}$$

where $\overline{x_i} = \frac{1}{n}\sum_t x_{it}$ and

$$\overline{x_j} = \frac{1}{n}\sum_t x_{jt}$$

and where t is the total number of time points in each run (6 in this case).

The similarity between any two runs (*i* and *j*) was defined as a linear combination of the similarity between each genes' profile:

$$s_{ij} = \sum_{g=1}^{G} s_{ij}^{g}$$

where G is the total number of present genes on the array. The resulting similarity matrix was subsequently used as an input for hierarchical clustering. Briefly, a hierarchical cluster tree was calculated from the distances in the matrix s using the nearest neighbor method. This uses the smallest distance between objects in the two clusters:

$$d(r,s) = \min(dist(x_{ri}, x_{sj})), i \in (i,\ldots,n_r), j \in (1,\ldots,n_s)$$

where n_r is the number of objects in cluster r and x_{ri} is the i^{th} object in cluster r.

In a parallel approach, the Jmp Genomics 3.2 software (SAS Institute, Cary, NC) was used to perform variance components analysis. Variance components models are used to determine the amount of variation in a dependent variable that is associated with one or more random-effects variables. In other words, this method estimates the contribution of each experimental variable to the variance in gene expression for all

expressed genes on the array, assuming each variable is a random effect. Briefly, a similarity matrix based on the gene expression values of each microarray sample is first calculated using the Pearson correlation metric. The resulting similarity matrix is used as an input for principal component analysis. A random effects model is then fitted to every principal component eigenvector:

$$y_N = Z\gamma + \varepsilon$$

where y_N is the Nth principal component's eigenvector

Z is a known design matrix, which captures the level of each random effect in every sample

γ is an unknown vector of random effects parameters, and

ε is an unknown random error vector.

From the parameters of the random effects model, the variance components of each effect (experimental variable) can be determined. The final variance of each effect is obtained by weighting the variance calculated in each principal component by the overall variance captured by each principal component.

5.3.5 Parameter Profile-Based Comparison of Process Runs

Our objective was to identify which process parameters where most affected by each of the three process parameter manipulations (i.e. temperature, pH, hydrolysate lot source) across all 12 characterized CHO fed-batch cultures. First, we compared the process parameter profiles between any two runs for each of the twelve measured parameters (viable cell density, antibody titer, etc.) and generated a similarity score for each process parameter. To do this, we compared the time profile of a process parameter (p) in runs i and j using Euclidean distance (d_{ij}^{p}):

$$d_{ij}^{p} = \left\| p_i - p_j \right\| = \sqrt{\sum_{k=1}^{t} (p_{ik} - p_{jk})^2}$$

where, p_{ik} is the measured value of the parameter in run i at time k; and t represents the total number of time points (15 in this case). This was calculated for every possible pair-wise run comparison, yielding a distance metric with N x N dimensions, where N is the total number of runs (12 in this case). In order to convert these distances to similarities, Euclidean distances were first scaled on a scale of 0 to 5, with 0 corresponding to the highest similarity, and 5 corresponding to the lowest similarity. Subsequently, the following transformation was applied:

$$s_{ij}^{p} = \exp(-d_{ij}^{p})$$

which results in similarity values ranging from 0 (dissimilar profiles) to 1 (identical profiles).

Three similarity matrices were then created for each experimental variable. These capture whether two runs were operated at the same experimental condition. In other words, if both runs were operated at 37°C, then their similarity is 1. Conversely, if one run was grown 37°C and the other at 34°C, their similarity is 0. Finally, a correlation coefficient was calculated for each process parameter by calculating the similarity between its similarity matrix and the experimental variable matrix (temperature shift, component lot source and pH shift), using both the Pearson and Spearman correlation calculations. The output can be interpreted as follows: if a process parameter's behavior across runs can be efficiently segregated by a particular experimental variable, then it will have a high similarity coefficient.

5.4 TRANSCRIPTIONAL IMPACT OF TEMPERATURE SHIFT

5.4.1 Results

We investigated the impact of reduced temperature cultivation on gene expression in CHO fed-batch cultures. Four cultures were inoculated at 6 x 10^5 cells/mL and maintained at 37°C; upon reaching 6 x 10^6 cells/mL, two of the four cultures were shifted to 34°C and maintained at this temperature throughout. All cultures were sampled daily for process parameter measurements. Samples were taken at Days 3, 6, 8, 10, 13 and 15 from all four cultures for microarray analysis.

5.4.1.1 Impact on Process Parameters

Temperature shift had a profound effect on a number of process parameters. Temperature-shifted cultures reached lower maximal viable cell densities than their constant temperature counterparts, but also sustained higher viabilities for longer durations (Figure 20). While the effect of temperature shift on IgG titer was mixed (Figure 21), there was a substantial change in specific productivity, as evidenced by the difference in slopes in Figure 22 (specific productivity is the slope of integral viable cell concentration versus antibody titer). In fact, temperature shift increased specific productivity by more than 27%. The slowdown in growth caused by reduced temperature cultivation also translated to reduced consumption of main carbon sources, including glucose (Figure 23) and glutamine (Figure 24). Interestingly, lactate consumption profiles were different under temperature-shifted conditions (Figure 25). Namely, cultures under a temperature shift consumed lactate much more readily.

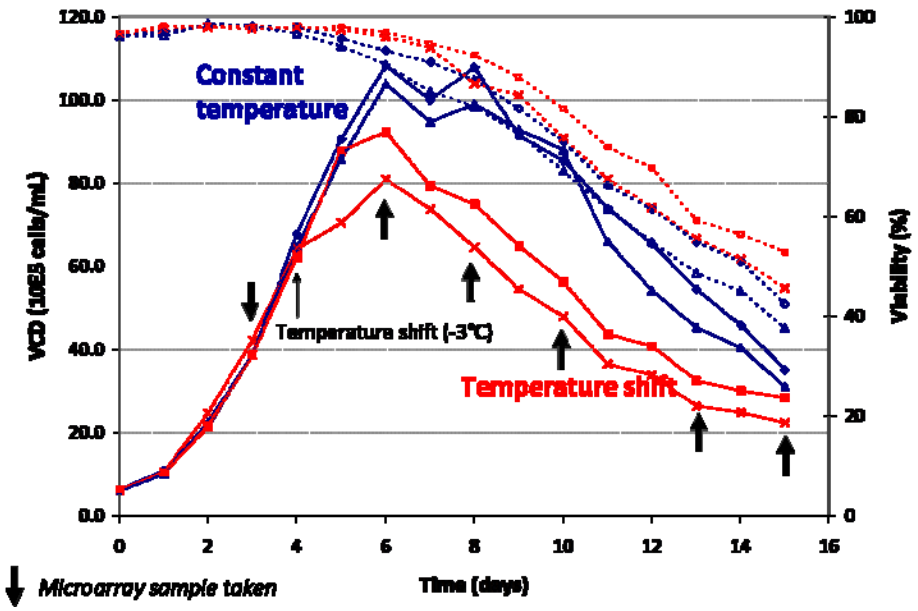


Figure 20 – Viable cell densities (VCD) and percent viability of cultures grown at constant and temperature-shifted conditions. Cultures maintained at 37°C are shown in blue, while temperature-shifted cultures are shown in red. Samples taken for microarray analysis are indicated with black arrows.

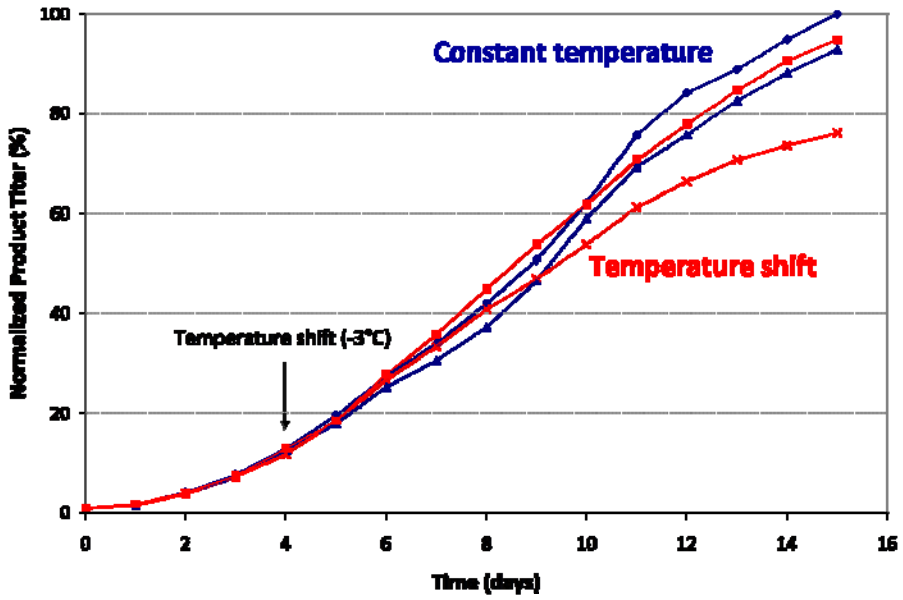


Figure 21 – Antibody titers of cultures grown at constant and temperature-shifted conditions. Titers are normalized to the highest value attained. Cultures maintained at 37°C are shown in blue, while temperature-shifted cultures are shown in red.

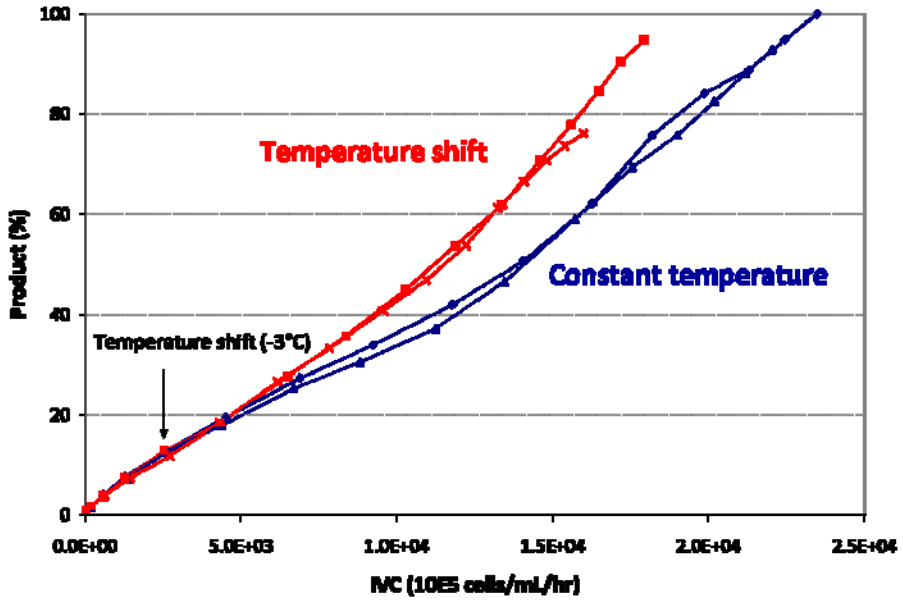


Figure 22 – Antibody titers plotted versus integral viable cell concentration (IVC) of cultures grown at constant and temperature-shifted conditions. Titers are normalized to the highest value attained. The slope of each line represents the specific antibody productivity. Cultures maintained at 37°C are shown in blue, while temperature-shifted cultures are shown in red.

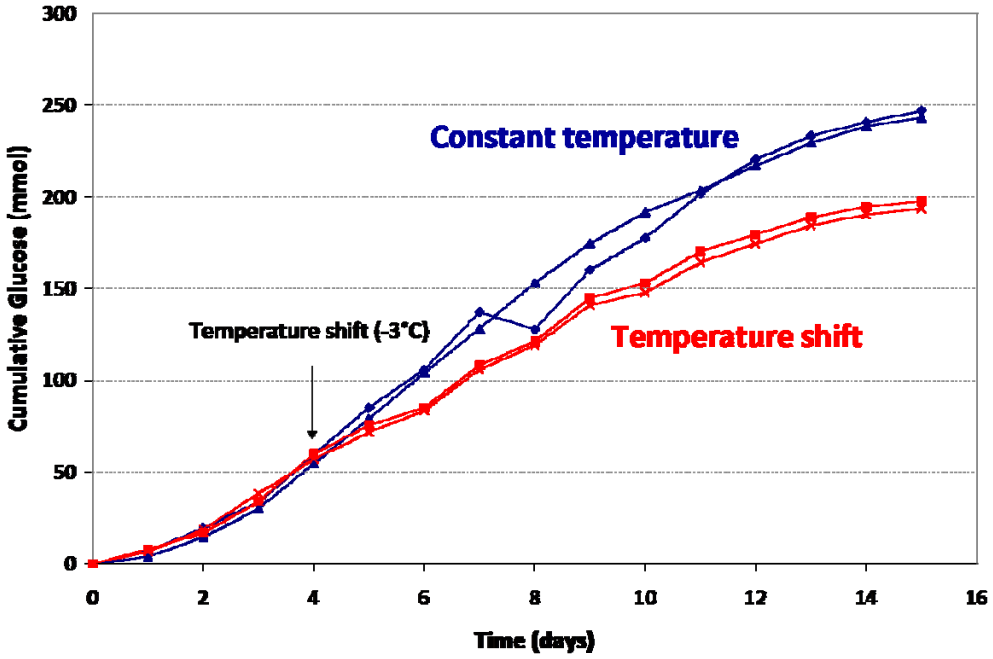


Figure 23 – Cumulative glucose consumption profiles of cultures grown at constant and temperature-shifted conditions. Cultures maintained at 37°C are shown in blue, while temperature-shifted cultures are shown in red.

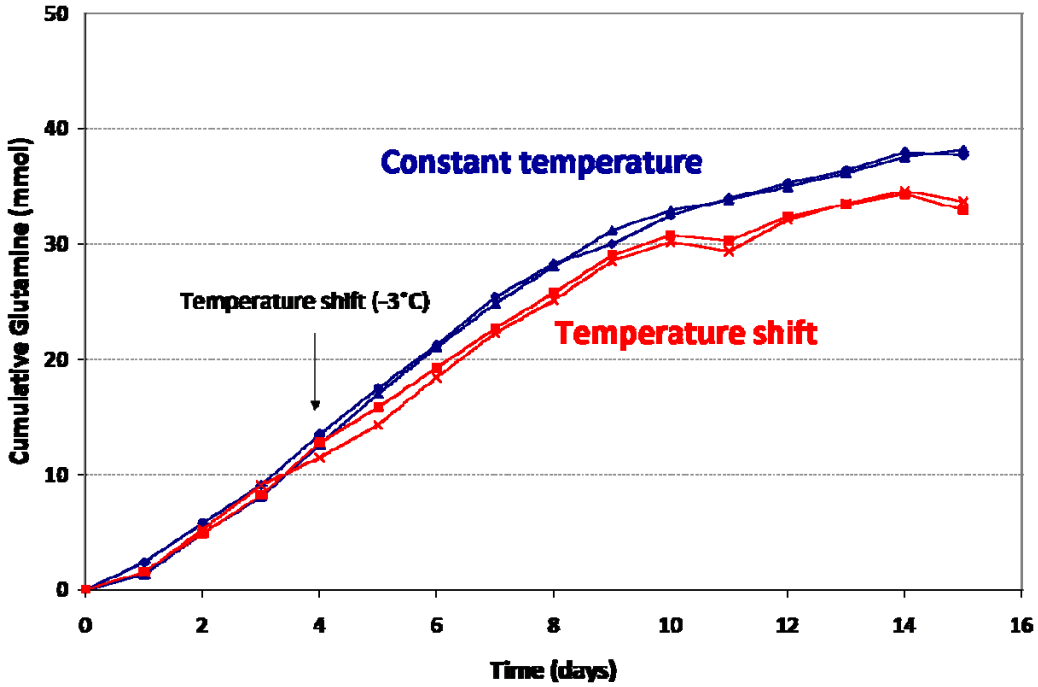


Figure 24 – Cumulative glutamine consumption profiles of cultures grown at constant and temperature-shifted conditions. Cultures maintained at 37°C are shown in blue, while temperature-shifted cultures are shown in red.

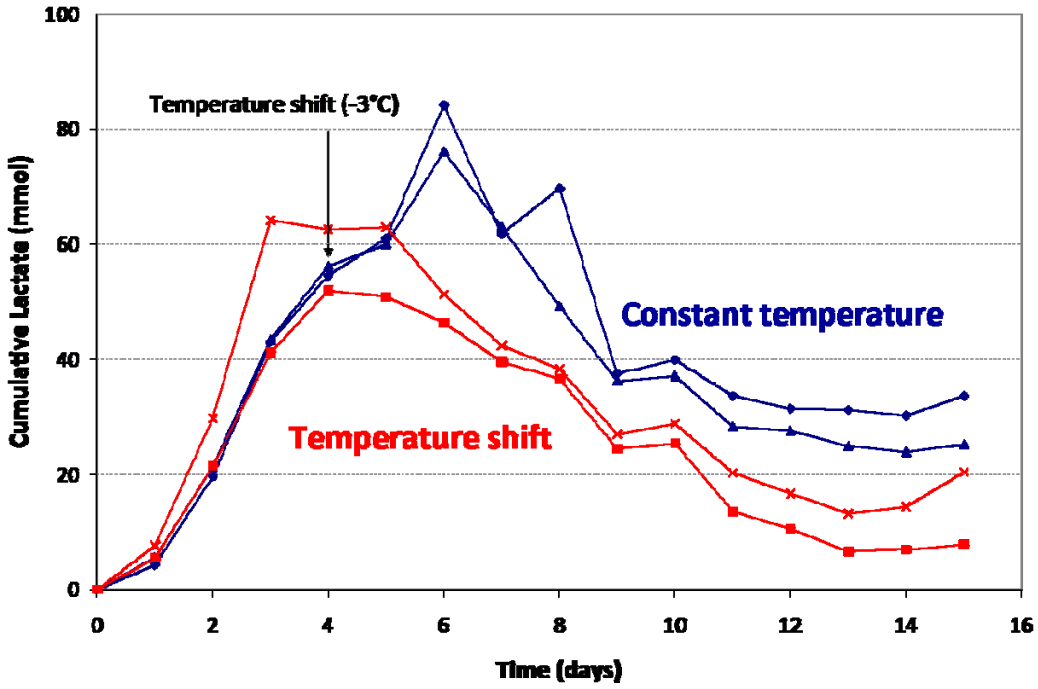


Figure 25 – Cumulative lactate profiles of cultures grown at constant and temperature-shifted conditions. Cultures maintained at 37°C are shown in blue, while temperature-shifted cultures are shown in red.

5.4.1.2 Identifying Kinetically Differentially Expressed Genes

We used a non-linear regression-based method to identify kinetically differentially expressed genes following temperature shift (Storey et al., 2005). 1072 genes were identified as kinetically differentially expressed at a q-value threshold of 0.05. This set of genes was further restricted using a Euclidean distance cutoff. Euclidean distance between control and temperature-shifted cultures was calculated on $\log_2$-transformed data (biological replicate cultures were first averaged before calculating Euclidean distance) to identify those genes with significantly different expression profiles between each condition. Genes were considered kinetically differentially expressed if they had a Euclidean distance ≥ 1.8, which represents two standard deviations above the mean of calculated Euclidean distances. This reduced the set of kinetically differentially expressed gene to 514. The top kinetically differentially expressed genes are shown in Figure 26.

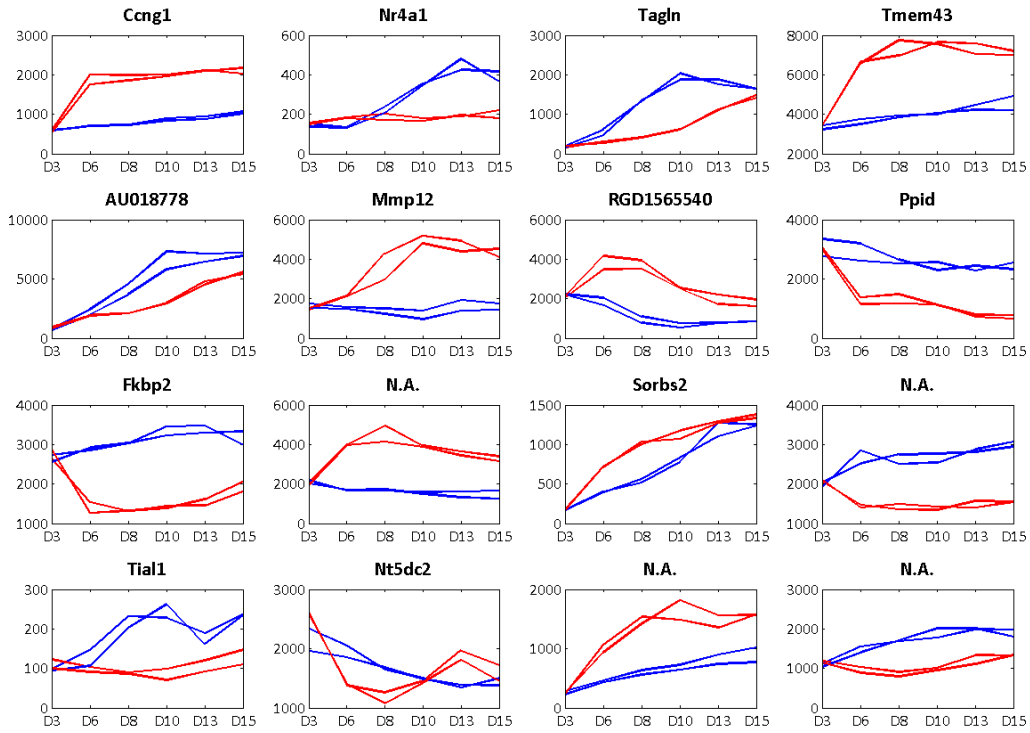


Figure 26 – Top kinetically differentially expressed genes identified through non-linear regression and further restricted using Euclidean distance (see text for description). Each subplot represents a different probeset on the array. The gene symbol is shown in the title, and intensity at all six time points is plotted.

5.4.1.3 Functional Analysis

In order to functionally interpret this set of kinetically differentially expressed genes, we performed gene ontology enrichment analysis. This type of analysis identifies those ontological terms which are present in greater than expected frequency amongst a set of differentially expressed genes. The analysis was performed using GenMAPP's MappFinder software (Doniger et al., 2003), which employs a hypergeometric distribution. Ontological classes with an enrichment p-value ≤ 0.05 were considered significant, and are shown in Table 10. Furthermore, two commercially available pathway analysis softwares, Ingenuity Pathway Analysis and Metacore were used to visualize kinetically differentially expressed genes in the context of canonical metabolic and signaling pathways.

As can be seen in Table 10, several functional classes were identified as significantly enriched amongst kinetically differentially expressed genes, and these are discussed in the sections below.

Table 10 – Significantly enriched gene ontology terms amongst kinetically differentially expressed genes.

Gene Ontology Term	Number of genes DE	Number of genes in GO term	Enrichment p-value
Endoplasmic reticulum	28	733	2.11e-6
Membrane-bound vesicle	15	569	0.013
GTPase activator activity	10	243	0.097
Metallopeptidase activity	12	167	0.0007
Microtubule cytoskeleton	23	392	5e-5
Cell cycle	62	878	2e-14
Fatty acid metabolism	6	103	0.01
Calcium transport	5	118	0.003

5.4.1.4 Cell cycle

Cell cycle genes were significantly enriched amongst kinetically differentially expressed genes. This is not surprising, as reduced temperature cultivation has been shown to induce a profound modulation of the cell cycle (Al-Fageeh et al., 2006). A number of mitosis-related genes were amongst the differentially expressed genes, and their expression profiles are shown in Figure 27. As can be seen, the majority of these genes decreased expression throughout the course of the culture. Furthermore, their expression in temperature-shifted cultures decreased much more rapidly, suggesting a possible slowdown in cell cycle progression under reduced temperatures. Chief amongst the genes identified were a number of cyclin genes, including Cyclin A2 (Ccna2), Cyclin B1 (Ccnb1), Cyclin F (Ccnf) and Cyclin G1 (Ccng1). All genes were downregulated throughout the course of the culture, with the exception of Cyclin G1, which was upregulated more than 2-fold upon temperature shift while remaining relatively constant throughout the 37°C culture (Figure 27). Cyclins are proteins which regulate progression through the cell cycle via their regulation of cyclin-dependent kinases (CDKs). Each

cyclin forms a complex with its partner CDK, leading to activation of the CDK kinase activity (Galderisi et al., 2003).

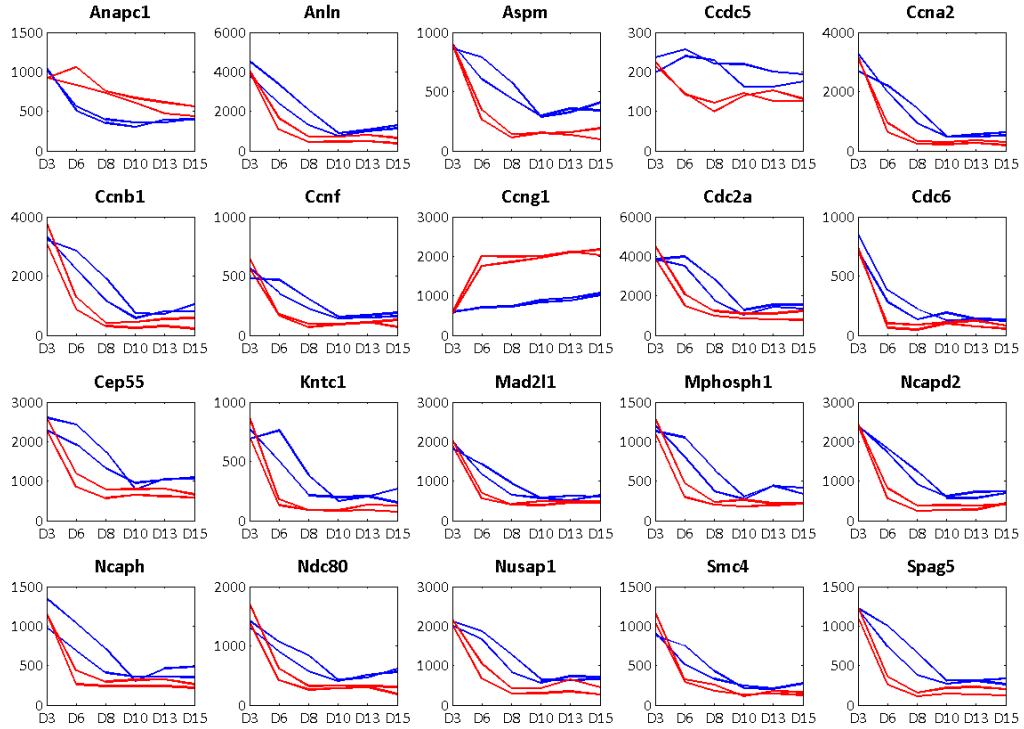


Figure 27 – Kinetically differentially expressed genes annotated as mitosis-related. Each subplot represents a different probeset on the array. The gene symbol is shown in the title, and intensity at all six time points is plotted.

5.4.1.5 Microtubule Cytoskeleton

Concomitant with changes in cell cycle genes, 23 microtubule cytoskeleton genes were identified amongst the kinetically differentially expressed genes. The modulation of microtubule cytoskeleton elements has been reported at sub-physiological temperatures (Stapulionis et al., 1997). All of the identified genes, with the exception of Stau2 (staufen homolog 2), were downregulated during the course of the culture. Furthermore, their expression was lower under reduced temperature conditions. The top five differentially expressed genes, as determined by those with the greatest Euclidean distance values, were Spag5 (Sperm associated antigen 5), Mphosph1 (M-phase phosphoprotein 1), Kntc1 (Kinetochore associated 1), Aspm (asp-like) and Nek2 (NIMA-related kinase 2). Spag5 has been associated with the mitotic spindle (Shao et al., 2001), while Kntc1 is an essential component of the mitotic checkpoint (Perez de Castro et al.,

2007). The crucial role these microtubule cytoskeletal genes play in components of the cell cycle again illustrates the profound transcriptional impact of reduced temperature cultivation on regulation and progression of the cell cycle.

5.4.1.6 Membrane-Bound Vesicle

Genes associated with membrane-bound vesicles were identified as kinetically differentially expressed under temperature shift. Rab13 was upregulated 1.6-fold during the course of the temperature-shifted culture, while being downregulated 1.4-fold in the constant temperature culture. This gene is a member of the small Rab GTPase family (Marzesco et al., 1998). These proteins are involved in vesicle trafficking and have been proposed to be required at every step of the process (Novick and Zerial, 1997). Also highly differentially expressed was Rho GTPase activating protein 22 (Arhgap22), which was upregulated 1.6-fold under temperature shift and downregulated 1.4-fold at constant temperature. ARHGAP family genes encode Rho/Rac/Cdc42-like GTPase activating proteins with RhoGAP domain (Sanz-Moreno et al., 2008).

5.4.2 Discussion

Temperature shift is a widely used technique to increase recombinant protein production. In fact, many industrial production processes use a biphasic cultivation mode and the concept has been reported on extensively in literature (Roessler et al., 1996; Furukawa and Ohsuye, 1999; Kaufmann et al., 1999; Schatz et al., 2003; Fox et al., 2004; Bollati-Fogolin et al., 2005). In this approach, an initial phase is carried out at 37°C for rapid cell growth and biomass accumulation. Subsequently, culture temperature is decreased, leading to reduced cell growth and a diversion of cellular resources to antibody production. In this study, we investigated the transcriptional impact of reduced temperature cultivation in CHO fed-batch cell culture processes. Replicate cultures were maintained at 37°C throughout, while a second pair of cultures was initially cultured at 37°C, but subsequently switched to 34°C upon reaching a viable cell concentration of 6 x 10^6 cells/mL. Time series transcriptome analysis was employed, and the microarray data obtained provides a very complete characterization of each culture's progression through

the lag, exponential, stationary and death phases. The effect of temperature shift on a number of process parameters was also characterized.

In this study, reduced temperature cultivation had a mixed effect on volumetric productivity. However, both temperature-shifted cultures showed a 27.5% increase in specific productivity. This is consistent with published results which have suggested that the effect of sub-optimal temperature cultivation on recombinant protein production is cell-line dependent (Al-Fageeh et al., 2006). For example, a study in an antibody-producing cell line cultured at 33°C saw a 1.2-fold increase in specific productivity, but no change in volumetric productivity (Yoon et al., 2003a). In contrast, cultivation of an interferon-gamma producing CHO cell line at 32°C led to increases in both specific and volumetric productivity (Fox et al., 2004). In the current study, cell growth was significantly stunted under reduced temperature cultivation. This was accompanied by a reduction in glucose and glutamine consumption at 34°C, which is consistent with a number of published reports (Sureshkumar and Mutharasan, 1991; Weidemann et al., 1994; Chuppa et al., 1997; Furukawa and Ohsuye, 1998; Fogolin et al., 2004).

A decrease in cell growth at reduced culture temperatures is a well-known phenomenon which has been observed in many different CHO cell lines (Furukawa and Ohsuye, 1999; Kaufmann et al., 1999; Yoon et al., 2003a; Yoon et al., 2003b; Fox et al., 2004). A variety of mechanisms have been postulated to explain this observation, including a possible role in the expression of cold-inducible RNA binding protein (Nishiyama et al., 1997). This particular transcript is absent from our CHO Affymetrix array, and its involvement can therefore be neither confirmed nor refuted. It has also been observed that a higher proportion of cells are in the G_0/G_1 phase at reduced temperature (Trummer et al., 2006b). Although no cell cycle analysis was conducted in this study, the observation that expression of mitotic genes is significantly reduced upon temperature-shift reinforces this observation.

Temperature shift was found to have a profound effect on the CHO transcriptome, with 514 genes identified as kinetically differentially expressed. These genes were identified using a combination of analytical techniques, including a distance-based approach and a regression-based approach. By considering the intersect of these two

different methods, one gains additional confidence in the significance of the identified genes. Using gene ontology enrichment analysis, a number of functional classes were identified as significantly enriched amongst kinetically differentially expressed genes. Foremost amongst these were cell cycle genes. As was alluded to in the previous paragraph, modulation of culture temperature has a profound effect on the cell's progression through the cell cycle, and the transcriptome results obtained help to reinforce the observations made at the process parameter level. Also significantly affected by temperature shift were microtubule cytoskeleton genes. This is consistent with a published report which investigated the transcriptional impact of reduced temperature cultivation in 3T3 cells and found a number of cytoskeleton genes to be differentially expressed (Beer et al., 2003).

Modulation of culture temperature also led to the temporal differential expression of a number of vesicle transport genes, notably those associated with membrane-bound vesicles. These results are consistent with the observations described in section 4.4, where temperature shift was combined with sodium butyrate treatment for a significant increase in antibody production. In that study, vesicle-mediated transport was identified as a crucial biological modulation in cells with increased specific recombinant protein production. The role of vesicle-mediated transport in facilitating increased recombinant protein production has also been reported by others in both CHO (Yee et al., 2008a; Yee et al., 2008b) and NS0 (Seth et al., 2007; Charaniya et al., 2008b) cells.

The time-series transcriptome analysis performed in this study represents the first reported transcriptome characterization of a CHO fed-batch culture system. This detailed transcriptional snapshot allowed us to gain biological insights not only into the cellular response to reduced temperature cultivation, but also into the transcriptional changes incurred during a typical CHO fed-batch protein production process. Indeed, it is surprising to see such a vast number of genes showing significant dynamic expression. It is commonly believed that mammalian culture systems exhibit modest transcriptional changes, especially when compared to the changes incurred in bacterial systems (Wlaschin et al., 2006). However, the results presented here suggest that gene expression

is highly dynamic throughout the course of a culture. As will be discussed in Chapter 7, this presents exciting opportunities for cell engineering applications.

5.5 TRANSCRIPTIONAL IMPACT OF PH SHIFT

5.5.1 Results

In this study, the transcriptional impact of a pH shift on CHO fed-batch cultures was investigated. Four fed-batch cultures were inoculated at 6 x 10^5 cells/mL and grown for 96 hr. At this point, the pH control setpoint of two of the four cultures was reduced by 0.15 pH units. 24 hr following the pH shift, all four cultures were subjected to a temperature shift of 3°C, and subsequently cultivated at 34°C until the end of the cultures. Microarray samples were taken at Days 3, 4, 6, 8, 10, 13 and 15 from all four cultures for microarray analysis.

5.5.1.1 Impact on Process Parameters

A number of process parameters were affected by pH shift. Of note, cumulative glucose (Figure 29), glutamine (Figure 31) and ammonia (Figure 32) profiles were significantly different upon pH shift, all showing reduced production and consumption profiles under low pH setpoint. Cultures grown at a lower pH reached a lower maximal viable cell density but sustained increased viabilities for longer as compared to the constant pH cultures (Figure 28). While pH-shifted cultures reached a lower final IgG titer (Figure 32), they also showed higher specific IgG production rates (Figure 33).

Plotting the relationship between two nutrients' cumulative consumption curves is frequently used to characterize cellular metabolism. The slopes of such curves are known as stoichiometric ratios and are characteristic for a given set of process conditions. Manipulations of process parameters can alter these stoichiometric ratios, reflecting an alteration in cellular metabolism. The process pH shift did not change the stoichiometric ratio of lactate to glucose (Figure 34), or glutamine to glucose (Figure 35). However, changes in the stoichiometric ratio of ammonia to glutamine were triggered upon pH shift, with pH-shifted cultures showing increased ammonia to glutamine ratios (Figure 36).

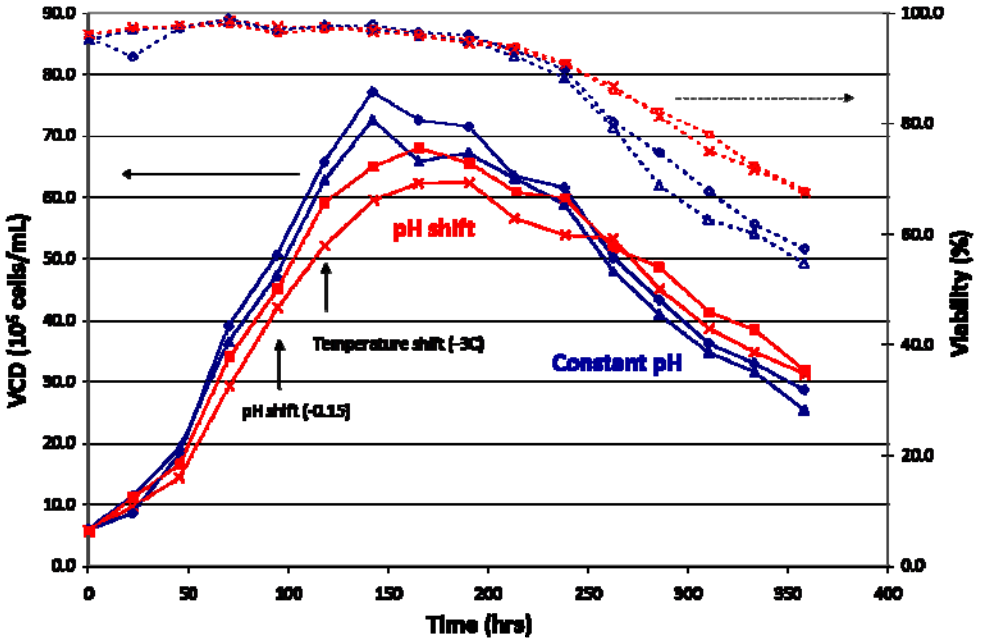


Figure 28 – Viable cell densities and cell viabilities of cultures grown at constant and pH-shifted conditions.

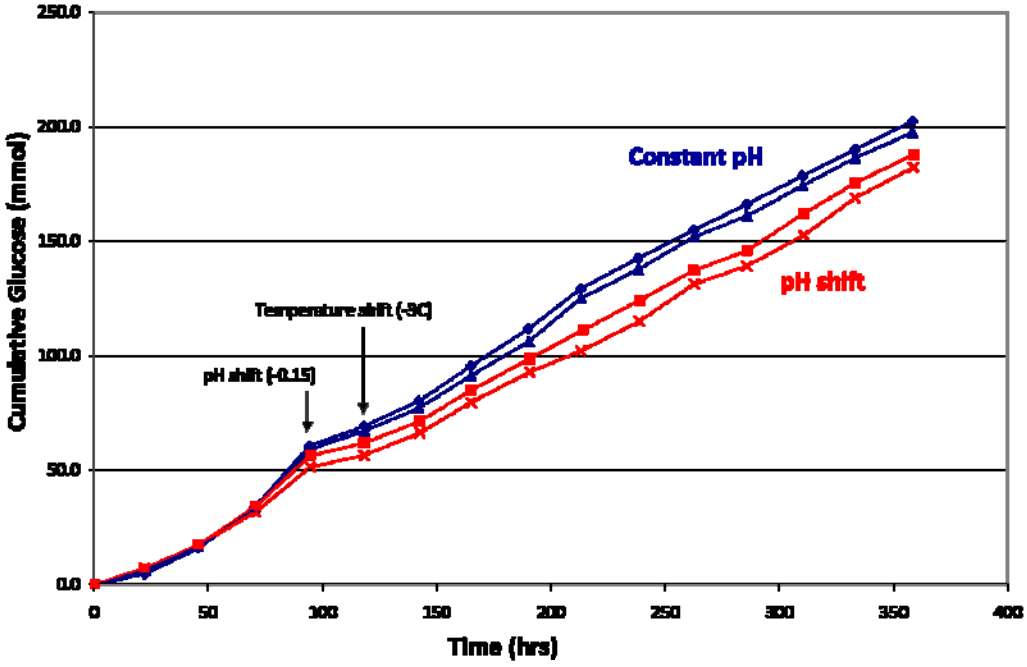


Figure 29 – Cumulative glucose consumption of cultures grown at constant and pH-shifted conditions.

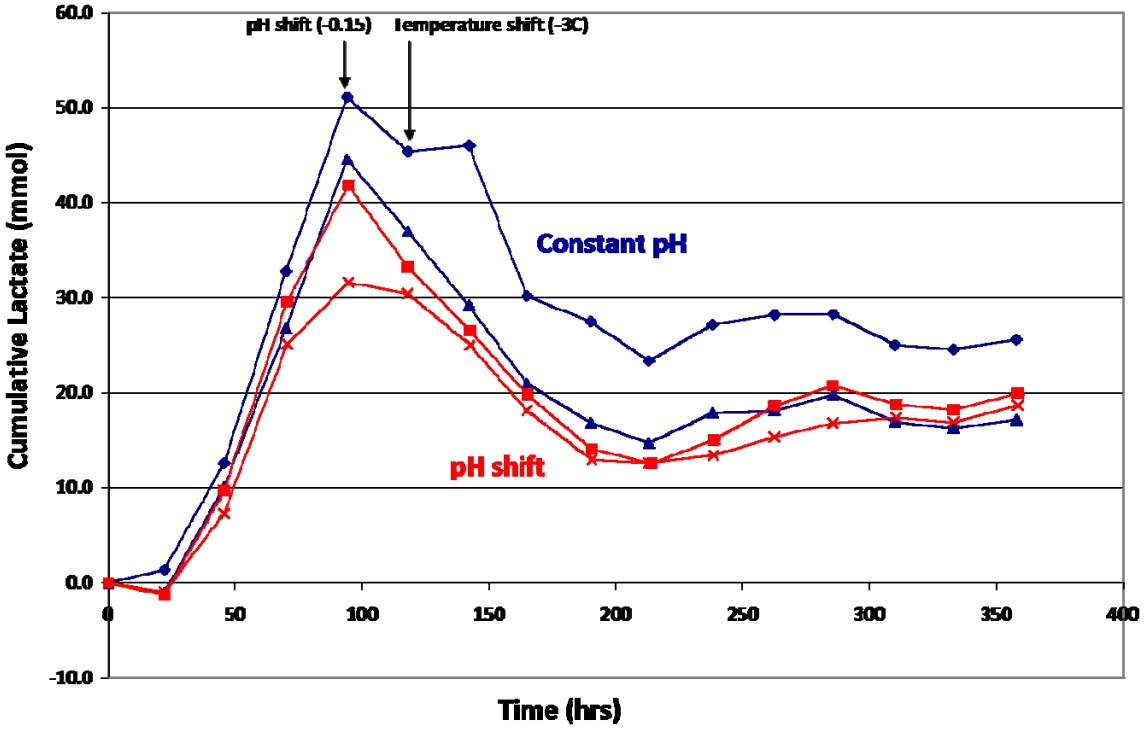


Figure 30 – Cumulative lactate production of cultures grown at constant and pH-shifted conditions.

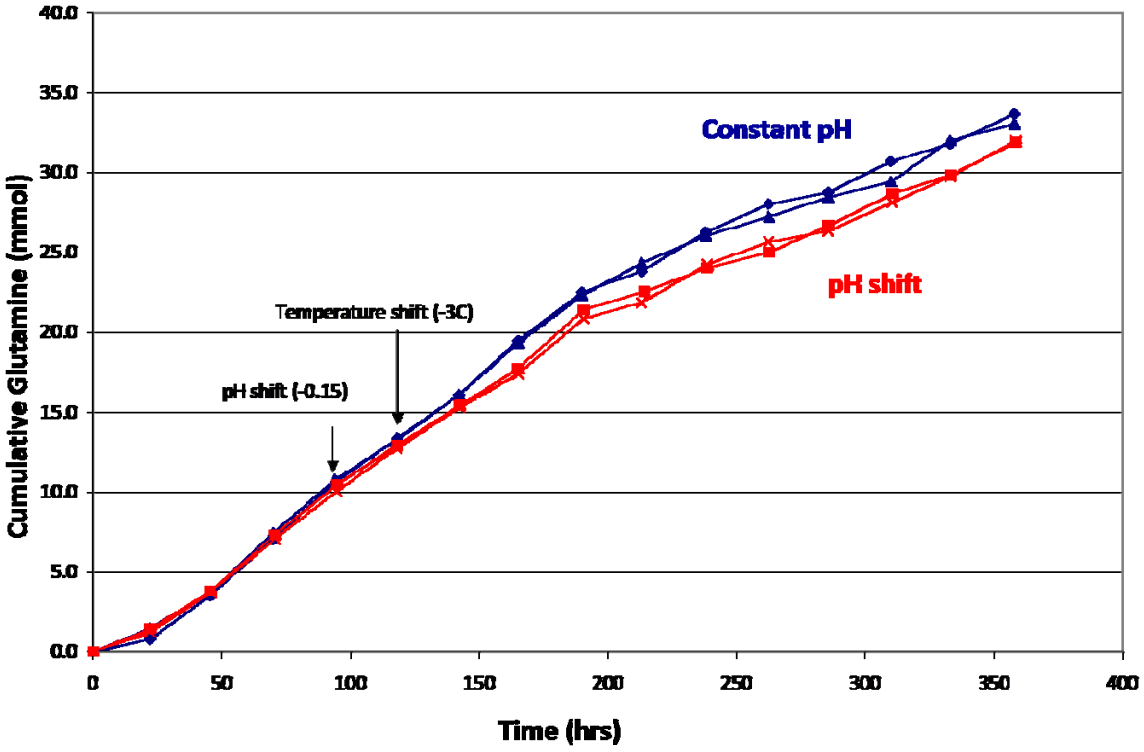


Figure 31 – Cumulative glutamine consumption of cultures grown at constant and pH-shifted conditions.

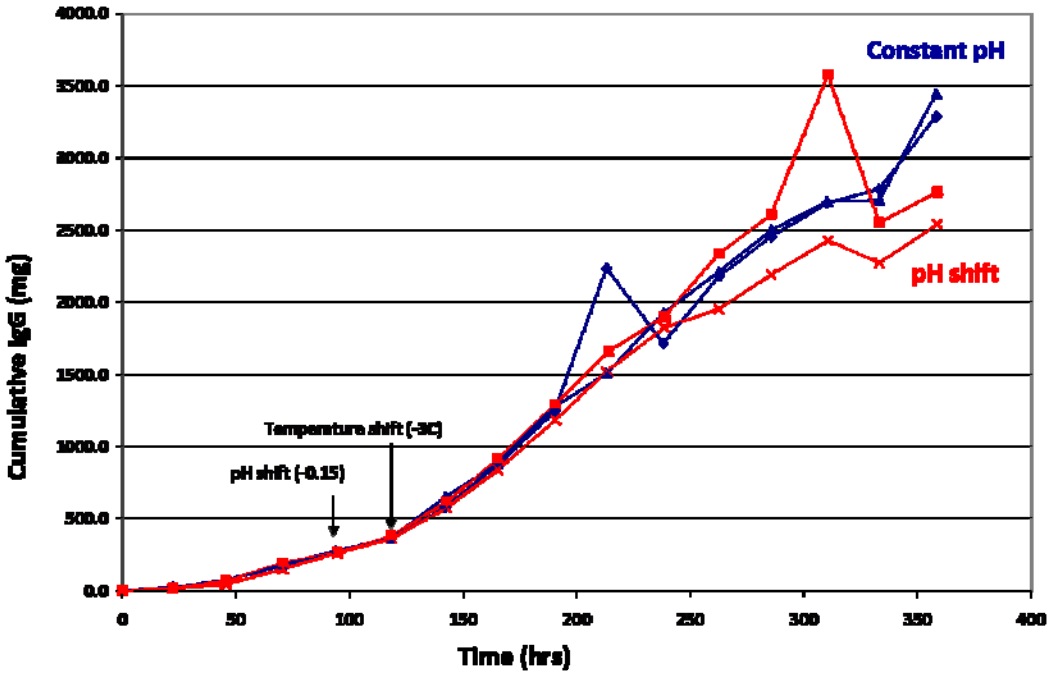


Figure 32 – Cumulative antibody production of cultures grown at constant and pH-shifted conditions.

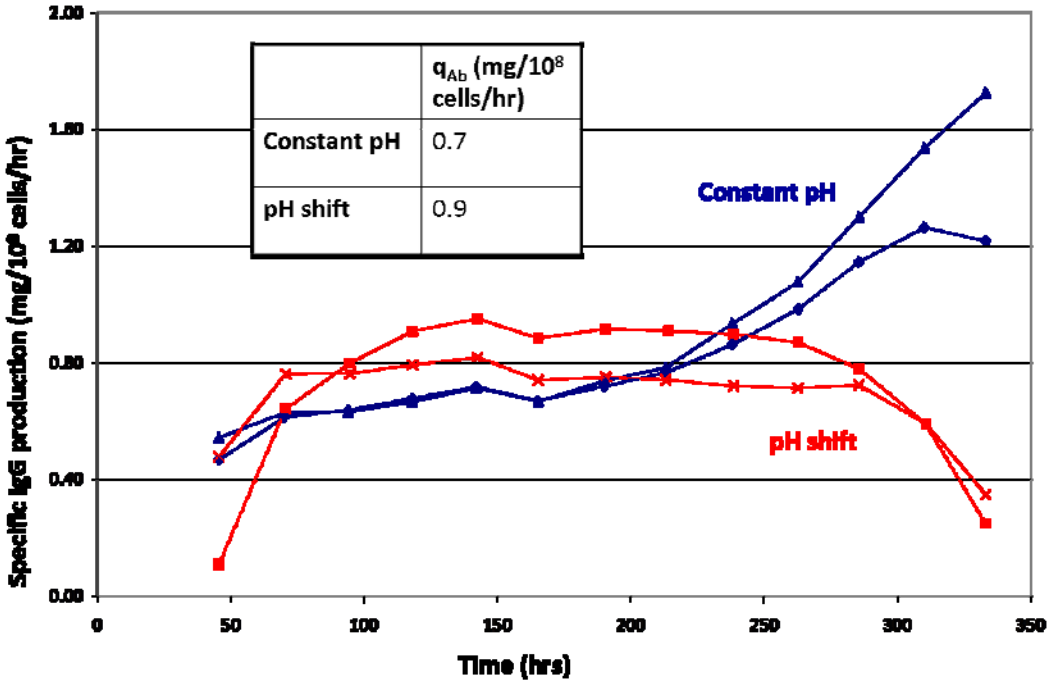


Figure 33 – Specific antibody production of cultures grown at constant and pH-shifted conditions.

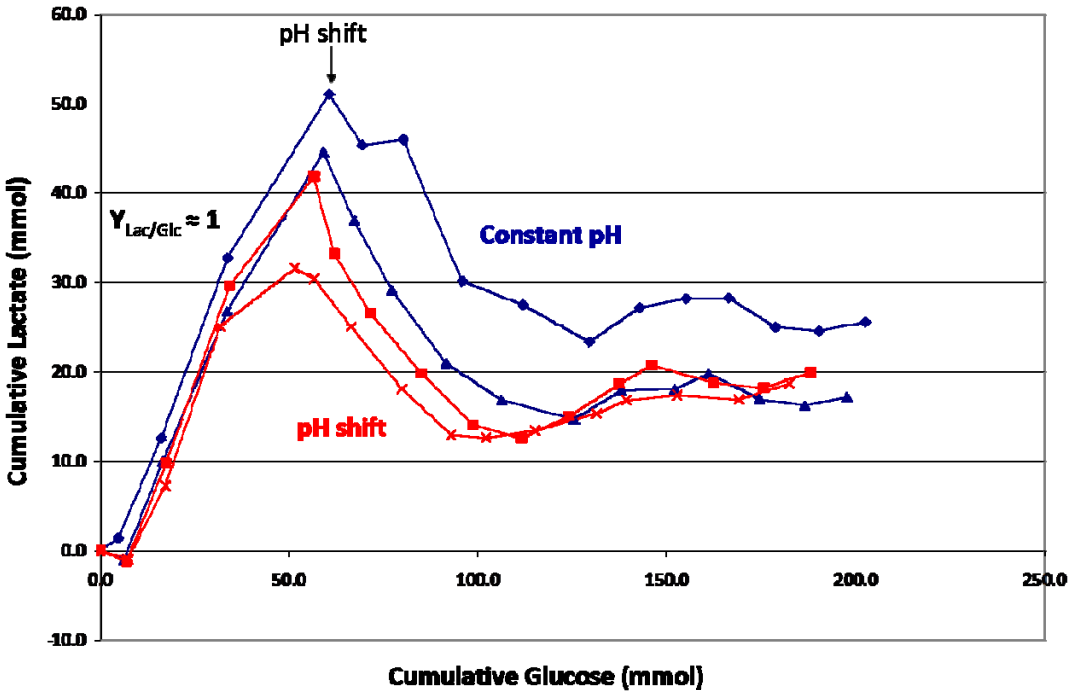


Figure 34 – $Y_{Lac/Glc}$ of cultures grown at constant and pH-shifted conditions.

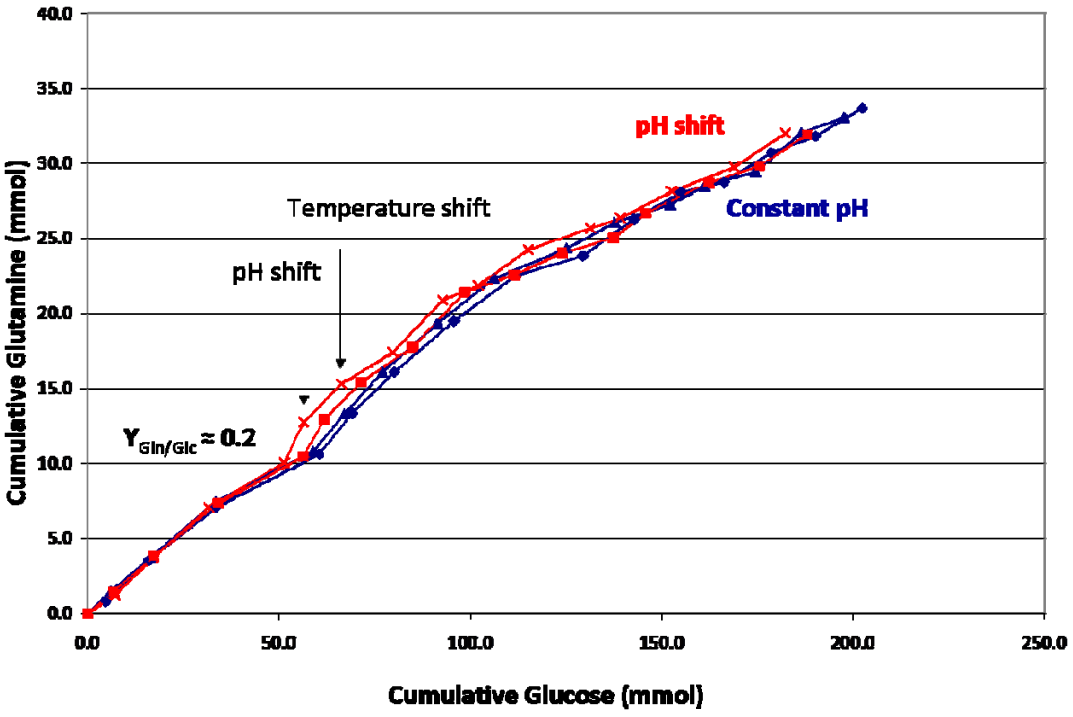


Figure 35 – $Y_{Gln/Glc}$ of cultures grown at constant and pH-shifted conditions.

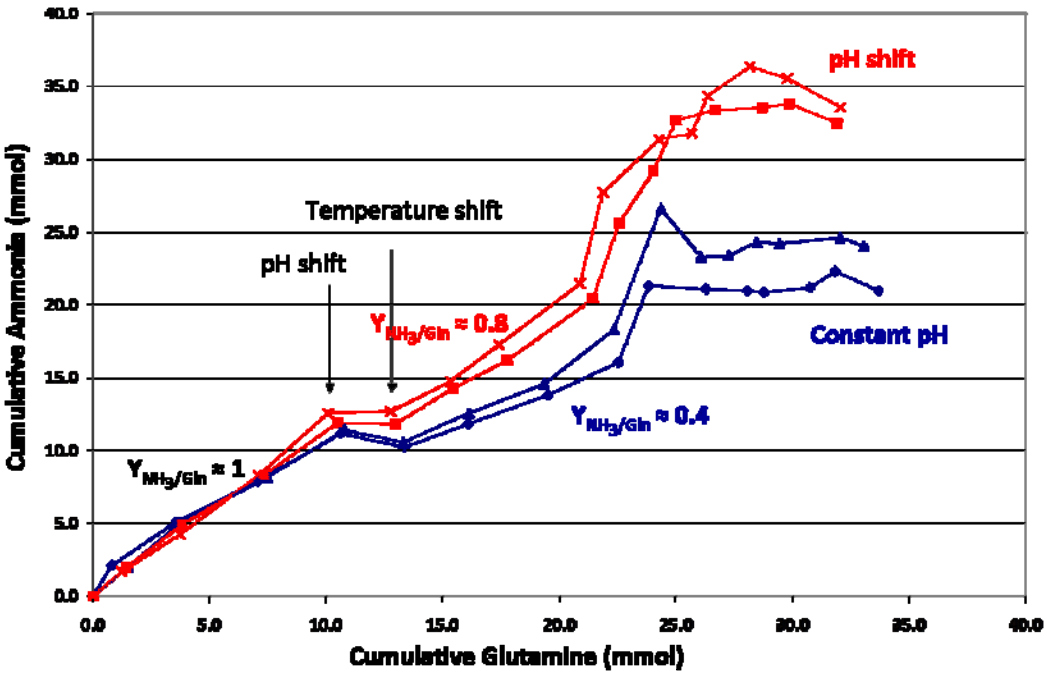


Figure 36 – $Y_{Amm/Gln}$ of cultures grown at constant and pH-shifted conditions.

5.5.1.2 Identifying Kinetically Differentially Expressed Genes

A non-linear regression-based method was used to identify kinetically differentially expressed genes between biological conditions, namely constant pH and pH shift. Using a q-value threshold of 10%, 1140 genes were identified as kinetically differentially expressed. To further restrict this set of genes, Euclidean distance was calculated for every gene between constant and pH-shifted cultures (expression profiles from biological replicates were first averaged before calculating distance). We examined the distribution of calculated Euclidean distances, and considered genes with a Euclidean distance of 2 standard deviations above the mean (Euclidean distance ≥ 1.1) as kinetically differentially expressed. This reduced the set of kinetically differentially expressed genes to 703. The top 20 genes identified through this analysis are shown in Figure 37 and a subset is further described in Table 11. It is also worthy to note that biological replicate cultures are highly reproducible in their expression profiles.

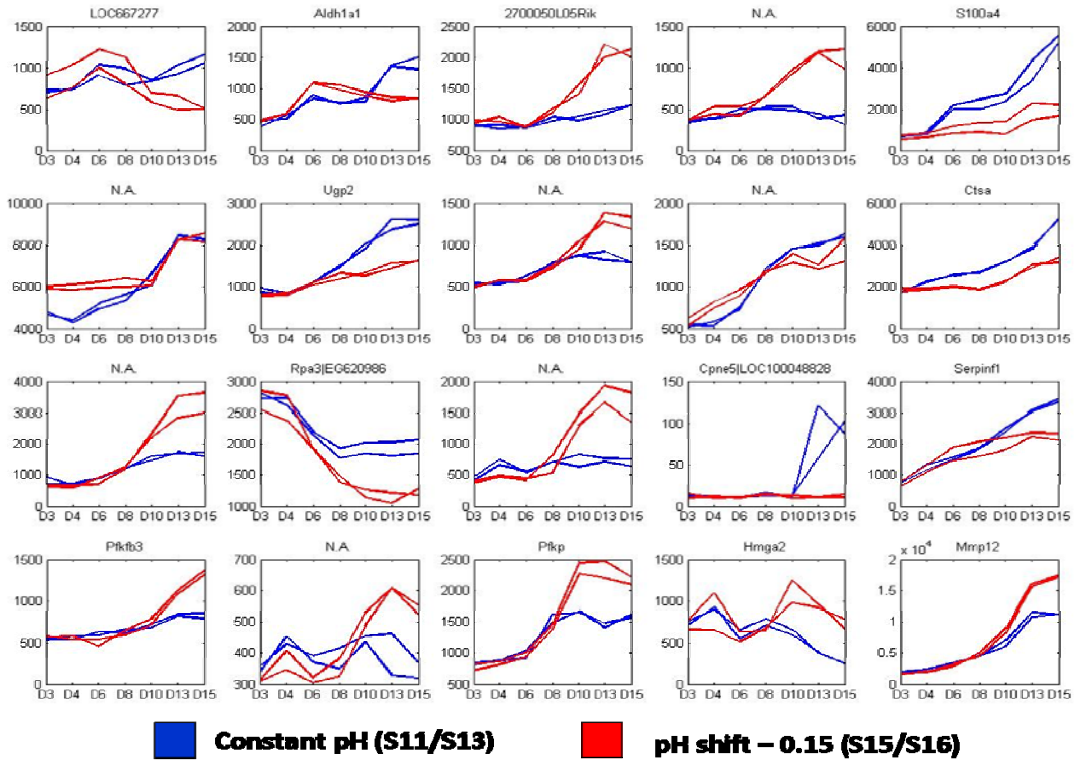


Figure 37 – Top kinetically differentially expressed genes identified through non-linear regression and further restricted using Euclidean distance cutoff (see text for description). Each subplot represents a different probeset on the array. The gene symbol is shown in the title, and intensity at all seven time points is plotted.

Table 11 – Description of top differentially expressed genes with annotation.

Gene Symbol	***Gene Description***	***Comments***
LOC667277	Complement component 1, r subcomponent	C1r is a member of the complement component 1, a part of the innate immune response. C1r is a protease, which together with another protease (C1s) forms a calcium-dependent tetramer.
Aldh1a1	Aldehyde dehydrogenase family 1, subfamily A1	Dehydrogenase enzyme which catalyzes the conversion of acetylaldehyde into acetic acid.
2700050L05Rik	Riken cDNA 2700050L05Rik gene	Also known as erythroid differentiation-related factor 1. This gene activates GATA-1, an erythroid transcription factor whose downstream targets include STAT3 and STAT4.

Gene Symbol	*Gene Description*	*Comments*
S100a4	S100 calcium binding protein A4	A Ca^{2+}-binding protein which has been implicated in cancer metastasis. It has been shown to regulate expression of matrix metalloproteinases.
Ugp2	UDP-glucose pyrophosphorylase 2	This enzyme catalyzes the reversible conversion of glucose-1-phosphate and UTP to UDP-glucose and pyrophosphatase.
Ctsa	Cathepsin A	This gene encodes a glycoprotein, known to associate with beta-galactosidase and neuraminidase, resulting in increased stability and activity.
Rpa3	Replication protein A3	Part of a heterotrimeric complex which binds ssDNA and plays a role in many aspects of nucleic acid metabolism.
Cpne5	Copine V	Ca^{2+}-dependent, phospholipid binding protein without known function. They are thought to be involved in membrane trafficking.
Serpinf1	Serine (or cysteine) peptidase inhibitor, clade F, member 1	A known inhibitor of angiogenesis which inhibits VEGF activation of AKT/PI3K and Src signaling pathways.
Pfkfb3	6-phosphofructo-2-kinase/fructose-2,6-bisphosphotase 3	A bifunctional enzyme which regulates the steady-state concentrations of fructose-2,6-bisphosphate. This isoform displays preferential phosphatase activity, leading to decreased fructose-2,6-bisphosphate levels, which reduces glycolytic flux.
Pfkp	Phosphofructokinase, platelet isoform	Glycolytic enzyme which converts fructose-6-phosphate to fructose-1,6-bisphosphate. It is allosterically inhibited by ATP, citrate, and its product. It is allosterically activated by AMP and fructose-2,6-bisphosphate.

5.5.1.3 Functional Analysis

The 703 genes which met the previously described differential expression thresholds were used as an input for functional analysis. A gene ontology enrichment analysis was performed using GenMAPP's MappFinder software. Ontological classes with an enrichment p-value ≤ 0.05 were considered significant, and are shown in Table 15. Expression profiles of a subset of cell cycle genes (Figure 38) and apoptosis genes (Figure 39) are shown. It is important to note that these functional classes encompass genes with vastly different expression profiles, both with time and across biological conditions. For example, gene members of the cell cycle ontological class are shown to both increase and decrease with time. A closer examination of the gene members of each ontological class is therefore necessary before any conclusions can be drawn.

Table 12 – Gene ontology classes found enriched in differentially expressed genes.

Gene ontology term	Number of genes DE	Number of genes in GO term	Enrichment p-value
Calcium binding	41	700	0
Extracellular matrix	24	386	0
Rho protein signal transduction	10	90	0.007
Cell cycle	46	878	0.008
Endonuclease activity	7	67	0.03
Ion transport	19	644	0.03
Regulation of apoptosis	17	292	0.05

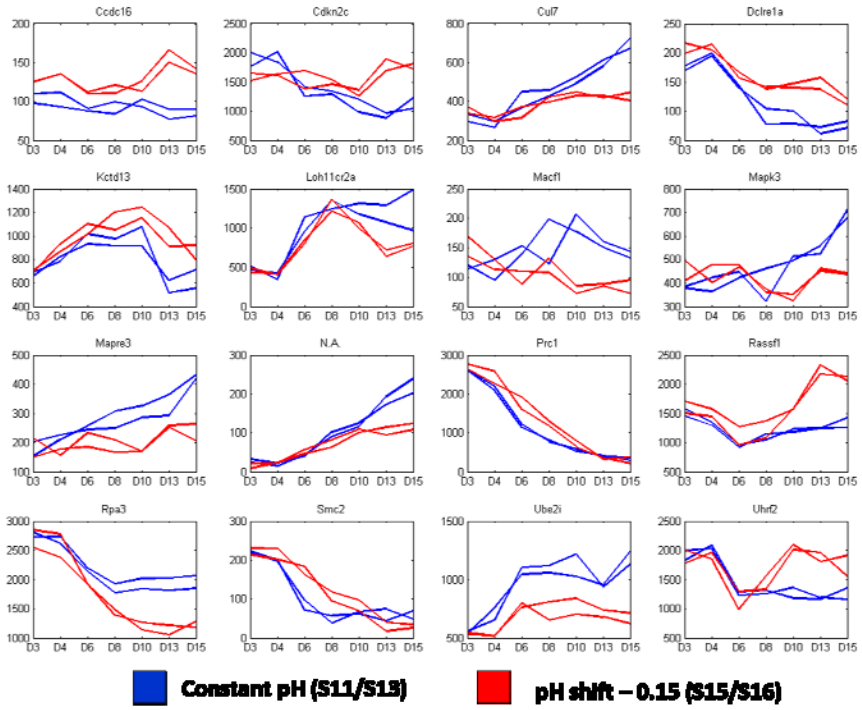


Figure 38 – Cell cycle genes.

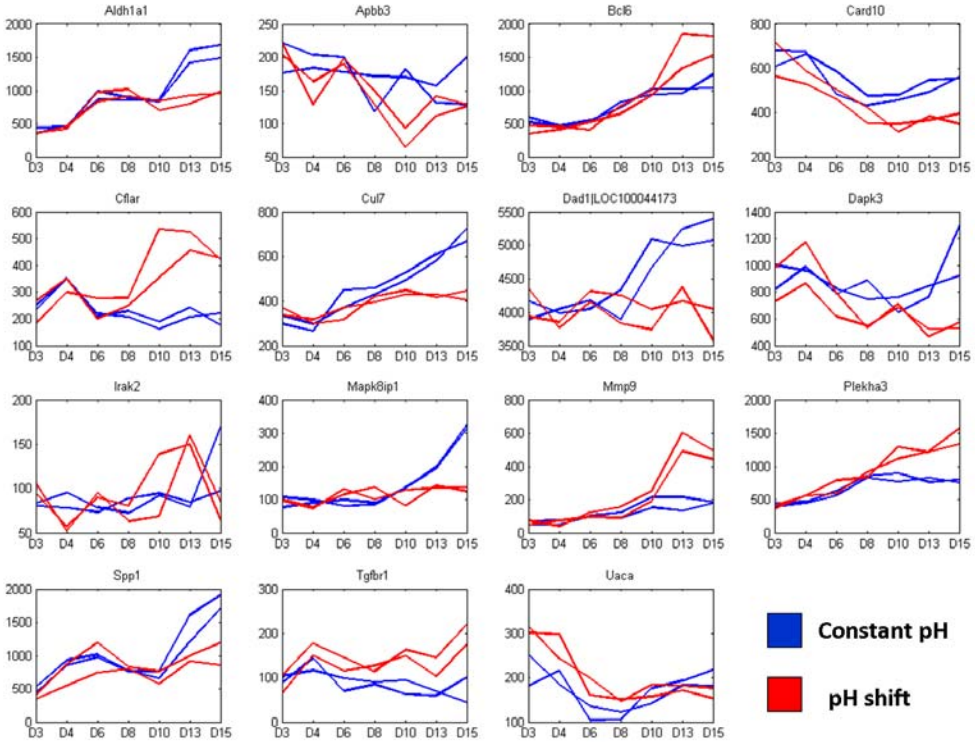


Figure 39 – Genes involved in the regulation of apoptosis.

5.5.1.4 Matrix Metalloproteinases

A number of matrix metalloproteinases were found to be differentially expressed upon pH shift (Figure 40). Matrix metalloproteinases (MMPs) are a group of peptidases involved in the degradation of extracellular matrix (ECM) components. The majority of MMPs are secreted from the cell, although a subset also exists as transmembrane proteins. MMPs are secreted as inactive proenzymes whose pro-peptide domain must be removed for activation. In addition to degrading ECM components, MMPs also cleave other proteinases, proteinase inhibitors, latent growth factors, chemotactic molecules, growth factor binding proteins, cell surface receptors and cell-cell adhesion molecules (Elenjord et al., 2008). MMP-9, MMP-10 and MMP-12 were all upregulated under pH-shifted conditions, and their expression was seen to increase only in the later stages of the culture (Day 8 onwards). MMP-9 degrades type IV and V collagens (Zeng et al., 1999), while MMP-10 degrades proteoglycans and fibronectin (Krampert et al., 2004). MMP-12 has been shown to degrade elastin (Suomela et al., 2001).

Matrix metalloproteinases are regulated by the family of tissue inhibitor of metalloproteinases (TIMPs), and their expression is also shown in Figure 40. Timp2 expression, which is interrogated by 3 probesets on the CHO array, was increased under constant pH conditions. This may explain why MMP expression was upregulated under pH-shifted conditions. An additional protein, S100A4, has been identified as a master regulator of MMPs and their inhibitors, TIMPs (Elenjord et al., 2008). This calcium-binding protein has also been shown to activate NFkB, providing a potential explanation for its regulation of MMPs. S100A4 was found to be kinetically differentially expressed under pH shift, showing increased expression under constant pH conditions. It is interesting to note that differential expression of S100A4 occurs immediately upon pH shift, while differences in the expression profiles of MMP-9, MMP-10, MMP-12 and TIMP-2 do not emerge until later in the cultures. This suggests that pH shift has an immediate effect on S100A4 expression, which then modulates gene expression changes in MMP and TIMP gene members.

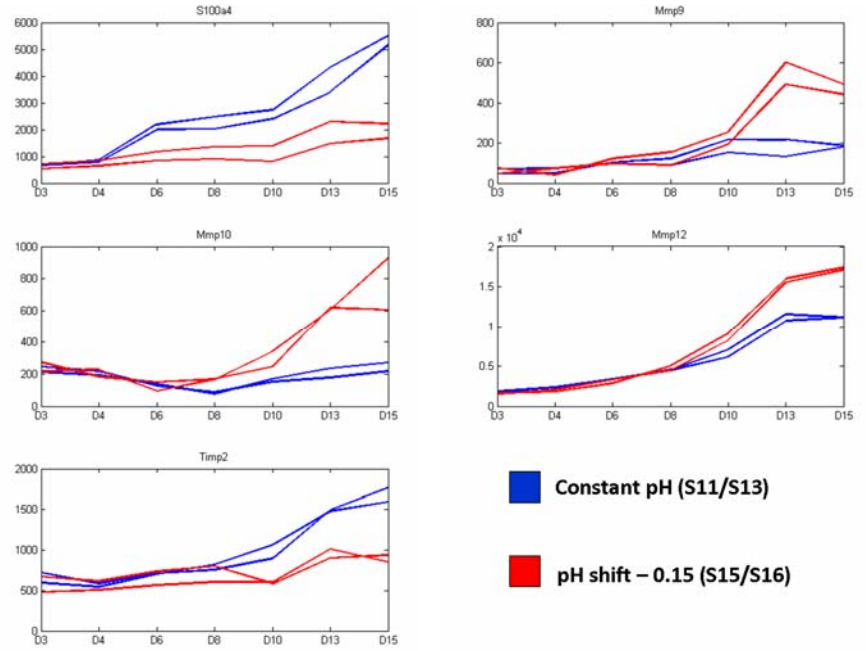


Figure 40 – Expression of differentially expressed members of the MMP and TIMP family, as well as S100A4.

5.5.1.5 Transporters

Ion transport was identified as significantly enriched amongst kinetically differentially expressed genes. Present amongst these were several subunits of the vacuolar H^+ ATPase, a highly conserved enzyme complex composed of multiple subunits. This enzyme plays a role in the acidification of multiple organelles, and is responsible for pumping protons across the plasma membrane, thus helping to maintain pH homeostasis. It is therefore not surprising to see that multiple subunits of this enzyme were affected by a pH shift in the culture. However, changes were observed across multiple subunits, and it is therefore difficult to draw a conclusion as to the precise effect of a pH shift on this enzyme.

We also examined the expression profiles of other transporters, and those affected by pH shift are shown in Figure 41. The primary substrate of each transporter (if known) is shown in Table 16. Slc25a12 is an isoform of the aspartate/glutamate transporter, one of the antiporter constituents of the malate-aspartate shuttle. This system plays a critical role in the cell, by enabling the translocation of electrons produced during glycolysis from the cytosol to the mitochondria for oxidative phosphorylation. This transporter was

slightly upregulated in the constant pH cultures. The other isoform's expression (Slc1a3) was unchanged. Its partner in the malate aspartate shuttle, the malate-alpha-ketoglutarate antiporter (Slc25a11) was also unchanged.

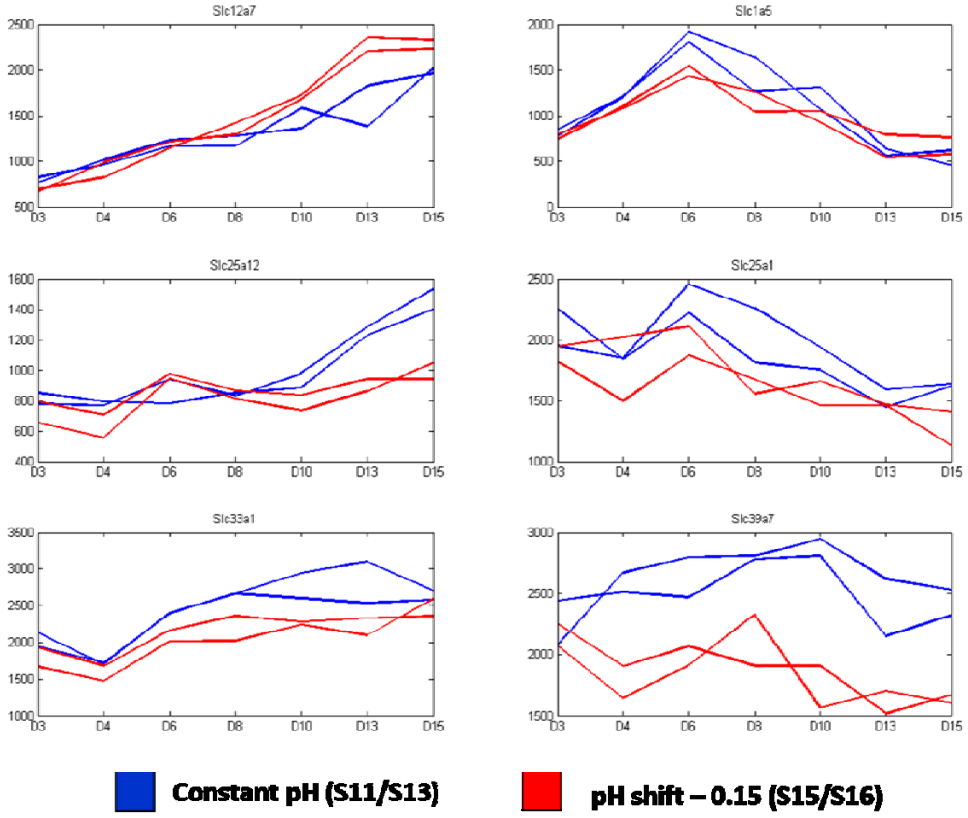


Figure 41 – Expression profiles of transporter genes found to be differentially expressed.

Table 13 – Predominant substrates and transport type of identified differentially expressed transporters.

Gene Symbol	*Predominant Substrate*	*Transport Type / Coupling Ions*
Slc12a7	Potassium, chloride	Cotransporter / K^+, Cl^-
Slc1a5	L-Ala, L-Ser, L-Thr, L-Cys, L-Gln	Cotransporter / Na^+
Slc25a12	Aspartate, glutamate	Exchanger / Aspartate, Glutamate
Slc25a1	Citrate, malate, PEP	Exchanger / Citrate – H^+, Malate
Slc33a1	Acetyl-CoA	Facilitated
Slc39a7	Manganese	Unknown

We next sought to examine whether phenotypic observations made based on process data could be explained at the transcriptome level. As discussed in Section 5.5.1.1, the shift in pH set point led to a number of observable differences in process parameters, including changes in glucose, glutamine and ammonia profiles. In an attempt to explain these observations, we took a closer look at the expression of genes involved in central metabolism.

5.5.1.6 Glycolysis and TCA cycle

Three genes involved in glycolysis and the TCA cycle were identified as kinetically differentially expressed: phosphofructokinase, platelet isoform, which was upregulated under pH-shifted cultures, lactate dehydrogenase C, which was upregulated in the control cultures and Slc25a12, which showed slight upregulation in the later stages of the control culture (Figure 42). Phosphofructokinase is a critical regulatory node in the glycolytic pathway and it is allosterically inhibited by ATP, citrate and fructose 1,6-bisphosphate, its reaction product, and allosterically activated by AMP and fructose 2,6-bisphosphate. At the gene expression level, while the muscle and liver isoforms of this enzyme have been extensively studied, little information is known about the platelet isoform. Interestingly, we observed kinetic differential expression of the C isoform of lactate dehydrogenase. We extended our comparison to include profiles from all ten temperature-shifted fed-batch runs described in this chapter, and compared LDH-C expression (Figure 43). Upon temperature-shift in constant pH conditions, we see a 2-fold upregulation of this gene, while expression remains relatively constant when temperature-shift and pH-shift are combined. Physiologically, this isoform is expressed exclusively in the testis, and has kinetic properties which favor the conversion of lactate to pyruvate (Boussouar and Benahmed, 2004). It is possible that this enzyme may play a role in the lactate consumption observed during the later stages of these cultures. The role of this enzyme will be further examined in Chapter 6. Overall, the few gene expression changes observed in the glycolytic pathway make it difficult to draw further conclusions about its modulation under pH shift.

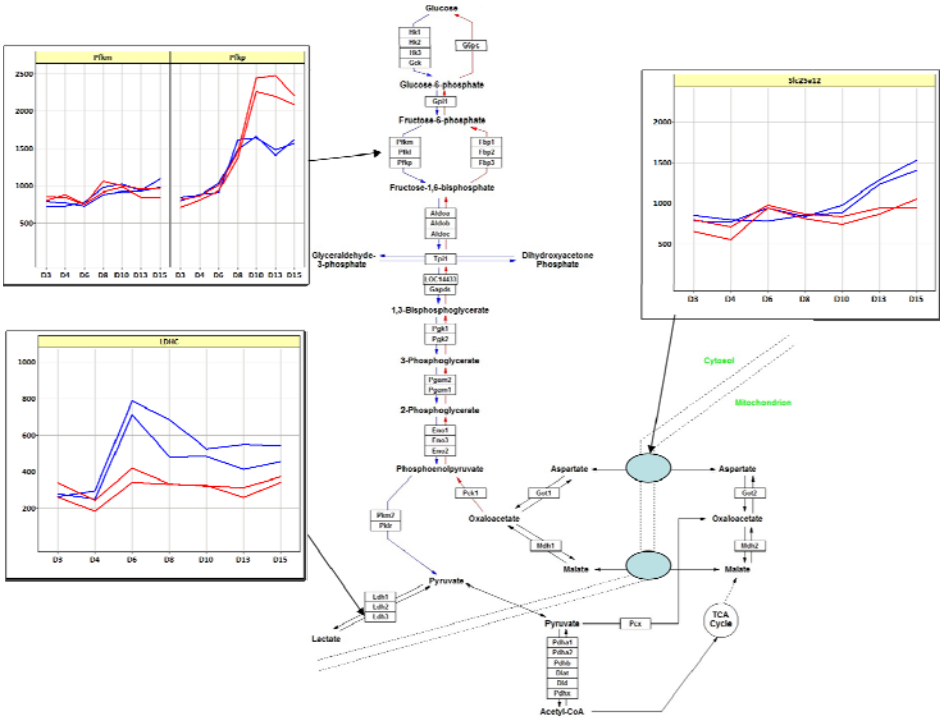


Figure 42 – Expression profiles of differentially expressed genes in the glycolytic pathway. Blue profiles represent control cultures, while red profiles represent pH-shifted cultures.

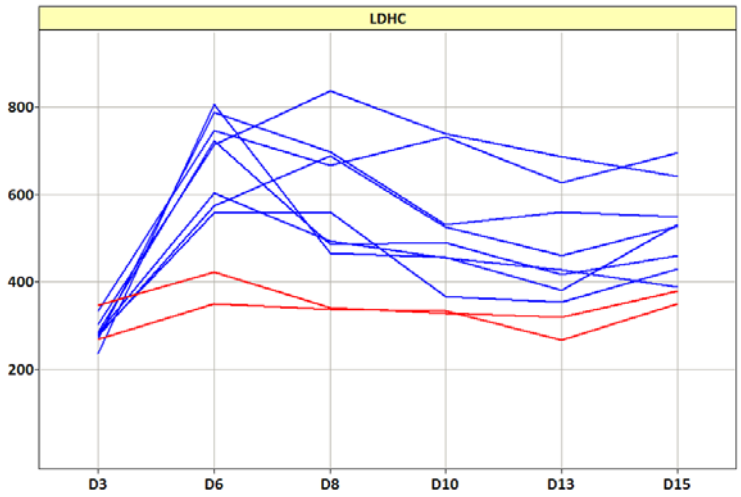


Figure 43 – Expression profile of LDH-C in 10 fed-batch cultures with temperature shift. Blue profiles represent constant pH cultures, while red profiles represent pH-shifted cultures.

5.5.1.7 Pentose Phosphate Pathway

The pentose phosphate pathway provides the cell with reducing equivalents, in the form of NADPH, and serves to synthesize 5-carbon sugars used in the synthesis of nucleotides and nucleic acids. Glycolytic intermediates can enter the pathway at many junctions, although the oxidative phase, which generates NADPH, begins with the irreversible dehydrogenation of glucose-6-phosphate. As can be seen in Figure 44, pH shift led to the upregulation of glucose-6-phosphate dehydrogenase, the enzyme which catalyzes the rate-limiting step in the oxidative phase of the pathway. This upregulation could have diverted part of the glycolytic flux to the pentose phosphate pathway, and could possibly explain the reduced cumulative glucose profiles observed upon pH shift. Another means for the cell to produce NADPH is through the action of malic enzyme 1, a cytosolic enzyme which catalyzes the reversible oxidative decarboxylation of malate to pyruvate (Pongratz et al., 2007). A mitochondrial isozyme, ME2, also exists, though it is not present on the CHO version 2 arrays. As can be seen in Figure 45, the expression of malic enzyme 1 was upregulated in the control culture in the later stages of the culture. This could perhaps provide the cell with additional reducing power, which may be lacking as a result of a possibly reduced pentose phosphate pathway flux in the control culture.

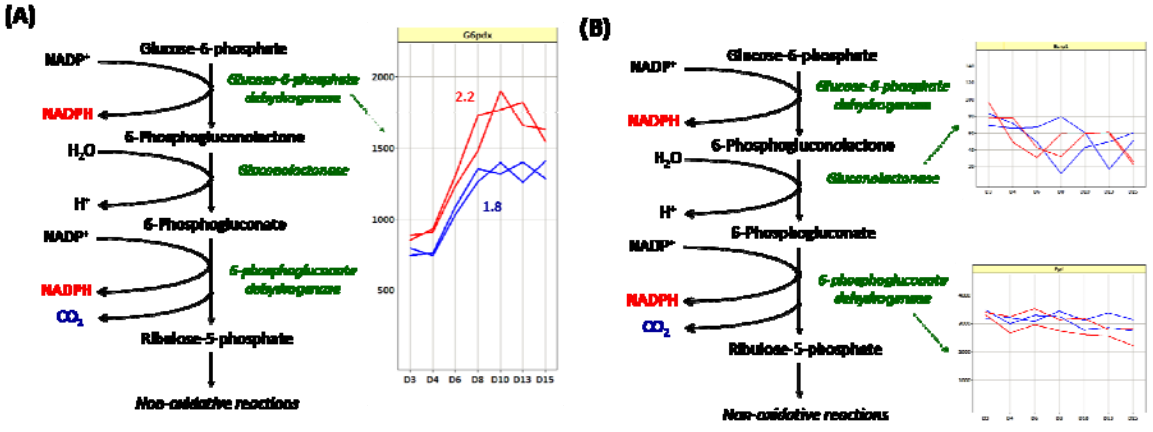


Figure 44 – Expression profiles of enzymes involved in the oxidative phase of the pentose phosphate pathway. (a) Expression profile of glucose-6-phosphate dehydrogenase, the limiting reaction in the pathway. The numbers in the graph represent the average fold change at each condition when comparing Day 15 to Day 3 expression. (b) Expression profiles of the remaining enzymes in the

oxidative phase of the pentose phosphate pathway. Blue profiles represent control cultures, while red profiles represent pH-shifted cultures.

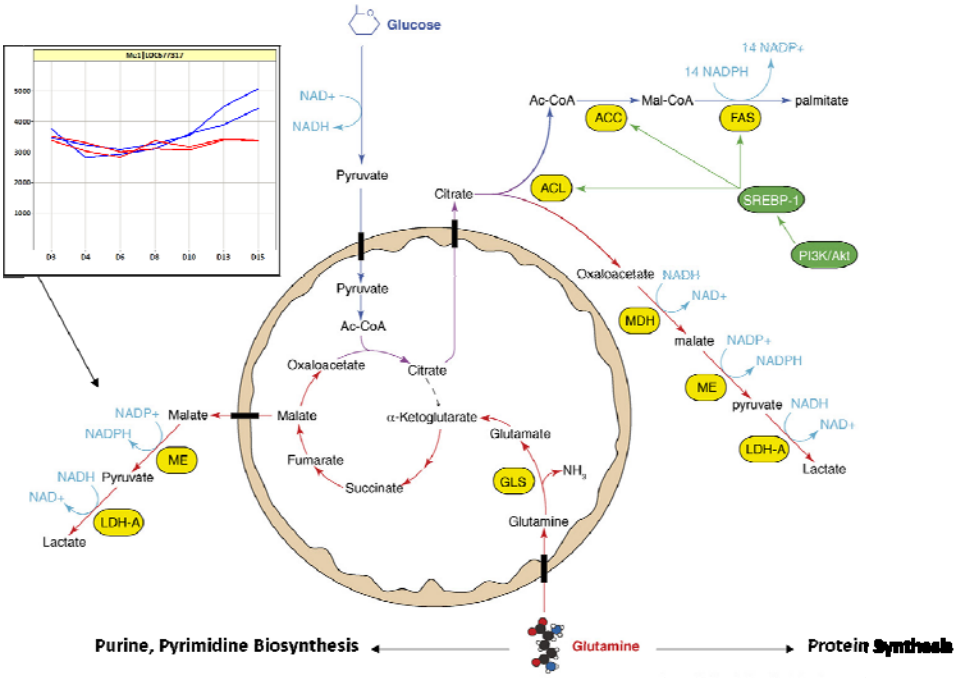


Figure 45 – An overview of glutamine metabolism, showing the expression of malic enzyme 2. Figure modified from (Deberardinis et al., 2008). Blue profiles represent control cultures, while red profiles represent pH-shifted cultures.

5.5.1.8 Glutaminolysis

Mammalian cells in culture are known to use glutamine to meet the anaplerotic demands of growth, in addition to its use as a precursor to purine and pyrimidine biosynthesis, as well as protein synthesis (Deberardinis et al., 2008). Glutamine enters the TCA cycle via catabolism in the mitochondrial matrix to yield glutamate and releasing an ammonium molecule. This reaction is catalyzed by a phosphate-dependent glutaminase. Glutamate is then converted to alpha-ketoglutarate, a reaction that can either be catalyzed by glutamate dehydrogenase, leading to the release of a second ammonia molecule, or an aminotransferase. Glutamine can also be synthesized by the cells in a reaction catalyzed by glutamine synthetase. The expression of these genes is shown in the figure below. No significant differential expression between conditions was observed, and can therefore not account for the increased ammonia production observed under pH-shifted conditions.

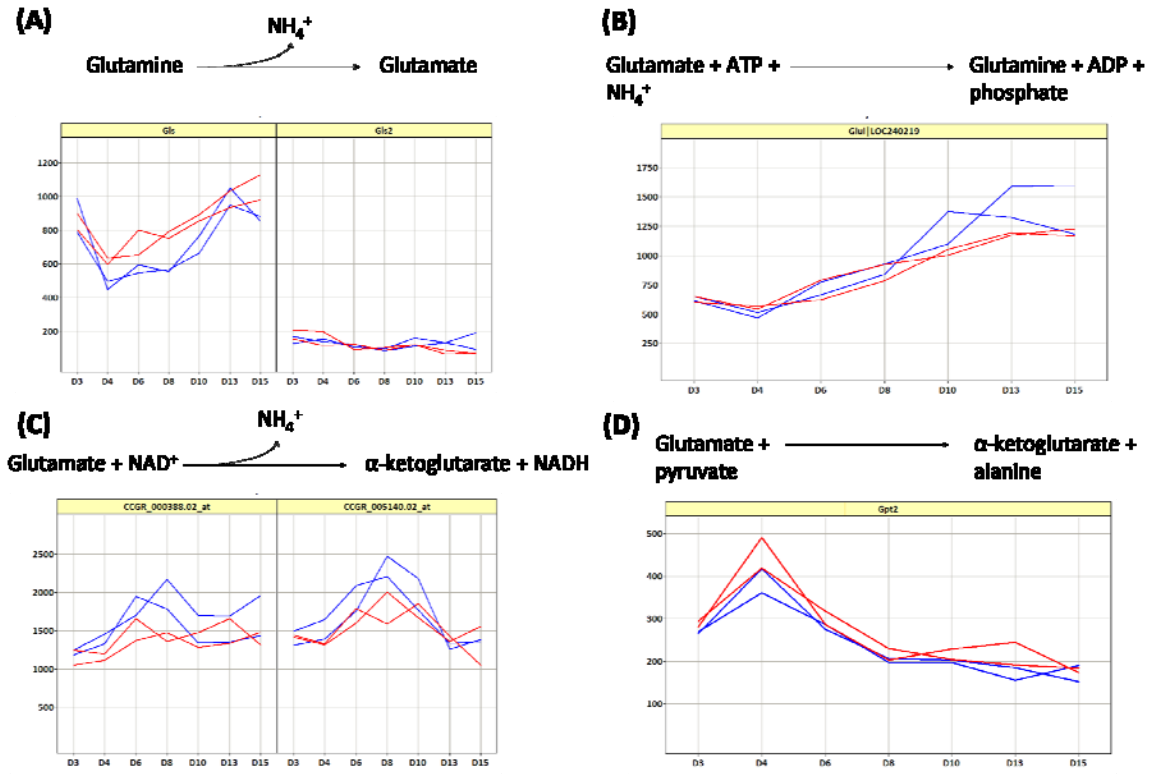


Figure 46 – Expression of genes involved in glutamine and glutamate metabolism. Blue profiles represent control cultures, while red profiles represent pH-shifted cultures. (A) Expression of glutaminase and glutaminase 2, which catalyzes the conversion of glutamine to glutamate, releasing ammonia. (B) Expression of glutamine synthetase, which catalyzes the conversion of glutamate to glutamine. (C) Expression of glutamate dehydrogenase. Each panel represents a different probeset on the array. (D) Expression of glutamic pyruvate transaminase.

5.5.2 Discussion

Manipulation of culture pH is a commonly used approach in cultivation of production mammalian cell lines. It has been observed that productivity can vary with pH and that the optimum pH for maximal productivity is cell line dependent (Yoon et al., 2004). However, the underlying mechanisms for this observation are poorly understood. In this study, we used transcriptome analysis to characterize the expression profiles of four CHO fed-batch cultures: two maintained under constant pH and two subjected to a reduction in pH setpoint of 0.15 units. This process change had a modest impact on a number of process parameters. A very modest decrease in achievable IgG titers was observed in pH-shifted cultures, although this was accompanied by a slight increase in specific IgG production. A more immediate impact was observed on cumulative glucose

and glutamine profiles. Perhaps most striking was a substantial change in glutamine to ammonia stoichiometric ratios.

A number of studies have been published which investigate the impact on process parameters of varying cultivation pH in several cell lines. An increase in specific productivity of erythropoietin with decreasing pH has been reported in CHO, accompanied with reduced specific glucose and glutamine consumption rates with decreasing pH (Yoon et al., 2004). Similar decreases in specific glucose consumption rates with decreasing pH were also reported in NS0 cells, but this was accompanied by a decrease in specific antibody production, reinforcing the notion that optimal pH for maximal productivity is cell line dependent (Osman et al., 2001). However, studies in both CHO and NS0 have shown that decreasing pH results in lower maximal viable cell densities and increased cell viabilities (Osman et al., 2001; Yoon et al., 2004; Trummer et al., 2006c). This is consistent with what we observed, where decreasing pH reduced achievable maximal cell densities by 13% and increased cell viabilities in the later stages of the culture by 17%.

At the transcript level, we found 703 genes to be kinetically differentially expressed. Through functional analysis, we identified a number of gene ontology terms as significantly enriched, including cell cycle genes, extracellular matrix components and ion transporters. We compared the results of our analysis with published reports investigating the global transcriptional response to pH shift. Bumke et al studied gene expression changes in human fibroblasts following a pH shift from 7.5 to 6.7 (Bumke et al., 2003). Although this study is different from this one in terms of experimental design, similar functional classes were identified amongst kinetically differentially expressed genes, namely cell cycle regulation and extracellular matrix components.

Matrix metalloproteinases have been the subject of intense study in the field of cancer biology. Specifically, it is known that tumors have an acidic extracellular pH (pH_e) and it has been observed that expression of certain MMPs is induced as a result of this acidic pH_e. Since MMPs are involved in the proteolytic degradation of ECM components, their expression is closely associated with tumor metastasis and angiogenesis. Kato et al showed that an acidic pH_e induces MMP-9 expression via

activation of phospholipase D (PLD) leading to NFkB activation. Furthermore, analysis of the promoter region of MMP-9 showed an NFkB binding site. They concluded that induction of MMP-9 expression was the result of NFkB activation (Kato et al., 2005). The observation that this class of gene is impacted by pH conditions is noteworthy, and should be examined further. It is possible that given their proteolytic nature, the expression of MMPs could destabilize a secreted antibody product. Consequently, this class of enzymes could represent interesting cell engineering targets to increase product stability.

It has been speculated that at increased pH, the activity of glycolytic enzymes could be elevated or changes in the membrane potential could allow for increased glucose transport into the cell (Ozturk and Palsson, 1991), accounting for reported increases in specific glucose consumption and specific lactate production (Osman et al., 2001; Yoon et al., 2004; Trummer et al., 2006c). While changes in specific lactate production were not observed in this study during the exponential growth phase, we did observe a decrease in specific glucose consumption rate with decreasing pH. The theory proposed above could possibly account for this observation, although transcriptional information cannot be used to support this claim. We did, however, observe increased expression of the first enzyme in the oxidative phase of the pentose phosphate pathway upon pH shift. This could signal an increased flux through the pentose phosphate pathway upon pH shift, leading to reduced specific glucose consumption.

At a transcriptional level, we did not observe any changes in the expression of enzymes involved in the conversion of glutamine to glutamate, and glutamate to alpha-ketoglutarate, both of which generate ammonia. This could therefore not account for the increased stoichiometric ratios of ammonia to glutamine production we observed. An increase in $Y_{Amm/Gln}$ at decreased pH has been reported by others (Yoon et al., 2004; Trummer et al., 2006b), however, no mechanism has been proposed to explain this observation.

The results presented here represent the first report of the transcriptional impact of modulating a fed-batch process pH setpoint in a recombinant protein-producing mammalian cell line. Through microarray analysis, we were able to establish some links

between transcript expression and corresponding phenotype at the process parameter level. However, given the fact that enzyme activity is well known to be modulated by environmental pH, additional experiments at the proteome and metabolome level are required to further understand the impact process pH has on mammalian cell culture processes.

5.6 TRANSCRIPTIONAL IMPACT OF HYDROLYSATE LOT SOURCE

5.6.1 Results

In this study, we investigated the transcriptional impact of cultivating CHO fed-batch cultures in growth medium prepared using different hydrolysate lots. Replicate fed-batch cultures were grown in medium prepared using one of two hydrolysate lots (Lot 23 and Lot 67). All cultures were inoculated at 6 x 10^5 cells/mL and sampled daily for process parameter measurements. All four cultures were subjected to a temperature shift of 3°C at Day 3. Samples were taken at Days 3, 6, 8, 10, 13 and 15 from all four cultures for microarray analysis.

5.6.1.1 Impact on Process Parameters

The cultivation of cells in different hydrolysate lots had an effect on a number of process parameters, including viable cell density (Figure 47), antibody titer (Figure 48), as well as cumulative glucose (Figure 49) and lactate (Figure 50) profiles. Growth in medium prepared using Lot 23 resulted in higher maximal viable cell densities, as well as elevated product titers. The difference in product titer between the best- and worst-producing cultures was approximately 20%, representing a significant difference. Furthermore, cultures grown in Lot 23 showed increased specific antibody production with a value of 0.97 mg/10^8 cell/hr, as compared to 0.79 mg/10^8 cell/hr in cultures grown in Lot 67, an improvement of almost 23%. Cultures grown in Lot 23 also showed higher cumulative glucose consumption, coupled with increased lactate consumption. Cultures grown in Lot 67 also showed increased osmolality from Day 8 onwards (Figure 51), probably a result of increased base addition necessary to maintain a constant pH.

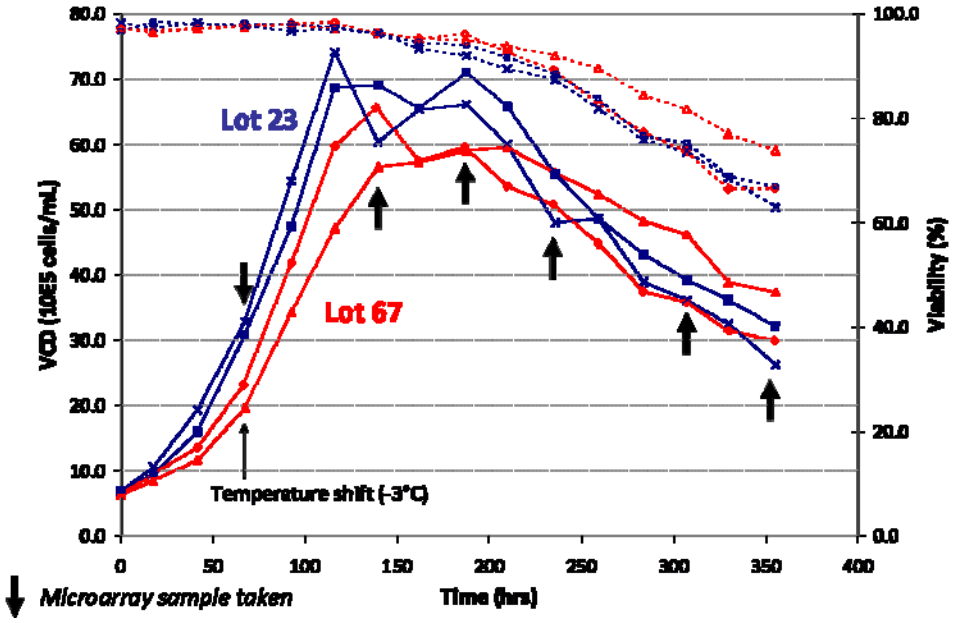


Figure 47 – Viable cell densities (VCD) and percent viability of cultures grown in different hydrolysate lots. Cultures grown in Lot 67 of the hydrolysate are shown in red, and cultures grown in Lot 23 are shown in blue. Samples taken for microarray analysis are indicated with black arrows.

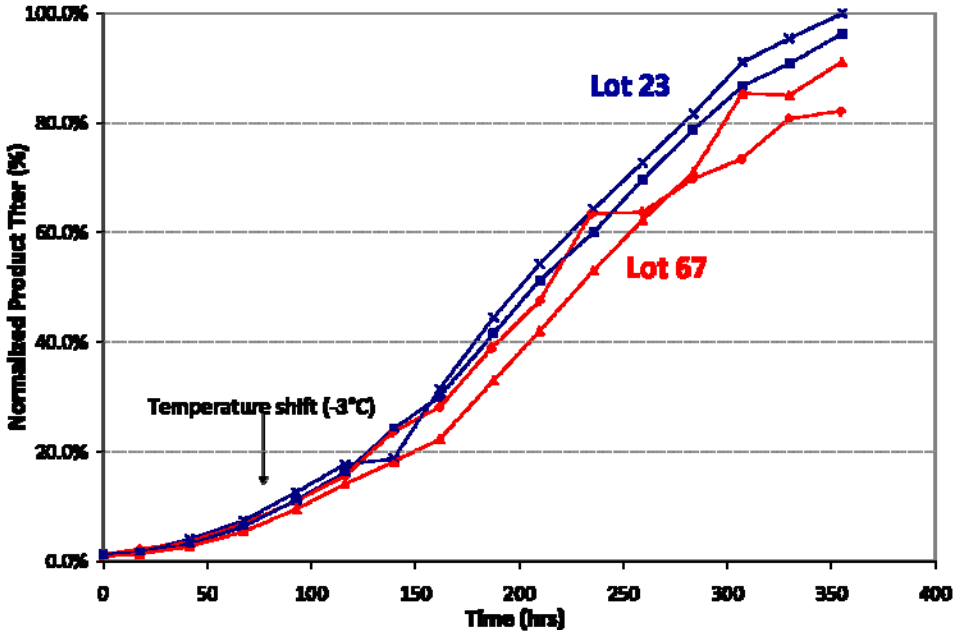


Figure 48 – Antibody titers of cultures grown in different hydrolysate lots. Titers are normalized to the highest value attained. Cultures grown in Lot 67 of the hydrolysate are shown in red, and cultures grown in Lot 23 are shown in blue.

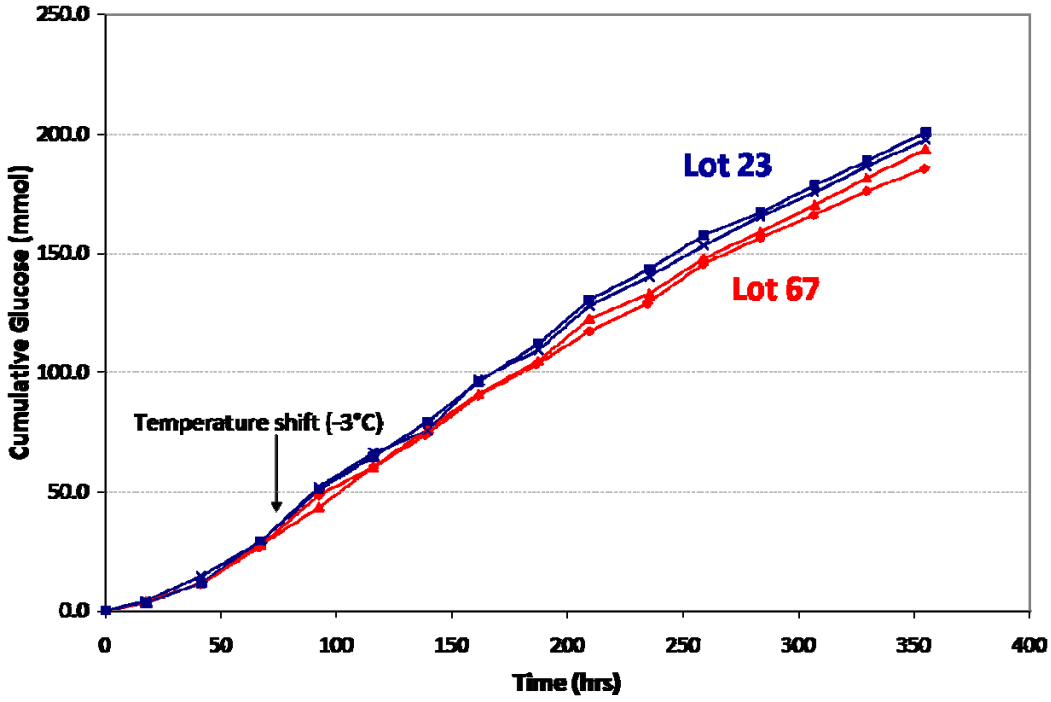


Figure 49 – Cumulative glucose profiles of cultures grown in different hydrolysate lots. Cultures grown in Lot 67 of the hydrolysate are shown in red, and cultures grown in Lot 23 are shown in blue.

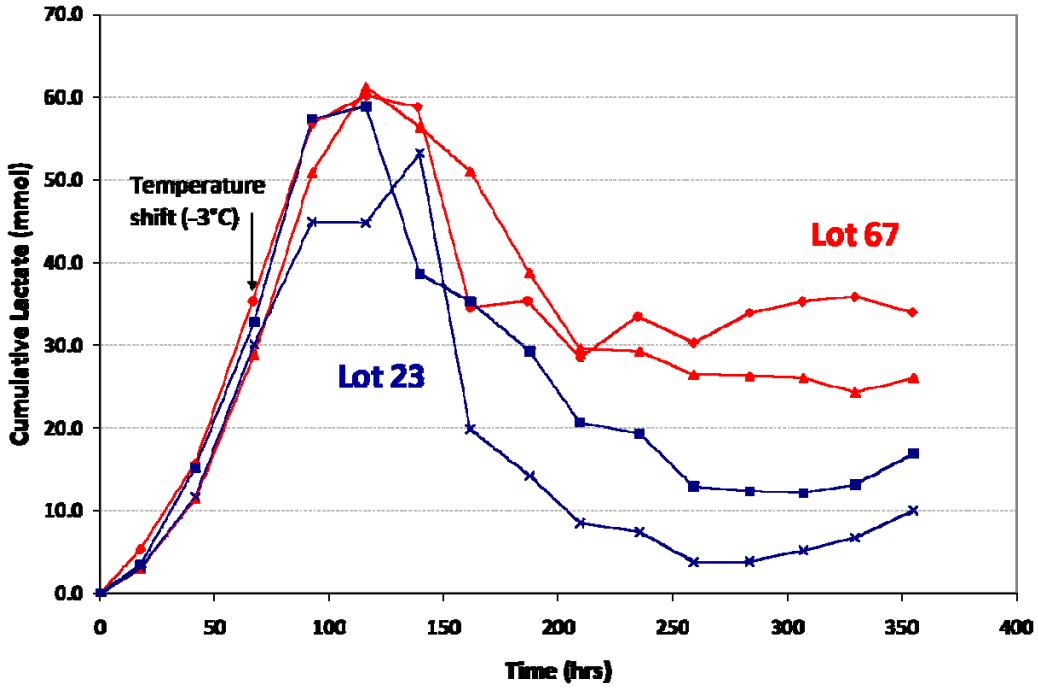


Figure 50 – Cumulative lactate profiles of cultures grown in different hydrolysate lots. Cultures grown in Lot 67 of the hydrolysate are shown in red, and cultures grown in Lot 23 are shown in blue.

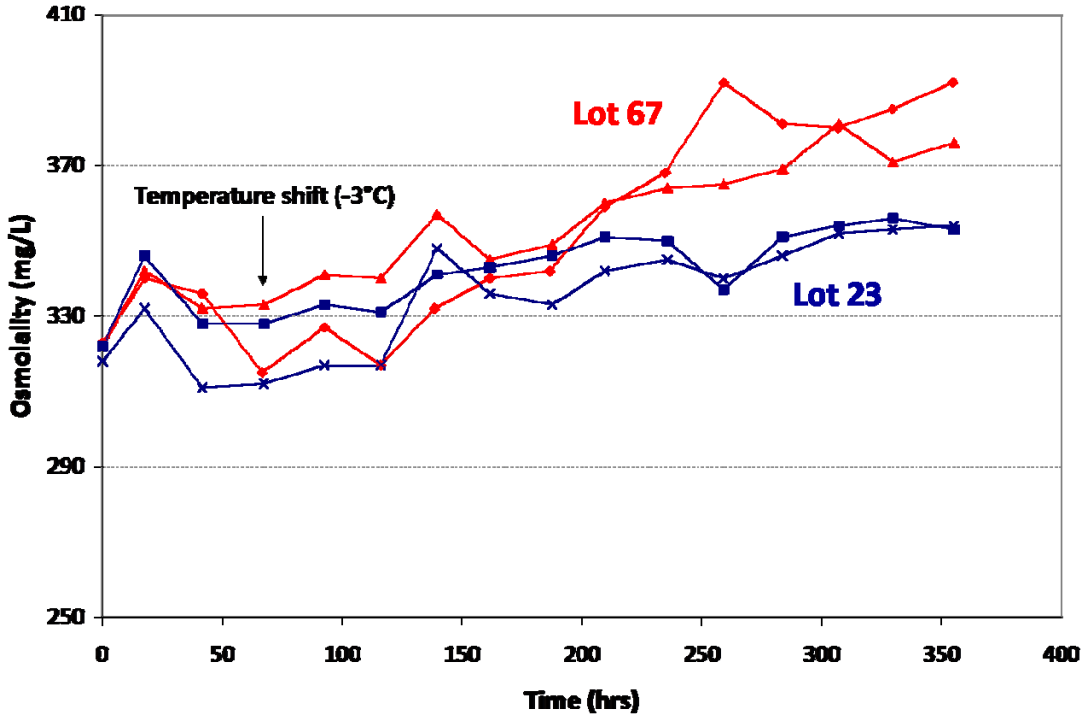


Figure 51 – Osmolality profiles of cultures grown in different hydrolysate lots. Cultures grown in Lot 67 of the hydrolysate are shown in red, and cultures grown in Lot 23 are shown in blue.

5.6.1.2 Identifying Kinetically Differentially Expressed Genes

The same non-linear regression-based method was again used to identify kinetically differentially expressed genes between different component lots (Storey et al., 2005). 1140 genes were identified as kinetically differentially expressed at a q-value threshold of 0.05. We further restricted this set of kinetically differentially expressed genes by calculating Euclidean distances for every gene between each component lot (gene intensities from replicate cultures were first averaged before calculating distances). We imposed a criterion of Euclidean distance ≥ 0.9, which represents two standard deviations above the mean. This reduced the set of kinetically differentially expressed genes to 538. The top 20 genes identified through this analysis are shown in Figure 52 and a subset is further described in Table 14.

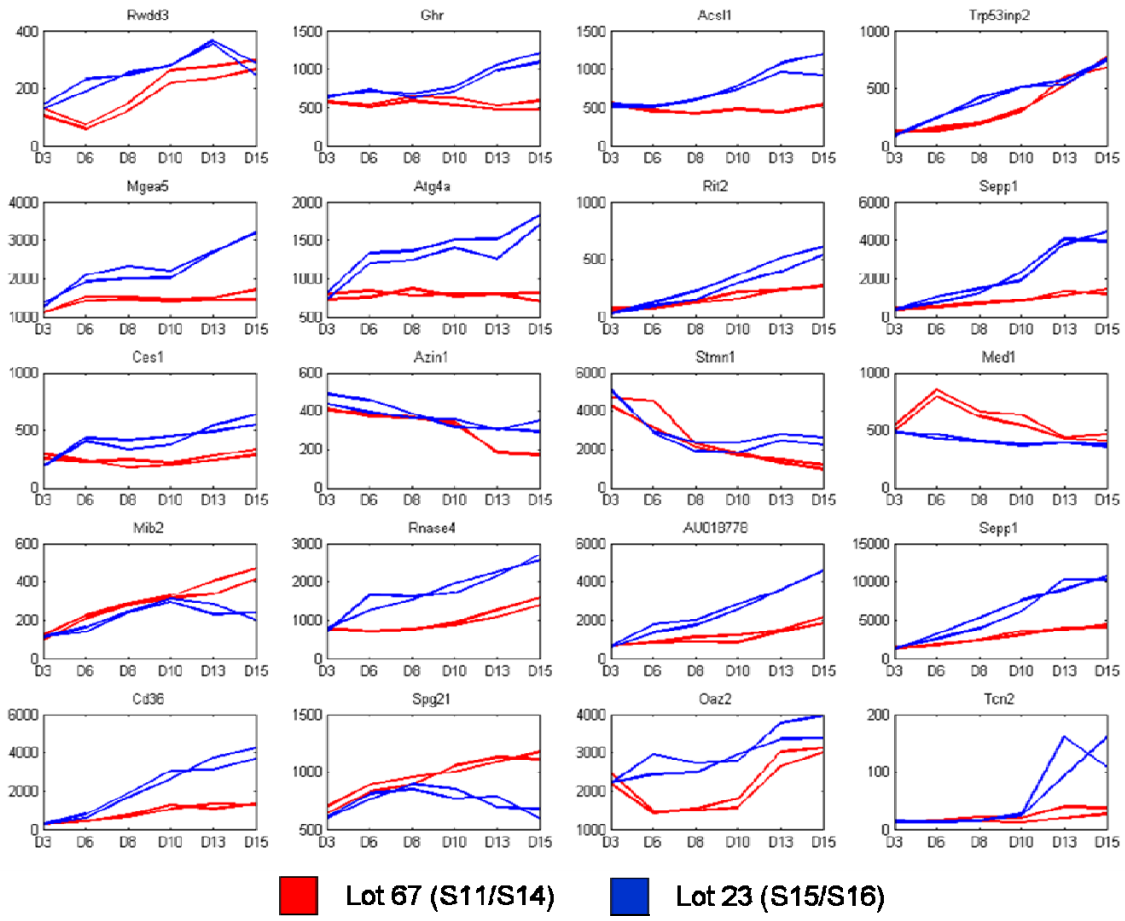


Figure 52 – Top kinetically differentially expressed genes identified through non-linear regression and further restricted using Euclidean distance (see text for description). Each subplot represents a different probeset on the array. The gene symbol is shown in the title, and intensity at all six time points is plotted. Cultures grown in Lot 67 are shown in red, while cultures grown in Lot 23 are shown in blue.

Table 14 – Description of top kinetically differentially expressed genes.

Gene Symbol	***Gene Description***	***Comments***
Rwdd3	RWD domain containing 3	RWD domain-containing proteins have been shown to play a role in sumoylation, a post-translational modification with many functions, including protein stability and transcription regulation.
Ghr	Growth hormone receptor	Transmembrane receptor for growth hormone. Binding leads to receptor dimerization and activation of signal transduction pathways leading to growth.

Gene Symbol	*Gene Description*	*Comments*
Acsl1	Acyl-CoA synthetase long-chain family member 1	Member of the long-chain fatty-acid CoA ligase family. Converts long-chain free fatty acids into fatty acyl-CoA esters.
Trp53inp2	P53 inducible nuclear protein 2	DR1 has been shown to regulate expression of this gene. The Trp53inp2 promoter region has been shown to have a putative CREB-1 binding site.
Mgea5	Meningioma expressed antigen 5	Also known as O-GlcNAcase: a glycosidase that removes O-linked N-acetylglucosamine modifications.
Atg4a	Autophagy-related cysteine endopeptidase 2	Autophagy is the process by which endogenous proteins and damaged organelles are destroyed intracellularly. Atg4a is known to activate Gabarpl1 through cleavage. Gabarpl1 was also upregulated in Lot 23.
Rit2	Ras-like without CAAX2	Belongs to the Ras superfamily of small GTPases.
Sepp1	Selenoprotein P, Plasma 1	Sepp1 has been implicated to act as an antioxidant in the extracellular space. Smarca4 as been shown to be involved in Sepp1 expression, and it is also differentially expressed in this dataset.
Ces1	Carboxylesterase 1	This enzyme hydrolyzes long chain fatty acid esters and thioesters.
Azin1	Antizyme inhibitor 1	Ornithine decarboxylase (ODC) catalyzes the conversion of ornithine to putrescine (the first and seemingly rate-limiting step in polyamine biosynthesis). Ornithine decarboxylase antizymes regulate this process by binding to and inhibiting ODC. This protein binds to ODC antizymes and stabilizes ODC, thus inhibiting antizyme-mediated ODC degradation.
Stmn1	Stathmin 1	The encoded protein is involved in the regulation of the microtubule filament system by destabilizing microtubules.

5.6.1.3 Functional Analysis

Kinetically differentially expressed genes were used as an input for functional analysis. A gene ontology enrichment analysis was performed using GenMAPP's MappFinder software. Ontological classes with an enrichment p-value of ≤ 0.05 were considered significant, and are shown in Table 15. As can be seen in Table 15, several

functional classes were identified as significantly enriched amongst kinetically differentially expressed genes. These are discussed in the sections below.

Table 15 – Gene ontology classes found enriched in differentially expressed genes.

Gene ontology term	Number of genes DE	Number of genes in GO term	Enrichment p-value
Cytoskeleton	32	923	0.007
Signal transducer activity	46	1580	0.03
Cell cycle	34	878	0.001
Polyamine metabolism	4	78	0.002
Lipid metabolism	27	519	0.05

5.6.1.4 Lipid metabolism

A number of lipid metabolism genes were identified as kinetically differentially expressed upon cultivation in a different hydrolysate lot. Their profiles are shown in Figure 53 and it is interesting to note that the most predominant expression profile observed amongst these genes is an increase in expression in Lot 23 cultures during the later stages of the culture. This may suggest that active lipid metabolism plays a role in increased productivity. Amongst the differentially expressed genes, Acaa2, Acadvl and Hadha are involved in mitochondrial long chain fatty acid oxidation, and all were upregulated in Lot 23 cultures from Day 6 onwards. Fatty acid beta-oxidation breaks down long-chain fatty acids to generate NADH, $FADH_2$ and Acetyl-CoA, which can subsequently enter the TCA cycle and be used for energy generation. Lpl, or lipoprotein lipase, was very drastically upregulated in Lot 23 cultures, going from an initial intensity of 6,000 to a maximum of more than 11,000, while in Lot 67 cultures, intensities peaked at approximately 9,000. This enzyme hydrolyzes triglycerides into free fatty acids and is responsible for providing a large portion of free fatty acids to the cell (Zechner, 1997).

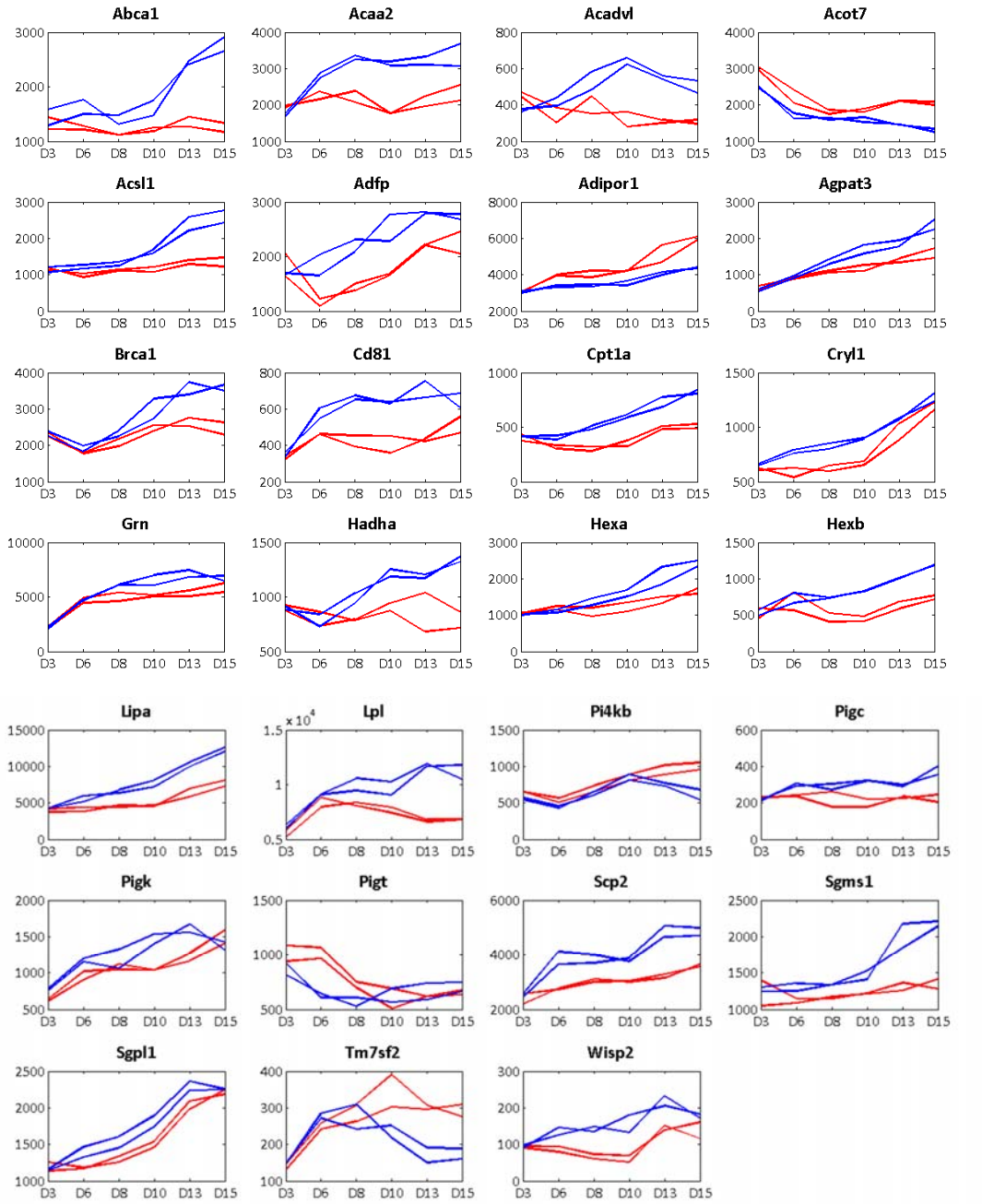


Figure 53 – Expression profiles of lipid metabolism genes identified as kinetically differentially expressed. Each subplot represents a different probeset on the array. The gene symbol is shown in the title, and intensity at all six time points is plotted. Cultures grown in Lot 67 are shown in red, while cultures grown in Lot 23 are shown in blue.

5.6.1.5 Polyamine metabolism

Functional analysis identified polyamine metabolism as one of the pathways affected by growth in different hydrolysate lots. Polyamines, which include spermine, spermidine, and their precursor, putrescine, are essential growth factors for eukaryotic cells (Gerner and Meyskens, 2004; Casero and Marton, 2007) and their metabolism is often dysregulated in cancers (Casero and Marton, 2007). The central reactions of polyamine metabolism are shown in Figure 54.

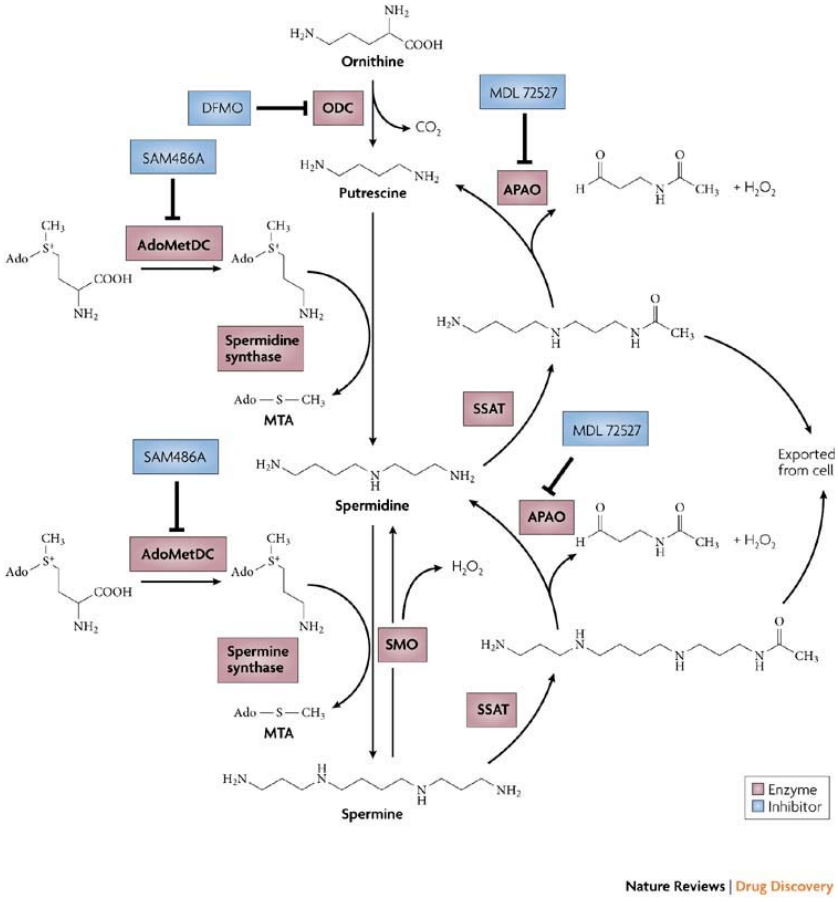


Figure 54 – Overview of polyamine metabolism. Figure reproduced from (Casero and Marton, 2007).

Ornithine decarboxylase, the enzyme which catalyzes the rate-limiting conversion of ornithine to putrescine, was upregulated in cultures grown in Lot 67 (Figure 55). Ornithine decarboxylase is inhibited by a family of antizymes, and they, in turn, are inhibited by a family of genes known as antizyme inhibitors. Collectively, the expression

profiles observed in this study point to a modulation of ornithine decarboxylase activity under growth in different hydrolysate lots (Figure 55).

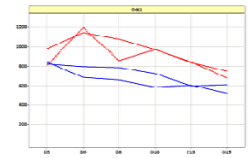
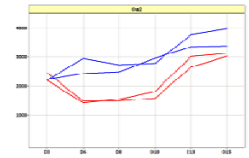
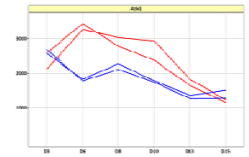
Gene Symbol	Description	Comments	Expression
Odc1	Ornithine decarboxylase 1	Catalyzes the conversion of ornithine to puterscine (the rate-limiting step of polyamine metabolism)	
Oaz2	Antizyme 2	This antizyme inhibits ornithine decarboxylase activity	
Azin1	Antizyme inhibitor 1	This enzyme inhibits the activity of Odc antizymes, thus leading to increased Odc activity	

Figure 55 – Expression of genes involved in polyamine metabolism. Cultures grown in Lot 23 are shown in blue and cultures grown in Lot 67 are shown in red.

5.6.1.6 Cell cycle

Cell cycle genes were identified as significantly enriched amongst kinetically differentially expressed genes. Interestingly, a closer examination of their gene expression profiles in different cultures revealed that their expression was different in one particular culture, Reactor S14 grown in Lot 67. Furthermore, this culture showed sustained viability during the later stages of the culture, as compared to the other three cultures (Figure 47). As shown in Figure 56, the expression of a number of cell cycle genes in Reactor S14 was higher as compared to the other three cultures. This could perhaps account for the difference in viability which was observed in this culture.

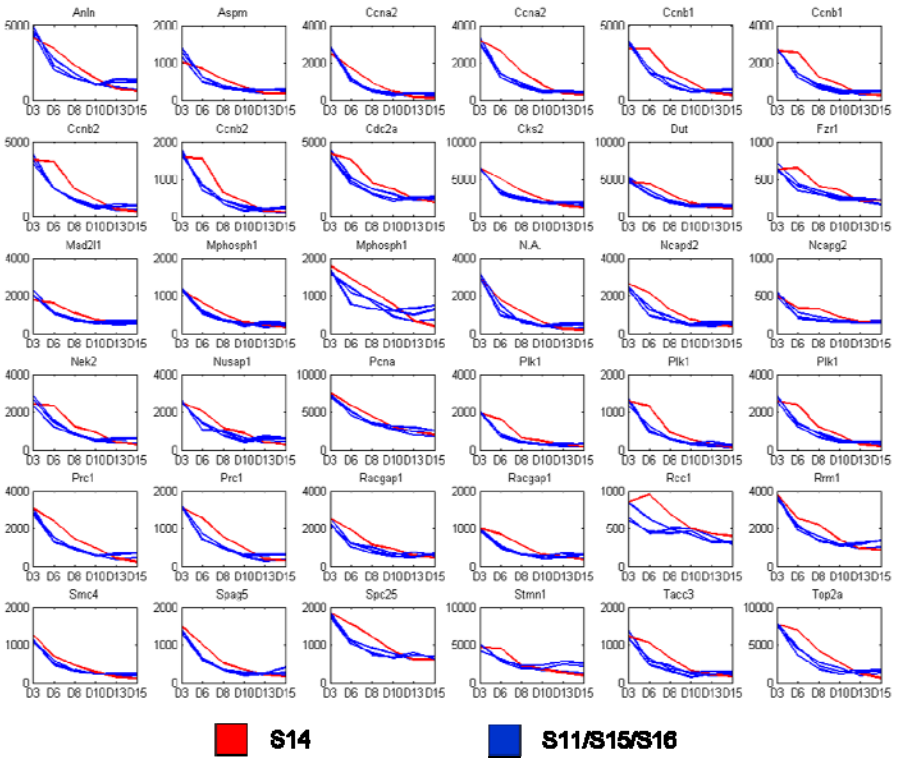


Figure 56 – Expression of cell cycle genes. Each subplot shows the expression of a particular cell cycle gene over time. The S14 culture is colored in red, while all other cultures are colored in blue.

5.6.1.7 Transporters

The expression of various transporters was also examined since different hydrolysate lots are known to have varying compositions of small molecules and peptides, which could trigger changes in transporter gene expression. Although this functional class was not identified as significantly enriched, a number of transporters display different expression profiles in varying component lots as shown in Figure 57. The primary substrate of each transporter (if known) is shown in Table 16. Of note, Slc45a1, a putative glucose transporter, was expressed at a higher level in cultures grown in Lot 67. This could possibly explain the increased glucose consumption rate of cultures grown in Lot 67 (Figure 58). However, the expression of other well-characterized glucose transporters, namely those in the GLUT family, would also have to be verified. Unfortunately, probes for these genes are currently absent on the version 2 CHO Affymetrix array.

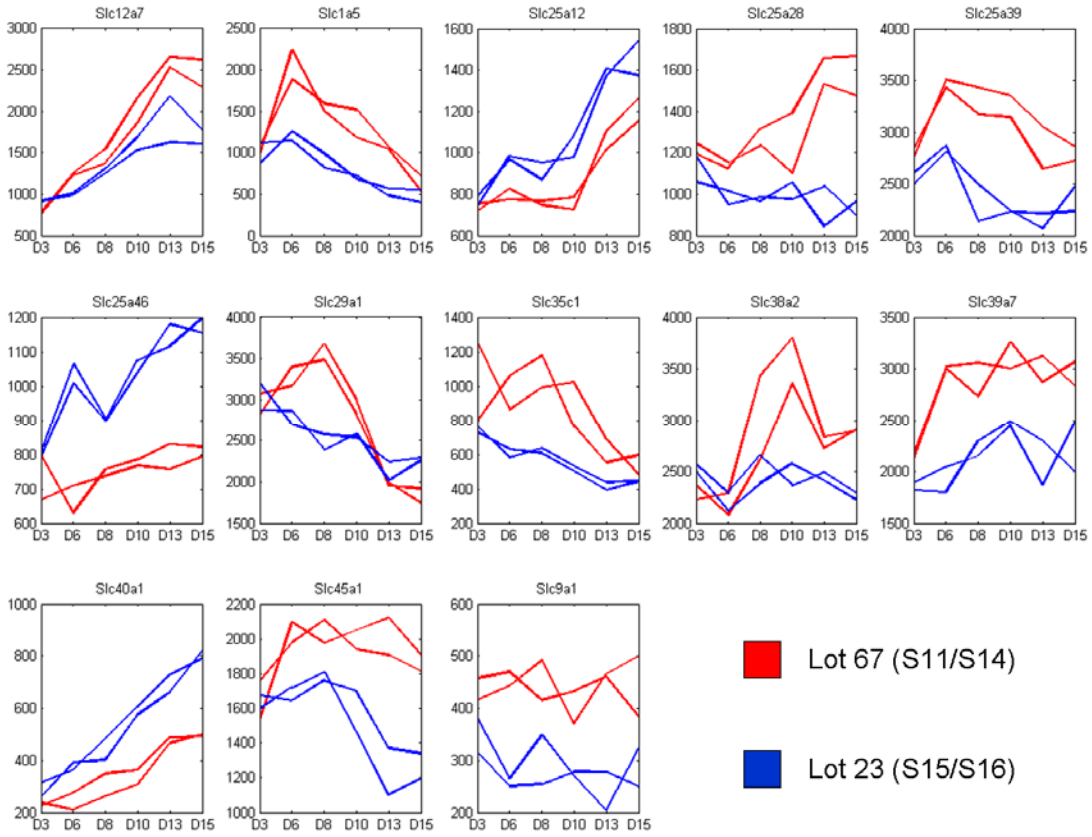


Figure 57 – Expression profiles of transporter genes found to be differentially expressed. Cultures grown in Lot 67 are shown in red, while cultures grown in Lot 23 are shown in blue.

Table 16 – Predominant substrates and transport type of identified differentially expressed transporters.

Gene Symbol	***Predominant Substrate***	***Transport Type / Coupling Ions***
Slc12a7	Potassium, chloride	Cotransporter / K^+, Cl^-
Slc1a5	L-Ala, L-Ser, L-Thr, L-Cys, L-Gln	Cotransporter / Na^+
Slc25a12	Aspartate, glutamate	Exchanger / Aspartate, Glutamate
Slc25a28	Unknown	Unknown
Slc25a39	Unknown	Unknown
Slc25a46	Unknown	Unknown
Slc29a1	Purine and pyrimidine nucleosides	Facilitated transport

Gene Symbol	*Predominant Substrate*	*Transport Type / Coupling Ions*
Slc35c1	GDP-fucose	Exchanger / GMP
Slc38a2	Alanine, aspargine, cysteine, glutamine, glycine, histidine, methionine, proline, serine	Cotransporter / Na^+
Slc39a7	Manganese	Unknown
Slc40a1	Ferrous ion	Facilitated transport
Slc45a1	Glucose	Unknown
Slc9a1	Na^+, H^+, Li^+, NH_4^+	Exchanger / Na^+, H^+

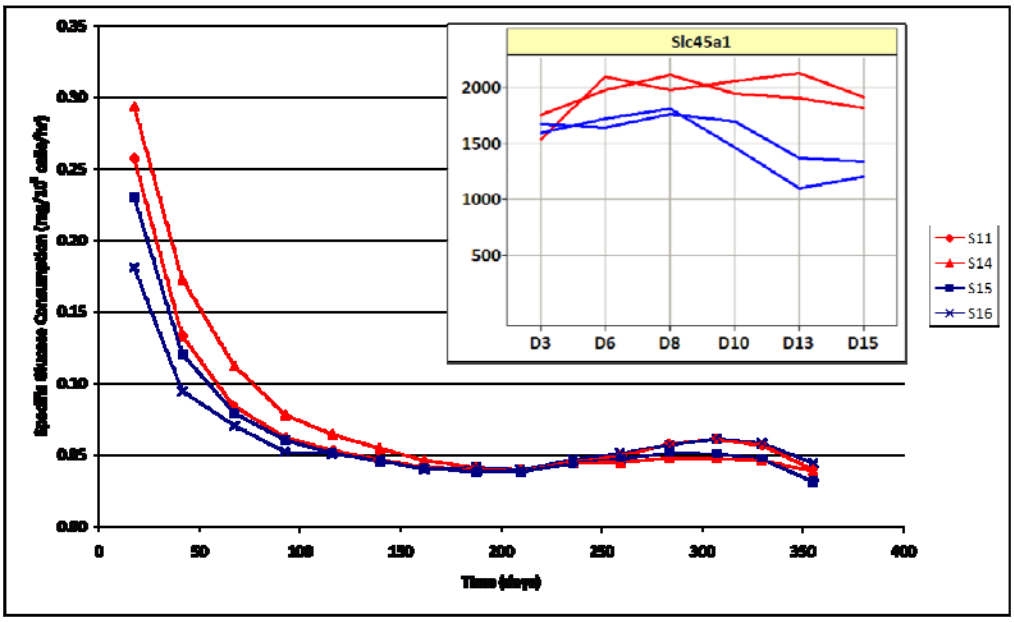


Figure 58 – Specific glucose consumption rates and expression of Slc45a1, a putative glucose transporter. Cultures grown in Lot 67 are shown in red, while cultures grown in Lot 23 are shown in blue.

5.6.2 Discussion

The use of plant hydrolysates as a serum substitute in cell culture medium formulations is used extensively in recombinant protein production processes. The presence of hydrolysates in the growth medium has been shown to enhance cell growth and lead to increased product titers (Franek et al., 2000; Burteau et al., 2003; Sung et al., 2004; Chun et al., 2005). However, these hydrolysates are derived from undefined raw materials and are consequently subject to extensive lot to lot variability (Luo and Chen, 2007). In this study, we examined the transcriptional impact of cultivating cells in different hydrolysate lots. Growth in different hydrolysate lots impacted a number of process parameters, including viable cell density, viability, glucose, lactate and antibody titer. In fact, a significant difference of almost 20% in antibody titer was seen between the best- and worst-producing cultures, each grown in a different hydrolysate lot. In a study screening multiple hydrolysate sources, Lu et al found a wide range in viable cell densities amongst cultures grown in varying hydrolysate sources, with more than a 6-fold change between the best- and worst-growing cultures (Lu et al., 2007).

At the transcript level, hydrolysate lot variability led to the identification of a large number of kinetically differentially expressed genes. Using both a regression-based method and a distance calculation-based method, we identified more than 500 genes as kinetically differentially expressed between cultures grown in alternative hydrolysate lots. Hydrolysates are known to contain multiple low-molecular weight components, including peptides, amino acids, vitamins and trace elements (Sung et al., 2004; Chun et al., 2005). The varying composition of these elements between different lots could exert an effect on multiple metabolic and signaling pathways within the cell. We therefore used functional analysis to identify the pathways significantly affected by hydrolysate lot variability.

Our analysis revealed that a number of genes involved in lipid metabolism had been significantly affected as a result of growth in different hydrolysate lots. A number of genes involved in fatty acid beta-oxidation were upregulated in cultures grown in Lot 23 (these cultures showed increased antibody titer). Fatty acid beta-oxidation is a highly

exergonic process, with each round producing one NADH, one $FADH_2$ and one acetyl-CoA molecule. It is possible that increased expression of these genes led to increases in metabolic energy, perhaps correlating with the observed increase in antibody titer. However, hydrolysates are known to contain varying concentrations of certain lipids (Gu et al., 1997; Franek et al., 2000). Therefore, it is possible that differences in lipid concentrations triggered the modulation of lipid metabolism genes observed under different hydrolysate lots.

Another class of genes affected by hydrolysate lot source was polyamine metabolism. The gene expression changes we observed collectively pointed to an increased expression of ornithine decarboxylase (ODC), the rate limiting enzyme in polyamine metabolism. Polyamines are essential for eukaryotic cell growth, and ornithine decarboxylase has been shown to be essential for cell survival during development (Pendeville et al., 2001). It has also been identified as a downstream target of MYC (Bello-Fernandez et al., 1993). A recent report of transcriptional analysis of CHO culture grown with and without hydrolysate also found polyamine metabolism to be modulated by hydrolysates (McCoy and Morris, 2008). The authors went on to show that similar growth and productivity could be obtained in the absence of hydrolysate by supplementing the growth medium with spermine, spermidine and putrescine. It is interesting to note that we also identified polyamine metabolism as modulated by hydrolysate, but in our case, the difference was observed simply as a result of hydrolysate lot variation. This observation opens up avenues for medium optimization, and it would be interesting to test whether the impact of hydrolysate lot variation can be mitigated by the addition of intermediates of polyamine metabolism.

Such an approach could be extended to study the impact of supplementing intermediates from all metabolic pathways identified. For instance, the addition of a lipid supplement could also help mitigate the influence of hydrolysate lot variability. Transcriptome analysis results could be used to guide medium optimization efforts, and help narrow the number of possible additives tested. In fact, a number of studies have been reported which use design of experiment approaches to systematically improve cell culture medium (Chun et al., 2003b; Sandadi et al., 2006; Huang et al., 2007). However,

in these studies, the choice of potential additives is not systematically addressed and is frequently empirically determined. The results of transcriptome analysis can therefore help guide such media improvement strategies.

We also observed a modulation in the expression of a number of transporter genes. It is known that peptones contained in hydrolysates are cleaved inside the cell by proteases, resulting in the release of free amino acids which can be used as nutrients or precursors for cellular components, such as other amino acids or nucleic acids (Heidemann et al., 2000). It is possible that variations in concentrations of these components in different hydrolysate lots were the cause, rather than the effect, of the modulation of transporter gene expression.

We observed a correlation between the expression of a number of cell cycle genes, and the behavior of one culture grown in Lot 67 (S14) which showed sustained viability for longer duration. The decreased expression of more than 36 cell cycle genes was delayed in this culture, and this could provide a potential explanation for the sustained viability observed at the phenotypic level in this culture. Although it is not feasible to genetically engineer each of the identified cell cycle genes in order to prevent cell death, an alternative may be to try and identify a master regulator of these genes and genetically engineer it in order to prolong culture viability. One approach to identify such a master regulator may be to compare the promoter sequences of the identified cell cycle genes and identify an over-represented motif which could be the target of a transcription factor. Such an approach as been used successfully to identify transcription factors regulating groups of co-expressed genes (Ho Sui et al., 2005).

This study revealed the profound effects that hydrolysate lot variation imparts on cellular transcriptomes. Identifying the cause of this variability is paramount, since it can have a major impact on a number of process parameters, including product titer. By using microarray analysis, we were able to pinpoint a number of metabolic pathways modulated by hydrolysate lot variability. The intermediates of these pathways represent attractive candidates for growth medium optimization. Indeed, it may be possible to mitigate the influence of hydrolysate lot variability by supplementing these compounds.

5.7 GLOBAL TRANSCRIPTOME ANALYSIS

We have presented the results of three microarray studies investigating the transcriptional effects of altering a number of process parameters, namely temperature, pH and hydrolysate lot source. While each individual study yielded insights into the transcriptional regulation of CHO fed-batch cultures, a global analysis of all microarray results can potentially provide additional information. Indeed, the power of microarray analysis as an analytical tool is greatly increased with increasing sample size and can allow for greater statistical significance. To this end, we performed a global analysis of all 12 CHO fed-batch cultures, both at the process parameter level and transcriptome level.

5.7.1 Global Process Data Analysis

In order to quantify which process parameters were most affected by each of the three experimental variables tested (temperature shift, pH shift and hydrolysate lot source), a correlation analysis was conducted which included all cell culture process data from all 12 runs. The first step in this methodology was to create a similarity matrix for each process parameter. To do this, each parameter's time series profile was organized in a 12 by 12 matrix and similarities were calculated for every process parameter using Euclidean distance as the similarity metric (see Section 0 for additional details). This is shown pictorially in Figure 59. If a process parameter is invariant to any type of process perturbation, then its similarity matrix will be made up of values close to 1 (a similarity of 1 means the profiles are identical in both runs). Three similarity matrices were then created for each experimental variable. These capture whether two runs were operated at the same experimental condition. In other words, if both runs were grown in Lot 23 of the hydrolysate, then their similarity is 1. Conversely, if one run was grown in Lot 23 and the other in Lot 67, their similarity is 0. Finally, a correlation coefficient was calculated for each process parameter by calculating the similarity between its similarity matrix and the experimental variable matrix (temperature shift, component lot source and pH shift), using both the Pearson and Spearman correlation calculations. The output can be interpreted as follows: if a process parameter's behavior across runs can be efficiently

segregated by a particular experimental variable, then it will have a high similarity coefficient.

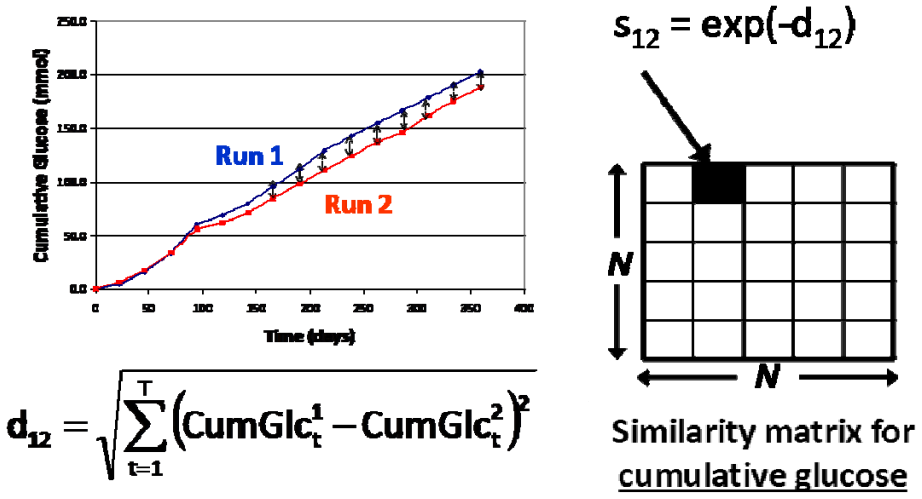


Figure 59 – Calculating similarity matrices for each process parameter. For each process parameter, a similarity matrix was created by calculating the Euclidean distance between the profiles of parameter in any two runs. This resulted in the creation of a 12 by 12 matrix for every parameter (N=number of runs).

An example of two process parameters with high (glucose concentration) and low (sodium concentration) correlation to pH shift, respectively, is shown in the figure below.

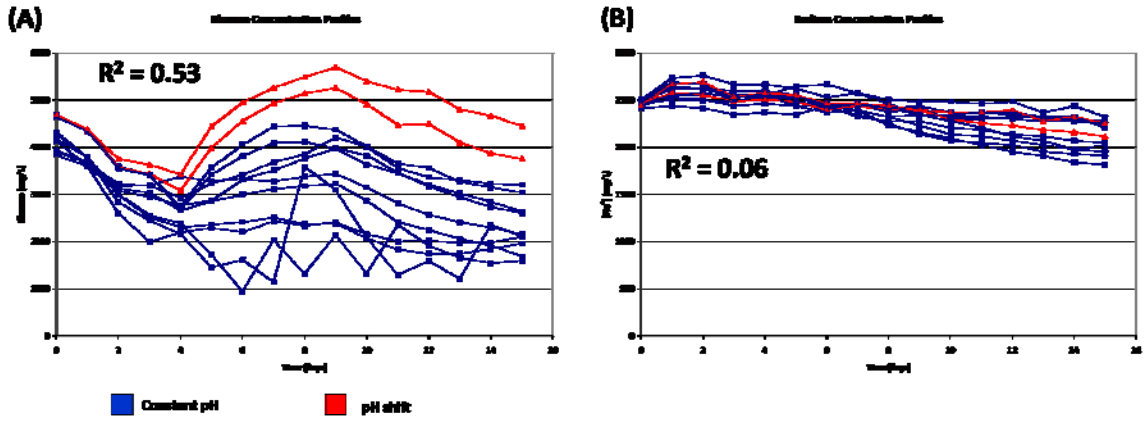


Figure 60 – Example of process parameters with (a) strong and (b) poor correlation to pH shift.

The results of this analysis, using Spearman rank correlation as the similarity metric, for all three experimental variables are summarized in Figure 61. Similar results were obtained using Pearson correlation as the similarity metric. As can be seen, temperature shift correlation coefficients are high for a number of process parameters,

indicating that this perturbation had the greatest effect on the process response. It should be noted that the sample size for this analysis is not always evenly distributed. For example, only 2 runs were performed at constant temperature while 10 runs had a temperature shift. Similarly, 2 runs had a pH shift while all remaining runs were cultured at a constant pH. It is interesting to note that antibody product or specific productivity was not strongly correlated to any of the experimental variables tested.

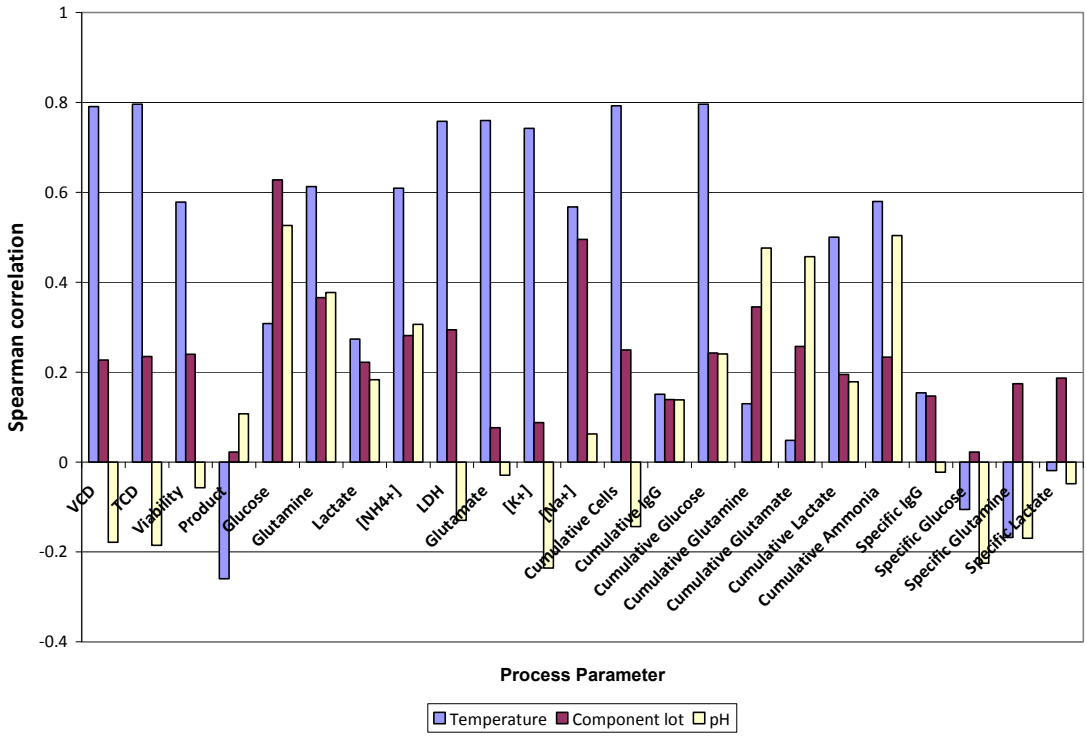


Figure 61 – Process parameter correlations to experimental variables.

5.7.2 Global Transcriptome Profiling

5.7.2.1 Identifying Sources of Variance in the Data

Variance Decomposition Analysis

In order to identify the major sources of variance in the combined set of microarray data, a variance components decomposition analysis was performed. This analysis was executed in Jmp Genomics 3.2. Each of the 12 reactor runs was classified on the basis of five experimental variables: temperature, component lot source, pH, reactor run and day. The relative contribution of each of these factors, as well as their

interactions, to the overall observed variance in the transcriptome data is shown in Figure 62. As expected, culture duration (i.e. day) accounts for the greatest proportion (51%) of observed variance. Interestingly, component lot emerges as the next determinant of variability in the dataset, accounting for 12.4%. This is surprising, as one would expect temperature shift to incur a greater transcriptional response than different hydrolysate lots. However, temperature shift only accounts for 8% of the variance. This analysis highlights the profound effect hydrolysate lot variability can have on the cellular transcriptional response.

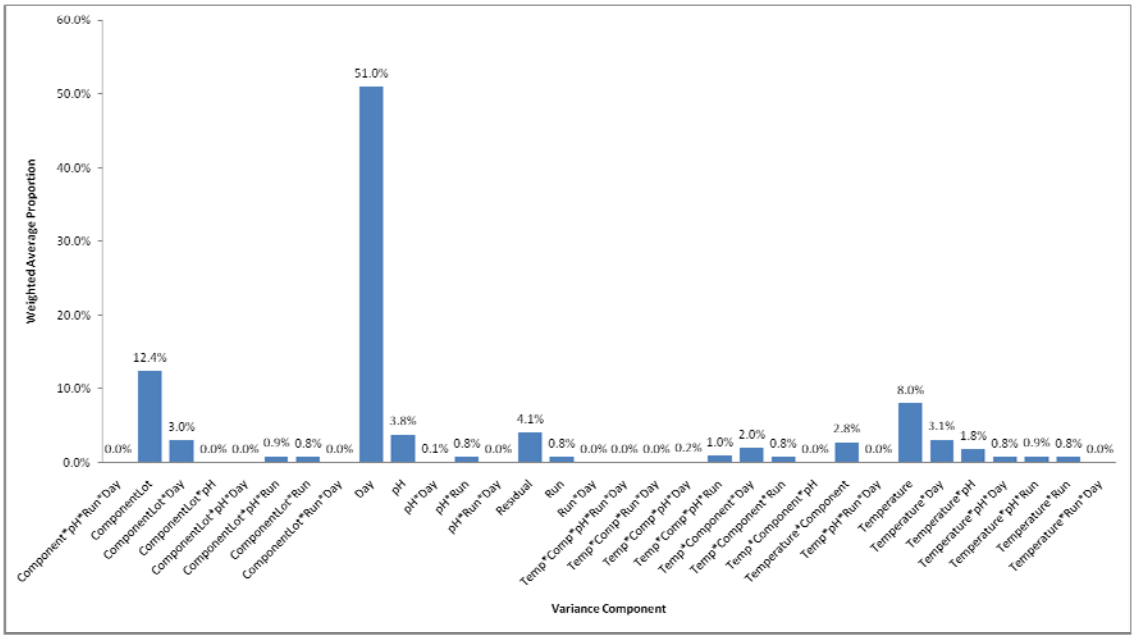


Figure 62 – Results of variance components decomposition analysis performed on all 12 time-series microarray datasets.

Unsupervised Clustering

In a parallel approach, unsupervised clustering was performed on the transcriptional response of each reactor run. Similarity matrices based on Pearson correlation and Spearman rank correlation were generated for each gene's temporal expression profiles in the twelve reactor runs. These were then used as the input for hierarchical clustering (see Section 5.3.4.3 for additional details). The results of the analysis, which are shown in Figure 63, confirm the observation made through the

variance decomposition analysis. Namely, that component lot can efficiently segregate microarray expression profiles. Identical dendrograms were obtained with both the Pearson correlation and Spearman rank correlation.

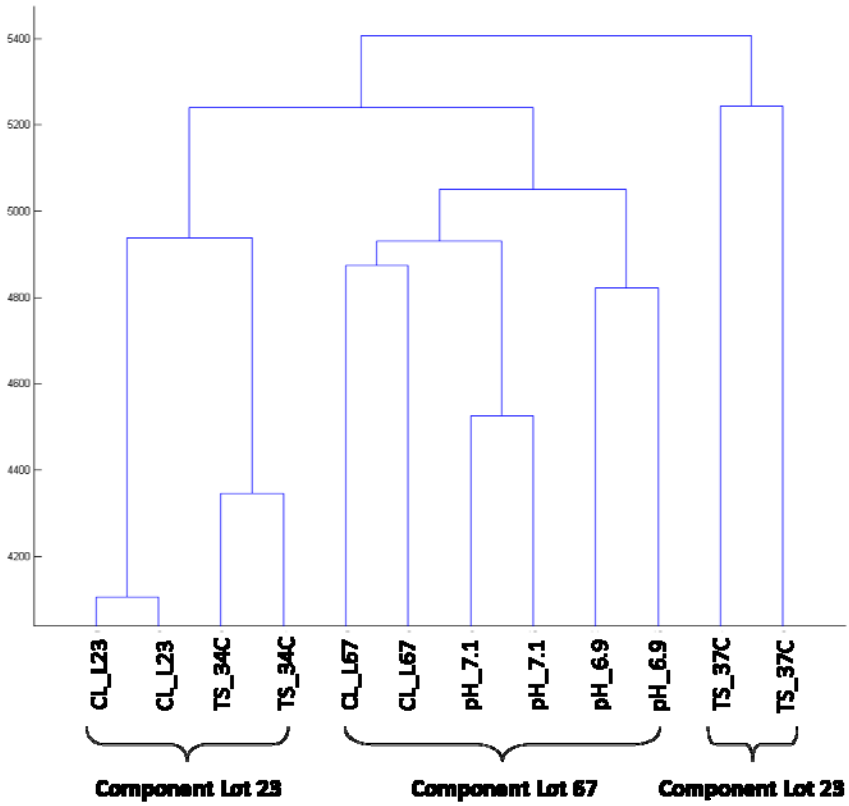


Figure 63 – Hierarchical clustering results of time series samples. Nomenclature: CL: component lot experiment, TS: temperature shift experiment, pH: pH shift experiment.

5.7.2.2 Hydrolysate Lot Analysis Revisited

The previously described variance decomposition analysis revealed that hydrolysate lot source accounts for more than 10% of the observed gene expression variability (Figure 62). Hierarchical clustering of gene expression profiles also confirmed these results. Consequently, we wanted to revisit this experiment by performing additional analysis on all twelve reactor runs. In order to identify a set of genes which we were confident was kinetically differentially expressed as a result of component lot source, we examined the intersect of genes identified as significantly differentially expressed across component lots between the initial dataset (consisting of 4

runs, described in section 5.6) and the remaining 8 runs, classified as either grown in Lot 23 or Lot 67 of the hydrolysate. The same regression-based approach was used to identify kinetically differentially expressed genes, and the false discovery rate (Q-value) was again set at 5%. As can be seen in the following Venn diagram, almost 70% of the genes identified from the initial data analysis are identified when the analysis is extended to use the remaining 8 runs. This reinforces the global effect component lot source has on gene expression.

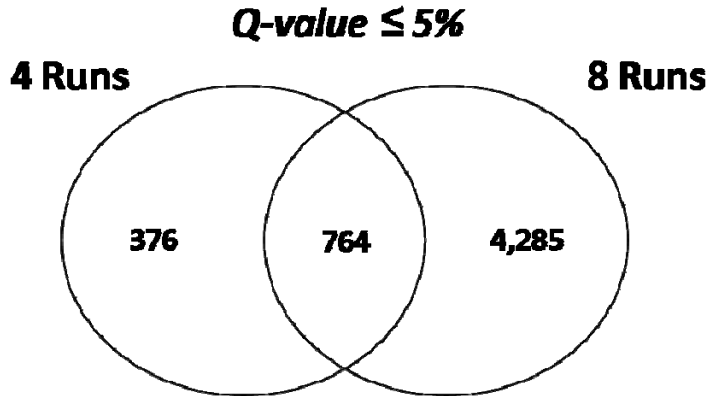


Figure 64 – Venn diagram of genes identified as kinetically differentially expressed across component lots at a q-value threshold of 5%.

Gene ontology enrichment analysis on genes found to be significantly differentially expressed (q-value ≤ 5%) in each dataset (4 runs or 8 runs) identified the same functional classes as significantly enriched. Those functional classes are shown in the following table.

Table 17 – Enriched functional classes identified amongst kinetically differentially expressed genes.

Enriched Functional Classes
Amine metabolism
Cell cycle
Cell redox homeostasis
Extracellular space
Hexosaminidase activity
Oxidoreductase activity
Response to oxidative stress
Vacuole

The top kinetically differentially expressed genes from the intersect of all analyses are shown in the figure below. These genes represent possible screening targets in the evaluation of component lot source.

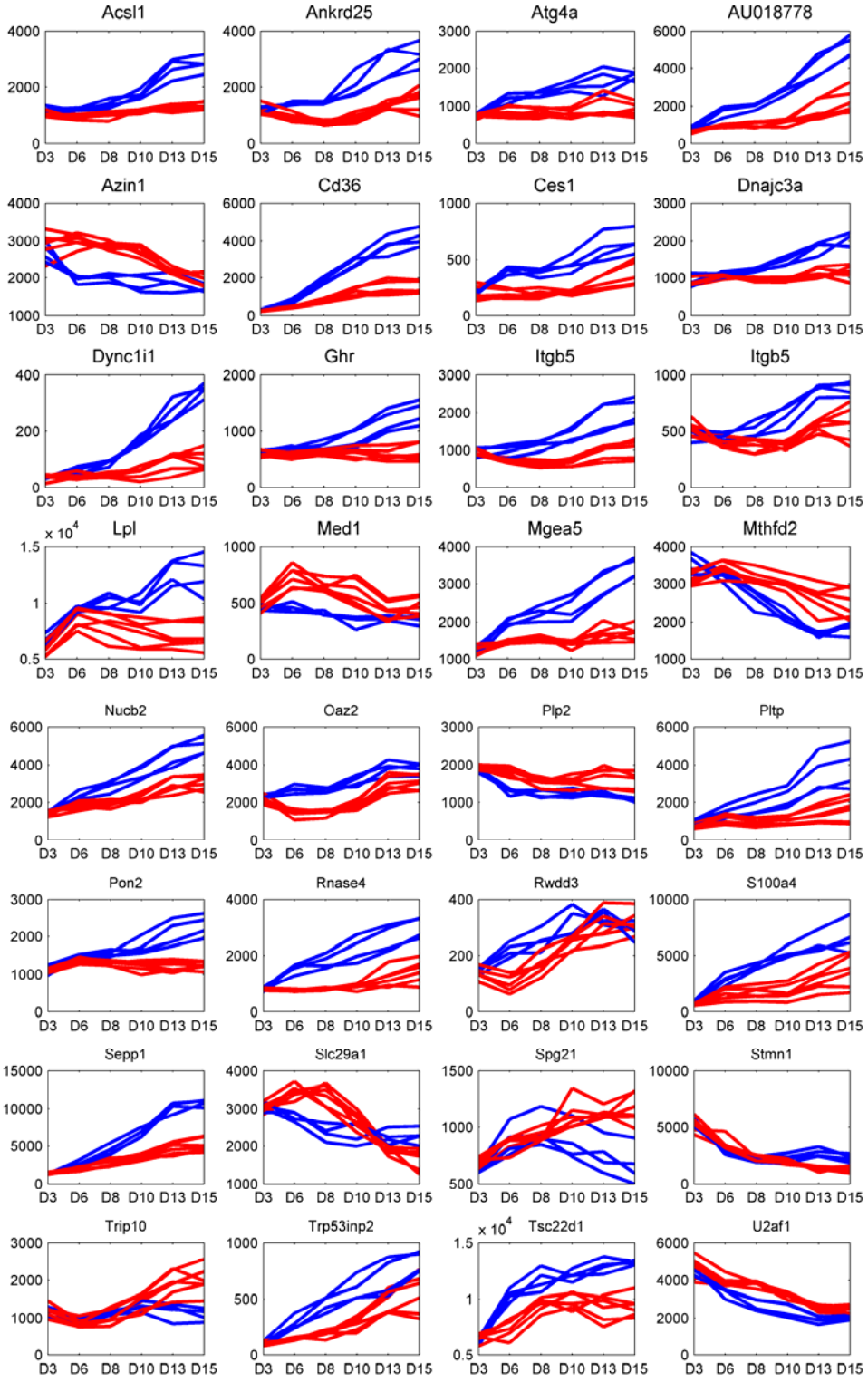


Figure 65 – Top kinetically differentially expressed genes across component lots. Blue profiles represent cultures grown in Lot 23 while red profiles represent cultures grown in Lot 67.

5.7.2.3 Reproducibility of Biological Replicates

This study afforded us the opportunity to evaluate the quality of our CHO Affymetrix array, by examining the reproducibility of biological replicates. Day 3 samples from all 12 reactor runs were compiled and segregated according to hydrolysate lot. A coefficient of variation based on intensity values across each set of 6 samples was calculated for every gene. As can be seen in Figure Blah, the distribution of coefficient of variation values is very narrow, suggesting little variation across biological replicate samples.

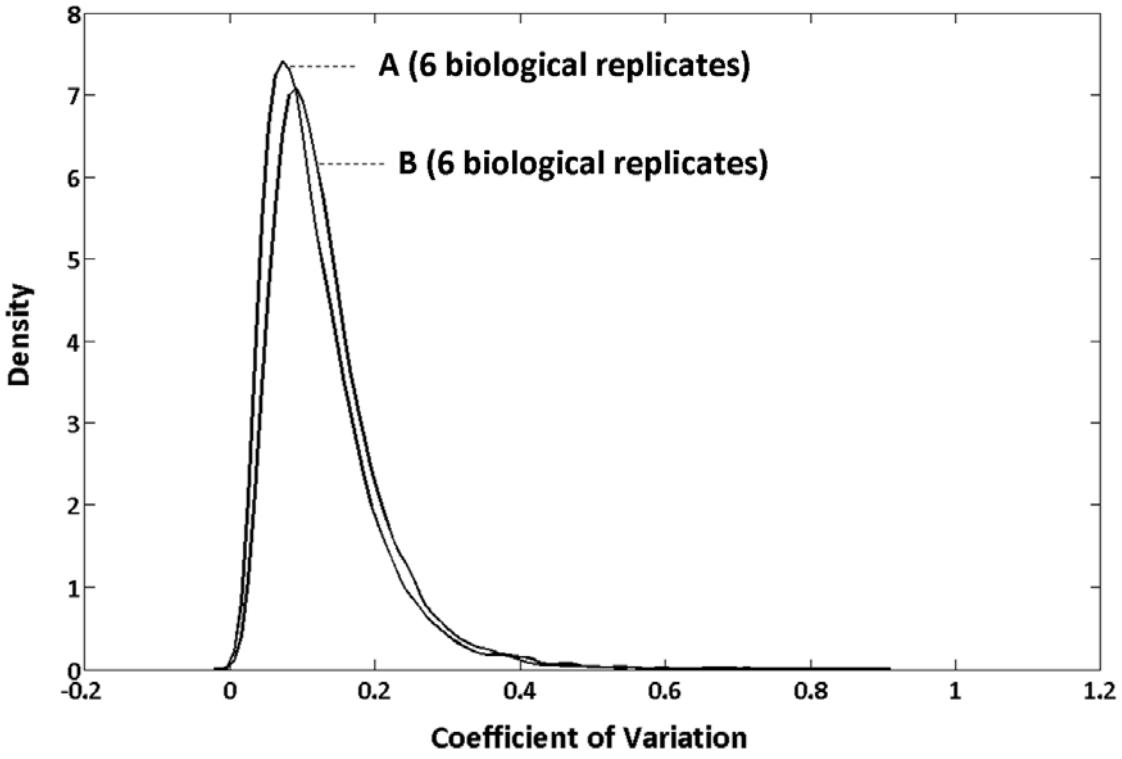


Figure 66 – Coefficient of variation for Day 3 sample intensities. Each curve shows the distribution of coefficients of variations for 6 biological replicates.

5.7.2.4 Differential Expression of Unannotated Genes

Finally, it is important to highlight that a number of unannotated genes were identified as significantly differentially expressed in each of the three studies described here. An example of kinetically differentially expressed genes without annotation identified in each of the three studies is shown in Figure 67. These examples highlight the importance of continuing to improve the annotation of our current set of CHO sequences, enabling new and interesting biological insights.

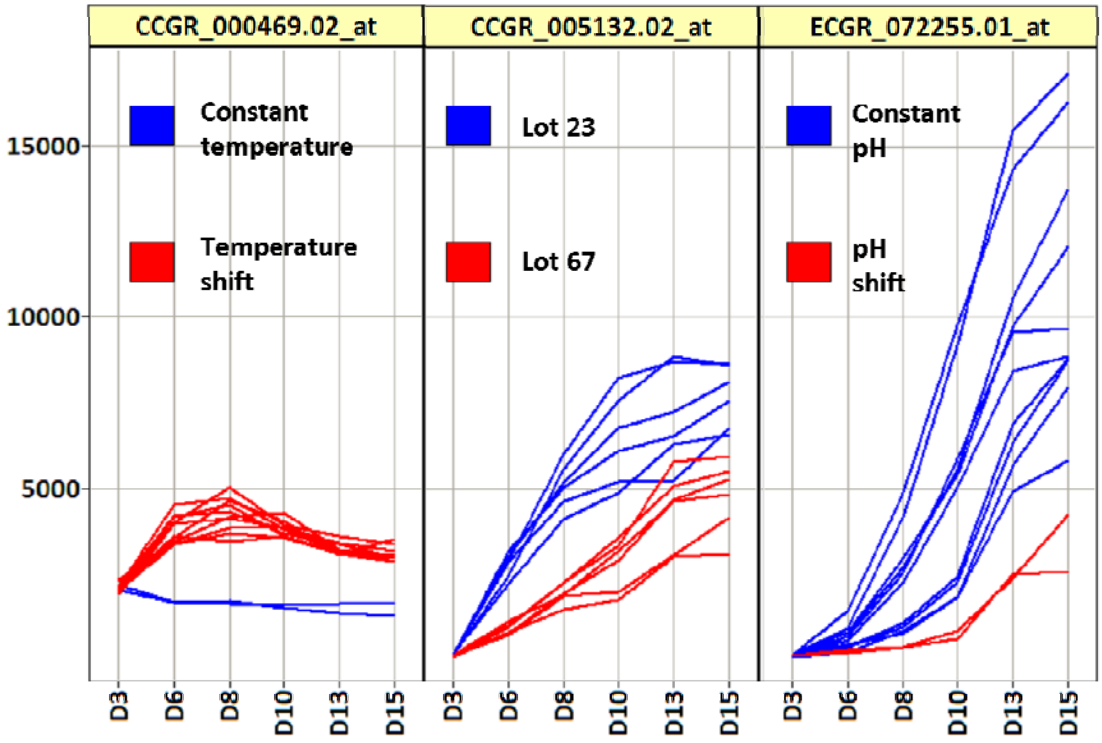


Figure 67 – Unannotated genes identified as kinetically differentially expressed in each of the three studies described in this chapter.

5.8 CONCLUSIONS

This study represents the first thorough transcriptional characterization of an antibody-producing CHO fed-batch system. Several interesting observations arose through this study. Firstly, gene expression in fed-batch cell culture systems was shown to be very dynamic, an observation which goes against the traditional notion that cultured mammalian cells have a lower magnitude of observed fold-changes as compared to other biological systems (Wlaschin et al., 2006). Secondly, we found that changes in process parameters had an appreciable impact on the cellular transcriptome, as evidenced by the number of genes identified as kinetically differentially expressed in each of the three studies. Such transcriptional changes were also observed to be reproducible across biological replicate cultures. It is interesting to note that in some instances, we were able to propose potential correlations between transcriptome and process parameter changes. For example, in the hydrolysate lot experiment, expression of a number of cell cycle genes was found to correlate with culture viability. Finally, by using data mining techniques on the entire accumulated set of time series microarray data, we were able to identify the profound transcriptional effect of hydrolysate lot variation. This information can potentially be used as a screening tool for hydrolysate lot evaluation by defining a pattern of gene expression which captures the favorable characteristics of a well-performing hydrolysate lot, or can help guide design of chemically defined medium through the identification of affected biological pathways. Overall, this study showed the utility of transcriptome analysis in process characterization.

6 ENGINEERING CELLULAR METABOLISM

6.1 SUMMARY

Mammalian cells in culture display enhanced glycolysis and reduced oxidative phosphorylation, as compared to their non-transformed counterparts. This inefficient nutrient utilization leads to lactate accumulation, which has been shown to be detrimental to cell growth, leading to reduced achievable product titers. To investigate this phenomenon, we compared the expression profiles of cultured cells and their tissue of origin. We found that cultured cells have upregulated a number of glycolytic enzymes, including lactate dehydrogenase A (LDH-A). Consequently, we used short interfering RNAs (siRNAs) to downregulate LDH-A in an antibody-producing NS0 cell line, and achieved more than 80% reduction at the enzyme activity level in one of the characterized clones. Interestingly, glucose and lactate profiles were unchanged in clones with LDH-A knockdown. Growth in low glucose concentration did not yield any appreciable differences. We next examined the effects of overexpressing the LDH-C isoform, whose kinetic properties favor conversion of lactate to pyruvate, and which could potentially result in a lactate consuming phenotype. However, neither LDH-C overexpression, nor combined LDH-C overexpression and LDH-A knockdown yielded any appreciable changes in central metabolisms. These results suggest that by the time stable engineered clones are selected, the effects of reduced LDH-A or increased LDH-C transcript levels may have already been mitigated by the cell.

6.2 INTRODUCTION

Cultured cells have been shown to be inefficient in their nutrient utilization and mainly metabolize glucose anaerobically. Therefore, the majority of pyruvate produced is converted to lactate by the catalytic action of lactate dehydrogenase, rather than entering the TCA cycle for energy generation. It has been shown that increasing lactate concentrations have an inhibitory effect on cell growth (Ozturk et al., 1992). Furthermore, the acidification of the medium caused by increased lactate concentration necessitates the addition of base for pH control, which in turn can be detrimental to the

cells (Irani et al., 1999). It is therefore beneficial to investigate the source of this inefficient nutrient utilization, and develop strategies to reduce byproduct accumulation.

Consequently, we used transcriptome analysis in an effort to gain insight into this phenomenon. Specifically, we compared gene expression profiles of cultured cells with expression data from their tissue of origin, which represent conditions were nutrient utilization has not been disrupted. Two cell lines were used in this analysis: Chinese hamster ovary cells and NS0, a mouse myeloma cell line. The NS0 cell line (Non-Secreting) was originally derived from an IgG-producing cancerous plasma cell line, therefore expression data from plasma cells was used as the tissue of origin data. CHO cell samples were compared to ovary microarray data. Gene expression changes between cultured cells and their tissue of origin were visualized using GenMAPP, which allows for microarray data to be overlaid onto canonical metabolic and signaling pathways. As can be seen in Figure 68, there is a general upregulation of the glycolytic pathway in both CHO and NS0 cell lines, as compared to their respective tissue of origin. This upregulation may partly explain the increased glycolytic activity observed in continuous cell lines. Furthermore, closer examination of the LDH node reveals that the LDH-A isoform is upregulated in both cell lines. In NS0 cells, the LDH-A subunit is predominantly expressed, with an average intensity of 5200, compared to an average intensity of 80 for the LDH-B subunit. The LDH-C subunit is not expressed in NS0 cells.

The lactate dehydrogenase (LDH) enzyme catalyzes the reversible conversion of pyruvate to lactate, accompanied by the oxidation of NADH to NAD^+. The enzyme exists as a tetramer with varying composition of two subunits: LDH-A and LDH-B. A third isoform, composed exclusively of subunits encoded by the LDH-C gene, is found exclusively in the testis (Table 18).

Table 18 – Summary of lactate dehydrogenase isoforms.

Gene	*Subunit*	*Isoform*
LDH-A	H_4 (Heart)	LDH-1
LDH-B	M_4 (Muscle)	LDH-5
LDH-C	C_4 (Testis)	LDH-C

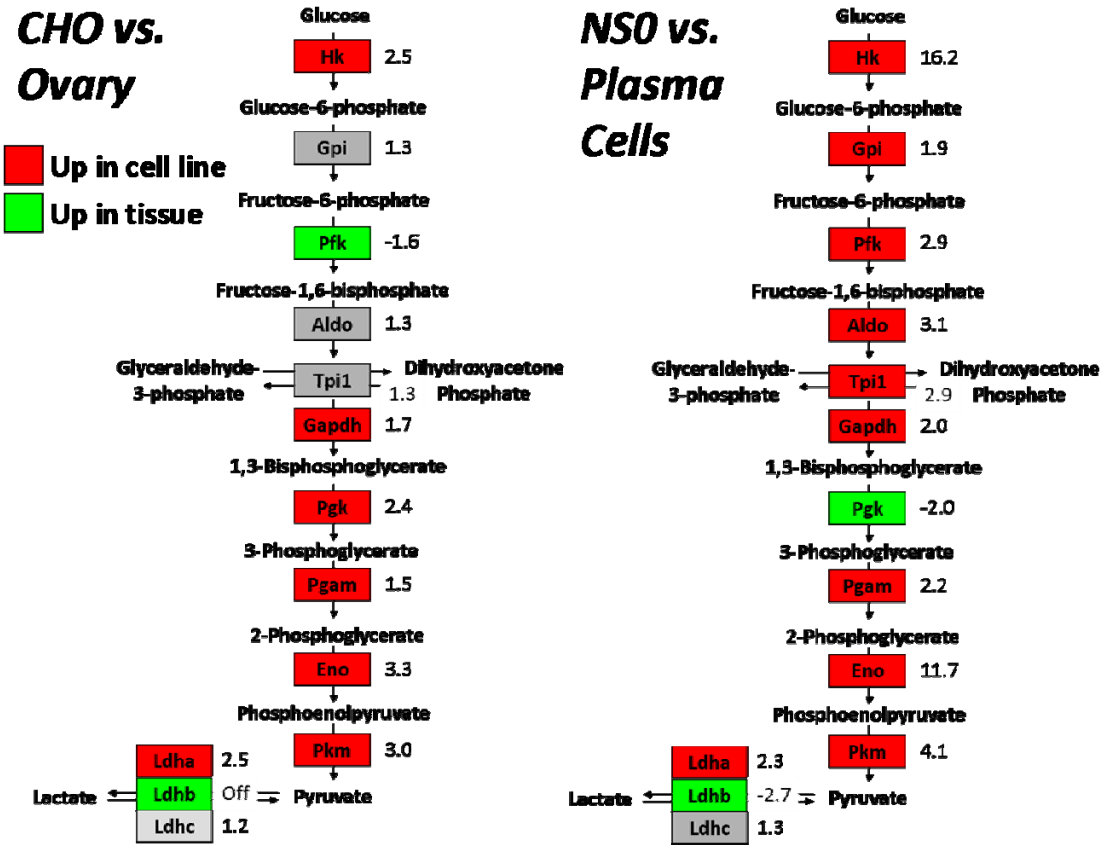


Figure 68 – Gene expression changes of glycolytic enzymes between cultured cells and their tissue of origin. In the left panel, CHO cells were compared to ovary tissue, while in the right panel, NS0 cells were compared to plasma cells. Each enzyme in the pathway is colored according to fold change: red if upregulated by at least 1.5-fold in the cell line and green if downregulated by at least 1.5-fold in the cell line. Enzymes are colored in grey if they do not meet the 1.5-fold threshold.

Each LDH isoform has distinct kinetic properties, as shown in Table 19. Consequently, there is tissue-specific distribution of each LDH isoform, according to the metabolic requirements of each tissue. For instance, the LDH-5 isoform, composed exclusively of LDH-B subunits, is the predominant subunit in mainly aerobic tissues, such as cardiac muscle (Read et al., 2001). In the testes, the LDH-C isoform is predominant and lactate plays a central role in the energy metabolism of Sertoli and germ cells (Boussouar and Benahmed, 2004). Sertoli cells, which predominantly express the LDH-1 isoform, convert pyruvate to lactate, which is subsequently transported into germ cells via monocarboxylate transporters. Lactate is then converted to pyruvate by the LDH-C enzyme, and pyruvate enters the TCA cycle for energy generation.

Table 19 – Kinetic properties of LDH isoforms (Read et al., 2001).

Isoform	K_m *(Pyruvate) [mM]*	K_m *(Lactate) [mM]*
LDH-1	0.12	5.5
LDH-5	0.25	20
LDH-C	0.12	4.9

Based on the observations of our transcriptome comparison of cultured mammalian cells to their native tissues, we decided to engineer cellular metabolism by modulating expression levels of the various LDH isoforms. Our first objective was to knockdown LDH-A and LDH-B transcript levels in cultured mammalian cells with the aim of reducing lactate formation and ameliorating nutrient utilization for increased energy production. To accomplish this, we used siRNA (short interfering RNA) technology, which has become the predominant methodology to achieve expression knockdown in mammalian systems (Dorsett and Tuschl, 2004). By generating stable transfectants, we were able to isolate clones with more than 8-fold downregulation of the LDH-A transcript, and greater than 80% reduction at the enzyme activity level. These were characterized in terms of their growth and nutrient consumption.

Our next objective was to characterize the effect of overexpressing the LDH-C transcript. The kinetic properties of this enzyme favor the conversion of lactate to pyruvate, and overexpression of LDH-C in the cell could potentially lead to lactate consumption, thereby reducing cellular concentrations of this inhibitory metabolite. The phenomenon of lactate consumption has been observed in mammalian cell culture, and a correlation has been observed between lactate consuming cultures and increased antibody titer (Charaniya et al., 2009). As model systems, we used a parental antibody-producing NS0 cell line, as well as one of the newly-generated LDH-A knockdown clones. The effects of LDH-C overexpression, as well as combined LDH-C overexpression and LDH-A knockdown on cellular growth and metabolism were characterized. The results are presented below.

6.3 MATERIALS AND METHODS

6.3.1 Microarray Data

Total RNA from two Chinese hamster ovary cell lines of different parental origin (DG44 and DXB11) was harvested from mid-exponential stage samples using the RNeasy Mini kit (Qiagen). Ovaries from late adolescent (4 month old) virgin female Chinese hamsters were used for total RNA isolation using Trizol (Invitrogen, Carlsbad, CA). RNA samples were labeled as per the protocol described in Section 4.3.2 and hybridized onto CHO version 1 Affymetrix arrays, containing 10,118 probesets. The average intensity on each array was normalized by global scaling to a target intensity of 500. Data were extracted using the Affymetrix Microarray Suite 5 (MAS5) algorithm. Genes with a maximum intensity ≤ 70 and a detection p-value ≥ 0.04 across all samples in a given study were called absent and excluded from further analysis. Total RNA from biological replicate cultures of mid-exponential growing NS0 cells expressing a recombinant IgG was extracted using the RNeasy Mini kit (Qiagen) and labeled using the Affymetrix One Sample labeling kit as described above. Samples were hybridized onto Affymetrix MOE 430 2.0 arrays containing 45,023 probesets. Intensity data was extracted as described above. Finally, microarray data from mouse plasma cells was downloaded from NCBI's Gene Expression Omnibus microarray data repository (GDS1695), and quadruplet samples hybridized onto MOE 420 2.0 arrays were used for this analysis. Data was scaled to an average intensity of 500 to be comparable across samples.

Fold changes between cultured cell lines and tissue samples were calculated for each system (CHO and NS0) and visualized in the context of metabolic and signaling pathways using GenMAPP (Dahlquist et al., 2002).

6.3.2 Cell Culture

The NS0 cell line used in this study produces a recombinant antibody product and has been described previously (Sauer et al., 2000). The parental NS0 cells and transfected clones were maintained in T-flasks and grown in a proprietary medium based on DMEM/F12 (1:1) containing glucose (21.7 mM), glutamine (6.25 mM), sodium

bicarbonate (29 mM), putrescine (0.7 μM), penicillin G (0.17 mM), streptomycin (68.6 μM), pluronic F68 (60 μM) and phenol red (19.9 μM).

6.3.3 Vector Construction, Transfection and Clone Isolation for LDH-A Knockdown

The Invitrogen Block-iT™ Pol II miR RNAi expression system (K4935-00) was used for knockdown. This system allows the expression of knockdown cassettes driven by RNA polymerase II (Pol II) promoters. miRNA sequences designed against the gene of interest are flanked by native miRNA sequences which allow for proper processing of the miRNA transcript. The Invitrogen RNAi designer software (https://rnaidesigner.invitrogen.com/rnaiexpress/) was used to design targeting sequences against the mouse LDH-A gene (NM_010699) and mouse LDH-B gene (NM_008492). The top two sequences against LDH-A and the top sequence against LDH-B, as ranked by the software, were used for targeting. The LDH-A targeting sequences were CAAGGACCAGCTGATTGTGAA (named LDH-A1) and ACGTGAACATCTTCAAGTTCA (named LDH-A2), while the LDH-B targeting sequence was AGTCTCCCTCCAGGAACTGAA (named LDH-B). Each targeting sequence was inserted into the targeting vector, pcDNA6.2-GW/EmGFP-miR (see Figure 69). Briefly, 200 μM of top strand oligo and 200 μM of bottom strand oligo were combined with annealing buffer to a total volume of 20 μL and incubated at 95°C for 4 min. This mixture was diluted 5,000-fold and used for ligation: 10 ng of linearized vector was combined with 10 nM double-stranded oligos, ligation buffer and 1 U T4 DNA ligase and the reaction was allowed to proceed for 5 min at room temperature. 2 μL of the ligation mixture was combined with one vial of OneShot® TOP10 chemically competent *E. coli* cells, incubated on ice for 5 min and heat-shocked for 30 sec at 42°C. 250 μL of room temperature S.O.C. medium was added to the cells and incubated for 1 hr at 37°C with shaking. 20 μL of bacterial culture was plated onto pre-warmed LB agar plates containing 50 μg/mL spectinomycin and incubated overnight at 37°C. Ten colonies for each expression construct were expanded, and plasmid DNA was purified using the QIAGEN Plasmid Mini Kit (Qiagen). Constructs were sequenced to verify the presence and correct orientation of the insert, as well as the sequence of the insert using

the forward sequencing primer provided with the kit. Upon sequence confirmation, one *E. coli* clone for each sequencing construct was expanded and frozen glycerol stocks were established. Large quantities of plasmid DNA were obtained using 1 L bacterial cultures and purified using the QIAGEN Plasmid Maxi kit. A negative control vector was also constructed in parallel using the same procedure as described above, and named pcDNA6.2-Neg. The negative control targeting sequence, which represents a scrambled sequence, is provided with the kit as a top and bottom strand oligo.

The RNAi expression system allows for up to three targeting sequences to be chained together on the same plasmid. Consequently, the three targeting sequences (LDH-A1, LDH-A2, LDH-B) were combined into one expression vector. Briefly, 2 μg of pcDNA6.2-LDH-A2 was digested with 10 U *Bam*H I and 10 U *Xho* I for 2 hr at 37°C. In parallel, 2 μg of pcDNA6.2-LDH-A1 was digested with 10 U *Bgl* II and 10 U *Xho* I for 2 hr at 37°C. Digests were run on 2% agarose gels, and the backbone and insert fragments were excised from the gel and purified using the QIAquick Gel Extraction kit (Qiagen). The purified backbone and insert were ligated at a 1:4 molar ratio using 10 U T4 DNA ligase. OneShot® TOP10 chemically competent *E. coli* cells were transformed as described above. Resultant colonies were expanded and plasmid DNA was purified using the QIAGEN Plasmid Mini Kit (Qiagen). Constructs were sequenced and verified. One *E. coli* clone was selected, expanded, and a frozen glycerol stock was established. The resulting construct was named pcDNA6.2-LDH-A1A2. To further append the LDH-B targeting construct, the same procedure as described above was used, using pcDNA6.2-LDH-B as the insert and pcDNA6.2-LDH-A1A2 as the backbone. The resulting construct was sequence verified and named pcDNA6.2-LDH-A1A2B.

40 μg of plasmid (pcDNA6.2-Neg, pcDNA6.2-LDH-A1, pcDNA6.2-LDH-A2, pcDNA6.2-LDH-B, pcDNA6.2-LDH-A1A2 or pcDNA6.2-LDH-A1A2B) was linearized by digestion with 10 U of *Pci* I, and purified using QIAQuick PCR purification spin columns (Qiagen, Valencia, CA). Ten million exponentially growing cells were washed twice in 10 ml of cold Opti-MEM medium (Invitrogen, Carlsbad, CA), and electroporated with 40 μg of linearized DNA in 1ml of cold Opti-MEM in 4 mm electroporation cuvettes (BioRad, Hurcules, CA). Electroporation was performed in the

BioRad Gene Pulser XceII (BioRad). The electroporation conditions used for NS0 cells were 300 V, 250 μF and infinite resistance. Transfected cells were transferred into pre-warmed growth medium supplemented with 10% FBS (Atlas Biological, Ft. Collins, CO). Transfection efficiency was determined using FACS by parallel transfection with an EGFP-containing plasmid.

Transfected cells were diluted in 96-well plates at 2000 cells/well, in 0.2 ml of maintenance medium supplemented with 4 μg/mL blasticidin. Plates were incubated for 10-12 days in a 37°C, 5% CO_2 environment. Clones were expanded for characterization in their selective medium.

6.3.4 Confirming knockdown

RNA was isolated using RNEasy columns according to the manufacturer's protocol, with on-column DNAse digestion (Qiagen). cDNA synthesis was performed from 5 μg of total RNA using Superscript III Reverse transcriptase (Invitrogen, Carlsbad, CA). Primers for the mouse 18s rRNA (control), LHD-A and LDH-B genes were designed using Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3) with a specified product size of 150 to 250 bp and melting temperature of 60°C. Quantitative real-time PCR was performed using the Stratagene Mx3000P (Stratagene) with SYBR Green I dye chemistry using the Full Velocity SYBR green QPCR kit (Stratagene, La Jolla, CA). PCR conditions were: 94°C for 10 min, followed by 40 cycles of 95°C for 30 sec, 60°C for 1 min, and 72 °C for 30 sec. Dissociation curves were determined after PCR by complete dissociation at 95°C, 1 min., followed and 30 sec annealing at 55°C and a rapid temperature ramp to 95°C. The Ct (threshold cycle number) values were determined at 0.2 of the reference dye normalized baseline value. Triplicate cDNA samples, a no RT-reaction control, and a no cDNA template control were run for each sample/primer pair. All PCR products were run on 2% agarose gels to confirm expected product sizes.

To confirm knockdown at the enzyme activity level, LDH activity was measured using a colorimetric LDH assay from Sigma (Product#: TOX7). This assay is based on the reduction of NAD+ by LDH, and subsequent stoichiometric conversion of a

tetrazolium dye by NADH. Briefly, 8 x 10^4 cells from each sample were isolated and resuspended in 100 μL growth medium. 0.1X volume of LDH assay lysis solution was added to each sample and incubated at 37°C for 45 min. Cells were spun to pellet debris, and supernatant was transferred to 96-well plate. LDH assay mixture was prepared by combining equal volumes of LDH assay substrate, cofactor prep and dye solution. 2X volume of this mixture was added to each sample and incubated at room temperature, in the dark for 30 min. The reaction was terminated by adding 0.1X volume of 1 N HCL. Absorbance was measured at 490 nm. A series of dilution was performed for each sample and a straight line was fit to each dilution vs. absorbance curve. Percent knockdown was determined based on percent change of the slope of the straight line fit.

6.3.5 Vector Construction, Transfection and Clone Isolation for LDH-C Overexpression

The full-length isoform C of the mouse LDH gene (GenBank accession: X04752.1) was obtained in the pDNR-Lib vector from Open Biosystems (Open Biosystems, Hunstville, AL). The coding sequence was subcloned into the pcDNA_3.1_bsd vector (Invitrogen) through restriction digest with *EcoR I* and *Xho I* to yield pcDNA_3.1_bsd_LDHC. The resulting construct was transfected into two cell lines in parallel: the parental NS0 cell line, as well as the LDH-A knockdown clone B4 (described in Section 6.4.1). A control transfection was also carried out in these two cells by transfecting the empty vector, pcDNA_3.1_bsd. Transfection and selection conditions used in this study were the same as those described in Section 6.3.3.

6.3.6 Confirming LDH-C Overexpression

Quantitative real-time PCR was used to confirm overexpression of the target gene in the isolated clones. Primers for the mouse LHD-C gene were designed using Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3) with a specified product size of 150 to 250 bp and melting temperature of 60°C. PCR conditions were the same as described in Section 6.3.4.

6.3.7 Batch and Fed-batch Cultures

Exponentially growing cells from T-flask seed cultures were inoculated at a concentration of 4 x 10^5 cells/mL. Samples were taken daily for the duration of the cultures. For each sample, cell concentration and viability were determined by counting with a hematocytometer using trypan blue staining. Lactate concentrations were measured using the YSI Model 27 industrial analyzer (Yellow Springs Instruments, Yellow Springs, OH). Glucose concentrations were determined in duplicate using Infinity Glucose Hexokinase Reagent (Thermo Electron Corporation, Waltham, MA). For glucose, concentrations were determined according to the manufacturer's protocol. Absorbance was read at 340 nm using a SpectraMax Plus 384 plate reader (Molecular Devices, Sunnyvale, CA). Antibody concentration was measured using ELISA as described in Section 4.3.5. For fed-batch cultures, feed medium was added in 1 mL ballust feedings from Day 3 onwards. The feed medium composition was ten-fold (10X) concentrated basal medium excluding bulk salts ($NaHCO_3$, NaCl, $CaCl_2$, KCl). Cumulative consumption and production profiles were obtained as described in Section 5.3.2. Specific consumption and production rates were determined by plotting the substrate or product concentrations against the time integral values of the growth curve and calculating the slope (Renard et al., 1988).

For batch cultures grown in low glucose concentrations, cells were seeded at 3.5 x 10^5 cells/mL in 6-well plates, and parallel cultures were initiated in growth medium containing either 0.1 g/L or 4 g/L glucose concentrations. Cells were cultured for 12 hr, after which all samples were counted, and glucose and lactate concentrations were measured, as described above.

6.4 RESULTS

6.4.1 Development of NS0 Cell Lines with Reduced LDH-A Activity

An NS0 cell line expressing a recombinant antibody product was transfected with multiple LDH-A and LDH-B knockdown constructs to achieve constitutive knockdown of LDH activity. A commercially-available RNA interference (RNAi) system based on miR-155 (Chung et al., 2006) was used as the expression vector (Figure 69). This system

allows transcription of the targeting siRNAs by RNA Polymerase II from a human CMV promoter. The targeting sequence is flanked by native miRNA sequences, which allow for proper processing of the pre-miRNA by the endogenous cellular machinery (Figure 70). A GFP reporter gene is also expressed polycistronically with the targeting construct and GFP expression should correlate with expression of the targeting sequence.

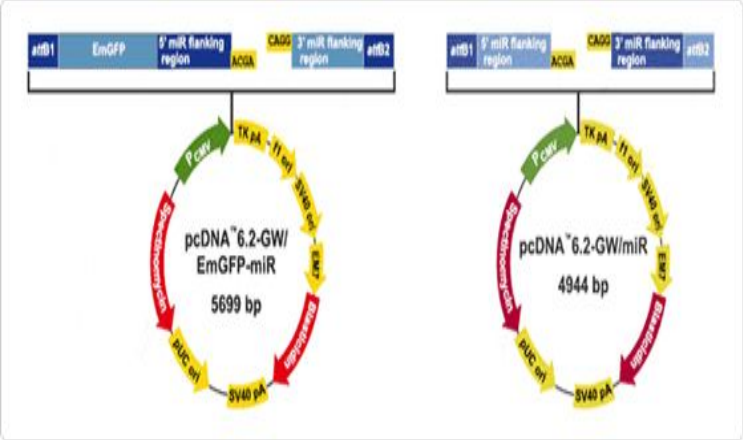


Figure 69 – RNAi expression vector used in this study. Figure reproduced from www.invitrogen.com.

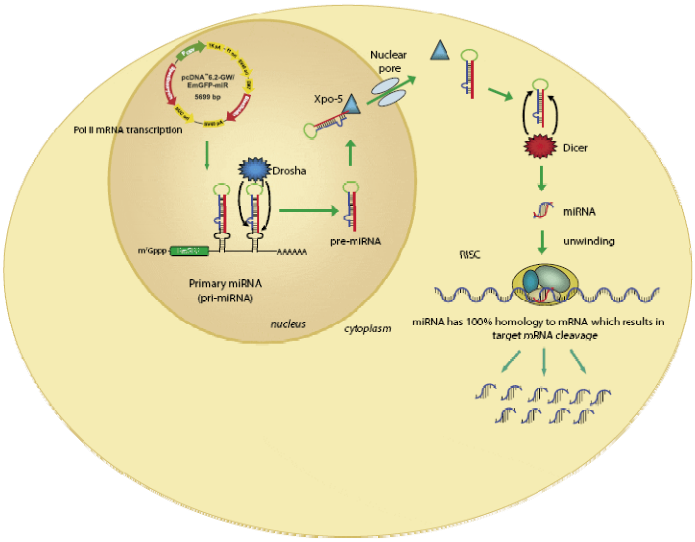


Figure 70 – Processing of siRNA targeting sequence makes use of the endogenous miRNA processing pathway. Figure reproduced from www.invitrogen.com.

A total of 5 targeting constructs were transfected, and these are shown in Table 20. Two control transfections were also performed in parallel: transfection with the pcDNA6.2-Neg plasmid, which contains a scrambled siRNA sequence and blasticidin

resistance gene, and electroporation in the absence of DNA. Stably transfected clones were isolated by limiting dilution cloning in 96 well plates in growth medium containing 4 μg/mL blasticidin for selection. Twenty-four clones from each of the targeting constructs were expanded for characterization, as well as twelve clones from the negative transfection. No clones were detected in the no DNA electroporation control.

Table 20 – Targeting constructs used for LDH knockdown.

Targeting Construct	*LDH Isoform Targeted*	*Targeted Region (bp from start)*
pcDNA6.2-LDH-A1	LDH-A	210
pcDNA6.2-LDH-A2	LDH-A	536
pcDNA6.2-LDH-B	LDH-B	724
pcDNA6.2-LDH-A1A2	LDH-A	210, 536
pcDNA6.2-LDH-A1A2B	LDH-A, LDH-B	See above

Quantitative Real-Time PCR (qRT-PCR) was used to quantify the expression level of the LDH-A and LDH-B transcripts in the expanded clones. The ratio of LDH-A expression in each clone to its expression in the negative control is shown in Table 21 for some of the expanded clones. Ct values for the LDH-B transcript revealed that this isoform was below detectable levels in all clones examined (data not shown). The use of an LDH-B targeting sequence was originally included as a precaution against the possibility that the cell, faced with reduced LDH-A levels, might upregulate the LDH-B transcript. However, this was not found to be the case. Interestingly, the only clones which showed any kind of significant reduction in LDH-A expression were those originating from the transfection with the pcDNA6.2-LDH-A1A2B construct; individual targeting constructs did not yield clones with any significant LDH-A knockdown (data not shown). It is known that the efficacy of siRNA targeting sequences can vary, and that factors such as targeting position within the target mRNA transcript can affect knockdown efficiency (Holen et al., 2002). Three of the characterized clones showed greater than 2-fold downregulation of the LDH-A transcript, namely clones B4, C5 and D3.

Table 21 – qPCR results of LDH-A knockdown in isolated clones.

Targeting Construct	*Clone*	*Ct value of 18s rRNA*	*Ct value of LDH-A*	*LDH-A Fold Change*
pcDNA6.2-Neg		8.53	17.88	
pcDNA6.2-LDH-A1A2B	A2	8.51	17.38	1.39
pcDNA6.2-LDH-A1A2B	A3	8.16	18.33	-1.77
pcDNA6.2-LDH-A1A2B	B1	8.22	18.44	-1.84
pcDNA6.2-LDH-A1A2B	B4	7.94	19.69	-8.04
pcDNA6.2-LDH-A1A2B	C5	8.39	18.87	-3.98
pcDNA6.2-LDH-A1A2B	C6	8.64	18.11	-1.25
pcDNA6.2-LDH-A1A2B	D1	8.19	17.44	1.07
pcDNA6.2-LDH-A1A2B	D3	8.14	19.35	-2.96

Next, LDH activity was measured using a colorimetric assay. The results are shown in the following table. The same three clones which were identified as promising from the qPCR characterization also showed decreases in LDH activity, and their rank in activity decrease was the same as that determined using qPCR.

Table 22 – LDH activity measurement results.

Targeting Construct	*Clone*	*LDH Activity Fold Change*
pcDNA6.2-Neg		1.04
pcDNA6.2-LDH-A1A2B	A2	1.12
pcDNA6.2-LDH-A1A2B	A3	1.13
pcDNA6.2-LDH-A1A2B	B1	1.11
pcDNA6.2-LDH-A1A2B	B4	-5.70
pcDNA6.2-LDH-A1A2B	C5	-2.46
pcDNA6.2-LDH-A1A2B	C6	1.06
pcDNA6.2-LDH-A1A2B	D1	1.12
pcDNA6.2-LDH-A1A2B	D3	-1.34

As an additional check, PCR primers were used to amplify a portion of the pcDNA6.2 construct, in order to confirm integration of the targeting construct into the host chromosome. The targeting construct plasmid DNA was used as a positive control, while the parental (untransfected) cell line was used as a negative control. Figure 71 shows that single bands were amplified from the both the plasmid and the NS0 genomic DNA. As a consequence of these analyses, Clones B4, C5 and D3 were chosen for further characterization.

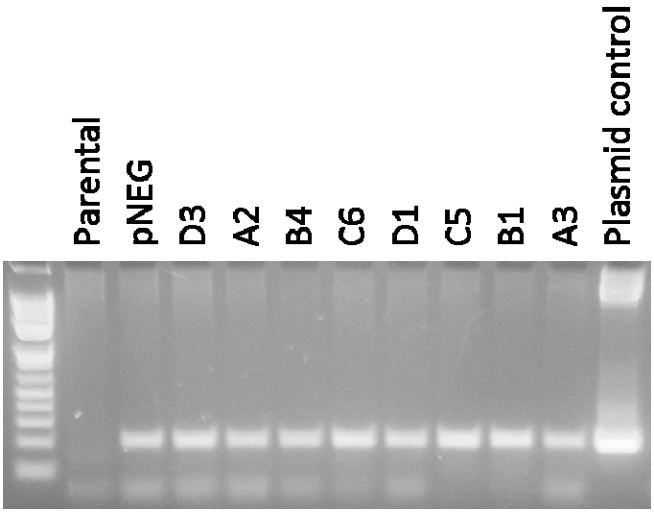


Figure 71 – PCR of gDNA DNA, confirming integration of targeting constructs into host genome.

6.4.2 Characterization of NS0 Cell Lines with Reduced LDH-A Activity

The three clones with the greatest reduction in LDH-A activity were characterized in fed-batch cultures, along with the parental cell line and negative control clone. The time profiles of cell concentration, glucose concentration and lactate concentration are shown in Figure 72 and Figure 73, respectively. Cumulative data is shown in order to account for the effects of dilution by the feed volume. The growth profiles of all three LDH knockdown clones were similar to those of the parental and negative control cultures (Figure 72). Interestingly, no significant differences in glucose or lactate profiles were observed in the LDH knockdown clones, as compared to the control clone or parental cell lines (Figure 73). Furthermore, no appreciable differences were observed in specific glucose consumption rates and specific lactate production rates, as demonstrated

by the similar profiles seen in Figure 74 (specific rate is the slope of the line of integral viable cell concentration vs. cumulative substrate).

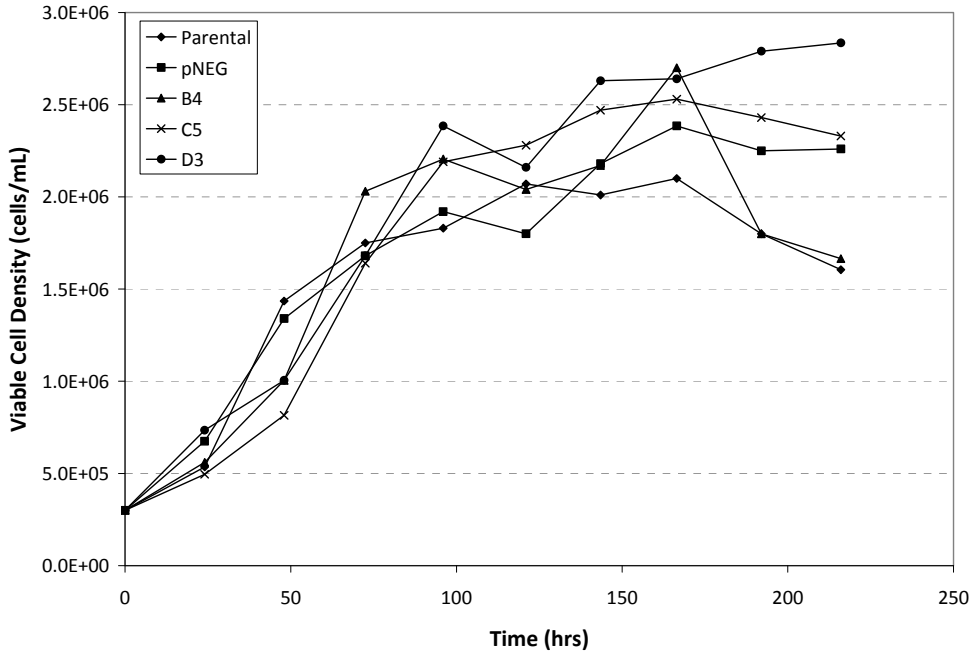


Figure 72 – Viable cell density in fed-batch culture characterization of reduced LDH-A activity clones. Growth performance of parental cell line and negative control (pNEG) is also shown.

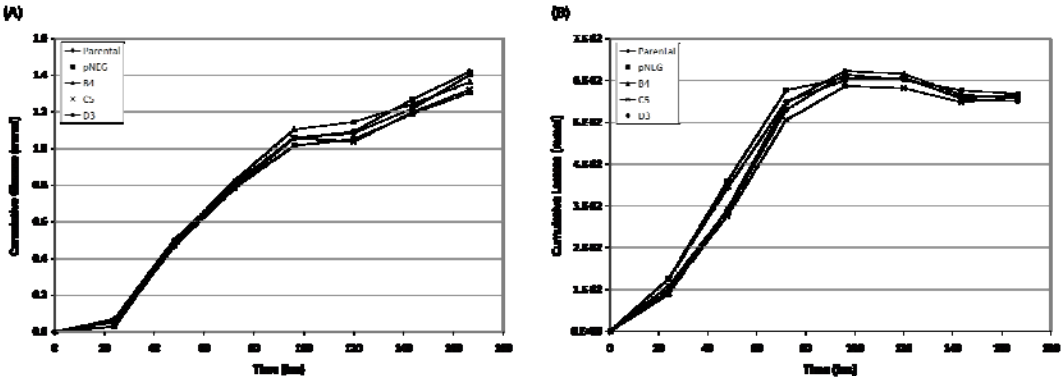


Figure 73 – Cumulative glucose (A) and lactate (B) profiles in fed-batch culture characterization of reduced LDH-A activity clones. Values are also shown for parental cell line and negative control (pNEG).

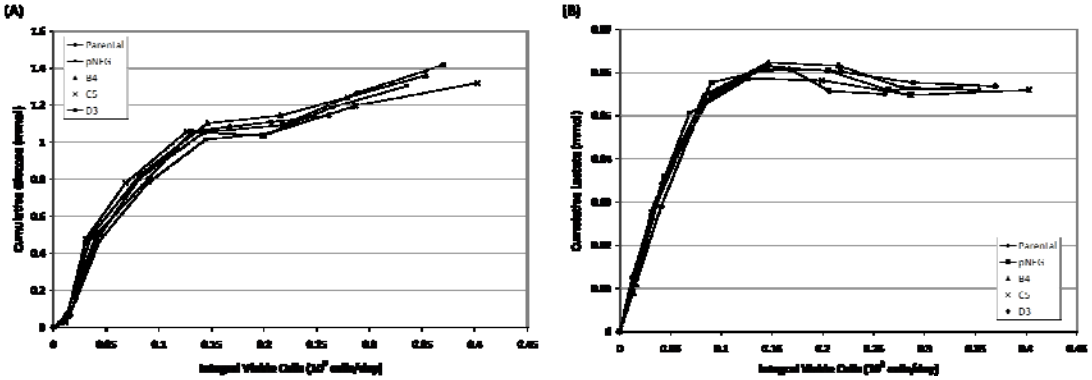


Figure 74 – Integral viable cell versus cumulative glucose (A) and cumulative lactate (B) for all cultures. The slope of the line represents the specific consumption and production rates.

Growth of LDH-A knockdown clones in typical glucose concentrations (4 g/L) did not yield any appreciable differences in their central metabolism as compared to controls. To further characterize these clones, cells were grown in low glucose concentrations for 12 hr. As can be seen in Figure 75, lactate concentrations were similar for all characterized clones at a given glucose concentration. Cell concentrations after 12 hr were similar for all cell clones analyzed at a given glucose concentration (data not shown).

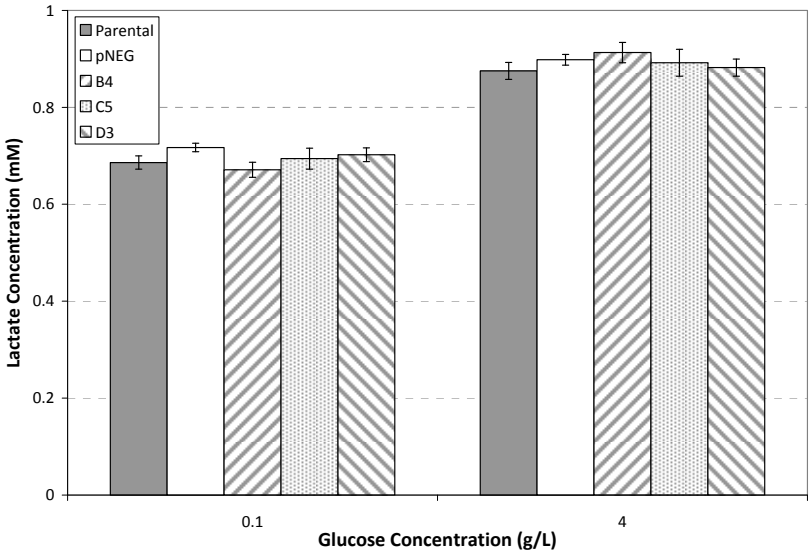


Figure 75 – Lactate concentrations of cells grown in normal (4 g/L) and low (0.1 g/L) glucose concentrations for 12 hr. Data from three LDH knockdown clones, as well as the negative control (pNEG) and parental cell line is shown.

6.4.3 Overexpression of LDH-C Isoform

Next, the C isoform of the murine lactate dehydrogenase gene was overexpressed in the same antibody-producing NS0 cell line used for LDH-A knockdown studies. In parallel, the same construct was transfected into the NS0 clone which exhibited the highest level of LDH-A knockdown, Clone B4 (described above). A control transfection was also performed in both cell lines using pcDNA_3.1_bsd. Stably transfected clones were isolated by limiting dilution cloning in 96 well plates in growth medium containing 4 μg/mL blasticidin for selection. Twenty-four clones from each transfection were expanded for characterization, as well as six clones from the negative control transfections. Overexpression of the target gene was confirmed using qRT-PCR. Expression levels of the LDH-A isoform were also verified and confirmed that the isolated clones derived from transfection of the B4 clone had maintained LDH-A knockdown (data not shown). Results for a subset of characterized clones are shown in Table 23. Significant overexpression was achieved in a number of characterized clones. It should be noted that the Ct values for LDH-C in the parental cell line, as well as each control transfection, are very high, indicating little or no expression of this isoform. These results confirm the microarray expression data presented earlier, which showed that the LDH-C isoform was not expressed in the NS0 cell line. The clones with the greatest degree of knockdown in each transfection (NS0_LDHC_2, renamed PL2 and B4_LDHC_2, renamed BL2) were selected for further characterization.

Table 23 – qPCR results of LDH-C overexpression in isolated clones.

Cell Line	*Clone Name*	*Ct value of GAPDH*	*Ct value of LDH-C*	*LDH-C Fold Change*
Parental cell line		17.22	28.88	
NS0_control	PC2	17.01	28.64	0.94
NS0_LDHC_1	PL1	17.27	26.29	6.23
NS0_LDHC_2	PL2	17.12	20.99	221.3
B4_control	BC2	17.13	28.80	0.99
B4_LDHC_1	BL1	17.12	22.37	84.5
B4_LDHC_2	BL2	17.19	19.69	570.7

Cell Line	*Clone Name*	*Ct value of GAPDH*	*Ct value of LDH-C*	*LDH-C Fold Change*
B4_LDHC_5	BL5	17.09	21.76	126.8
B4_LDHC_12	BL12	16.72	21.02	163.5

6.4.4 Characterization of Isolated Clones

Clones with the greatest degree of LDH-C overexpression in each transfected cell line (parental and LDH-A knockdown) were characterized using fed-batch cultures. Clones isolated from each control transfection were also included in these characterization studies. The time profiles of cell concentration, cumulative glucose and cumulative lactate are shown in Figure 76 and Figure 77, respectively. Viable cell density profiles were similar amongst examined clones (error bars represent data from triplicate cultures), although cell concentration in the control transfection in the parental culture (PC2) declined more readily than other characterized clones. Cumulative glucose profiles were similar, with the exception of the combined LDH-A knockdown and LDH-C overexpressing clone (BL2). No statistically significant differences were observed in lactate profiles (Figure 77B), and all four clones consumed lactate at similar rates in the later stages of the culture (Figure 77D).

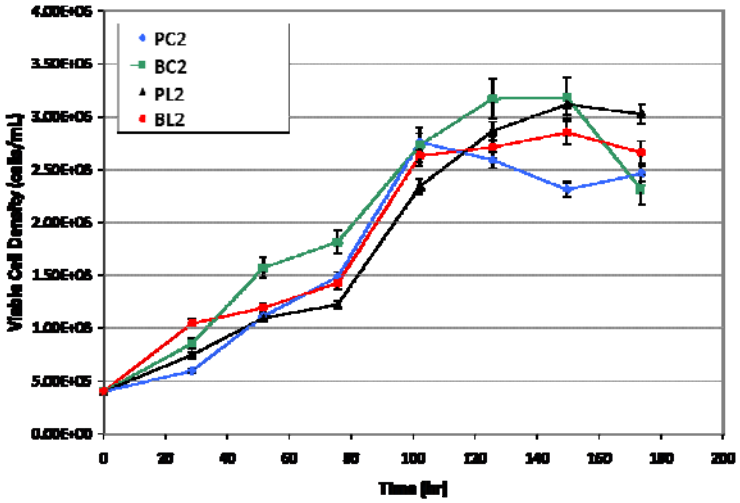


Figure 76 – Viable cell densities in fed-batch culture characterization of LDH-C overexpressing clones.

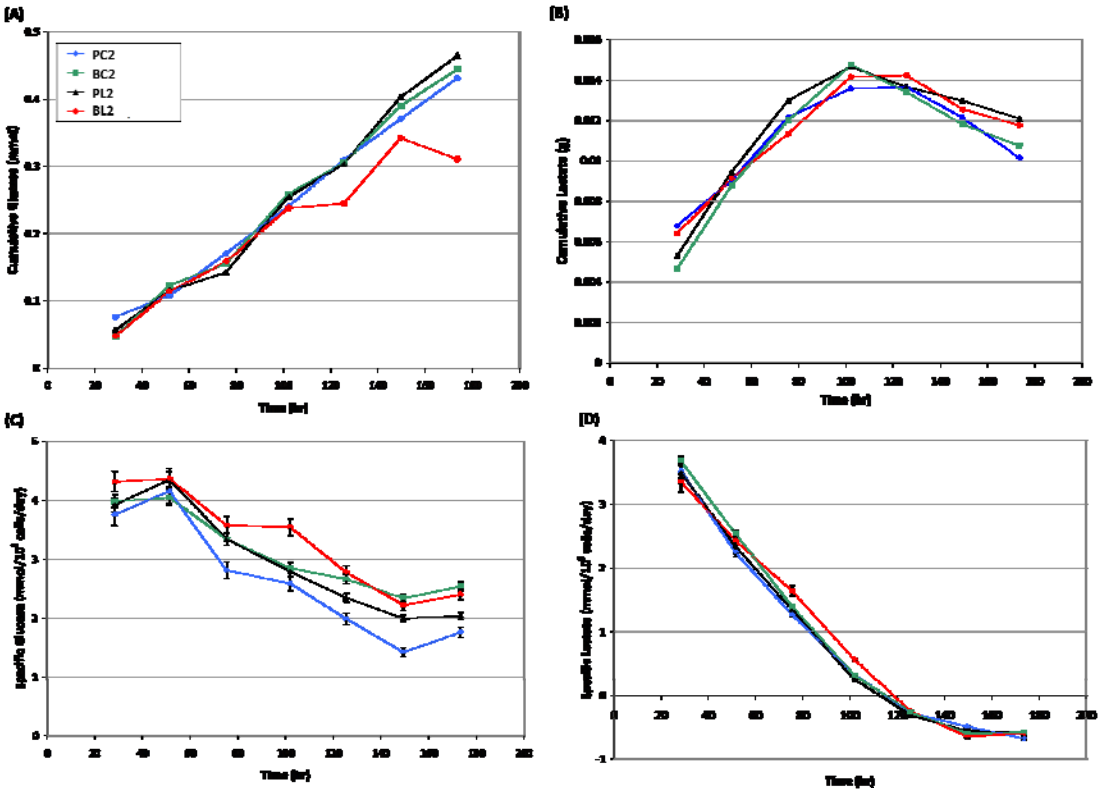


Figure 77 – (A) Cumulative glucose, (B) cumulative lactate, (C) specific glucose consumption, (D) specific lactate production in LDH-C overexpressing clones.

6.5 DISCUSSION

Nutrient utilization in mammalian cells is known to be inefficient, with reports of as much as 88% of the glycolytic flux being diverted to lactate production rather than entering the TCA cycle for energy generation (Brand and Hermfisse, 1997). Several approaches have been employed in an effort to reduce lactate accumulation. One such approach has been the use of a control feeding strategy to maintain glucose concentration at low levels (Wlaschin and Hu, 2006). By controlling glucose concentration at 0.1 g/L, Zhou et al were able to reduce lactate production by more than 7-fold (Zhou et al., 1995). Cell engineering approaches have also been employed to reduce lactate accumulation, including the introduction of the yeast pyruvate carboxylase enzyme into BHK-21 and CHO cells (Irani et al., 1999; Fogolin et al., 2004), and the introduction of the fructose transporter, GLUT5, into CHO cells for use of alternative main carbon sources (Wlaschin and Hu, 2007a).

In this study, we first examined the effects of knocking-down LDH-A mRNA transcripts in an antibody-producing NS0 cell line using a microRNA-based RNA interference platform. This approach allowed us to isolate clones with more than 8-fold downregulation of the LDH-A transcript (Table 22). However, characterization of these clones did not reveal any significant changes in central carbon metabolism, as measured by glucose consumption rates and lactate production rates (Figure 74). This is contrary to published reports where LDH-A was genetically engineered in hybridomas cells (Chen et al., 2001) and CHO cells (Jeong et al., 2001; Kim and Lee, 2007). In the first study, LDH-A activity was disrupted in hybridoma cells using homologous recombination. A 30% decrease in activity was achieved, which resulted in a 50% reduction in specific lactate production (Chen et al., 2001). Jeong et al used antisense mRNAs to reduce LDH-A activity by 29% in CHO cells, leading to a 50% reduction in lactate concentrations (Jeong et al., 2001). Finally, Kim and Lee used siRNAs to downregulate LDH-A activity in CHO cells. They achieved 75 – 89% reduction in mRNA levels, which translated to 45 – 79% decreases in specific lactate production (Kim and Lee, 2007).

It is interesting to speculate as to why a decrease of more than 80% at the LDH enzyme activity level did not translate into any appreciable differences in glucose consumption and lactate production profiles. One possible explanation may be that LDH enzyme levels are present in the cell in such excess, that a reduction of more than 80% still leaves the cell with sufficient enzyme levels to adequately process the incoming glycolytic flux. In order to test this hypothesis, cells were grown in reduced glucose concentrations. Although growth in low glucose conditions did reduce the amount of lactate produced by the cells, no appreciable differences between LDH-A knockdown clones and control were observed. It may be that the level of knockdown achieved in this system was insufficient and that knockout of LDH-A may be required to yield any appreciable changes in central metabolism.

Another possibility could be that in order to capture the impact of LDH-A knockdown, one must examine the cell's transient response to reduced LDH-A transcript levels. It may be that by the time clones with stable knockdown are selected and

expanded (a period of 4 to 6 weeks), the cell may have already compensated for the reduced LDH-A levels. This is a distinct possibily, in light of the role that LDH plays in maintaining the NADH/NAD+ ratio in the cell. The conversion of pyruvate to lactate by LDH leads to the formation of NAD+ from NADH, and this crucial coenzyme is required for multiple reactions in the cell. It has been shown that maintaining the proper NADH/NAD+ ratio is crucial for the cell, and multiple processes are known to be regulated by the availability of these cofactors (Pollak et al., 2007). It is possible that the effects of downregulating LDH-A have already been mitigated by the cell by the time stable clones are isolated. In order to examine this hypothesis, downregulation of LDH-A should be placed under the control of an inducible or dynamic promoter.

We next examined the effects of overexpressing the LDH-C isoform on cellular metabolism. Neither LDH-C overexpression alone, nor combined LDH-C overexpression and LDH-A knockdown yielded any appreciable changes in cellular metabolism. Physiologically, LDH-C is expressed exclusively in the testis (Goldberg, 1990), and recent work has also found a number of LDH-C splice variants to be expressed in human cancers (Koslowski et al., 2002). In germ cells, LDH-C catalyzes the conversion of lactate to pyruvate, which is subsequently used for energy generation. In light of its physiological function, our objective was to evaluate the impact of overexpressing this isoform on central metabolism. In particular, could lactate consumption be enhanced through overexpression of LDH-C? The characterization studies, shown in Figure 76 and Figure 77, did not uncover significant changes in central metabolism. It may be that in order for LDH-C to have an effect, it should be the sole isoform expressed by the cell, as it is in male germ cells (Li et al., 1989). Even though LDH-C was overexpressed in a clone with significant LDH-A knockdown, a comparison of their expression levels based on qRT-PCR results shows that both isoforms have approximately equal Ct values (Ct for LDH-C = 19.69 and Ct for LDH-A = 18.98), and thus approximately equal transcript levels. Further studies in LDH-A knockout cell lines would be required to confirm this hypothesis. No reports currently exist in the literature detailing the effects of LDH-C overexpression in cultured mammalian cells.

6.6 CONCLUSIONS

In this study, cellular metabolism was engineered via manipulation of the LDH enzyme. In a first approach, a siRNA-based methodology was used to reduce LDH-A transcript levels in an antibody-producing NS0 cell line. While an 80% reduction in LDH activity was observed, this did not translate into any appreciable difference in glucose and lactate profiles among knockdown clones. The effects of overexpressing LDH-C, an isoform of lactate dehydrogenase whose kinetic properties favor the conversion of lactate to pyruvate, were next examined. Neither overexpression of LDH-C alone, nor combined overexpression of LDH-C and downregulation of LDH-A yielded any appreciable differences in cellular metabolism. It is possible that during the process of selecting stable clones, changes at the cellular level had already occurred which mitigated the modulation of LDH. As a consequence, it may instructive to observe the transient response of the cell to a reduction in LDH-A transcript copy numbers or an increase in LDH-C transcript levels. This requires the use of dynamic cellular engineering, a topic which is addressed in the following chapter.

7 ENGINEERING GENE EXPRESSION DYNAMICS

7.1 SUMMARY

Genetic engineering of cultured mammalian cells has been used successfully to alter cellular properties, as well as to increase recombinant protein production. Efforts have been focused on introducing metabolic, anti-apoptotic and glycosylation-related genes and significant progress has been reported. Most of these reports have relied on the use of strong, constitutive expression systems. However, dynamic regulation of genetically engineered genes could potentially improve on these efforts by delivering varying gene dosage levels. While inducible expression systems can be used to fine-tune target gene expression, the addition of external inducer molecules could potentially be detrimental. The use of endogenous promoters which display dynamic behavior can circumvent this problem. To this end, we have used transcriptome data mining of time series microarray data to identify candidate promoters which display dynamic gene expression. Clustering of gene expression patterns revealed hundreds of genes with varying dynamics across a wide intensity range. The promoters of genes identified in this analysis can potentially provide a more flexible and dynamic control of the transgenes. Following identification of these dynamic genes, we selected one candidate promoter from the thioredoxin interacting protein (Txnip) gene, and validated its expression dynamics in CHO using a promoter-reporter construct. We found that reporter gene expression patterns closely matched those obtained from microarray experiments. This approach illustrates a novel concept in metabolic engineering which can potentially be used to achieve dynamic control of cellular behaviors for enhanced process characteristics.

7.2 INTRODUCTION

Genetic engineering of mammalian cells can be used to modulate multiple aspects of cellular behavior, including growth, metabolism and secretion. In fact, a number of studies have been published which apply genetic engineering to recombinant protein-producing cells, often with the ultimate aim of improving product titers, or imparting

some other beneficial phenotypic trait. Notable examples include the overexpression of anti-apoptotic factors Aven and e1B-19K in CHO cells to prolong culture duration, leading to increased antibody titers (Figueroa et al., 2007), and the downregulation of alpha1,6-fucosyltransferase in CHO cells to reduce antibody fucosylation, which has been shown to decrease the effectiveness of the resulting antibody product by decreasing ADCC (antibody-dependent cell-mediated cytotoxicity) activity (Mori et al., 2004).

Typically, target genes are manipulated using strong, constitutive promoters, either overexpressing the target gene, or overexpressing a small interfering RNA sequence to downregulate the target gene. However, it may be advantageous to dynamically control gene targets in order to study the temporal response of cellular behavior to the genetic manipulation. Furthermore, certain cell engineering modulations could be more beneficial to the cell if implemented only in certain growth phases, such as the use of cell cycle arrest for a biphasic production strategy (Fussenegger et al., 1998; Watanabe et al., 2002).

A number of inducible gene regulation systems which rely on the addition of an external inducer molecule have been developed and successfully applied in multiple applications, including cell engineering (reviewed in (Weber and Fussenegger, 2007)) and gene therapy (reviewed in (Toniatti et al., 2004)). However, the addition of such external inducers could potentially trigger unwanted changes in cellular behavior. Consequently, the use of endogenous promoters which naturally display dynamic behavior would be well-suited for cell engineering applications. One approach to identify candidate promoter genes is to mine historical time series microarray data. The use of data mining techniques, such as principal component analysis, can identify gene expression trends, and quantify their reproducibility. Once candidate genes which display the desired expression trend are identified, the next step is to isolate the promoter region of the target gene. The core promoter region is defined as the minimal portion of the promoter required for transcription initiation, and encompasses the RNA polymerase binding site. Primary regulatory elements, such as specific transcription factor binding sites, are usually contained within the proximal promoter region, which extends beyond

the core promoter, upstream of the gene of interest. Additional regulatory elements may also be present further upstream.

One potential biological function which could be targeted for dynamic control is apoptosis. It is known that commonly used industrial cell lines such as CHO and NS0 have conserved many components of the apoptotic machinery, and that apoptosis is a major contributor to cell death in bioreactor cultures (Arden and Betenbaugh, 2004). Anti-apoptotic cell engineering approaches are therefore attractive options to prolong culture duration with the ultimate aim of increasing product titer. However, in such an approach, it would be desirable to overexpress an anti-apoptotic gene only in the later stages of the culture, with the onset of apoptosis. Consequently, the expression of such an anti-apoptotic gene would ideally be driven by a dynamic promoter which shows minimal expression in the lag and exponential phases of the culture, and whose expression is turned on in the stationary phase.

In this study, we mined historical microarray data accrued from multiple CHO fed-batch cultures in order to identify genes whose expression is dynamic with time and reproducible across multiple cultures and experimental conditions. In order to validate the concept of dynamic cellular engineering, the promoter region of an identified candidate gene was used to drive expression of a reporter gene, thus confirming the predicted dynamic gene expression behavior. This proof of concept study represents a first step in dynamic cell engineering.

7.3 MATERIALS AND METHODS

7.3.1 Microarray Data Mining

A total of 72 microarray samples from twelve fed-batch CHO cultures were used to identify genes with temporally dynamic behavior. These samples, hybridized on CHO version 2 Affymetrix arrays, have been described in Chapter 5. All samples were normalized using quantile normalization, where the target intensity distribution was the average intensity across all samples for every gene. Genes with a maximum intensity $\leq$ 70 and a detection p-value $\geq$ 0.04 across all samples were removed from further analysis. Genes with a coefficient of variation of $\leq$ 20% were also removed from further analysis.

Gene expression values were preprocessed to have the following format: for each gene, a 12 by 6 matrix of $\log_2$-mean centered expression values was created, representing the 12 reactor runs, each with six time points. Principal component analysis was performed on every gene using the algorithm described by (Alter et al., 2000), using a total of 3 principal components. For every gene, the values of the three principal components were extracted, as well as the variance captured by each principal component. Genes for which the first principal component captured less than 80% of the variance were excluded from further analysis.

Next, k-means clustering was performed (Everitt, 1974) using each gene's first principal component eigenvector as an input. A total of 6 clusters were specified. Finally, gene ontology enrichment analysis, as implemented in GenMAPP's MappFinder (see Section 3.3.4 for further details) was used to identify those functional classes significantly enriched amongst each of the six identified clusters. Gene ontology terms with an enrichment p-value ≤ 0.05 were considered significant.

7.3.2 Promoter Construct

Thioredoxin interacting protein (Txnip) was identified as the first candidate promoter target. The human promoter sequence of Txnip, consisting of regions 142926719:142927671 on Human Chromosome 1, was obtained from SwitchGear Genomics (Menlo Park, CA). This 952 bp fragment drives the expression of a destabilized luciferase reporter gene, luc2P (Promega, Madison, WI). This construct was named pSGG_HS_Txnip. This construct lacks a mammalian selectable marker. Consequently, an additional plasmid was used during transfection to provide the mammalian selectable marker, pcDNA3.1_bsd (Invitrogen, Carslbad, CA), which encodes a blasticidin resistance gene under the control of an EM7 promoter.

7.3.3 Stable Clone Generation

The CHO cell line used in this study produces a recombinant antibody product and has been described elsewhere (Figueroa et al., 2007). To generate stable clones, this cell line was co-transfected with both constructs in a 6:1 (pSGG_HS_Txnip:pcDNA3.1_bsd) molar ratio. Briefly, 20 μg of reporter plasmid

(pSGG_HS_Txnip) and 3.3 μg of pcDNA3.1_bsd were linearized by digestion with *BamH I* and *Pci* I, respectively, and purified using QIAQuick PCR purification spin columns (Qiagen, Valencia, CA). Five million exponentially growing cells were transferred to 1 mL of EX-CELL CHO Cloning media (Sigma-Aldrich, St. Louis, MO), and electroporated in 4 mm electroporation cuvettes (BioRad, Hurcules, CA). Electroporation was performed in the BioRad Gene Pulser XceII (BioRad). The electroporation conditions used were 300 V, 250 μF and infinite resistance. Transfected cells were transferred into pre-warmed EX-CELL CHO cloning media (Sigma-Aldrich, St. Louis, MO). Transfection efficiency was determined using FACS by parallel transfection with an EGFP-containing plasmid.

Twenty-four hours following transfection, cells were diluted in 96-well plates at 2000 cells/well, in 0.2 ml of growth medium (CHO SSFM-II, Invitrogen, Carlsbad, CA) supplemented with 4 μg/mL blasticidin. Plates were incubated for 10 – 14 days in a 37°C, 5% CO_2 environment. Clones were expanded for characterization in their selective medium.

7.3.4 Clone Characterization

Stable clones were characterized for their luciferase activity by batch cultivation in shaker flasks. Briefly, 3 x 10^5 cells were seeded in growth medium, and samples were taken daily for luciferase quantification and cell counts. For luciferase quantification, 1 x 10^5 cells were aliquoted and washed twice with phosphate-buffered saline solution. Cells were lysed with 250 μL passive lysis buffer (Promega). Lysates were incubated at room temperature for 15 min and centrifuged at 4°C (1500 rpm for 30 s). Luciferase reporter gene activities were determined using the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's instructions. Light intensity was measured using a Synergy HT luminometer.

7.4 RESULTS AND DISCUSSION

7.4.1 Identifying Candidate Promoters

Our first objective was to identify genes which showed reproducible dynamic behavior during a typical fed-batch culture, irrespective of experimental conditions. In other words, we wanted to pinpoint which genes were invariant in all experimental conditions tested. The promoters of such genes represent attractive promoter candidates, given that we observed their expression dynamics to be highly reproducible across multiple experimental conditions. In order to identify these genes, we used multiple iterations of principal component analysis. As described in the materials and methods, we first performed principal component analysis for every gene, where the input was a 12 by 6 expression data matrix, representing the 12 reactor runs, each with six time points. We then examined the percent variance captured for each gene by the first principal component. If a gene's expression is similar in all runs, then the first principal component will capture a significant portion of that gene's variance. Consequently, we filtered out genes where the first principal component did not capture at least 80% of the observed variability. This then allowed us to perform unsupervised clustering, using the k-means algorithm, to identify the predominant expression patterns in this set of principal components. A pictorial representation of the methodology is shown in Figure 78.

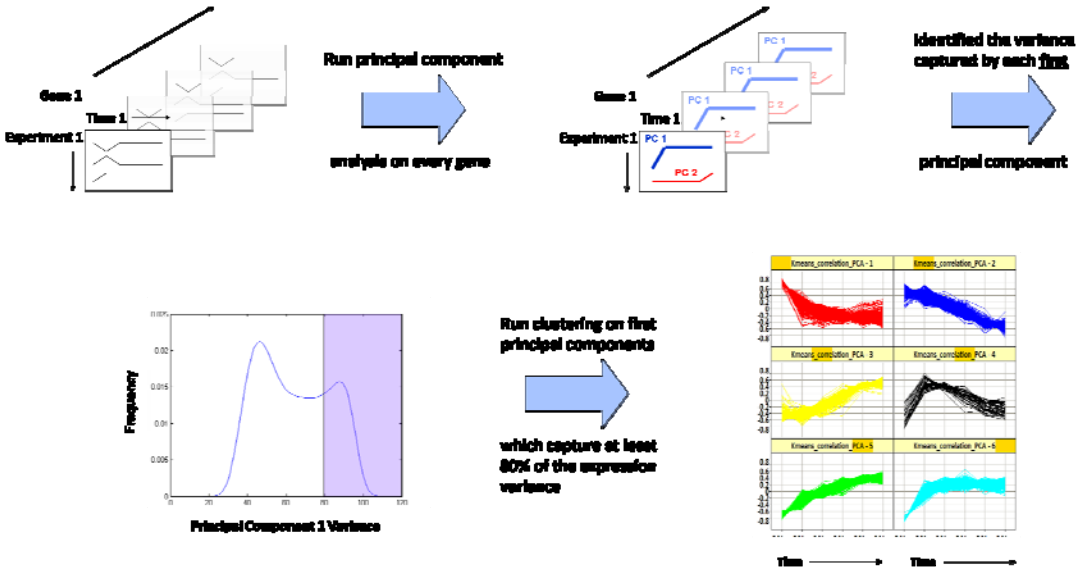


Figure 78 – Principal component analysis-based methodology to identify gene expression variability.

The results of the k-means clustering analysis are shown in Figure 79. Functional analysis of the resulting clusters was performed using gene ontology enrichment analysis. As can be seen in Table 24, a number of functional classes are significantly enriched within each cluster. Of note, cell cycle genes display a monotonic decreasing trend with time, while the expression of apoptotic genes is increased as the cells enter the stationary phase.

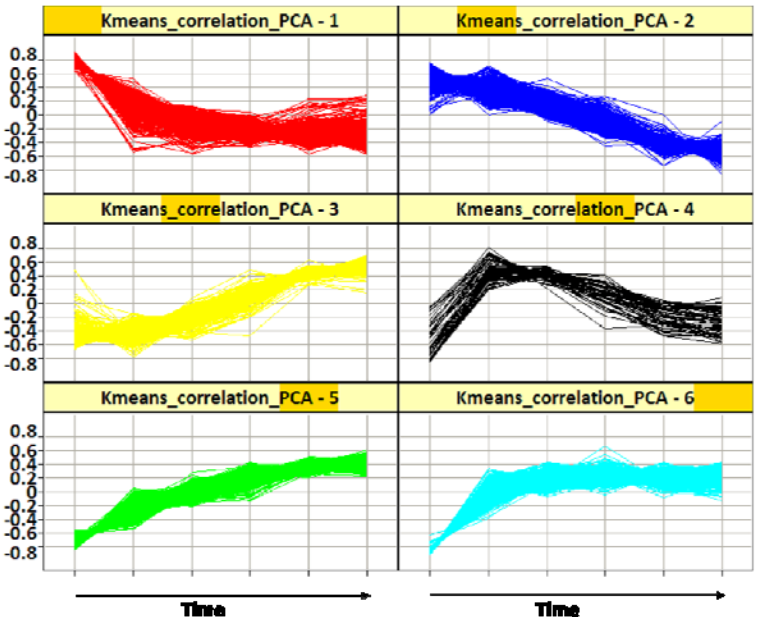


Figure 79 – K-means clustering results.

Table 24 – Enriched gene ontology terms in identified k-means clusters.

K-Means Cluster	***No. of genes in cluster***	***Enriched Gene Ontology Classes***
1	683	DNA binding, Methyltransferase activity, Microtubule motor activity, Chromosome, Cytoskeleton, Cell cycle including DNA replication, M phase, regulation of cell cycle, mRNA transport, Lipid biosynthesis, mRNA processing, RNA splicing, DNA repair
2	697	Lyase activity, Ribosome biogenesis, Cell cycle including DNA replication, M phase, regulation of

K-Means Cluster	*No. of genes in cluster*	*Enriched Gene Ontology Classes*
		cell cycle, mRNA transport, cholesterol metabolism, amine metabolism, mRNA processing, positive regulation of transcription
3	429	Actin cytoskeleton, vacuole, vesicle-mediated transport, fatty acid metabolism, sphingolipid metabolism,
4	54	No significantly enriched terms
5	757	Glutathione transferase activity, cation transporter, endoplasmic reticulum, vacuole, Notch signaling pathway, positive regulation of apoptosis, cation homeostasis, intracellular protein transport, proteolysis and peptidolysis
6	395	Apoptosis

Once we had identified the major temporal gene expression patterns present in the data, we focused on isolating genes with our desired expression profile: low expression levels in the lag and exponential phases and significantly higher expression in the stationary phase. Consequently, we took a closer look at the gene members of k-means clusters 5 and 6. Genes included in these clusters span a wide range of intensities and induction fold-changes and an example of genes spanning low, mid and high intensity ranges is shown in Figure 80. By considering gene expression intensity, it may be possible to select candidate promoters which will yield varying gene expression levels.

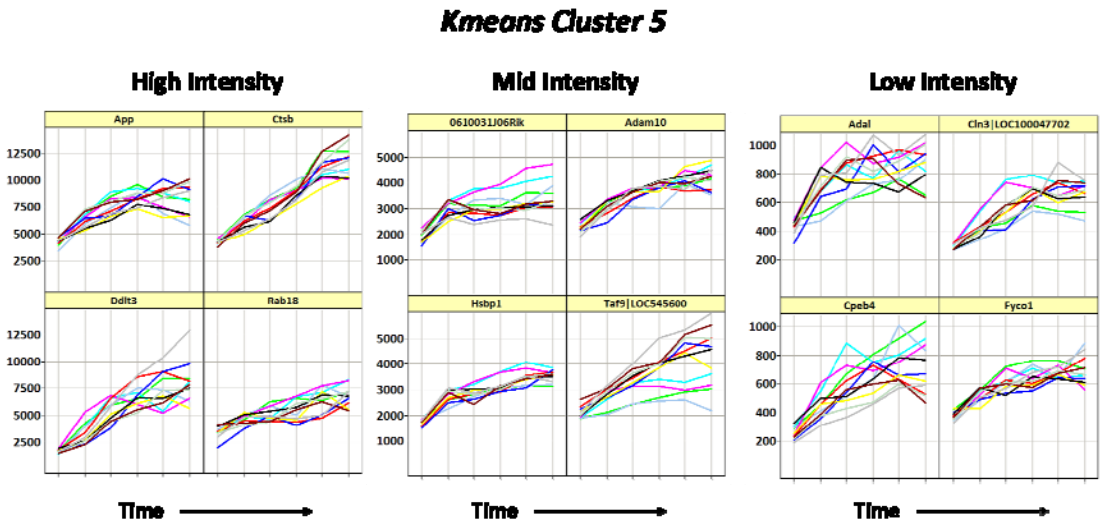


Figure 80 – Example of genes in Kmeans cluster 5 spanning various intensities in their dynamic range. Each subplot shows the expression of a different gene, and each colored line represents the expression of that gene in a particular time course experiment.

As a first candidate promoter, we chose thioredoxin interacting protein (Txnip), whose expression profile is shown in Figure 81. The expression of this gene is reproducible across multiple fed-batch cultures, and is also highly expressed, as judged by the high intensity values. Therefore, the use of this gene's promoter region to drive expression of a target gene should result in high expression levels.

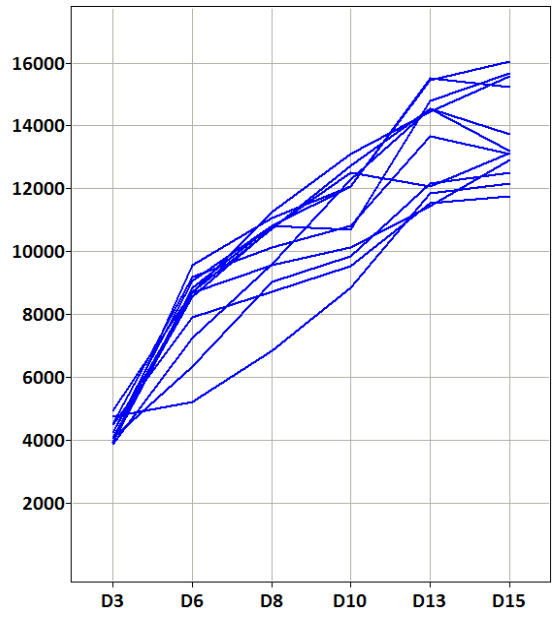


Figure 81 – Temporal expression profiles of Txnip gene across multiple CHO fed-batch cultures.

7.4.2 Characterization of CHO Cell Lines Stably Expressing the Txnip/Luciferase Reporter Construct

Once the candidate promoter had been identified through microarray data mining, our next objective was to demonstrate its ability to drive expression of a reporter gene and to investigate whether the same expression profiles which were observed in microarray experiments could be seen in the reporter gene. As an initial proof of concept study, we chose to use the promoter region of the human Txnip gene, rather than isolate the Txnip promoter region from CHO. This decision was done in part to accelerate the process of validating the proposed concept, and was further supported by evidence in the literature that the human Txnip promoter sequence was active in hamster cultured cell lines (Minn et al., 2005). A reporter construct, with the 952 bp fragment immediately upstream of the human Txnip gene driving the expression of a destabilized luciferase gene, was obtained (SwitchGear Genomics). Stable CHO clones were generated by co-transfecting the reporter construct with a second plasmid providing a mammalian resistance marker, followed by limited dilution cloning and growth in selective media.

Stable clones were characterized for their growth and luciferase activity, and the results for two of these clones are shown in Figure 82. Ideally, the specific luciferase activity, shown in red in the figure, should closely follow that of the gene expression profiles shown in Figure 81. Indeed, both clones show the desired profiles, namely an increase in activity in the later stages of the culture. However, it should be noted that the absolute luciferase activity (shown in blue) shows an order of magnitude difference between the two clones. Such clonal variability is inevitable when analyzing stable clones isolated from a single transfection. One source of variability is the location of transgene integration. It is well-known that integration site can significantly affect the resulting gene expression levels (Wurm and Petropoulos, 1994), and could explain the varying luciferase levels observed in these clones. Therefore, cell clone screening becomes a necessary component when using dynamic promoters for cell engineering applications.

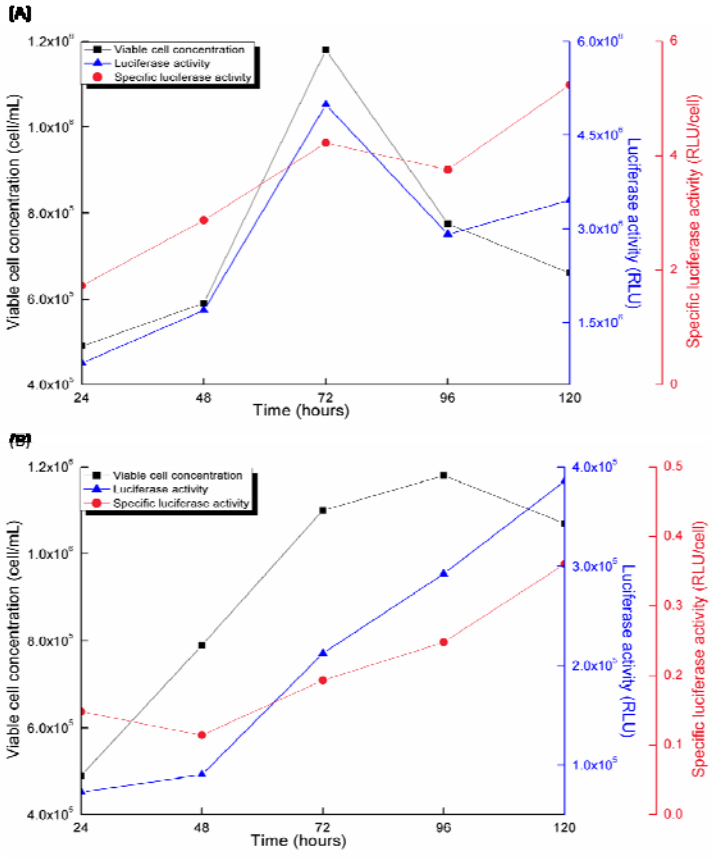


Figure 82 – Cell growth, absolute and specific luciferase activity in two CHO clones (panels (A) and (B)) stably expressing the human Txnip luciferase reporter construct. These figures were generated by Huong Le.

7.5 CONCLUSIONS

This study represents a proof-of-concept experiment for gene expression under the control of dynamic promoters. Through transcriptome data mining, we were able to identify a number of candidate promoters, based on the expression profiles of the genes they express. It is interesting to note that these show a wide dynamic range during the course of a culture, and are reproducible across a number of experimental conditions. Further experiments are currently underway to use this promoter to drive expression of an anti-apoptotic gene (Aven) in CHO cells. Work is also being done in parallel to isolate the Chinese hamster Txnip promoter region and compare its performance to that of the human promoter. This study shows the utility transcriptome data mining and we believe that dynamic gene regulation for cellular engineering applications is an area that holds much promise.

8 THE USE OF NEXT-GENERATION TECHNOLOGIES FOR TRANSCRIPTOME ANALYSIS

8.1 SUMMARY

While high throughput sequencing technologies have only recently become widely available, they have already altered the landscape of transcriptome characterization. In this chapter, we describe the use of the Illumina Solexa GAII platform to characterize the transcriptome of an antibody-producing Chinese hamster ovary (CHO) cell line. More than 55 million sequencing reads were generated and mapped to an existing set of CHO unigenes derived from expressed sequence tags (ESTs), as well as several public sequence databases. The extensive depth of sequencing revealed the wide dynamic range of transcripts inside the cell, spanning more than six orders of magnitude. A large fraction of sequences unearthed have not been previously reported or seen in our existing sequence repertoire. With the depth that high throughput sequencing methods can reach, one can expect that the entire transcriptome of this industrially important organism will be decoded in the near future.

8.2 INTRODUCTION

Advances in genomic technology are causing transformational changes in health care, agriculture and many other sectors of industry. Gradually, their impact on the bioprocessing industry is also becoming apparent. Transcriptome analysis and genome-level characterization are increasingly being used to capture process dynamics and identify cellular features related to key performance parameters. Genome sequencing of producing organisms is progressively being employed to identify key genetic traits for genetic manipulation and generation of superior producers. To date, efforts on genomic applications in bioprocess research have been largely confined to microbial species. Other than cells of human or mouse origin, genome sequencing has not been feasible on mammalian cell producers because of their large genome size and the associated cost of sequencing. A notable example are Chinese hamster ovary (CHO) cells, for which little genomic information is available, despite their economic importance due to the value of the therapeutic proteins produced in these cells.

Current efforts to develop genomic tools for CHO cells have focused on sequencing expressed sequence tags (ESTs), including the project undertaken by us in conjunction with the Consortium on Chinese Hamster Ovary Genomics (Jayapal et al., 2007b). Through traditional phage libraries and Sanger sequencing-based methodologies, we have established a collection of more than 70,000 ESTs which collapse to more than 28,000 unigenes (Kantardjieff et al., 2009b). This sequence repertoire has enabled us to construct both cDNA- and oligonucleotide-based microarrays, which have been used to analyze transcriptome dynamics of CHO cells under different culture conditions which affect growth and productivity (Yee et al., 2008a; Kantardjieff et al., 2009a). Similar EST sequencing efforts in CHO cells have generated valuable resources. However, if the number of different transcripts found in mouse (Carninci et al., 2005) is used as a reference, current CHO EST sequence repertoires still only capture a portion of the total transcripts expressed in Chinese hamster. This slow pace in reaching the depth of the transcriptome is likely to drastically accelerate as high throughput sequencing technologies become increasingly affordable. Indeed, ultra high throughput sequencing methodologies, such as those commercialized by Illumina Solexa and ABI Solid, can now deliver tremendous sequencing depth at a fraction of the cost of traditional Sanger sequencing.

Given their ability to reach high sequencing depths, these sequencing methodologies are increasingly being applied to assess transcript abundance levels in the entire transcriptome by recording the frequency of occurrence of sequence fragments from each transcript, a technique commonly referred to as RNA-Seq (Mortazavi et al., 2008). Reports of RNA-Seq applications include transcriptome studies in mouse (Mortazavi et al., 2008; Rosenkranz et al., 2008), human (Marioni et al., 2008; Mudge et al., 2008; Sultan et al., 2008; Wang et al., 2008) and yeast (Nagalakshmi et al., 2008; Wilhelm et al., 2008). In this study, we used RNA-Seq to examine the transcriptome of a CHO cell line producing high levels of recombinant IgG molecules. The unprecedented sequencing depth that was reached revealed much that was not previously known in CHO cells. These insights are likely to spur other researchers to take a similar route in further exploring the transcriptome and genome of CHO cells.

8.3 MATERIALS AND METHODS

8.3.1 Cell Culture and RNA Preparation

The same antibody-producing CHO cell line was used for both Illumina sequencing and Affymetrix hybridizations. This cell line overexpresses two anti-apoptotic genes and has been described elsewhere (Figueroa et al., 2007). Cells were grown in CHO SSFM II media (Invitrogen, Carlsbad, CA) and harvested in mid-exponential phase. Total RNA was isolated using the RNeasy mini kit (QIAGEN, Valencia, CA) according to the manufacture's protocol, using the optional on-column DNAse digestion step. RNA quality was verified using a 2100 Bioanalyzer RNA Nanochip (Agilent, Santa Clara, CA) and all samples had a RIN (RNA integrity number) greater than 9.

8.3.2 Illumina Library Preparation and Sequencing

Total RNA samples were sent to the National Center for Genome Resources (Santa Fe, NM) for Illumina Solexa sequencing following the protocol for cDNA sample sequencing. Briefly, poly A+ RNA was isolated from 10 μg total RNA via two rounds of oligo-dT selection (Invitrogen, Carlsbad, CA) and reverse transcribed using random hexamers. Adapters complementary to sequencing primers were annealed to cDNA fragment ends and the resultant cDNA library was size fractionated using agarose gel electrophoresis. 200 bp fragments were excised and subjected to 15 rounds of PCR amplification using primers complementary to the previously-ligated adaptors. Single-stranded cDNA fragments were then annealed to the flow cell surface in a cluster station (Illumina, San Diego, CA). 46 cycles of sequencing-by-synthesis was performed and high quality sequences (those passing default quality filtering parameters in the Illumina GA Pipeline GERALD stage) were retained for further analysis. FASTQ files were converted to FASTA files and filtered to remove platform-specific primer sequences.

8.3.3 Mapping and Transcript Abundance Calculation

High quality reads were aligned to reference assemblies using the RMAP software (Smith et al., 2008) with a seed size of 12 bp and allowing up to 2 mismatches.

Reference assemblies used in this study include our current set of previously generated CHO EST sequences containing 28,914 unique sequences (Kantardjieff et al., 2009b), Ensembl v48 cDNA collections for mouse, human, rat and squirrel, the FANTOM3 collection of mouse cDNAs (Carninci et al., 2005), the Repbase Update database of eukaryotic repetitive sequences (Jurka et al., 2005) and the NONCODE collection of noncoding RNAs (He et al., 2008). Reads that mapped to multiple genes in the reference sequence set (known as multireads) were treated as follows: the short read abundance for each gene was first determined based on unambiguously mapped sequences. Multireads were subsequently allocated based on the proportion of unique Solexa sequences identified for each gene, following the method described by (Mortazavi et al., 2008). In other words, if the same Solexa read x is mapped to gene A and gene B, and if gene A is mapped to uniquely by a total of 40 reads and gene B is mapped to uniquely by a total of 60 reads, then read x will be counted 0.4 times for gene A and 0.6 times for gene B. Abundance was reported as either an absolute count of frequency or a length normalized count, defined as the total number of reads aligned to a target sequence divided by the length of the target sequence. The RMAP algorithm was run by Faraaz Yusufi at the Bioprocessing Technology Institute in Singapore.

8.3.4 Affymetrix Array Hybridization and Processing

Aliquots from the same total RNA sample used for Illumina sequencing were hybridized to custom CHO Affymetrix version 2 arrays with 23,070 probesets. Briefly, biotin-labeled cRNA was prepared from 10 μg of total RNA using the one-cycle target labeling kit (Affymetrix, Santa Clara, CA) according to the manufacturer's instructions. Labeled, fragmented cRNA was delivered to the University of Minnesota Affymetrix Microarray Core Facility for hybridization. CEL files were processed using the GeneData Expressionist Refiner module (GeneData, San Francisco, CA), which was used to assess overall array quality and to obtain condensed single intensity values for each probeset using the Microarray Analysis Suite Statistical Algorithm (MAS 5.0). Mean intensity values were linearly scaled to a value of 500.

8.3.5 Functional Annotation

Gene ontology annotations were extracted from each gene identifier mapped to by an Illumina read using Ensembl (http://www.ensembl.org). Gene ontology enrichment analysis was performed for genes transcribed at different levels using the Gene Ontology Browser tool in Spotfire Decision Site 9.1. Gene ontology definition files dated 09/08 were downloaded from the Gene Ontology Consortium website (http://www.geneontology.org) and used for the analysis. A hypergeometric test was performed using default parameter values between the test group and all annotated genes to obtain enrichment p-values. Gene ontology terms with a p-value ≤ 0.05 were considered as significantly enriched.

8.3.6 Visualizing Data in a Genomic Context

High quality reads were mapped to Ensembl v48 mouse cDNA sequences, as described above. Annotations and mouse genomic coordinates were extracted for all sequences which were confidently mapped to by RNA-Seq tags (≤ 2 mismatches). Data was visualized along mouse genome coordinates using Spotfire Decision Site 9.1.

8.4 Results and Discussion

A total of 2.45 Gbp from 49.8 million 46 bp reads from a full sequencing run and 6 million 36 bp reads from a titration run were obtained. A large percentage (48.2%) of these reads mapped to ENSEMBL mouse cDNAs. Among these, 41.2% also mapped to our collection of 28,914 CHO unigenes. Additionally, 17.6% mapped to our CHO unigene set but not to ENSEMBL mouse. Another 5% aligned to various other databases detailed in the Materials and Methods, while the remaining 30% of reads were unmapped.

To characterize the transcriptome dynamics in CHO, we examined the mapping frequency of Illumina Solexa reads to our set of 28,914 CHO unigenes. More than 99% of CHO unigenes were mapped to by at least one Illumina Solexa read. Of the remaining 391 unmapped CHO unigenes, 80.1% were originally sequenced from either a tissue-derived cDNA library (brain or spleen), or a cDNA library prepared from chemically-

treated RNA (e.g. tunicamycin, sodium butyrate or 5-azacytidine). This suggests that these transcripts may be absent under normal culture conditions, and only expressed following drastic transcriptional changes incurred by chemical treatment, or they may only be present in tissues. Figure 83 shows the frequency distribution of mapped Illumina Solexa reads to CHO unigenes. Approximately 20% of CHO unigenes (5783 sequences) were mapped to by only one Illumina Solexa read. Of these, 85.2% had been identified only once (i.e. singletons) in previous Sanger EST sequencing. Therefore, both sequencing methodologies identified these as low abundance transcripts. 17.2% (994 of 5783) of CHO unigenes mapped to by one Illumina Solexa read had an ortholog in human, mouse or rat. Gene ontology terms assigned to these low abundance orthologous transcripts were examined. Genes localized to the plasma membrane, as well as genes with extracellular products were highly represented. Furthermore, more than 20% (229 genes) were classified as transcription factors. This observation is consistent with the notion that transcription factors are typically found in low abundance (Holland, 2002).

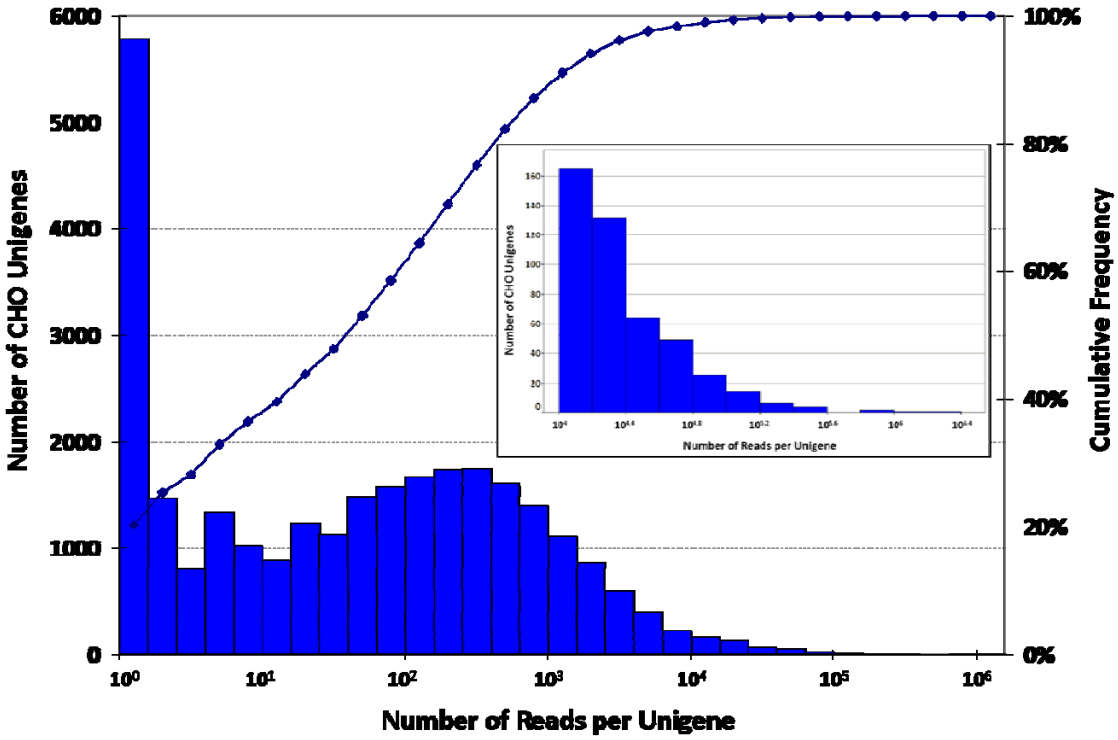


Figure 83 – Histogram of number of mapped Illumina Solexa reads per CHO unigene. Note that the bins are plotted in logarithm scale; within each logarithm bin, size is the same. However, bin size increases by ten fold in each logarithm increment. The inset shows an expanded view of unigenes with a mapping frequency greater than 10^4. The cumulative frequency is shown on the right y-axis.

The multiplicity of reads derived from a unigene (i.e. number of reads per unigene) is indicative of its transcript abundance level. Since the length of each unigene varies, some have reported abundance as number of reads per kilobase per million mapped reads (RPKM) to allow for normalization to transcript size (Mortazavi et al., 2008). The number of reads mapped to each unigene in our repertoire was found to vary over a wide dynamic range, from one to more than 1.6 million. The most abundant transcripts as assessed by reads per unigene are listed in Table 25 (including three sequences having no known orthologs in human, mouse or rat). The most abundant unigene identified was the heavy chain of recombinant IgG, with more than 1.6 million reads or 3.1×10^4 RPKM. The light chain of recombinant IgG, with an abundance level of 2.9×10^4 RPKM, closely follows the heavy chain. As a comparison, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and beta-actin have 1.9×10^4 RPKM and 5.4×10^3 RPKM, respectively. The high abundance level of both heavy and light chain transcripts in hyperproducing cells is consistent with similar observations in both CHO (Lattenmayer, 2007) and mouse myeloma cells (Dorai, 2006); suggesting that abundance of heavy and light chain mRNA transcripts may be important to high productivity. Ontology terms of the 500 most abundant CHO unigenes, as determined by Illumina Solexa read mapping, were analyzed and found to be highly represented in ribosomal genes, as well as members of glycolysis and the electron transport chain, lysosomal genes and nucleosome assembly genes.

Table 25 – Most abundant CHO unigenes identified through mapping of Illumina Solexa reads to CHO unigene dataset.

Gene Description	***Illumina Read Abundance***	***EST Sequence Length (bp)***
Recombinant immunoglobulin heavy chain C gene segment	1,644,460	1,964
N.A.[1]	1,003,265	253
Recombinant immunoglobulin Kappa light chain C gene segment	753,593	962
Glyceraldehyde-3-phosphate dehydrogenase	670,311	1,301
cDNA sequence BC005624	297,190	1,290
N.A.[1]	292,815	1,265
Heat shock protein 5	247,121	2,529
Eukaryotic translation elongation factor 1 alpha 1	226,264	1,834
Ferritin heavy chain 1	184,459	934

Gene Description	*Illumina Read Abundance*	*EST Sequence Length (bp)*
Cricetulus griseus Csr1	181,157	1,148
N.A.[1]	165,898	890
Pyruvate kinase, muscle	162,868	865
Actin, beta, cytoplasmic	151,084	1,026
Enolase 1, alpha non-neuron	132,826	1,762
Calreticulin	132,136	1,917

[1] *N.A.: No annotation available*

To assess the performance of Illumina sequencing for transcriptome analysis, the same RNA sample was used for sequencing and hybridization onto CHO Affymetrix arrays. The number of reads mapped to each unigene is plotted against the intensity of the corresponding probe sets on the Affymetrix CHO microarray (Figure 84). Overall, a good correlation is seen between the abundance levels assessed by the two methods, with a Spearman correlation coefficient of 0.78. This degree of correlation is consistent with published results (Marioni et al., 2008; Rosenkranz et al., 2008; Wilhelm et al., 2008). The correlation in the high intensity and high reads region (top 1% of Affymetrix present calls: absolute intensity $\geq$ 11,000) is greatly diminished, an observation which could be attributed to a deviation from linearity in the microarray assay in this region, while read count is not prone to saturation effects (Wilhelm et al., 2008). We also compared gene ranks based on abundance, as determined by each method. When comparing the top 500 most abundant genes from microarray and sequencing, 60% of genes are common between the two methodologies. We observed decreased correlation between array and sequencing data (Spearman correlation coefficient of 0.26) in the lowest quartile of Affymetrix present calls (absolute intensity $\leq$ 180 and detection p-value $\leq$ 0.04). The discrepancy observed in this region of low abundance genes may be a consequence of decreased sensitivity in DNA microarrays (Liu et al., 2007; Mortazavi et al., 2008), whereby their low intensity signals can be impossible to distinguish from background. Furthermore, when CHO samples are hybridized to CHO Affymetrix arrays, 50% of probesets derived solely from tissue sequences generally have an absence call, indicating that these transcripts are exceedingly low or nil. However their detection through Illumina Solexa sequencing suggests that deep sequencing can capture transcripts which

were previously thought to be absent. It is also interesting to note that the dynamic range observed through Illumina Solexa, which spans six orders of magnitude, is far greater than what was obtained using Affymetrix arrays (approximately three orders of magnitude).

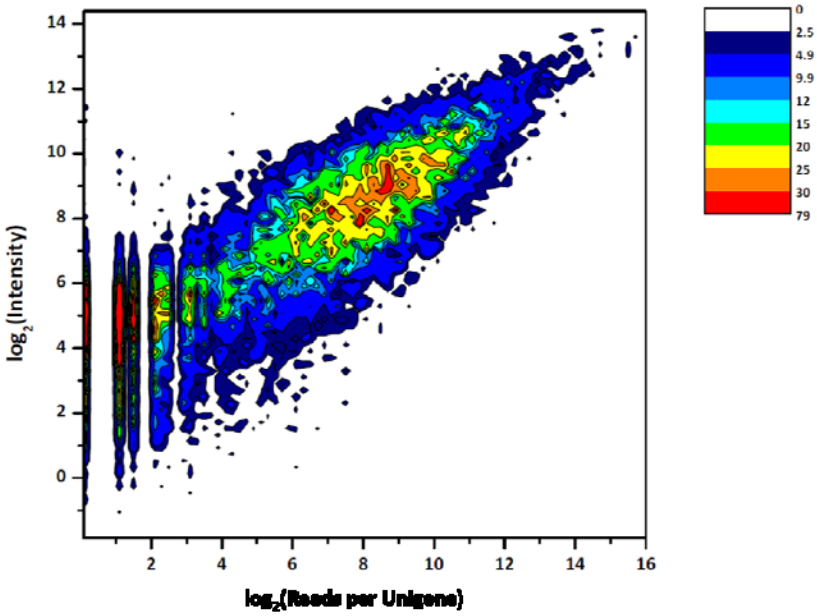


Figure 84 – Correlation of reads mapped per CHO unigene and Affymetrix signal intensities. The contour plot shows the log_2-transformed count of Illumina Solexa reads mapped to CHO unigenes versus the log_2-transformed Affymetrix intensity of corresponding CHO unigenes. The contour plot color legend is shown on the right.

While the vast majority of CHO unigenes in our existing dataset were identified in this RNA-Seq study, we were also able to discover new transcripts by mapping to orthologous sequence databases. More than 26 million reads or 48.2% of total reads were mapped to 14,070 unique mouse genes, representing an increase of nearly 4000 genes over our previous report of 10,707 orthologous mouse genes (Wlaschin and Hu, 2007b). Of the close to 4000 additional mouse genes that could be mapped by CHO Illumina Solexa reads, a large number are rare, as judged by the number of mapped Illumina Solexa reads. These mouse orthologs are possibly expressed only at low levels, or have low homology to their CHO counterparts in regions that are not mapped. Functional class assignment of this subset of newly identified mouse orthologs showed an increase in representation of signal transduction, receptor activity, cytoskeleton and nucleosome positioning related classes as compared to the existing CHO EST dataset.

The abundance of Illumina Solexa reads which mapped to mouse orthologs was next examined in a genomic context. This can allow for rapid identification of highly transcribed regions, as well as putative gene deserts. RNA-Seq read counts were summed over 1 Mb bins along mouse chromosomes and subsequently divided by the number of unique genes within each bin. These normalized abundances are plotted along mouse genomic coordinates (Figure 85). Color thresholds were set based on the distribution of normalized mapped read frequencies to mouse orthologs. Each color threshold captures 20% of the normalized mapped read count distribution. Clusters of highly expressed transcripts as well as low abundance regions on each chromosome can be seen. Approximately 25% of the mouse genome (i.e. 25% of all 1 MB bins) was not mapped to by CHO Illumina Solexa reads Such gaps could represent regions which are not expressed in the cell line studied, possibly as a result of gene silencing mechanisms (Paulin et al., 1998), or regions where sequence homology between mouse and Chinese hamster is low, preventing confident mapping. It is also possible that these regions are absent, or have been rearranged in the Chinese hamster genome or during the course of laboratory cultivation and manipulation of CHO cells, which are known to be highly aneuploid (Chasin and Urlaub, 1975).

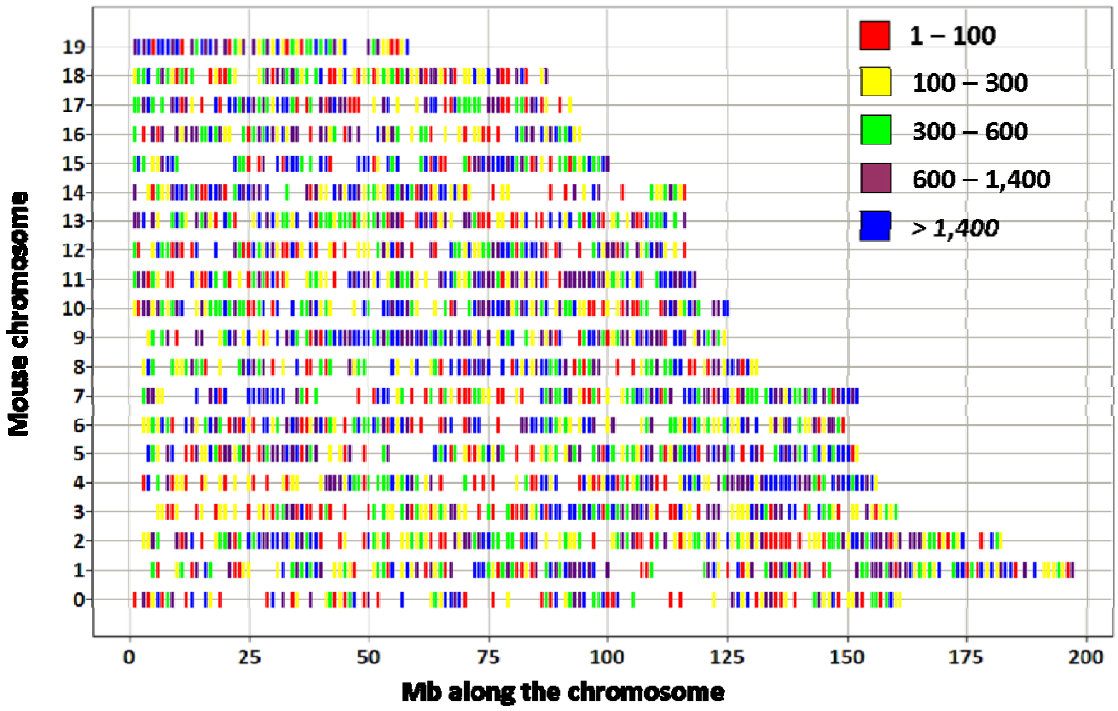


Figure 85 – Normalized abundance of Illumina Solexa reads mapped to orthologous mouse genes along mouse genomic coordinates. Normalized abundances were determined by creating 1 MB bins along each chromosome and summing the number of mapped read for all genes within each bin, then dividing by the total number of genes within each bin. Each color threshold captures 20% of the normalized abundance distribution.

8.5 CONCLUSIONS

There is no doubt that the information generated in this study will yield additional insights once a reference genome is available for the Chinese hamster. With the availability of a genome map, RNA-Seq data can be mined to identify novel transcripts. Indeed, studies in a range of species have used transcriptome sequencing data to identify putative novel transcripts: using RNA-Seq data generated from multiple mouse tissues, Mortazavi et. al. identified 596 novel candidate transcript models (Mortazavi et al., 2008). Similarly, RNA-Seq data generated in yeast identified 453 novel transcripts (Wilhelm et al., 2008). Transcriptome data obtained from this study and others like it will aid in the annotation and characterization of the Chinese hamster genome. Several algorithms have already been developed which make use of RNA-Seq data for genome annotation (Denoeud et al., 2008) as well as for a number of other useful characterizations, such as the determination of transcription start and stop sites, and the identification of novel splice variants.

Sequencing-based approaches for expression profiling will continue to become more affordable and no doubt replace hybridization-based methodologies in the future. This continued decrease in cost, coupled with increases in achievable read length will facilitate *de novo* sequencing of the Chinese hamster genome. The availability of such a resource will enable the study of previously unknown mechanisms in this organism, including the prevalence of DNA methylation and the characterization of transcriptionally active regions. This knowledge has the potential to revolutionize the field of recombinant protein production in cultured mammalian cells.

9 CONCLUSIONS AND FUTURE DIRECTIONS

The development of genomic tools has enabled the global analysis of biological systems, and has already begun to exert its impact on the field of bioprocessing. Through intensive efforts in our laboratory, we have developed a set of resources which has allowed us to use transcriptome analysis to study the process of recombinant protein production in Chinese hamster ovary (CHO) cells. This thesis describes the application of microarray technology to the multiple facets of mammalian cell culture. The insights gained here not only deepen our understanding of protein production in such cells, but also provides avenues and opportunities for process enhancement strategies and cell line engineering efforts.

Over the past several years, our compendium of CHO sequence information has grown to include more than 28,000 unique sequences. This information was the basis for our design of custom CHO Affymetrix arrays. While our dataset has yet to encompass the entire CHO transcriptome, our annotation and characterization has revealed it is well-represented in a number of key functional classes. Through the application of next-generation sequencing technologies, we are continuing to expand our sequence collection and as sequencing costs continue to plummet, we can now set genome sequencing in our sights.

The availability of the Chinese hamster genome sequence will greatly expand our ability to understand the process of recombinant protein production at a fundamental level. Furthermore, previously accumulated data will be able to be re-analyzed in a genomic context and will now doubt yield additional insights. As an example, more than 15% of the RNA-Seq data described in Chapter 8 could not be mapped to any sequence database. There is no doubt that re-examining this data within the context of the Chinese hamster genome will allow us to identify novel splice variants, and may discover additional transcriptional hotspots. Furthermore, genome data will allow us to study the epigenetic mechanisms involved in recombinant protein production, which have already been shown to play a role in mouse cell lines (Seth et al., 2006).

This work has demonstrated the wide applicability of transcriptome analysis, and has surveyed the multiple facets of bioprocessing in cultured mammalian cells. As the field of mammalian cell culture becomes increasingly mature, a greater understanding of the biological mechanisms of recombinant protein production will emerge as the most direct way to improve these processes and continue to provide these life-saving therapeutics.

10 REFERENCES

Adams, M. D., et al. (1991). "Complementary DNA sequencing: expressed sequence tags and human genome project." Science **252**(5013): 1651-6.

Aggarwal, S. (2008). "What's fueling the biotech engine-2007." Nat Biotechnol **26**(11): 1227-33.

Al-Fageeh, M. B., et al. (2006). "The cold-shock response in cultured mammalian cells: harnessing the response for the improvement of recombinant protein production." Biotechnol Bioeng **93**(5): 829-35.

Alahari, A. (2009). "Implementing cost reduction strategies for HuMab manufacturing processes." BioProcess Int. **7**(Suppl. 1): 48-54.

Alter, O., et al. (2000). "Singular value decomposition for genome-wide expression data processing and modeling." Proc Natl Acad Sci U S A **97**(18): 10101-6.

Arden, N. and M. J. Betenbaugh (2004). "Life and death in mammalian cell culture: strategies for apoptosis inhibition." Trends Biotechnol. **22**(4): 174-180.

Baik, J. Y., et al. (2006). "Initial transcriptome and proteome analyses of low culture temperature-induced expression in CHO cells producing erythropoietin." Biotechnol Bioeng **93**(2): 361-71.

Baldin, V., et al. (1993). "Cyclin D1 is a nuclear protein required for cell cycle progression in G1." Genes Dev **7**(5): 812-21.

Barnabe, N. and M. Butler (1994). "Effect of temperature on nucleotide pools and monoclonal antibody production in a mouse hybridoma." Biotechnol Bioeng **44**(10): 1235-45.

Barnes, L. M., et al. (2000). "Advances in animal cell recombinant protein production: GS-NS0 expression system." Cytotechnology **32**(2): 109-23.

Bebbington, C. R., et al. (1992). "High-level expression of a recombinant antibody from myeloma cells using a glutamine synthetase gene as an amplifiable selectable marker." Bio/Technology **10**(2): 169-75.

Beer, C., et al. (2003). "Gene expression analysis of murine cells producing amphotropic mouse leukaemia virus at a cultivation temperature of 32 and 37 DegC." J. Gen. Virol. **84**(7): 1677-1686.

Bello-Fernandez, C., et al. (1993). "The ornithine decarboxylase gene is a transcriptional target of c-Myc." Proc. Natl. Acad. Sci. U. S. A. **90**(16): 7804-8.

Bild, A. H., et al. (2006). "Oncogenic pathway signatures in human cancers as a guide to targeted therapies." Nature **439**(7074): 353-7.

Birch, J. R. and A. J. Racher (2006). "Antibody production." Adv Drug Deliv Rev **58**(5-6): 671-85.

Bollati-Fogolin, M., et al. (2005). "Temperature Reduction in Cultures of hGM-CSF-expressing CHO Cells: Effect on Productivity and Product Quality." Biotechnol. Prog. **21**(1): 17-21.

Borys, M. C., et al. (1993). "Culture pH affects expression rates and glycosylation of recombinant mose placental lactogen proteins by Chinese hamster ovary (CHO) cells." Bio/Technology **11**(6): 720-4.

Boussouar, F. and M. Benahmed (2004). "Lactate and energy metabolism in male germ cells." Trends Endocrinol Metab **15**(7): 345-50.

Brand, K. A. and U. Hermfisse (1997). "Aerobic glycolysis by proliferating cells: a protective strategy against reactive oxygen species." Faseb J **11**(5): 388-95.
Brenner, S., et al. (2000). "Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays." Nat Biotechnol **18**(6): 630-4.
Brigstock, D. R. (2002). "Regulation of angiogenesis and endothelial cell function by connective tissue growth factor (CTGF) and cysteine-rich 61 (CYR61)." Angiogenesis **5**(3): 153-65.
Brown, M. E., et al. (1992). "Process development for the production of recombinant antibodies using the glutamine synthetase (GS) system." Cytotechnology **9**(1-3): 231-6.
Brusick, D. (1986). "Genotoxic effects in cultured mammalian cells produced by low pH treatment conditions and increased ion concentrations." Environ Mutagen **8**(6): 879-86.
Bumke, M. A., et al. (2003). "Modulation of gene expression by extracellular pH variations in human fibroblasts: a transcriptomic and proteomic study." Proteomics **3**(5): 675-88.
Burteau, C. C., et al. (2003). "Fortification of a protein-free cell culture medium with plant peptones improves cultivation and productivity of an interferon-gamma-producing CHO cell line." In Vitro Cell Dev Biol Anim **39**(7): 291-6.
Bush, K. T., et al. (1994). "Induction of the FK506-binding protein, FKBP13, under conditions which misfold proteins in the endoplasmic reticulum." Biochem J **303 (Pt 3)**: 705-8.
Candido, E. P., et al. (1978). "Sodium butyrate inhibits histone deacetylation in cultured cells." Cell **14**(1): 105-13.
Carninci, P., et al. (2005). "The transcriptional landscape of the mammalian genome." Science **309**(5740): 1559-63.
Casero, R. A., Jr. and L. J. Marton (2007). "Targeting polyamine metabolism and function in cancer and other hyperproliferative diseases." Nat Rev Drug Discov **6**(5): 373-90.
Charaniya, S., et al. (2008a). "Mining bioprocess data: opportunities and challenges." Trends Biotechnol **26**(12): 690-9.
Charaniya, S., et al. (2008b). "Mining transcriptome data for function-trait relationship of hyper productivity of recombinant antibody." Biotechnol Bioeng.
Charaniya, S., et al. (2009). "Mining manufacturing data for discovery of high productivity process characteristics." Biotechnology & Bioengineering **in review**.
Chartrain, M. and L. Chu (2008). "Development and production of commercial therapeutic monoclonal antibodies in Mammalian cell expression systems: an overview of the current upstream technologies." Curr Pharm Biotechnol **9**(6): 447-67.
Chasin, L. A. and G. Urlaub (1975). "Chromosome-wide event accompanies the expression of recessive mutations in tetraploid cells." Science **187**(4181): 1091-3.
Chen, K., et al. (2001). "Engineering of a mammalian cell line for reduction of lactate formation and high monoclonal antibody production." Biotechnol. Bioeng. **72**(1): 55-61.

Cherlet, M. and A. Marc (2000). "Stimulation of monoclonal antibody production of hybridoma cells by butyrate: Evaluation of a feeding strategy and characterization of cell behavior." Cytotechnology **32**: 17-29.
Chotigeat, W., et al. (1994). "Role of environmental conditions on the expression levels, glycoform pattern and levels of sialyltransferase for hFSH produced by recombinant CHO cells." Cytotechnology **15**: 217-221.
Chun, B.-H., et al. (2003a). "Enhanced production of recombinant B-domain deleted factor VIII from Chinese hamster ovary cells by propionic and butyric acids." Biotechnol. Lett. **25**(4): 315-319.
Chun, B. H., et al. (2005). "Use of plant protein hydrolysates for varicella virus production in serum-free medium." Biotechnol Lett **27**(4): 243-8.
Chun, C., et al. (2003b). "Application of factorial design to accelerate identification of CHO growth factor requirements." Biotechnol. Prog. **19**(1): 52-57.
Chung, K. H., et al. (2006). "Polycistronic RNA polymerase II expression vectors for RNA interference based on BIC/miR-155." Nucleic Acids Res **34**(7): e53.
Chuppa, S., et al. (1997). "Fermentor temperature as a tool for control of high-density perfusion cultures of mammalian cells." Biotechnol. Bioeng. **55**(2): 328-338.
Chusainow, J., et al. (2009). "A study of monoclonal antibody-producing CHO cell lines: what makes a stable high producer?" Biotechnol Bioeng **102**(4): 1182-96.
Coe, N. R. and D. A. Bernlohr (1998). "Physiological properties and functions of intracellular fatty acid-binding proteins." Biochim Biophys Acta **1391**(3): 287-306.
Coquelle, A., et al. (1997). "Expression of fragile sites triggers intrachromosomal mammalian gene amplification and sets boundaries to early amplicons." Cell **89**(2): 215-25.
Crow, J. A. and E. F. Retzel (2005). "Diogenes - Reliable prediction of protein-encoding regions in short genomic sequences."
D'Souza-Schorey, C. and P. Chavrier (2006). "ARF proteins: roles in membrane traffic and beyond." Nat Rev Mol Cell Biol **7**(5): 347-58.
Dahlquist, K. D., et al. (2002). "GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways." Nat Genet **31**(1): 19-20.
de Hostos, E. L. (1999). "The coronin family of actin-associated proteins." Trends Cell Biol **9**(9): 345-50.
De Leon Gatti, M., et al. (2007). "Comparative transcriptional analysis of mouse hybridoma and recombinant Chinese hamster ovary cells undergoing butyrate treatment." J Biosci Bioeng **103**(1): 82-91.
Deberardinis, R. J., et al. (2008). "Brick by brick: metabolism and tumor cell growth." Curr Opin Genet Dev **18**(1): 54-61.
Denoeud, F., et al. (2008). "Annotating genomes with massive-scale RNA sequencing." Genome Biol **9**(12): R175.
Doniger, S. W., et al. (2003). "MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data." Genome Biol **4**(1): R7.
Doolan, P., et al. (2008). "Transcriptional profiling of gene expression changes in a PACE-transfected CHO DUKX cell line secreting high levels of rhBMP-2." Mol Biotechnol **39**(3): 187-99.

Dorsett, Y. and T. Tuschl (2004). "siRNAs: applications in functional genomics and potential as therapeutics." Nat Rev Drug Discov **3**(4): 318-29.
Doumas, S., et al. (2005). "Anti-inflammatory and antimicrobial roles of secretory leukocyte protease inhibitor." Infect Immun **73**(3): 1271-4.
Dyring, C., et al. (1994). "Observations on the influence of glutamine, asparagine and peptone on growth and t-PA production of Chinese hamster ovary (CHO) cells." Cytotechnology **16**: 37-42.
Elenjord, R., et al. (2008). "Dysregulation of matrix metalloproteinases and their tissue inhibitors by S100A4." Connect Tissue Res **49**(3): 185-8.
Enard, W., et al. (2002). "Intra- and interspecific variation in primate gene expression patterns." Science **296**(5566): 340-3.
Everitt, B. (1974). Cluster Analysis. London.
Fann, C. H., et al. (2000). "Limitations to the amplification and stability of human tissue-type plasminogen activator expression by Chinese hamster ovary cells." Biotechnol Bioeng **69**(2): 204-12.
Figueroa, B., Jr., et al. (2007). "Enhanced cell culture performance using inducible anti-apoptotic genes E1B-19K and Aven in the production of a monoclonal antibody with Chinese hamster ovary cells." Biotechnol Bioeng **97**(4): 877-92.
Flintoff, W. F. and K. Essani (1980). "Methotrexate-resistant Chinese hamster ovary cells contain a dihydrofolate reductase with an altered affinity for methotrexate." Biochemistry **19**(18): 4321-7.
Fogolin, M. B., et al. (2004). "Impact of temperature reduction and expression of yeast pyruvate carboxylase on hGM-CSF-producing CHO cells." J Biotechnol **109**(1-2): 179-91.
Fox, S. R., et al. (2004). "Maximizing interferon-gamma production by Chinese hamster ovary cells through temperature shift optimization: Experimental and modeling." Biotechnol. Bioeng. **85**(2): 177-184.
Franek, F., et al. (2000). "Plant protein hydrolysates: preparation of defined peptide fractions promoting growth and production in animal cells cultures." Biotechnol Prog **16**(5): 688-92.
Furukawa, K. and K. Ohsuye (1998). "Effect of culture temperature on a recombinant CHO cell line producing a C-terminal alpha -amidating enzyme." Cytotechnology **26**(2): 153-164.
Furukawa, K. and K. Ohsuye (1999). "Enhancement of productivity of recombinant alpha -amidating enzyme by low temperature culture." Cytotechnology **31**(1-2): 85-94.
Furuno, M., et al. (2006). "Clusters of internally primed transcripts reveal novel long noncoding RNAs." PLoS Genet **2**(4): e37.
Fussenegger, M., et al. (1998). "Controlled proliferation by multigene metabolic engineering enhances the productivity of Chinese hamster ovary cells." Nat Biotechnol **16**(5): 468-72.
Galderisi, U., et al. (2003). "Cell cycle regulation and neural differentiation." Oncogene **22**(33): 5208-19.
Gerner, E. W. and F. L. Meyskens, Jr. (2004). "Polyamines and cancer: old molecules, new understanding." Nat Rev Cancer **4**(10): 781-92.
Goldberg, E. (1990). "Developmental expression of lactate dehydrogenase isozymes during spermatogenesis." Prog Clin Biol Res **344**: 49-52.

Gramer, M. J. and C. F. Goochee (1993). "Glycosidase activities in Chinese hamster ovary cell lysate and cell culture supernatant." Biotechnol. Prog. **9**(4): 366-73.
Grant, B. D. and S. Caplan (2008). "Mechanisms of EHD/RME-1 Protein Function in Endocytic Transport." Traffic.
Griffiths-Jones, S., et al. (2008). "miRBase: tools for microRNA genomics." Nucleic Acids Res **36**(Database issue): D154-8.
Grillari, J., et al. (2001). "Analysis of alterations in gene expression after amplification of recombinant genes in CHO cells." J Biotechnol **87**(1): 59-65.
Gu, X., et al. (1997). "Influence of Primatone RL supplementation on sialylation of recombinant human interferon-gamma produced by Chinese hamster ovary cell culture using serum-free media." Biotechnol Bioeng **56**(4): 353-60.
Gupta, P. and K. H. Lee (2007). "Genomics and proteomics in process development: opportunities and challenges." Trends Biotechnol **25**(7): 324-30.
Harbers, M. and P. Carninci (2005). "Tag-based approaches for transcriptome research and genome annotation." Nat Methods **2**(7): 495-502.
Harding, H. P., et al. (1999). "Protein translation and folding are coupled by an endoplasmic-reticulum-resident kinase." Nature **397**(6716): 271-4.
Hayduk, E. J. and K. H. Lee (2005). "Cytochalasin D can improve heterologous protein productivity in adherent Chinese hamster ovary cells." Biotechnol Bioeng **90**(3): 354-64.
He, S., et al. (2008). "NONCODE v2.0: decoding the non-coding." Nucleic Acids Res **36**(Database issue): D170-2.
Heidemann, R., et al. (2000). "The use of peptones as medium additives for the production of a recombinant therapeutic protein in high density perfusion cultures of mammalian cells." Cytotechnology **32**(2): 157-167.
Hendrick, V., et al. (2001). "Increased productivity of recombinant tissular plasminogen activator (t-PA) by butyrate and shigt of temperature: A cell cycle phases analysis." Cytotechnology **36**: 71-83.
Ho Sui, S. J., et al. (2005). "oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes." Nucleic Acids Res **33**(10): 3154-64.
Holen, T., et al. (2002). "Positional effects of short interfering RNAs targeting the human coagulation trigger Tissue Factor." Nucleic Acids Res **30**(8): 1757-66.
Holland, M. J. (2002). "Transcript abundance in yeast varies over six orders of magnitude." J Biol Chem **277**(17): 14363-6.
Huang, E. P., et al. (2007). "Development of Super-CHO protein-free medium based on a statistical design." J. Chem. Technol. Biotechnol. **82**(5): 431-441.
Hubbard, T., et al. (2002). "The Ensembl genome database project." Nucleic Acids Res **30**(1): 38-41.
Irani, N., et al. (1999). "Improvement of the primary metabolism of cell cultures by introducing a new cytoplasmic pyruvate carboxylase reaction." Biotechnol. Bioeng. **66**(4): 238-246.
Jayapal, K. P., et al. (2007a). "Recombinant Protein Therapeutics from CHO Cells-20 Years and Counting." Chemical Engineering Progress **103**(10): 40.
Jayapal, K. P., et al. (2007b). "Recombinant protein therapeutics from CHO cells - 20 years and counting." Chem Eng Prog **103**(10): 40.

Jeong, D., et al. (2001). "Blocking of acidosis-mediated apoptosis by a reduction of lactate dehydrogenase activity through antisense mRNA expression." Biochem Biophys Res Commun **289**(5): 1141-9.
Jones, P. A. and S. M. Taylor (1980). "Cellular differentiation, cytidine analogs and DNA methylation." Cell **20**(1): 85-93.
Jun, S. C., et al. (2005). "Selection strategies for the establishment of recombinant Chinese hamster ovary cell line with dihydrofolate reductase-mediated gene amplification." Appl. Microbiol. Biotechnol. **69**(2): 162-169.
Jurka, J., et al. (2005). "Repbase Update, a database of eukaryotic repetitive elements." Cytogenet Genome Res **110**(1-4): 462-7.
Kantardjieff, A., et al. (2009a). "Transcriptome and proteome analysis of Chinese hamster ovary cells under low temperature and butyrate treatment." Journal of Biotechnology **in press**.
Kantardjieff, A., et al. (2009b). "Developing genomic platforms for Chinese hamster ovary cells." Biotechnology Advances.
Kato, Y., et al. (2005). "Acidic extracellular pH induces matrix metalloproteinase-9 expression in mouse metastatic melanoma cells through the phospholipase D-mitogen-activated protein kinase signaling." J Biol Chem **280**(12): 10938-44.
Kaufman, R. J., et al. (1985). "Coamplification and coexpression of human tissue-type plasminogen activator and murine dihydrofolate reductase sequences in Chinese hamster ovary cells." Mol Cell Biol **5**(7): 1750-9.
Kaufmann, H., et al. (1999). "Influence of low temperature on productivity, proteome and protein phosphorylation of CHO cells." Biotechnol Bioeng **63**(5): 573-82.
Khoo, S. H., et al. (2007). "A genome-wide transcriptional analysis of producer and non-producer NS0 myeloma cell lines." Biotechnol Appl Biochem **47**(Pt 2): 85-95.
Kim, N. S., et al. (1998a). "Clonal variability within dihydrofolate reductase-mediated gene amplified Chinese hamster ovary cells: stability in the absence of selective pressure." Biotechnol. Bioeng. **60**(6): 679-688.
Kim, N. S. and G. M. Lee (2000). "Overexpression of bcl-2 inhibits sodium butyrate-induced apoptosis in Chinese hamster ovary cells resulting in enhanced humanized antibody production." Biotechnol Bioeng **71**(3): 184-93.
Kim, S. H. and G. M. Lee (2007). "Down-regulation of lactate dehydrogenase-A by siRNAs for reduced lactic acid formation of Chinese hamster ovary cells producing thrombopoietin." Appl Microbiol Biotechnol **74**(1): 152-9.
Kim, S. J., et al. (1998b). "Characterization of chimeric antibody producing CHO cells in the course of dihydrofolate reductase-mediated gene amplification and their stability in the absence of selective pressure." Biotechnol. Bioeng. **58**(1): 73-84.
Klickstein, L. B. (2001). "Production of a complete cDNA library." Curr Protoc Mol Biol **Chapter 5**: Unit5 8A.
Koslowski, M., et al. (2002). "Multiple splice variants of lactate dehydrogenase C selectively expressed in human cancer." Cancer Res **62**(22): 6750-5.
Krampe, B., et al. (2008). "Transcriptome and proteome analysis of antibody-producing mouse myeloma NS0 cells cultivated at different cell densities in perfusion culture." Biotechnol Appl Biochem **50**(Pt 3): 133-41.

Krampert, M., et al. (2004). "Activities of the matrix metalloproteinase stromelysin-2 (MMP-10) in matrix degradation and keratinocyte organization in wounded skin." Mol Biol Cell **15**(12): 5242-54.
Kurano, N., et al. (1990). "Growth behavior of Chinese hamster ovary cells in a compact loop bioreactor: 1. Effects of physical and chemical environments." J. Biotechnol. **15**: 101-11.
Leek, J. T., et al. (2006). "EDGE: extraction and analysis of differential gene expression." Bioinformatics **22**(4): 507-8.
Li, S. S., et al. (1989). "Differential activity and synthesis of lactate dehydrogenase isozymes A (muscle), B (heart), and C (testis) in mouse spermatogenic cells." Biol Reprod **40**(1): 173-80.
Link, T., et al. (2004). "Bioprocess development for the production of a recombinant MUC1 fusion protein expressed by CHO-K1 cells in protein-free medium." J. Biotechnol. **110**(1): 51-62.
Lister, R., et al. (2008). "Highly integrated single-base resolution maps of the epigenome in Arabidopsis." Cell **133**(3): 523-36.
Liu, F., et al. (2007). "Comparison of hybridization-based and sequencing-based gene expression technologies on biological replicates." BMC Genomics **8**: 153.
Lockhart, D. J., et al. (1996). "Expression monitoring by hybridization to high-density oligonucleotide arrays." Nature Biotechnology **14**(13): 1675-80.
Lu, C., et al. (2007). "A T-flask based screening platform for evaluating and identifying plant hydrolysates for a fed-batch cell culture process." Cytotechnology **55**(1): 15-29.
Luo, Y. and G. Chen (2007). "Combined approach of NMR and chemometrics for screening peptones used in the cell culture medium for the production of a recombinant therapeutic protein." Biotechnol Bioeng **97**(6): 1654-9.
MacDougald, O. A. and C. F. Burant (2007). "The rapidly expanding family of adipokines." Cell Metab **6**(3): 159-61.
Makowski, L. and G. S. Hotamisligil (2005). "The role of fatty acid binding proteins in metabolic syndrome and atherosclerosis." Curr Opin Lipidol **16**(5): 543-8.
Mardis, E. R. (2008). "The impact of next-generation sequencing technology on genetics." Trends Genet **24**(3): 133-41.
Marioni, J. C., et al. (2008). "RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays." Genome Res **18**(9): 1509-17.
Marzesco, A. M., et al. (1998). "The rod cGMP phosphodiesterase delta subunit dissociates the small GTPase Rab13 from membranes." J Biol Chem **273**(35): 22340-5.
McCoy, R. E. and A. Morris (2008). Identification and optimization of peptone-regulated metabolic pathways to develop high-performance chemically-defined processes. Abstracts of Papers, 236th ACS National Meeting, Philadelphia, PA, United States, August 17-21, 2008.
Mimura, Y., et al. (2001). "Butyrate increases production of human chimeric IgG in CHO-K1 cells whilst maintaining function and glycoform profile." J Immunol Methods **247**(1-2): 205-16.

Minn, A. H., et al. (2005). "Thioredoxin-interacting protein is stimulated by glucose through a carbohydrate response element and induces beta-cell apoptosis." Endocrinology **146**(5): 2397-405.
Moore, A., et al. (1997). "Effects of temperature shift on cell cycle, apoptosis and nucelotide pools in CHO cell batch cultures." Cytotechnology **23**: 47-54.
Mori, K., et al. (2004). "Engineering Chinese hamster ovary cells to maximize effector function of produced antibodies using FUT8 siRNA." Biotechnol Bioeng **88**(7): 901-8.
Mortazavi, A., et al. (2008). "Mapping and quantifying mammalian transcriptomes by RNA-Seq." Nat Methods **5**(7): 621-8.
Mortensen, R. M. and R. E. Kingston (2009). "Selection of transfected mammalian cells." Curr Protoc Mol Biol **Chapter 9**: Unit9.5.
Mudge, J., et al. (2008). "Genomic convergence analysis of schizophrenia: mRNA sequencing reveals altered synaptic vesicular transport in post-mortem cerebellum." PLoS ONE **3**(11): e3625.
Mutskov, V. and G. Felsenfeld (2004). "Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9." Embo J. **23**(1): 138-149.
Nagalakshmi, U., et al. (2008). "The transcriptional landscape of the yeast genome defined by RNA sequencing." Science **320**(5881): 1344-9.
Nevins, J. R. and A. Potti (2007). "Mining gene expression profiles: expression signatures as cancer phenotypes." Nat Rev Genet **8**(8): 601-9.
Nishiyama, H., et al. (1997). "A glycine-rich RNA-binding protein mediating cold-inducible suppression of mammalian cell growth." J. Cell Biol. **137**(4): 899-908.
Nissom, P. M., et al. (2006). "Transcriptome and proteome profiling to understanding the biology of high productivity CHO cells." Mol Biotechnol **34**(2): 125-40.
Novick, P. and M. Zerial (1997). "The diversity of Rab proteins in vesicle transport." Curr Opin Cell Biol **9**(4): 496-504.
Oh, S. K., et al. (1993). "Substantial overproduction of antibodies by applying osmotic pressure and sodium butyrate." Biotechnol Bioeng **42**(5): 601-10.
Osman, J. J., et al. (2001). "The response of GS-NS0 myeloma cells to pH shifts and pH perturbations." Biotechnol Bioeng **75**(1): 63-73.
Oyadomari, S. and M. Mori (2004). "Roles of CHOP/GADD153 in endoplasmic reticulum stress." Cell Death Differ **11**(4): 381-9.
Ozturk, S. S. and B. O. Palsson (1991). "Growth, metabolic, and antibody production kinetics of hybridoma cell culture: 2. Effects of serum concentration, dissolved oxygen concentration, and medium pH in a batch reactor." Biotechnol. Prog. **7**(6): 481-94.
Ozturk, S. S., et al. (1992). "Effects of ammonia and lactate on hybridoma growth, metabolism, and antibody production." Biotechnol Bioeng **39**(4): 418-31.
Pan, Q., et al. (2008). "Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing." Nat Genet **40**(12): 1413-5.
Paulin, R. P., et al. (1998). "Gene silencing by DNA methylation and dual inheritance in Chinese hamster ovary cells." Genetics **149**(2): 1081-8.
Pendeville, H., et al. (2001). "The ornithine decarboxylase gene is essential for cell survival during early murine development." Mol. Cell. Biol. **21**(19): 6549-6558.

Perez de Castro, I., et al. (2007). "A census of mitotic cancer genes: new insights into tumor cell biology and cancer therapy." Carcinogenesis **28**(5): 899-912.
Pollak, N., et al. (2007). "The power to reduce: pyridine nucleotides--small molecules with a multitude of functions." Biochem J **402**(2): 205-18.
Pongratz, R. L., et al. (2007). "Cytosolic and mitochondrial malic enzyme isoforms differentially control insulin secretion." J Biol Chem **282**(1): 200-7.
Pop, M. and S. L. Salzberg (2008). "Bioinformatics challenges of new sequencing technology." Trends Genet **24**(3): 142-9.
Post, P. L., et al. (1998). "Human myosin-IXb is a mechanochemically active motor and a GAP for rho." J Cell Sci **111 (Pt 7)**: 941-50.
Puck, T. T. (1957). "The genetics of somatic mammalian cells." Adv Biol Med Phys **5**: 75-101.
Puck, T. T., et al. (1958). "Genetics of somatic mammalian cells. III. Long-term cultivation of euploid cells from human and animal subjects." J Exp Med **108**(6): 945-56.
Read, J. A., et al. (2001). "Structural basis for altered activity of M- and H-isozyme forms of human lactate dehydrogenase." Proteins **43**(2): 175-85.
Reinartz, J., et al. (2002). "Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms." Brief Funct Genomic Proteomic **1**(1): 95-104.
Renard, J. M., et al. (1988). "Evidence that monoclonal antibody production kinetics is related to the integral of viable cells in batch systems." Biotechnology Letters **10**: 91-96.
Richards, E. J. and S. C. Elgin (2002). "Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects." Cell **108**(4): 489-500.
Rodriguez, J., et al. (2005). "Enhanced production of monomeric interferon-beta by CHO cells through the control of culture conditions." Biotechnol Prog **21**(1): 22-30.
Roessler, B., et al. (1996). "Temperature: a simple parameter for process optimization in fed-batch cultures of recombinant Chinese hamster ovary cells." Enzyme Microb. Technol. **18**(6): 423-7.
Ron, D. and J. F. Habener (1992). "CHOP, a novel developmentally regulated nuclear protein that dimerizes with transcription factors C/EBP and LAP and functions as a dominant-negative inhibitor of gene transcription." Genes Dev **6**(3): 439-53.
Rosenkranz, R., et al. (2008). "Characterizing the mouse ES cell transcriptome with Illumina sequencing." Genomics **92**(4): 187-94.
Rossman, K. L., et al. (2005). "GEF means go: turning on RHO GTPases with guanine nucleotide-exchange factors." Nat Rev Mol Cell Biol **6**(2): 167-80.
Sandadi, S., et al. (2006). "Application of Fractional Factorial Designs to Screen Active Factors for Antibody Production by Chinese Hamster Ovary Cells." Biotechnol. Prog. **22**(2): 595-600.
Sanz-Moreno, V., et al. (2008). "Rac activation and inactivation control plasticity of tumor cell movement." Cell **135**(3): 510-23.
Sauer, P. W., et al. (2000). "A high-yielding, generic fed-batch cell culture process for production of recombinant antibodies." Biotechnol. Bioeng. **67**(5): 585-597.
Schatz, S. M., et al. (2003). "Higher expression of fab antibody fragments in a CHU cell line at reduced temperature." Biotechnol. Bioeng. **84**(4): 433-438.

Schena, M., et al. (1995). "Quantitative monitoring of gene expression patterns with a complementary DNA microarray." Science **270**(5235): 467-70.
Schroder, M. and R. J. Kaufman (2005). "ER stress and the unfolded protein response." Mutat Res **569**(1-2): 29-63.
Schwartz, H., et al. (2002). Quality control of a wheat gluten hydrolysate for use as a raw material in cell culture media. Pittcon, New Orleans, LA.
Schwenk, R. W., et al. (2008). "Regulation of sarcolemmal glucose and fatty acid transporters in cardiac disease." Cardiovasc Res **79**(2): 249-58.
Sealy, L. and R. Chalkley (1978). "The effect of sodium butyrate on histone modification." Cell **14**(1): 115-21.
Seth, G., et al. (2006). "Reverting cholesterol auxotrophy of NS0 cells by altering epigenetic gene silencing." Biotechnol Bioeng **93**(4): 820-7.
Seth, G., et al. (2007). "Molecular portrait of high productivity in recombinant NS0 cells." Biotechnol Bioeng **97**(4): 933-51.
Shao, X., et al. (2001). "Testicular protein Spag5 has similarity to mitotic spindle protein Deepest and binds outer dense fiber protein Odf1." Mol Reprod Dev **59**(4): 410-6.
Shen, D. and S. T. Sharfstein (2006). "Genome-wide analysis of the transcriptional response of murine hybridomas to osmotic shock." Biotechnol Bioeng **93**(1): 132-45.
Simpson, F., et al. (1997). "Characterization of the adaptor-related protein complex, AP-3." J Cell Biol **137**(4): 835-45.
Sirotnak, F. M., et al. (1981). "Relative frequency and kinetic properties of transport-defective phenotypes among methotrexate-resistant L1210 clonal cell lines derived in vivo." Cancer Res **41**(11 Pt 1): 4447-52.
Smith, A. D., et al. (2008). "Using quality scores and longer reads improves accuracy of Solexa read mapping." BMC Bioinformatics **9**: 128.
Smith, M. L., et al. (1994). "Interaction of the p53-regulated protein Gadd45 with proliferating cell nuclear antigen." Science **266**(5189): 1376-80.
Sprong, H., et al. (2003). "Association of the Golgi UDP-galactose transporter with UDP-galactose:ceramide galactosyltransferase allows UDP-galactose import in the endoplasmic reticulum." Mol Biol Cell **14**(8): 3482-93.
Stamnes, M. A. and J. E. Rothman (1993). "The binding of AP-1 clathrin adaptor particles to Golgi membranes requires ADP-ribosylation factor, a small GTP-binding protein." Cell **73**(5): 999-1005.
Stapulionis, R., et al. (1997). "Efficient mammalian protein synthesis requires an intact F-actin system." J. Biol. Chem. **272**(40): 24980-24986.
Stark, G. R., et al. (1989). "Recent progress in understanding mechanisms of mammalian DNA amplification." Cell **57**(6): 901-8.
Stetler-Stevenson, W. G. and D. W. Seo (2005). "TIMP-2: an endogenous inhibitor of angiogenesis." Trends Mol Med **11**(3): 97-103.
Storey, J. D., et al. (2005). "Significance analysis of time course microarray experiments." Proc Natl Acad Sci U S A **102**(36): 12837-42.
Su, X. and N. A. Abumrad (2009). "Cellular fatty acid uptake: a pathway under construction." Trends Endocrinol Metab **20**(2): 72-7.

Subramanian, A., et al. (2005). "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles." Proc Natl Acad Sci U S A **102**(43): 15545-50.
Sultan, M., et al. (2008). "A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome." Science **321**(5891): 956-60.
Sung, Y. H., et al. (2004). "Yeast hydrolysate as a low-cost additive to serum-free medium for the production of human thrombopoietin in suspension cultures of Chinese hamster ovary cells." Appl Microbiol Biotechnol **63**(5): 527-36.
Suomela, S., et al. (2001). "Metalloelastase (MMP-12) and 92-kDa gelatinase (MMP-9) as well as their inhibitors, TIMP-1 and -3, are expressed in psoriatic lesions." Exp Dermatol **10**(3): 175-83.
Sureshkumar, G. K. and R. Mutharasan (1991). "The influence of temperature on mouse-mouse hybridoma growth and monoclonal antibody production." Biotechnol. Bioeng. **37**(3): 292-5.
Swiderek, H. and M. Al-Rubeai (2007). "Functional genome-wide analysis of antibody producing NS0 cell line cultivated at different temperatures." Biotechnol Bioeng **98**(3): 616-30.
Takizawa, C. G. and D. O. Morgan (2000). "Control of mitosis by changes in the subcellular location of cyclin-B1-Cdk1 and Cdc25C." Curr Opin Cell Biol **12**(6): 658-65.
Tanaka, S., et al. (1994). "C3G, a guanine nucleotide-releasing protein expressed ubiquitously, binds to the Src homology 3 domains of CRK and GRB2/ASH proteins." Proc Natl Acad Sci U S A **91**(8): 3443-7.
Toniatti, C., et al. (2004). "Gene therapy progress and prospects: transcription regulatory systems." Gene Ther. **11**(8): 649-657.
Trummer, E., et al. (2008). "Transcriptional profiling of phenotypically different Epo-Fc expressing CHO clones by cross-species microarray analysis." Biotechnol J **3**(7): 924-37.
Trummer, E., et al. (2006a). "Process parameter shifting: part I. Effect of DOT, pH, and temperature on the performance of Epo-Fc expressing CHO cells cultivated in controlled batch bioreactors." Biotechnol. Bioeng. **94**(6): 1033-1044.
Trummer, E., et al. (2006b). "Process parameter shifting: Part I. Effect of DOT, pH, and temperature on the performance of Epo-Fc expressing CHO cells cultivated in controlled batch bioreactors." Biotechnol Bioeng **94**(6): 1033-44.
Trummer, E., et al. (2006c). "Process parameter shifting: Part II. Biphasic cultivation-A tool for enhancing the volumetric productivity of batch processes using Epo-Fc expressing CHO cells." Biotechnol Bioeng **94**(6): 1045-52.
Tusher, V. G., et al. (2001). "Significance analysis of microarrays applied to the ionizing radiation response." Proc Natl Acad Sci U S A **98**(9): 5116-21.
Urlaub, G. and L. A. Chasin (1980). "Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity." Proc Natl Acad Sci U S A **77**(7): 4216-20.
Urlaub, G., et al. (1983). "Deletion of the diploid dihydrofolate reductase locus from cultured mammalian cells." Cell **33**(2): 405-12.
Urlaub, G., et al. (1986). "Effect of gamma rays at the dihydrofolate reductase locus: deletions and inversions." Somat Cell Mol Genet **12**(6): 555-66.

Velculescu, V. E., et al. (1995). "Serial analysis of gene expression." Science **270**(5235): 484-7.
Wang, E. T., et al. (2008). "Alternative isoform regulation in human tissue transcriptomes." Nature **456**(7221): 470-6.
Wang, M. D., et al. (2002). "Erythropoietin production from CHO cells grown by continuous culture in a fluidized-bed bioreactor." Biotechnol Bioeng **77**(2): 194-203.
Wang, Z., et al. (2009). "RNA-Seq: a revolutionary tool for transcriptomics." Nat Rev Genet **10**(1): 57-63.
Watanabe, S., et al. (2002). "Regulation of cell cycle and productivity in NS0 cells by the over-expression of p21CIP1." Biotechnol Bioeng **77**(1): 1-7.
Waterston, R. H., et al. (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature **420**(6915): 520-62.
Weber, W. and M. Fussenegger (2007). "Inducible product gene expression technology tailored to bioprocess engineering." Curr Opin Biotechnol **18**(5): 399-410.
Weidemann, R., et al. (1994). "Low temperature cultivation--a step towards process optimisation." Cytotechnology **15**(1-3): 111-6.
Wheeler, L. W., et al. (2008). "Cyclin A-CDK activity during G1 phase impairs MCM chromatin loading and inhibits DNA synthesis in mammalian cells." Cell Cycle **7**(14): 2179-88.
Whiteman, E. L., et al. (2002). "Role of Akt/protein kinase B in metabolism." Trends Endocrinol Metab **13**(10): 444-51.
Wilhelm, B. T., et al. (2008). "Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution." Nature **453**(7199): 1239-43.
Wilson, C., et al. (1990). "Position effects on eukaryotic gene expression." Annu Rev Cell Biol **6**: 679-714.
Wlaschin, K. F. and W. S. Hu (2006). "Fedbatch culture and dynamic nutrient feeding." Adv Biochem Eng Biotechnol **101**: 43-74.
Wlaschin, K. F. and W. S. Hu (2007a). "Engineering cell metabolism for high-density cell culture via manipulation of sugar transport." J Biotechnol **131**(2): 168-76.
Wlaschin, K. F. and W. S. Hu (2007b). "A scaffold for the Chinese hamster genome." Biotechnol Bioeng **98**(2): 429-39.
Wlaschin, K. F., et al. (2005). "EST sequencing for gene discovery in Chinese hamster ovary cells." Biotechnol Bioeng **91**(5): 592-606.
Wlaschin, K. F., et al. (2006). "Toward genomic cell culture engineering." Cytotechnology **50**(1-3): 121-140.
Wong, D. C., et al. (2006). "Transcriptional profiling of apoptotic pathways in batch and fed-batch CHO cell cultures." Biotechnol Bioeng **94**(2): 373-82.
Wurm, F. M. (2004). "Production of recombinant protein therapeutics in cultivated mammalian cells." Nat Biotechnol **22**(11): 1393-8.
Wurm, F. M., et al. (1986). "Inducible overproduction of the mouse c-myc protein in mammalian cells." Proc Natl Acad Sci U S A **83**(15): 5414-8.
Wurm, F. M. and M. Jordan (2003). "Gene transfer and gene amplification in mammalian cells." New Compr. Biochem. **38**: 309-335.
Wurm, F. M. and C. J. Petropoulos (1994). "Plasmid integration, amplification and cytogenetics in CHO cells: questions and comments." Biologicals **22**(2): 95-102.

Xiao, H., et al. (1997). "Sodium butyrate induces NIH3T3 cells to senescence-like state and enhances promoter activity of p21WAF/CIP1 in p53-independent manner." Biochem Biophys Res Commun **237**(2): 457-60.
Yee, J. C., et al. (2008a). "Genomic and proteomic exploration of CHO and hybridoma cells under sodium butyrate treatment." Biotechnol Bioeng **99**(5): 1186-204.
Yee, J. C., et al. (2008b). "Comparative transcriptome analysis to unveil genes affecting recombinant protein productivity in mammalian cells." Biotechnol Bioeng.
Yee, J. C., et al. (2009). "Comparative transcriptome analysis to unveil genes affecting recombinant protein productivity in mammalian cells." Biotechnol Bioeng **102**(1): 246-63.
Yoon, S. K., et al. (2004). "Effect of culture pH on erythropoietin production by Chinese hamster ovary cells grown in suspension at 32.5 and 37.0 DegC." Biotechnol. Bioeng. **89**(3): 345-356.
Yoon, S. K., et al. (2003a). "Effect of low culture temperature on specific productivity and transcription level of anti-4-1BB antibody in recombinant Chinese hamster ovary cells." Biotechnol Prog **19**(4): 1383-6.
Yoon, S. K., et al. (2003b). "Effect of low culture temperature on specific productivity, transcription level, and heterogeneity of erythropoietin in Chinese hamster ovary cells." Biotechnol Bioeng **82**(3): 289-98.
Yuan, Z., et al. (2006). "Genomic organization, promoter activity, and expression of the human choline transporter-like protein 1." Physiol Genomics **26**(1): 76-90.
Zechner, R. (1997). "The tissue-specific expression of lipoprotein lipase: implications for energy and lipoprotein metabolism." Curr Opin Lipidol **8**(2): 77-88.
Zeng, Z. S., et al. (1999). "Loss of basement membrane type IV collagen is associated with increased expression of metalloproteinases 2 and 9 (MMP-2 and MMP-9) during human colorectal tumorigenesis." Carcinogenesis **20**(5): 749-55.
Zhou, W., et al. (1997). "Fed-batch culture of recombinant NS0 myeloma cells with high monoclonal antibody production." Biotechnol Bioeng **55**(5): 783-92.
Zhou, W., et al. (1995). "High viable cell concentration fed-batch cultures of hybridoma cells through on-line nutrient feeding." Biotechnol Bioeng **46**(6): 579-87.
Zhuang, B., et al. (2009). "FARP1 promotes the dendritic growth of spinal motor neuron subtypes through transmembrane Semaphorin6A and PlexinA4 signaling." Neuron **61**(3): 359-72.

11 APPENDIX

Table I – List of differentially expressed genes at Day 15 between control population and mAb-producing clones in the functional class 'Regulation of cell cycle.'

Gene Symbol	Description	Log F.C.[a]	Avg. int. IgG[b]	Avg. int. C[c]
Ereg	Epiregulin	2.39	1398.9	266.1
N.A.	Dual specificity protein phosphatase 6 (EC 3.1.3.48)	1.86	465.9	128.1
Ccnd1	Cyclin D1	1.82	1466.3	416.0
E2f8	E2F transcription factor 8	1.40	78.7	29.8
Tsc2	Tuberous sclerosis 2	1.38	67.1	25.8
PKD2	Polycystin-2 (Polycystic kidney disease 2 protein homolog) (Autosomal dominant polycystic kidney disease type II protein) (Polycystwin) (R48321).	1.20	213.4	92.8
Sertad1	SERTA domain containing 1	1.16	413.3	185.0
E2f6	E2F transcription factor 6	1.15	76.3	34.4
Reck	Reversion-inducing-cysteine-rich protein with kazal motifs	1.14	198.8	90.4
Loh11cr2a	Loss of heterozygosity, 11, chromosomal region 2, gene A homolog (human)	1.13	1046.0	478.0
Tgfb1	Transforming growth factor, beta 1	1.11	177.0	82.1
Cdc2l5	Cell division cycle 2-like 5 (cholinesterase-related cell division controller)	-1.23	49.9	116.7
Gadd45a	Growth arrest and DNA-damage-inducible 45 alpha	-1.30	1214.2	2997.4
Ddit3	DNA-damage inducible transcript 3	-1.56	648.9	1912.5
Atf5	Cyclic AMP-dependent transcription factor ATF-5	-1.61	482.3	1474.6

[a] Log F.C.: $\log_2$ fold change, $\log_2\left(\frac{average\,(I)\_IgG\ clones}{average\,(I)\_control}\right)_{Day15}$

[b] Avg. int. IgG: average intensity of all IgG clones at Day 15

[c] Avg. int. C: average intensity of control population at Day 15

Table II – List of differentially expressed genes at Day 15 between control population and mAb-producing clones in the functional class 'Response to stress.'

Gene Symbol	Description	Log F.C.[a]	Avg. int. IgG[b]	Avg. int. C[c]
Fabp4	Fatty acid binding protein 4, adipocyte	5.43	371.7	8.6
Rarres2	Retinoic acid receptor responder (tazarotene induced) 2	4.23	1645.9	87.8
Spp1	Pogo transposable element with ZNF domain	3.47	112.3	10.1
N.A.	Myeloid cell nuclear differentiation antigen	3.23	304.3	32.5
Smad7	MAD homolog 7 (Drosophila)	2.45	108.4	19.9
Ereg	Epiregulin	2.39	1398.9	266.1
Ccl2	Chemokine (C-C motif) ligand 2	2.16	2972.8	665.9
THBS1	Thrombospondin-1 precursor.	2.03	281.8	69.0
Errfi1	ERBB receptor feedback inhibitor 1	1.90	1223.9	329.0
Ercc1	Excision repair cross-complementing rodent repair deficiency, complementation group 1	1.26	1329.2	555.9
Zfp36	Zinc finger protein 36	1.25	494.4	207.6
Fn1	Fibronectin 1	1.12	1354.5	624.8
Tgfb1	Transforming growth factor, beta 1	1.11	177.0	82.1
N.A.	Putative ISG12(a) protein	1.06	85.7	41.2
Cyp4v3	Cytochrome P450, family 4, subfamily v, polypeptide 3	-1.10	72.0	154.9
Ptx3	Pentraxin related gene	-1.14	338.0	746.5
Sqstm1	Sequestosome 1	-1.20	1493.1	3429.7
Ccrl2	Chemokine (C-C motif) receptor-like 2	-1.21	243.1	561.3
Gadd45a	Growth arrest and DNA-damage-inducible 45 alpha	-1.30	1214.2	2997.4
Ddit3	DNA-damage inducible transcript 3	-1.56	648.9	1912.5

[a] F.C.: fold change, $\left(\frac{average\,(I)\_IgG\ clones}{average\,(I)\_control} \right)_{Day15}$

[b] Avg. int. IgG: average intensity of all IgG clones at Day 15

[c] Avg. int. C: average intensity of control population at Day 15

Table III – List of differentially expressed genes at Day 15 between control population and mAb-producing clones in the functional class 'Protease inhibitor activity.'

Gene Symbol	Description	Log F.C.[a]	Avg. int. IgG[b]	Avg. int. C[c]
Serpine1	Serine (or cysteine) peptidase inhibitor, clade E, member 1	2.31	72.1	14.5
Timp2	Tissue inhibitor of metalloproteinase 2	2.18	890.5	196.4
THBS1	Thrombospondin-1 precursor.	2.03	281.8	69.0
Slpi	Secretory leukocyte peptidase inhibitor	1.77	1065.5	312.9
CRIM1	Cysteine-rich motor neuron 1 protein precursor (CRIM-1) (Cysteine-rich repeat-containing protein S52).	1.25	592.5	248.9
Reck	Reversion-inducing-cysteine-rich protein with kazal motifs	1.14	198.8	90.4

[a] F.C.: fold change, $\left(\frac{average\,(I)\_IgG\ clones}{average\,(I)\_control} \right)_{Day15}$

[b] Avg. int. IgG: average intensity of all IgG clones at Day 15

[c] Avg. int. C: average intensity of control population at Day 15

Table IV – List of differentially expressed genes at Day 15 between control population and mAb-producing clones in the functional class 'Rho guanine nucleotide exchange factors.'

Gene Symbol	Description	Log F.C.[a]	Avg. int. IgG[b]	Avg. int. C[c]
Farp1	FERM, RhoGEF (Arhgef) and pleckstrin domain protein 1 (chondrocyte-derived)	1.85	329.6	91.3
E130112L23Rik	RIKEN cDNA E130112L23 gene (E130112L23Rik), mRNA	1.15	477.0	215.4
Plekhg2	Pleckstrin homology domain containing, family G (with RhoGef domain) member 2	1.01	71.6	35.6

[a] F.C.: fold change, $\left(\frac{average\,(I)\_IgG\ clones}{average\,(I)\_control} \right)_{Day15}$

[b] Avg. int. IgG: average intensity of all IgG clones at Day 15

[c] Avg. int. C: average intensity of control population at Day 15

Table V – List of differentially expressed genes at Day 15 between control population and mAb-producing clones in the functional class 'tRNA aminoacylation.'

Gene Symbol	Description	Log F.C.[a]	Avg. int. IgG[b]	Avg. int. C[c]
Tars	Threonyl-tRNA synthetase	-1.07	588.4	1233.9
Lars	Leucyl-tRNA synthetase	-1.17	684.5	1541.0
Yars	Tyrosyl-tRNA synthetase	-1.18	616.6	1395.5
Sars	Seryl-aminoacyl-tRNA synthetase	-1.20	1096.0	2523.2
Nars	Asparaginyl-tRNA synthetase	-1.36	810.3	2079.3

[a] F.C.: fold change, $\left(\frac{average\,(I)\_IgG\ clones}{average\,(I)\_control} \right)_{Day15}$

[b] Avg. int. IgG: average intensity of all IgG clones at Day 15

[c] Avg. int. C: average intensity of control population at Day 15

Table VI – List of differentially expressed genes at Day 15 between control population and mAb-producing clones in the functional class 'Extracellular matrix.'

Gene Symbol	Description	Log F.C.[a]	Avg. int. IgG[b]	Avg. int. C[c]
Col6a1	Orocollagen, type VI, alpha 1	4.68	507.0	19.8
Spp1	Pogo transposable element with ZNF domain	3.47	112.3	10.1
Timp2	Tissue inhibitor of metalloproteinase 2	2.18	890.5	196.4
Col18a1	Procollagen, type XV	2.01	78.0	19.3
Ecm1	Extracellular matrix protein 1	1.42	164.0	61.4
Col5a1	Procollagen, type V, alpha 1	1.34	622.1	245.6
Col4a1	Procollagen, type IV, alpha 1	1.18	63.3	27.9
Agrn	Agrin	1.14	402.5	182.7
Fn1	Fibronectin 1	1.12	1354.5	624.8
Tgfb1	Transforming growth factor, beta 1	1.11	177.0	82.1
GAD2	Glutamate decarboxylase 2	-1.27	72.9	175.8
DCN	Decorin precursor (Bone proteoglycan II)	-3.02	73.2	593.1

[a] F.C.: fold change, $\left(\frac{average\,(I)\_IgG\ clones}{average\,(I)\_control} \right)_{Day15}$

[b] Avg. int. IgG: average intensity of all IgG clones at Day 15

[c] Avg. int. C: average intensity of control population at Day 15

Table VII – List of differentially expressed genes at Day 15 between control population and mAb-producing clones in the functional class 'Hydrolase activity, acting on ester bonds.'

Gene Symbol	Description	Log F.C.[a]	Avg. int. IgG[b]	Avg. int. C[c]
N.A.	Calcitonin receptor precursor (CT-R) (C1A C1B).	2.21	81.0	17.5
Pde1b	Phosphodiesterase 1B, Ca2+-calmodulin dependent	2.02	100.8	24.9
Ppap2b	Phosphatidic acid phosphatase type 2B	1.85	1807.3	502.7
Acpl2	Acid phosphatase-like 2	1.61	480.8	157.7
Endod1	Endonuclease domain containing 1	1.58	895.6	300.0
N.A.	PREDICTED: similar to Dual specificity phosphatase 14	1.40	314.5	118.8
TNS3	Tensin-like SH2 domain containing 1	1.40	404.8	153.6
Camk2g	Calcium calmodulin-dependent protein kinase II gamma	1.24	361.4	153.2
PDE4B	cAMP-specific 3 ,5 -cyclic phosphodiesterase 4B	1.14	246.4	111.5
Mtmr6	myotubularin related protein 6	1.04	1007.9	489.2
Pfkfb3	6-phosphofructo-2-kinase fructose-2,6-biphosphatase 3	-1.15	239.6	532.6
Dnase1l3	Deoxyribonuclease 1-like 3	-1.17	121.4	273.5
Ptp4a1	Protein tyrosine phosphatase 4a1	-1.18	48.7	110.2
Rpp38	Ribonuclease P MRP 38 subunit (human)	-1.22	207.9	484.0
N.A.	Dual specificity protein phosphatase 6 (EC 3.1.3.48)	1.86	465.9	128.1
Ercc1	Excision repair cross-complementing rodent repair deficiency, complementation group 1	1.26	1329.2	555.9

[a] F.C.: fold change, $\left(\frac{average\,(I)\_IgG\ clones}{average\,(I)\_control} \right)_{Day15}$

[b] Avg. int. IgG: average intensity of all IgG clones at Day 15

[c] Avg. int. C: average intensity of control population at Day 15

Table VIII – List of differentially expressed genes following methotrexate treatment identified in all mAb-producing clones.

Gene Symbol	Description	Log F.C.[a]	Avg. int. Day 0[b]	Avg. int. Day 15[c]
N.A.	Mus musculus BAC clone RP24-325G24 from chromosome 14	1.57	40.4	103.7
Zfp53	zinc finger protein 53	1.47	46.9	115.1
Dhfr	dihydrofolate reductase	1.35	527.6	1352.9
2610101N10Rik	U2-associated SR140 protein	1.32	63.3	144.2
Ugcgl2	UDP-glucose ceramide glucosyltransferase-like 2	1.22	51.4	115.1
5730406M06Rik	8 days embryo whole body cDNA, RIKEN full-length enriched library, clone:5730525O18 product:SR rich protein homolog	1.20	121.5	239.6
Nufip2	nuclear fragile X mental retardation protein interacting protein 2	1.08	403.0	765.7
Gas2	growth arrest specific 2	1.01	188.2	328.6
Sfrs5	splicing factor, arginine serine-rich 5 (SRp40, HRS)	0.96	564.1	1048.1
Rock2	Rho-associated coiled-coil forming kinase 2	0.77	83.0	138.4
N.A.	Mouse DNA sequence from clone RP23-271B13 on chromosome 2	0.76	127.6	215.6
Dst	dystonin	0.72	165.3	264.1

[a] F.C.: fold change, $\left(\frac{average\,(I)\_Day\,15}{average\,(I)\_Day\,0} \right)_{IgG\ Clones}$

[b] Avg. int. Day 0: average intensity of all IgG clones at Day 0

[c] Avg. int. Day 15: average intensity of all IgG clones at Day 15

Table XI: List of differentially expressed genes grouped by enriched functional classes identified solely through gene ontology enrichment.

Gene Symbol	Description	Euclidean distance	Average intensity (22 hr)	Fold change 37°C (168 hr)/ 37°C (22 hr)	Fold change 33°C + NaBu[1] (168 hr)/33°C + NaBu (22 hr)
Golgi apparatus					
Ap4s1	Adaptor-related protein complex AP-4, sigma 1	5.62	349	-1.8	-15.0
Ganab	Alpha glucosidase 2 alpha neutral subunit	5.55	570	1.5	-1.2
Tapbp	TAP binding protein	5.54	297	1.9	-1.4
M6prbp1	Mannose-6-phosphate receptor binding protein 1	4.36	119	1.4	-1.3
Pkmyt1	Protein kinase, membrane associated tyrosine threonine 1	3.96	290	-12.0	-110.0
Copg2	Coatamer protein complex, subunit gamma 2	3.61	622	1.0	2.2
Ap3m1	Adaptor-related protein complex 3, mu 1	3.23	111	-1.2	1.2
Chst12	Carbohydrate sulfotransferase 12	2.61	481	-1.3	2.4
Ap1g2	Adaptor protein complex AP-1, gamma 2 subunit	2.51	55	1.1	2.0
Slc35a2	Solute carrier family 35, member 2	2.50	24	1.0	7.6
Ap3d1	Adaptor-related protein complex 3, delta 1 subunit	1.71	390	1.1	2.9
Ap2a2	Adaptor protein complex AP-2, alpha 2 subunit	1.17	562	1.1	2.2
Fatty acid beta-oxidation					
Cav1	Caveolin	4.00	1052	1.1	-1.8
Ptgs1	Prostaglandin-endoperoxidase synthase 1	3.55	85	1.0	-2.4
Cd36	CD36 antigen	3.47	794	3.7	1.4
Hadhsc	L-3-hydroxyacyl-CoA dehydrogenase, short chain	3.15	356	1.5	8.6
Acaa2	Acetyl-CoA acyltransferase 2	2.50	901	1.4	6.2
Fasn	Fatty acid synthase	2.48	992	-1.1	-3.6
Acadm	Acetyl-CoA dehydrogenase, medium chain	2.30	272	1.1	3.7
Mecr	Mitochondrial trans-2-enoyl-CoA reductase	2.24	322	-1.2	1.8
Hadhb	Hydroxyacy-CoA dehydrogenase 3-ketoacyl-CoA thiolase enoyl-CoA hydratase, subunit B	1.77	1031	1.3	3.8
Acsl1	Acyl-CoA synthetase long-chain family member 1	1.60	478	2.3	5.7

Gene Symbol	Description	Euclidean distance	Average intensity (22 hr)	Fold change 37°C (168 hr)/ 37°C (22 hr)	Fold change 33°C + NaBu[1] (168 hr)/33°C + NaBu (22 hr)
Hadha	Hydroxyacyl-CoA dehydrogenase 3-ketoacyl-CoA thiolase enoyl-CoA hydratase, subunit A	1.56	444	1.9	3.6
Acsl4	Acyl-CoA synthetase long-chain family member 4	1.41	1807	-1.2	1.6
Regulation of small GTPase-mediated signaling					
Kndc1	Kinase non-catalytic C-lobe domain containing 1	7.25	15	1.0	14.0
Ect2	Ect2 oncogene	4.36	1492	-19.2	-198.4
Rab5c	RAB5C	4.31	263	-1.2	1.6
Arfip2	ADP-ribosylation factor interacting protein 2	3.49	23	2.0	2.5
Trim23	Tirpartite motif protein 23	3.44	215	1.6	-1.3
Ngef	Neuronal guanine exchange factor	2.79	424	-2.0	-4.1
Gnbpg	Guanine nucleotide binding protein G	2.75	434	-1.4	2.8
Rapgef1	Rap guanine nucleotide exchange factor (GEF) 1	2.43	96	-1.6	2.9
Plekhg2	Pleckstrin homology domain containing 2	2.41	34	1.0	2.4
Arl4c	ADP-ribosylation factor-like 4C	2.28	345	-2.4	-1.2
Isn1	Intersectin 1	2.28	142	-2.4	2.1
M9B	Myosin-9B	2.23	380	-1.2	2.5
Rsu1	Ras suppressor protein 1	2.04	735	1.2	2.7
Rnd2	Rho family GTPase 2	2.01	60	1.0	2.0
Rras	Harvey rat sarcoma oncogene, subgroup R	1.57	419	1.7	2.5
Grb2	Growth factor receptor bound protein 2	1.56	682	-1.2	2.1
C330021 A05Rik	Rab-related GTP-binding protein	1.53	68	1.0	2.6
Sar1b	SAR1 gene homolog B	1.43	262	1.9	2.4
Ulk1	Unc-51 like kinase 1	1.37	132	2.1	4.2
Cdc42	Cell division cycle 42 homolog	1.35	4132	1.0	2.0
Kifap3	Kinesin-associated protein 3	1.20	447	1.0	1.6
Arl3	ADP-ribosylation factor-like 3	1.13	282	1.6	2.9

[1] NaBu: sodium butyrate

Table X: List of differentially expressed genes grouped by enriched functional classes identified through both gene set enrichment analysis and gene ontology enrichment analysis.

Gene Symbol	Description	Euclidean distance	Average intensity (22 hr)	Fold change 37°C (168 hr)/ 37°C (22 hr)	Fold change 33°C + NaBu[1] (168 hr)/33°C + NaBu (22 hr)
Cell cycle					
Chek2	CHK2 checkpoint homolog	5.42	144	-3.3	-23.7
Pkmyt1	Protein kinase, membrane associated tyrosine threonine 1	3.96	290	-12.1	-111.0
Atr	Ataxia telangiectasia and Rad3 related	3.85	194	1.0	-16.4
Orc1l	Origin recognition complex, subunit 1-like	3.76	343	-6.4	-13.6
Ccnb1	Cyclin B1	3.70	3166	-17.5	-99.7
Cdc20	Cell division cycle 20 homolog	3.66	2201	-14.9	-88.9
Mcm5	Minichromosome maintenance deficient 5, cell division cycle 46	3.62	1004	-20.9	-146.2
Cdc2a	Cell division cycle 2 homolog A	3.50	3014	-5.2	-19.8
Hdac7a	Histone deacetylase 7A	3.35	87	-2.9	-5.7
Ccnb2	Cyclin B2	3.28	1554	-13.8	-65.5
Mcm3	Minichromosome maintenance deficient 3	2.87	11165	-7.3	-20.1
Bub1	Budding uninhibited by benzimidazoles 1 homolog	2.85	104	-3.8	-5.7
Ywhag	3-monooxygenase tryptophan 5-monoxygenase activation protein	2.72	454	-1.1	-5.5
Plk1	Polo-like kinase 1	2.64	956	-12.4	-21.5
Ccna2	Cyclin A2	2.41	1546	-25.0	-59.4
Prkdc	Protein kinase, DNA activated, catalytic polypeptide	2.39	108	2.3	1.1
E2f6	E2F transcription factor 6	2.38	88	1.4	2.6
Mcm7	Minichromosome maintenance deficient 7	2.38	2359	-3.4	-9.3
Extracellular matrix					
Col4a5	Procollagen, type IV, alpha 5	4.71	154	2.3	-1.4
Sgcg	Gamma sarcoglycan	3.57	107	5.4	2.3
Efemp2	Epidermal growth factor-containing fibulin-like extracellular matrix protein 2	3.34	304	1.9	1.1
Spp1	Pogo transposable element with ZNF domain	3.22	461	2.9	10.3
Nid1	Nidogen 1	3.16	1922	2.4	1.0
Fn1	Fibronectin 1	2.70	232	1.7	9.3
Lama2	Laminin, alpha 2	2.70	248	-1.8	-2.9
Mmp14	Matrix metallopeptidase 14	2.50	338	2.1	1.0
Col12a1	Procollagen, type XII, alpha 1	2.44	120	1.0	2.7

Gene Symbol	Description	Euclidean distance	Average intensity (22 hr)	Fold change 37°C (168 hr)/ 37°C (22 hr)	Fold change 33°C + NaBu[1] (168 hr)/33°C + NaBu (22 hr)
DCN	Decorin precursor	2.35	82	2.3	6.3
Timp2	Tissue inhibitor of metalloproteinase 2	2.26	1024	2.0	-1.5
GABRA3	Gamma-aminobutyric-acid receptor alpha-3 subunit precursor	2.21	28	1.0	7.3
Ctgf	Connective tissue growth factor	2.20	247	1.3	4.1
Growth factor activity					
Arts1	Type 1 tumor necrosis factor receptor shedding aminopeptidase regulator	3.65	450	1.2	-5.7
Bmp1	Bone morphogenic protein 1	3.45	127	1.8	-2.3
Spp1	Pogo transposable element with ZNF domain	3.22	461	2.9	10.3
Cxcl12	Chemokine (C-X-C motif) ligand 12	3.09	84	-2.7	1.1
Hdgf	Hepatoma-derived growth factor	2.79	1039	-1.1	-5.5
Fgf13	Fibroblast growth factor 13	2.75	278	-1.2	-2.6
Vegfa	Vascular endothelial growth factor A	2.46	1030	1.7	-2.1
Ion transport					
Slco4a1	Solute carrier organic anion transporter family, member 4a1	5.98	884	1.7	-6.2
Clcn2	Chloride channel 2	4.81	32	1.0	3.6
Cab3	Voltage-dependent L-type calcium channel beta-3 subunit	2.85	35	1.0	2.9
Svop	SV2 related protein	2.40	26	1.0	4.6
GABRA3	Gamma-aminobutyric-acid receptor alpha-3 subunit precursor	2.21	28	1.0	7.37
Cytoplasmic membrane-bound vesicles					
Ehd3	EH-domain containing 3	4.46	86	-1.3	3.6
Ehd1	EH-domain containing 1	4.32	243	-2.3	3.3
Rab5c	RAB5C, member RAS oncogene family	4.31	527	-1.2	1.6
Anp32e	Acidic (leucine0rich) nuclear phosphoprotein 32 family, member E	3.60	405	-1.7	-1.9
Slc17a5	Solute carrier family 17, member 5	3.55	63	1.1	4.3
Syp	Synaptophysin	2.67	30	1.0	6.9
Ap2a2	Adaptor protein complex AP-2, alpha 2 subunit	1.17	562	1.1	2.2
Cytoskeletal protein binding					
Mapk8ip1	Mitogen activated protein kinase 9 interacting protein 1	4.96	13	1.0	7.8
C130076O07Rik	Neuronal cell adhesion molecule precursor	4.82	20	1.0	10.8

Gene Symbol	Description	Euclidean distance	Average intensity (22 hr)	Fold change 37°C (168 hr)/ 37°C (22 hr)	Fold change 33°C + NaBu[1] (168 hr)/33°C + NaBu (22 hr)
Nusap1	Nucleolar and spindle associated protein 1	2.81	1095	-9.6	-21.5
Myo9b	Myosin-9b	2.22	380	-1.2	2.5
Ktn1	Kinectin 1	2.21	36	1.0	5.5
Coro1c	Coronin, actin binding protein 1C	2.21	645	-1.7	2.7
Myh9	Myosin, heavy polypeptide 9, non-muscle	2.16	309	-1.4	4.3
Flnb	Filamin, beta	1.47	248	1.3	3.6

[1] NaBu: sodium butyrate

Table XI: Other differentially expressed genes which are not a member of any identified enriched functional class.

Gene Symbol	Description	Euclidean distance	Average intensity (22 hr)	Fold change 37°C (168 hr)/37°C (22 hr)	Fold change 33°C + NaBu[1] (168 hr)/33°C + NaBu (22 hr)
Glycosylation					
Alg8	Asparagine-linked glycosylation 8 homolog	2.45	309	-1.7	-3.2
Ganab	Alpha glucosidase 2 alpha neutral subunit	5.50	570	1.5	-5.9
Protein transport					
Ap4s1	Adaptor-related protein complex, AP-4, sigma subunit 1	5.61	349	-1.8	-15.0
Arl4c	ADP-ribosylation factor-like 4C	2.28	347	-2.4	-1.2
Cadps	Ca^{2+}-dependent activator protein for secretion	3.14	59	4.0	1.0
Copg2	Coatamer protein complex, subunit gamma 2	3.61	622	1.0	2.2
Cpne5	Copine V	3.50	26	1.0	8.6
Rpl23	Ribosomal protein L23	4.35	35	2.0	-1.3
Tmem48	Transmembrane protein 48	2.36	390	-1.9	-5.2
Protein folding					
Dnajc9	DnaJ (Hsp40) homolog, subfamily C, member 9	2.31	499	-1.5	-3.0
MDN1	Midasin	2.74	180	1.5	-1.6
Ppid	Peptidylprolyl isomerase D	3.09	772	1.4	-5.0
Serpinh1	Serine (or cysteine) peptidase inhibitor	4.70	3042	1.6	-4.0
Lipid metabolism					
Bbs4	Bardet-Beidl 4 homolog	2.22	159	1.6	1.1
Gm2a	GM2 ganglioside activator protein	2.35	158	-1.3	2.7
Hmgcr	3-hydroxy-3-methylglutaryl-Coenzyme A reductase	2.87	584	-2.5	-9.6
Hmgcs1	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1	2.23	2920	-3.5	-8.4
Isyna1	Myo-inositol 1-phosphate synthase A1	2.42	173	-2.2	1.5
Mecr	Mitochondrial trans-2-enoyl-CoA reductase	2.24	170	-1.2	1.9
Npc1	Niemann Pick type C1	2.29	295	1.9	2.5
Nr2f2	Nuclear receptor subfamily 2, group F, member 2	2.47	307	-1.3	-2.6
Nsmaf	Neutral sphigomyelinase activation associated factor	2.64	290	-1.3	-2.9
Osbpl8	Oxysterol binding protein-like 8	2.24	112	-1.1	2.4
Pitpna	Phosphotadylinositol transfer protein, alpha	3.98	496	-2.3	2.5

Gene Symbol	Description	Euclidean distance	Average intensity (22 hr)	Fold change 37°C (168 hr)/37°C (22 hr)	Fold change 33°C + NaBu[1] (168 hr)/33°C + NaBu (22 hr)
Pitpnm1	Phosphotadylinositol membrane-associated 1	2.69	37	3.1	1.0
Psap	Prosaposin	6.60	581	2.4	1.2
Ptpmt1	Protein tyrosine phosphatase, mitochondrial 1	2.59	624	1.0	-2.8
Rarres2	Retinoic acid receptor responder 2	3.65	30	1.0	4.5
Phosphoinositide metabolism					
Ptpmt1	Protein tyrosine phosphatase, mitochondrial	2.59	624	1.0	-2.8
Regulation of DNA transcription					
4933413G19Rik	Forkhead box M1	2.57	301	-2.6	-5.8
AFF3	AF4 FMR2 family member 3	3.45	8	1.0	11.4
ARID1A	AT-rich interacting domain-containing protein 1A	2.34	513	-1.4	-3.6
Anks3	Ankryn repeat and sterile alpha motif domain containing 3	2.25	75	1.9	-1.3
Btg2	B-cell translocating gene 2	2.60	80	1.1	8.2
ETV5	ETS translocation variant 5	3.24	244	1.0	-5.1
Esx1	Extraembryonic, spermatogenesis, homeobox 1	3.71	1364	-4.2	1.4
Gtf2i	General transcription factor II i	2.50	374	1.4	-2.0
Jun	Jun oncogene	2.41	1247	-1.8	1.2
Keap1	Kelch-like ECH-associated protein 1	2.27	360	1.5	1.2
Maz	MYC-associated zinc finger protein	2.57	280	-3.5	-2.6
Msrb2	Methionine sulfoxide reductase B2	2.98	256	1.1	-2.1
Nr2f2	Nuclear receptor subfamily 2, group F	2.47	307	1.9	1.0
PPP1R12B	Protein phosphatase 1 regulatory subunit 12B	3.66	133	2.1	1.0
Ssbp2	Single-stranded DNA binding protein 2	2.58	56	1.7	7.7
Sub1	SUB1 homolog	2.41	144	-1.6	2.5
Zf	HCF-binding transcription factor Zhangfei	2.47	145	-1.3	-2.9
Zfhx1a	Zinc finger homeobox 1a	2.28	582	1.8	-1.3
Zfp326	Zinc finger protein 326	2.60	342	1.0	-2.0
Zfp367	Zinc finger protein 367	2.87	267	-8.0	-2.1
Apoptosis					
Fis1	Fission 1 homolog	2.23	249	1.4	-3.0
Gulp1	GULP, engulfment adaptor PTB domain containing 1	2.26	570	1.3	-1.6
Itgb3bp	Expressed sequence AU022583	3.20	65	1.6	-3.2
Nupr1	Nuclear protein 1	3.15	1535	4.2	-1.4
Pcbp4	Poly(rC) binding protein 4	2.54	160	-1.1	4.0

Gene Symbol	Description	Euclidean distance	Average intensity (22 hr)	Fold change 37°C (168 hr)/37°C (22 hr)	Fold change 33°C + NaBu[1] (168 hr)/33°C + NaBu (22 hr)
Txnip	Thioredoxin interacting protein	4.06	615	2.5	23.3
Other genes					
4930583H14Rik	Ovary-specific acidic protein	5.30	60	1.8	34.3
Cbara1	Calcium binding atopy-related autoantigen 1	2.44	400	-1.1	2.5
Cd68	CD68 antigen	2.59	102	6.6	8.6
Gstm2	Glutathione S-transferase, mu 2	2.20	528	1.3	4.7

[1] NaBu: sodium butyrate