Published in final edited form as:
Biotechnol Adv. 2012 ; 30(3): 629–638. doi:10.1016/j.biotechadv.2011.10.010.

Genomics in mammalian cell culture bioprocessing

Diane M. Wuest[a], **Sarah W. Harcum**[b], and **Kelvin H. Lee**[a,*]

[a]Chemical Engineering and Delaware Biotechnology Institute, University of Delaware, 15 Innovation Way, Newark, DE 19711, USA

[b]Bioengineering, Clemson University, 301 Rhodes Research Center, Clemson, SC 29634, USA

Abstract

Explicitly identifying the genome of a host organism including sequencing, mapping, and annotating its genetic code has become a priority in the field of biotechnology with aims at improving the efficiency and understanding of cell culture bioprocessing. Recombinant protein therapeutics, primarily produced in mammalian cells, constitute a $108 billion global market. The most common mammalian cell line used in biologic production processes is the Chinese hamster ovary (CHO) cell line, and although great improvements have been made in titer production over the past 25 years, the underlying molecular and physiological factors are not well understood. Confident understanding of CHO bioprocessing elements (e.g. cell line selection, protein production, and reproducibility of process performance and product specifications) would significantly improve with a well understood genome. This review describes mammalian cell culture use in bioprocessing, the importance of obtaining CHO cell line genetic sequences, and the current status of sequencing efforts. Furthermore, transcriptomic techniques and gene expression tools are presented, and case studies exploring genomic techniques and applications aimed to improve mammalian bioprocess performance are reviewed. Finally, future implications of genomic advances are surmised.

Keywords

Biologics; Genomics; Mammalian cell culture bioprocessing; Next-generation genomic sequencing; Proteomics; Recombinant DNA technology; Recombinant protein production; Transcriptomics

1. Introduction

A genome contains all of the information necessary to construct an organism, and the particular arrangement of the nucleotides that comprise genomic deoxyribonucleic acid (DNA) specifies its uniqueness. Explicitly identifying an organism's genome is highly valuable for biological research and has become a priority in the field of biotechnology following the development of genomic sequencing tools. One of the most ambitious goals in life science is to derive a relationship between the genome sequence of a cell and its non-linear cellular protein dynamics. If an understanding of the link between genome sequence and protein expression can be obtained, a quantitative description of cellular phenomena can follow. A holistic analysis of gene function and expression at the systems level is more complex than the genome sequencing itself and requires the concurrent implementation of high-throughput global genomic and proteomic technologies. These technologies foster

*Corresponding author. Tel.:+1 302 831 0344; fax: +1 302 831 4841. KHL@udel.edu (K.H. Lee).

exploration and innovation in biotechnology, specifically in cell culture bioprocessing where cells are engineered to produce therapeutics. Variations in genomic sequence (via DNA deletion, translocation, amplification, and mutation) and gene expression (via transcription, transcript processing, translation, and post-translational modification) must be identified and harnessed to effectively engineer a desired cell phenotype.

Developing desirable phenotypes, such as high productivity and robust health and longevity in harsh environments, is advantageous in commercial cell culture bioprocesses. Recombinant DNA technology, which employs genetic engineering to isolate, manipulate, and reintroduce DNA into a cell, modifies a cell to become a factory for the production of foreign protein products and can subsequently alter its phenotype. The first genetically engineered recombinant protein approved by the US Food and Drug Administration (FDA) was synthetic human insulin, a recombinant hormone developed in 1977 by Herbert Boyer (founder of Genentech Inc.) and later acquired by Eli Lilly and Company. Today, sixteen categories of medications produced through biological processes involving recombinant DNA (referred to as biologics, biopharmaceuticals, or recombinant proteins) constitute a $108 billion market and are listed in Table 1 with respective 2010 sales (La Merie Business Intelligence, 2010). For a complete financial analysis of the many biologics encompassed in Table 1, see reference La Merie Business Intelligence (2010).

Although the rate of approval of new biologics has decreased because of the growing prevalence of biosimilars and the need for cost containment as a result of the late 2000s global financial crisis, biologic production remains a significant and growing sector of the overall pharmaceutical market (Walsh, 2010). The high potency and ability of recombinant proteins to combat diseases beyond the range of traditional small molecule drugs give the industry opportunity for future growth. Therefore, the importance of biologics in the pharmaceutical sector drives cost-effective advances in cell culture bioprocess manufacturing.

2. Mammalian cells in bioprocessing

Recombinant proteins are primarily produced in mammalian cells (versus microbial fermentation) because the necessary and often complex tasks of proper assembly, folding, and post-translational modifications (PTMs) can be only efficiently carried out with mammalian cell machinery. The majority of therapeutic proteins display at least one PTM which greatly influences the biochemical and therapeutic properties of the protein *in vivo*. This unique advantage in using mammalian cells outweighs the costs of their longer development and manufacturing times compared to bacterial cells such as *Escherichia coli* (*E. coli*). The first approved biologic from a mammalian bioprocess platform was tissue plasminogen activator (tPA), produced in 1987 by Genentech Inc. Today, production of biologics in mammalian cells dominates: of the 58 biopharmaceutical products approved between 2006 and 2010, 32 were produced in mammalian cells, 17 were produced in *E. coli,* four in yeast, three in transgenic animals, and two in insect cultures (Walsh, 2010).

The most common mammalian cell line used in biologic production processes is the Chinese hamster ovary (CHO) cell line, which was isolated from the Chinese hamster in the 1950s (Tjio and Puck, 1958). In the early 1980s, under the direction of Dr. Chasin at Columbia University, two derivatives of the CHO cell line, CHO-K1 and CHO pro-3, gave rise to the two most commonly used cell lines in bioprocessing today, DUKX-X11 and DG44. Both cell lines were engineered to be deficient in dihydrofolate reductase (DHFR) activity: chemical mutagenesis was used in DUKX-X11 to delete one DHFR allele and mutate the other, and ionizing radiation was used in DG44 to delete both alleles (Lee et al., 2010; Wurm and Hacker, 2011). A cell line deficient in DHFR activity, which requires the

addition of glycine, hypoxanthine, and thymidine (GHT) in the medium for survival, allows for the implementation of a selection system based on the insertion of a cloned *dhfr* gene in combination with the gene of interest. When grown in GHT deficient media, only the cells containing the recombinant vector can survive and produce the protein of interest. It is not surprising that CHO cells were used to produce tPA in 1987 and are still the most commonly used host today because, over the past 25 years, research has focused on manipulating established host cell lines, not on searching for a better host system. With a small number of host cell lines used, including CHO, NS0 (mouse myeloma derived), human embryonic kidney, and baby hamster kidney cell lines, it is more efficient for scientists to build upon established resources to genetically introduce desired characteristics and develop optimized standard production processes in a familiar host that is considered safe by regulatory agencies. Since 1987, mammalian cell culture manufacturing processes have effectively evolved to a single platform of large-scale, single-cell suspension cultures grown in fed-batch, stirred tank bioreactors.

These commonly used host cell lines have the important capability of adapting to suspension growth in the absence of serum (and animal products) and can be continuously grown to produce biologics for a few months at a time. Driven by the increasing demand for biopharmaceutical therapeutics since their inception, there has been a 20-fold increase in specific and volumetric productivity as a consequence of effective media and bioprocess optimization efforts. In the 1980s, a typical biologic manufacturing process extended over 7 days and achieved 1–2 million cells/mL that were able to produce 50–100 mg/L titer. Today, a typical process extends over 21 days and achieves 10–15 million cells/mL that are able to produce 1–5 g/L titer (Hacker et al., 2009). The gain in titer can be attributed to the optimization of clone screening and selection strategies to identify optimal phenotype characteristics and of media composition, feeding strategies, and bioprocess conditions (e.g. temperature, pH, hyperosmotic pressure, and the addition of small chemical compounds, such as butyrate, zinc sulfate, cytochalasin D, and ammonium ions) to sustain higher cell densities with increased production capabilities.

Although great improvements have been made in production titer, the underlying molecular and physiological factors are not well understood, particularly on the systems-level. Some cellular pathways (e.g. cellular metabolism, growth, and PTMs) have been explored using microarray technology to screen cell lines, leading to biologic production and performance improvements. DNA microarray technology, which has been used in many disciplines, such as disease characterization, diagnostics development, and forensics, is one of the most versatile genomic tools for bioprocess development because it is capable of identifying gene expression levels of an entire genome in one experiment (Jaluria et al., 2007). For example, experiments can be performed to compare the abilities of cell lines to produce different levels of recombinant protein or to compare the effects of media components or environmental conditions on productivity. These types of experiments have revealed a wide range of important genes and proteins from functional classes such as protein folding, secretion, cell metabolism, cytoskeletal architecture, cell growth, and apoptosis. However, the challenge of translating these findings into altered cellular phenotypes with improved productivity and product quality still remains. It is often hard to interpret data from the large and complex experimental results and several cellular functions likely need to be targeted together to achieve an optimum. Substantial progress has been achieved in understanding system-level physiology in bacterial and yeast cultures, and the methodologies used to set up and analyze microbial experiments offer valuable insight for mammalian cell culture experimentation.

In spite of the economic and industrial importance of CHO host cell lines, little is known about their genetic makeup, which is a result of the difficulty and high cost in sequencing a

complex mammalian genome. Sequencing the human genome required world-wide collaboration to pool sequencing, analyzing, and monetary resources for its completion in 2001, and other clinically important model organisms were sequenced soon thereafter via similar collaborative efforts (e.g. mouse in 2002 and rat in 2004). Compared to the size of bacterial genomes (e.g. 4 million base pair *E. coli* genome), mammalian genomes are significantly larger (e.g. 3 billion base pair mouse genome) and arranged in multiple chromosomes that contain introns (intervening segments of DNA that do not code for a product). Furthermore, CHO cell lines undergo genetic rearrangements during successive generations of cell lines and gene amplification, ultimately requiring the need to sequence individually cloned cell lines, not just the Chinese hamster genome. Without the CHO genetic code available, transcriptome and proteome profiling experimentation has relied on comparative information from related species (e.g. mouse) and expressed sequence tag (EST) databases. Although advances have been made in CHO cell culture research, experimental results using comparative genomes or EST databases are limited.

3. Importance of obtaining CHO cell line genetic sequences

Confident understanding of CHO bioprocessing elements (e.g. cell line selection, protein production, and reproducibility of process performance and product specifications) would significantly improve with the sequencing, mapping, and annotating of the CHO cell genome. In the production of standard monoclonal antibody therapeutics, improved cell line selection strategies, based on known genetic information, could lead to more efficient bioprocessing platforms. Traditional methods of clone selection are low-throughput due to the large amount of time and effort required to empirically screen thousands of cells in search of a single high-producing clone and then cultivate it for several months to exclude problems that may appear over time. Once a superior growing and producing clone is selected and passaged for a length of time, a decline in specific productivity often occurs due to low- or non-producing subpopulations overtaking the high-producing cells. Theoretically, all cells within a population should be genetically and phenotypically identical, but in fact, there is a large amount of inherent variability of growth characteristics and specific productivity within populations. This inconsistency of growth and productivity is unpredictable and can compromise regulatory approval (Barnes et al., 2003). Factors important to maintain culture stability, or cells that can maintain a constant level of protein production over long culture times, include the incorporation site of the recombinant gene into the CHO genome and phenotypic differences between cells due to protein expression noise.

Currently, genes are inserted into the cellular chromosome by random transgene integration events leading to a variety of chromosomal integration sites and copy number. Expression of the gene is more dependent on the position of integration than the copy number, and stability is increased when the gene is located in the telomeric regions of the chromosome (Kuystermans et al., 2007). With a known genetic sequence, researchers could control the insertion of the expression system by targeting known genes in the telomeric region. Once the gene is incorporated into the genome for active transcription, stochastic gene fluctuations readily occur to allow cells to adapt to short-lived environmental changes without accumulation of potentially harmful mutations. Prediction of these fluctuations would be difficult even if epigenetic markers were known because these events are driven by random promoter activation bursts and the molecular characterization of individual clones would be challenging. However, with a known genome, strategies used to implement factors to decrease promoter noise (e.g. use of low noise promoters, noise-suppressing endogenous genes, or recruitment of chromatin-opening elements) would be easier to accomplish (Pilbrough et al., 2009). Therefore, targeted insertion of the recombinant gene to enhance population stability and subsequent selection strategies could help decrease the high cost of

current selection strategies and overcome the bottleneck in moving a candidate cell line from development into manufacturing.

If cell line selection could be improved by identifying a stable clone to be utilized across all clinical phases of product development, instead of creating multiple cell lines per project, the efficiency of standard monoclonal antibody bioprocessing platforms would improve. Less development time would be necessary to optimize different clones and increase cell performance. Furthermore, high product yields are now common for standard monoclonal antibody product launches and provide adequate material in the manufacturing process, so large development efforts to increase titer are no longer accentuated. Instead, research and development efforts should focus on understanding process fundamentals, such as controlling product quality and process consistency (Kelley, 2009). Knowledge of genetic mechanisms behind these process fundamentals (e.g. genes that initiate or control product quality) would be of great benefit for manufacturing processes as a complete understanding of product characteristics is important in regulatory control. When compared to standard monoclonal antibodies, characteristics of non-platform monoclonal antibodies (e.g. antibody-toxin drug conjugates and bispecific antibodies) and other recombinant proteins are less well-understood. In the development processes of these products, improving titer is emphasized, and experiments designed with the aid of a known genome sequence would expedite increased protein production achievements.

System-wide profiling experiments performed to clarify process fundamentals of standard monoclonal antibodies and increase titer of unconventional monoclonal antibodies can be best performed referencing the known genetic code of the specific CHO cell line in use. However, if the genome sequence of the specific cell line is unknown, using the Chinese hamster genome as a reference would also be beneficial. Using this more broadly based genome could provide the foundation to explore gene and protein expression and facilitate whole-genome sequencing of CHO cell lines. In 2007, Wlaschin and Hu created a scaffold for the Chinese hamster genome to aid biomolecular research and future efforts to sequence the Chinese hamster genome (Wlaschin and Hu, 2007). The complete sequencing and assembly of the Chinese hamster would facilitate comparative genomic experiments in CHO cells by allowing researchers to study the homologous genomic regions of this model organism. Currently it is unknown exactly how much genomic differences exist between different CHO cell lines and in comparison to the Chinese hamster genome, and once this information is known, researchers will be able to better design sequencing and expression experiments.

Acquiring these important genetic codes would enable DNA sequence-reliant techniques, including micro-ribonucleic acid (miRNA) and DNA microarray chips, to study gene structure, regulation, and expression on a system-wide level. Microarray experiments could identify global gene expression profiles (i.e. the transcriptome) of each experimental case or clone, providing valuable data needed to engineer and improve cellular physiology or understand poorly characterized phenotypes. There are also multiple enzyme-catalyzed pathways, constituting cellular metabolism, involved in cellular processes, and these pathways could be modeled and manipulated to control the accumulation of waste products. Once holistic understanding of the cell is gained, the design of targeted genetic manipulations can advance cell line engineering and ultimately bioprocess protein production.

4. Current status of CHO genome sequencing

Developments in DNA sequencing technology have drastically improved over the past 10 years advancing from the conventional Sanger method to a variety of methods referred to as

next-generation sequencing (NGS). The Sanger method, performed via automated capillary-based Sanger biochemistry (Shendure and Hanlee, 2008), dominated the industry from its inception in 1975 until 2005 and was used to sequence the human genome and several subsequent model organism genomes (e.g. mouse, rat, and dog). Around 2005, NGS became available from several companies (e.g. 454 sequencing from Roche Applied Science, Solexa technology from Illumina, and SOLiD platform from Applied Biosystems) and the technology used to increase sequence output relied on cyclic-array sequencing. Currently available NGS platforms are conceptually similar as they all produce a library by random fragmentation of DNA, amplify the library fragments, and perform automated sequencing reactions, but they differ in specific sequencing chemistry and data output (Table 2). Variation in data output creates a challenge when comparing platforms and forming tailored assembly programs. Most NGS platforms under development target single molecule sequencing technologies based on fluorescence, nano-technology, electronic detection, electron microscopy, and ion sensing. For a complete explanation and discussion of currently available and developing NGS technologies, refer to Mardis (2009, 2011), Metzker (2010), and Zhang et al. (2011).

Researchers interested in sequencing the CHO genome have the ability to implement these commercial NGS platforms and are able to produce enormous volumes of sequence data at a fraction of the time and cost of Sanger sequencing. However, further computational resources are needed for the assembly and annotation of a complete genome, which is accomplished by reference-guided assembly of a closely related organism or *de novo* assembly. Both methods have limitations. In the reference-guided approach, the quality and relatedness to the sequenced species are important factors in selecting a reference genome, and sequence variants may not be detected. In *de novo* assembly, the shorter read lengths from NGS technologies makes assembly difficult, especially in repetitive regions of the genome.

The complete sequencing and assembling of individual cell lines are necessary to fully understand and utilize genome information. An example of reference-guided assembly using mouse and rat reference genomes in an initial genomic analysis of a high-producing CHO cell line was recently demonstrated by the efforts of a single research laboratory (Hammond et al., 2011). In August of 2011, the first draft genomic sequence of an ancestral cell line, CHO-K1, using the *de novo* approach was reported by an international consortium comprised of BGI-Shenzhen (Shenzhen, People's Republic of China), GT Life Sciences (San Diego, CA), Peking University Shenzhen Hospital (Shenzhen, People's Republic of China), University of Delaware (Newark, DE), Technical University of Denmark (Lyngby, Denmark), Stanford University (Stanford, CA), Johns Hopkins University (Baltimore, MD), Technical University of Denmark (Hørsholm, Denmark), and University of Copenhagen (Copenhagen, Denmark) (Xu et al., 2011). The researchers involved generated 2.45 Gb of genomic sequence with approximately 24,000 genes using Illumina sequencing and provided comprehensive annotation. This breakthrough study provides a foundation for future complete CHO cell line genomic sequencing and genomic-based cell line engineering efforts.

Prior to this sequencing milestone, only a few comprehensive collaborations had been forged to advance CHO genomics. For example, a collaboration was formed between the University of Minnesota and the Bioprocessing Technology Institute of Singapore to build a CHO sequence database for complementary DNA (cDNA)-based microarray technology in 2002. In 2006, based on promising results, they formed the Consortium on CHO Cell Genomics in partnership with the Society of Biological Engineers and corporate partners to accelerate the development of CHO genomic resources. In a recent effort to make CHO genomic information publicly accessible, the University of Delaware, in collaboration with

The Johns Hopkins University and the University of Natural Resources and Applied Life Sciences in Vienna (BOKU), created a website (CHOgenome.org) to facilitate the organization and distribution of CHO genome data and resources. The website will host the latest updates of CHO genomes, provide genome browsing tools, and provide links to other genomic databases. For example, the recent draft of the annotated CHO-K1 sequence provided by Xu et al. (2011) is hosted on the website as a starting point for public access. Future plans for the website include integrating transcriptomic, proteomic, and metabolomic data, hosting a user forum with community features, and providing support for the development of sequence-specific user tools.

Other resources required for genome assembly and annotation are also being developed independently that will help in generating complete understanding of the CHO genome. First, the assignment of DNA fragments to chromosome locations via physical mapping is central to the completion of the whole genome sequence and is facilitated by the use of bacterial artificial chromosome (BAC) libraries. This clone-based approach involves generating an overlapping series of clones that represents the entire genome. Progress toward constructing a CHO BAC library, consisting of 122,281 clones, was performed by Omasa et al. (2009) to determine the genetic structure of the 20 individual chromosomes in a CHO-DG44 cell line. Second, annotating the structural arrangement and biological function of genomic elements is required. Annotation is achieved through the use of automated gene prediction programs, which use genome context information, similarity scores, experimental data, and integrations to identify gene matches. Third, understanding which genes are expressed and their expression levels establishes the connection back to the mapped and annotated genome. The set of actively expressed genes includes all RNA molecules (e.g. messenger, ribosomal, transfer, and other non-coding, small RNA) and can vary with cell lines and external environmental conditions. Recently, to identify, annotate, and profile miRNA expression in CHO cells, Hackl et al. (2011) sequenced the small RNA transcriptome of six CHO cell lines and developed a novel method for miRNA identification and annotation in the absence of genomic sequence information. Finally, developing an infrastructure for storing, processing, and analyzing large data sets from high-throughput experiments is crucial in all of the above mentioned components. The reduced cost of NGS technologies has made high-throughput experiments accessible to individual labs (rather than large sequencing centers) that may not have the necessary computational infrastructure. There has been and will continue to be significant growth in the field of bioinformatics, which will help implement an integrated, global view of genomic and proteomic profiles once genomic sequences are fully defined and characterized.

5. Applied genomics in bioprocessing

Typically, applied genomics in the clinical setting is thought of as a means to leverage the information and technologies derived from the Human Genome Project to gain a better understanding of human diseases and to develop therapeutic entities to treat these diseases. Applied genomics in bioprocessing likewise utilizes the same information structure and technologies. However, the focus in bioprocessing is to gain better control over the methods used to manufacture biologics by determining the relationship between the transcriptome and the production pathway within the organism. System-wide transcriptome characterization enables functional genomic studies based on global gene expression patterns, single nucleotide polymorphism (SNP) surveys, methylation assays, small RNA sequencing, and splice variant analysis (Bouck and Vision, 2007; Rudd, 2003). Improvements have been gained in bioprocess development by utilizing applied genomics, and the following sub-sections will highlight tools available to assess the transcriptome, summarize comparative studies that utilized applied genomics, and describe tools that can be used to alter the transcriptome to obtain a desired phenotype.

5.1. Transcriptomic tools

Central to understanding the transcriptional landscape are gene expression analysis tools, including real-time quantitative reverse transcriptase polymerase chain reaction (qRT-PCR), identification of expressed genes via ESTs, DNA microarrays, and transcriptome sequencing (termed RNA-Seq). qRT-PCR is a tool for assessing gene expression of a particular gene of interest by quantifying mRNA levels. Just as in standard PCR where the DNA sequence to be amplified must be known, the mRNA sequence must also be known in qRT-PCR to design primers specific to each measured gene (Nolan et al., 2006). Due to limited genome information precluding a global assessment, the qRT-PCR tool (a gene-by-gene tool) has been widely used to assess the transcriptome in CHO cells under bioprocessing conditions.

Additional information about mRNA expression can be gained through the use of ESTs, which are short partial sequences (500–800 base pairs) of transcribed cDNA that represent part of a full mRNA sequence. Assembly of fragments is necessary to identify full mRNA sequences, and this process is not trivial and can yield artifacts (e.g. contigs that contain two distinct gene products). The assembly process is much simpler when a complete genome sequence is available and annotated because it becomes possible to circumvent EST contig assembly and directly match transcripts. ESTs for unknown genes are often included on DNA microarrays in addition to the known genes.

DNA microarrays are based on the same principle as Southern blots, where two cDNA strands hybridize together to form a double stranded DNA molecule. Unlike Southern blots where only one DNA sequence is examined per blot, a DNA microarray allows for thousands of DNA sequences to be probed simultaneously. Thus, DNA microarray analysis is the most comprehensive and powerful tool regularly used to assess the global gene expression state of an organism. It is a costly technique per sample yet inexpensive per gene examined, relative to qRT-PCR. Additionally, DNA microarrays are not very sensitive to small changes in gene expression, which necessitates the confirmation of large gene expression changes with an orthogonal method, such as qRT-PCR (Canales et al., 2006). Similar to the previous techniques mentioned, this tool is most effective when implemented for species with a known genetic sequence.

RNA-Seq is a process where nearly the entire transcriptome of an organism, under a particular set of conditions, is queried at once using high-throughput DNA sequencing technologies (Wang et al., 2009; Wold and Myers, 2008). Cost-effective massively parallel sequencing is the foundation of the RNA-Seq technique (Wilhelm and Landry, 2009). The technical reproducibility and accuracy of this technique were validated by comparison with gene expression arrays when Marioni et al. (2008) were unable to detect significant technical variation of transcript sequence data between human liver and kidney RNA samples. Sequence data has been determined to be sufficient for identification and analysis of differentially expressed genes, such as detection of low-expressional levels, alternative splice variants, and novel transcripts (Fu et al., 2009; Marioni et al., 2008; Mortazavi et al., 2008; Nagalakshmi et al., 2008). The high cost of RNA-Seq procedures relative to DNA microarrays per sample analyzed, in addition to the unidentified biases due to the ligation steps and known biases for high abundance transcripts has limited its widespread application (Huse et al., 2007; Mane et al., 2009; Margulies et al., 2005; Moore et al., 2006). It is anticipated that RNA-Seq analysis of transcriptomes from relevant bioprocess conditions will increase as sequencing costs decrease.

In the recent release of the draft CHO-K1 genome, RNA-Seq data was obtained, and gene expression levels were queried with respect to the glycosylation pathway (the enzymatic process that attaches glycans to proteins) and viral resistance (Xu et al., 2011). Results indicated that only approximately 50% of the known glycosylation genes were expressed but

included many critical glycosylation genes, such as N-glycan transferases, mannosyltransferases, and sugar nucleotide synthesis genes. Conversely, transcripts for sulfotransferase, fucosyltransferase, and N-acetylgalactosamine were not detected (Xu et al., 2011). These non-detected genes and their related enzymes correlate well with the typical glycosylation profiles observed in CHO cells, which lack these glycans (Jenkins et al., 2009). Additionally, the RNA-Seq results provided insight into the observed viral resistance observed in many CHO cell host cell lines. Specifically, transcripts for viral receptor genes were undetected in the CHO-K1 transcriptome, yet were present on the draft CHO-K1 genome (Xu et al., 2011).

To supplement qRT-PCR, DNA microarray, and RNA-Seq transcriptomic techniques, there are several additional gene expression tools that are gaining utility in bioprocessing. These tools, briefly described in the following paragraphs, include suppression subtractive hybridization (SSH), comparative genomic hybridization (CGH), serial analysis of gene expression (SAGE), and differential display (DD).

SSH identifies genes that are differentially expressed by removing common transcripts between a test and a control sample. The mRNA is converted to cDNA and then is trimmed with restriction enzymes and ligated to unique sequences. The test and control cDNA samples are mixed under conditions that allow for hybridization and PCR amplification. The control sample is labeled so it can be removed before cDNA from the test sample is efficiently amplified and identified (Bui et al., 2005; Rebrikov and Kogan, 2003).

CGH, also referred to as chromosomal microarray analysis, is a method to analyze copy number changes in a cell. CGH only detects unbalanced chromosomal changes, not balanced reciprocal translocations or inversions since the copy number remains the same for these gene alterations (Tan et al., 2007). Most applications of this technique to date have been to identify genetic abnormalities in humans (Aradhya and Cherry, 2007; Darilek et al., 2008; Gokcumen and Lee, 2009; Sher et al., 2009).

SAGE is a method for gene expression pattern analysis. SAGE does not require preexisting clones and thus can be used to identify and quantify unknown as well as known genes. Short sequence tags, 10–14 base pairs, are formed from long serial molecules, and the number of times a particular tag is observed is proportional the gene expression level (Velculescu et al., 1995).

DD systematically amplifies the 3′ terminal end of mRNA species using arbitrary primers paired with anchored poly-T primers. The DNA fragments are then run on DNA sequencing gels to provide direct side-by-side visualization of relative mRNA abundance (Liang et al., 1994). The mRNA sequence can be recovered from the DNA sequencing gel by PCR amplification and subsequent sequencing. Northern blots of the sequences can be compared against the samples to confirm differential expression (Appel et al., 1999; Chen and Harcum, 2007).

5.2. Comparative studies

To better understand the cellular and transcriptomic responses to environmental effects, recent studies have analyzed gene expression of mammalian cells under stressful bioreactor conditions. Early studies were limited to Northern blots to assess gene expression. For example, Kim and Lee (2002) examined the effects of hyperosmolarity on the gene expression of a humanized antibody, while Yoon et al. (2003) examined the effect of low temperature on productivity of CHO cells expressing an antibody. As better tools were developed, qRT-PCR was incorporated to examine cell responses to differing bioprocessing conditions. For example, CHO cells expressing tPA were examined for sialidase gene

expression levels, and it was observed that sialidase gene expression followed the intracellular sialidase activity (Clark et al., 2004). Later, Clark et al. (2005) compared the gene expression levels in CHO cells producing a recombinant IL-4/13 cytokine trap fusion protein cultured in normal glucose media to a media also containing galactose. It was observed that the degradation gene encoding sialidase was elevated, while the intracellular biosynthetic gene expression levels remained unchanged by the galactose addition.

Glycosylation has been studied by various research groups to understand the underlying genes involved and the effects of various environmental parameters. Examination of 12 glycosylation genes using qRT-PCR demonstrated that ammonium affected the genes responsible for the glycosylation reactions (Chen and Harcum, 2006). Specifically, this study observed that genes less sensitive to elevated ammonium were mainly involved in the early glycosylation reactions associated with the endoplasmic reticulum or cytosol. The genes most affected by elevated ammonium were mainly involved in the branching reactions, associated with the Golgi. Later, DD was used to identify genes affected by ammonium that were previously unassociated with ammonium stress and to identify several new CHO cell genes (Chen and Harcum, 2007). More recently, Wong et al. (2010a) examined the expression of 24 genes in CHO cells involved in glycosylation by qRT-PCR to better understand recombinant glycoprotein glycosylation processes in a bioreactor. Twenty-one of these genes had significant regulation in fed-batch cultures, and decreased sialylation of the product correlated with decreased gene expression of CMP-sialic acid synthesis genes and increased sialidase gene expression. These researchers further expanded their examination of glycosylation gene expression to 79 genes and applied this tool to nucleotide sugar precursor feeding effects on intracellular glycosylation activities in CHO cells. Although the most pronounced effect was an increase in intra-cellular nucleotide sugar pool levels, they also were able to correlate these feeding strategies to improved sialylation and increased β-1, 4-galactosyltransferase and β-galactoside α-2, 3-sialyltransferase gene expression levels (Wong et al., 2010b).

To understand the mechanisms behind recombinant protein expression, a mouse DNA microarray study compared non-producing (parental) and producing (monoclonal antibody) strains of NS0 cells (Khoo et al., 2007). These researchers were able to identify 104 genes with differential gene expression, and the functional annotation of these genes determined that the differentially expressed genes were involved in protein synthesis, lipid metabolism, and cell cycle regulation. However, none of the affected genes were glycosylation-related. To gain further insight into culture production decline, transcriptome and proteome studies were conducted on continuous NS0 cultures expressing a monoclonal antibody (Krampe et al., 2008). They determined that cell cycle arrest, apoptosis, and several stress-related genes were up-regulated and interferon regulatory factor 4 and protein folding related genes were down-regulated later in the cultures (Krampe et al., 2008). Swiderek and Al-Rubeai (2007) and Swiderek et al. (2008) investigated the effects of culture temperature and hypoxia on NS0 cells using DNA microarrays. The primary effect of reduced culture temperature was the down-regulation of genes controlling central metabolism, e.g. glycolysis, gluconeogenesis, pentose phosphate pathway, and inositol metabolism. The primary response of NS0 cells to hypoxia was up-regulation of glycolysis genes and the down-regulation of purine and pyrimidine synthesis genes (Swiderek et al., 2008). Further work examined the effect of Bcl-2 overexpression in NS0 cells expressing a monoclonal antibody cultured in chemostats at two dilution rates. Overall, the Bcl-2 NS0 cells had elevated transcription gene expression levels, highly elevated anti-apoptosis gene expression levels, and slightly elevated pro-apoptosis gene expression levels (Krampe et al., 2011).

Gene expression studies have also been conducted on both NS0 and hybridoma cell lines under hyperosmotic pressure. Wu et al. (2004) examined antibody production in NS0 cells

using a two-color mouse DNA microarray containing 7445 genes. They focused their analysis on the cell physiology and reported that only three carbohydrate genes were up-regulated and one carbohydrate gene was down-regulated. Shen and Sharfstein (2006) investigated gene expression changes in OKT3 mouse hybridoma cells due to hyperosmolarity using Affymetrix Gene Chip MOE430A, which contained over 14,000 genes and 4000 EST probe sets. They observed that genes associated with metabolism, cell cycle regulation, apoptosis, regulation of transcription and translation, transport, and signaling pathways were regulated, yet very few were stress-response genes. Taking this work further, the transcriptional responses to increased osmolarity in CHO cells were investigated using a proprietary CHO DNA microarray, and the overall transcriptomic responses of the CHO and NS0 cell lines were significantly different. However, several functional groups had common responses (Shen et al., 2010).

As in the work of Shen et al. (2010), investigators with access to a proprietary CHO DNA microarray have been able to expand understanding of the effects of environmental conditions on the transcriptome in CHO cells. An early study compared the transcriptome of CHO cells cultured in batch and fed-batch modes (Wong et al., 2006). They were able to determine that apoptosis gene regulation played a critical role in prolonged cultures. Furthermore, DNA microarray studies performed by Yee et al. (2009), have determined that genes associated with vesicle trafficking, endocytosis and cytoskeletal elements play a critical role in antibody production in CHO cells. Furthermore, Nissom et al. (2006) observed that the apoptotic pathways in CHO cells were signaled by death receptor- and mitochondria-mediated pathways instead of endoplasmic reticulum-mediated pathways. DNA microarray studies performed by Yee et al. (2009) determined that genes associated with vesicle trafficking, endocytosis and cytoskeletal elements play a critical role in antibody production in CHO cells.

The reported positive effect of sodium butyrate on productivity (Kumar et al., 2007) was investigated using DNA microarrays by Gatti et al. (2007) and determined to be due to changes in histone modifications, chaperones, lipid metabolism, and protein processing. Yee et al. (2008) coupled genomic and proteomic profiling to assess CHO and hybridoma cells under sodium butyrate stress. They observed that cell cycle and apoptosis genes were affected in both cell lines. Additionally, genes associated with protein processing and secretion, as well as redox activity genes were up-regulated (Yee et al., 2008). Furthermore, Kantardjieff et al. (2010) observed that combining sodium butyrate and low culture temperatures likely improved productivity of immunoglobulins due to increased expression of Golgi apparatus, cyto-skeleton, and signal transduction genes.

In addition to the species-specific DNA microarray experiments, cross-species DNA microarray hybridization has been studied in both CHO and mouse hybridoma cells (Gatti et al., 2007). Studies were conducted to corroborate physiology and gene expression observations under sodium butyrate treatments, and both in-house DNA microarrays contained glycosylation-related genes (Korke et al., 2004; Wlaschin et al., 2005). Interestingly, their gene expression analysis did not identify any glycosylation-related genes with significant expression changes due to sodium butyrate in the CHO and hybridoma cell lines (Gatti et al., 2007). This observation was surprising, as it has been observed that sodium butyrate-additions alter glycosylation (Chung et al., 2001; Etchevarry and Ryll, 2008; Sung and Lee, 2005). Finally, the CHO cells exhibited improved protein productivity at higher sodium butyrate levels than the hybridoma cells, which could not be explained by their gene expression data (Gatti et al., 2007).

DNA microarray analysis of bioprocessing conditions has become a valuable tool to troubleshoot production issues. For example, DNA microarrays were used to investigate the

cause of incomplete protein processing of recombinant human bone morphogenic protein 2 (rhBMP-2) expressed in CHO cells in large-scale bioprocessing settings (Doolan et al., 2008). A modifying enzyme was cloned into the rhBMP-2 CHO cell line, resulting in improved protein processing, and gene expression analysis of the original and enzyme containing clones identified the Unfolded Protein Response as differentially expressed. Additionally, several endoplasmic reticulum and Golgi protein encoding genes were identified as possibly playing a role. Taking the analysis of bioprocessing differences further, Doolan et al. (2010) coupled DNA microarrays with proteome analysis to compare fast and slow growing CHO cells, and the valosin-containing protein (VCP) was identified as an important factor in the fast growing phenotype. Overexpression and siRNA were both used to confirm the importance of VCP in the fast growing phenotype (Doolan et al., 2010). In a separate study, gene expression profiles from a data set of 70 stationary-phase CHO cells expressing monoclonal antibodies or Fc-fusions were examined and used to formulate a model to predict productivity (Clarke et al., 2011). The model identified 287 genes, many of which were associated with protein metabolism, and the genes were validated using independent data sets. Although these studies provide great insight into CHO cell bioprocessing, using a publicly sequenced CHO genome (such as the draft CHO-K1 genome) will allow for improved protein predictions, splice variant predictions, promoter/enhancer region studies, and better sequence design for RNA interference (RNAi).

5.3. Tools to alter the transcriptome

Once transcriptomic information has been identified to potentially improve a phenotype, there exist numerous tools to alter gene expression patterns, such as classical cDNA overexpression, chromosomal deletion techniques, and RNAi (Kim and Eberwine, 2010; Krämer et al., 2010; Vilgelm et al., 2006). Overexpression of genes, a method by which the target gene product is expressed in cultured mammalians cells, to improve productivity has a long history of success (Dietmair et al., 2011). There are many standardized methods for inserting genes into mammalian cells, which are encompassed by two categories, transient and stable (Kim and Eberwine, 2010). Transient clones are faster to generate relative to stable clones, but they have significantly lower yields and are therefore not used in commercial manufacturing (Baldi et al., 2007). Instead, transient clones are primarily used to rapidly investigate the effect of a particular gene on the target product's productivity. For example, transient overexpression of the X-box binding protein 1 in CHO cells expressing a monoclonal antibody increased productivity (Codamo et al., 2011). There are also examples of stable expressing clones, for example the stable overexpression of Mcl-1 in CHO cells expressing a monoclonal antibody increased cell viability (Majors et al., 2009), and the stable overexpression of Akt in CHO cells expressing a monoclonal antibody was observed to decrease apoptosis (Hwang and Lee, 2009).

RNAi is a naturally occurring system in cells used to control gene expression levels, and there are two major types of RNAi systems: miRNA and small interfering RNA (siRNA) (Huang et al., 2011; Li and Shen, 2009; Lwa et al., 2010). These small RNAs can bind to specific mRNA and decrease the translation of these sequences by either decreasing the level of the target mRNA or halting translation of the target mRNA (Barron et al., 2011; Koh et al., 2009). Other variations of RNAi include short hairpin RNA (shRNA), which have been used to modify glycosylation genes in CHO cells (Beuger et al., 2009; Zhang et al., 2010). With respect to bioprocessing, Johnson et al. (2011) identified 350 miRNA species in CHO cells from four cultures representing four different cell lines, two different growth temperatures, and one culture under sodium butyrate stress. Hackl et al. (2011) identified 387 miRNA species in CHO cells by examining six strains, and they validated 26 target cDNA sequences, which should provide a framework for RNAi tool development in CHO cells.

Zinc-finger nucleases (ZFNs) are one method that can be used to create a gene knockout in mammalian cells (Krämer et al., 2010). Santiago et al. (2008) first demonstrated the utility of ZFNs in CHO cells by recreating the *dhfr* gene deletions already commonly used in CHO cell cloning. The three clones they created grew as expected and were generated faster than homologous recombination methods. Cost et al. (2010) deleted both the *bak* and *bax* genes in CHO cells using ZFNs resulting in cells that grew normally; however, when subjected to apoptotic stimuli they could not produce caspases. Liu et al. (2010) used ZFNs to demonstrate that three independent gene deletions in CHO cells corresponding to glutamine synthetase, DHFR, and α-1, 6-fucosyltransferase, could be obtained in a single strain. Other methods of gene knockout exist; however, these other methods have not been widely examined under bioprocessing conditions.

6. Applied proteomics in bioprocessing

Complementary to the study of the genome to advance understanding of gene presence, expression, and control, is the analysis of the proteome to uncover global protein expression. More specifically, proteomics is the quantification of proteins within a cell at a specific time point under explicitly defined conditions. Proteins are implicated in nearly all cellular activities, so an examination of their presence, properties, and functions provides a more detailed representation of the behavior of a system compared to examining genes or mRNA. The proteome contains a larger number of proteins than the number of genes in the genome due to alternative splicing of mRNAs and a large range of PTMs. There is a non-linear correlation between mRNA and protein expression levels due to varying stabilities and life times of these molecules; mRNA has a half life on the order of minutes to hours, and protein has a half life on the order of minutes to days. Furthermore, the proteome itself varies dynamically with the conditions and environment of the system.

These factors contribute to the analytical challenges associated with proteomics in separating, detecting, quantifying, and identifying the wide range of protein expression levels ($1–10^6$) in a cell. Even though there is no single method that can accurately measure protein expression over this wide range or an amplification method available to aid in rare protein detection (like PCR in genomics), progress in developing proteomic methods has been made. Various methods, including gel-based and mass spectrometry (MS)-based techniques, differ in their respective resolution and application (for a detailed review of methods, see May et al., 2011). Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) is the most widely used separation technique, which allows for the simultaneous analysis of a large number of proteins separated by charge and mass. Although this technique allows for the separation of up to 10,000 protein species, it is limited in its ability to detect proteins of low-abundance, high hydrophobicity, low or high molecular weight, extreme isoelectric point, or proteins that co-migrate. The creation of two-dimensional difference in gel electrophoresis (2D-DIGE) improved the accuracy in quantifying different proteins by implementing internal standards for better protein spot identification and eliminated gel variability for small sample numbers. To overcome some limitations in gel-based techniques, multidimensional peptide separation coupled with isotope labeling and automated tandem MS analysis has been implemented, and several different approaches exist (Link, 2002).

Although data analysis of proteomic experiments can be laborious, the results can be very rewarding. Experiments can identify leads for understanding cell function or gene expression, and when used in connection with gene, mRNA, and metabolic profiling analysis, the system can be more fully understood on a global level. Differential proteomic analysis is widely used to analyze and detect distinct differences in protein expression between cultures exposed to different nutrients and environments. Effects from various

types and levels of nutrients in basal and feeding media, specific culture additives (e.g. zinc sulfate and sodium butyrate), and pH and temperature ranges have been studied (see Jenkins et al., 2009 for a more detailed review). A wide range of genes and proteins have been identified from a variety of functional classes, such as protein folding, secretion, and synthesis; cell architecture, metabolism, and proliferation; and signal transduction. With genetic sequences available, more detailed information can be obtained from these experiments with more confidence in the accuracy and meaning of their results.

7. Future opportunities with a known CHO genetic sequence

The field of mammalian cell culture engineering, including genomic and proteomic experimentation, is poised to employ CHO genetic sequences in DNA sequence-reliant techniques, and many researchers have been anticipating this advancement for over 10 years. Cellular engineering focuses on manipulating genes to create a product of interest, and with a known genome, experimental designs can be better defined and outcomes better analyzed. Metabolic engineering provides the framework for analyzing genome-wide differential gene expression, protein content, and intracellular chemical reactions. Although progress has been made to elucidate desired and undesired metabolic activities in mammalian cell cultures using metabolic flux analysis (see reference Quek et al., 2010 for a complete review), metabolic engineering has had limited utility in CHO cell culture because of the lack of genetic information. Similarly, systems biology and synthetic biology approaches to mammalian cellular engineering have been restricted. Utilizing a systems biology approach, which employs a quantitative analysis of biological systems via integration of whole genome scale data sets to create a quantitative phenotypic description, will be crucial to attain future achievements in biotechnology. Furthermore, the emerging field of synthetic biology will advance biotechnology progress through its aim to enable a systematic forward engineering of biology for improved and novel applications.

These approaches have already been applied to bacterial and yeast cultures because of the availability of their genetic sequences, and the knowledge gained from their results can provide a basis for CHO cell culture applications. For example, inverse metabolic engineering methods have been developed to understand and implement desirable phenotypes in cellular engineering (Bailey et al., 1996). Direct comparisons of wild-type strains to engineered mutants have been accomplished with readily available, high-quality reference genomes and the reduced cost of resequencing (Warner et al., 2009). Additionally, a variety of microorganisms have been subjected to detailed multi-omic analysis to identify proteins and mechanisms involved in specific phenotypes. Experiments of this type focus on revealing cellular and metabolic changes in response to genetic perturbations or environmental stress (for a detailed list of integrated-omic micro-organism studies see Han et al., 2011).

More complete cellular understanding acquired from holistically designed studies will benefit mammalian cell bioprocess development. In the near future, better performing clones and more efficient production processes could be developed with the advancement of genomic understanding. To develop superior clones, recombinant DNA insertion can be directed to stable chromosome sites, and resulting clones can be quickly and quantitatively compared via sequence and expression profiling. Researchers may be able to reliably identify differences between clones (using the correct reference genome), further improve desired phenotypes, and implement one clone for all phases of clinical development. Biological production processes have matured over the past 20 years, and cell culture efforts are focused on optimizing and streamlining current technologies. Only reliable manufacturing processes that produce batches of product with consistent quality are successful. With more information available in genomics, transcriptomics, and proteomics

of CHO cell lines, underlying phenomena that contribute to product quality characteristics can be controlled more tightly.

For companies that produce biosimilars or generics, it is possible that better-defined genetics and clone attributes may provide advantages for developing their products. Obtaining this knowledge will depend on the decisions of the innovators to patent genomic information and methods or keep them as trade secrets. During the time period of 2006–2010, 14 biosimilar products were approved in the European Union (EU) following the groundbreaking approval of Omnitrope (a single chain polypeptide growth hormone biosimilar product) in Australia in 2005 and then in the EU and subsequently the US in 2006 (Walsh, 2006, 2010). Over the same four years, only 25 new biological entities and 14 reformulation products were approved in the United States or European Union. Thus, the biosimilar market is large and will continue to grow as many first-generation products with good market value (e.g. Herceptin, Rituxan, Remicade, and Humira) are reaching the end of their patent protection (Walsh, 2010). Regulatory agencies may implement additional requirements for all approved products (new molecular entities, biosimilars, and generics) based on the new genetic information that will dictate clone design. Regulatory agencies may prefer well-defined genetic information, such as complete genome sequences on individual cell lines over time, just as they prefer well-defined media and consistent manufacturing processes. Generating genomic data (e.g. genome sequences and -omics data) for regulatory submission may require more validated analytical efforts and increase the cost of manufacturing. These uncertainties along with the anticipation of implementing genomic-based technologies dictate that the future of mammalian cell culture bioprocessing is bright and will continue to grow and develop over the coming years.

Acknowledgments

This work was funded by the National Science Foundation and the National Institute of Standards and Technology.

References

Appel M, Bellstedt DU, et al. Differential display of eukaryotic mRNA: meeting the demands of the new millennium. J Plant Physiol. 1999; 154:561–70.

Aradhya S, Cherry AM. Array-based comparative genomic hybridization: clinical contexts for targeted and whole-genome designs. Genet Med. 2007; 9(9):553–9. [PubMed: 17873642]

Bailey JE, Sburlati A, et al. Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes. Biotechnol Bioeng. 1996; 52(1):109–21. [PubMed: 18629857]

Baldi L, Hacker DL, et al. Recombinant protein production by large-scale transient gene expression in mammalian cells: state of the art and future perspectives. Biotechnol Lett. 2007; 29(5):677–84. [PubMed: 17235486]

Barnes LM, Bentley CM, et al. Stability of protein production from recombinant mammalian cells. Biotechnol Bioeng. 2003; 81(6):631–9. [PubMed: 12529877]

Barron N, Sanchez N, et al. MicroRNAs: tiny targets for engineering CHO cell phenotypes? Biotechnol Lett. 2011; 33(1):11–21. [PubMed: 20872159]

Beuger V, Kunkele KP, et al. Short-hairpin-RNA-mediated silencing of fucosyltransferase 8 in Chinese-hamster ovary cells for the production of antibodies with enhanced antibody immune effector function. Biotechnol Appl Biochem. 2009; 53:31–7. [PubMed: 19032154]

Bouck A, Vision T. The molecular ecologist's guide to expressed sequence tags. Mol Ecol. 2007; 16(5):907–24. [PubMed: 17305850]

Bui LC, Leandri RD, et al. SSH adequacy to preimplantation mammalian development: scarce specific transcripts cloning despite irregular normalisation. BioMed Cent Genomics. 2005; 6:155.

Canales RD, Luo YL, et al. Evaluation of DNA microarray results with quantitative gene expression platforms. Nat Biotechnol. 2006; 24(9):1115–22. [PubMed: 16964225]

Chen PF, Harcum SW. Effects of elevated ammonium on glycosylation gene expression in CHO cells. Metab Eng. 2006; 8(2):123–32. [PubMed: 16380282]

Chen PF, Harcum SW. Differential display identifies genes in Chinese hamster ovary cells sensitive to elevated ammonium. Appl Biochem Biotechnol. 2007; 141(2–3):349–59. [PubMed: 18025561]

Chung BS, Jeong YT, et al. Effect of sodium butyrate on glycosylation of recombinant erythropoietin. J Microbiol Biotechnol. 2001; 11(6):1087–92.

Clark KJ, Chaplin FW, et al. Temperature effects on product-quality-related enzymes in batch CHO cell cultures producing recombinant tPA. Biotechnol Prog. 2004; 20(6):1888–92. [PubMed: 15575729]

Clark KJ, Griffiths J, et al. Gene-expression profiles for five key glycosylation genes for galactose-fed CHO cells expressing recombinant IL-4/13 cytokine trap. Biotechnol Bioeng. 2005; 90(5):568–77. [PubMed: 15818560]

Clarke C, Doolan P, et al. Predicting cell-specific productivity from CHO gene expression. J Biotechnol. 2011; 151(2):159–65. [PubMed: 21115077]

Codamo J, Hou JJC, et al. Efficient mAb production in CHO cells incorporating PEI-mediated transfection, mild hypothermia and the co-expression of XBP-1. J Chem Technol Biotechnol. 2011; 86(7):923–34.

Cost GJ, Freyvert Y, et al. BAK and BAX deletion using zinc-finger nucleases yields apoptosis-resistant CHO cells. Biotechnol Bioeng. 2010; 105(2):330–40. [PubMed: 19777580]

Darilek S, Ward P, et al. Pre- and postnatal genetic testing by array-comparative genomic hybridization: genetic counseling perspectives. Genet Med. 2008; 10(1):13–8. [PubMed: 18197052]

Dietmair S, Nielsen LK, et al. Engineering a mammalian super producer. J Chem Technol and Biotechnol. 2011; 86:905–14.

Doolan P, Melville M, et al. Transcriptional profiling of gene expression changes in a PACE-transfected CHO DUKX cell line secreting high levels of rhBMP-2. Mol Biotechnol. 2008; 39(3): 187–99. [PubMed: 18240028]

Doolan P, Meleady P, et al. Microarray and proteomics expression profiling identifies several candidates, including the valosin-containing protein (VCP), involved in regulating high cellular growth rate in production CHO cell lines. Biotechnol Bioeng. 2010; 106(1):42–56. [PubMed: 20091739]

Etchevarry, T.; Ryll, T. Mammalian cell culture process. US Patent. 5,705,364. 2008 Jan 6.

Fu X, Fu N, et al. Estimating accuracy of RNA-Seq and microarrays with proteomics. BioMed Central Genomics. 2009; 10:161. [PubMed: 19371429]

Gatti MD, Wlaschin KF, et al. Comparative transcriptional analysis of mouse hybridoma and recombinant Chinese hamster ovary cells undergoing butyrate treatment. J Biosci Bioeng. 2007; 103(1):82–91. [PubMed: 17298905]

Gokcumen O, Lee C. Copy number variants (CNVs) in primate species using array-based comparative genomic hybridization. Methods. 2009; 49(1):18–25. [PubMed: 19545629]

Hacker DL, De Jesus M, et al. 25 years of recombinant proteins from reactor-grown cells — where do we go from here? Biotechnol Adv. 2009; 27:1023–7. [PubMed: 19463938]

Hackl M, Jakobi T, et al. Next-generation sequencing of the Chinese hamster ovary micro RNA transcriptome: identification, annotation and profiling of microRNAs as targets for cellular Engineering. J Biotechnol. 2011; 153(1–2):62–75. [PubMed: 21392545]

Hammond S, Swanberg JC, et al. Genomic sequencing and analysis of a Chinese hamster ovary cell line using Illumina sequencing technology. BioMed Central Genomics. 2011; 12:67. [PubMed: 21269493]

Han MJ, Lee JW, et al. Understanding and engineering of microbial cells based on proteomics and its conjunction with other omics studies. Proteomics. 2011; 11:721–43. [PubMed: 21229587]

Huang Y, Shen XJ, et al. Biological functions of microRNAs: a review. J Physiol Biochem. 2011; 67(1):129–39. [PubMed: 20981514]

Huse SM, Huber JA, et al. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol. 2007; 8(7):R143. [PubMed: 17659080]

Hwang SO, Lee GM. Effect of Akt overexpression on programmed cell death in antibody-producing Chinese hamster ovary cells. J Biotechnol. 2009; 139(1):89–94. [PubMed: 18983878]

Jaluria P, Konstantopoulos K, et al. A perspective on microarrays: current applications, pitfalls, and potential uses. Microb Cell Fact. 2007; 6:4. [PubMed: 17254338]

Jenkins N, Meleady P, et al. Strategies for analyzing and improving the expression and quality of recombinant proteins made in mammalian cells. Biotechnol Appl Biochem. 2009; 53:73–83. [PubMed: 19397493]

Johnson KC, Jacob NM, et al. Conserved microRNAs in Chinese hamster ovary cell lines. Biotechnol Bioeng. 2011; 108(2):475–80. [PubMed: 20830683]

Kantardjieff A, Jacob NM, et al. Transcriptome and proteome analysis of Chinese hamster ovary cells under low temperature and butyrate treatment. J Biotechnol. 2010; 145(2):143–59. [PubMed: 19770009]

Kelley B. Industrialization of mAb production technology — the bioprocessing industry at a crossroads. MAbs. 2009; 1(5):443–52. [PubMed: 20065641]

Khoo SH, Falciani F, et al. A genome-wide transcriptional analysis of producer and non-producer NS0 myeloma cell lines. Biotechnol Appl Biochem. 2007; 47:85–95. [PubMed: 17223793]

Kim TK, Eberwine JH. Mammalian cell transfection: the present and the future. Anal Bioanal Chem. 2010; 397(8):3173–8. [PubMed: 20549496]

Kim NS, Lee GM. Response of recombinant Chinese hamster ovary cells to hyperosmotic pressure: effect of Bcl-2 overexpression. J Biotechnol. 2002; 95(3):237–48. [PubMed: 12007864]

Koh TC, Lee YY, et al. Identification and expression analysis of miRNAs during batch culture of HEK-293 cells. J Biotechnol. 2009; 140(3–4):149–55. [PubMed: 19428708]

Korke R, Gatti MD, et al. Large scale gene expression profiling of metabolic shift of mammalian cells in culture. J Biotechnol. 2004; 107(1):1–17. [PubMed: 14687967]

Krämer O, Klausing S, et al. Methods in mammalian cell line engineering: from random mutagenesis to sequence-specific approaches. Appl Microbiol Biotechnol. 2010; 88(2):425–36. [PubMed: 20689950]

Krampe B, Swiderek H, et al. Transcriptome and proteome analysis of antibody-producing mouse myeloma NS0 cells cultivated at different cell densities in perfusion culture. Biotechnol Appl Biochem. 2008; 50:133–41. [PubMed: 18302537]

Krampe B, Fagan A, et al. Chemostat-based transcriptional analysis of growth rate change and Bcl-2 over-expression in NS0 cells. Biotechnol Bioeng. 2011; 108(7):1603–15. [PubMed: 21337334]

Kumar N, Gammell P, et al. Proliferation control strategies to improve productivity and survival during CHO based production culture — a summary of recent methods employed and the effects of proliferation control in product secreting CHO cell lines. Cytotechnology. 2007; 53(1–3):33–46. [PubMed: 19003188]

Kuystermans D, Krampe B, et al. Using cell engineering and omic tools for the improvement of cell culture processes. Cytotechnology. 2007; 53(1–3):3–22. [PubMed: 19003186]

La Merie Business Intelligence. [accessed August 23, 2011] Top 30 biologics. 2010. Report released March 3, 2011http://www.pipelinereview.com/free-downloads/TOP_30_-Biologics_2010_RDPN_Special_Mar_2011.pdf

Lee JS, Park HJ, et al. Protein reference mapping of dihydrofolate reductase-deficient CHO DG44 cell lines using 2-dimensional electrophoresis. Proteomics. 2010; 10:2292–302. [PubMed: 20391528]

Li LM, Shen Y. Overcoming obstacles to develop effective and safe siRNA therapeutics. Expert Opin Biol Ther. 2009; 9(5):609–19. [PubMed: 19392577]

Liang P, Zhu W, et al. Differential display using one-base anchored oligo-dT primers. Nucleic Acids Res. 1994; 22:5763–4. [PubMed: 7838734]

Link AJ. Multidimensional peptide separations in proteomics. Trends Biotechnol. 2002; 20(12):S8–S13. [PubMed: 12570153]

Liu PQ, Chan EM, et al. Generation of a triple-gene knockout mammalian cell line using engineered zinc-finger nucleases. Biotechnol Bioeng. 2010; 106(1):97–105. [PubMed: 20047187]

Lwa TR, Tan CH, et al. Identification of cellular genes critical to recombinant protein production using a Gaussia luciferase-based siRNA screening system. J Biotechnol. 2010; 146(4):160–8. [PubMed: 20188772]

Majors BS, Betenbaugh MJ, et al. Mcl-1 overexpression leads to higher viabilities and increased production of humanized monoclonal antibody in Chinese hamster ovary cells. Biotechnol Prog. 2009; 25(4):1161–8. [PubMed: 19551877]

Mane SP, Evans C, et al. Transcriptome sequencing of the Microarray Quality Control (MAQC) RNA reference samples using next generation sequencing. BioMed Central Genomics. 2009; 10:264. [PubMed: 19523228]

Mardis ER. New strategies and emerging technologies for massively parallel sequencing: applications in medical research. Genome Med. 2009; 1(40):40.1–4. [PubMed: 19435481]

Mardis ER. A decade's perspective on DNA sequencing technology. Nature. 2011; 470(7333):198–203. [PubMed: 21307932]

Margulies M, Egholm M, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005; 437(7057):376–80. [PubMed: 16056220]

Marioni JC, Mason CE, et al. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 2008; 18(9):1509–17. [PubMed: 18550803]

May C, Brosseron F, et al. Instruments and methods in proteomics. Methods Mol Biol. 2011; 696:3–28. [PubMed: 21063938]

Metzker ML. Sequencing technologies — the next generation. Nat Rev Genet. 2010; 11:31–46. [PubMed: 19997069]

Moore MJ, Dhingra A, et al. Rapid and accurate pyrosequencing of angiosperm plastid genomes. BioMed Cent Plant Biol. 2006; 6:17.

Mortazavi A, Williams BA, et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008; 5(7):621–8. [PubMed: 18516045]

Nagalakshmi U, Wang Z, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science. 2008; 320(5881):1344–9. [PubMed: 18451266]

Nissom PM, Sanny A, et al. Transcriptome and proteome profiling to understanding the biology of high productivity CHO cells. Mol Biotechnol. 2006; 34:125–40. [PubMed: 17172658]

Nolan T, Hands RE, et al. Quantification of mRNA using real-time RT-PCR. Nat Protoc. 2006; 1(3): 1559–82. [PubMed: 17406449]

Omasa T, Cao Y, et al. Bacterial artificial chromosome library for genome-wide analysis of Chinese hamster ovary cells. Biotechnol Bioeng. 2009; 104:986–94. [PubMed: 19575438]

Pilbrough W, Munro TP, et al. Intraclonal protein expression heterogeneity in recombinant CHO cells. PLoS One. 2009; 4(12):e8432. [PubMed: 20037651]

Quek LE, Dietmair S, et al. Metabolic flux analysis in mammalian cell culture. Metab Eng. 2010; 12:161–71. [PubMed: 19833223]

Rebrikov DV, Kogan YN. The use of suppression subtractive hybridization to identify strain-specific sequences in bacterial genomes. Russ J Genet. 2003; 39(10):1112–5.

Rudd S. Expressed sequence tags: alternative or complement to whole genome sequences? Trends Plant Sci. 2003; 8(7):321–9. [PubMed: 12878016]

Santiago Y, Chan E, et al. Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. Proc Natl Acad Sci U S A. 2008; 105(15):5809–14. [PubMed: 18359850]

Shen D, Sharfstein ST. Genome-wide analysis of the transcriptional response of murine hybridomas to osmotic shock. Biotechnol Bioeng. 2006; 93(1):132–45. [PubMed: 16196057]

Shen DA, Kiehl TR, et al. Transcriptomic responses to sodium chloride-induced osmotic stress: a study of industrial fed-batch CHO cell cultures. Biotechnol Prog. 2010; 26(4):1104–15. [PubMed: 20306541]

Shendure J, Hanlee J. Next-generation DNA Sequencing. Nat Biotechnol. 2008; 26:1135–43. [PubMed: 18846087]

Sher G, Keskintepe L, et al. Genetic analysis of human embryos by metaphase comparative genomic hybridization (mCGH) improves efficiency of IVF by increasing embryo implantation rate and

reducing multiple pregnancies and spontaneous miscarriages. Fertil Steril. 2009; 92(6):1886–94. [PubMed: 19135663]

Sung YH, Lee GM. Enhanced human thrombopoietin production by sodium butyrate addition to serum-free suspension culture of Bcl-2-overexpressing CHO cells. Bio-technol Prog. 2005; 21(1): 50–7.

Swiderek H, Al-Rubeai M. Functional genome-wide analysis of antibody producing NS0 cell line cultivated at different temperatures. Biotechnol Bioeng. 2007; 98(3):616–30. [PubMed: 17421047]

Swiderek H, Logan A, et al. Cellular and transcriptomic analysis of NS0 cell response during exposure to hypoxia. J Biotechnol. 2008; 134(1–2):103–11. [PubMed: 18291545]

Tan DSP, Lambros MBK, et al. Getting it right: designing microarray (and not 'micro-awry') comparative genomic hybridization studies for cancer research. Lab Invest. 2007; 87(8):737–54. [PubMed: 17558419]

Tjio JH, Puck TT. Genetics of somatic mammalian cells. II. Chromosomal constitution of cells in tissue culture. J Exp Med. 1958; 108(2):259–68. [PubMed: 13563760]

Velculescu VE, Zhang L, et al. Serial analysis of gene-expression. Science. 1995; 270(5235):484–7. [PubMed: 7570003]

Vilgelm AE, Chumakov SP, et al. RNA interference: biology and prospects of application in biomedicine and biotechnology. Mol Biol. 2006; 40(3):339–54.

Walsh G. Biopharmaceutical benchmarks 2006. Nat Biotechnol. 2006; 24:769–78. [PubMed: 16841057]

Walsh G. Biopharmaceutical benchmarks 2010. Nat Biotechnol. 2010; 28:917–24. [PubMed: 20829826]

Wang Z, Gerstein M, et al. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009; 10(1):57–63. [PubMed: 19015660]

Warner JR, Patnaik R, et al. Genomics enabled approaches in strain engineering. Curr Opin Microbiol. 2009; 12:223–30. [PubMed: 19467921]

Wilhelm BT, Landry JR. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. Methods. 2009; 48(3):249–57. [PubMed: 19336255]

Wlaschin KF, Hu WS. A scaffold for the Chinese hamster genome. Biotechnol Bioeng. 2007; 98(2): 429–39. [PubMed: 17390381]

Wlaschin KF, Nissom PM, et al. EST sequencing for gene discovery in Chinese hamster ovary cells. Biotechnol Bioeng. 2005; 91(5):592–606. [PubMed: 16003777]

Wold B, Myers RM. Sequence census methods for functional genomics. Nat Methods. 2008; 5(1):19–21. [PubMed: 18165803]

Wong DCF, Wong KTK, et al. Transcriptional profiling of apoptotic pathways in batch and fed-batch CHO cell cultures. Biotechnol Bioeng. 2006; 94(2):373–82. [PubMed: 16570314]

Wong DCF, Wong NSC, et al. Profiling of N-glycosylation gene expression in CHO cell fed-batch cultures. Biotechnol Bioeng. 2010a; 107(3):516–28. [PubMed: 20521304]

Wong NSC, Wati L, et al. An investigation of intracellular glycosylation activities in CHO cells: effects of nucleotide sugar precursor feeding. Biotechnol Bioeng. 2010b; 107(2):321–36. [PubMed: 20506284]

Wu M-H, Dimopoulos G, et al. The effect of hyperosmotic pressure on antibody production and gene expression in the GS-NS0 cell line. Biotechnol Appl Biochem. 2004; 40(1):41–6. [PubMed: 15270706]

Wurm FM, Hacker D. First CHO genome. Nat Biotechnol. 2011; 29(8):718–20. [PubMed: 21822249]

Xu X, Nagarajan H, et al. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nat Biotechnol. 2011; 29(8):735–41. [PubMed: 21804562]

Yee JC, Gatti MD, et al. Genomic and proteomic exploration of CHO and hybridoma cells under sodium butyrate treatment. Biotechnol Bioeng. 2008; 99(5):1186–204. [PubMed: 17929327]

Yee JC, Gerdtzen ZP, et al. Comparative transcriptome analysis to unveil genes affecting recombinant protein productivity in mammalian cells. Biotechnol Bioeng. 2009; 102(1):246–63. [PubMed: 18726962]

Yoon SK, Kim SH, et al. Effect of low culture temperature on specific productivity and transcription level of anti-4-1BB antibody in recombinant Chinese hamster ovary cells. Biotechnol Prog. 2003; 19(4):1383–6. [PubMed: 12892507]

Zhang M, Koskie K, et al. Enhancing glycoprotein sialylation by targeted gene silencing in mammalian cells. Biotechnol Bioeng. 2010; 105(6):1094–105. [PubMed: 20014139]

Zhang J, Chiodini R, et al. The impact of next-generation sequencing on genomics. J Genet Genomics. 2011; 38:95–109. [PubMed: 21477781]

Table 1

2010 sales of biologics (Adapted from La Merie 2011 with permission).

Biologic category	2010 sales ($ billions)
Cancer antibodies	21.98
Anti-TNF antibodies	20.95
Insulin and insulin analogs	15.50
Erythropoietins	9.25
Coagulation factors	6.51
Interferon beta	6.48
G-CSF	5.38
Anti-inflammatory antibodies	4.08
Ophthalmic antibody	3.11
Human growth hormone	3.01
Interferon alpha	2.72
Enzyme replacement	2.58
Follicle stimulating hormone	1.20
Antiviral antibody	1.04
Other antibodies	0.47
Other proteins	3.53

Table 2

Current next-generation sequencing platforms. Adapted from Zhang et al., 2011, with permission.

Platform	Amplification	Read length	Throughput	Sequencing chemistry
Roche/GS-FLX Titanium	Emulsion PCR	400–600 bp	500 Mbp/run	Pyrosequencing
Illumina/HiSeq 2000, HiScan	Bridge PCR (cluster PCR)	2×100 bp	200 Gpb/run	Reversible terminators
ABI/SOLiD 5500×1	Emulsion PCR	50–100 bp	>100 Gpb/run	Sequencing-by-ligation (octamers)
Polonator/G.007	Emulsion PCR	26 bp	8–10 Gbp/run	Sequencing-by-ligation (monomers)
Helicos/Heliscope	No	35 bp	21–37 Gbp/run	True single-molecule sequencing
Pacific BioSciences/RS	No	1000 bp	N/A	Single-molecule real time