ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

Mining Transcriptome Data for Function–Trait Relationship of Hyper Productivity of Recombinant Antibody

Salim Charaniya,[1] George Karypis,[2] Wei-Shou Hu[1]

[1]Department of Chemical Engineering and Materials Science, University of Minnesota, 421 Washington Avenue SE, Minneapolis, Minnesota 55455-0132; telephone: 612-625-0546; fax: 612-626-7246; e-mail: acre@cems.umn.edu

[2]Department of Computer Science and Engineering, University of Minnesota, Minneapolis, Minnesota

Received 3 March 2008; revision received 13 October 2008; accepted 4 November 2008

Published online 18 November 2008 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bit.22210

ABSTRACT: In the past decade we have witnessed a drastic increase in the productivity of mammalian cell culture-based processes. High-producing cell lines that synthesize and secrete these therapeutics have contributed largely to the advances in process development. To elucidate the productivity trait in the context of physiological functions, the transcriptomes of several NS0 cell lines with a wide range of antibody productivity were compared. Gene set testing (GST) analysis was used to identify pathways and biological functions that are altered in high producers. Three complementary tools for GST—gene set enrichment analysis (GSEA), gene set analysis (GSA), and MAPPFinder, were used to identify groups of functionally coherent genes that are up- or downregulated in high producers. Major functional classes identified include those involved in protein processing and transport, such as protein modification, vesicle trafficking, and protein turnover. A significant proportion of genes involved in mitochondrial ribosomal function, cell cycle regulation, cytoskeleton-related elements are also differentially altered in high producers. The observed correlation of these functional classes with productivity suggests that simultaneous modulation of several physiological functions is a potential route to high productivity.

Biotechnol. Bioeng. 2009;102: 1654–1669.

© 2008 Wiley Periodicals, Inc.

KEYWORDS: recombinant antibody production; transcriptome; data mining; pathway analysis; cell culture

Introduction

Many antibodies, including over a dozen that have been successfully introduced to clinical applications in the past decade are important therapeutics for cancer, arthritis and other diseases. Many of those antibodies are produced in quantities exceeding thousands of kilograms annually; the increased demand has prompted increasing efforts in cell line and process development. Advances have led to a drastic increase in productivity in the past few years. By and large, the pursuit for higher productivity has been conducted by systematic yet empirical screening. The desire to better understand the complex trait of hyper productivity has led to a number of studies comparing the transcriptome (Khoo et al., 2007; Seth et al., 2007b) or proteome (Alete et al., 2005; Dinnis et al., 2006; Seth et al., 2007b; Smales et al., 2004) of cell lines with varying productivities. With their large-scale surveying power, transcriptome and proteome analyses certainly hold promise for discerning the genotypic characteristics of hyper productivity (for review, see Gupta and Lee, 2007; Seth et al., 2007a).

The transcriptomic investigations on the productivity of recombinant mammalian cells have largely employed statistical criteria as the primary means of identifying genes that are differentially expressed between two cell lines with different productivities (Khoo et al., 2007), or two classes of cell lines with different range of productivities (Seth et al., 2007b). Unlike many other biological traits, such as disease states of cancer development or the differentiation of stem cells to various lineages, high productivity is not a well-characterized trait. The final product titer, which is typically used to characterize the productivity of a cell line, results from a composite effect of cell concentration, balance between growth and death rate, as well as the protein secretory capability. Many alternative routes probably exist for achieving the same productivity (Seth et al., 2007b). It is likely that even among high-producing cell lines multiple "physiological classes" exist and their transcriptomes may also reflect such diversity.

In a previous study, the transcriptome profiles of seven high and four low recombinant IgG-producing NS0 cell

Correspondence to: W.-S. Hu

© 2008 Wiley Periodicals, Inc.

lines were analyzed to identify genes which are significantly different between the two groups (Seth et al., 2007b). As in other studies, the high and low producers were classified heuristically based on final titer of IgG in culture. It was not clear whether all the producers share common features in their transcriptome. Nevertheless, through differential expression analysis, it was highlighted that several genes related to protein synthesis and cell cycle were differentially expressed between high and low producers.

In this study we undertook a different but complementary approach of transcriptome-based investigation of high productivity trait. We surveyed pathway analysis tools in an effort to identify the biological functions, which are significantly modulated in cell lines with high productivity. Biological interpretation of genes that are differentially expressed between two or more phenotypes is often facilitated by grouping them into few functional classes. However, the list of differentially expressed genes depends on the stringency of the statistical threshold used. The difficulty in identifying genes which are truly differentially expressed is further compounded by the observation that, for mammalian cells in culture, changes in productivity levels are accompanied by only modest alterations in the expression levels of individual genes (De Leon Gatti et al., 2007; Seth et al., 2007b; Smales et al., 2004; Yee et al., 2008). This is in contrast to gene expression changes seen in microbial populations or in stem cells in early stages of differentiation.

We have employed gene set testing (GST) tools to assess gene expression alteration at functional class level rather than at individual gene level (for a review on GST, see Goeman and Buhlmann, 2007). In this approach, functionally related genes are combined a priori into gene sets and transcriptome data are evaluated in terms of these gene sets instead of individual genes. The correlation of a gene set to a phenotype is evaluated by comparing the observed number of genes in a gene set, which have altered expression level with the expected number under a null hypothesis. Assessment of multiple GST tools revealed that a large percentage of genes in functions related to protein processing and secretion, such as Golgi apparatus, the cytoskeletal network, and protein degradation were altered between high and low producers. Differential transcript changes were also observed in cell cycle-related genes. The relatively modest changes at individual gene level between the two groups suggest that the expression changes are not localized but a broad range of functional modulation is likely to accompany the process of high producer selection during cell line development.

Materials and Methods

Cells and Sample Preparation

The 11 GS-NS0 cell lines and their cultural conditions and have been described previously (Seth et al., 2007b). The average productivity of the seven high producers is approximately five times the average productivity of the four low producers. In addition, biological replicates of four high producers (H1–H4) and two low producers (L1–L2) were performed under the same culture and sample preparations conditions, as reported previously.

Microarray Hybridization

GeneChip® Mouse genome 430A 2.0 (MOE430A 2.0) (Affymetrix, Santa Clara, CA) was used for assaying the transcriptome of the six biological replicate cultures. MOE430A 2.0 contains 22,690 probes representing approximately 14,000 well-characterized mouse genes. Biotinylated cRNA was prepared as per the protocol described in the Affymetrix Technical Manual. Fifteen micrograms of biotinylated cRNA was used for hybridization. The arrays were scanned at University of Minnesota Biomedical Image Processing Laboratory.

Microarray Data Processing

The raw intensity data from each array was normalized using Affymetrix Microarray Suite (MAS) version 5.0, which includes background correction, perfect match (PM) adjustment, and calculation of expression summary from 11 probe pairs using one-step Tukey's Biweight method for estimation of robust mean. The probe intensities from each array were scaled to an average of 500. Further, a quantile normalization procedure was employed at probe level to ensure that probe intensities from different arrays have the same distribution. Using a one-sided Wilcoxon signed rank test, the MAS 5.0 algorithm also determines a "detection" P-value for every probe. A P-value <0.04 was used as the criterion to call a transcript "present." Transcripts with absolute intensity, averaged across all the samples, less than 60 were discarded before further analysis.

Differential Expression Analysis

Significance analysis of microarrays (SAM) version 3.0 was used to identify genes that are differentially expressed between the high producers and low producers (Tusher et al., 2001). SAM combines a d-statistic with repeated sample permutations to determine the percentage of genes that are identified as differentially expressed by chance, that is, false discovery rate (FDR). A threshold of 10% FDR was used in this study. SAM outputs a q-value for every probe, which is an estimate of the FDR incurred when that probe, and all the probes with a lower q-value are called significantly differentially expressed. In this study, all the probes with q-value $\leq 10\%$ were considered as differentially expressed.

Functional Analysis

GST was performed on 242 gene sets to identify those that correlate with the phenotype distinction between the high producer and the low producer groups. A set of genes involved in the same biological function is defined as a gene set. Three different GST tools—MAPPFinder (Doniger et al., 2003), gene set enrichment analysis (GSEA) (Mootha et al., 2003; Subramanian et al., 2005), and gene set analysis (GSA) (Efron and Tibshirani, 2007), were used. For MAPPFinder, which is built into the software package Gene Map Annotator and Pathway Profiler (GenMAPP) (Dahlquist et al., 2002), the criteria of q-value $\leq$10% and two different fold change thresholds (1.2 and 1.4) were used to identify differentially expressed genes as input. For each method, the null distribution was estimated by 1,000 permutations. The enrichment of every gene set is characterized by a P-value. In this study, gene sets with a P-value $\leq$0.06 in at least two of the three GST methods were identified as significantly enriched. GSEA was used as a module in GenePattern (Reich et al., 2006) and GSA was available in SAM version 3.0 (Tusher et al., 2001).

Results

Transcriptome Analysis of Antibody-Producing NS0 Cells

In a recent study we compared the transcriptome and proteome profiles of 11 GS-NS0 cell lines with different productivities, broadly classified into two groups as 7 high and 4 low producers (Seth et al., 2007b). Several differentially expressed transcripts were involved in protein synthesis and cell cycle-related pathways suggesting a direct or indirect correlation between those biological functions and high productivity. Here, we extend the analysis by functional investigation measures to systematically explore the physiological traits that impart high productivity.

The seven high producers have IgG productivity ranging between 2 and 11 times the average productivity of the low producers, suggesting that productivity is a continuous distribution rather than a discrete binary class. The average productivity of four high producers (H1–H4) is 5.3 times the average productivity of the four low producers and is also 50% higher than the average of the other three high producers (H5–H7). Additionally, as described earlier, the composite nature of high productivity is likely to result in substantial diversity in the expression signatures that may be associated with high productivity. Upon examining the expression profiles of the seven high producers, it was observed that the four high producers (H1–H4) have a substantially high mutual similarity in their gene expression levels compared to the other three high producers (H5–H7) as well as the low producers. The observation was also strengthened by a supervised classification method (see Discussion Section) used to group the producers into "High" or "Low" class based on their transcriptome profiles. Analysis based on this small set of producers suggested that four high producers (H1–H4) can be distinguished from the other three high producers (H5–H7) and the low producers. These observations reflect the differences in the transcriptome fingerprint of the various high producers.

To discern possible physiological functions that confer the high productivity trait, we set out to identify functionally related genes that are altered between the two groups with different productivities. The four high producers (H1–H4), which are distinguishable from the low producers with respect to IgG productivity and transcriptome profiles, were used for subsequent functional analysis. The three producers H5–H7 were not included in further analysis.

Further, biological replicates of six of the remaining producers (H1–H4 and L1–L2) were performed. M-A plots of various binary combinations of samples (arrays) were examined to ensure consistent normalization from replicates. No intensity-dependent bias in fold change was observed.

Functional Analysis

Pathway analysis was performed using three different GST tools—MAPPFinder, GSEA, and GSA. A total of 242 MicroArray Pathway Profiles (MAPPs) for the mouse genome were obtained from GenMAPP (http://www.genmapp.org/HTML_MAPPs/Mouse/MAPPIndex_Mm_Contributed.htm). Each MAPP, a collection of genes involved in the same biological function or pathway, was considered as a gene set. The functional classes are categorized according to the organizing principles of Gene Ontology (GO)—cellular component, biological process, and molecular function.

Each of the three tools uses a different methodology to test the null hypothesis for every gene set. GSEA and GSA rank all the transcripts in the dataset using a class-correlation metric. Signal-to-noise ratio was used for GSEA, whereas GSA uses t-statistic as the class-correlation metric. Thus, for example, the genes which are upregulated in high producers are ranked high, whereas those that are downregulated in the high producers are ranked at the bottom of the rank-ordered list. The ranking scheme thus does not explicitly require a threshold for differential expression. Genes in a particular gene set or functional class are mapped on this ranked list. GSEA uses a Kolmogorov-Smirnov-like statistic for every gene set to test the enrichment of genes at the extremes of the ranked list (Subramanian et al., 2005), whereas GSA employs a maxmean statistic (Efron and Tibshirani, 2007). In contrast, MAPPFinder uses a list of differentially expressed genes as input. The overrepresentation of differentially expressed transcripts in a gene set is tested using the hypergeometric distribution. A z-score computes the difference between the fraction of genes differentially up- or downregulated in a gene set and the overall fraction expected in the population (Doniger et al., 2003). A gene set with a

higher-than-expected proportion of differentially expressed genes has a high z-score and hence a higher likelihood of a correlation to productivity phenotype. The results of MAPPFinder are, however, dependent on the user-defined threshold for differential expression. To that end, we employed SAM (Tusher et al., 2001) to identify differentially expressed genes. The transcriptome analysis on cultured mammalian cell in the past few years has generally demonstrated that the degree of differential expression observed is not of very large magnitude. Whether the hyper productivity trait is the manifestation of vast number of genes, each altering at a relatively minute level, or large but localized expression changes in a small number of genes, is still an open question. We thus employed two different criteria for differential expression of individual transcripts: (1) q-value $\leq$10% (10% FDR) and at least a 1.2-fold change (2) q-value $\leq$10% and a fold change of at least 1.4.

A distinction between GSEA, GSA and MAPPFinder is thus the reliance of MAPPFinder on a user-defined differential expression criterion. A potential drawback of this user-dependence is that the fraction of genes in a gene set that are called differentially expressed can change depending on the criterion, which in turn can affect the results of such a "discrete" method (Ben-Shaul et al., 2005). Furthermore, change in the activity of a biological pathway can be effected by a modest change in a large number of genes involved in the pathway, many of which may not satisfy a stringent differential expression criterion. Modest changes can be identified more readily by quantifying the *shift* in distribution of a differential expression metric for a set of functionally related genes, compared to the overall distribution for all genes (Ben-Shaul et al., 2005; Mootha et al., 2003). GSEA proposes such a "continuous" methodology whereby all the genes in a dataset are ranked according to a class-correlation metric, and enrichment of a gene set is based on the non-random positioning of its members in the ranked list (Mootha et al., 2003; Subramanian et al., 2005). GSA method proposed potential improvements to GSEA algorithm. GSA uses a different enrichment statistic and a modified procedure for estimating the null distribution (Efron and Tibshirani, 2007).

The significance of each functional class or a gene set is characterized by an enrichment P-value. A low P-value for a gene set suggests that a significant fraction of transcripts in that set have altered expression levels between the two phenotypic groups. Since the three GST tools use different methodologies, the gene sets identified can also differ. The gene sets identified as significant (P-values $\leq$0.06) by each method were cross-compared. Eight sets which were identified in at least two of the three methods are listed in Table I. A positively enriched gene set is one in which many genes in that functional class are upregulated in the high producers. Similarly, in a negatively enriched gene set, genes downregulated in high producers are overrepresented.

The biological process of cell cycle (GO:0007049) was the only functional class identified as significantly enriched by GSEA and MAPPFinder (P-value $\leq$0.06) and marginally enriched by GSA (P-value = 0.062) (Table I). Among the genes involved in cell cycle progression that are represented on the MOE430A array, 28% were differentially expressed, which includes 24 upregulated and 9 downregulated. Other functional classes that constitute different molecular functions such as isomerase activity (GO:0016853), structural constituent of ribosome (GO:0003735), GTPase regulator activity (GO:0030695), and ligase activity (GO:0016874) were also correlated to the phenotypic difference between the high and low producer groups. Golgi apparatus (GO:0005794), cytoskeleton (GO:0005856), and chromatin (GO:0000785) are the ontological classes under cellular component that were identified as altered between the high and low producer groups. In the ensuing sections, several of these functional classes are elaborated with emphasis on differentially expressed genes in each class.

Table I. Functional gene sets identified by different gene set testing methods.

Gene set (functional class)	No. genes in gene set	% present on MOE430A	% of genes D.E.[a]	Enrichment P-value for GSEA	GSA	MAPPFinder (D.E. criteria) 1[a]	MAPPFinder (D.E. criteria) 2[b]	Direction of alteration
Cellular component								
Golgi apparatus	336	69	19	0.150	0.053	0.035	0.008	Up
Cytoskeleton	189	68	20	0.027	0.042	0.231	0.119	Down
Chromatin	118	59	26	0.002	0.034	0.162	0.005	Down
Biological process								
Cell cycle	166	72	28	0.039	0.062	0.037	0.027	Up
Molecular function								
Isomerase activity	126	73	19	0.016	0.113	0.016	0.285	Up
Structural constituent of ribosome	164	77	19	0.069	0.059	0.024	0.331	Up
GTPase regulatory activity	174	66	20	0.045	0.106	0.014	0.028	Down
Ligase activity	160	67	27	0.078	0.057	0.021	0.173	Down

[a]D.E. criteria 1: q-value $\leq$10% and fold change $\geq$1.2.
[b]D.E. criteria 1: q-value $\leq$10% and fold change $\geq$1.4.

Genes Enriched in High and Low Producer Classes

GSEA identifies a subset of genes (called leading-edge subset) in each gene set that are key contributors to the enrichment of the functional class. These genes reside at the top or bottom region of the rank-ordered list. A significant proportion of these leading edge genes also meet the criteria used to identify differentially expressed genes by SAM. In six of the eight gene sets, at least 60% of the genes in the leading-edge subset also satisfy the differential expression criteria (q-value $\leq$10% and fold change $\geq$1.2). For every gene set, the percentage of gene probes that are represented on the microarray is also shown in Table I. At least 19% of genes present on the array are differentially expressed in each functional class.

These enriched genes and their extent of differential expression for different functional classes are listed in Tables II–VII. For each gene, the average hybridization intensity from the four high producers is also listed. To gauge the range of signal intensities from every array, the 25th, 50th, and 75th percentiles correspond to signal intensities of 49, 164, and 595, respectively. Following is a discussion on the possible role of these functional classes in conferring the hyper productivity trait. The focus is on the gene class rather than individual genes. A brief annotation on the functions of key genes is included in these tables.

Golgi Apparatus

Golgi apparatus (GO:0005794) was identified as significantly upregulated in two of the three GST tools. Two hundred thirty-three genes present on MOE430A array are annotated as belonging to Golgi apparatus gene set according to the GenMAPP database. Genes in this functional class encode proteins that are involved in post-translational modification and protein trafficking and secretion. Among the 37 transcripts differentially expressed in this gene set by at least 1.4-fold (Table II), as many as 27 are upregulated. All the 27 upregulated genes are in the leading-edge subset identified by GSEA further suggesting that upregulation of these transcripts in high producers plays a role in enhancing productivity. Some of these differentially expressed genes localized in Golgi can be clustered into categories of similar functions, notably vesicle transport and glycosylation. These categories are described below.

Vesicle transport. Ten genes, nine of which are upregulated, are related to vesicle transport. Protein processing in Golgi involves transport of cargos via membrane vesicles from one Golgi compartment to another as well as from Golgi to other cellular destinations. The correct delivery of membrane vesicles to their receiving targets is mediated by two complementary sets of transmembrane proteins: vesicle SNARE (v-SNARE) proteins and target membrane-specific SNARE (t-SNARE) proteins. v-SNARE on the membrane vesicle and t-SNARE on the target membrane interact to form a trans-SNARE complex that facilitates fusion of the vesicle to the target membrane. At least 30 different SNARE proteins are present in mammalian cells. Two of the nine upregulated transcripts encode the SNARE proteins, Gosr1 and Gosr2. The assembly of COPI-coated vesicles is initiated by activation of a small G protein, Arf1 on the Golgi membrane, followed by Arf1-mediated recruitment of the preassembled heptameric COPI coat complex. GTPase-activating protein Arfgap1 stimulates GTP hydrolysis of Arf1 thereby contributing to COPI vesicle budding (Bigay et al., 2003). The transcript of Arfgap1 is upregulated by 1.5-fold in high producers. Interestingly, although the Arf1 transcript, with q-value of 25%, did not meet our statistical criterion, its expression level in each of the four high producers is $\sim$1.2-fold higher than the average of low producers. Notably, the average signal intensity of Arf1 is 5629, which corresponds to the 98th percentile of intensities on every microarray. This suggests that Arf1 is one of the most highly abundant transcripts in the cell, and a 20% upregulation, if true, may alter cellular capacity for vesicle transport.

Protein glycosylation. Eight differentially expressed transcripts listed in Table II encode glycosyltransferase enzymes. These membrane-bound enzymes reside mainly in different Golgi compartments (*cis*, *medial*, *trans*) and catalyze the transfer of various sugar moieties to the newly formed protein transported from endoplasmic reticulum (ER). The signal intensity for Man2a1 transcript is 2340 (93rd percentile) suggesting that the transcript is abundant and its upregulation by 70% in high producers is quite significant. However, not all the glycosyltransferase enzymes probed were upregulated. Genes encoding four glycosylation enzymes are downregulated by twofold or greater in high producers (Table II). There appears to be a change in the ratio of α-2,6-, α-2,8-, and α-2,3-sialyltransferase, and an enhanced mannosidase transcript level in high-producing cells. Alpha-mannosidase I and II are responsible for the trimming of high mannose glycan on glycoproteins before glycan extension. These are the first steps of glycan processing in Golgi, whereas sialations are the final steps of glycan synthesis for glycoproteins. It is also interesting to note that in high-producing cells the transcript of α-1,6-fucosyltransferase is twofold lower. The product of fucosyltransferase, fucosylated *N*-glycan, has been reported to confer immunoglobulin G with lower antibody-dependent cellular cytotoxcity (ADCC) compared to the unfucosylated product (Shields et al., 2002). It will be interesting to examine whether the glycoform of the high antibody producing cells indeed has altered glycans.

Cytoskeleton

Twenty percent of the genes in this functional category are differentially expressed. Members of the cytoskeleton class encode proteins that associate with one or more filamentous elements that form the cellular scaffold essential for maintaining cell shape, exerting and distributing mechanical force and various other functions such as exocytosis, endocytosis, mitosis, and cell motility. They also play an

Table II. List of differentially expressed genes in the functional class "Golgi apparatus."

Gene symbol	Gene title	F.C.[a]	*q*-val. (%)[b]	Avg. int.[c]	Core subset[d]	Annotation
Vesicle transport						
Cog4	Component of oligomeric Golgi complex 4	(+) 2.3	0.48	311.7	Yes	A component of octameric COG complex, that plays a crucial role in tethering transport vesicles from late Golgi or early endosomes to *cis*-Golgi (Ungar et al., 2002)
Ap1g1	Adaptor protein complex AP-1, gamma 1 subunit	(+) 1.9	0.34	666.9	Yes	γ-subunit of adaptor-related protein complex 1, which is important for formation of clathrin-coated pit on vesicles of *trans*-Golgi network (TGN)
Ap3s2	Adaptor-related protein complex 3, sigma 2 subunit	(+) 1.7	1.68	721.9	Yes	σ_2 subunit of clathrin-related adaptor complex 3 (AP3). AP3 complex likely to be associated with TGN and peripheral endosome-like structures involved in protein sorting (Dell'Angelica et al., 1997)
Gosr1	Golgi SNAP receptor complex member 1	(+) 1.6	3.96	280.5	Yes	A v-SNARE that is a key component of Golgi 20S SNARE complex involved in vesicle transport from ER to *cis*- and *medial*-Golgi (Subramaniam et al., 1996)
Copb2	Coatomer protein complex, subunit beta 2 (beta prime)	(+) 1.5	6.62	1149.5	Yes	A subunit of the coatomer protein complex that forms the coat of non-clathrin coated vesicles
Arfgap1	ADP-ribosylation factor GTPase activating protein 1	(+) 1.5	1.68	431.8	Yes	GTPase activating protein, which is involved in COPI vesicle-mediated protein transport between Golgi cisternae and retrograde transport from *cis*-Golgi to ER
Cog8	Component of oligomeric Golgi complex 8	(+) 1.4	6.62	707.9	Yes	Component of COG complex that plays a crucial role in determining the protein transport capacity of Golgi
Gosr2	Golgi SNAP receptor complex member 2	(+) 1.4	3.10	396.0	Yes	A Golgi SNARE (GS27), which participates in protein transport from *medial*-Golgi to *trans*-Golgi and TGN (Lowe et al., 1997)
Arfrp1	ADP-ribosylation factor related protein 1	(+) 1.4	1.68	243.4	Yes	
Napg	*N*-ethylmaleimide sensitive fusion protein attachment protein gamma	(−) 1.4	6.62	460.7	No	γ-isoform of *N*-ethylmaleimide-sensitive factor (NSF) attachment protein. Involved in disassembly of T-SNARE/V-SNARE/SNAP25 complex after vesicle fusion to facilitate their recycling (Sollner et al., 1993)
Protein glycosylation						
St3gal2	ST3 beta-galactoside alpha-2,3-sialyltransferase 2	(+) 2.0	0.00	313.1	Yes	Enzyme that can transfer sialylic acid residue to glycoproteins that have Galβ1,3GalNAc as the terminal disaccharide (Lee et al., 1993)
Man2a1	Mannosidase 2, alpha 1	(+) 1.7	5.47	2339.5	Yes	α-mannosidase II enzyme found primarily in *medial* Golgi. Catalyzes the cleavage of α(1,3) and α(1,6)-mannose residues from the high mannose glycan resulting in the formation of a core glycan structure—$glcNAc_2Man_3$, that is common to all *N*-glycans
Man1a2	Mannosidase, alpha, class 1A, member 2	(+) 1.5	6.62	686.7	Yes	α-mannosidase IA, which cleaves α(1,2)-mannose residues from the high mannose oligosaccharide
Man1a	Mannosidase 1, alpha	(+) 1.5	3.09	696.3	Yes	
St6gal1	Beta galactoside alpha 2,6 sialyltransferase 1	(−) 2.4	5.47	204.2	No	
St8sia4	ST8 alpha-*N*-acetyl-neuraminide alpha-2,8-sialyltransferase 4	(−) 2.3	1.68	324.4	No	
Fut8	Fucosyltransferase 8	(−) 2.1	2.15	1198.5	No	Enzyme that can transfer fucose from GDP-fucose to first galactose of *N*-glycan through α-1,6-linkage
Galnt11	UDP-*N*-acetyl-alpha-D-galactosamine: polypeptide *N*-acetylgalactosaminyltransferase 11	(−) 2.0	0.00	227.5	No	
Others						
Sgpp1	Sphingosine-1-phosphate phosphatase 1	(+) 1.9	5.47	148.5	Yes	
Slc35a5	Solute carrier family 35, member A5	(+) 1.8	3.09	583.4	Yes	
Emid1	EMI domain containing 1	(+) 1.7	0.48	1147.8	Yes	
Nsg2	Neuron specific gene family member 2	(+) 1.7	8.25	1178.8	Yes	

Table II. (*Continued*)

Gene symbol	Gene title	F.C.[a]	*q*-val. (%)[b]	Avg. int.[c]	Core subset[d]	Annotation
Mbtps1	Membrane-bound transcription factor peptidase, site 1	(+) 1.7	0.48	1174.9	Yes	Involved in proteolytic activation of activating transcription factor 6 (Atf6)—a key transducer or ER stress. Mbtps1 also important for achieving cholesterol homeostasis by activation of two sterol regulatory element binding proteins (SREPB1 and SREPB2) (DeBose-Boyd et al., 1999)
Chst12	Carbohydrate sulfotransferase 12	(+) 1.6	5.47	336.7	Yes	
Tmed2	Transmembrane emp24 domain trafficking protein 2	(+) 1.6	0.71	1694.2	Yes	
Adam10	A disintegrin and metallopeptidase domain 10	(+) 1.5	2.15	1098.7	Yes	
Golph3	Golgi phosphoprotein 3	(+) 1.5	5.47	1935.5	Yes	
Gopc	Golgi associated PDZ and coiled-coil motif containing	(+) 1.5	6.62	238.0	Yes	
Bicd2	Bicaudal D homolog 2	(+) 1.4	1.68	277.1	Yes	
Golga4	Golgi autoantigen, golgin subfamily a, 4	(+) 1.4	6.62	291.0	Yes	
Chrnb1	Cholinergic receptor, nicotinic, beta polypeptide 1	(+) 1.4	6.62	200.0	Yes	
Gabarapl2	Gamma-aminobutyric acid (GABA-A) receptor-associated protein-like 2	(+) 1.4	2.15	2037.4	Yes	
Aph1a	Anterior pharynx defective 1a homolog (*C. elegans*)	(−) 1.7	0.90	363.8	No	
Lman2	Lectin, mannose-binding 2	(−) 1.6	1.24	798.3	No	
Psen2	Presenilin 2 (PS2)	(−) 1.6	6.62	467.5	No	
Clcn3	Chloride channel 3	(−) 1.5	8.25	1251.2	No	
Rnpep	Arginyl aminopeptidase (aminopeptidase B)	(−) 1.4	3.10	387.3	No	

[a]Fold change (F.C.), (+) upregulated in high producer (H), (−) downregulated in high producer.
[b]*q*-value calculated for each probe using SAM.
[c]Average intensity of high producers on Affymetrix array (MOE430A).
[d]Core member of the functional gene set identified as the leading-edge subset in GSEA.

essential role in intracellular protein transport, especially vesicle transport (reviewed in Ross et al., 2008; Stamnes, 2002). Among the 23 genes that are differentially expressed in this functional class by at least 1.4-fold, 8 genes that are downregulated (Table III). The most significant is Hook homolog 2 (Hook2), which was not detected (detection *P*-value >0.04) in any high producer in contrast to an average intensity of 646 in low producers. Hook proteins attach to microtubules at their N-terminal domains and the C-terminal domain associates with different organelles. Myo9b, which serves as a molecular motor to drive intracellular cargo on actin filaments, is also downregulated by 1.4-fold at transcript level. With the diverse functions that cytoskeletal elements are involved in, it is difficult to point the exact cellular functions that may have been altered. But combined with the altered functional classes identified, it seems logical to speculate that the change in cytoskeletal elements may be related to vesicle trafficking and protein secretion. Interestingly, a recent 2D-gel based proteomic investigation of four IgG_4-producing NS0 cell lines by linear regression analysis of functionally related proteins also indicated a correlation between cytoskeleton-related proteins and antibody productivity (Dinnis et al., 2006).

Chromatin

The third cellular component ontological class—chromatin, comprises of proteins involved in packaging DNA to form the condensed chromosome. The transcript levels of a large percentage (26%) of genes in this gene set were altered between high and low producer groups.

Seventy genes from this functional class are present on the MOE430A array as annotated by the GenMAPP database, of which 16 are differentially expressed by at least 1.4-fold (Table IV). The most notable gene in this class encodes tripartite motif 28 (Trim28). Genes encoding proteins involved in transcriptional silencing, such as DNA methyltransferase 3A (Dnmt3a) and methyl-CpG binding protein 1 (Mbd1), were also downregulated in high producers by more than 1.5-fold. Four other differentially expressed genes in this class encode histone proteins, which form the building blocks of chromatin. Modifications of these proteins alter DNA accessibility thereby regulating cellular processes such as transcription and replication.

Cell Cycle Progression

Among the 119 genes involved in cell cycle progression that are present on MOE430A array, 29 were identified as differentially expressed with a *q*-value and fold change threshold of 10% and 1.4, respectively (Table V). Among these, 21 are upregulated and 8 are downregulated. Thirteen of the upregulated genes encode products involved in mitotic (M) phase of cell cycle progression. One of the highly expressed and upregulated genes encodes a component of anaphase-promoting complex/cyclosome (APC/C).

Table III. List of differentially expressed genes involved in cytoskeleton function.

Gene symbol	Gene title	F.C.[a]	*q*-val. (%)[b]	Avg. int.[c]	Core subset[d]	Annotation
Actin-binding						
Tpm2	Tropomyosin 2, beta	(+) 2.1	3.09	307.3	No	Members of an actin-binding protein family that stabilize actin filaments and regulate access to other actin-binding proteins
Tpm1	Tropomyosin 1, alpha	(+) 1.8	0.00	159.3	No	
Tmsb10	Thymosin, beta 10	(+) 1.8	3.97	457.0	No	
Coro1c	Coronin, actin binding protein 1C	(+) 1.7	1.68	207.6	No	
Shrm	Shroom	(+) 1.6	1.68	95.2	No	
Tmod3	Tropomodulin 3	(+) 1.6	3.96	866.1	No	
Rdx	Radixin	(+) 1.5	6.62	348.2	No	
Sntb2	Syntrophin, basic 2	(+) 1.4	0.71	238.8	No	
Hip1r	Huntingtin interacting protein 1 related	(+) 1.4	8.25	414.4	No	
Dst	Dystonin	(−) 1.9	3.09	270.0	Yes	
Myo9b	Myosin IXb	(−) 1.4	5.47	244.3	Yes	Encodes an actin-based motor
Others						
Elmo1	Engulfment and cell motility 1	(+) 2.1	0.34	328.8	No	
Pxn	Paxillin	(+) 1.5	1.68	174.1	No	
Nf2	Neurofibromatosis 2	(+) 1.5	3.09	311.4	No	
Bicd2	Bicaudal D homolog 2	(+) 1.4	1.68	277.1	No	
Ctnnb1	Catenin (cadherin associated protein), beta 1	(+) 1.4	3.97	1167.3	No	An adherens junction protein that mediates cell-cell communication. Also interacts with TCF/LEF family of transcription factors to activate cyclin D1 transcription for G1/S phase transition (Tetsu and McCormick, 1999)
Sirt2	Sirtuin 2	(+) 1.4	3.96	595.8	No	A tubulin deacetylase that regulates exit from the mitotic phase of the cell cycle (Dryden et al., 2003)
Hook2	Hook homolog 2 (Drosophila)	(−) 17.4	0.00	37.2	Yes	A Hook protein which attaches to microtubules at its N-terminal domain and the C-terminal domain associates with different organelles
Jak2	Janus kinase 2	(−) 1.9	0.00	245.0	Yes	
Arpc5l	Actin related protein 2/3 complex, subunit 5-like	(−) 1.9	0.00	682.3	Yes	
Add3	Adducin 3 (gamma)	(−) 1.7	0.00	582.8	Yes	
Sspn	Sarcospan	(−) 1.5	8.25	161.4	Yes	
Ptpn21	Protein tyrosine phosphatase, non-receptor type 21	(−) 1.4	3.96	186.4	Yes	

[a]Fold change (F.C.), (+) upregulated in high producer (H), (−) downregulated in high producer.
[b]*q*-value calculated for each probe using SAM.
[c]Average intensity of high producers on Affymetrix array (MOE430A).
[d]Core member of the functional gene set identified as the leading-edge subset in GSEA.

APC/C associates with cell division cycle 20 (Cdc20) and other E1 ubiquitin-activating and E2 ubiquitin-conjugating enzymes during various stages of mitosis. APC/C regulates progression through M phase by 26S proteasome-mediated degradation of cyclin A and cyclin B, that arrest cell cycle. Ras-association domain family protein 1 (Rassf1), whose transcript is also upregulated in high producers by 2.2-fold, binds to Cdc20 during prometaphase to inhibit APC/C activity (Song et al., 2004).

Cell growth and death is the outcome of a concerted effect of several exogenous and endogenous factors. Low growth rate and high cell viability during the production phase is desirable in a high-producing cell line. Previous attempts to tap cell cycle regulation have focused on inducing expression of factors such as cyclin-dependent kinase inhibitors that can arrest cell growth (for review, see Seth et al., 2006). The upregulation of negative effectors of cell proliferation, Rassf1 and Mad2l1, which arrest G2/M phase progression, and Gmnn (Geminin), which inhibits transition from G2 to S phase, may reflect subtle changes in growth regulation in high-producing cells.

While these results suggest that transcript level alteration in several cell cycle-related genes is correlated with productivity, it is likely that several other unaccounted gene products also modulate cell proliferation at various levels of regulatory hierarchy. GST methods such as GSEA and GSA are motivated by the correlation of a pathway with one phenotype or another, that is, the genes involved in a common biological function are upregulated in one phenotype with respect to another, or vice versa. Alteration of many physiological processes can be invoked by differential up or downregulation of several genes involved.

Table IV. List of differentially expressed genes involved in the gene set "Chromatin"

Gene symbol	Gene title	F.C.[a]	*q*-val. (%)[b]	Avg. int.[c]	Core subset[d]	Annotation
Histones						
Hist3h2a	Histone 3, H2a	(+) 1.7	5.47	180.6	No	
H2afy3	H2A histone family, member Y3	(−) 2.4	6.62	139.7	Yes	
H2afy	H2A histone family, member Y	(−) 2.0	6.62	1690.8	Yes	
H2afz	H2A histone family, member Z	(−) 1.5	0.28	7553.1	Yes	
Transcriptional regulation						
Trim28	Tripartite motif protein 28	(−) 3.1	0.00	2322.4	Yes	Member of TRIM family of transcription factors that negatively regulate transcription from RNA polymerase II promoters
Dnmt3a	DNA methyltransferase 3A	(−) 2.6	2.15	114.4	Yes	Involved in de novo methylation of DNA
Mbd1	Methyl-CpG binding domain protein 1	(−) 1.6	0.40	1373.2	Yes	Member of methyl-CpG-binding domain proteins that interact with histone deacetylases to form transcriptional repressor complexes. Mbd1 can repress transcription from methylated gene promoters
Others						
Hmgb1	High mobility group box 1-like	(+) 2.1	3.97	1809.8	No	
Cbx2	Chromobox homolog 2 (Drosophila Pc class)	(+) 1.7	6.62	213.2	No	
Baz1b	Bromodomain adjacent to zinc finger domain, 1B	(+) 1.7	2.15	1040.1	No	
Cbx8	Chromobox homolog 8 (Drosophila Pc class)	(+) 1.5	0.71	276.7	No	
4930548G07Rik	RIKEN cDNA 4930548G07 gene	(+) 1.5	8.25	127.2	No	
Smarcc1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily c, member 1	(+) 1.4	8.25	869.0	No	
Suv39h2	Suppressor of variegation 3–9 homolog 2 (Drosophila)	(+) 1.4	8.25	363.3	No	
Asf1b	ASF1 anti-silencing function 1 homolog B (*S. cerevisiae*)	(−) 2.0	0.00	591.4	Yes	
Cbx1	Chromobox homolog 1 (Drosophila HP1 beta)	(−) 1.6	8.25	154.2	Yes	

[a]Fold change (F.C.), (+) upregulated in high producer (H), (−) downregulated in high producer.
[b]*q*-value calculated for each probe using SAM.
[c]Average intensity of high producers on Affymetrix array (MOE430A).
[d]Core member of the functional gene set identified as the leading-edge subset in GSEA.

However, regulation of cell cycle is a homeostatic balance of many positive and negative factors. For regulatory networks involving intricate interaction of positive and negative elements, the state of the functional class may not be easily identified as up- or downregulated.

Ribosomal Constituents

Several genes in this class have modest, albeit significant changes in expression level. Among the 14 genes differentially expressed by at least 1.4-fold, 10 are upregulated and 4 are downregulated (Table VI). Eight of the 10 upregulated transcripts encode mitochondrial ribosomal proteins. These mitochondrial ribosomal proteins are encoded in the nucleus and are responsible for translation of mitochondrial genes. Mitochondrial ribosome consists of a small 28S subunit and a larger 39S subunit. Among the eight upregulated transcripts, six encode 39S subunit proteins and two encode 28S subunit proteins. Although mitochondria harbor between 5% and 15% of eukaryotic proteome, few of the proteins are synthesized in mitochondria (Jensen et al., 2004). The mitochondrial genome has 13 protein-coding genes, which predominantly encode enzymes involved in the oxidative phosphorylation pathway for ATP synthesis (Bibb et al., 1981). These genes are not represented on the MOE430A array. Hence, the effect of the upregulation of a large number of mitochondrial ribosomal proteins on protein synthesis cannot be confirmed.

Ligase Activity

The set of genes encoding enzymes involved in ligation was also enriched (Table VII). Among the 20 differentially expressed, 8 are upregulated and 12 are downregulated. All the 12 downregulated genes were also identified in GSEA as members of the leading-edge subset. Although the genes in this functional category are involved in the molecular function of ligation, they are not all involved in the same biological function. However, among the 20 differentially expressed genes, 8 are involved in ubiquitin ligation. The ubiquitin-proteasome system plays a major role in cellular protein turnover. The ubiquitin-mediated proteasomal degradation pathway is comprised of several steps during which the ubiquitin moiety is activated by attachment to ubiquiting-activating enzyme (E1) and transferred to

Table V. List of differentially expressed genes involved in cell cycle progression.

Gene symbol	Gene title	F.C.[a]	*q*-val. (%)[b]	Avg. int.[c]	Core subset[d]	Annotation
G2/M phase transition						
Rassf1	Ras association (RalGDS/AF-6) domain family 1	(+) 2.2	0.71	703.8	Yes	Binds to Cdc20 to inhibit the activity of anaphase-promoting complex/cyclosome (APC/C)—a large multisubunit E3 ubiquitin ligase, during prometaphase of mitosis (Song et al., 2004). Also inhibits cellular progression from G1 to S phase by post-translational inhibition of cyclin D1 (Shivakumar et al., 2002)
Anapc5	Anaphase-promoting complex subunit 5	(+) 1.8	0.90	2043.3	Yes	Component of the APC/C complex. In active form, from prometaphase to telophase, APC/C promotes degradation of proteins, such as cyclin A and cyclin B, that arrest cell cycle
Mad2l1	MAD2 (mitotic arrest deficient, homolog)-like 1 (yeast)	(+) 1.7	2.15	2156.0	Yes	A component of the mitotic spindle checkpoint, that binds to Cdc20 and APC/C during metaphase to inhibit APC/C ligase activity (Fang et al., 1998; Li and Benezra, 1996)
Chfr	Checkpoint with forkhead and ring finger domains	(+) 1.7	2.15	445.4	Yes	Functions as a checkpoint for entry into mitotic phase of cell cycle
Stag1	Stromal antigen 1	(+) 1.6	3.09	554.6	Yes	
Smc4l1	SMC4 structural maintenance of chromosomes 4-like 1	(+) 1.6	1.24	1489.7	Yes	
Rad21	RAD21 homolog (*S. pombe*)	(+) 1.6	2.15	1512.8	Yes	
Rbl1	Retinoblastoma-like 1 (p107)	(+) 1.5	1.24	216.9	Yes	
Chek1	Checkpoint kinase 1 homolog (*S. pombe*)	(+) 1.5	5.47	871.0	Yes	
Nek2	NIMA (never in mitosis gene a)-related expressed kinase 2	(+) 1.5	8.25	543.0	Yes	
Cdc23	CDC23 (cell division cycle 23, yeast, homolog)	(+) 1.5	8.25	435.8	No	
Sirt2	Sirtuin 2 (*S. cerevisiae*)	(+) 1.4	3.96	595.8	Yes	
Sept11	Septin 11	(+) 1.4	1.67	362.6	Yes	
Spag5	Sperm associated antigen 5	(−) 1.5	3.09	621.5	No	Involved in regulation of mitotic spindle apparatus
Others						
Ccnd2	Cyclin D2	(+) 1.8	0.00	1054.9	Yes	Interacts with cyclin dependent kinases Cdk4 and Cdk6 for G1/S phase transition
Ccnc	Cyclin C	(+) 1.7	8.25	130.3	Yes	
Mapk6	Mitogen-activated protein kinase 6	(+) 1.7	3.09	753.2	Yes	
Gmnn	Geminin	(+) 1.7	0.48	4444.5	Yes	Accumulates during S and G2 phases and inhibits DNA replication by interacting with Cdt1, a replication initiation factor (Wohlschlegel et al., 2000)
Calm2	Calmodulin 2	(+) 1.6	5.47	3099.9	Yes	
Ccng2	Cyclin G2	(+) 1.6	3.96	698.3	Yes	
Calm3	Calmodulin 3	(+) 1.4	5.47	685.3	Yes	
Cdc7	Cell division cycle 7 (*S. cerevisiae*)	(+) 1.4	0.90	500.1	Yes	
Siah1a	Seven in absentia 1A	(−) 1.8	0.40	152.2	No	Encodes a E3 ubiquitin ligase involved in protein degradation
Lzts2	Leucine zipper, putative tumor suppressor 2	(−) 1.6	6.62	269.5	No	
Ahr	Aryl-hydrocarbon receptor	(−) 1.5	6.62	71.3	No	
Pdcd4	Programmed cell death 4	(−) 1.5	3.09	671.5	No	
Cspg6	Chondroitin sulfate proteoglycan 6	(−) 1.4	3.96	1343.9	No	
Txnl4	Thioredoxin-like 4	(−) 1.4	8.25	160.9	No	
Chaf1b	Chromatin assembly factor 1, subunit B (p60)	(−) 1.4	6.62	272.9	No	

[a]Fold change (F.C.), (+) upregulated in high producer (H), (−) downregulated in high producer.
[b]*q*-value calculated for each probe using SAM.
[c]Average intensity of high producers on Affymetrix array (MOE430A).
[d]Core member of the functional gene set identified as the leading-edge subset in GSEA.

ubiquitin-conjugating enzyme (E2). A third component, ubiquitin-ligase (E3) acts in association with E2 enzyme and binds to specific protein degrading signals on the target protein and a polyubiquitin chain is attached to the protein. This acts as a recognition signal for the 26S proteasome for protein degradation. Among the eight differentially expressed genes in this gene set, five encode E2 enzymes, and three correspond to E3 enzymes.

Three differentially expressed genes in this functional category encode tRNA synthetases, which are enzymes that

Table VI. List of differentially expressed genes in the functional class "structural constituent of ribosome."

Gene symbol	Gene title	F.C.[a]	*q*-val. (%)[b]	Avg. int.[c]	Core subset[d]	Annotation
Mitochondrial ribosomal proteins						
Mrpl1	Mitochondrial ribosomal protein L1	(+) 1.6	3.96	547.9	Yes	
Mrpl37	Mitochondrial ribosomal protein L37	(+) 1.6	6.62	1531.3	Yes	
Mrpl52	Mitochondrial ribosomal protein L52	(+) 1.5	3.09	841.2	Yes	
Mrpl19	Mitochondrial ribosomal protein L19	(+) 1.5	5.47	885.2	Yes	These genes encode components of the large
Mrpl38	Mitochondrial ribosomal protein L38	(+) 1.4	3.09	113.8	Yes	39S subunit of mitochondrial ribosomes
Mrpl27	Mitochondrial ribosomal protein L27	(+) 1.4	1.24	2263.0	Yes	
Mrpl44	Mitochondrial ribosomal protein L44	(−) 1.5	0.40	1497.2	No	
Mrpl43	Mitochondrial ribosomal protein L43	(−) 1.4	2.15	1343.7	No	
Mrps23	Mitochondrial ribosomal protein S23	(+) 1.6	5.47	952.2	Yes	These genes encode components of the small
Mrps5	Mitochondrial ribosomal protein S5	(+) 1.5	5.47	522.8	Yes	28S subunit of mitochondrial ribosomes
Others						
Nola2	Nucleolar protein family A, member 2	(+) 1.5	3.09	5299.2	Yes	
Rps10	Ribosomal protein S10	(+) 1.4	8.25	114.4	Yes	
Rpl5	Ribosomal protein L5	(−) 1.8	0.00	253.0	No	
Mrp63	Mitochondrial ribosomal protein 63	(−) 1.4	6.62	132.2	No	

[a]Fold change (F.C.), (+) upregulated in high producer (H), (−) downregulated in high producer.
[b]*q*-value calculated for each probe using SAM.
[c]Average intensity of high producers on Affymetrix array (MOE430A).
[d]Core member of the functional gene set identified as the leading-edge subset in GSEA.

read the trinucleotide sequence on the corresponding tRNAs and ligate the appropriate amino acid. The three differentially expressed transcripts encode tRNA synthetases corresponding to phenylalanine, glycine, and alanine. An additional five transcripts coding for tRNA synthases of leucine, glutamate, glutamine, and histidine are differentially expressed with fold change between 1.2- and 1.4-fold. Additionally, four differentially expressed genes with ligase

Table VII. List of differentially expressed genes involved in the functional class "Ligase activity."

Gene symbol	Gene title	F.C.[a]	*q*-val. (%)[b]	Avg. int.[c]	Core subset[d]	Annotation
Ubiquitin ligation						
Ube2s	Ubiquitin-conjugating enzyme E2S	(+) 1.4	8.25	6047.6	No	These genes encode members of ubiquitin-conjugating enzyme (E2) family
Ube2d2	Ubiquitin-conjugating enzyme E2D 2	(−) 1.7	0.00	546.2	Yes	
Ube2j1	Ubiquitin-conjugating enzyme E2, J1	(−) 1.6	3.96	362.8	Yes	
Ube2f	Ubiquitin-conjugating enzyme E2F (putative)	(−) 1.5	0.90	578.7	Yes	
Ube2t	Ubiquitin-conjugating enzyme E2T (putative)	(−) 1.4	8.25	1121.6	Yes	
Wwp2	WW domain containing E3 ubiquitin protein ligase 2	(+) 1.4	1.68	422.6	No	These genes encode members of E3 ubiquitin ligase family
Siah1a	Seven in absentia 1A	(−) 1.8	0.40	152.2	Yes	
Wwp1	WW domain containing E3 ubiquitin protein ligase 1	(−) 1.5	3.09	234.9	Yes	
tRNA synthetases						
Aars	Alanyl-tRNA synthetase	(+) 1.5	0.34	1915.5	No	
Farsla	Phenylalanine-tRNA synthetase-like, alpha subunit	(−) 1.7	1.24	265.8	Yes	
Gars	Glycyl-tRNA synthetase	(−) 1.4	6.62	1827.8	Yes	
Acyl-CoA synthetases						
Acsl6	Acyl-CoA synthetase long-chain family member 6	(+) 1.7	8.25	101.3	No	
Acsl1	Acyl-CoA synthetase long-chain family member 1	(+) 1.6	3.96	189.2	No	
Acss1	Acyl-CoA synthetase short-chain family member 1	(−) 1.5	3.09	156.0	Yes	
Acsl5	Acyl-CoA synthetase long-chain family member 5	(−) 1.4	3.09	2753.8	Yes	
Others						
Brap	BRCA1 associated protein	(+) 1.8	2.15	146.3	No	
Lig3	Ligase III, DNA, ATP-dependent	(+) 1.8	1.67	386.7	No	
Chfr	Checkpoint with forkhead and ring finger domains	(+) 1.7	2.15	445.4	No	
Gclm	Flutamate-cysteine ligase, modifier subunit	(−) 1.6	2.15	1454.4	Yes	
Rnf14	Ring finger protein 14	(−) 1.4	3.09	667.5	Yes	

[a]Fold change (F.C.), (+) upregulated in high producer (H), (−) downregulated in high producer.
[b]*q*-value calculated for each probe using SAM.
[c]Average intensity of high producers on Affymetrix array (MOE430A).
[d]Core member of the functional gene set identified as the leading-edge subset in GSEA.

activity encode acyl-CoA synthetases, two of which are upregulated. Members of this family of enzyme convert long chain fatty acids into fatty acyl-CoA esters. These enzymes are required for synthesis of cellular lipids and also fatty acid degradation.

Discussion

Identification of Molecular Signature for Productivity Trait

There is a profound interest in understanding the foundation of biological variability in the productivity of recombinant mammalian cells. It is customary to classify producing cells into high or low productivity groups. However, productivity is not an exactly defined trait. At least three elements may affect the final titer of recombinant protein profoundly: (i) the specific recombinant protein secretion rate, (ii) the growth rate and the growth extent, and (iii) the duration of sustained viability upon reaching maximum cell concentration. A super producer may have acquired some elements of all those positive characteristics, whereas a moderate producer may have only some of those positive characteristics. Even within each functional characteristic, there exist multiple routes to achieve the same superior features. For example, an elevated energy metabolism can potentially be accomplished through enrichment of mitochondria in each cell, enhanced expression of selective genes in mitochondrial metabolism, etc. Furthermore, the range of productivity is likely to be a continuum, rather than an arbitrary cut-off of high and low-producing classes.

Given the heterogeneous nature of cells in the same productivity class and the relatively arbitrary nature of conventional classification of high and low producers, the transcriptomes of cells in the same class may not all share common features. In an attempt to corroborate the differences in productivity with transcriptomic changes, the high and low producers were compared using support

Table VIII. Differentially expressed genes involved in early secretion pathway at nodes 4 and 5.

Gene symbol	Gene title	F.C.[a]	q-val. (%)[b]	Avg. int.[c]	Annotation
(a) Node 4: COPII vesicle-mediated ER-to-Golgi transport					
Sec31	SEC31-like 1 (*S. cerevisiae*)	(+) 1.5	3.96	1138.0	Component of the Sec13-Sec31 heterotetramer which is the outer structural layer of COPII coat
Trappc5	Trafficking protein particle complex 5	(+) 1.3	3.09	167.8	A subunit of heptameric TRAPPI tethering complex that activates Rab1 through GTP exchange (Sacher et al., 1998)
Rab1	RAB1, member RAS oncogene family	(+) 1.3	6.62	5875.2	A member of the Rab family of monomeric GTPases. The GTP-bound form of Rab1 regulates tethering and fusion of COPII vesicles to the target membrane (Allan et al., 2000)
Sar1a	SAR1 gene homolog A (*S. cerevisiae*)	(+) 1.3	10.44	2604.1	A small GTPase that is recruited to the ER membrane, which in turn recruits several components of COPII coat for vesicle formation and cargo selection
Vdp	Vesicle docking protein p115	(+) 1.2	10.44	1229.8	A coiled-coil tethering protein that is recruited to COPII coat in a Rab1-dependent manner and interacts with a subset of v-SNAREs to promote vesicle tethering and fusion (Allan et al., 2000)
Sec22b	SEC22 vesicle trafficking protein homolog B (*S. cerevisiae*)	(−) 1.6	0.40	353.0	*Sec22b* encodes a t-SNARE protein for ER-Golgi transport
Rab2	RAB2, member RAS oncogene family	(−) 1.2	6.62	1442.0	Another member of Rab family that is essential for vesicle-mediated transport from ER to Golgi or pre-Golgi compartments (Tisdale and Balch, 1996)
(b) Node 5: COPI vesicle-mediated intra-Golgi transport and retrograde transport from Golgi to ER					
Arf3	ADP-ribosylation factor 3	(+) 2.0	0.48	211.2	A member of the Class I family of Arf proteins that regulate assembly of several coat proteins including COPI and clathrin coats
Wasl	Wiskott-Aldrich syndrome-like (human) (N-WASP)	(+) 1.5	0.90	334.1	N-WASP regulates actin assembly on COPI vesicles by stimulating the Arp2/3 (actin-related protein 2/3) complex (Fucini et al., 2002)
Arf2	ADP-ribosylation factor 2	(+) 1.4	10.44	167.5	Another member of class I family of Arf proteins
Arhgap21	Rho GTPase activating protein 21	(+) 1.3	6.62	825.7	Serves as a GTPase-activating protein for the Rho-family GTPase, Cdc42, which in turn activates N-WASP for actin polymerization

[a]Fold change (F.C.), (+) upregulated in high producer (H), (−) downregulated in high producer.
[b]q-value calculated for each probe using SAM.
[c]Average intensity of high producers on Affymetrix array (MOE430A).

vector machines (SVM) (Vapnik, 1995, 1998). SVM is a pattern classification method that can identify the diverse gene expression attributes associated with high productivity. Classification models constructed from these attributes can also be used to interrogate other "unknown" producers based on their gene expression profiles. A differential expression-based gene selection approach was combined with leave-one-out cross-validation to construct and evaluate SVM models for binary classification of high and low producers (Baker and Kramer, 2006; Joachims, 1999; Zhang et al., 2006). The models highlighted the expression differences between high and low producers as well as the expression differences amongst the high producers. Four of the seven high producers had a substantially high congruity in their gene expression profiles and could be contrasted from the other three high producers as well as the low producers. A limitation of models based on a small number of producers is their restricted predictive ability. Nonetheless, a comparison based on few producers can provide useful insights to the physiology that underlies the productivity trait. A small sample size is almost an inherent problem of comparative transcriptome analysis of high-low producers. All efforts in cell line development aim to generate high-producing clones; few low-producing lines are kept, let alone characterized. Even the number of high-producing cell lines selected for characterization is usually small. This is a drastic contrast to studies of different disease phenotypes, which often have a very large patient base with tens to hundreds of samples for identification of molecular signatures (Golub et al., 1999; Ramaswamy et al., 2003).

Significance Testing for Functional Analysis

In search of genes conferring hyper productivity through transcriptome analysis, one invariably resorts to statistical testing by setting criteria of fold differential expression, FDR, etc. More stringent criteria imply a lower risk of false positive calls. In seeking pivotal genes responsible for a complex trait, a question inevitably arises whether the trait is caused by colossal alterations in a small number of master genes or by minute variations globally distributed in many functional classes. Small changes in transcript levels can be physiologically significant, especially when many genes involved in the same pathway or functional class change simultaneously. We thus employed multiple methods for functional class analysis. Two of the three methods used do not rely on setting an explicit criterion to preselect differentially expressed genes. Rather a differential expression metric is used to rank all the genes in the dataset, and the distribution of ranks of a set of functionally related genes is used to characterize alteration of a functional class.

The functional class analysis identified several classes in cellular component, biological process, and molecular function ontology as significantly changed, as described earlier. However, many differentially expressed genes that do not fall into one of the identified functional classes may also play a significant role. Genes which are differentially expressed but do not fall into one of the significant classes were further examined. It was noted that many are involved in protein trafficking (Table VIII). Together with the leading edge genes in identified functional classes, an overall picture of changes observed can be broadly depicted in Figure 1a. Functionally, the leading edge genes in the three cellular component classes (Golgi apparatus, cytoskeleton and chromatin) are involved in protein synthesis, processing and transport. Spatially, the enriched genes in the classes related to biological processes and molecular functions are distributed in cytosol, mitochondria, and ER. The three dimensions of functional class analysis (cellular component, biological process, and molecular function) are closely interconnected, as expected. Node 1 comprises members of transcriptional regulation that are present in chromatin functional class (Table IV). Genes involved in protein synthesis, namely the genes encoding ribosomal proteins

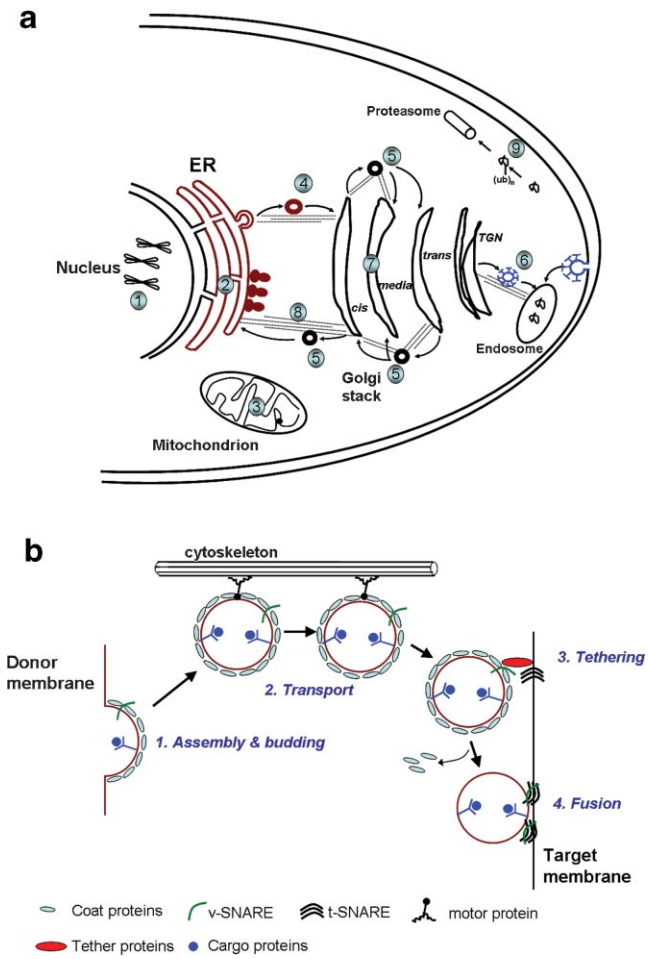

Figure 1. Genes differentially expressed between high and low producing NS0 cell lines grouped according to intracellular function. **a**: Each node depicts an intracellular process with large number of differentially expressed genes as identified by GST. **b**: Schematic of the steps involved in vesicle-mediated transport (nodes 4, 5, and 6) (modified from Cai et al., 2007).

and tRNA synthetases in cytoplasm (node 2) and mitochondrial ribosomal proteins (node 3) are differentially expressed between high and low producers (Tables VI and VII). Nodes 4–8 depict functions related to protein processing and secretion. Nodes 4, 5, and 6 represent various components of protein transport mediated by COPII, COPI, and clathrin-coated vesicles, respectively. Node 7 describes constituents of Golgi apparatus including glycosyltransferase enzymes (Table II). The cytoskeletal network (node 8) comprises actin and microtubule filaments and molecular motors that drive vesicular cargo from one organelle to another (Table III). Lastly, node 9 depicts the machinery for protein degradation, which includes several members of ubiquitin-mediated protein degradation pathway (Table VII).

Nodes 4, 5, and 6, which describe vesicle-mediated transport in early secretion pathway, are shown in expanded form in Figure 1b. The four essential steps in this trafficking include cargo selection followed by vesicle assembly and budding, vesicle transport to the target organelle by cytoskeletal motors, tethering of vesicles to the target membrane, and the final step involves fusion of the vesicular and target membranes. COPII vesicles are required for ER to early-Golgi transport. Transcripts of several genes involved in COPII vesicle transport are differentially expressed (Table VIIIa). Sar1a, a small GTPase, which is activated by nucleotide exchange on the ER membrane, is crucial for vesicle assembly. Rab1 and Rab2 are monomeric GTPases that regulate tethering and fusion of COPII vesicles to early Golgi or ER-Golgi intermediate compartment (ERGIC). Vdp and Trappc5 are among several Rab effectors that interact with Rab1 to ensure specificity of vesicle tethering to target membrane. Both Vdp and Trappc5 are upregulated in high producers. According to a recent report, Vdp is also upregulated in sodium butyrate-treated mouse hybridoma cells (MAK) and CHO cells by 1.3- and 1.5-fold, respectively (Yee et al., 2008). Interestingly, Vdp is also among the genes identified as Xbp1 targets during B cell differentiation (Shaffer et al., 2004). The final step of vesicle fusion is facilitated by ER-to-Golgi SNARE proteins, Gosr1 (Table II) and Sec22b. Upregulation of these protein trafficking components in high producers is consistent with an earlier report suggesting that ER to *cis*-Golgi transport is the rate-limiting step in mammalian as well as insect cells (Hooker et al., 1999). Several components of COPI vesicle-mediated intra-Golgi transport and retrograde transport from Golgi to ER (node 5) are also differentially expressed at the transcript level (Table VIIIb). Of particular note are the genes encoding N-Wasp and Arhgap21 that coordinate actin assembly on COPI vesicles in an Arf1-dependent manner. Others members including components of vesicle tethering complex COG (Cog4 and Cog8), Arf1-activating protein (Arfgap1), and a SNARE protein (Gosr2) that facilitate COPI-mediated vesicle transport are also upregulated in high producers at the transcript level (Table II). The transcript of Napg, which encodes the γ-SNAP protein involved in disassembly of SNARE complex for vesicle membrane recycling, is also differentially expressed (Table II).

Concluding Remarks

This and the previous studies both examined the hyper productivity trait by comparing high and low-producing NS0 cell lines. The previous study used transcriptome and iTRAQ^{TM}-based proteomic profiles (Seth et al., 2007b) and took an inferential statistics approach. Several genes involved in protein processing were differentially altered between the two classes at transcriptome as well as proteome levels. Furthermore, genes related to cell cycle also appeared to be modified between the two classes at transcriptome level. In this study, pathway-level analysis was performed on a subset of high and low producers to identify physiological functions that are differentially altered between the two groups. GST tools were employed to discern statistical significance at a functional level, rather than at individual gene level. Analysis based on GST indicates that several functional classes, including protein processing constituents in the Golgi apparatus, cytoskeleton-related, and cell cycle-related functions were altered. At an individual gene level the results of this and the previous study were not entirely overlapping. This may not be surprising as the dataset used was somewhat different and the methods employed are not the same. However, at functional level the results of the two studies bear many similarities. With a trait as complex as high recombinant protein productivity, the dissection of transcriptome data is inherently a complex task. Gene-level differential expression analysis and GST are complementary approaches that are revealing in somewhat different ways. This study demonstrates the value of using GST as complementary to gene-level differential expression analysis.

Taken together our data and previous reports on transcriptome and proteome analysis of high and low producers seem to suggest that a number of functional classes are involved in enhanced productivity, including protein processing, vesicle trafficking and cell growth regulation. The study of hyper productivity is likely to benefit from combining high–low producer comparisons with comparative investigations on culture conditions that increase productivity. As such results begin to accumulate in the near future we can expect our understanding of this complex trait to expand and our ability to direct cells towards hyper productivity to greatly advance.

Informatics support from Minnesota Supercomputing Institute is gratefully acknowledged. We also thank Dr. Gargi Seth and Anne Kantardjieff for valuable discussions.

References

Alete DE, Racher AJ, Birch JR, Stansfield SH, James DC, Smales CM. 2005. Proteomic analysis of enriched microsomal fractions from GS-NS0

murine myeloma cells with varying secreted recombinant monoclonal antibody productivities. Proteomics 5(18):4689–4704.
Allan BB, Moyer BD, Balch WE. 2000. Rab1 recruitment of p115 into a cis-SNARE complex: Programming budding COPII vesicles for fusion. Science 289(5478):444–448.
Baker SG, Kramer BS. 2006. Identifying genes that contribute most to good classification in microarrays. BMC Bioinformatics 7(1):407.
Ben-Shaul Y, Bergman H, Soreq H. 2005. Identifying subtle interrelated changes in functional gene categories using continuous measures of gene expression. Bioinformatics 21(7):1129–1137.
Bibb MJ, Van Etten RA, Wright CT, Walberg MW, Clayton DA. 1981. Sequence and gene organization of mouse mitochondrial DNA. Cell 26(2 Pt 2):167–180.
Bigay J, Gounon P, Robineau S, Antonny B. 2003. Lipid packing sensed by ArfGAP1 couples COPI coat disassembly to membrane bilayer curvature. Nature 426(6966):563–566.
Cai H, Reinisch K, Ferro-Novick S. 2007. Coats, tethers, Rabs, and SNAREs work together to mediate the intracellular destination of a transport vesicle. Dev Cell 12(5):671–682.
Dahlquist KD, Salomonis N, Vranizan K, Lawlor SC, Conklin BR. 2002. GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. Nat Genet 31(1):19–20.
De Leon Gatti M, Wlaschin KF, Nissom PM, Yap M, Hu WS. 2007. Comparative transcriptional analysis of mouse hybridoma and recombinant Chinese hamster ovary cells undergoing butyrate treatment. J Biosci Bioeng 103(1):82–91.
DeBose-Boyd RA, Brown MS, Li WP, Nohturfft A, Goldstein JL, Espenshade PJ. 1999. Transport-dependent proteolysis of SREBP: Relocation of site-1 protease from Golgi to ER obviates the need for SREBP transport to Golgi. Cell 99(7):703–712.
Dell'Angelica EC, Ohno H, Ooi CE, Rabinovich E, Roche KW, Bonifacino JS. 1997. AP-3: An adaptor-like protein complex with ubiquitous expression. EMBO J 16(5):917–928.
Dinnis DM, Stansfield SH, Schlatter S, Smales CM, Alete D, Birch JR, Racher AJ, Marshall CT, Nielsen LK, James DC. 2006. Functional proteomic analysis of GS-NS0 murine myeloma cell lines with varying recombinant monoclonal antibody production rate. Biotechnol Bioeng 94(5):830–841.
Doniger SW, Salomonis N, Dahlquist KD, Vranizan K, Lawlor SC, Conklin BR. 2003. MAPPFinder: Using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. Genome Biol 4(1):R7.
Dryden SC, Nahhas FA, Nowak JE, Goustin AS, Tainsky MA. 2003. Role for human SIRT2 NAD-dependent deacetylase activity in control of mitotic exit in the cell cycle. Mol Cell Biol 23(9):3173–3185.
Efron B, Tibshirani R. 2007. On testing the significance of sets of genes. Ann Appl Stat 1(1):107–129.
Fang G, Yu H, Kirschner MW. 1998. The checkpoint protein MAD2 and the mitotic regulator CDC20 form a ternary complex with the anaphase-promoting complex to control anaphase initiation. Genes Dev 12(12): 1871–1883.
Fucini RV, Chen JL, Sharma C, Kessels MM, Stamnes M. 2002. Golgi vesicle proteins are linked to the assembly of an actin complex defined by mAbp1. Mol Biol Cell 13(2):621–631.
Goeman JJ, Buhlmann P. 2007. Analyzing gene expression data in terms of gene sets: Methodological issues. Bioinformatics 23(8):980–987.
Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. 1999. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. Science 286(5439):531–537.
Gupta P, Lee KH. 2007. Genomics and proteomics in process development: Opportunities and challenges. Trends Biotechnol 25(7):324–330.
Hooker AD, Green NH, Baines AJ, Bull AT, Jenkins N, Strange PG, James DC. 1999. Constraints on the transport and glycosylation of recombinant IFN-gamma in Chinese hamster ovary and insect cells. Biotechnol Bioeng 63(5):559–572.
Jensen RE, Dunn CD, Youngman MJ, Sesaki H. 2004. Mitochondrial building blocks. Trends Cell Biol 14(5):215–218.
Joachims T. 1999. Making large-scale SVM learning practical. In: Scholkofp B, Burges CJC, Smola AJ, editors. Advances in kernel methods—Support vector learning. Cambridge, MA: M.I.T. Press. p 169–184.
Khoo SH, Falciani F, Al-Rubeai M. 2007. A genome-wide transcriptional analysis of producer and non-producer NS0 myeloma cell lines. Biotechnol Appl Biochem 47(Pt 2):85–95.
Lee YC, Kurosawa N, Hamamoto T, Nakaoka T, Tsuji S. 1993. Molecular cloning and expression of Gal beta 1,3GalNAc alpha 2,3-sialyltransferase from mouse brain. Eur J Biochem 216(2):377–385.
Li Y, Benezra R. 1996. Identification of a human mitotic checkpoint gene: hsMAD2. Science 274(5285):246–248.
Lowe SL, Peter F, Subramaniam VN, Wong SH, Hong W. 1997. A SNARE involved in protein transport through the Golgi apparatus. Nature 389(6653):881–884.
Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, et al. 2003. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 34(3): 267–273.
Ramaswamy S, Ross KN, Lander ES, Golub TR. 2003. A molecular signature of metastasis in primary solid tumors. Nat Genet 33(1):49–54.
Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. 2006. GenePattern 2.0. Nat Genet 38(5):500–501.
Ross JL, Ali MY, Warshaw DM. 2008. Cargo transport: Molecular motors navigate a complex cytoskeleton. Curr Opin Cell Biol 20(1):41–47.
Sacher M, Jiang Y, Barrowman J, Scarpa A, Burston J, Zhang L, Schieltz D, Yates JR III, Abeliovich H, Ferro-Novick S. 1998. TRAPP, a highly conserved novel complex on the cis-Golgi that mediates vesicle docking and fusion. EMBO J 17(9):2494–2503.
Seth G, Hossler P, Yee JC, Hu WS. 2006. Engineering cells for cell culture bioprocessing—Physiological fundamentals. Adv Biochem Eng Biotechnol 101:119–164.
Seth G, Charaniya S, Wlaschin KF, Hu WS. 2007a. In pursuit of a super producer-alternative paths to high producing recombinant mammalian cells. Curr Opin Biotechnol 18(6):557–564.
Seth G, Philp RJ, Lau A, Jiun KY, Yap M, Hu WS. 2007b. Molecular portrait of high productivity in recombinant NS0 cells. Biotechnol Bioeng 97(4):933–951.
Shaffer AL, Shapiro-Shelef M, Iwakoshi NN, Lee AH, Qian SB, Zhao H, Yu X, Yang L, Tan BK, Rosenwald A, et al. 2004. XBP1, downstream of Blimp-1, expands the secretory apparatus and other organelles, and increases protein synthesis in plasma cell differentiation. Immunity 21(1):81–93.
Shields RL, Lai J, Keck R, O'Connell LY, Hong K, Meng YG, Weikert SH, Presta LG. 2002. Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fcgamma RIII and antibody-dependent cellular toxicity. J Biol Chem 277(30):26733–26740.
Shivakumar L, Minna J, Sakamaki T, Pestell R, White MA. 2002. The RASSF1A tumor suppressor blocks cell cycle progression and inhibits cyclin D1 accumulation. Mol Cell Biol 22(12):4309–4318.
Smales CM, Dinnis DM, Stansfield SH, Alete D, Sage EA, Birch JR, Racher AJ, Marshall CT, James DC. 2004. Comparative proteomic analysis of GS-NS0 murine myeloma cell lines with varying recombinant monoclonal antibody production rate. Biotechnol Bioeng 88(4): 474–488.
Sollner T, Bennett MK, Whiteheart SW, Scheller RH, Rothman JE. 1993. A protein assembly-disassembly pathway in vitro that may correspond to sequential steps of synaptic vesicle docking, activation, and fusion. Cell 75(3):409–418.
Song MS, Song SJ, Ayad NG, Chang JS, Lee JH, Hong HK, Lee H, Choi N, Kim J, Kim H, Kim JW, Choi EJ, Kirschner MW, Lim DS. 2004. The tumour suppressor RASSF1A regulates mitosis by inhibiting the APC-Cdc20 complex. Nat Cell Biol 6(2):129–137.
Stamnes M. 2002. Regulating the actin cytoskeleton during vesicular transport. Curr Opin Cell Biol 14(4):428–433.
Subramaniam VN, Peter F, Philp R, Wong SH, Hong W. 1996. GS28, a 28-kilodalton Golgi SNARE that participates in ER-Golgi transport. Science 272(5265):1161–1163.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102(43): 15545–15550.

Tetsu O, McCormick F. 1999. Beta-catenin regulates expression of cyclin D1 in colon carcinoma cells. Nature 398(6726):422–426.

Tisdale EJ, Balch WE. 1996. Rab2 is essential for the maturation of pre-Golgi intermediates. J Biol Chem 271(46):29372–29379.

Tusher VG, Tibshirani R, Chu G. 2001. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 98(9):5116–5121.

Ungar D, Oka T, Brittle EE, Vasile E, Lupashin VV, Chatterton JE, Heuser JE, Krieger M, Waters MG. 2002. Characterization of a mammalian Golgi-localized protein complex, COG, that is required for normal Golgi morphology and function. J Cell Biol 157(3):405–415.

Vapnik VN. 1995. The nature of statistical learning theory. New York: Springer. xv, 188 p.

Vapnik VN. 1998. Statistical learning theory. New York: Wiley. xxiv, 736 p.

Wohlschlegel JA, Dwyer BT, Dhar SK, Cvetic C, Walter JC, Dutta A. 2000. Inhibition of eukaryotic DNA replication by geminin binding to Cdt1. Science 290(5500):2309–2312.

Yee JC, de Leon Gatti M, Philp RJ, Yap M, Hu WS. 2008. Genomic and proteomic exploration of CHO and hybridoma cells under sodium butyrate treatment. Biotechnol Bioeng 99(5):1186–1204.

Zhang X, Lu X, Shi Q, Xu XQ, Leung HC, Harris LN, Iglehart JD, Miron A, Liu JS, Wong WH. 2006. Recursive SVM feature selection and sample classification for mass-spectrometry and microarray data. BMC Bioinformatics 7:197.