iTRAQ-based Quantitative Proteomic Analysis of Mammalian Cell Lines

by

Vida Talebian

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Biology

Waterloo, Ontario, Canada, 2018

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

High-throughput proteomics approaches have been successfully employed in the inclusive and unbiased characterization of cellular proteome. In this thesis, we are aiming to comprehensively compare changes in proteome expression in mammalian cell lines under two different conditions. The first is an analysis of mAb producing CHO cells with a focus on comparing the productivity of inducible cells, and the second investigates human colorectal cancer cells with a focus on chemo-resistant cells. To this end, the quantitative proteomics technique iTRAQ was performed on the samples.

The production of monoclonal antibodies as therapeutic medicines has gained increasing attention and applications in the last decades. Improving production levels is an important issue that can economically impact the industry. Chinese hamster ovary (CHO) cells are the most commonly used mammalian cell lines for the production of therapeutic mAbs and have been genetically modified for this purpose. Monitoring the effects of genetic modification systems on the cells is very important to maintain the quality and consistency of the cellular products. The possible changes or stresses forced on the cells after these modifications can be observed by analyzing the cells' proteome and can enable the development of optimized expression systems. Also, the biological characteristics of high-producing CHO cells could be obtained by proteomic comparison of these cells.

We quantified the expression levels of 1,616 unique proteins from an iTRAQ MS analysis and identified 81 proteins with p-values lower than 0.05 where the highest fold-change belong to the mAb sequence. Overall, proteins did not show any large changes based on the fold-change and it is likely that the induction system with cumate gene switch did not cause large metabolic changes or stress responses in the CHO cells. With these results we can suggest that the system has generated a stable cell line with a higher capacity for mAb production without inducing any major stress responses in the cells, and the cell line could have the potential for high mAb production using a stronger promoter system.

The second topic of interest in this thesis is therapy resistance colorectal cancer, which is either caused by a selection or induction of drug resistance in a subset of colorectal cancer cells. Drug resistance is the first reason for treatment failure or metastasis, therefore understanding the biological features of resistant cells can be helpful. To this end, the cancer cell line, HT-29, and a SN-38 drug-resistant HT-29 cell line were explored by iTRAQ quantitative proteomics analysis. SN-38 is the active metabolite of the chemotherapy drug, irinotecan, a common mediation for treating colon cancer.

The expression levels of 1,665 proteins were quantified and 21 proteins had significant changes with more than 1 log-fold-change. PAGE4 is the most up-regulated protein in the resistant HT-29 cells; it shows to have a role in cell viability and survival. Among the down-regulated proteins, cytokeratin (KRT20) had a significant low expression in resistant HT-29 cells that can show the change in cell migration because it is known for circulating CRC cells. Biosynthesis of antibiotics is identified as the most enriched KEGG pathway within the up-regulated genes; also, ribosome pathway is observed as the most enriched pathway among the down-regulated gene list. Overall, a large number of the differentially regulated genes were known in the stem cancer cells, EMT processes and metastasis. The main functions among the differentially regulated proteins is related to cell adhesion and migration, cell proliferation, and the ribosome.

Acknowledgements

I would like to sincerely thank my supervisor Professor Brendan McConkey for his continuous support, guidance and encouragement throughout the course of my research. I would also like to express thanks to the University of Waterloo and the Department of Biology for the opportunity to pursue graduate studies. I would like to thank Professors Andrew Doxey and Mungo Marsden for agreeing to be join my MSc. committee and for reviewing and evaluating my thesis.

I would like to acknowledge MAbNet, part of the project was related to the larger NSREC Strategic Network MAbNet grant. I would like to thank Dr. Durocher and their lab members at the Biotechnology Research Institute (BRI, Montréal, Canada) for supporting part of the research experiments. I would also like to thank Dr. Jonathan Blay lab at the School of Pharmacy at the University of Waterloo and Ms. Julia Fux for the helps in the second part of thesis. Thank you to all members of McConkey lab during my time in the lab, thanks to Zohreh, Janet, Laura, Mark, Monica, Owen, and Karsten. Thanks to Dr. Steve Mooney for all the support and mentorship in the lab.

Thank you to my husband Ali for supporting me in pursue of my dreams and sticking with me; Mom, Dad, and my friends for all their support and love.

Table of contents

List of figures

List of tables

Abbreviations

Acronym	Name
ALC	Average Local Confidence
AML	Acute myeloid leukemia
BSA	Bovine serum albumin
CHO	Chinese hamster ovary
CID	Collision induced dissociation
CMV	Cytomegalovirus
CRC	Colorectal cancer
Da	Dalton
DAVID	The database for annotation, visualization, and integrated discovery
DHFR	Dihydrofolate reductase
ECM	Extracellular matrix
ESI	Electrospray ionization
FASP	Filter-aided sample preparation
FC	Fold-change
FDA	Food and Drug Administration
FDR	False discovery rate
GO	Gene Ontology
GS	Glutamine Synthetase
GSH	Glutathione
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IAM	Iodoacetamide
iTRAQ	Isobaric tag for relative and absolute quantitation
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC	Liquid chromatography
MAb	Monoclonal antibody
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
P-gp	P-glycoprotein
ppm	Parts per million
PSM	Peptide Spectrum Match
PTM	Post translational modification
SN-38	7-ethyl-10-hydroxycamptothecin
SPIDER	Software for Protein Identification from Sequence Tags Containing De Novo Sequencing Error
TCEP	tris(2-carboxyethyl)phosphine
TEAB	Triethylammonium bicarbonate
TLC	Total Local Confidence
UA	Urea

Chapter 1

Introduction

1.1 Proteomics study of CHO cells

1.1.1 An introduction to monoclonal antibodies

At the beginning of 20th century the Nobel laureate Paul Ehrlich introduced the concept of antibodies; he coined the term "magic bullet" to describe 'drugs that go straight to their intended cell-structural targets' [1]. Ehrlich's works inspired many scientists, and has led to discovery of successful treatments including novel chemotherapeutics. In 1975, Köhler and Milstein made the first hybridoma cell for producing monoclonal antibodies (mAb) by fusion of "a mouse myeloma and mouse spleen cells from an immunized donor" [2]. This discovery brought them the Nobel prize for Physiology or Medicine in 1984 [3]. Early applications of mAbs were primarily for diagnostic purposes. Therapeutic applications of mAbs were not evident until 1986, when the first mAb product, Orthoclone OKT3 (muromonab-CD3), was approved for use in treating and preventing graft versus host disease in transplant patients [4].

The approval of recombinant proteins and mAbs as treatments for different autoimmune diseases and cancers has brought more attention to the production technology. In 2014, six mAbs were listed in the top 10 best-selling drugs, and overall sale of these mAb products has reached 60 billion USD per year [5]. Currently, around fifty monoclonal antibodies are approved and available on the market, and it is anticipated that around seventy monoclonal antibodies will be on the market by 2020 [6]. In the recent years, worldwide sales of all monoclonal antibody products have had a higher growth compared to other recombinant therapeutic proteins. For example, from 2008 to 2013 mAbs market had an increase of about 90% (from 39 billion dollars to about 75 billion dollars), while other recombinant protein therapeutics had only 26% in that time period [6]. Most of monoclonal antibodies in the market are produced using mammalian cell lines. Chinese hamster ovary cells (CHO), baby hamster kidney (BHK), murine myeloma derived (NS0), human

embryonic kidney (HEK-293), and the human retina-derived PerC6 are the main mammalian host cell lines used in the industry. Among these cell lines, CHO cells are the host cell line for nearly 70% of recombinant protein therapeutics [7].

Subsequently, the competition in the field of biopharmaceuticals has increased. This is partly because some of the patents for the recombinant protein therapeutics including mAbs have expired or are due to expire soon, therefore other manufacturers in countries like India and China have entered the market to produce the generic version of these biologic drugs, also called biosimilars. From other aspects, the access to bioinformatics tools has increased and the release of the first draft of genome sequence of CHO cells in 2011 has facilitated the study of these cells in depth [8]. Gaining more information about the cells growth and protein production mechanisms could lead to developing and improving the methods of optimizing production of mAbs.

1.1.2 Host cell lines

A first step in producing recombinant proteins including mAbs is to find the right host cells that could be used to express the protein of interest. As a rule of thumb, smaller proteins (smaller than 30 KDa) that do not have complicated post-translational modifications could be produced in model prokaryotes such as *Escherichia coli*. These bacterial systems are cheaper as the bacteria grows faster and the medium used for growing the bacteria is much cheaper to the ones used for growing eukaryotic cells, especially for mammalian cells. Normally larger proteins (greater than approximately 100 kDa) cannot be expressed in bacterial systems. Eukaryotic cells such as mammalian cells, fungi or baculovirus systems could be used as a host for expressing these large proteins. In addition to the size of the protein other important factors such as required post-translational modification for the protein of interest also should be considered in choosing the right expression system. For example glycosylation, which is an important post-translational modification in mAbs, cannot be done by prokaryotic systems such as *E. coli*. Also, disulfide-rich proteins cannot be expressed properly in *E. coli*. Sometimes even proteins of interest that are normally expressed in eukaryotes could be toxic to prokaryotic cells and therefore cannot be produced by these cells [9]. Other

expression systems for production of recombinant proteins also exist, such as using plants and transgenic animals, such as expression of mAbs in tobacco plant and production of recombinant proteins in milk, egg white, blood or urine of transgenic animals like mice, pigs, or rabbits, but they have not gained as much as popularity as mammalian cell expression systems so far [9, 10].

An antibody is constructed of two heavy chains and two light chains that are connected through hydrogen bonding, van der Waals interactions, and disulfide bonds. A general overview of the structure is shown in figure 1.1. Light chains are comprised of a variable domain and a constant domain, while heavy chains are have a variable domain and three constant domains. The antigen binding region of a mAb (Fab region) contains the heavy chain and light chain variable domains; this fragment determines the specificity of the antibody. The fragment crystallizable (Fc region) is made up of constant heavy chain domains that are glycosylated, and is responsible for effector functions and the extended half-life of the molecule in the body (Figure 1.1) [11].

Monoclonal antibodies are complex glycoproteins; after their expression by the cell they go through important post-translational modification called glycosylation. During this modification, glycans that are certain carbohydrate moieties, bind to the proteins to build the mAb's structure. Glycosylation plays an important role in the structure and the reaction of the body to the mAb. The glycosylation pattern could highly affect the protein stability, biological activity e.g. antibody-dependent cellular cytotoxicity (ADCC), and also it affects the pharmacokinetics of the monoclonal antibodies [12].

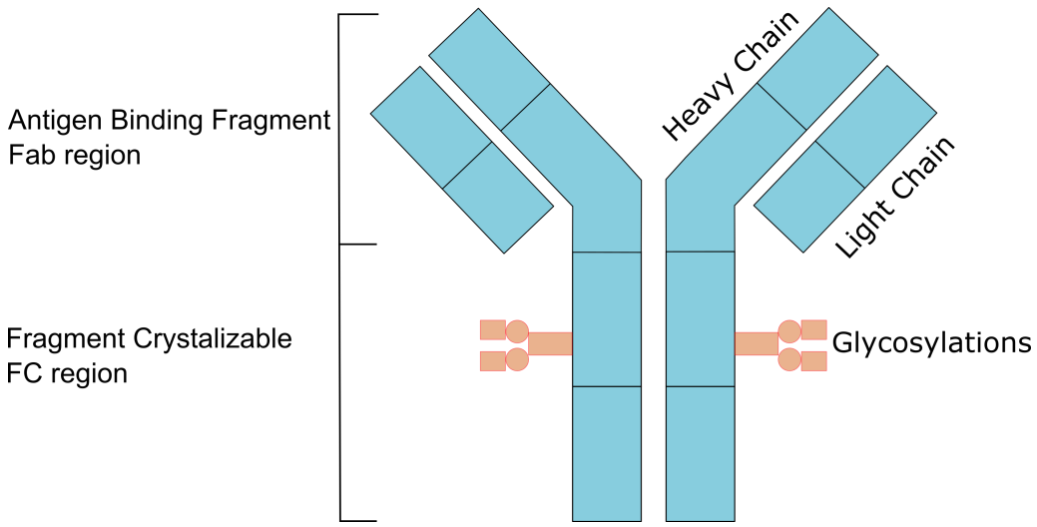

Figure 1.1: A generalized structure of a glycosylated mAb.

The correct glycosylation pattern is critical for having functional glycoforms that do not have immunogenic effects and do not clear quickly from the body [13]. Factors such as the enzymatic pathways of the host cell and the culture conditions impact the glycosylation patterns of secreted protein of interest [14]. To this end, the host cells for mAb production that determine the characteristics of product are the important key to the biopharmaceutical products. For example, insect cells normally are not able to produce these glycosylation patterns that are similar to the patterns in human antibodies. Among the different eukaryotic cells mammalian cells could express proteins with more similarity to the ones expressed in human cells with proper folding and glycosylation forms without the need for renaturing the product. Among the mammalian cells Chinese hamster ovary (CHO) cells have become the workhorse for producing complex therapeutic proteins such as mAbs [9]. In the next section, I will review these cells and the process of production.

1.1.3 Chinese Hamster Ovary Cells

In 1919, Chinese hamsters were first used as a substitute for mice in pneumococcal typing. Later on, low number of chromosomes in Chinese hamsters made them a model for radiation cytogenetics and for cell culture studies [15]. In the late twentieth century, female Chinese hamsters became vital tools for the development of molecular biology. The mammalian genetics studies had progressive improvements by research on the donated ovary cells, and they continued to be an essential part of biotechnological science and industry [13]. The first CHO cell line was developed in the late 1950s by Dr. Theodore Puck at the University of Colorado [16]. Since then these cells were used in *in-vitro* studies and many different modifications on them was implemented for different application purposes from lab-studies to being a cell factory for producing biologics [15]. The first protein therapeutic produced by CHO cells as the host was human tissue plasminogen activator (tPA, Activase) (Genentech, San Francisco, CA, USA). This product was approved for human use at 1986 by the FDA. In this section, I summarize the process of making protein therapeutics in CHO cells. Normally for the production of every recombinant therapeutic protein three overall steps should be undertaken, as described below.

As a first step, the gene which is responsible for production of the protein of interest should be expressed by the cells. To introduce the gene of interest, expression vectors are designed that will insert the gene into the host cells and ensure that a suitable transcription factor system is in place. After the gene of interest is cloned into the optimized expression vector, the host cells are transfected for integration of the gene in the host genome. Finally, as not all of these clones will have the desired characteristics, selecting the best clones is the next step by which we can optimize the expression levels of the protein [13]. This is the up-stream process, in which an engineered cell line for producing the protein of interest is created and should be preserved in a cell bank for further use [16].

At the second step of production, the CHO cells prepared during the upstream process are cultured and extended through culture flasks, small bioreactors, and lastly to a large bioreactor setup that has the capacity to support producing enough of the target protein. Production CHO cell lines are mostly adapted to grow in suspended cell culture so that it is easier to grow them in stirred tank bioreactors. The protein of interest is typically released in the culture medium. This step is usually called fermentation or mid-stream; sometimes the first two steps together are called the up-stream process [16].

The down-stream or purification step includes the use of separation techniques such as filtration and different types of chromatography techniques. In this part of the process the desired protein is purified and any impurities such as remnants of bacterial and viral contamination are removed, de-activated or excluded from the product. The result of this step is normally called Active Pharmaceutical Ingredient (API) that can be stored for the formulation and finishing step in which additives such as sugars (sorbitol) are added to the API to gain the proper specification for use as an effective, safe drug in humans [17].

In manufacturing mAbs for therapeutics, the CHO cells and their productivity acquire a special attention. These cell lines are considered suitable for scale up in bioreactors because the well-developed selection and gene amplification systems allow us to produce stable mAb-producing clones. Also by producing glycan structures similar to those of human mAbs, CHO cells are one of the most suitable and preferred host cells

for mAb production [18]. To ascertain the consistency between produced mAbs, accurate screenings of production cell lines have to be performed. These screening steps are costly and time consuming which lead to an expensive final products and delay in introducing the products to the market. Major advances for increasing the CHO cells' productivity have been achieved in operational processes, such as culture media optimization and better feeding strategies [19]. However, further studies are required to improve our understanding of the intrinsic features of these cells. That being said, bioinformatics studies of the cells can help us to understand the specific characteristics that change the productivity of cells.

1.1.4 Development of cell lines with high productivity

In generating CHO cell line that can produce a specific protein, typically clonal selection is performed after transfection of the gene of interest. Some of the most common modifications in developing CHO cell line is to increase the performance in selection through amplification systems using dihydrofolate reductase (DHFR) and glutamine synthetase (GS) genes [20, 21]. The DHFR system is provides methotrexate (MTX) resistance and the GS system confers methionine sulfoxamine (MSX) resistance.

Methotrexate binds to DHFR and stops it from catalyzing folate to tetrahydrofolate. Tetrahydrofolate is a necessary precursor in the synthesis of purines, pyrimidines and glycine [21]. Therefore, if there are low levels of DHFR cells are deprived of nucleosides and die. The CHO cells will be transformed with recombinant DNA including the gene of interest and the DHFR gene. The cells are cultured in increasing levels of MTX for gene amplification, and only CHO cells with high expression of the DHFR gene can grow in this condition. In the GS system, the product of GS enzymatic reaction with glutamate and ammonia is glutamine. Methionine sulfoxamine (MSX) prevents the glutamine production by binding to the GS enzyme. Similar to the DHFR/MTX system, gene amplification happens with subjecting to increasing concentrations of MSX.

1.1.5 Gene expression systems and vectors

Normally, in the development of recombinant CHO cell lines the random integration of a vector containing the gene of interest (GOI) occurs and then in a selection process, cells that carry the GOI will be selected [22]. This random integration of the GOI usually causes the cells to have non-homogenous phenotypes that some of these phenotypes might be unwanted [22]. Usually an extensive cell screening should be taken to find the clone with the desired specification for the protein production. For example in a study 120,000 CHO clones were screened for selecting the healthy ones [23]. To overcome this issue, site specific integration of transgene targeting could be a useful strategy [7, 22]. Genes encoding proteins such as zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeats (CRISPR)-associated (Cas) RNA guided nucleases are common examples of tools for targeted insertion/deletion (indel) mutations or an accurate sequence alterations in a cell [22].

Vectors are DNA molecules that can replicate and we use them for transferring DNA fragments to cells. Vectors derived from mammalian viruses are the most commonly used vectors for expressing heterologous genes in mammalian cells, such as, simian viruses 40 (SV40), polyoma virus, and herpes virus [20]. SV40 or a promoter from cytomegalovirus (CMV) is the most famous vector for transient gene expression with a strong viral promoter [24]. Adenovirus is medium sized (90–100 nm), non-enveloped virus holding a double stranded DNA genome. One advantage of using this virus as a vector is that this virus can be propagated in suspension cell cultures [24]. Vaccinia vectors are other common viral vectors for transferring genes to mammalian cells. Cells infected with the vaccinia virus produces large amount of virus particles per cell (up to 5000), which makes the vaccinia system being able to support high levels of recombinant protein expression. Companies such as Immuno AG have used the vaccinia system for large scale (e.g. ≥ 1000 L reactor) protein production.

1.1.5.1 Lentiviral vectors

A time consuming and costly procedure in the development of CHO cell lines is the cloning and selection of the best-performing clones. One of the approaches for dealing with this issue is transient transfection of vectors into CHO cells. This is a rapid method that has been used for producing recombinant proteins in research or preclinical studies. A group of researchers at the Biotechnology Research Institute of the National Research Council of Canada in Montreal have been working on a new approach based on infection of cells by lentiviral vectors (LVs).

LVs are a type of retroviruses that are more complex and able to infect dividing and non-dividing cells. They incorporate their genome into the chromosome of the cells for stable transgene expression [25]. In 2007, Gaillet et al. developed a new approach for improving the transient expression of proteins in CHO cells. The method is based on the cell infection by viral vectors and the cell line used in this study was made by the system that was developed from this method.

A lentiviral vector includes the gene of interest and required regulatory sequences, also, plasmids that encode enzymes and proteins for making virions are often used for production of LVs [22]. However, the method of co-transfection is costly and difficult to scale up, and production of a large quantity of LVs is easier with packaging cell lines or vector producing cell lines (VPCLs) [26]. Packaging cell lines can stably express all proteins for making virions [25]. In 2008, Broussau et al. created a new inducible packaging cell line that efficiently produces high quantities of the lentiviral vector. As assessed using green fluorescent protein (GFP), the production was able to reach titers greater than $3.0*10^7$TU/ml, though this will vary based on the size and the sequence of the transgene.

1.1.6 Development of Inducible Cells by Transduction System

An important controlling factor of recombinant protein production is transcription promoters and enhancers. For example, the human cytomegalovirus (CMV) is a common promoter used for expression of human recombinant protein and CMV5 is a version of CMV, which carries an enhancer sequence and some other

features such as splicing sites present in the Ad5 MLP [18]. Furthermore, this promoter has the ability for efficient mRNA translation which results in a much higher protein expression as compared to a standard CMV promoter [18]. CMV5 is the optimized version of CMV promoter that is usually used in CHO or other mammalian cell expression systems. The new inducible expression system demonstrated that CHO cells with CR5 promoter are stronger than CMV5 in either of the vectors, adenovirals or lentivirals [26].

In the cumate gene-switch system, the cumate transactivator (cTA) or the reverse cTA (rcTA) binds to the CR5 promoter and initiates transcription (Figure 1.2). Addition of cumate, the small non-toxic molecule, prevents or induces binding of cTA or rcTA to the CR5 promoter, respectively [27]. The capacity of this new system has been evaluated by constructing several LVs carrying different cDNAs.

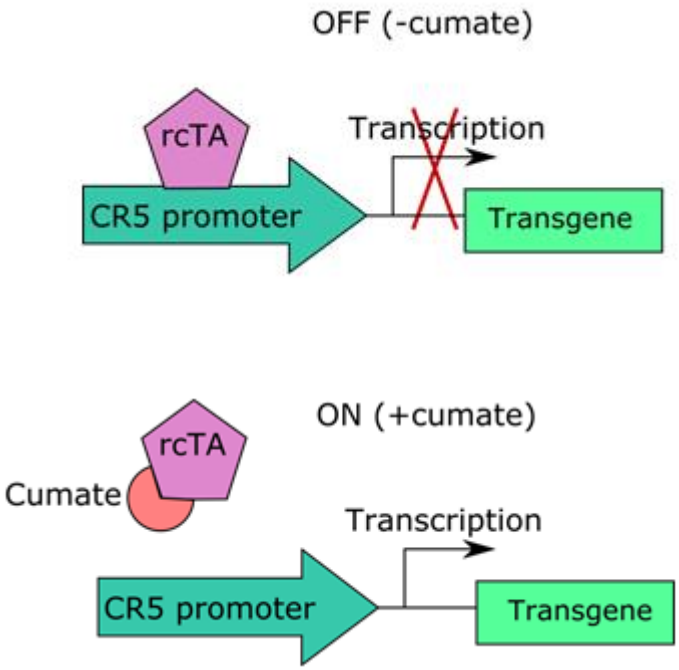

Figure 1.2: The cumate gene switch mechanism. In the reverse activator configuration, with the presence of cumate and binding with reverse cumate transactivator (rcTA), the transcription will be initialized from the CR5 promoter.

1.1.7 Application of OMICs tools to MAb production

OMIC tools are useful techniques for understanding the mechanism of cell growth and protein production at the different level of genome, and proteome, metabolome. Since the introduction of these methods, there have been different studies exploring CHO cell features, however the CHO genome sequence was not available until 2011. Thus researchers had to rely on the available data from similar species such as mouse and rat [28]. In 2011, the genomic sequence of CHO-K1 cells was published, facilitating high-throughput studies on CHO cell lines [8]. In the genomic and transcriptomic study performed by this group, 24,383 predicted genes and 11,099 transcripts were identified in CHO-K1 cell line [29]. Several studies afterwards improved on the identified CHO protein lists, and the collection of this information assists interpretation and brings more confidence in analysis of proteomic results from CHO cells [29, 30].

In a proteomics study by Orellana and colleagues, CHO cells that were producing higher amount of mAb have shown to have higher content of the antioxidant glutathione [31]. Glutathione (GSH) potentially has several benefits in the process of producing high amounts of mAb in a cell. GSH works as a storage component for cysteine, which may be used during protein translation and modification. It can also reduce cysteine disulfides and as the result providing more cysteine available for protein synthesis. In hyperoxidizing environments, correction of protein disulfide bond formation is another beneficial role of GSH [31].

In a metabolome analysis study, Chong et al. [32] explored differences in biological pathways of CHO cell lines producing different levels of mAb. They discovered that the amount of interacellular NAD (nicotinamide adenine dinucleotide) and FAD (flavin adenine dinucleotide) were higher in a high producer cell line compared to a low producer cell line. The high producer cells need more energy for protein synthesis. Mitochondria provides energy in the cells by the oxidation of sugars, fatty acids, and amino acids. NAD and FAD (reduced to NADH and $FADH_2$) carry electrons though the electron transport chain.

The above examples of OMICs studies show that bioinformatics approaches could be instrumental in our understanding of the biologic systems such cell lines use in the production of therapeutic recombinant proteins at high levels. The findings demonstrate that using OMICs approaches we would be able to identify factors leading to increased MAb production in CHO cells.

1.2 Proteomics of colorectal cancer HT-29 cells

1.2.1 Introduction to Human Colorectal Cancer

Human colorectal cancer (CRC) is the third most common cancer worldwide. According to the American cancer society report 2017, it is the second leading cause of cancer death in men and the third in women. During the past decade, the mortality rate of CRC had about 3% decrease each year (American cancer society, CRC facts and figures, 2017). While there have been major developments in cancer therapies and increases in taking screening tests, CRC is still one of the deadliest cancers in the world. One of the important issues in treating patients with metastatic tumors is that cancer cells can become resistant over time to several options for treatment [33].

Tumor excision by surgery is the first option for treatment, but relapse occurs in almost half of the patients after removing localized tumors [33]. The first set of standard chemotherapy treatments involves folinic acid, 5-fluorouracil, and irinotecan or oxaliplatin which both have a broad equivalent effect [34]. In addition, some combinations of these therapeutic agents have shown better overall response and survival rates in patients [34]. In this thesis, the focus is on a CRC cell line that became resistant to treatment by SN-38, the active metabolite of irinotecan.

In the early 1970s, camptothecins (CPT) failed in first phase of clinical trials because of excessive toxicity. However, the interest in camptothecin as an antitumor drug renewed after identification of the molecular mechanism of this action [35]. Camptothecins inhibit DNA synthesis by inducing fragmentation of chromosomal DNA, which is partially reversible. Inhibition of RNA synthesis is also reversible by drug

removal [36]. Potential anticancer agents, topotecan and irinotecan are derived from CPTs and later approved by FDA for treating ovarian and lung cancer, and colorectal cancer respectively [36].

In the 1990s, camptothecin analogs became effective treatments for several malignant cancers, especially colorectal cancer, for which only a few treatment options were available at the time [37]. Irinotecan is a camptothecin derivative (CPT-11) that is biologically inactive. This compound is catalyzed in a carboxylesterase-mediated hydrolysis reaction, which converts it to its active metabolite, SN-38 [38]. In humans, carboxylesterases are present in liver, intestine, and plasma, and they have an important role in drug metabolism [38].

To understand the mechanism of treatment by the inhibitors the role of TOP1 is briefly described. This enzyme cleaves DNA to make a single strand break and stays covalently bound to the strand [39]. This action allows DNA to rotate and relax and then TOP1 ligates the DNA strand and releases the relaxed DNA [39]. Irinotecan or CPT-11 interrupts DNA synthesis by preventing the activity of topoisomerase I and makes protein-linked DNA breaks at replication forks [39]. This single strand breakage leads to a double strand break at replication forks, subsequently activating DNA repair systems. When a cell fails to repair the DNA due to inefficient repair pathways, the cell goes through cell-cycle arrest and apoptosis [35, 39]. Therefore, the effect of TOP1 inhibitors on cancer cells depends on levels and activity of TOP1 enzyme in cells and efficiency of DNA repair process [36].

Camptothecin (CPT) derivatives target the nuclear enzyme DNA topoisomerase I (TOP1). This mechanism has been successfully utilized in cancer treatment. During the S phase of the cell cycle the replication fork will make permanent breaks and cause S phase cell cycle arrest [35]. This cytotoxic process happens in two steps – first, the drug will cause accumulation of cleavage complexes containing reversible DNA single strand breaks, then these complexes will be turned into permanent DNA single and double strand breaks [35].

That being said, the effectiveness of this treatment may be altered by other elements that are involved. There is the potential that tumors can become resistant to the therapeutic agents [40]. Drug resistance is known to be the most common reason in failed cancer therapies, and more than 90% of metastatic colorectal cancer patients have shown resistance to chemotherapy drugs [40].

1.2.2 Therapeutics for colorectal cancer

One of the major advances in treating colorectal cancer patients is irinotecan, along with fluorouracil (FU), used as a single drug or a combination with other drugs it was approved in the 1990s. Irinotecan was valuable for treating patients and has become a standard of treatment [37]. The pharmacokinetics of this class of drugs has been complicated to understand. Like drugs that undergo hepatic disposition, the excretion of SN-38 is happening by glucuronidation; also, SN-38 can be oxidized by the cytochrome P450 family [41]. Glucuronide formation (SN-38G) can also be reversed by bacteria in the intestine. A small amount of the irinotecan will be converted to SN-38 in the patient's body and the excess of drug will be metabolized, excreted by liver or renal transport [41]. The effectiveness of irinotecan may also be influenced by expression levels of other proteins such as transporters and DNA repair proteins, discussed below.

In vitro studies have shown that drug efflux can change via ATP-binding cassette (ABC) transporter proteins, which are involved in passage of drugs through cells and tissues [42]. The expression level of proteins such as P-glycoprotein (P-gp) and multi-drug resistance protein (MRP) are correlated with chemotherapy resistance in cancer cells. Research on these proteins has shown P-gp has very high expression levels in drug-resistant cell lines [43, 44].

TDP1, tyrosyl-DNA phosphodi-esterase 1, is a DNA repair enzyme that in various situations can remove covalent adducts from DNA [45]. Meisenberg et al. (2015) studied the role of TDP1 and TOP1 in cellular response to TOP1-inhibitor treatments. TDP1 removes the TOP1 peptide from TOP1–DNA breaks and this happens at the beginning of the repair process, so it is critical in starting the DNA repair. In addition, this

enzyme has a role in the resolution of DNA damage through TOP1 [46, 47]. Therefore, changes in the TDP1's activity can improve efficacy of treatments by TOP1 targeting chemotherapies [48]. Meisenberg et al. showed that polypeptides of both TOP1 and TDP1 had various expression levels in CRC cell lines and clinical tumor samples [49]. Depletion of TDP1 has outcomes that make TOP1 inhibitors a potential treatment option in colorectal cancer [49]. In the case of TDP1 depletion, DNA breakage increases and irinotecan sensitivity, depending TOP1 levels, can significantly increase [49].

New strategies have been developed to overcome some of the known causes of resistance, such as P-gp inhibitors [43]. Elacridar, tariquidar, and zosuquidar are some of these products that effectively enhance chemotherapy sensitivity in clinical studies [40]. The intrinsic mechanism underlying the drug resistant cells still needs to be explored from different perspectives. The knowledge on the cellular mechanism of resistance in cancer cells could assist physicians to decide whether TOP1 inhibitors or a combination with other inhibitors might be beneficial in treating patients based on the features of their tumors. Incorporation of tumor-specific features into treatment plans could have a substantial effect in increasing survival rate.

1.2.3 HT-29 cell line background

Colorectal cancer tumors can be categorized into two groups based on the type of genomic instability. The majority of CRC patients' tumors have chromosome instability (CIN), where chromosomes might be duplicated or deleted completely or partially, so number or structure of chromosomes is abnormal. In some cases, tumors have microsatellite instability (MSI) at the nucleotide level, which is caused by impaired DNA mismatch repair [50]. The HT-29 cell line with chromosomal instability and HCT-116 cell line with microsatellite instability are the most common cell lines to be used in research studies.

Researchers in Dr. Blay's lab at the University of Waterloo have generated a derivative of HT-29 cell line by treating with the irinotecan active metabolite, SN-38, that is SN-38 resistant and studied dynamics of the cell cycle as they become resistant to the drug [51].

1.2.4 Classical resistance mechanisms

A drugs' efficacy depends on the drugs accumulation in target tissues, which could be limited by how much it enters or leaves cells. The absorption and accumulation of substances with potential toxicity is limited by active efflux mechanisms in the plasma membrane. Drug resistance in cancer cells could occur through these efflux mechanisms, by up-regulation of proteins, such as P-glygoprotein (P-gp, MDR1) and multi-drug resistance protein, MRP2. These efflux pump proteins transport compounds (hydrophillic or charged) from cell's cytosol to extracellular matrix or into lysosomes. By removing drug substances, efflux proteins can increase resistance to drugs and therefore reduce efficacy. This mechanism of drug resistance has been shown *in vitro* by analysis of resistant cell lines and clinically in patients at the time of failed treatments and relapse. For example, the RNA levels of BCRP were significantly over-expressed in leukemia patients between the diagnosis time and the relapse time at a refractory state in comparison to the diagnosis time [52].

Expression of drug efflux pumps are associated with irinotecan resistance, ABC transporter superfamily proteins ABCG2/BCRP and ABCB1/p-glycoprotein (Figure 1.3) [53, 54]. CRC cells resistant to SN-38 had significantly over-expressed ABCG2 protein [55]. Physiological characteristics of cells, such as growth and dynamics of cell cycle, have also shown changes in SN-38-resistant CRC cells. SN-38 resistant cells showed a slower growth, the S phase fraction of cell cycle was reduced, while the G2/M phase showed an increase [56]. One of the results of a longer G2 phase is the accumulation of cyclin B1 in the cytoplasm during this phase. There is also confirmation of this result *in vivo*, with observed slower growth and reduced DNA synthesis.

In addition, genetic polymorphism in transporters can change the drugs metabolism and dis-positioning, therefore individuals can show variable pharmacokinetics [39]. These chemo-refractory tumors are active in adapting and advance with alternative drug-resistance programs; as a result, ABC transporter inhibitors have typically not had promising results [57].

Overall, changes in the efflux and influx of the molecules out of and into the cells are the most important factors in the drug resistance process. In addition, in the irinotecan treatment activation of DNA repair mechanism and drug detoxifying proteins, and disruption in signaling pathways for apoptosis (Figure 1.4). Here, we are interested to examine the proteome of our SN-38-resistant colorectal cancer cells to identify the components involved in the resistance mechanism.

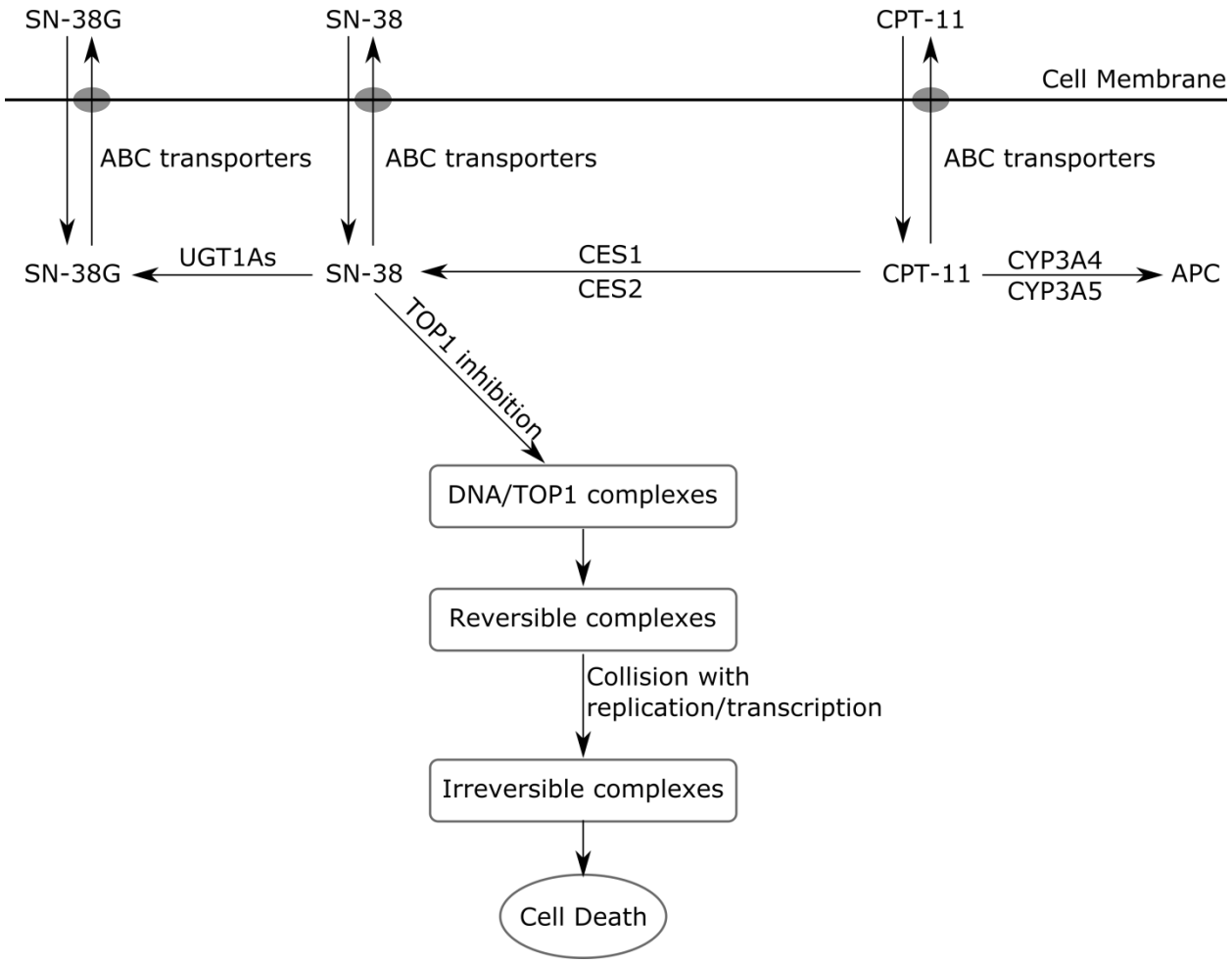

Figure 1.3: An overview of the pathway of irinotecan disposition in and out of cells [58].

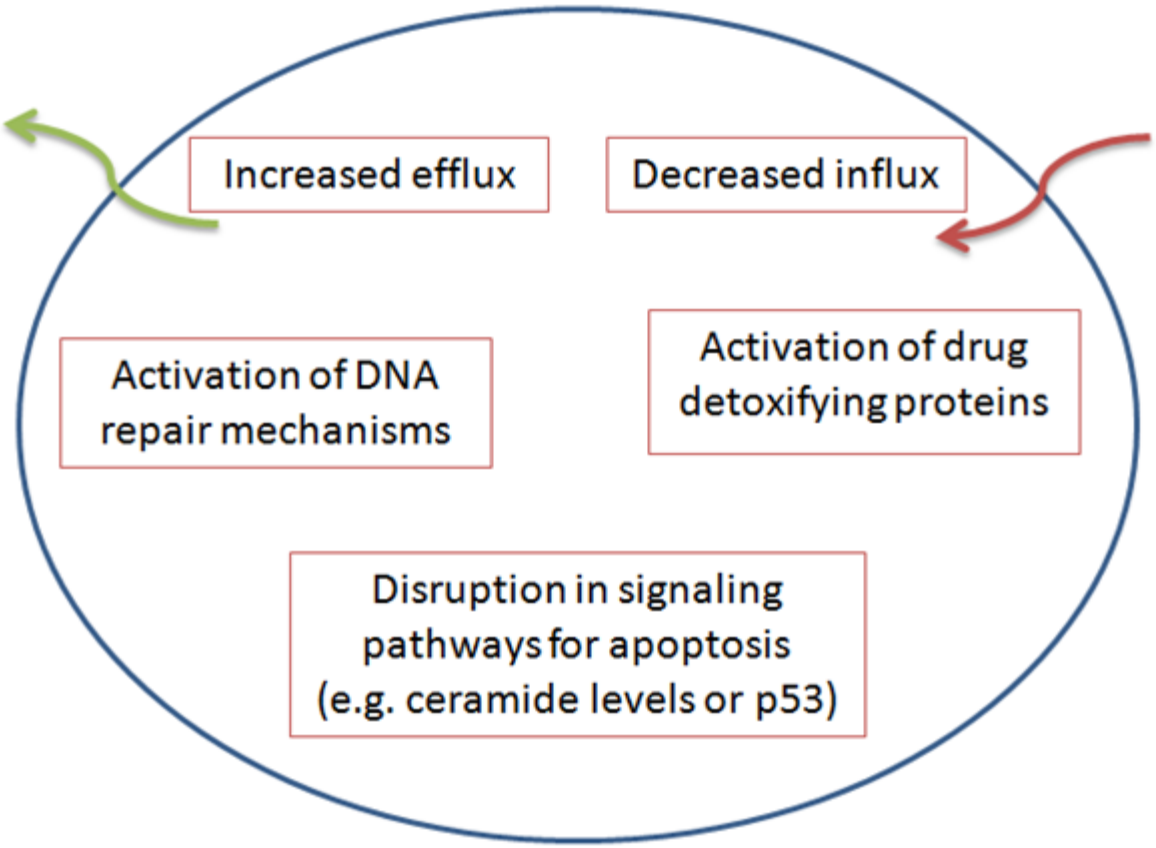

Figure 1.4: Some of the cellular factors that are involved in drug resistance mechanisms.

1.2.5 The role of fibronectin and integrins

Fibronectin was discovered in 1973, as a surface antigen protein that is produced by fibroblasts in chicken [59]. In another publication, it was identified as a large extracellular protein in hamster and human fibroblast cells by experiments using radio-labeling of surface proteins [59, 60]. Later, the nature of these extracellular proteins and actin microfilaments was understood to be functionally linked together. Now we know that the integrins are a family of cell surface receptors that make the link between extracellular proteins and cytoskeleton [60, 61].

There are 18 integrin α subunits known in humans. The integrin subunits, α and β, form heterodimers and make receptors with unique structures that have particular functions. By attaching cells to the ECM substrates, these proteins are providing information about the extracellular environment to the cell as well as returning signaling information [62]. This connection can regulate intracellular signaling pathways and behavior of cells. Also, with these cell interactions information from a cell could be transmitted to the ECM environment. Cells can change integrin dimer structures that alter the similarity of the receptor and ligands [62, 63].

When they have low similarity they are in the inactive state in which intra-membrane bonds keep β-subunit in a stable tilted position towards α-subunit, whereas the active receptor forms in the high affinity of the two, when β-subunit is shifted away from the α-subunit [62]. The integrins' active or inactive state can be affected by some adaptor proteins that are in interaction with the β-integrins [62, 64]. The expression level of proteins such as kindling kinases, talin, alpha-actinin, and filamin could change the activation process of integrins [64]. In conclusion, the activity of integrins can be influenced from cell signaling as much as it can influence cells' behavior.

Many of the commonly known functions of integrins are among the important functions that help tumors grow aggressively. For example, integrins can promote cell migration and survival through interactions with the ECM ligands. It has been shown that integrin receptors α2β1 and α5β1 have a role in promoting

colorectal cancer metastasis to the liver, mediated through their interactions with collagen and fibronectin [65, 66]. However, some integrins can function as a metastasis promoter or an anti-cancer substance, or both in different conditions. For example, α5β1 can also promote cancer invasion and metastasis by secreting matrix metaloproteinases (MMPs) and this process can be prevented by another integrin, α4β1, which also binds to fibronectins at different site [67, 68].

Among all of the positive or negative effects that integrins can have on cancer cells, the effects that interactions of integrin and ECM have on cell survival can have very important role in the resistance of cancer cells. Anoikis is a form of apoptosis that happens in epithelial cells when they become detached from the ECM [69]. Therefore, anoikis that occurs by signaling provided from interactions between integrin and ECM can stop tumors from expanding cancer and spreading the cancer cells to other tissues in the body. However, cancer cells could reach a protection against other stimulus and find an opportunity to have adhesion to ECM and resist the toxicity of chemotherapy agents with cell adhesion-mediated drug resistance [70]. For example, it has been shown that resistance to chemotherapy drugs increases when multiple myeloma cells are attached to fibronectins [70]. In a study on acute myeloid leukemia (AML) cells with exogenous fibronectin showed more resistance to a chemotherapy drug through a PI3K/GSK3β pathway that is recognized with the α4β1 and α5β1integrins [71]. In this case, drug efflux pumps were not involved in the resistance process of AML cells. AML patients with α4β1-negative tumors had higher survival rate without cancer than α4β1-positive patients [71]. In both *in vitro* and *in vivo* xenograft experiments, it was demonstrated that SN-38 resistant cells had a slower doubling time, a shorter S phase and a longer G2 phase fraction [56]. These changes to the cell cycle could affect tumor progression and resistance to other antitumor treatments.

1.3 Objectives and hypothesis

The goal of first part of this thesis is to apply proteomic techniques in order to identify differences in the proteome of an inducible CHO cell line that can produce higher amounts of mAb with cumate. To distinguish between these two cell lines, they are termed low-producer and high-producer cells. In this thesis, the inducible CHO cells constructed with the vector containing cDNA of the mAb EG2-hFC1 produced high levels of mAb with the presence of cumate and very low levels of mAb without cumate. A proteomics approach can reveal the possible stresses or cell responses at the protein level when the production of a recombinant protein (mAb) is high. This type of analysis can suggest potential targets for optimizing the cell inducing technique to gain optimized mAb production.

The second topic of interest is drug-resistance in human colorectal cancer. A derivative of HT-29 cell line with SN-38 resistance was generated by prolonged exposure to SN-38, which is the active metabolite of the chemotherapeutic drug irinotecan. The iTRAQ proteomics method was similarly applied to explore the changes in the cells that became resistant to the treatment. By exploring the cell proteome, our goal is to determine the key pathways and metabolic features that change in CRC cells that become resistant. This can help to better elucidate the process by which cancer cells undergo in a path to drug resistance.

The first chapter of this thesis includes an introduction to the CHO and HT-29 cell lines. Background information on the proteomics technique used in this thesis and the methods applied in data analysis are described in chapter two, and the results of the analyses are reported and discussed in chapters three and four. In this thesis, I report and discuss the proteomic experiments and results of HT-29 cells in comparison to HT-29S cells with induced resistance to SN-38. The objective is a quantitative comparison of the whole cells' proteome to discover differentially expressed proteins to understand their role in the development of SN-38 resistance in an HT-29 cell line.

Chapter 2

Quantitative Proteomics Methods

2.1 Introduction

In this thesis, quantitative proteomic technique by iTRAQ labeling reagents is used to quantitatively compare target samples with a set of control samples to understand the intrinsic features of the sample under the desired conditions. In one part, we have the inducible CHO cell line CHO-1A7 expressing the monoclonal antibody EG2-hFc1 that is generated using lentiviral vectors and the cumate gene-switch; in the second analysis, we have the colorectal cancer cell line HT-29 and HT-29S that is resistant to the chemotherapy treatment SN-38. In summary, Figure 2.1 shows the workflow of an iTRAQ experiment consisting of the sample preparation, labeling steps, and mass spectrometry analysis.

2.2 iTRAQ Labeling

The iTRAQ technique has been developed to compare proteins of up to 8 different complex samples. Samples can be labeled by iTRAQ reagents and the mixture of labeled proteins can be analyzed simultaneously by mass spectrometry methods [72]. This proteomics method enables identification and quantitation of peptides/proteins in addition to their post-translational modifications of the samples mixture [72].

For an iTRAQ experiment four or eight reagents can be used to tag peptide samples. The structure of the reagent labels (Figure 2.2) consists of a uniquely charged reporter group that is identifiable in mass spectrum, a peptide reactive group that interacts with primary amines, e.g. lysines and peptide N-termini, and a neutral balance portion that maintains the overall mass of the reagent molecule and falls off in the fragmentation step [72].

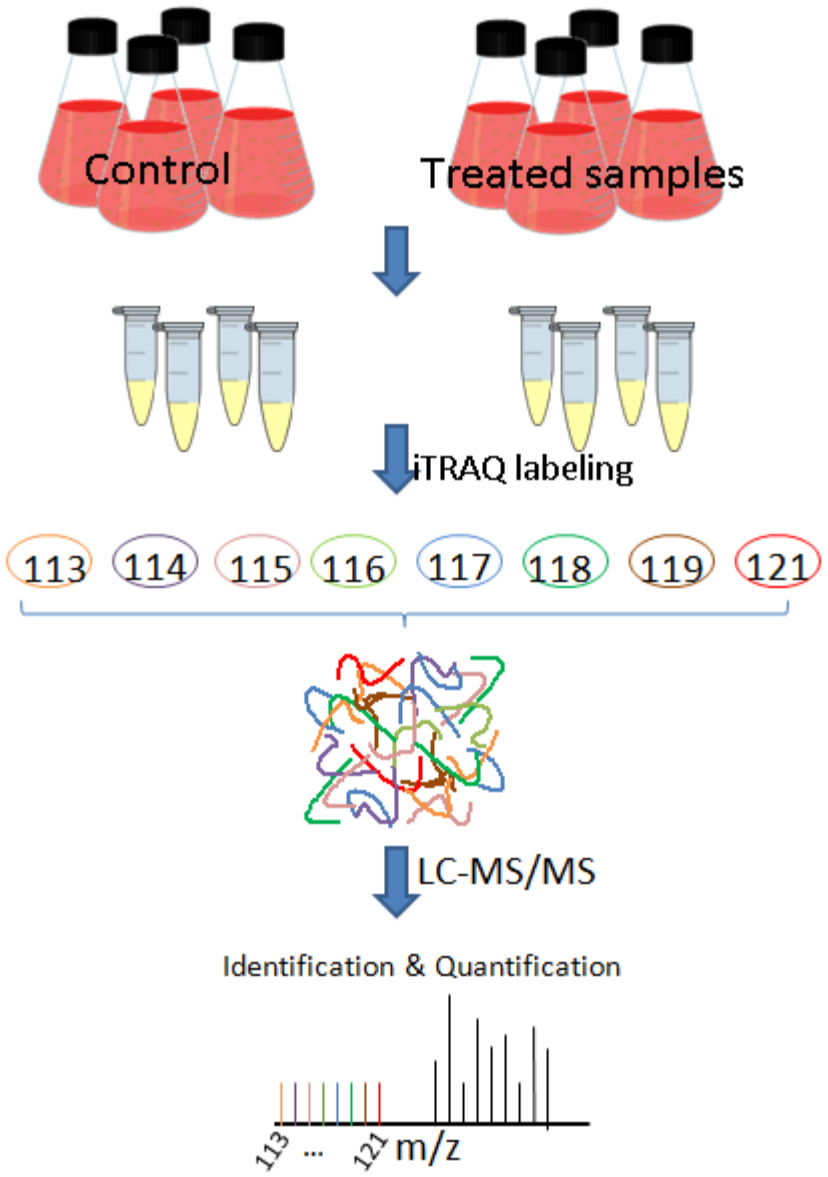

Figure 2.1: The workflow of sample preparation for an iTRAQ experiment. Samples are four control and four treated replicates of cells. After protein extraction and digestion following the filter-aided sample preparation method, samples are labeled with 8 iTRAQ reagents and the peptide pool is prepared for mass spectrometry. Finally, data analysis is done for identification and quantitation of the proteins.

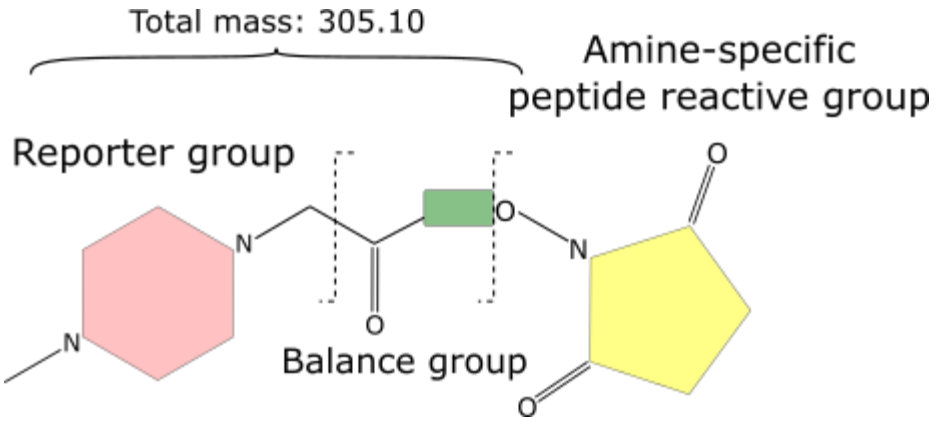

Figure 2.2: **The iTRAQ reagent structure.** Four iTRAQ 4-plex reagents have the same total mass of 145 Da with reporter group mass of 114-117 Da and balance group with mass 31-28 Da [73]. The total mass of isobaric tags among all 8-plex reagents is 305 Da, reporter groups have 113-119 and 121 Da mass and balance groups have a mass of 192-186, and 184 Da.

In an iTRAQ experiment, equal amounts of total peptide from each sample is labeled by 8 labeling agents according to iTRAQ manufacturer's protocol. To label the samples efficiently, pH has to be adjusted and kept between 7.5 and 8.5. The labeled peptide samples are mixed and prepared for mass spectrometry analysis by a peptide clean-up step. Then, the sample will be sent to a mass spectrometry facility for analysis and generating MS raw data that will be used in further data analysis.

2.2.4 Mass spectrometry data analysis

Samples consisting of pooled, iTRAQ-labeled peptides were analyzed using a Q-Exactive mass spectrometer at the SPARC BioCentre (SickKids hospital, Toronto). Raw MS and MS/MS data includes the data from ionized peptides based on their mass to charge ratio (m/z) and the masses of ions made from fragmented peptides on the spectra. In order to identify peptide sequences from the complex mass spectra, expected ionization and fragmentation patterns for peptide sequences can be identified and compared using MS software. The fragment ions observed in the spectra are made based on the peptide sequences, the ion's internal energy, and the charge. The mechanism of fragmentation of protonated peptides is well known. The fragment ions that carry the charge on N-terminal are called a, b, or c, and when the charge is on the C-terminal, they are named as x, y, or z (Figure 2.3) [74]. Ions b and y are the most stable ones appearing as peaks in the spectra. These peaks represent specific masses in the spectra and can be searched through protein databases with predicted masses of fragmented peptides to identify peptide sequences in the samples.

The data produced by a combination of LC and MS/MS can be seen as thousands of points represented by retention time (RT), intensity, and mass to charge ratio (m/z). The process of inferring proteins sequences from the MS/MS data can be done by several software packages. PEAKS Studio 7.5 (Bioinformatics Solutions Inc., Waterloo, ON, Canada) is a powerful software tool that incorporates a *de novo* sequencing method into its database search algorithm. This algorithm achieves higher accuracy and sensitivity, and better time efficiency than other common software packages [75]. PEAKS Studio 7.5 is used for

identification and quantification of proteins in the mass spectrometry data. The general workflow of MS data analysis in PEAKS is represented in Figure 2.4.

PEAKS includes a *de novo* sequencing algorithm that can identify proteins independently from the sequence database. *De novo* means from the beginning in general usage; it means making computational predictions using a model without incorporating any of the existing data. *De novo* peptide sequencing is allocation of fragment ions from the tandem mass spectrum. This *de novo* algorithm finds the best set of amino acid sequences among all of the different possible arrangements, and allows the software to cover the new peptides not included in databases of known proteins. PEAKS performs these computations efficiently using an advanced dynamic programming algorithm [75]. The *de novo* algorithm determines a score for each peptide sequence that represents the best match for the b and y peaks in the spectrum [75]. Then, the algorithm reports the distribution of average local confidence scores for each identified peptide. The average local confidence is the total local confidence divided by the number of amino acids in the peptide sequence, independent of peptides' size.

The target-decoy method is widely used to estimate the identification quality as a false discovery rate. The software searches the concatenation of a target and a decoy database that have the same sizes. The false search results should become evenly distributed in target and decoy, when the decoy is constructed correctly. We can estimate the FDR since all the decoys are false, FDR will be estimated as the ratio of the decoy hits over the target hits.

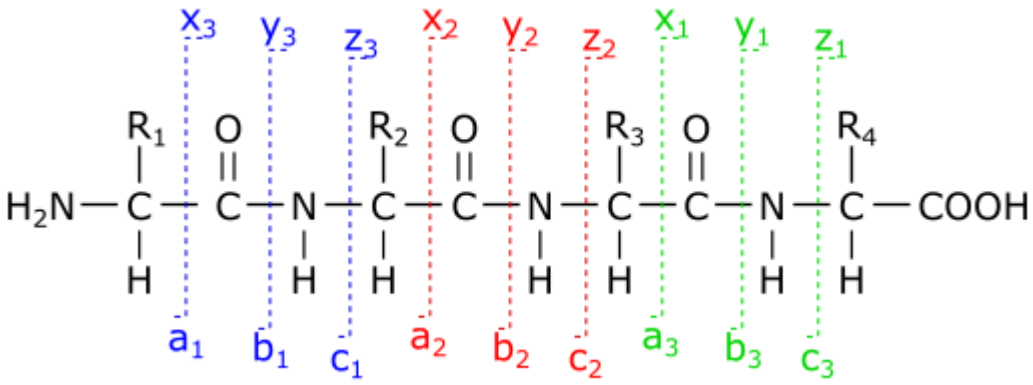

Figure 2.3: Fragmentation ions in peptide backbone. In peptide backbone fragmentation a, b, and c ions extend from the amino terminus, and x, y, and z ions extend from the carboxyl terminus [76].

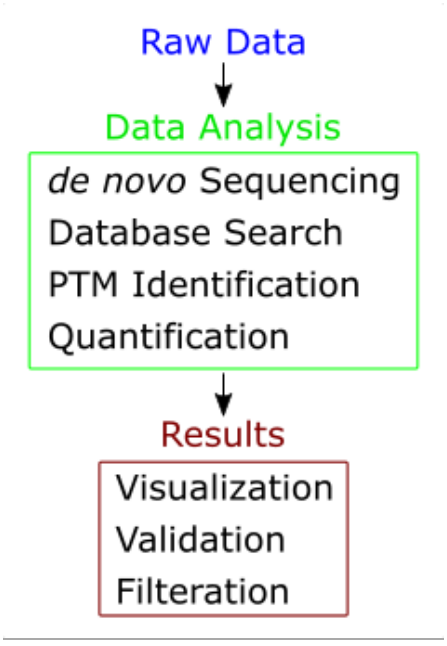

Figure 2.4: The overall workflow of raw MS data analysis using PEAKS

(http://www.bioinfor.com/peaks-studio)

PEAKS database search tool (PEAKS DB) incorporates *de novo* sequencing results into the database search to optimally explain the spectrum. Therefore, PEAKS DB works efficiently by limiting the search items with shortlists of proteins and peptides based on de novo results. The search scoring function finds the similarities between *de novo* results and the database; it also normalizes them to be comparable across the spectra. Then it confirms the results with a false discovery rate (FDR) threshold chosen by users, such as 1% false positive identifications. Finally, using peptides with the highest confidence score, it infers and groups protein sequences [77].

The iTRAQ method from PEAKS Q module was used for protein quantification analysis. This module detects features of the mass spectrometry data and calculates each labeled peptide ratio based on peak areas in the spectra. The algorithm uses total intensity of the top-3 unique peptides for each protein and consequently makes a clustered set of proteins with the detected raw intensities and ratios.

2.2.5 Differential expression analysis

Data normalization is necessary to ensure bias-free comparison of the intensities across different labeled samples. Several normalization methods that are not based on internal standards have been used for MS data, including quantile normalization, lowess, and global scaling [78]. Quantile normalization is a frequently used algorithm with this type of data, which sorts and scales the data to match a certain quantile among samples.

Quantile has a simple and quick algorithm that uses the complete raw intensities data, to create the normalization relations and could be applied to the data using R bioconductor packages, e.g. “limma” or “preprocessCore” [79, 80]. This normalization method aligns the distribution of intensities for each column across the dataset, or equivalently across samples so that they are similar. The idea is that the quantile-quantile plot shows the distribution of two sets of data as a straight diagonal line. It can be extended to n vectors of data to have the same distribution. In this dataset there are 8 groups of data from each samples labeled with a unique iTRAQ reagent. To transform the quantiles, projection of quantiles onto the unit

diagonal should be performed. This procedure is taking the mean quantile and replacing it as the value of the data in original raw dataset [79].

After normalizing the data from the eight labeled samples, variation between data from the two conditions should be explained. If the data cannot be explained by random variation, we conclude that that the two distributions have some significant differences present and it is not just through random chance. The test of significance is a method of statistical inference performed to accept or reject a claim based on sample data. In order to compare the abundances of proteins of the two cell lines, a hypothesis test should be implemented. An unpaired two sample t-test can be performed, testing the null hypothesis states that the averages of the intensities of samples are equal to each other. The test generates a p-value that shows the probability of the sample mean difference in protein expression of the two groups would be the same as or larger than the actual observation, when the null hypothesis is true (there is no change between the groups). Therefore, the lower the p-values are, the less likely that the mean difference occurred by chance alone and the difference is significant.

The result can be one of the following four situations, when a null hypothesis is true and test has not rejected (not significant), when the null is true but t-test rejected it (type I error), when a null hypothesis is false and test rejected the null (significant), and it is false but the test has not rejected the null (type II error). To generate meaningful results, the statistical power is a key prerequisite in a successful quantitative study. When we have β as the false-negative rate, the statistical power is $1-\beta$. This is the probability that the result is a type II error [81].

There are multiple comparisons and null hypothesis tests stating that the averages of the intensities of each sample group are equal to each other. By testing each hypotheses separately with a significance level $\alpha=0.05$, commonly chosen for two standard deviations from the mean, there is a high chance of error in several hypothesis tests. The probability of observing a significant result by only chance can be high when n is a large number:

$P = 1-(1- \alpha)^n \approx n\alpha$

A multiple testing correction procedure is required to adjust a suitable statistical confidence measures based on the number of hypothesis tests. One way is decreasing the error threshold for each test by the number of tests to have an overall α with the desired limit. This correction, the Bonferroni technique, requires α/n for each test to result in an overall of α allowable error for n number of comparisons. This method is too strict for high-throughput experimental data and has been replaced by Benjamini-Hochberg FDR method or similar techniques to control the false discovery rate [82].

The False Discovery Rate (FDR) is the expected proportion of false positives within the significant tests and estimates the rejection region. Benjamini-Hochberg method for a given q, say q = 0.05 or q = 0.10, calculates linear step-up values as expected to be the significant p-values. While the p-value is lower than the expected value the tests are rejected. The Benjamini-Hochberg procedure is as follows,

$P_i : P_1 \leq P_2 \leq \ldots \leq P_m$, (m: number of tests)

$k = \text{Max}(i, \text{such that } P_i \leq \frac{i}{m} q)$

The procedure rejects null hypotheses P_i to P_k, produces a list of significant tests.

Since there have been many studies regarding the problem of finding differentially expressed genes in microarray experiments, many resources are available for the statistical analysis. R programming packages R stats, multtest, and LIMMA library (Linear Models for Microarray Data, v 3.22.1) [80] of the bioconductor package [83] are used for statistical analysis.

2.2.6 Functional analysis of differentially expressed proteins

After analysis and comparing the protein expressions with a suitable significance level, a list of proteins will be identified with significant differential expressions. Therefore, studying their cellular functions and molecular mechanisms is essential. For this purpose bioinformatics tools and databases, such as COG, KEGG, DAVID, and GO have been developed that help in functional analysis of lists of genes and proteins.

This level of analysis is similar to genomics data annotation, so gene ontology tools can be useful. The database for annotation, visualization, and integrated discovery, DAVID, [84] is an integrated analytical and biological knowledge-based tool is used to extract biological meanings from gene or protein lists. DAVID can provide enrichment categories of genes in the list.

DAVID is a web-based bioinformatics tool available at http://david.abcc.ncifcrf.gov/ is developed to annotate large gene lists into meaningful biological information [85, 86]. It is designed to map the gene list to their associated biological annotation or gene ontology terms, and to identify those genes that are enriched (overrepresented) [85]. In addition, DAVID clusters related terms into groups to facilitate the issue of redundancy in biological annotation terms [86]. The output includes a *p*-value calculated by a modified Fisher's exact test (EASE score), which shows the significance of enrichment of terms. The threshold for enrichment score is 1.3, which is equivalent to 0.05 [85].

The list of genes corresponding to differentially expressed proteins is subjected to DAVID's analytical modules to identify functional groups within enriched genes. We can explore possible links between function and the cells productivity, and more thoroughly understand the cells response to the treatment. In case that the cells were under a stress or had an abnormal condition, genes that relate to the process in the cells would be identified in the enriched groups. Then, the enrichment score of gene groups is used for ranking the importance of annotation terms [85]. Using DAVID 6.8 analytical tools, we can identify the gene enrichment in the biological process and functions in gene ontology, and the KEGG pathway database most relevant biological terms associated with the gene list.

Chapter 3

Comparative proteomic analysis of CHO cells

3.1 Introduction

Chinese hamster ovary (CHO) cells are a suitable host for production of therapeutic mAbs that are complex multimeric and glycosylated proteins [7]. Typically, a population of recombinant CHO cells clones with a range of specific productivity is generated by transfection of cells with the gene of interest, followed by clonal selection, and amplification. This process for having the best-performing clones is time-consuming (taking 4 to 8 months) and costly. Expression systems using strong viral promoters helps in improving mAb production. The inducible CHO cell line in this study was developed using CR5 promoter in lentiviral vectors [26] with cumate gene switch for inducing high levels of mAb secretion [26, 27]. This mAb product is used in treating a variety of cancers because it targets cells that express epidermal growth factor receptor (EGFR). EGFR is a tyrosine kinase receptor that plays a role in the control of important cellular processes such as angiogenesis, cell growth and proliferation. Various types of human cancer cells have mutations in the EGFR gene [87].

Transduction of cell lines with lentiviral vectors is a gene expression system for optimizing the production of recombinant antibodies [26]. One of the key elements of a good expression system is the use of a strong promoter. An inducible system with the cumate gene switch promoter (CR5) with the cumate transactivator (cTA) and the reverse cTA promotes expression of the gene of interest (explained in details in section 1.1.6) [26, 27].

We used a proteomics approach in this thesis to investigate the possible stress or cell responses in the inducible cells when they produce high amounts of MAb. Using the iTRAQ proteomics technique, we quantitatively compare the cells at the protein expression level. The CHO cell line was generated by transducing the cells with lentiviral vectors including cumate gene switch promoter elements in Dr.

Durocher`s lab. By adding the cumate treatment these cells were able to have close to 6 fold change increase in mAb production. The proteomic analysis helps us to compare protein expression within induced and uninduced cells and find possible targets for optimizing the technique for optimizing mAb production.

3.2 Methods

Four control replicates of inducible cell line were cultured, and another four replicates were treated with cumate to increase mAb production. The collected cells were subjected to cell lysis, protein extraction and digestion. These eight peptide samples were labeled using 8-plex iTRAQ isobaric labels for differential expression analysis. Mass spectrometry (MS) was performed and I used the generated MS data for protein identification and quantitation, and statistical analysis of protein expression. In this chapter, the results of the data analysis is presented and discussed.

Using the iTRAQ method, protein expression levels of eight samples can be compared in one mass spectrometry run. The iTRAQ reagents label lysine residues and all free amino terminal groups of peptides via covalently bound tags. In the CID stage of tandem mass spectrometry, the balance group of the tags is lost, and in the analysis of the peptide fragmentation (MS/MS) the reporter groups are detected. The MS/MS spectra includes tags from all samples containing the given peptide, and contributions of each sample is measured by the reporter ion peak's intensity. In this study, the spectrum is analyzed using PEAKS Studio 7.5 and then peptides and proteins are quantified to identify the differentially expressed proteins.

3.2.1 Cell culture

The cell lines were developed using a transduction system by lentiviral vectors coupled with the cumate gene-switch [26]. Transduced CHO samples were obtained after treating with cumate and a control set without treatment. The mAb titer in the culture media was measured at the harvest time and cells were frozen at -80°C prior to iTRAQ experiment. Protein samples were extracted from whole cell lysates, then proteins were digested and peptides were labeled using iTRAQ 8-plex reagents.

Cell lines were kindly provided by Dr. Yves Durocher at the Biotechnology Research Institute (BRI) (Montréal, Canada). The CHO-1A7 cell line was engineered for the stable expression of the chimeric heavy chain antibody EG2-hFc1. The process of the cell culture and induction was contributed by Adeline Poulain from Dr. Yves Durocher's lab [88]. The replicates of cell line that were developed from a stable transduction LV system, were cultured in suspension with working volumes of 250ml shaker flasks. Cells were incubated at 5% CO2 in a humid environment at 32° Celsius. Four of these batches were treated with 2ug/ml cumate 5 days post-transduction (6-9 million cells/ml). The cumate induced CHO cells and control samples were harvested at the exponential phase of growth 5 days after the treatment, and the monoclonal antibody product of each replicate was measured. Cell pellets were stored at -80°C prior to proteomic analysis.

3.2.2 Protein sample preparation and labeling

In mass spectrometry experiments, many detergents and common chemicals are incompatible with instruments and can introduce contamination that will distort the results. Polyethyleneglycol from plastic ware and keratins from skin are the most abundant contaminants of protein samples. To prevent contamination procedures were done in biosafety cabinet. In addition, in all steps of the procedure HPLC-grade solvents were used and all pipette tips and tubes were rinsed twice with 100% methanol and let dry overnight in the biosafety cabinet.

Whole cell protein extraction from cell pellets that were stored at -80°C was started after quickly thawing the samples and lysing them. Lysis buffer constitutes of 1% Triton X-100, 100 mM HEPES (238.3 g/mol) adjusted to pH 7.5 by NaOH/HCl, 50 mM tris(2-carboxyethyl)phosphine (TCEP from iTRAQ kit), and 5 ul protease inhibitor cocktail (Sigma, P8340) was prepared and for each sample 700 ul of the lysis buffer was used for 30 min at 4°C. Once the cells are disrupted proteases can degrade proteins; therefore protease inhibitors must be in the lysis buffer. Also, TCEP should be added in the lysis buffer as the reducing agent as this agent is necessary for cleaving disulfide bonds, helping to keep proteins stable and to prevent protein aggregates. TCEP replaces the commonly used reducing agent dithiothreitol (DTT) with several advantages [89], such as a wider effective pH range, more hydrophilic characteristics, stability and reactivity for a

longer time at room temperature as it is more resistant to oxidation in air. These features ensure the effectiveness of the reducing agent on the protein samples.

Next, the cell lysates were centrifuged at 10,000x g to extract the supernatant. After each step during the protocol the protein concentration was measured by Bio-Rad microplate protein assay (Bio-Rad Laboratories Ltd., Mississauga, Canada). This assay is based on Bradford method [90], which requires constructing a standard curve at each protein assay run. Bovine serum albumin (BSA) (Sigma, ON, Canada) with 0.0, 0.2, 0.4, 0.6, 0.8, and 1.0 mg/ml concentration was used for the linear standard curve. The corresponding spectrophotometer absorbance was read at 595 nm by a Multiskan UV microplate spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). The protein samples were diluted at each step to prevent any interference of the buffers with the Bradford agent and to increase sensitivity of the measurement at high protein concentrations.

Following the measurement of protein concentration, from each sample 1 mg of protein lysates were precipitated using a Calbiochem ProteoExtract kit (Cat# 539180, EMD Millipore). The precipitation agent was prepared according to the kit's instructions and the wash solution was prepared by adding 150ml high quality ethanol to the bottle of wash solution provided in the kit. These agents were stored and chilled at -20°C prior to precipitating the samples. From each sample, 1 mg of the protein lysate was precipitated by addition of a 4x volume of the precipitant agent following the manufacturer's protocol. Samples were incubated at -20°C for 60 minutes, and then proteins were pelleted by centrifugation at 10,000x g at room temperature and the supernatant was discarded. Following the Calbiochem protocol, pellets of proteins were washed 2 times with 500µl of cold (-20°C) wash solution.

Subsequently, the precipitated protein samples were re-suspended in UA buffer, containing 8M Urea and 100mM HEPES at pH 8.6. At this step, another Bradford microplate assay was performed to measure the concentration of solubilized proteins. From each sample 250µg of completely solubilized proteins was used in filter-aided protein sample preparation procedure [91].

In this study, after testing two 3k and 10k filters with the protein samples, for the best results I chose to use Amicon Ultra centrifugal filters with relative molecular mass cut-off of 3k. In the filter-aided sample preparation the whole proteome is completely solubilized in chaotropic detergents, then the buffer is exchanged in the filter purification step [92].

Next, proteins were alkylated to block cysteines by adding 50mM iodoacetamide (IAM) to the UA buffer. Alkylation was performed with a limited quantity of IAM and the buffer was slightly alkaline with pH of approximately 8 to make sure that is not affecting residues other than cysteines. Samples were in 0.5 M triethylammonium bicarbonate pH 8.5 (TEAB) buffer prior to digestion with lyophilized trypsin. In-filter digestion was performed using 1:50 trypsin to protein ratio. Samples were trypsinized for approximately 8 hours at 37 °C. For deactivating trypsin's activity and concentrating the peptides, samples were dried using Thermo Scientific Savant Speed Vac vacuum centrifuge and re-suspended in 30uL TEAB. Finally, prior to labeling, peptide concentration of samples was estimated using the A280 method [93] on a NanoDrop™ 2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA).

Equal amounts of peptide from each sample were labeled using the eight labeling agents according to iTRAQ manufacturer's protocol. To label the samples efficiently, pH was verified using Hydrion pH papers and adjusted to be between 7.5 and 8.5. The labeled peptide samples were combined into one tube and dried using vacuum centrifuge in a Speed Vac® from Savant (Thermo Fisher Scientific, Waltham, MA, USA). Then, the dried sample was resuspended in the ACN buffer, comprised of 300ul of 5% acetonitrile, 0.1% formic acid for peptide clean up step.

Peptide clean up was performed using Empore™ C18 solid phase extraction cartridge (4115SD) according to the cartridge column's protocol. The washing step was done by adding deionized water and 0.1% formic acid solution to the sample and letting it pass through the particle bed of the C18 cartridge by gravity. After 3 times repeating the washing step, sample was eluted in two steps of adding 200ul buffer of 70% acetonitrile and 0.1% formic acid to the cartridge. Peptide concentration was estimated using

Nanodrop, and mass spectrometry analysis was performed at the mass spectrometry facility of SPARC BioCentre at the SickKids hospital, Toronto, as described in chapter 2.

3.2.3 Mass spectrometry data analysis

The raw mass spectrometry data is processed by PEAKS Studio. First, the de novo analysis of the software identifies the peptide sequences, and using the results the PEAKS database search tool searches through the database that we constructed for the dataset. After filtering the results of protein identification analysis, PEAKS quantitation tool is used to calculate the expression data for the proteins. The differential expression analysis of the dataset is performed using R packages stats within the CRAN project (R Core Team, https://cran.r-project.org/) and limma within the Bioconductor project (https://bioconductor.org/packages/release/bioc/html/limma.html) [80, 94].

Known post translational modifications were selected in PEAKS to assist the accurate identification of sequences. Carbamidomethylation of cysteines (+57.02 Da) and iTRAQ modified N-terminal and lysine-side chain of tryptic peptides (+304.2 Da) were selected as fixed modifications; oxidation of methionine (+15.99 Da) and iTRAQ modified tyrosine (+304.2 Da) were set as variable modifications in the analysis. Only 1 missed cleavage was allowed, majority of peptides were identified with no missed cleavage.

3.3 Results

Following four days of cumate treatment cells were harvested. Corresponding EG2 quantification data was provided by Dr. Durocher, Biotechnology Research Institute (BRI), Montréal, Canada. The expression level of mAb in the induced samples was increased by 5.7 fold change on average (Table 3.1).

Table 3.1: mAb titers of CHO cell line replicates in the two treatment conditions

Replicates of Inducible Cells CHOBRI55E1-EG2-Fc-X1	Cumate treatment	mAb Titer (mg/l)	Fold change
Control 1	-	51	
Control 2	-	48	
Control 3	-	42	
Control 4	-	51	
1	+	295	5.8
2	+	283	5.9
3	+	266	6.3
4	+	254	5.0

3.3.1 Protein identification and quantitation by PEAKS

The raw data results from the LC-MS/MS system were analyzed by PEAKS Studio 7.5 (Bioinformatics Solutions Inc.). From the total of 84,416 tandem mass spectrometry scans, 19,544 peptide sequences were identified after filtering by the *de novo* sequencing algorithm. Figure 3.1 represents the average local confidence scores of peptides vs. mass errors in ppm. The mass error of the peptides as stated in PEAKS user manual is calculated based on "a ratio of observed mass error (difference between observed mass and theoretical mass) and the theoretical mass and is expressed in ppm" [75].

The Refseq protein database from chogenome.org with the addition of the produced antibody sequence with a total of 22,527 entries was used in PEAKS DB search. PEAKS database search could match 12,771 peptide sequences with minimum 15 confidence score (initial cutoff for peptide matches). The reported result with a FDR cut-off at 5% included 11,523 peptide sequences and 3,410 proteins identified in the search against this database. Figure 3.2 shows the PSM (peptide spectrum match) score distribution. We observe very few decoy matches in the region with high scores, this demonstrates the high confidence in the search result and the peptide -10lgP score threshold is of high confidence. The software estimated FDR using the tartget-decoy method and the figure 3.3 shows the FDR curve for the PSMs.

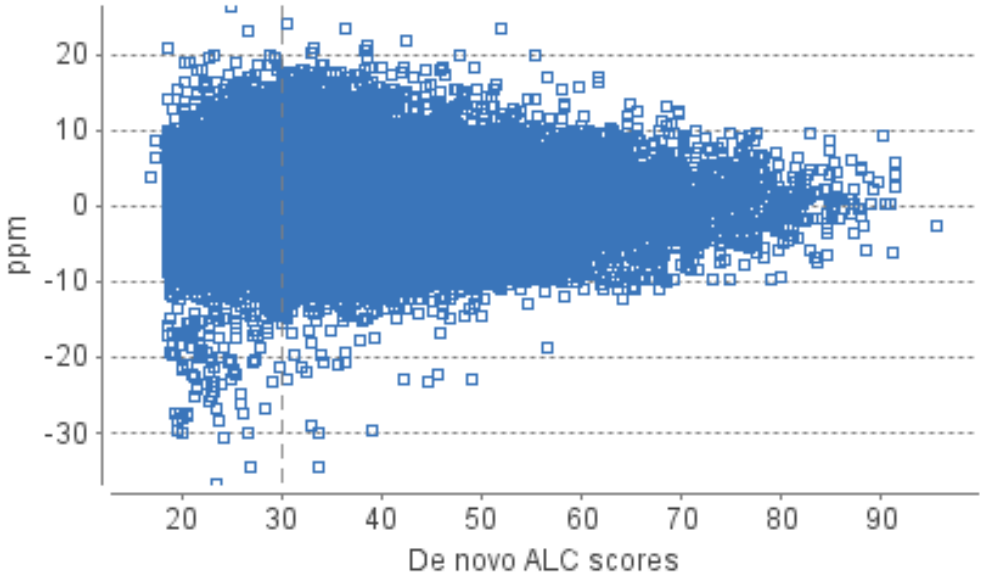

Figure 3.1: Average local confidence (ALC) scores of *de novo* peptides are plotted against mass errors of the identified peptides in ppm. Peptides with ALC scores above 30 was used in the next steps of analysis Mass errors (in ppm) of the identified peptides are plotted against their ALC scores.

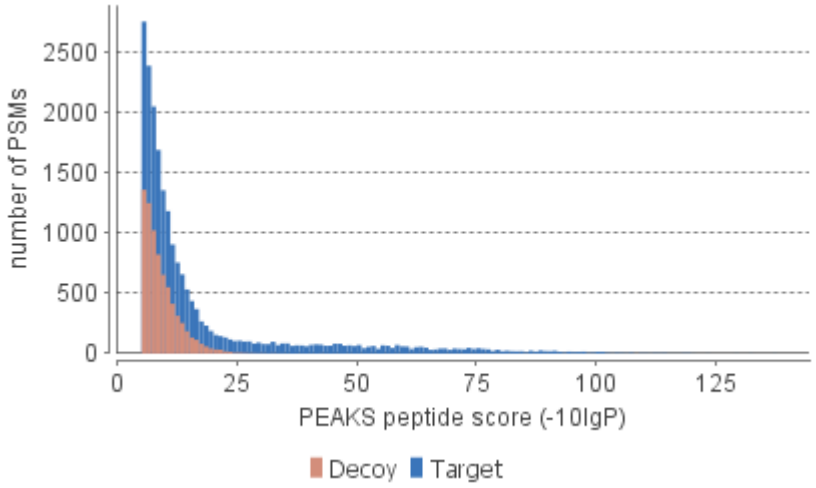

Figure 3.2: The histogram of PEAKS peptide scores. Histogram shows the number of peptide-spectrum matches with the corresponding peptide score calculated in the PEAKS DB search. Using the target-decoy method, decoy hits shown in brown and target hits in blue, the FDR is estimated by the software.

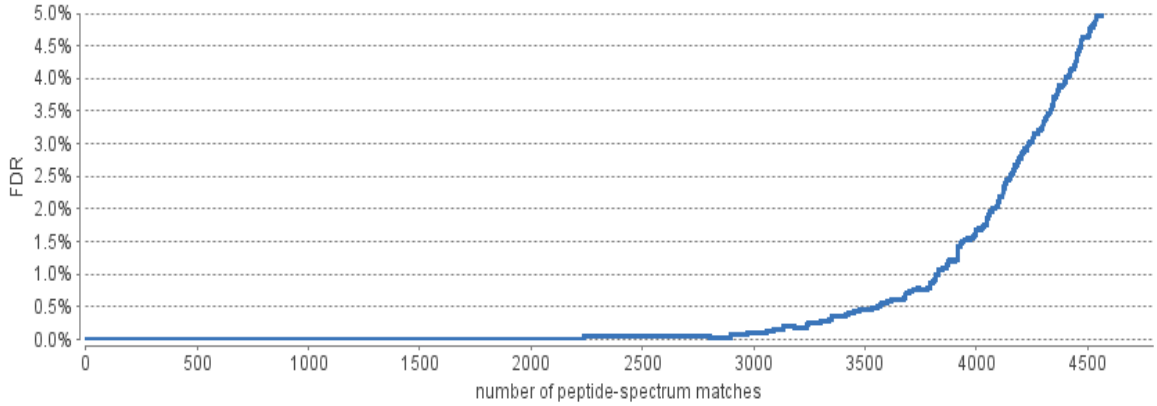

Figure 3.3: The FDR curve for the identified peptide spectrum matches. The figure shows the FDR value corresponding with the number of PSMs that are kept in the final result.

3.3.2 Differential expression analysis

Raw intensities of all proteins were exported from PEAKS software. The software represents protein expressions as an intensity value for each label. Based on our experimental design, data was separated into two groups of control samples and samples with cumate treatment. To reduce bias due to differences in sequencing depth between samples, the expression values were subjected to quantile normalization as described in section 2.2.5. Once the data is grouped and normalized, they were compared using unpaired student's t-test with unequal variance in R stats and limma package [80]. Next, threshold for significant differences was set by applying the Benjamini-Hochberg correction method to the generated list of p-values [82]. Also, fold changes of proteins were calculated and added to the table.

In protein identification by PEAKS, groups of proteins are identified by a common set of peptides. In this dataset we had a total number of 3,410 proteins identified. The protein groups contain similar sequences; therefore, the top protein from each group that has the highest score and number of peptides identified is used in the data analysis steps; at the end 1,616 proteins which have a valid expression value for all of the samples (more than zero) were tested. For a more clear representation and identifying significantly regulated proteins with large differences, a logarithmic scale of p-values was plotted versus fold-change, known as the volcano plot (Figure 3.5). Among this set of unique protein sequences 82 proteins had differences with raw p-values lower than 0.05 (Appendix A). Then we removed proteins that had minimal change in expression levels in all samples with a fold-change threshold close to 1 (-0.1< logFC <0.1). Using Benjamini-Hochberg method for multiple hypotheses testing correction with q equal to 0.05, the p-value threshold was set to 2.6e-04 (0.00026), which will not accept any of the proteins as a significant result (Figure 3.4). FDR level with q as 0.25 will set the p-value threshold to 0.004 and 9 top proteins will be accepted with a significant difference. Table 3.2 shows the proteins with the lowest p-values which have a high increasing or decreasing fold change among the list; appendix A provides the complete list of 82 proteins with p-values below 0.05. The top protein that is accepted in the hypothesis testing is the mAb sequence produced by the cells. Based on the experimental data, the antibody titer in

the culture media had an average increase of 5.7 fold, and the proteomic data is consistent with this, as the mAb had almost a 3 fold-change in the whole cell proteome.

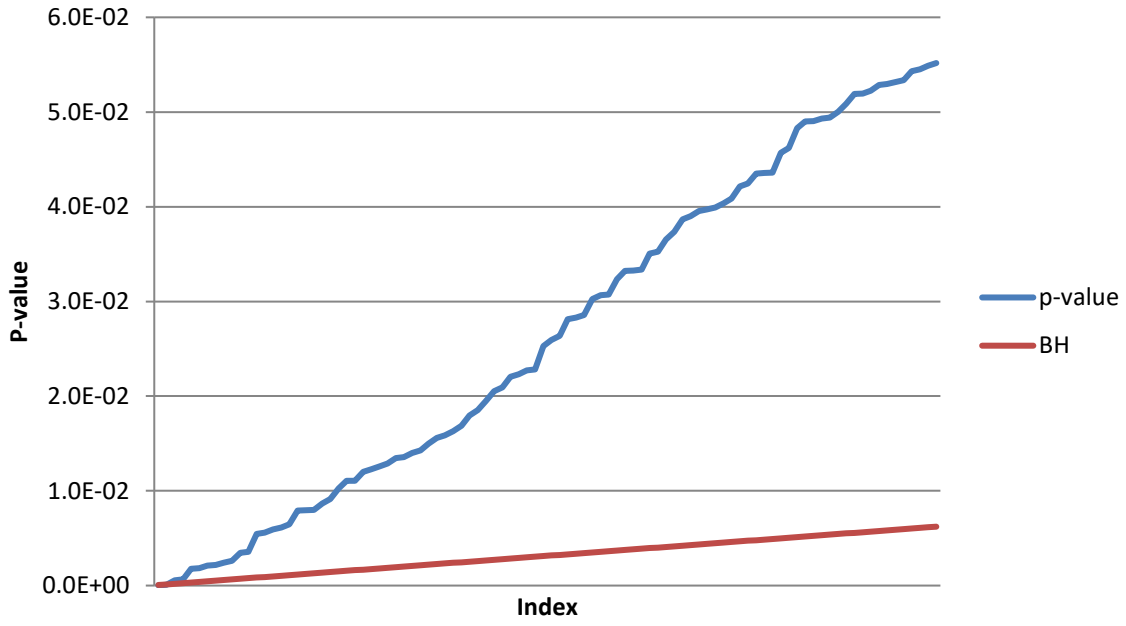

Figure 3.4: Sorted p-values vs. BH adjusted values. The blue curve represents the p-values lower than 0.05 and the red is the Benjamini-Hochberg adjusted values. The distribution of p-values shows that there are not any significant results with this threshold.

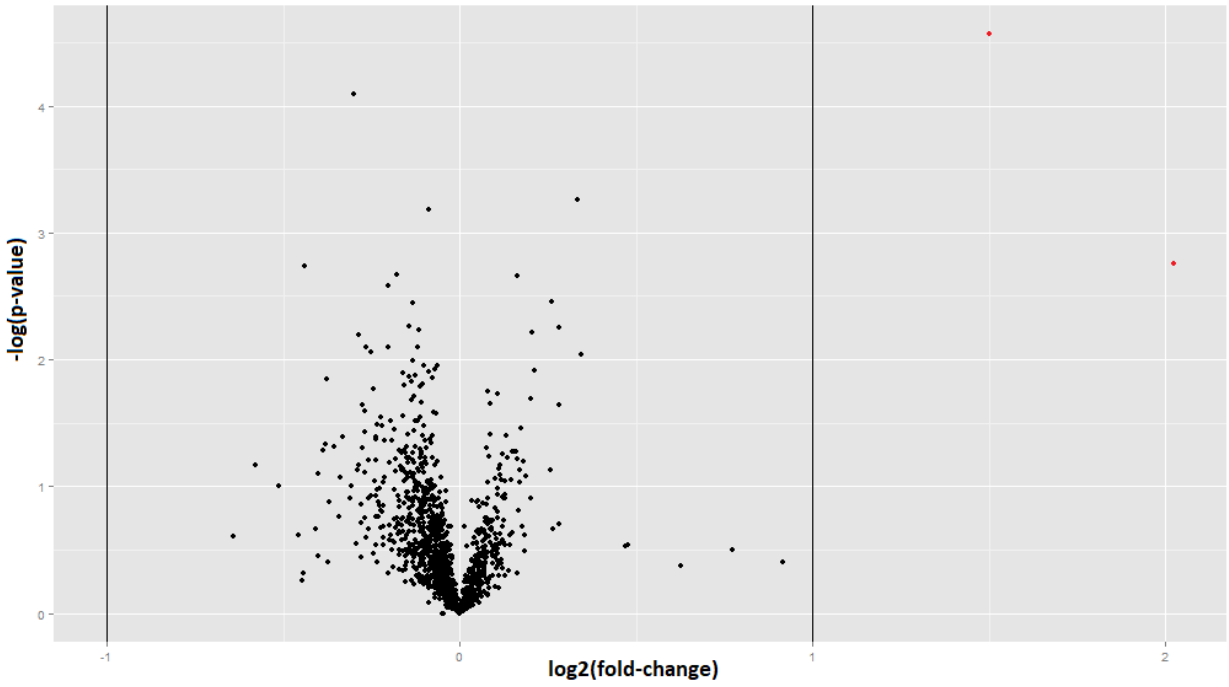

Figure 3.5: Volcano plot of CHO protein expression data. The negative logarithm of p-value vs. log fold-change of CHO proteins. The points shown in red color are sequences of the mAb product and the sequence of antibody's heavy-chain.

Table 3.2: List of candidate genes with the highest fold-change showing the highest increasing/decreasing expression levels in the high-producer CHO cells and the significant p-values with two FDR cut-offs at 5% and 25%. Table includes the log fold-changes above 0.2 and below -0.2 in addition to the significant p-values with FDR 25%, although they have small changes.

Gene	Fold Change	log FC	p-value	Significant FDR (5%)	Significant FDR (25%)	Protein ID	Protein
EG2-hFc1	2.83	1.50	0.00003	Y	Y	mAb sequence	EG2-hFc1 mAb product
TRMT112	1.26	0.33	0.001	N	Y	354498436	tRNA methyltransferase
PPP1R12A	0.94	-0.09	0.001	N	Y	354492600	protein phosphatase 1 regulatory subunit 12A-like
RSP7	1.12	0.16	0.002	N	Y	354495303	40S ribosomal protein S7-like
MRP-L55	0.74	-0.43	0.002	N	Y	354482471	Ribosomal protein L55 mitochondrial
GSTP1	0.88	-0.18	0.002	N	Y	350537543	glutathione S-transferase P 1
FKB11	1.20	0.26	0.003	N	Y	354497358	Peptidyl-prolyl cis-trans isomerase
Lsm3	0.87	-0.20	0.003	N	Y	354465501	snRNA-associated Sm-like
PUF60	0.91	-0.14	0.004	N	Y	354491098	poly(U)-binding-splicing factor PUF60-like
TRA2B	1.22	0.29	0.006	N	N	354484188	Splicing factor
GANAB	1.15	0.20	0.006	N	N	354493306	Glucosidase alpha neutral
LAMC1	0.82	-0.29	0.006	N	N	354481424	Laminin subunit gamma-1
VIM	0.83	-0.27	0.008	N	N	354482483	Vimentin
HN1L	0.84	-0.25	0.009	N	N	354466573	Hematological neurological expressed 1
RPS6	1.16	0.21	0.012	N	N	354507225	Ribosomal protein S6
GLUL	1.22	0.29	0.023	N	N	354481440	Glutamine synthetase
UBE2O	0.83	-0.27	0.023	N	N	354489489	Ubiquitin-conjugating enzyme E2O
IGF2BP2	0.83	-0.27	0.025	N	N	354484186	mRNA-binding protein 2
CN166	0.86	-0.22	0.028	N	N	354477974	RNA-binding protein
Slc27a4	0.85	-0.23	0.032	N	N	354499491	Long-chain fatty acid transport
Icam1	0.86	-0.22	0.033	N	N	350539683	Intercellular adhesion molecule 1
HSPG2	0.85	-0.23	0.040	N	N	354483018	Basement membrane-specific heparan sulfate proteoglycan
CTSM	0.85	-0.23	0.042	N	N	354468711	Cathepsin M
CC115	0.86	-0.22	0.044	N	N	354491621	Coiled-coil domain
GALM	0.83	-0.27	0.049	N	N	354488923	Aldose 1-epimerase

3.3.3 Functional analysis

The top protein with the highest fold-change is the multifunctional methyltransferase (trmt112). It is involved in methylation of proteins and tRNAs, and heterodimers of this methyltransferase subunit with some other methyltransferases function in efficient translation termination by catalyzing a methylation in translation factor 1. Transformer-2 beta (TRA2B) is a splicing factor and also reacts in the translation process, particularly in controlling pre-mRNA splicing. Its activity can suppress or cause exon inclusion. These splicing patterns can change by acting as antagonist of other splicing regulators, like RNA binding motif protein.

Among other proteins with increasing expression level, GLUL is the protein-coding gene for glutamine synthetase or glutamate-ammonia ligase. Amino acid synthesis and glutamine uptake are among the related pathways to GLUL. In the conditions that cells produce high levels of antibody the metabolism and energy uptake is expected to be higher.

The alpha subunit of glycosidase II (Ganab) codes a protein that has a role in protein folding and quality control. In the process of translation in endoplasmic reticulum, this enzyme involves in cleaving glucose residues from immature glycoproteins. The presence of this protein is required in modifying glycosylation patterns since it is involved in N-glycan metabolism. The peptidyl-prolyl cis-trans isomerase, expressed by the gene FKB11, accelerates the folding process of proteins during protein synthesis. These processes in cells with a high rate of mAb production, which have large and complex structures to fold, are expected to be increased.

Among the proteins showing a decrease in abundance are vimentins. They are intermediate filaments attached to the nucleus, endoplasmic reticulum, and mitochondria. This protein plays a major role in cytoskeletal structure of a cell and along with other filaments, such as MYH10, vimentin involves in stabilizing collagen mRNAs as well [95].

Laminin gamma1 (LAMc1) is one of the chains in the extracellular matrix glycoprotein that is involved in biological processes, such as cell signaling, cell adhesion and migration. Heparan sulfate proteoglycan-2 (HSPG2) is a proteoglycan that binds many extracellular components and surface molecules. Both of these proteins that are involved in cell signaling and adhesion had slightly lower expression levels in high-producer cells, although without a significant p-value.

Although individual genes were not strongly supported, it is possible to look for enriched function within the identified gene list as a group. The list of genes in table 3.2 and from the list of all proteins with raw p-value below 0.05, a list of the genes with down-regulated protein expression, as well as a list of the genes with up-regulated protein expression values was uploaded on DAVID analysis tool. In the gene list manager section, Chinese hamster (*Cricetulus griseus*) was selected as the species and gene population background. If DAVID knowledgebase did not have a gene's identifiers for Chinese hamster and their biological annotations the mouse data was used instead. In the functional analysis of the list, available information of mouse homolog was used. Therefore, the annotations were limited to the two species Chinese hamster (*Cricetulus griseus*) and mouse (*Mus musculus*) selected in the gene enrichment analysis.

In the analysis reports from DAVID, no molecular function category was identified to be significantly enriched in the gene list. The biological processes and molecular functions commonly shared among the candidate gene list were the RNA splicing factors, RNA splicing via spliceosome and transesterification reaction. However, the terms were not significantly enriched above the threshold of 1.3 enrichment score (Table 3.3).

Similarly, using the two lists of genes with positive and negative fold-changes and raw p-values below 0.05, the KEGG pathway database search results were reported. Table 3.4 is a list of the pathways with significant enrichment scores using the gene list with negative changes. No pathways were enriched in the analysis of the list of genes with positive fold-changes. Two KEGG pathways were also identified with high enrichment scores, ECM-receptor interaction (cge04512) and glyoxylate and dicarbozylate

metabolism (cge00630). In the former pathway all the genes in the list had negative fold-changes LAMC1, COL6A, CD44, FN1, ITGA5. LAMC1 had the largest change, -0.29 fold-change, which had the lowest fold-change and p-value among the involved genes. The laminin gamma 1 is a subunit of the extracellular matrix glycoprotein. In the second identified KEGG pathway three genes were listed that are involved in the biosynthesis of glycoxylate and carbohydrates pathway. These genes had very small changes although glutamine synthetase (GLUL) had a 0.28 fold-change increase in the high-producer cells. Since this protein is also involved in the synthesis of amino acids and glutamine uptake, the increase in the induced cell line is expected.

Table 3.3: Molecular functions of the table 3.2 gene list

Molecular Function	Genes
RNA binding	LSM3
Catalytic activity , Hydrolase activity	Ganab
Cysteine-type endopeptidase activity	Ctsm
Aldose 1-epimerase activity	GALM
Nucleotide binding	IGF2BP2, TRA2B, Ube2o, Slc27a4
Integrin binding	ICAM1
Extracellular matrix structural constituent	lamc1
Protease binding	Hspg2
mRNA binding, structural constituent of ribosome	RPS6
Protein methyltransferase activity	TRMT112
Glycoprotein binding, structural molecule activity	VIM

Table 3.4: KEGG pathway database search results. The gene list of proteins with a negative fold-change was used in DAVID analysis tool.

Term	Genes	Fold Enrichment	Benjamini
cge04512 ECM-receptor interaction	CD44, ITGA5, COL6A2, COL6A1, LAMC1, FN1	10.5	2.4E-02
cge00630 Glyoxylate and dicarboxylate metabolism	LOC100760000, GLUL, GCSH	20.3	4.7E-02

3.4 Discussion

In this work a quantitative proteomic analysis of inducible CHO cells with cumate gene switch was performed. The cell line produces the MAb EG2-hFc1 using a CR5 promoter and cumate gene-switch, and could produce over 200ug/ml of mAb in the presence of cumate. With the presence of cumate in cell culture medium, these CHO cells were able to produce almost 6 times the amount of mAb produced by uninduced cells. Baseline production of MAb without induction is somewhat higher than expected at 48 ug/ml, indicating a reasonable amount of 'leakage' with this promoter system. In this study we were aiming to capture the differences in proteome level between these two different conditions. Our goal is to show whether producing high amounts of mAb is putting cells under a significant stress or if this expression system has the potential to generate a stable cell line with higher amounts of mAb being produced.

Using the iTRAQ 8-plex reagents and LC-MS/MS analysis, expression levels of over a thousand unique proteins were identified. In this analysis 82 proteins had a p-value lower than 0.05 but by performing the FDR test at q=0.25, only nine proteins were identified as having statistically significant changes. Of these, we would expect two to be false positives. The highest fold-changes are observed in proteins involved in protein processing and RNA binding and splicing. As shown in table 3.3, there are enzymes involved in metabolic pathways, precursor metabolites and some are related to energy uptake in CHO.

Knowing that producing high amounts of protein can put cells under increased stress [32], we were expecting that some of the identified differentially regulated proteins may be involved in stress response processes. For example, Meleady et al. 2012 in a proteomics study found heat shock proteins and proteins involved in the response to oxidative stress are changing in high producer cells [30].

Although the cumate induced cell line produced higher mAbs titers on average, we did not identify any known stress response proteins with statistically significant changes in abundance. The number of replicates used in this study was quite low with four treatments and four controls, so increasing the number of replicates would likely result in increased size of dataset and numbers of statistically significant changes.

However, based on fold-change there are not many proteins showing large changes in abundance, therefore it is likely that the cumate induction system is not causing larger metabolic shifts or stress responses within the cell. The lack of large significant differences in the proteome of cells in the two conditions can show that the system has generated a stable cell line with higher production capacity. As stress responses are not being induced in the cell line there is likely the potential to increase mAb production using a stronger promoter system. The lack of stress response suggests that there is excess capacity for exogenous protein production that could still be utilized.

For future studies, considering the scale of culture and production conditions in this experiment, the harvesting time, or the complexity and structure of the produced mAb could affect the cells conditions and stressors. Therefore, scaling up the production using this cell line might influence the cells and further evaluations would be required to analyze the cells condition. Furthermore, cells can be analyzed at several intervals during the production process to understand the metabolic changes.

Chapter 4

Comparative proteomics analysis of HT-29 cells

4.1 Introduction

Colorectal cancer has one of the highest mortality rates, and is the third leading cause of death in the world (American cancer society, CRC facts and figures, 2017). In the recent years, survival rate of CRC has increased but this is mostly due to the improvements in the screening and diagnosis of tumors at the early stage rather than improvements in therapies. Drug resistance is one of the main reasons for treatment failure or patient relapses with metastasis [33]. Irinotecan is one of the main therapeutic agents in treating CRC patients as a single agent therapy or in combination therapies. Irinotecan is a camptothecin derivative and a member of topoisomerase I inhibitor family of anti-tumor agents. Irinotecan is biologically inactive and it needs to be catalyzed to its active metabolite SN-38.

The target of SN-38, Topoisomerase I, cuts and ligates a DNA strand to change the linking numbers by one. It is an essential enzyme for relaxing DNA supercoils that are made during transcription, replication processes, and chromatin remodeling. TOP1 inhibitors can interfere during this cleavage reaction; interference with the topoisomerases activity results in formation of cleavable complexes with DNA. Cellular sensitivity to TOP1 inhibitors could be determined by the deficiencies in both the checkpoint pathways and DNA-repair pathways [42]. DNA repair mechanisms and activity of TOP1 enzyme in the cells affects inhibitors efficiency [36].

By exploring the protein expression data, we would like to determine the key pathways and metabolic features that have changed in HT-29S cells, to better elucidate the process by which cancer cells undergo in a path to drug resistance. To this end HT-29 cells were treated with prolonged exposure to SN-38, which made them resistant to the drug.

4.2 Methods

4.2.1 Cell culture

A derivative of the CRC cell line HT-29 has been generated in Dr. Blay's lab that is highly resistant to SN-38 treatment. This cell line, termed HT-29S, has been developed by sequentially increasing cell line exposures to the drug in the culture media. Briefly, the cell line was developed starting with the HT-29 cells in the culture media, and gradually increasing the SN-38 concentration in 6 steps from 0.1 nM in the first week to 100nM after 3 weeks [51].

The HT-29 and the HT-29S resistant cells were cultured in Nunc T25 culture flasks (Thermo Scientific Fisher) at 37°C in a humidified incubator with 5% CO_2. Hyclone DMEM containing 4mM L-glutamine and 4500 mg/L L-glucose, 1mM sodium pyruvate, and 10% FBS was used as well. In addition, HT-29S cell cultures had 30nM SN-38 (Sigma Aldrich) added to the growth media. Cultures that were reaching confluency were passaged after suspension by several minutes' exposure to 1mL TrypLE Express (Life Technologies) at 37°C. Subsequently, 10-15% of the cells were seeded in a new culture flask with fresh media.

Cell samples were prepared for the iTRAQ experiment by Ms. Julia Fux in Dr. Blay's lab. The HT-29 and HT-29S cells were grown to 80-90 % confluency in T75 culture flasks. Then, cells were suspended using 0.25% Trypsin-EDTA (Gibco) and transferred into 15 ml tubes. After centrifugation at 300x g at 4°C, the media was removed and the remaining cell pellets were washed with cold PBS to completely remove the media and trypsin. The washing step was repeated for 5 times and cells were stored at -80°C.

4.2.2 Protein sample preparation and iTRAQ labeling

The whole cell proteome extraction was done using the stored frozen cell pellets. Cells were quickly thawed on ice and 700 µl of the lysis buffer was added to the samples in microtubes. The lysis buffer was freshly prepared using 100mM HEPES at pH 7.5, 1% Triton X-100, 100mM dithiothreitol (DTT), and 5 µl of the protease inhibitor cocktail from Sigma Aldrich (P8340). The samples were incubated on ice at

approximately 4°C for 30min. Once the lysis buffer starts to disrupt the cells, proteases may degrade the proteins, and the protease inhibitor cocktail prevents this unwanted process. To ensure that proteins are not aggregated and extra disulfide bonds are not created, it is necessary to use a reducing agent like DTT or TCEP. The reducing agent cleaves disulfide bonds and prevents formation of protein aggregates. The lysate samples were centrifuged at 10,000x g and the supernatant was extracted leaving the cell debris to discard.

The protein concentration was measured at this step using Bio-Rad microplate protein assay [91] (Bio-Rad Laboratories Ltd., Mississauga, Canada). Following the Bradford method, a set of standard protein samples was prepared using bovine serum albumin (BSA) (Sigma-Aldrich, ON, Canada). Six standard BSA samples containing 0, 0.2, 0.4, 0.6, 0.8, and 1 mg/ml of protein were used to create the linear standard curve. A Multiskan UV microplate spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA) was used to read the corresponding absorbance of the samples at 595nm.

According to the protein concentration of each sample, a volume equivalent of 1mg protein was separated from the cell lysates. Next, the protein lysate samples were precipitated using the Calbiochem ProteoExtract kit from EMD Millipore (Cat# 539180). The kit provided the precipitation agent and the wash solution that were prepared following the manufacturer's protocol and stored at -20°C prior to the process. Samples consisting of 1 mg of the protein lysate were precipitated using 4 times volume of the precipitation agent and incubated at -20°C for a minimum of 1 hour. After centrifugation at 10,000x g speed in 4°C protein pellets were collected and the supernatant was discarded. Then, protein pellets were washed by adding 500 µl of wash solution (at -20°C) and quickly disrupting the pellet. The wash step was repeated for two times and without disrupting the pellets the wash solution was discarded. The protein samples were re-suspended in UA buffer, containing 8M Urea and 100mM HEPES buffer at pH 8.6. The protein concentration of each sample was measured by Bradford microplate assay and 250 µg of completely solubilized proteins from each sample was used in the next step.

Filter-aided protein sample preparation was used for protein purification and digestion [93]. Amicon Ultra centrifugal filters (UFC500324, Merck Millipore, MA, USA) with relative molecular mass cut-off of 3k were used in this experiment following the manufacturer's protocol. Protein samples were loaded on the wetted filters in the chaotropic UA buffer, keeping the proteins completely solubilized. Then, proteins were alkylated to block cysteines by adding 50 mM iodoacetamide (IAM) to the UA buffer. The buffer was changed to 0.5 M triethylammonium bicarbonate pH 8.5 (TEAB) prior to trypsin digestion. Lyophilized trypsin was used with an approximate ratio of 1:50 trypsin to protein, with 20 μg trypsin used for each sample. Samples were stored at 37°C in a bead bath for approximately 8 hours overnight. The digested samples were dried in a Thermo Scientific Savant Speed Vac vacuum centrifuge to deactivate trypsin and re-suspended in 30 μl of TEAB buffer. Concentration of the digested proteins was estimated at this stage using the A280 method [93] on a NanoDrop™ 2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA).

Following iTRAQ manufacturer's protocol, equal amounts of 8 peptide samples were labeled using the eight isotopic labeling reagents in the kit. To assure efficiency of the labeling process, pH was tested by Hydrion pH papers and adjusted with TEAB buffer to stay between 7.5 and 8.5. After one hour incubation at room temperature, labeled samples were mixed into one microtube and evaporated using a Speed Vac® centrifuge from Savant (Thermo Fisher Scientific, Waltham, MA, USA). Then, the sample was resuspended in a buffer, containing 70% acetonitrile and 0.1% formic acid. Then the sample was sent for mass spectrometry analysis at the mass spectrometry facility of SPARC BioCentre at the SickKids hospital, Toronto, as described in section 2.2.4.

4.3 Results

4.3.1 Protein identification and quantitation by PEAKS

The iTRAQ mass spectrometry data was consisted of 79,278 MS/MS scans, and after filtering a total of 19,544 peptide sequences were identified by the *de novo* sequencing analysis in PEAKS. The average local confidence scores for de novo sequenced peptides are shown in figure 4.1. The database search identifies peptide sequences that match the spectrum using de novo analysis, and the identified proteins containing these peptides will be reported in groups of closely similar sequences. A PEAKS database search through the NCBI non-redundant (nr) human protein sequences could match 11,798 peptide sequences with a minimum confidence score of 15. A total of 9,168 proteins were identified within 1,665 protein groups, and the top protein within a group with the highest score and highest number of peptides identified is reported as a unique protein and is used in the expression analysis. Figure 4.2 shows the PSM (peptide spectrum match) score distribution and figure 4.3 shows the FDR curve for the PSMs.

Quantitation within PEAKS 7.5 identified 1,665 groups of proteins with at least one unique peptide hit identified for each of the 1665 proteins. Expression data of these 1,665 unique proteins were tested and 370 proteins had statistically significant differences in expression level between the HT-29 and the HT-29S resistant cells. Based on the Benjamini-Hochberg method for multiple hypotheses testing correction with q as 0.05, the p-value threshold was set to 0.016.

4.3.2 Differential expression analysis of HT-29 vs. HT-29S proteome

Differential expression analysis of the high-throughput mass spectrometry data provides a near-comprehensive look at the proteome of the HT-29 cells and the resistant cells. Relative abundances of identified proteins in the HT-29 and HT-29S samples were compared and plotted as statistical significance vs. fold change (Figure 4.4). The lists of genes that are significantly up- or down-regulated respectively in the resistant cells are presented in Tables 4.1 and 4.2.

The most significantly up-regulated gene is PAGE4 (cancer/testis antigen prostate-associated family member 4) and a member of GAGE family (Table 4.1). The cancer/testis antigens (CTAs) are heterogeneous proteins expressed in the normal tissue of testis. A variety of tumors and some reproductive tissues express GAGE genes. Based on several studies on the function of PAGE4, it is an intrinsically disordered protein (IDP) and highly expressed in fetal prostate and benign or malignant prostate disease cells as well as cancer cells in other reproductive tissues [96]. Importantly, PAGE4 is also a stromal marker that may be associated with an epithelial to mesenchymal transition (EMT) [97].

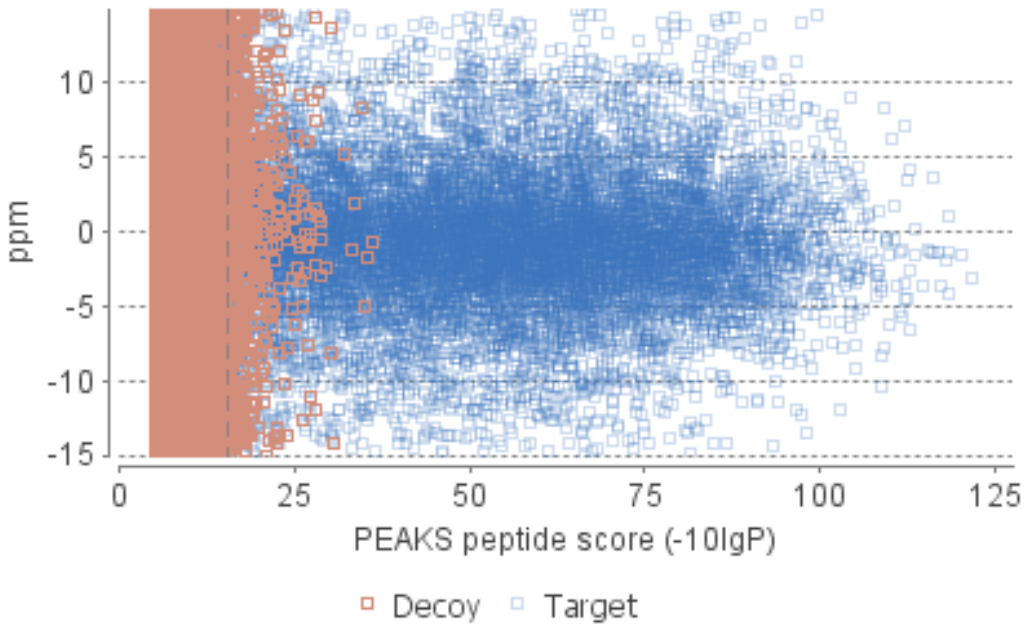

Figure 4.1: Peptide confidence score distribution vs. mass error in ppm. The target-decoy method is used in PEAKS to estimate the FDR, decoy hits shown in brown and target hits in blue. For high confidence, we expect to have few decoys with high scores and a large number of search results.

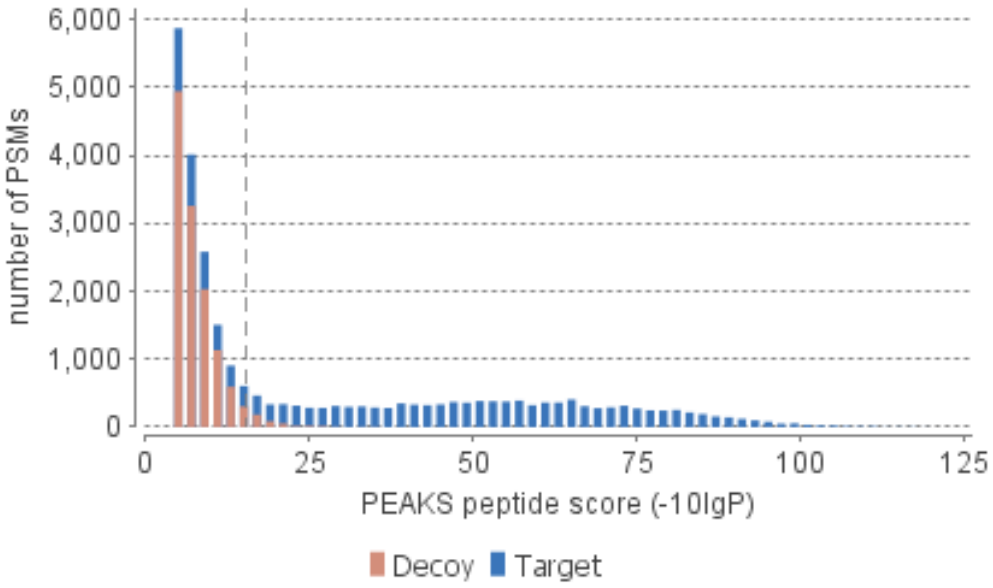

Figure 4.2: The distribution of the identified peptide scores for the peptide spectrum matches (PSMs) shown in Figure 4.1. The majority of correctly identified target peptide residues have high local confidence scores, whereas the majority of decoy peptides have low scores. The dashed line is represents a default threshold. A threshold of 20 was used for further analysis to reduce the number of false positives.

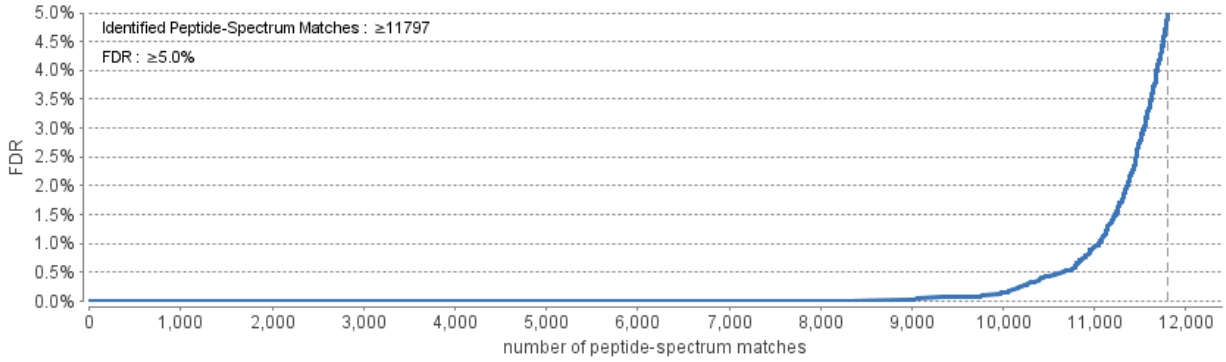

Figure 4.3: The false discovery rate curve for peptide spectrum matches (PSM) used in protein identifications. The curve represents the FDR associated to the number of peptide-spectrum matches that are in the database search result. The vertical dashed line is indicating the score threshold based on the selected score threshold for the FDR lower than 5%.

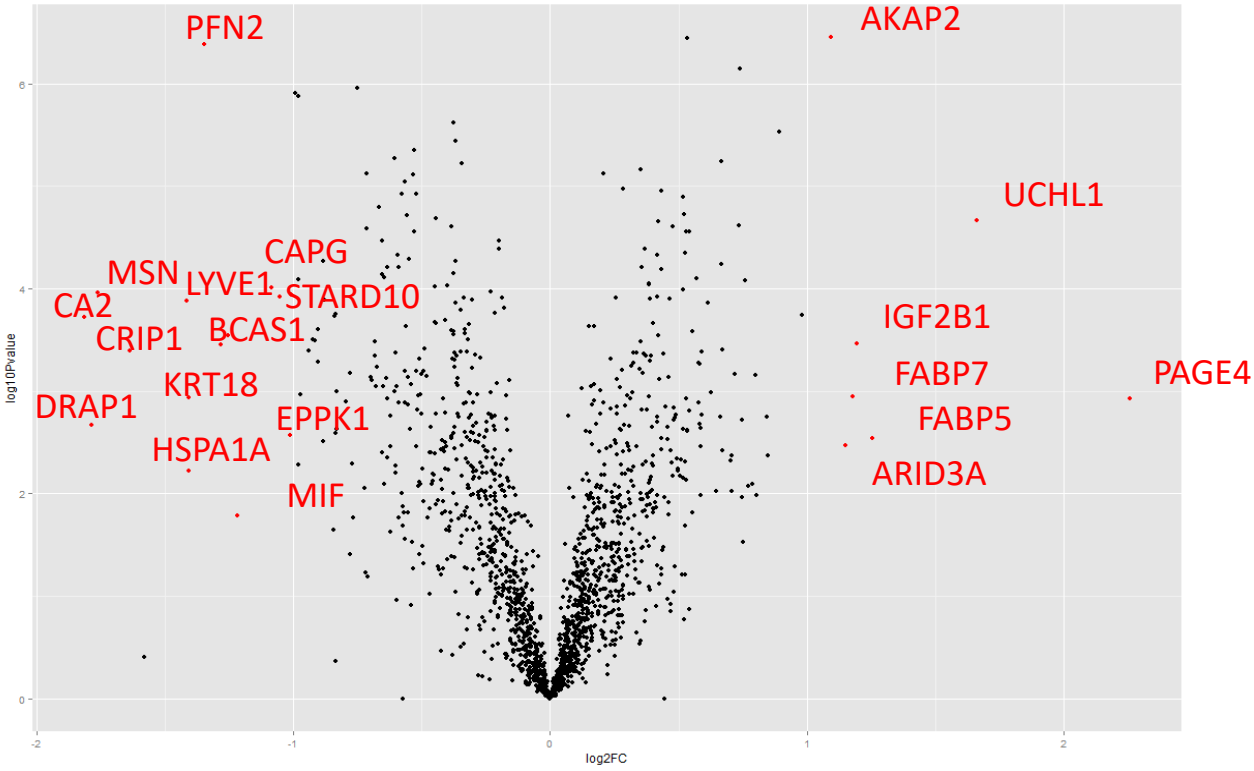

Figure 4.4: Volcano plot $\log_2$(FC) vs. –log(*p*-value), the red dots are labeled and representing the significantly regulated proteins using an FDR threshold of q = 0.05 and | logFC | ≥ 1.0.

Table 4.1: List of genes and protein ids that are significantly up-regulated, passing an FDR threshold of $q < 0.05$ and with greater than 2-fold change in expression ($|\log FC| \geq 1$) in HT-29S cells. The full list of proteins passing FDR $q < 0.05$ ($p < 0.016$) is given in Appendix A.

Gene Symbol	logFC	P-value	Genbank GI	Protein
PAGE4	2.3	1.2E-03	5901986	Prostate associated gene, P antigen family member 4
UCHL1	1.7	2.2E-05	136681	Ubiquitin carboxyl-terminal hydrolase isozyme L1
FABP5	1.3	2.8E-03	4557581	fatty acid-binding protein, epidermal
IGF2B1	1.2	3.5E-04	56237027	insulin-like growth factor 2 mRNA-binding protein 1
FABP7	1.2	1.1E-03	4557585	fatty acid-binding protein, brain
ARID3A	1.2	3.3E-03	56799575	AT-rich interaction domain 3A, DRIL1
AKAP2	1.1	3.5E-07	310772255	A-kinase anchor protein 2

Table 4.2: List of genes/protein Ids that were down-regulated in HT-29S cell line (passing FDR threshold of q = 0.05) and with greater than 2-fold change in expression (| logFC | ≥ 1)

Gene Symbol	logFC	P-value	Genbank GI	Protein
CA2	-1.8	1.9E-04	245850	S-100P=Ca(2+)-binding protein
DRAP1	-1.8	2.2E-03	18426973	dr1-associated corepressor
MSN	-1.8	1.1E-04	4505257	Moesin
CRIP1	-1.6	4.0E-04	4503047	Cysteine-rich protein 1
LYVE1	-1.4	1.3E-04	40549451	Lymphatic vessel endothelial hyaluronic acid receptor 1 precursor
KRT18	-1.4	1.1E-03	30311	Cytokeratin 18
KRT20	-1.4	5.9E-03	27894337	Keratin type I cytoskeletal 20
PFN2	-1.3	4.1E-07	16753215	Profilin-2 isoform a
BCAS1	-1.3	3.5E-04	530418484	Breast carcinoma-amplified sequence 1 isoform X6
HSPA1A	-1.3	2.9E-04	194388088	heat shock 70 kDa protein
CAPG	-1.1	9.9E-05	1002819390	macrophage-capping protein isoform 3
STARD10	-1.1	1.2E-04	119595276	START domain containing 10 isoform CRA_c
EPPK1	-1.0	2.7E-03	119602582	epiplakin 1
SERPINA1	-1.0	1.1E-04	6855601	PRO0684

IDPs can be involved in the cellular processes where binding with high specificity and low affinity is important, for example they are involved in signaling and regulating transcription processes. There is often an association between overexpression of IDPs and human diseases including cancer. So, IDPs such as tau protein, p53, and BRCA1 are possible targets for drugs that modulate protein-protein interactions [96]. Zeng et al. explored the DNA binding preferences of PAGE4 and revealed that it binds to GC-rich sequences and possibly PAGE4 is functioning as a transcription factor or a sequence-specific DNA-binding protein. This protein is involved in stress responses, preventing DNA damage, inhibition of apoptosis caused by stresses, and suppressing reactive oxygen species [96, 98].

In a study by Iavarone et al., PAGE4 was identified as potentially down-regulating the transcription of lipoprotein lipase, suggesting a transcription factor function for this gene [99]. High expression levels of some CTAs including PAGE4 have shown associations with drug-resistant cancer cells [96, 100]. Cells over-expressing PAGE4 had a significant increase in cell viability and its down-regulation had shown to decrease cell survival [96].

PAGE4 suppression resulted in significant increase of protein p21 and phospho-p53, which suggests that PAGE4 is playing a role in protecting cells against stress and apoptosis. Protein p21 inhibits cyclin kinase activities by binding to the CDK2 and CDK4 complexes, and regulates progression of the cell cycle at G1. The tumor suppressor protein p53 also controls p21 expression and is involved in p53-dependent cell cycle arrest at G1 phase in reaction to stresses [96, 101].

UCHL1 gene is coding an enzyme that modifies ubiquitins at specific sites. According to pathway networks, regulation of Uchl1 is related to Chks in checkpoint regulation pathway that is related to DNA repair pathways as well as the G2-M phase transition.

Insulin-like growth factor binding protein (IGF2B1) regulates several transcripts by binding to mRNAs, such as beta-actin (ACTB) mRNA translation that is essential for determining cell polarity and cell migration. IGF2 has interactions with PTGS2 mRNA and stabilizes this transcript during colonic wound

healing. It can also bind to and stabilize the CD44 transcript, thereby promoting cell adhesion and invasiveness in cancer cells [102, 103]. Up-regulation of IGF2 is associated with faster tumor progression in various cancers [104].

GFBPs can effect adhesion and migration of cells and enhance the action of IGFs [105]. IGFBP-1 contains a recognition sequence for integrin receptors in its primary structure, which allows it to specifically bind the fibronectin receptor, α5β1 integrin [106].

Fatty acid binding protein (FABP) has an affinity for lipid ligands and is known for numerous biological functions such as modulation of cell growth and proliferation. Also, in a study on drug-resistant pancreatic adenocarcinoma cells FABP, cofilin, and stratifin were significantly overexpressed, suggesting that overexpression of FABP proteins could be possibly involved in removal of cytotoxic drugs [107]. Arid3A is a transcription factor involved in controlling the progression of cell cycle by RB1/E2F1 pathway. Although ARID3A has very low expression level in the colon and stomach tissue, it is strongly over-expressed in stomach tumors and is suggested as a cancer marker [108]. Increased expression of actinin-4 induces cell migration and motility [109].

The protein coding BSG, also known as CD147, is also significantly up-regulated, although with a lower fold change of log FC = 0.52 (see Appendix B) in the resistant cells. CD147 regulates synthesis of hyaluronan (HA) and interacts with the HA receptors, CD44, and LYVE-1 [110]. This protein is also involved in production of matrix metalloproteinases (MMPs) by stimulating adjacent fibroblasts. MMPs are known as targets in cancer therapy and are linked to chemoresistant cancer cells through regulation of autophagy and apoptosis pathways [111].

The main function of cytokeratins is in stabilizing the cells and other possible functions as well. Among the intermediate filament gene family, cytokeratins are the most complex genes with acidic and basic types based on their charge, with the KRT9-20 being in the acidic subdivision and KRT1-8 in the basic type [112]. Cytokeratin genes are positive and highly investigated markers in colorectal cancer. Cytokeratin 20

(KRT20) is found to be expressed in 95% of CRC primary and metastatic tumors and it is a marker for detection of circulating CRC cells, because normally epithelial cells that express KRT20 are not in circulation [113]. A statistically significant trend has been detected in KRT20 positive patients with increasing stages of disease [113]. The expression profile of KRT20 is a factor in selecting treatment options for CRC patients, and Kust et al. suggested that more aggressive treatments could be used for patients with positive KRT20 after operation. KRT18-20 and KRT80 were observed in the protein data significantly down-regulated in the resistant cells, as well as KRT7 which is up-regulated with a fold-change of 0.67.

Breast carcinoma amplified sequence 1 (BCAS1) is over-expressed in most breast cancer cells, but down-regulated in colorectal tumors. It is also expressed in many organs like brain, colon and prostate. In our drug-resistant HT-29S cell line, BCAS1 had a significant down-regulation.

Another protein that has significant decrease in HT-29S cells is profilin-2, which is known that is involved in several pathways related to regulation of actin cytoskeleton and cytoskeletal remodeling. It is a ubiquitous actin binding protein in the profilin family that has a crucial role in cellular motility, along with several other genes that are known as actin binding components and were down-regulated in the resistant cell line (Table 4.3). Profilin-2 reacts to extracellular signals by binding to actin and regulating actin polymerization and structure of the cytoskeleton. The polymerization of actin increases when the level of profilin-2 is in low concentrations. It has been studied that how regulation of the PFN2 gene can affect the migration, invasion, and stemness features of HT-29 cells [114]. Profilin-2 regulates the expression of E-cadherin and snail that are EMT markers as well as the expression of some stem cell marker proteins, CD133, SOX2, and β-catenin. Kim et al. 2015 showed that knockdown of profilin-2 increased expression of E-cadherin and suppressed the expression of snail [114].

In the iTRAQ analysis of the HT-29S cell line, expression level of CKB was significantly decreased with a fold-change difference of 1.0 (see Appendix B). Creatine kinases (CK) are involved in reversible catalysis of the transfer of phosphate between ATP and creatine phosphate. Creatine kinase brain (CKB) is a

cytosolic isoform of creatine kinase family of enzymes. CKB is mainly expressed in the brain and it is overexpressed in many cancer cells, but in colon cancer cells it is may be downregulated. By subcellular fractionation the overall levels of CKB showed a decrease in colon cancers, whereas nuclear matrix proteins from normal and colon cancer samples had overexpressed CKB. CKB has shown to be expressed localized in the nuclear matrix. In a study by Mooney et al. (2011), which looked into the role of CKB in tumorigenesis in colon cancer cells, downregulation of CKB appears to promote epithelial-to-mesenchymal transition (EMT) in these cells [115]. Along with the qPCR analysis they found that the protein and mRNA levels had significantly decreased in colon cancer.

Knowing that CKB may be downregulated in colon cancer cells and this enzyme is involved in energy transduction, cells are likely under metabolic stress. Lower expression level of CKB in the resistant cells could be a sign of more aggressive cancer cells (that is putting cells under stress).

Also, previous results have indicated that with downregulation of CKB the expression levels of PAGE4 and SNAIL were increased [115]. Consistent with this study, protein expression level of PAGE4 in the resistant cell line HT-29S significantly increased by 2.26 fold within the current study as well.

Moesin (MSN) functions as cross-linker for plasma membranes and actin-based cytoskeletons. Moesin is important for cell-cell signaling and movement. It was shown in studies on human lymphoid cells that moesin could have a role in regulating cell proliferation, migration and cell adhesion, as well as immunologic synapse formation [116].

4.4. Functional analysis of differentially regulated proteins

Functional analysis of pathways was done with DAVID [84], using the list of genes (Appendix B) that had significant differential expression of proteins in the high-producing cells independent of fold change. Using DAVID analytical tool [84], enriched gene ontology (GO) terms and KEGG pathways and identified and their biological annotations are used to classify key functions and pathways that could be linked to drug resistance in the treated cells. The tables reported below include enriched biological processes, molecular functions, and cellular components, the enriched pathways from KEGG database search, and the enriched transcription factor binding cites.

In this section, only the significantly enriched results from functional analysis using DAVID are shown and discussed. Two sets were prepared; a gene set of the significantly down-regulated proteins and a gene set of significantly up-regulated proteins were uploaded on DAVID. Human genome (*Homo sapiens*) was selected as the species and gene population background at the gene list manager section.

Table 4.3 contains the significantly enriched GO terms from the down-regulated gene set, with the genes related to the top annotation cluster shown in table 4.4. The cluster with the highest enrichment score are functions related to cell-cell adhesion and among other highly enriched terms, features are common with the aggressive types of cancer cells and cancer stem cells. Table 4.5 shows the significantly enriched GO terms from the up-regulated gene set, and Table 4.6 shows the gene members of the corresponding GO terms. Lastly, Table 4.7 shows the enriched KEGG pathways from the up- and down- regulated gene sets.

Table 4.3: Significantly enriched functional annotation clusters observed from the down-regulated genes in the resistant cell line

Cluster	Enrichment Score	Enriched Terms	GO terms IDs	# of Genes Involved	FDR
1	19.86	Cadherin binding involved in cell-cell adhesion	GO:0098641	31	5.0E-18
		cell-cell adherens junction	GO:0005913	31	1.8E-17
		cell-cell adhesion	GO:0098609	29	8.6E-17
2	7.37	Actin-binding	UniProt Keyword	22	1.1E-10
		Actin filament binding	GO:0051015	10	1.7E-02
		Actin cytoskeleton	GO:0015629	11	9.6E-02
3	5.73	Ribosomal protein	UniProt Keyword	14	2.0E-05
		Translational initiation	GO:0006413	13	3.6E-05
		SRP-dependent cotranslational protein targeting to membrane	GO:0006614	11	8.8E-05
		Ribonucleoprotein	UniProt Keyword	16	1.2E-04
		Ribosome	GO:0005840	13	1.5E-04
		Viral transcription	GO:0019083	11	4.7E-04
		nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	GO:0000184	11	8.2E-04
		Structural constituent of ribosome	GO:0003735	14	1.1E-03
		Ribosome Pathway	hsa 03010	12	3.3E-03
		rRNA processing	GO:0006364	13	4.6E-03
		Translation	GO:0006412	14	4.6E-03
4	3.12	prefoldin complex	GO:0016272	4	4.4E-02

Table 4.4: The top annotation cluster of differentially down-regulated proteins and the group of genes related to each term.

Category	Term	Genes	Fold Enrichment	FDR
Molecular Function	Cadherin binding involved in cell-cell adhesion **GO:0098641**	LDHA, LIMA1, PDLIM5, HSPA1A, SFN, TAGLN2, EFHD2, PAK2, RPL6, RANBP1, LRRFIP1, EHD1, PLEC, DBNL, LAD1, ANXA1, SLC9A3R2, VASP, TNKS1BP1, CORO1B, KRT18, ITGA6, CGN, LASP1, CAPG, TMOD3, EPS8L2, TJP2, DBN1, EPS8L1, SPTAN1	9.91	5.0E-18
Cellular Component	Cell-cell adherens junction **GO:0005913**	LDHA, LIMA1, PDLIM5, HSPA1A, SFN, TAGLN2, EFHD2, PAK2, RPL6, RANBP1, LRRFIP1, EHD1, PLEC, DBNL, LAD1, ANXA1, SLC9A3R2, VASP, TNKS1BP1, CORO1B, KRT18, ITGA6, CGN, LASP1, CAPG, TMOD3, EPS8L2, TJP2, DBN1, EPS8L1, SPTAN1	9.51	1.8E-17
Biological Process	Cell-cell adhesion **GO:0098609**	LDHA, LIMA1, PDLIM5, HSPA1A, SFN, TAGLN2, EFHD2, PAK2, RPL6, RANBP1, LRRFIP1, EHD1, PLEC, DBNL, LAD1, SLC9A3R2, VASP, TNKS1BP1, CORO1B, KRT18, CGN, LASP1, CAPG, TMOD3, EPS8L2, TJP2, DBN1, EPS8L1, SPTAN1	10.10	8.6E-17

Table 4.5: The most enriched gene ontology terms from the list of differentially up-regulated proteins with the group of genes related to each term.

Category	Term	GO terms IDs	# of Genes Involved	Fold Enrichment	FDR
Molecular Function	Poly(A) RNA binding	GO:0044822	41	3.52	3.8E-09
	Protein binding	GO:0005515	130	1.44	1.2E-06
Biological Process	Cell proliferation	GO:0008283	15	4.17	0.02
Cellular Component	Extracellular exosome	GO:0070062	73	2.78	1.9E-14
	Cytosol	GO:0005829	73	2.36	1.8E-10
	Cytoplasm	GO:0005737	88	1.81	4.1E-07
	Nucleoplasm	GO:0005654	58	2.23	2.3E-06
	Myelin sheath	GO:0043209	12	8.46	2.3E-04
	Mitochondrial matrix	GO:0005759	15	4.92	0.003
	Cell-cell adherens junction	GO:0005913	14	4.65	0.01
	Spliceosomal complex	GO:0005681	8	9.12	0.04

Table 4.6: The group of genes related to the most enriched terms listed in table 4.5

Category	Term	Genes	Fold Enrichment	FDR
Molecular Function	Poly(A) RNA binding **GO:0044822**	XRCC5, PUS1, FKBP4, SNRPD1, SNRPD2, IGF2BP1, CCT3, SERPINH1, SRRT, SAFB, C1ORF52, TPT1, QKI, HSPE1, UBAP2L, CCAR2, RBM26, GNL3, ENO1, CLNS1A, ALDH6A1, EEF1A1, NOC4L, LGALS3, TRIM28, FSCN1, HLA-A, PRPF3, HNRNPA1, PRPF6, EIF4B, APEH, RBPMS, CCT4, ILF2, VCP, THRAP3, MAPRE1, FARSA, POP7, DUT	3.52	3.8E-09
Biological Process	Cell proliferation **GO:0008283**	XRCC5, BCAT1, NASP, FSCN1, UCHL1, SRA1, SRRT, KIF2C, PPP1R8, TACSTD2, RAP1B, YAP1, MAPRE1, LRP2, GNL3	4.17	0.02
Cellular Component	Extracellular exosome **GO:0070062**	ALAD, ATP5B, PITPNA, UCHL1, SNRPD2, CCT3, CD2AP, PDHB, VCL, MTHFD1, TPT1, IMPDH2, ALDH6A1, BSG, PTPRF, ERP29, HLA-A, PROSC, CTSV, SPAG9, NPC2, CLIC3, PRDX6, CLIC4, TACSTD2, CFL2, PDCD6IP, CA2, LCP1, UGP2, FKBP4, SNX2, DAG1, PPT1, SNX3, ATP6V1B2, UBAC1, ACAT2, SERPINH1, ACP1, TPM4, PFAS, ATIC, KRT7, HSPE1, LAMB1, ENO1, EEF1A1, PARD6B, LGALS3, VTA1, FSCN1, GARS, SOD1, HNRNPA1, ANXA3, ERP44, APEH, PSMC6, CCT5, CCT4, VCP, HEBP1, THRAP3, FABP3, PHGDH, QPRT, RAP1B, LRP2, PHPT1, AHSA1, FABP5, DUT	2.78	1.9E-14

Table 4.7: Enriched KEGG pathways in the candidate genes list.

Gene list	KEGG pathway	Genes	Benjamini Adjusted p-value
Up-regulated genes	Biosynthesis of antibiotics	BCAT1, PGM3, ATIC, ARG2, PHGDH, HMGCS1, ACAT2, PCCB, UGP2, PFAS, PDHB, ENO1	0.01
Down-regulated genes	Ribosome	RPL19, RPS17, MRPL14, RPL13A, RPL13, RPL6, RPL21, RPS15, RPL11, RPL12, RPS4X, RPS23	5.0E-04

4.5 Discussion

Cancer stem cells have been found in primary and malignant CRC tumors and hypothesized to be resistant to chemotherapy treatments [117]. In addition, stem cell markers are enriched in recurrent CRCs after chemotherapies. There are experimental evidence that carcinoma associated fibroblasts (CAFs) play significant roles in affecting the sensitivity to a several chemotherapy agents [118]. Also it has been shown that targeting the fibroblasts can considerably decrease the chemoresistance [118]. These carcinoma associated fibroblasts increase chemo-resistance through CAF-derived exosomes.

Extracellular exosome is the top enriched cellular component, and exosomes are evidently known as a mark of metastatic cells. Their involvement in the expelling of drugs by carrying ABC transporters supports the exosomes function in drug resistance [119]. A number of Rab family proteins are among the key regulators of exosome secretion [119].

The only significantly enriched KEGG pathway using the list of up-regulated gene is the biosynthesis of antibiotics and for the list of down-regulated genes the ribosome pathway is enriched (Table 4.7). This could show the significant changes in the drug efflux in cells that may have resulted in the resistance. Ribosomal proteins are linked to proliferation, progression, and metastasis of colorectal cancer cells [120].

Several proteins with significant changes in the resistant cells, such as PAGE4 and cytokeratins have been known for malignant types of cancer, cancer stem cells, or links to EMT in cancer cells. Also, the cells skeletal features, adhesion and migration showed significant differences in the drug-resistant state. Exploring the proteome allowed us to recognize the mechanisms and pathways that were differentially regulated. Proteins that are related to cell-cell adhesion, actin binding, and ribosomal pathway have decreased expression levels (Table 4.3). Whereas, proteins related to the cell proliferation processes and biosynthesis of antibiotics pathway are increased in the resistant cell line (Table 4.5). While cells are more aggressively growing, it can be expected to have a decrease in RNA processes, protein folding, translation, and ribosome pathway.

Chapter 5

Conclusions

5.1 Quantitative proteomic analysis of inducible CHO cells

In this thesis, a quantitative proteomic analysis using iTRAQ technique was performed to compare the inducible CHO cells with cumate treatment while producing high amounts of the heavy chain antibody, EG2-hFc1. The inducible CHO cell line with the CR5 promoter and cumate gene switch were able to produce almost six times the amount of mAb product in presence of cumate. The successful iTRAQ labelling and mass spectrometry analysis allowed us to quantify the expression level of 3,410 proteins. A list of 1,616 unique proteins with a valid expression levels was used in differential expression analysis, which identified 82 proteins with a p-value lower than 0.05. After the corrections for multiple hypothesis testing an FDR threshold of 0.25 would detect 8 proteins that have significantly regulated in addition to the mAb product. Three proteins had an increase in the expression levels, and five of them had a decrease, but none of the significant changes were large.

The differentially regulated proteins were precursor metabolites, involved in metabolic pathways and energy uptake. Biological processes and functions of the candidate genes included translation, mRNA splicing, protein folding and protein modifications. However, these proteins did not show large changes in the cumate induced cells and all of the proteins had a log-fold change lower than 0.3 and greater than -0.4. Therefore it is likely that the cumate induction system is not causing larger metabolic shifts or stress responses within the cell. However, further investigation could help in understanding the system better, for example by testing the metabolites during the cell culture and mAb production at different scales, testing the cells and metabolites at several time intervals would also add more information about the induced cells.

Overall, the cells did not show to have a strong stress response in producing high levels of mAb, and small changes in the metabolism and translational pathways are identified. The two conditions did not have

any large significant difference in the proteome that can show this inducible expression system has successfully generated a stable cell line with higher production capacity.

My work in this thesis explores the changes in quantitative protein expression levels of the inducible CHO cell line associated with increased EG2-hFc1 production. The outcome and results can be a reference for any future studies performed using this cell line or the inducible expression system. Based on this proteomic study, the increase in production of this mAb did not significantly change the cells main metabolisms to induce stress responses. Other bioinformatics analysis of cells with this inducible system, or with various levels of production, or with another mAb product would provide more insights into the cellular changes that may be changing with the increase in protein secretion levels. Additionally, proteomic and metabolomics study of the cells at different time frames and different scales of production will provide extensive information on the changes related to growth and production process. In addition to further experimental studies that could help understanding the cells process, more analysis on the published data such as genes with unknown functions could be beneficial.

5.2 Quantitative proteomic analysis of drug-resistant HT-29 cell line

In the second part of the thesis, a quantitative proteomic analysis was performed for comparison of the HT-29 and HT-29S cell lines. The HT-29S cells are developed to become resistant to SN-38 treatment, and this analysis would help us in understanding the details of the resistance mechanism in the CRC cells. The cell line was generated with an increasing exposure to the drug treatment during the culture.

Using the iTRAQ labeling technique, peptide samples from 4 replicates of the HT-29 cell line and 4 replicates of the resistant HT-29S cell line were labeled and prepared for MS analysis. Then we were able to test the expression of 1,665 unique proteins for any differences between the two groups and 370 proteins were identified with statistically significant differences and a fold-change greater than 0.3.

Within the up-regulated gene list biosynthesis of antibiotics pathway was significantly enriched with 12 related genes, and the ribosome pathway was identified as a significantly enriched pathway among the down-regulated genes list (Table 4.7). Since the identification of several ribosomal proteins that had been down-regulated in the resistant cell line, regulation of ribosomal proteins is associated with the progression of CRC. The protein, PAGE4 is identified as the most significantly up-regulated protein in the resistant cells, this protein was previously known for its functions in the prostate cancer cells, but this is the first identification in HT-29 cells. These results could be further investigated by comparing to information from patients' tumor studies.

Transcription factors are known to have a role in regulating a number of cancer cell processes, for example the EMT process, but this method may be weak in including and quantifying some types of proteins, such as the transcription factors. A future study on the proteome or transcriptome of the resistant cells could be focused on considering transcription factors. Further analysis on the proteins with unknown functions or the network analysis could be done on the published data, as well as integrating information from external studies.

References

1. Strebhardt, K. and A. Ullrich, *Paul Ehrlich's magic bullet concept: 100 years of progress.* Nat Rev Cancer, 2008. **8**(6): p. 473-480.
2. Kohler, G. and C. Milstein, *Continuous Cultures of Fused Cells Secreting Antibody of Predefined Specificity.* Nature, 1975. **256**(5517): p. 495-497.
3. Weiner, G.J., *Building better monoclonal antibody-based therapeutics.* Nat Rev Cancer, 2015. **15**(6): p. 361-370.
4. Liu, J.K., *The history of monoclonal antibody development - Progress, remaining challenges and future innovations.* Ann Med Surg (Lond), 2014. **3**(4): p. 113-6.
5. EvaluatePharma, *World Preview 2014, Outlook to 2020.* Evaluate, Ltd.: Boston, MA, 2014, 2014.
6. Ecker, D.M., S.D. Jones, and H.L. Levine, *The therapeutic monoclonal antibody market.* Mabs, 2015. **7**(1): p. 9-14.
7. Kim, J.Y., Y.G. Kim, and G.M. Lee, *CHO cells in biotechnology for production of recombinant proteins: current state and further potential.* Appl Microbiol Biot, 2012. **93**(3): p. 917-930.
8. Xu, X., et al., *The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line.* Nat Biotechnol, 2011. **29**(8): p. 735-41.
9. Demain, A.L. and P. Vaishnav, *Production of recombinant proteins by microbes and higher organisms.* Biotechnol Adv, 2009. **27**(3): p. 297-306.
10. Ko, K., *Expression of recombinant vaccines and antibodies in plants.* Monoclon Antib Immunodiagn Immunother, 2014. **33**(3): p. 192-8.
11. Chiu, M.L. and G.L. Gilliland, *Engineering antibody therapeutics.* Curr Opin Struct Biol, 2016. **38**: p. 163-73.
12. Sola, R.J. and K. Griebenow, *Glycosylation of Therapeutic Proteins An Effective Strategy to Optimize Efficacy.* Biodrugs, 2010. **24**(1): p. 9-21.
13. Datta, P., R.J. Linhardt, and S.T. Sharfstein, *An 'omics approach towards CHO cell engineering.* Biotechnol Bioeng, 2013. **110**(5): p. 1255-1271.
14. Werner, R.G., et al., *Appropriate mammalian expression systems for biopharmaceuticals.* Arzneimittel-Forsch, 1998. **48**(8): p. 870-880.

15. Jayapal, K.R., et al., *Recombinant protein therapeutics from CHO cells - 20 years and counting.* Chem Eng Prog, 2007. **103**(10): p. 40-47.

16. Tjio, J.H. and T.T. Puck, *Genetics of somatic mammalian cells. II. Chromosomal constitution of cells in tissue culture.* J Exp Med, 1958. **108**(2): p. 259-68.

17. Shire, S.J., *Formulation and manufacturability of biologics.* Curr Opin Biotech, 2009. **20**(6): p. 708-714.

18. Rita Costa, A., et al., *Guidelines to cell engineering for monoclonal antibody production.* Eur J Pharm Biopharm, 2010. **74**(2): p. 127-38.

19. Wuest, D.M., S.W. Harcum, and K.H. Lee, *Genomics in mammalian cell culture bioprocessing.* Biotechnol Adv, 2012. **30**(3): p. 629-38.

20. Cockett, M.I., C.R. Bebbington, and G.T. Yarranton, *High level expression of tissue inhibitor of metalloproteinases in Chinese hamster ovary cells using glutamine synthetase gene amplification.* Biotechnology (N Y), 1990. **8**(7): p. 662-7.

21. Alt, F.W., et al., *Selective multiplication of dihydrofolate reductase genes in methotrexate-resistant variants of cultured murine cells.* J Biol Chem, 1978. **253**(5): p. 1357-70.

22. Lee, J.S., et al., *Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway.* Sci Rep-UK, 2015. **5**.

23. Malphettes, L., et al., *Highly Efficient Deletion of FUT8 in CHO Cell Lines Using Zinc-Finger Nucleases Yields Cells That Produce Completely Nonfucosylated Antibodies.* Biotechnol Bioeng, 2010. **106**(5): p. 774-783.

24. Khan, K.H., *Gene expression in Mammalian cells and its applications.* Adv Pharm Bull, 2013. **3**(2): p. 257-63.

25. Broussau, S., et al., *Inducible packaging cells for large-scale production of lentiviral vectors in serum-free suspension culture.* Mol Ther, 2008. **16**(3): p. 500-7.

26. Gaillet, B., et al., *High-level recombinant protein production in CHO cells using lentiviral vectors and the cumate gene-switch.* Biotechnol Bioeng, 2010. **106**(2): p. 203-15.

27. Mullick, A., et al., *The cumate gene-switch: a system for regulated expression in mammalian cells.* BMC Biotechnol, 2006. **6**: p. 43.

28. Yee, J.C., et al., *Genomic and proteomic exploration of CHO and hybridoma cells under sodium butyrate treatment.* Biotechnol Bioeng, 2008. **99**(5): p. 1186-204.

29. Baycin-Hizal, D., et al., *Proteomic analysis of Chinese hamster ovary cells.* J Proteome Res, 2012. **11**(11): p. 5265-76.

30. Meleady, P., et al., *Utilization and evaluation of CHO-specific sequence databases for mass spectrometry based proteomics.* Biotechnol Bioeng, 2012. **109**(6): p. 1386-94.

31. Orellana, C.A., et al., *High-antibody-producing Chinese hamster ovary cells up-regulate intracellular protein transport and glutathione synthesis.* J Proteome Res, 2015. **14**(2): p. 609-18.

32. Chong, W.P., et al., *LC-MS-based metabolic characterization of high monoclonal antibody-producing Chinese hamster ovary cells.* Biotechnol Bioeng, 2012. **109**(12): p. 3103-11.

33. Hammond, W.A., A. Swaika, and K. Mody, *Pharmacologic resistance in colorectal cancer: a review.* Ther Adv Med Oncol, 2016. **8**(1): p. 57-84.

34. Seymour, M.T., et al., *Different strategies of sequential and combination chemotherapy for patients with poor prognosis advanced colorectal cancer (MRC FOCUS): a randomised controlled trial.* Lancet, 2007. **370**(9582): p. 143-52.

35. Hsiang, Y.H., M.G. Lihou, and L.F. Liu, *Arrest of replication forks by drug-stabilized topoisomerase I-DNA cleavable complexes as a mechanism of cell killing by camptothecin.* Cancer Res, 1989. **49**(18): p. 5077-82.

36. Pommier, Y., *Topoisomerase I inhibitors: camptothecins and beyond.* Nat Rev Cancer, 2006. **6**(10): p. 789-802.

37. Wall, M.E., *Camptothecin and taxol: discovery to clinic.* Med Res Rev, 1998. **18**(5): p. 299-314.

38. Cummings, J., et al., *Factors influencing the cellular accumulation of SN-38 and camptothecin.* Cancer Chemother Pharmacol, 2002. **49**(3): p. 194-200.

39. Hsiang, Y.H., et al., *Camptothecin induces protein-linked DNA breaks via mammalian DNA topoisomerase I.* J Biol Chem, 1985. **260**(27): p. 14873-8.

40. Longley, D.B. and P.G. Johnston, *Molecular mechanisms of drug resistance.* J Pathol, 2005. **205**(2): p. 275-92.

41. Cummings, J., et al., *Enhanced clearance of topoisomerase I inhibitors from human colon cancer cells by glucuronidation.* Biochem Pharmacol, 2002. **63**(4): p. 607-13.

42. Gottesman, M.M., T. Fojo, and S.E. Bates, *Multidrug resistance in cancer: role of ATP-dependent transporters.* Nat Rev Cancer, 2002. **2**(1): p. 48-58.

43. Goldman, B., *Multidrug resistance: can new drugs help chemotherapy score against cancer?* J Natl Cancer Inst, 2003. **95**(4): p. 255-7.

44. Thomas, H. and H.M. Coley, *Overcoming multidrug resistance in cancer: an update on the clinical strategy of inhibiting p-glycoprotein.* Cancer Control, 2003. **10**(2): p. 159-65.

45. Chiang, S.C., J. Carroll, and S.F. El-Khamisy, *TDP1 serine 81 promotes interaction with DNA ligase IIIalpha and facilitates cell survival following DNA damage.* Cell Cycle, 2010. **9**(3): p. 588-595.

46. Zhou, T., et al., *Tyrosyl-DNA phosphodiesterase and the repair of 3'-phosphoglycolate-terminated DNA double-strand breaks.* DNA Repair (Amst), 2009. **8**(8): p. 901-11.

47. Alagoz, M., et al., *DNA repair and resistance to topoisomerase I inhibitors: mechanisms, biomarkers and therapeutic targets.* Curr Med Chem, 2012. **19**(23): p. 3874-85.

48. Walker, S., et al., *Development of an oligonucleotide-based fluorescence assay for the identification of tyrosyl-DNA phosphodiesterase 1 (TDP1) inhibitors.* Anal Biochem, 2014. **454**: p. 17-22.

49. Meisenberg, C., et al., *Clinical and cellular roles for TDP1 and TOP1 in modulating colorectal cancer response to irinotecan.* Mol Cancer Ther, 2015. **14**(2): p. 575-85.

50. Lengauer, C., K.W. Kinzler, and B. Vogelstein, *Genetic instabilities in human cancers.* Nature, 1998. **396**(6712): p. 643-9.

51. Berg, S., *Alterations in Colorectal Cancer Cell-extracellular Matrix Interactions Upon Acquisition of Chemotherapy Resistance*. 2015, University of Waterloo: Waterloo. p. 125.

52. van den Heuvel-Eibrink, M.M., et al., *Increased expression of the breast cancer resistance protein (BCRP) in relapsed or refractory acute myeloid leukemia (AML).* Leukemia, 2002. **16**(5): p. 833-9.

53. Chan, L.M., S. Lowes, and B.H. Hirst, *The ABCs of drug transport in intestine and liver: efflux proteins limiting drug absorption and bioavailability.* Eur J Pharm Sci, 2004. **21**(1): p. 25-51.

54. Mathijssen, R.H., et al., *Irinotecan pathway genotype analysis to predict pharmacokinetics.* Clin Cancer Res, 2003. **9**(9): p. 3246-53.

55. Candeil, L., et al., *ABCG2 overexpression in colon cancer cells resistant to SN-38 and in irinotecan-treated metastases.* Int J Cancer, 2004. **109**(6): p. 848-54.

56. Petitprez, A., et al., *Acquired irinotecan resistance is accompanied by stable modifications of cell cycle dynamics independent of MSI status.* Int J Oncol, 2013. **42**(5): p. 1644-53.

57. Yu, M., A. Ocana, and I.F. Tannock, *Reversal of ATP-binding cassette drug transporter activity to modulate chemoresistance: why has it failed to provide clinical benefit?* Cancer Metastasis Rev, 2013. **32**(1-2): p. 211-27.

58. Ballesta, A., et al., *A combined experimental and mathematical approach for molecular-based optimization of irinotecan circadian delivery.* PLoS Comput Biol, 2011. **7**(9): p. e1002143.

59. Ruoslahti, E., et al., *Fibroblast surface antigen: a new serum protein.* Biochim Biophys Acta, 1973. **322**(2): p. 352-8.

60. Hynes, R.O., *The emergence of integrins: a personal and historical perspective.* Matrix Biol, 2004. **23**(6): p. 333-40.

61. Hynes, R.O., *Integrins: bidirectional, allosteric signaling machines.* Cell, 2002. **110**(6): p. 673-87.

62. Kim, C., et al., *Basic amino-acid side chains regulate transmembrane integrin signalling.* Nature, 2011. **481**(7380): p. 209-13.

63. Kim, C., F. Ye, and M.H. Ginsberg, *Regulation of integrin activation.* Annu Rev Cell Dev Biol, 2011. **27**: p. 321-45.

64. Shattil, S.J., C. Kim, and M.H. Ginsberg, *The final steps of integrin activation: the end game.* Nat Rev Mol Cell Biol, 2010. **11**(4): p. 288-300.

65. Bartolome, R.A., et al., *Cadherin-17 interacts with alpha2beta1 integrin to regulate cell proliferation and adhesion in colorectal cancer cells causing liver metastasis.* Oncogene, 2014. **33**(13): p. 1658-69.

66. Pelillo, C., et al., *Colorectal Cancer Metastases Settle in the Hepatic Microenvironment Through alpha5beta1 Integrin.* J Cell Biochem, 2015. **116**(10): p. 2385-96.

67. Huhtala, P., et al., *Cooperative signaling by alpha 5 beta 1 and alpha 4 beta 1 integrins regulates metalloproteinase gene expression in fibroblasts adhering to fibronectin.* J Cell Biol, 1995. **129**(3): p. 867-79.

68. Yang, B., et al., *Clinicopathological and prognostic significance of alpha5beta1-integrin and MMP-14 expressions in colorectal cancer.* Neoplasma, 2013. **60**(3): p. 254-61.

69. Frisch, S.M. and H. Francis, *Disruption of epithelial cell-matrix interactions induces apoptosis.* J Cell Biol, 1994. **124**(4): p. 619-26.

70. Damiano, J.S., et al., *Cell adhesion mediated drug resistance (CAM-DR): role of integrins and resistance to apoptosis in human myeloma cell lines.* Blood, 1999. **93**(5): p. 1658-67.

71. De Toni, F., et al., *A crosstalk between the Wnt and the adhesion-dependent signaling pathways governs the chemosensitivity of acute myeloid leukemia.* Oncogene, 2006. **25**(22): p. 3113-22.

72. Ross, P.L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.* Mol Cell Proteomics, 2004. **3**(12): p. 1154-69.

73. Zieske, L.R., *A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies.* J Exp Bot, 2006. **57**(7): p. 1501-8.

74. Johnson, R.S., et al., *Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine.* Anal Chem, 1987. **59**(21): p. 2621-5.

75. Ma, B., et al., *PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry.* Rapid Commun Mass Spectrom, 2003. **17**(20): p. 2337-42.

76. Roepstorff, P. and J. Fohlman, *Proposal for a common nomenclature for sequence ions in mass spectra of peptides.* Biomed Mass Spectrom, 1984. **11**(11): p. 601.

77. Zhang, J., et al., *PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification.* Mol Cell Proteomics, 2012. **11**(4): p. M111 010587.

78. Karpievitch, Y.V., et al., *Normalization of peak intensities in bottom-up MS-based proteomics using singular value decomposition.* Bioinformatics, 2009. **25**(19): p. 2573-80.

79. Bolstad, B.M., et al., *A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.* Bioinformatics, 2003. **19**(2): p. 185-93.

80. Smyth, G.K., *Linear models and empirical bayes methods for assessing differential expression in microarray experiments.* Stat Appl Genet Mol Biol, 2004. **3**: p. Article3.

81. Faul, F., et al., *Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses.* Behav Res Methods, 2009. **41**(4): p. 1149-60.

82. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.* Journal of the Royal Statistical Society. Series B (Methodological), 1995. **57**(1): p. 289-300.

83. Gentleman, R.C., et al., *Bioconductor: open software development for computational biology and bioinformatics.* Genome Biol, 2004. **5**(10): p. R80.

84. Dennis, G., Jr., et al., *DAVID: Database for Annotation, Visualization, and Integrated Discovery.* Genome Biol, 2003. **4**(5): p. P3.

85. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.

86. Huang, D.W., et al., *DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W169-75.

87. Ciardiello, F. and F. De Vita, *Epidermal growth factor receptor (EGFR) inhibitors in cancer therapy.* Prog Drug Res, 2005. **63**: p. 93-114.

88. Poulain, A., et al., *Rapid protein production from stable CHO cell pools using plasmid vector and the cumate gene-switch.* J Biotechnol, 2017. **255**: p. 16-27.

89. Han, J.C. and G.Y. Han, *A procedure for quantitative determination of tris(2-carboxyethyl)phosphine, an odorless reducing agent more stable and effective than dithiothreitol.* Anal Biochem, 1994. **220**(1): p. 5-10.

90. Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.* Anal Biochem, 1976. **72**: p. 248-54.

91. Wisniewski, J.R., A. Zougman, and M. Mann, *Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome.* J Proteome Res, 2009. **8**(12): p. 5674-8.

92. Wisniewski, J.R., et al., *Universal sample preparation method for proteome analysis.* Nat Methods, 2009. **6**(5): p. 359-62.

93. Desjardins, P., J.B. Hansen, and M. Allen, *Microvolume protein concentration determination using the NanoDrop 2000c spectrophotometer.* J Vis Exp, 2009(33).

94. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.

95. Challa, A.A. and B. Stefanovic, *A novel role of vimentin filaments: binding and stabilization of collagen mRNAs.* Mol Cell Biol, 2011. **31**(18): p. 3773-89.

96. Zeng, Y., et al., *The cancer/testis antigen prostate-associated gene 4 (PAGE4) is a highly intrinsically disordered protein.* J Biol Chem, 2011. **286**(16): p. 13985-94.

97. Jia, D., et al., *Phenotypic Plasticity and Cell Fate Decisions in Cancer: Insights from Dynamical Systems Theory.* Cancers (Basel), 2017. **9**(7).

98. Kulkarni, P., et al., *Prostate-associated gene 4 (PAGE4), an intrinsically disordered cancer/testis antigen, is a novel therapeutic target for prostate cancer.* Asian J Androl, 2016. **18**(5): p. 695-703.

99. Iavarone, C., et al., *PAGE4 is a cytoplasmic protein that is expressed in normal prostate and in prostate cancers.* Mol Cancer Ther, 2002. **1**(5): p. 329-35.

100. Duan, Z., et al., *Overexpression of MAGE/GAGE genes in paclitaxel/doxorubicin-resistant human cancer cell lines.* Clin Cancer Res, 2003. **9**(7): p. 2778-85.

101. Rajagopalan, K., et al., *The Stress-response protein prostate-associated gene 4, interacts with c-Jun and potentiates its transactivation.* Biochim Biophys Acta, 2014. **1842**(2): p. 154-63.

102. Kato, T., et al., *Increased expression of insulin-like growth factor-II messenger RNA-binding protein 1 is associated with tumor progression in patients with lung cancer.* Clin Cancer Res, 2007. **13**(2 Pt 1): p. 434-42.

103. Vikesaa, J., et al., *RNA-binding IMPs promote cell adhesion and invadopodia formation.* EMBO J, 2006. **25**(7): p. 1456-68.

104. Livingstone, C., *IGF2 and cancer.* Endocr Relat Cancer, 2013. **20**(6): p. R321-39.

105. Clemmons, D.R., *Insulin-like growth factor binding proteins and their role in controlling IGF actions.* Cytokine Growth Factor Rev, 1997. **8**(1): p. 45-62.

106. Jones, J.I., et al., *Insulin-like growth factor binding protein 1 stimulates cell migration and binds to the alpha 5 beta 1 integrin by means of its Arg-Gly-Asp sequence.* Proc Natl Acad Sci U S A, 1993. **90**(22): p. 10553-7.

107. Sinha, P., et al., *Increased expression of epidermal fatty acid binding protein, cofilin, and 14-3-3-sigma (stratifin) detected by two-dimensional gel electrophoresis, mass spectrometry and microsequencing of drug-resistant human adenocarcinoma of the pancreas.* Electrophoresis, 1999. **20**(14): p. 2952-60.

108. Samyesudhas, S.J., L. Roy, and K.D. Cowden Dahl, *Differential expression of ARID3B in normal adult tissue and carcinomas.* Gene, 2014. **543**(1): p. 174-80.

109. Honda, K., et al., *Actinin-4 increases cell motility and promotes lymph node metastasis of colorectal cancer.* Gastroenterology, 2005. **128**(1): p. 51-62.

110. Grass, G.D., et al., *CD147: regulator of hyaluronan signaling in invasiveness and chemoresistance.* Adv Cancer Res, 2014. **123**: p. 351-73.

111. Shay, G., C.C. Lynch, and B. Fingleton, *Moving targets: Emerging roles for MMPs in cancer progression and metastasis.* Matrix Biol, 2015. **44-46**: p. 200-6.

112. Moll, R., et al., *Cytokeratin 20 in human carcinomas. A new histodiagnostic marker detected by monoclonal antibodies.* Am J Pathol, 1992. **140**(2): p. 427-47.

113. Kust, D., et al., *Cytokeratin 20 positive cells in blood of colorectal cancer patients as an unfavorable prognostic marker.* Acta Clin Belg, 2016. **71**(4): p. 235-43.

114. Kim, M.J., et al., *Profilin 2 promotes migration, invasion, and stemness of HT-29 human colorectal cancer stem cells.* Biosci Biotechnol Biochem, 2015. **79**(9): p. 1438-46.

115. Mooney, S.M., et al., *Creatine kinase brain overexpression protects colorectal cells from various metabolic and non-metabolic stresses.* J Cell Biochem, 2011. **112**(4): p. 1066-75.

116. Lagresle-Peyrou, C., et al., *X-linked primary immunodeficiency associated with hemizygous mutations in the moesin (MSN) gene.* J Allergy Clin Immunol, 2016. **138**(6): p. 1681-1689 e8.

117. O'Brien, C.A., et al., *A human colon cancer cell capable of initiating tumour growth in immunodeficient mice.* Nature, 2007. **445**(7123): p. 106-10.

118. Loeffler, M., et al., *Targeting tumor-associated fibroblasts improves cancer chemotherapy by increasing intratumoral drug uptake.* J Clin Invest, 2006. **116**(7): p. 1955-62.

119. Azmi, A.S., B. Bao, and F.H. Sarkar, *Exosomes in cancer development, metastasis, and drug resistance: a comprehensive review.* Cancer Metastasis Rev, 2013. **32**(3-4): p. 623-42.

120. Lai, M.D. and J. Xu, *Ribosomal proteins and colorectal cancer.* Curr Genomics, 2007. **8**(1): p. 43-9.

Appendix A

Table A: List of CHO proteins identified with p-values smaller than 0.05

Gene Symbol	Protein GI	Description	p-value	log FC
-	-	eg2-hfc1 [Lama glama-Homo sapiens]	0.000	1.502
Upar	354492547	urokinase plasminogen activator surface receptor partial	0.000	-0.300
Trmt112	354498436	tRNA methyltransferase 112 homolog	0.001	0.333
LOC100754718	354492600	protein phosphatase 1 regulatory subunit 12A-like	0.001	-0.088
Mrpl55	354482471	39S ribosomal protein L55 mitochondrial-like	0.002	-0.439
Gstp1	350537543	glutathione S-transferase P 1	0.002	-0.178
Rps7	354495303	40S ribosomal protein S7-like	0.002	0.163
Lsm3	354465501	U6 snRNA-associated Sm-like protein LSm3-like	0.003	-0.203
Fkbp11	354497358	peptidyl-prolyl cis-trans isomerase FKBP11-like	0.003	0.260
LOC100758296	354491098	poly(U)-binding-splicing factor PUF60-like	0.004	-0.134
Dnaja1	354498827	LOW QUALITY PROTEIN: dnaJ homolog subfamily A member 1	0.005	-0.143
Tra2b	354484188	transformer-2 protein homolog beta-like	0.006	0.281
Bgn	354488869	biglycan-like	0.006	-0.117
Ganab	354493306	neutral alpha-glucosidase AB isoform 1	0.006	0.205
Lamc1	354481424	laminin subunit gamma-1	0.006	-0.287
Synj1	354466272	synaptojanin-1 isoform 2	0.008	-0.120
Fam3a	354500495	protein FAM3A-like	0.008	-0.204
Vim	354482483	vimentin	0.008	-0.266
LOC100756788	354466573	hematological and neurological expressed 1 protein-like	0.009	-0.253
LOC100772255	354481494	stress-associated endoplasmic reticulum protein 1-like partial	0.009	0.345
Wdr61	354471439	WD repeat-containing protein 61-like isoform 2	0.010	-0.135
G3bp2	354499120	ras GTPase-activating protein-binding protein 2 isoform 1	0.011	-0.064
Psma1	354488597	proteasome subunit alpha type-1-like	0.011	-0.103
Gfm1	354496448	elongation factor G mitochondrial-like	0.012	-0.072
LOC100753707	354507225	40S ribosomal protein S6-like partial	0.012	0.212
Epb4.1l2	354494678	band 4.1-like protein 2	0.013	-0.087
Col6a1	354476788	collagen alpha-1(VI) chain	0.013	-0.162
Gcsh	354493382	glycine cleavage system H protein mitochondrial-like	0.013	-0.126
LOC100755252	354474224	thymidylate kinase-like	0.014	-0.144
Mesd	354499297	LDLR chaperone MESD-like	0.014	-0.078
Asna1	354479529	ATPase Asna1	0.014	-0.376
Hmgcs1	354485654	hydroxymethylglutaryl-CoA synthase cytoplasmic-like	0.015	-0.137
Rbm12	354477974	RNA-binding protein 12	0.016	-0.105

Tuba1c	345842428	tubulin alpha-1C chain	0.016	-0.158
LOC100766608	354486253	26S protease regulatory subunit 6B-like	0.016	-0.113
LOC100771019	354488039	polyadenylate-binding protein 2-like	0.017	-0.244
Rrbp1	354468128	ribosome-binding protein 1	0.018	0.079
LOC100774114	354492313	60S ribosomal protein L28-like	0.019	0.106
Mthfd1l	354466920	monofunctional C1-tetrahydrofolate synthase mitochondrial	0.019	-0.130
LOC100763481	354477579	hypothetical protein LOC100763481	0.021	0.199
Col6a2	354476782	collagen alpha-2(VI) chain isoform 1	0.021	-0.138
Fh	354475349	fumarate hydratase mitochondrial-like	0.022	-0.107
Tkt	354467411	transketolase-like	0.022	0.085
Glul	354481440	glutamine synthetase-like	0.023	0.282
Ube2o	354489489	ubiquitin-conjugating enzyme E2 O	0.023	-0.277
LOC100755568	354484186	insulin-like growth factor 2 mRNA-binding protein 2-like isoform 2	0.025	-0.270
LOC100767827	354507734	UDP-glucose 6-dehydrogenase-like partial	0.026	-0.076
LOC100760000	354481833	acetyl-CoA acetyltransferase cytosolic-like	0.026	-0.069
Sh3pxd2a	354500209	SH3 and PX domain-containing protein 2A	0.028	-0.160
LOC100769327	354499764	UPF0568 protein C14orf166 homolog	0.028	-0.224
Rbck1	354498892	ranBP-type and C3HC4-type zinc finger-containing protein 1 partial	0.029	-0.112
LOC100774828	354504685	methylosome protein 50-like	0.030	-0.196
Tsr2	354476053	pre-rRNA-processing protein TSR2 homolog	0.031	-0.119
Aip	354495724	AH receptor-interacting protein	0.031	-0.126
Slc27a4	354499491	long-chain fatty acid transport protein 4	0.032	-0.233
Itga5	354497226	integrin alpha-5	0.033	-0.102
Lta4h	354478523	leukotriene A-4 hydrolase	0.033	-0.100
Icam1	350539683	intercellular adhesion molecule 1	0.033	-0.220
Anxa1	354496812	annexin A1-like	0.035	0.172
Dnase2	354479537	deoxyribonuclease-2-alpha-like	0.035	-0.186
Fn1	354494385	fibronectin isoform 3	0.037	-0.131
COX2	91176206	cytochrome c oxidase subunit II	0.037	-0.270
Cd44	354470417	CD44 antigen isoform 3	0.039	-0.146
Srsf3	354484028	serine/arginine-rich splicing factor 3-like	0.039	0.086
Rpl31	354482378	60S ribosomal protein L31-like	0.040	0.130
LOC100762308	354473557	protein-L-isoaspartate(D-aspartate) O-methyltransferase-like	0.040	-0.076
Pam16	354488489	coronin-7	0.040	-0.104
Hspg	354483018	LOW QUALITY PROTEIN: basement membrane-specific heparan sulfate proteoglycan core protein	0.040	-0.237
Erbb2ip	354474057	protein LAP2	0.041	-0.330
LOC100757077	354503831	protein CTLA-2-beta-like	0.042	-0.238
Ddt	354492004	D-dopachrome decarboxylase-like	0.042	-0.083

LOC100762458	354495283	nuclear mitotic apparatus protein 1-like partial	0.043	-0.099
Ccdc115	354491621	coiled-coil domain-containing protein 115-like	0.044	-0.212
LOC100751694	354502651	14-3-3 protein theta-like	0.044	-0.193
Sf3b5	354475529	splicing factor 3B subunit 5-like	0.046	-0.080
Dctpp1	354506130	LOW QUALITY PROTEIN: dCTP pyrophosphatase 1-like	0.046	-0.379
Pea15	349501078	astrocytic phosphoprotein PEA-15	0.048	-0.151
Tom1	354499827	target of Myb protein 1	0.049	-0.356
Tubb4b	346644707	tubulin beta-2C chain	0.049	-0.125
Ltv1	354475527	protein LTV1 homolog	0.049	0.076
Galm	354488923	aldose 1-epimerase-like	0.049	-0.275
LOC100756352	354486710	hypothetical protein LOC100756352	0.050	-0.096

Appendix B

Table B: List of significantly differentially expressed proteins in HT-29S cell line with 5% Benjamini-Hochberg FDR correction.

Gene Symbol	Protein GI	Description	P-value	logFC
PAGE4	5901986	P antigen family member 4	1.2E-03	2.256
UCHL1	189067502	unnamed protein product	2.2E-05	1.660
FABP5	4557581	fatty acid-binding protein epidermal	2.8E-03	1.255
IGF2BP1	56237027	insulin-like growth factor 2 mRNA-binding protein 1 isoform 1	3.5E-04	1.196
FABP7	4557585	fatty acid-binding protein brain isoform 1	1.1E-03	1.177
ARID3A	56799575	DRIL3	3.3E-03	1.151
AKAP2	194385574	unnamed protein product	3.5E-07	1.093
EPB41L3	767997596	band 4.1-like protein 3 isoform X11	1.8E-04	0.981
PDLIM1	189054550	Carboxyl Terminal LIM Domain Protein 1	3.0E-06	0.892
ERVH48-1	819231746	suppressyn precursor	4.3E-03	0.847
TPM1	578827355	tropomyosin alpha-1 chain isoform X5	1.8E-03	0.846
CXorf67	42821110	uncharacterized protein CXorf67	1.0E-02	0.804
SET	119608226	SET translocation (myeloid leukemia-associated) isoform CRA_c	6.9E-04	0.800
ARG2	4502215	arginase-2 mitochondrial precursor	8.0E-03	0.789
MRFAP1	194384244	unnamed protein product	8.5E-03	0.770
QPRT	1060907	quinolinate phosphoribosyl transferase	8.2E-05	0.760
GINS4	14150122	DNA replication complex GINS protein SLD5	1.9E-03	0.749
KLRG2	51094790	FLJ44186 protein	1.1E-02	0.748
NENF	7019545	neudesin precursor	7.2E-07	0.739
MPP1	119593065	membrane protein palmitoylated 1 55kDa isoform CRA_e	2.4E-05	0.734
FSCN1	14043297	Similar to singed (Drosophila)-like (sea urchin fascin homolog like) partial	6.9E-04	0.719
PITPNA	767992345	phosphatidylinositol transfer protein alpha isoform isoform X1	9.5E-03	0.708
CLIC4	194376454	unnamed protein product	4.3E-03	0.708
PTK7	22902128	inactive tyrosine-protein kinase 7 isoform c precursor	4.7E-03	0.703
LCP1	62898171	L-plastin variant	3.8E-03	0.672
KRT7	767974195	keratin type II cytoskeletal 7 isoform X1	3.9E-04	0.671
ATP5B	32189394	ATP synthase subunit beta mitochondrial precursor	5.8E-06	0.667
RAE1	4506399	mRNA export factor	5.8E-05	0.666
TRIM28	5032179	transcription intermediary factor 1-beta	1.8E-03	0.661
ALAD	248841	delta-aminolevulinate dehydratase	9.4E-03	0.649

ATIC	20127454	bifunctional purine biosynthesis protein PURH	1.0E-03	0.628
SH2D4A	292658785	SH2 domain-containing protein 4A isoform b	1.4E-04	0.616
BCAT1	296010902	branched-chain-amino-acid aminotransferase cytosolic isoform 3	4.1E-04	0.591
CTSV	189053498	unnamed protein product	1.7E-03	0.589
TACSTD2	49457514	TACSTD2	3.5E-03	0.588
NPC2	189065149	unnamed protein product	2.3E-03	0.588
ALDH6A1	515870067	methylmalonate-semialdehyde dehydrogenase [acylating] mitochondrial isoform 3	1.0E-02	0.586
LRP2	767918153	low-density lipoprotein receptor-related protein 2 isoform X1	2.1E-03	0.584
PALLD	260656032	palladin isoform 3	5.4E-04	0.581
PHGDH	767903710	D-3-phosphoglycerate dehydrogenase isoform X1	5.3E-04	0.580
PLIN2	194382232	unnamed protein product	1.3E-03	0.578
C1orf52	19550363	gm117 form A	7.9E-05	0.572
CNPY2	7657176	protein canopy homolog 2 isoform 1 precursor	1.5E-02	0.556
SNX2	2827434	sorting nexin 2	2.6E-04	0.554
LY6E	119602691	lymphocyte antigen 6 complex locus E isoform CRA_a	1.6E-03	0.543
LAMB1	186837	laminin B1	2.7E-05	0.541
UBAC1	5759309	putative glialblastoma cell differentiation-related protein	2.4E-03	0.536
RAP1B	354459356	ras-related protein Rap-1b isoform 4	2.4E-03	0.535
TPT1	15214610	Tumor protein translationally-controlled 1	3.6E-07	0.532
AKAP12	60219476	hypothetical protein	2.8E-05	0.530
CCT4	375477430	T-complex protein 1 subunit delta isoform b	7.4E-03	0.529
RNF114	8923898	E3 ubiquitin-protein ligase RNF114	1.2E-03	0.529
AHSA1	6912280	activator of 90 kDa heat shock protein ATPase homolog 1 isoform 1	4.5E-05	0.524
ACAT2	12653279	ACAT2 protein	4.8E-03	0.524
IMPDH2	307066	inosine-5'-monophosphate dehydrogenase (EC 1.1.1.205)	1.9E-05	0.523
BSG	31076333	CD147	1.3E-05	0.519
HEBP1	20336761	heme-binding protein 1	2.3E-03	0.519
MTHFD1	14602585	Methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1 methenyltetrahydrofolate cyclohydrolase formyltetrahydrofolate synthetase	1.0E-04	0.518
HSPE1	4504523	10 kDa heat shock protein mitochondrial	7.1E-03	0.518
PLS3	62897161	plastin 3 variant	4.7E-03	0.509
SNRPD1	32959908	small nuclear ribonucleoprotein Sm D1	6.7E-03	0.506
DFFA	14043461	DNA fragmentation factor 45kDa alpha polypeptide	6.0E-03	0.500
SLC39A14	16041779	SLC39A14 protein	5.7E-03	0.500
SNRPD2	237649049	small nuclear ribonucleoprotein Sm D2 isoform 2	4.5E-03	0.494
ANXA3	4826643	annexin A3	4.6E-03	0.492
SKA3	260763912	spindle and kinetochore-associated protein 3 isoform 2	1.7E-03	0.490

KRAS	15488883	V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog	1.8E-03	0.487
ILF2	284261	interleukin enhancer-binding factor ILF-2 - human	1.9E-03	0.483
PGM3	119569054	phosphoglucomutase 3 isoform CRA_c	1.4E-03	0.482
REPS1	18203725	REPS1 protein	2.5E-05	0.480
SUGT1	119572429	SGT1 suppressor of G2 allele of SKP1 (S. cerevisiae) isoform CRA_b	6.7E-03	0.478
SPAG9	76779789	SPAG9 protein	5.0E-03	0.474
KIF2C	663071031	kinesin-like protein KIF2C isoform 2	3.8E-03	0.469
PSMC3	62896529	proteasome 26S ATPase subunit 3 variant	1.3E-04	0.467
POP7	153791431	ribonuclease P protein subunit p20	1.2E-02	0.463
QKI	7542353	QUAKING isoform 4	7.3E-04	0.461
HNRNPA1	194382016	unnamed protein product	1.6E-02	0.461
PRDX6	4758638	peroxiredoxin-6	2.2E-03	0.458
RPRD1A	119621775	hypothetical protein FLJ10656 isoform CRA_d	3.0E-03	0.458
NASP	767904447	nuclear autoantigenic sperm protein isoform X2	1.4E-03	0.448
SOD1	4507149	superoxide dismutase [Cu-Zn]	1.1E-02	0.444
UBQLN1	12060171	DA41	4.3E-04	0.443
CLNS1A	909618069	methylosome subunit pICln isoform d	1.1E-05	0.435
SNX3	4507143	sorting nexin-3 isoform a	2.9E-03	0.435
FABP3	4758328	fatty acid-binding protein heart isoform 2	3.7E-03	0.434
RPRD1B	21708061	Regulation of nuclear pre-mRNA domain containing 1B	6.5E-05	0.434
RPA3	119614009	replication protein A3 14kDa isoform CRA_a	7.6E-04	0.426
TPM4	4507651	tropomyosin alpha-4 chain isoform Tpm4.2cy	2.9E-04	0.421
FARSA	194386000	unnamed protein product	2.2E-05	0.420
CCAR2	767951088	cell cycle and apoptosis regulator protein 2 isoform X1	2.0E-03	0.419
LUC7L	14336684	putative RNA binding protein	4.7E-05	0.419
NELFA	119602953	Wolf-Hirschhorn syndrome candidate 2 isoform CRA_c	1.0E-02	0.417
KIAA1143	33468965	uncharacterized protein KIAA1143 isoform 1	1.8E-03	0.417
PFAS	31657129	phosphoribosylformylglycinamidine synthase	1.2E-04	0.417
DUT	4503423	deoxyuridine 5'-triphosphate nucleotidohydrolase mitochondrial isoform 2	1.4E-02	0.415
GARS	943350815	glycine--tRNA ligase isoform 2	1.2E-02	0.411
THRAP3	1009287627	thyroid hormone receptor-associated protein 3 isoform 2	1.4E-02	0.404
PMPCB	40226469	PMPCB protein partial	2.2E-04	0.403
FAM207A	948284426	protein FAM207A isoform 7	5.4E-03	0.402
CHORDC1	6581056	CHORD containing protein-1	1.4E-02	0.400
CA2	922664887	carbonic anhydrase 2 - Chain A Surface Lysine Acetylated Human Carbonic Anhydrase Ii In Complex With A Sulfamate-based Inhibitor	4.3E-03	0.400
VCP	6005942	transitional endoplasmic reticulum ATPase	1.5E-02	0.400
ATP6V1B2	13938355	ATP6V1B2 protein partial	9.3E-03	0.398

VTA1	556503352	vacuolar protein sorting-associated protein VTA1 homolog isoform b	2.2E-03	0.397
HPDL	14249394	4-hydroxyphenylpyruvate dioxygenase-like protein	5.1E-03	0.396
PDE12	767922620	2' 5'-phosphodiesterase 12 isoform X1	6.9E-03	0.395
SHTN1	385198097	shootin-1 isoform e	4.6E-04	0.395
FDX1	182747	ferredoxin partial	1.3E-02	0.391
PARD6B	62955042	partitioning defective 6 homolog beta	9.8E-03	0.391
PRPF6	119595583	chromosome 20 open reading frame 14 isoform CRA_d	7.4E-03	0.389
MCM3	20384693	cervical cancer proto-oncogene 5	1.3E-04	0.389
SNRPGP15	205829943	RecName: Full=Putative small nuclear ribonucleoprotein G-like protein 15	5.0E-04	0.388
MAPRE1	194386362	Microtubule Associated Protein RP/EB Family Member 1	8.9E-05	0.387
WDR77	951233338	methylosome protein 50 isoform 4	7.9E-03	0.385
RBPMS	5803141	RNA-binding protein with multiple splicing isoform A	1.1E-03	0.384
CNOT3	6599188	hypothetical protein	9.1E-05	0.384
NOC4L	119618951	nucleolar complex associated 4 homolog (S. cerevisiae) isoform CRA_b	1.1E-02	0.383
APEH	530372380	acylamino-acid-releasing enzyme isoform X6	1.3E-02	0.379
EIF4B	194388916	unnamed protein product	1.0E-02	0.378
DIS3	190014623	exosome complex exonuclease RRP44 isoform a	1.5E-03	0.376
ERP29	5803013	endoplasmic reticulum resident protein 29 isoform 1 precursor	4.5E-04	0.374
RBM26	767977967	RNA-binding protein 26 isoform X16	4.9E-03	0.373
AIM1L	767904873	absent in melanoma 1-like protein isoform X4	4.1E-05	0.370
XAGE2	19747283	X antigen family member 2	1.4E-02	0.369
PROSC	119583764	proline synthetase co-transcribed homolog (bacterial) isoform CRA_a	4.3E-04	0.365
LGALS3	28071074	unnamed protein product	6.6E-04	0.363
DAG1	294997282	dystroglycan preproprotein	5.2E-03	0.362
GNL3	119585654	guanine nucleotide binding protein-like 3 (nucleolar) isoform CRA_b	1.5E-02	0.360
OTUD5	209977067	OTU domain-containing protein 5 isoform c	1.1E-02	0.359
HLA-A	156601507	MHC class I antigen	6.2E-05	0.358
YAP1	530788252	transcriptional coactivator YAP1 isoform 9	1.5E-03	0.356
CLIC3	40288290	chloride intracellular channel protein 3	6.8E-06	0.354
UGP2	449441	UDP-glucose pyrophosphorylase	1.1E-02	0.353
CD2AP	530381720	CD2-associated protein isoform X1	3.3E-04	0.353
WASF1	4507913	wiskott-Aldrich syndrome protein family member 1	9.9E-03	0.351
SERPINH1	32454741	serpin H1 precursor	6.0E-04	0.350
RAD23B	194374237	unnamed protein product	3.5E-03	0.337
PSMC6	195539395	26S protease regulatory subunit 10B	4.2E-04	0.335
PUS1	767975389	tRNA pseudouridine synthase A mitochondrial isoform X2	4.3E-04	0.326
VCL	24657579	Vinculin	5.5E-03	0.325

EEF1A1	62896661	eukaryotic translation elongation factor 1 alpha 1 variant	5.3E-04	0.325
PDCD6IP	22027538	programmed cell death 6-interacting protein isoform 1	8.5E-03	0.323
SNX6	33337751	MSTP010	1.0E-03	0.321
HMGCS1	767934415	hydroxymethylglutaryl-CoA synthase cytoplasmic isoform X2	7.5E-03	0.317
PPP1R12A	767974349	protein phosphatase 1 regulatory subunit 12A isoform X13	5.7E-04	0.316
CCT5	12804225	Unknown (protein for IMAGE:3543711) partial	5.9E-03	0.315
ACP1	179661	cytoplasmic phosphotyrosyl protein phosphatase partial	1.2E-03	0.315
LMNB2	16306859	Lamin B2	1.2E-02	0.312
SH3GLB2	9910352	endophilin-B2 isoform b	3.8E-03	0.309
NFATC2IP	51873914	NFATC2IP protein partial	1.1E-02	0.309
CLPX	767983141	ATP-dependent Clp protease ATP-binding subunit clpX-like mitochondrial isoform X1	2.2E-03	0.309
UBAP2L	568214245	ubiquitin-associated protein 2-like isoform c	3.5E-03	0.305
PMPCA	194387808	unnamed protein product	5.8E-03	0.304
SAFB	221045036	unnamed protein product	9.7E-03	0.304
TMED1	5803040	transmembrane emp24 domain-containing protein 1 precursor	3.2E-03	0.302
PDHB	119585776	pyruvate dehydrogenase (lipoamide) beta isoform CRA_b	1.1E-03	0.299
PCCB	221044434	propionyl-CoA carboxylase	1.4E-03	0.293
PTPRF	530363076	receptor-type tyrosine-protein phosphatase F isoform X9	6.4E-03	0.291
ENO1	4503571	alpha-enolase isoform 1	1.3E-02	0.291
PPP1R8	20336239	nuclear inhibitor of protein phosphatase 1 isoform beta/delta	2.1E-03	0.288
CCDC43	119571973	coiled-coil domain containing 43 isoform CRA_a	3.9E-03	0.286
MCM4	193785697	unnamed protein product	1.1E-05	0.286
CCT3	194374631	unnamed protein product	6.2E-04	0.282
PRPF3	194386352	unnamed protein product	1.3E-02	0.282
XRCC5	62988844	unknown	6.4E-03	0.281
ERP44	119579326	thioredoxin domain containing 4 (endoplasmic reticulum) isoform CRA_a partial	6.2E-04	0.281
SEPHS1	307219240	selenide water dikinase 1 isoform 3	5.2E-03	0.278
SRA1	9930614	steroid receptor RNA activator isoform 3	2.4E-03	0.276
PPT1	194377188	unnamed protein product	2.4E-03	0.274
PHPT1	12052814	hypothetical protein	1.0E-02	0.271
BCL9L	530398008	B-cell CLL/lymphoma 9-like protein isoform X3	1.3E-02	0.271
FKBP4	767971249	peptidyl-prolyl cis-trans isomerase FKBP4 isoform X1	1.5E-02	0.271
DNMT1	4503351	DNA (cytosine-5)-methyltransferase 1 isoform b	9.3E-04	0.271
PPP6C	183603931	serine/threonine-protein phosphatase 6 catalytic subunit isoform c	8.7E-03	0.267
SRRT	194382242	unnamed protein product	1.1E-02	0.266

SMAP1	194382546	unnamed protein product	1.4E-03	0.260
LMNB1	5031877	lamin-B1 isoform 1	4.2E-03	0.260
CFL2	14719392	cofilin-2 isoform 1	7.6E-04	0.259
PBDC1	32306539	protein PBDC1 isoform 1	1.1E-02	0.259
NUP107	530400597	nuclear pore complex protein Nup107 isoform X1	1.7E-03	0.258
MRPL14	974005264	39S ribosomal protein L14 mitochondrial isoform d	1.1E-03	-0.312
PAK2	984305	hPAK65 partial	3.0E-03	-0.313
EWSR1	119580187	Ewing sarcoma breakpoint region 1 isoform CRA_d	2.2E-04	-0.314
RAB11A	4758984	ras-related protein Rab-11A isoform 1	1.0E-03	-0.315
EIF3J	3264861	eukaryotic translation initiation factor eIF3 p35 subunit	8.8E-03	-0.316
TSTA3	4507709	GDP-L-fucose synthase	1.8E-03	-0.316
XPNPEP1	34783912	XPNPEP1 protein partial	8.3E-03	-0.318
GRSF1	530376963	G-rich sequence factor 1 isoform X2	5.0E-03	-0.319
HNRNPAB	55956921	heterogeneous nuclear ribonucleoprotein A/B isoform b	8.8E-03	-0.319
LMNA	119573385	lamin A/C isoform CRA_d	3.1E-04	-0.321
BLVRB	4502419	flavin reductase (NADPH)	4.1E-04	-0.322
NAA50	817478488	N-alpha-acetyltransferase 50 isoform 2	1.4E-02	-0.322
NDUFA5	47682777	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 5 13kDa	2.8E-03	-0.323
IGBP1	4557663	immunoglobulin-binding protein 1	4.5E-03	-0.326
FBP1	182311	fructose-1 6-bisphosphatase	1.3E-03	-0.327
RPL35P5	119628326	hCG1983332	1.5E-02	-0.331
NIPSNAP1	530420475	protein NipSnap homolog 1 isoform X1	2.8E-04	-0.332
HSPA4L	119625602	heat shock 70kDa protein 4-like isoform CRA_a	2.5E-03	-0.333
LSM7	7706423	U6 snRNA-associated Sm-like protein LSm7	2.5E-04	-0.334
DBNL	62198235	drebrin-like protein isoform b	1.2E-04	-0.335
RPL21	18104948	60S ribosomal protein L21	1.3E-03	-0.337
BCAR1	282398125	breast cancer anti-estrogen resistance protein 1 isoform 8	6.0E-06	-0.342
NQO1	554790420	NAD(P)H dehydrogenase [quinone] 1 isoform d	3.4E-03	-0.343
CTSD	4503143	cathepsin D preproprotein	6.7E-03	-0.344
CHCHD2	7705851	coiled-coil-helix-coiled-coil-helix domain-containing protein 2 precursor isoform 2	1.6E-04	-0.349
SDF4	189054914	unnamed protein product	4.1E-04	-0.350
CAMK2D	1010228622	calcium/calmodulin-dependent protein kinase type II subunit delta isoform 16	1.8E-03	-0.356
PSMF1	578835639	proteasome inhibitor PI31 subunit isoform X2	8.8E-04	-0.359
TPM1	530406412	tropomyosin alpha-1 chain isoform X10	1.6E-03	-0.361
ECHS1	194097323	enoyl-CoA hydratase mitochondrial	7.4E-04	-0.362
RPL6	21410970	Ribosomal protein L6	8.4E-04	-0.362
VBP1	125987848	RecName: Full=Prefoldin subunit 3; AltName: Full=HIBBJ46; AltName: Full=Von Hippel-Lindau-binding protein 1; Short=VBP-1; Short=VHL-binding protein 1	4.7E-03	-0.363

NAP1L4	5174613	nucleosome assembly protein 1-like 4	1.3E-02	-0.365
SSBP1	4507231	single-stranded DNA-binding protein mitochondrial precursor	4.5E-04	-0.366
RANBP9	39645650	RANBP9 protein	1.4E-04	-0.367
P4HB	20070125	protein disulfide-isomerase precursor	3.6E-06	-0.369
SLC9A3R2	71897229	Na(+)/H(+) exchange regulatory cofactor NHE-RF2 isoform b	5.4E-05	-0.370
MISP	27735067	mitotic interactor and substrate of PLK1	3.6E-03	-0.370
CAPG	21730367	capping actin protein, gelsolin like - Chain A Ca2+-Binding Mimicry In The Crystal Structure Of The Eu3+-Bound Mutant Human Macrophage Capping Protein Cap G	4.2E-04	-0.371
MRPL46	26667177	39S ribosomal protein L46 mitochondrial	4.9E-04	-0.371
HNRNPUL2	118601081	heterogeneous nuclear ribonucleoprotein U-like protein 2	2.2E-03	-0.371
DBN1	119605397	drebrin 1 isoform CRA_c partial	5.8E-04	-0.372
CHRAC1	8393116	chromatin accessibility complex protein 1	2.4E-06	-0.374
ETFA	957949341	electron-transfer-flavoprotein alpha polypeptide isoform 3 partial	2.6E-04	-0.374
MRPL49	4826649	39S ribosomal protein L49 mitochondrial	2.5E-03	-0.375
AK2	312836806	adenylate kinase 2 mitochondrial isoform c	7.1E-05	-0.376
VIL1	189053947	unnamed protein product	1.0E-02	-0.376
VDAC2	296317337	voltage-dependent anion-selective channel protein 2 isoform 1	2.8E-04	-0.377
RELA	15079553	RELA protein	4.8E-04	-0.382
PIN4	119592218	protein (peptidylprolyl cis/trans isomerase) NIMA-interacting 4 (parvulin)	2.8E-03	-0.383
EPS8L2	21264616	epidermal growth factor receptor kinase substrate 8-like protein 2	2.5E-05	-0.383
ZNF207	119600638	zinc finger protein 207 isoform CRA_b	2.9E-03	-0.384
TST	62898774	thiosulfate sulfurtransferase variant	2.8E-03	-0.385
AK1	4502011	adenylate kinase isoenzyme 1 isoform 1	2.5E-04	-0.386
LRRFIP1	212276104	leucine-rich repeat flightless-interacting protein 1 isoform 5	1.5E-02	-0.388
CORO1B	119595032	coronin actin binding protein 1B isoform CRA_c	5.8E-03	-0.397
PFDN2	12408675	prefoldin subunit 2	9.3E-05	-0.399
GSN	221045102	unnamed protein product	3.4E-03	-0.403
PFDN6	7657162	prefoldin subunit 6	1.6E-03	-0.406
FKBP9	33469985	peptidyl-prolyl cis-trans isomerase FKBP9 isoform 1 precursor	3.2E-03	-0.407
PDRG1	38197489	PDRG1 protein partial	5.7E-03	-0.407
PEBP1	4505621	phosphatidylethanolamine-binding protein 1	1.5E-02	-0.408
MTAP	386642469	methylthioadenosine phosphorylase	2.0E-04	-0.409
PLIN4	194379754	unnamed protein product	3.9E-03	-0.412
LDHA	62897717	lactate dehydrogenase A variant	3.6E-03	-0.414
RPS17	4506693	40S ribosomal protein S17	1.2E-02	-0.417
ETFB	119592410	electron-transfer-flavoprotein beta polypeptide	1.2E-03	-0.418

EFHD2	20149675	EF-hand domain-containing protein D2	1.6E-02	-0.420
AKR1B10	3150035	aldose reductase-like peptide	9.0E-03	-0.422
PMVK	5729980	phosphomevalonate kinase	3.0E-03	-0.422
TCEA3	767905824	transcription elongation factor A protein 3 isoform X5	3.7E-03	-0.423
RPL11	19851919	CLL-associated antigen KW-12	7.4E-03	-0.426
TALDO1	48257056	TALDO1 protein partial	1.2E-02	-0.438
RPL12	4506597	60S ribosomal protein L12	2.2E-04	-0.444
C9orf16	5002581	EST00098 protein	2.0E-05	-0.444
NHS	38093637	Nance-Horan syndrome protein isoform 1	1.3E-02	-0.445
RPS4X	48376549	RPS4X protein partial	1.7E-03	-0.446
LIMA1	62897503	epithelial protein lost in neoplasm beta variant	2.1E-04	-0.449
CGN	119573820	cingulin isoform CRA_b	9.5E-05	-0.449
VASP	957950351	vasodilator-stimulated phosphoprotein isoform 3 partial	4.1E-03	-0.450
KLK6	4506155	kallikrein-6 isoform A preproprotein	6.5E-03	-0.453
EPS8	119616761	epidermal growth factor receptor pathway substrate 8 isoform CRA_b	7.5E-03	-0.455
EML4 ALK	161176980	EML4/ALK fusion protein variant 3	6.4E-03	-0.455
TPPP3	56676375	tubulin polymerization-promoting protein family member 3	8.1E-03	-0.459
PODXL	16550641	unnamed protein product	2.3E-03	-0.463
RALB	194386002	unnamed protein product	4.0E-03	-0.464
TATDN1	14042943	putative deoxyribonuclease TATDN1 isoform a	4.0E-03	-0.467
CSRP1	193783759	unnamed protein product	1.4E-02	-0.467
ACTN4	2804273	alpha actinin 4	7.1E-04	-0.482
SNAP29	4759154	synaptosomal-associated protein 29	3.9E-04	-0.489
RPL13	15431295	60S ribosomal protein L13 isoform 1	1.2E-02	-0.492
RPS23	4506701	40S ribosomal protein S23	6.4E-04	-0.493
GSTO1	300360532	glutathione S-transferase omega-1 isoform 2	5.8E-03	-0.498
GRN	194382178	unnamed protein product	6.8E-04	-0.501
BOLA2	46577124	RecName: Full=BolA-like protein 2	1.7E-03	-0.502
MPST	432376	rhodanese	8.0E-03	-0.505
RPS15	4506687	40S ribosomal protein S15 isoform 2	1.1E-03	-0.505
PLEC	530389159	plectin isoform X13	4.8E-04	-0.508
ARHGDIA	4757768	rho GDP-dissociation inhibitor 1 isoform a	7.7E-03	-0.510
STAM	21040514	STAM protein	1.1E-03	-0.511
ANP32A	578827458	acidic leucine-rich nuclear phosphoprotein 32 family member A isoform X1	8.4E-03	-0.515
SERPINB1	194388748	unnamed protein product	1.5E-03	-0.515
ANXA4	1703319	RecName: Full=Annexin A4; AltName: Full=35-beta calcimedin; AltName: Full=Annexin IV; AltName: Full=Annexin-4; AltName: Full=Carbohydrate-binding protein p33/p41; AltName: Full=Chromobindin-4;	1.2E-05	-0.520
EXOSC4	9506689	exosome complex component RRP41	3.5E-03	-0.521

GSR	305410793	glutathione reductase mitochondrial isoform 4 precursor	6.3E-04	-0.522
COMT	6466450	catechol O-methyltransferase isoform S-COMT	4.5E-06	-0.528
TXNDC5	30354488	TXNDC5 protein	2.8E-05	-0.529
PYCR2	4960118	pyrroline 5-carboxylate reductase isoform	7.6E-06	-0.534
KRT19	386803	40-kDa keratin protein partial	8.6E-04	-0.543
CMPK1	12644008	RecName: Full=UMP-CMP kinase; AltName: Full=Deoxycytidylate kinase; Short=CK; Short=dCMP kinase; AltName: Full=Nucleoside-diphosphate kinase; AltName: Full=Uridine monophosphate/cytidine monophosphate kinase;	2.3E-03	-0.547
TMOD3	6934244	tropomodulin 3	5.2E-05	-0.548
TJP2	221044090	unnamed protein product	7.3E-04	-0.553
PCBD1	348903	4a-carbinolamine dehydratase partial	1.5E-02	-0.554
YBX1	340419	Y box binding protein-1	1.4E-03	-0.555
GMDS	3258631	GDP-D-mannose-4 6-dehydratase	1.9E-05	-0.559
RPL19	62897725	ribosomal protein L19 variant	2.4E-04	-0.562
PDAP1	7657441	28 kDa heat- and acid-stable phosphoprotein	6.4E-04	-0.563
IRF2BP2	32306875	interferon regulatory factor-2 binding protein 2B	9.0E-06	-0.566
RNASET2	5231228	ribonuclease T2 precursor	1.3E-03	-0.567
POLR3K	597709807	DNA-directed RNA polymerase III subunit RPC10	1.5E-02	-0.571
ARPC5	5031593	actin-related protein 2/3 complex subunit 5 isoform 1	1.3E-02	-0.572
CALM2	4502549	calmodulin isoform 2	9.8E-03	-0.576
SPTAN1	179106	nonerythroid alpha-spectrin	1.2E-05	-0.577
PDIA3	860986	protein disulfide isomerase	6.2E-05	-0.589
SFN	16306737	SFN protein	1.3E-03	-0.591
SNCG	119600721	synuclein gamma (breast cancer-specific protein 1) isoform CRA_a	4.7E-05	-0.594
MARCKS	187387	myristoylated alanine-rich C-kinase substrate	9.0E-04	-0.599
RPL13A	111494155	RPL13A protein	6.3E-03	-0.600
EHD1	194386692	unnamed protein product	5.3E-03	-0.601
POF1B	530422196	protein POF1B isoform X2	4.2E-04	-0.602
LYZ	157834713	Chain A Contribution Of Hydrophobic Effect To The Conformational Stability Of Human Lysozyme	1.0E-03	-0.604
LGALS3BP	119609949	lectin galactoside-binding soluble 3 binding protein isoform CRA_a	5.3E-06	-0.604
TAGLN2	4507357	transgelin-2 isoform b	3.5E-03	-0.620
EIF1AY	519666785	eukaryotic translation initiation factor 1A Y-chromosomal isoform 2	1.7E-03	-0.621
LGALS1	754496395	lectin galactoside-binding soluble 1 partial	4.4E-03	-0.634
KRT80	31873640	hypothetical protein	1.2E-03	-0.634
ITGA6	578804287	integrin alpha-6 isoform X2	6.2E-05	-0.635
MVP	194389978	unnamed protein product	7.4E-04	-0.638
FKBP10	119581172	FK506 binding protein 10 65 kDa isoform CRA_b	7.7E-05	-0.646

CST3	364505964	Chain A Crystal Structure Of L68v Mutant Of Human Cystatin C	8.0E-03	-0.646
TRMT2A	194384404	unnamed protein product	9.0E-04	-0.649
C11orf54	37782472	LP4947	7.3E-05	-0.654
TNKS1BP1	110556636	182 kDa tankyrase-1-binding protein	3.4E-05	-0.656
DNPH1	673541350	Chain A Crystal Structure Of Human Dnph1 (rcl) With 6-naphthyl-purine- Riboside-monophosphate	4.0E-03	-0.656
ANXA1	4502101	annexin A1	1.6E-05	-0.667
LASP1	5453710	LIM and SH3 domain protein 1 isoform a	5.8E-04	-0.676
PPA1	11056044	inorganic pyrophosphatase	8.9E-04	-0.679
TOLLIP	6048243	TOLLIP protein	3.3E-04	-0.682
TUBA4A	514052659	tubulin alpha-4A chain isoform 2	4.5E-04	-0.683
CFL1	5031635	cofilin-1	7.8E-04	-0.696
PDLIM5	221044868	Carboxyl Terminal LIM Domain Protein 5	7.3E-04	-0.698
TSTD1	163965377	thiosulfate sulfurtransferase/rhodanese-like domain-containing protein 1 isoform 1	2.6E-05	-0.714
IDH2	588282794	isocitrate dehydrogenase [NADP] mitochondrial isoform 3	7.5E-06	-0.716
CSPG4	34148711	melanoma chondroitin sulfate proteoglycan	8.8E-03	-0.723
SCP2	302344767	non-specific lipid-transfer protein isoform 8	1.1E-06	-0.751
MUC5AC	748983076	mucin-5AC precursor	5.1E-03	-0.771
NAPRT	33991172	NAPRT1 protein	6.7E-04	-0.777
RRBP1	23822112	RecName: Full=Ribosome-binding protein 1; AltName: Full=180 kDa ribosome receptor homolog; Short=RRp; AltName: Full=ES/130-related protein; AltName: Full=Ribosome receptor protein	1.3E-03	-0.795
GPD2	193787465	unnamed protein product	1.0E-03	-0.830
AGR2	5453541	anterior gradient protein 2 homolog precursor	2.4E-03	-0.830
TESC	767974680	calcineurin B homologous protein 3 isoform X1	1.8E-04	-0.834
SPEF2	767934681	sperm flagellar protein 2 isoform X4	2.6E-03	-0.835
ARVCF	119623413	armadillo repeat gene deletes in velocardiofacial syndrome isoform CRA_b	1.8E-04	-0.839
SERPINB5	453369	maspin	1.3E-04	-0.878
AKR1C3	1815604	3-alpha-hydroxysteroid dehydrogenase	3.1E-03	-0.882
C1orf116	12653943	Chromosome 1 open reading frame 116	5.4E-05	-0.882
ACAA2	189069404	unnamed protein product	5.2E-04	-0.903
UGDH	151567925	Chain A Structure Of Human Udp-glucose Dehydrogenase Complexed With Nadh And Udp-glucose	2.5E-04	-0.905
EPS8L1	33341720	PP10566	3.2E-04	-0.916
G6PD	452269	glucose-6-phosphate dehydrogenase	3.1E-04	-0.924
CKB	119602226	creatine kinase brain isoform CRA_a	4.0E-04	-0.941
LAD1	119611770	ladinin 1 isoform CRA_b	1.1E-03	-0.970
SH3BGRL3	13775198	SH3 domain-binding glutamic acid-rich-like protein 3	8.0E-05	-0.979
MIF	4505185	macrophage migration inhibitory factor	5.2E-03	-0.980
RANBP1	119623391	RAN binding protein 1 isoform CRA_a	1.3E-06	-0.981

CRIP2	62898063	cysteine-rich protein 2 variant	1.2E-06	-0.994
SERPINA1	6855601	PRO0684	1.1E-04	-1.010
EPPK1	119602582	epiplakin 1	2.7E-03	-1.011
STARD10	119595276	START domain containing 10 isoform CRA_c	1.2E-04	-1.054
CAPG	1002819390	macrophage-capping protein isoform 3	9.9E-05	-1.086
HSPA1A	194388088	unnamed protein product	2.9E-04	-1.253
BCAS1	530418484	breast carcinoma-amplified sequence 1 isoform X6	3.5E-04	-1.283
PFN2	16753215	profilin-2 isoform a	4.1E-07	-1.347
KRT20	27894337	keratin type I cytoskeletal 20	5.9E-03	-1.406
KRT18	30311	cytokeratin 18 (424 AA)	1.1E-03	-1.408
LYVE1	40549451	unnamed protein product	1.3E-04	-1.416
CRIP1	4503047	cysteine-rich protein 1	4.0E-04	-1.635
MSN	4505257	moesin	1.1E-04	-1.763
H2A	15826398	Chain A Crystal Structure Of The Nc2-Tbp-Dna Ternary Complex	2.2E-03	-1.785
CA2	245850	S-100P=Ca(2+)-binding protein	1.9E-04	-1.813