Identification of Oncostatin M Target Genes by RNA-Seq in Mouse Primary Myotube Cells

by

Aynur Erkin Kashgari

A Dissertation Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the degree of

Master of Science

in

Molecular Biology and Genetics

KOÇ
ÜNİVERSİTESİ

July 14, 2021

Identification of Oncostatin M Target Genes by RNA-Seq in Mouse Primary Myotube Cells

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Aynur Erkin Kashgari

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Asst. Prof. Serkan Kır (Advisor)

_____

Asst. Prof. İsmail Sağlam

_____

Assoc. Prof. Özlen Konu Karakayalı

Date:

ABSTRACT

Cancer cachexia is a debilitating wasting syndrome marked by inflammation and loss of adipose and skeletal muscle tissues. While about 20% of cancer deaths are caused by cachexia, there are no effective treatments available, and molecular mechanisms underlying cancer cachexia remain elusive. Oncostatin M, belonging to the gp130/IL6 cytokine family, is mainly released by macrophages, monocytes, and T-cells and exert pleiotropic activities in cell differentiation, proliferation, and inflammatory network. In various cell and cancer types, OSM was paradoxically reported to both promote and inhibit inflammation. Even though cytokine IL-6 is known to be associated with muscle atrophy, the functional significance of cytokine OSM in skeletal muscle cells is poorly understood. The advent of high-throughput sequencing technologies presented an unprecedented approach to study RNA. Therefore, it has become the prime choice to characterize and quantify cell transcripts. Here, we aim to investigate tumor signaling to muscle tissue by identifying cytokine Oncostatin M target genes in skeletal muscle cells using high-throughput RNA sequencing (RNA-seq) technology. Using "Tailored Pipeline," we have analyzed IL6 family cytokines, specifically OSM, IL6, and LIF. We treated mouse primary myotube cells with these cytokines and deciphered the differentially expressed genes that were significantly up-and down-regulated and show association with muscle atrophy.

RNA sequencing showed widespread alterations in cytokine treated samples compared to control samples, with the largest differences in OSM treated samples compared to other conditions. OSM significantly induced muscle atrophy-related genes such as *Ampd3, Sln, Murf-1, Atrogin-1*, and *Serpina3n*. Consistent with this, these signature genes were validated using qRT-PCR. GSEA analysis revealed that top commonly enriched pathways include hypoxia, inflammatory response pathway, glycolysis, and IL6-JAK-STAT3 pathways. These results provide a global view of OSM, IL6, and LIF expression responses in mouse primary myotube cells with cytokine-specific response patterns. Our results suggest these cytokines may use similar signaling mechanisms in mouse primary myotube cells. Compared to IL6 and LIF, OSM potently promotes cellular atrophy in cultured myotubes. Elucidating cytokine-induced mechanisms may establish potential therapeutic targets for the treatment of the cachectic syndrome. Our findings support an important role for OSM in cellular muscle atrophy and suggest that OSM could be targeted as a novel factor in fighting cachexia associated with cancer or other chronic diseases.

ÖZETÇE
Fare Birincil Miyotüp Hücrelerinde RNA-Seq ile Oncostatin M Hedef Genlerinin Tanımlanması

Aynur E. Kashgari

Moleküler Biyoloji ve Genetik, Yüksek Lisans

12 Temuz 2021

Kanser kaşeksisi , iltihap ve yağ ile iskelet kası dokularının kaybıyla belirtiliran bir zayıflatıcı yıkım sendromudur. Kanser ölümlerinin yaklaşık %20'si kaşeksiden kaynaklanırken, mevcut etkili tedaviler yoktur ve kanser kaşeksisinin altında yatan moleküler mekanizmalar belirsizliğini korumaktadır.Gp130/IL6 sitokin ailesine ait olan onkostatin M, temel olarak makrofajlar, monositler ve T hücreleri tarafından salınır ve hücre farklılaşması, proliferasyonu ve inflamatuar ağda pleiotropik aktiviteler uygular. Çeşitli hücre ve kanser türlerinde, OSM'nin paradoksal olarak inflamasyonu hem desteklediği hem de inhibe ettiği bildirildi. Sitokin IL-6'nın kas atrofisi ile ilişkili olduğu bilinmesine rağmen, sitokin OSM'nin iskelet kası hücrelerindeki fonksiyonel önemi yeterince bilinmemektedir.Yüksek verimli sıralama teknolojilerinin ortaya çıkışı, RNA yı incelemek için benzeri görülmemiş bir yaklaşım sundu.Bu nedenle, hücre transkriptlerini karakterize etmek ve ölçmek için birincil seçim haline geldi. Burada, yüksek verimli RNA dizileme (RNA-seq) teknolojisini kullanarak iskelet kası hücrelerinde sitokin Oncostatin M hedef genlerini tanımlayarak kas dokusuna tümör sinyalini araştırmayı amaçlıyoruz. " Özelleştirilmiş Boru Hattı" kullanarak IL6 ailesi sitokinlerini, özellikle OSM, IL6 ve LIF'yi analiz ettik.Fare birincil miyotüp hücrelerini bu sitokinlerle tedavi ettik ve önemli ölçüde yukarı ve aşağı regüle olan ve kas atrofisi ile ilişki gösteren farklı şekilde eksprese edilen genleri deşifre ettik.

RNA dizilimi, sitokin ile muamele edilmiş numunelerde kontrol numunelerine kıyasla yaygın değişiklikler gösterdi, diğer koşullara kıyasla OSM ile muamele edilmiş numunelerde en büyük farklılıklar. OSM, Ampd3, Sln, Murf-1, Atrogin-1 ve Serpina3n gibi kas atrofisi ile ilgili genleri önemli ölçüde indükledi. Bununla tutarlı olarak, bu imza genleri qRT-PCR kullanılarak onaylandı.GSEA analizi en yaygın olarak zenginleştirilmiş yolların hipoksi, inflamatuar yanıt yolu, glikoliz ve IL6-JAK-STAT3 yollarını içerdiğini ortaya koydu. Bu sonuçlar, sitokine özgü yanıt modelleriyle fare birincil miyotüp hücrelerinde OSM, IL6 ve LIF ekspresyon yanıtlarının küresel bir görünümünü sağlar. Sonuçlarımız, bu sitokinlerin fare birincil miyotüp hücrelerinde benzer sinyal mekanizmalarını kullanabileceğini gösteriyor. IL6 ve LIF ile karşılaştırıldığında, OSM, kültürlenmiş miyotüplerde hücresel atrofiyi güçlü bir şekilde teşvik eder. Sitokin kaynaklı mekanizmaların aydınlatılması kaşektik sendromun tedavisi için potansiyel terapötik hedefler oluşturabilir. Bulgularımız, hücresel kas atrofisinde OSM'nin önemli bir rolünü desteklemektedir ve OSM'nin kanser veya diğer kronik hastalıklarla ilişkili kaşeksi ile mücadelede yeni bir faktör olarak hedeflenebileceğini düşündürmektedir

ACKNOWLEDGEMENTS

First and foremost, I like to express my sincere gratitude to my supervisor, **Asst. Prof.Serkan Kır** for providing me the opportunity to pursue my passion for working with a project where bioinformatics merges with life science under his supervision.

I also would like to thank him for his invaluable advice, continuous support, and patience during my MSc. study.

I would also like to thank my thesis committee member, **Asst.Prof.İsmail Sağlam** and **Assoc. Prof.Ozlen Konu**, for providing invaluable feedback for my thesis project.

At the same time, this project was supported by **TÜBİTAK ARDEB** project no **118Z791**. I would like to thank the Scientific and Technological Research Council of Turkey and Koç University for their financial supports throughout my Master of Science education and research.

I would like to thank all the members of the **Cancer and Metabolism Laboratory**. The friendly and collaborative environment in the lab was excellent support. Their kind help and support have made my study and life in Turkey a wonderful time. My special thanks go to my project mate- **Aylin Domaniku**, for her continuous and timely help.

I want to thank my steadfast husband, **Ayper**, for his enduring love, and he has made countless sacrifices to help me reach this point. In addition, my son, **Omar,** has inspired me well with his love and caring, even as he bore the brunt of my moods and absences.

My parents and siblings deserve special thanks for their continued support and encouragement. Without such a team, I suspect that I would be in this place today.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

OSM	Oncostatin M
RNA-Seq	RNA sequencing
IL6	Interleukin 6
gp130	Glycoprotein 130
LIF	Leukemia inhibitory factor
GSEA	Gene set enrichment analysis
JAK-STAT	Janus Kinase/Signal Transducer and Activator of Transcription
qRT-PCR	Quantitative Real-Time Polymerase Chain Reaction
TNF-α	Tumor necrosis factor-alpha
IL-1	Interleukin-1
IFN-γ	Interferon-gamma
NF-κB	Nuclear factor-kappa
MuRF1	Muscle RING finger 1
Atrogin-1	Muscle atrophy F box
HTS	High-throughput sequencing
mRNA	Messenger ribonucleic acid
DGE	Differential gene expression
cDNA	complementary deoxyribonucleic acid
QC	quality control
DMEM	Dulbecco's Modified Eagle Medium
FBS	Fetal bovine serum
bFGF	basic fibroblast growth factor
PCA	Principal component analysis

Chapter 1:
INTRODUCTION

1.1 Cancer Cachexia

1.1.1 Epidemiology

Cancer Cachexia affects around 80% of cancer patients who are in their advanced stage. It is estimated that at least 20% of all deaths related to cancer are indirectly caused by cachexia. More than 80% of gastric or pancreatic cancer patients suffer from cachexia. In contrast, around 50% of patients diagnosed with lung or colon cancer are affected. About 40% of patients with breast cancer or acute leukemia develop the wasting syndrome (Argilés et al., 2014).

1.1.2 Definition

Cachexia is not a new clinical problem, and it has been known for centuries. The Greek Physician already wrote the earliest report which described cachexia Hippocrates who explained the complex relationship between cachexia and chronic heart failure, noting that: "the flesh is consumed and becomes water…the abdomen fills with water, the feet and legs swell, the shoulders, clavicles, the chest, and thighs melt away….. This illness is fatal" (A.M& PB, 1962). Etymologically, the word cachexia originated from the Greek word κακός kakos, and it is the combination of the words kakós (bad) and hexis (condition or appearance) which adequately reports the patients' overall composition. There is significant heterogeneity in its presence and severity, so its definition is somewhat controversial. However, recently a group of international experts came to a consensus defining cachexia. Based on the International Consensus Statement, cachexia is defined as a “multifactorial syndrome characterized by the ongoing loss of skeletal muscle mass that cannot be fully reversed by conventional support” (Fearon et al., 2011). Besides being correlated with cancer, cachexia is also associated with acute and chronic inflammatory diseases (e.g., chronic heart failure (A.M & PB,1962), autoimmune deficiency syndrome (AIDS) (Gottlieb et al.,1983), and chronic obstructive pulmonary disease (COPD) (Calder et al.,2018).

1.1.3 Cachexia as a multi-organ syndrome

Cachexia is an intricate and multifaceted metabolic wasting syndrome that is associated with underlying illness. The disorder is characterized by elevated resting energy expenditure (REE), excess systemic inflammation, increased insulin resistance, metabolic processes, anorexia, and anomia, depending on the cancer type (Evans et al., 2008). . Impacted cachectic patients often develop serious feebleness and faintness, followed by decreased physical activities and life expectancy (Dodson et al., 2011) (Figure 1.1). The progressing metabolic aberrations altogether lead to whole-body mass loss because of elevated catabolism of muscle protein and adipolysis in fat tissue. Skeletal muscle mass loss with or without body fat loss is crucial in cancer cachexia pathology. All aforementioned clinical features of cachexia contribute to the reduced tolerance of patients to several cancer therapies and the diminished life quality.

Due to its contribution to more than 40% of body weight, skeletal muscle is the principal organ involved in the wasting process developing through cancer cachexia. Furthermore, muscle wasting is the most prominent phenotypic characteristic of cancer cachexia. However, recent studies show that organs and tissues such as adipose tissue, liver, gut, heart, and brain directly participate in the progress of cachexia and may have been linked with muscle loss. For instance, aberrations in heart function of cachectic patients, modifications of liver protein synthesis, hypothalamus mediators, and induction of brown fat tissue wasting (Argilés et al., 2014). Hence, cachexia can be regarded as a multi-organ wasting syndrome. Figure 1.2 illustrates some of the systemic consequences of cachexia.

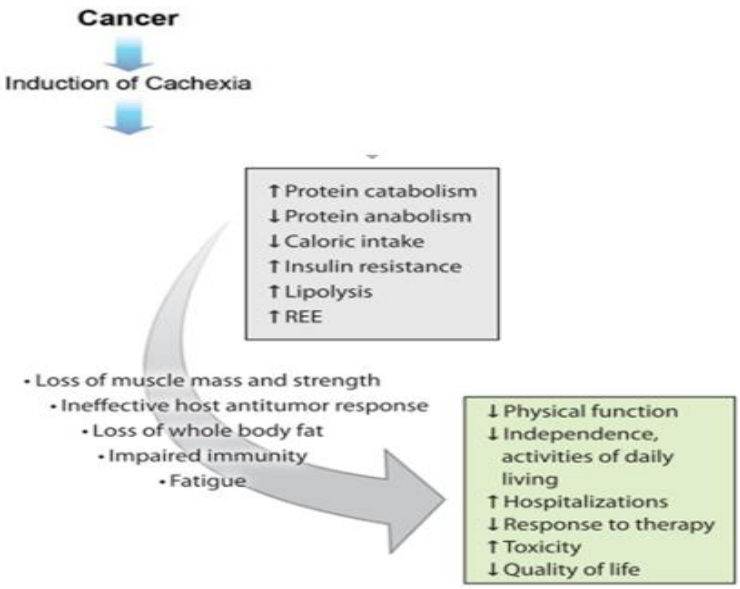

Figure 1.1: Cancer cachexia pathophysiology. The scheme was extracted from Dodson et al.,2011.

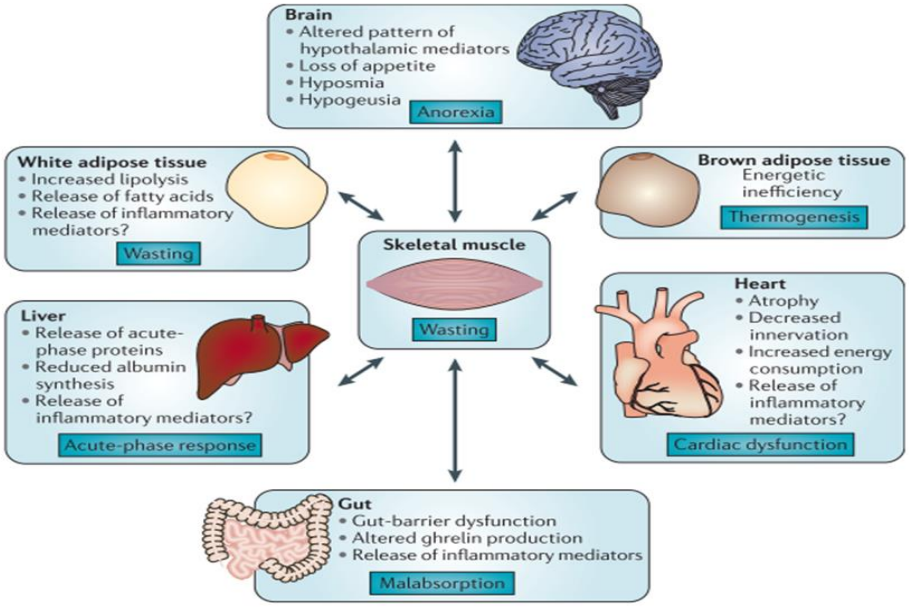

Figure 1.2: Cachexia consequences in various organs. Adapted from Argilés et al., 2014.

1.1.4 Systemic Mediators and Signaling

Nutritional interventions alone cannot reverse the cachectic wasting process, so lack of nutrients alone cannot explain the severe metabolic changes during cachexia (Tisdale, 1997). Even though patients are already in the refractory stage, the tumor mass only accounts for less than 1% of their body weight, so tumor energy demands cannot be the only cause of wasting energy (Fearon et al., 2012). Parabiotic transfer experiments using tumor-bearing rats demonstrated that circulating agents cause energy-wasting (Moley et al., 1985). Nonetheless, the etiology of cancer cachexia, whether it is triggered by the tumor or caused by the tumor's host response, is yet to be explored. Over the years, researchers have investigated the perplexing drivers of cancer cachexia, intending to develop innovative regimens that either impede or stop tumor-triggered fat and skeletal muscle loss. Generally, the origin of these mediators is presumed to be from immune cells, tumor cells, or the targeted mesenchymal tissues, which are undertaking a wasting process. Due to the dependencies of cancer cells on the secretion of inflammatory inducing mediators, chronic inflammation plays a significant role in carcinogenesis. These mediators defend cancer cells against apoptosis and growth. So, it is widely believed that inflammatory cytokines exert the primary driving force behind the cachectic process. A considerable number of studies using animals demonstrated a correlation between elevated cytokines and the extent of weight loss or decreased life expectancy.

The host response-derived factors tumor necrosis factor-alpha (TNF-a.), which was also known as “cachectin,” was among the remarkable proinflammatory cytokines that were identified in cachectic conditions (Beutler & Cerami, 1986). Subsequently, this cytokine was shown to promote cachexia in mice. Other prominent proinflammatory cytokines include interleukin-1 (IL-1), interferon-gamma (IFN-γ), and the tumor-derived protein-interleukin-6 (IL-6). Treating with a single cytokine or mixture of cytokines mentioned above sufficiently reduced food eating and increased other specific traits of the cachexia syndrome (Suzuki et al., 2013). Nuclear factor-kappa B (NF-κB), a nuclear transition activator, exerts a significant function in the upregulation of inflammatory gene expression, including TNF-a, IL-1, and IL-6. NF-κB is one of the predominant transcription factors in transducing TNF-a signals to the cells. Furthermore, activation of NF-κB also positively affects tumor growth

by accelerating inflammation, increasing tumor cellular proliferation, and blocking apoptosis (Suzuki et al., 2013).

1.2 Muscle Atrophy and Cancer Cachexia

Patients with cancer cachexia suffer most from intense impacts of skeletal muscle wasting despite losing body fat and muscle mass, and muscle wasting affect body function and performing daily physical activities. Muscle atrophy denotes dystrophy of striated muscle, in which muscle fiber becomes thin and unexpectedly disappears. At least in part, cancer cachexia is considered to be caused by interactions of the host with the tumor. Components of this interaction are then thought to drive molecular pathways resulting in muscle atrophy. Cancer cachexia-associated skeletal muscle atrophy is mediated by reducing protein production and concurrent activation of protein catabolism. These phenomena can emerge either directly (such as with proteolysis-inducing factor (PIF) stimulating the ubiquitin-proteasome pathway) or indirectly via other cascades (such as proinflammatory cytokines contributing to the induction of the acute phase response). Various hormones also have been implicated in muscle atrophy. So, various mechanisms and alterations contribute to muscle atrophy, including protein synthetic, myogenic, autophagic, proteasomal, inflammatory, and mitochondrial pathways (Suzuki et al., 2013). The following section will analyze current research in these areas.

1.2.1 Protein Synthetic Pathways

In cancer cachexia-associated muscle atrophy, compared to healthy controls, throughout the progression of cancer cachexia, the initiation of cachexia initially decreased the rate of protein synthesis by ~ 19%. It progressed toward ~50% with further weight loss (White et al., 2011). Some factors that affect protein synthesis are exercise, anabolic and catabolic hormones, energy balance, and amino acids (Kimball et al., 2002). The mTOR-related signal transduction pathways mainly regulate protein synthesis. The AKT/mTOR pathway is the utmost predominant regulator of translation initiation (Kimball et al., 2002). A primary activator of this pathway is the anabolic factor insulin-like growth factor-1 (IGF-1). It suppresses the ubiquitin-proteasome system (UPS). It promotes protein synthesis through the mammalian target of rapamycin (mTOR)-p70S6 kinase (p70 S6K) signaling pathway, and therefore, a lowered IGF-1 level was measured in the cancer cachexia (Suzuki

et al., 2013). Also, both the synthesis of IGF-1 and serum IGF-1 from skeletal muscle is diminished in cancer cachexia. (Attard-Montalto et al., 1998). AMPK inhibits mTOR by preventing the interaction with both p70S6K and 4E-BP1. Inflammatory cytokines, such as IL-6, TNF-a, also reduce mTOR activation directly and indirectly through phosphorylation of AMPK (Yang et al., 2018). Moreover, activation of Akt also represses transcription factors of FoxO family proteins through phosphorylation. This way, FoxO members cannot enter the nucleus and induce transcription of the muscle-specific ubiquitin ligases (MAFbx/Atrogin-1 and MuRF1) (Fearon et al., 2012).

1.2.2 . Protein Degradation

Previous reports show three pathways that affect the breakdown of intracellular proteins in skeletal muscle: the calcium-dependent pathway, the lysosomal pathway, and the ubiquitin-proteasome pathway (UPP). Lysosomal and calcium-dependent proteases have been found to involve 15%-20% of total protein breakdown in muscle. Diversely, UPP is believed to be fundamental and plays a significant function in the process of muscle atrophy in experimental and human models of cancer cachexia (Biolo et al., 1982). Moreover, it is believed that the proteolysis in skeletal muscle has primarily resulted from induction of the adenosine triphosphate-dependent (ATP) ubiquitin-proteosome system (UPS), and it is estimated that 20% of the bodyweight loss is owing to this pathway (Khal et al., 2005). UPS is triggered by prime factors of the cachexia, which include two muscle-specific E3 ubiquitin ligases, muscle RING finger 1 (MuRF1) and muscle atrophy F box (MAFbx/Atrogin-1). MuRF1 facilitates the ubiquitylation of the thick filament of the sarcomere and other thick filament elements (Glass, 2010). In addition, ATROGIN1, upregulated in many various models of cachexia, acts as a reliable indicator of acute muscle atrophy (Fearon et al., 2012). Polyubiquitination of myofibrillar proteins spotted for breakdown is stimulated by the proteins mentioned above. Forkhead-box O (FoxO) transcription factors control the expression of the two E3 ubiquitin ligases, and customarily they are phosphorylated and inactivated by the phosphatidylinositol three kinases (PI3K)-AKT pathway. Proteolysis-inducing factor (PIF) originated from the tumor is found in many cancer patients (Todorov et al., 1996), and TNF-a triggers this signaling pathway. Together with these agents, IFN-γ also promotes muscle atrophy by reducing protein production by inhibiting the eukaryotic

initiation factor 2 alpha (elF2a), part of the protein translation machinery. Also, Atrogin-1 induces ubiquitination of this factor. Aside from these, a different mechanism that stimulates muscle wasting is the blocking of muscle differentiation due to reduced myogenic differentiation (MyoD) expressions, and IFN-γ mediates this effect via the NF-κB and p38 MAPK pathways (Fearon et al., 2012). Transforming growth factor-beta (TGF-β) family member myostatin is also one of the widely explored cytokines. It is mainly synthesized in skeletal muscle cells and signals via the activin type II receptor by recruiting an ALK family kinase, activating a Smad2 and Smad3 transcription factor complex, reducing AKT-mediated protein synthesis. This initiated signaling cascade also increases Atrogin-1 and MuRF1, the ensuing breakdown of myofibrillar proteins. Through this signaling cascade, myostatin, which is detected in high levels in cachexia, promotes skeletal muscle atrophy. Under physiological conditions, myostatin is inhibited by IGF-1-mediated inhibition of pertinent transcription factors (Suzuki et al., 2013) (Figure 1.3).

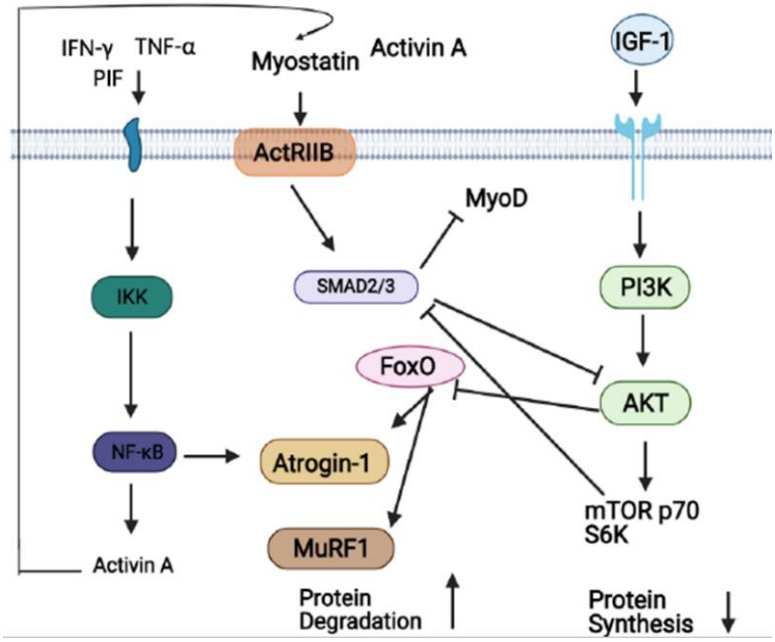

Figure 1.3: Protein synthesis and degradation pathways in muscle influenced by cancer cachexia, adapted from Fearon et al.,2012.

1.3 IL-6 Family of Cytokines and Signaling Pathway

Cytokines play essential roles in autocrine, paracrine, and endocrine signaling. The interleukin-6-type (IL-6) cytokines are defined as cytokines that typically signal through signaling receptor subunit glycoprotein 130 kDa (gp130). This group of cytokines consists of IL-6, IL-11, leukemia inhibitory factor (LIF), ciliary neurotrophic factor (CNTF), and Oncostatin M (OSM). The IL-6 family of cytokines regulates acute phase response to injury and infection. In addition, these cytokines exert critical activities in immune response, hematopoiesis, neuronal regeneration, and embryogenesis. IL-6 cytokines receptors are classified into non-signaling receptors (IL6, IL-11T, and CNTFR) and signal-transducing receptors (GP130, LIFR, OSMR). The signal-transducing receptors are associated with JAKs and are tyrosine phosphorylated by cytokine stimulation. Each member of this cytokine family has its unique combination of receptors; however, all of these receptors require the co-receptor gp130 for signaling transduction via the JAK-STAT pathway, primarily via JAK1, JAK2, and Tyk2, and downstream STAT1, STAT3, and STAT5 (Heinrich et al., 2003). The binding of IL6 and IL-11 to their perspective receptor subunits triggers GP130 homodimers to be recruited to transduce the signals of IL6 and IL-11. The other members signal through either heterodimer assembled from GP130 and LIFR (LIF, CNTF) or heterodimers of GP130 and OSMR (mouse OSM). For example, human OSM can transduce its signals via both GP130-LIFR and GP130-OSMR heterodimers. LIF and OSM recruit their receptor subunits directly, not needing other receptor subunits (Rose-John, 2018).

Depending on the cell and tissue type, IL6 cytokines receptors activate the JAK/STAT pathway and the p38 mitogen-activated protein kinase (MAPK) cascade. The activation of MAPK requires the SHP2 (SH2domain-containing tyrosine phosphate) binding site on gp130 and LIFR. It is also reported that nuclear factor-kappa B (NF-κB) affects these cytokines` myogenic activities (Heinrich et al., 2003). The SOCS (suppressor of cytokine signaling) family proteins exert a blockage function to activate these cytokine signaling pathways. The SOCS proteins are stimulated through the JAK/STAT pathway, ensuing STAT-mediated signal transduction inhibition (Heinrich et al., 2003). The overexpression of SOCS-3 mRNA was detected in the IL-6-administered muscles due to the elevated

STAT3 activity. In addition, SOCS-3 is reported to arbitrate the weakening of the signaling via the IL-6 receptor (Haddad et al., 2005).

This family of cytokines has pleiotropic effects on muscle tissues. LIF regulates myogenesis both in vitro and in vivo. LIF was demonstrated to stimulate proliferation of myoblasts in vitro, and it was also reported that LIF activates the JAK1/STAT1/STAT3 to regulate the proliferation and differentiation of C2C12 cells (Sun et al., 2007). Delivery of LIF to injured muscle promote muscle regeneration (Sun et al., 2007). Similarly, IL-6 was also shown to trigger myoblasts' proliferation in vitro, and IL-6 mRNA was detected in regenerating muscle (Kami & Senba, 1998). Circulating IL-6 concentrations were increased in experimental cachexia models, and neutralizing antibody inhibition of IL-6 signaling was demonstrated to exert a defensive activity on body weight loss. On the contrary, exogenous IL-6 was shown to have no role in the proteolysis of rat and murine myotubes.

What is more, blocking Il-6 signaling through a mouse IL-6R antibody exhaustively countered the muscular alterations displayed in IL-6 transgenic mice and diminished muscle atrophy in tumor-bearing wild-type mice (Muñoz-Cánoves et al., 2013); Therefore, it was advised that IL-6 might possess a nonrestrictive function during the progress of bodyweight reduction and skeletal muscle degradation accompanied by the presence of other circulating factors.

1.4 Oncostatin M

1.4.1 Biochemical and Genetic Profile of OSM

Oncostatin M (OSM), first identified in1986 by Zarling et al., belongs to the interleukin-6 family. Human OSM (hOSM) originally was discovered by its activity to block the proliferation of A375 melanoma cells and other human tumor cells (Zarling et al., 1986). hOSM is a secreted glycoprotein of 28 kDa and was initially extracted from phorbol 12-myristate 13-acetate (PMA) -stimulated human histiocytic lymphoma U937 cells. The hOSM cDNA encodes a protein of 252 amino acid residues (Malik et al., 1989). The hOSM gene is positioned in human chromosome 22q12 with three exons, and its expression is controlled by C/EBP and GC-rich elements and triggered by the STAT (Malik et al., 1989). Mouse OSM shows 48% similarity with human OSM. In addition, the mOSM cDNA was

found as an early gene triggered by IL-2, IL-3, and EPO. The mOSM gene is positioned in the vicinity of LIF in mouse chromosome11, and mOSM is found in hematopoietic tissues, including bone marrow, thymus, and spleen, while expressed in activated T cells, monocytes, and neutrophils (Malik et al., 1989).

1.4.2 The OSM Receptor

While OSM can weakly bind to gp130 directly, it seems this interaction is insufficient to transmit its signals for all OSM-mediated activities. In humans, OSM uses two kinds of functional receptors. The type 1 OSM receptor consists of gp130 and LIFR, and the type II OSM receptor is composed of gp130 and OSMR. The open reading frame of hOSMR has 979 amino acids, showing low similarity with hLIFR and gp130. Expression of hOSMR is detected in many cell types, including placental cells, muscle cells, skin cells, and multiple tumor cell lines. Biological activities vary among human and murine OSM. For example, a 30-fold higher amount of mOSM was needed to inhibit the growth of M1 cells in contrast to hOSM.

Moreover, hOSM is more robust than mOSM in blocking the differentiation of mouse ES cells. mOSM only transmits signals via its distinct receptor dimer consisted of gp130 and OSMR. Interestingly, hOSM cannot transduce signals through the mouse OSMR but transduces through LIFR (Lindberg et al., 1998).

1.4.3 OSM Signaling Pathway

OSM binds to receptor subunits (gp130 and OSMR or LIFR) and triggers receptor subunits to be dimerized, phosphorylating, and activating multiple JAKs. OSM exerts its activities through distinct signaling pathways such as the JAK/STAT pathway, the ERK1/ERK2, c-Jun N-terminal kinase (JNK), and p38, protein kinase C delta (PKCd), and phosphatidylinositol-3-kinase (PI3K/Akt) pathways. Depending on various cell types, respective signaling pathways may be activated under various OSM amounts. The intracellular part of OSMR encompasses the conserved proline-rich box1 region (CYPDIPDP), which recruits the JAK1 and JAK2 proteins. Once cytokines bind to respective receptors, they consequently activate inactive JAKs by phosphorylation. The phosphorylated receptors by JAKs sequentially serve as an anchoring spot for STAT

transcription factors. JAKs phosphorylate receptor-localized STATs and dissociate them from the receptor and translocate to the nucleus, stimulating cytokine-responsive gene expression (Masjedi et al., 2021). OSM is reported to trigger STAT1, STAT3, and STAT5 in multiple cell lines (Heinrich et al., 2003).

OSM activates phosphorylation of STAT1 and STAT3 in human astrocytes and astro-glioma cells (Schaefer et al., 2000). Intriguingly, OSM treatment leads to different gene expression profiles between low and high-passage astrocytes by modulating the fractions of STAT3 and STAT1. In low-passage astrocytes, OSM induced both STAT1 and STAT3 dimes; however, on the contrary, OSM only induced STAT3 dimers in high-passage cells (Schaefer et al., 2000).

1.4.4 Functions of OSM

OSM has pleiotropic properties and (in humans) signals through the Leukemia Inhibitory Factor receptor (LIFRβ) besides OSMRβ. Its effects have been shown to be context- and tissue-specific. Depending on target cells, it modulates tumor growth and nontumor cells either positively or negatively. OSM hinders the growth of various tumor cells, for example, solid tissue tumor cells, breast cancer cells, lung cancer cells, melanoma cells (Zarling et al., 1986), and glioma cells (Halfter et al., 1998).On the contrary, OSM also stimulates the growth of AIDS-related Kaposi's sarcoma cells (Block et al., 1995) and myeloma cells (Zhang et al., 1994). In addition, mOSM induces the growth of endothelial-like cells in primary culture of the aorta–gonad–mesonephros (AGM) originated cells and hematopoiesis (Mukouyama et al., 1998). Even though OSM regulates the growth of many cell types in vitro, OSMR defective mice grow normally without any abnormalities. However, bovine OSM (*bOSM*) overexpression using the keratin-14 promoter resulted in embryonic lethality (Malik et al., 1995). Comparable results are observed in mice overexpressing *bOSM* under the Control of metallothionein promoter, indicating that excess OSM is deleterious in mouse development (Malik et al., 1995).

Besides affecting growth, OSM, as a cytokine, is expressed and secreted by activated T cells and monocytes. Notably, OSM is shown to exhibit both pro-and anti-inflammatory activities due to various cellular microenvironments. These contrary effects of OSM could

result from activating concrete signaling pathways downstream of OSMRβ or LIFRβ. Human OSM triggers both OSMRβ and LIFRβ and cannot bind to murine OSMRβ but can transmit signals through murine LIFRβ. However, only specific murine OSMRβ enables OSM to transduce signals. Overall, OSM exerts anti-inflammatory functions in mouse models employing human OSM, indicating that these activities rely on LIFRβ complex stimulation. In a mouse model of RA, treatment with human OSM diminished the inflammatory activities.

Furthermore, it aided the reestablishment of tissue homeostasis by inhibiting the LPS-induced activation of TNF-α in a dose-dependent way (M. Christopher, 2016). The anti-inflammatory effect of OSM was also revealed in a diabetic wound healing model in mice. However, OSM's role on skeletal muscle is yet to be explored.

1.5 Gene Expression Analysis by RNA-Sequencing Technology

1.5.1 Methods to Analyze Genome-wide Gene Expression

It is worth mentioning that high-throughput sequencing (HTS) has become a paramount choice for detecting and measuring gene expression to obtain an insight into the transcriptional behavior of biological systems. Hence, the identification and analysis of differential genes are becoming a fundamental paradigm in various fields of biology and medicine. It can be applied to determine significantly differentially expressed genes between biological samples (Schena et al., 1995). Differential gene expression analysis plays a pivotal role in classifying heterogeneous diseases such as cancer (Bhattacharjee et al., 2002) and understanding the link between gene profile and survival or tumor aggressiveness (Veer et al., 2002). Furthermore, differential gene expression is essential to diagnose diseases (Heller et al., 1997) and discover new therapeutic drugs (Pagliarulo et al., 2002).

Gene expression analysis may include two parts: genome-wide and target-based, depending on what it is expected to study. The data is obtained at the biological system level in the absence of essential target genes of interest. Therefore genome-wide approaches such as microarrays (Augenlicht & Kobrin, 1982) or RNA-Seq (Mortazavi et al., 2008) appeared as a powerful tool for detecting differential gene expression, enabling researchers to measure the abundance of RNA species in a particular biological system. The entire set of transcripts

defines transcriptomic profiling of a cell, and their quantity at a definite time point is the foremost approach to measure differential gene expression. Thus, knowledge of the transcriptome has become very useful for connecting a gene's expression change and the cell's physical manifestation (phenotype). On the other hand, polymerase chain reaction (qPCR) can be implemented only if the target genes are already identified (Murphy et al., 1990).

Many technologies have been advanced over the years to perform transcriptomic analysis. However, beyond all these methods, high-throughput-sequencing (HTS) technology has presented a distinctive aspect of the transcriptional landscape of an organism and become the paradigm to measure RNA expression levels. With the advent of sequencing technology, it is now attainable to profile gene expression levels in each field of life sciences and become an indispensable clinical use technique. In general, this technology is applied to identify the differential gene expression, gene isoform, post-transcriptional modifications to explore phenotypic variation (Rapaport et al., 2013).

A microarray was an important tool to detect gene expression to understand the large number of gene expression levels in the 1990s. However, the power of this technology has several disadvantages. For instance, due to cross-hybridization, the expression measurements have a high background level. Hence, in order to identify the prior sequences, the probe sequences must be pre-specified. In addition, because of both environment and overload signals, the accuracy of expression measurements is constrained (Rapaport et al., 2013).

Contrary to microarray technology, some sequence-dependent methods are also essential to sequencing cDNA. For example, Sanger sequencing is performed to determine the cDNA sequence (Sanger et al., 1990). However, Sanger sequencing has low throughput and is not quantitative, and in order to overcome these limitations, tag-based methods were also advanced. Tag-based methods are comprised of serial analysis of gene expression (SAGE) (Velculescu et al., 1995), cap analysis of gene expression (CAGE) (Shiraki et al., 2003), and massively parallel signature sequencing (MPSS) (Shiraki et al., 2003). However,

tag-based sequencing technology also has its drawbacks. For example, short tags can be aligned to multiple locations on the reference genome, resulting in ambiguous transcripts.

Finally, RNA-Seq, a high-throughput sequencing technology based on next-generation sequencing (NGS) platform for differential gene expression, has emerged to overcome the constraints of both microarrays and tag-based methodologies. Remarkably, RNA-Seq technology is more reliable to arrays and employed to map and quantify transcriptome over all kinds of cells, alterations, and conditions (Brenner et al., 2000). This high-throughput sequencing technique is employed to quantify the transcriptome in a biological sample and analyze transcriptional changes between various experimental conditions to determine which genes or pathways may be affected by the specific treatments. Apart from differential gene expression (DGE), RNA-Seq can be employed to annotate and discover transcripts and identify polyadenylation and alternative splicing. Quantitative data from this novel technique is used for measuring differentially expressed genes, alternative transcription start sites (TSS), polyadenylation sites, and alternative splicing under different conditions. Qualitative data enables identifying transcripts, transcription start sites, 5` and 3` untranslated regions (UTR), and exon-intron boundaries (Carlson et al., 2012).

Transcriptomics is used to identify transcript function (small RNAs, non-coding RNAs, and messenger RNAs), to discern a gene's transcriptional structure (splicing mechanisms, start sites, and post-transcriptional modifications), and to quantify changes at each expressed transcript level in a given sample (Wang et al., 2009). For instance, studies have used RNA-Seq to identify alternative splicing patterns (Modified et al., 1985) to discover novel transcripts, non-coding RNAs, and mRNAs (Guttman et al., 2010).

1.5.2 RNA-Sequencing Experiment Workflow

Even though different technologies require different experimental procedures for sequencing study, principally, any high-throughput sequencing technology can be employed for RNA-Seq study. Illumina has emerged as one of the broadly applied NGS methods for both DNA and RNA-Seq. In the RNA-Seq study, after converting long RNAs to a library of cDNAs fragments with adapters attached to one or both ends, the library is sequenced in a high-throughput way to generate shortened reads (Wang et al., 2009). The generated tens of

millions of reads are mapped to a reference genome or transcriptome to get the genomic landscape of transcriptional or post-transcriptional gene expression.

Comparable to other high-throughput sequencing technologies, RNA-Seq also encounters critical issues such as data storage capacity and large amounts of data processing. However, RNA-Seq has advanced us to make an unprecedented large-scale over-view of the transcriptome despite the challenges. This thesis focuses on the RNA-Seq analysis of mRNAs and the custom pipeline, which is proper for NGS data generated from the Illumina platform.

A standard RNA-Seq study begins with isolating RNAs from tissues or cells of interest. Then a subset of RNAs is purified from the total RNA to analyze transcriptome expression. In particular, for protein-coding RNAs, this enrichment analysis is conducted by selecting poly (A) + molecules using oligo-deoxythymidine (dT) associated with cellulose beads or magnetic. In cells, the most abundant RNA-Ribosomal RNA (rRNA) constitutes about 90% of cellular RNAs (Wilhelm & Landry, 2009). Therefore, the rRNA is removed to allow increased sequencing coverage of the lower abundance mRNAs due to little interest in rRNAs in most studies. RNA fragmentation, the most commonly used RNA-Seq library preparation technique, takes place after rRNA depletion. First, RNA samples are subjected to a fragmentation process to get a certain size range. Then, after purification, the larger RNAs are fragmented using RNA hydrolysis.

Reverse transcriptase requiring primers to be hybridized into the RNA chain then converts the RNA fragments into DNA through reverse transcription. These primers can be oligo-dT or a sequence of random primers. Random primers are preferred to use as they have the potential hybridization capacity to random RNA molecule sites. Eventually, after synthesizing the first strand of cDNA, the RNA template gets eliminated and generated a second cDNA by using DNA pol I. Finally, a double-stranded cDNA molecule is generated.

During RNA-Seq library construction, the desired size of cDNAs has been generated through reverse transcription (RT) of fragmented RNAs with random hexamer primers. Blunt ends are generated from 3′ of the cDNA overhangs using a specific enzyme during the

adapter ligation process. Next, a shortened RNA ligase II is applied for a series of 3′ ligations. In contrast, RNA ligase I ligated the 5′adapter (Hrdlickova et al., 2017). Variance in 5′and 3`adapters sequence conserves strand specificity of RNA. In order to prepare cDNA fragments for ligation, the 3′sapped end containing a free T base over-hanged is appended with an A base. Eventually, distinctive adapters are appended to each strand of 3′ends of the double-ended cDNA. The gel exclusion method can be used for the selection of desired DNA length.

Additionally, adapters that are appended to each other or unligated to cDNAs are detached through this procedure. Finally, two primers recombine with the adapter`s end, and PCR amplifies collections of cDNA. The ds cDNAs are converted into ssDNA templates, which formed clonal clusters through bridge amplification before sequencing.

The double-stranded cDNA molecules that resulted from PCR amplification need to be denatured into single strands molecules so that single-stranded DNA templates are bridged-amplified to construct clonal clusters before sequencing. During this process, sequencing tags that enable multiplexing (barcodes or indices) can also be used. High throughput sequencing begins with the hybridizing sequencing primers to each single-stranded DNSs from the clonal clusters. The concurrent reverse complementation of DNA templates occurs applying fluorescent-labeled nucleotides. A laser excites clusters after adding a nucleotide, making the last incorporated base fluorescent. This cycle is typically repeated 30 to 200 times to remove the new nucleotide's fluorescence dye and blocking group. The cycle generated a sequence of images containing newly incorporated nucleotides where the fluorescence-labeled signal of each cluster is detected. As a result, the color of the fluoresced spot marks a distinct nucleotide base. The nucleotide sequences for each cluster can be obtained by combining the acquired sequence of images.

Lastly, a FASTQ format file, which has become de facto for NGS analysis, is generated to save this sequencing information. FASTQ file encloses a distinct ID for identifying the read, the raw sequence of nucleotides, and the quality scores per base in the read. The read quality at each base is defined by the Phred score ranging from 0 to 60 on a

logarithmic scale. The Phred quality score is denoted and calculated as Q=-10log(e). FASTQ files are always preprocessed to check the quality controls and remove any adapters from the sample preparation process. Moreover, it is worth noting that RNA sequencing can be single-end sequencing or double-end sequencing. As an example, a yeast ORF with one intron is illustrated in Figure 1.4.

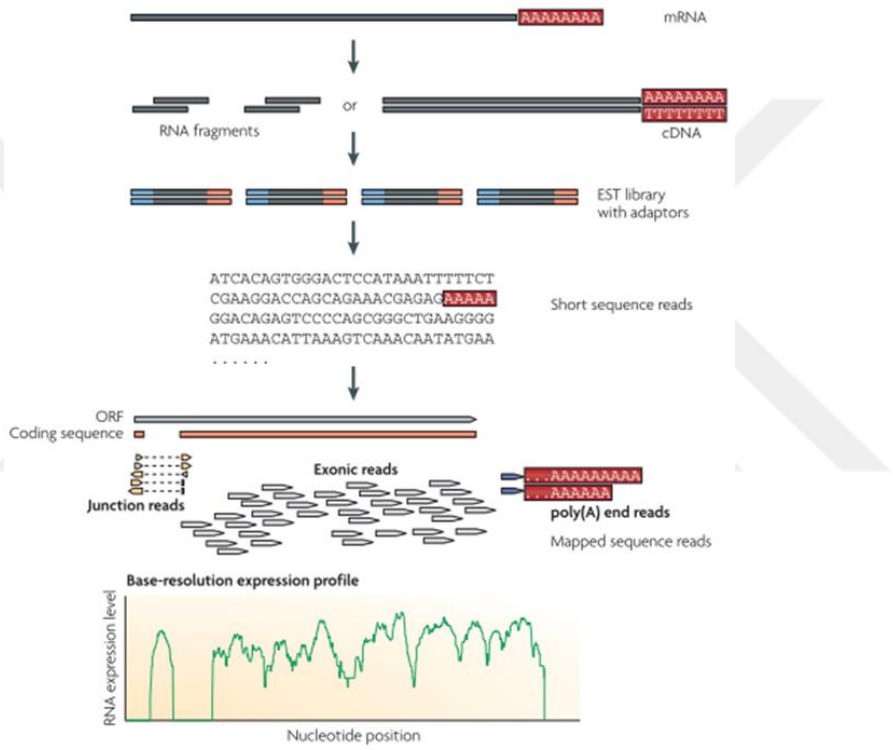

Figure 1.4: A standard RNA-seq protocol. Adapted from Wang et al.,2009.

1.5.3 Quality Control of Sequencing Data

Next-generation sequencing (NGS) technologies have substantially broadened the span of genomic research. As previously mentioned, high-throughput sequences generate an enormous quantity of data in one run. Before analyzing the obtained sequence and extracting biological results, assessing the library quality, and sequencing performance are imperative. Hence, before mapping or assembling any reads, the poor-quality reads must be filtered. NGS can also adversely impact raw data quality for downstream analysis because of various artifacts generated during the library preparation and sequencing process.

Until recently, several software tools have been developed to underline NGS data issues, such as poor base quality, impurity due to adapter sequences, and in base distribution (Trivedi et al., 2014) To assess the intrinsic quality of the raw reads generated by the sequencing platform is the primary step of the quality control (QC) process, and FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) is a widely used tool. Quality control is to consider the quality of alignments by mapping the reads to a genome. Several metrics, for instance, depth of coverage, contamination of rRNA, continuity of coverage, and GC bias, are used to assess the quality of alignments (Andrews, 2010). In this type of analysis, the low sequencing quality bases need to be removed to ensure high data quality. Finally, the cellular behaviors are identified, and biological interpretations are made.

1.5.4 Aligning Reads & Expression Quantification

The short sequence reads must be aligned with a reference genome or transcriptome after aberrant reads are removed. This alignment step aims to find the genomic positions of the acquired RNA sequences on a given reference genome. It can be very demanding to match the reads with the reference genome because sequencing generated enormous, small reads that needed to be aligned to a very large reference genome. What is more, the presence of non-coding regions makes aligning reads to the genome a much longer process than aligning to the transcriptome. Thus, read mapping is also the first computational time-demanding process in RNA-Seq data analysis. Currently, there are many alignment tools available, for instance, Bowtie (Langmead & Salzberg, 2012), BWA (Li & Durbin, 2009), and STAR (Dobin et al., 2013), and they are specialized for mapping short reads to a reference genome or transcriptome. In all situations, a reference genome or transcriptome barcode built at the beginning using hash tables or Burrows-Wheeler transform (BWT allows fast retrieval of the collection of the spots in the reference sequence where the reads are most probably to match (Li & Homer, 2010). After those locations are determined, alignment occurs in the candidate regions. The alignment's precision and speed depend upon the algorithms used in the alignment (Li & Homer, 2010).

BAM files are generated from mapping output, and it has all the data of overlapped and non-overlapped reads. Especially, overlapped encompasses the knowledge about the genomic location of aligned reads and its specific quality record (Li et al., 2009). After reads

have been aligned, the numbers of reads mapped to each unit of a specified genomic feature are counted, estimating the corresponding expression levels. It can be done using HTSeq (Anders et al., 2015) or featureCounts (Liao et al., 2014), among other options. Generally, the primary goal of many RNA-seq experiments is to identify genes that have an expression at different levels across the experimental samples. Plenty of algorithms have recently been developed, particularly to identify differentially expressed genes (DEGs) from RNA-seq data, including DESeq2 (Love et al., 2014), edgeR (Robinson & Oshlack, 2010), and Cuffdiff2 (Robinson & Oshlack, 2010). However, although numerous groups have performed various exhaustive comparative studies of different statistical approaches to identify DEGs, there is no standard unity even after many assessments.

Furthermore, it is critical to assess the overall quality of the data by investigating if the replicates are reproducible and whether non-biological factors affected the data. Generally, there should be high reproducibility among technical replicates (Mortazavi et al., 2008); however, there are no criteria for biological replicates. It varies according to how heterogenous the biological entity is. When there is a variance in gene expression between treatment conditions, there will be an expectation that samples with the same treatment will group in a Principal Component Analysis (PCA). Below Figure 1.5 is the outline of the mRNA-Seq data analysis pipeline.

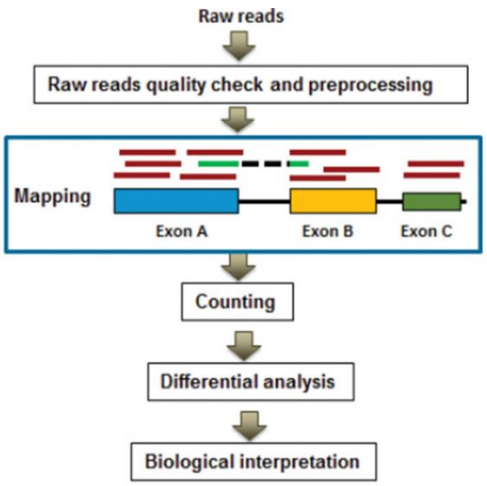

Figure 1.5: Flow chart of the mRNA-Seq data analysis pipeline. Adapted from Zhao et al., 2016.

1.5.5 Pathway Analysis

Biologists face the challenge of interpreting large-scale data, especially with the advent of high-throughput technology. The limitation usually lies in understanding the meaning of the array of genes to reveal the underlying molecular mechanism of the phenotype. Finally, a set of DEGs can be classified into customary pathways to identify significantly enriched pathways. The main goal of this pathway enrichment analysis is to identify ultimate possibilities of the hidden connections between genetic and molecular codes and physical traits of a biological entity in the research (Emmert-Streib & Glazko, 2011). For individual genes, the variation in gene expression relies on a particular mild or even imperceptible condition. Hence the pathway enrichment analysis of a group of gene variations could signify variances among different phenotypes (Emmert-Streib & Glazko, 2011), assess differentially expressed genes' function, and perform different cellular pathways. Functional enrichment analyses can be performed based on annotation databases, for instance, Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. The array of single-gene identification is transformed into pathways which are then concatenated into these databases of gene expression arrangement.

1.6 Purpose

The purpose of this thesis will be to study the molecular mechanism of cancer cachexia utilizing transcriptomic methodologies to identify differentially expressed genes and significantly modulated pathways. By doing so, I will fill major knowledge gaps by examining cancer cachexia molecular mechanisms and utilizing RNA-sequencing methods to identify novel altered genes (transcripts) and affected pathways in response to cytokine treatments in mouse primary myotubes.

1.6 Objectives of Data Analysis

In order to further elucidate the mechanisms behind cancer cachexia, this investigation will determine the number of significantly differentially expressed genes upon treating mouse primary myotubes with various cytokines and determine the top canonical

pathways impacted by identified differentially expressed genes in comparison of control and treatment groups.

Chapter 2:
METHODS

2.1 Cytokine Treatment

Primary myoblasts isolated from male mice were used in this experiment. Firstly, DMEM containing 20% FBS (Sigma) was used to expand the primary myoblasts for 48 hours with supplementation of basic fibroblast growth factor (bFGF) (Promega) and 1% penicillin-streptomycin (Gibco). Then, myoblasts were split into 12 well plates, and in order to extract RNA, cells become differentiated into mature myotubes in DMEM with 5% horse serum and 1% penicillin-streptomycin (Gibco) (~24 hours) and then recombinant OSM, IL6, LIF (250ng/µl) are administered to the mature myotubes. The differentiation medium was changed daily.

2.2 RNA Extraction and Isolation

Total RNA was isolated from cultured cells using the Direct-zolTM RNA MiniPrep kit Zymo Research (Irvine, CA 92614) based on the manufacturer`s protocol. The concentration and purity of RNA samples were assessed using a Nanodrop analyzer.

2.3 Library Preparation and Sequencing

The RNA samples were submitted to Macrogen (Europe) for library construction and sequencing. The template size distribution was checked by running on an Agilent Technologies 2100 Bioanalyzer using a DNA 1000 chip to verify the size of PCR enriched fragments. Samples were formulated using the Illumina® TruSeq Stranded mRNA RNA Library PrepKit following the manufacturer's instructions. A different adapter was used to tag each sequencing library, and an Agilent 2100 Bioanalyzer was used to evaluate library quality. The indexed library samples were normalized and pooled in equimolar concentrations before sequencing. Pair-end sequencing of RNA-Seq libraries was generated through the Illumina HiSeq technology with 30 million 2x101bp paired-end reads per sample. The Illumina sequencer generated raw images utilizing sequencing control software for system control and base calling through an integrated primary analysis software called

RTA (Real-Time Analysis). The BCL (base calls) binary was converted into FASTQ applying the Illumina package bcl2fastq. Adapters are not removed from the reads. The raw sequencing reads were stored as fastq files, including the read summary of each sample.

2.4 Processing of Raw Data

Figure 2.4 illustrates a flowchart of a tailored bioinformatics pipeline incorporating open-source tools and codes written by me to identify DEGs. Raw data is collected, then the reads are handled using FastQC (Andrews, S. (2010)). FastQC is a Java software that provides tools to conduct a QC evaluation of the sequencing dataset. With the FastQC software (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), the read quality was assessed to determine sequencing errors, contamination, and sequencing or PCR artifacts. FastQC reports include per base sequence quality, per sequence quality scores, per base sequence content, per base GC content, base N content, sequence length distribution, duplication level, and Kmer content. Problems detected by a QC analysis could originate from the sequencer or the starting library material. Reads with low-quality scores and lengths less than ten bp were removed. Also, 3' adapter sequence and repetitive elements were trimmed using Trimmomatic (v0.39). After filtering, clean reads with a length greater than 10bp were aligned to the mouse reference genome (GRCM28.P6 assembly).

2.5 Bioinformatics

The filtered reads mapped to the mouse reference genome (GRCM28.P6 assembly) with the Ensemble gene annotation (Mus_musculus.GRCm38.100.chr.gtf) using STAR-Aligner (v2.7.0a). Following alignment, MultiQC was used to summarize mapped reads, and Qualimap software (v2.2.1) was used to assess the coverage of the alignment data. The Python package HTSeq (0.13.5, https://htseq.readthedocs.io/en/master/overview.html), including an integrated script, was used to agglomerate raw counts of aligned reads and quantified unique fragments which are mapped in every gene using the intersection-nonempty mode, which decides how to process reads that are overlapped across the genes. Only uniquely aligned fragments are considered to form a gene quantification table.

Only raw read counts are not enough to study expressional levels across the experimental groups. These metrics can be influenced by transcript length, the total number

of reads, and sequencing biases. For instance, there is a greater likelihood of longer genes being detected as more reads, while differential expression analysis yields more DE expressed genes as sequencing depth increases. There are different normalization methods to overcome many affected factors, including FPKM (Fragments Per Kilobase Million), TPM (Transcripts per Kilobase Million), and RPKM (Reads Per Kilobase of transcripts per Million mapped reads) (Pachter, 2011). DEG analysis necessitates the gene expression metrics that need to be investigated comparatively across the samples. So, following quantification of gene expression, a differential expression (DE) test is conducted across the RNA-Seq samples by applying statistical methods and comparing the normalized expression levels.

DESeq2, a R/Bioconductor package (Love et al., 2014), was utilized in further analysis. This algorithm applies a negative binomial distribution to model the gene expression distribution. It requires raw data as input due to its internal normalization. It outputs a table with mean expression level as a combined estimate from both conditions and estimates separately for each condition, the fold change from the first to the second condition, the logarithm of the fold change, the p-value for statistical significance of this change, and the adjusted p-value (q-value) with Benjamin-Hochberg procedure (Benjamini & Hochberg, 1995) that determine the percentage of false positives among all the rejected hypotheses (FDR).

It is difficult to distinguish between no expression and low expression of a gene between samples, according to RNA-Seq data. However, a cutoff of the total row sum of read counts of 10 has been used in this work in order to separate expressed genes from the not expressed genes. Thus, only the sum of total genes across the samples of at least ten was considered when determining significant differential expression. Tests for statistical significance (P-values) accounted for multiple testing by computing the False Discovery Rate p-value (FDR p-value) using the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995). The FDR adjusts p-values for the number of hypothesis tests performed to control type1 error rate1. Differentially Expressed (DE) genes were defined as surpassing threshold levels of FDR=0.05 and Log2FC = 1. DESeq2 was chosen over other analysis

packages due to its accuracy, log2 normalization technique, acceptance in the literature, and ease of use. Principal component analysis was conducted to check the clustering among biological replicates.

Interpreting sequencing data can be difficult due to its high dimensionality. There are many genes and many samples. PCA is a data scaling method to reduce and compact high dimensional data into fewer dimensions to interpret data. PCA analysis was done in R with the DESeq2 package. PCA is a quick way to visualize variation between expression analysis samples. Hierarchical clustering was done as another exploratory analysis to expand and contrast the PCA. According to normalized gene expression values, hierarchical clustering of the samples was done to investigate if there are subgroups of samples with similarities in their expression profile. It is a complementary method to identify robust arrangements of the data and possible deviations. The heatmap shows the correlation of gene expression for all pairwise aggregations of samples in the data. The color shades represent substructure in the data, and clustering of replicates is expected for each condition. In addition, we expect to observe samples grouped similarly to the clusters seen in a PCA plot. For heatmap, clustering was done across samples after trimming low expressed or unexpressed genes by setting the sum of gene counts across samples to be a minimum of ten.

GSEA (Gene-Set Enrichment Analysis) (Subramanian et al., 2005) is one of the principal methodologies to identify previously known functional gene sets (GO terms) from the list of differential expression genes that are over-represented either in top or bottom of the ranked list based upon enrichment scores. This approach attains its strength by attending to gene sets- genes that share common biological functions, chromosomal location, or regulation. Enrichment scores (ES) were given by walking along with the ranked list of genes, incrementing a running sum when a gene is in the gene-set, and decrementing when it is not present. Thus, a positive ES score implies gene-set enrichment at the topmost of the ranked list, and a negative ES score implies gene-set enrichment at the lowermost of the ranked list. In this study, GSEA was performed to identify potential effects of cytokine OSM signaling through known functions of genes using the pre-ranked mode using FGSEA (Subramanian et al., 2005), where the set of genes was sorted based on log2 fold change

values. Rather than p/q-values, fold-change values were considered to detect the enrichment peak.

Figure 2.1 shows the tailored pipeline used in this study. Boxes with a red outline shape indicate the task and means to do it, and boxes without an outline indicate the input/output files.

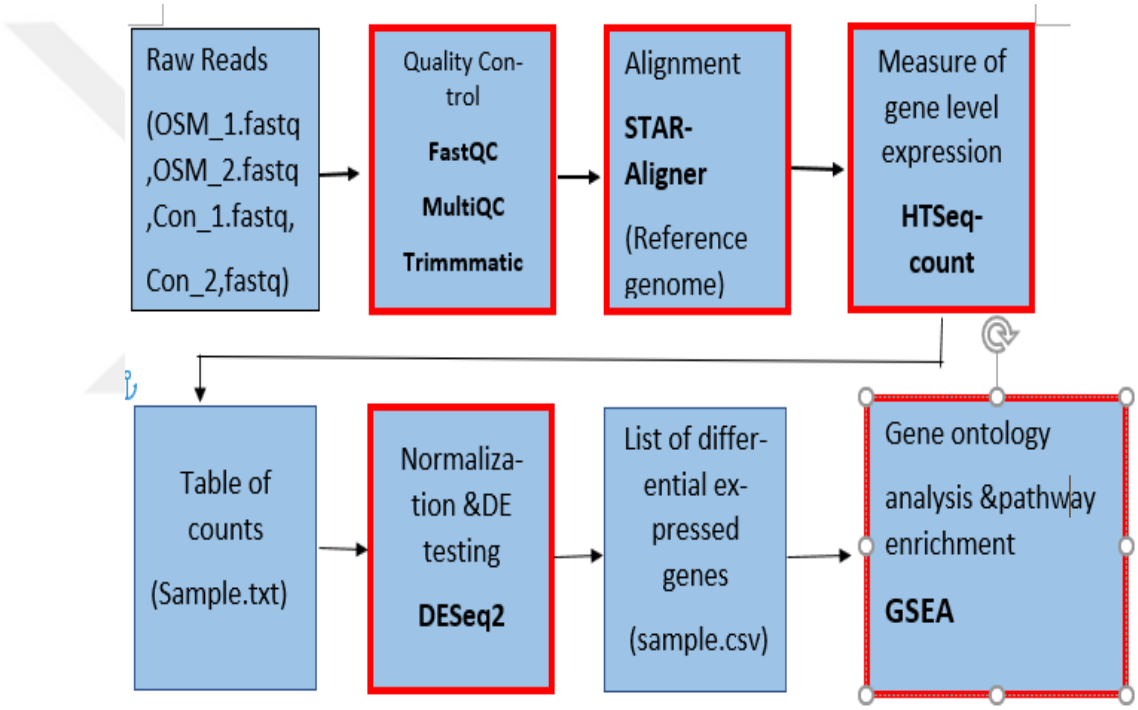

Figure 2.1: Tailored RNA-Seq analysis pipeline flow chart.

Chapter 3:
RESULTS

3.1 Quality Check of Raw Reads

To study the effects of cytokines (OSM, LIF, IL6) on muscle cells, primary myoblasts isolated from male mice were cultured and differentiated into myotubes, then treated with recombinant proteins (OSM, IL6, LIF), then collected RNA for sequencing. There are two biological replicates for each condition and eight samples in total. Sequencing data is first checked for quality before performing differential analysis. Using the FastQC tool, the quality of all eight samples read files containing pair-ended 101bp reads were evaluated. As a result, all the fastq files had a similar overall good quality. As a representative example, Figure 3.1 illustrates the FastQC result from the OSM treated sample. In FastQC, under the "Mean Quality Scores" module, a summary diagram indicates the quality of all sequences in the file. The X-axis in this diagram shows the position for each base in each sequence in the file, and the y-axis shows the Phred quality scores (Q value), which range from 0 to 40. A higher Phred quality score indicates a higher base calling accuracy.

Furthermore, summary per sequence and per-base quality plots were generated with the MultiQC program. For example, according to Figure 3.1a, it can be seen that the sequencing has very good quality since all quality values are above 36. The green zone illustrates a high accuracy for the base calling at each position in this area where the quality score is higher than 36. Furthermore, the median quality is high for all sequences, which means base calling has high confidence and high quality.

The module "Adapter Contents" is a summary diagram indicating adapter content per base position for all sequences in the file. The X-axis in this diagram shows the position for each base in each sequence in the file, and the y-axis in this diagram shows % of sequences in the file. For example, according to Figure 3.1b, it can be seen that there are adapters around the 80th base pair position which need to be trimmed before downstream analysis.

The module "Per sequence GC content" shows the proportion of GC content across the whole length of all reads. A normal distribution of the GC content is expected and means that GC content in the file should not deviate from the normal distribution. According to Figure 3.1c, the GC content calculated distribution is almost identical to normal modeled distributions, illustrating a sequencing run. a. "Mean quality scores" is a summary diagram indicating the quality for all sequences in the file. It can be seen that all quality values are over 36, and sequencing has a very good quality and shows high accuracy. b. the plot "adapter content" displays the summative percentage count of the proportion of the library which has seen each of the adapter sequences at each location. c, the plot of "Per sequence GC content" gives the calculation GC content across the whole length of each sequence in the file. The line indicates a normal GC content assignment nearly identical to the normal calculated distribution and illustrates a good quality sequencing run.

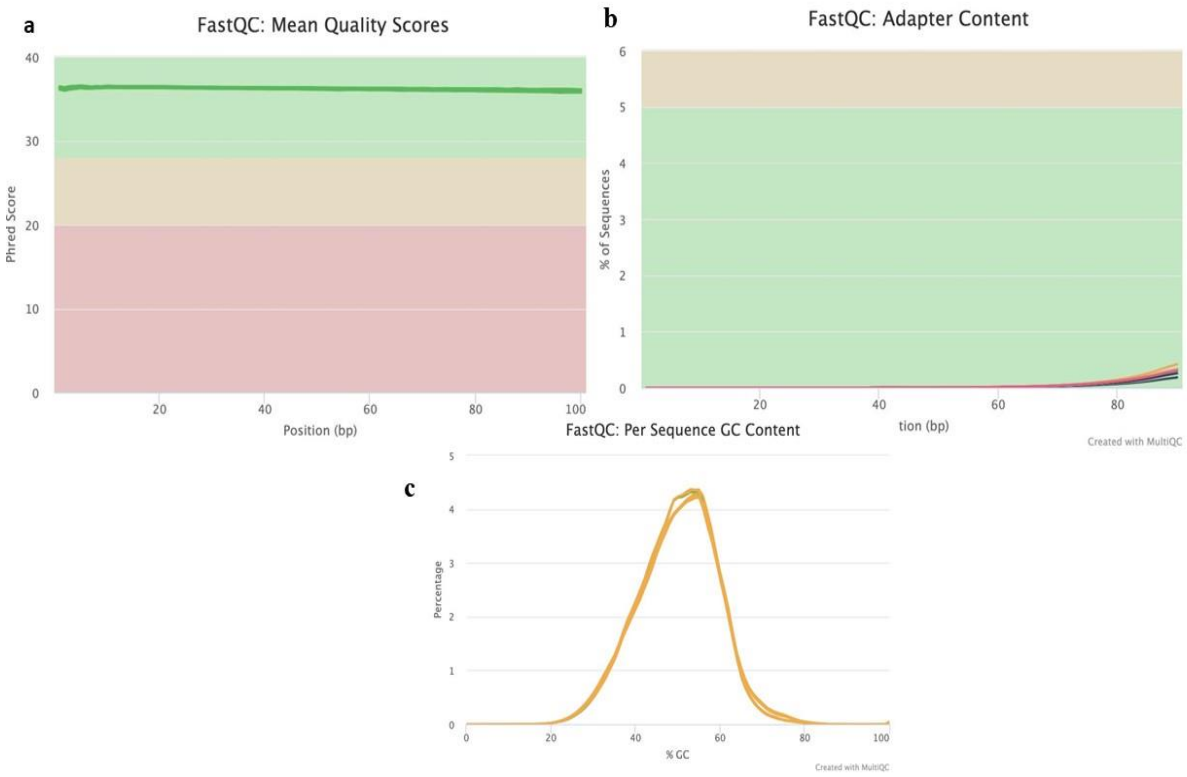

Figure 3.1: Plots generated FastQC & MultiQC tool from OSM treated mouse primary myotube cells.

3.2 Trimming & Quality control of files

As shown in the Figure 3.1b plot, there are adapters near the 80th base position. Next, each FASTQ file was trimmed using Trimmomatic to filter short, fragmented adapters and low-quality reads. This trimming step improves the read quality by deleting/trimming sequences below a certain threshold, leading to a higher quality. After Trimmomatic, trimmed data were checked again using FastQC, and summary per sequence and per-base quality plots were generated employing the MultiQC program and shown below. These plots demonstrate improved data quality after the preprocessing step and indicate that Truseq adaptors had already been trimmed and low-quality reads were removed.

According to Figure 3.2a, it can be seen that no portion of the reads contains N content indicating a good base call was made as the pipelines were able to decipher the data sufficiently. Figure 3.2b "**Mean Quality Scores**" plot shows is a summary diagram indicating the quality of all sequences in the file. The X-axis in this diagram shows the position for each base in each sequence in the file, and the y-axis shows the Phred quality scores (Q value), which range from 0 to 40. According to Figure 3.2b, it can be seen that the sequencing has very good quality since all quality values are above 36. The green zone illustrates a high accuracy for the base calling at each position in this area where the quality score is higher than 36. The median quality is high for all sequences, which means base calling has performed with high confidence and high quality. Consistent with the previous plot, after trimming, the sequencing file still has a very high quality.

Figure 3.2c **"Per sequence quality scores"** graph shows the mean quality score on the x-axis and the number of sequences with that means on the y-axis. According to Figure 3.2c, it can be seen that after trimming, the sequence still has high average quality scores. Thus, overall, the loss of information inherent in trimming was not significant.

a "**Per Base N Content"** shows the percentage of base calls at each position where N call was made. b, "**Mean Quality Scores"** depicts a ketch of the range of quality values across all bases at every location. c, the plot of "**Per sequence Quality Scores.**" Illustrates if a subgroup of the sequences has a low-quality score universally.

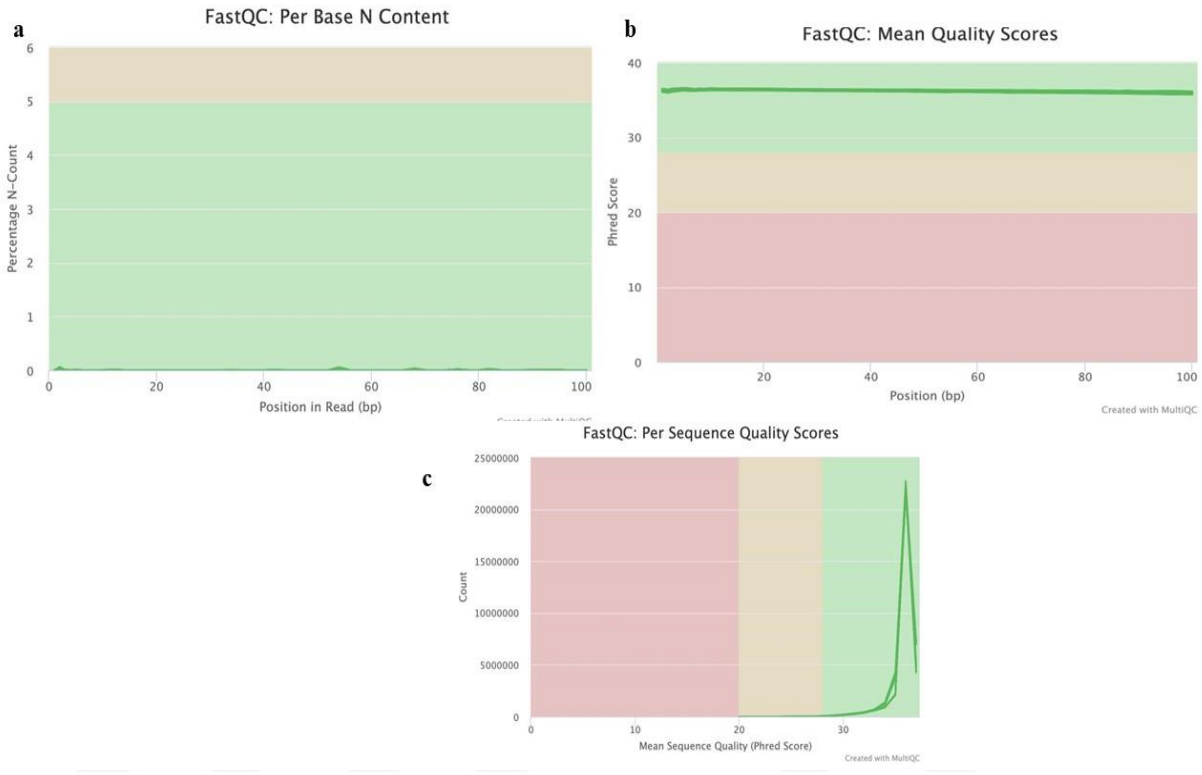

Figure 3.2: Plots generated by FastQC & MultiQC tool from OSM treated mouse primary myotube cells.

3.3 Mapping reads-STAR-Aligner, QualiMap & MultiQC

The filtered reads were then aligned to a mouse reference genome. The reference file was obtained from the Ensembles (Flicek et al., 2013) website, and especially, the used file accords to the build 38 patch release number 6. The reference file was indexed using STAR-Aligner and used STAR default setting to map the RNA-Seq reads to the reference mouse genome. Next, QualiMap calculates the different quality scores, for instance, DNA or rRNA contamination, 5'-3' biases, and coverage biases. Based on the mapping, QualiMap creates a plain text file and a quality report (pdf-format) for each alignment file of the dataset. These files contain alignment-specific parameters, such as the counts of aligned reads and the origin of the reads (i.e., exonic, intronic, intergenic, and overlapping exons).

Figure 3.3 illustrates the STAR-Aligner score generated using MultiQC. It shows that more than 30M reads across all samples are uniquely mapped to the reference genome, and very few unmapped reads, indicating a good mapping quality throughout the genome.

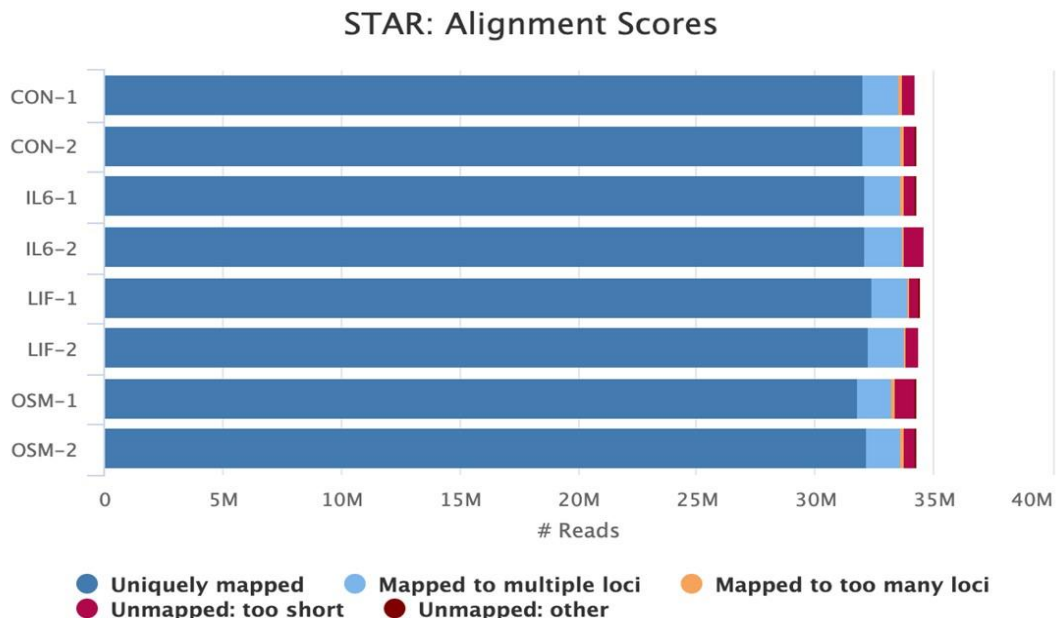

Figure 3.3: STAR Alignment Score generated by MultiQC.

Table 3.1 QualiMap indicates that approximately 93% of all reads across the samples are uniquely mapped. The percentage of mapped reads is a general indicator of complete sequencing certainly. As expected, a high percentage of reads were aligned to the genome uniquely, and equally, a small fraction of reads are mapped to multiple regions in the reference genome. The high number of aligned reads indicates that raw data generation and preprocessing were successful. On the other hand, a low percentage of aligned reads could indicate a problem with the RNA extraction because the raw data could be, for instance, contaminated with bacterial DNA.

Table 3.2 displays the number of alignments mapped into exonic, intronic, and intergenic regions. The exonic region contains 5'UTR, protein-coding region, and 3'UTR region. Among the RNA-seq reads, approximately 88% of reads were mapped within annotated exonic regions of the genome, and 10% are aligned into intergenic regions for

each condition. Less than 2% mapped within intergenic regions. Thus, in contrast with poly-A enrichment, a greater intronic mapping is expected for rRNA depletion.

Overall, this measurement provides information about sample quality and helps to identify any biases in the data. However, the summaries extracted from these QC results may imply that it is necessary to rectify these biases by going back and modifying the parameters for mapping accordingly.

Table 3.1: QualiMap output results of mapping the RNA-Seq reads to the mouse reference genome.

Sample	# of clean reads	# of nucleotides	mappped percentage	multiple alignment
Control1	32058182	6475752764	93.64%	4.46%
Control2	32056147	6475341694	93.50%	4.65%
Il6-1	32116412	6487515224	93.66%	4.45%
IL6-2	32095739	6483339278	92.72%	4.63%
LIF-1	32385411	6541853022	94.05%	4.41%
LIF-2	32262030	6516930060	93.79%	4.37%
OSM-1	31793241	6422234682	92.74%	4.26%
OSM-2	32167106	6497755412	93.85%	4.30%

Table 3.2:Percentage of read alignment to the mouse genomic regions reported by QualiMap.

Sample	Exonic (%)	Intronic(%)	Intergenic(%)
Control1	88.07%	10.17%	1.77%
Control2	88.13%	10.14%	1.72%
Il6-1	87.67%	10.32%	2.01%
IL6-2	88.60%	9.70%	1.70%
LIF-1	88.76%	9.45%	1.79%
LIF-2	88.37%	9.81%	1.82%
OSM-1	88.66%	9.37%	1.97%
OSM-2	88.32%	9.78%	1.90%

3.4 Expression quantification-HTSeq-count

It takes a file with aligned sequencing reads (BAM file), a list of genomic features (genes), and counts how many reads uniquely map to each feature (gene). Then, it generates a count matrix, with genes as rows and samples are columns. Figure 3.4: HTSeq gene count assignments generated by MultiQC.is the count assignment shows more than 30M reads across the samples uniquely assigned to genes.

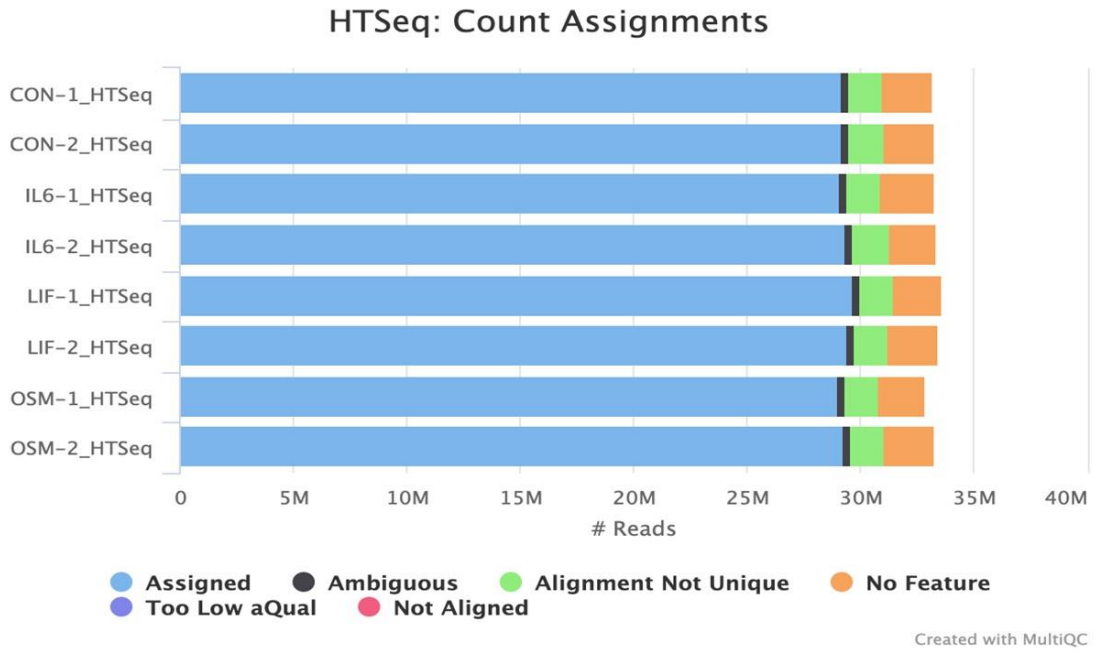

Figure 3.4: HTSeq gene count assignments generated by MultiQC.

3.5 Differential expression -DESeq2(R-Bioconductor)

We aim to compare cytokine-treated samples with control samples with the dataset constructed from experimental groups and corresponding treatments. For this purpose, tables from the preceding steps were joined with the gene quantification table of control samples. Then, DESeq2 estimates each gene's dispersion from these concatenated tables and analyzes if the genes are expressed differentially between the specified treatments (e.g., comparison between OSM-Control, LIF-Control, IL6-Control).

Table 3.3 indicates the total number of DEGs, upregulated, downregulated genes when comparing OSM vs. Control, IL6 vs. Control, and LIF vs. Control with adjusted p-value < 0.05. In the first column, the nonzero total read count indicates the total number of

detected genes with a nonzero count by DESeq2 for each sample type. The second and third columns give information about upregulated and downregulated genes, representing increasing and decreasing the quantity of RNA under a specific p-value, respectively. All these subjects were achieved using the DESeq2 package by running the summary function from dds with an adjusted p-value. The table shows that the number of nonzero totals reads counts in OSM treated samples were more than in IL6 and LIF treated samples, and there were no outliers in all samples. In OSM treated samples, the number of upregulated and downregulated genes is much greater than IL6 and LIF treated samples.IL6 treated samples have the least DEGs indicating IL6 treated samples are very similar to control groups.

Table 3.3: Statistics from DESeq2 (q-values < 0.05) results.

Sample	Total read count	Total DEGs	Upregulated	Downregulated
OSM	15152	3289	650	576
IL6	15152	17	1	16
LIF	15152	1149	178	225

3.4.1 Scatter plot

In RNA-Seq data, compared to other multidimensional data, variance grows with the mean, and DESeq2 offers the regularized-logarithm transformation or rlog for short to solve this. Regarding high gene counts, the rlog transformation does not vary a lot from an ordinary log2 transformation. Nevertheless, the values are contracted towards the genes` averages over all samples for genes with lower counts. Figure 3.5- Scatter plots of biological replicates across samples show a robust correlation between biological replicates for each condition.

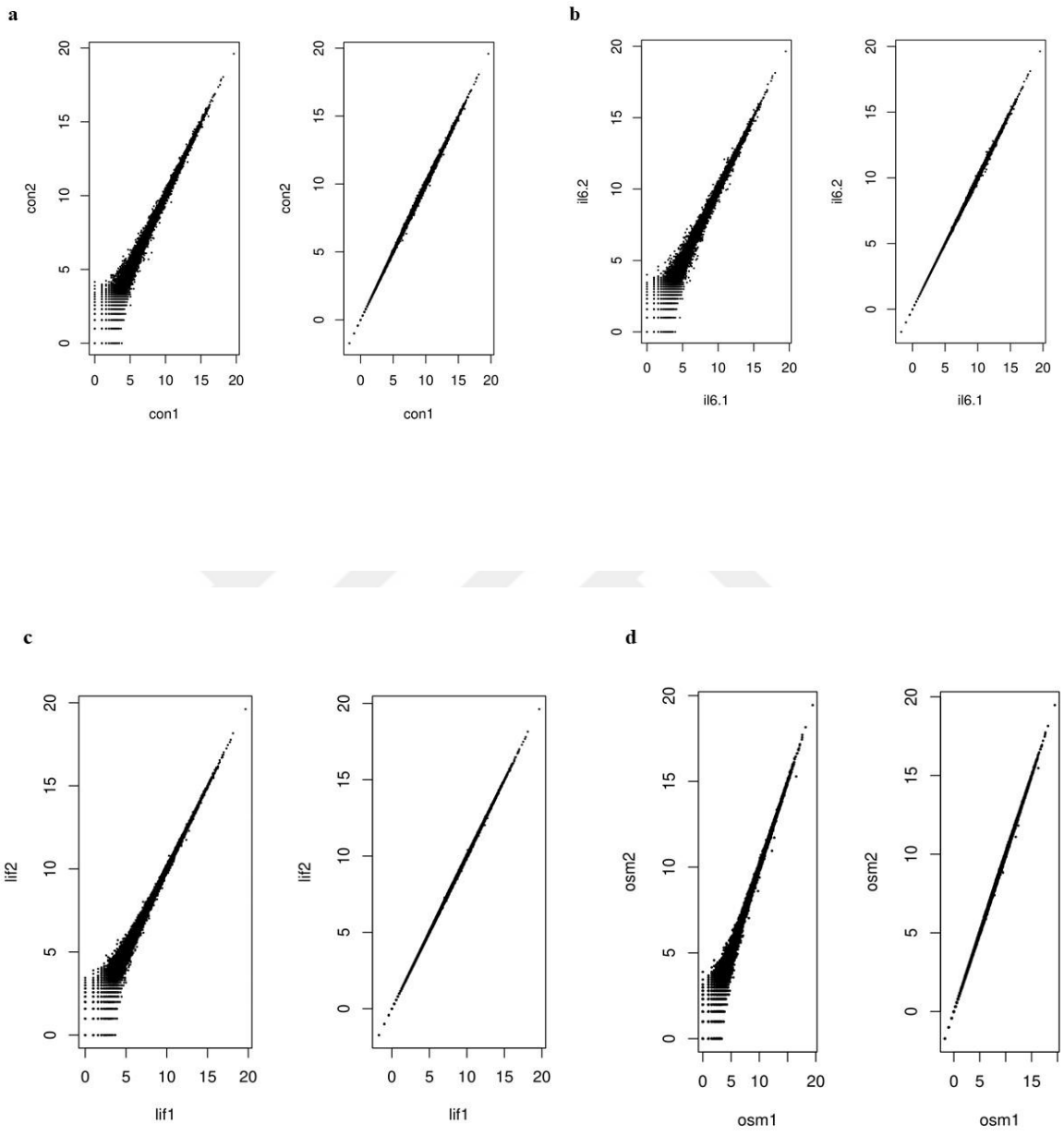

Figure 3.5: Scatter plot of biological replicates for each condition. Left: using an ordinary log2 transformation. Right: Using the rlog transformation.

3.4.2 MA plot

The MA (ratio intensity) plot displays the log-fold change against the expression, where "M" shows the log of the ratio of level counts for each gene between samples and "A" indicates the average level counts for each gene between two samples. MA plot can be used to analyze the normalization of the data and visualize reproducibility between samples. MA plot is used for representing log-fold change versus normalized mean expression between two groups of samples. Genes with similar expression levels in two samples will be presented around the horizontal line. Along the y-axis, data points with maximum values, when falling above the threshold, indicate the genes that have greatly differential expression levels and indicate the number of genes being upregulated. On the contrary, high levels of downregulated genes are shown as data points below the -1 threshold.

Figure 3.6: MA plot indicating the log fold-change (y-axis) versus mean expression (y-axis) created by R software using plotMA function in DESeq2.There were barely any significantly DEGs between IL6 and Control samples. Almost all genes were represented around the horizontal line, which means that the samples were expressed similarly. There were comparatively few blue points in the plot of LIF compared to OSM indicating there were fewer differential expressed genes in LIF samples than OSM samples. However, many dots are located around the y-axis and do not drift apart; therefore, a good dataset normalization assumption can be made. A small amount of significant DE genes can also be observed with the volcano plot. Most data points are close to zero on the y-axis, indicating treatment samples are similar to control groups in e expression patterns. Data points have been colored blue if the adjusted p-value is less than 0.05..

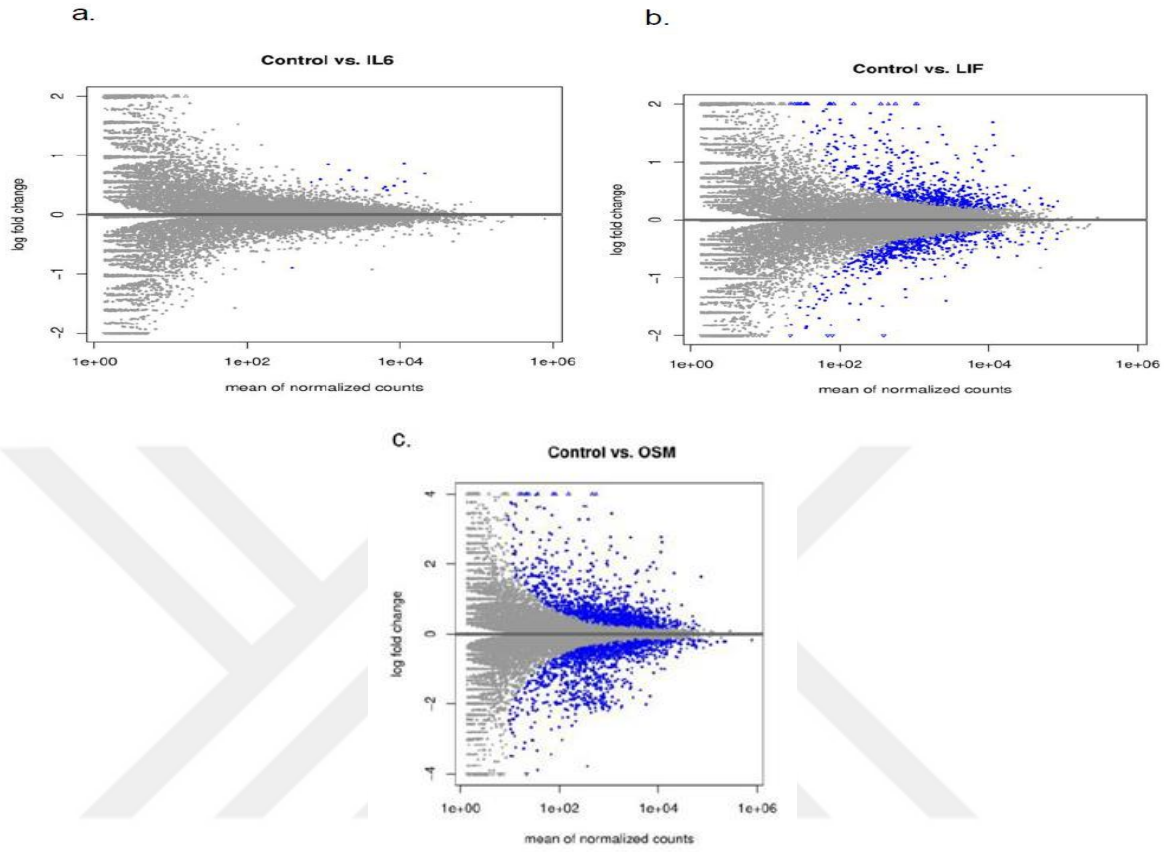

Figure 3.6: MA plot indicating the log fold-change (y-axis) versus mean expression (x-axis)created by R software using plotMA function in DESeq2.

3.4.3 .PCA plot

In order to view sample to sample distance, PCA (principal component analysis) plots were created. In a PCA, the data points (samples) are projected onto the 2D plane such that they spread out in the two directions that show most of the variances. In PCA analysis, only components one and two are usually considered. PC1 indicates the values of the samples on the x-axis direction that separates the data points the most. The y-axis separates the data the second-most, and the values of the samples in this direction are written PC2. Y-axis direction must be orthogonal to the first direction. For creating a PCA plot, the function of plotPCA was used in the DESeq2 package. This visualization method is a common way of detecting the similarity in gene expression between treatment groups. In a PCA, samples located close to each have more correlation and are more similar than other samples located far away. Clusters of samples in a PCA plot represent a strong correlation and are more similar than other samples located far away. Thus, clusters of samples in a PCA plot

represent a robust correlation of gene expression signals. If gene expression differences exist among the different treatment groups, it is foreseen that biological replicates of the same treatment will be grouped in PCA. Consistently, the PCA shows distinct clustering of the two replicates of each treatment condition, suggesting a substantial change in gene expression between each treatment but not much change in between replicates (Figure 3.7).

Interestingly, IL6 and Control replicate cluster closely together, suggesting much fewer changes in gene expression among the two conditions. On the other hand, OSM replicates cluster distinctly far from control replicates, indicating larger variances in gene expression among the two conditions. Analysis showed a perfect separation between different treatments. In Figure 3.7, the Control group overlapped with the IL6 group.

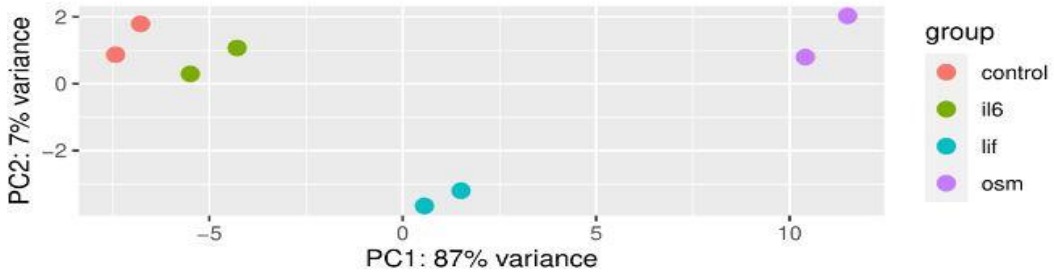

Figure 3.7: Principal component analysis (PCA) of the gene abundance in each condition.

3.4.4 Sample Correlation Heatmap

In order to view the correlations across samples and assess how the samples would create clusters and group together, heatmap with the default clustering method using Euclidean distance were generated for all four conditions using pheatmap function in R software DESeq2 package. The color in the heatmap provides information about similarity. The diagonal has only red colors, which illustrates the distance of the sample with itself. The red color means short distance and high similarity between samples. According to the legend at the top of the right side in the heatmap, the red color shows short distance and high correlations among samples.

On the other hand, the dendrogram (hierarchical clustering) shows how samples cluster together. The length of the vertical lines that connect to the samples shows the similarity between samples. Short connecting lines illustrate high similarity. Figure 3.8, the heatmap, illustrates the sample's correlation of Control, il6, lif, osm separately. In the

heatmap, samples were presented in both rows and columns, and the dendrogram at the top and left indicate how samples were clustered together. It can be seen that il6 and Control 2 cluster together, which was not expected.

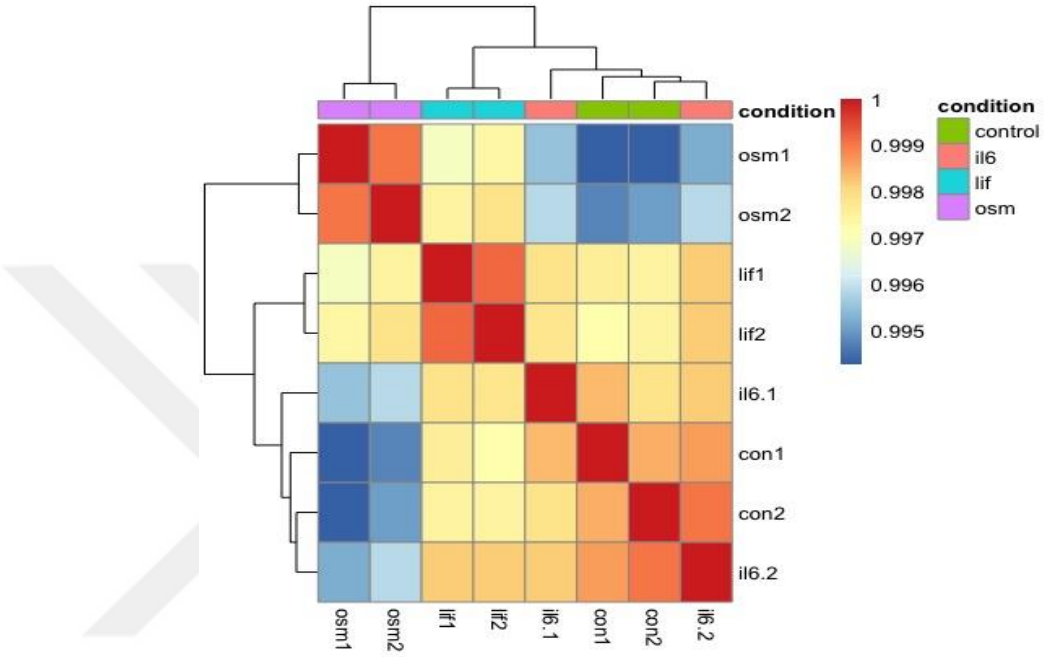

Figure 3.8: Heatmap of Euclidean sample distances after vst transformation.:

3.4.5 Volcano Plot

A volcano plot is a 2-dimensional (2D) scatter plot having a shape like a volcano. It is employed to visualize and identify statistically significant gene expression alteration between two experimental conditions regarding log fold change (x-axis) and p-value (Y-axis). For example, Figure 3.9 shows volcano plots of three conditions (il6, lif, osm). It can be seen that the IL6 sample has the least differential expressed genes and the OSM sample have the most significantly expressed genes as expected. In Figure 3.9, the significance is represented in the negative $\log_{10}$ of the p-value so that genes with greater significance are plotted higher on the y-axis. All differently expressed genes ($p<0.05$), $|\text{fold change}| >= 1.5$ are represented by the red (down-regulated) and green (up-regulated) points. Genes in gray are non-significant. Volcano plots are generated *bioinfokit.*

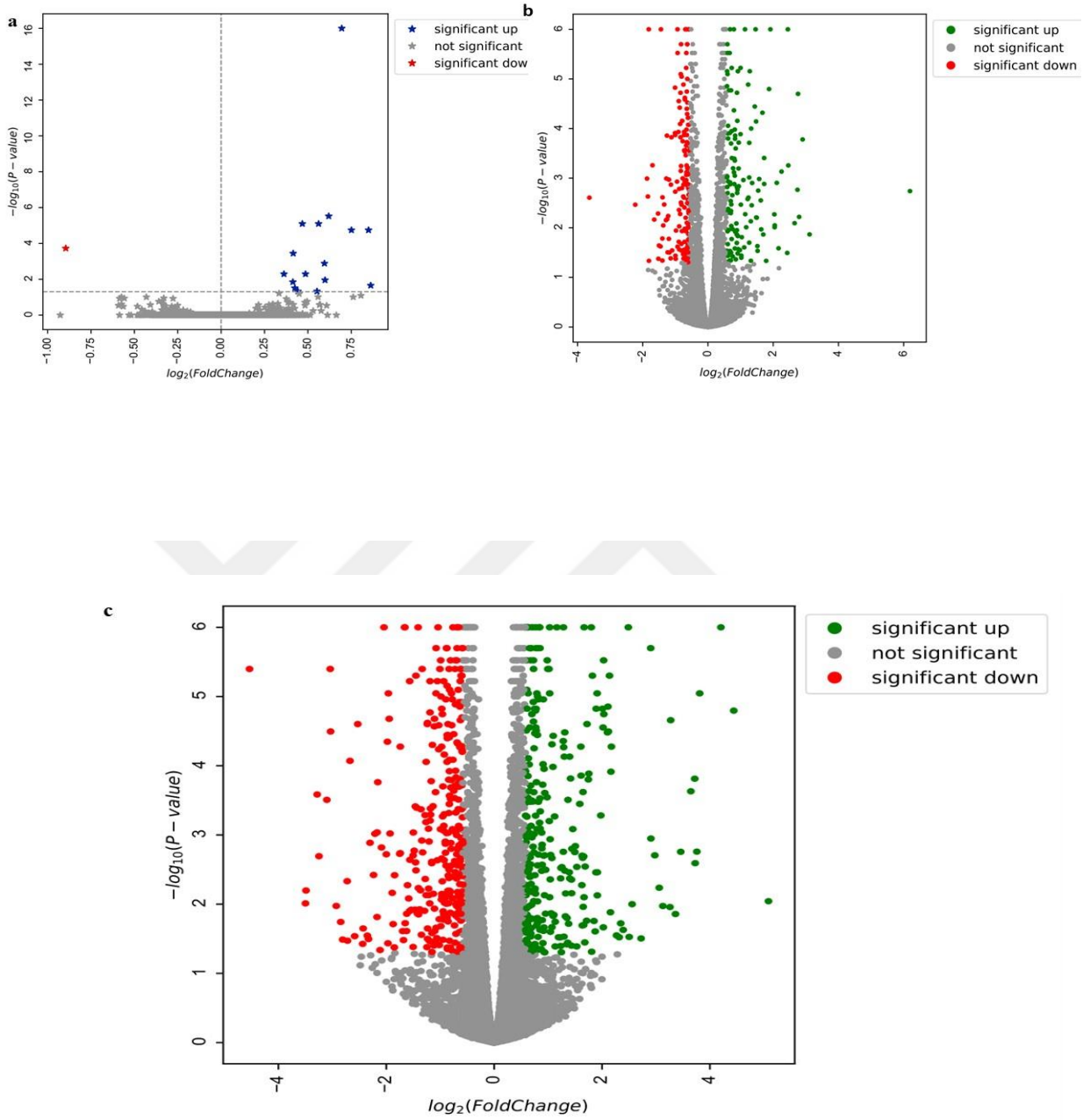

Figure 3.9: Volcano plot of genes. Genes are illustrated in terms of their measured expression change (x-axis)and the significance of the changes (y-axis).

3.4.6 Heatmap of Differentially Expressed Genes

The heatmap can be combined with clustering methods that group genes based on their gene expression pattern similarity. The data is displayed in a grid in heat maps where each row shows a gene, and each column shows a sample. The color and magnitude of the boxes are used to indicate changes (not absolute values) in gene expression. For example, in

below Figure 3.10, green represents upregulated genes, and red implies down-regulated genes. Black indicates unchanged expression. It can be seen from the heatmap; OSM has the most altered genes compared to control samples, and the gene expression pattern of IL6 is very similar to the Control group, and the gene expression pattern of LIF lies in between OSM and control groups.

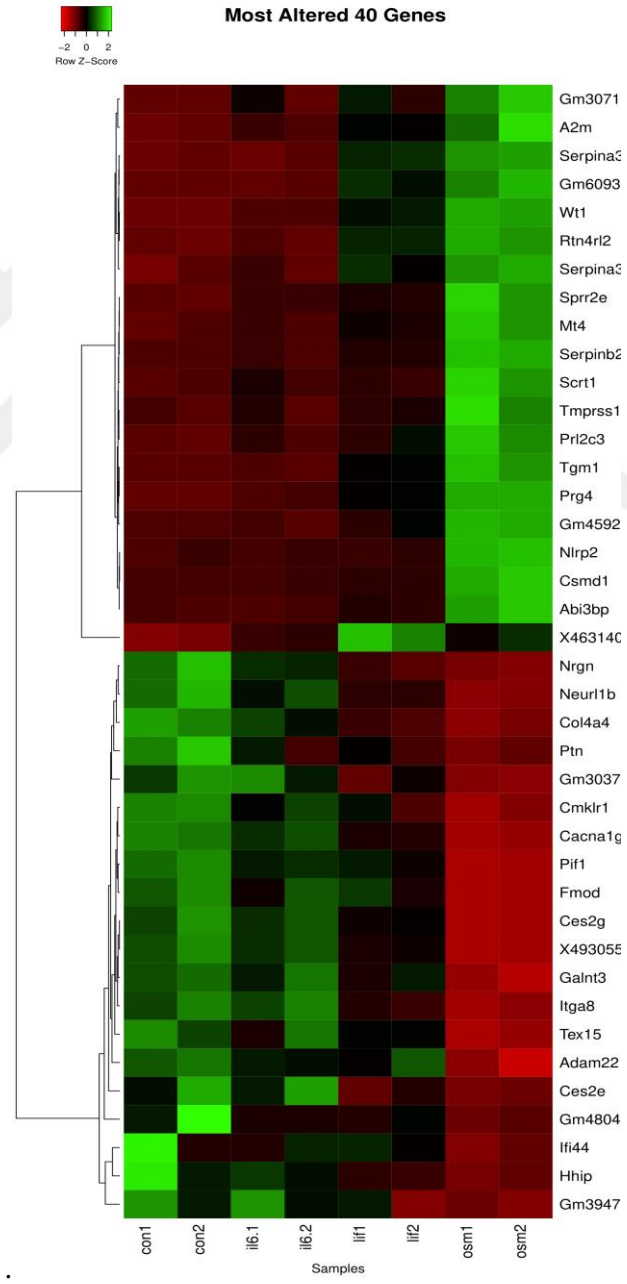

Figure 3.10: Heatmap of the most varied 40 genes across the samples.:

3.4.7 Veen diagram

Figure 3.11 illustrates the comparison between the DEGs detection (adj.p-value < 0.05) results across the samples (OSM, LIF, IL6). As shown in the below Venn diagram, 3007 transcript genes were stimulated by OSM, LIF, or IL6 compared to controls. 1886 (62.7%) were differentially expressed consequently to OSM treatment only, 267 (8.9%) by LIF treatment only. DESeq2 analysis of gene transcripts differentially expressed consequently to OSM, LIF, and IL6-treatment is expressed as a Venn diagram. Thus, the total 3007 transcripts stimulated by OSM, LIF, and IL6, 1886 (62.7%) were differentially expressed to OSM treatment only, 267 (8.9%) by LIF treatment is no IL6 treatment unique genes. Out of 3007 DEGs, 407 (64.4%) significantly upregulated genes are specific to OSM treated samples, and 56 (8.9%) genes are significantly upregulated specifically to LIF treated samples. There are no significantly upregulated genes specific to IL6 treated samples only. 509 (74.1%) genes are significantly downregulated specifically to OSM treated samples, and 37 (5.4%) genes are significantly downregulated specific to LIF treated samples. This shows that most DEGs are shared between samples indicating higher similarity in gene expression induced by these cytokines.

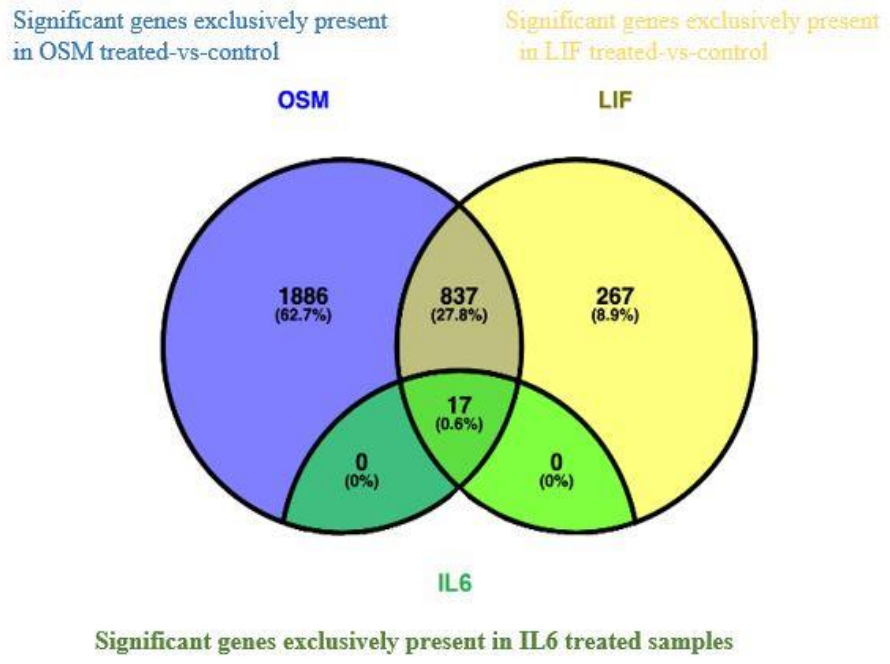

Figure 3.11: Veen Diagram of OSM, LIF, and IL6-Induced Transcriptomes.

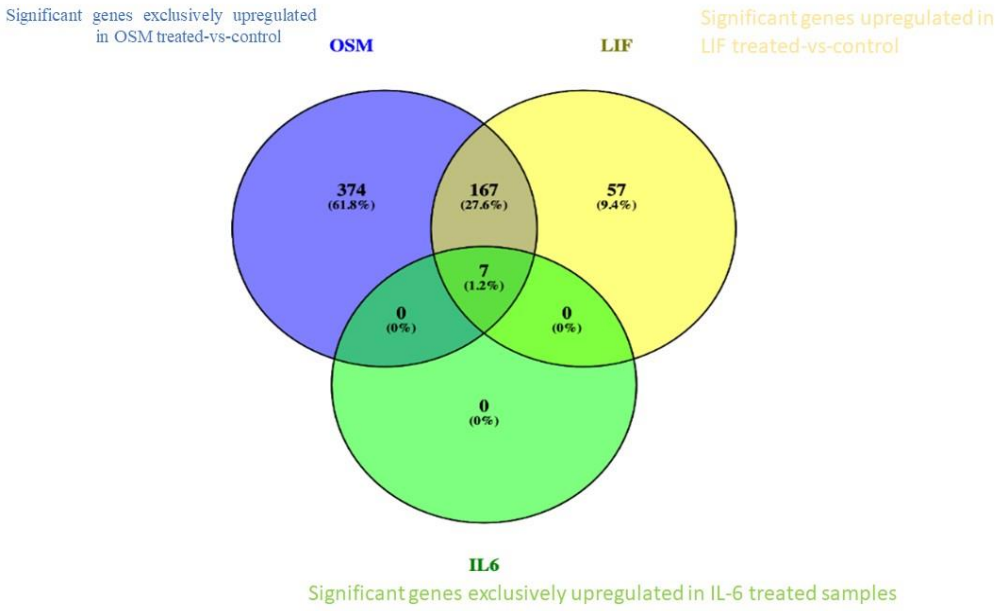

Figure 3.12: Veen Diagram of OSM, LIF, and IL6-Induced Significantly Upregulated genes.

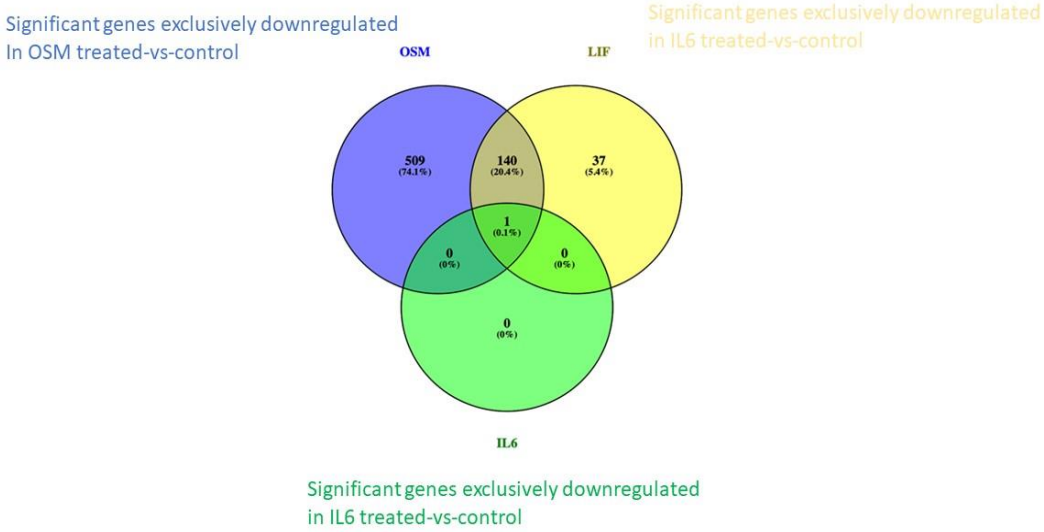

Figure 3.13: Veen Diagram of OSM, LIF, and IL6 Induced Significantly Down-regulated genes.

3.5 Gene Set Enrichment Analysis with GSEA

Gene set enrichment analysis was conducted using the fgsea package to resolve if specified lists (or sets) of genes indicate a statistically significant bias in their allocation within a ranked gene list (http://www.broadinstitute.org/gsea/; Subramanian A., 2005). It calculates an enrichment score for the test group by examining the ranking of its member genes using any suitable metric and determines the significance of the enrichment score by randomly permuting sample labels. For GSEA analysis, unfiltered genes were used without assigning significance, as advised by the authors. Instead, the approach utilizes an enrichment score (ES) statistic, computed based on a vector of gene-level signed statistics, for instance, t-statistic from a differential expression test. Results showed that top positively enriched pathways common to these cytokines include hypoxia, inflammatory response, IL6-JAK-STAT3, and glycolysis. However, compared to IL6 and LIF, and OSM showed the highest variance for these pathways. Hallmark pathways are shown below:

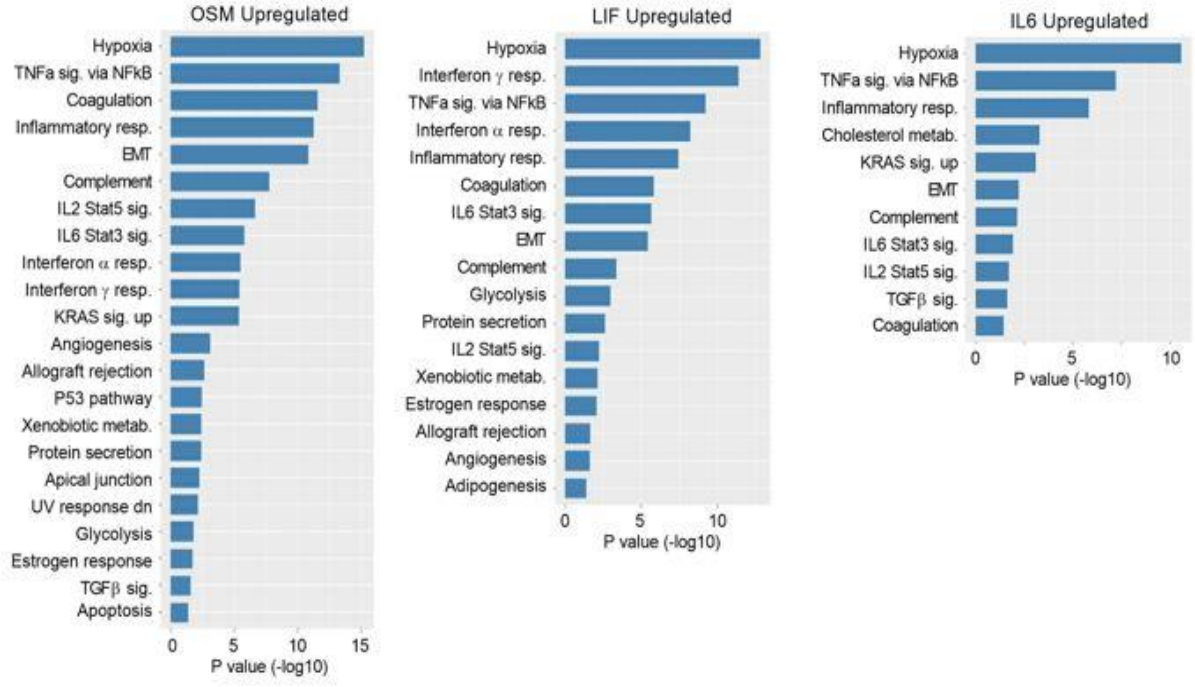

Figure 3.14: Positively Enriched Pathways Across the Different Treatment Groups.

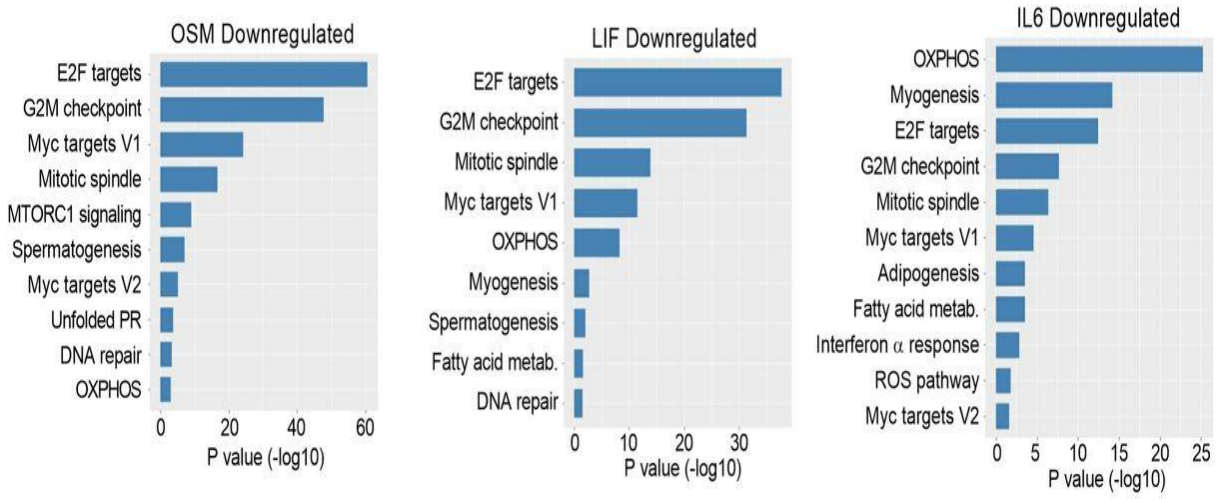

Figure 3.15: Negatively Enriched Pathways Across Different Treatment Groups.

3.5.1 Gene Set Enrichment Plot

GSEA method evaluates whether genes from the pathway of interest are randomly distributed through the expression profile of the experiment. For enriched pathways, it is expected that genes will be concentrated at the utmost or bottom-ranked genes from the experiment. Black vertical bars indicate the rank of genes in the gene set in the sorted list metric. The green curve corresponds to the enrichment score (ES). The more the green ES curve is shifted to the upper left of the graph, the more enriched the gene set in the first group.

Contrarily, a shift of the ES curve to the lower right corresponds to more enrichment in the second group. Figure 3.16 shows that genes in the OSM treatment samples are overexpressed with respect to the ranked gene list of hypoxia. Figure 3.17 shows the profile of the running ES score and gene set members` positions, and it displays that genes in the OSM treatment samples are overexpressed with respect to the ranked gene list of the IL6-JAK-STAT3 signaling pathway. It indicates that in the context of muscle atrophy, OSM may signal through these pathways.

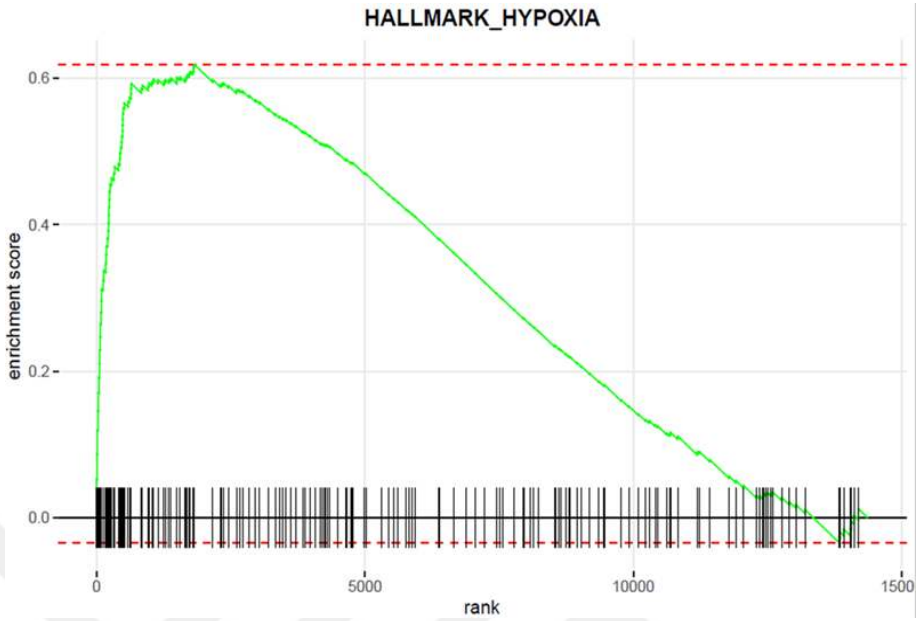

Figure 3.16: The enrichment plot of hallmark hypoxia.

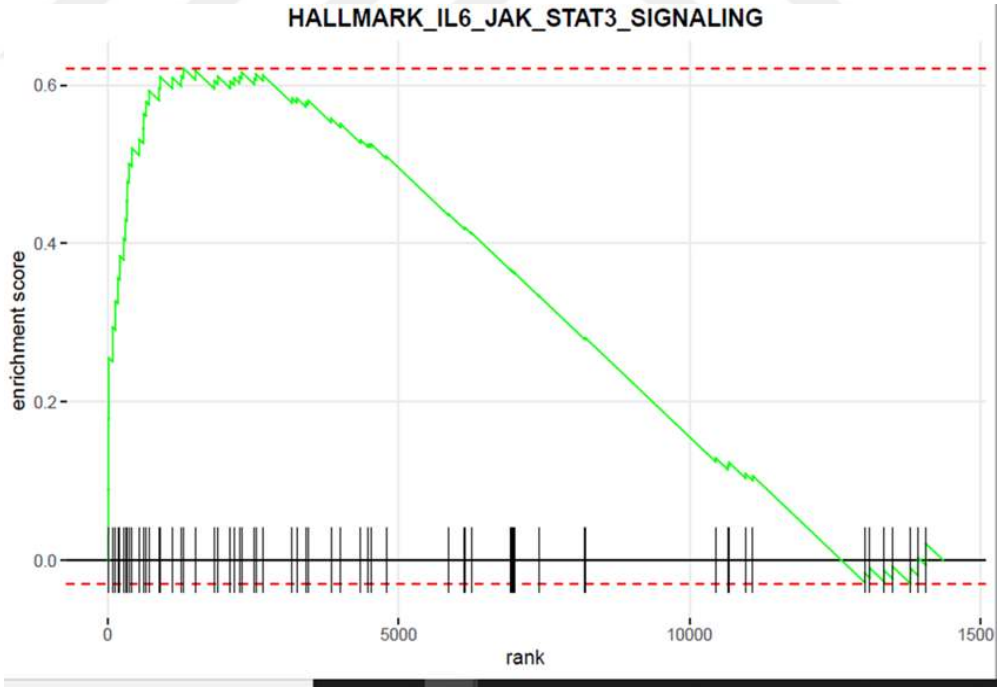

Figure 3.17: The enrichment plot of hallmark IL6 JAK STAT3 signaling.

3.5.2 Visualization with the Integrative Genomics Viewer (IGV)

The Integrative Genomics Viewer (IGV) is a visualization tool that facilitates immediate real-time inspection of various big-scale genomic data on a computer. With the

help of IGV, the indexed and sorted BAM file on IGV can be visualized to see the aligned regions. We utilized the Integrated Genome Viewer (browser) to view RNA-seq data coverage throughout the mouse genome. Coverage plots illustrate our analysis for cachectic marker genes (*Ampd3, Trim63/Murf1*). Aligned reads for *Ampd3* were uniformly aligned to the genomic location of chromosome 7 under all conditions, and aligned reads for *Trim 63/Murf1* were uniformly aligned to the genomic position of chromosome 4 under all conditions. Below, red represents peaks for OSM treated samples, and blue represents control samples. Consistent with previous results, overexpression of *Ampd3* and *Murf1* were also illustrated with higher red peaks compared to blue peaks.

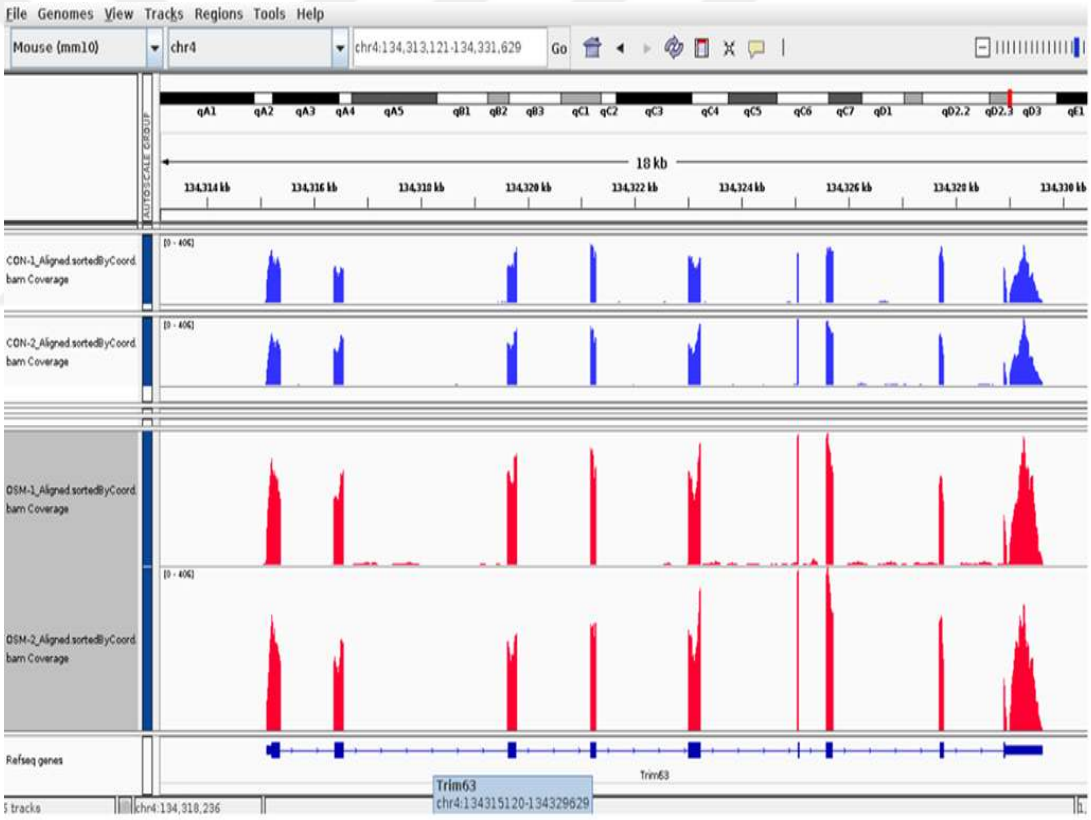

Figure 3.18: Integrated Genomics Viewer (IGV) illustrates base coverage for transcript Trim63 contrasting Control versus OSM samples.

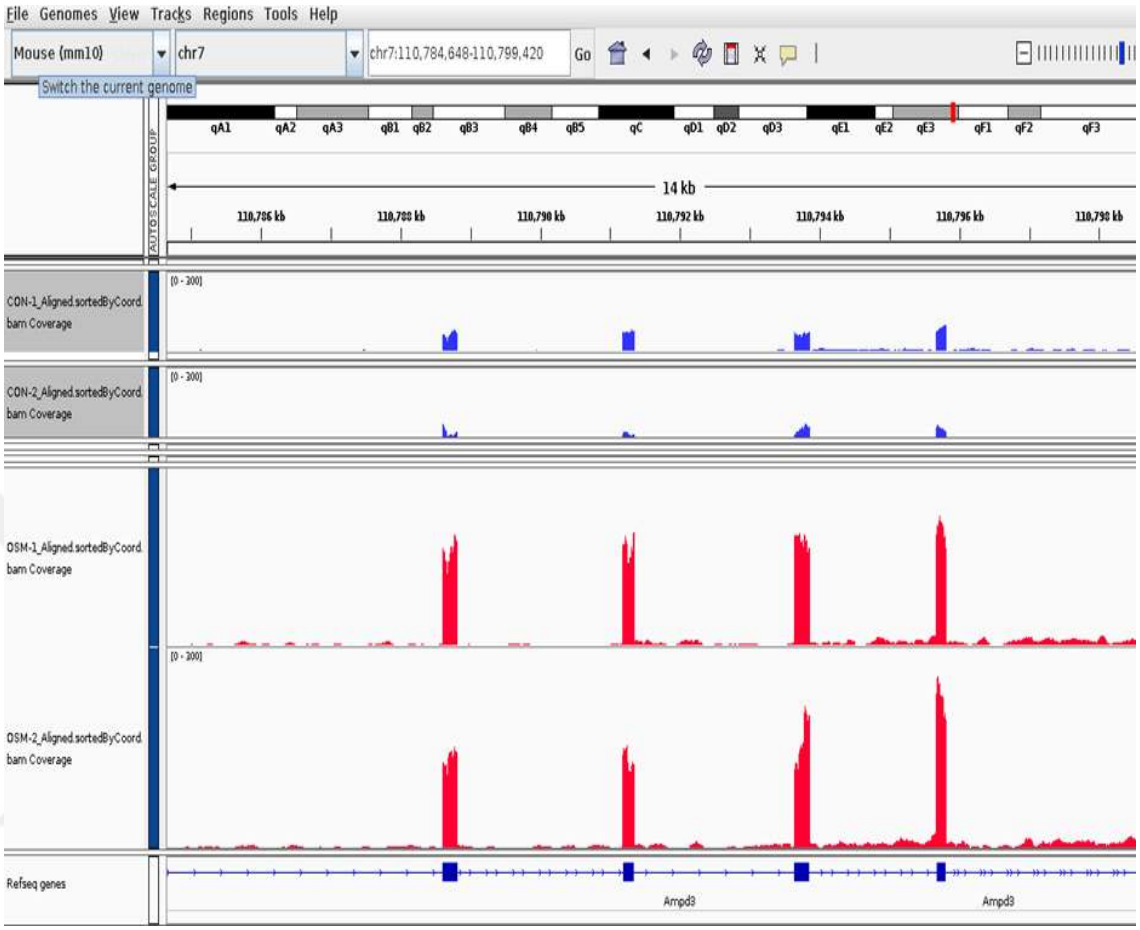

Figure 3.19: Integrated Genomics Viewer (IGV) illustrates base coverage for transcript Ampd3 contrasting Control versus OSM samples.

3.6 qRT-PCR Validation of RNA-Seq Results

To verify the accuracy of the expression profiles constructed by RNA-Seq and DEGs analysis, qRT-PCR was used to investigate upregulated muscle atrophy-related genes (*Atrogin-1, Ampd3, Sln, Serpina3n*) and top OSM target genes(*Serpinb2*, *Prg4, Wt1, Abi3bp, Sprr2e, Il33)*. Gene expression levels of these muscle atrophy-related genes of mouse primary myotube cells are measured in response to OSM treatment. It is displayed in Figure 3.20 (performed by Aylin Domaniku) that *Atrogin-1, Ampd3, Sln, Serpina3n* are significantly overexpressed in OSM treated samples compared to other groups, and it indicates that muscle atrophy is induced, suggesting that OSM is a stronger muscle atrophy inducer compared to IL6, and LIF. Furthermore, consistent with RNA-Seq results, top OSM

target genes (*Serpinb2*, *Prg4, Wt1, Abi3bp, Sprr2e, Il33)* are significantly highly overexpressed compared to IL6 and LIF conditions(Figure 3.21).

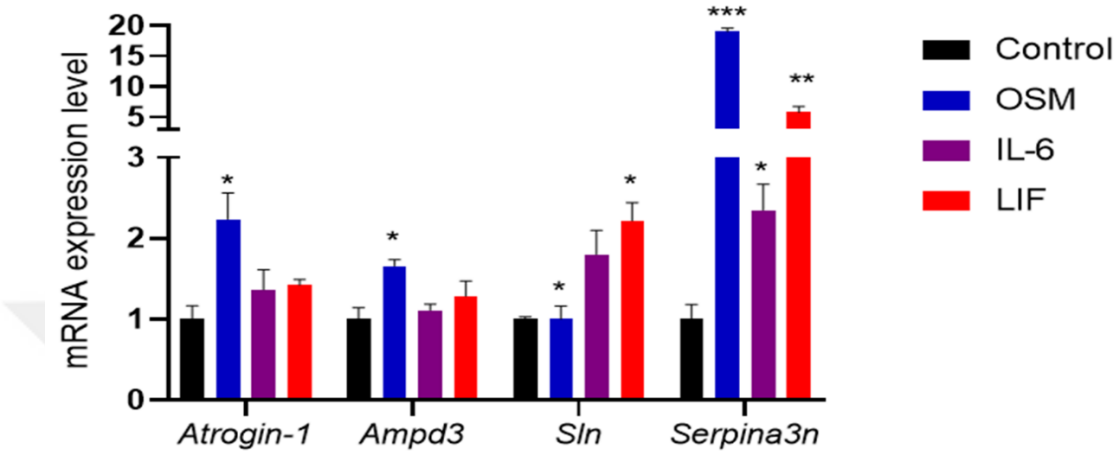

*Figure 3.20: Cytokine-induced atrophy related gene expression were verified using RT-qPCR. Error bars represent standard error. The values are mean ± SEM. *p < 0.05, **p < 0.005, ***p < 0.0005. Adapted from Aylin Domaniku.*

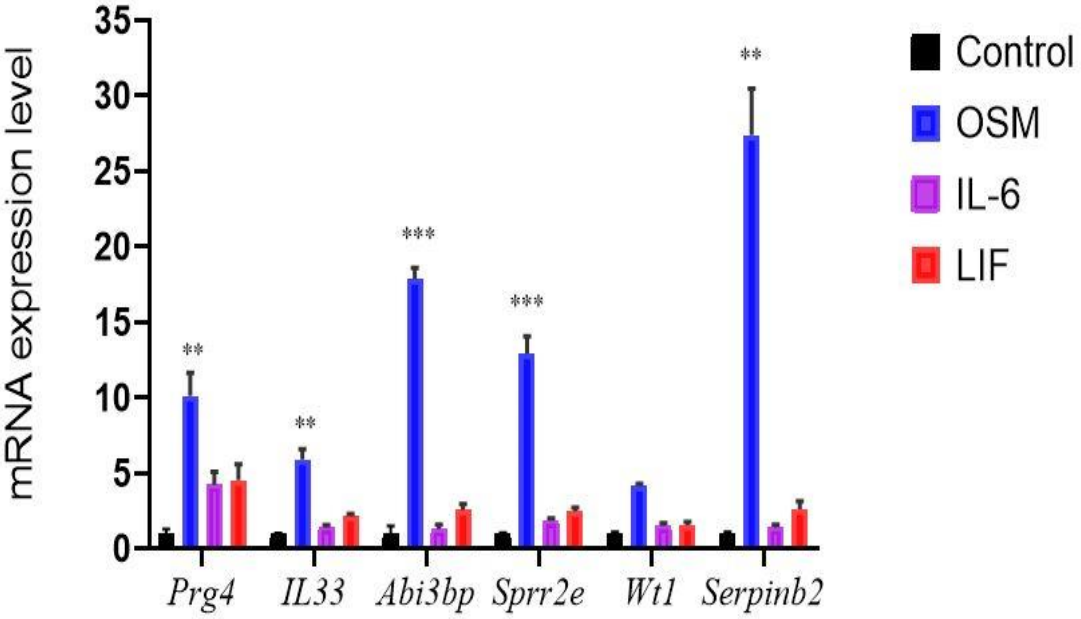

*Figure 3.21: Cytokine induced top OSM target gene expression were verified using RT-qPCR. Error bars represent standard error. The values are mean ± SEM. *p < 0.05, **p < 0.005, ***p < 0.0005. Data from Aylin Domaniku.*

Chapter 4:
CONCLUSIONS/DISCUSSION

With the advancement of high-throughput sequencing methods, gene-expression analysis by microarrays is being replaced by RNA-seq technology. This project used eight samples to profile gene expression and determine gene expression alterations of mouse primary myotube cells treated by IL-6 family cytokines (OSM, IL6, and LIF). We are interested in identifying upregulated transcripts/genes that may be implicated in muscle atrophy in response to cytokine stimulations. RNA-seq experiments create enormous amounts of data, and statistical analysis is required to reduce false-positive estimations and obtain desired results. Gene expression analysis by Microarray has many packages and well-defined methods to deal with differential expression data. However, such approaches cannot be directly employed for RNA-seq. Methods based on the negative binomial distribution (DESeq2, EdgeR) were developed to deal with RNA-seq data. However, these linear modeling methods cannot be used for data without biological replicates. We retrieved an average of 30 million reads per sample. An average of 93% of reads aligned to the reference mouse genome, with expression profiling efficiency (% exonic reads) of 88%, and expression of 15152 protein-coding genes were uncovered on average among samples. Clustering analysis (sample distance heatmap) showed that primary myotube cells treated with cytokine IL6 for 24 h grouped with control samples; however, myotube cells treated with cytokine OSM for 24 h clustered apart from control samples, signifying much potent response with cytokine OSM treatment, which is consistent with PCA analysis when samples are plotted against PC axes.

In the analysis, p-values were calculated from the raw counts (row sum) greater than 10 to avoid very low counts. A total of 3007 genes was identified as differentially expressed across the samples. The differentially expressed genes are filtered based on q-values and fold-change cutoff. The Venn diagrams and tab-delimited text files clearly show the differentially expressed genes and their related information. Out of 3007 DEGs induced by OSM treatment, 576 genes are significantly down-regulated, and 659 genes were significantly upregulated, including *Atrogin-1* and *Murf-1*. In an effort to identify more atrophy-related genes, a literature search showed that a hind limb immobilization study by

Tomiya et al.(2019) showed that skeletal muscle atrophy upregulates *Sln* expression (Tomiya et al., 2019). Another study by Gueugneau et al. (2018) reported increased expression of *Serpina3n* in muscle atrophy (Gueugneau et al., 2018). A study by Davis et al.(2015) further shows that overexpression of *Ampd3* initiates muscle atrophy in C2C12 myotubes((Davis, 2015). In agreement with these findings, our RNA-Seq analysis also revealed upregulation of *Sln, Ampd3,* and *Serpina3n* in cytokine treated samples, indicating that OSM induced muscle atrophy in primary myotube cells. Thus, these atrophy marker genes` expression is also validated using qPCR. In addition, DEGs were analyzed to assess enrichment pathways. Our results show shared enriched pathways, including hypoxia, glycolysis, inflammatory response, and IL6-JAK-STAT3 pathways; however, the magnitude of enrichment of these pathways varies among the samples, and OSM treated samples showed the greatest magnitude of enrichment. This finding implies that as an IL-6 family cytokine (OSM, LIF, IL-6), they may induce the same signaling pathways in cultured myotubes, which need to be validated with further studies.

Furthermore, Vollmer reported that the hypoxia-inducible factor 1alpha gene is upregulated in hepatocytes and hepatoma cells in response to OSM treatment (Vollmer et al., 2009). This increase is mediated via the JAK/STAT3 signaling pathway, suggesting a possible cross-talk between OSM and hypoxia signaling. Our analysis also uncovered the positive enrichment of the hypoxia signaling pathway upon OSM treatment; therefore, the possible link between OSM signaling and hypoxia signaling in muscle myotubes needs to be examined further.IL6 involvement in muscle atrophy is widely known. However, our findings suggest that compared to IL6, OSM is a potent inducer of cellular atrophy, evidenced by highly significant upregulation of muscle atrophy marker genes by OSM treatment but not by IL6 treatment. Therefore, we propose cytokine OSM as a novel potent muscle atrophy inducer, and targeting OSM can be a novel therapeutic target for muscle atrophy-related chronic diseases.

This master thesis work is only a part of long-term research focused on studying the mechanisms regulating skeletal muscle atrophy. Further in-vitro/in-vivo experiments are needed to identify and prove the novel candidate cachectic factors and genes. Network enrichment analysis (NEA) can be done on DE genes to identify signaling pathways

regulated by overrepresented genes. Besides, the genes in the system are also influenced by other interactions such as mRNA co-expression, protein-protein interactions, which are involved in myotube response to OSM treatment. To find such links, Funcoup analysis can be done on OSM signaling-related genes identified by NEA. The analysis helps to identify the nearby genes that were related to DEGs. In addition, research is ongoing to overcome the problems handling the RNA-seq data. However, the approach used in this project to identify differentially expressed genes and visualize the differential expressed genes from RNA-seq is promising and efficient in deriving the solution. With the help of statistical analysis, many differentially expressed genes were narrowed down, and a significant number of genes associated with skeletal muscle atrophy were emphasized. This enables biologists to focus on a shortlist of genes for further analysis and reduce the labor devoted to experiments.

Furthermore, by GSEA, signaling pathways, such as hypoxia, glycolysis, angiogenesis, and IL6-Jak-STAT3 pathways, were associated with myotube cell response to OSM treatment. RNA-Seq data analysis also revealed cachectic markers such as *Atrogin1, Murf1, Ampd3, Serpina3n,* and *Sln*. Future studies should address the roles of the pathways and genes in the development of muscle atrophy.

BIBLIOGRAPHY

KATZ, A. M., & KATZ, P. B. (1962). Diseases of the heart in the works of Hippocrates. *British heart journal*, *24*(3), 257–264. https://doi.org/10.1136/hrt.24.3.257

Argilés, J. M., Busquets, S., Stemmler, B., & López-Soriano, F. J. (2014). Cancer cachexia: Understanding the molecular basis. *Nature Reviews Cancer*, *14*(11), 754–762. https://doi.org/10.1038/nrc3829

Fearon, K., Strasser, F., Anker, S. D., Bosaeus, I., Bruera, E., Fainsinger, R. L., Jatoi, A., Loprinzi, C., MacDonald, N., Mantovani, G., Davis, M., Muscaritoli, M., Ottery, F., Radbruch, L., Ravasco, P., Walsh, D., Wilcock, A., Kaasa, S., & Baracos, V. E. (2011). Definition and classification of cancer cachexia: An international consensus. *The Lancet Oncology*, *12*(5), 489–495. https://doi.org/10.1016/S1470-2045(10)70218-7

Argilés, J. M., Busquets, S., Stemmler, B., & López-Soriano, F. J. (2014). Cancer cachexia: Understanding the molecular basis. *Nature Reviews Cancer*, *14*(11), 754–762.

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Argilés, J. M., Busquets, S., Stemmler, B., & López-Soriano, F. J. (2014). Cancer cachexia: Understanding the molecular basis. *Nature Reviews Cancer*, *14*(11), 754–762. https://doi.org/10.1038/nrc3829

Attard-Montalto, S. P., Camacho-Hübner, C., Cotterill, A. M., D'Souza-Li, L., Daley, S., Bartlett, K., Halliday, D., & Eden, O. B. (1998). Changes in protein turnover, IGF-I and IGF binding proteins in children with cancer. *Acta Paediatrica, International Journal of Paediatrics*, *87*(1), 54–60. https://doi.org/10.1080/08035259850157877

Augenlicht, L. H., & Kobrin, D. (1982). Cloning and Screening of SequencesExpressedin a MouseColon Tumor1 The abbreviations used are poly-A @, polyadenylic avian virus reverse (a. *Cancer Research*, *42*(March).

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Beutler, B., & Cerami, A. (1986). Cachectin and tumor necrosis factor as two sides of the same biological coin. *Nature*, *320*(6063), 584–588. https://doi.org/10.1038/320584a0

Bhattacharjee, A., Richards, W. G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., Loda, M., Weber, G., Mark, E. J., Lander, E. S., Wong, W., Johnson, B. E., Golub, T. R., Sugarbaker, D. J., Meyerson, M., … Friend, S. H. (2002). Discovery and analysis of inflammatory disease-related genes using cDNA

microarrays. *Proceedings of the National Academy of Sciences of the United States of America*, *415*(6871), 2150–2155. https://doi.org/10.1073/pnas.94.6.2150

Biolo, G., Bosutti, A., Iscra, F., Toigo, G., Gullo, A., & Guarnieri, G. (1982). *Contribution of the Ubiquitin-Proteasome Pathway to Overall Muscle Proteolysis in Hypercatabolic Patients Ubiquitin-Proteasome Pathway to Overall Muscle Proteolysis in Hypercatabolic Patients*. *XXXI*(1), 1–5.

Block, C., Ks, O., Growth, C., Binding, O. M., & Antibodies, A. (1995). *AIDS-associated Kaposi's Sarcoma (KS) Cells Express Oncostatin M (OM) -specific Receptor but Not Leukemia Inhibitory Factor / OM Receptor or Interleukin-6 Receptor*. 1319–1327.

Brenner, D. A., Waterboer, T., Choi, S. K., Lindquist, J. N., Stefanovic, B., Burchardt, E., Yamauchi, M., Gillan, A., & Rippe, R. A. (2000). *New aspects of hepatic fibrosis Journal of Hepatology*. *32*(1), 32–38.

Carlson, M., Obenchain, V., Pages, H., Shannon, P., Tenenbaum, D., & Morgan, M. (2012). *High-throughput sequence analysis with R and*. 83.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dodson, S., Baracos, V. E., Jatoi, A., Evans, W. J., Cella, D., Dalton, J. T., & Steiner, M. S. (2011). Muscle wasting in cancer cachexia: clinical implications, diagnosis, and emerging treatment strategies. *Annual Review of Medicine*, *62*, 265–279. https://doi.org/10.1146/annurev-med-061509-131248

Emmert-Streib, F., & Glazko, G. V. (2011). Pathway analysis of expression data: Deciphering functional building blocks of complex diseases. *PLoS Computational Biology*, *7*(5), 1–6. https://doi.org/10.1371/journal.pcbi.1002053

Evans, William. J, Morley, J. . (2008). Cachexia A new definition.pdf. *Clinical Nutrition*, *27*, 793–799. https://doi.org/10.1016/j.clnu.2008.06.013

Fearon, K., Glass, D., & Guttridge, D. C. (2012). Cancer cachexia: Mediators, signaling, and metabolic pathways. *Cell Metabolism*, *16*(2), 153–166. https://doi.org/10.1016/j.cmet.2012.06.011

Fearon, K., Strasser, F., Anker, S. D., Bosaeus, I., Bruera, E., Fainsinger, R. L., Jatoi, A., Loprinzi, C., MacDonald, N., Mantovani, G., Davis, M., Muscaritoli, M., Ottery, F., Radbruch, L., Ravasco, P., Walsh, D., Wilcock, A., Kaasa, S., & Baracos, V. E. (2011). Definition and classification of cancer cachexia: An international consensus. *The Lancet Oncology*, *12*(5), 489–495. https://doi.org/10.1016/S1470-2045(10)70218-7

Flicek, P., Ahmed, I., Amode, M. R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., Gil, L., García-Girón, C., Gordon, L., Hourlier, T., Hunt, S., Juettemann, T., Kähäri, A. K., Keenan, S., … Searle, S. M. J.

(2013). Ensembl 2013. *Nucleic Acids Research*, *41*(D1), 48–55. https://doi.org/10.1093/nar/gks1236

Glass, D. J. (2010). Signaling pathways perturbing muscle mass. *Current Opinion in Clinical Nutrition and Metabolic Care*, *13*(3), 225–229. https://doi.org/10.1097/MCO.0b013e32833862df

Guttman, M., Garber, M., Levin, J. Z., Donaghey, J., Robinson, J., Adiconis, X., Fan, L., Koziol, M. J., Gnirke, A., Nusbaum, C., Rinn, J. L., Lander, E. S., & Regev, A. (2010). Erratum: Ab initio reconstruction of cell type-specific transcriptomes in mice reveals the conserved multi-exonic structure of lincRNAs (Nat. Biotechnol. (2010) 28 (503-510)). *Nature Biotechnology*, *28*(7), 756. https://doi.org/10.1038/nbt0710-756b

Haddad, F., Zaldivar, F., Cooper, D. M., & Adams, G. R. (2005). IL-6-induced skeletal muscle atrophy. *Journal of Applied Physiology*, *98*(3), 911–917. https://doi.org/10.1152/japplphysiol.01026.2004

Halfter, H., Lotfi, R., Westermann, R., Young, P., Bernd Ringelstein, E., & Stögbauer, F. T. (1998). Inhibition of growth and induction of differentiation of glioma cell lines by Oncostatin M (OSM). *Growth Factors*, *15*(2), 135–147. https://doi.org/10.3109/08977199809117189

Heinrich, P. C., Behrmann, I., Haan, S., Hermanns, H. M., Müller-Newen, G., & Schaper, F. (2003). Principles of interleukin (IL)-6-type cytokine signalling and its regulation. *Biochemical Journal*, *374*(1), 1–20. https://doi.org/10.1042/BJ20030407

Heller, R. A., Schena, M., Chai, A., Shalon, D., Bedilion, T., Gilmore, J., Woolley, D. E., & Davis, R. W. (1997). Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(6), 2150–2155. https://doi.org/10.1073/pnas.94.6.2150

Hrdlickova, R., Toloue, M., & Tian, B. (2017). RNA-seq methods for transcriptome analysis. *Journal of Cellular Biochemistry*, *8*(1), 1–24. https://doi.org/10.1002/wrna.1364.RNA-Seq

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44–57. https://doi.org/10.1038/nprot.2008.211

Kami, K., & Senba, E. (1998). Localization of leukemia inhibitory factor and interleukin-6 messenger ribonucleic acids in regenerating rat skeletal muscle. *Muscle and Nerve*, *21*(6), 819–822. https://doi.org/10.1002/(SICI)1097-4598(199806)21:6<819:AID-MUS20>3.0.CO;2-M

Khal, J., Wyke, S. M., Russell, S. T., Hine, A. V., & Tisdale, M. J. (2005). Expression of the ubiquitin-proteasome pathway and muscle loss in experimental cancer cachexia. *British Journal of Cancer*, *93*(7), 774–780. https://doi.org/10.1038/sj.bjc.6602780

Kimball, S. R., Farrell, P. A., & Jefferson, L. S. (2002). Exercise effects on muscle insulin signaling and action invited to review: Role of insulin in translational control of protein synthesis in skeletal muscle by amino acids or exercise. *Journal of Applied Physiology*, *93*(3), 1168–1180. https://doi.org/10.1152/japplphysiol.00221.2002

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, H., & Homer, N. (2010). A survey of sequence alignment algorithms for next-generation sequencing. *Briefings in Bioinformatics*, *11*(5), 473–483. https://doi.org/10.1093/bib/bbq015

Liao, Y., Smyth, G. K., & Shi, W. (2014). FeatureCounts: An efficient general-purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Lindberg, R. A., Juan, T. S.-C., Welcher, A. A., Sun, Y., Cupples, R., Guthrie, B., & Fletcher, F. A. (1998). Cloning and Characterization of a Specific Receptor for Mouse Oncostatin M. *Molecular and Cellular Biology*, *18*(6), 3357–3367. https://doi.org/10.1128/mcb.18.6.3357

Love, M., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. In *bioRxiv*. https://doi.org/10.1101/002832

M. Christopher, A. M. L. S. (2016). Oncostatin M and its role in fibrosis. *Physiology & Behavior*, *176*(1), 100–106. https://doi.org/10.1080/03008207.2018.1500558.Oncostatin

Malik, N., Haugen, H. S., Modrell, B., Shoyab, M., & Clegg, C. H. (1995). Developmental abnormalities in mice transgenic for bovine Oncostatin M. *Molecular and Cellular Biology*, *15*(5), 2349–2358. https://doi.org/10.1128/mcb.15.5.2349

Malik, N., Kallestad, J. C., Gunderson, N. L., Austin, S. D., Neubauer, M. G., Ochs, V., Marquardt, H., Zarling, J. M., Shoyab, M., & Wei, C. M. (1989). Molecular cloning, sequence analysis, and functional expression of a novel growth regulator, Oncostatin M. *Molecular and Cellular Biology*, *9*(7), 2847–2853. https://doi.org/10.1128/mcb.9.7.2847

Masjedi, A., Hajizadeh, F., Beigi Dargani, F., Beyzai, B., Aksoun, M., Hojjat-Farsangi, M., Zekiy, A., & Jadidi-Niaragh, F. (2021). Oncostatin M: A mysterious cytokine in cancers. *International Immunopharmacology*, *90*(July), 107158. https://doi.org/10.1016/j.intimp.2020.107158

Modified, A., Fibers, W., & Alkaline, U. (1985). *Alternative Isoform Regulation in Human Tissue Transcriptomes*. *456*(7221), 470–476. https://doi.org/10.1038/nature07509.Alternative

Moley, J. F., Green, M. V., Carson, R. E., Morrison, S. D., & Norton, J. A. (1985). Parabiotic Transfer of Cancer Anorexia/Cachexia in Male Rats. *Cancer Research*, *45*(November), 5547–5552.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, *5*(7), 621–628. https://doi.org/10.1038/nmeth.1226

Mukouyama, Y. S., Hara, T., Xu, M. J., Tamura, K., Donovan, P. J., Kim, H. J., Kogo, H., Tsuji, K., Nakahata, T., & Miyajima, A. (1998). In vitro expansion of murine multipotential hematopoietic progenitors from the embryonic aorta-gonad-mesonephros region. *Immunity*, *8*(1), 105–114. https://doi.org/10.1016/S1074-7613(00)80463-X

Muñoz-Cánoves, P., Scheele, C., Pedersen, B. K., & Serrano, A. L. (2013). Interleukin-6 myokine signaling in skeletal muscle: A double-edged sword? *FEBS Journal*, *280*(17), 4131–4148. https://doi.org/10.1111/febs.12338

Murphy, L. D., Herzog, C. E., Rudick, J. B., Fojo, A. T., & Bates, S. E. (1990). Use of the Polymerase Chain Reaction in the Quantitation of mdr-1 Gene Expression. *Biochemistry*, *29*(45), 10351–10356. https://doi.org/10.1021/bi00497a009

Pachter, L. (2011). *Models for transcript quantification from RNA-Seq*. 1–28. http://arxiv.org/abs/1104.3889

Pagliarulo, V., Datar, R. H., & Cote, R. J. (2002). Role of genetic and expression profiling in pharmacogenomics: The changing face of patient management. *Current Issues in Molecular Biology*, *4*(4), 101–110. https://doi.org/10.21775/cimb.004.101

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C. E., Socci, N. D., & Betel, D. (2013). Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biology*, *14*(9). https://doi.org/10.1186/gb-2013-14-9-r95

Robinson, M. D., & Oshlack, A. (2010). *A scaling normalization method for differential expression analysis of RNA-seq data*. 1–12.

Rose-John, S. (2018). Interleukin-6 family cytokines. *Cold Spring Harbor Perspectives in Biology*, *10*(2), 1–18. https://doi.org/10.1101/cshperspect.a028415

Sanger, F., Nicklen, S., & Coulson, A. (1990). DNA sequencing with chain-terminating inhibitors. *Obstetrics and Gynecology*, *75*(4), 59S-76S. https://doi.org/10.1097/00006250-199004001-00013

Schaefer, L. K., Wang, S., & Schaefer, T. S. (2000). Oncostatin M activates stat DNA binding and transcriptional activity in primary human fetal astrocytes: Low- and high-

passage cells have distinct patterns of stat activation. *Cytokine*, *12*(11), 1647–1655. https://doi.org/10.1006/cyto.2000.0774

Schena, M., Shalon, D., Davis, R. W., & Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, *270*(5235), 467–470. https://doi.org/10.1126/science.270.5235.467

Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., & Fukuda, S. (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(26), 15776–15781.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

Sun, L., Ma, K., Wang, H., Xiao, F., Gao, Y., Zhang, W., Wang, K., Gao, X., Ip, N., & Wu, Z. (2007). JAK1-STAT1-STAT3, a key pathway promoting proliferation and preventing premature differentiation of myoblasts. *Journal of Cell Biology*, *179*(1), 129–138. https://doi.org/10.1083/jcb.200703184

Suzuki, H., Asakawa, A., Amitani, H., Nakamura, N., & Inui, A. (2013). Cancer cachexia - Pathophysiology and management. *Journal of Gastroenterology*, *48*(5), 574–594. https://doi.org/10.1007/s00535-013-0787-0

Tisdale, M. J. (1997). Biology of cachexia. *Journal of the National Cancer Institute*, *89*(23), 1763–1773. https://doi.org/10.1093/jnci/89.23.1763

Todorov, P., Cariuk, P., McDevitt, T., Coles, B., Fearon, K., & Tisdale, M. (1996). Characterization of a cancer cachectic factor. In *Nature* (Vol. 379, Issue 6567, pp. 739–742). https://doi.org/10.1038/379739a0

Trivedi, U. H., Cézard, T., Bridgett, S., Montazam, A., Nichols, J., Blaxter, M., & Gharbi, K. (2014). Quality control of next-generation sequencing data without a reference. *Frontiers in Genetics*, *5*(MAY), 1–13. https://doi.org/10.3389/fgene.2014.00111

Veer, van`t laura, Dai, H., Vijver, M., He, Y., Hart, A., & Mao, M. (2002). Gene expression profiling predicts the clinical outcome of breast cancer. *Nature*, *415*(6871), 530–536. https://doi.org/10.1038/415530a

Velculescu, V., Zhang, L., Vogelstein, B., & Kinzler, K. (1995). Serial analysis of gene expression. *Nature Protocols*, *270*(4), 1743–1760. https://doi.org/10.1038/nprot.2006.269

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics in Western Equatoria State. *Nature Reviews Genetics*, *10*(1), 57.

White, J. P., Baynes, J. W., Welle, S. L., Kostek, M. C., Matesic, L. E., Sato, S., & Carson, J. A. (2011). The regulation of skeletal muscle protein turnover during the progression of cancer cachexia in the Apc Min/+ mouse. *PLoS ONE*, *6*(9). https://doi.org/10.1371/journal.pone.0024650

Wilhelm, B. T., & Landry, J. R. (2009). RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods*, *48*(3), 249–257. https://doi.org/10.1016/j.ymeth.2009.03.016

Yang, J., Cao, R. Y., Li, Q., & Zhu, F. (2018). Muscle atrophy in cancer. *Advances in Experimental Medicine and Biology*, *1088*, 329–346. https://doi.org/10.1007/978-981-13-1435-3_15

Zarling, J. M., Shoyab, M., Marquardt, H., Hanson, M. B., Lioubin, M. N., & Todaro, G. J. (1986). Oncostatin M: A growth regulator produced by differentiated histiocytic lymphoma cells. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(24), 9739–9743. https://doi.org/10.1073/pnas.83.24.9739

Zhang, X. G., Gu, J. J., Lu, Z. Y., Yasukawa, K., Yancopoulos, G. D., Turner, K., Shoyab, M., Taga, T., Kishimoto, T., Bataille, R., & Klein, B. (1994). Ciliary neurotrophic factor, interleukin 11, leukemia inhibitory factor, and Oncostatin M are growth factors for human myeloma cell lines using the interleukin six signal transducer GP130. *Journal of Experimental Medicine*, *179*(4), 1337–1342. https://doi.org/10.1084/jem.179.4.1337

Argilés, J. M., Busquets, S., Stemmler, B., & López-Soriano, F. J. (2014). Cancer cachexia: Understanding the molecular basis. *Nature Reviews Cancer*, *14*(11), 754–762. https://doi.org/10.1038/nrc3829

Gueugneau, M., d'Hose, D., Barbé, C., de Barsy, M., Lause, P., Maiter, D., Bindels, L. B., Delzenne, N. M., Schaeffer, L., Gangloff, Y. G., Chambon, C., Coudy-Gandilhon, C., Béchet, D., & Thissen, J. P. (2018). Increased Serpina3n release into circulation during glucocorticoid-mediated muscle atrophy. *Journal of Cachexia, Sarcopenia and Muscle*, *9*(5), 929–946. https://doi.org/10.1002/jcsm.12315

Tomiya, S., Tamura, Y., Kouzaki, K., Kotani, T., Wakabayashi, Y., Noda, M., & Nakazato, K. (2019). Cast immobilization of hindlimb upregulates sarcolipin expression in atrophied skeletal muscles and increases thermogenesis in C57BL/6J mice. *American Journal of Physiology - Regulatory Integrative and Comparative Physiology*, *317*(5), R649–R661. https://doi.org/10.1152/ajpregu.00118.2019

Vollmer, S., Kappler, V., Kaczor, J., Flügel, D., Rolvering, C., Kato, N., Kietzmann, T., Behrmann, I., & Haan, C. (2009). Hypoxia-inducible factor 1α is up-regulated by oncostatin M and participates in oncostatin M signaling. *Hepatology*, *50*(1), 253–260. https://doi.org/10.1002/hep.22928

Argilés, J. M., Busquets, S., Stemmler, B., & López-Soriano, F. J. (2014). Cancer cachexia: Understanding the molecular basis. *Nature Reviews Cancer*, *14*(11), 754–762. https://doi.org/10.1038/nrc3829

Gueugneau, M., d'Hose, D., Barbé, C., de Barsy, M., Lause, P., Maiter, D., Bindels, L. B., Delzenne, N. M., Schaeffer, L., Gangloff, Y. G., Chambon, C., Coudy-Gandilhon, C., Béchet, D., & Thissen, J. P. (2018). Increased Serpina3n release into circulation during glucocorticoid-mediated muscle atrophy. *Journal of Cachexia, Sarcopenia and Muscle*, *9*(5), 929–946. https://doi.org/10.1002/jcsm.12315

Tomiya, S., Tamura, Y., Kouzaki, K., Kotani, T., Wakabayashi, Y., Noda, M., & Nakazato, K. (2019). Cast immobilization of hindlimb upregulates sarcolipin expression in atrophied skeletal muscles and increases thermogenesis in C57BL/6J mice. *American Journal of Physiology-Regulatory Integrative and Comparative Physiology*, *317*(5), R649–R661. https://doi.org/10.1152/ajpregu.00118.2019

Vollmer, S., Kappler, V., Kaczor, J., Flügel, D., Rolvering, C., Kato, N., Kietzmann, T., Behrmann, I., & Haan, C. (2009). Hypoxia-inducible factor 1α is up-regulated by oncostatin M and participates in oncostatin M signaling. *Hepatology*, *50*(1), 253–260. https://doi.org/10.1002/hep.22928

Davis, P., Witczak, C. and Brault, J. (2015), AMP Deaminase 3 Accelerates Protein Degradation in C2C12 Myotubes. The FASEB Journal, 29: 825.2. https://doi.org/10.1096/fasebj.29.1_supplement.825.2