University of Alberta

Study of molecular and metabolic changes in skeletal muscle in response to cancer

by

Cynthia Stretch

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Oncology

Fall 2013
Edmonton, Alberta

Dedicated with love to:

My amazing parents:
Betty and Julio

And to

My awesome husband:
Robin

Abstract

Cancer cachexia is a multifactorial syndrome characterized by involuntary weight loss, wasting of skeletal muscle driven by reduced food intake and abnormal metabolism. Cachexia has a negative impact on quality of life, response to chemotherapy and survival. Cachexia research is undeveloped with respect to understanding molecular changes involved and its classification / diagnostic criteria (there are no clinically useful predictors and diagnostic tests). The purpose of this research was to take advantage of gene expression (transcriptomic) and metabolite (metabolomic) profiling to address these gaps.

Patients with cancer consented to provide skeletal muscle biopsy (n=134) for gene expression array or plasma and urine (n=93) for nuclear magnetic resonance spectroscopy and mass spectrometry. Omic data output was examined in relation to different dimensions of cachexia phenotype; gene expression was examined in relation to weight loss, muscle mass and muscle radiation attenuation and metabolites were examined in relation to muscle loss, muscle and fat mass, metabolic rate and food intake. Statistical analysis included standard statistical tests and machine learning methods.

Muscle gene expression varied strongly in relation to muscle attenuation, and to a much lesser degree with weight loss and muscle mass. Differential expression suggests low attenuation muscle has persistent inflammation, increased degradation, altered energy metabolism, increased extracellular matrix components and altered growth signalling. Urinary metabolites reflected muscle mass and to a lesser extent fat mass, and could be used to predict muscle mass and

rate of muscle loss with 98% and 82% accuracy, respectively. Urinary metabolites related to muscle mass and muscle loss were associated with amino acid and ATP synthesis.

Overall, transcriptomics work revealed a molecular signature for low muscle attenuation, which parallels many gene expression changes observed during aberrant muscle repair and metabolic syndrome. This explorative transcriptomic study provides multiple potentially crucial pathways that have yet to be studied in detail in cachexia. Metabolomics work revealed that urine metabolites are most reflective of muscle mass and its change. This work suggests that it may be possible to develop a metabolomics-based tool to assess skeletal muscle mass in cancer. Validation of urine metabolomics to predict muscle mass loss is warranted.

Acknowledgements

This thesis was made possible from financial support from the Department of Oncology, University of Alberta, Queen Elizabeth II Masters and Doctoral Scholarship, the Faculty of Medicine and the University of Alberta Graduate Students' Association.

I would not have been able to complete this thesis without assistance from several important people.

Thank you to my supervisor Dr. Vickie Baracos. I am honored and grateful you took me on as a doctoral student, I could not have asked for a better supervisor and mentor. I am especially grateful for our career and life discussions, your generosity and your willingness to accommodate my need to always put family first. Knowingly or unknowingly you have taught me the importance of diplomacy, collaboration, patience, composure and mentorship. However, the most important lesson, one that will remain with me for the rest of my professional and personal life, is to "keep calm at all times".

Thank you to the other members of my committee: Dr. Linda McCargar, Dr. Ron Ball and Dr. Michael Sawyer for your guidance and expertise. I would also like to thank Dr. Russel Greiner for your invaluable input and collaboration.

I would also like to acknowledge the contribution of several colleagues, without whom this research would not have been possible: Dr. Russ Greiner's current and former lab members Tom Eastman, Roman Eisner, Sheehan Khan and Nasimeh Asgarian, Dr. Oliver Bathe in Calgary and Dr. Helen Steed who collected skeletal muscle samples for transcriptomic studies, Drs. Carla Prado and Marina Mourtzakis who enrolled and collected samples for the metabolomic studies, Dr. Sambasivarao Damaraju for his invaluable guidance on microarray studies, Dr. David Wishart for guidance on metabolomic studies, Soenke Moehn for assisting me on urine metabolite validation by high performance liquid chromatography, and Abha Hoedl for assistance in the lab.

Thank you to the many former and current graduate students I have been fortunate to meet, especially members of Cachexia Central. I also have a special thanks to Dr. Janice Kapty for her guidance, friendship and our venting/coffee breaks.

Thank you to my best friend and amazing husband, Robin Stretch. I could not have done this without your support and encouragement. You were there after every PhD-related "learning experience" and success. Thank you to mom and dad who instilled in me the importance of education and hard work and to my siblings Kevin, Carla, Javier and Julian who are constantly teaching me how to have fun, relax and grow.

Finally, thank you the patients and their families who participated in these studies.

Table of Contents

CHAPTER 2: Research plan

CHAPTER 3: Sex and sample size affect skeletal muscle gene expression microarray experiments

CHAPTER 4: Skeletal muscle in cancer cachexia is characterized by features of wasting, pathological lipid infiltration, inflammation and aberrant regeneration processes

CHAPTER 5: Prediction of skeletal muscle and fat mass in patients with advanced cancer using a metabolomic approach

CHAPTER 6: Learning to predict cancer-associated skeletal muscle wasting from 1H-NMR profiles of urinary metabolites

CHAPTER 7: Final Discussion

APPENDIX

List of Tables

List of Figures

List of Abbreviations

ALST	appendicular lean soft tissue
AR	androgen receptor
ARE	androgen response element (s)
BMI	body mass index
BMP	bone morphogenetic proteins
cDNA	complementary DNA
CI	confidence interval
cm^2	centimeter squared (referes to surface area)
COPD	chronic obstructive pulmonary disease
CT	computerized tomography
d	day(s)
D2O	deuterium oxide
ddH_2O	distilled deionized water
DSS-d6	2,2-dimethyl-2-silapentane-5-sulfonate d6
DXA	dual energy x-ray absorptiometry
ECM	extracellular matrix
eIF-4E	eukaryotic translation initiation factor 4E
F	female
FDR	false discovery rate
FFA	free fatty acid
FFM	fat-free mass

FoxO	forkhead box O
g	gram
GEO	gene expression omnibus
h	hour(s)
HCl	hydrochloric acid
HPLC	high performace liquid chromatography
Hsc70	heat shock cognate 70
HU	Hounsfield unit
IFN-γ	interferon gamma
IGF-1	insulin like growth factor 1
IGF1R	insulin like growth factor 1 receptor
IL-1β	interleukin 1 beta
IL-6	interleukin 6
IPA	ingenuity pathway analysis
IκBs	inhibitors of NF- κB
kcal	kilocalorie(s)
kDa	kilo Dalton
KEGG	Kyoto encyclopedia of genes and genomes
kg	kilogram
L	liter
L3	3rd lumbar vertebrae
LASSO	least absolute shrinkage selection operator
LOOCV	leave one out cross validation

LST	lean soft tissue
M	male
MAFbx	muscle atrophy F-box
MHz	mega hertzs
MI	mutual information
mM	milli molar
MRI	magnetic imaging resonance
MS	mass spectrometry
mTOR	mammalian target of rapamycin
mTORC1	mTOR complex 1
MuRF1	muscle ring finger1
NaOH	sodium hydroxide
NCBI	National Center for Biotechnology Information
NF-κB	nuclear factor kappa B
NMR	nuclear magnetic resonance
PI3K	phosphatidylinositol-3,4,5-triphosphate kinase
PIA	pathway informed analysis
PLS-DA	partial least squares discriminant analysis
PPARα	peroxisome proliferator-activated receptor alpha
ppm	parts per million
r	correlation coeffiecient
R^2	coefficient of determination
REE	resting energy expenditure

ref	reference
RIN	RNA integrity numbers
RQ	respiratory quotient
SD	standard deviation
SMI	skeletal muscle index
SVM	support vector machine
TAN	tree-augmentented naive bayes
TCA cycle	tricarboxylic acid cycle
TFM	total fat mass
TGF	transforming growth factor
TGF- β	tumor growth factor beta
TIM	translocase of the inner membrane
TNF-α	tumore necrosis factor alpha
TOM	translocase of the outer membrane
Ub	ubiquitin
µL	micro liter
µM	micro molar
^{1}H-NMR	proton NMR

CHAPTER 1: Introduction and literature review

1.1 Purpose

The purpose of this chapter is to provide an overview of cancer associated changes in body composition, with a focus on skeletal muscle, and to describe how a better understanding of molecular and metabolic changes involved may improve patient outcome.

1.2 Introduction

Cancer is currently the leading cause of death in Canada; approximately one in four Canadians will die of cancer (1). Cancer cachexia is a syndrome that accompanies cancer and results in a poorer response to treatment (2), quality of life (3) and prognosis (4). Recently, an international panel of experts in cancer cachexia research participated in a formal consensus process and published the following consensus definition "*cancer cachexia as a multifactorial syndrome defined by an ongoing loss of skeletal muscle mass (with or without loss of fat mass) that cannot be fully reversed by conventional nutritional support and leads to progressive functional impairment. Its pathophysiology is characterised by a negative protein and energy balance driven by a variable combination of reduced food intake and abnormal metabolism*" (5). Currently in oncology clinics, cancer cachexia is rarely recognized or assessed (6) and treatment options are limited (7). This can be attributed to lack of fast and easily accessible screening tools and incomplete knowledge of molecular mechanisms involved in cancer cachexia development and progression.

1.3 Cancer cachexia

Cachexia has been recognized as a paraneoplastic syndrome for a long time (8). Much like tumor progression, cachexia develops progressively through various stages of severity: precachexia, cachexia and lastly refractory cachexia (5). Not all patients traverse through all three stages. Patients with precachexia would have early clinical and metabolic signs that may occur prior to involuntary wasting (e.g. anorexia) (5). Patients with cachexia would present with muscle depletion (defined in detail below) and be actively losing energy stores (5). Finally, patients with refractory cachexia would have a life expectancy of less than 3 months (i.e. have very advanced cancer) (5). Wasting becomes exponential during the last 3 months of life (9). Ideally, by identifying patients in early stages the syndrome can be managed prior to reaching the refractory / untreatable stage.

Severe muscle depletion, termed sarcopenia, is a result of aberrant control of muscle mass and denotes a muscle mass less than 2 standard deviations below that of typical healthy adults (of the same sex) (10). Sarcopenia is directly responsible for functional impairment (11), increased risk of fractures (12), increased length of hospital stay (13) and shorter survival (14) in non-malignant disease. Various conditions lead to sarcopenia including aging, disuse, starvation, denervation, and cancer cachexia. In malignancy, sarcopenia has been associated with shorter time to tumour progression (15) and dose-limiting toxicities from several different types of chemotherapy resulting in dose-reduction or termination of treatment (15-17). Weight loss ($\geq$ 5% loss over the past 6 months in the absence of simple starvation) and low body mass index (BMI <18.5kg/m^2 or

<20kg/m^2) are commonly used to diagnose cancer cachexia in the clinic. These criteria would easily identify a stereotypical cachectic patient who is emaciated upon examination but would ignore obese patients or those losing muscle while gaining fat or retaining water and therefore appear weight stable. Indeed, it is body composition that needs to be assessed to detect muscle depletion and ongoing loss.

Cachexia progression depends on: cancer type, stage and response to anticancer therapy, the presence of systemic inflammation, and low food intake (5). Cancer cachexia appears to be most prevalent amongst patients with solid tumors (e.g. lung and gastrointestinal cancers) (18). Further, it is well established that advanced stage, multiple sites of metastasis and progressive disease are risk factors for weight loss (9). Tumor tissue is metabolically demanding (estimated 200-300 kcal/day per kg of tumor tissue). If caloric intake is not increased to meet the increased resting metabolic rate, mobilization of fat and protein from energy stores (adipose and muscle tissue) is a likely outcome. Systemic inflammation also results in increased metabolic rate and is common in cancer cachexia (19, 20). As detailed in later sections, inflammatory cytokines also play roles in activating muscle catabolism and inhibiting muscle anabolism. Finally, reduced food intake is also common in cancer cachexia (21). Reduced food intake may be due to a decreased central drive to eat (which may be altered by inflammation) (22, 23), taste and smell changes (24) and decreased gastrointestinal tract motility (21). Because of multiple pro-wasting factors that exist in cachexia, cancer cachexia is said to be a "*multifactorial syndrome*" (5).

Most previous attempts at treating cachexia have focused on unimodal approaches which address one factor at a time. For example, progestational agents (25) and cortecosteroids (e.g. dexamethasone) (26) have been used to improve appetite and oral nutritional supplements to increase intake (27). However, none of the unimodal approaches have been shown to prevent or reverse muscle wasting, the key feature of cancer cachexia. Due to multiple factors promoting wasting, experts suggest a multimodal approach is needed to treat cancer cachexia (28). Currently, there is no standardized treatment for cancer cachexia. This may be attributed to incomplete understanding of molecular mechanisms responsible for wasting (28, 29). Though helpful, most studies aimed at understanding mechanisms have been conducted on animal models and focused on a handful of genes and proteins at a time. In fact little is known about what is happening in human tissues in response to cancer cachexia. It is necessary to validate what is found in animal models in humans, a step which is complicated by the heterogeneity of humans.

Though cancer alone can result in muscle loss, many patients have other conditions that further aggravate wasting. The typical cancer patient is elderly (1), has 3 comorbid conditions (30) (e.g. cardiovascular disease, hypertension and insulin resistance), is sedentary (31), and is overweight or obese (10). Exclusive of cancer, all these conditions are associated with skeletal muscle wasting and / or changes in muscle tissue composition associated with negative outcomes (32-36) and therefore add levels of complexity to diagnosing, understanding molecular mechanisms and treating cancer cachexia.

1.4 Skeletal muscle

As emphasized above, muscle loss is central in cancer cachexia. Skeletal muscle is the largest organ and protein store in the body. In addition to its roles in mobility, skeletal muscle plays a central role in the control of whole-body metabolism; it accounts for 80% of glucose disposal, and in the rested state skeletal muscle fatty acid oxidation contributes about 90% of the energy requirements (37). These two roles (functional and metabolic) are intimately linked and determine composition, turnover rates, contractile proteins and energy usage. Protein turnover, the balance between protein synthesis and protein degradation rates, plays a major role in maintaining adult muscle mass (38). Cellular turnover appears to play a minor role (39). Most details about the pathways involved in protein and cellular turnover in cancer are the result of work conducted on animal models.

1.4.1 Protein turnover: control of protein synthesis

Although investigations of signaling pathways controlling synthesis and degradation are ongoing, certain pathways are known to be involved (Figure 1-1 (synthesis) and Figure 1-2 (degradation)). Insulin like growth factor 1 (IGF-1) and insulin are among the most studied growth promoting factors in muscle (40). Ligand binding to IGF-1 receptors results in downstream activation of phosphatidylinositol-3,4,5-triphosphate kinase (PI3K)/Akt/mammalian target of rapamycin (mTOR) signaling (41). mTOR complex 1 (mTORC1) is currently recognized as the major pathway regulating protein synthesis in adult muscle (38). Activated mTOR results in increased protein synthesis by ultimately activating

eukaryotic translation initiation factor 4E (eIF-4E). In addition to activating protein synthesis, Akt signaling is suggested to inhibit induction of atrophy signaling by phosphorylating the atrophic Forkhead box O proteins (FoxO) family of transcription factors thereby excluding them from the nucleus. When chronically deactivated, this pathway plays a role in muscle atrophy by reducing protein synthesis (42). In cancer cachexia, often associated with insulin resistance, activity of PI3K is decreased and inhibition of FoxO and expression of components of the ubiquitin-proteosome system are increased.

Androgens also promote protein synthesis and muscle growth. Androgens diffuse to skeletal muscle and bind to androgen receptors (AR) (43). Ligand bound ARs then homodimerize and bind to androgen response elements on the genome to affect rates of transcription of various genes (43, 44). Anabolic steroids also increase AR mRNA and IGF-1 production (45-47) and may reduce systemic pro-inflammatory cytokines tumor necrosis factor alpha (TNF-α), interleukin 6 (IL-6) and interleukin 1 beta (IL-1β) (48, 49). Post menopause testosterone depletion can lead to decreases in bone mass, bone marrow activity and muscle strength in both men and women (50, 51). Evidence suggests that patients with advanced cancer often have low androgen levels and this is associated with cachexia (52).

Protein synthesis is also negatively regulated by myostatin, a member of the tumor growth factor β (TGF- β) family which is expressed and secreted mainly in skeletal muscle (53). Myostatin influences MyoD expression, a major growth transcription factor in muscle, via inhibition of Smad2/3 transcription

factors (54). Furthermore, myostatin antagonizes protein synthesis by regulation of the Akt-FoxO pathway (55). In tumor bearing animals, myostatin expression and bioactivity are upregulated (56) and levels of follistatin, a myostatin inhibitor, are reduced (57). Notably, IGF-1 dominantly blocks effects of myostatin *in vitro* (58) providing another pathway by which insulin resistance may be blunting protein synthesis.

1.4.2 Protein turnover: control of protein breakdown

The following pathways are available for protein breakdown in muscle: ubiquitin-proteasome, lysosomal, calcium dependent and caspase dependent pathways. The ubiquitin (Ub)-proteasome pathway is the main mechanism of muscle protein breakdown in cancer cachexia (59, 60). This system consists of concerted actions of enzymes that link chains of polypeptide co-factor, Ub, onto proteins to mark them for degradation. A multicatalytic protease complex, the 26S proteosome, degrades ubiquitinated proteins into peptides. Three enzymatic components are needed to ubiquitinate proteins: E1 (Ub-activating ezyme), E2 (Ub-carrier or conjugating proteins), and E3 (Ub-protein ligase). E3 is the key enzyme in this group and varies among tissues and physiologic states. E3 recognizes a specific protein substrate and catalyzes the transfer of activated Ub to it. Muscle Atrophy F-box (MAFbx) and Muscle Ring Finger1 (MuRF1), are specific constituents of muscle and their expression increases dramatically (eight – to- 20 fold) in cancer cachexia (60, 61). Knockdown of both MAFbx and MuRF-1 in tumor-bearing animals results in decreased muscle mass loss (62). Inflammation appears to an important trigger of the Ub-proteosome system.

Evidence from animal models suggests that inhibiting pro-inflammatory cytokines (e.g. IL-6 and TNF-α) results in downregulation of the Ub-proteasome system and reduced muscle protein loss (63).

Although skeletal muscle contains few lysosomes, the lysosomal system contributes to muscle protein breakdown via macroautophagy (64). Autophagy is required for a variety of homeostatic functions such as clearance of protein aggregates, turnover of ribosomes, mitochondria and endoplasmic reticulum and plays a role in programmed cell death (apoptosis) (65). Lysosomal processes are stimulated during cancer cachexia (59, 64, 66) partly due to pro-inflammatory cytokines TNF-α and IL-6 (64, 67). Furthermore, suppression of the IGF-1/AKT/FoxO pathway activates autophagy (68).

Calpains are non-lysosomal, calcium-dependent cysteine proteases inactive under basal conditions. Activated calpains cleave myofibrillar cytoskeletal proteins resulting in disruption of the sarcomere and release of myofilaments that are subsequently ubiquitinated and degraded by the 26S proteosome (69). Activated calpains also inhibit Akt activity, which in turn results in activation of FoxO transcription factors (70). Like the Ub-proteasome system, calpains are influenced by inflammatory factors; calpain activity is reduced when activity of pro-inflammatory cytokine TNF-α is blocked (63).

Caspases are proteins involved in apoptosis and possibly breakdown via non-apoptotic pathways. It remains unclear what role caspases play in cancer associated wasting. It is possible that a limited form of apoptosis of nuclei without cell death may be activated in effort to preserve size of myonuclear domains

within atrophied myofibrils (65). It may also be possible that apoptosis is a driving pathological process in muscle atrophy (65).

As suggested above, pro-inflammatory cytokines can directly activate catabolism via different proteolytic pathways in muscle (71). Binding of TNF-α, IL1- β and IL-6 to surface receptors can act via various downstream pathways which ultimately activate nuclear factor κB (NF-κB), by inactivation/degradation of inhibitors of NF- κB (IκBs). Active NF-κB translocates to the nucleus and alters expression of MyoD (suppressed), Ub-proteosome pathway proteins, tissue-degrading enzymes such as metalloproteinases (72). Inhibition of NF- κB prevented muscle loss in tumor bearing animals (73).

1.4.3 Cellular turnover

It is well established that cellular turnover plays a major role during muscle development in embryo. During development, mononucleated muscle progenitors fuse to form nondividing multinucleated myofibers. Some progenitors remain associated to adult myofibers as satellite cells (i.e. skeletal muscle stem cells) located beneath the basal lamina of myofibers (74). In adult muscle, satellite cell nuclei account for 3-6% of all nuclei contained within the basal lamina. These remain mitotically quiescent and activate for regeneration (after injury), hypertrophy or atrophy (75-78). However, what role satellite cells play in wasting conditions remains controversial and requires further study (79).

1.5 Adipose tissue

Lipid metabolism and mobilization are altered in cancer cachexia (80). Adipose tissue loss is common in cancer patients (81, 82). Similar to skeletal

muscle, the prevalence and severity of adipose tissue loss increases approaching death (81). Free fatty acids (FFAs), derived mainly from triglyceride stored in adipose tissue are the main energy source for muscle at rest. In addition to its fuel storage function adipose tissue is an active endocrine organ. Adipose tissue derived pro-inflammatory and anti-inflammatory molecules (adipokines) act systemically participating in physiologic and pathologic processes. The role of adipose tissue in cross-talk with skeletal muscle during cachexia is unknown.

1.6 Ongoing study of molecular mechanisms in wasting

As mentioned there are gaps in our understanding of molecular mechanisms involved in cancer cachexia in human muscle. Until a decade and half ago, work was done in a "top-down" manner: novel proteins were isolated, followed by extensive biochemical purification, and protein and DNA sequencing. Therefore, insights occurred one gene at a time. Development of gene expression profiling using high-throughput microarray-based methods allow for concurrent analysis of the expression level for thousands of genes in a sample. Microarray technology is a potentially crucial tool for understanding mechanisms involved in regulating the pathological changes in muscle.

1.6.1 Gene expression microarray

Array technology evolved from Southern blotting, where fragmented DNA is attached to a substrate and then probed with a known DNA sequence. At the very basic level, array technology follows the central dogma of biology: a gene (DNA) encodes an mRNA, which in turn encodes a protein. Thus, altered expression of mRNA would translate to changes in amounts of protein present and

available for biological activity (unless altered by post translational modifications). In the mid 1990s, as the database of complementary DNA (cDNA) sequences was expanding, efforts to develop methods for determining expression of many sequences simultaneously ultimately resulted in the first microarray (83).

Microarray experiments typically have the following objectives: to identify differentially expressed oligonucleotide sequences (features which are segments of gene transcripts) or to identify molecular markers that can be used as tools for disease diagnosis and prognosis or as predictors of clinical outcomes. Most microarray studies focus on the former objective in efforts to identify genes / pathways involved in a particular phenotype and to discover potential molecular targets for treatment development.

Soon after this technology was introduced researchers began applying it to muscle tissues in wasting conditions (59, 84-86). The study by Lecker *et al.* was the first to examine muscle gene expression from preclinical models of cancer cachexia (59). These early studies had limitations: at the time there were no rat microarray chips (cancer cachexia models used were rat models) and microarray chips contained fewer sequences (16K compared to 40-60K in today's chips). Nearly a decade passed until microarray analysis of muscle in cancer was revisited. The study by Stephens *et al.* was the first to examine global gene expression in human skeletal muscle from patients losing and not losing weight (87). This group suggested that available preclinical models do not accurately reflect the molecular characteristics of human muscle from cancer cachexia

patients (87). They found no significant differences in expression of the ubiquitin ligases MURF1 and MAFbx; however weight loss may not be reflective of muscle catabolism *per se*.

1.6.2 Gene expression microarray: methodological considerations

Despite almost immediate application after microarray introduction, methodological issues related to microarray chips, sample preparation, data normalization and data analysis are still being worked out. Consequently, results from different studies are inconsistent and irreproducible. As articulated by Abdullah-Sayani and colleagues "*the quantity of articles that document the discovery of new gene profiles is as plentiful as the number of publications that scrutinize their interpretation*" (88). None of the steps in the microarray experimental process (chip production, probe hybridization, image quantification, normalization and data interpretation) have an agreed upon, standardized protocol.

Modern microarrays are commercially produced by several companies (e.g. Agilent and Affimetrix), each using different manufacturing techniques, labeling methods, hybridization protocols, probe lengths, and probe sequences (89). All of these factors may affect microarray performance (89). Since platforms do not share the same set of probe sequences, data obtained from different platforms cannot be directly compared. Lack of intra-platform standardization protocols is particularly problematic when trying to re-use data made available in public databases such as Gene Expression Omnibus (GEO) (90). While various groups have presented cross-platform normalization methods (89), this is still an active area of research.

The value of microarray is the ability to measure gene expression for thousands of genes simultaneously; however, it is this highly dimensional output that presents the biggest obstacles. As with all experiments, it is important to start with a good study design: adequate sample sizes, matched experimental variables of cases and controls, biologically homogenous sample populations, and samples handled uniformly through the course of the entire experiment. Due to large amounts of data produced, typical sample size calculations cannot be used (91). Further, little is known about the consequences of using different sample sizes on statistical analysis of microarray data (92). Previously conducted microarray studies of human muscle during cancer-associated atrophy had sample sizes ranging from n=18 to n=21 (87, 93). It remains unclear what an appropriate sample size is, particularly for a free living human population.

Development of new methods to analyze microarray data is an active area of research in the fields of bioinformatics, biostatistics and computing science. To identify differentially expressed genes, the most common goal of microarray studies, fold-change and t-test results are most commonly reported (94). These methods are not without criticism: arbitrary cutoffs are used to determine what is considered significantly differentially expressed, low gene expression may result in large fold changes despite having no biological significance, genes with low variability may be missed (when calculating fold-change) and results are rarely tested for robustness or validated. To make predictive models, expression data is used to predict what class (e.g. disease versus healthy) patients belong to. This is called a classifier. This classifier can be built using different algorithms such as

support vector machines (SVM) and least absolute shrinkage selection operator (LASSO). SVM and LASSO are machine learning algorithms; machine learning, a branch of artificial intelligence in the field of computing science, plays an important role in the development of microarray data analysis methods.

1.7 Measuring body composition

Proper assessment of body composition and body composition changes are essential for diagnosing, staging and ultimately managing cancer cachexia. There are essentially three types of methods that can be used to measure or estimate the amount of fat and /or muscle in the body: anthropometrics, image-based methods and metabolically-based methods. As previously mentioned, anthropometrics such as weight, BMI and skinfold measures are convenient, non-invasive and inexpensive but inadequate because they do not differentiate between different components (muscle and fat).

1.7.1 Measuring body composition: image-based methods

Dual energy X-ray absorptiometry (DXA) (95), computerized tomography (CT) imaging analysis (2) and magnetic imaging resonance (MRI) can differentiate between lean and fat mass. These are considered gold-standard methodologies for body composition assessment due to low precision error of 2% (96), but they do have limitations. DXA cannot distinguish between different lean tissues (muscle, organs, skin and tumor) or different fat tissue depots (intramuscular, visceral and subcutaneous). CT and MRI imaging can discriminate muscle, adipose tissue depots, bone, organs and tumor (*87)*. However, both MRI and CT imaging instruments are expensive to purchase and to

use, and CT exposes participants to radiation. Further, these methods are less useful for assessment of body composition change than of tissue mass *per se*. Tissue mass change may be subtle on a daily basis but additive over time. For example, a loss of body fat of 2% in one month, is in fact a catastrophic rate of loss, because if sustained it would add up to a 24% loss of whole body fat in one year. A loss of 2% in a given month is however undetectable within the precision error of even the most precise techniques available (CT, MRI or DXA) (9).

In the oncology setting, diagnostic imaging using CT and MRI is part clinical care; these are used for diagnostic and follow up purposes to assess tumor progression and response to therapy. Body composition researchers have taken advantage of routine imaging to estimate body composition from as few as one cross-sectional image (typically at the level of the 3rd lumbar vertebrae (L3) (9, 17, 97). However, despite availability of CT and MRI images in the clinic, image-based body composition assessment has not transitioned into clinical practice or been incorporated into clinical trials. In fact, body composition is rarely, if ever, assessed in the oncology clinic.

1.7.2 Measuring body composition: metabolically-based methods

Most body composition efforts have focused on imaging techniques. However, easily accessible biofluids, such as urine and plasma, have also been used to predict certain elements of body composition. This method uses knowledge of biochemical pathways and relatively simple analytical techniques to measure metabolites indicative of body components. Use of biofluids is particularly appealing because it is relatively non-invasive.

In 1919, Burger suggested that total daily urinary creatinine excretion was proportional to body muscle mass (98). Since then, close correlations between creatinine and muscle mass have been confirmed and validated, r^2 of 0.72 to 0.90 in healthy individuals (99-101). Metabolites resulting from tissue catabolism are also likely to be sensitive indicators of the initiation of a catabolic state. Tissues normally undergo turnover with a low daily rate of synthesis and breakdown. During catabolic conditions, when tissue mobilization occurs, characteristic metabolites of lipid and muscle protein catabolism are produced and represented in various biological pools, such as plasma and urine. For example, normal adults fasted for 24 h, demonstrate large changes in plasma concentrations of β-hydroxybutyrate (2 fold-increase), acetetoacetate, acetone, free fatty acids and glycerol (102). Similarly, muscle catabolism generates free amino acids, including also a post –translationally methylated histidine (3 – methylhistidine) that is unique to actomyosin, many secondary metabolites of these compounds, urea and creatinine (103-105). Despite documented relationships between metabolite concentrations in easily accessible biofluids and body composition, this has not been exploited in studying or assessing cancer cachexia.

1.7.2.1 Metabolic profiling

The idea that metabolite changes in biological fluids are indicative of changes in metabolism has been around for centuries (106). Advancements in instrumentation and quantification methods now allow researchers to measure hundreds of metabolites simultaneously. These advances include: increased sensitivity of nuclear magnetic resonance (NMR) spectroscopy and mass

spectrometry (MS) and development of multivariate data analysis methods (106). With these technologies we are now able to glimpse into the metabolome (the complete set of small-molecule metabolites in a biological sample) to study changes in tissues and fluids. Metabolite profiling, termed metabolomics, may provide a valuable tool for studying body composition and its change.

Much like microarray analysis, metabolomic analysis was applied soon after its introduction despite unresolved issues. First studies focused on toxicology (107) but quickly expanded to other fields such as diabetes (108), exercise (109) and tumor development (110). Like microarray analysis, there are no standardized methods for metabolomic studies. Studies looking at the same disease / problem may use different: biofluids (urine, plasma, serum, saliva, bile, cerebral spinal fluids, etc.), platforms (NMR or MS based), and statistical analysis methodology. Prior to using metabolomics as a tool for body composition assessment it is necessary to identify the best platform, biofluid and statistical method for the task.

Based on knowledge of metabolism and prior experiments looking at single metabolites (creatinine, 3-methylhistidine, free fatty acids, and glycerol) urine and plasma are good candidates for studying body composition and its change. Though it is currently impossible to measure all metabolites using a single platform, NMR and MS are the most often used and accessible. MS is attractive due to its high sensitivity; however sample preparation and analysis will destroy samples. Kits to analyze samples for metabolomics studies using MS are commercially available (currently only from Biocrates by Life Sciences (http://www.biocrates.com/)). Most metabolites identified with these kits are lipid

molecules and may therefore be more suited to study plasma. NMR is not as sensitive (concentrations > 1 μM) but requires minimal sample preparation, is nondestructive and is fast (about 5 minutes per sample). Proton NMR is particularly appealing for urine analysis because it can be used to quantify creatine and creatinine, many amino acids, amino acid derivatives, tricarboxylic acid cycle (TCA cycle) intermediates, urea cycle intermediates and sugars (i.e. metabolites that may be produced by muscle).

Due to large amounts of data generated by metabolomic analysis, typical statistical methods cannot be used. Researchers have integrated multivariate statistical methods originally used in engineering and economics, as well as machine learning algorithms to analyze metabolomic data. What analysis is conducted depends upon what the goal of the metabolomics experiment is. The goal may be to determine what metabolites are correlated with the outcome (e.g. what metabolites are related to losing or not losing muscle); in which case the output would provide valuable biological insights relating to what pathways may be at play. The goal may be to build a predictor (e.g. predict if future patients are losing or not losing muscle) which could act as a screening or diagnostic tool that could be used in the clinic.

To add another level of complexity to metabolomics research, it is important to note that biofluids are very complex. The identity and concentration of metabolites in such fluids is dictated by exogenous and endogenous sources. Exogenous sources, such as nutrient intake, can vary greatly in a population of free-living individuals (111). Endogenous sources, such as resting energy

expenditure and rates of protein turnover can also vary, particularly in the cancer population (5, 9). It remains to be determined what potential sources of variation in are present in different biological pools.

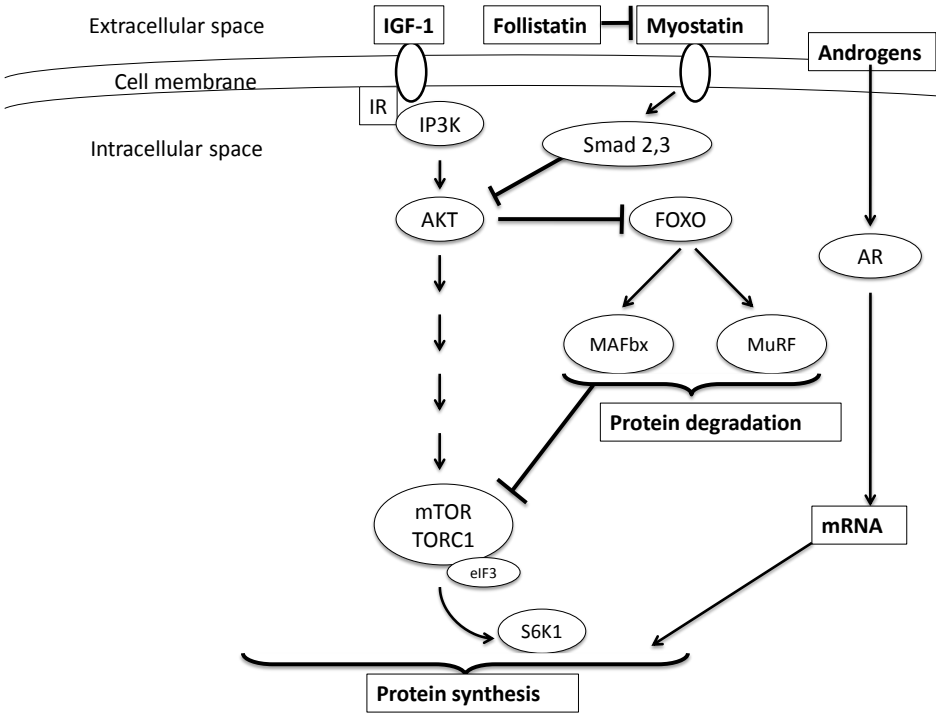
1.8 Summary

From this review it is evident that cancer cachexia is a complex and multifactorial syndrome. At the very basic level there is a need to study molecular mechanisms involved in cachexia development and progression in humans. Gene expression microarray is a potentially useful tool to accomplish this. However, it is clear that there are methodological issues that need to be addressed before this technology can be used properly.

Research also shows that despite the prevalence and negative outcomes associated with cachexia it is seldom assessed and diagnosed. This is partly due to the lack of body composition assessment tools in the clinical setting. It may be possible to use knowledge of metabolism and metabolic profiling technology to develop a useful screening tool that could be used in the clinical setting. Like microarray technology, before metabolomics can be used for this purpose it is necessary to address methodological issues.

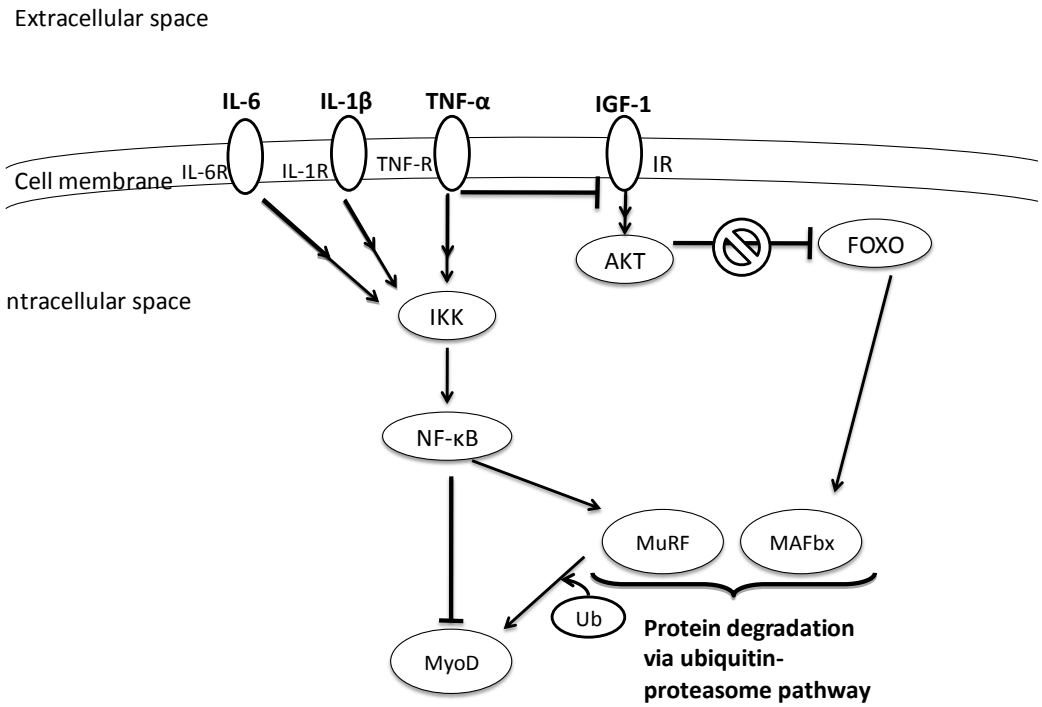
Figures

Figure 1-1. Simplified pathways involved in protein synthesis



Growth factors (GF), insulin like growth factor 1 (IGF-1), myostatin and androgens all ultimately regulate protein synthesis. Mammalian target of rapamycin (mTOR)/target of rapamycin complex 1 (TORC1) signaling is a major regulation point for protein synthesis. IGF-1 also inhibits upregulation of proteins involved in protein breakdown (MAFbx and MuRF). Conversely, upregulation of MAFbx and MuRF inhibit protein synthesis via mTOR/TORC1.

Figure 1-2. Simplified pathways involved in protein degradation



Inflammatory cytokines interleukin 1β (IL-1β), interleukin 6 (IL-6) and tumor necrosis factor (TNF-α) all ultimately regulate protein degradation. Nuclear factor κB (NF-κB) signaling is a major regulation point for protein degradation leading to upregulation of ubiquitin-proteasome pathway proteins MuRF and MAFbx. MuRF and MAFbx are involved in ubiquitination of various proteins including MyoD which is involved in regeneration and protein synthesis. TNF-α also inhibits protein synthesis by inhibiting AKT activation.

References

1. Canadian Cancer Society's Steering Committee on Cancer Statistics. Canadian Cancer Statistics 2012. Toronto, ON: Canadian Cancer Society 2012.

2. Prado CM, Antoun S, Sawyer MB, Baracos VE. Two faces of drug therapy in cancer: drug-related lean tissue loss and its adverse consequences to survival and toxicity. Curr Opin Clin Nutr Metab Care. 2011.

3. Tan BH, Fearon KC. Cachexia: prevalence and impact in medicine. Current opinion in clinical nutrition and metabolic care. 2008;11:400-7.

4. Bachmann J, Heiligensetzer M, Krakowski-Roosen H, Buchler MW, Friess H, Martignoni ME. Cachexia worsens prognosis in patients with resectable pancreatic cancer. Journal of gastrointestinal surgery : official journal of the Society for Surgery of the Alimentary Tract. 2008;12:1193-201.

5. Fearon K, Strasser F, Anker SD, Bosaeus I, Bruera E, Fainsinger RL, et al. Definition and classification of cancer cachexia: an international consensus. The lancet oncology. 2011;12:489-95.

6. Churm D, Andrew IM, Holden K, Hildreth AJ, Hawkins C. A questionnaire study of the approach to the anorexia-cachexia syndrome in patients with cancer by staff in a district general hospital. Supportive care in cancer : official journal of the Multinational Association of Supportive Care in Cancer. 2009;17:503-7.

7. Yavuzsen T, Davis MP, Walsh D, LeGrand S, Lagman R. Systematic review of the treatment of cancer-associated anorexia and weight loss. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2005;23:8500-11.

8. Doehner W, Anker SD. Cardiac cachexia in early literature: a review of research prior to Medline. International journal of cardiology. 2002;85:7-14.

9. Lieffers JR, Mourtzakis M, Hall KD, McCargar LJ, Prado CM, Baracos VE. A viscerally driven cachexia syndrome in patients with advanced colorectal cancer: contributions of organ and tumor mass to whole-body energy demands. The American journal of clinical nutrition. 2009;89:1173-9.

10. Prado CM, Lieffers JR, McCargar LJ, Reiman T, Sawyer MB, Martin L, et al. Prevalence and clinical implications of sarcopenic obesity in patients with solid tumours of the respiratory and gastrointestinal tracts: a population-based study. The lancet oncology. 2008;9:629-35.

11. Janssen I, Heymsfield SB, Ross R. Low relative skeletal muscle mass (sarcopenia) in older persons is associated with functional impairment and physical disability. J Am Geriatr Soc. 2002;50:889-96.

12. Baumgartner RN, Koehler KM, Gallagher D, Romero L, Heymsfield SB, Ross RR, et al. Epidemiology of sarcopenia among the elderly in New Mexico. Am J Epidemiol. 1998;147:755-63.

13. Pichard C, Kyle UG, Morabia A, Perrier A, Vermeulen B, Unger P. Nutritional assessment: lean body mass depletion at hospital admission is associated with an increased length of stay. Am J Clin Nutr. 2004;79:613-8.

14. Metter EJ, Talbot LA, Schrager M, Conwit R. Skeletal muscle strength as a predictor of all-cause mortality in healthy men. J Gerontol A Biol Sci Med Sci. 2002;57:B359-65.

15. Prado CM, Baracos VE, McCargar LJ, Reiman T, Mourtzakis M, Tonkin K, et al. Sarcopenia as a determinant of chemotherapy toxicity and time to tumor progression in metastatic breast cancer patients receiving capecitabine treatment. Clinical cancer research : an official journal of the American Association for Cancer Research. 2009;15:2920-6.

16. Prado CM, Baracos VE, McCargar LJ, Mourtzakis M, Mulder KE, Reiman T, et al. Body composition as an independent determinant of 5-fluorouracil-based chemotherapy toxicity. Clinical cancer research : an official journal of the American Association for Cancer Research. 2007;13:3264-8.

17. Antoun S, Baracos VE, Birdsell L, Escudier B, Sawyer MB. Low body mass index and sarcopenia associated with dose-limiting toxicity of sorafenib in patients with renal cell carcinoma. Ann Oncol. 2010;21:1594-8.

18. Dewys WD, Begg C, Lavin PT, Band PR, Bennett JM, Bertino JR, et al. Prognostic effect of weight loss prior to chemotherapy in cancer patients. Eastern Cooperative Oncology Group. The American journal of medicine. 1980;69:491-7.

19. Falconer JS, Fearon KC, Plester CE, Ross JA, Carter DC. Cytokines, the acute-phase response, and resting energy expenditure in cachectic patients with pancreatic cancer. Ann Surg. 1994;219:325-31.

20. Staal-van den Brekel AJ, Schols AM, ten Velde GP, Buurman WA, Wouters EF. Analysis of the energy balance in lung cancer patients. Cancer Res. 1994;54:6430-3.

21. Yavuzsen T, Walsh D, Davis MP, Kirkova J, Jin T, LeGrand S, et al. Components of the anorexia-cachexia syndrome: gastrointestinal symptom correlates of cancer anorexia. Supportive care in cancer : official journal of the Multinational Association of Supportive Care in Cancer. 2009;17:1531-41.

22. Braun TP, Zhu X, Szumowski M, Scott GD, Grossberg AJ, Levasseur PR, et al. Central nervous system inflammation induces muscle atrophy via activation of the hypothalamic-pituitary-adrenal axis. The Journal of experimental medicine. 2011;208:2449-63.

23. Schwartz MW, Woods SC, Porte D, Jr., Seeley RJ, Baskin DG. Central nervous system control of food intake. Nature. 2000;404:661-71.

24. Hutton JL, Baracos VE, Wismer WV. Chemosensory dysfunction is a primary factor in the evolution of declining nutritional status and quality of life in patients with advanced cancer. Journal of pain and symptom management. 2007;33:156-65.

25. Berenstein EG, Ortiz Z. Megestrol acetate for the treatment of anorexia-cachexia syndrome. Cochrane Database Syst Rev. 2005:CD004310.

26. Loprinzi CL, Kugler JW, Sloan JA, Mailliard JA, Krook JE, Wilwerding MB, et al. Randomized comparison of megestrol acetate versus dexamethasone versus fluoxymesterone for the treatment of cancer anorexia/cachexia. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 1999;17:3299-306.

27. Ovesen L, Allingstrup L, Hannibal J, Mortensen EL, Hansen OP. Effect of dietary counseling on food intake, body weight, response rate, survival, and quality of life in cancer patients undergoing chemotherapy: a prospective, randomized study. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 1993;11:2043-9.

28. Fearon KC. Cancer cachexia: developing multimodal therapy for a multidimensional problem. Eur J Cancer. 2008;44:1124-32.

29. Madeddu C, Mantovani G. An update on promising agents for the treatment of cancer cachexia. Current opinion in supportive and palliative care. 2009;3:258-62.

30. Extermann M, Overcash J, Lyman GH, Parr J, Balducci L. Comorbidity and functional status are independent in older cancer patients. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 1998;16:1582-7.

31. Healthy aging: physical activity and older adults. Public Health Agency of Canada; 2002.

32. Freiberger E, Sieber C, Pfeifer K. Physical activity, exercise, and sarcopenia - future challenges. Wien Med Wochenschr. 2011;161:416-25.

33. Ryan AS, Buscemi A, Forrester L, Hafer-Macko CE, Ivey FM. Atrophy and intramuscular fat in specific muscles of the thigh: associated weakness and hyperinsulinemia in stroke survivors. Neurorehabilitation and neural repair. 2011;25:865-72.

34. Ikezoe T, Mori N, Nakamura M, Ichihashi N. Effects of age and inactivity due to prolonged bed rest on atrophy of trunk muscles. European journal of applied physiology. 2012;112:43-8.

35. Johannsen DL, Conley KE, Bajpeyi S, Punyanitya M, Gallagher D, Zhang Z, et al. Ectopic lipid accumulation and reduced glucose tolerance in elderly adults are accompanied by altered skeletal muscle mitochondrial activity. The Journal of clinical endocrinology and metabolism. 2012;97:242-50.

36. Goodpaster BH, Theriault R, Watkins SC, Kelley DE. Intramuscular lipid content is increased in obesity and decreased by weight loss. Metabolism: clinical and experimental. 2000;49:467-72.

37. Kelley DE, Mokan M, Simoneau JA, Mandarino LJ. Interaction between glucose and free fatty acid metabolism in human skeletal muscle. The Journal of clinical investigation. 1993;92:91-8.

38. McCarthy JJ, Esser KA. Anabolic and catabolic pathways regulating skeletal muscle mass. Current opinion in clinical nutrition and metabolic care. 2010;13:230-5.

39. Sandri M. Signaling in muscle atrophy and hypertrophy. Physiology (Bethesda). 2008;23:160-70.

40. Musaro A, McCullagh K, Paul A, Houghton L, Dobrowolny G, Molinaro M, et al. Localized Igf-1 transgene expression sustains hypertrophy and regeneration in senescent skeletal muscle. Nature genetics. 2001;27:195-200.

41. Rommel C, Bodine SC, Clarke BA, Rossman R, Nunez L, Stitt TN, et al. Mediation of IGF-1-induced skeletal myotube hypertrophy by PI(3)K/Akt/mTOR and PI(3)K/Akt/GSK3 pathways. Nature cell biology. 2001;3:1009-13.

42. Bodine SC, Stitt TN, Gonzalez M, Kline WO, Stover GL, Bauerlein R, et al. Akt/mTOR pathway is a crucial regulator of skeletal muscle hypertrophy and can prevent muscle atrophy in vivo. Nat Cell Biol. 2001;3:1014-9.

43. Li J, Al-Azzawi F. Mechanism of androgen receptor action. Maturitas. 2009;63:142-8.

44. Wyce A, Bai Y, Nagpal S, Thompson CC. Research Resource: The androgen receptor modulates expression of genes with critical roles in muscle development and function. Mol Endocrinol. 2010;24:1665-74.

45. Montano M, Flanagan JN, Jiang L, Sebastiani P, Rarick M, LeBrasseur NK, et al. Transcriptional profiling of testosterone-regulated genes in the skeletal muscle of human immunodeficiency virus-infected men experiencing weight loss. J Clin Endocrinol Metab. 2007;92:2793-802.

46. Urban RJ. Effects of testosterone and growth hormone on muscle function. J Lab Clin Med. 1999;134:7-10.

47. Sheffield-Moore M, Urban RJ, Wolf SE, Jiang J, Catlin DH, Herndon DN, et al. Short-term oxandrolone administration stimulates net muscle protein synthesis in young men. J Clin Endocrinol Metab. 1999;84:2705-11.

48. Khosla S, Atkinson EJ, Dunstan CR, O'Fallon WM. Effect of estrogen versus testosterone on circulating osteoprotegerin and other cytokine levels in normal elderly men. J Clin Endocrinol Metab. 2002;87:1550-4.

49. Malkin CJ, Pugh PJ, Jones RD, Kapoor D, Channer KS, Jones TH. The effect of testosterone replacement on endogenous inflammatory cytokines and lipid profiles in hypogonadal men. J Clin Endocrinol Metab. 2004;89:3313-8.

50. Zitzmann M, Nieschlag E. Hormone substitution in male hypogonadism. Molecular and cellular endocrinology. 2000;161:73-88.

51. Lobo RA. Androgens in postmenopausal women: production, possible role, and replacement options. Obstetrical & gynecological survey. 2001;56:361-76.

52. Del Fabbro E, Hui D, Nooruddin ZI, Dalal S, Dev R, Freer G, et al. Associations among hypogonadism, C-reactive protein, symptom burden, and survival in male cancer patients with cachexia: a preliminary report. Journal of pain and symptom management. 2010;39:1016-24.

53. Jespersen J, Kjaer M, Schjerling P. The possible role of myostatin in skeletal muscle atrophy and cachexia. Scand J Med Sci Sports. 2006;16:74-82.

54. Langley B, Thomas M, Bishop A, Sharma M, Gilmour S, Kambadur R. Myostatin inhibits myoblast differentiation by down-regulating MyoD expression. J Biol Chem. 2002;277:49831-40.

55. McFarlane C, Plummer E, Thomas M, Hennebry A, Ashby M, Ling N, et al. Myostatin induces cachexia by activating the ubiquitin proteolytic system through an NF-kappaB-independent, FoxO1-dependent mechanism. J Cell Physiol. 2006;209:501-14.

56. Gilson H, Schakman O, Kalista S, Lause P, Tsuchida K, Thissen JP. Follistatin induces muscle hypertrophy through satellite cell proliferation and inhibition of both myostatin and activin. Am J Physiol Endocrinol Metab. 2009;297:E157-64.

57. Costelli P, Muscaritoli M, Bonetto A, Penna F, Reffo P, Bossola M, et al. Muscle myostatin signalling is enhanced in experimental cancer cachexia. Eur J Clin Invest. 2008;38:531-8.

58. Trendelenburg AU, Meyer A, Rohner D, Boyle J, Hatakeyama S, Glass DJ. Myostatin reduces Akt/TORC1/p70S6K signaling, inhibiting myoblast differentiation and myotube size. American journal of physiology Cell physiology. 2009;296:C1258-70.

59. Lecker SH, Jagoe RT, Gilbert A, Gomes M, Baracos V, Bailey J, et al. Multiple types of skeletal muscle atrophy involve a common program of changes in gene expression. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2004;18:39-51.

60. Lecker SH, Goldberg AL, Mitch WE. Protein degradation by the ubiquitin-proteasome pathway in normal and disease states. J Am Soc Nephrol. 2006;17:1807-19.

61. Latres E, Amini AR, Amini AA, Griffiths J, Martin FJ, Wei Y, et al. Insulin-like growth factor-1 (IGF-1) inversely regulates atrophy-induced genes via the phosphatidylinositol 3-kinase/Akt/mammalian target of rapamycin (PI3K/Akt/mTOR) pathway. The Journal of biological chemistry. 2005;280:2737-44.

62. Bodine SC, Latres E, Baumhueter S, Lai VK, Nunez L, Clarke BA, et al. Identification of ubiquitin ligases required for skeletal muscle atrophy. Science. 2001;294:1704-8.

63. Costelli P, Bossola M, Muscaritoli M, Grieco G, Bonelli G, Bellantone R, et al. Anticytokine treatment prevents the increase in the activity of ATP-ubiquitin- and Ca(2+)-dependent proteolytic systems in the muscle of tumour-bearing rats. Cytokine. 2002;19:1-5.

64. Deval C, Mordier S, Obled C, Bechet D, Combaret L, Attaix D, et al. Identification of cathepsin L as a differentially expressed message associated with skeletal muscle wasting. Biochem J. 2001;360:143-50.

65. Wing SS, Lecker SH, Jagoe RT. Proteolysis in illness-associated skeletal muscle atrophy: from pathways to networks. Critical reviews in clinical laboratory sciences. 2011;48:49-70.

66. Baracos VE, DeVivo C, Hoyle DH, Goldberg AL. Activation of the ATP-ubiquitin-proteasome pathway in skeletal muscle of cachectic rats bearing a hepatoma. Am J Physiol. 1995;268:E996-1006.

67. Tsujinaka T, Fujita J, Ebisui C, Yano M, Kominami E, Suzuki K, et al. Interleukin 6 receptor antibody inhibits muscle atrophy and modulates proteolytic systems in interleukin 6 transgenic mice. J Clin Invest. 1996;97:244-9.

68. Mammucari C, Milan G, Romanello V, Masiero E, Rudolf R, Del Piccolo P, et al. FoxO3 controls autophagy in skeletal muscle in vivo. Cell metabolism. 2007;6:458-71.

69. Goll DE, Neti G, Mares SW, Thompson VF. Myofibrillar protein turnover: the proteasome and the calpains. J Anim Sci. 2008;86:E19-35.

70. Smith IJ, Lecker SH, Hasselgren PO. Calpain activity and muscle wasting in sepsis. Am J Physiol Endocrinol Metab. 2008;295:E762-71.

71. Argiles JM, Busquets S, Toledo M, Lopez-Soriano FJ. The role of cytokines in cancer cachexia. Curr Opin Support Palliat Care. 2009;3:263-8.

72. Li H, Malhotra S, Kumar A. Nuclear factor-kappa B signaling in skeletal muscle atrophy. J Mol Med. 2008;86:1113-26.

73. Wyke SM, Russell ST, Tisdale MJ. Induction of proteasome expression in skeletal muscle is attenuated by inhibitors of NF-kappaB activation. Br J Cancer. 2004;91:1742-50.

74. Mauro A. Satellite cell of skeletal muscle fibers. The Journal of biophysical and biochemical cytology. 1961;9:493-5.

75. Schiaffino S, Bormioli SP, Aloisi M. The fate of newly formed satellite cells during compensatory muscle hypertrophy. Virchows Archiv B: Cell pathology. 1976;21:113-8.

76. Bruusgaard JC, Johansen IB, Egner IM, Rana ZA, Gundersen K. Myonuclei acquired by overload exercise precede hypertrophy and are not lost on detraining. Proceedings of the National Academy of Sciences of the United States of America. 2010;107:15111-6.

77. Kadi F, Charifi N, Denis C, Lexell J, Andersen JL, Schjerling P, et al. The behaviour of satellite cells in response to exercise: what have we learned from human studies? Pflugers Archiv : European journal of physiology. 2005;451:319-27.

78. Jejurikar SS, Kuzon WM, Jr. Satellite cell depletion in degenerative skeletal muscle. Apoptosis : an international journal on programmed cell death. 2003;8:573-8.

79. Pallafacchina G, Blaauw B, Schiaffino S. Role of satellite cells in muscle growth and maintenance of muscle mass. Nutrition, metabolism, and cardiovascular diseases : NMCD. 2012.

80. Bing C. Lipid mobilization in cachexia: mechanisms and mediators. Current opinion in supportive and palliative care. 2011;5:356-60.

81. Murphy RA, Baracos VE, Mazurak VC, Reiman T, Bess B, Antoun S, et al. Lung cancer cachexia, *circa* 2011. In Preparation. 2011.

82. Fearon KC. The Sir David Cuthbertson Medal Lecture 1991. The mechanisms and treatment of weight loss in cancer. Proc Nutr Soc. 1992;51:251-65.

83. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. 1995;270:467-70.

84. Lee CK, Klopp RG, Weindruch R, Prolla TA. Gene expression profile of aging and its retardation by caloric restriction. Science. 1999;285:1390-3.

85. Jagoe RT, Lecker SH, Gomes M, Goldberg AL. Patterns of gene expression in atrophying skeletal muscles: response to food deprivation. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2002;16:1697-712.

86. Yechoor VK, Patti ME, Saccone R, Kahn CR. Coordinated patterns of gene expression for substrate and energy metabolism in skeletal muscle of diabetic mice. Proceedings of the National Academy of Sciences of the United States of America. 2002;99:10587-92.

87. Stephens NA, Gallagher IJ, Rooyackers O, Skipworth RJ, Tan BH, Marstrand T, et al. Using transcriptomics to identify and validate novel biomarkers of human skeletal muscle cancer cachexia. Genome medicine. 2010;2:1.

88. Abdullah-Sayani A, Bueno-de-Mesquita JM, van de Vijver MJ. Technology Insight: tuning into the genetic orchestra using microarrays--limitations of DNA microarrays in clinical practice. Nature clinical practice Oncology. 2006;3:501-16.

89. Rudy J, Valafar F. Empirical comparison of cross-platform normalization methods for gene expression data. BMC bioinformatics. 2011;12:467.

90. Barrett T, Edgar R. Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. Methods in enzymology. 2006;411:352-69.

91. Hirakawa A, Hamada C, Yoshimura I. Sample size calculation through the incorporation of heteroscedasticity and dependence for a penalized t-statistic in microarray experiments. Journal of biopharmaceutical statistics. 2012;22:260-75.

92. Klebanov L, Yakovlev A. Is there an alternative to increasing the sample size in microarray studies? Bioinformation. 2007;1:429-31.

93. Gallagher IJ, Stephens NA, Macdonald AJ, Skipworth RJ, Husi H, Greig CA, et al. Suppression of skeletal muscle turnover in cancer cachexia: evidence from the transcriptome in sequential human muscle biopsies. Clinical cancer research : an official journal of the American Association for Cancer Research. 2012;18:2817-27.

94. Verducci JS, Melfi VF, Lin S, Wang Z, Roy S, Sen CK. Microarray analysis of gene expression: considerations in data mining and statistical treatment. Physiological genomics. 2006;25:355-63.

95. Fouladiun M, Korner U, Bosaeus I, Daneryd P, Hyltander A, Lundholm KG. Body composition and time course changes in regional distribution of fat and lean tissue in unselected cancer patients on palliative care--correlations with food intake, metabolism, exercise capacity, and hormones. Cancer. 2005;103:2189-98.

96. Mitsiopoulos N, Baumgartner RN, Heymsfield SB, Lyons W, Gallagher D, Ross R. Cadaver validation of skeletal muscle measurement by magnetic resonance imaging and computerized tomography. J Appl Physiol. 1998;85:115-22.

97. Mourtzakis M, Prado CM, Lieffers JR, Reiman T, McCargar LJ, Baracos VE. A practical and precise approach to quantification of body composition in cancer patients using computed tomography images acquired during routine care. Appl Physiol Nutr Metab. 2008;33:997-1006.

98. Burger MZ. I. The meaning of creatinine coefficient for the quantitative measurement of muscle mass and body composition. II. Creatine and creatinine excretion: relationship to muscle mass. Z Gesamte Exp Med. 1919:361-99.

99. Heymsfield SB, Arteaga C, McManus C, Smith J, Moffitt S. Measurement of muscle mass in humans: validity of the 24-hour urinary creatinine method. The American journal of clinical nutrition. 1983;37:478-94.

100. Wang ZM, Gallagher D, Nelson ME, Matthews DE, Heymsfield SB. Total-body skeletal muscle mass: evaluation of 24-h urinary creatinine excretion by computerized axial tomography. The American journal of clinical nutrition. 1996;63:863-9.

101. Welle S, Thornton C, Totterman S, Forbes G. Utility of creatinine excretion in body-composition studies of healthy men and women older than 60 y. The American journal of clinical nutrition. 1996;63:151-6.

102. Cahill GF, Jr. Fuel metabolism in starvation. Annual review of nutrition. 2006;26:1-22.

103. Giesecke K, Magnusson I, Ahlberg M, Hagenfeldt L, Wahren J. Protein and amino acid metabolism during early starvation as reflected by excretion of urea and methylhistidines. Metabolism: clinical and experimental. 1989;38:1196-200.

104. Lowry SF, Horowitz GD, Jeevanandam M, Legaspi A, Brennan MF. Whole-body protein breakdown and 3-methylhistidine excretion during brief fasting, starvation, and intravenous repletion in man. Annals of surgery. 1985;202:21-7.

105. Fryburg DA, Barrett EJ, Louard RJ, Gelfand RA. Effect of starvation on human muscle protein metabolism and its response to insulin. The American journal of physiology. 1990;259:E477-82.

106. Nicholson JK, Lindon JC. Systems biology: Metabonomics. Nature. 2008;455:1054-6.

107. Robertson DG, Watkins PB, Reily MD. Metabolomics in toxicology: preclinical and clinical applications. Toxicological sciences : an official journal of the Society of Toxicology. 2011;120 Suppl 1:S146-70.

108. Mihalik SJ, Michaliszyn SF, de las Heras J, Bacha F, Lee S, Chace DH, et al. Metabolomic profiling of fatty acid and amino acid metabolism in youth with obesity and type 2 diabetes: evidence for enhanced mitochondrial oxidation. Diabetes care. 2012;35:605-11.

109. Lewis GD, Farrell L, Wood MJ, Martinovic M, Arany Z, Rowe GC, et al. Metabolic signatures of exercise in human plasma. Science translational medicine. 2010;2:33ra7.

110. Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, et al. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. Nature. 2009;457:910-4.

111. Hutton JL, Martin L, Field CJ, Wismer WV, Bruera ED, Watanabe SM, et al. Dietary patterns in patients with advanced cancer: implications for anorexia-cachexia therapy. The American journal of clinical nutrition. 2006;84:1163-70.

CHAPTER 2: Research plan

2.1 Rationale and overall hypothesis

Cancer cachexia is a common syndrome characterized by muscle wasting. Wasting in cachexia is associated with immobility and mortality. Development of proper management strategies and treatment development is hampered by: 1) unknown mechanisms involved in humans and 2) lack of fast, accurate, non-invasive and clinically available screening tools in the clinical setting to diagnose cachexia. The overall hypothesis that unites this work is that high throughput technologies, namely gene expression profiling and metabolomic profiling, can be used to detect altered transcription and metabolite concentrations in patients with cancer to ultimately improve patient care. Various methodological considerations were addressed throughout this work in order to properly use these technologies.

2.2 Objectives and Hypotheses

Investigated in Chapter 3: Effects of sample size on differential gene expression profiles, robustness criteria for top ranked genes and discriminative models using human skeletal muscle as an example

The objective was to assess how sample size affects microarray data analysis. Specifically, to examine how sample size variation from n= 10 (5 per group) to n=120 (60 per group) affects the significance and rank order assigned to top differentially expressed genes and the ability to build a classifier to predict the phenotype of future samples. Sex was chosen as the phenotype because it is an unambiguous variable that remains unchanged despite environmental factors or

pathological states. It was hypothesized that sample sizes previously used (n= 2 to 30) to analyze human skeletal muscle gene expression using microarray result in inconsistent and unreliable rank order which may explain the lack of concordance between differentially expressed gene lists generated in different studies looking at the same phenotype.

Investigated in Chapter 4: Skeletal muscle in cancer cachexia is characterized by features of wasting, pathological lipid infiltration, inflammation and aberrant regeneration processes

Decreased muscle mass, decreased muscle attenuation (an indicator of increased fat infiltration) and increased weight loss are associated with negative outcomes in cancer. The objective of this chapter was to explore skeletal muscle gene expression from cancer patients with altered muscle characteristics and weight maintenance. It was hypothesized that expression of genes encoding proteins involved in growth, inflammatory signaling and degradation would be altered in the studied phenotypes.

Investigated in Chapter 5: Prediction of skeletal muscle and fat mass in patients with advanced cancer using a metabolomic approach

The objective was to determine how wide variations of lean and fat mass, dietary intake, metabolic rate and fuel metabolism present in patients with advanced cancer alter the urinary and plasma metabolome. It was hypothesized that urinary and plasma metabolomes are defined, in part; by varying mass of tissues (e.g.

adipose and skeletal muscle) as these produce tissue-specific metabolites in the course of their turnover / metabolism.

Investigated in Chapter 6: Learning to predict cancer-associated skeletal muscle wasting from ^{1}H-NMR profiles of urinary metabolites

The objective was to determine if a random spot urine sample could be used to detect if patients were losing or not losing muscle using a metabolomic approach. A second objective was to determine what classification algorithm would be most appropriate for analyzing the multivariate data that results from metabolomic analysis. It was hypothesized that metabolites produced from tissue breakdown are likely to be a sensitive indicator of muscle wasting and may provide a new diagnostic approach.

CHAPTER 3: Sex and sample size affect skeletal muscle gene expression microarray experiments

3.1 Introduction

Microarray technology has been adopted to gain a comprehensive picture of gene expression differences. In human studies, the sample size is often limited because microarray technology is quite costly and the required tissue biopsies may be invasive. For example, in the quest to understand sexual dimorphism in human skeletal muscle gene expression, the early report by Roth *et al.* (1) studied pooled samples from 5 men and 5 women on 4K arrays (Invitrogen). Later, several other groups studied samples from 6 to 15 participants per sex on 45K arrays (Affimetrix) (2-4). Such sample sizes are not unusual in gene array studies on human tissues (5).

A lack of concordance is evident in gene lists generated in studies that compared the same phenotypes. For example, amongst the top 20 - 30 differentially expressed genes reported in the two studies cited above (by Welle *et al.* and by Maher *et al.*), only 5 were common to both lists: ALDH4A1, DAAM2, INSR, IRX3, TPD52. The issue of poor overlap of gene lists across studies has raised doubts about their reliability and robustness of gene signatures in general (6).

A version of this chapter has been accepted to for publication by *Plosone*. Cynthia Stretch; Sheehan Khan; Nasimeh Asgarian; Roman Eisner; Saman Vaisipour; Sambasivarao Damaraju; Kathryn Graham; Oliver F. Bathe; Helen Steed; Russell Greiner; Vickie E. Baracos. Effects of sample size on differential gene expression, rank order and prediction accuracy of a gene signature. PLoS One. 2013 Jun 3;8(6):e65380

Microarray studies are conducted either: (1) to identify differentially expressed genes between groups (e.g. towards understanding underlying biological mechanisms) and/or (2) to identify patterns of gene expression that can be used to develop a predictor with high accuracy (e.g. for diagnosis of a disease) (7). Researchers typically report the top differentially expressed genes and these are often credited with high importance, however reproducibility of the identity and rank order (i.e. 1st or 50th most differentially expressed) is usually not addressed.

Sample size is proposed to be an important determinant of the number of differentially expressed genes reliably detected as well as the accuracy of a predictor (8-12). Some prior studies have considered what sample size is required to ensure that genes associated with a phenotype can be discovered with a minimal false discovery rate (13); others explore effects of sample size on the overlap of gene lists (8, 9); and yet others have investigated the effect of sample size on the likelihood of identifying true associations among the top ranked genes (14). In general, these analyses consider various sub-samples of a given large initial dataset, to determine how well each size of subsamples approximates findings made using the entire dataset. Because of a general paucity of large datasets, authors either used computer-simulated datasets (8, 9), or created data pools by combining independent datasets (8, 15). However, simulated data does not necessarily reflect biological variation and pooling of data from different studies by different investigators introduces batch effects and thereby increase variability (5, 16). We can avoid these problems by using a single large dataset

acquired on the same platform, lab and experimental condition. It is also important that the class label (phenotype) be unambiguous. An objective class label (e.g. male vs. female) rather than subjective (e.g. estrogen receptor status, subject to measurement error and based on the subjective opinion of an individual pathologist (17)) would be ideal. A subjective class label may contaminate the dataset with incorrectly labeled instances and therefore introduce variation.

Here, we used sexual dimorphism in human skeletal muscle gene expression using a single large (n=134) dataset with 41K Agilent arrays, as a model to assess effects of sample size on differential expression, rank order and prediction tasks. For the association analyses, our goal was to determine consistency of the rank orderings of genes, from one size-n sample to another; this is different from other studies that attempt to determine how many of the top biomarkers are "correct" (14).

3.2 Methods

3.2.1 Ethics Statement

This study was approved by the Alberta Cancer Research Ethics Committee. Patients provided written consent. Tissues were stored at the Alberta Cancer Research Biorepository/Canadian Breast Cancer Foundation Tumor Bank and the University of Calgary HPB/GI Tumor Tissue Bank.

3.2.2 Participants and acquisition of muscle samples

Adult (>18 yrs) cancer patients underwent open abdominal surgery as part of their clinical care. Biopsies of *rectus abdominis* muscle (0.5 – 1 g) were taken

from the site of incision, at the start of surgery using sharp dissection and without the use of electrocautery. Biopsies were immediately frozen in liquid nitrogen and stored in liquid nitrogen vapor phase until analysis. Age and cancer type information were abstracted from medical charts.

3.2.3 Computed tomography image analysis

Digital axial computed tomography (CT) scans, completed for the purpose of planning the surgery, were used to quantify skeletal muscle area as in our prior studies (18). Briefly, images at the 3rd lumbar vertebra (L3) were analyzed for total muscle cross-sectional area (cm^2) within a specified Hounsfield Unit range (-29 to +150) using Slice-O-Matic software (v.4.3, Tomovision, Montreal, Canada). Muscle area was normalized for stature and reported as lumbar skeletal muscle index (SMI, cm^2/m^2). To express muscle in conventional units, whole body fat-free mass (FFM) was estimated from a regression equation that has been applied in several different cancer populations (18-20):

Whole body skeletal muscle volume = 0.166 * [skeletal muscle 5 cm above L4-L5 (cm^2)] + 2.142; r^2=0.855

To determine if there was a difference in muscle mass rate of change between men and women, we determined the mean tissue area for 2 consecutive images (taken ~ 100 days apart) prior to biopsy. We expressed this rate as percentage change per 100 days, to take into account minor variation in the number of days between scans for different individuals.

3.2.4 Microarray analysis

Total RNA was isolated using Trizol (Sigma-Aldrich, Oakville, ON, CAN) and purified using Qiagen RNeasy columns (Mississauga, ON, CAN) according to the manufacturer's protocols. RNA was quantified using a NanoDrop 1000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) and its integrity evaluated using a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA) according to manufacturer's protocols. RNA samples with RNA Integrity Numbers (RIN) greater than 7.0 were used.

RNA was subjected to linear amplification and Cy3 labeling and Hybridization to Agilent Whole Human Genome Arrays using Agilent kits (One Color Low RNA Input Linear Amplification Kit Plus, One Color RNA Spike-In Kit and Gene Expression Hybridization Kit) according to the manufacturer's protocols. Arrays were scanned using an Agilent Scanner, the data was extracted and quality was evaluated using Feature Extraction Software 10.5.1 (Agilent). Data was normalized using GeneSpring GX 11.5.1 (Agilent). The data used in this publication have been deposited in the U.S. National Center for Biotechnology Information (NCBI) Gene Expression Omnibus25 and are accessible through GEO series accession number GSE41726.

3.2.5 Statistical analysis

There were a total of 41,000 oligonucleotide sequences (i.e. transcripts) on each microarray chip. This produces a dataset that describes each of 134 participants (69 men and 65 women), using 41,000 transcripts (each a real

number) and sex (either M or F). Microarray intensity values were log transformed prior to analyses.

3.2.5.1 Effect of sample size on differentially expressed gene lists

For each sample size considered (n = 10 (5♀, 5♂), 20 (10♀, 10♂), … 120 (60♀, 60♂)), we randomly selected a size-n subsample (containing equal numbers of men and women) from our dataset of n=134. For each of these size-n subsamples, we computed the t-test on the (log transformed) intensities over the set of males vs. the set of females. We repeated this procedure 50 times for each sample size n and then for each gene, averaged the p-values computed over these 50 trials. Mean p-values were then sorted from lowest to highest to determine top 100 transcripts for each sample size. We also evaluated how the specific rank order of top genes was affected by sample size. For each size n subsample we assigned a rank value (1 to 100) to each gene, based on its p-value. We then sorted the gene based on its mean rank (for each sample size), based on all 50 repeats. As our main focus was this ranking, we simply used the p-values from the t-tests, rather than any multiplicity-corrected variant (such as the Benjamini-Hogeberg correction (21)). If we had used a multiplicity correction, enforced monotonicity would have been required to ensure the ranking of adjusted p-values remain unchanged. A method of enforced monotinicity was presented by Yekutieli and Benjamini (22).

3.2.5.2 Effect of sample size on prediction accuracy

As XY chromosome transcripts (1,548 transcripts) are obviously highly related to sex, a single XY transcript may be sufficient to build a classifier that

could predict sex perfectly. To generate a more typically physiological prediction problem we therefore excluded these transcripts when building classifiers. We used the LASSO algorithm (implemented using R, glmnet package) (23). Given a training dataset, LASSO produces a classifier that predicts the class label (sex) of a new patient from his/her microarray data. In general, the quality of a classifier is its predictive accuracy (% correct classification) on novel subjects; we used 10-fold cross validation to internally validate the model. To determine how sample size of the training dataset affects sex prediction accuracy, we trained classifiers using randomly selected sub-samples of our data (n=10 (5♀, 5♂) to n= 110 (55♀, 55♂)). We repeated this 50 times for each n.

To externally validate our model, we used publicly available datasets that used the same tissue (i.e. skeletal muscle) and platform (i.e. Agilent), for which the sex was known: dataset GSE24215 included microarray data from 10 healthy, young men and dataset GSE23697 included 34 healthy, adult men. To determine how sample size of the training dataset affects sex prediction accuracy on these external datasets, we trained classifiers using randomly selected subsamples of our data (n=10 (5♀, 5♂) to n= 110 (55♀, 55♂)) then used these learned classifiers to predict sex on the external datasets. We repeated this 50 times for each n.

3.3 Results

Gene expression microarray analysis was conducted on 134 rectus abdominus muscle biopsies (69♂, 65♀). Characteristics of the study participants are shown in Table 3-1. As expected, men were 26% more muscular than women

(t-test, $p<0.0001$). Mean age and number of patients undergoing chemotherapy did not differ between the men and women in this study.

3.3.1 Effect of sample size on differential expression

The full dataset was checked for differential gene expression revealed 717 differentially expressed transcripts with a p-value < 0.0001 (Appendix 1). Note that the biological interpretation of these differentially expressed genes is not the goal of this study.

This analysis was repeated for random samples of n=10 (5♀, 5♂) to n=120 (60♀, 60♂) increasing the sample size by increments of 5♀ and 5♂ (Figure 3-1, top panels). At n=10 (5♀, 5♂), no genes were significant at p-value < 0.0001 whereas at n=120 (60♀, 60♂), there were 472 differentially expressed transcripts at the same p-value cutoff. Of course, the variance of these measurements become less meaningful for large subsamples, as the different size-n subsamples will have high overlap since they are all drawn from our dataset of 134 (65♀, 69♂). The variances, however, are fairly accurate for small values of n. To assess the similarity of sample sets, we calculated the median Jaccard score over 1000 randomly generated pairs of subsamples of size n. The Jaccard score of two sets A and B is the size of the intersection divided by the size of the union, i.e. $J(A,B) = |A \cap B| / |A \cup B|$. Note that the Jaccard score is always between 0 and 1; the score of 0 means the two sets are disjoint, while the score of 1 means they are identical. As the median Jaccard score for two n =30 (15♀, 15♂) subsamples is around 0.1, the overlap is very small. Such sizes are the most relevant, as they reflect the sizes of many earlier human microarray studies. Below we consider n=60 (30♀,

30♂); we consider the observed variations relevant as the Jaccard scores here are still under 0.3.

We then explored whether the ranking of genes were reproducible over different sample sizes. From the previous analysis, for each of the 50 random samplings, each gene was given a ranking based on the p-value of its t-test (e.g. if a gene is ranked 4th, or 25th, or 120th), see Figure 1, bottom panels. In Figure 3-1 (bottom right-hand panel), in a large sample (n=60 (30♀, 30♂)) the top three genes (PRKY, DDX3Y, UTY) were reproducibly identified in the top 3 ranks in all 50 iterations of sampling. By contrast, the p-value of 10th ranked gene was very close to its immediate neighbors; while on average it ranked 10th, its rank ranged from 5th to 17th. As we decreased n, the ranking of any given gene became more and more variable, in that the rank of every gene had a larger range (e.g. at sample size n=30 (15♀, 15♂), the gene whose average rank was 10th , ranged in rank from 1st to 127th in the different random subsamples).

3.3.2 Effect of sample size on prediction accuracy

Microarray data are sometimes used to make a prediction (i.e. to determine the phenotype of a future subject (e.g. healthy or disease), based on a classifier produced from prior subjects. While there is no clinical need to predict a person's sex using muscle gene expression array, our data does provide the opportunity to explore the relationship between n and the ability to build a robust predictor. The classifier based on all n=134 participants used only 92 genes of the complete set of 41,000 and could predict sex with mean 92.5 ± 7.3 % (10-fold) cross-validation accuracy. We then explored the predictive accuracy of this model on publicly

available muscle expression array data obtained in unrelated investigations conducted on the same platform (Agilent). In two such external datasets, this model had excellent accuracy: correctly predicted sex for 9/10 participants and for 35/35 participants in dataset GSE24215 and dataset GSE23697, respectively.

Figure 3-2 shows the mean internal cross-validation accuracy of sex prediction as we varied the sample size of the training data from n=10 (5♀, 5♂) to n=110 (55♀, 55♂). When the training sample had n=10 (5♀, 5♂), the classifier was unable to predict sex any better than chance (~50% cross validation accuracy). This accuracy increased as we increased n; we achieved predictive accuracy above 90% when training on a sample of at least 80 individuals (40 of each sex). This trend of increased accuracy with increased n was also seen when we used different subsamples of size n=10 (5♀, 5♂) to n=110 (55♀, 55♂) from our dataset to predict sex on the external datasets mentioned above (Figure 3-3).

3.4 Discussion

Our empirical evidence suggests that small sample sizes often typical of microarray studies negatively affect their interpretation, whether used to determine differential gene expression or to accurately predict future instances. The relatively high cost of analyses and the invasiveness of sampling tissues such as skeletal muscle in humans often dictate rather small sample sizes (24, 25) but our results suggest that efforts to increase n may well be justified.

Researchers in biology attribute great importance to top ranked gene(s) in differential expression analyses (26). This is the first study examining the effect of sample size on gene rank using one large dataset and a biologically unambiguous

label. We show that any given gene may have a wide range of ranks, especially for small sample sizes. For example, in 50 subsamples of size n= 20 (10♀, 10♂), the gene that had the highest average rank, sometimes appeared in rank 200. By contrast, at n=60 (30♀, 30♂) the top three ranked genes were constant. These observations explain the lack of concordance between the findings of two prior studies of sexual dimorphism (2, 4) with each other, and with our results. Those two earlier studies had 6 to 15 of each sex in their analysis and the 5 genes that they had in common with each other did not all rank in our top 100 differentially expressed gene list (these genes ranked 49th (IRX3), 62nd (DAAM2), 67th (TPD52), 147th (ALDH4A1) or was not significant (p=0.3, rank= 18854) (INSR) at n=134 (our full dataset).

Microarray analysis is often used to identify gene signatures that can be used to develop a predictor. In agreement with previous studies looking at this methodological issue, we conclude that small sample sizes (e.g. n < 20 per class label) will often result in poor predictors (8, 10, 15, 27), but the accuracy improves with increased sample size in the training dataset. This was the case both within our data and, as shown in Figures 3-3, when trying to make predictions on external datasets. In our study, a LASSO predictor, trained on our full dataset (n=134) returned 90-100% accuracy on publicly available data (external validation). This excellent predictive accuracy also suggests that our findings of sex-related gene expression are not confounded by the use of a cancer patient sample, because the predictor based on these patients was accurate on data obtained on healthy men and women.

Sample size is not the only factor that can influence microarray analysis. Indeed, incomplete annotation of the genome and probes targeted to different regions of the encoding gene, stringency of hybridization conditions, commercially available arrays vs. in-house built, pre-analytical variables in the tissue accrual including induced hypoxia concomitant post de-vitalization of tissue and temperature and duration of storage of tissues should also be considered when comparing previous or designing future microarray experiments. Here, we focused on sample size while maintaining the tissue collection method, microarray platform and storage conditions constant for all samples. Our analysis suggests principles that dictate how ranking and prediction accuracy can vary, in relation to the biological label (sex) that we chose to study. Studies with larger inherent variance in the data (e.g. due to batch effects introduced by pooling several datasets) or different effect sizes may require considerably larger sample sizes than we report here (8-10, 28). By contrast in animal experiments which permit extensive control of many sources of variation, smaller sample sizes may be sufficient to test similar experimental questions. Thus, it is not possible to state how many genes will be reliable / reproducible at different sample sizes for other datasets *a priori*. However, it would be beneficial to assess how sample size may affect ranking and prediction tasks, as we did here, by examining the robustness of top-ranked genes and mean and variance of cross-validated results for different subsamples of varying n, respectively. Even if a dataset is deemed to have a sufficient sample size, there are other methodological considerations that were not addressed here but which are important to properly interpret the data. For

example, using different normalization and multiplicity correction methods lead to different differential expression results (29-30). Researchers need to evaluate what normalization method is best suited for their data and carefully consider what multiplicity correction should be used, which depends on the properties of the dataset in questions (e.g. the normality of the data).

3.5 Conclusion

We conclude that gene signatures generated from small datasets should be interpreted with caution as they may not be reproducible and that prediction models built using small sample sizes result in poor prediction accuracy. While we cannot recommend specific sample sizes, outside the problem that we studied, our analysis shows that the sample size n= 10 (5♀, 5♂) was not useful for either prediction (which was not better than chance) nor for association (the probability of finding reproducible top 10 genes was negligible).

Tables

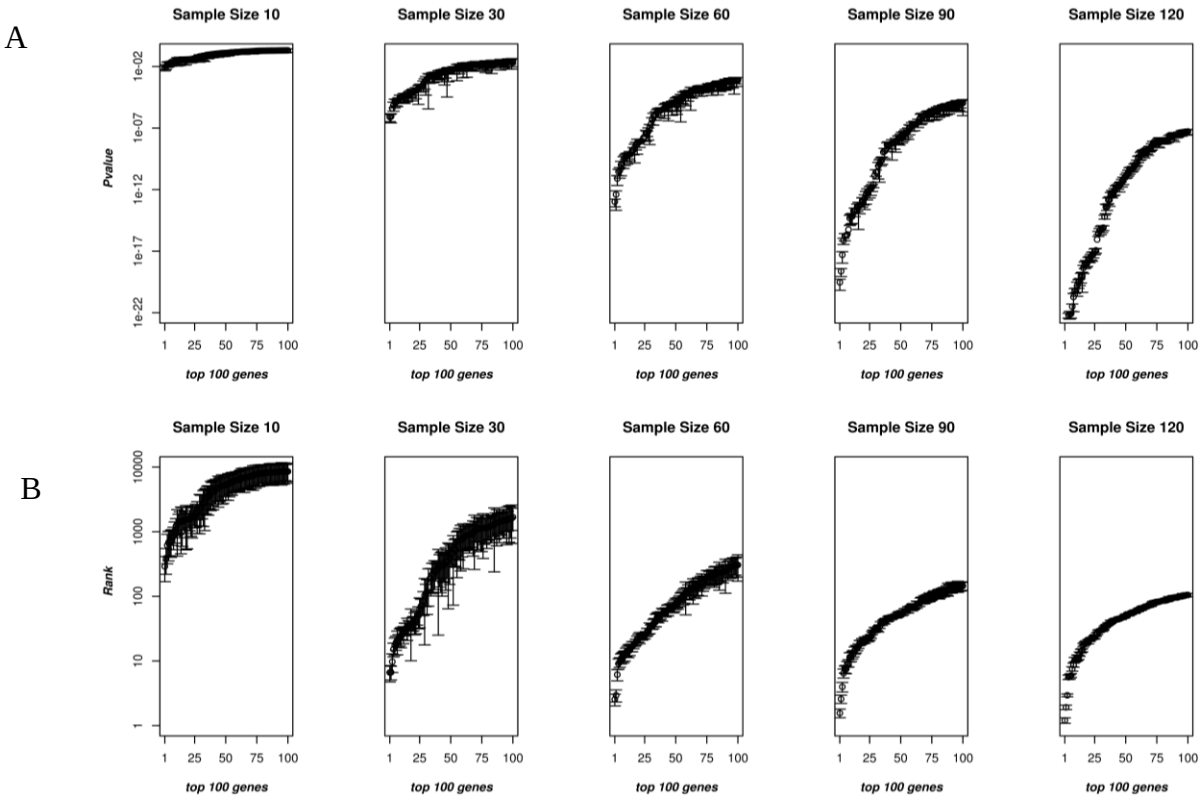
Table 3-1: Patient characteristics

	Men	Women
Total, n	69	65
Age, mean years ± SD	59 ± 13	63 ± 13
Muscle, mean ± SD		
Skeletal muscle index (cm^2/m^2)	52.9 ± 7.8	41.9 ± 8.3*
[1]Estimated whole body skeletal muscle, kg	27.0 ± 4.8	17.9 ± 3.7*
Muscle rate of change, %/100d	-4.4 ± 10.9	-4.5 ± 12.5
Diagnosis at surgery, %		
Benign neoplasm	13	18
Cancer, liver or intrahepatic bile ducts	17	14
Cancer, gastrointestinal tract	46	22
Cancer, pancreas	19	25
Cancer, ovary or uterus	0	17
Cancer, head and neck	3	2
Cancer, skin	0	2
Cancer, kidney	1	0

*Different from men, $P < 0.0001$. [1]Derived regression equations by Shen *et al.* (ref # 20)

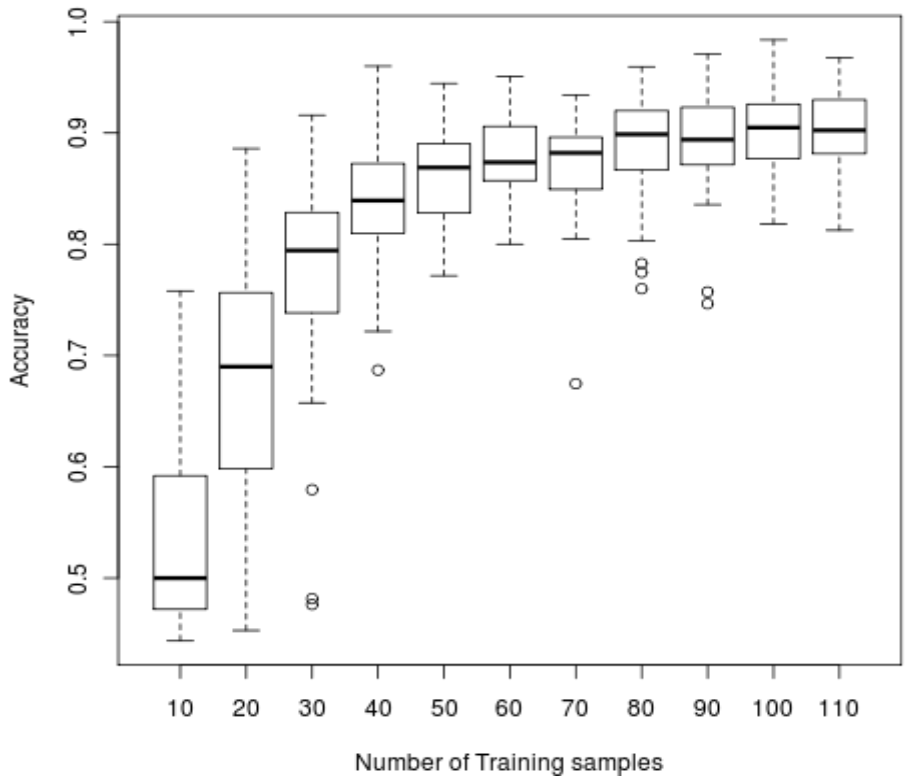
Figures

Figure 3-1: The effect of sample size on p-values and rank order for the top 100 transcripts



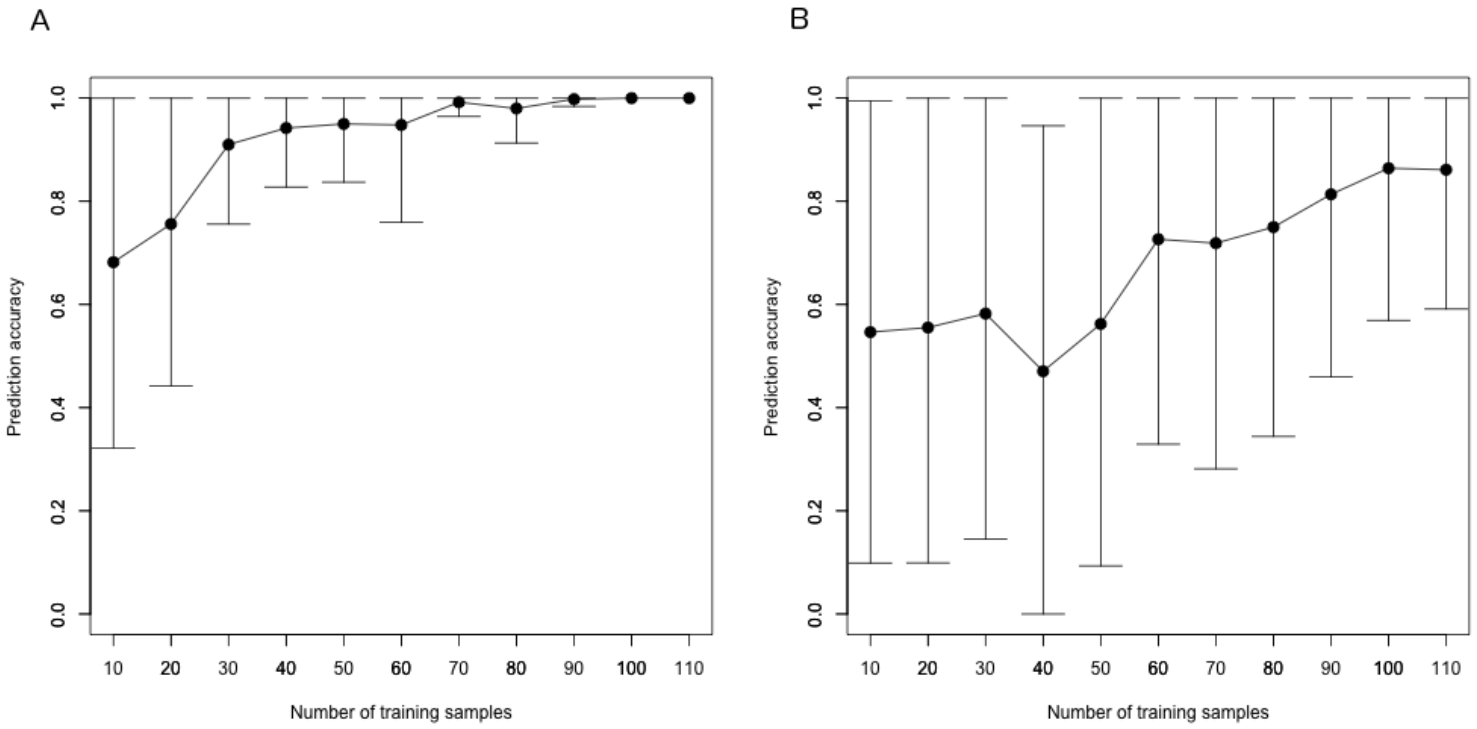
For each n tested (n= 10 (5♀, 5♂), 30 (15♀, 15♂), 60 (30♀, 30♂), 90 (45♀, 45♂) and 120 (60♀, 60♂) are shown here), n samples were randomly selected from our dataset of n=134 participants, 50 different times. Note log on the y-axes.
Top panels: The average and 0.95 SD of the p-values for the top 100 transcripts. As sample size was increased, the average p-value decreased and became less variable.
Bottom panels: The average and 0.95 SD (log) rank for the top 100 transcripts. As sample size was increased, the average rank decreased and became less variable.

Figure 3-2: Box-and-whiskers plot showing the mean internal cross-validation accuracy of sex prediction for different sample sizes



Sample sizes tested ranged from n=10 (5♀, 5♂) to n=110 (55♀, 55♂). To calculate the mean 10-fold cross validation prediction accuracy, for each n (=10…110), we built classification models using a randomly selected size-n subsamples of our full dataset of n=134. This was repeated 50 times and the median prediction accuracy for each sample was calculated. As sample size increased, so did prediction accuracy.

Figure 3-3: **Plots showing the mean and standard deviation accuracy of sex prediction on two external datasets using a predictor trained using different sample sizes from our dataset.**



We built predictors using different training sample sizes ranging from n=10 (5♀, 5♂) to n=110 (55♀, 55♂) from our full dataset. We then calculated the prediction accuracy, for each n (=10…110) on two external datasets (A. Dataset GSE24215 and B. Dataset GSE23697). This was repeated 50 times and the mean and standard deviation prediction accuracy for each sample size was calculated. As the training sample size increased, so did prediction accuracy on the external datasets.

References

1. Roth SM, Ferrell RE, Peters DG, Metter EJ, Hurley BF, Rogers MA. Influence of age, sex, and strength training on human muscle gene expression determined by microarray. Physiological genomics. 2002;10:181-90.

2. Welle S, Tawil R, Thornton CA. Sex-related differences in gene expression in human skeletal muscle. PloS one. 2008;3:e1385.

3. Liu D, Sartor MA, Nader GA, Gutmann L, Treutelaar MK, Pistilli EE, et al. Skeletal muscle gene expression in response to resistance exercise: sex specific regulation. BMC Genomics. 2010;11:659.

4. Maher AC, Fu MH, Isfort RJ, Varbanov AR, Qu XA, Tarnopolsky MA. Sex differences in global mRNA content of human skeletal muscle. PloS one. 2009;4:e6335.

5. Klebanov L, Yakovlev A. Is there an alternative to increasing the sample size in microarray studies? Bioinformation. 2007;1:429-31.

6. Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. Lancet. 2005;365:488-92.

7. Simon R. Diagnostic and prognostic prediction using gene expression profiles in high-dimensional microarray data. British journal of cancer. 2003;89:1599-604.

8. Kim SY. Effects of sample size on robustness and prediction accuracy of a prognostic gene signature. BMC bioinformatics. 2009;10:147.

9. Ein-Dor L, Zuk O, Domany E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. Proceedings of the National Academy of Sciences of the United States of America. 2006;103:5923-8.

10. van Vliet MH, Reyal F, Horlings HM, van de Vijver MJ, Reinders MJ, Wessels LF. Pooling breast cancer datasets has a synergetic effect on classification performance and improves signature stability. BMC Genomics. 2008;9:375.

11. Ramasamy A, Mondry A, Holmes CC, Altman DG. Key issues in conducting a meta-analysis of gene expression microarray datasets. PLoS Med. 2008;5:e184.

12. Pusztai L, Mazouni C, Anderson K, Wu Y, Symmans WF. Molecular classification of breast cancer: limitations and potential. Oncologist. 2006;11:868-77.

13. Matsui S, Oura T. Sample sizes for a robust ranking and selection of genes in microarray experiments. Statistics in medicine. 2009;28:2801-16.

14. Kuo CL, Zaykin DV. Novel rank-based approaches for discovery and replication in genome-wide association studies. Genetics. 2011;189:329-40.

15. Dobbin KK, Zhao Y, Simon RM. How large a training set is needed to develop a classifier for microarray data? Clinical cancer research : an official journal of the American Association for Cancer Research. 2008;14:108-14.

16. Bolstad BM, Collin F, Simpson KM, Irizarry RA, Speed TP. Experimental design and low-level analysis of microarray data. Int Rev Neurobiol. 2004;60:25-58.

17. Diaz LK, Sneige N. Estrogen receptor analysis for breast cancer: current issues and keys to increasing testing accuracy. Advances in anatomic pathology. 2005;12:10-9.

18. Eisner R, Stretch C, Eastman T, Xia J, Hau D, Damaraju S, et al. Learning to predict cancer-associated skeletal muscle wasting from 1H-NMR profiles of urinary metabolites. Metabolomics. 2011;7:25-34.

19. Murphy RA, Mourtzakis M, Chu QS, Baracos VE, Reiman T, Mazurak VC. Nutritional intervention with fish oil provides a benefit over standard of care for weight and skeletal muscle mass in patients with nonsmall cell lung cancer receiving chemotherapy. Cancer. 2011;117:1775-82.

20. Shen W, Punyanitya M, Wang Z, Gallagher D, St-Onge MP, Albu J, et al. Total body skeletal muscle and adipose tissue volumes: estimation from a single abdominal cross-sectional image. J Appl Physiol. 2004;97:2333-8.

21. Benjamini Y, Hogchberg, Y. Controlling the False Discovery Rate: a practical and powerful approach to multiple testing. J Royal Stat Soc Ser B. 1995;1:289-300.

22. Yekutieli D, Benjamini Y. Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. Journal of Statistical Planning and Inference. 1999;82:171-96.

23. Tibshirani R. Regression shrinkage and selection via the lasso. J R Statist Soc B. 1996;58:267-88.

24. Timmons JA, Sundberg CJ. Oligonucleotide microarray expression profiling: human skeletal muscle phenotype and aerobic exercise training. IUBMB life. 2006;58:15-24.

25. Virtanen C, Takahashi M. Muscling in on microarrays. Applied physiology, nutrition, and metabolism = Physiologie appliquee, nutrition et metabolisme. 2008;33:124-9.

26. Fluck M, Dapp C, Schmutz S, Wit E, Hoppeler H. Transcriptional profiling of tissue plasticity: role of shifts in gene expression and technical limitations. J Appl Physiol. 2005;99:397-413.

27. Popovici V, Chen W, Gallas BG, Hatzis C, Shi W, Samuelson FW, et al. Effect of training-sample size and classification difficulty on the accuracy of genomic predictors. Breast cancer research : BCR. 2010;12:R5.

28. Damavandi B. Estimating the overlap of top instances in lists ranked by correlation to label. Edmonton, Alberta: University of Alberta; 2012.

29. Ding Y, Wilkins D. The effect of normalization on microarray data analysis. DNA Cell Biology. 2004;10:635-642.

30. Bouaziz M, Jeanmougin M, Guedj M. Multiple testing in large-scale genetic studies. Methods Mol Biol. 2012;888:213-33.

CHAPTER 4: Skeletal muscle in cancer cachexia is characterized by features of wasting, pathological lipid infiltration, inflammation and aberrant regeneration processes

4.1 Introduction

Cancer patients often develop cachexia which is defined by muscle wasting (with or without loss of fat) (1). Muscle loss is an important component of the pathophysiology of cancer as it is associated with fatigue, decreased quality of life, decreased response to treatment and decreased survival (2, 3).

Mechanisms involved in muscle wasting are not completely understood. However, much of the research has had a focus on protein turnover and cell proliferation and apoptosis, since these are considered *a priori* to be involved. Specific processes considered include protein degradation (e.g. the ubiquitin (Ub)-proteasome pathway), apoptosis, autophagy and cell cycle arrest which favour wasting, and protein synthesis, transcription, translation, cell cycle and progenitor cell activation and differentiation which oppose wasting (4). Figure 4-1 part A shows a conceptual framework of different inputs from distant organs/tissues and local/neighbouring cells and part B of the figure shows a conceptual framework of pathways that are suggested to affect muscle in cancer cachexia. These different pathways are activated through ligand-receptor downstream signalling (e.g. pro-inflammatory cytokines, glucocorticoids, growth factors, growth hormone, insulin and androgens) as well as via neural or mechanical activation of muscle. Considering the obvious nature of these signals and pathways in wasting, it is easy

to understand the considerable emphasis on them in the cancer cachexia literature (5-17). Generally, there is an increase in catabolism and a decrease in anabolism during cancer-associated wasting (11, 12).

Gene array has been attempted to understand events leading to wasting, beyond the involvement of the pathways described above. Lecker *et al.* used 10k Affymetrix gene arrays to identify a common set of transcriptional adaptations responsible for muscle wasting in laboratory rodents in four different disease states (cancer, diabetes, renal failure and starvation) (18). This common transcriptional program, identified upregulation of protein degradation, downregulation of factors involved in muscle protein transcription and translation as key *atrophy genes* but additionally identified downregulation of pathways of energy metabolism, extracellular matrix (ECM) remodelling as well as several miscellaneous pathways(18). Braun *et al.* used 41k Agilent arrays and found many aspects of the Lecker transcriptional *atrophy gene* program in response to central (intracerebroventricular) interleukin 1β injection in mice, including transcription, protein degradation, ECM and energy metabolism (19). Braun also identified myogenic differentiation, inflammatory signalling, and oxidative stress in their model (19). Collectively these works have begun to reveal a wider understanding of muscle wasting in cachexia, however they have yet to be extensively validated in either human or animal studies.

Most of the understanding of cancer-associated changes in skeletal muscle comes from over 400 studies conducted in rodent tumour models. However animal studies may or may not provide a basis for the design of therapeutic

strategies unless it is can be verified that homologous events occur in humans. Only 18 studies used muscle biopsies from patients with cancer as of July 2012 (Table 4-1). These few studies used small, heterogeneous samples and focused on just a few of the observations and pathways identified in the animal models such as decreased fiber size (20-22), increased fat content (23), increased proteolysis (24-29), decreased protein synthesis (30, 31), apoptosis and regeneration processes (32, 33).

Involuntary weight loss is the primary clinical feature of cachexia, as healthy adults are highly resistant to weight loss. Weight loss is expediently measured (1). This explains why weight loss is the most commonly used criterion for cachexia classification (Table 4-1). This has the inherent flaw that weight loss may or may not be related to muscle loss (34). Arguably, weight loss is not an ideal way to classify patients for cachexia research designed to probe molecular events and pathways in skeletal muscle. Muscle loss is considered the defining characteristic of cachexia (34). Very recently, computed tomography (CT) has been applied to quantification and characterization of skeletal muscle in cancer patients (35-37), providing for the first time an opportunity to use direct measures of muscle in clinical cachexia research.

Recently, one research group published studies using a gene expression analysis of skeletal muscle to explore cachexia mechanisms (38, 39). However, these previous studies are limited by the use of weight loss to classify patients as cachexic or not. Also, it may be reasonably assumed that samples of mixed sex and small n would detract from their ability to detect significant differences in

gene expression (see Chapter 3). Our aim was to identify muscle gene expression signatures in relation to muscle characteristics detectable on CT images (muscle mass index, muscle radiation attenuation (an index of fat infiltration)) using a relatively large dataset of a single sex.

4.2 Methods

This study was approved by the Alberta Cancer Research Ethics Committee. Patients provided written informed consent to banking of the tumour, buffy coat, and skeletal muscle with the Canadian Breast Cancer Foundation Tumour Bank, a province wide tumour bank within the Cancer Care domain, Alberta Health Services. Release of 69 muscle samples from the bank for microarray analysis and collection of patient information for this study was conducted under Research Ethics Protocol ETH-21709.

4.2.1 Participants and acquisition of muscle samples

To avoid a potential confounder we opted to focus on only one sex. Compared to women, men with cancer have been reported to experience a faster rate of weight loss (40), be more likely to present with sarcopenia (i.e. low muscle mass) (35) and report poorer global quality of life (associated with reduced muscle quality) and physical function and increased fatigue (41). Based on the foregoing we opted to focus on a single sex (men).

Acquisition of samples was conducted in the same manner as in Chapter 3. Briefly, biopsies of *rectus abdominis* muscle (0.5 – 1 g) were taken at the start of open abdominal surgery using sharp dissection and immediately frozen in liquid

nitrogen and stored in liquid nitrogen until analysis. Using scheduled surgeries allowed us to avoid causing additional pain and burden to patients, as would be the case if we had used the usual percutaneous needle biopsy method. Our approach was perceived by patients as minimally invasive and as a result allowed us to obtain an unusually large number of human muscle samples (see Table 4-1 to compare sample sizes from other human studies).

Chart review was used to identify details of diagnosis, computed tomography (CT) images (analysis detailed below) and history of weight loss. Patients were considered to be losing weight if they had lost more than 5% of weight over the 6 months prior to the biopsy.

4.2.2 Body composition analysis

Digital axial CT scans done as part of clinical care were used to quantify skeletal muscle area as in our prior studies (34, 42). Different tissues are indentified by an individual trained in anatomical radiology and quantified based on their attenuation characteristics which are a function of their composition (43). Radiation attenuation is measured in Hounsfield units (HU) using water [0 HU] and air [-1000 HU] as reference. In this study, images at the 3rd lumbar vertebra (L3) were analyzed for total muscle cross-sectional area (cm^2) within a specified Hounsfield Unit range (-29 to +150) using Slice-O-Matic software (v.4.3, Tomovision, Montreal, Canada). Muscle area was normalized for stature and reported as lumbar skeletal muscle index (SMI, cm^2/m^2). Muscle attenuation is considered to be related to muscle fat content (44). Specifically, a lower mean

attenuation indicates increased lipid content. We measured the mean muscle attenuation for the entire cross-sectional area at L3.

4.2.4 Microarray analysis

Microarray analysis was conducted as described in Chapter 3. Briefly, total RNA was isolated using Trizol (Sigma-Aldrich, Oakville, ON, CAN) and purified using QiagenRNeasy columns (Mississauga, ON, CAN) according to the manufacturer's protocols. RNA was quantified using a NanoDrop 1000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) and its integrity evaluated using a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA) according to manufacturer's protocols. RNA samples with RNA Integrity Numbers (RIN) greater than 7.0 were used. RNA was subjected to linear amplification and Cy3 labelling and Hybridization to Agilent Whole Human Genome Arrays using Agilent kits (One Color Low RNA Input Linear Amplification Kit Plus, One Color RNA Spike-In Kit and Gene Expression Hybridization Kit) according to the manufacturer's protocols. Arrays were scanned using an Agilent Scanner, the data was extracted and quality was evaluated using Feature Extraction Software 10.5.1 (Agilent). The data was normalized using GeneSpring GX 11.5.1 (Agilent). Each microarray chip had of 41,000 oligonucleotide sequences or transcripts.

4.2.5 Statistical analysis

4.2.5.1 Differential gene expression

To run the differential expression analysis patients were first classified as low (Class 1) or high (Class 2) for skeletal muscle index, muscle attenuation and

weight loss. The approach taken to split patients into the different classes is often referred to as *extreme phenotype classification* and often used in gene expression-type studies (45-48). This method consists of comparing patients with very high or very low values for SMI or muscle attenuation or weight loss and excluding patients in the middle. This approach exploits extremes of phenotype to maximize the chance of observing major differences in gene expression. Also, when a population is split across a cut point into 2 classes, individuals on either side of the cut point who are not different from one another within the error term of the measurement are placed in different classes; this problem is avoided with our approach. To conduct this classification we first sorted patients based on their SMI or attenuation values (low to high). After sorting, patients in the lowest tertile were classified as Class 1, patients in the highest tertile were classified as Class 2 and patients with values in the middle were excluded. Figure 4-2 uses muscle attenuation as an example of how this is conducted.

For each phenotype assessed, T-test analysis was conducted on the log transformed intensities from the microarray data (Class 1 vs. Class 2).

4.2.5.1 Pathway analysis

Differentially expressed genes were analyzed using Ingenuity Pathway Analysis (IPA) (Ingenuity® Systems, www.ingenuity.com). Specifically, we uploaded the Agilent ID and nominal p-value from T-test analysis described above for each differentially expressed transcript with a p-value ≤ 0.01 into IPA. Based on the updated IPA database, each transcript was assigned its corresponding gene name. Non-coding or unknown genes in the IPA software

were not included in IPA analysis. IPA provided information about common biological functions and canonical pathways of differentially expressed genes based on the information in the IPA database.

4.3 Results

4.3.1 Patient characteristics

From our group of 69 patients, 64 patients had weight loss information or CT derived data (i.e. SMI and HU values). This clinical information was used to assign patients into Class 1 or Class 2 as detailed in the methods section. Patient characteristics for Class 1 and Class 2 for each variable are shown in Table 4-2. Though the sample sizes in Table 4-2 may appear suboptimal based on the results from the previous Chapter, it is important to note that the classification method used in Chapter 3 differs from what was used here (i.e. the extreme phenotype classification method). By using this method we exploit extremes of each phenotype to maximize the chance of observing major differences in gene expression. The three phenotype characteristics were not mutually exclusive and tended to overlap to some degree (Figure 4-3). Muscle attenuation and SMI were somewhat correlated ($r=0.33$). There was little to no correlation between the percent of weight lost and SMI and muscle attenuation, $r=0.11$ and $r=-0.01$, respectively. Comparing the distal ends of the weight loss spectrum, there were no significant differences in directly measured muscle parameters (index or attenuation). Patients with low muscle attenuation were significantly older than patients with high muscle attenuation ($p<0.0001$).

4.3.2 Differential expression analysis

T-test analysis was conducted for each phenotype. For each of the 3 phenotypes, there were different numbers of differentially expressed genes at given p-values. In Table 4-3 the numbers of differentially expressed transcripts are given for each phenotype and p-value, and both the total number and the number that exceeded 1.5-fold in magnitude are given. For muscle attenuation there were 5, 52, 440, 1644 and 2715 differentially expressed genes at $p < 0.00001$, $p < 0.0001$, $p < 0.001$, $p < 0.005$ and $p < 0.01$, respectively. Numbers of differentially expressed genes using each of the p-value cutoffs are considerably higher than for weight loss and SMI as shown in Table 4-3. Weight loss and SMI had relatively weak signatures (few differentially expressed genes); these phenotypic characteristics had 10-fold to 20-fold fewer differentially expressed genes than observed for muscle attenuation. Thus, we decided to focus on muscle attenuation classification for our pathway analysis and further discussion. Of differentially expressed genes according to weight loss and SMI, only 10-14% overlapped with genes differentially expressed according to muscle attenuation at $p < 0.01$. Because patients in the low muscle attenuation class were significantly older than those in high muscle attenuation class we conducted t-test analysis using age as a phenotype after using the same extreme phenotype classification method used for muscle attenuation, SMI and weight loss. We identified 1404 differentially expressed transcripts at $p < 0.01$ according to age, however, only 3% of these transcripts overlapped with differentially expressed transcripts according to muscle attenuation at $p < 0.01$ (data not shown).

4.3.3 Pathway analysis

Of 2715 differentially expressed transcripts according to muscle attenuation, 390 were not mapped by IPA. The 2325 transcripts that were mapped constituted 2104 unique genes. Pathway analysis provided a starting point for interpreting differentially expressed genes by identifying canonical pathways found to be represented based on the 2104 genes. Table 4-4 shows the list of identified canonical pathways. The list in Table 4-4 is organized according to categories: inflammation, degradation, apoptosis, growth and proliferation, transcription, ATP production, lipid metabolism, intracellular structure and vesicle transport, cellular adhesion and extracellular matrix and motor unit. About 70% of the canonical pathways listed belong to the inflammation, growth and proliferation or ATP production categories. It is important to note that many of the canonical pathways on Table 4-4 were identified based on common molecules located downstream in signalling cascades. For example, under the category *Growth and Proliferation*, IPA identified insulin signalling, insulin-like growth factor 1 (IGF-1) signalling and mTOR signalling. These three canonical pathways have many common molecules in their signalling cascade (e.g. phosphoinositide-3-kinase (PI3K), mammalian target of rapamycin (mTOR) and 3-phosphoinositide dependent protein kinase-1 (PDPK1)). These common downstream molecules were differentially expressed and largely responsible for the identification of these canonical pathways by IPA. By contrast, cell surface receptors that activate these signalling cascades, insulin receptors, and IGF-1 receptors) were not differentially expressed. Thus, canonical pathways identified by IPA were used to supplement

the more thorough independent study of differentially expressed genes assigned to each category which is described below.

We were able identify a function for 37% of the differentially expressed genes with a fold change ≥ 1.5 and ≤ -1.5 by conducting an extensive review of the literature. Many genes had limited published information regarding their function. We grouped differentially expressed into different categories (Tables 4-5 to 4-19), as we did for canonical pathways in Table 4-4. We acknowledge that some genes have diverse functions and therefore could be assigned to more than one category. However, for the purpose of presenting the results, each gene was limited to one category. In Tables 4-5 to 4-19 (discussed below) the negative and positive fold change values indicate genes are downregulated or upregulated in low attenuation class, respectively.

4.3.4 Inflammation

Downregulated genes indicate decreased activation of nuclear factor kappa B (NFκB) coupled with chemoattraction of lymphocytes (Table 4-5). NFκB is a downstream regulator of pro-inflammatory cytokines. Downregulated TNIP2 and ZFP91 encode genes which positively regulate NFκB-dependent transcription (50, 51). MTPN, which encodes myotrophin, was also downregulated. Myotrophin is thought to promote hypertrophy, at least partly, through the activation of NFκB (52). Toll-like receptor 4 (TLR4) is integral to innate immunity and is also known to activate NFκB (53) was downregulated. A commonality with other downregulated genes is their involvement in chemoattraction (IL16, IL23A,

CCR1 and CCRL1) (54-56) and activation (ZFPM1) (57) of lymphocytes, particularly CD4 cells.

Upregulated genes suggest involvement of the complement system and immune cell – muscle cell interaction (Table 4-6). A marked increase in mRNA levels for complement system proteins was identified; C2, CF1, C1R, C1S, CFH and C7 were all upregulated. Complement activation also has been shown to contribute to inflammation after skeletal muscle ischemia/reperfusion (58). Upregulation of SPON2, essential to the initiation of the innate immune response / inflammatory cell recruitment (59, 60), was also identified. IL21R (interleukin 21 receptor) was also upregulated. IL21R is involved in regulating immunoglobulin expression (61); we did identify a marked upregulation in immunoglobulin expression (IGK@, IGHG4, IGKC and IGLC1 were all upregulated). CD28 was also upregulated as has been shown in muscle in response to stimulation with pro-inflammatory cytokines IL-1α and IFN-γ compared to unstimulated muscle (62). CD80 and CD86 are ligands for CD28 and are expressed by macrophages, monocytes, dendritic cells, B cells, and activated T cells (62) suggesting a role in immune cell recruitment.

4.3.5 Degradation and cell death

A list of downregulated genes associated with the Ub-proteasome pathway is shown in Table 4-7. Downregulated genes included UBE2G2, UBE2N and UBE2K which encode Ub-conjugating enzymes. Ub-conjugating enzymes encoded by these genes suggest decreased lipid droplet degradation (UBE2G2) (63), decreased growth hormone receptor endocytosis (UBE2N) (63) and

decreased suppression of apoptosis (UBE2K) (64). We also identified several downregulated genes encoding Ub protein ligases and members of the F-box protein family which form a complex with Skp and Cullin to form the Skp, Cullin, F-box containing complex (SCF complex) which acts as a multi-protein ubiquitin ligase complex. Expression of Ub ligase encoded by HERC5 has been shown to be induced by pro-inflammatory cytokines IL-1β and TNFα via NF-kB (65) which is consistent with the decreased activation of NF-kB suggested by differentially expressed genes under the *Inflammation* category. The remaining downregulated genes encoding Ub-ligases suggest increased cell cycle progression and stem cell renewal (PJA1 and FBXO4) (66, 67), decreased proliferation (UBE3A and FBXL17) (68, 69) and increased motility (RNF5) (70). RING finger proteins play a critical role in mediating the transfer of Ub to the protein targeted for degradation; RING finger proteins RC3H1 and RNF31 were both downregulated. Finally, we identified PSMB7 to be downregulated. PSMB7 encodes a subunit of the proteasome and downregulation of this gene has been correlated with inflammation (71).

A list of upregulated genes associated with the Ub-proteasome pathway is shown in Table 4-8. Among upregulated genes was UBE2Z which encodes a Ub-conjugating enzyme which has not been extensively studied but is not generally highly expressed in skeletal muscle (72). BTRC, FBXO32, FBXL20 and WWP1 were among the upregulated genes encoding Ub protein ligases. BTRC encodes a ligase involved in negatively regulating vascular endothelial growth factor (VEGF) receptor 2 accumulation (73), negatively regulating anti-inflammatory

interleukin 10 receptor accumulation (74). FBXO32 encodes a muscle-specific ubiquitin ligase known in the literature as F-box protein 32 or Atrogin-1 or muscle atrophy F-box (MAFbx). MAFbx is known to be induced in catabolic states including cancer cachexia (5) and is thought to be involved in decreasing rates of protein synthesis by targeting the myogenic differentiation factor MyoD for degradation and inhibiting differentiation of myoblasts (75). FBXL20 encodes an F-box protein with unknown targets. Finally, WWP1 encodes a ubiquitin ligase that negatively regulates TGF-β signaling in cooperation with Smad7 (76).

The bottom of Tables 4-7 and 4-8 list genes associated with lysosomal protein degradation. TRAK1 was downregulated and encodes trafficking protein, kinesin binding 1. Knockdown of TRAK1 has recently been found to result in decreased degradation of internalized epidermal growth factor receptors through a block in endosome-to-lysosome trafficking in HeLa cells (77). Genes encoding lysosomal cysteine proteases cathepsin S and cathepsin C were upregulated. Under inflammatory conditions, cathepsin S is involved in controlling the level of class II Major histocompatibility complex-antigens on the surface of muscle cells (78). Cathepsin C plays a key role in neutrophil-dominated inflammatory diseases by activating neutrophil-derived serine proteases which can lead to tissue damage and chronic inflammation (79). Together differential expression of genes encoding lysosomal degradation-related proteins suggest increased promotion of an inflammatory state.

Genes encoding both apoptosis and autophagy were differentially expressed (Table 4-9). With the exception of BAX which encodes the apoptotic

activator BCL2-associaed X protein, all other downregulated genes identified encode inhibitors of apoptosis (DKK3, BBC3, SATB2, CARD16 and BIRC5 (80-84)). Upregulated genes included mostly activators of apoptosis (APAF1, DIABLO, EDARADD and PPIF). BFAR was also upregulated, this gene encodes bifunctional apoptosis regulator and has been found to associate with Bcl-2 in endoplasmic reticulum membranes and inhibits apoptosis (85). Differential expression of apoptosis-related genes suggests a decrease in apoptosis inhibition coupled by an increase in apoptosis activation. We identified upregulated but not downregulated autophagy-associated genes. These included ATG16L1, ATG4D, BECN1 and ATG7 which encode proteins involved in promoting autophagy. In summary, these genes suggest increased apoptosis and autophagy in low attenuation muscle.

4.3.6 Growth

Differentially expressed genes involved in growth and proliferation were further subcategorized in Table 4-10. Taken together, differential expression in this category supports increased differentiation, cell cycle progression and cytoskeleton reorganization and decreased mTOR associated signalling.

Under the subheading *Growth factors,* GFER was downregulated. This gene encodes the augmenter of liver regeneration protein which supports cell proliferation by acting as an anti-apoptotic factor and protecting cells from oxidative damage (86). Growth factor receptors IGF2R and midkine were upregulated and both are upregulated during muscle regeneration promoting muscle differentiation and growth (87, 88). Though IGF2R encodes an insulin-

like growth factor receptor it is structurally unrelated to either insulin like growth factor 1 receptor (IGF1R) or insulin receptor which are known to be decreased during cancer-associated atrophy (89).

Differential expression under the transforming growth factor (TGF) family subheading shows downregulation of two genes encoding bone morphogenetic proteins (BMP). BMP signalling plays a critical role in coordinating differentiation of satellite cells from a proliferative to committed state during regeneration (90). Specifically, BMPs induce satellite cell commitment to the differentiation program and accelerate differentiation into fused myotubes (90). The most upregulated TGF family gene was TGFB2 (transforming growth factorβ2). Transforming growth factorβ2 is involved in controlling differentiation of adult skeletal myoblasts (91).

Decrease in expression of Ras and Rho GTPases has been observed in muscle atrophy (92). However, we observed downregulation of some Rho/RasGTPases and upregulation of others. Of the downregulated genes, HRAS is the most well studied in skeletal muscle; specifically, the H-Ras protein encoded by this gene is known to inhibit myogenesis and differentiation (93, 94). Of the upregulated genes, RAC1 and RHOQ are both involved in glucose uptake and actin reorganisation in response to insulin signalling (95, 96). Moreover, RHOQ is also involved in myofibril organization during differentiation (97).

Genes encoding non-Ras/Rho growth factor signalling proteins are also listed in Table 4-10 under the heading *Growth factor downstream regulators*. Most of these were downregulated and are associated with the mTOR complex;

MTOR itself, as well as ICK, PIK3C2A and PLD1 involved in mTOR activation and belong to the well known PI3K/AKT/mTOR signalling cascade often downregulated in cancer cachexia models (98-100). MKNK2, encoding MAP kinase interacting serine/threonine kinase 2, is a negative downstream regulator of the elongation factor eIF4E which is downstream of mTOR (101) and was also downregulated. HSPA8 which encodes heat shock cognate 70 (Hsc70) was also downregulated. HSPA8 is known to be downregulated in diabetes and its expression responds to insulin administration (102). Upregulated genes under this subheading were MAPKAP1 and MAP2K6. MAPKAP1 encodes mitogen-activated protein kinase associated protein 1 which is important for mTORC2-mediated phosphorylation (mTORC2 is a regulator of the cytoskeleton) (103). MAP2K6 is a downstream regulator of RAC1 signalling in response to TGF signalling.

The estrogen receptor gene (ESR1, also known as ERα) was upregulated. ERα signalling is closely linked to insulin sensitivity and insulin metabolic signalling in skeletal muscle (104). WDR77 which encodes a steroid receptor coactivator that enhances androgen receptor and estrogen receptor-mediated transcriptional activity (105) was downregulated. Activated androgen receptor acts as DNA-binding transcription factor by binding to androgen response elements (ARE) on the DNA (106). Many differentially expressed gene have AREs (denoted by the ‡symbol in Tables 4-4 to 4-19) and are responsive to androgen signalling (106). We identified slightly more upregulated genes with AREs (n=75) compared to downregulated genes with AREs (n=65).

The *Wnt signalling pathway* category was interesting since we identified upregulation of the receptor as well as many of the downstream components of this pathway (FZD1, SFRP2 and DVL1). This is suggestive of an increase in Wnt signalling. Wnt signalling is involved in altering cell cycle progression, stem cell fate, growth and may modulate insulin resistance in muscle (107-109).

We identified differential expression of genes encoding proteins involved in *cell cycle regulation*, such as cyclins. Downregulated genes under this subheading include genes which inhibit cell cycle progression promoting senescence (e.g. CDKN2A (110)) and genes that promote cell cycle progression (e.g. CENP-Fand FBXO11 (111, 112)). With the exception of TOB2 (113), upregulated genes included genes involved in promoting cell cycle progression at different parts of the cell cycle.

Lastly, we included a miscellaneous subheading labelled *Other growth/proliferation related genes*. Downregulated genes under this subheading included genes involved in promoting proliferation (e.g. TSPAN1 (114)) and myogenic differentiation (e.g. AVP (115)). Upregulated genes included genes involved in muscle development (e.g. SHOX2 (116) and PROX1 (117)), myogenic differentiation (e.g. RBM38 (118) and MEF2D (119)) and growth (e.g. STAT5B (120)).

4.3.7 Transcription and translation

Many genes encoding transcription factors, transcription regulators, ribosomal subunits and translation factors were differentially expressed. However,

no evidence was obtained for the simplistic view that the low muscle attenuation class had general repression of muscle gene expression (Tables 4-11 and 4-12).

Downregulated genes included transcription factors such as ASCL2, USF2, MAFA and TFEC; ASCL2, USF2 and TFEC. These genes are associated with promoting transcription of genes that ultimately result in growth (121, 122). PARP10, which encodes a protein that interacts with the transcription factor myc, was also downregulated. There were also downregulated transcription coactivators (e.g. CTRTC3 (123) and TAF10 (124)). In addition to the above, there was a marked downregulation of ribosomal components.

Upregulated genes included POLR3B encoding a polymerase III which synthesizes ribosomal RNA and BRF2 encodes a protein that is part of the RNA polymerase III transcription factor complex. Upregulated POLR2J2 and POLR2K encode components of polymerase II which synthesize messenger RNA. Transcription factors ATF4, RUNX2 and GTF2H5 were upregulated. ATF4 is involved in increasing amino acid transporter expression during amino acid deprivation; this is thought to be a mechanism of increasing amino acid transport out of the myocyte for survival (125). RUNX2 stimulates transdifferentiation of satellite cells into mineralizing osteoblastic type cells (126) and there is evidence to suggest that GTF2H5 may be involved in DNA repair (127).

4.3.8 ATP production and mitochondrial function

Most differentially expressed genes in this category were upregulated (Tables 4-13 and 4-14). Downregulated genes included genes encoding proteins involved in glucose metabolism (FBP1, HK1, and GPD2) and the tricarboxylic

acid cycle (IDH3A). FBP1 and IDH3A have been documented to be downregulated during inflammation (128, 129) and loss of HK1 was recently shown to increase TNF-dependent cell death by modifying caspase-driven apoptosis (130). Genes encoding complex I (ND4, ND3 and ND2) and ATP synthase (ATP5G2) of the electron transport chain were also downregulated. Notably, the downregulated complex I genes are transcribed from mitochondrial DNA; no complex I genes transcribed from nuclear DNA were downregulated. Upregulated genes included mostly genes encoding proteins of the electron transport chain (NADH dehydrogenase complex components, cytochrome c oxidase subunits, succinate dehydrogenase complex subunits and ATP synthase complex subunit). The most upregulated gene was THRB which encodes thyroid hormone receptor β. Thyroid hormone receptor β is preferentially expressed in the liver under normal healthy conditions where it plays a major role in cholesterol and lipoprotein metabolism (131). In skeletal muscle, knockout of THRB in rodents results in decreased fatigue resistance compared to wild-type controls (132) suggesting that in skeletal muscle this hormone receptor is involved in energy metabolism. Together, the aforementioned differentially expressed genes suggest decreased glycolysis and disrupted respiratory ATP synthesis. However, the mix of upregulated and downregulated genes encoding proteins participating in the electron transport chain it is not clear if ATP synthesis is increased or decreased.

There were a few genes downregulated encoding proteins involved in protecting cells from oxidative stress (TXNRD1, GSTA2, SOD3, and PRDX2).

TXNRD1 and GSTA2 are known to be downregulated in response to hypoxia and glucocorticoid treatment, respectively (133, 134). Glucocorticoid signalling was identified by pathway analysis and has been associated with cancer cachexia (135). SOD encodes superoxide dismutase 3 which inhibits reactive oxygen species-induced trafficking of toll-like receptor 4 signalling (136) suggesting that its downregulation may be associated with increased toll-like receptor 4 inflammatory signalling. Finally, downregulation of PRDX2 is known to contribute to angiotensin II-mediated apoptosis in the kidneys (137). Together downregulated genes suggest increased increased glucocorticoid signalling increased inflammation via toll-like receptor signalling and increased oxidative stress that likely makes muscle cells more prone to apoptosis. There was also upregulation of genes involved in protecting cells from oxidative stress (TXNRD3, GSTK1, OXR1, RDH13 and PINK1). OXR1, RDH13 and PINK1 encode proteins which act on the mitochondria to protect from oxidative stress (138-140). GSTK1 encodes a protein suggested to have a role in the detoxication of lipid peroxides generated in peroxisomes (141). Upregulated genes suggest increased protection of oxidative stress in the mitochondria and from reactive oxygen species generated during lipid metabolism.

In line with the decreased expression of NADH dehydrogenase complex subunits from mitochondrial DNA, we identified decreased expression of the mitochondrial RNA polymerase encoded by POLRMT (Table 4-15). We also identified decreased expression of the translocase of the outer membrane (TOM) complex subunit encoded by TOMM7 and the translocase of the inner membrane

(TIM) complex protein encoded by TIMM17B. The TOM and TIM complexes are required to import mitochondrial proteins transcribed from nuclear DNA (142). A decrease in protein import has been documented during chronic disuse atrophy (142). We identified both downregulated and upregulated mitochondrial ribosomal proteins; it is unclear if this is associated with increased or decreased ribosomal function. However, there was a marked upregulation in aminoacyl tRNA synthetases HARS2, TARS2, PARS2 and FARS2. Aminoacyl tRNA synthetases catalyze the ligation of amino acids to their cognate tRNAs. Despite this common task phenotypes caused by mutant alleles in different aminoacyl tRNA synthetases vary in tissue specificity and clinical presentation (143) suggesting that these upregulated genes may be involved in other roles in addition to mitochondrial translation that are still unknown.

4.3.9 Lipid metabolism

There were fewer downregulated genes than upregulated genes in this category (Table 4-16). Downregulated genes included LEPR which encodes the leptin receptor. In skeletal muscle leptin reduces lipid accumulation, increases lipid oxidation and promotes hypertrophy (144-146). Hypertrophy occurs partly via activation of the PI3K/AKT signalling pathway (147) which, as previously mentioned, also appears to be downregulated. FAR2, which encodes a fatty acyl CoA reductase that converts saturated C16 and C18 fatty acids into fatty alcohols (148), was also markedly downregulated. Lastly, APOOL and SLC37A3 were also downregulated, however little is known about the function of these two genes. Together downregulated genes suggest increased lipid accumulation.

Among upregulated genes were ADIPOR2, LPIN2 and PPARA which are connected in lipid metabolism pathways. Adiponectin signalling is involved in peroxisome proliferator-activated receptor α (PPARα) activation (149). Lipins, such as lipin 2 encoded by LPIN2, also interact with PPARα to regulate transcriptional activity (150) and are associated with lipid accumulation (151). Finally, we identified upregulation of lipid droplet associated proteins (BSCL2 and PLIN5) which also suggest increased intracellular lipid storage (152, 153). Like downregulated genes, these upregulated lipid-related genes also suggest increased lipid accumulation in low attenuation muscle.

4.3.10 Intra- and extra- cellular structure components

Most genes encoding proteins involved in intracellular structure and vesicle transport were downregulated (Table 4-17). Taken together differentially expressed proteins in this category suggest decreased maintenance of cellular structure and increased intracellular reorganization. Downregulated genes included genes involved in vesicle and organelle trafficking (e.g. KIF23, CLINT, MYO5C and MYOF) (154-157), maintaining cellular structure (e.g. PKP3, CEP68 and TUBA3C) (158, 159) and structural changes required for cell division (e.g. BRK1 and ARHGAP33) (160, 161). Upregulated genes included SYNPO which encodes synaptopodin. Synaptopodin regulates the actin-bundling activity and may be required for cell movement as it is involved in actin cytoskeleton plasticity (162). JAKMIP2 and JAKMIP3 were also upregulated; these encode

proteins that regulate membrane traffic events occurring in the Golgi apparatus (163, 164).

Among downregulated ECM genes were genes encoding cadherins and laminins, and ECM intracellular signalling components catenins and calmodulins. These are all components of adherens junctions which connect the basement membrane to the intracellular space (165). Many of these downregulated genes are also associated with the neuromuscular junction (e.g. CASK (166), LAMA4 (167), CTTN (168) and CTNND2 (169)). Among other downregulated ECM genes are others that encode proteins also involved in adhesion (LPHN1, TGFBI, HEPACAM, and LGALS3). Upregulated ECM genes included ECM genes typically upregulated during fibrosis and ECM remodelling: COL8A1, COL6A3, FN1, FBN2, FAP and MMP2 (170-172). Notably, TGFB2 (introduced in section 4.3.6 above) is known to be upregulated in injured muscle and involved in fibrosis (173). Other upregulated ECM genes are involved in promoting cell-cell interactions (e.g. NRCAM, MTSS1, VCAM1, MYO9A, VCAN and FBLN5 (174-176)).

4.3.11 Contractile and neuromuscular junction components

Differential expression under this category suggests disrupted actin stabilization and also disruption of the neuromuscular junction. Only NEBL encoding nebulette, involved in stabilizing actin (177), was downregulated in this category (Table 4-19). Upregulation of MYL4 encoding myosin light chain 4 may be indicative of proliferation of myoprogenitors (178). TNNT2 which encodes a troponin molecule and SNTA1 which encodes syntrophin α 1 were also

upregulated. Syntrophin α 1 plays a role in decreasing Ca2+ in muscles in the presence of laminin, linking the ECM with the actin cytoskeleton and membrane stabilization/contraction by interacting with the dystrophin-associated protein complex (179).

Decreased expression of extracellular matrix components associated with the NMJ (at the synaptic basal lamina) was mentioned above. The bottom part of Table 4-19 includes mostly downregulated genes encoding proteins of the postsynaptic terminal such as .CHRNB4, GRIN2D and DRD4 which encode subunits of the acethylcholine, glutamate and dopamine receptors. SYN1 was also downregulated; since this gene is only known to be expressed by neurons (180), it is unclear what function it plays in skeletal muscle.

4.4 Discussion

4.4.1 Muscle characteristics during cancer

Here we present a relatively large dataset of muscle microarray that allowed us to investigate molecular differences related to skeletal muscle index, muscle attenuation and weight loss. This explorative study indicates that among the three aforementioned phenotypes, muscle attenuation gives the strongest molecular signature. Aberrant (low) muscle attenuation is a poor prognostic factor for cancer patients, and this feature has a very strong signature of differential gene expression. Though patients in the two muscle attenuation classes (high and low muscle attenuation) were significantly different according to age, differential analysis according to age results in a very different list of differentially expressed genes. This suggests that differential expression according to muscle attenuation

is distinct and specific to extremes in muscle attenuation phenotype. While weight loss has to date been the main phenotypic characteristic used in clinical classification of patients as cachectic or not, our findings make it very clear that weight loss has a trivial differential gene expression signature compared with skeletal muscle attenuation. The fact that patients with the very highest degrees of weight loss did not have significantly lower muscle index or significantly lower muscle attenuation than those with minimal weight loss, makes it clear that weight change overall does not represent changes in skeletal muscle characteristics. This is perhaps not surprising, given that body weight loss can easily include alterations in fat mass, or the mass of any other organ or body constituent.

It is also notable that skeletal muscle index (i.e. the absolute muscularity of the men in our population) showed comparatively little gene expression signature being the least of the 3 phenotypic characteristics considered. In this regard, it should be noted that the stable skeletal muscle mass of normal healthy adults shows a considerable degree of inherent variation (181, 182), and this variation may not necessarily be anticipated to be associated with strong variation in gene expression as it does not concern pathological variation. While some cancer patients may be losing or may have lost a good deal of muscle, at least some patients with low muscle mass had low muscle mass before the onset of illness.

Our findings of a strong gene expression signature associated with low muscle attenuation underscores a number of recent observations related to this characteristic of muscle. Reduced radiation attenuation of muscle is a relatively newly characterized and distinctive abnormality. The generally accepted boundary

of low attenuation muscle is <30 HU (183-185); nearly all of the patients in our low muscle attenuation class had values below this boundary. The muscle attenuation range in this study was comparable to a large cancer population (n=1473) from the Cross Cancer Institute (186) indicating that we had a representative sample. Low attenuation is related to deficits in physical functioning, altered metabolism and poor prognosis. Decreased skeletal muscle attenuation was linked with a decrease in muscle strength and performance (187-189) and increased risk of hip fracture (190). Reduced muscle attenuation has been associated with accumulation of lipid (43, 191). Skeletal muscles normally contain only small amounts of fat used as a source of energy during aerobic work. Excess infiltration of fat has emerged as an important factor associated with insulin resistance and Type II diabetes (192-194). Recent reports suggest that low muscle attenuation is common in cancer, positively correlated with wasting and an independent prognostic of poor survival in cancer patients (21, 23, 186, 195). This is consistent with relationships between low SMI and low attenuation in our study. Despite this, it is not in the usual repertoire of oncologic radiology to report quantifiable dimensions of muscles such as cross-sectional area or attenuation. There may be merit in quantifying skeletal muscle of cancer patients at standard vertebral landmarks, with a view to identify individuals affected by muscle wasting and altered attenuation.

4.4.2 Comparison with prior array studies of skeletal muscle during atrophy

Analyses of molecular events occurring during cancer cachexia have, for the most part, been limited to looking at a few genes or proteins at a time. Lecker

et al. were the first to investigate skeletal muscle changes during atrophy using gene arrays (18). This seminal work aimed to identify a common transcriptional program active during atrophy regardless of the cause (cancer, diabetes, renal failure and starvation). Only differentially expressed genes in all four states were included as part of the atrophy program (18). This approach designed to detect common downstream pathways, but deliberately excluded any gene expression events not common to all 4 states. For example, inflammation-related genes may have been differentially expressed in cancer but not in starvation and thus not included. Focusing only on genes differentially expressed in all four states also made it more likely to select genes encoding proteins located downstream in signalling cascades. Compared to Lecker *et al.* we identified more genes situated upstream of the final common transcriptional events. In common with Lecker *et al.* our analysis detected the pathways of protein degradation, ATP production and substrate metabolism, transcription factors, extracellular matrix proteins, translational control, oxidative stress and growth and differentiation (18). Most importantly, our findings within these categories agree with the findings from Lecker *et al.*. Specifically, we both showed increased expression of Ub-proteasome pathway proteins involved in promoting skeletal muscle atrophy, decreased glycolysis, suppressed of cell growth via the mTOR pathway and changes in the expression of extracellular matrix proteins. Our analysis focused only in patients with cancer which is likely why we also identified a large number of differentially expressed genes suggesting increased inflammation as well as genes indicating disrupted lipid metabolism and mitochondrial function. Our list

of categories was more extensive than that of Lecker *et al.* but this is to be expected considering the study design (i.e. genes had to be differentially expressed in all disease states) and the fact that they conducted their study nearly a decade ago with 10K arrays which is limited in comparison to the 41K arrays used in our study. The recently conducted study by Braun *et al.* explored gene expression changes in response to central IL-1β injection in mice and found many of the same altered pathways as Lecker *et al.* despite being a different model (19). Braun *et al.* further strengthened the presence of a common transcriptional program but also introduced differentially expressed genes in skeletal muscle in the presence of an inflammatory response. This is important as inflammation is thought to be an important driver of skeletal muscle change in cancer (196). Braun *et al.* identified differentially expressed genes related to inflammatory signalling, protein degradation, growth factor signalling, nutrient signalling, intermediary metabolism, adipokine signalling, oxidative stress, myogenic differentiation and extracellular matrix (19). In our study we also identified differentially expressed genes in these categories. Many of the differentially expressed genes presented by Braun *et al*. showed transient gene expression following acute cell stimulation. For example, genes encoding p-selectin and NFκB showed a marked upregulation immediately (2 h) after acute IL-1β injection followed by a transient decrease in expression and ultimately downregulation at 8 h post injection (19). In our study, genes encoding p-selectin and NFκB were downregulated which may be reflective of the chronic state of cancer patients. Transient gene expression following acute cell stimulation might not be comparable to gene expression changes during

chronic illness. The sampled human cases in this study are from a fixed time point (along a disease trajectory). Regardless of limitations associated with the aforementioned rodent studies, similarities with our results from human muscle are encouraging and serve to suggest results from rodent muscle can be extended to human muscle, *albeit* in a restricted manner.

4.4.3 A molecular signature in muscle during cancer

Low muscle attenuation means that a lower than usual proportion of muscle cross sectional area had normal attenuation values (i.e. < 30 HU). Patients with low attenuation also had significantly reduced overall muscle cross sectional area (SMI) compared with normal attenuation individuals. While low attenuation was coupled with significantly reduced in cross sectional area, in addition it had altered patterns of inflammation, growth / proliferation, cellular architecture and energy metabolism, based on differential gene expression. Muscle attenuation variation appears to be the result of multiple interacting pathways which are also altered during muscle atrophy (Figure 4-1). While pathway analysis revealed many different possible cytokine- and hormone-activated signalling pathways (many of which shown in Figure 4-1 part A), we did not identify many genes encoding cytokines/hormones or cytokine/hormone receptors. Most of the differentially expressed genes associated with ligand-activated signalling pathways (e.g. IL-1, IL-8, IL-2 and insulin signalling pathways) encode proteins which interact with ligand receptors (upstream) or are downstream in the signalling cascades shown in Figure 4-1 part B. Pathway analysis was used to supplement the more detailed study of individual differentially expressed genes

within each category. Below we will focus on a general view of how differentially expressed genes identified in the present study may work in concert to result in the low muscle attenuation phenotype.

4.4.3.1 Molecular signal indicates altered energy metabolism

Skeletal muscle plays a major role in whole body energy metabolism; it is the main site for glucose disposal and fatty acid oxidation. Diabetes and insulin resistance, in the absence of cancer, are the main clinical conditions associated with increased fat infiltration in muscle that have been studied to date (44). Most studies looking at increased fat infiltration focus solely on pathways associated with lipid synthesis and storage (44). All the lipid metabolism-related differentially expressed genes identified in our study support increased lipid accumulation. Upregulation of genes encoding lipid droplet associated proteins such as perilipin 5 and upregulation of adiponectin receptor, lipin and PPARα all suggest altered fat metabolism promoting lipid storage. Promotion of lipid storage is expected in low attenuation muscle since low attenuation is associated with increased lipid accumulation. PPARα is a major regulator of lipid metabolism involved in the transport proteins to facilitate fatty acid uptake , acyl CoA synthases, fatty acid-binding proteins that facilitate delivery of fatty acids to cellular compartments, mitochondrial carnitine system and fatty acid oxidation (197). We suggest that the role of adipokines, adipokine receptors and their downstream signalling mediators (PPARs) play a role during cancer-associated the fatty infiltration into muscle.

Altered muscle energy metabolism has been documented in muscle during cancer with a phenotype akin to metabolic syndrome, specifically insulin resistance (198). Thus, differential expression of genes encoding proteins involved in energy metabolism was expected. Our results suggested mitochondrial dysfunction. Disrupted mitochondrial function has been associated with increased lipid content in the muscle of elderly subjects (199, 200) and more extensively studied in the liver of patients with nonalcoholic fatty liver disease (NAFLD) which characterized by fat accumulation (201). Mitochondrial abnormalities associated with NAFLD include depletion of mitochondrial DNA which can result from increases in reactive oxygen species, decreased activity of respiratory chain complexes, and impaired mitochondrial β-oxidation (201). In our study of skeletal muscle, differential expression analysis did identify upregulation of nuclear NADH dehydrogenase (complex I) genes, complex II genes and complex IV genes and downregulation of mitochondrial NADH dehydrogenase (complex I) genes. However, it is not clear if the activity of the different complexes, known to be strong correlates of muscle oxidative capacity (202), was altered. Mitochondrial dysfunction, including decreased oxidative phosphorylation capacity and disrupted mitochondrial dynamics, has been reported in response to systemic inflammation and skeletal muscle wasting (203) and so further analysis of mitochondrial function (e.g. assessing respiration rate and mitochondrial membrane potential) is warranted in muscles of cancer patients.

4.4.3.2 Molecular signal parallels aberrant repair

Muscle damage can occur in response to disease, exposure to myotoxic agents, sharp or blunt trauma, ischemia, exposure to hot or cold temperatures and exercise. Regardless of the cause the process of repair follows a common set of events summarized by two stages, degeneration and regeneration (204). These well-orchestrated and interrelated phases of repair involve interactions between muscle cells, immune cells, progenitor cells (e.g. satellite cells), fibroblasts, neurons, adipocytes, and endothelial cells (204). Degeneration involves cell death, disruption of the myofiber sarcolemma and thus increased myofiber permeability, infiltration of pro-inflammatory immune cells, phagocytosis of cellular debris by macrophages and activation of progenitor cells (204). Regeneration involves activation of cell proliferation, growth and differentiation pathways to replace damaged tissue. Aberrant repair may be defined by deregulation of degeneration and/or regeneration processes such that muscle is incapable of recovering from insult or injury. Abarrant repair has been documented in chronic muscle diseases and conditions including muscular dystrophies, aging and myositis and is characterized by persistent inflammation, degeneration of myofibers and fibrosis (205). Based on differential expression it appears that low attenuation muscle is in a state of degeneration and regeneration similar to aberrant repair often seen in chronic tissue damage (205).

Differential expression of molecules involved in immune cell-muscle cell interaction suggests immune cell recruitment and activation. These are hallmarks of aberrant repair (205). This is evidenced by the increased expression of

immunoglobulins and adhesion molecules typically involved in immune cell extravasation (e.g. VCAM-1 and SPON2). P-selectin, a molecule important in the initial recruitment of leukocytes, was recently found to be markedly upregulated in muscle immediately after (2 h) inflammation stimuli followed by a decrease in expression 8 h after inflammation stimulation (19). The gene encoding p-selectin was differentially expressed (p=1.67E-03) but had a very small fold change (FC= -1.1) in our study. This was expected since p-selectin shows an acute and transient early increase in expression; we do not expect its expression to remain increased during chronic inflammation. Increased immunoglobulin expression was recently identified in muscle from patients with myositis and associated with B cell infiltration (206, 207). B cells activate the complement system which is consistent with the upregulation of complement genes observed here. Differential expression was also suggestive of a decrease in T cell associated proteins. It is unclear what populations of immune cells are present in patients with cancer-associated changes. However, based on the information from the regeneration literature, it is clear that the identity of immune cells associated with muscle can lead to altered metabolism and cell survival (208).

Persistent inflammation during aberrant repair is associated with protein degradation (208). Consistent with findings from Lecker *et al.* which looked at the atrophy program in animals (18) and numerous other studies on humans (Table 4-1), our findings suggest involvement of the Ub-proteasome pathway in muscle during cancer. However, unlike Lecker *et al.* not all differentially expressed genes associated with the Ub-proteasome pathway were upregulated. We identified

upregulated Ub-proteasome pathway genes that play roles in decreasing growth factor signalling, decreasing anti-inflammatory cytokine signalling and decrease myogenic differentiation. Decreased myogenic differentiation is suggested by the increase in FBXO32 which encodes one of the most well studied Ub ligases in skeletal muscle wasting conditions. In addition, we also identified many downregulated Ub proteasome pathway genes involved in increased cell cycle progression and stem cell renewal, decreased proliferation and increased motility. Despite identifying different differentially expressed genes our Ub-proteasome pathway proteins support the general conceptual framework of cancer cachexia which indicates increased catabolism and decreased anabolism.

Differential expression of genes related with apoptosis and autophagy identified in our study agree with animal models which suggest these pathways are increased in cancer cachexia (32, 209). Apoptosis is a part of the degeneration process and its constant activation may be indicative of an aberrant repair-like state. Constant activation of factors promoting differentiation of progenitor cells is also indicative of aberrant repair. Of the different myogenic signalling pathways identified, Wnt signalling stood out since several genes encoding proteins of this pathway were upregulated. Wnt signalling can alter the adipogenic potential of myoblasts and is involved in the conversion of myogenic cells into non-myogenic fibrogenic cells (107, 210-212).

Remodelling of the extracellular matrix (ECM) is an important step in normal repair process. ECM remodelling facilitates immune cell infiltration, accommodates for new myocytes in response to inflammatory and growth factors

and guide formation of neuromuscular junctions (205). This remodelling involves systematic and timely breakdown and repair of the different ECM layers in muscle in order to avoid fibrosis (205). Based on the CT scans alone, it is not possible to determine if patients in the low attenuation class had fibrotic muscle, and further histological work is required. Based on the differential expression (upregulation of ECM components) we would suggest that fibrosis is a possibility in low attenuation muscle.

4.5 Conclusion

This explorative study is a first step to understanding the etiology of cancer-associated muscle changes patients with cancer. Future work is needed to examine how stage, medication, anti-cancer treatment and functional status affects muscle gene expression of patients with low and high muscle attenuation. Unfortunately, the current patient group was not large enough to reliable analyze the data to examine how the aforementioned factors may affect gene expression. Based on the analysis of the present patient group we suggest that the association between inflammation (particularly identifying the identity of infiltrating immune cells), fibrogenesis and lipid metabolism (particularly related to mitochondrial function and PPAR signalling) in cancer patients with varying muscle attenuation are good starting points for future validation and follow-up studies.

Tables

Table 4-1: Prior studies using human skeletal muscle from patients suffering from cancer-associated muscle pathology

Reference	Muscle biopsied	Phenotypes examined	Used to study	Major findings
Bossola et al. 2001 (ref #24)	Rectus abdominis	- Cancer, gastric, weight losing(n=11♂,9♀) - Controls: non-cancer, abdominal disease, weight stable (n=6♂,4♀))	Ub-proteasome pathway	- Ub mRNA was significantly cancer patients than controls. - Ub mRNA higher in cancer patients with higher disease stage and greater weight loss
Bossola et al. 2003 (ref #25)	Rectus abdominis	- Cancer, gastric, weight losing (n=14♂,9♀)) - Controls: non-cancer, undergoing surgery, weight stable (n=9♂,5♀)	Ub-proteasome pathway	- Proteasome proteolytic activities (CTL, chymotrypsin-like; TL, trypsin-like; PGP, peptidyl-glutamyl-peptidase) significantly increased in gastric cancer patients with respect to controls
DeJong et al. 2005 (ref #26)	Rectus abdominis	- Cancer, pancreatic, weight losing (n=15) - Controls: non-cancer, weight stable (n=11)	Ub-proteasome pathway and inflammation	- Weight loss in pancreatic cancer is associated with systemic inflammation and increased mRNA expression for ubiquitin but not uncoupling proteins in skeletal muscle.
Jagoe et al. 2002 (ref #27)	Latissimusd orsi	- Cancer, lung (n=27♂,9♀) - Controls: non-cancer (n=4♂,6♀)	Ub-proteasome pathway and lysosomalproteolytic pathway	- mRNA levels for cathepsin B (involved in lysosomal proteolysis), but not for components of the Ub-proteasome pathway, were higher in patients with cancer compared with controls. Among lung cancer patients, cathepsin B mRNA levels correlated with fat-free mass index and tumour stage
Op den Kamp et al. 2012 (ref #28)	Vastuslatera lis	- Cancer, lung with <10% weight loss (n=15♂,1♀) - Controls: healthy, weight stable (n=7♂,3♀)	Ub-proteasome pathway activity, muscle nuclear factor kappa B (NF-κB) expression and systemic inflammation	- Patients with weight loss showed increased plasma levels of C-reactive protein (CRP), soluble Tumor Necrosis Factor receptor 1 (sTNF-R1), fibrinogen and decreased levels of albumin - No changes in fat free body mass or skeletal muscle NF-κB and ubiquitin proteasome system activity were observed
Williams et	Rectus	- Cancer (n=4♂,2♀)	Ub-proteasome	- mRNA levels for ubiquitin and the 20S

al. 1999 (ref #29)	abdominis	- Controls: non-cancer, (n=5♂,1♀)	pathway	proteasome subunits were higher in muscle from patients with cancer than in muscle from control patients
Busquets et al. 2007 (ref #32)	Rectus abdominis	- Cancer, gastrointestinal, weight losing (n=16) - Controls: non-cancer, weight stable (n=11)	Apoptosis signalling	-Muscle from weight-losing cancer patients showed a significant increase in muscle DNA fragmentation compared with control participants. - Muscle from weight losing cancer patients had a decrease in MyoD (a myogenic factor) protein content
Aversa et al. 2012 (ref #31)	Rectus abdominis	- Cancer, lung or gastric, weight stable (n= 26♂,7♀) - Controls: non-cancer, age matched, undergoing surgery, weight stable (n= 8♂,8♀)	Myostatin signalling	- Myostatin signalling is altered in non-weight-losing cancer patients
Banduseela et al. 2007 (ref #30)	Tibialis anterior or deltoid	- Cancer, lung, muscle loss confirmed by MR images (n=1♂) - Controls: - Malnutrition, weight losing (n=1♀) - Acute quadriplegic myopathy (n=1♂) - Hereditary motor and sensory neuropathy (n=2♀) - Amyotrophic lateral sclerosis (n=1♂) - Non-cancer, healthy (n=2♂)	Muscle paralysis and myosin loss	- A significant preferential loss of the motor protein myosin together with a downregulation of protein synthesis at the transcriptional level was observed in the patient with cancer-associated muscle wasting
Pessina et al. 2010 (ref #33)	Rectus abdominis	- Cancer, gastric, weight losing (n=17♂,13♀) - Controls: non-cancer, age-matched, weight stable (n=5♂,3♀)	Genes involved in muscle regeneration	- The expression of the genes involved in muscle regeneration (Pax7, MyoD, necdin) was increased with respect to the controls. - There was no difference in Myf5 expression or of the neonatal isoform of Myosin Heavy Chain (nMHC) expression between patients and controls.
Ueyama et al. 1998 (ref # 214)	Quadriceps femoris	-Case study: 1 male patient with lymphoma	Muscle lymphoma	- Immunohistochemical analysis of the biopsied muscle and the subcutaneous tumor led to the final diagnosis of true histiocytic lymphoma (an extremely rare condition).
Weber et al.	Vastuslatera	- Cancer, with >20% weight loss (n=9♂,8♀)	Myoglobin plasma	- Cancer patients had decreased plasma

2007 (ref #20)	lis	- Controls: (n=14♂,13♀)	levels as a marker of muscle mass and fiber composition	myoglobin concentrations, maximal quadriceps muscle cross-sectional area as assessed by magnetic resonance imaging, body cell mass and maximal oxygen uptake (VO(2)max) compared to controls.
Weber et al. 2009 (ref #21)	Vastuslatera lis	- Cancer, >10% weight loss (n=10♂,9♀) - Controls: healthy, age-, gender-, and body-height-matched, weight stable (n=10♂,9♀)	Muscle fiber size and capillarization	- Cancer patients had lowerbody mass index, muscle cross sectional area, total fiber size and VO(2max) compared to controls. - Absolute strength of quadriceps muscle was reduced in cancer compared to controls but was identical when normalized on muscle cross sectional area
Zampieri et al. 2010 (ref #22)	Rectus abdominis	- Cancer, colorectal (n=10) - Controls: non-cancer (n=7)	Morphometric analyses, ATPase histochemistry and immunohistochemical analysis of makers of muscle denervation and injury-induced muscle regeneration	- Muscle from patients had a higher percentage of myofibers with internalized or central nuclei compared to controls. - In 30% of patients, small myofibers expressing the MHC-emb were identified - In 50% of patients, larger fibers positive for N-CAM were identified - Among the 10,000 analysed myofibers in control biopsies, no MHC-emb and N-CAM-positive muscle fibers were detected.
Shaw et al. 1991 (ref # 213)		- Cancer, weight stable (n=26) - Cancer, weigh losing (n=21) - Controls: non-cancer, weight stable (n=18)	Leucine kinetics to study protein metabolism at the whole-body and tissue level	- Patients losing weight had a significant elevationwhole-body protein catabolism compared with the other two groups based on. - Whole-body protein synthesis was also elevated (to a lesser extent) weight losing patients.
Gallagher et al. 2012 (ref #39)	Quadriceps femoris	- Cancer, gastrointestinal, weight losing (n=10♂,2♀) - Controls: healthy, weight stable (n=4♂,2♀)	Global mRNA expression of sequential human muscle biopsies	- Depression of muscle turnover in patients with cancer-associated weight loss
Stephens et al. 2010	Rectus abdominis	- Cancer, gastrointestinal, weight losing (n=12♂,6♀) - Controls: non-cancer, weight stable, undergoing surgery (n=2♂,1♀)	Global mRNA expression (microarray)	- CaMKIIbeta correlated positively with weight loss in all muscle groups and CaMKII protein levels were elevated in rectus abdominis in

(ref #38)				cancer patients - TIE1 was also positively associated with weight loss. - Candidates selected from the pre-clinical literature, including FOXO protein and ubiquitin E3 ligases, were not related to weight loss in this human clinical study.
Stephens et al. 2011 (ref #23)	Rectus abdominis	- Cancer, (n=11♂,8♀) - Controls: non-cancer, undergoing surgery (n=2♂,4♀)	Intramyocellular lipid droplets	- Compared with controls, cancer patients had increased lipid droplet number and diameter. - Mean lipid droplet count correlated positively with the severity of weight loss

Table 4-2: Demographics and anthropometrics of samples in Class 1 and Class 2 for skeletal muscle index (SMI), muscle attenuation and weight loss

	All samples	**Class 1 Low SMI**	**Class 2 High SMI**	**Class 1 Low muscle attenuation**	**Class 2 High muscle attenuation**	**Class 1 Weight stable**	**Class 2 Weight losing**
Total, n	69	18	18	18	18	18	18
Age, mean years ± SD	59 ± 13	63 ± 16	54 ± 14	68 ± 10	48 ± 13***	60 ± 11	60 ± 10
Muscle, mean ± SD							
Skeletal muscle surface area (cm^2)	162 ± 28	143 ± 10	191 ± 20***	155 ± 25	176 ± 23*	147 ± 28	158 ± 35
Skeletal muscle index (cm^2/m^2)	53 ± 8	**45 ± 2**	**57 ± 8*** **	49 ± 7	57 ± 8*	48 ± 8	54 ± 8
Muscle attenuation (HU)	36 ± 9	32 ± 9	40 ± 11*	**26 ± 5**	**47 ± 4*** **	30 ± 9	35 ± 9
Percent weight loss, %	-4 ± 7	-4 ± 7	-3 ± 5	-3 ± 5	-2 ± 5	**1 ± 3**	**-11 ± 7**
Diagnosis at surgery, n							
Cancer, liver or intrahepatic bile ducts	12	6	2	4	3	2	2
Cancer, gastrointestinal tract	35	10	6	9	8	13	8
Cancer, pancreas	13	2	7	4	6	1	3
Benign growth	9	1	3	1	1	2	4
Previous chemotherapy exposure, %	29	28	33	30	33	50	30

Asterisks denote p-values for t-test comparing high and low groups *$p \leq 0.05$, ***$p \leq 0.0001$

4-3: Number of differentially expressed transcripts for the three different classifiers (weight loss, skeletal muscle index (SMI) and muscle attenuation) at different p-value cutoffs

	p<0.00001	p<0.0001	p< 0.001	p<0.005	p<0.01
Weight loss	1	1	21	98	212
Weight loss with a fold change increases >1.5 or decrease < -1.5	1	1	11	51	93
Skeletal Muscle Index (SMI)	0	0	7	34	103
SMI with a fold change increases >1.5 or decrease < -1.5					
Muscle attenuation	5	52	440	1644	2715
Muscle attenuation with a fold change increases >1.5 or decrease < -1.5	1	25	196	821	1403

4-4: Canonical pathways associated with differentially expressed genes associated with muscle attenuation identified by Ingenuity Pathway Analysis

Ingenuity Canonical Pathways	**p-value**
Inflammation-related pathways	
IL-1 Signalling	2.24E-03
fMLP Signalling in Neutrophils	2.45E-03
HMGB1 Signalling	3.31E-03
LPS-stimulated MAPK Signalling	3.89E-03
Glucocorticoid Receptor Signalling	5.62E-03
IL-8 Signalling	5.13E-03
iNOS Signalling	8.32E-03
Regulation of IL-2 Expression in Activated and Anergic TLymphocytes	8.32E-03
NF-κB Signalling	1.05E-02
B Cell Receptor Signalling	1.07E-02
IL-2 Signalling	1.15E-02
Toll-like Receptor Signalling	1.74E-02
Leukocyte Extravasation Signalling	1.86E-02
iCOS-iCOSL Signalling in T Helper Cells	2.19E-02
MSP-RON Signalling Pathway	2.69E-02
PI3K Signalling in B Lymphocytes	3.24E-02
Lymphotoxin β Receptor Signalling	3.31E-02
Complement System	3.63E-02
IL-3 Signalling	3.98E-02
CD28 Signalling in T Helper Cells	4.47E-02
Role of NFAT in Regulation of the Immune Response	4.79E-02
Degradation-related pathways	
Protein Ubiquitination Pathway	1.45E-02
14-3-3-mediated Signalling	3.89E-02
Apoptosis-related pathways	
CeramideSignalling	3.98E-02
Nucleotide Excision Repair Pathway	4.17E-02
Growth and proliferation-related pathways	
Signalling by Rho Family GTPases	4.57E-05
Androgen Signalling	1.66E-03
Rac Signalling	1.32E-03
Wnt/β-catenin Signalling	2.63E-03
p53 Signalling	3.28E-03
IGF-1 Signalling	4.68E-03
G Beta Gamma Signalling	8.91E-03
EGF Signalling	1.15E-02
PDGF Signalling	1.51E-02
PAK Signalling	1.95E-02
mTORSignalling	2.09E-02
Factors Promoting Cardiogenesis in Vertebrates	3.63E-02
Insulin Receptor Signalling	4.27E-02
Aryl Hydrocarbon Receptor Signalling	4.79E-02
Transcription and translation-related pathways	
FLT3 Signalling in Hematopoietic Progenitor Cells	3.98E-02
Urate Biosynthesis/Inosine 5'-phosphate Degradation	3.98E-02
ERK/MAPK Signalling	4.47E-02

ATP production-related pathways	
Mitochondrial Dysfunction	2.77E-06
L-cysteine Degradation I	9.12E-04
Retinoic acid receptor Activation	8.91E-03
Aspartate Degradation II	7.76E-03
Glutamate Degradation II	9.33E-03
Valine Degradation I	2.51E-02
Leucine Degradation I	3.55E-02
TCA Cycle II (Eukaryotic)	3.80E-02
Guanosine Nucleotides Degradation III	2.29E-02
Protein Kinase A Signalling	3.02E-02
Phenylalanine Degradation I (Aerobic)	4.90E-02
Lipid metabolism-related pathways	
PPARα/RXRα Activation	1.41E-03
Glycerol-3-phosphate Shuttle	9.33E-03
Glycerol Degradation I	2.63E-02
Molybdenum Cofactor Biosynthesis	4.90E-02
Intracellular structure and vesicle transport-related pathways	
MacropinocytosisSignalling	5.13E-03
Gα12/13 Signalling	5.25E-03
RAN Signalling	1.51E-02
Cellular adhesion and extracellular matrix -related pathways	
ILK Signalling	9.55E-03
PaxillinSignalling	2.63E-02
Tight Junction Signalling	5.37E-03
Geranylgeranyldiphosphate Biosynthesis	4.90E-02
Motor unit-related pathways	
Amyloid Processing	4.38E-02

[1] Fisher's exact test was used to calculate a p-value by IPA to determine the probability that the association between the genes in the dataset and the pathways is explained by chance alone. Low p-value suggests that the association between genes belonging to a particular pathway are unlikely to be associated with a function merely by chance.

Table 4-5: Shortlist of genes downregulated in low versus high attenuation muscle encoding proteins involved in inflammation

Agilent ID	Gene Symbol	Entrez Gene Name	p-value[1]	FC[2]
A_24_P363615	MTPN	myotrophin	7.04E-03	-3.1
A_23_P147098	MTPN	myotrophin	5.55E-03	-2.5
A_24_P148717	CCR1	chemokine (C-C motif) receptor 1	4.62E-03	-2.4
A_23_P54573	ZFPM1	zinc finger protein, multitype 1	3.61E-03	-2.3
A_23_P10232	BANK1‡	B-cell scaffold protein with ankyrin repeats 1	2.54E-03	-2.2
A_23_P88781	CIAPIN1	cytokine induced apoptosis inhibitor 1	2.88E-03	-2.1
A_24_P229936	CIITA‡	class II, major histocompatibility complex, transactivator	3.06E-03	-2.1
A_23_P72737	IFITM1	interferon induced transmembrane protein 1	7.52E-03	-2.0
A_23_P258418	TNIP2	TNFAIP3 interacting protein 2	8.03E-03	-1.9
A_24_P56052	ZFP91‡	zinc finger protein 91 homolog	2.85E-03	-1.9
A_32_P41496	LOC100132831	A20-binding inhibitor of NF-kappaB activation 2 pseudogene	6.92E-03	-1.9
A_24_P406334	STEAP1	six transmembrane epithelial antigen of the prostate 1	9.32E-03	-1.8
A_24_P936272	HLA-B	major histocompatibility complex, class I, B	3.43E-03	-1.7
A_23_P6909	CCRL1	chemokine (C-C motif) receptor-like 1	6.54E-03	-1.6
A_23_P13382	LSP1	lymphocyte-specific protein 1	1.81E-03	-1.6
A_23_P21495	FCGBP	Fc fragment of IgG binding protein	5.24E-03	-1.6
A_23_P109235	RALY‡	RNA binding protein, autoantigenic (hnRNP-associated with lethal yellow homolog (mouse))	3.37E-03	-1.6
A_23_P152620	TNFSF13	tumor necrosis factor (ligand) superfamily, member 13	8.60E-03	-1.6
A_24_P772061	PPIA	peptidylprolylisomerase A	2.49E-03	-1.6
A_23_P331928	CD109‡	CD109 molecule	2.15E-03	-1.6
A_24_P73599	IL16‡	interleukin 16	2.20E-03	-1.5
A_23_P76078	IL23A	interleukin 23, α subunit p19	6.67E-03	-1.5
A_24_P158903	IRAK4	IL-1 receptor-associated kinase 4	1.69E-03	-1.5
A_32_P66881	TLR4‡	toll-like receptor 4	1.33E-03	-1.5
A_23_P74928	MR1‡	major histocompatibility complex, class I-related	2.18E-03	-1.5
A_23_P101992	MARCO	macrophage receptor with collagenous structure	6.31E-03	-1.5

[1] t-test p-value

[2]FC: fold change. Fold change values that are positive are higher in patients with low muscle attenuation

‡ Androgen response elements identified by Wyce et al. in the genome of skeletal muscle cells

Table 4-6: Shortlist of genes upregulated in low versus high attenuation muscle encoding proteins involved in inflammation

Agilent ID	Gene Symbol	Entrez Gene Name	p-value[1]	FC[2]
A_23_P64898	KLRG1	killer cell lectin-like receptor subfamily G, member 1	8.55E-03	1.5
A_23_P133916	C2	complement component 2	7.96E-03	1.5
A_23_P121533	SPON2‡	spondin 2, extracellular matrix protein	5.26E-03	1.5
A_23_P214627	AIF1 (includes EG:11629)	allograft inflammatory factor 1	3.65E-03	1.5
A_23_P110777	LECT2	leukocyte cell-derived chemotaxin 2	5.37E-03	1.6
A_23_P10873	TLR1	toll-like receptor 1	3.08E-03	1.7
A_32_P157927	IGK@	immunoglobulin kappa locus	5.06E-03	1.8
A_23_P7212	CFI	complement factor I	4.37E-03	1.8
A_24_P354800	HLA-DOA	major histocompatibility complex, class II, DO α	6.83E-03	1.8
A_24_P373312	NFATC3‡	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 3	1.36E-04	2.0
A_23_P125423	C1R‡	complement component 1, r subcomponent	1.21E-03	2.0
A_23_P78092	EVI2A‡	ecotropic viral integration site 2A	4.10E-03	2.1
A_23_P2492	C1S‡	complement component 1, s subcomponent	2.29E-04	2.2
A_24_P24371	IGHG4	immunoglobulin heavy constant gamma 4 (G4m marker)	9.02E-03	2.2
A_24_P227927	IL21R‡	interleukin 21 receptor	4.34E-03	2.2
A_24_P273972	CFH‡	complement factor H	4.23E-03	2.3
A_24_P88696	SCG2‡	secretogranin II	9.77E-03	2.8
A_23_P91095	CD28‡	CD28 molecule	8.08E-03	3.0
A_23_P21260	IGKC	immunoglobulin kappa constant	8.56E-03	3.1
A_23_P96191	IGK@	immunoglobulin kappa locus	5.23E-03	3.1
A_23_P213857	C7	complement component 7	8.83E-03	3.2
A_24_P318990	IGLC1‡	immunoglobulin lambda constant 1 (Mcg marker)	7.15E-03	3.6

[1] t-test p-value

[2]FC: fold change. Fold change values that are positive are higher in patients with low muscle attenuation

‡ Androgen response elements identified by Wyce et al. in the genome of skeletal muscle cells

Table 4-7: Shortlist of genes downregulated in low versus high attenuation muscle encoding proteins involved in protein degradation

Agilent ID	Gene Symbol	Entrez Gene Name	p-value[1]	FC[2]
A_23_P406385	FBXL16	F-box and leucine-rich repeat protein 16	9.05E-05	-4.4
A_32_P213755	PSMB7‡	proteasome (prosome, macropain) subunit, beta type, 7	7.39E-03	-4.0
A_32_P20288	FBXL17‡	F-box and leucine-rich repeat protein 17	2.64E-03	-2.9
A_23_P11244	PJA1	praja ring finger 1, E3 ubiquitin protein ligase	6.47E-03	-2.5
A_23_P103334	RC3H1	ring finger and CCCH-type domains 1	5.55E-03	-2.5
A_32_P185149	RNF5	ring finger protein 5, E3 ubiquitin protein ligase	4.17E-03	-2.3
A_32_P114348	RNF5	ring finger protein 5, E3 ubiquitin protein ligase	4.72E-03	-2.2
A_23_P211179	UBE2G2	ubiquitin-conjugating enzyme E2G 2	4.57E-03	-2.0
A_23_P110196	HERC5	HECT and RLD domain containing E3 ubiquitin protein ligase 5	2.68E-03	-1.9
A_23_P8095	RNF5	ring finger protein 5, E3 ubiquitin protein ligase	3.91E-03	-1.9
A_23_P413634	ZNF329	zinc finger protein 329	2.85E-03	-1.9
A_23_P341471	UBE2N	ubiquitin-conjugating enzyme E2N	9.40E-03	-1.7
A_24_P381803	UBE3A‡	ubiquitin protein ligase E3A	9.92E-04	-1.7
A_24_P286013	UBE2K	ubiquitin-conjugating enzyme E2K	1.03E-03	-1.6
A_24_P171873	FBXO4	F-box protein 4	4.74E-03	-1.5
A_23_P99632	RNF31	ring finger protein 31	4.61E-03	-1.5
A_23_P132536	TRAK1‡	trafficking protein, kinesin binding 1	4.65E-03	-2.3

[1] t-test p-value

[2]FC: fold change. Fold change values that are positive are higher in patients with low muscle attenuation

‡ Androgen response elements identified by Wyce et al. in the genome of skeletal muscle cells

Table 4-8: Shortlist of genes upregulated in low versus high attenuation muscle encoding proteins involved in protein degradation

Agilent ID	Gene Symbol	Entrez Gene Name	p-value[1]	FC[2]
A_24_P47681	CAND1	cullin-associated and neddylation-dissociated 1	9.84E-03	1.5
A_23_P46819	BTRC‡	beta-transducin repeat containing E3 ubiquitin protein ligase	1.70E-03	1.5
A_24_P218259	FBXO32‡	F-box protein 32	8.52E-03	1.5
A_24_P339305	UBE2Z	ubiquitin-conjugating enzyme E2Z	6.02E-03	1.6
A_23_P91468	PSMA7	proteasome (prosome, macropain) subunit, α type, 7	1.29E-04	1.7
A_23_P141146	FBXL20	F-box and leucine-rich repeat protein 20	1.77E-03	1.7
A_23_P146990	WWP1	WW domain containing E3 ubiquitin protein ligase 1	9.89E-03	1.7
A_23_P65254	POMP	proteasome maturation protein	1.20E-04	1.8
A_23_P83438	UBE2Z	ubiquitin-conjugating enzyme E2Z	7.83E-03	1.8
A_23_P322562	NEURL	neuralized homolog (Drosophila)	2.44E-03	2.5
A_32_P179572	UBFD1	ubiquitin family domain containing 1	6.82E-03	3.3
A_23_P169934	RILPL1	Rab interacting lysosomal protein-like 1	7.64E-03	1.8
A_23_P46141	CTSS‡	cathepsin S	2.87E-03	1.9
A_23_P1552	CTSC‡	cathepsin C	8.74E-03	2.1
A_24_P567408	CLN3	ceroid-lipofuscinosis, neuronal 3	2.16E-03	2.6

[1] t-test p-value

[2]FC: fold change. Fold change values that are positive are higher in patients with low muscle attenuation

‡ Androgen response elements identified by Wyce et al. in the genome of skeletal muscle cells

Table 4-9: Shortlist of genes differentially expressed according muscle attenuation encoding proteins involved in apoptosis and autophagy

Agilent ID	Gene Symbol	Entrez Gene Name	p-value[1]	FC[2]
A_24_P218620	DKK3‡	dickkopf 3 homolog (Xenopuslaevis)	9.63E-03	-2.4
A_23_P90189	BBC3	BCL2 binding component 3	7.01E-03	-2.2
A_23_P346309	BAX	BCL2-associated X protein	2.01E-03	-1.9
A_32_P174365	SATB2‡	SATB homeobox 2	1.93E-05	-1.6
A_23_P64173	CARD16	caspase recruitment domain family, member 16	4.30E-03	-1.6
A_23_P118815	BIRC5	baculoviral IAP repeat containing 5	4.76E-03	-1.6
A_23_P36611	APAF1‡	apoptotic peptidase activating factor 1	7.86E-03	1.5
A_23_P65963	BFAR	bifunctional apoptosis regulator	4.61E-04	1.6
A_23_P47800	DIABLO	diablo, IAP-binding mitochondrial protein	7.41E-03	1.6
A_23_P503233	EDARADD	EDAR-associated death domain	9.23E-03	1.7
A_23_P202104	PPIF	peptidylprolylisomerase F	2.67E-03	1.9
A_32_P113508	ATG16L1	autophagy related 16-like 1 (S. cerevisiae)	3.24E-04	1.5
A_23_P101342	ATG4D	autophagy related 4D, cysteine peptidase	2.65E-03	1.7
A_23_P89410	BECN1	beclin 1, autophagy related	3.09E-03	1.9
A_23_P143987	ATG7‡	autophagy related 7	7.94E-03	2.0
A_24_P313597	BECN1	beclin 1, autophagy related	9.00E-03	3.1

[1] t-test p-value

[2]FC: fold change. Fold change values that are positive are higher in patients with low muscle attenuation

‡ Androgen response elements identified by Wyce et al. in the genome of skeletal muscle cells

Table 4-10: Shortlist of genes differentially expressed according attenuation muscle encoding proteins involved in growth and proliferation

Agilent ID	Gene Symbol	Entrez Gene Name	p-value[1]	FC[2]
Growth factor				
A_23_P206484	GFER	growth factor, augmenter of liver regeneration	1.87E-03	-2.1
A_23_P334021	IGF2R‡	insulin-like growth factor 2 receptor	2.45E-04	1.8
A_23_P116235	MDK	midkine	8.72E-03	3.2
Transforming growth factor (TGF) family				
A_23_P68487	BMP7‡	bone morphogenetic protein 7	1.94E-03	-3.1
A_23_P115118	BMP8B‡	bone morphogenetic protein 8b	2.38E-03	-2.5
A_32_P146394	TGFBR1‡	transforming growth factor, beta receptor 1	6.68E-03	-2.1
A_23_P33768	ZFYVE9‡	zinc finger, FYVE domain containing 9	2.75E-03	-1.8
A_24_P264790	LTBP3	latent transforming growth factor β binding protein 3	6.29E-03	-1.8
A_23_P218144	LTBP2‡	latent transforming growth factor beta binding protein 2	8.46E-03	1.5
A_23_P405129	LTBP2‡	latent transforming growth factor beta binding protein 2	9.28E-04	1.6
A_24_P402438	TGFB2‡	transforming growth factor, beta 2	3.38E-03	2.1
Ras/Rho family				
A_24_P24856	LOC440461	Rho GTPase activating protein 27 pseudogene	4.91E-03	-2.7
A_23_P321855	ARHGEF7‡	Rho guanine nucleotide exchange factor (GEF) 7	8.79E-03	-2.0
A_23_P502747	RASAL2‡	RAS protein activator like 2	3.19E-03	-2.0
A_23_P258151	FGD5	FYVE, RhoGEF and PH domain containing 5	3.16E-03	-1.7
A_24_P44916	CDC42EP5‡	CDC42 effector protein (Rho GTPase binding) 5	7.27E-03	-1.7
A_23_P98183	HRAS	Harvey rat sarcoma viral oncogene homolog	6.03E-03	-1.6
A_23_P69738	RASL11B ‡	RAS-like, family 11, member B	9.96E-03	1.7
A_23_P72968	ARHGAP36	Rho GTPase activating protein 36	9.00E-04	1.9
A_32_P49844	RHOQ	ras homolog family member Q	4.02E-03	2.1
A_32_P222695	ARHGEF37	Rho guanine nucleotide exchange factor 37	5.30E-03	2.5
A_32_P217709	RAC1	ras-related C3 botulinum toxin substrate 1	6.34E-03	6.2
Growth factor downstream regulators				
A_24_P17917	ICK	intestinal cell (MAK-like) kinase	5.73E-04	-6.0
A_23_P36166	PIK3C2A	phosphoinositide-3-kinase, class 2, α polypeptide	4.01E-03	-5.0
A_32_P99715	HSPA8	heat shock 70kDa protein 8	3.27E-03	-4.6
A_23_P105524	PLCZ1	phospholipase C, zeta 1	2.95E-03	-4.1
A_23_P142304	MKNK2‡	MAP kinase interacting	4.04E-03	-2.6

		serine/threonine kinase 2		
A_24_P938135	MKNK2‡	MAP kinase interacting serine/threonine kinase 2	3.16E-03	-2.4
A_23_P34606	MTOR	mechanistic target of rapamycin	7.17E-04	-2.1
A_23_P50773	CRTC1‡	CREB regulated transcription coactivator 1	4.45E-03	-2.0
A_32_P143048	ZFYVE9‡	zinc finger, FYVE domain containing 9	2.86E-03	-1.8
A_32_P112331	PLD1	phospholipase D1, phosphatidylcholine-specific	1.23E-04	-1.5
A_24_P416489	MAP2K6	mitogen-activated protein kinase kinase 6	6.30E-04	1.7
A_24_P43884	MAPKAP1‡	mitogen-activated protein kinase associated protein 1	3.66E-03	1.8
Steroid hormone				
A_23_P115149	WDR77	WD repeat domain 77	7.28E-03	-2.4
A_24_P383478	ESR1‡	estrogen receptor 1	2.58E-03	1.7
A_23_P309739	ESR1‡	estrogen receptor 1	8.16E-03	2.8
Wntsignalling pathway				
A_23_P52986	VWCE‡	von Willebrand factor C and EGF domains	7.98E-04	-1.7
A_24_P208513	WNT6	wingless-type MMTV integration site family, 6	9.87E-03	-1.6
A_24_P261417	DKK3‡	dickkopf 3 homolog (Xenopuslaevis)	5.78E-03	-1.5
A_24_P38276	FZD1	frizzled family receptor 1	8.44E-03	1.5
A_23_P81103	SFRP2	secreted frizzled-related protein 2	8.53E-03	1.6
A_23_P347432	DVL1	dishevelled, dsh homolog 1 (Drosophila)	7.96E-03	2.0
Cell cycle regulation				
A_23_P43484	CDKN2A	cyclin-dependent kinase inhibitor 2A	3.95E-04	-3.6
A_23_P401	CENPF‡	centromere protein F, 350/400kDa (mitosin)	3.69E-03	-2.9
A_24_P357536	FBXO11‡	F-box protein 11	4.40E-03	-2.2
A_23_P65757	CCNB2	cyclin B2	7.23E-04	-1.8
A_23_P253446	GAP43‡	growth associated protein 43	5.37E-04	-1.8
A_24_P82466	GAS7‡	growth arrest-specific 7	3.97E-05	-1.8
A_23_P138435	ZMIZ1‡	zinc finger, MIZ-type containing 1	9.34E-04	-1.6
A_24_P765784	GTF2I‡	general transcription factor IIi	9.42E-03	1.6
A_23_P216679	CDC14B‡	CDC14 cell division cycle 14 homolog B	1.75E-03	1.6
A_24_P913227	CDC23	cell division cycle 23 homolog	2.88E-03	1.6
A_23_P88083	CDC16‡	cell division cycle 16 homolog	4.05E-03	1.7
A_23_P380010	STARD9	StAR-related lipid transfer (START) domain containing 9	9.90E-03	1.7
A_32_P37143	GAS2L3	growth arrest-specific 2 like 3	1.70E-03	1.8
A_23_P335813	TOB2	transducer of ERBB2, 2	9.31E-03	1.9
A_23_P97064	FBXO6	F-box protein 6	1.77E-04	2.1
A_23_P87575	CCNT1	cyclin T1	2.85E-03	6.9
Other growth/proliferation related genes				
A_23_P160167	TSPAN1	tetraspanin 1	2.19E-03	-9.8
A_23_P109133	AVP	arginine vasopressin	8.13E-03	-4.0

A_32_P138586	GLI4	GLI family zinc finger 4	2.66E-03	-2.5
A_23_P114349	XAGE3	X antigen family, member 3	2.76E-03	-2.2
A_24_P89701	IMPDH1	IMP (inosine 5'-monophosphate) dehydrogenase 1	6.34E-03	-2.0
A_23_P119992	VRK2‡	vaccinia related kinase 2	5.79E-04	-1.6
A_24_P830690	PDPK1	3-phosphoinositide dependent protein kinase-1	3.32E-03	-1.5
A_23_P47885	LRIG3‡	leucine-rich repeats and immunoglobulin-like domains 3	4.24E-03	1.5
A_23_P2097	TRIM68	tripartite motif containing 68	1.43E-03	1.5
A_23_P51679	MEF2D‡	myocyte enhancer factor 2D	2.92E-03	1.6
A_23_P100788	STAT5B	signal transducer and activator of transcription 5B	7.04E-03	1.6
A_23_P345212	BOD1L2	biorientation of chromosomes in cell division 1-like 2	5.00E-04	2.4
A_23_P17430	RBM38	RNA binding motif protein 38	6.38E-04	2.9
A_23_P137634	PROX1‡	prosperohomeobox 1	5.61E-03	4.4
A_23_P124384	SHOX2	short stature homeobox 2	3.42E-03	4.9

[1] t-test p-value
[2]FC: fold change. Positive are higher in patients with low muscle attenuation
‡ Androgen response elements identified by Wyce et al. in the genome of skeletal muscle cells

Table 4-11: Shortlist of genes downregulated in low versus high attenuation muscle encoding proteins involved in transcription and translation

Agilent ID	Gene Symbol	Entrez Gene Name	p-value[1]	FC[2]
A_32_P171061	ASCL2	achaete-scute complex homolog 2	5.79E-03	-5.8
A_23_P20927	TNKS	tankyrase, TRF1-interacting ankyrin-related ADP-ribose polymerase	3.02E-03	-4.9
A_23_P84762	PARP10	poly (ADP-ribose) polymerase family, member 10	3.55E-03	-3.0
A_32_P211799	USF2	upstream transcription factor 2, c-fos interacting	3.36E-03	-2.4
A_23_P77228	CRTC3‡	CREB regulated transcription coactivator 3	8.09E-03	-2.4
A_23_P203463	TAF10	TAF10 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 30kDa	3.69E-03	-2.2
A_24_P127719	MAFA‡	v-mafmusculoaponeuroticfibrosarcoma oncogene homolog A (avian)	9.21E-03	-2.2
A_24_P37519	LZTFL1‡	leucine zipper transcription factor-like 1	9.31E-03	-2.2
A_24_P921781	DMRT3‡	doublesex and mab-3 related transcription factor 3	2.51E-03	-2.1
A_23_P325100	DMRT3‡	doublesex and mab-3 related transcription factor 3	3.99E-03	-1.8
A_23_P83298	PRRX2‡	paired related homeobox 2	2.53E-04	-1.8
A_32_P184394	TFEC	transcription factor EC	9.84E-03	-1.8
A_23_P41292	CTBP1‡	C-terminal binding protein 1	9.75E-05	-1.7
A_24_P49106	TCEAL7	transcription elongation factor A (SII)-like 7	4.38E-03	-1.6
A_23_P63289	SSU72‡	SSU72 RNA polymerase II CTD phosphatase homolog (S. cerevisiae)	2.86E-03	-1.5
A_24_P385119	TAF2	TAF2 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 150kDa	8.10E-03	-1.5
A_23_P90383	RPL18A	ribosomal protein L18a	6.38E-03	-2.7
A_24_P7629	RPL32P3	ribosomal protein L32 pseudogene 3	2.51E-03	-2.7
A_24_P221366	RPS15A	ribosomal protein S15a	3.55E-03	-2.4
A_32_P207231	RPL7	ribosomal protein L7	9.39E-03	-2.4
A_23_P24763	RPS13	ribosomal protein S13	3.74E-03	-2.3
A_23_P97021	EIF2C3	eukaryotic translation initiation factor 2C, 3	3.20E-03	-2.2
A_32_P220127	RPL34	ribosomal protein L34	9.46E-03	-1.6
A_32_P193288	RPL18A	ribosomal protein L18a	6.98E-03	-1.6
A_23_P74097	TCEB3	transcription elongation factor B (SIII), polypeptide 3 (110kDa, elongin A)	5.19E-03	-1.5

[1] t-test p-value

[2]FC: fold change. Fold change values that are positive are higher in patients with low muscle attenuation

‡ Androgen response elements identified by Wyce et al. in the genome of skeletal muscle cells

Table 4-12: Shortlist of genes upregulated in low versus high attenuation muscle encoding proteins involved in transcription and translation

Agilent ID	Gene Symbol	Entrez Gene Name	p-value[1]	FC[2]
A_23_P434442	TCEAL3	transcription elongation factor A (SII)-like 3	3.44E-03	1.5
A_32_P192545	TCEAL6	transcription elongation factor A (SII)-like 6	3.51E-03	1.5
A_24_P196117	GTF2H5‡	general transcription factor IIH, polypeptide 5	5.46E-03	1.5
A_23_P35148	TAF13	TAF13 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 18kDa	2.51E-03	1.5
A_23_P161615	POLA2‡	polymerase (DNA directed), α 2, accessory subunit	2.71E-03	1.7
A_32_P161762	RUNX2‡	runt-related transcription factor 2	1.94E-03	1.7
A_23_P68547	MCM8‡	minichromosome maintenance complex component 8	5.95E-03	1.8
A_23_P120941	ATF4‡	activating transcription factor 4 (tax-responsive enhancer element B67)	4.76E-04	1.8
A_32_P18470	TCEAL5	transcription elongation factor A (SII)-like 5	9.17E-03	1.9
A_24_P142442	POLDIP2	polymerase (DNA-directed), delta interacting protein 2	1.81E-03	2.0
A_23_P20480	BRF2	BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like	2.96E-03	2.1
A_23_P157452	POLR2K	polymerase (RNA) II (DNA directed) polypeptide K, 7.0kDa	8.71E-03	2.2
A_23_P348281	TCEANC2	transcription elongation factor A (SII) N-terminal and central domain containing 2	7.92E-03	2.5
A_23_P215253	POLR2J2‡	polymerase (RNA) II (DNA directed) polypeptide J2	3.39E-03	2.5
A_24_P321626	POLR3B	polymerase (RNA) III (DNA directed) polypeptide B	3.37E-03	2.7
A_32_P47701	EEF1A1‡	eukaryotic translation elongation factor 1 α 1	1.17E-03	1.5
A_24_P356015	EIF2S1	eukaryotic translation initiation factor 2, subunit 1 α, 35kDa	1.90E-04	2.5

[1] t-test p-value

[2]FC: fold change. Fold change values that are positive are higher in patients with low muscle attenuation

‡ Androgen response elements identified by Wyce et al. in the genome of skeletal muscle cells

Table 4-13: Shortlist of downregulated genes in low versus high attenuation muscle encoding proteins involved in ATP production and reactive oxygen species

Agilent ID	Gene Symbol	Entrez Gene Name	p-value[1]	FC[2]
A_24_P927850	IDH3A	isocitrate dehydrogenase 3 (NAD+) α	9.91E-03	-2.7
A_24_P710024	ND4[§]	NADH dehydrogenase, subunit 4 (complex I)	4.97E-03	-2.6
A_23_P119812	GPD2[‡]	glycerol-3-phosphate dehydrogenase 2	8.01E-04	-2.3
A_23_P257111	FBP1	fructose-1,6-bisphosphatase 1	4.13E-03	-2.3
A_24_P117528	PRPS2	phosphoribosyl pyrophosphate synthetase 2	1.38E-05	-2.1
A_23_P205959	ALDH1A3[‡]	aldehyde dehydrogenase 1 family, member A3	9.15E-03	-2.0
A_23_P360209	ND3[§]	NADH dehydrogenase, subunit 3 (complex I)	2.75E-03	-2.0
A_23_P431853	ND2[§]	mitochondrially encoded NADH dehydrogenase 2	1.82E-03	-2.0
A_23_P161297	OGDHL	oxoglutarate dehydrogenase-like	8.15E-03	-1.9
A_24_P367965	HK1[‡]	hexokinase 1	7.88E-03	-1.9
A_23_P163161	SDR39U1	short chain dehydrogenase/reductase family 39U, member 1	4.90E-03	-1.6
A_23_P87616	ATP5G2	ATP synthase, H+ transporting, mitochondrial Fo complex, subunit C2 (subunit 9)	4.57E-03	-1.5
A_32_P149416	TXNRD1[‡]	thioredoxinreductase 1	7.53E-03	-3.1
A_23_P214300	GSTA2	glutathione S-transferase α 2	8.19E-03	-2.1
A_23_P254741	SOD3	superoxide dismutase 3, extracellular	5.61E-03	-1.5
A_32_P227525	PRDX2[‡]	peroxiredoxin 2	6.29E-03	-1.5

[1] t-test p-value
[2]FC: fold change. Positive values are higher in patients with low muscle attenuation
[§]Gene encoded from mitochondrial genome
[‡] Androgen response elements identified by Wyce et al. in the genome of skeletal muscle cells

Table 4-14: Shortlist of upregulated genes in low versus high attenuation muscle encoding proteins involved in ATP production and reactive oxygen species

Agilent ID	Gene Symbol	Entrez Gene Name	p-value[1]	FC[2]
A_23_P51009	NDUFB3	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 3, 12kDa	6.47E-03	1.5
A_23_P52639	COX8A	cytochrome c oxidase subunit VIIIA	1.58E-03	1.5
A_23_P130418	NDUFV2‡	NADH dehydrogenase (ubiquinone) flavoprotein 2, 24kDa	5.07E-03	1.5
A_23_P162982	DHRS4	dehydrogenase/reductase (SDR family) member 4	1.08E-03	1.5
A_23_P154832	ATP5J	ATP synthase, H+ transporting, mitochondrial Fo complex, subunit F6	1.62E-03	1.5
A_23_P106544	CMC2	COX assembly mitochondrial protein 2 homolog	6.02E-05	1.6
A_23_P55123	COX10‡	COX10 homolog	1.10E-04	1.7
A_24_P416951	NDUFV3	NADH dehydrogenase (ubiquinone) flavoprotein 3, 10kDa	7.17E-03	1.7
A_23_P140960	NDUFAB1	NADH dehydrogenase (ubiquinone) 1, α/beta subcomplex, 1, 8kDa	3.33E-03	1.7
A_23_P159650	COX7B	cytochrome c oxidase subunit VIIb	3.42E-03	1.7
A_23_P141032	COX4I1	cytochrome c oxidase subunit IV isoform 1	9.04E-04	1.8
A_23_P345942	NDUFAF2	NADH dehydrogenase (ubiquinone) complex I, assembly factor 2	4.95E-03	1.8
A_32_P117016	ALDH1L2	aldehyde dehydrogenase 1 family, member L2	2.87E-03	1.8
A_24_P128020	NDUFS4	NADH dehydrogenase (ubiquinone) Fe-S protein 4, 18kDa	2.79E-04	2.2
A_23_P106575	GOT2	glutamic-oxaloacetic transaminase 2, mitochondrial	2.82E-03	2.2
A_23_P129313	IVD	isovaleryl-CoA dehydrogenase	7.26E-03	2.2
A_24_P58944	SDHAP1	succinate dehydrogenase complex, subunit A, flavoproteinpseudogene 1	8.64E-03	2.4
A_24_P206047	SLC25A4	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 4	8.35E-03	2.5
A_23_P138967	SDHD	succinate dehydrogenase complex, subunit D, integral membrane protein	6.63E-03	2.6
A_32_P67259	SDHA	succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	4.65E-03	2.9
A_23_P157569	ADHFE1‡	alcohol dehydrogenase, iron containing, 1	3.50E-04	4.4
A_32_P170925	TXNRD3‡	thioredoxinreductase 3	5.59E-03	1.7
A_23_P20107	GSTK1	glutathione S-transferase kappa 1	5.98E-03	1.7
A_23_P94204	OXR1	oxidation resistance 1	3.73E-04	1.8
A_23_P39185	RDH13	retinol dehydrogenase 13	7.08E-03	1.9
A_23_P23194	PINK1‡	PTEN induced putative kinase 1	1.27E-03	2.2

[1] t-test p-value

[2]FC: fold change. Positive values are higher in patients with low muscle attenuation

‡ Androgen response elements identified by Wyce et al. in the genome of skeletal muscle cells

Table 4-15: Shortlist of differentially expressed according to muscle encoding proteins involved in mitochondrial transcription and translation

Agilent ID	Gene Symbol	Entrez Gene Name	p-value[1]	FC[2]
A_32_P143596	TOMM7	translocase of outer mitochondrial membrane 7 homolog (yeast)	6.55E-03	-3.2
A_23_P33886	TIMM17B	translocase of inner mitochondrial membrane 17 homolog B (yeast)	1.56E-03	-2.1
A_23_P169629	SHMT2	serine hydroxymethyltransferase 2	9.34E-03	-2.0
A_24_P371194	MRPL53	mitochondrial ribosomal protein L53	2.11E-04	-1.9
A_23_P122233	MRPL22	mitochondrial ribosomal protein L22	5.53E-03	-1.9
A_24_P283294	MRPS10‡	mitochondrial ribosomal protein S10	5.32E-03	-1.7
A_23_P114808	MECR‡	mitochondrial trans-2-enoyl-CoA reductase	9.20E-03	-1.7
A_24_P569294	MRPS12	mitochondrial ribosomal protein S12	2.17E-03	-1.7
A_23_P131096	POLRMT‡	polymerase (RNA) mitochondrial (DNA directed)	5.06E-03	-1.6
A_23_P102258	MRPL53	mitochondrial ribosomal protein L53	2.00E-03	-1.6
A_23_P383688	AARS2	alanyl-tRNAsynthetase 2, mitochondrial (putative)	1.15E-03	-1.6
A_23_P9582	TUFM	Tu translation elongation factor, mitochondrial	5.68E-04	-1.5
A_23_P152353	EARS2	glutamyl-tRNAsynthetase 2, mitochondrial (putative)	2.59E-03	-1.5
A_32_P111609	TOMM7	translocase of outer mitochondrial membrane 7 homolog (yeast)	8.48E-03	-1.5
A_23_P86182	MRPS21	mitochondrial ribosomal protein S21	4.89E-03	1.5
A_23_P114826	MRPS15	mitochondrial ribosomal protein S15	8.39E-03	1.5
A_24_P12932	MRPS16	mitochondrial ribosomal protein S16	7.03E-03	1.5
A_23_P49768	MRPL27	mitochondrial ribosomal protein L27	2.92E-03	1.5
A_23_P33720	FARS2‡	phenylalanyl-tRNAsynthetase 2, mitochondrial	3.48E-03	1.6
A_23_P71464	DECR1‡	2,4-dienoyl CoA reductase 1, mitochondrial	5.02E-03	1.7
A_24_P351304	IMMT	inner membrane protein	3.22E-03	1.7
A_23_P413721	GPD1	glycerol-3-phosphate dehydrogenase 1	7.57E-03	1.7
A_23_P72138	MRPS22	mitochondrial ribosomal protein S22	4.27E-03	1.9
A_32_P209989	MRPL46‡	mitochondrial ribosomal protein L46	1.08E-03	1.9
A_24_P125690	MRPL34	mitochondrial ribosomal protein L34	4.25E-03	1.9
A_23_P157352	MRPS33	mitochondrial ribosomal protein S33	4.03E-04	1.9
A_23_P25348	ACAD10	acyl-CoA dehydrogenase family, member 10	8.63E-04	2.7
A_23_P41588	HARS2	histidyl-tRNAsynthetase 2, mitochondrial (putative)	9.15E-04	3.0
A_23_P553	TARS2	threonyl-tRNAsynthetase 2, mitochondrial (putative)	1.80E-03	3.4
A_23_P51291	PARS2	prolyl-tRNAsynthetase 2, mitochondrial (putative)	2.17E-03	3.7

[1] t-test p-value
[2]FC: fold change. Positive are higher in patients with low muscle attenuation
‡ Androgen response elements identified by Wyce et al. in the genome of skeletal muscle cells

Table 4-16: Shortlist of differentially expressed according to muscle encoding proteins involved in lipid metabolism

Agilent ID	Gene Symbol	Entrez Gene Name	p-value[1]	FC[2]
A_23_P161135	LEPR‡	leptin receptor	6.20E-04	-3.0
A_23_P150903	FAR2	fatty acyl CoA reductase 2	8.52E-04	-2.9
A_23_P136986	APOOL‡	apolipoprotein O-like	4.05E-03	-1.7
A_23_P95130	SLC37A3	solute carrier family 37 (glycerol-3-phosphate transporter), member 3	6.24E-03	-1.5
A_24_P231104	LEPR‡	leptin receptor	1.01E-03	-1.5
A_24_P13376	ADIPOR2	adiponectin receptor 2	5.65E-04	1.5
A_24_P301557	LPIN2‡	lipin 2	2.17E-04	1.6
A_24_P151920	TMEM97	transmembrane protein 97	2.75E-04	1.6
A_24_P244442	BSCL2	Berardinelli-Seip congenital lipodystrophy 2 (seipin)	2.13E-04	1.8
A_23_P148919	CPT2	carnitinepalmitoyltransferase 2	1.04E-03	1.9
A_23_P206945	ACOX1	acyl-CoA oxidase 1, palmitoyl	3.49E-03	1.9
A_24_P570049	PPARA‡	peroxisome proliferator-activated receptor α	4.61E-03	1.9
A_23_P210900	ACSS2‡	acyl-CoA synthetase short-chain family member 2	1.33E-04	2.1
A_24_P272222	PLIN5	perilipin 5	4.22E-03	2.2
A_23_P80449	THRB‡	thyroid hormone receptor, beta	1.07E-03	3.5

[1] t-test p-value
[2] FC: fold change. Positive are higher in patients with low muscle attenuation
‡ Androgen response elements identified by Wyce et al. in the genome of skeletal muscle cells

Table 4-17: Shortlist of differentially expressed genes according to muscle attenuation encoding proteins involved in intracellular structure and vesicle transport

Agilent ID	Gene Symbol	Entrez Gene Name	p-value[1]	FC[2]
A_23_P95810	PKP3	plakophilin 3	4.42E-04	-9.3
A_23_P48835	KIF23‡	kinesin family member 23	4.85E-03	-4.6
A_24_P157342	BRK1	BRICK1, SCAR/WAVE actin-nucleating complex subunit	7.70E-04	-3.6
A_24_P10674	ARHGAP33	Rho GTPase activating protein 33	4.77E-03	-3.4
A_23_P133345	CLINT1	clathrininteractor 1	4.03E-05	-2.6
A_23_P210323	CEP68‡	centrosomal protein 68kDa	1.09E-03	-2.5
A_23_P140434	MYO5C‡	myosin VC	5.55E-03	-2.0
A_23_P56736	TUBA3C	tubulin, α 3c	8.27E-03	-1.9
A_24_P259819	MZT2A	mitotic spindle organizing protein 2A	8.51E-03	-1.9
A_23_P420361	BRK1	BRICK1, SCAR/WAVE actin-nucleating complex subunit	1.40E-03	-1.9
A_23_P137532	PLOD1‡	procollagen-lysine, 2-oxoglutarate 5-dioxygenase 1	9.26E-03	-1.8
A_24_P184799	COCH	coagulation factor C homolog, cochlin (Limulus polyphemus)	3.33E-03	-1.7
A_23_P354387	MYOF	myoferlin	8.43E-05	-1.7
A_23_P389102	MYO1D‡	myosin ID	6.57E-03	-1.7
A_24_P184803	COCH	coagulation factor C homolog, cochlin (Limulus polyphemus)	2.85E-03	-1.7
A_23_P63789	ZWINT	ZW10 interactor	6.27E-03	-1.6
A_23_P109171	BFSP1‡	beaded filament structural protein 1, filensin	1.38E-03	-1.5
A_23_P69326	CADPS‡	Ca++-dependent secretion activator	6.82E-03	-1.5
A_32_P24122	STMN3	stathmin-like 3	5.88E-03	-1.5
A_23_P101193	MYO5B‡	myosin VB	1.53E-03	-1.5
A_24_P532589	MZT1	mitotic spindle organizing protein 1	7.08E-04	1.5
A_24_P405992	SYNPO‡	synaptopodin	6.35E-03	1.5
A_24_P220058	MAPRE1‡	microtubule-associated protein, RP/EB family, member 1	6.81E-03	1.9
A_24_P323522	JAKMIP3	Janus kinase and microtubule interacting protein 3	9.64E-03	1.9
A_23_P70509	ATAT1	α tubulin acetyltransferase 1	1.41E-03	2.2
A_23_P91293	VAPB‡	VAMP (vesicle-associated membrane protein)-associated protein B and C	4.08E-04	2.2
A_23_P156390	JAKMIP2	janus kinase and microtubule interacting protein 2	4.50E-03	2.7
A_23_P213798	SYNPO‡	synaptopodin	8.05E-03	3.9

[1] t-test p-value

[2]FC: fold change. Fold change values that are positive are higher in patients with low muscle attenuation

‡ Androgen response elements identified by Wyce et al. in the genome of skeletal muscle cells

Table 4-18: Shortlist of differentially expressed according to muscle encoding proteins involved in cellular adhesion and extracellular structure

Agilent ID	Symbol	Entrez Gene Name	p-value[1]	FC[2]
A_23_P140797	CDH8‡	cadherin 8, type 2	7.89E-03	-5.7
A_23_P124095	CALML5	calmodulin-like 5	2.89E-03	-3.9
A_23_P25790	CDH24	cadherin 24, type 2	2.76E-03	-3.0
A_32_P168464	CASK‡	calcium/calmodulin-dependent serine protein kinase	6.99E-05	-2.8
A_23_P391926	LPHN1	latrophilin 1	3.83E-03	-2.7
A_23_P156327	TGFBI‡	transforming growth factor, beta-induced, 68kDa	3.11E-03	-2.2
A_23_P16944	SDC1	syndecan 1	2.55E-03	-2.2
A_24_P213944	HEPACAM‡	hepatic and glial cell adhesion molecule	3.15E-03	-2.1
A_23_P128919	LGALS3	lectin, galactoside-binding, soluble, 3	6.48E-03	-2.1
A_23_P133656	LAMA4‡	laminin, α 4	7.41E-03	-2.0
A_23_P55716	BCAM	basal cell adhesion molecule	1.75E-03	-2.0
A_32_P142088	MPZL1‡	myelin protein zero-like 1	1.07E-04	-1.9
A_32_P210390	CAMKK2	Ca2+/calmodulin-dependent kinase kinase 2,β	2.42E-03	-1.8
A_32_P452655	LGALS9C	lectin, galactoside-binding, soluble, 9C	1.22E-03	-1.8
A_23_P110624	CTNND2	catenin, delta 2	1.43E-04	-1.7
A_23_P7397	PCDHB10	protocadherin beta 10	2.89E-04	-1.6
A_24_P409126	FNDC3A‡	fibronectin type III domain containing 3A	9.33E-03	-1.6
A_23_P202823	CTTN	cortactin	1.87E-03	-1.6
A_23_P44291	CRTAP‡	cartilage associated protein	3.88E-04	-1.6
A_23_P160286	PRG4	proteoglycan 4	5.61E-04	-1.6
A_23_P27315	EMILIN2	elastin microfibrilinterfacer 2	2.26E-04	-1.5
A_23_P63557	CNTN2	contactin 2 (axonal)	5.97E-03	-1.5
A_24_P71661	CRTAP‡	cartilage associated protein	1.65E-03	1.5
A_24_P353905	MXRA8‡	matrix-remodelling associated 8	9.72E-04	1.5
A_24_P139152	COL8A1‡	collagen, type VIII, α 1	6.33E-03	1.6
A_23_P56746	FAP‡	fibroblast activation protein, α	5.55E-03	1.6
A_23_P138137	OMA1	OMA1 zinc metallopeptidase homolog	8.51E-05	1.6
A_24_P302506	AMIGO1	adhesion molecule with Ig-like domain 1	2.15E-03	1.6
A_24_P759477	ITGB8‡	integrin, beta 8	4.92E-03	1.7
A_23_P94030	LAMB1‡	laminin, beta 1	7.11E-03	1.7
A_24_P85539	FN1‡	fibronectin 1	8.50E-03	1.7
A_23_P138139	OMA1	OMA1 zinc metallopeptidase homolog	7.09E-05	1.8
A_23_P131614	COL6A3‡	collagen, type VI, α 3	5.90E-03	1.8
A_32_P5040	NOTCH2NL‡	notch 2 N-terminal like	9.30E-03	1.8
A_23_P151805	FBLN5	fibulin 5	4.74E-03	1.9
A_23_P329573	ITGB2‡	integrin, beta 2	7.66E-03	2.0
A_23_P152305	CDH11‡	cadherin 11, type 2, OB-cadherin	6.40E-04	2.0
A_23_P130961	ELANE	elastase, neutrophil expressed	8.71E-03	2.1
A_23_P163787	MMP2‡	matrix metallopeptidase 2	1.27E-03	2.2
A_23_P144959	VCAN‡	Versican	8.12E-04	2.2
A_23_P205841	MYO9A‡	myosin IXA	3.50E-03	2.4
A_23_P151267	LIMA1‡	LIM domain and actin binding 1	9.68E-03	2.4
A_23_P34345	VCAM1‡	vascular cell adhesion molecule 1	2.65E-03	2.8
A_23_P214026	FBN2	fibrillin 2	6.45E-03	3.5
A_23_P347632	MTSS1	metastasis suppressor 1	9.39E-03	4.0
A_24_P252364	NRCAM‡	neuronal cell adhesion molecule	1.45E-03	4.8

[1] t-test p-value

[2]FC: fold change. Positive are higher in patients with low muscle attenuation

‡ Androgen response elements identified by Wyce et al. in the genome of skeletal muscle cells

Table 4-19: Shortlist of differentially expressed according to muscle attenuation encoding proteins involved in muscle and neural function

Agilent ID	Gene Symbol	Entrez Gene Name	p-value[1]	FC[2]
A_24_P398147	NEBL‡	nebulette	3.11E-03	-5.5
A_24_P257022	TNNT2‡	troponin T type 2 (cardiac)	9.09E-03	2.2
A_24_P322709	SNTA1	syntrophin, α 1	6.87E-04	2.6
A_24_P927304	TNNT2‡	troponin T type 2 (cardiac)	1.42E-03	2.6
A_24_P188218	MYL4‡	myosin, light chain 4, alkali; atrial, embryonic	3.97E-03	3.9
A_24_P224488	MAPT	microtubule-associated protein tau	8.17E-03	-3.5
A_23_P500824	SYN1	synapsin I	9.52E-03	-3.2
A_23_P332789	CHRNB4	cholinergic receptor, nicotinic, beta 4	4.38E-03	-3.0
A_23_P35534	NEUROG3	neurogenin 3	1.98E-03	-2.9
A_23_P153549	GRIN2D‡	glutamate receptor, ionotropic, N-methyl D-aspartate 2D	7.06E-03	-2.8
A_23_P21570	NPAS3‡	neuronal PAS domain protein 3	2.61E-03	-2.4
A_23_P86801	RAPSN	receptor-associated protein of the synapse	5.23E-04	-2.2
A_23_P96072	GRIN1	glutamate receptor, ionotropic, N-methyl D-aspartate 1	5.52E-03	-2.2
A_24_P295465	LRRTM3	leucine rich repeat transmembrane 3	3.00E-03	-2.2
A_23_P150162	DRD4‡	dopamine receptor D4	9.30E-03	-2.1
A_23_P212608	CLSTN2‡	calsyntenin 2	4.72E-03	-1.8
A_23_P344451	HDGFRP3‡	hepatoma-derived growth factor, related protein 3	2.90E-03	1.6
A_23_P211522	SYNGR1	synaptogyrin 1	1.17E-03	2.0

[1] t-test p-value
[2] FC: fold change. Positive are higher in patients with low muscle attenuation
‡ Androgen response elements identified by Wyce et al. in the genome of skeletal muscle cells

Figures

Figure 4-1: A conceptual framework of the different inputs from distant organs and tissues (part A, this page) and pathways that are suggested to affect muscle in cancer cachexia (part B, next page)

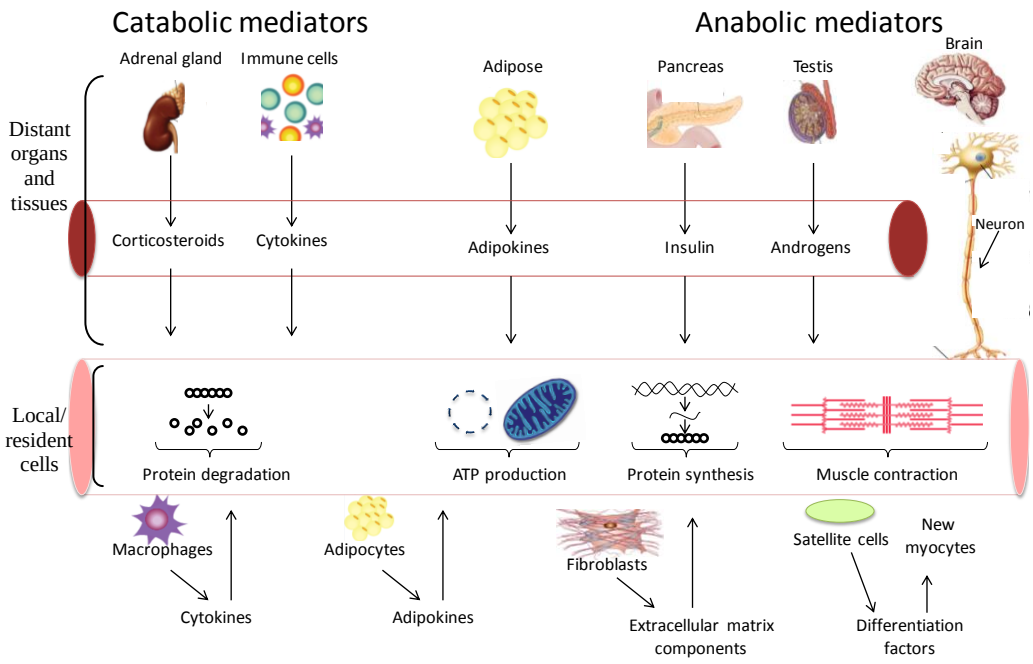


Figure 4-1 part A: Skeletal muscle receives signals from many distant organs and tissues, as well as from neighboring/resident cells within the muscle tissue. Some of these signals will promote catabolism (left side of figure) and others will promote anabolism (right side of figure).

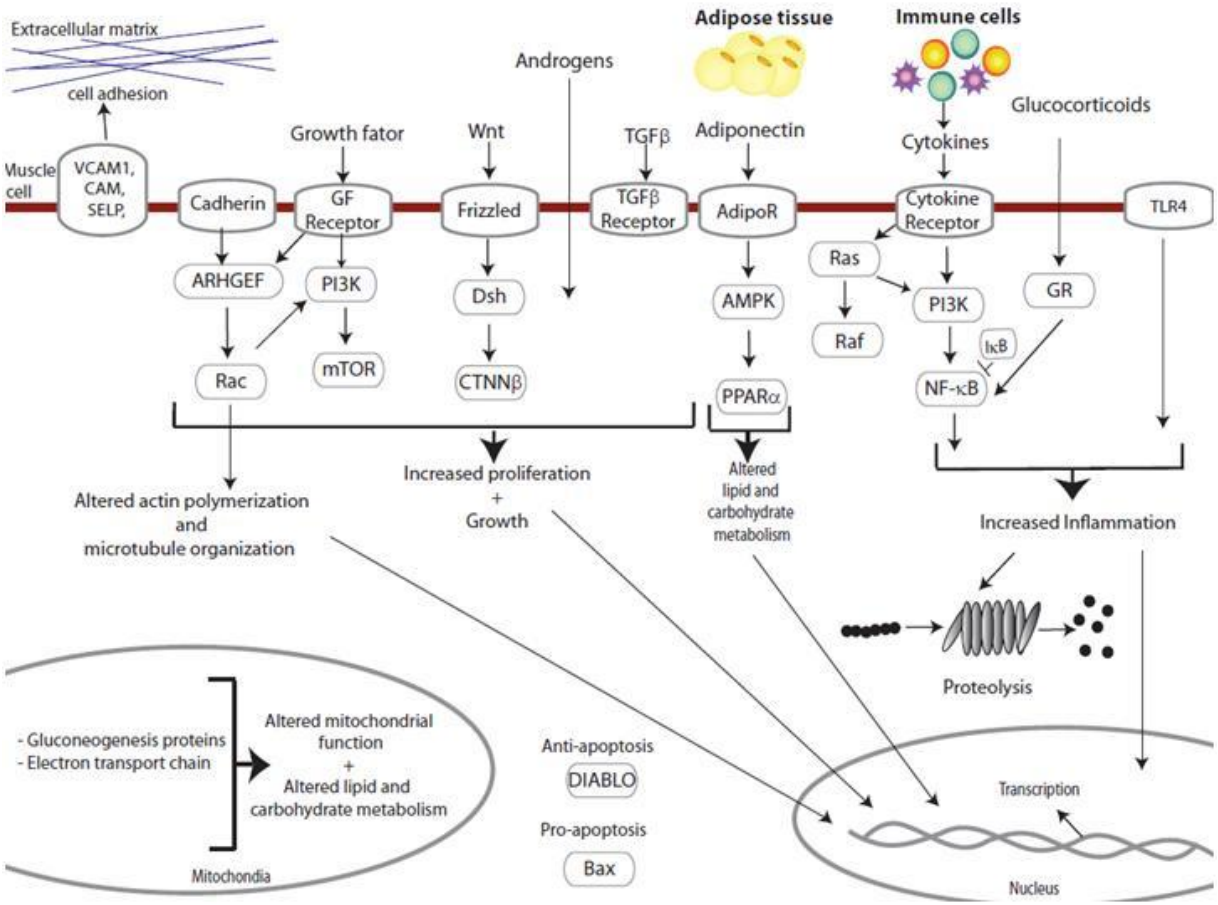
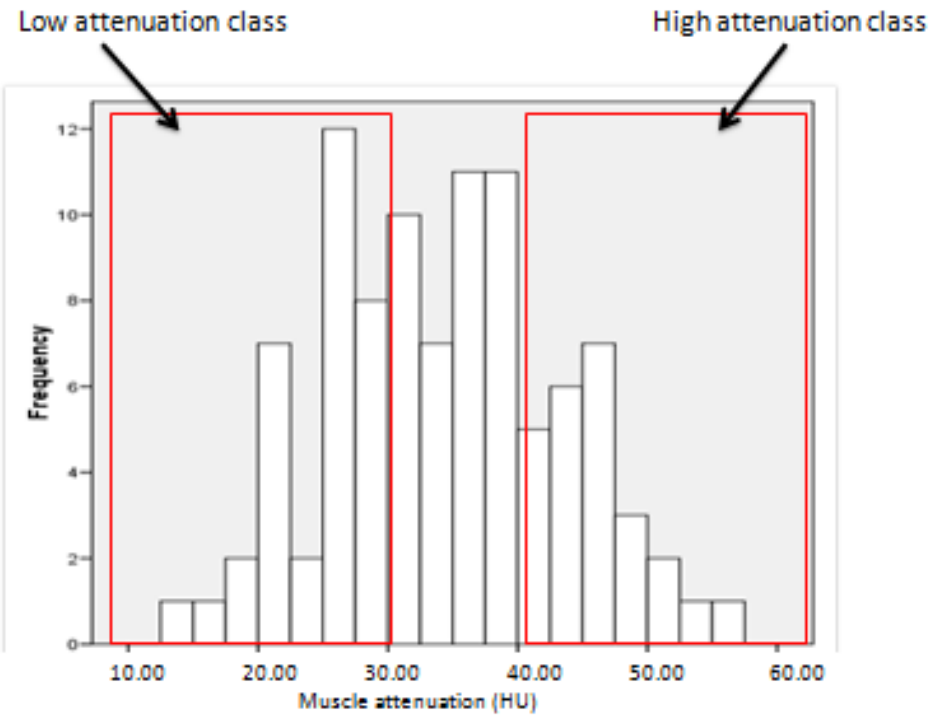


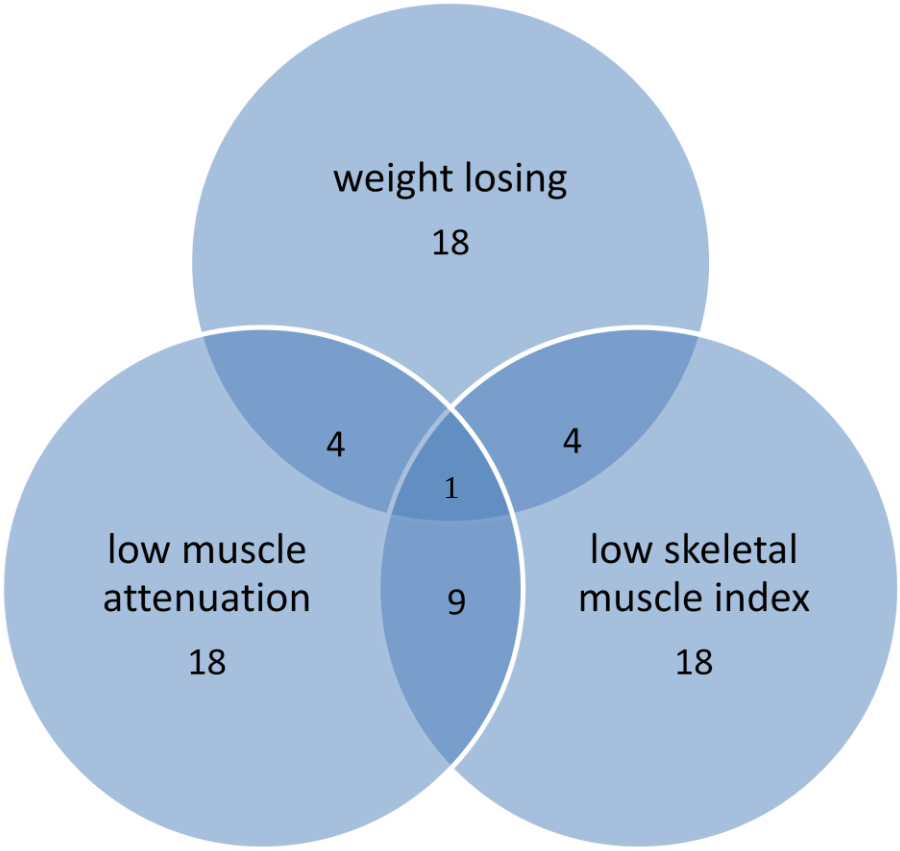
Figure 4-1 part B: The different signals that reach myocytes will ultimately determine if the muscle will undergo catabolism or anabolism. This simplified schematic includes some signalling pathways known to affect skeletal muscle metabolism.

Figure 4-2: An example of how the extreme phenotype classification method was conducted using the distribution of muscle attenuation values for men in this study



The extremes of the phenotypes were selected by selecting the third lowest and highest values of the measured values for each phenotype (in this case muscle attenuation) and excluding the patients in the middle third.

Figure 4-3: Venn diagram showing overlaps between patients classified as being weight losing, having low muscle attenuation and low skeletal muscle index.



Of the patients classified in the weight losing class, only also had low muscle attenuation or low skeletal muscle index. However, there was greater overlap between the patients with low attenuation and low skeletal muscle index.

References

1. Fearon K, Strasser F, Anker SD, Bosaeus I, Bruera E, Fainsinger RL, et al. Definition and classification of cancer cachexia: an international consensus. The lancet oncology. 2011;12:489-95.

2. Esper DH, Harb WA. The cancer cachexia syndrome: a review of metabolic and clinical manifestations. Nutr Clin Pract. 2005;20:369-76.

3. Prado CM, Baracos VE, McCargar LJ, Mourtzakis M, Mulder KE, Reiman T, et al. Body composition as an independent determinant of 5-fluorouracil-based chemotherapy toxicity. Clinical cancer research : an official journal of the American Association for Cancer Research. 2007;13:3264-8.

4. Sandri M. Signaling in muscle atrophy and hypertrophy. Physiology (Bethesda). 2008;23:160-70.

5. Wing SS, Lecker SH, Jagoe RT. Proteolysis in illness-associated skeletal muscle atrophy: from pathways to networks. Critical reviews in clinical laboratory sciences. 2011;48:49-70.

6. Baltgalvis KA, Berger FG, Pena MM, Mark Davis J, White JP, Carson JA. Activity level, apoptosis, and development of cachexia in Apc(Min/+) mice. J Appl Physiol. 2010;109:1155-61.

7. Paul PK, Kumar A. TRAF6 coordinates the activation of autophagy and ubiquitin-proteasome systems in atrophying skeletal muscle. Autophagy. 2011;7:555-6.

8. Schwarzkopf M, Coletti D, Sassoon D, Marazzi G. Muscle cachexia is regulated by a p53-PW1/Peg3-dependent pathway. Genes & development. 2006;20:3440-52.

9. Argiles JM, Busquets S, Toledo M, Lopez-Soriano FJ. The role of cytokines in cancer cachexia. Current opinion in supportive and palliative care. 2009;3:263-8.

10. Menconi M, Fareed M, O'Neal P, Poylin V, Wei W, Hasselgren PO. Role of glucocorticoids in the molecular regulation of muscle wasting. Critical care medicine. 2007;35:S602-8.

11. Glass DJ. Signaling pathways perturbing muscle mass. Current opinion in clinical nutrition and metabolic care. 2010;13:225-9.

12. McCarthy JJ, Esser KA. Anabolic and catabolic pathways regulating skeletal muscle mass. Current opinion in clinical nutrition and metabolic care. 2010;13:230-5.

13. Pallafacchina G, Blaauw B, Schiaffino S. Role of satellite cells in muscle growth and maintenance of muscle mass. Nutrition, metabolism, and cardiovascular diseases : NMCD. 2012.

14. Rommel C, Bodine SC, Clarke BA, Rossman R, Nunez L, Stitt TN, et al. Mediation of IGF-1-induced skeletal myotube hypertrophy by PI(3)K/Akt/mTOR and PI(3)K/Akt/GSK3 pathways. Nature cell biology. 2001;3:1009-13.

15. Burney BO, Hayes TG, Smiechowska J, Cardwell G, Papusha V, Bhargava P, et al. Low testosterone levels and increased inflammatory markers in patients with cancer and relationship with cachexia. The Journal of clinical endocrinology and metabolism. 2012;97:E700-9.

16. Braun TP, Marks DL. Hypothalamic regulation of muscle metabolism. Current opinion in clinical nutrition and metabolic care. 2011;14:237-42.

17. Pasiakos SM. Exercise and amino acid anabolic cell signaling and the regulation of skeletal muscle mass. Nutrients. 2012;4:740-58.

18. Lecker SH, Jagoe RT, Gilbert A, Gomes M, Baracos V, Bailey J, et al. Multiple types of skeletal muscle atrophy involve a common program of changes in gene expression. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2004;18:39-51.

19. Braun TP, Zhu X, Szumowski M, Scott GD, Grossberg AJ, Levasseur PR, et al. Central nervous system inflammation induces muscle atrophy via activation of the hypothalamic-pituitary-adrenal axis. The Journal of experimental medicine. 2011;208:2449-63.

20. Weber MA, Kinscherf R, Krakowski-Roosen H, Aulmann M, Renk H, Kunkele A, et al. Myoglobin plasma level related to muscle mass and fiber composition: a clinical marker of muscle wasting? J Mol Med (Berl). 2007;85:887-96.

21. Weber MA, Krakowski-Roosen H, Schroder L, Kinscherf R, Krix M, Kopp-Schneider A, et al. Morphology, metabolism, microcirculation, and strength of skeletal muscles in cancer-related cachexia. Acta Oncol. 2009;48:116-24.

22. Zampieri S, Doria A, Adami N, Biral D, Vecchiato M, Savastano S, et al. Subclinical myopathy in patients affected with newly diagnosed colorectal cancer at clinical onset of disease: evidence from skeletal muscle biopsies. Neurological research. 2010;32:20-5.

23. Stephens NA, Skipworth RJ, Macdonald AJ, Greig CA, Ross JA, Fearon KC. Intramyocellular lipid droplets increase with progression of cachexia in cancer patients. Journal of cachexia, sarcopenia and muscle. 2011;2:111-7.

24. Bossola M, Muscaritoli M, Costelli P, Bellantone R, Pacelli F, Busquets S, et al. Increased muscle ubiquitin mRNA levels in gastric cancer patients. American journal of physiology Regulatory, integrative and comparative physiology. 2001;280:R1518-23.

25. Bossola M, Muscaritoli M, Costelli P, Grieco G, Bonelli G, Pacelli F, et al. Increased muscle proteasome activity correlates with disease severity in gastric cancer patients. Annals of surgery. 2003;237:384-9.

26. DeJong CH, Busquets S, Moses AG, Schrauwen P, Ross JA, Argiles JM, et al. Systemic inflammation correlates with increased expression of skeletal muscle ubiquitin but not uncoupling proteins in cancer cachexia. Oncology reports. 2005;14:257-63.

27. Jagoe RT, Redfern CP, Roberts RG, Gibson GJ, Goodship TH. Skeletal muscle mRNA levels for cathepsin B, but not components of the ubiquitin-proteasome pathway, are increased in patients with lung cancer referred for thoracotomy. Clin Sci (Lond). 2002;102:353-61.

28. Op den Kamp CM, Langen RC, Minnaard R, Kelders MC, Snepvangers FJ, Hesselink MK, et al. Pre-cachexia in patients with stages I-III non-small cell lung cancer: systemic inflammation and functional impairment without activation of skeletal muscle ubiquitin proteasome system. Lung Cancer. 2012;76:112-7.

29. Williams A, Sun X, Fischer JE, Hasselgren PO. The expression of genes in the ubiquitin-proteasome proteolytic pathway is increased in skeletal muscle from patients with cancer. Surgery. 1999;126:744-9; discussion 9-50.

30. Banduseela V, Ochala J, Lamberg K, Kalimo H, Larsson L. Muscle paralysis and myosin loss in a patient with cancer cachexia. Acta myologica : myopathies and cardiomyopathies : official journal of the Mediterranean Society of Myology / edited by the Gaetano Conte Academy for the study of striated muscle diseases. 2007;26:136-44.

31. Aversa Z, Bonetto A, Penna F, Costelli P, Di Rienzo G, Lacitignola A, et al. Changes in myostatin signaling in non-weight-losing cancer patients. Annals of surgical oncology. 2012;19:1350-6.

32. Busquets S, Deans C, Figueras M, Moore-Carrasco R, Lopez-Soriano FJ, Fearon KC, et al. Apoptosis is present in skeletal muscle of cachectic gastro-intestinal cancer patients. Clin Nutr. 2007;26:614-8.

33. Pessina P, Conti V, Pacelli F, Rosa F, Doglietto GB, Brunelli S, et al. Skeletal muscle of gastric cancer patients expresses genes involved in muscle regeneration. Oncology reports. 2010;24:741-5.

34. Prado CM, Lieffers JR, McCargar LJ, Reiman T, Sawyer MB, Martin L, et al. Prevalence and clinical implications of sarcopenic obesity in patients with solid tumours of the respiratory and gastrointestinal tracts: a population-based study. The lancet oncology. 2008;9:629-35.

35. Baracos VE, Reiman T, Mourtzakis M, Gioulbasanis I, Antoun S. Body composition in patients with non-small cell lung cancer: a contemporary view of cancer cachexia with the use of computed tomography image analysis. The American journal of clinical nutrition. 2010;91:1133S-7S.

36. Di Sebastiano KM, Mourtzakis M. A critical evaluation of body composition modalities used to assess adipose and skeletal muscle tissue in cancer. Applied physiology, nutrition, and metabolism = Physiologie appliquee, nutrition et metabolisme. 2012;37:811-21.

37. Prado CM, Birdsell LA, Baracos VE. The emerging role of computerized tomography in assessing cancer cachexia. Current opinion in supportive and palliative care. 2009;3:269-75.

38. Stephens NA, Gallagher IJ, Rooyackers O, Skipworth RJ, Tan BH, Marstrand T, et al. Using transcriptomics to identify and validate novel biomarkers of human skeletal muscle cancer cachexia. Genome medicine. 2010;2:1.

39. Gallagher IJ, Stephens NA, Macdonald AJ, Skipworth RJ, Husi H, Greig CA, et al. Suppression of skeletal muscle turnover in cancer cachexia: evidence from the transcriptome in sequential human muscle biopsies. Clinical cancer research : an official journal of the American Association for Cancer Research. 2012;18:2817-27.

40. Palomares MR, Sayre JW, Shekar KC, Lillington LM, Chlebowski RT. Gender influence on weight-loss pattern and survival of nonsmall cell lung carcinoma patients. Cancer. 1996;78:2119-26.

41. Stephens NA, Gray C, Macdonald AJ, Tan BH, Gallagher IJ, Skipworth RJ, et al. Sexual dimorphism modulates the impact of cancer cachexia on lower limb muscle mass and function. Clin Nutr. 2012;31:499-505.

42. Roman Eisner CS, Thomas Eastman, Jianguo Xia, David Hau, Sambasivarao Damaraju, Russell Greiner, David S. Wishart and Vickie E. Baracos. Learning to predict cancer-associated skeletal muscle wasting from 1H-NMR profiles of urinary metabolites. Metabolomics. 2011;7:25-34.

43. Goodpaster BH, Kelley DE, Thaete FL, He J, Ross R. Skeletal muscle attenuation determined by computed tomography is associated with skeletal muscle lipid content. J Appl Physiol. 2000;89:104-10.

44. Miljkovic I, Zmuda JM. Epidemiology of myosteatosis. Current opinion in clinical nutrition and metabolic care. 2010;13:260-4.

45. Perez-Gracia JL, Gloria Ruiz-Ilundain M, Garcia-Ribas I, Maria Carrasco E. The role of extreme phenotype selection studies in the identification of clinically relevant genotypes in cancer research. Cancer. 2002;95:1605-10.

46. Nebert DW. Extreme discordant phenotype methodology: an intuitive approach to clinical pharmacogenetics. European journal of pharmacology. 2000;410:107-20.

47. Perez-Gracia JL, Gurpide A, Ruiz-Ilundain MG, Alfaro Alegria C, Colomer R, Garcia-Foncillas J, et al. Selection of extreme phenotypes: the role of clinical observation in translational research. Clinical & translational oncology : official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico. 2010;12:174-80.

48. Li D, Lewinger JP, Gauderman WJ, Murcray CE, Conti D. Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies. Genetic epidemiology. 2011;35:790-9.

50. Leotoing L, Chereau F, Baron S, Hube F, Valencia HJ, Bordereaux D, et al. A20-binding inhibitor of nuclear factor-kappaB (NF-kappaB)-2 (ABIN-2) is an activator of inhibitor of NF-kappaB (IkappaB) kinase alpha (IKKalpha)-mediated NF-kappaB transcriptional activity. The Journal of biological chemistry. 2011;286:32277-88.

51. Jin HR, Jin X, Lee JJ. Zinc-finger protein 91 plays a key role in LIGHT-induced activation of non-canonical NF-kappaB pathway. Biochemical and biophysical research communications. 2010;400:581-6.

52. Adhikary G, Gupta S, Sil P, Saad Y, Sen S. Characterization and functional significance of myotrophin: a gene with multiple transcripts. Gene. 2005;353:31-40.

53. Frisard MI, McMillan RP, Marchand J, Wahlberg KA, Wu Y, Voelker KA, et al. Toll-like receptor 4 modulates skeletal muscle substrate metabolism. American journal of physiology Endocrinology and metabolism. 2010;298:E988-98.

54. Kimura N, Itoh S, Nakae S, Axtell RC, Velotta JB, Bos EJ, et al. Interleukin-16 deficiency suppresses the development of chronic rejection in murine cardiac transplantation model. The Journal of heart and lung transplantation : the official publication of the International Society for Heart Transplantation. 2011;30:1409-17.

55. Tosi G, Bozzo L, Accolla RS. The dual function of the MHC class II transactivator CIITA against HTLV retroviruses. Frontiers in bioscience : a journal and virtual library. 2009;14:4149-56.

56. Francis JN, Jacobson MR, Lloyd CM, Sabroe I, Durham SR, Till SJ. CXCR1+CD4+ T cells in human allergic disease. J Immunol. 2004;172:268-73.

57. Kitamura N, Mori A, Tatsumi H, Nemoto S, Hiroi T, Kaminuma O. Zinc finger protein, multitype 1, suppresses human Th2 development via downregulation of IL-4. International archives of allergy and immunology. 2011;155 Suppl 1:53-6.

58. Kyriakides C, Austen W, Jr., Wang Y, Favuzza J, Kobzik L, Moore FD, Jr., et al. Skeletal muscle reperfusion injury is mediated by neutrophils and the complement membrane attack complex. The American journal of physiology. 1999;277:C1263-8.

59. Jia W, Li H, He YW. The extracellular matrix protein mindin serves as an integrin ligand and is critical for inflammatory cell recruitment. Blood. 2005;106:3854-9.

60. He YW, Li H, Zhang J, Hsu CL, Lin E, Zhang N, et al. The extracellular matrix protein mindin is a pattern-recognition molecule for microbial pathogens. Nature immunology. 2004;5:88-97.

61. Borte S, Pan-Hammarstrom Q, Liu C, Sack U, Borte M, Wagner U, et al. Interleukin-21 restores immunoglobulin production ex vivo in patients with common variable immunodeficiency and selective IgA deficiency. Blood. 2009;114:4089-98.

62. Nagaraju K, Raben N, Villalba ML, Danning C, Loeffler LA, Lee E, et al. Costimulatory markers in muscle of patients with idiopathic inflammatory myopathies and in cultured muscle cells. Clin Immunol. 1999;92:161-9.

63. Spandl J, Lohmann D, Kuerschner L, Moessinger C, Thiele C. Ancient ubiquitous protein 1 (AUP1) localizes to lipid droplets and binds the E2 ubiquitin conjugase G2 (Ube2g2) via its G2 binding region. The Journal of biological chemistry. 2011;286:5599-606.

64. Bae Y, Kho CW, Lee SY, Rhim H, Kang S. Hip2 interacts with and destabilizes Smac/DIABLO. Biochemical and biophysical research communications. 2010;397:718-23.

65. Kroismayr R, Baranyi U, Stehlik C, Dorfleutner A, Binder BR, Lipp J. HERC5, a HECT E3 ubiquitin ligase tightly regulated in LPS activated endothelial cells. Journal of cell science. 2004;117:4749-56.

66. Zoabi M, Sadeh R, de Bie P, Marquez VE, Ciechanover A. PRAJA1 is a ubiquitin ligase for the polycomb repressive complex 2 proteins. Biochemical and biophysical research communications. 2011;408:393-8.

67. Lin DI, Barbash O, Kumar KG, Weber JD, Harper JW, Klein-Szanto AJ, et al. Phosphorylation-dependent ubiquitination of cyclin D1 by the SCF(FBX4-alphaB crystallin) complex. Molecular cell. 2006;24:355-66.

68. Srinivasan S, Nawaz Z. E3 ubiquitin protein ligase, E6-associated protein (E6-AP) regulates PI3K-Akt signaling and prostate cell growth. Biochimica et biophysica acta. 2011;1809:119-27.

69. Xiao GG, Zhou BS, Somlo G, Portnow J, Juhasz A, Un F, et al. Identification of F-box/LLR-repeated protein 17 as potential useful biomarker for breast cancer therapy. Cancer genomics & proteomics. 2008;5:151-60.

70. Didier C, Broday L, Bhoumik A, Israeli S, Takahashi S, Nakayama K, et al. RNF5, a RING finger protein that regulates cell motility by targeting paxillin ubiquitination and altered localization. Molecular and cellular biology. 2003;23:5331-45.

71. Krawiec P, Gluszko P, Kwasny-Krochin B, Undas A. Decreased proteinZ levels in patients with rheumatoid arthritis: links with inflammation. Thrombosis and haemostasis. 2011;106:548-50.

72. Jin J, Li X, Gygi SP, Harper JW. Dual E1 activation systems for ubiquitin differentially regulate E2 enzyme charging. Nature. 2007;447:1135-8.

73. Shaik S, Nucera C, Inuzuka H, Gao D, Garnaas M, Frechette G, et al. SCF(beta-TRCP) suppresses angiogenesis and thyroid cancer cell migration by promoting ubiquitination and destruction of VEGF receptor 2. The Journal of experimental medicine. 2012;209:1289-307.

74. Jiang H, Lu Y, Yuan L, Liu J. Regulation of interleukin-10 receptor ubiquitination and stability by beta-TrCP-containing ubiquitin E3 ligase. PloS one. 2011;6:e27464.

75. Lagirand-Cantaloube J, Cornille K, Csibi A, Batonnet-Pichon S, Leibovitch MP, Leibovitch SA. Inhibition of atrogin-1/MAFbx mediated MyoD proteolysis prevents skeletal muscle atrophy in vivo. PloS one. 2009;4:e4973.

76. Komuro A, Imamura T, Saitoh M, Yoshida Y, Yamori T, Miyazono K, et al. Negative regulation of transforming growth factor-beta (TGF-beta) signaling by WW domain-containing protein 1 (WWP1). Oncogene. 2004;23:6914-23.

77. Webber E, Li L, Chin LS. Hypertonia-associated protein Trak1 is a novel regulator of endosome-to-lysosome trafficking. Journal of molecular biology. 2008;382:638-51.

78. Wiendl H, Lautwein A, Mitsdorffer M, Krause S, Erfurth S, Wienhold W, et al. Antigen processing and presentation in human muscle: cathepsin S is critical for MHC class II expression and upregulated in inflammatory myopathies. Journal of neuroimmunology. 2003;138:132-43.

79. Laine DI, Busch-Petersen J. Inhibitors of cathepsin C (dipeptidyl peptidase I). Expert opinion on therapeutic patents. 2010;20:497-506.

80. Jung IL, Kang HJ, Kim KC, Kim IG. Knockdown of the Dickkopf 3 gene induces apoptosis in a lung adenocarcinoma. International journal of molecular medicine. 2010;26:33-8.

81. Sandow JJ, Jabbour AM, Condina MR, Daunt CP, Stomski FC, Green BD, et al. Cytokine receptor signaling activates an IKK-dependent phosphorylation of PUMA to prevent cell death. Cell death and differentiation. 2012;19:633-41.

82. Wei JD, Lin YL, Tsai CH, Shieh HS, Lin PI, Ho WP, et al. SATB2 participates in regulation of menadione-induced apoptotic insults to osteoblasts. Journal of orthopaedic research : official publication of the Orthopaedic Research Society. 2012;30:1058-66.

83. Wang X, Narayanan M, Bruey JM, Rigamonti D, Cattaneo E, Reed JC, et al. Protective role of Cop in Rip2/caspase-1/caspase-4-mediated HeLa cell death. Biochimica et biophysica acta. 2006;1762:742-54.

84. Dai D, Liang Y, Xie Z, Fu J, Zhang Y, Zhang Z. Survivin deficiency induces apoptosis and cell cycle arrest in HepG2 hepatocellular carcinoma cells. Oncology reports. 2012;27:621-7.

85. Zhang H, Xu Q, Krajewski S, Krajewska M, Xie Z, Fuess S, et al. BAR: An apoptosis regulator at the intersection of caspases and Bcl-2 family proteins. Proceedings of the National Academy of Sciences of the United States of America. 2000;97:2597-602.

86. Polimeno L, Pesetti B, De Santis F, Resta L, Rossi R, De Palma A, et al. Decreased expression of the augmenter of liver regeneration results in increased apoptosis and oxidative damage in human-derived glioma cells. Cell death & disease. 2012;3:e289.

87. Armand AS, Lecolle S, Launay T, Pariset C, Fiore F, Della Gaspera B, et al. IGF-II is up-regulated and myofibres are hypertrophied in regenerating soleus of mice lacking FGF6. Experimental cell research. 2004;297:27-38.

88. Hu J, Higuchi I, Yoshida Y, Shiraishi T, Osame M. Expression of midkine in regenerating skeletal muscle fibers and cultured myoblasts of human skeletal muscle. European neurology. 2002;47:20-5.

89. Trobec K, von Haehling S, Anker SD, Lainscak M. Growth hormone, insulin-like growth factor 1, and insulin signaling-a pharmacological target in body wasting and cachexia. Journal of cachexia, sarcopenia and muscle. 2011;2:191-200.

90. Friedrichs M, Wirsdoerfer F, Flohe SB, Schneider S, Wuelling M, Vortkamp A. BMP signaling balances proliferation and differentiation of muscle satellite cell descendants. BMC cell biology. 2011;12:26.

91. Stewart JD, Masi TL, Cumming AE, Molnar GM, Wentworth BM, Sampath K, et al. Characterization of proliferating human skeletal muscle-derived cells in vitro: differential modulation of myoblast markers by TGF-beta2. Journal of cellular physiology. 2003;196:70-8.

92. Chockalingam PS, Cholera R, Oak SA, Zheng Y, Jarrett HW, Thomason DB. Dystrophin-glycoprotein complex and Ras and Rho GTPase signaling are altered in muscle atrophy. American journal of physiology Cell physiology. 2002;283:C500-11.

93. Vaidya TB, Weyman CM, Teegarden D, Ashendel CL, Taparowsky EJ. Inhibition of myogenesis by the H-ras oncogene: implication of a role for protein kinase C. The Journal of cell biology. 1991;114:809-20.

94. Scholz ME, Meissner JD, Scheibe RJ, Umeda PK, Chang KC, Gros G, et al. Different roles of H-ras for regulation of myosin heavy chain promoters in satellite cell-derived muscle cell culture during proliferation and differentiation. American journal of physiology Cell physiology. 2009;297:C1012-8.

95. Chiu TT, Jensen TE, Sylow L, Richter EA, Klip A. Rac1 signalling towards GLUT4/glucose uptake in skeletal muscle. Cellular signalling. 2011;23:1546-54.

96. Kanzaki M. Insulin receptor signals regulating GLUT4 translocation and actin dynamics. Endocrine journal. 2006;53:267-93.

97. Coisy-Quivy M, Touzet O, Bourret A, Hipskind RA, Mercier J, Fort P, et al. TC10 controls human myofibril organization and is activated by the sarcomeric RhoGEF obscurin. Journal of cell science. 2009;122:947-56.

98. Wu D, Chapman JR, Wang L, Harris TE, Shabanowitz J, Hunt DF, et al. Intestinal cell kinase (ICK) promotes activation of mTOR complex 1 (mTORC1) through phosphorylation of Raptor Thr-908. The Journal of biological chemistry. 2012;287:12510-9.

99. Fang Y, Park IH, Wu AL, Du G, Huang P, Frohman MA, et al. PLD1 regulates mTOR signaling and mediates Cdc42 activation of S6K1. Current biology : CB. 2003;13:2037-44.

100. Fearon K, Arends J, Baracos V. Understanding the mechanisms and treatment options in cancer cachexia. Nature reviews Clinical oncology. 2013;10:90-9.

101. Hu SI, Katz M, Chin S, Qi X, Cruz J, Ibebunjo C, et al. MNK2 inhibits eIF4G activation through a pathway involving serine-arginine-rich protein kinase in skeletal muscle. Science signaling. 2012;5:ra14.

102. Chen HS, Jia J, Su HF, Lin HD, Chen JW, Lin SJ, et al. Downregulation of the constitutively expressed Hsc70 in diabetic myocardium is mediated by insulin deficiency. The Journal of endocrinology. 2006;190:433-40.

103. Lu M, Wang J, Ives HE, Pearce D. mSIN1 protein mediates SGK1 protein interaction with mTORC2 protein complex and is required for selective activation of the epithelial sodium channel. The Journal of biological chemistry. 2011;286:30647-54.

104. Manrique C, Lastra G, Habibi J, Mugerfeld I, Garro M, Sowers JR. Loss of Estrogen Receptor alpha Signaling Leads to Insulin Resistance and Obesity in Young and Adult Female Mice. Cardiorenal medicine. 2012;2:200-10.

105. Ligr M, Patwa RR, Daniels G, Pan L, Wu X, Li Y, et al. Expression and function of androgen receptor coactivator p44/Mep50/WDR77 in ovarian cancer. PloS one. 2011;6:e26250.

106. Wyce A, Bai Y, Nagpal S, Thompson CC. Research Resource: The androgen receptor modulates expression of genes with critical roles in muscle development and function. Mol Endocrinol. 2010;24:1665-74.

107. Brack AS, Conboy MJ, Roy S, Lee M, Kuo CJ, Keller C, et al. Increased Wnt signaling during aging alters muscle stem cell fate and increases fibrosis. Science. 2007;317:807-10.

108. Armstrong DD, Esser KA. Wnt/beta-catenin signaling activates growth-control genes during overload-induced skeletal muscle hypertrophy. American journal of physiology Cell physiology. 2005;289:C853-9.

109. Zhou D, Strakovsky RS, Zhang X, Pan YX. The skeletal muscle Wnt pathway may modulate insulin resistance and muscle development in a diet-induced obese rat model. Obesity (Silver Spring). 2012;20:1577-84.

110. Baker DJ, Wijshake T, Tchkonia T, LeBrasseur NK, Childs BG, van de Sluis B, et al. Clearance of p16Ink4a-positive senescent cells delays ageing-associated disorders. Nature. 2011;479:232-6.

111. Abida WM, Nikolaev A, Zhao W, Zhang W, Gu W. FBXO11 promotes the Neddylation of p53 and inhibits its transcriptional activity. The Journal of biological chemistry. 2007;282:1797-804.

112. Bomont P, Maddox P, Shah JV, Desai AB, Cleveland DW. Unstable microtubule capture at kinetochores depleted of the centromere-associated protein CENP-F. The EMBO journal. 2005;24:3927-39.

113. Winkler GS. The mammalian anti-proliferative BTG/Tob protein family. Journal of cellular physiology. 2010;222:66-72.

114. Chen L, Yuan D, Zhao R, Li H, Zhu J. Suppression of TSPAN1 by RNA interference inhibits proliferation and invasion of colon cancer cells in vitro. Tumori. 2010;96:744-50.

115. Moresi V, Garcia-Alvarez G, Pristera A, Rizzuto E, Albertini MC, Rocchi M, et al. Modulation of caspase activity regulates skeletal muscle regeneration and function in response to vasopressin and tumor necrosis factor. PloS one. 2009;4:e5570.

116. Vickerman L, Neufeld S, Cobb J. Shox2 function couples neural, muscular and skeletal development in the proximal forelimb. Developmental biology. 2011;350:323-36.

117. Risebro CA, Searles RG, Melville AA, Ehler E, Jina N, Shah S, et al. Prox1 maintains muscle structure and growth in the developing heart. Development. 2009;136:495-505.

118. Miyamoto S, Hidaka K, Jin D, Morisaki T. RNA-binding proteins Rbm38 and Rbm24 regulate myogenic differentiation via p21-dependent and -independent regulatory pathways. Genes to cells : devoted to molecular & cellular mechanisms. 2009;14:1241-52.

119. Penn BH, Bergstrom DA, Dilworth FJ, Bengal E, Tapscott SJ. A MyoD-generated feed-forward circuit temporally patterns gene expression during skeletal muscle differentiation. Genes & development. 2004;18:2348-53.

120. Klover P, Chen W, Zhu BM, Hennighausen L. Skeletal muscle growth and fiber composition in mice are regulated through the transcription factors STAT5a/b: linking growth hormone to the androgen receptor. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2009;23:3140-8.

121. Zhu R, Yang Y, Tian Y, Bai J, Zhang X, Li X, et al. Ascl2 knockdown results in tumor growth arrest by miRNA-302b-related inhibition of colon cancer progenitor cells. PloS one. 2012;7:e32170.

122. Kim KS, Jung HS, Chung YJ, Jung TS, Jang HW, Lee MS, et al. Overexpression of USF increases TGF-beta1 protein levels, but G1 phase arrest was not induced in FRTL-5 cells. Journal of Korean medical science. 2008;23:870-6.

123. Than TA, Lou H, Ji C, Win S, Kaplowitz N. Role of cAMP-responsive element-binding protein (CREB)-regulated transcription coactivator 3 (CRTC3) in the initiation of mitochondrial biogenesis and stress response in liver cells. The Journal of biological chemistry. 2011;286:22047-54.

124. Hao H, d'Alincourt-Salazar M, Kelley KM, Shatnawi A, Mukherjee S, Shah YM, et al. Estrogen-induced and TAFII30-mediated gene repression by direct recruitment of the estrogen receptor and co-repressors to the core promoter and its reversal by tamoxifen. Oncogene. 2007;26:7872-84.

125. Luo JQ, Chen DW, Yu B. Upregulation of amino acid transporter expression induced by l-leucine availability in L6 myotubes is associated with ATF4 signaling through mTORC1-dependent mechanism. Nutrition. 2012.

126. Gersbach CA, Byers BA, Pavlath GK, Garcia AJ. Runx2/Cbfa1 stimulates transdifferentiation of primary skeletal myoblasts into a mineralizing osteoblastic phenotype. Experimental cell research. 2004;300:406-17.

127. Coin F, Proietti De Santis L, Nardo T, Zlobinskaya O, Stefanini M, Egly JM. p8/TTD-A as a repair-specific TFIIH subunit. Molecular cell. 2006;21:215-26.

128. Sanderson LM, Boekschoten MV, Desvergne B, Muller M, Kersten S. Transcriptional profiling reveals divergent roles of PPARalpha and PPARbeta/delta in regulation of gene expression in mouse liver. Physiological genomics. 2010;41:42-52.

129. Feingold KR, Wang Y, Moser A, Shigenaga JK, Grunfeld C. LPS decreases fatty acid oxidation and nuclear hormone receptors in the kidney. Journal of lipid research. 2008;49:2179-87.

130. Schindler A, Foley E. A functional RNAi screen identifies hexokinase 1 as a modifier of type II apoptosis. Cellular signalling. 2010;22:1330-40.

131. Gullberg H, Rudling M, Salto C, Forrest D, Angelin B, Vennstrom B. Requirement for thyroid hormone receptor beta in T3 regulation of cholesterol metabolism in mice. Mol Endocrinol. 2002;16:1767-77.

132. Johansson C, Lannergren J, Lunde PK, Vennstrom B, Thoren P, Westerblad H. Isometric force and endurance in soleus muscle of thyroid hormone receptor-alpha(1)- or -beta-deficient mice. American journal of physiology Regulatory, integrative and comparative physiology. 2000;278:R598-603.

133. Naranjo-Suarez S, Carlson BA, Tsuji PA, Yoo MH, Gladyshev VN, Hatfield DL. HIF-independent regulation of thioredoxin reductase 1 contributes to the high levels of reactive oxygen species induced by hypoxia. PloS one. 2012;7:e30470.

134. Ki SH, Cho IJ, Choi DW, Kim SG. Glucocorticoid receptor (GR)-associated SMRT binding to C/EBPbeta TAD and Nrf2 Neh4/5: role of SMRT recruited to GR in GSTA2 gene repression. Molecular and cellular biology. 2005;25:4150-65.

135. Tan BH, Fladvad T, Braun TP, Vigano A, Strasser F, Deans DA, et al. P-selectin genotype is associated with the development of cancer cachexia. EMBO molecular medicine. 2012;4:462-71.

136. Kwon MJ, Han J, Kim BH, Lee YS, Kim TY. Superoxide dismutase 3 suppresses hyaluronic acid fragments mediated skin inflammation by inhibition of toll-like receptor 4 signaling pathway: superoxide dismutase 3 inhibits reactive oxygen species-induced trafficking of toll-like receptor 4 to lipid rafts. Antioxidants & redox signaling. 2012;16:297-313.

137. Hsu HH, Hoffmann S, Di Marco GS, Endlich N, Peter-Katalinic J, Weide T, et al. Downregulation of the antioxidant protein peroxiredoxin 2 contributes to angiotensin II-mediated podocyte apoptosis. Kidney international. 2011;80:959-69.

138. Elliott NA, Volkert MR. Stress induction and mitochondrial localization of Oxr1 proteins in yeast and humans. Molecular and cellular biology. 2004;24:3180-7.

139. Belyaeva OV, Korkina OV, Stetsenko AV, Kedishvili NY. Human retinol dehydrogenase 13 (RDH13) is a mitochondrial short-chain dehydrogenase/reductase with a retinaldehyde reductase activity. The FEBS journal. 2008;275:138-47.

140. Pridgeon JW, Olzmann JA, Chin LS, Li L. PINK1 protects against oxidative stress by phosphorylating mitochondrial chaperone TRAP1. PLoS biology. 2007;5:e172.

141. Morel F, Rauch C, Petit E, Piton A, Theret N, Coles B, et al. Gene and protein characterization of the human glutathione S-transferase kappa and evidence for a peroxisomal localization. The Journal of biological chemistry. 2004;279:16246-53.

142. Joseph AM, Hood DA. Plasticity of TOM complex assembly in skeletal muscle mitochondria in response to chronic contractile activity. Mitochondrion. 2012;12:305-12.

143. Pierce SB, Chisholm KM, Lynch ED, Lee MK, Walsh T, Opitz JM, et al. Mutations in mitochondrial histidyl tRNA synthetase HARS2 cause ovarian dysgenesis and sensorineural hearing loss of Perrault syndrome. Proceedings of the National Academy of Sciences of the United States of America. 2011;108:6543-8.

144. Muoio DM, Dohm GL, Fiedorek FT, Jr., Tapscott EB, Coleman RA. Leptin directly alters lipid partitioning in skeletal muscle. Diabetes. 1997;46:1360-3.

145. Ceddia RB, William WN, Jr., Curi R. The response of skeletal muscle to leptin. Frontiers in bioscience : a journal and virtual library. 2001;6:D90-7.

146. Olmedillas H, Sanchis-Moysi J, Fuentes T, Guadalupe-Grau A, Ponce-Gonzalez JG, Morales-Alamo D, et al. Muscle hypertrophy and increased expression of leptin receptors in the musculus triceps brachii of the dominant arm in professional tennis players. European journal of applied physiology. 2010;108:749-58.

147. Uddin S, Bu R, Ahmed M, Hussain AR, Ajarim D, Al-Dayel F, et al. Leptin receptor expression and its association with PI3K/AKT signaling pathway in diffuse large B-cell lymphoma. Leukemia & lymphoma. 2010;51:1305-14.

148. Cheng JB, Russell DW. Mammalian wax biosynthesis. I. Identification of two fatty acyl-Coenzyme A reductases with different substrate specificities and tissue distributions. The Journal of biological chemistry. 2004;279:37789-97.

149. Yamauchi T, Nio Y, Maki T, Kobayashi M, Takazawa T, Iwabu M, et al. Targeted disruption of AdipoR1 and AdipoR2 causes abrogation of adiponectin binding and metabolic actions. Nature medicine. 2007;13:332-9.

150. Finck BN, Gropler MC, Chen Z, Leone TC, Croce MA, Harris TE, et al. Lipin 1 is an inducible amplifier of the hepatic PGC-1alpha/PPARalpha regulatory pathway. Cell metabolism. 2006;4:199-210.

151. Ryu D, Seo WY, Yoon YS, Kim YN, Kim SS, Kim HJ, et al. Endoplasmic reticulum stress promotes LIPIN2-dependent hepatic insulin resistance. Diabetes. 2011;60:1072-81.

152. Cartwright BR, Goodman JM. Seipin: from human disease to molecular mechanism. Journal of lipid research. 2012;53:1042-55.

153. Bosma M, Minnaard R, Sparks LM, Schaart G, Losen M, de Baets MH, et al. The lipid droplet coat protein perilipin 5 also localizes to muscle mitochondria. Histochemistry and cell biology. 2012;137:205-16.

154. Takahashi S, Fusaki N, Ohta S, Iwahori Y, Iizuka Y, Inagawa K, et al. Downregulation of KIF23 suppresses glioma proliferation. Journal of neuro-oncology. 2012;106:519-29.

155. Miller SE, Collins BM, McCoy AJ, Robinson MS, Owen DJ. A SNARE-adaptor interaction is a new mode of cargo recognition in clathrin-coated vesicles. Nature. 2007;450:570-4.

156. Jacobs DT, Weigert R, Grode KD, Donaldson JG, Cheney RE. Myosin Vc is a molecular motor that functions in secretory granule trafficking. Molecular biology of the cell. 2009;20:4471-88.

157. Posey AD, Jr., Demonbreun A, McNally EM. Ferlin proteins in myoblast fusion and muscle growth. Current topics in developmental biology. 2011;96:203-30.

158. Kundu ST, Gosavi P, Khapare N, Patel R, Hosing AS, Maru GB, et al. Plakophilin3 downregulation leads to a decrease in cell adhesion and promotes metastasis. International journal of cancer Journal international du cancer. 2008;123:2303-14.

159. Graser S, Stierhof YD, Nigg EA. Cep68 and Cep215 (Cdk5rap2) are required for centrosome cohesion. Journal of cell science. 2007;120:4321-31.

160. Escobar B, de Carcer G, Fernandez-Miranda G, Cascon A, Bravo-Cordero JJ, Montoya MC, et al. Brick1 is an essential regulator of actin cytoskeleton

required for embryonic development and cell transformation. Cancer research. 2010;70:9349-59.

161. Rosario M, Franke R, Bednarski C, Birchmeier W. The neurite outgrowth multiadaptor RhoGAP, NOMA-GAP, regulates neurite extension through SHP2 and Cdc42. The Journal of cell biology. 2007;178:503-16.

162. Asanuma K, Kim K, Oh J, Giardino L, Chabanis S, Faul C, et al. Synaptopodin regulates the actin-bundling activity of alpha-actinin in an isoform-specific manner. The Journal of clinical investigation. 2005;115:1188-98.

163. Cruz-Garcia D, Diaz-Ruiz A, Rabanal-Ruiz Y, Peinado JR, Gracia-Navarro F, Castano JP, et al. The Golgi-associated long coiled-coil protein NECC1 participates in the control of the regulated secretory pathway in PC12 cells. The Biochemical journal. 2012;443:387-96.

164. Cruz-Garcia D, Vazquez-Martinez R, Peinado JR, Anouar Y, Tonon MC, Vaudry H, et al. Identification and characterization of two novel (neuro)endocrine long coiled-coil proteins. FEBS letters. 2007;581:3149-56.

165. Angst BD, Marcozzi C, Magee AI. The cadherin superfamily: diversity in form and function. Journal of cell science. 2001;114:629-41.

166. Sanford JL, Mays TA, Varian KD, Wilson JB, Janssen PM, Rafael-Fortney JA. Truncated CASK does not alter skeletal muscle or protein interactors. Muscle & nerve. 2008;38:1116-27.

167. Nishimune H, Valdez G, Jarad G, Moulson CL, Muller U, Miner JH, et al. Laminins promote postsynaptic maturation by an autocrine mechanism at the neuromuscular junction. The Journal of cell biology. 2008;182:1201-15.

168. Madhavan R, Gong ZL, Ma JJ, Chan AW, Peng HB. The function of cortactin in the clustering of acetylcholine receptors at the vertebrate neuromuscular junction. PloS one. 2009;4:e8478.

169. Rodova M, Kelly KF, VanSaun M, Daniel JM, Werle MJ. Regulation of the rapsyn promoter by kaiso and delta-catenin. Molecular and cellular biology. 2004;24:7188-96.

170. Sabatelli P, Gualandi F, Gara SK, Grumati P, Zamparelli A, Martoni E, et al. Expression of collagen VI alpha5 and alpha6 chains in human muscle and in Duchenne muscular dystrophy-related muscle fibrosis. Matrix biology : journal of the International Society for Matrix Biology. 2012;31:187-96.

171. Morales MG, Vazquez Y, Acuna MJ, Rivera JC, Simon F, Salas JD, et al. Angiotensin II-induced pro-fibrotic effects require p38MAPK activity and

transforming growth factor beta 1 expression in skeletal muscle cells. The international journal of biochemistry & cell biology. 2012;44:1993-2002.

172. Brinckmann J, Hunzelmann N, Kahle B, Rohwedel J, Kramer J, Gibson MA, et al. Enhanced fibrillin-2 expression is a general feature of wound healing and sclerosis: potential alteration of cell attachment and storage of TGF-beta. Laboratory investigation; a journal of technical methods and pathology. 2010;90:739-52.

173. Smith CA, Stauber F, Waters C, Alway SE, Stauber WT. Transforming growth factor-beta following skeletal muscle strain injury in rats. J Appl Physiol. 2007;102:755-61.

174. Dawson JC, Bruche S, Spence HJ, Braga VM, Machesky LM. Mtss1 promotes cell-cell junction assembly and stability through the small GTPase Rac1. PloS one. 2012;7:e31141.

175. Jesse TL, LaChance R, Iademarco MF, Dean DC. Interferon regulatory factor-2 is a transcriptional activator in muscle where It regulates expression of vascular cell adhesion molecule-1. The Journal of cell biology. 1998;140:1265-76.

176. Omelchenko T, Hall A. Myosin-IXA regulates collective epithelial cell migration by targeting RhoGAP activity to cell-cell junctions. Current biology : CB. 2012;22:278-88.

177. Pappas CT, Bliss KT, Zieseniss A, Gregorio CC. The Nebulin family: an actin support group. Trends in cell biology. 2011;21:29-37.

178. Nakao K, Yasue H, Fujimoto K, Jougasaki M, Yamamoto H, Hitoshi Y, et al. Increased expression and regional differences of atrial myosin light chain 1 in human ventricles with old myocardial infarction. Analyses using two monoclonal antibodies. Circulation. 1992;86:1727-37.

179. Bhat HF, Baba RA, Bashir M, Saeed S, Kirmani D, Wani MM, et al. Alpha-1-syntrophin protein is differentially expressed in human cancers. Biomarkers : biochemical indicators of exposure, response, and susceptibility to chemicals. 2011;16:31-6.

180. Bykhovskaia M. Synapsin regulation of vesicle organization and functional pools. Seminars in cell & developmental biology. 2011;22:387-92.

181. Shen W, Punyanitya M, Wang Z, Gallagher D, St-Onge MP, Albu J, et al. Total body skeletal muscle and adipose tissue volumes: estimation from a single abdominal cross-sectional image. J Appl Physiol. 2004;97:2333-8.

182. Janssen I, Heymsfield SB, Wang ZM, Ross R. Skeletal muscle mass and distribution in 468 men and women aged 18-88 yr. J Appl Physiol. 2000;89:81-8.

183. Goodpaster BH, Kelley DE, Wing RR, Meier A, Thaete FL. Effects of weight loss on regional fat distribution and insulin sensitivity in obesity. Diabetes. 1999;48:839-47.

184. Goodpaster BH, Thaete FL, Kelley DE. Thigh adipose tissue distribution is associated with insulin resistance in obesity and in type 2 diabetes mellitus. The American journal of clinical nutrition. 2000;71:885-92.

185. Goodpaster BH, Thaete FL, Kelley DE. Composition of skeletal muscle evaluated with computed tomography. Annals of the New York Academy of Sciences. 2000;904:18-24.

186. Martin L, Birdsell, L., MacDonald, N., Reiman, T., Clandinin, M. T., McCargar, L., Murphy, R., Gosh, S., Sawyer, M. B., Baracos, V. E. Cancer cachexa in the age of obesity: skeletal muscle wasting is a powerful prognostic factor, independent of body mass index. Journal of Clinical Oncology. 2012.

187. Visser M, Kritchevsky SB, Goodpaster BH, Newman AB, Nevitt M, Stamm E, et al. Leg muscle mass and composition in relation to lower extremity performance in men and women aged 70 to 79: the health, aging and body composition study. Journal of the American Geriatrics Society. 2002;50:897-904.

188. Goodpaster BH, Carlson CL, Visser M, Kelley DE, Scherzinger A, Harris TB, et al. Attenuation of skeletal muscle and strength in the elderly: The Health ABC Study. J Appl Physiol. 2001;90:2157-65.

189. Rasch A, Bystrom AH, Dalen N, Martinez-Carranza N, Berg HE. Persisting muscle atrophy two years after replacement of the hip. The Journal of bone and joint surgery British volume. 2009;91:583-8.

190. Lang T, Cauley JA, Tylavsky F, Bauer D, Cummings S, Harris TB. Computed tomographic measurements of thigh muscle cross-sectional area and attenuation coefficient predict hip fracture: the health, aging, and body composition study. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. 2010;25:513-9.

191. Larson-Meyer DE, Smith SR, Heilbronn LK, Kelley DE, Ravussin E, Newcomer BR. Muscle-associated triglyceride measured by computed tomography and magnetic resonance spectroscopy. Obesity (Silver Spring). 2006;14:73-87.

192. Goodpaster BH, He J, Watkins S, Kelley DE. Skeletal muscle lipid content and insulin resistance: evidence for a paradox in endurance-trained athletes. The Journal of clinical endocrinology and metabolism. 2001;86:5755-61.

193. Perseghin G, Scifo P, De Cobelli F, Pagliato E, Battezzati A, Arcelloni C, et al. Intramyocellular triglyceride content is a determinant of in vivo insulin resistance in humans: a 1H-13C nuclear magnetic resonance spectroscopy assessment in offspring of type 2 diabetic parents. Diabetes. 1999;48:1600-6.

194. Miljkovic-Gacic I, Gordon CL, Goodpaster BH, Bunker CH, Patrick AL, Kuller LH, et al. Adipose tissue infiltration in skeletal muscle: age patterns and association with diabetes among men of African ancestry. The American journal of clinical nutrition. 2008;87:1590-5.

195. Murphy RA, Mourtzakis M, Chu QS, Baracos VE, Reiman T, Mazurak VC. Nutritional intervention with fish oil provides a benefit over standard of care for weight and skeletal muscle mass in patients with nonsmall cell lung cancer receiving chemotherapy. Cancer. 2011;117:1775-82.

196. Tsoli M, Robertson G. Cancer cachexia: malignant inflammation, tumorkines, and metabolic mayhem. Trends in endocrinology and metabolism: TEM. 2012.

197. Duncan JG. Peroxisome proliferator activated receptor-alpha (PPARalpha) and PPAR gamma coactivator-1alpha (PGC-1alpha) regulation of cardiac metabolism in diabetes. Pediatric cardiology. 2011;32:323-8.

198. Honors MA, Kinzig KP. The role of insulin resistance in the development of muscle wasting during cancer cachexia. Journal of cachexia, sarcopenia and muscle. 2012;3:5-11.

199. Crane JD, Devries MC, Safdar A, Hamadeh MJ, Tarnopolsky MA. The effect of aging on human skeletal muscle mitochondrial and intramyocellular lipid ultrastructure. The journals of gerontology Series A, Biological sciences and medical sciences. 2010;65:119-28.

200. Conley KE, Jubrias SA, Esselman PC. Oxidative capacity and ageing in human muscle. The Journal of physiology. 2000;526 Pt 1:203-10.

201. Wei Y, Rector RS, Thyfault JP, Ibdah JA. Nonalcoholic fatty liver disease and mitochondrial dysfunction. World journal of gastroenterology : WJG. 2008;14:193-9.

202. Larsen S, Nielsen J, Hansen CN, Nielsen LB, Wibrand F, Stride N, et al. Biomarkers of mitochondrial content in skeletal muscle of healthy young human subjects. The Journal of physiology. 2012;590:3349-60.

203. White JP, Baltgalvis KA, Puppa MJ, Sato S, Baynes JW, Carson JA. Muscle oxidative capacity during IL-6-dependent cancer cachexia. American journal of physiology Regulatory, integrative and comparative physiology. 2011;300:R201-11.

204. Karalaki M, Fili S, Philippou A, Koutsilieris M. Muscle regeneration: cellular and molecular events. In Vivo. 2009;23:779-96.

205. Mann CJ, Perdiguero E, Kharraz Y, Aguilar S, Pessina P, Serrano AL, et al. Aberrant repair and fibrosis development in skeletal muscle. Skeletal muscle. 2011;1:21.

206. Greenberg SA. A gene expression approach to study perturbed pathways in myositis. Current opinion in rheumatology. 2007;19:536-41.

207. Bradshaw EM, Orihuela A, McArdel SL, Salajegheh M, Amato AA, Hafler DA, et al. A local antigen-driven humoral response is present in the inflammatory myopathies. J Immunol. 2007;178:547-56.

208. Tidball JG, Villalta SA. Regulatory interactions between muscle and the immune system during muscle regeneration. American journal of physiology Regulatory, integrative and comparative physiology. 2010;298:R1173-87.

209. Argiles JM, Lopez-Soriano FJ, Busquets S. Mechanisms and treatment of cancer cachexia. Nutrition, metabolism, and cardiovascular diseases : NMCD. 2012.

210. Taylor-Jones JM, McGehee RE, Rando TA, Lecka-Czernik B, Lipschitz DA, Peterson CA. Activation of an adipogenic program in adult myoblasts with age. Mechanisms of ageing and development. 2002;123:649-61.

211. Joe AW, Yi L, Natarajan A, Le Grand F, So L, Wang J, et al. Muscle injury activates resident fibro/adipogenic progenitors that facilitate myogenesis. Nature cell biology. 2010;12:153-63.

212. Uezumi A, Fukada S, Yamamoto N, Takeda S, Tsuchida K. Mesenchymal progenitors distinct from satellite cells contribute to ectopic fat cell formation in skeletal muscle. Nature cell biology. 2010;12:143-52.

213. Shaw JH, Humberstone DA, Douglas RG, Koea J. Leucine kinetics in patients with benign disease, non-weight-losing cancer, and cancer cachexia: studies at the whole-body and tissue level and the response to nutritional support. Surgery. 1991;109:37-50.

214. Ueyama H, Kumamoto T, Johno M, Mita S, Tsuda T. Localized muscle wasting as an initial symptom of skeletal muscle lymphoma. J Neurol Sci. 1998;154(1):113-5.

CHAPTER 5: Prediction of skeletal muscle and fat mass in patients with advanced cancer using a metabolomic approach

5.1 Introduction

Recent progress in high-throughput analytical technologies and in bioinformatics now permits simultaneous analysis of hundreds of compounds constituting the metabolome. Metabolomic analyses give complex fingerprints that appear to be characteristic of a given metabolic phenotype or diet. While many have suggested that metabolomics has the potential to change how nutrition research is conducted (1-3), much of this potential remains unrealized (4). A surprisingly small number of metabolomic studies have been conducted in human nutrition to date, and progress is hampered by a number of unsolved problems, most notably by the lack of well-established, standardized methods for collecting, measuring, analysing and reporting metabolomic data (1, 5).

One important prerequisite for effective use of metabolomic approaches is to understand how variability in endogenous (e.g. tissue metabolism) and exogenous (e.g. diet) metabolite sources affect metabolomic profiles. A conceptual framework for these contributions includes multiple elements (Figure 5-1). Food intake may be the largest contribution to diurnal variation in metabolites. Diet is also a source of elements characteristic of specific foods: phytochemicals (ex. coffee (6), tea (7), cocoa (8), almonds (9)) or amines (e.g.

A version of this chapter has been published in *Journal of Nutrition*.
Stretch C, Eastman T, Mandal R, Eisner R, Wishart DS, Mourtzakis M, Prado CM, Damaraju S, Ball RO, Greiner R, Baracos VE. Prediction of skeletal muscle and fat mass in patients with advanced cancer using a metabolomic approach. J Nutr. 2012 Jan;142(1):14-21.

fish (10)). Metabolomic assessment in the postabsorptive state would generally limit the immediate influence of meals on substrate flux (11).

Individuals also may have widely divergent body proportions of organs (12), fat and skeletal muscle (13). As illustrated in Figure 5-1, relatively high body fat mass would result in a disproportionately low contribution of muscle-derived metabolites to the metabolome overall. Beside the well-known relation of urinary creatinine excretion to skeletal muscle mass (14), the amount of body lean and fat mass remains an unexplored source of variation in human metabolite profiling studies.

Based on the above, the following hypothesis was explored: in the postabsorptive state, the metabolome is defined, in part; by the varying proportions of tissues (e.g. adipose tissue, skeletal muscle) as these produce tissue-specific metabolites in the course of their turnover / metabolism. For example, adipose tissue is the origin of fatty acids and hence of ketones; and creatinine and 3-methylhistidine originate in skeletal muscle. Such tissue-associated metabolites would therefore provide good variables for predicting varying proportions of tissues. This will be tested in a population of patients with advanced (stage IIIB and IV) cancer.

Detection of nutritional and metabolic alterations that accompany the progression of cancer is a crucial part of patient care. Patients with advanced lung and colorectal cancers are known to have wide variations in lean and fat mass, dietary intake, metabolic rate and fuel metabolism due to the disease (12, 13, 15). Metabolomics is a particularly attractive technology to detect such variations as

this is a vulnerable patient population since it is fast, relatively inexpensive and most importantly non-invasive. To begin to understand the potential utility of metabolomic profiling in this patient population blood and urine metabolites were quantified using proton NMR and MS. Metabolite patterns were assessed in the post-absorptive state, using multivariate statistics and machine learning approaches to detect metabolite signatures of these features.

5.2 Methods

5.2.1 Study Design

Approval was provided by the Research Ethics Board of the Alberta Cancer Board. Eligible participants were recruited between 1/2005 and 10/2006 and included men and women with advanced stage (IV) non-small cell lung or colorectal cancer, >18 years of age, able to communicate freely in English and able to provide informed consent. At this advanced disease stage, both lung and colorectal cancer patients show similar characteristics (12, 13). Study participants were receiving therapy appropriate to their disease and stage. Patients with creatinine clearance below 60 ml/min (n=11), radiation to the kidneys (n=2), or bladder metastasis resulting in blood in the urine (n=3) were excluded as these independently affect urinary excretion, making the total number of included patients n= 55. While there is not an explicit formula for sample size calculation for metabolomic studies, previous work from our group (16) and others (9, 17-19) have been able to discriminate dietary or metabolic types with samples of this size.

5.2.2 Assessments

Patients collected diet records under the supervision of a dietitian for 3 days including one weekend day (20). To reduce the confounding acute effect of meals, participants were studied after a 12 h fast spanning the night of the 3rd day to the following morning. While this fasting could be expected to reduce the influence of prior intake, as protein intake during the day is related to the rate of amino acid oxidation during the night (21), a potential influence of prior protein intake may be expected.

Participants attended the Human Nutrition Research Unit for sampling and metabolic evaluation. While full details regarding assessments on these patients have been published (22), a brief summary is provided below.

Height and weight were measured with participants barefoot and in a hospital gown. DXA employed a LUNAR Prodigy High Speed Digital Fan Beam X-Ray-Based Densitometer, (General Electric) with enCORE 9.20 software for analysis of total fat mass (TFM) and lean soft tissue (LST). Appendicular lean soft tissue (ALST) was calculated by summing the LST from the limbs (arms and legs) (23) and is a measure of appendicular skeletal muscle (24). Percent muscle and fat mass is often used to describe body composition phenotypes. To test body composition phenotype in addition to absolute lean and fat mass with relation to metabolites, this was also calculated (% lean or fat mass = LST or TFM (kg) / body weight (kg)).

Nutrient intakes were estimated using the Canadian Nutrient File Database (FOOD PROCESSOR II nutrient analysis software, version 9.0; Esha Research,

Salem, OR). Total energy intake and macronutrient intakes (total protein, fat, carbohydrate and sugar) were calculated. Since the body weight (range 48-142 kg) and composition (range % body fat 13.8-56.2%) was variable, both the absolute or per kg body weight expression of energy intake and expenditure would be difficult to interpret. These data were thus expressed per kg lean soft tissue as assessed by DXA (25).

Resting energy expenditure (REE) and respiratory quotient (RQ) were determined by indirect calorimetry (VMax 29N, SensorMedics, Yorba Linda, CA) as detailed in Prado *et al.* (22).

Urine and plasma samples were collected immediately upon arrival to the research unit. Sodium azide was added to urine samples to a final concentration of ~0.02% to prevent bacterial growth. Whole blood was collected and plasma was isolated by a clinical laboratory provider (Dynacare Kasper Medical Laboratories, certified by: College of American Pathologists and College of Physicians& Surgeons of Alberta). Urine and plasma samples were stored at -80 °C until ready for analysis.

5.2.3 NMR spectroscopy

Urine samples were prepared by and analyzed according to a recently published procedure (16). Blood was prepared by removing high molecular weight compounds by ultrafiltration using Nanosep 3kDa microcentrifuge filter tubes. Prior to filtration, microcentrifuge filter tubes were washed using distilled deionized water (ddH_2O) to remove glycerol used as a preservative in the filters. Ultrafiltrate volumes ranged from 250 µL to 400 µL. Ultrafiltrates were then

brought to volume (585 µL) using ddH_2O. As with the urine samples, plasma was combined with 65 µL of internal standard (Chenomx Inc, Edmonton, Alberta) (consisting of ~5 mM sodium 2,2-dimethyl-2-silapentane-5-sulfonate d6 (DSS-d6), 0.2% sodium azide in 99% D_2O) and pH corrected to pH 6.75 ± 0.05 using small amounts of NaOH or HCl. A 600 µL aliquot of prepared sample was placed in a 5 mm NMR tube for NMR spectral acquisition.

One-dimensional NMR spectra were acquired using the standard NOESY pulse sequence on a four-channel Varian Inova-600 MHz NMR spectrometer with a triax-gradient 5-mm HCN probe. Quantification of metabolites by targeted profiling was performed using Chenomx NMRSuite 4.6 (Chenomx Inc. Edmonton, Canada).

Two analysts independently used Chenomx software to identify metabolite concentrations; only consensus assignments agreed upon by both analysts were used in statistical analysis. Laboratory analyses were also conducted to verify creatinine concentrations and amino acid peak assignments. A more complete description of these additional laboratory analyses has been published (16).

5.2.4 Mass spectrometry

Direct flow injection MS using AbsoluteIDQ™ Kit (BIOCRATES Life Sciences AG (Austria)) was used for the analysis of plasma and urine samples. This kit assay in combination with a 4000 QTrap (Applied Biosystems/MDS Sciex) mass spectrometer permits the identification and quantification of up to 160 metabolites in urine and plasma. Samples were prepared according to manufacturer's instructions. A standard flow injection protocol consisting of two

20 μL injections (one for the positive and one for the negative ion detection mode) was applied for all measurements. Multiple reaction monitoring detection was used for quantification. MetIQ software (BIOCRATES Life Sciences AG (Austria)) was used to control assay workflow, including sample registration, and calculation of metabolite concentrations.

5.2.5 Metabolite data preprocessing

Concentration values for metabolites can range over several orders of magnitude, both within and between patients. This was addressed by using the natural log of concentration values. The issue of different urine dilutions was also addressed. Various methods have been proposed to address this, including normalization by creatinine concentration (26), by total peak area (27), and by probability quotient (28). We earlier considered these options and found that they reduced predictive accuracy of various algorithms compared to no data normalization (16). Therefore, our main analysis used only log transformation as the only pre-processing step. In addition, normalization of urine metabolite concentrations to whole – body lean soft tissue mass was also conducted attempted.

5.2.6 Statistical methods

The objective of the statistical analysis is to relate the patient's plasma/urine metabolites with the patient's dietary/physiological assessments (i.e. class). As there is no single accepted method for statistical treatment of quantitative metabolomic data, several methods previously used were compared. Data were analyzed by partial least squares discriminant analysis (PLS-DA) (29),

support vector machines (SVM)) (30) and least absolute shrinkage and selection operator (LASSO) (31). Each of these algorithms uses a labelled data set (i.e. data describing a set of patients, along with the *class label* for each patient – e.g. patient has high energy intake versus low energy intake) to produce a *classifier* that can predict the class label of a new patient. Here classes were defined as the distal ranges of values (highest and lowest) for each assessment where an instance (patient) was labelled *high* if his/her measurement on this assessment was at least 0.5 SD above the median and labelled *low* if the assessment was at least 0.5 SD below the median. For example, only individuals with high and low energy intake were included, leaving out an intermediate group representing a band approximately the width of the measurement error of this variable.

The effectiveness of each learning algorithm was assessed, i.e. how accurately the classifier classifies novel patients, by leave-one-out cross-validation (LOOCV) and permutation testing. The baseline accuracy rate was compared with the LOOCV accuracy results obtained by PLS-DA, SVM and LASSO. The baseline accuracy rate is the frequency of the most common class, expressed as a percentage (i.e. if 60 patients are in class A and 40 patients in class B, then the baseline accuracy would be 60%). Note that metabolomic information is not used in calculating the baseline accuracy. Thus, if the metabolic profile data contains any signal with respect to a particular classification task then it would increase the classification accuracy above the baseline accuracy rate; the maximum accuracy being 100%. Predictors were subjected to permutation testing (1000 permutations), to determine whether the predictive cross-validation

accuracies of these classifiers are statistically significant. An acceptable model was that very few (< 50 out of 1000; i.e. $p<0.05$) permuted models outperform the original model.

PLS-DA is commonly used to build predictors using eigenvalues (29, 32). LOOCV analysis by PLS-DA was conducted using SIMCA P11.0 (Umetrics, Umeå, Sweden). SVM (30) views each instance as a vector in multi-dimensional space, and seeks the maximally separating hyperplane between the classes in this space. SVM analysis (with a linear kernel) using LOOCV was conducted using the WEKA machine learning package (33). LASSO is a linear classifier based on a form of regularized regression, which incorporates a penalty into the least squares objective function when learning a set of regression coefficients. LASSO implicitly performs variable selection because it sets some of the regression coefficients to zero; hence the associated variables (here, metabolite concentrations) will not contribute to the model. This technique was implemented using R and used the glmnet package to perform LASSO regression using LOOCV (34). Though the above three algorithms differ, they all work by finding a hyperplane that separates two classes in multidimensional space. At the most basic level, future patients would be predicted to belong to one class or another based on where their metabolite concentrations place them in multidimensional space relative to that hyperplane.

Many researchers interested in nutrition and metabolism may ask: *which metabolites best discriminate the classes?* This was addressed by using mutual information to quantify the dependence between each metabolite and the class

outcome (the two groups for each assessment – e.g., high muscle mass versus low muscle mass); see (16). Mutual information analysis yields unit-less values, where larger values indicate a higher degree of dependence.

5.3 Results

5.3.1 Distribution of measured dietary/physiological assessments in advanced cancer patients

Of the 55 cancer patients included in this study, 58% were male, 45% had lung cancer and the overall median age was 61 ± 11 yrs. The median and variation for each measured assessment as well as characteristics of classes are shown (Table 5-1).

5.3.2 Urine - Metabolites identified and quantified

Using NMR we quantified 71 metabolites. However 8 metabolites were excluded because they were: drug metabolites or constituents of vehicle for drug administration (ibuprofen, acetaminophen, salicylurate, propionate, propylene glycol and mannitol), belonged to microbial metabolism or aspartame consumption (methanol) (40, 41) or had unreliable quantification (urea). Urea has unreliable quantification since suppression of the NMR signal by pre-saturation may lead to resonant suppression of the urea peak due to proton exchange with water, thereby making its quantification unreliable (35). A list of the 63 remaining metabolites can be found in Table 5-2. NMR-measured concentrations of creatinine were confirmed using laboratory tests (intraclass correlation of 0.949

with a 95% confidence interval of 0.907 to 0.971). Spike-in experiments provided positive confirmation of peak assignments for amino acids (16) (data not shown).

Mass spectrometry identified 117 metabolites including 12 amino acids, 37 acyl carnitines, 55 glycerophospholipids, 12 sphingolipids and the combined concentration of hexose sugars.

5.3.3 Urine - Results of statistical analyses

The urine data analysis summary (Table 5-3) presents the accuracies for the best predictive models using all three methods (i.e. SVM, LASSO and PLS-DA). The most accurate predictors were for appendicular skeletal muscle mass: SVM (LOOCV accuracy = 98%), LASSO (LOOCV accuracy = 90%) and PLS-DA (LOOCV accuracy = 85%), compared with a baseline accuracy rate of 54%. Lean soft tissue (which includes skeletal muscle, soft lean tissues, organs and skin) was also accurately predicted with all three algorithms (Table 5-3). Similar accuracies of these models may be explained by the high correlation (Pearson correlation = 0.98) between total lean and appendicular lean tissues. Satisfactory predictive models were achieved for total fat mass using SVM (LOOCV accuracy = 79%), LASSO (LOOCV accuracy = 82%) and PLS-DA (LOOCV accuracy = 79%), compared with the baseline accuracy rate of 50%. Median concentration and standard deviation of urinary metabolites quantified by NMR for the two classes (high and low) of lean soft tissue and fat mass are listed in Tables 5-4 and 5-5, respectively. High and low percent lean and fat mass did not produce predictive models that were any better than could be obtained by chance.

The distribution of cancer types (lung, colorectal) was not different between the high and low fat mass groups or between the high and low lean tissue mass groups. However, the low class for lean tissues was composed mainly of women (95%) and the high class was composed of 100% men. By itself, sex was predicted by urinary metabolite profiles, with a base model, SVM, LASSO and PLS-DA providing respectively, 58%, 91%, 85% and 78% LOOCV accuracy. However, it seems likely that sex was discriminated by nothing other than the fact that men are generally larger and more muscular than women, leading to differential production of muscle-specific metabolites. This is consistent with the observation that creatine, a muscle-specific metabolite, was in the top two metabolites that contributed to the discrimination of sex in the three statistical analyses, as well as the mutual information (see below).

Neither total energy, carbohydrate, sugar nor fat intake could be predicted accurately from the NMR urine metabolite data; this was true when the classes were determined based on absolute (total) intake, intake / kg body weight or intake / kg lean soft tissue. However, protein intake did result in a satisfactory but relatively weak model with a baseline accuracy of 53% and LOOCV accuracies using SVM (70%), LASSO (73%) and PLS-DA (73%). Protein intake also produced several randomly permuted models that appeared more accurate than the model learned using the original data (48 times for SVM and 28 times for LASSO). To test whether variation in lean tissue was confounding chances of building predictive models based on macronutrient intakes, urine metabolite concentrations were normalized to the whole body LST mass for each individual,

and reran all of the classifiers. This normalization made no difference in the accuracy of the predictive models for any macronutrient (e.g., the LOOCV for SVM for protein intake was 73% with this normalization compared with 70% (see above)).

The RQ classes included a group with substantially fat oxidation (RQ =0.7) and a class with a slightly higher RQ of 0.8 reflecting more mixed oxidation. RQ did not produce a predictive model any better than could be obtained by chance, as determined by permutation testing. REE classes (Table 5-1) were developed for each sex and then aggregated as this is a sex-dependent variable (36). The median energy intake per kg body weight or per kg LST for patients in the low and high REE classes was not different (47 ± 10 kcal / kg LST; 29 ± 8 kcal / kg body weight for both). REE also did not produce a predictive model any better than could be obtained by chance, regardless of the basis of normalization of this value (total, per kg body weight, per kg lean soft tissue).

Classifiers built using urine metabolites measured by MS alone or NMR and MS pooled together resulted in no improvements in LOOCV accuracy and, in fact, for most models the accuracy decreased.

5.3.4 Urine - metabolites related to lean and fat mass

Bivariate analysis allows for the ranking of metabolites according to their mutual information for ALST and TFM. As LST and ALST share the same top 30 metabolites, albeit in slightly different rank order, Table 5-6 shows only ALST. This mutual information analysis for ALST and LST further supports the suggestion that the discrimination of sex is nothing more than the discrimination

of lean mass, because 17 of the top 20 metabolites (including creatine) in the list of mutual information for sex were identical with the metabolites discriminating muscle and lean tissue, (not shown).

5.3.5 Plasma - Metabolites identified and quantified

Mass spectrometry identified 143 metabolites including 15 amino acids, 25 acyl carnitines, 87 glycerophospholipids, 15 sphingolipids and the combined concentration of all hexose sugars (Table 5-7). Quantitative NMR analysis identified the 31 metabolites listed in Table 5-8.

5.3.6 Plasma - Results of data analyses

Plasma NMR data analysis resulted in poor predictive models (i.e. not different from the baseline accuracy rate) for lean and fat mass, percent lean and fat mass, total energy and macronutrient intake and energy metabolism. Plasma MS data resulted in satisfactory prediction using SVM (71%), LASSO (88%) and PLS-DA (79%) of total body fat, compared to baseline of 50%. Median concentration and standard deviation of plasma metabolites quantified by MS for the two fat mass groups (high and low) are listed in Table 5-9. Predictive models built using plasma metabolites measured by NMR and MS pooled together resulted in no improvements in LOOCV accuracy and, in fact, for most models the accuracy decreased.

5.3.7 Plasma - Metabolites related to total body fat

Bivariate analysis was used to rank metabolites according to their mutual information for TFM; the top 30 are shown (Table 5-6). All of the metabolites

included in the top 30 are lipid molecules (acylcarnitines, phosphatidylcholines, lysophosphatidylcholines, and sphingomyelins).

5.4 Discussion

The present quantitative metabolomic study supports the hypothesis that body lean and fat mass have distinctive metabolic profiles. Broad categories (high and low) of muscle mass quantity were accurately predicted from metabolite concentrations in easily obtained physiological fluids. This level of discrimination lends itself to the identification of occult sarcopenia (i.e. absolute muscle mass >2 standard deviations less than for normal healthy adults), a clinically important condition (13). The above results also suggest the potential of metabolomic approaches to further studies of body components *per se*. These findings suggest that lean / muscle mass can be easily predicted using urinary metabolite profiles, which is advantageous over other means as it can be obtained non-invasively. People who are not candidates for other forms of imaging (too large, CT or MRI contraindications), or in remote locations, or are too frail to undergo assessment could be possible candidate populations for a metabolomics-based screening test.

Variation in lean and fat mass inherent in patients with cancer and possibly other chronic conditions may confound metabolomic studies intended to look at diet, metabolic disorders or diseases. This variation could be eliminated by assessing patients only within predefined lean and fat mass ranges by ensuring that lean and fat mass was used as a basis for matching participants in different

treatment groups, or by explicitly including these factors when building predictors for various conditions.

5.4.1 Sources of variation in metabolite profiles

To eliminate *acute effects of dietary intake* samples were collected under standardized conditions of overnight fasting. A diet record completed over the 3 days preceding sampling and lean and fat mass measurements provided the estimates of energy and macronutrient intake. The fasting period would appear to have largely eliminated effects of foods eaten in the 3 days preceding the measurements, as neither total energy intake, nor that of any energy macronutrient, was associated with a clear-cut metabolite profile.

Protein intake resulted in a relatively weak predictive model with cross validation accuracy of up to 73%. There are many possible reasons for the appearance of a metabolic profile, when we stratified for protein intake. All of the elements of macronutrient intake (sugar, fat, total carbohydrate and protein) were correlated with one another (Pearson r =0.62 to 0.82). The individuals within the low protein class here had protein-energy malnutrition, with median intakes of 0.8 g protein and 24.9 kcal/kg body weight /d and protein:energy ratio of 0.034 g protein /kcal. This was different ($p<0.001$) from the high protein class with protein intake of 1.6 g/kg body weight/d, energy intake of 37.0 kcal/kg body weight/d and protein:energy ratio of 0.047 g protein /kcal, and can be compared with recommended values of recommended intakes of 1.2-2.0 g protein and 30-35 kcal /kg body weight /d for this patient population (37). Thus in overnight

fasted individuals, the presence of quite severe malnutrition could be detected but with a relatively weak signal. It will be useful to further understand why the patients with the low energy intakes chose and consumed foods with a low protein:energy ratio, as it may not be in their advantage.

Analysis could not produce a satisfactory predictor for *RQ and REE*. The rationale for including RQ as a possible source of variation was that oxidation of these fuels may result in changes in plasma and urine metabolites. For example, increased plasma acylcarnitine concentrations have been correlated with fatty acid oxidation (42) and changes in urinary amino acid concentrations may be reflective of protein catabolism. It seems likely that the fasting protocol contributed to a low level of variation in RQ. REE was expressed in relation to whole body LST, to limit the variation contributed by variability in body weight and in % body fat. Here, the mean REE of the two classes (27.7 ± 1.9 versus 37.2 ± 3.2 kcal / kg LST / day) can be compared with values for healthy individuals across the entire lifespan. With reference to the meta-analysis reported by Weinsier *et al*. (36), 27.7 kcal / kg LST / day is a normal metabolic rate for healthy individuals from ages 50-70 years. By contrast, 37.2 / kg LST / day (30% higher) is hypermetabolic and REE in this range is not expected beyond adolescence. The lack of discrimination of REE in our models suggests that the higher overall metabolic throughput is not necessarily associated with alterations in patterns of metabolites.

Previously published population-based data demonstrated high variability in muscle and fat mass in cancer patients (38). Individuals with the same body weight and BMI may be up to 2.5-fold different in amount of lean and muscle

tissue. Previous work also indicated that even patients with metastatic cancers were likely to be overweight or obese (13, 38), with severe muscle wasting (sarcopenia) concurrently present in a significant proportion of the population. Like many elderly people, cancer patients may be affected by age, limited mobility, and their primary disease as well as comorbid conditions, each resulting in muscle loss.

There has been interest in testing for metabolic signature of cancer. The present results suggest that, to identify specific metabolic discriminates of cancer, as opposed to cancer-associated variations in lean and fat mass and food intake (which are not specific to this disease), studies should be conducted with the following controls. 1) Participants should be in the fasted state because it reduces the effect of diet, 2) participants could be stratified for protein intake or provided a standardized protein intake prior to the measurement to remove its effect (39) and 3) as suggested above, assessing patients only within predefined categories of lean and fat mass, or by ensuring that lean and fat mass is used as a basis for matching participants in cancer versus noncancer groups. Many diseases in addition to cancer are associated with wasting of the lean tissues (COPD, chronic heart failure, chronic renal failure, diabetes). In the absence of suitable controls, a metabolic signature apparently due to one of these diseases may merely reflect the presence of wasting (a non-specific effect) as opposed to a specific disease-related process.

5.4.2 Methodological considerations

Particular attention was given to characterizing the patient population using precise measures of lean and fat mass, supplemented with estimates of metabolic rate, RQ and macronutrient intake. Empirical results showed that SVM was the most accurate classifier of the three algorithms that we tested; as this is the second time that superior predictive accuracy is obtained with SVM (16) compared with multiple other methods, we suggest it may be a good method for future work.

No single analytical platform captures all metabolites in a biological sample. Proton NMR is less sensitive than MS but easily captures amino acids and their intermediates, TCA cycle intermediates and other metabolites involved in energy metabolism (e.g. glucose). Both skeletal muscle and organ tissues metabolize compounds producing end products that are ultimately excreted from the body via urine. Thus it was not surprising that metabolites mainly responsible for the discrimination between low and high lean and muscle tissues are metabolites known to originate mainly from muscle (creatine), amino acid metabolism (trigonelline, guanidoacetate) and intermediary metabolites (2-oxoglutarate, succinate, fumarate, pyruvate). The sensitivity of MS enabled the quantification of lipid molecules in plasma that was not possible with NMR and proved to be a superior platform to build a predictive model for total fat mass.

5.5 Conclusion

In conclusion, the above results can inform future studies using metabolomic approaches in human nutrition and metabolism. Our findings raise the possibility of a non-invasive test for lean and fat mass based on urine sampling. This may have applications in screening or in individuals who cannot undergo diagnostic imaging by DXA, CT or MRI. Secondly, our findings will assist in the design of future studies, to assist with minimization of sources of variation which may confound the interpretation of results.

Tables

Table 5-1: Variation in lean and fat mass, energy intake and energy metabolism for all cancer patients as well as after patients were divided into two classes (low and high)

	All Patients (n= 55)		**Low Class[1]**	**High Class**
	Median ± SD	Range	Median ± SD	Median ± SD
Lean and fat mass				
TFM, *kg*	28 ± 9.5	11.4 - 62.5	17 ± 3.5	37.5 ± 7.6
LST, *kg*	46.4 ± 10.6	29.3 - 65.8	35.7 ± 3.2	57.8 ± 4.6
ALST, *kg*	19.3 ± 4.8	12.4 - 28.5	14.7 ± 1.2	24.8 ± 2.3
Dietary Intake				
Total Energy Intake, *kcal/kg LST*	46.5 ± 12.8	27.4 - 85.1	32.5 ± 3.8	62.6 ± 10.7
CHO Intake, *g/kg LST*	6 ± 2	3 - 11.9	4.3 ± 0.6	7.8 ± 1.4
Sugar Intake, *g/kg LST*	1.9 ± 0.9	0.5 - 4.6	1.1 ± 0.2	2.9 ± 0.6
Fat Intake, *kcal/kg LST*	14.7 ± 6	1.7 - 35.3	9 ± 2.8	20.4 ± 4.5
Protein Intake, *g/kg LST*	1.9 ± 0.6	1 - 3.6	1.4 ± 0.2	2.5 ± 0.4
Energy Metabolism				
REE, *kcal /kg LST/day*	33 ± 5	25 - 46	29 ± 4	38 ± 4
RQ	0.8 ± 0.1	0.7 - 0.9	0.7 ± 0	0.8 ± 0

[1] Low and High classes were determined by using the cutoff of median ± 0.5 SD for all the patients in the study, as detailed in the methods section.

Table 5-2: Concentrations of 63 urine metabolites quantified by NMR for all patients included in data analysis[1]

Urine metabolite	**^{1}H chemical shift (ppm) and coupling[2]**	***µmol/L***
1,6-Anhydro-β-D-glucose	3.53(m), 3.67(m), 3.69(m), 3.75(dd), 4.09(dd), 4.62(dd), 5.45(m)	38.1 ± 9.79
1-Methylnicotinamide	4.47(s), 8.17(t), 8.88(d), 8.95(d), 9.26(s)	44.5 ± 5.26
2-Aminobutyrate	0.97(t), 1.89(m), 3.71(dd)	9.74 ± 4.36
2-Hydroxyisobutyrate	1.35(s)	35.3 ± 3.88
2-Oxoglutarate	2.44(t), 3.00(t)	84.8 ± 52.5
3-Aminoisobutyrate	1.19(d), 2.60(m), 3.03(dd), 3.1(dd)	41.1 ± 16.1
3-Hydroxybutyrate	1.19(d), 2.29(dd), 2.40(dd), 1.14(m)	11.8 ± 3.86
3-Hydroxyisovalerate	1.26(s),2.36(s)	7.85 ± 6.71
3-Indoxylsulfate	7.18(m), 7.26(m), 7.36(s), 7.49(d), 7.70(d), 10.10(s)	171 ± 21.4
4-Hydroxyphenylacetate	3.44(s), 6.85(m), 7.16(m)	59.3 ± 16.0
Acetate	1.91(s)	25.0 ± 31.4
Acetone	2.22(s)	7.73 ± 3.67
Adipate	1.54(m), 2.19(m)	8.45 ± 5.04
Alanine	1.47(d), 3.78(qt)	169 ± 31.9
Asparagine	2.86(dd), 2.95(dd), 3.99(dd), 6.91(s), 7.62(s)	42.7 ± 6.98
Betaine	3.26(s), 3.89(s)	64.1 ± 16.8
Carnitine	2.41(dd), 2.45(dd), 3.22(s), 3.40(m), 3.43(m), 4.56(m)	23.4 ± 7.09
Citrate	2.53(d), 2.69(d)	2140 ± 347
Creatine	3.02(s), 3.92(s)	42.0 ± 43.1
Creatinine	3.03(s), 4.05(s)	7980 ± 749.0
Dimethylamine	2.72(s)	298 ± 31.0
Ethanolamine	3.14(m), 3.82(m)	220 ± 32.9
Formate	8.45(s)	75.3 ± 10.6

Fucose	1.20(d), 1.24(d), 3.44(dd), 3.56(t), 3.63(dd), 3.74(dd), 3.76(dd), 3.79(m), 3.80(m), 3.85(dd), 3.86(m), 3.87(m), 3.97(dd), 4.01(m), 4.03(m), 4.07(dd), 4.18(m), 4.55(d), 5.20(d), 5.22(d), 5.27(d)	75.5 ± 11.6
Fumarate	6.51	4.13 ± 2.19
Glucose	3.24(dd), 3.40(m), 3.41(m), 3.47(m), 3.49(m), 3.53(dd), 3.71(m), 3.72(dd), 3.74(m), 3.82(m), 3.84(m), 3.90(dd), 4.64(d), 5.23(d)	189 ± 31.8
Glutamine	2.12(m), 2.15(m), 2.43(m), 2.47(m), 3.76(t), 6.87(s), 7.58(s)	232 ± 38.5
Glycine	3.56(s)	666 ± 327
Glycolate	3.95(s)	154 ± 28.3
Guanidoacetate	3.79(s)	83.3 ± 9.62
Hippurate	3.96(d), 7.54(m), 7.55(m), 7.63(t), 7.82(m), 7.83(m), 8.52(s)	975 ± 379
Histidine	3.15(dd), 3.25(dd), 3.99(qt), 7.11(s), 7.92(s)	178 ± 33.5
Hypoxanthine	8.18(s), 8.20(s)	48.8 ± 7.48
Isoleucine	0.93(t), 1.00(d), 1.25(m), 1.46(m), 1.97(m), 3.66(d)	7.83 ± 2.04
Lactate	1.32(d), 4.11(d)	68.1 ± 15.0
Leucine	0.95(d), 0.96(d), 1.67(m), 1.70(m), 1.73(m), 3.73(m)	21.6 ± 2.66
Lysine	1.43(m), 1.50(m), 1.72(m), 1.88(m), 1.91(m), 3.02(t), 3.75(t)	88.1 ± 18.9
Methylamine	2.6(s)	14.5 ± 1.90
Methylguanidine	2.83(s)	14.5 ± 2.81
N,N-Dimethylglycine	2.92(s), 3.72(s)	21.0 ± 3.97
O-Acetylcarnitine	2.14(s), 2.50(dd), 2.63(dd), 3.19(s), 3.60(dd), 3.84(dd), 5.59(m)	9.84 ± 4.39
Pantothenate	0.89(s), 0.92(s), 2.41(t), 3.39(d), 3.43(qt), 3.44(qt), 3.51(d), 3.98(s), 8.00(dd)	23.8 ± 4.09
Pyroglutamate	2.03(m), 2.38(m), 2.41(m), 2.50(m), 4.17(dd)	147 ± 21.2

Pyruvate	2.37(s)	19.9 ± 4.22
Quinolinate	7.46(dd), 8.00(d), 8.45(d)	43.4 ± 5.65
Serine	3.84(dd), 3.94(dd), 3.98(dd)	124 ± 28.6
Succinate	2.4(s)	25.3 ± 5.86
Sucrose	3.47(t), 3.55(dd), 3.66(s), 3.68(m), 3.76(m), 3.80(m), 3.82(m), 3.83(m), 3.84(m), 3.88(m), 4.04(t), 4.21(d), 5.40(d)	23.8 ± 35.6
Tartrate	4.33(s)	12.9 ± 15.6
Taurine	3.25(t), 3.41(t)	253 ± 56.6
Threonine	1.32(d), 3.59(d), 4.26(m)	82.4 ± 11.0
Trigonelline	4.43(s), 8.07(dd), 8.82(m), 8.83(m), 9.11(s)	85.1 ± 23.7
Trimethylamine N-oxide	3.26(s)	403 ± 100
Tryptophan	3.30(dd), 3.47(dd), 4.05(q), 7.19(m), 7.27(m), 7.31(s), 7.52(d), 7.72(d)	52.5 ± 7.14
Tyrosine	3.05(dd), 3.19(dd), 3.94(q), 6.88(m), 7.20(m)	50.5 ± 8.09
Uracil	5.79(d), 7.52(d)	35.3 ± 4.00
Valine	0.98(d), 1.03(d), 2.26(m), 3.61(d)	31.5 ± 3.34
Xylose	3.22(dd), 3.31(dd), 3.43(t), 3.52(dd), 3.60(m), 3.62(m), 3.65(m), 3.67(m), 3.69(m), 3.92(dd), 4.57(d), 5.19(d)	48.9 ± 8.79
cis-Aconitate	3.12(d), 5.75(m)	125 ± 37.9
myo-Inositol	3.27(t), 3.53(dd), 3.62(dd), 4.05(m)	76.6 ± 22.9
trans-Aconitate	3.44(s), 3.59(s)	21.1 ± 2.19
1-Methylhistidine	3.21(dd), 3.29(dd), 3.74(s), 3.95(dd), 7.13(s), 8.10(s)	245 ± 72.5
3-Methylhistidine	3.08(dd), 3.16(dd), 3.70(s), 3.96(dd), 7.03(s), 7.70(s)	67.3 ± 10.6
[1]Values are median ± SE, *n*=55. [2]s=singlet, d=doublet, dd=doublet-doublet, t=triplet, q=quartet, m=multiplet		

Table 5-3: Best predictive models based on SVM, LASSO and PLS-DA analysis and permutation testing

	Analysis from urine metabolites[1]			**Analysis from plasma metabolites[2]**
	LST	ALST	TFM	TFM
Baseline accuracy, %	54	51	50	50
SVM				
LOOCV accuracy, %	90	98	79	71
Permuted models better than original data, n (p-value)	0 (< 0.00001)	0 (< 0.00001)	15(0.015)	62 (0.062)
LASSO				
LOOCV accuracy, %	87	90	82	88
Permuted models better than original data, n (p-value)	1 (0.001)	1 (0.001)	10 (0.01)	1 (0.001)
PLS-DA				
LOOCV accuracy, %	85	85	79	79
Permuted models better than original data, n (p-value)	0 (< 0.00001)	1 (0.001)	17 (0.017)	27 (0.027)

[1] Urine metabolites include only metabolites analyzed by nuclear magnetic resonance (n=63 metabolites).
[2] Plasma metabolites included only metabolites analyzed by mass spectrometry (n=143 metabolites).

Table 5-4: Concentration of 63 urinary metabolites quantified by NMR from cancer patients included in the high and low lean soft tissue classes[1, 2]

Urine metabolite	Patients with low lean soft tissue	Patients with high lean soft tissue
	μmol/L	
1,6-Anhydro-β-D-glucose	34.5 ± 15.4	36.3 ± 18.1
1-Methylnicotinamide	43.5 ± 12.4	50.7 ± 6.51
2-Aminobutyrate	6.56 ± 2.76	15.5 ± 2.07
2-Hydroxyisobutyrate	27.7 ± 4.79	39.8 ± 10.0
2-Oxoglutarate	144 ± 124	84.8 ± 30.5
3-Aminoisobutyrate	47.2 ± 18.0	35.4 ± 27.6
3-Hydroxybutyrate	11.6 ± 5.00	16.4 ± 6.01
3-Hydroxyisovalerate	5.98 ± 3.97	7.07 ± 20.6
3-Indoxylsulfate	203 ± 36.1	119 ± 37.7
4-Hydroxyphenylacetate	54.5 ± 37.0	67.5 ± 10.5
Acetate	29.5 ± 6.80	21.2 ± 5.13
Acetone	5.4 ± 0.58	8.2 ± 4.2*
Adipate	8.42 ± 12.0	8.98 ± 1.95
Alanine	118 ± 48.1	227 ± 78.1
Asparagine	37.4 ± 9.57	57.9 ± 12.6
Betaine	42.7 ± 13.7	127 ± 43.3*
Carnitine	16.9 ± 11.7	28.5 ± 14.8
Citrate	1720 ± 427.0	2410 ± 705.0
Creatine	86.5 ± 108	36.6 ± 8.82*
Creatinine	7430 ± 1300	8300 ± 1340
Dimethylamine	333 ± 56.9	268 ± 46.4
Ethanolamine	184 ± 41.8	234 ± 77.0
Formate	48.1 ± 9.16	128 ± 16.2**
Fucose	82.3 ± 18.5	69.5 ± 15.4
Fumarate	4.70 ± 3.84	3.89 ± 1.08
Glucose	171 ± 66.0	201 ± 40.0
Glutamine	260 ± 53.3	245 ± 59.8
Glycine	688 ± 176	891 ± 994
Glycolate	133 ± 34.3	242 ± 60.5
Guanidoacetate	83.5 ± 16.8	63.4 ± 16.8
Hippurate	903 ± 942	1030 ± 410.0
Histidine	146 ± 36.8	323 ± 74.0*
Hypoxanthine	40.7 ± 10.9	42.8 ± 13.1
Isoleucine	8.72 ± 1.16	7.72 ± 6.20
Lactate	110 ± 19.3	71.4 ± 27.9
Leucine	16.5 ± 4.48	22.6 ± 4.31
Lysine	89.5 ± 24.7	118 ± 47.6
Methylamine	13.4 ± 3.05	10.6 ± 3.76
Methylguanidine	17.4 ± 3.50	13.2 ± 3.62

N,N-Dimethylglycine	16.0 ± 4.72	26.5 ± 10.5
O-Acetylcarnitine	6.96 ± 1.84	15.4 ± 3.10*
Pantothenate	20.9 ± 7.29	24.3 ± 8.99
Pyroglutamate	162 ± 49.7	137 ± 20.5
Pyruvate	20.4 ± 8.82	14.7 ± 7.54
Quinolinate	38.7 ± 10.8	41.8 ± 6.55
Serine	118 ± 33.9	128 ± 49.2
Succinate	26.3 ± 13.7	26.6 ± 4.95
Sucrose	22.5 ± 97.8	25.3 ± 13.1
Tartrate	11.2 ± 8.82	15.3 ± 14.6
Taurine	222 ± 104	336 ± 87.2
Threonine	84.4 ± 15.8	95.4 ± 23.4
Trigonelline	89.5 ± 32.1	89.8 ± 35.0
Trimethylamine N-oxide	459 ± 246	403 ± 87.2
Tryptophan	39.9 ± 9.52	66.6 ± 14.7
Tyrosine	39.7 ± 8.01	64.5 ± 14.5*
Uracil	54.6 ± 8.63	39.2 ± 6.13
Valine	27.2 ± 4.62	32.2 ± 7.64
Xylose	58.8 ± 20.3	48.0 ± 12.6
cis-Aconitate	157 ± 47.4	108 ± 32.5
myo-Inositol	35.2 ± 19.5	87.6 ± 53.2
trans-Aconitate	21.0 ± 2.72	19.6 ± 3.36
π-Methylhistidine	232 ± 98.3	197 ± 164
τ-Methylhistidine	48.3 ± 15.6	73.0 ± 19.8

[1]Values are median ± SE, Low class n=21, High class n=18.
[2]Low and High class included patients who had values 0.5 SD below and above the median for lean soft tissue, respectively.
Asterisks indicate different from low lean soft tissue: *P<0.05, **P<0.01, ***$P\leq$0.001 based on Mann-Whitney nonparametric statistical analysis

Table 5-5: Concentration of 63 urinary metabolites quantified by NMR from cancer patients included in the high and low total fat mass classes[1, 2]

Urine metabolite	Patients with low fat mass	Patients with high fat mass
	µmol/L	
1,6-Anhydro-β-D-glucose	45.6 ± 20.1	32.6 ± 17.2
1-Methylnicotinamide	44.5 ± 5.74	52.4 ± 9.50
2-Aminobutyrate	10.3 ± 4.47	7.04 ± 7.47
2-Hydroxyisobutyrate	32.5 ± 4.78	43.7 ± 6.95
2-Oxoglutarate	108 ± 22.4	84.8 ± 186
3-Aminoisobutyrate	70.6 ± 25.6	21.5 ± 28.8
3-Hydroxybutyrate	13.7 ± 6.14	10.4 ± 4.87
3-Hydroxyisovalerate	14.4 ± 4.79	7.07 ± 4.35
3-Indoxylsulfate	276 ± 37.4	107 ± 32.4**
4-Hydroxyphenylacetate	62.4 ± 19.0	69.9 ± 25.6
Acetate	35.4 ± 7.57	23.9 ± 9.58
Acetone	7.8 ± 0.80	8.23 ± 1.75
Adipate	8.82 ± 1.39	11.2 ± 4.18
Alanine	176 ± 50.5	236 ± 53.3
Asparagine	32.4 ± 12.8	52.0 ± 12.4
Betaine	68.7 ± 23.4	123 ± 27.4
Carnitine	27.9 ± 6.18	46.3 ± 18.2*
Citrate	1350 ± 424.0	2240 ± 625.0
Creatine	47.8 ± 57.1	40.0 ± 26.1
Creatinine	8940 ± 1550	7780 ± 1670
Dimethylamine	333 ± 74.8	243 ± 57.8
Ethanolamine	289 ± 53.7	193 ± 81.5
Formate	77.6 ± 17.8	93.2 ± 21.1
Fucose	90.3 ± 28.7	66.7 ± 22.5
Fumarate	4.6 ± 0.63	3.91 ± 5.61
Glucose	178 ± 94.9	195 ± 39.0
Glutamine	232 ± 62.9	310 ± 61.9
Glycine	550 ± 177	541 ± 167
Glycolate	211 ± 66.2	189 ± 62.0
Guanidoacetate	83.4 ± 25.8	70.0 ± 21.0
Hippurate	1910 ± 544.0	608 ± 512
Histidine	151 ± 52.2	138 ± 69.8
Hypoxanthine	64.8 ± 14.6	57.7 ± 17.3
Isoleucine	8.43 ± 1.35	8.42 ± 1.48
Lactate	76.1 ± 16.5	74.6 ± 29.8
Leucine	18.9 ± 6.13	20.8 ± 3.29
Lysine	56.0 ± 33.0	90.3 ± 31.5
Methylamine	21.6 ± 3.14	9.83 ± 1.35*
Methylguanidine	11.6 ± 3.13	9.86 ± 4.15

N,N-Dimethylglycine	16.6 ± 5.12	21.6 ± 6.72
O-Acetylcarnitine	10.5 ± 1.08	17.6 ± 3.94*
Pantothenate	20.7 ± 7.65	21.1 ± 6.34
Pyroglutamate	178 ± 67.9	137 ± 30.6
Pyruvate	22.2 ± 5.06	14.7 ± 12.9
Quinolinate	44.1 ± 13.1	39.0 ± 11.1
Serine	126 ± 54.6	139 ± 36.7
Succinate	28.7 ± 13.2	25.3 ± 15.5
Sucrose	22.5 ± 7.86	26.0 ± 18.7
Tartrate	11.9 ± 19.4	10.8 ± 2.68
Taurine	213 ± 152	243 ± 81.8
Threonine	75.5 ± 17.9	87.7 ± 19.7
Trigonelline	156 ± 50.4	73.1 ± 41.4
Trimethylamine N-oxide	460 ± 363	341 ± 126
Tryptophan	59.5 ± 8.48	42.0 ± 17.5
Tyrosine	46.7 ± 12.6	59.7 ± 17.9
Uracil	49.2 ± 5.13	27.2 ± 9.85
Valine	31.2 ± 5.37	33.2 ± 5.47
Xylose	70.3 ± 29.3	46.4 ± 16.8
cis-Aconitate	139 ± 46.9	129 ± 73.1
myo-Inositol	87.3 ± 39.8	79.2 ± 68.2
trans-Aconitate	20.8 ± 5.10	21.2 ± 4.07
π-Methylhistidine	245 ± 125	190 ± 116
τ-Methylhistidine	78.2 ± 14.2	107 ± 23.4

[1]Values are median ± SE, Low class n=14, High class n=14.
[2]Low and High class included patients who had values 0.5 SD below and above the median for total fat mass, respectively.
Asterisks indicate different from low total fat mass: *P<0.05, **P<0.01, ***$P\leq$0.001 based on Mann-Whitney nonparametric statistical analysis

Table 5-6: Bivariate analysis: top 30 metabolites related to the best predictive models

	Urinary metabolites		**Plasma metabolites** [2]
	ALST	TFM	TFM
Rank [1]	Metabolite (mutual information value)	Metabolite (mutual information value)	Metabolite (mutual information value)
1	Fumarate (0.162)	3-Indoxylsulfate (0.167)	lysoPC.a.C26:0 (1.126)
2	Creatine (0.133)	Uracil (0.149)	C8 (1.012)
3	4-Hydroxyphenylacetate (0.074)	Fumarate (0.125)	C10 (0.849)
4	Quinolinate (0.058)	O-Acetylcarnitine (0.122)	C14 (0.716)
5	2-Oxoglutarate (0.056)	Methylamine (0.121)	C5 (0.449)
6	Adipate (0.056)	Acetone (0.099)	C8:1 (0.262)
7	Sucrose (0.048)	Taurine (0.095)	C3 (0.221)
8	Betaine (0.045)	Tartrate (0.088)	C0 (0.21)
9	Trigonelline (0.042)	Glycolate (0.084)	PC.aa.C40:4 (0.188)
10	Formate (0.038)	3-Aminoisobutyrate (0.073)	SM.C24:1 (0.181)
11	Glycolate (0.037)	Hypoxanthine (0.069)	PC.aa.C38:3 (0.177)
12	Taurine (0.03)	Trigonelline (0.058)	PC.aa.C40:5 (0.148)
13	O-Acetylcarnitine (0.03)	Carnitine (0.055)	PC.aa.C38:4 (0.14)
14	1-Methylnicotinamide (0.027)	1-Methylnicotinamide (0.054)	PC.aa.C34:4 (0.133)
15	Xylose (0.026)	Tryptophan (0.052)	C16 (0.124)
16	Glucose (0.024)	Adipate (0.047)	C18:2 (0.12)
17	2-Aminobutyrate (0.022)	Dimethylamine (0.043)	PC.aa.C36:4 (0.115)
18	Guanidoacetate (0.022)	Trimethylamine-N-oxide (0.041)	C4 (0.109)

19	3-Aminoisobutyrate (0.022)	2-Oxoglutarate (0.04)	PC.aa.C30:2 (0.108)
20	Methylguanidine (0.022)	Creatinine (0.039)	SM.C16:1 (0.105)
21	Succinate (0.02)	3-Methylhistidine (0.037)	PC.aa.C38:5 (0.095)
22	Tartrate (0.02)	Hippurate (0.036)	lysoPC.a.C14:0 (0.093)
23	Tryptophan (0.019)	Pantothenate (0.035)	PC.aa.C36:2 (0.087)
24	Lactate (0.019)	2-Hydroxyisobutyrate (0.035)	lysoPC.a.C18:0 (0.084)
25	cis-Aconitate (0.019)	Pyroglutamate (0.033)	PC.aa.C24:0 (0.084)
26	Hippurate (0.018)	3-Hydroxybutyrate (0.032)	PC.aa.C4:3 (0.082)
27	Pyroglutamate (0.018)	Threonine (0.031)	SM.C18:1 (0.081)
28	Tyrosine (0.017)	Ethanolamine (0.028)	C12 (0.074)
29	Acetone (0.016)	Lactate (0.026)	PC.aa.C36:1 (0.074)
30	3-Indoxylsulfate (0.016)	Succinate (0.026)	PC.aa.C36:3 (0.072)

[1] We use mutual information to quantify the dependence between two variables (see the Methods section); this allows us to rank metabolites according to the degree of dependence with the two different classes (low and high). In this case mutual information is high when a particular metabolite is highly correlated with ALST or TFM.

[2] Cx:y; acylcarnitine (x= number of carbons in acyl chain, y= location of double bond); PC.aa, phosphatidylcholine diacyl; lysoPC a, lysoPhosphatidylcholine acyl; SM, sphingomyelin.

Table 5-7: Concentration of 143 plasma metabolites quantified by MS for all patients included in data analysis.

Plasma metabolite[2]	***µmol/L***	**Plasma metabolite**	***µmol/L***
glycerophospholipids		acyl carnitines	
PC aa C24:0	0.11 ± 0.010	C0	30.1 ± 1.21
PC aa C28:1	3.01 ± 0.107	C10	0.27 ± 0.018
PC aa C30:0	5.28 ± 0.241	C10:1	0.15 ± 0.013
PC aa C30:2	5.32 ± 0.179	C10:2	0.04 ± 0.004
PC aa C32:0	16.8 ± 0.595	C12	0.1 ± 0.006
PC aa C32:1	26.0 ± 1.68	C12:1	0.000 ± 0.014
PC aa C32:2	7.78 ± 0.353	C14	0.05 ± 0.002
PC aa C32:3	0.74 ± 0.034	C14:1	0.1 ± 0.006
PC aa C34:1	273 ± 10.7	C14:1-OH	0.0 ± 0.001
PC aa C34:2	387 ± 12.1	C14:2	0.04 ± 0.002
PC aa C34:3	21.6 ± 1.32	C16	0.1 ± 0.004
PC aa C34:4	1.87 ± 0.138	C16:2	0.009 ± 0.001
PC aa C36:0	8.37 ± 0.298	C18	0.05 ± 0.002
PC aa C36:1	80.4 ± 2.87	C18:1	0.2 ± 0.007
PC aa C36:2	267 ± 7.23	C18:2	0.05 ± 0.002
PC aa C36:3	176 ± 5.96	C2	6.32 ± 0.420
PC aa C36:4	183 ± 7.73	C3	0.27 ± 0.014
PC aa C36:5	34.0 ± 2.29	C4	0.18 ± 0.012
PC aa C36:6	1.3 ± 0.078	C4-OH (C3-DC)	0.0 ± 0.007
PC aa C38:0	3.92 ± 0.154	C5	0.1 ± 0.007
PC aa C38:1	9.46 ± 0.351	C5:1	0.0 ± 0.002
PC aa C38:3	68.7 ± 2.40	C5:1-DC	0.02 ± 0.001
PC aa C38:4	114 ± 4.06	C7-DC	0.04 ± 0.003
PC aa C38:5	86.0 ± 3.04	C8	0.25 ± 0.010
PC aa C38:6	92.4 ± 4.35	C8:1	0.17 ± 0.012
PC aa C40:1	0.52 ± 0.030		
PC aa C40:2	0.49 ± 0.015		
PC aa C40:3	1.0 ± 0.028	sphingolipids	
PC aa C40:4	5.39 ± 0.219	SM (OH) C14:1	6.98 ± 0.217
PC aa C40:5	17.4 ± 0.727	SM (OH) C16:1	3.97 ± 0.131
PC aa C40:6	33.1 ± 1.74	SM (OH) C22:1	15.7 ± 0.555
PC aa C42:0	0.60 ± 0.042	SM (OH) C22:2	14.9 ± 0.467
PC aa C42:1	0.28 ± 0.0164	SM (OH) C24:1	2.3 ± 0.089
PC aa C42:2	0.2 ± 0.008	SM C16:0	125 ± 3.03
PC aa C42:4	0.2 ± 0.007	SM C16:1	19.7 ± 0.609
PC aa C42:5	0.41 ± 0.016	SM C18:0	27.6 ± 0.939
PC aa C42:6	0.58 ± 0.018	SM C18:1	14.4 ± 0.464
PC ae C30:0	0.41 ± 0.017	SM C20:2	1.9 ± 0.075
PC ae C30:1	2.0 ± 0.061	SM C22:3	19.2 ± 1.24
PC ae C32:1	3.41 ± 0.135	SM C24:0	28.1 ± 1.03

PC ae C32:2	0.62 ± 0.024	SM C24:1	89.1 ± 2.38
PC ae C34:0	2.7 ± 0.092	SM C26:0	0.42 ± 0.014
PC ae C34:1	11.5 ± 0.407	SM C26:1	0.86 ± 0.024
PC ae C34:2	11.4 ± 0.484		
PC ae C34:3	6.54 ± 0.377	Amino acids	
PC ae C36:0	1.6 ± 0.051	Arginine	88.9 ± 2.73
PC ae C36:1	9.53 ± 0.334	Glutamine	664 ± 16.5
PC ae C36:2	14.8 ± 0.497	Glycine	238 ± 19.7
PC ae C36:3	10.3 ± 0.360	Histidine	88.8 ± 2.43
PC ae C36:4	16.3 ± 0.691	Methionine	26.3 ± 0.774
PC ae C36:5	10.2 ± 0.496	Ornithine	50.8 ± 2.01
PC ae C38:0	3.08 ± 0.138	Phenylalanine	51.2 ± 1.36
PC ae C38:1	5.88 ± 0.203	Proline	174 ± 6.46
PC ae C38:2	6.56 ± 0.195	Serine	82.9 ± 2.34
PC ae C38:3	5.76 ± 0.177	Threonine	108 ± 3.33
PC ae C38:4	14.8 ± 0.553	Tryptophan	65.6 ± 1.62
PC ae C38:5	20.4 ± 0.702	Tyrosine	73.6 ± 2.39
PC ae C38:6	8.86 ± 0.357	Valine	223 ± 7.76
PC ae C40:0	0.000 ± 1.11	Leucine/Isoleucine	175 ± 5.39
PC ae C40:1	1.9 ± 0.067		
PC ae C40:2	1.8 ± 0.060	Hexose sugars	4650 ± 137.0
PC ae C40:3	1.2 ± 0.042		
PC ae C40:4	2.46 ± 0.103		
PC ae C40:5	4.18 ± 0.173		
PC ae C40:6	5.65 ± 0.233		
PC ae C42:0	0.49 ± 0.031		
PC ae C42:1	0.36 ± 0.012		
PC ae C42:2	0.65 ± 0.024		
PC ae C42:3	0.82 ± 0.039		
PC ae C42:4	1.1 ± 0.067		
PC ae C42:5	2.12 ± 0.150		
PC ae C44:3	0.1 ± 0.007		
PC ae C44:4	0.54 ± 0.038		
PC ae C44:5	1.89 ± 0.135		
PC ae C44:6	1.2 ± 0.086		
lysoPC a C14:0	6.3 ± 0.086		
lysoPC a C16:0	105 ± 3.32		
lysoPC a C16:1	3.49 ± 0.151		
lysoPC a C17:0	2.1 ± 0.090		
lysoPC a C18:0	33.4 ± 1.16		
lysoPC a C18:1	27.2 ± 1.23		
lysoPC a C18:2	25.9 ± 1.47		
lysoPC a C20:3	2.98 ± 0.119		
lysoPC a C20:4	6.29 ± 0.333		
lysoPC a C26:0	0.65 ± 0.024		

lysoPC a C28:0	0.51 ± 0.016		
lysoPC a C28:1	0.73 ± 0.019		

[1]Values are median ± SE, *n*=55.

[2]Cx:y; acylcarnitine (x= number of carbons in acyl chain, y= location of double bond); PC.aa, phosphatidylcholine diacyl; lysoPC a, lysoPhosphatidylcholine acyl; SM, sphingomyelin.

Table 5-8: Concentration of 31 plasma metabolites quantified by NMR for all patients included in data analysis[1].

Plasma metabolite	^{1}H chemical shift (*ppm*) and coupling[2]	*µmol/L*
2-Aminobutyrate	0.97(t), 1.89(m), 3.71(dd)	20 ± 1.2
2-Hydroxybutyrate	1.35(s)	38.8 ± 2.26
3-Hydroxybutyrate	1.19(d), 2.29(dd), 2.40(dd), 1.14(m)	43.0 ± 11.1
Acetate	1.91(s)	44.4 ± 2.56
Acetoacetate	2.27(m), 3.43(m)	18.3 ± 3.26
Acetone	2.22(s)	8.26 ± 1.79
Alanine	1.47(d), 3.78(qt)	384 ± 13.4
Arginine	1.68(m), 1.90(m), 3.23(t), 3.76(t)	74.4 ± 3.97
Asparagine	2.86(dd), 2.95(dd), 3.99(dd), 6.91(s), 7.62(s)	41.9 ± 3.93
Betaine	3.26(s), 3.89(s)	42.6 ± 2.24
Carnitine	2.41(dd), 2.45(dd), 3.22(s), 3.40(m), 3.43(m), 4.56(m)	42.4 ± 1.84
Citrate	2.53(d), 2.69(d)	85.8 ± 4.93
Creatine	3.02(s), 3.92(s)	25.1 ± 4.28
Creatinine	3.03(s), 4.05(s)	69.0 ± 3.01
Glucose	3.24(dd), 3.40(m), 3.41(m), 3.47(m), 3.49(m), 3.53(dd), 3.71(m), 3.72(dd), 3.74(m), 3.82(m), 3.84(m), 3.90(dd), 4.64(d), 5.23(d)	4660 ± 185.0
Glutamine	2.12(m), 2.15(m), 2.43(m), 2.47(m), 3.76(t), 6.87(s), 7.58(s)	517 ± 21.9
Glycerol	3.55(m), 3.64(m), 3.78(m)	69.9 ± 5.34
Glycine	3.56(s)	199 ± 12.4
Isoleucine	0.93(t), 1.00(d), 1.25(m), 1.46(m), 1.97(m), 3.66(d)	65.5 ± 2.21
Lactate	1.32(d), 4.11(d)	1200 ± 56.00
Leucine	0.95(d), 0.96(d), 1.67(m), 1.70(m), 1.73(m), 3.73(m)	106 ± 3.76
Lysine	1.43(m), 1.50(m), 1.72(m), 1.88(m), 1.91(m), 3.02(t), 3.75(t)	129 ± 6.22
Methionine	2.16(m), 2.63(t), 3.85(dd)	15 ± 0.70
Ornithine	1.73(m), 1.83(m), 1.93(m), 3.05(t), 3.78(t)	43.2 ± 2.70
Phenylalanine	3.19(m), 3.98(dd), 7.32(d), 7.36(m), 7.42(m)	56.6 ± 3.61
Proline	1.99(m), 2.06(m), 2.34(m), 3.33(dt), 3.41(dt), 4.12(dd_	152 ± 8.32
Pyruvate	2.37(s)	39.6 ± 4.46
Serine	3.84(dd), 3.94(dd), 3.98(dd)	104 ± 6.73
Threonine	1.32(d), 3.59(d), 4.26(m)	55.6 ± 5.44
Tyrosine	3.05(dd), 3.19(dd), 3.94(q), 6.88(m), 7.20(m)	54.7 ± 5.45
Valine	0.98(d), 1.03(d), 2.26(m), 3.61(d)	220 ± 7.16

[1]Values are median ± SE, *n*=55.

[2]s=singlet, d=doublet, dd=doublet-doublet, t=triplet, q=quartet, m=multiplet

Table 5-9: Concentration of 142 plasma metabolites quantified by MS from cancer patients included in the high and low total fat mass classes[1, 2]

Plasma metabolite[3]	Patients with low fat mass	Patients with high fat mass
	µmol/L	
Arginine	97.8 ± 5.91	82.3 ± 4.21
Glutamine	682 ± 28.4	655 ± 43.2
Glycine	268 ± 13.24	221 ± 10.8*
Histidine	92.6 ± 5.79	96.8 ± 4.23
Methionine	25.4 ± 1.31	27.9 ± 2.16
Ornithine	50.5 ± 3.95	51.9 ± 3.80
Phenylalanine	52.2 ± 2.70	57.6 ± 3.24
Proline	165 ± 17.5	176 ± 11.1
Serine	88.3 ± 4.15	84.2 ± 4.77
Threonine	113 ± 5.49	98.5 ± 8.49
Tryptophan	65.0 ± 3.45	70.0 ± 3.62
Tyrosine	70.6 ± 4.39	82.6 ± 6.19*
Valine	220± 14.9	263± 15.2
Leucine/Isoleucine	170 ± 12.0	190 ± 11.1
C0	25.9 ± 1.90	32.5 ± 1.57**
C10	0.3 ± 0.04	0.26 ± 0.016
C10:1	0.17 ± 0.030	0.15 ± 0.020
C10:2	0.0 ± 0.009	0.05 ± 0.008
C12	0.15 ± 0.013	0.1 ± 0.009
C12:1	0.0 ± 0.03	0.0 ± 0.03
C14	0.05 ± 0.005	0.05 ± 0.004
C14:1	0.12 ± 0.010	0.10 ± 0.011
C14:1-OH	0.008 ± 0.003	0.02 ± 0.002
C14:2	0.05 ± 0.005	0.04 ± 0.004
C16	0.1 ± 0.009	0.1 ± 0.006
C16:2	0.01 ± 0.002	0.009 ± 0.002
C18	0.05 ± 0.003	0.05 ± 0.003
C18:1	0.17 ± 0.011	0.2 ± 0.009
C18:2	0.05 ± 0.004	0.05 ± 0.003
C2	6.65 ± 0.403	7.09 ± 0.565
C3	0.24 ± 0.014	0.34 ± 0.033**
C4	0.13 ± 0.012	0.17 ± 0.021*
C4-OH (C3-DC)	0.00 ± 0.015	0.00 ± 0.014
C5	0.080 ± 0.011	0.12 ± 0.019*
C5:1	0.0 ± 0.005	0.0 ± 0.007
C5:1-DC	0.0 ± 0.003	0.009 ± 0.003
C7-DC	0.04 ± 0.007	0.04 ± 0.006
C8	0.28 ± 0.027	0.24 ± 0.011
C8:1	0.12 ± 0.014	0.20 ± 0.020*

PC aa C24:0	0.056 ± 0.017	0.109 ± 0.017
PC aa C28:1	2.90 ± 0.266	3.36 ± 0.194
PC aa C30:0	5.595 ± 0.409	5.67 ± 0.402*
PC aa C30:2	4.76 ± 0.247	6.28 ± 0.414
PC aa C32:0	17 ± 1.012	17.9 ± 1.19
PC aa C32:1	25.9 ± 2.84	28.6 ± 3.77
PC aa C32:2	7.93 ± 0.699	8.14 ± 0.626
PC aa C32:3	0.65 ± 0.064	0.73 ± 0.046
PC aa C34:1	247± 17.9	285 ± 25.4
PC aa C34:2	414 ± 23.8	443 ± 25.6
PC aa C34:3	21.8 ± 2.89	24.3 ± 2.20
PC aa C34:4	1.75 ± 0.328	2.14 ± 0.159
PC aa C36:0	8.07 ± 0.534	9.14 ± 0.678
PC aa C36:1	78.7 ± 5.07	87.5 ± 6.89
PC aa C36:2	267 ± 15.5	286 ± 14.7
PC aa C36:3	172 ± 12.4	196 ± 13.6
PC aa C36:4	174 ± 16.0	200 ± 14.7
PC aa C36:5	30.0 ± 3.14	32.8 ± 3.14
PC aa C36:6	1.22 ± 0.12	1.07 ± 0.116
PC aa C38:0	3.30 ± 0.370	3.42 ± 0.256
PC aa C38:1	9.55 ± 0.675	9.94 ± 0.763
PC aa C38:3	60.8 ± 4.05	75.1 ± 5.44**
PC aa C38:4	108 ± 7.84	120 ± 9.06*
PC aa C38:5	74.4 ± 6.34	78.1 ± 6.12
PC aa C38:6	72.5 ± 7.00	79.0 ± 8.65
PC aa C40:1	0.48 ± 0.069	0.478 ± 0.057
PC aa C40:2	0.51 ± 0.033	0.52 ± 0.031
PC aa C40:3	1.04 ± 0.067	1.1 ± 0.068
PC aa C40:4	4.84 ± 0.429	5.53 ± 0.473*
PC aa C40:5	14.9 ± 1.27	17.7 ± 1.39*
PC aa C40:6	25.8 ± 2.21	33.0 ± 3.38
PC aa C42:0	0.59 ± 0.085	0.54 ± 0.050
PC aa C42:1	0.27 ± 0.036	0.25 ± 0.021
PC aa C42:2	0.17 ± 0.018	0.16 ± 0.016
PC aa C42:4	0.21 ± 0.017	0.23 ± 0.012
PC aa C42:5	0.39 ± 0.031	0.40 ± 0.026
PC aa C42:6	0.54 ± 0.036	0.56 ± 0.038
PC ae C30:0	0.48 ± 0.038	0.43 ± 0.026
PC ae C30:1	1.99 ± 0.161	2.1 ± 0.097
PC ae C32:1	3.34 ± 0.350	3.52 ± 0.253
PC ae C32:2	0.62 ± 0.056	0.68 ± 0.051
PC ae C34:0	2.80 ± 0.216	2.81 ± 0.151
PC ae C34:1	12.2 ± 0.920	12.0 ± 0.789
PC ae C34:2	14.3 ± 1.20	11.6 ± 0.936
PC ae C34:3	8.07 ± 0.771	6.67 ± 0.936

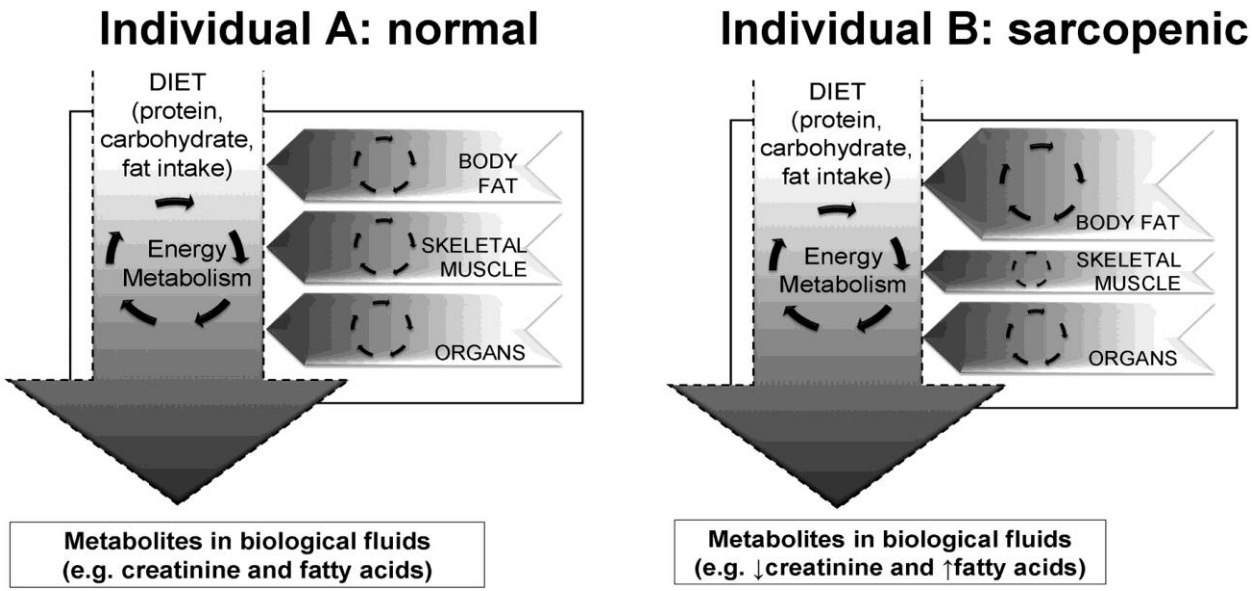
PC ae C36:0	1.58 ± 0.103	1.74 ± 0.111
PC ae C36:1	10.2 ± 0.798	10.1 ± 0.591
PC ae C36:2	16.3 ± 1.21	15.5 ± 0.883
PC ae C36:3	11.0 ± 0.944	10.4 ± 0.764
PC ae C36:4	15.9 ± 1.90	18.5 ± 1.42
PC ae C36:5	10.3 ± 1.28	10.6 ± 1.11
PC ae C38:0	2.88 ± 0.233	2.74 ± 0.276
PC ae C38:1	5.97 ± 0.444	6.32 ± 0.372
PC ae C38:2	6.66 ± 0.428	6.77 ± 0.379
PC ae C38:3	5.93 ± 0.523	6.03 ± 0.353
PC ae C38:4	13.6 ± 1.49	16.2 ± 1.18
PC ae C38:5	18.9 ± 1.80	22.6 ± 1.46
PC ae C38:6	8.09 ± 0.917	8.80 ± 0.668
PC ae C40:0	0.000 ± 1.71	0.000 ± 2.05
PC ae C40:1	1.87 ± 0.134	1.79 ± 0.154
PC ae C40:2	1.89 ± 0.119	1.90 ± 0.144
PC ae C40:3	1.28 ± 0.093	1.3 ± 0.073
PC ae C40:4	2.23 ± 0.244	2.68 ± 0.175
PC ae C40:5	3.69 ± 0.425	4.17 ± 0.308
PC ae C40:6	4.46 ± 0.544	5.11 ± 0.416
PC ae C42:0	0.44 ± 0.075	0.45 ± 0.056
PC ae C42:1	0.36 ± 0.028	0.37 ± 0.026
PC ae C42:2	0.61 ± 0.045	0.66 ± 0.043
PC ae C42:3	0.83 ± 0.073	0.83 ± 0.066
PC ae C42:4	1.15 ± 0.131	1.0 ± 0.091
PC ae C42:5	2.17 ± 0.288	2.03 ± 0.264
PC ae C44:3	0.16 ± 0.010	0.1± 0.009
PC ae C44:4	0.55 ± 0.062	0.51 ± 0.048
PC ae C44:5	1.78 ± 0.216	1.79 ± 0.175
PC ae C44:6	1.29 ± 0.143	1.22 ± 0.116
lysoPC a C14:0	5.92 ± 0.195	6.16 ± 0.104
lysoPC a C16:0	108 ± 7.45	108 ± 7.27
lysoPC a C16:1	3.56 ± 0.257	3.37 ± 0.384
lysoPC a C17:0	2.14 ± 0.193	1.94 ± 0.152
lysoPC a C18:0	32.0 ± 2.47	34.0 ± 2.19
lysoPC a C18:1	27.9 ± 1.87	25.6 ± 2.54
lysoPC a C18:2	29.2 ± 2.76	25.6 ± 2.98
lysoPC a C20:3	2.53 ± 0.213	3.24 ± 0.237
lysoPC a C20:4	6.20 ± 0.441	6.31 ± 0.587
lysoPC a C26:0	0.69 ± 0.027	0.79 ± 0.083
lysoPC a C28:0	0.51 ± 0.022	0.52 ± 0.029
lysoPC a C28:1	0.70 ± 0.043	0.73 ± 0.033
SM (OH) C14:1	7.01 ± 0.529	7.21 ± 0.372
SM (OH) C16:1	4.46 ± 0.300	4.35 ± 0.259
SM (OH) C22:1	16.0 ± 1.15	16.1 ± 0.961

SM (OH) C22:2	14.8 ± 0.907	15.1 ± 0.943
SM (OH) C24:1	2.47 ± 0.176	2.41 ± 0.199
SM C16:0	127 ± 5.03	129 ± 7.05
SM C16:1	18.2 ± 0.852	22.5 ± 1.45*
SM C18:0	31.4 ± 1.48	28.0 ± 2.28
SM C18:1	13.7 ± 0.710	15.4 ± 1.12
SM C20:2	1.85 ± 0.175	2.00 ± 0.106
SM C22:3	17.7 ± 1.76	19.1 ± 1.76
SM C24:0	28.1 ± 1.74	31.2 ± 2.19
SM C24:1	88.3 ± 2.37	94.3 ± 5.92
SM C26:0	0.42 ± 0.024	0.43 ± 0.026
SM C26:1	0.87 ± 0.044	0.94 ± 0.053
Hexose	4520 ± 142.0	4710 ± 251.0

[1]Values are median ± SE, Low class *n*=14, High class *n*=14.
[2]Low and High class included patients who had values 0.5 SD below and above the median for total fat mass, respectively.
[3]Cx:y; acylcarnitine (x= number of carbons in acyl chain, y= location of double bond); PC.aa, phosphatidylcholine diacyl; lysoPC a, lysoPhosphatidylcholine acyl; SM, sphingomyelin. Asterisks indicate different from low total fat mass: **P*<0.05, ***P*<0.01, ****P*≤0.001 based on Mann-Whitney nonparametric statistical analysis

Figures

Figure 5-1: A conceptual framework: contributions of multiple elements (diet, lean and fat mass, energy metabolism) to the metabolome of different biofluids



Considering two individuals of equal body weight and equal macronutrient intake, but with different muscle and fat mass (A. normal, B high relative fat mass), it would be expected that metabolic fluxes in B would have a proportionately higher amount of fat-derived metabolites and a lower proportion of muscle-derived metabolites.

References

1. Gibney MJ, Walsh M, Brennan L, Roche HM, German B, van Ommen B. Metabolomics in human nutrition: opportunities and challenges. Am J Clin Nutr. 2005;82:497-503.

2. Zivkovic AM, German JB. Metabolomics for assessment of nutritional status. Curr Opin Clin Nutr Metab Care. 2009;12:501-7.

3. Zeisel SH, Freake HC, Bauman DE, Bier DM, Burrin DG, German JB, et al. The nutritional phenotype in the age of metabolomics. J Nutr. 2005;135:1613-6.

4. Wishart DS. Metabolomics: Applications to Food Science and Nutrition Research. . Trends in Food Science & Technology. 2008;19:482-93.

5. Scalbert A, Brennan L, Fiehn O, Hankemeier T, Kristal BS, van Ommen B, et al. Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. Metabolomics. 2009;5:435-58.

6. Stalmach A, Mullen W, Barron D, Uchida K, Yokota T, Cavin C, et al. Metabolite profiling of hydroxycinnamate derivatives in plasma and urine after the ingestion of coffee by humans: identification of biomarkers of coffee consumption. Drug Metab Dispos. 2009;37:1749-58.

7. van Velzen EJ, Westerhuis JA, van Duynhoven JP, van Dorsten FA, Grun CH, Jacobs DM, et al. Phenotyping tea consumers by nutrikinetic analysis of polyphenolic end-metabolites. J Proteome Res. 2009;8:3317-30.

8. Llorach R, Urpi-Sarda M, Jauregui O, Monagas M, Andres-Lacueva C. An LC-MS-based metabolomics approach for exploring urinary metabolome modifications after cocoa consumption. J Proteome Res. 2009;8:5060-8.

9. Llorach R, Garrido I, Monagas M, Urpi-Sarda M, Tulipani S, Bartolome B, et al. Metabolomics study of human urinary metabolome modifications after intake of almond (Prunus dulcis (Mill.) D.A. Webb) skin polyphenols. J Proteome Res. 2010;9:5859-67.

10. Wang Y, Tang H, Nicholson JK, Hylands PJ, Sampson J, Holmes E. A metabonomic strategy for the detection of the metabolic effects of chamomile (Matricaria recutita L.) ingestion. J Agric Food Chem. 2005;53:191-6.

11. Primrose S, Draper J, Elsom R, Kirkpatrick V, Mathers JC, Seal C, et al. Metabolomics and human nutrition. Br J Nutr. 2011:1-7.

12. Lieffers JR, Mourtzakis M, Hall KD, McCargar LJ, Prado CM, Baracos VE. A viscerally driven cachexia syndrome in patients with advanced colorectal cancer: contributions of organ and tumor mass to whole-body energy demands. The American journal of clinical nutrition. 2009;89:1173-9.

13. Prado CM, Lieffers JR, McCargar LJ, Reiman T, Sawyer MB, Martin L, et al. Prevalence and clinical implications of sarcopenic obesity in patients with solid tumours of the respiratory and gastrointestinal tracts: a population-based study. The lancet oncology. 2008;9:629-35.

14. Heymsfield SB, Arteaga C, McManus C, Smith J, Moffitt S. Measurement of muscle mass in humans: validity of the 24-hour urinary creatinine method. The American journal of clinical nutrition. 1983;37:478-94.

15. Hutton JL, Martin L, Field CJ, Wismer WV, Bruera ED, Watanabe SM, et al. Dietary patterns in patients with advanced cancer: implications for anorexia-cachexia therapy. The American journal of clinical nutrition. 2006;84:1163-70.

16. Eisner R, Stretch C, Eastman T, Xia J, Hau D, Damaraju S, et al. Learning to predict cancer-associated skeletal muscle wasting from 1H-NMR profiles of urinary metabolites. Metabolomics. 2010;7:25-34.

17. Chorell E, Moritz T, Branth S, Antti H, Svensson MB. Predictive metabolomics evaluation of nutrition-modulated metabolic stress responses in human blood serum during the early recovery phase of strenuous physical exercise. J Proteome Res. 2009;8:2966-77.

18. Miccheli A, Marini F, Capuani G, Miccheli AT, Delfini M, Di Cocco ME, et al. The influence of a sports drink on the postexercise metabolism of elite athletes as investigated by NMR-based metabolomics. J Am Coll Nutr. 2009;28:553-64.

19. Walsh MC, Brennan L, Pujos-Guillot E, Sebedio JL, Scalbert A, Fagan A, et al. Influence of acute phytochemical intake on human urinary metabolomic profiles. Am J Clin Nutr. 2007;86:1687-93.

20. Posner BM, Martin-Munley SS, Smigelski C, Cupples LA, Cobb JL, Schaefer E, et al. Comparison of techniques for estimating nutrient intake: the Framingham Study. Epidemiology. 1992;3:171-7.

21. Price GM, Halliday D, Pacy PJ, Quevedo MR, Millward DJ. Nitrogen homeostasis in man: influence of protein intake on the amplitude of diurnal cycling of body nitrogen. Clin Sci (Lond). 1994;86:91-102.

22. Prado CM, Mourtzakis M, Baracos VE, Reiman T, Sawyer MB, McCargar LJ. Overweight and obese patients with solid tumors may have sarcopenia, poor

prognosis and early features of cachexia. International Journal of Body Composition Research. 2010;8:7-15.

23. Mourtzakis M, Prado CM, Lieffers JR, Reiman T, McCargar LJ, Baracos VE. A practical and precise approach to quantification of body composition in cancer patients using computed tomography images acquired during routine care. Appl Physiol Nutr Metab. 2008;33:997-1006.

24. Heymsfield SB, Smith R, Aulet M, Bensen B, Lichtman S, Wang J, et al. Appendicular skeletal muscle mass: measurement by dual-photon absorptiometry. Am J Clin Nutr. 1990;52:214-8.

25. Cao DX, Wu GH, Zhang B, Quan YJ, Wei J, Jin H, et al. Resting energy expenditure and body composition in patients with newly detected cancer. Clin Nutr. 2010;29:72-7.

26. Holmes E, Foxall PJ, Nicholson JK, Neild GH, Brown SM, Beddell CR, et al. Automatic data reduction and pattern recognition methods for analysis of 1H nuclear magnetic resonance spectra of human urine from normal and pathological states. Anal Biochem. 1994;220:284-96.

27. Craig A, Cloarec O, Holmes E, Nicholson JK, Lindon JC. Scaling and normalization effects in NMR spectroscopic metabonomic data sets. Analytical chemistry. 2006;78:2262-7.

28. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. Analytical chemistry. 2006;78:4281-90.

29. Westerhuis JA, Hoefsloot HCJ, Smit S, Vis DJ, Smilde AK, van Velzen EJJ, et al. Assessment of PLSDA cross validation. Metabolomics. 2008;4:81-9

30. Mahadevan S, Shah SL, Marrie TJ, Slupsky CM. Analysis of metabolomic data using support vector machines. Anal Chem. 2008;80:7562-70.

31. Friedman J, Hastie T, Hofling H, Tibshirani R. Pairwise coordinate optimization. The Annals of Applied Statistics. 2007;1:302-32.

32. Gavaghan CL, Wilson ID, Nicholson JK. Physiological variation in metabolic phenotyping and functional genomic studies: use of orthogonal signal correction and PLS-DA. FEBS Lett. 2002;530:191-6.

33. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. SIGKDD Explorations. 2009;11:10-8.

34. Friedman J, Hastie T, Tibshirani R. Regularized paths for generalized linear models via coordinate descent. Journal of Statistical Software. 2010;33.

35. Saude EJ, Sykes BD. Urine stability for metabolomic studies: effects of preparation and storage. Metabolomics. 2007;3:19-27.

36. Weinsier RL, Schutz Y, Bracco D. Reexamination of the relationship of resting metabolic rate to fat-free mass and to the metabolically active components of fat-free mass in humans. Am J Clin Nutr. 1992;55:790-4.

37. Arends J, Bodoky G, Bozzetti F, Fearon K, Muscaritoli M, Selga G, et al. ESPEN Guidelines on Enteral Nutrition: Non-surgical oncology. Clin Nutr. 2006;25:245-59.

38. Baracos VE, Reiman T, Mourtzakis M, Gioulbasanis I, Antoun S. Body composition in patients with non-small cell lung cancer: a contemporary view of cancer cachexia with the use of computed tomography image analysis. Am J Clin Nutr. 2010;91:1133S-7S.

39. Thorpe JM, Roberts SA, Ball RO, Pencharz PB. Prior protein intake may affect phenylalanine kinetics measured in healthy adult volunteers consuming 1 g protein. kg-1. d-1. J Nutr. 1999;129:343-8.

40. Siragusa RJ, Cerda JJ, Baig MM, Burgin CW, Robbins FL. Methanol production from the degradation of pectin by human colonic bacteria. Am J Clin Nutr. 1988;47:848-51.

41. Stegink LD, Brummel MC, Filer LJ, Baker GL. Blood methanol concentrations in one-year-old infants administered graded doses of aspartame. J Nutr. 1983;113:1600-6.

42. Ferrer I, Ruiz-Sala P, Vicente Y, Merinero B, Pérez-Cerdá C, Ugarte M. Separation and identification of plasma short-chain acylcarnitine isomers by HPLC/MS/MS for the differential diagnosis of fatty acid oxidation defects and organic acidemias. J Chromatogr B Analyt Technol Biomed Life Sci. 2007;860(1):121-6.

CHAPTER 6: Learning to predict cancer-associated skeletal muscle wasting from ^{1}H-NMR profiles of urinary metabolites

6.1 Introduction

Metabolic, neuronal and hormonal controls normally ensure that body weight and composition are maintained constant during adult life. Involuntary weight gains or losses are significant perturbations of this precise control. The focus of our research is cancer-associated muscle wasting. Muscle depletion is associated with poor functional status, treatment toxicity and shorter life expectancy (1-5). Prado *et al.* (2008), and others have shown that muscle loss may occur independently of changes in fat mass, and that muscle wasting may be an early or occult phenomenon that is difficult to detect against the background of overall body weight and body weight change, especially in overweight or obese individuals. A recent consensus definition of cachexia (6) makes a distinction between behavior of skeletal muscle and adipose tissue: "cachexia is a complex metabolic syndrome associated with underlying illness and *characterized by loss of muscle with or without loss of fat mass*…". Muscle wasting may go unnoticed in its early stages, if progressing slowly, or if it is obscured by changes in other tissues. Improved approaches for detecting the onset and evolution of muscle wasting would help manage wasting syndromes and facilitate early intervention. Wasting has a cumulative nature. For example, muscle loss at a rate of 0.07% in

A version of this chapter has been published in *Metabolomics*.
Roman Eisner†, Cynthia Stretch†, Thomas Eastman, Jianguo Xia, David Hau, Sambasivarao Damaraju, Russell Greiner, David S. Wishart, Vickie Baracos. Learning to predict cancer-associated skeletal muscle wasting from ¹H-NMR profiles of urinary metabolites. *Metabolomics*. 2011;7(1):25-34
† Contributed equally to this work

one day appears trivial, but if sustained equals 7% loss over 100 days and 25% loss over a year – a quantity that would have important physiological consequences for most individuals. Dual energy X-ray absorptiometry (DXA), computed tomography (CT) and magnetic resonance imaging (MRI) are considered the most precise measures of adipose and muscle tissues currently available (7-10), but have several limitations. Images must be repeated over time to detect loss, access and cost may be limitations and their analysis may be time-consuming and labor-intensive and DXA and CT expose patients to radiation. Clinicians are keen to find new diagnostic approaches for identifying and monitoring muscle loss that are faster, cheaper, safer and more accessible.

We hypothesized that metabolites produced from tissue breakdown are likely to be a sensitive indicator of muscle wasting and may provide a new diagnostic approach. Muscle breakdown generates amino acids and their various catabolites, as well as urea and creatinine. Several of these end products are detectable in physiological fluids using NMR spectroscopy (11, 12). Coupled with recent advances in machine learning and multivariate statistics, metabolomic approaches have led to the identification of biomarkers for several diseases (i.e. celiac disease, prostate cancer) (13, 14). Based on these ideas we investigated whether we could detect muscle wasting using metabolomic data from urine samples from patients with cancer. Urine was selected as the biofluid of choice, since several end products of muscle catabolism (i.e. creatinine, methylhistidine) are specifically excreted in urine. We applied machine-learning techniques to

identify patterns of urinary metabolic profiles that discriminate the condition of muscle loss.

6.2 Methods

6.2.1 Study design and sample collection

This study was approved by the Alberta Cancer Board Research Ethics Board. All participants provided written informed consent and had advanced cancer of the colon or lung, defined as locally recurrent or metastatic. Patients with prior radiation to the kidneys or malignancy of the kidney or urinary tract were excluded as these independently alter the ability to concentrate urine normally. Urine was selected as the biofluid of choice. Patients donated a urine sample taken at random (i.e. not controlled for time of day, or food intake) during a routine visit to the cancer center (n=93). We did not undertake a 24 h urine collection, as the patients' ages and medical conditions (life-limiting illness) limit their inclination to commit to this additional burden. Preliminary data from our group suggests that urine volumes do not differ between cancer patients similar to those studied here (n=17, mean 24 hour urine volume 0.025 ± 0.009 L/kg body weight) and age and sex matched healthy controls (n=25, 0.024 ± 0.011 L/kg body weight). After adding sodium azide to a final concentration of 0.02% to prevent bacterial growth, samples were stored frozen at -80 °C until NMR data acquisition.

6.2.2 CT image analysis

We compared our predictive model, obtained by applying machine learning methods to urine metabolites, with muscle loss quantified in serial CT images (15). Images were analyzed for total skeletal muscle tissue cross-sectional area (cm^2) at the 3^{rd} lumbar vertebra using Slice-O-Matic software V4.3 (Tomovision, Montreal, Canada). Further details of image analysis can be found elsewhere (9, 10, 16) and prior studies in the Baracos laboratory (1, 17). During routine clinical care tumor progression is assessed by CT at intervals of ~100 days. Two scans (preceding and following the urine sample) were selected. Muscle area in the CT image preceding the urine collection was used as a reference (baseline) to compute the % loss or gain. We expressed this rate as % change per 100 days, to take into account minor variation in the number of days between scans for different individuals.

6.2.3 NMR spectra acquisition and targeted profiling

Urine samples were prepared by adding 65 μL of an internal chemical shift standard (supplied by Chenomx Inc., Edmonton, Canada consisting of 5 mM sodium 2,2-dimethyl-2- silapentane-5-sulfonate-d6 (DSS-d6) and 0.2% sodium azide in 99% D_2O) to 585 μL of urine. Using small amounts of NaOH or HCl, the sample was adjusted to pH 6.75 ± 0.05. A 600 μL aliquot of prepared sample was placed in a 5-mm NMR tube (Wilmad, Buena, NJ). All one-dimensional NMR spectra of urine samples were acquired using the first increment of the standard NOESY pulse sequence using a 600 MHz Varian INOVA NMR spectrometer

(Varian Inc., Palo Alto, CA) equipped with a triple axis-gradient 5-mm HCN probe.

We used a *targeted profiling* or quantitative metabolomic approach (12, 18) to identify and quantify metabolites from the resulting NMR spectra using Chenomx NMRSuite 4.6 (Chenomx Inc. Edmonton, Canada). Quantitative approaches are more interpretable than spectral binning and are also more robust with respect to compound overlap, and variability in solution conditions (i.e. pH and ionic strength) (12, 19). Two analysts (DDH, CS) independently analyzed the spectra and we included only those compounds and concentrations agreed upon by both analysts. Compound spiking with authentic standards from the Human Metabolome Library (20) was used to confirm the identity of difficult-to-assign compounds. As a further check, additional (non-NMR) laboratory analyses were conducted to verify creatinine concentrations and amino acid peak assignments and concentrations. Creatinine was determined colorimetrically (SpectraMAX 190 using SoftMax Pro V5 software) with two different commercial kits based on Jaffè's basic picrate method (Stanbio Creatinine Liquicolor Kit, Cat No. SB 0420-250 and Cedarlane, Creatinine Assay Kit, Cat. No. 500701-480). Amino acid assignments and concentrations were verified by a spike-in experiment with a solution containing Ala, Asn, Gln, Gly, His, Ile, Leu, Lys, Phe, Ser, Tau, Thr, Trp, Tyr, Val, 1-Methylhistidine and 3-Methylhistidine (3-MH). Spiked samples were quantified by NMR as described above and by reverse-phase HPLC using Waters pico-tag® method (Waters Co., MA, USA) (21).

6.2.4 Statistical methods

6.2.4.1 Data preprocessing

Many statistical procedures assume that variables are normally distributed and a significant violation of this assumption can seriously increase the chances of making an analytical error. Data can appear non-normal if some values are extreme outliers relative to the rest of the sample. This frequently happens in urine samples as metabolite concentrations can vary up to several hundred–fold. To correct for this problem, we transformed the data by taking the natural log of the concentration values.

Water intake during the day can alter concentration of metabolites in urine. We employed three approaches to correct for this effect, including a) *normalization to creatinine concentration* in each sample (22)), b) *normalization by total peak area of each sample*; this assumes that the integrated area under an NMR spectrum is a linear function of the detectable metabolite concentrations in the samples (23, 24)) and c) *probability quotient normalization* (25), which calculates a most probable dilution factor (the median) by examining the distribution of the quotients of the amplitudes of a test spectrum by those of a reference spectrum.

6.2.4.2 Development of a Classifier

Metabolomic researchers (11, 13) compute how each individual compound correlates with the outcome — i.e. muscle loss or muscle gain. While such bivariate analyses typically provide valuable biological insights, they do not directly help clinicians who are primarily interested in making a diagnosis. As

our primary goal was to develop a tool that could predict whether a patient is losing muscle based on their urine metabolite concentrations, we considered different analytical tools. For diagnostic applications, it is useful to have a *classifier* that returns a prediction: given the metabolic profile m_r of patient r, determine whether this patient r is losing muscle or not, $c_r \in$ (losing, gaining). A classifier can base its prediction on a potentially complicated combination of all metabolite concentrations.

A sample of historical data (i.e. a collection of patient metabolic profiles, along with their respective muscle loss/gain values ($[m_r, c_r])_r$ over a set of patients r), is used as a starting point. We can use the machine learning approach to computationally *learn* a classifier, from this historical data. The classifier can then be used to predict the status of future patients. We summarize below a number of machine learning approaches.

6.2.4.3 Classifiers considered

We examined classification performance using 8 different standard statistical and machine-learning approaches:

(a) Naïve Bayes – a Bayesian classifier that assumes that metabolite concentrations are all independent of each other, for each of the two classes $C \in$ (losing, gaining) (26).

(b) Tree-augmented Naïve Bayes (TAN) – a Bayesian classifier allowing a simplified set of conditional dependencies between pairs of metabolites (forming a tree structure), for the overall distribution (27).

(c) multi-TAN – identical to TAN except that the tree structure for the two classes is allowed to differ from one another (27).

(d) Full dependence model – a Bayesian network classifier allowing any metabolite concentration to depend on any other metabolite concentration.

(e) Partial Least Squares-Discriminant Analysis (PLS-DA) – a common approach in metabolomics studies that uses an eigenvalue-based approach to create a classifier.

(f) Decision Trees (also called recursive partitioning systems) (28) sequentially decide which feature to examine, based on the observed values of the features already examined, until having enough information to return a class value (26).

(g) Support Vector Machines (14) view each instance as a vector in k-dimensional space, and seek the maximally separating hyperplane between the classes in this space (26). We use a SVM with a linear kernel.

(h) Pathway Informed Analysis (PIA) – A Bayesian classifier using biological knowledge in the form of metabolic pathways, extracted from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (29). PIA treats each pathway as a graph structure, similar to the "substrate graph" structure of Fell & Wagner (2001), where each node of the graph represents a specific metabolite and each edge connects a pair of metabolites that participate in the same reaction (i.e., malate and fumarate) (30). Incorporating pathway knowledge into classifiers represents a confluence of statistical and biological expertise that could improve predictive power (note that none of the other 7 learning algorithms use biological information).

6.2.4.4 Prediction accuracy of classifiers

We used cross–validation and permutation tests to assess the accuracy of our classifiers. The quality of a classifier is defined by how well it performs on novel test instances that were not part of the training set. Such an evaluation could be based on an external validation set – i.e., new data that the learner has not previously seen. Here, we used cross-validation (31) to approximate an external validation set. This involves partitioning the training data into k=5 subsets; then k times we first produce a classifier based on (k-1) subsets of the data, which we then test of the remaining subset. We then use these k evaluation scores to estimate the mean and variance of the accuracy (on novel data) that we would obtain using a classifier built using the entire training set.

Permutations tests are particularly useful for confirming robustness of the classifier and for ensuring it has not been over-trained (32). We first randomly permute the labels (muscle loss status) for the training data, then run the entire cross-validation evaluation process on this newly re-labeled data. As permutation removes any correlation between data and label, we should get just "noise" on the permuted data. We then compared the diagnostic accuracy on the original un-permuted data, with the distribution of the accuracy obtained using the various permuted datasets. This allowed us to estimate the likelihood that results from un-permutated data were due to chance.

6.2.4.4 Bivariate analysis

A standard approach to analyzing quantitative metabolomic data is bivariate analysis -- i.e. finding the degree to which the primary outcome depends

on each individual metabolite. Each highly-dependent metabolite is a feature that is associated with the biological process of interest (here, muscle wasting). We focus on binary classification, labeling each patient as either losing or gaining muscle and use *mutual information* (33) to quantify the dependence between the binary class outcome $C \in$ (losing, gaining) and the real-valued concentration of each of the 63 metabolites $M \in$ *(*fumarate, malate, oxaloacetate *...)*, which we assume follows a Gaussian distribution. This involves computing:

$$MI(M,C) = \sum_c \int_{-\infty}^{+\infty} p(m,c) \log \frac{p(m,c)}{p(m)P(c)} dm$$

where P(c) is the probability that the class C=c (here, we set P(C=losing) to be the observed frequency of "muscle losing" patients in the data sample) and $p(m) = 1/\sqrt{2\pi\sigma_m{}^2} \exp\left(-(m-\mu_m)^2/\left(2\sigma_m{}^2\right)\right)$ is the Gaussian probability density function, which is based on the mean μ_m and variance $\sigma_m{}^2$ estimated from the data sample. We use a similar function for *p(m,c),* using estimated mean and variance that depends on whether the class is *C=losing* or *C=gaining*. Notice that MI(M,C) is 0 if the metabolite *M* is completely independent of C; larger values indicate a higher degree of correlation.

6.3 Results

6.3.1 Muscle loss continuum in advanced cancer patients

Figure 6-1 shows the distribution of muscle loss and gain for the 93 samples in our study. Patients in the 2 classes (Table 6-1) did not differ in age,

sex, cancer site or stage and these features were uncorrelated with one another and also uncorrelated with muscle loss/gain. Because the measurements of muscle change are only precise to about 1.5%/100 days, we adopted the following simple classification rule: patients were designated as not losing muscle (anabolic) if the change in muscle mass exceeded +0.75% / 100days; patients were designated as losing muscle (catabolic) if muscle was lost over time and exceeded -0.75% / 100days. Using this classification scheme we excluded the 20 patients whose change was between -0.75% and +0.75% / 100 days (shaded area in Figure 1). We classified 44 patients as muscle losing (Mean -4.71% / 100 d; SD = 5.13) and 29 patients as not losing muscle (Mean +3.91%/ 100 d; SD = 2.33). These two groups of patients with known muscle change status (loss or gain) were used to build predictive models using urinary metabolites.

6.3.2 Metabolites detected and used in statistical approaches

We assigned and quantified 71 metabolites in each sample. Creatinine concentrations assessed by NMR were within 95% (95% confidence interval of 91% - 97%) of the values confirmed by laboratory tests. Spike-in experiments provided positive confirmation of peak assignments for Ala, Asn, Gln, Gly, His, Ile, Leu, Lys, Phe, Ser, Tau, Thr, Trp, Tyr, Val, 1-Methylhistidine and 3-Methylhistidine. We excluded drug metabolites or drug vehicle constituents (ibuprofen, acetaminophen, salicylurate, propionate, propylene glycol, mannitol) from statistical analyses. Methanol (a microbial (non human metabolite) was excluded as unlikely to be related to muscle loss. Urea was excluded since suppression of the NMR signal by pre-saturation may lead to resonant suppression

of the urea peak due to proton exchange with water, thereby making its quantification unreliable (19). The remaining 63 metabolites were used in subsequent analyses.

Median concentration and concentration ranges for these 63 metabolites are shown (Table 6-2). Numerous individual metabolites, as well as the total concentration (i.e. the sum of all 63 metabolites), were increased in the patients losing muscle (Mann-Whitney test). Levels of creatinine were higher in patients with muscle loss ($P<0.001$).

None of the methods of data normalization (by creatinine concentration, by total peak area, probability quotient normalization) proved helpful and all three methods *reduced* the predictive accuracy of the classifiers, compared with no data normalization. The use of a log transformation was ultimately found to be the only preprocessing step for the metabolite concentrations that improved the predictors' performance compared with raw concentration values.

6.3.3 A classifier for muscle loss based on urinary metabolites

Of all tested algorithms (Table 6-3), SVM was the most accurate classifier, and predicted muscle loss status with a (5-fold cross validation) accuracy of 82.2% (σ = 7.45%). Although PIA produced a classifier with the same accuracy, we focus on the SVM model because it is more familiar. Figure 6-1 also identifies the patients who were misclassified by the SVM classifier. While 6 misclassified patients had muscle losses/gains of less than 2%, there were 7 misclassified patients with losses/gains up to 5%.

In 1,000 permutation tests (32) on the SVM model, which produced an average accuracy of 53.8%, no permutation performed better than the SVM classifier. This non-parametric test suggests that chance alone would not have produced this SVM result – i.e., the probability that a classifier (here SVM) could produce a score of 82.2%, if there was no real pattern in the data, is $P < 0.001$.

6.3.4 Urine metabolites related to muscle loss

Mutual information was used to quantify dependence between each of the 63 metabolites the class outcome (losing muscle vs. not losing muscle), Table 6-4. Larger values indicate a higher degree of dependence between the metabolite and the class outcome.

6.4 Discussion

Cancer-associated muscle wasting is associated with reduction in functional status, in response to treatment and in life expectancy. Methods currently used to assess muscle loss involve diagnostic imaging techniques such as computed tomography (CT), which are costly, inconvenient, invasive, time consuming and have limited ability to detect early or slowly evolving wasting. We present a novel approach using single time-point urinary metabolite profiles to determine whether a patient is experiencing muscle wasting; ^{1}H-NMR analysis of a single random urine sample may be a fast, cheap, safe and inexpensive tool to screen and monitor muscle loss, and that useful classifiers for predicting related metabolic conditions are possible with the methodology presented.

6.4.1 Building an accurate classifier using metabolomic data

Here we present the first steps towards diagnostic markers of cancer-associated skeletal muscle wasting using metabolic profiling. Human populations are variable with respect to age, gender, ethnicity, diet, drug intake, and health status and some, but not all, of these features can be controlled in research studies. Against this background, we tried to determine whether urine could be used to diagnose patients with skeletal muscle wasting, an important physiological component of the negative nitrogen balance characteristic of wasting syndromes. Muscle wasting may go unnoticed in its early stages, if progressing slowly, or if it is obscured by changes in other tissues. This is why a robust classifier that can achieve an accuracy of 82%, using metabolites in a single randomly collected urine sample could be considered a significant advance compared with a standard approach that requires several months and the acquisition of at least 2 diagnostic images.

In efforts to find the most accurate classifier we tested 8 different classifiers. Our validated results (cross validation and permutation testing) support the claim that SVM found a meaningful pattern within the data. SVM performance was superior to PLS-DA, a method often used in metabolomic studies (34). PLS-DA reduces the dimensionality of the data in a way that increases the separation with respect to the topic of interest (e.g., a disease state, or other variable under study). While PLS-DA can overfit (34), our results show that PLS-DA performed competitively with the top predictors.

We can envisage an even better predictor and more precise diagnostic test based on an extended metabolite profile. In NMR spectra, it is typically possible to detect only those compounds with concentrations >1 μM. Analysis of blood plasma or more sensitive or comprehensive metabolomic methods (i.e. MS-based methods) may reveal additional metabolites related to muscle loss. Different analytic approaches could also permit detection of compounds involved in lipid metabolism to shed light on fat loss and gain. Furthermore, serial urine sampling and CT image analysis over time would take advantage of repeated measures within individuals and would enable the ability to define biochemical changes early and late in disease and pathways implicated in disease pathogenesis. Finally, it will be important to account for some of the presently unexplained sources of variation, which may limit the precision of this diagnostic test. It is not obvious why certain patients were misclassified by our classifier (Figure 6-1); it could be due to some undetected underlying condition (i.e. kidney dysfunction) or to inherent limitations of the data, given that the minimum interval over which gain or loss can be detected by CT (months) was represented by a single point in time sample.

6.4.2 Metabolomic signature of muscle loss

Different metabolites appear in urine in function of processes of passive diffusion, active transport and reuptake and are not a representative sample of all of intermediary metabolism. Owing to the specific nature of urine, metabolites associated with amino acid metabolism, urea cycle, intermediary metabolism (glycolysis, TCA cycle, 1-carbon metabolism) and creatine metabolism were

prominent. Nevertheless, that certain urinary metabolites are related to muscle loss (Table 6-4) does suggest some of the underlying biology of muscle wasting. Several of the metabolites are constituents, breakdown products or metabolites formed in muscle. Creatinine is a degradation product of creatine, a phosphorylated molecule specific to muscle energy metabolism; both of these compounds were related to muscle loss. Creatinine is known to be raised when muscle is broken down (35). This likely explains why creatinine normalization during the prepossessing step did not perform well; creatinine normalization would only work if the assumption that creatinine concentration is only related to urine dilution. Muscle proteins contain a higher proportion of branched chain amino acids compared with proteins in other tissues, and muscle is the predominant site of their catabolism. Thus the association of valine, leucine and of a decarboxylation product of leucine, 3-OH-isovalerate, with muscle loss is not surprising. This is not exclusive to the branched chain amino acids; during muscle protein breakdown, all of its constituent amino acids enter oxidative pathways. Increased levels of several metabolites in urine are possibly indicative of increased flux of amino acids (Leu, Ile, Val, Ala, Thr, Tyr, Gln, Ser), and of amino acid carbon through intermediary metabolism (succinate, trans-aconitate) and 1-carbon metabolism (betaine, trigonelline). Finally, in relation to the suggestion that insulin resistance may be a prominent feature of cancer-associated muscle wasting (36, 37), urinary glucose was elevated in patients with muscle loss (397 μM) compared with patients who were not losing muscle (93 μM) ($p<0.001$). Elevated blood and urine glucose levels are associated with insulin-resistant

states, and while the urine glucose levels in patients with muscle wasting were below levels typical of diabetes, this may be an early sign of insulin resistance.

6.5 Conclusion

Our work is the first attempt to use metabolomics to diagnose muscle wasting occurring as a result of cancer cachexia in humans. We developed a single time-point urine test using concentrations of 63 urinary metabolites to diagnose muscle wasting. This minimally invasive test is rapid, robust, quite accurate (82.2%), and able to detect a small but physiologically relevant rate of muscle loss (outside of ±0.75% loss /100 days). Metabolites related to muscle wasting include a variety of compounds likely to originate from catabolism of this tissue and may also shed some light on the underlying metabolic aberrations that lead to muscle loss.

Tables

Table 6-1: Characteristics of study participants

	Patients with muscle gain >0.75%/100 d	**Patients with muscle loss >0.75%/100 d**	**All patients**
Age (mean ± SD)	64 ± 11	62 ± 10	63 ± 10
Gender			
% Male	48	66	59
Cancer Type			
% Lung	28	23	25
% Colorectal	72	77	75
Staging % (N)			
Stage 1	NA	NA	NA
Stage 2	NA	2 (1)	1 (1)
Stage 3	31 (9)	25 (11)	27 (20)
Stage 4	69 (20)	73 (32)	71 (52)

Table 6-2: Median concentration and concentration ranges of 63 urine metabolites included in the statistical analyses

Metabolite	^{1}H chemical shift (ppm) and coupling[1]	Median concentration (range) (µM)		
		Patients with muscle gain >0.75%/100 d	**Patients with muscle loss <-0.75%/100 d**	**All patients**
1,6-Anhydro-β-D-glucose	3.53(m), 3.67(m), 3.69(m), 3.75(dd), 4.09(dd), 4.62(dd), 5.45(m)	34.7 (9.4 - 191.9)	98.5 (4.7 - 687.3) **	47.5 (4.7 - 942.6)
1-Methylnicotinamide	4.47(s), 8.17(t), 8.88(d), 8.95(d), 9.26(s)	18.2 (6.4 - 1036)	51.7 (6.9 - 474.1) **	36.4 (4.6 - 1036)
2-Aminobutyrate	0.97(t), 1.89(m), 3.71(dd)	7.4 (1.3 - 28.9)	15.3 (2.1 - 173) **	10.5 (1.3 - 173)
2-Hydroxyisobutyrate	1.35(s)	18.8 (4.9 - 66.2)	42.1 (7.8 - 85.4) ***	33.7 (4.5 - 193.4)
2-Oxoglutarate	2.44(t), 3.00(t)	25.3 (5.6 - 987.2)	69.3 (5.5 - 2467) *	63 (5.5 - 2467)
3-Aminoisobutyrate	1.19(d), 2.60(m), 3.03(dd), 3.1(dd)	21.3 (3.1 - 209.2)	30.7 (2.6 - 1481)	29.8 (2.6 - 1481)
3-Hydroxybutyrate	1.19(d), 2.29(dd), 2.40(dd), 1.14(m)	6.5 (2.2 - 34.3)	25.6 (1.7 - 176.6) ***	11.8 (1.7 - 176.6)
3-Hydroxyisovalerate	1.26(s),2.36(s)	5.3 (0.9 - 57.6)	21.3 (2.5 - 164.4) ***	13 (0.9 - 359.7)
3-Indoxylsulfate	7.18(m), 7.26(m), 7.36(s), 7.49(d), 7.70(d), 10.10(s)	104.9 (27.8 - 613.8)	202.4 (34.9 - 1038) ***	165.6 (27.8 - 1038)
4-Hydroxyphenylacetate	3.44(s), 6.85(m), 7.16(m)	48.2 (15.5 - 799.6)	93 (17.6 - 430.9) **	70.1 (15.5 - 799.6)
Acetate	1.91(s)	16.2 (3.5 - 202.5)	71.5 (9.9 - 410.6) ***	34.9 (3.3 - 410.6)
Acetone	2.22(s)	6.8 (2.3 - 23.8)	8.2 (2.3 - 206.5)	7.6 (2.1 - 206.5)
Adipate	1.54(m), 2.19(m)	6.2 (1.6 - 19.2)	16.1 (3.1 - 325.6) ***	11.1 (1.6 - 325.6)
Alanine	1.47(d), 3.78(qt)	78.6 (16.8 - 601)	320.1 (26.8 - 1314) ***	195.7 (13.9 - 1447)
Asparagine	2.86(dd), 2.95(dd), 3.99(dd), 6.91(s), 7.62(s)	29.2 (6.7 - 152.6)	64.4 (8 - 272.9) ***	42.3 (6.7 - 272.9)
Betaine	3.26(s), 3.89(s)	27.3 (2.3 - 312.4)	112.5 (4.1 - 391.7) ***	63.8 (2.3 - 788.8)
Carnitine	2.41(dd), 2.45(dd), 3.22(s), 3.40(m), 3.43(m), 4.56(m)	19 (2.7 - 206.5)	31.7 (4.5 - 488.1) **	24.6 (2.7 - 488.1)

Citrate	2.53(d), 2.69(d)	1014 (59.6 - 4214)	2336 (80.9 - 13636) **	1790 (59.6 - 13636)
Creatine	3.02(s), 3.92(s)	18.6 (2.8 - 393.6)	87.4 (7.9 - 1863) ***	48.8 (1.1 - 1862.7)
Creatinine	3.03(s), 4.05(s)	3616 (1003 - 15116)	10003 (1256 – 33944) ***	8032 (868.2 - 33944)
Dimethylamine	2.72(s)	148.8 (41.3 - 496.2)	370.6 (46.9 - 1562) ***	306.2 (27.4 - 1562)
Ethanolamine	3.14(m), 3.82(m)	113.9 (21.5 - 907.8)	270.7 (16.1 - 1439) **	212.4 (16.1 - 1439)
Formate	8.45(s)	61.4 (6.4 - 294.4)	136.1 (27.7 - 1476) **	91.7 (6.4 - 1476)
Fucose	1.20(d), 1.24(d), 3.44(dd), 3.56(t), 3.63(dd), 3.74(dd), 3.76(dd), 3.79(m), 3.80(m), 3.85(dd), 3.86(m), 3.87(m), 3.97(dd), 4.01(m), 4.03(m), 4.07(dd), 4.18(m), 4.55(d), 5.20(d), 5.22(d), 5.27(d)	37.7 (5.7 - 196.2)	90.2 (13.6 - 408.4) ***	68.4 (5.7 - 408.4)
Fumarate	6.51	3.1 (0.8 - 36.2)	6.6 (1.1 - 96.6) ***	4.2 (0.8 - 96.6)
Glucose	3.24(dd), 3.40(m), 3.41(m), 3.47(m), 3.49(m), 3.53(dd), 3.71(m), 3.72(dd), 3.74(m), 3.82(m), 3.84(m), 3.90(dd), 4.64(d), 5.23(d)	92.9 (26.9 - 337.5)	397 (43.9 - 8724.8) ***	190 (26.9 - 8724.8)
Glutamine	2.12(m), 2.15(m), 2.43(m), 2.47(m), 3.76(t), 6.87(s), 7.58(s)	112.9 (23.3 - 862.1)	401.3 (26.8 - 1684) ***	226.4 (15.1 - 1684)
Glycine	3.56(s)	382.2 (38.3 - 2281)	690.2 (52.6 - 5073) **	560 (38.3 - 18195)
Glycolate	3.95(s)	66 (5.4 - 439.9)	179.8 (10.9 - 682.8) **	126.3 (5.4 - 885.5)
Guanidoacetate	3.79(s)	45.6 (7 - 301.1)	96.1 (18.2 - 563.5) **	72.1 (4.6 - 563.5)
Hippurate	3.96(d), 7.54(m), 7.55(m), 7.63(t), 7.82(m), 7.83(m), 8.52(s)	574.8 (122.7 - 6667)	21816 (93.1 - 19263) ***	1274 (93.1 - 19263)

Histidine	3.15(dd), 3.25(dd), 3.99(qt), 7.11(s), 7.92(s)	78.4 (16.3 - 616.1)	326 (14.1 - 1869) ***	182.8 (14 - 1868.8)
Hypoxanthine	8.18(s), 8.20(s)	31.8 (3.8 - 161.7)	44.8 (4.2 - 265.3)	42.6 (3.7 - 265.3)
Isoleucine	0.93(t), 1.00(d), 1.25(m), 1.46(m), 1.97(m), 3.66(d)	4.2 (1.8 - 18)	8.4 (2 - 40.2)*	7.7 (1.8 - 117.5)
Lactate	1.32(d), 4.11(d)	39.4 (7.3 - 199.3)	110.4 (17.5 - 3659) ***	87.3 (7.3 - 3659)
Leucine	0.95(d), 0.96(d), 1.67(m), 1.70(m), 1.73(m), 3.73(m)	9 (2.5 - 31.4)	24.3 (3.5 - 103.8) ***	19.1 (2.5 - 103.8)
Lysine	1.43(m), 1.50(m), 1.72(m), 1.88(m), 1.91(m), 3.02(t), 3.75(t)	34.9 (10.5 - 787.5)	106.7 (15.2 - 464.6) ***	75.6 (4.3 - 787.5)
Methylamine	2.6(s)	5.1 (1.5 - 44.6)	20.4 (1.8 - 52.3) ***	15.5 (1.5 - 108.8)
Methylguanidine	2.83(s)	6.8 (1.7 - 36.6)	10.4 (2.1 - 141.6)	8.8 (1.7 - 141.6)
N,N-Dimethylglycine	2.92(s), 3.72(s)	9.2 (1.2 - 52.5)	30.4 (3.4 - 119.9) ***	21.3 (1.2 - 169.5)
O-Acetylcarnitine	2.14(s), 2.50(dd), 2.63(dd), 3.19(s), 3.60(dd), 3.84(dd), 5.59(m)	6.1 (1.2 - 43.9)	14.2 (1.6 - 254.6) **	11.6 (1.2 - 254.6)
Pantothenate	0.89(s), 0.92(s), 2.41(t), 3.39(d), 3.43(qt), 3.44(qt), 3.51(d), 3.98(s), 8.00(dd)	14.4 (3.1 - 691.4)	26.3 (2.6 - 187.5) *	22.6 (1.7 - 691.4)
Pyroglutamate	2.03(m), 2.38(m), 2.41(m), 2.50(m), 4.17(dd)	82.7 (21.4 - 442.1)	251.7 (37.6 - 1066) ***	155.5 (18 - 1066)
Pyruvate	2.37(s)	6.5 (0.9 - 66.6)	21.4 (1.8 - 184.8) ***	15.4 (0.9 - 184.8)
Quinolinate	7.46(dd), 8.00(d), 8.45(d)	26.7 (5.2 - 163.6)	76 (16.2 - 260.7) ***	51 (5.2 - 260.7)
Serine	3.84(dd), 3.94(dd), 3.98(dd)	90.5 (16.2 - 269.8)	218.3 (32.6 - 1245) ***	136.8 (16.2 - 1245)
Succinate	2.4(s)	8.6 (1.7 - 221)	50.2 (6.4 - 587.8) ***	29.3 (1.2 - 587.8)
Sucrose	3.47(t), 3.55(dd), 3.66(s), 3.68(m), 3.76(m), 3.80(m), 3.82(m), 3.83(m), 3.84(m), 3.88(m), 4.04(t), 4.21(d), 5.40(d)	19.2 (6.5 - 600.6)	67.6 (10.2 - 2081) ***	41.1 (6.5 - 2081)

Tartrate	4.33(s)	10.7 (2.2 - 273.1)	16.3 (3 - 834.5)	12.8 (2.2 - 834.5)
Taurine	3.25(t), 3.41(t)	176.2 (17.9 - 1513)	407.5 (55.3 - 4285) **	280.3 (13.5 - 4284)
Threonine	1.32(d), 3.59(d), 4.26(m)	39.1 (9.1 - 250.5)	102.6 (8.2 - 448.5) ***	67.6 (4.4 - 448.5)
Trigonelline	4.43(s), 8.07(dd), 8.82(m), 8.83(m), 9.11(s)	74.6 (10.1 - 564.5)	190.2 (10.2 - 2257) **	97.2 (10.1 - 2257)
Trimethylamine N-oxide	3.26(s)	243 (55.7 - 1533)	542.8 (66.8 - 5460) **	403.2 (14.9 - 5460)
Tryptophan	3.30(dd), 3.47(dd), 4.05(q), 7.19(m), 7.27(m), 7.31(s), 7.52(d), 7.72(d)	21.3 (10.5 - 185.4)	82.1 (9.9 - 260.3) ***	56.7 (9.9 - 260.3)
Tyrosine	3.05(dd), 3.19(dd), 3.94(q), 6.88(m), 7.20(m)	23.9 (4.2 - 180)	86.8 (14 - 537.2) ***	58.5 (4.2 - 537.2)
Uracil	5.79(d), 7.52(d)	20.2 (3.1 - 138)	29.5 (4.2 - 179.2)	28.1 (3.1 - 179.2)
Valine	0.98(d), 1.03(d), 2.26(m), 3.61(d)	13.2 (4.1 - 53.3)	39.8 (4.3 - 160.1) ***	30.8 (4.1 - 160.1)
Xylose	3.22(dd), 3.31(dd), 3.43(t), 3.52(dd), 3.60(m), 3.62(m), 3.65(m), 3.67(m), 3.69(m), 3.92(dd), 4.57(d), 5.19(d)	32.8 (10.1 - 259.4)	71.3 (16.6 - 2158) ***	51.2 (9.8 - 2155)
cis-Aconitate	3.12(d), 5.75(m)	54.1 (12.9 - 298.1)	235.1 (15.1 - 1862) ***	128.7 (12.9 - 1862)
myo-Inositol	3.27(t), 3.53(dd), 3.62(dd), 4.05(m)	30.5 (11.6 - 315.5)	131.9 (22 - 850.4) ***	78.2 (8.1 - 850.4)
trans-Aconitate	3.44(s), 3.59(s)	13.6 (4.9 - 181.2)	45.3 (7.9 - 216.3) ***	26.9 (4.9 - 639.3)
1-Methylhistidine	3.21(dd), 3.29(dd), 3.74(s), 3.95(dd), 7.13(s), 8.10(s)	73 (11.4 - 1186)	245 (16.6 - 2694) **	199.8 (11.4 - 2694)
3-Methylhistidine	3.08(dd), 3.16(dd), 3.70(s), 3.96(dd), 7.03(s), 7.70(s)	29.7 (8.6 - 184.6)	82.8 (8 - 317) ***	71.4 (8 - 317)
Total metabolites		29.2 (0.8 - 15116)	82.8 (1.1 – 33944) ***	56.7 (0.8 – 33944)

* ($P<0.05$), **($P<0.01$), ***($P\leq0.001$) P-values were obtained using Mann-Whitney nonparametric statistical analysis comparing patients with versus without muscle loss.

[1]s=singlet, d=doublet, dd=doublet-doublet, t=triplet, q=quartet, m=multiplet

Table 6-3: Predictive performance for muscle loss of 8 machine learning approaches, averaged over the 5 folds of cross-validation

Classifier	Mean Accuracy	Standard Deviation
Support Vector Machines	82.2 %	7.5
Pathway Informed Analysis	82.2 %	7.2
Multi-Tree Augmented Naïve Bayes	76.8 %	8.0
PLS-DA	76.7 %	14.8
Tree Augmented Naïve Bayes	76.7 %	9.6
Full dependence model	75.1 %	6.8
Naïve Bayes model	75.1 %	9.7
J48 Decision tree	65.8 %	9.1
Random permutation test (SVM)	53.8%	6.5

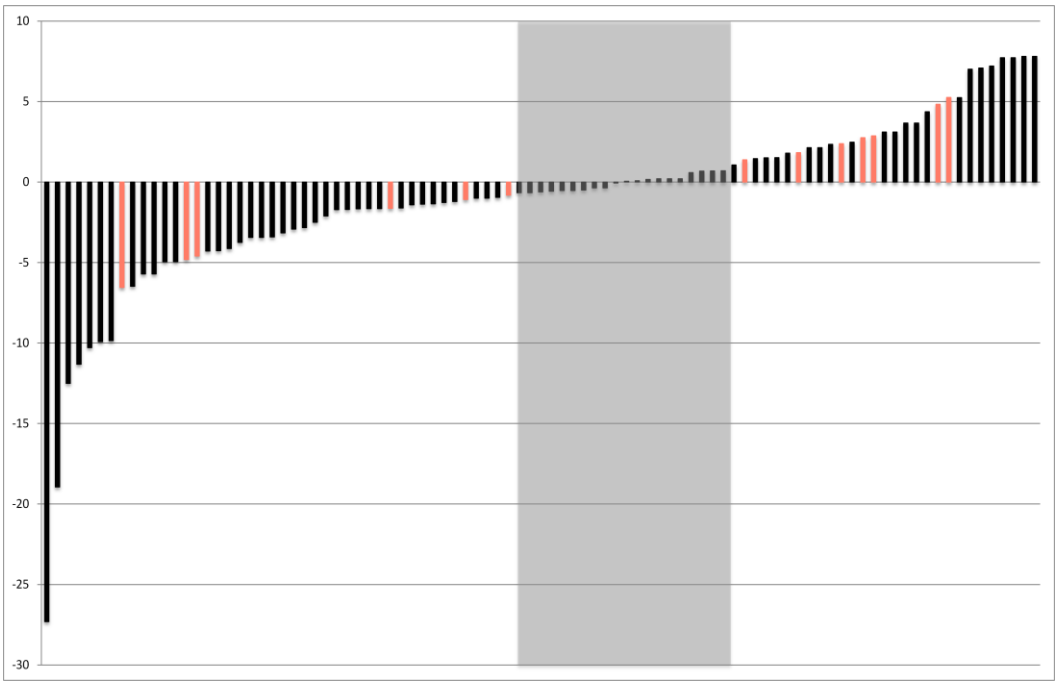
Table 6-4: Bivariate analysis: top 30 urine metabolites related to skeletal muscle loss

Metabolite	Mutual Information
Adipate	0.213524065
Glucose	0.203986953
Quinolinate	0.196946368
myo-Inositol	0.17310194
Valine	0.165888472
Succinate	0.164968427
Betaine	0.16477065
Leucine	0.16401522
N,N-Dimethylglycine	0.157109136
3-Hydroxyisovalerate	0.15366801
Creatine	0.147918127
Acetate	0.14713014
Alanine	0.145640343
Pyroglutamate	0.142823206
3-Hydroxybutyrate	0.142709618
Glutamine	0.141498304
cis-Aconitate	0.134050157
Methylamine	0.130789947
Tryptophan	0.130351855
Dimethylamine	0.126891503
Xylose	0.125333817
Creatinine	0.125318846
Formate	0.123810009
Tyrosine	0.116765473
trans-Aconitate	0.110643466
Lactate	0.109694272
Sucrose	0.108986623
2-Hydroxyisobutyrate	0.107381302
Serine	0.106181068
Threonine	0.101972559

[1]*Mutual information* is a way to quantify dependence between two variables. We computed the mutual information between each of the 63 metabolites the class outcome, C. Here we have a binary outcome variable (losing muscle vs. not losing muscle) and a continuous metabolite concentration variable that we assume follows a Gaussian distribution. Mutual Information computed as described under statistical methods and yields unit-less values, larger values indicate a higher degree of dependence.

Figures

Figure 6-1: Percentage muscle change continuum in cancer patients as determined by computed tomography image analysis



The boxed region highlights the patients excluded from analysis. Light colored columns indicate those samples that were misclassified by the SVM during cross-validation.

Using serial computed tomography images, patients' muscle change (loss or gain) was computed. The boxed region indicates patients whose muscle change fell within a minimal margin of ± 0.75% / 100 d. These patients were not included in the analysis as their calculated muscle change is within the precision error of the imaging method. The remaining patients were classified as losing or not losing muscle. These two groups represent the distal ends of the muscle change continuum and statistically different from each other ($P<0.001$).

References

1. Prado CM, Lieffers JR, McCargar LJ, Reiman T, Sawyer MB, Martin L, et al. Prevalence and clinical implications of sarcopenic obesity in patients with solid tumours of the respiratory and gastrointestinal tracts: a population-based study. The lancet oncology. 2008;9:629-35.

2. Tan BH, Fearon KC. Cachexia: prevalence and impact in medicine. Current opinion in clinical nutrition and metabolic care. 2008;11:400-7.

3. Antoun S, Baracos VE, Birdsell L, Escudier B, Sawyer MB. Low body mass index and sarcopenia associated with dose-limiting toxicity of sorafenib in patients with renal cell carcinoma. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO. 2010;21:1594-8.

4. Prado CM, Baracos VE, McCargar LJ, Mourtzakis M, Mulder KE, Reiman T, et al. Body composition as an independent determinant of 5-fluorouracil-based chemotherapy toxicity. Clinical cancer research : an official journal of the American Association for Cancer Research. 2007;13:3264-8.

5. Prado CM, Baracos VE, McCargar LJ, Reiman T, Mourtzakis M, Tonkin K, et al. Sarcopenia as a determinant of chemotherapy toxicity and time to tumor progression in metastatic breast cancer patients receiving capecitabine treatment. Clinical cancer research : an official journal of the American Association for Cancer Research. 2009;15:2920-6.

6. Evans WJ, Morley JE, Argiles J, Bales C, Baracos V, Guttridge D, et al. Cachexia: a new definition. Clin Nutr. 2008;27:793-9.

7. Heymsfield SB, Wang Z, Baumgartner RN, Ross R. Human body composition: advances in models and methods. Annu Rev Nutr. 1997;17:527-58.

8. Mitsiopoulos N, Baumgartner RN, Heymsfield SB, Lyons W, Gallagher D, Ross R. Cadaver validation of skeletal muscle measurement by magnetic resonance imaging and computerized tomography. J Appl Physiol. 1998;85:115-22.

9. Shen W, Punyanitya M, Wang Z, Gallagher D, St-Onge MP, Albu J, et al. Total body skeletal muscle and adipose tissue volumes: estimation from a single abdominal cross-sectional image. J Appl Physiol. 2004;97:2333-8.

10. Mourtzakis M, Prado CM, Lieffers JR, Reiman T, McCargar LJ, Baracos VE. A practical and precise approach to quantification of body composition in cancer patients using computed tomography images acquired during routine care. Applied physiology, nutrition, and metabolism = Physiologie appliquee, nutrition et metabolisme. 2008;33:997-1006.

11. Slupsky CM, Rankin KN, Wagner J, Fu H, Chang D, Weljie AM, et al. Investigations of the effects of gender, diurnal variation, and age in human urinary metabolomic profiles. Anal Chem. 2007;79:6995-7004.

12. Wishart DS. Quantitative metabolomics using NMR. Trends Analyt Chem. 2008;27:228-37.

13. Bertini I, Calabro A, De Carli V, Luchinat C, Nepi S, Porfirio B, et al. The metabonomic signature of celiac disease. J Proteome Res. 2009;8:170-7.

14. Mahadevan S, Shah SL, Marrie TJ, Slupsky CM. Analysis of metabolomic data using support vector machines. Anal Chem. 2008;80:7562-70.

15. Ross R. Advances in the application of imaging methods in applied and clinical physiology. Acta diabetologica. 2003;40 Suppl 1:S45-50.

16. Shen W, Punyanitya M, Wang Z, Gallagher D, St-Onge MP, Albu J, et al. Visceral adipose tissue: relations between single-slice areas and total volume. Am J Clin Nutr. 2004;80:271-8.

17. Lieffers JR, Mourtzakis M, Hall KD, McCargar LJ, Prado CM, Baracos VE. A viscerally driven cachexia syndrome in patients with advanced colorectal cancer: contributions of organ and tumor mass to whole-body energy demands. The American journal of clinical nutrition. 2009;89:1173-9.

18. Weljie AM, Newton J, Mercier P, Carlson E, Slupsky CM. Targeted profiling: quantitative analysis of 1H NMR metabolomics data. Anal Chem. 2006;78:4430-42.

19. Saude EJ, Sykes BD. Urine stability for metabolomic studies: effects of preparation and storage. Metabolomics. 2007;3:19-27.

20. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, et al. HMDB: the Human Metabolome Database. Nucleic acids research. 2007;35:D521-6.

21. Bidlingmeyer BA, Cohen SA, Tarvin TL. Rapid analysis of amino acids using pre-column derivatization. J Chromatogr. 1984;336:93-104.

22. Holmes E, Foxall PJ, Nicholson JK, Neild GH, Brown SM, Beddell CR, et al. Automatic data reduction and pattern recognition methods for analysis of 1H nuclear magnetic resonance spectra of human urine from normal and pathological states. Anal Biochem. 1994;220:284-96.

23. Bollard ME, Stanley EG, Lindon JC, Nicholson JK, Holmes E. NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition. NMR Biomed. 2005;18:143-62.

24. Craig A, Cloarec O, Holmes E, Nicholson JK, Lindon JC. Scaling and normalization effects in NMR spectroscopic metabonomic data sets. Analytical chemistry. 2006;78:2262-7.

25. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. Analytical chemistry. 2006;78:4281-90.

26. Witten IH, Frank E. Data mining : practical machine learning tools and techniques. 2nd ed. Amsterdam ; Boston, MA: Morgan Kaufman; 2005.

27. Friedman N, Geiger, D., Goldszmidt, M. Bayesian network classifiers. Machine learning. 1997;29:131-63.

28. Quinlan JR. C4.5 : programs for machine learning. San Mateo, Calif.: Morgan Kaufmann Publishers; 1993.

29. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, et al. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 2008;36:D480-4.

30. Eastman T. A disease classifier for metaboli profiles based on metabolic pathway knowledge. Edmonton, Alberta: University of Alberta; 2010.

31. Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning : data mining, inference, and prediction : with 200 full-color illustrations. New York: Springer; 2001.

32. Pesarin F. Multivariate permutation tests : with applications in biostatistics. Chichester ; New York: J. Wiley; 2001.

33. Cover TM, Thomas JA. Elements of information theory. 2nd ed. Hoboken, N.J.: Wiley-Interscience; 2006.

34. Westerhuis JA, Hoefsloot HCJ, Smit S, Vis DJ, Smilde AK, van Velzen EJJ, et al. Assessment of PLSDA cross validation. Metabolomics. 2008;4:81-9.

35. Akcay MN, Akcay G, Solak S, Balik AA, Aylu B. The effect of growth hormone on 24-h urinary creatinine levels in burned patients. Burns. 2001;27:42-5.

36. Asp ML, Tian M, Wendel AA, Belury MA. Evidence for the contribution of insulin resistance to the development of cachexia in tumor-bearing mice. Int J Cancer. 2009.

37. Wang X, Hu Z, Hu J, Du J, Mitch WE. Insulin resistance accelerates muscle protein degradation: Activation of the ubiquitin-proteasome pathway by defects in muscle cell signaling. Endocrinology. 2006;147:4160-8.

CHAPTER 7: Final Discussion

7.1 Introduction

The purpose of this research was to apply transcriptomic and metabolomic technologies to gain a better understanding of transcriptional changes that occur in skeletal muscle, develop a non-invasive tool to detect muscle mass changes in cancer and gain insight into the metabolic alterations underlying cancer cachexia. It was clear early on in this research that various methodological considerations would have to be addressed (Figure 7-1). The work presented here extends beyond cachexia research to other fields that may use metabolomic and transcriptomic technologies. This final discussion outlines key points and concepts related to the research presented in the previous chapters and makes recommendations for future research.

7.2 The concept of classification in cancer cachexia studies

At no point in this thesis were patients classified as cachectic or non-cachectic, specifically. Cancer cachexia is a complex and multifactorial syndrome (1) and patients with cachexia may present with varying degrees of muscle and fat mass, weight loss, muscle loss, myocellular fat infiltration, dietary intake and/or resting metabolic rate (2, 3). I aimed to better understand molecular and metabolic differences in relation to these variations. This required classification of patients according to not one criterion but multiple criteria (Chapters 4 - 6). This differs from a historical approach which was limited to a single criterion, percentage

weight loss (as outlined in Table 4-1 of Chapter 4). Accumulating evidence suggests that each unit of weight lost does not have constant composition (varying proportions of lean and fat tissue) (4, 5). Thus, if a study aims to understand molecular mechanisms in relation to weight loss alone it will be unclear if molecular differences identified are related with muscle loss, fat loss, or a combination of the two. This is a point raised repeatedly throughout this thesis and is something that should be considered when planning future cachexia studies. In this thesis, multiple classification criteria were possible due to the particularly rich datasets used in Chapters 4-6. Such datasets are rare but results from this thesis suggest that future studies could benefit from considering relevant sources of variation when studying this multifactorial syndrome.

A typical method of classification involves dichotomizing a continuous variable by splitting patient populations based on a single cutoff value for which there may be no statistical justification (6). For example, using a cutoff of 5% weight loss, patients would be classified as weight losing or weight stable depending on whether they experienced weight loss greater than or less than the cutoff, respectively (5). This method forces patients that may not be different from each other (e.g. a patient with 5.1% and a patient with 4.9% weight loss) (7) into different classes; therefore decreasing the chances of identifying differences in molecular or metabolic signatures. The method used in Chapters 4-6 involved splitting patients based on excluding patients with values on either side of the cut point (as shown in Figure 6-1). The band of patients excluded should be at least as wide as the measurement error of the phenotype in question. A wider zone of

exclusion permits comparisons of patients at the extremes of phenotypes, which have a greater chance of carrying the characteristic molecular signatures associated with a particular factor. Though not often used in cancer cachexia studies, this approach is not new to medicine and has been recommended for OMIC studies (8-11). Using this classification method allowed patients to be split into distinct and clinically relevant groups. For example, when looking at skeletal muscle index (SMI) in Chapter 4, the values for SMI in the high SMI class were within the range expected for healthy men of the same age whereas the average SMI values in the low SMI class were within the range expected for sarcopenic men. Sarcopenia in cancer is clinically relevant as it is associated with poor outcomes of surgery including increased length of stay, infectious complications and increased requirement for inpatient rehabilitation (12, 13); altered clinical response to opioid analgesics (14), increased chemotherapy toxicity (2), poorer prognosis (15) and decreased survival (16-18).

It is important to note that classification made to study molecular and metabolic changes in response to phenotypic changes that occur in response to cachexia is a distinct task from classification made in the clinic for diagnostic purposes. The classification approach used in this thesis (i.e. looking at the extremes for each phenotype) is not intended for clinical diagnostic purposes. A year prior to the completion of this thesis, experts introduced a new classification system indicating severity (precachexia, cachexia, and refractory cachexia) (1); this classification system was intended to diagnose and treat cancer cachexia. The

specific diagnostic criteria of these stages were deemed to require the future acquisition and analyses of large data sets (1).

7.3 Detecting small, early changes in muscle mass

Underdiagnosis of cachexia has generally been attributed to disagreement and ambiguity surrounding its definition (19, 20). Experts recently defined cancer cachexia as "*a multifactorial syndrome characterised by an ongoing loss of skeletal muscle mass (with or without loss of fat mass) that cannot be fully reversed by conventional nutritional support and leads to progressive functional impairment*" and emphasized the importance of muscle as marker of cachexia and advocated for routine assessment of muscle mass (1). Despite this, muscle mass is rarely assessed in the clinic. Chapter 6 presents a novel, fast and non-invasive method of detecting very low rates of muscle loss (0.75% change/100d) using a randomly collected urine sample. The ability to detect low rates of muscle wasting suggests that this method has the potential to detect even early stage cachexia (precachexia). Experts have recently emphasized that "*every effort should be concentrated on the recognition of preclinical cachexia (precachexia)*"(21). Most previously conducted clinical trials have focused on treating the advanced (refractory) cachectic patient, in other words, they focused on stereotypical cachectic patients who exhibit the 'skin and bones' phenotype or lost uncharacteristic amounts of body weight (21). By definition, this is not a stage that would respond to treatments or interventions. Ideally, treatments or interventions should target the prechachexia stage as this would theoretically

prevent or delay progression to the refractory stage (21). The metabolomics-based method from Chapter 6 could potentially be used to focus on patients at a clinically relevant stage in the cancer cachexia continuum.

In Chapter 6, skeletal muscle rate of loss was measured using two consecutive computed tomography images taken as part of routine clinical care over ~100 days. Future validation studies should consider short-term fluctuations in urinary metabolite concentrations within the time frame of these two images. It is crucial that the metabolomic-based method withstand small-fluctuations in metabolite concentrations and correctly identify patients as losing or not losing muscle within that time frame. In other words, the false positive/false negative rate of this assessment must be determined. This would require obtaining urine samples periodically and assess the classification accuracy at each point in time using the metabolomic profile. Validation of the metabolomic-based method is warranted considering its potential.

7.4 Advancing our understanding of cancer cachexia mechanisms

Chapter 4, which focused on molecular changes in muscle from cancer patients, included a large number of muscle samples compared to previous studies using human muscle (see Table 4-1). This was possible because samples were obtained by conducting tissue collection during clinically scheduled surgeries which appears to be perceived as minimally invasive by patients. Though some groups have taken advantage of surgery to obtain muscle biopsies (22-25), the invasive percutaneous biopsy method appears to be most often used (26-28).

Surgery is a major part of cancer diagnosis, treatment and palliative care and presents many opportunities to obtain tissue samples; this approach should be considered in future cancer cachexia studies.

There are no prior reports looking at the relationship between muscle mass and muscle attenuation variation in relation to muscle gene expression profiles in cancer patients or animal models. Thus, it was not possible to relate findings from Chapter 4 with previously conducted studies. Interpretation was limited to other conditions for which muscle gene expression was examined. The concept of tissue cross-talk, such as immune cell – muscle cell interactions, emerged from the results in Chapter 4. This concept is a relatively new way of thinking about cancer cachexia. Myocytes interact with immune cells, fibroblasts, stem cells, adipocytes, neurons and endothelial cells (Figure 4-1). Differential gene expression in Chapter 4 revealed many genes supporting myocyte interaction with these other cells. Future studies, should appreciate the importance of cell-cell interaction. For example, in the case of *in vitro* studies it may be beneficial to study co-cultures as opposed to myocytes only.

Chapter 4 was a hypothesis-generating study which identified numerous avenues that warrant further study. Namely, the association between muscle attenuation and muscle atrophy with inflammation, degradation, mitochondrial dysfunction and fibrosis.

- Systemic inflammation, as indicated by increased plasma acute phase proteins (e.g. C-reactive protein) concentrations, is often observed in cancer patients (29). Though a single measure of inflammation may not

reflect the status of inflammation at the tissue level, it is unfortunate that this measure was not available for participants in Chapter 4. Having plasma acute phase protein concentrations would have allowed us to check for a relationship between muscle attenuation and systemic inflammation. Regardless, gene expression data in Chapter 4 did suggest immune cell infiltration at the tissue level. This could be confirmed by conducting a flow cytometry experiment by labeling immune cells. Such experiments have been conducted in rodent models of other conditions such as notexin-induced myoinjury, acute phase of *Trypanosoma cruzi* infection and regeneration (30-32) but not during cancer-associated muscle changes. Future cancer cachexia studies should focus on the role of infiltrating/local immune cells in cachexia development and progression.

- Protein degradation has been a major theme in cancer cachexia. Our work supports prior reports stating that the ubiquitin-proteasome system is a major proteolytic pathway during cancer. Perhaps the most important question regarding degradation is, what turns it on?
- Muscle energy metabolism is altered in cancer cachexia and our findings pointing to altered mitochondrial function are in line with this. Mitochondrial function has recently become a focus within cachexia research (33-37). Accumulating reactive oxygen species (38), altered insulin sensitivity (39) and inflammation (33, 40) are all associated with mitochondrial dysfunction in muscle and were suggested in low attenuation muscle based on differential expression in Chapter 4.

Mitochondrial function has yet to be studied even at a basic level in human skeletal muscle from cancer patients with cachexia. Identifying the number and shape (e.g. is there swelling) might be a good starting point to studying mitochondrial change. More sophisticated studies might include examining mitochondrial function. Experiments to measure mitochondrial function in muscle cells may include measuring mitochondrial proton current to assess the respiration rate, measuring mitochondrial membrane potential and measuring coupling efficiency in muscle cells (41). Other experiments may be conducted in isolated muscle mitochondria such as measuring mitochondrial proton current using the Clark oxygen electrode method or measuring mitochondrial respiratory control index (the ratio of respiration to phosphorylation of ADP to ATP) (41).

- The presence and significance of fibrosis in skeletal muscle in cancer cachexia is unknown. A search on Pubmed using the terms "cancer cachexia and fibrosis" yields 5 studies, all but one dealing with fibrosis in adipose tissue. Studies on other conditions that result in cachexia, such as chronic heart failure and HIV infection, do indicate that skeletal muscle fibrosis accompanies muscle atrophy (42, 43). Chapter 4 provides evidence for the presence of fibrosis-related pathways in low attenuation/low muscle mass index muscle during cancer. Examining the presence of fibrosis through histological analysis and examining the presence and activity of fibroblasts and fibroblast progenitors over the

course of the cachexia trajectory would be a great start to further our understanding of fibrosis in cachexia.

7.5 Methodological considerations

To use transcriptomic and metabolomic approaches to study cancer cachexia, I was forced to address some methodological issues. Figure 7-1 shows the general workflow of transcriptomic and metabolomic studies and identifies what issues I addressed. Methods for sample processing, sample analysis and data processing were not addressed for gene expression studies as these steps were followed according to manufacturer (Agilent) instructions. Results from Chapters 3 and 4 can be easily compared to other studies which use Agilent microarray technology. Other groups have compared microarray platforms (44), though this area is still under investigation suggesting that cross platform comparisons should be avoided. Likewise, urine collection and storage methodology was not addressed, for example, we did not study the impact of different collection containers or study the effect of sample freeze thaw cycles which may affect metabolomic profiles (54). However, the urine collection method was kept the same in both Chapter 5 and 6.

7.5.1 Sample size in gene expression studies

Unlike univariate statistical problems there are no standard sample size calculations for microarray and developing a sample size calculation method is an ongoing endeavor (45, 46). Sample sizes range drastically from one microarray study to another and may be as low as one pooled sample per group (47). Chapter

3 demonstrates how sample size can greatly affect the identification of differentially expressed genes and therefore interpretation of microarray studies. This work was done prior to the work in Chapter 4 specifically because it was unclear how sample size would affect results. Results from Chapter 3 allowed us to confidently proceed with work done in Chapter 4 and can be used by others analyzing gene expression data.

7.5.2 Data analysis and interpretation in gene expression studies

Gene expression microarray analysis offers a lot of information and as a result interpretation of results can be difficult. For example, in Chapter 4, thousands of genes were differentially expressed. For interpretation, genes were grouped into categories according to their function. Of course, most genes have more than one function and how genes are grouped may be contested. Further, many differentially expressed transcripts have unknown functions and though these may be important, they cannot be categorized and therefore remain excluded from interpretation. Transcripts of unknown function have been the focus of a decade long project which involved 32 institutes, 442 consortium members, and 1649 experiments called the Encyclopedia of DNA Elements (ENCODE) project (48). 80% of transcripts previously considered "junk" has been assigned biochemical functions including switching transcription on or off and regulating the degree of transcription (48). Incorporation of recently released ENCODE results to results from Chapter 4 may indicate what roles the differentially expressed transcripts of unknown function have in cancer cachexia.

Gene expression may not always translate to protein content or protein activity. This is an inherent limitation of any gene expression analysis. Once transcribed and then translated, post translational modifications (e.g. phosphorylation, methylation, ubiquitination) can enhance or inhibit protein function. To examine how the function of proteins encoded by differentially expressed genes from Chapter 4 may be altered it will be necessary to examine each protein independently. This caveat should always be considered when conducting explorative microarray studies.

7.5.3 Sample processing and analysis in metabolomics studies

Proton nuclear magnetic resonance (H^1-NMR) was used as the primary method of metabolite quantification in this research. Sample processing for NMR is relatively simple. For urine samples there are 3 steps: 1) a fixed amount of urine is aliquoted, 2) internal standard is added and 3) samples are brought to the same pH prior to analysis to prevent unpredictable chemical shifts. Since in metabolomics NMR is used as an analytical technique, error at any of these three steps will alter results. Quantification of metabolite concentrations is based on the intensity of NMR signals versus the frequency in reference to the internal standard. If measuring the concentration of only one metabolite in a homogenous solvent, then only the signals from that one metabolite would be present in the NMR spectrum and the NMR spectrum would be easy to read. In complex biofluids signals from different metabolites often overlap with each other making quantification difficult. To ensure samples were processed properly and spectra

were read correctly, extensive validation steps were undertaken: only metabolites identified by two different analysts were considered, amino acid concentrations were validated using high performance liquid chromatography (HPLC), and creatinine concentration was validated using commercially available kits. Such extensive validation is unique to the work presented here and provides a high-quality dataset for statistical analysis.

7.5.4 Data processing and analysis in metabolomics studies

Urine was selected as a biofluid of study in Chapter 5 and 6 because it is a waste pool for metabolites generated during protein turnover. It is also preferred because its collection is non-invasive, does not require trained personnel and is easy to do. However, urine is prone to dilution effects due to water intake, water retention and kidney function. Dilution is not usually considered an issue in plasma where metabolite concentrations are highly regulated and small changes observed in pathological situations are typically significant (49). To address dilution raw urine metabolomic data is often normalized to make the data from all samples directly comparable with each other (49-51). No normalization method was used prior to analysis in Chapters 5 and 6 since all normalization methods tested resulted in decreased prediction accuracy. This same approach was taken by other metabolomic researchers (52, 53). However, it does not mean that the issue of dilution is resolved. Additional studies would be required to resolve this issue and develop a new and reliable urine normalization method to address dilution variation. Perhaps by injecting an exogenous metabolite at a constant rate that is

not used by human tissues and is not reabsorbed by the kidneys (e.g. inulin) it would be possible to correct for dilution effects.

Metabolomics, like microarray, may be used to build a predictive classifier (e.g. to predict if new patients are losing or not losing muscle) or to identify a biomarker (e.g. elevated levels of metabolite A may indicate the presence of cachexia). These represent two different statistical problems. Chapters 5 and 6 mainly dealt with the building of predictive classifiers. Chapter 6 compared the different statistical and machine learning classification methods. Machine learning classifiers, support vector machines (SVM) and LASSO and the commonly used partial least squares discriminant analysis (PLS-DA) were the best performing classifiers out of the ones tested. Chapter 5 employed these three classification methods to explore potential sources of variation in urine and plasma. Body composition gave a metabolomic signature in both biofluids. This is a particularly important finding as it suggests that future metabolomic studies using urine, and to a lesser extent plasma, should consider body composition when selecting patients and making interpretations. For example, if comparing urine from early stage cancer patients with urine from late stage cancer patients in order to identify markers of cancer stage, it may be beneficial to assess muscle mass since it may be a confounding factor.

7.5.5 Considerations for future transcriptomic and metabolomic studies

The crux of OMIC technologies is the quantity of data produced. At present there are no standard methods of statistically dealing with either gene

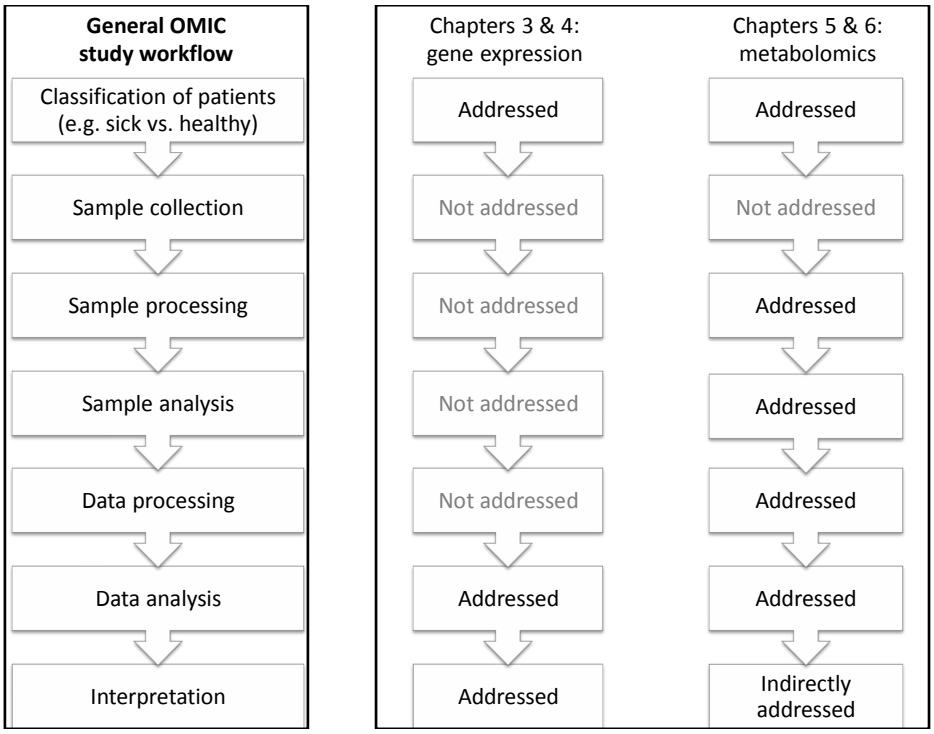
expression or metabolomic data. In fact, a relatively new field of research, the field of bioinformatics, emerged specifically to address such issues. Methodological results from this work contribute to this field. Future transcriptomic studies may use the results from Chapter 3 to make decisions about sample size. Future metabolomic studies may draw from our work from Chapters 5 and 6 regarding potential sources of variation in urine and plasma, data normalization and available predictive classifiers.

7.5 Conclusions

This research emphasizes the need to address methodological issues in order to use novel techniques in a way that yields reliable and worthwhile results. Throughout this work every step of the gene expression and metabolomic study process was scrutinized. The focus on methodological considerations should not be considered a distraction from cancer cachexia research; addressing these issues allowed the author to use these methods with a greater level of confidence. Gene expression profiling studies identified numerous avenues for future research and the metabolomics studies indicated that the cancer patient urine metabolome could be used to predict low muscle mass and muscle mass change.

Figures

Figure 7-1: Methodological issues associated with steps in gene expression profiling and metabolomics workflow addressed (or not addressed) in this thesis



References

1. Fearon K, Strasser F, Anker SD, Bosaeus I, Bruera E, Fainsinger RL, et al. Definition and classification of cancer cachexia: an international consensus. The lancet oncology. 2011;12:489-95.

2. Prado CM, Lieffers JR, McCargar LJ, Reiman T, Sawyer MB, Martin L, et al. Prevalence and clinical implications of sarcopenic obesity in patients with solid tumours of the respiratory and gastrointestinal tracts: a population-based study. The lancet oncology. 2008;9:629-35.

3. Bye A, Jordhoy MS, Skjegstad G, Ledsaak O, Iversen PO, Hjermstad MJ. Symptoms in advanced pancreatic cancer are of importance for energy intake. Supportive care in cancer : official journal of the Multinational Association of Supportive Care in Cancer. 2012.

4. Baracos VE, Reiman T, Mourtzakis M, Gioulbasanis I, Antoun S. Body composition in patients with non-small cell lung cancer: a contemporary view of cancer cachexia with the use of computed tomography image analysis. The American journal of clinical nutrition. 2010;91:1133S-7S.

5. Blum D, Omlin A, Baracos VE, Solheim TS, Tan BH, Stone P, et al. Cancer cachexia: a systematic literature review of items and domains associated with involuntary weight loss in cancer. Critical reviews in oncology/hematology. 2011;80:114-44.

6. Blum D, Strasser F. Cachexia assessment tools. Current opinion in supportive and palliative care. 2011;5:350-5.

7. MacDonald N. Terminology in cancer cachexia: importance and status. Current opinion in clinical nutrition and metabolic care. 2012;15:220-5.

8. Perez-Gracia JL, Gloria Ruiz-Ilundain M, Garcia-Ribas I, Maria Carrasco E. The role of extreme phenotype selection studies in the identification of clinically relevant genotypes in cancer research. Cancer. 2002;95:1605-10.

9. Nebert DW. Extreme discordant phenotype methodology: an intuitive approach to clinical pharmacogenetics. European journal of pharmacology. 2000;410:107-20.

10. Perez-Gracia JL, Gurpide A, Ruiz-Ilundain MG, Alfaro Alegria C, Colomer R, Garcia-Foncillas J, et al. Selection of extreme phenotypes: the role of clinical observation in translational research. Clinical & translational oncology : official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico. 2010;12:174-80.

11. Li D, Lewinger JP, Gauderman WJ, Murcray CE, Conti D. Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies. Genetic epidemiology. 2011;35:790-9.

12. Lieffers JR, Bathe OF, Fassbender K, Winget M, Baracos VE. Sarcopenia is associated with postoperative infection and delayed recovery from colorectal cancer resection surgery. British journal of cancer. 2012;107:931-6.

13. Pausch T, Hartwig W, Hinz U, Swolana T, Bundy BD, Hackert T, et al. Cachexia but not obesity worsens the postoperative outcome after pancreatoduodenectomy in pancreatic cancer. Surgery. 2012;152:S81-8.

14. Naito T, Tashiro M, Yamamoto K, Ohnishi K, Kagawa Y, Kawakami J. Impact of cachexia on pharmacokinetic disposition of and clinical responses to oxycodone in cancer patients. European journal of clinical pharmacology. 2012;68:1411-8.

15. Tandon P, Ney M, Irwin I, Ma MM, Gramlich L, Bain VG, et al. Severe muscle depletion in patients on the liver transplant wait list - its prevalence and independent prognostic value. Liver transplantation : official publication of the American Association for the Study of Liver Diseases and the International Liver Transplantation Society. 2012.

16. Parsons HA, Baracos VE, Dhillon N, Hong DS, Kurzrock R. Body composition, symptoms, and survival in advanced cancer patients referred to a phase I service. PloS one. 2012;7:e29330.

17. Sebastiano KM, Yang L, Zbuk K, Wong RK, Chow T, Koff D, et al. Accelerated muscle and adipose tissue loss may predict survival in pancreatic cancer patients: the relationship with diabetes and anaemia. The British journal of nutrition. 2012:1-11.

18. Utech AE, Tadros EM, Hayes TG, Garcia JM. Predicting survival in cancer patients: the role of cachexia and hormonal, nutritional and inflammatory markers. Journal of cachexia, sarcopenia and muscle. 2012.

19. Argiles JM, Anker SD, Evans WJ, Morley JE, Fearon KC, Strasser F, et al. Consensus on cachexia definitions. Journal of the American Medical Directors Association. 2010;11:229-30.

20. Fox KM, Brooks JM, Gandra SR, Markus R, Chiou CF. Estimation of Cachexia among Cancer Patients Based on Four Definitions. Journal of oncology. 2009;2009:693458.

21. Lucia S, Esposito M, Rossi Fanelli F, Muscaritoli M. Cancer Cachexia: From Molecular Mechanisms to Patient's Care. Critical reviews in oncogenesis. 2012;17:315-21.

22. Stephens NA, Gallagher IJ, Rooyackers O, Skipworth RJ, Tan BH, Marstrand T, et al. Using transcriptomics to identify and validate novel biomarkers of human skeletal muscle cancer cachexia. Genome medicine. 2010;2:1.

23. Stephens NA, Skipworth RJ, Macdonald AJ, Greig CA, Ross JA, Fearon KC. Intramyocellular lipid droplets increase with progression of cachexia in cancer patients. Journal of cachexia, sarcopenia and muscle. 2011;2:111-7.

24. Smith IJ, Aversa Z, Hasselgren PO, Pacelli F, Rosa F, Doglietto GB, et al. Calpain activity is increased in skeletal muscle from gastric cancer patients with no or minimal weight loss. Muscle & nerve. 2011;43:410-4.

25. Sun YS, Ye ZY, Qian ZY, Xu XD, Hu JF. Expression of TRAF6 and ubiquitin mrna in skeletal muscle of gastric cancer patients. Journal of experimental & clinical cancer research : CR. 2012;31:81.

26. Aversa Z, Bonetto A, Penna F, Costelli P, Di Rienzo G, Lacitignola A, et al. Changes in myostatin signaling in non-weight-losing cancer patients. Annals of surgical oncology. 2012;19:1350-6.

27. Phillips BE, Smith K, Liptrot S, Atherton PJ, Varadhan K, Rennie MJ, et al. Effect of colon cancer and surgical resection on skeletal muscle mitochondrial enzyme activity in colon cancer patients: a pilot study. Journal of cachexia, sarcopenia and muscle. 2012.

28. Gallagher IJ, Stephens NA, Macdonald AJ, Skipworth RJ, Husi H, Greig CA, et al. Suppression of skeletal muscle turnover in cancer cachexia: evidence from the transcriptome in sequential human muscle biopsies. Clinical cancer research : an official journal of the American Association for Cancer Research. 2012;18:2817-27.

29. Stephens NA, Skipworth RJ, Fearon KC. Cachexia, survival and the acute phase response. Current opinion in supportive and palliative care. 2008;2:267-74.

30. Brigitte M, Schilte C, Plonquet A, Baba-Amer Y, Henri A, Charlier C, et al. Muscle resident macrophages control the immune cell reaction in a mouse model of notexin-induced myoinjury. Arthritis and rheumatism. 2010;62:268-79.

31. Cardillo F, Postol E, Nihei J, Aroeira LS, Nomizo A, Mengel J. B cells modulate T cells so as to favour T helper type 1 and CD8+ T-cell responses in the acute phase of Trypanosoma cruzi infection. Immunology. 2007;122:584-95.

32. Arnold L, Henry A, Poron F, Baba-Amer Y, van Rooijen N, Plonquet A, et al. Inflammatory monocytes recruited after skeletal muscle injury switch into antiinflammatory macrophages to support myogenesis. The Journal of experimental medicine. 2007;204:1057-69.

33. White JP, Puppa MJ, Sato S, Gao S, Price RL, Baynes JW, et al. IL-6 regulation on skeletal muscle mitochondrial remodeling during cancer cachexia in the ApcMin/+ mouse. Skeletal muscle. 2012;2:14.

34. Wang X, Pickrell AM, Zimmers TA, Moraes CT. Increase in muscle mitochondrial biogenesis does not prevent muscle loss but increased tumor size in a mouse model of acute cancer-induced cachexia. PloS one. 2012;7:e33426.

35. Crouser ED. Peroxisome proliferator-activated receptors gamma coactivator-1alpha: master regulator of mitochondrial biogenesis and survival during critical illness? American journal of respiratory and critical care medicine. 2010;182:726-8.

36. Remels AH, Gosker HR, Schrauwen P, Hommelberg PP, Sliwinski P, Polkey M, et al. TNF-alpha impairs regulation of muscle oxidative phenotype: implications for cachexia? FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2010;24:5052-62.

37. Constantinou C, Fontes de Oliveira CC, Mintzopoulos D, Busquets S, He J, Kesarwani M, et al. Nuclear magnetic resonance in conjunction with functional genomics suggests mitochondrial dysfunction in a murine model of cancer cachexia. International journal of molecular medicine. 2011;27:15-24.

38. Siegel MP, Wilbur T, Mathis M, Shankland EG, Trieu A, Harper ME, et al. Impaired adaptability of in vivo mitochondrial energetics to acute oxidative insult in aged skeletal muscle. Mechanisms of ageing and development. 2012;133:620-8.

39. Dela F, Helge JW. Insulin resistance and mitochondrial function in skeletal muscle. The international journal of biochemistry & cell biology. 2013;45:11-5.

40. Meng SJ, Yu LJ. Oxidative stress, molecular inflammation and sarcopenia. International journal of molecular sciences. 2010;11:1509-26.

41. Brand MD, Nicholls DG. Assessing mitochondrial dysfunction in cells. The Biochemical journal. 2011;435:297-312.

42. Filippatos GS, Kanatselos C, Manolatos DD, Vougas B, Sideris A, Kardara D, et al. Studies on apoptosis and fibrosis in skeletal musculature: a comparison of

heart failure patients with and without cardiac cachexia. International journal of cardiology. 2003;90:107-13.

43. Kusko RL, Banerjee C, Long KK, Darcy A, Otis J, Sebastiani P, et al. Premature expression of a muscle fibrosis axis in chronic HIV infection. Skeletal muscle. 2012;2:10.

44. Fan X, Shao L, Fang H, Tong W, Cheng Y. Cross-platform comparison of microarray-based multiple-class prediction. PloS one. 2011;6:e16067.

45. Lin WJ, Hsueh HM, Chen JJ. Power and sample size estimation in microarray studies. BMC bioinformatics. 2010;11:48.

46. Hirakawa A, Hamada C, Yoshimura I. Sample size calculation through the incorporation of heteroscedasticity and dependence for a penalized t-statistic in microarray experiments. Journal of biopharmaceutical statistics. 2012;22:260-75.

47. Roth SM, Ferrell RE, Peters DG, Metter EJ, Hurley BF, Rogers MA. Influence of age, sex, and strength training on human muscle gene expression determined by microarray. Physiological genomics. 2002;10:181-90.

48. Cracking ENCODE. Lancet. 2012;380:950.

49. Craig A, Cloarec O, Holmes E, Nicholson JK, Lindon JC. Scaling and normalization effects in NMR spectroscopic metabonomic data sets. Analytical chemistry. 2006;78:2262-7.

50. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. Analytical chemistry. 2006;78:4281-90.

51. Warrack BM, Hnatyshyn S, Ott KH, Reily MD, Sanders M, Zhang H, et al. Normalization strategies for metabonomic analysis of urine samples. Journal of chromatography B, Analytical technologies in the biomedical and life sciences. 2009;877:547-52.

52. Slupsky CM, Rankin KN, Fu H, Chang D, Rowe BH, Charles PG, et al. Pneumococcal pneumonia: potential for diagnosis through a urinary metabolic profile. Journal of proteome research. 2009;8:5550-8.

53. Slupsky CM, Steed H, Wells TH, Dabbs K, Schepansky A, Capstick V, et al. Urine metabolite analysis offers potential early diagnosis of ovarian and breast cancers. Clinical cancer research : an official journal of the American Association for Cancer Research. 2010;16:5835-41.

54. Want EJ, Wilson ID, Gika H, Theodoridis G, Plumb RS, Shockcor J, Holmes E, Nicholson JK. Global metabolic profiling procedures for urine using UPLC-MS. Nature Protocols. 2010;6:1005-18.

Appendix 1: Differential gene expression according to sex

The full data set used in Chapter 3 included 134 participants (69 men and 65 women) and 41,000 oligonucleotide sequences for each subject. Differential expression was calculated using t-test on the log transformed intensities over the set of males vs. the set of females. Below is a list of the 717 differentially expressed oligunucleotide sequences with a p-value < 0.0001 and a false discovery rate < 0.003. Each oligonucleotide was mapped to its corresponding gene using Ingenuity Pathway Analysis (IPA) (Ingenuity® Systems, www.ingenuity.com). From the 717 sequences, 527 unique genes were mapped in IPA (see Gene Symbol column in Table AP1). While many of the top differentially expressed features were on the X and Y chromosomes, these only accounted for 10% and 4% of the 717 features with a p-value <0.0001, respectively.

Table AP1: Differentially expressed features according to sex

ID	Gene Symbol	Chromosome	p-value	FC	FDR
A_23_P137248	PRKY*	Y	3.09E-31	-0.7	1.27E-26
A_24_P130936	DDX3Y*	Y	1.32E-29	-0.5	2.70E-25
A_23_P160004	UTY*†	Y	7.47E-29	-1.6	1.02E-24
A_23_P364792	TXLNG2P*	Y	1.45E-26	-1.4	1.48E-22
A_23_P96658	TXLNG2P*	Y	3.09E-26	-1.4	2.53E-22
A_23_P251232	TTTY14*†	Y	4.18E-26	-1.2	2.86E-22
A_24_P66233	TTTY14*†	Y	1.21E-24	-1.3	6.32E-21
A_23_P148629	EIF1AY*	Y	1.23E-24	-0.6	6.32E-21
A_24_P307993	LOC100509121	Y	2.90E-24	-2.5	1.32E-20
A_23_P121441	NLGN4Y†	Y	5.17E-24	-1.3	2.12E-20
A_23_P11408	PRY2 (human)	Y	1.46E-23	-0.8	5.20E-20
A_32_P111701	GYG2*†	X	1.52E-23	-0.9	5.20E-20
A_32_P109165	GINS3*	16	1.95E-23	-1.2	6.16E-20
A_24_P186030	PRKY*	Y	4.13E-23	-1.0	1.21E-19
A_23_P137238	KDM5D	Y	8.57E-23	-1.1	2.34E-19
A_24_P216625	NCRNA00185*	Y	4.15E-22	-1.6	1.06E-18
A_24_P348861	TTTY15	Y	4.53E-22	-3.6	1.09E-18

A_24_P500584	XIST	X	6.21E-22	0.6	1.41E-18
A_23_P217507	ZBED1*†	X	1.84E-21	-1.1	3.96E-18
A_23_P137226	USP9Y	Y	2.74E-21	-1.5	5.61E-18
A_24_P942743	ZFY	Y	6.10E-21	-1.3	1.19E-17
A_23_P62446	HSFY1/HSFY2*	Y	6.90E-21	-2.9	1.29E-17
A_23_P73848	NCRNA00185*	Y	9.36E-21	-1.6	1.60E-17
A_23_P324384	RPS4Y2	Y	1.31E-20	-1.3	2.15E-17
A_24_P237511	EIF1AY*	Y	2.23E-20	-1.6	3.52E-17
A_23_P217797	DDX3Y*	Y	1.31E-19	-1.6	1.99E-16
A_23_P259314	RPS4Y1	Y	2.18E-18	-1.4	2.98E-15
A_23_P254842	HDHD1†	X	3.35E-18	1.1	4.44E-15
A_23_P137876	EIF1AX*	X	5.80E-18	1.1	7.43E-15
A_23_P307346	CA5B*†	X	5.45E-17	1.3	6.77E-14
A_23_P3934	RNF43	17	1.84E-16	-1.0	2.22E-13
A_24_P36745	CXorf38*	X	2.66E-16	1.1	3.11E-13
A_24_P134626	TXLNG†	X	5.48E-16	1.3	6.24E-13
A_32_P86623	ZBED1*	X	2.77E-15	-0.7	3.06E-12
A_32_P25737	CHIC1	X	4.06E-15	-0.6	4.38E-12
A_23_P35194	EIF1AX*	X	6.22E-15	1.2	6.54E-12
A_23_P85640	INPP5B†	1	6.93E-15	1.1	7.11E-12
A_23_P370027	GGT7*	20	7.99E-15	-1.4	7.99E-12
A_23_P253896	NPNT†	4	8.39E-15	0.6	8.19E-12
A_24_P378987	DHRSX†	X	1.43E-14	-0.7	1.37E-11
A_23_P152136	GINS3*	16	1.47E-14	-1.00	1.37E-11
A_23_P61886	TSPAN5*†	4	2.30E-14	-1.65	2.10E-11
A_24_P126060	DDX3X*†	X	5.41E-14	1.24	4.72E-11
A_23_P217304	KDM6A	X	7.89E-14	1.26	6.74E-11
A_23_P340148	ZNF711*	X	8.16E-14	1.11	6.83E-11
A_24_P586712	TPRG1	3	4.36E-13	2.00	3.46E-10
A_23_P317654	DDX3X*	X	4.39E-13	1.61	3.46E-10
A_23_P217297	ZNF711*	X	7.13E-13	1.12	5.51E-10
A_23_P152235	IRX3*	16	7.90E-13	-0.81	6.00E-10
A_23_P323930	TSPAN5*	4	1.00E-12	-0.80	7.45E-10
A_23_P325093	GGT7*	20	1.02E-12	-1.05	7.45E-10
A_23_P140748	NDRG4	16	2.09E-12	-2.05	1.47E-09
A_24_P45005	NPEPL1*†	20	2.68E-12	-0.60	1.86E-09
A_23_P57236	GGT7*	20	3.93E-12	-1.18	2.69E-09
A_23_P55586	CDH20†	18	8.96E-12	-2.21	6.02E-09
A_23_P255535	ASMT	X	1.19E-11	0.84	7.87E-09
A_23_P125656	DDX3X*	X	1.28E-11	1.18	8.31E-09
A_23_P149019	BAI2†	1	2.16E-11	1.25	1.38E-08
A_24_P68222	CD99P1	Y	4.21E-11	-1.18	2.58E-08
A_23_P339582	CXorf38*	X	6.68E-11	0.81	4.03E-08
A_24_P375761	STGC3	3	7.02E-11	0.75	4.17E-08
A_32_P52785	DAAM2†	6	9.05E-11	-1.46	5.26E-08
A_23_P137031	EIF2S3*	X	9.10E-11	1.87	5.26E-08
A_24_P196419	GGT7*	20	1.27E-10	-0.88	7.21E-08
A_24_P73769	C16orf89*	16	1.50E-10	1.00	8.45E-08

A_32_P144421	ZNF518B*	4	1.58E-10	-0.94	8.76E-08
A_23_P216257	TPD52*†	8	2.53E-10	-0.84	1.38E-07
A_23_P145514	IL20RA†	6	2.67E-10	-0.50	1.44E-07
A_23_P31921	ASS1†	9	3.11E-10	1.35	1.65E-07
A_24_P57631	GPC3†	X	3.32E-10	0.83	1.74E-07
A_23_P153945	GTDC1†	2	3.89E-10	-0.88	2.02E-07
A_24_P286114	SLC1A3*†	5	4.12E-10	0.74	2.11E-07
A_24_P237389	EIF1AX*	X	4.51E-10	1.40	2.28E-07
A_23_P85053	ZRSR2	X	5.02E-10	0.87	2.51E-07
A_24_P331560	STS†	X	6.09E-10	1.51	3.01E-07
A_23_P167509	CYFIP2*†	5	7.31E-10	1.24	3.57E-07
A_32_P171328	UBE2S*	19	8.15E-10	-0.62	3.91E-07
A_32_P34516	XKR6†	8	8.51E-10	-1.43	4.01E-07
A_23_P70991	AIMP2	7	1.05E-09	-1.19	4.85E-07
A_32_P81514	LOC100506725	7	1.05E-09	0.91	4.85E-07
A_23_P26640	C16orf89*	16	1.09E-09	0.74	4.97E-07
A_23_P371729	GJA5	1	1.15E-09	0.83	5.16E-07
A_24_P299193	GGT7*	20	1.28E-09	-1.42	5.73E-07
A_32_P197340	LOC285141	2	1.42E-09	-0.72	6.22E-07
A_23_P217411	SMC1A	X	1.81E-09	1.15	7.72E-07
A_24_P403561	LRP4	11	1.81E-09	1.15	7.72E-07
A_23_P306215	FAM84A	2	1.85E-09	-0.76	7.84E-07
A_24_P332647	SSH1 (includes EG:231637)†	12	2.35E-09	-1.66	9.84E-07
A_23_P359043	AKAP2/PALM2-AKAP2†	9	2.53E-09	1.73	1.05E-06
A_23_P217704	GYG2*	X	2.70E-09	0.83	1.11E-06
A_32_P182299	C1orf168†	1	3.00E-09	-1.10	1.22E-06
A_24_P121631	ZNF764	16	3.37E-09	-0.61	1.35E-06
A_24_P925505	CD36*	7	3.66E-09	0.59	1.46E-06
A_23_P315212	NTSR2	2	4.35E-09	1.47	1.71E-06
A_23_P212126	COLQ*	3	4.38E-09	0.74	1.71E-06
A_23_P111583	CD36*	7	4.45E-09	0.98	1.72E-06
A_23_P53198	DGAT2*	11	5.76E-09	0.76	2.21E-06
A_23_P254226	OFD1	X	6.85E-09	1.28	2.60E-06
A_24_P288848	CXorf36*	X	7.17E-09	1.13	2.70E-06
A_24_P96474	LDOC1L*	22	8.35E-09	-0.75	3.11E-06
A_23_P215883	NCALD†	8	9.27E-09	1.01	3.42E-06
A_23_P320622	TTTY10	Y	1.21E-08	-3.10	4.41E-06
A_23_P132308	C22orf23*	22	1.33E-08	-0.74	4.74E-06
A_23_P125639	ZFX*†	X	1.34E-08	1.58	4.74E-06
A_23_P121215	CAMK1	3	1.45E-08	1.73	5.07E-06
A_24_P26160	COLQ*	3	1.46E-08	0.63	5.07E-06
A_23_P211007	NRIP1	21	1.69E-08	-0.79	5.82E-06
A_23_P48676	PYGL	14	1.81E-08	1.08	6.19E-06
A_23_P362770	CCDC36*	3	1.84E-08	0.80	6.23E-06
A_23_P339240	PLCH1	3	2.06E-08	-4.41	6.93E-06
A_32_P72447	UBE2S*	19	2.50E-08	-0.57	8.34E-06
A_23_P114221	RBBP7†	X	2.52E-08	0.83	8.34E-06

A_23_P329835	UTY*	Y	2.66E-08	-3.76	8.73E-06
A_32_P184933	UBE2S*	19	2.74E-08	-0.53	8.93E-06
A_24_P132518	IKBKB†	8	2.93E-08	-1.25	9.46E-06
A_23_P250963	SLC1A3*	5	2.97E-08	0.55	9.51E-06
A_24_P340128	P2RY8	X	3.21E-08	-0.66	1.01E-05
A_23_P387656	EPB41L4B*†	9	3.27E-08	0.70	1.02E-05
A_23_P127565	LAYN†	11	3.33E-08	1.35	1.03E-05
A_23_P85250	CD24	Y	3.42E-08	-0.98	1.05E-05
A_32_P151557	EIF2S3*	X	3.66E-08	0.74	1.12E-05
A_24_P156576	GEMIN8†	X	3.87E-08	1.34	1.18E-05
A_23_P12755	LOXL4*	10	4.24E-08	1.36	1.27E-05
A_23_P309361	HENMT1	1	4.31E-08	-0.70	1.28E-05
A_24_P367645	MAP7D2	X	4.40E-08	1.04	1.30E-05
A_32_P100464	LOC100507588	19	4.68E-08	0.47	1.37E-05
A_24_P399680	FAM210B	20	4.96E-08	-1.22	1.43E-05
A_23_P58647	CTNNA1*†	5	4.98E-08	1.80	1.43E-05
A_24_P346431	TNS3*†	7	5.00E-08	1.64	1.43E-05
A_23_P210109	CYP26B1*	2	5.06E-08	2.05	1.44E-05
A_23_P503182	ABR†	17	5.59E-08	0.76	1.57E-05
A_23_P217379	COL4A6	X	6.02E-08	0.54	1.68E-05
A_23_P147423	ADAMTS9*†	3	6.12E-08	1.06	1.70E-05
A_32_P55241	SHISA2*	13	6.22E-08	-0.77	1.71E-05
A_24_P797678	ZC3H7B	11	6.79E-08	-1.02	1.84E-05
A_24_P207195	IRX3*	16	6.85E-08	-1.00	1.85E-05
A_24_P772488	PLXNA4	7	7.31E-08	0.67	1.96E-05
A_23_P127495	BBOX1	11	7.54E-08	2.09	2.01E-05
A_24_P295791	DGAT2*	11	7.61E-08	0.91	2.01E-05
A_23_P137209	UBA1*	X	7.98E-08	1.27	2.10E-05
A_23_P154526	GRB14†	2	8.12E-08	1.09	2.11E-05
A_23_P372255	ITPKB†	1	8.65E-08	1.44	2.23E-05
A_23_P52793	PPP2R1B*†	11	9.23E-08	1.52	2.36E-05
A_32_P461386	LOC100131840	3	9.58E-08	0.90	2.44E-05
A_23_P344481	STOX1*†	10	1.01E-07	1.80	2.51E-05
A_23_P165608	SEMA4F*	2	1.09E-07	0.74	2.71E-05
A_23_P83098	ALDH1A1†	9	1.11E-07	-1.02	2.75E-05
A_23_P162466	PKP2 (includes EG:287925)*	12	1.12E-07	-0.50	2.75E-05
A_24_P166807	TPD52*	8	1.21E-07	-0.58	2.95E-05
A_24_P944588	ZNF682	19	1.36E-07	-0.77	3.27E-05
A_32_P32391	OR7E156P*	13	1.37E-07	-0.99	3.28E-05
A_23_P435018	UNC45B*	17	1.58E-07	-0.67	3.75E-05
A_23_P206612	USP31*†	16	1.58E-07	-1.56	3.75E-05
A_23_P321703	BCL2A1	15	1.64E-07	-0.51	3.85E-05
A_32_P58407	KCND3*†	1	1.69E-07	0.82	3.96E-05
A_23_P353014	CACNA2D4	12	1.70E-07	1.61	3.97E-05
A_23_P25525	GTF3A	13	1.78E-07	-0.46	4.12E-05
A_23_P131723	YWHAQ*†	2	1.83E-07	1.70	4.22E-05
A_32_P197698	LOC153546	5	1.88E-07	-0.79	4.31E-05
A_24_P307974	TAF8 (includes	6	1.92E-07	-1.00	4.37E-05

	EG:129685)				
A_23_P429959	PRX	19	1.93E-07	1.56	4.37E-05
A_24_P390583	USP31*	16	2.05E-07	-1.39	4.61E-05
A_23_P131428	VAX2	2	2.14E-07	0.66	4.78E-05
A_24_P336137	C22orf23*	22	2.23E-07	-0.83	4.97E-05
A_32_P104063	CRNDE	16	2.25E-07	-0.71	4.98E-05
A_32_P224522	SLC25A23*	19	2.27E-07	-1.34	5.01E-05
A_24_P634530	CPPED1*	16	2.31E-07	-1.01	5.07E-05
A_24_P133288	PKP2 (includes EG:287925)*	12	2.79E-07	-0.72	6.05E-05
A_32_P51119	STOX1*	10	2.97E-07	1.30	6.40E-05
A_23_P156117	CYFIP2*	5	3.04E-07	1.24	6.53E-05
A_23_P206733	CES1	16	3.10E-07	1.09	6.58E-05
A_23_P346390	CXorf36*	X	3.18E-07	1.81	6.73E-05
A_23_P120953	SERHL2	22	3.36E-07	1.07	7.06E-05
A_23_P145895	TP53TG1*	7	3.41E-07	-0.79	7.12E-05
A_24_P677525	PLXNB2*	22	3.54E-07	1.94	7.34E-05
A_24_P4426	INPP5F*†	10	3.55E-07	1.30	7.34E-05
A_23_P11341	FAM104B*	X	3.56E-07	-0.45	7.34E-05
A_32_P156851	RCAN2†	6	3.64E-07	-0.90	7.45E-05
A_24_P161973	ATP11A	13	3.88E-07	-1.26	7.91E-05
A_23_P141992	HSD11B1L	19	3.92E-07	-0.76	7.95E-05
A_24_P284093	DACT1*	14	4.09E-07	0.96	8.26E-05
A_23_P26697	TRIM47	17	4.18E-07	1.04	8.36E-05
A_23_P416142	DLG1*†	3	4.18E-07	2.84	8.36E-05
A_23_P13442	MICAL2*†	11	4.33E-07	1.27	8.61E-05
A_23_P307682	FAM78B	1	5.07E-07	-0.53	0.00010
A_23_P65518	DACT1*	14	5.08E-07	1.50	0.00010
A_24_P193582	DEF8*	16	5.16E-07	-0.93	0.00010
A_24_P326660	MCAM*	11	5.42E-07	1.85	0.00011
A_24_P320796	FKBP9L	7	5.56E-07	1.71	0.00011
A_23_P159325	ANGPTL4	19	5.82E-07	0.70	0.00011
A_24_P929083	MAGI2-AS3	7	6.11E-07	-1.06	0.00012
A_24_P236235	FLRT2†	14	6.12E-07	1.05	0.00012
A_23_P54283	EID1*†	15	6.19E-07	-0.95	0.00012
A_24_P201531	ARCN1†	11	6.58E-07	1.30	0.00012
A_23_P3602	NUDT7	16	6.70E-07	-0.82	0.00012
A_23_P103588	HMGCS2	1	6.76E-07	0.71	0.00013
A_24_P162073	BCR*†	22	6.93E-07	1.12	0.00013
A_23_P52531	FAM24B†	10	6.95E-07	-1.09	0.00013
A_23_P200160	CFH	1	6.99E-07	1.09	0.00013
A_24_P230938	MORN4	10	7.44E-07	-1.01	0.00014
A_24_P20292	B3GNT7	2	7.52E-07	0.92	0.00014
A_23_P404965	GNL1	6	7.87E-07	-0.72	0.00014
A_23_P54758	GDE1 (includes EG:393213)	16	8.12E-07	-0.60	0.00015
A_23_P218918	FGF2†	4	8.40E-07	1.41	0.00015
A_23_P162171	MCAM*	11	8.61E-07	1.48	0.00015
A_24_P339858	C21orf90	21	8.70E-07	1.68	0.00015

A_23_P24843	MICAL2*	11	8.76E-07	1.09	0.00015
A_32_P177955	LOC441461	9	8.89E-07	-0.68	0.00016
A_23_P36745	ALDH2*	12	8.92E-07	0.55	0.00016
A_23_P210100	CYP26B1*	2	9.77E-07	2.18	0.00017
A_23_P134139	FABP7	6	9.97E-07	0.89	0.00017
A_24_P188071	TUBA1C†	12	1.04E-06	-1.03	0.00018
A_23_P60120	GSDMC	8	1.05E-06	0.53	0.00018
A_23_P57570	A4GALT	10	1.06E-06	-1.13	0.00018
A_23_P166023	PFDN4†	20	1.11E-06	-0.80	0.00019
A_23_P49677	UNC45B*	17	1.12E-06	-0.85	0.00019
A_32_P192594	LOC400099	13	1.14E-06	-1.18	0.00019
A_24_P622186	BMS1	10	1.14E-06	-0.84	0.00019
A_23_P113523	GTPBP6*	X	1.14E-06	-0.73	0.00019
A_23_P144807		5	1.17E-06	1.03	0.00019
A_23_P168551	SLC29A4	7	1.18E-06	0.69	0.00019
A_24_P111106	FGF1*†	5	1.19E-06	0.65	0.00019
A_24_P274987	TMEFF1†	9	1.20E-06	1.92	0.00020
A_32_P184279	CCDC6†	10	1.22E-06	-0.88	0.00020
A_24_P687131	LOC285033	2	1.25E-06	-1.19	0.00020
A_23_P3221	SQRDL	15	1.27E-06	-0.63	0.00020
A_23_P106322	CPEB1	15	1.27E-06	-0.62	0.00020
A_24_P216765	TOMM20†	1	1.28E-06	-0.69	0.00020
A_23_P7791	OGFRL1†	6	1.29E-06	-0.81	0.00020
A_32_P35220	CBWD5†	14	1.29E-06	-0.90	0.00020
A_23_P120488	NPEPL1*	9	1.29E-06	-0.58	0.00020
A_23_P44244	SMARCA1†	X	1.41E-06	0.77	0.00022
A_23_P401774	ELMOD1†	11	1.41E-06	-0.79	0.00022
A_24_P941038	VSTM4	10	1.41E-06	1.13	0.00022
A_24_P179225	MATN2*†	8	1.42E-06	0.71	0.00022
A_32_P166422	HECW1	7	1.44E-06	-0.80	0.00022
A_32_P196263	ADAMTS9*	3	1.46E-06	1.51	0.00022
A_23_P35684	INPP5F*	10	1.46E-06	1.93	0.00022
A_23_P131255	SPATS2L	2	1.47E-06	-0.84	0.00022
A_23_P120270	MCFD2†	4	1.52E-06	1.16	0.00023
A_23_P137514	IVNS1ABP	1	1.58E-06	-1.33	0.00024
A_23_P502731	PRRX1†	1	1.64E-06	0.69	0.00025
A_24_P393372	PACS2	5	1.67E-06	0.80	0.00025
A_32_P194025	FAM104B*	X	1.74E-06	-0.52	0.00026
A_32_P226646	LOC100129781†	16	1.83E-06	-0.74	0.00027
A_24_P175783	ARHGEF12†	11	1.87E-06	2.59	0.00028
A_24_P184803	COCH	14	2.00E-06	0.90	0.00029
A_23_P400465	GTF3C6	6	2.04E-06	-0.85	0.00030
A_23_P142815	ATP6V1B1	2	2.06E-06	1.24	0.00030
A_23_P416813	ZFP82	19	2.16E-06	0.67	0.00031
A_24_P274842	TP53TG1*	7	2.17E-06	-0.58	0.00031
A_23_P1492	AVPI1	10	2.27E-06	1.31	0.00033
A_23_P148600	INE1	X	2.27E-06	1.25	0.00033
A_24_P290527	ZFX*	X	2.37E-06	1.16	0.00034

A_23_P71328	MATN2*†	8	2.55E-06	1.02	0.00036
A_24_P152404	C10orf76†	10	2.58E-06	-0.60	0.00036
A_23_P99614	BTBD6	14	2.59E-06	-0.75	0.00036
A_23_P209669	NRP2†	2	2.59E-06	0.77	0.00036
A_24_P286527	RABL5*	7	2.63E-06	-0.65	0.00037
A_23_P53126	LMO2	11	2.83E-06	-1.53	0.00040
A_23_P31124	COL21A1†	6	2.85E-06	1.49	0.00040
A_24_P295999	CD4	12	2.90E-06	-0.60	0.00040
A_23_P126623	PGD†	1	2.92E-06	0.79	0.00040
A_23_P5435	KRTCAP3*	2	2.95E-06	-1.43	0.00041
A_23_P60079	ANGPT2	8	2.98E-06	1.56	0.00041
A_32_P180315	C9orf174†	9	3.00E-06	-0.77	0.00041
A_23_P203115	TMEM25	11	3.02E-06	-0.59	0.00041
A_23_P308954	BHLHB9	X	3.02E-06	-0.46	0.00041
A_32_P221799	HIST1H2AG (includes others)†	6	3.04E-06	0.88	0.00041
A_23_P155556	CLDND1	3	3.13E-06	1.09	0.00042
A_24_P134765	CCDC36*	3	3.23E-06	0.55	0.00043
A_23_P22926	GNB1	1	3.28E-06	1.86	0.00044
A_32_P122754	C9orf30*†	9	3.50E-06	1.57	0.00047
A_32_P195850	DPY19L2	12	3.51E-06	-0.95	0.00047
A_24_P70888	PLXNB2*	22	3.53E-06	1.78	0.00047
A_24_P335202	CHRDL2*	11	3.75E-06	-1.53	0.00049
A_24_P734953	TRNP1	1	3.76E-06	0.73	0.00049
A_23_P92073	PARP3	3	3.77E-06	-0.65	0.00049
A_23_P16110	OR7E24	19	3.85E-06	-1.47	0.00050
A_23_P20752	CDK20	9	3.87E-06	-0.87	0.00050
A_24_P384018	OR7E156P*	13	4.03E-06	-1.11	0.00052
A_23_P36496	RBMS1†	2	4.03E-06	1.22	0.00052
A_32_P48397	PLXNB2*	22	4.13E-06	2.26	0.00053
A_23_P314222	LEO1†	15	4.21E-06	-0.61	0.00054
A_24_P262688	LAIR1	19	4.26E-06	2.25	0.00055
A_23_P94159	FBXO25	8	4.35E-06	-0.72	0.00056
A_23_P377882	KCNH2	7	4.39E-06	1.07	0.00056
A_24_P650482	PCBP1-AS1	2	4.42E-06	-0.99	0.00056
A_23_P54116	DAAM1†	14	4.43E-06	1.21	0.00056
A_24_P242688	HADHA	2	4.53E-06	0.93	0.00057
A_23_P381172	MRAP*	21	4.54E-06	1.57	0.00057
A_23_P208259	ZNF665 (includes others)	19	4.63E-06	0.60	0.00058
A_23_P44257	COMMD8†	4	4.69E-06	-0.84	0.00058
A_23_P164047	MMD†	17	4.69E-06	1.67	0.00058
A_24_P607880	IPW	15	4.77E-06	-0.94	0.00059
A_23_P98446	SC5DL*	11	4.78E-06	-1.76	0.00059
A_24_P367776	ACSM5†	16	4.88E-06	-0.52	0.00060
A_23_P346337	SFXN1	5	4.90E-06	-1.07	0.00060
A_23_P215642	TNS3*	7	5.08E-06	0.90	0.00062
A_23_P31584	RABL5*	7	5.11E-06	-0.70	0.00062
A_23_P47616	FOLH1	11	5.12E-06	1.12	0.00062

A_23_P156890	TCF21	6	5.14E-06	0.65	0.00062
A_24_P196469	TCEANC	X	5.20E-06	0.77	0.00063
A_32_P171747	ZNF518B*	4	5.29E-06	-0.82	0.00064
A_23_P71268	AZGP1*	7	5.32E-06	1.11	0.00064
A_23_P122976	GNAI1†	7	5.32E-06	1.37	0.00064
A_32_P155035	LOC100506411	14	5.39E-06	-0.95	0.00064
A_32_P195065	SEMA4F*	2	5.44E-06	1.23	0.00065
A_24_P127235	BCR*	22	5.52E-06	1.02	0.00065
A_23_P68978	EFCAB6†	22	5.65E-06	-0.52	0.00067
A_23_P110212	ACSL1†	4	5.68E-06	0.99	0.00067
A_23_P348737	NR2F1	5	5.78E-06	1.42	0.00068
A_23_P71270	AZGP1*	7	5.83E-06	1.88	0.00068
A_23_P23584	CTNNBIP1†	1	5.83E-06	-0.91	0.00068
A_23_P95823	NSMCE1†	16	5.89E-06	-0.41	0.00068
A_24_P223518	FKBP9	7	5.93E-06	1.79	0.00069
A_32_P61729	RTN3*	11	5.96E-06	2.37	0.00069
A_24_P85085	SYTL2†	11	5.97E-06	1.96	0.00069
A_23_P68700	DNAJC28	21	5.98E-06	-0.85	0.00069
A_24_P153643	DOCK3†	3	6.01E-06	-0.73	0.00069
A_24_P126210	MRAP*	21	6.16E-06	0.49	0.00070
A_24_P335221	RTN3*	11	6.35E-06	1.36	0.00072
A_32_P181527	C8orf85	8	6.44E-06	-2.51	0.00073
A_24_P347480	NEK9†	14	6.50E-06	-0.68	0.00073
A_23_P36753	ALDH2*	12	6.55E-06	0.56	0.00074
A_23_P207999	PMAIP1†	18	6.63E-06	-0.95	0.00074
A_23_P73982	TMEM48†	1	6.64E-06	-0.86	0.00074
A_24_P406754	LOXL4*	10	6.72E-06	1.67	0.00075
A_24_P160225	CXorf36*	X	6.82E-06	0.24	0.00076
A_23_P11841	ATP2B4†	1	6.85E-06	2.26	0.00076
A_23_P211878	FLNB†	3	6.89E-06	1.25	0.00076
A_23_P416656	MYO1C	17	7.05E-06	1.37	0.00078
A_32_P323	LRRC40	1	7.35E-06	1.18	0.00081
A_24_P237374	UBA1*	X	7.39E-06	2.39	0.00081
A_24_P199905	YWHAQ*	2	7.50E-06	1.27	0.00082
A_23_P257335	KHDRBS3†	8	7.60E-06	0.80	0.00083
A_23_P88893	DEF8*	16	7.72E-06	-1.18	0.00084
A_23_P129085	SPESP1†	15	7.79E-06	-0.33	0.00085
A_32_P49748	B4GALNT3	12	7.82E-06	0.80	0.00085
A_23_P144054	PRKCD	3	7.90E-06	3.75	0.00085
A_23_P54576	KIFC3	16	7.95E-06	1.41	0.00086
A_23_P13548	CHRDL2*	11	7.97E-06	-0.89	0.00086
A_24_P304549	LAMP1	13	8.00E-06	0.88	0.00086
A_32_P147149	FLJ33630	5	8.18E-06	-1.00	0.00087
A_24_P80633	CTNNA1*	5	8.18E-06	2.40	0.00087
A_23_P213336	FGF1*	5	8.23E-06	0.48	0.00087
A_23_P144020	CNTN4	3	8.47E-06	1.37	0.00089
A_32_P228067	C1orf101	1	8.50E-06	-0.39	0.00089
A_23_P212034	DLG1*	3	8.50E-06	1.33	0.00089

A_24_P89708	IMPDH1	7	8.68E-06	1.10	0.00091
A_23_P500381	HTR7†	10	8.82E-06	-1.85	0.00092
A_23_P151209	CSRNP2	12	8.87E-06	1.34	0.00092
A_23_P14708	ZNF280D†	15	9.03E-06	1.32	0.00094
A_23_P75749	GLYAT	11	9.14E-06	0.77	0.00095
A_24_P297078	C20orf3	20	9.23E-06	1.26	0.00095
A_23_P62270	KDM5C	X	9.36E-06	1.18	0.00096
A_23_P394836	INF2	14	9.49E-06	1.52	0.00098
A_23_P129956	DUSP3†	17	9.58E-06	-0.83	0.00098
A_23_P25487	SLC38A4*†	12	9.59E-06	0.92	0.00098
A_32_P232883	LOC100507165	11	9.61E-06	-0.72	0.00098
A_24_P195785	COL4A5	X	9.74E-06	0.92	0.00099
A_23_P39076	RRAS*	19	9.93E-06	1.18	0.00101
A_23_P65918	ITPKA†	15	9.99E-06	-1.70	0.00101
A_23_P103256	CFHR3	1	9.99E-06	1.22	0.00101
A_32_P232865	BCR*	22	1.00E-05	0.85	0.00101
A_24_P239988	ITSN2	2	1.02E-05	1.02	0.00102
A_23_P155477	C3orf18	3	1.03E-05	-0.64	0.00103
A_23_P3204	MAPK6†	15	1.04E-05	-1.06	0.00104
A_23_P250800	ST3GAL6†	3	1.08E-05	0.63	0.00108
A_24_P386622	ARRB1	11	1.10E-05	1.01	0.00109
A_24_P728215	ZBTB20-AS1	3	1.10E-05	-1.05	0.00109
A_23_P253677	TMEM192	4	1.13E-05	-0.77	0.00112
A_24_P111547	MDH1B	2	1.13E-05	-0.43	0.00112
A_24_P97849	DBN1*	5	1.16E-05	0.87	0.00114
A_23_P118246	GINS2	16	1.21E-05	-0.59	0.00118
A_23_P19657	LRP11	6	1.22E-05	0.56	0.00119
A_24_P393496	LOC100288798	12	1.23E-05	-1.02	0.00120
A_32_P89730	LOH12CR2*	12	1.23E-05	-0.98	0.00120
A_23_P380990	CLEC4F	2	1.28E-05	2.53	0.00124
A_23_P209246	GLI2†	2	1.29E-05	1.73	0.00124
A_23_P77066	SNRPN†	15	1.30E-05	-0.74	0.00125
A_23_P127915	STK33	9	1.30E-05	-0.48	0.00125
A_23_P41917	HOMER1	5	1.34E-05	-0.51	0.00129
A_23_P84995	MTMR8	X	1.35E-05	0.73	0.00129
A_24_P321919	IQGAP1†	15	1.35E-05	1.28	0.00129
A_23_P75283	RBP4	10	1.36E-05	0.66	0.00129
A_23_P39074	RRAS*	19	1.38E-05	1.52	0.00131
A_32_P169406	LOC400043	12	1.39E-05	1.88	0.00131
A_23_P61623	SLCO3A1†	12	1.39E-05	1.29	0.00131
A_23_P78782	CA11	19	1.40E-05	-0.91	0.00131
A_24_P933704	PAM†	5	1.40E-05	1.71	0.00131
A_23_P5757	TPRKB	2	1.41E-05	-0.38	0.00132
A_23_P161439	C10orf116	10	1.41E-05	1.77	0.00132
A_23_P43226	KCTD9†	8	1.43E-05	-0.94	0.00134
A_23_P4536	EPB41L3	18	1.44E-05	1.20	0.00134
A_23_P163258	PARP6†	15	1.44E-05	-0.94	0.00134
A_24_P30806	EID1*	15	1.47E-05	-1.12	0.00136

A_23_P215525	OSBPL3†	7	1.50E-05	1.01	0.00138
A_23_P316612	GLIS1†	1	1.54E-05	1.46	0.00142
A_23_P358394	FAM65B*	6	1.55E-05	0.69	0.00142
A_24_P196774	SNX8	7	1.56E-05	-0.66	0.00143
A_24_P174313	KANSL3*	2	1.59E-05	-0.60	0.00144
A_23_P104798	IL18 (includes EG:16173)†	11	1.65E-05	1.16	0.00150
A_23_P350617	KLB*	4	1.68E-05	0.85	0.00152
A_23_P111260	NT5E†	6	1.68E-05	1.17	0.00152
A_23_P433132	KRTCAP3*	2	1.70E-05	-0.95	0.00153
A_23_P76015	ARHGEF17	11	1.74E-05	1.23	0.00156
A_23_P382045	TULP4†	6	1.75E-05	-1.11	0.00157
A_23_P104563	CPT1A	11	1.78E-05	1.30	0.00159
A_23_P363954	THRSP*	11	1.82E-05	0.66	0.00162
A_24_P109214	APOC1*	19	1.84E-05	-1.27	0.00163
A_23_P200984	PTPRF*†	1	1.86E-05	1.76	0.00165
A_23_P40657	GCAT	22	1.87E-05	-0.66	0.00165
A_23_P11295	MTCP1NB	X	1.88E-05	-0.82	0.00166
A_24_P756657	C6orf225	6	1.89E-05	-0.78	0.00166
A_24_P171268	RASSF5†	1	1.90E-05	-0.73	0.00167
A_32_P198810	LOH12CR2*	12	1.90E-05	-0.91	0.00167
A_24_P221903	LRRC3DN	21	1.94E-05	0.98	0.00170
A_24_P339071	CDR2	16	1.95E-05	1.78	0.00171
A_23_P204736	GPD1	12	1.97E-05	-0.43	0.00171
A_32_P134167	CUL5	11	2.00E-05	-0.79	0.00174
A_23_P29023	C21orf119	21	2.01E-05	-0.64	0.00174
A_24_P270814	CRK*†	17	2.05E-05	2.05	0.00177
A_24_P284523	MAP3K10	19	2.12E-05	-1.17	0.00183
A_32_P213330	RGNEF†	5	2.13E-05	3.14	0.00183
A_24_P100627	ATRNL1†	10	2.13E-05	-0.30	0.00183
A_23_P94338	ENPP2†	8	2.15E-05	1.34	0.00184
A_23_P111672	TES†	7	2.19E-05	0.86	0.00187
A_24_P383609	NANOS1	10	2.20E-05	-0.59	0.00188
A_24_P382533	NNAT†	20	2.25E-05	1.29	0.00191
A_32_P211621	SLC25A24	1	2.27E-05	-0.86	0.00192
A_32_P143000	FAM189A1	15	2.27E-05	0.94	0.00192
A_23_P98455	VWA5A	11	2.32E-05	1.09	0.00196
A_23_P348227	ZNF135*	19	2.36E-05	0.68	0.00199
A_23_P204640	NANOG	12	2.39E-05	-1.28	0.00201
A_23_P217339	PRKX†	X	2.40E-05	2.42	0.00201
A_23_P163458	EHD4†	15	2.41E-05	1.77	0.00201
A_23_P36562	ITGA5†	12	2.41E-05	2.25	0.00201
A_23_P103690	FAM189B*	1	2.41E-05	-0.51	0.00201
A_23_P258310	PXDNL	8	2.44E-05	1.56	0.00203
A_23_P208450	SLC25A23*	19	2.44E-05	-1.34	0.00203
A_32_P5480	CERS6†	2	2.48E-05	-0.78	0.00206
A_32_P85330	C15orf37	15	2.50E-05	-0.45	0.00207
A_32_P4814	TMEM185A	X	2.51E-05	-0.74	0.00207
A_23_P20558	CDC37L1†	9	2.51E-05	-1.19	0.00207

A_32_P210872	HEPN1	11	2.53E-05	8.46	0.00208
A_23_P4082	CCT6B	17	2.54E-05	-1.21	0.00208
A_24_P234701	APBB2†	4	2.54E-05	2.39	0.00208
A_23_P252075	AHCYL2	7	2.56E-05	0.82	0.00210
A_23_P72187	MED14	X	2.58E-05	1.15	0.00210
A_32_P88240	KBTBD12	3	2.58E-05	-0.78	0.00210
A_23_P69531	KLB*	4	2.61E-05	0.91	0.00213
A_23_P110882	TSPYL4†	6	2.63E-05	-0.92	0.00213
A_23_P119857	TTC32†	2	2.64E-05	-1.82	0.00214
A_23_P372888	SC5DL*	11	2.65E-05	-0.48	0.00214
A_23_P371145	ADPRHL1*†	13	2.71E-05	-0.86	0.00219
A_24_P167030	AP3M1	10	2.72E-05	1.66	0.00219
A_24_P355145	DNAJC5B	8	2.73E-05	0.52	0.00220
A_24_P297166		22	2.76E-05	0.84	0.00221
A_24_P83738	ASTN2†	9	2.83E-05	1.76	0.00227
A_23_P130677	C19orf80	19	2.85E-05	0.79	0.00228
A_32_P224253	DNAH11	7	2.88E-05	1.24	0.00229
A_23_P11685	PLA2G4A	1	2.92E-05	0.84	0.00232
A_24_P136683	CA5BP1	X	3.01E-05	1.63	0.00239
A_32_P427150	SHISA2*	13	3.01E-05	-0.68	0.00239
A_23_P412186	ZNF252	8	3.02E-05	-0.45	0.00239
A_23_P356484	RPS10	6	3.03E-05	-0.50	0.00239
A_24_P934755	LOC100507948	2	3.04E-05	-0.59	0.00239
A_23_P422115	C9orf116	9	3.04E-05	-0.61	0.00239
A_23_P431360	ZNF219*†	14	3.06E-05	-0.78	0.00240
A_32_P56713	BCR*	22	3.10E-05	1.64	0.00242
A_32_P171530	LIFR-AS1	5	3.10E-05	-0.46	0.00242
A_24_P321581	SLC38A4*	12	3.11E-05	1.31	0.00242
A_23_P255263	STOM†	9	3.15E-05	1.58	0.00245
A_23_P408249	PCK1 (includes EG:18534)*	20	3.16E-05	1.03	0.00245
A_23_P45579	HSFY1/HSFY2*	Y	3.17E-05	-3.62	0.00246
A_23_P392126	C17orf108	17	3.26E-05	-0.44	0.00252
A_23_P36658	MGST1	6	3.28E-05	0.79	0.00253
A_32_P149298	KIAA1841	2	3.29E-05	0.63	0.00253
A_23_P400378	GPBAR1	2	3.34E-05	1.39	0.00256
A_23_P167096	VEGFC†	14	3.39E-05	-1.01	0.00259
A_24_P71781	TMEM108†	3	3.42E-05	1.28	0.00261
A_23_P16451	UBXN6	19	3.44E-05	-0.68	0.00261
A_24_P772103	PITPNC1†	17	3.44E-05	2.01	0.00261
A_23_P369237	ADIPOQ	3	3.47E-05	0.56	0.00263
A_23_P53397	SP1	12	3.53E-05	1.04	0.00267
A_23_P137470	SIPA1L2†	1	3.56E-05	0.40	0.00269
A_24_P233995		1	3.60E-05	2.08	0.00271
A_23_P217168	CXorf36*	X	3.61E-05	1.12	0.00271
A_23_P79289	COBLL1*†	2	3.64E-05	1.21	0.00273
A_23_P79251	EHD3	2	3.67E-05	-1.02	0.00275
A_23_P63010	CERS2	1	3.69E-05	1.27	0.00276
A_23_P426305	AOC3	17	3.73E-05	1.89	0.00278

A_23_P168105	EGFL8	6	3.75E-05	1.28	0.00279
A_23_P108437	FZD5*	2	3.77E-05	2.48	0.00280
A_23_P6708	LOC152217	3	3.80E-05	-0.47	0.00282
A_23_P48438	ADPRHL1*	13	3.82E-05	-0.94	0.00283
A_23_P136116	AGMO	7	3.83E-05	1.56	0.00283
A_23_P76799	BAZ1A†	14	3.85E-05	1.66	0.00284
A_23_P215009	FAM65B*	6	3.86E-05	1.01	0.00284
A_23_P77401	CPPED1*	16	3.96E-05	-0.90	0.00291
A_23_P4649	APOC1*	19	3.98E-05	-1.40	0.00291
A_24_P382187	IGFBP4	17	3.98E-05	2.19	0.00291
A_24_P341222	C4orf52	4	3.99E-05	-0.19	0.00291
A_24_P308851	PCDHGA7†	5	4.00E-05	0.82	0.00292
A_32_P3385	FLJ37798	6	4.04E-05	-0.52	0.00294
A_24_P272225	LOC645676†	1	4.10E-05	-0.38	0.00297
A_24_P107291	PPP2R1B*	1	4.11E-05	-0.86	0.00297
A_23_P258698	MANBA†	11	4.11E-05	3.32	0.00297
A_23_P152082	SPTBN5†	15	4.17E-05	0.75	0.00301
A_23_P105212	THRSP*	11	4.20E-05	0.68	0.00302
A_23_P368259	EID2B	19	4.21E-05	-0.86	0.00303
A_24_P73738	RPL13†	16	4.24E-05	-0.67	0.00304
A_23_P56197	CRLF1	19	4.25E-05	-0.73	0.00305
A_23_P201156	CADM3	1	4.26E-05	1.04	0.00305
A_23_P251293	SNCG	10	4.29E-05	3.03	0.00306
A_23_P210623	PCK1 (includes EG:18534)*	20	4.32E-05	3.68	0.00307
A_24_P174316	KANSL3*	2	4.32E-05	-0.66	0.00307
A_23_P94141	RAD54B	8	4.36E-05	-0.92	0.00309
A_24_P932939	LOC401052	3	4.41E-05	1.30	0.00312
A_23_P215735	ST7†	7	4.42E-05	1.65	0.00312
A_23_P28258	PRKAG3†	2	4.42E-05	-0.51	0.00312
A_24_P389415	PNMA2	8	4.46E-05	0.89	0.00314
A_23_P136433	FGF1*	5	4.52E-05	0.58	0.00317
A_32_P219279	ELFN2†	22	4.56E-05	1.13	0.00319
A_23_P63101	FAM189B*	1	4.57E-05	-0.60	0.00320
A_32_P106646	FAM36A†	1	4.61E-05	-0.70	0.00322
A_32_P157391	FOLH1B	11	4.68E-05	0.34	0.00326
A_24_P924862	RAPH1†	2	4.83E-05	-0.79	0.00334
A_24_P364087	SERGEF	11	4.84E-05	-0.99	0.00335
A_23_P42036	LYRM2	6	4.85E-05	-0.42	0.00335
A_24_P328657	PEX5L	3	4.86E-05	1.14	0.00335
A_24_P148503	FZD5*	2	4.87E-05	1.33	0.00335
A_23_P255851	RNF38	9	4.89E-05	2.59	0.00335
A_23_P142055	C19orf38	19	4.96E-05	-0.81	0.00339
A_23_P126186	DEGS1	1	4.96E-05	0.95	0.00339
A_23_P121564	GUCY1B3†	4	5.06E-05	0.61	0.00344
A_23_P357207	MRAP2	6	5.08E-05	-1.66	0.00344
A_23_P101084	SPATA22	17	5.08E-05	1.09	0.00344
A_23_P214739	FBXL4	6	5.09E-05	-0.69	0.00344
A_24_P35891	ZNF219*	14	5.11E-05	-0.89	0.00345

A_32_P224302	ZNF135*	19	5.14E-05	0.91	0.00346
A_24_P254532	PGK1†	X	5.20E-05	-1.47	0.00350
A_23_P17695	SLC37A1	21	5.33E-05	-0.95	0.00356
A_32_P208654	PIWIL2†	8	5.34E-05	-0.33	0.00356
A_23_P17786	PITPNB†	22	5.35E-05	1.44	0.00356
A_32_P55462	ZFR2	19	5.37E-05	-0.40	0.00357
A_32_P144018	LOC100506451	12	5.38E-05	-1.33	0.00357
A_23_P85783	PHGDH	1	5.43E-05	1.37	0.00360
A_23_P146584	C9orf30*	9	5.49E-05	2.39	0.00363
A_32_P228804	COBLL1*	2	5.51E-05	1.58	0.00364
A_23_P147711	NPR1 (includes EG:18160)	1	5.56E-05	1.46	0.00367
A_23_P111517	WBSCR17†	7	5.58E-05	1.78	0.00367
A_32_P205637	PARD6B†	20	5.62E-05	-0.56	0.00368
A_24_P222835	S100PBP	1	5.62E-05	3.70	0.00368
A_23_P340149	ZNF711*	X	5.62E-05	1.45	0.00368
A_23_P2414	C12orf39	12	5.64E-05	0.70	0.00368
A_23_P108075	SLC7A10	19	5.67E-05	0.95	0.00369
A_23_P327307	PAK2†	3	5.68E-05	2.31	0.00369
A_32_P42075	CCZ1/CCZ1B	7	5.74E-05	-1.10	0.00373
A_23_P5550	PUM2†	2	5.86E-05	2.86	0.00380
A_23_P105138	CAT	11	5.87E-05	0.99	0.00380
A_23_P4922	C19orf68	19	5.89E-05	-0.96	0.00380
A_23_P39871	SLC19A3	2	5.91E-05	1.15	0.00381
A_24_P54174	TNFRSF1B†	1	5.96E-05	-1.05	0.00383
A_23_P21324	TWIST2	2	5.97E-05	0.93	0.00383
A_32_P447001	FLJ27352	15	6.01E-05	-0.70	0.00385
A_24_P260122	PHLDB2†	3	6.02E-05	1.64	0.00385
A_23_P216556	EPB41L4B*	9	6.03E-05	2.28	0.00385
A_32_P140268	KCND3*	1	6.07E-05	0.86	0.00387
A_23_P202837	CCND1	11	6.22E-05	1.13	0.00396
A_23_P254165	RAI2	X	6.36E-05	-1.43	0.00404
A_24_P943575	CHD6†	20	6.46E-05	-2.36	0.00409
A_24_P74374	CTSA	20	6.63E-05	2.54	0.00419
A_23_P162211	MANSC1†	12	6.68E-05	0.60	0.00422
A_24_P358381	GTPBP6*	X	6.71E-05	-0.97	0.00423
A_23_P64721	HCAR3	12	6.79E-05	0.87	0.00427
A_23_P51699	ARHGEF2†	1	6.80E-05	1.79	0.00427
A_23_P363472	NDFIP2	13	6.80E-05	-0.80	0.00427
A_24_P861009	BRWD1†	21	6.92E-05	-0.66	0.00433
A_23_P422212	SLC35F3	1	6.95E-05	1.96	0.00434
A_23_P141044	ZNF688	16	6.97E-05	-0.69	0.00435
A_24_P941268	CA5B*	X	6.98E-05	0.93	0.00435
A_23_P159053	RAD17 (includes EG:19356)	5	6.99E-05	-0.44	0.00435
A_24_P385313	PTPRF*	1	7.02E-05	4.26	0.00436
A_23_P157963	CNTLN	9	7.06E-05	-0.53	0.00437
A_24_P100830	AMN1 (includes EG:196394)†	12	7.14E-05	-0.37	0.00442

A_23_P132365	LDOC1L*	22	7.22E-05	-0.37	0.00446
A_23_P106675	PLCG2†	16	7.27E-05	1.61	0.00447
A_23_P10121	SFRP1	8	7.27E-05	1.26	0.00447
A_23_P500353	KCNN2	5	7.28E-05	0.62	0.00447
A_32_P34920	FOXD1†	5	7.30E-05	1.66	0.00448
A_23_P171232	AMOT	X	7.34E-05	0.97	0.00450
A_24_P56557	ATPBD4	15	7.50E-05	-1.04	0.00458
A_23_P99980	HMGB1	13	7.54E-05	-1.04	0.00460
A_23_P203751	TMEM135	11	7.65E-05	1.24	0.00466
A_23_P330788	IQSEC2	X	7.67E-05	-0.69	0.00466
A_32_P36694	JAZF1†	7	7.67E-05	1.25	0.00466
A_23_P256542	FAM162A	3	7.68E-05	-1.23	0.00466
A_23_P133665	FRK†	6	7.70E-05	0.75	0.00467
A_23_P56328	PLVAP†	19	7.80E-05	1.06	0.00472
A_24_P354615	MTMR12	5	8.02E-05	1.01	0.00484
A_32_P223059	SLC45A1†	1	8.06E-05	-1.43	0.00486
A_23_P83556	CRK*	17	8.11E-05	1.25	0.00488
A_23_P503115	BCR*	22	8.12E-05	0.73	0.00488
A_23_P305759	ABHD3	18	8.18E-05	1.22	0.00491
A_23_P35114	PLEKHO1†	1	8.25E-05	-1.28	0.00495
A_23_P208812	ZNF507†	19	8.43E-05	-1.15	0.00504
A_32_P206949	TMEM17	2	8.49E-05	-0.68	0.00506
A_23_P48585	SALL2	14	8.52E-05	-0.49	0.00507
A_23_P206310	KIAA0513	16	8.57E-05	-0.88	0.00509
A_23_P11192	UBE2E3†	2	8.71E-05	-0.68	0.00517
A_23_P129188	CALML4	15	8.84E-05	1.48	0.00524
A_23_P434212	SULT1A1	16	8.88E-05	3.04	0.00525
A_23_P92025	CIDEC†	3	8.98E-05	1.56	0.00531
A_24_P719579	CISD3	17	9.01E-05	-0.97	0.00532
A_23_P30126	FGFBP1	4	9.17E-05	9.27	0.00539
A_23_P44836	NT5DC2	3	9.19E-05	2.98	0.00540
A_23_P212675	NME9	3	9.24E-05	-1.16	0.00542
A_24_P62833	ADAMTSL4†	1	9.35E-05	0.78	0.00547
A_23_P141484	C17orf63†	17	9.38E-05	-0.95	0.00547
A_23_P250914	ATP6V1C2	2	9.39E-05	0.94	0.00547
A_23_P432947	GREM1†	15	9.41E-05	-0.31	0.00548
A_23_P156284	DBN1*	5	9.50E-05	1.12	0.00552
A_23_P57709	PCOLCE2 (includes EG:26577)†	3	9.51E-05	0.59	0.00552
A_23_P501276	TUBB2A†	6	9.56E-05	1.01	0.00554
A_24_P59643	KIAA1456	8	9.59E-05	0.77	0.00555
A_24_P62615	CAP1†	1	9.63E-05	1.68	0.00557
A_24_P218814	RDH5†	12	9.69E-05	1.09	0.00559
A_24_P98249	TACC1†	8	9.81E-05	1.19	0.00564
A_32_P166031	LOC100507568	15	9.82E-05	1.76	0.00564
A_23_P93009	SRP19	5	9.86E-05	0.59	0.00565
A_23_P158041	AQP7	9	9.98E-05	1.88	0.00571
A_32_P209250	Not mapped by IPA	Y	9.08E-21	-1.25	0.00000
A_32_P35165	Not mapped by IPA	X	8.52E-19	-3.46	0.00000

A_32_P216715	Not mapped by IPA	X	1.84E-18	0.85	0.00000
A_24_P642758	Not mapped by IPA	X	3.78E-14	1.20	0.00000
A_24_P238386	Not mapped by IPA	16	1.36E-13	-3.24	0.00000
A_24_P703642	Not mapped by IPA	2	1.23E-12	1.32	0.00000
A_32_P11325	Not mapped by IPA	9	2.79E-11	-0.95	0.00000
A_32_P117185	Not mapped by IPA	2	3.26E-11	0.88	0.00000
A_24_P332541	Not mapped by IPA	12	8.21E-10	1.60	0.00000
A_24_P33055	Not mapped by IPA	7	1.43E-09	1.76	0.00000
A_24_P178444	Not mapped by IPA	11	1.24E-08	1.59	0.00000
A_24_P306814	Not mapped by IPA	5	1.26E-08	1.62	0.00000
A_24_P913576	Not mapped by IPA	14	3.19E-08	-1.22	0.00001
A_24_P929818	Not mapped by IPA	9	4.22E-08	1.11	0.00001
A_24_P255836	Not mapped by IPA	2	5.54E-08	-0.83	0.00002
A_24_P101742	Not mapped by IPA	5	6.52E-08	1.76	0.00002
A_32_P182395	Not mapped by IPA	9	8.12E-08	-0.69	0.00002
A_32_P119165	Not mapped by IPA	9	1.00E-07	0.61	0.00003
A_24_P358606	Not mapped by IPA	5	1.00E-07	1.04	0.00003
A_32_P232682	Not mapped by IPA	7	1.14E-07	-0.80	0.00003
A_32_P231493	Not mapped by IPA	13	2.72E-07	-0.76	0.00006
A_24_P691826	Not mapped by IPA	17	3.06E-07	-0.59	0.00007
A_23_P170719	Not mapped by IPA	19	5.20E-07	0.63	0.00010
A_24_P332292	Not mapped by IPA	15	5.61E-07	-0.75	0.00011
A_24_P900555	Not mapped by IPA	Y	6.52E-07	1.12	0.00012
A_23_P350754	Not mapped by IPA	11	8.19E-07	-1.69	0.00015
A_32_P86616	Not mapped by IPA	X	8.94E-07	2.32	0.00016
A_24_P7330	Not mapped by IPA	22	1.06E-06	1.13	0.00018
A_32_P17484	Not mapped by IPA	20	1.29E-06	-1.03	0.00020
A_24_P367100	Not mapped by IPA	2	1.52E-06	-1.25	0.00023
A_32_P225667	Not mapped by IPA	16	1.66E-06	-0.84	0.00025
A_24_P418028	Not mapped by IPA	14	1.67E-06	-1.12	0.00025
A_24_P488927	Not mapped by IPA	1	2.08E-06	0.50	0.00030
A_24_P5994	Not mapped by IPA	20	2.55E-06	1.13	0.00036
A_32_P186725	Not mapped by IPA	8	3.28E-06	-0.66	0.00044
A_32_P40424	Not mapped by IPA	6	4.95E-06	-0.76	0.00061
A_32_P194563	Not mapped by IPA	Y	5.26E-06	-0.95	0.00063
A_24_P110201	Not mapped by IPA	4	5.63E-06	-0.23	0.00067
A_24_P93425	Not mapped by IPA	1	6.07E-06	-0.55	0.00069
A_32_P18838	Not mapped by IPA	8	6.17E-06	-0.81	0.00070
A_23_P317413	Not mapped by IPA	14	8.17E-06	-0.45	0.00087
A_32_P86533	Not mapped by IPA	2	8.84E-06	1.03	0.00092
A_32_P58705	Not mapped by IPA	7	1.17E-05	0.69	0.00115
A_32_P148616	Not mapped by IPA	4	1.18E-05	-0.55	0.00115
A_24_P170203	Not mapped by IPA	11	1.30E-05	-0.77	0.00125
A_32_P105397	Not mapped by IPA	15	1.39E-05	-0.88	0.00131
A_32_P30760	Not mapped by IPA	5	1.49E-05	-1.52	0.00138
A_32_P201212	Not mapped by IPA	4	1.51E-05	-1.36	0.00138
A_23_P90780	Not mapped by IPA	2	1.62E-05	-1.14	0.00148
A_32_P109181	Not mapped by IPA	21	1.73E-05	-0.71	0.00155

A_24_P222432	Not mapped by IPA	9	1.96E-05	-1.33	0.00171
A_24_P478362	Not mapped by IPA	5	2.08E-05	-1.32	0.00180
A_24_P110242	Not mapped by IPA	16	2.87E-05	3.67	0.00229
A_32_P23525	Not mapped by IPA	12	3.28E-05	1.15	0.00253
A_24_P707530	Not mapped by IPA	17	3.35E-05	-0.74	0.00257
A_32_P142459	Not mapped by IPA	4	3.39E-05	0.76	0.00259
A_23_P28307	Not mapped by IPA	2	3.60E-05	1.11	0.00271
A_32_P214340	Not mapped by IPA	4	4.11E-05	1.07	0.00298
A_32_P13612	Not mapped by IPA	3	4.33E-05	-0.48	0.00308
A_24_P490877	Not mapped by IPA	8	4.45E-05	1.05	0.00314
A_32_P105539	Not mapped by IPA	9	4.66E-05	-0.70	0.00325
A_32_P31963	Not mapped by IPA	5	4.75E-05	0.86	0.00330
A_24_P489639	Not mapped by IPA	11	4.81E-05	0.93	0.00334
A_32_P49284	Not mapped by IPA	8	4.87E-05	-0.83	0.00335
A_32_P144999	Not mapped by IPA	6	4.98E-05	1.20	0.00339
A_32_P56463	Not mapped by IPA	14	5.05E-05	-0.48	0.00343
A_24_P721828	Not mapped by IPA	14	5.05E-05	-0.90	0.00343
A_23_P114582	Not mapped by IPA	1	5.26E-05	1.12	0.00353
A_24_P818529	Not mapped by IPA	Y	5.26E-05	-0.27	0.00353
A_24_P823514	Not mapped by IPA	17	5.35E-05	-0.72	0.00356
A_32_P161554	Not mapped by IPA	X	5.58E-05	-0.83	0.00367
A_24_P170874	Not mapped by IPA	2	5.68E-05	1.19	0.00369
A_24_P533142	Not mapped by IPA	11	5.94E-05	-0.68	0.00383
A_32_P144281	Not mapped by IPA	X	6.41E-05	0.75	0.00407
A_32_P212920	Not mapped by IPA	8	6.85E-05	1.33	0.00429
A_32_P144629	Not mapped by IPA	8	7.04E-05	-0.71	0.00437
A_32_P185881	Not mapped by IPA	2	7.38E-05	-0.75	0.00452
A_32_P8653	Not mapped by IPA	3	8.44E-05	1.37	0.00504
A_24_P307395	Not mapped by IPA	14	8.46E-05	1.61	0.00505
A_24_P281009	Not mapped by IPA	9	9.10E-05	0.94	0.00536
A_32_P40615	Not mapped by IPA	7	9.26E-05	-0.52	0.00542
A_32_P115663	Not mapped by IPA	5	9.70E-05	-0.71	0.00559
A_32_P169353	Not mapped by IPA	2	9.78E-05	2.39	0.00563
A_23_P435390	Not mapped by IPA		9.86E-05	1.39	0.00565

FC, fold change (men versus women); FDR, false discovery rate
*An asterisk indicates that a given gene is represented in the microarray set with multiple identifiers
† Androgen response elements identified by Wyce et al. in the genome of skeletal muscle cells (Wyce A, Bai Y, Nagpal S, Thompson CC. Research Resource: The androgen receptor modulates expression of genes with critical roles in muscle development and function. Mol Endocrinol. 2010;24:1665-74.)